이학석사 학위논문

The origins and genetic diversity of the pinewood nematode (*Bursaphelenchus xylophilus*) in South Korea

한국 소나무재선충의 기원 및 유전적 다양성 분석

February 2023

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Chan Wook Koh

The origins and genetic diversity of the pinewood nematode (*Bursaphelenchus xylophilus*) in South Korea

Advised by Professor Choongwon Jeong

Submitting a master's thesis of
Biological Sciences

February 2023

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Chan Wook Koh

Confirming the master's thesis written by
Chan Wook Koh
February 2023

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

Abstract

Since the pinewood nematode, *Bursaphelenchus xylophilus*, was suspected the causal agent of pine wilt disease to diverse *Pinus* species inhabiting East Asia and Europe, widespread research about this invasive parasite has been conducted. In South Korea, *B. xylophilus* was first reported in 1988, and dispersed throughout the Korean Peninsula for 3 decades, while the number of introduction events and their detailed spreading route has remained unclear. A population study would provide convincing information for tracking the origin and dispersal routes of *B. xylophilus*, but the individual-level whole-genome scale approach so far has been obstructed, due to the lack of genetic material attained from the single nematode. Instead, researchers utilized a small part of the genome or artificially produced inbred lines, both of which provide partial information. In this study, the novel genotyping protocol which utilizes the whole genome amplification was developed and using this approach the genome of 359 *B. xylophilus*, 42 *B. mucronatus*, and 6 *B.doui* individuals were sequenced. High-quality variant panels were produced with about 2 million bi-allelic SNPs throughout the whole genome of *B. xylophilus*, which were used for individual-level genotyping. The genome-scale genetic information at the individual level revealed 5 distinct *B. xylophilus* lineages spread in South Korea, one of which is related to the Japanese and Portuguese isolates. Multiple introduction events of *B. xylophilus* into South Korea were proposed, while the detailed route of dispersal remained unclear due to the unexpectedly similar genetic profile of nematode individuals within each lineage. Also, by comparing *B. xylophilus* and its relative species *B. mucronatus*, low genetic diversity and uniform profile of *B. xylophilus* compared to its allied species were discovered. The low degree of heterozygosity and sustainability of its genetic profile during the propagation process might be associated with the pathogenicity of *B. xylophilus*. This study would provide the foundation to track the worldwide dispersal of *B. xylophilus* and

disclose the key elements of its pathogenicity.

Keywords: pinewood nematode, invasive species, population genetics, whole genome sequencing, *Bursaphelenchus xylophilus*, phylogenetic analysis

Student Number: 2021-23775

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1. Pinewood nematode as a global threat to forest ecosystems

Pinewood nematode, *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934) Nickle, is a causative agent of pine wilt disease in several species of pine trees (Kiyohara and Tokushige, 1971). The origin of *B. xylophilus* is North America including Canada, the USA, and Mexico (Dwinell, 1997), where pine tree species dominant are resistant to pine wilt disease (Dwinell and Nickle, 1989). Due to human-mediated long-distance dispersal, in the 1900s these pathogenic parasites have been spread throughout other continents, including East Asia and Europe (Nunes da Silva et al., 2015). Major *Pinus* species in invaded regions, such as *P. sylvestris*, P. densiflora, *P. thunbergii*, and *P. massoniana* are vulnerable to pine wilt disease (Mamiya, 1989; Bakke et al., 1991). From the first report in Japan (1905), the outbreaks of pine wilt disease have been reported in China (1982), Taiwan (1985), South Korea (1988), and Portugal (1989) (Yi et al., 1989; Mota et al., 1999). Recently, *B. xylophilus* was reported near the border between Portugal and Spain, and pinewood nematodes are threatening European forestry where *Pinus* spp. susceptible to pine wilt disease inhabit (Robertson et al., 2011; Vicente et al., 2013.; Silva et al., 2015). Pinewood nematode is considered a serious invasive pathogen, and studies about its virulency, spreading process, and prevention of its dispersal have been conducted in its introduced and threatened countries.

The complicated life cycle of *B. xylophilus* hinders the prevention of its dispersal and eradication of pine wilt disease. *B. xylophilus* is known to be transported by its vector, *Monochamus* species (Morimoto and Iwasaki, 1972; Naves et al., 2007; LI et al., 2020). Pinewood nematodes are introduced into uninfected pine trees by their carrying beetles when they look for food resources. Inoculated nematodes then breed and proliferate explosively inside the trees,

killing host trees by blocking the transport of water. Then the longhorn beetles lay the eggs on the dead tree. Nematodes gather around the pupa while hatched larvae of longhorn beetles pupate. And imago beetles emerging leave dead trees with pinewood nematodes in their bodies, searching other healthy pine trees for food intake. Once the pine trees are infected, there is no cure to recover. For this reason, disease control of pine wilt disease is mainly concentrated on preventing the dispersal of pinewood nematode.

1.2. Previous study about pinewood nematode

Since *B. xylophilus* came out as the cause of pine wilt disease in various *Pinus* species, factors that affect the pathogenicity of pinewood have been broadly studied (Kikuchi et al., 2011; Kiyohara et al., 1990). It is suggested that there is a correlation between generation time and the pathogenicity of *B. xylophilus* (Filipiak et al., 2021). Altitude, temperature, and precipitation also affect the spread of pine wilt disease (Lee et al., 2017). Especially, the temperature seems to play a key role in pathogenicity (Lee et al., 2017; Li et al, 2022). As parasitic nematodes, a comparison to other non-pathogenic nematodes was also performed (Sultana et al., 2013).

It is also of interest about interactions between pinewood nematodes and their associating species, such as host pine trees, vector beetles, and symbiotic bacteria. The relationship and interactions of *B. xylophilus* with its interactive species have been studied broadly. (Jones et al., 2008). In the host pine tree, pinewood nematodes mount onto *Monochamus* pupa so that they are moved to another pine tree along with eclosed imago when leaving the tree. Researchers discovered that pinewood nematodes in different stages inside the host tree are attracted to move to the pupal chamber of vector beetles according to the composition of volatiles released from larvae (Zhao et al., 2007). Bacteria associated with *B. xylophilus* seems to have a role in the pathogenicity of *B. xylophilus* (Vicente et al., 2012). The interaction of *B. xylophilus* with its associated microbes also have been of interest and studied (Zhao et al., 2014;

Wang et al., 2022). Pine trees respond to the invasion of pinewood nematode and react to protect themselves. Infested host tree releases volatiles to eliminate parasitic nematodes (Zas et al., 2015; Lee et al., 2019). Meanwhile, pinewood nematodes also respond against the defense mechanism of their host pine trees. To survive against the defensive response of the infected host tree, associating genes of *B. xylophilus* including flavonoid biosynthesis, and oxidation-reduction, are upregulated when infiltrated (Shin et al., 2009; Li et al., 2019; Zhang et al., 2020).

Though it has been suspected of pine wilt disease for decades, studying the genetic profile of *B. xylophilus* has been challenged by several limitations (Mallez et al., 2013). Each nematode individual has a small body size and consequently genomic material for individual-scale analysis has been insufficient. For these reasons, a population study of *B. xylophilus* has utilized small portions against the whole genome, such as amplified fragment length polymorphism (AFLP), internal space marker (ITS), effector genes, or microsatellite (Zhou et al., 2007; Cheng et al., 2008; Valadas et al., 2013; Figueiredo et al., 2013; Mallez et al., 2015). Alternatively, another study utilized genomic DNA extraction from hundreds of nematodes (Ding et al., 2021), which is inefficient when searching rare variants, within-population scale or individual-level resolution is required. Otherwise, researchers bypassed the lack of genomic material by producing inbred lines and multiplying nematodes with clonal genetic profiles (Shinya et al., 2012; Palomares-Rius et al., 2015; Ekino et al., 2018). This labor-intensive and time-consuming process leads to the loss of genetic diversity and distorts the original genetic and phenotypic profiles due to repeated inbreeding and laboratory cultivation (Tanaka et al, 2017). Despite some concerns, researchers have thrived to reveal the population structure and long-distance spreading of pinewood nematodes (Cheng et al., 2008; Mallez et al., 2021).

1.3. Pinewood nematode in South Korea

The first report on *B. xylophilus* in South Korea was in Busan in 1989 (Yi et al., 1989). It was suggested that the first introduced pinewood nematode might be transferred from Japan, as Busan is geographically close to Japan, and trade between the two countries was active. However, recent studies reported that both China and Japan have associated with B.xylophilus populations in South Korea (Jung et al., 2010a). Thus, more than once *B. xylophilus* possibly invaded South Korea. The Korean Peninsula is the favored environment for *B. xylophilus*, since *P. densiflora, P. thunbergii*, and other Korean pines were vulnerable to pine wilt disease, and two vector beetles in *Monochamus* spp., *M. saltuarius,* and *M. alternatus* inhabit Korea (Han et al., 2008; Kim et al., 2020). For three decades, pinewood nematode has been spread throughout the Korean Peninsula. In 2021−2022, 135 sites in South Korea had outbreaks of pine wilt disease.

Despite the great interest and many studies, it is still unclear how many times pinewood nematodes had been introduced and whether the dispersal of *B. xylophilus* in South Korea occurred by natural spreading through insect vectors or human−mediated transport (Robinet et al, 2009). Meanwhile, population analysis was conducted using amplified fragment length polymorphisms (AFLPs) patterns to describe genetic relationships between *B. xylophilus* and *B. mucronatus*, which misled suspicious conclusion that *B. xylophilus* and *B. mucronatus* samples are not even distinguished from each other in the phylogenetic tree (Jung et al., 2010a). Another study detected the relationships between *B. xylophilus* in South Korea and neighboring countries (Jung et al., 2010b). They concluded that two *B. xylophilus* populations close to those in China and Japan, respectively, were in South Korea. This study had the limitation of low resolution, owing to relatively less informative microsatellite markers compared to whole genome data.

1.4. Pinewood nematode and its relative species

Among about 200 species in *Bursaphelenchus* spp., *B.*

mucronatus is one of the closest species to *B. xylophilus* (Kanazaki and Giblin-Davis, 2018). These nematodes share the food resources, host tree species, and vectors with *B. xylophilus*. Morphologically similar to *B. xylophilus*, a hybrid of *B. xylophilus* and *B. mucronatus* can be produced, and even it is found in the field (Y. Li et al., 2021). Despite the resemblance between the two species, there are distinct differences in molecular level between them (Zheng et al., 2003; Matsunaga et al., 2019). Especially, *B. mucronatus* don't have pathogenicity to their host trees (Akbulut et al., 2007; Vincent et al., 2008). For these reasons, *B. mucronatus* have been compared to *B. xylophilus* for searching the cause of pathogenicity (Lee et al., 2019), and the genetic relationship between them also has been studied (Sultana et al., 2013; Pereira et al., 2013; Zhou et al., 2017). In South Korea, 2 types of *B. mucronatus* were reported: the East Asian type and the European type (Han et al., 2008).

1.5. Purpose of this study

Pinewood nematode is an optimal case for studying the spread of invasive species and their effect on the introduced environment (Estoup and Guillemaud, 2010). The origin of *B. xylophilus* is established well and its dispersal route is updated in real-time. For example, *B. xylophilus* population reported from Portugal originated from Japan, not directly from the USA, and it is spreading toward Spain and European inland (Burgermeister et al., 1999; Abelleira et al., 2011; Fonseca et al., 2012). As the occurrence of pine wilt disease is visually detected and isolates are readily sampled from the infected trees, the change in the genetic profile of the introduced population during the dispersal progress also can be analyzed and the impact of invasive species on the introduced environment also can be evaluated.

Considering its devastating impact on the newly introduced ecosystem, it is important to study the characteristic of *B. xylophilus* and track its spreading routes for the efficient prevention of pine wilt

disease. Tracking the origin and dispersal routes of pinewood nematodes has been an intriguing and important topic (Estoup and Guillemaud, 2010; Mallez et al., 2021). The genome-wide approach could be an appropriate tool to provide more precise and informative research to study the origin and the population structure of *B. xylophilus* than utilizing a relatively small subset of the genome (Zhou et al., 2007; Cheng et al., 2008; Valadas et al., 2013; Figueiredo et al., 2013; Mallez et al., 2015). Unfortunately, the small body size of a single nematode disturbed it to attain enough amount of genomic material for whole-genome scale research, which made individual-level analysis of pinewood nematode impossible. Researchers bypassed this issue by producing genetically clonal inbred lines consuming substantial labor and time, remaining the issue of losing its original genetic property.

In this study, through a newly developed experiment procedure, which is called single nematode sequencing protocol, the lack of genomic DNA was overcome and individual-level whole-genome sequencing was conducted. Using this technique, it is expected that the population structure and concealed dispersal routes of *B. xylophilus* in South Korea will be described in detail. Also, relationships between Korean pinewood nematodes and those overseas would be revealed. Finally, genetic differences between *B. xylophilus* and its relative species would be useful hints to discover the source of the pathogenicity of *B. xylophilus*.

2. Materials and Methods

2.1. Isolate collection and individual selection

LineF and lineJ inbred lines were provided (Woo, 2022), and two types of F_1 hybrid between two inbred lines: mJ−fF (offspring of lineJ male and lineF female), and fJ−mF (offspring of lineJ female lineF male) were produced. Both F_1 progenies were back−crossed with lineJ individuals. Korean *B. xylophilus* and relative species isolates were collected from 173 infested trees sampled from all over the Korean Peninsula in 2020 and 2021 and provided by the National Institute of Forest Science (Appendix 1). GP−1~4 were collected from the body of *M. saltuarius*, captured from a mountain in Gapyeong−si. Nematodes from each isolate are cultured on the mycelia of *Botrytis cinerea* grown on potato dextrose agar. The subculture of each isolate was conducted once or twice. Reared nematodes were harvested for further experiments. By pouring 5 mL of deionized water, cultured mixed−stage nematodes were floated on the surface of the agar and then moved onto the new microtubes. Collected nematode solutions were inverted, and kept for 3 minutes. After nematodes were precipitated, supernatants were removed and tubes were filled with deionized water to repeat the washing step 3 times. 0.5% sodium hypochlorite solution was used instead of deionized water, for the additional washing to remove bacteria on the surface of the nematode's body. Washed nematode solutions were incubated by shaking overnight.

For each isolate, 20 μL of nematode solution was applied to the new potato dextrose agar. To avoid uncontrolled input sources of genetic material from eggs conceived, 1 to 3 adult male individuals were selected from each isolate, and each nematode was moved onto a 0.2 mL PCR tube with 15 μL of deionized water. Genomic DNA was extracted from the nematode by repeated flash−freezing on liquid nitrogen and thawing the body in a 37℃ water bath. gDNA amplification was performed using a Repli−g single cell kit (Cat. No.

150343, Qiagen, USA). Each extracted gDNA sample was applied 1 μL of concentrated D2 buffer (DLB buffer: 1M dithiothreitol = 3:1) and incubated for 3 minutes at room temperature. 3 μL of stop solution was added, and each sample was stored on a chilling plate. Repli-g master mix was produced, with 29 μL of reaction buffer, 0.25 μL of 1M dithiothreitols, and 2 μL of repli-g DNA polymerase for each sample. Mixed samples were incubated at 30℃ for 2 hours. DNA concentration was measured using a Qubit 4 fluorometer and Qubit™ 1X dsDNA HS Assay Kits (Cat. No. Q33231, ThermoFisher Scientific, USA).

2.2. Whole genome sequencing

DNA libraries were constructed using amplified genomic DNA of each pinewood nematode individual and QIAseq FX DNA Library Kit (Cat. No. 180479, Qiagen, USA). 1 μL of amplified genomic DNA sample was used for each nematode individual and diluted to 1/10. 40 μL of fragmentation mix was added and samples were incubated at 32℃ for 10 minutes. 1 μL of Unique-dual Adapter was added for each sample with 49 μL of adapter ligation mix. DNA fragment size selection and impurity cleanup were conducted using AMPure XP Beads (Cat. No. A63881, Beckman Coulter, USA). Selected gDNA fragments were amplified. DNA purification was conducted using MinElute PCR purification kit (Cat. No. 28006, Qiagen, USA). Additional size selection was conducted, to select 300-1000bp DNA fragments. Constructed DNA libraries were sequenced on Illumina HiSeq X / NovaSeq 6000 platform using 151 bp paired-end reads.

For the relative species *B. mucronatus* and *B. doui*, we sequenced 42 individuals of *B. mucronatus* lineages (Asian and European type) and 6 individuals of *B. doui*. Isolate cultivation, washing, individual selection, genomic DNA amplification, and library construction followed the same protocol with *B. xylophilus* samples. DNA libraries were sequenced on Illumina HiSeq X / NovaSeq 6000 platform using 151 bp paired-end reads.

2.3. Sequenced reads mapping and quality filtering

Sequenced raw reads of Korean *B. xylophilus* samples and sequence reads of 8 inbred lines were aligned to the *B. xylophilus* reference genome (Ka4C1) (Dayi et al., 2020), and relative species individuals were aligned to the *B. mucronatus* reference genome (ZJ-2014) (Wu et al., 2020). Inbred line strains data were downloaded from the National Center for Biotechnology Information Sequence Read Archive (NCBI-SRA) under the accession number PRJDB3459 (Cotton et al., 2016). BWA-mem v0.7.17-r1198-dirty (Li and Durbin, 2010) was used. Using samtools v1.9 (H. Li et al., 2009), read pairs properly mapped were selected by applying "samtools view -bh -f 0x0003 -F 0x0004" option. PCR duplicates were removed using Picard MarkDuplicate v.2.27.1. Finally, reads with Phred-scale mapping quality score higher than 30 were kept to generate BAM files using samtools v1.9 with "samtools view -bh -q30" option. To check whether there is bacterial DNA or *B. cinerea* genome proportion in sequenced reads, Kraken2 v2.1.2 (Wood et al. 2009) was used.

2.4. Variant calling and filtering

From the generated individual BAM files, we produced gVCF files using Genome Analysis ToolKit (GATK) v3.8.1.0 HaplotypeCaller module (Auwera, et al. 2013) with "--ERC GVCF –variant_index_type LINEAR –variant_index_parameter 128000 –rf BadCigar –mbq 30" option. We produced 3 variant panels; 367 individuals of the whole *B. xylophilus* samples, 359 Korean *B. xylophilus* samples, and 48 relative species samples (42 *B. mucronatus* and 6 *B. doui* individuals). Using GATK-CombineGVCFs and GATK-GenotypeGVCFs modules, gVCF files were merged and a VCF file containing high-quality variant calls was produced. Variant filtering was performed to remove low-quality variants, using GATK-SelectVariants and VariantsFitleration module with "-filter "QUAL < 30.0 || DP < 1680.0 || DP > 42000.0 || QD < 2.0 || SOR

> 3.0 || FS > 60.0 || MQ <= 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0" option. DP options were adjusted considering the mean sequencing depth of each variant panel, "DP < 1500.0 || DP > 37500.0" and "DP < 200.0 || DP > 5000.0" for the Korean sample panel and relative species panel, respectively. Filtered variants were sorted into bi−allelic SNP, bi−allelic InDel, and multi−allelic SNPs/InDels.

2.5. Genotype refinement

Bi−allelic SNP panels were chosen for the following refinement and analysis. For 2 *B. xylophilus* bi−allelic SNP panels, genotype imputation was performed using BEAGLE v4.1 (Browning and Browning, 2016) with "lowmem=true window=150000 overlap=10000 niterations=10" option. The individual genotypes were called in EIGENSTRAT format, with genotype probability >0.99 as the missingness cutoff. EIGENSTRAT format genotype files were converted to plink format files using convertf v4510 (Patterson et al., 2006) and PLINK v1.90b6.26 (Chang, et al. 2015). Monomorphic SNPs and SNPs with missingness > 0.05 were filtered out ("−−geno 0.05 –maf 0.001"). For relative species, genotype refinement using Beagle was not conducted. Instead, the genotype was converted into EIGENSTRAT format with missingness cutoff GL>0.99. Among all SNPs, loci with a missingness count > 12 were excluded for further analysis.

2.6. Principal component analysis

Principal component analysis (PCA) was performed with smartpca v16000 (Patterson et al., 2006). For *B. xylophilus* diversity panels, "lsqproject: YES" and "numchrom: 6" options were used, while for the relative species panel "numchrom: 6" option was not applied. 8 inbred line individuals were projected on the PCA plots.

2.7. Genetic distance and F_{st} calculation

The genetic distance between 14 *B. xylophilus* populations was estimated using F_2 distance (Patterson et al., 2012). The square of the allele frequency difference between two populations for each locus was calculated and averaged. Distance heatmap was drawn with a $(1-F_2)$ matrix. F_{st} calculation was followed by Wright's F_{st} (Wright, 1950; Chen et al., 2015). The allele frequency of the total population ($\bar{p}$) was calculated as the mean of allele frequencies for each population (p). F_{st} was calculated as follows:

$$\frac{\sum_{i=1}^{r}(p-\bar{p})^2}{r\bar{p}(1-\bar{p})}$$

where r is the number of lineages.

2.8. Phylogenetic tree analysis

A phylogenetic tree was constructed for 14 *B. xylophilus* lineages (excluding progenies of lineJ and lineF inbred lines) using OrientAGraph v1.0 (Molloy et al., 2021). C14 was set to root, adding up to 4 migration edges with "−m 0−4". The number of SNPs per block was set to 20000 with "−k 20000", and a global rearrangement option was applied with "−global". Allele frequencies were calculated for each lineage from plink output files.

2.9. Subpopulation detection

Relatedness among *B. xylophilus* individuals in the Korean lineage was estimated using principal component analysis (PCA), pairwise mismatch rate (PMR), rare variant sharing rate, and genotype similarity. "lsqproject: YES" and "numchrom: 6" options were used to perform smartpca. The numerator of the PMR value for each individual pair was calculated as follows:

$$\sum_{k=1}^{n}\left(\frac{g_{ik}}{2}\times\frac{(1-g_{jk})}{2}+\frac{(1-g_{ik})}{2}\times\frac{g_{jk}}{2}\right)$$

where n is the total number of SNPs used, and g_i, g_j is the genotype of individual i and j, coded with the copy number of derived alleles in kth SNP. The denominator was the total number of SNPs. For rare variant sharing rate, the numerator was calculated as follows:

$$\sum_{k=1}^{n} \sqrt{\frac{g_{ik} \times g_{jk}}{2}}$$

where g_i, g_j is the genotype of individual i and j, coded with the copy number of minor alleles in kth SNP. The denominator was the total number of SNPs used. For pairwise genotype similarity, the numerator was calculated as follows:

$$\sum_{k=1}^{n} \left(\sqrt{(2 - g_{ik}) \times (2 - g_{jk})} \times \log p_{kanc} + \sqrt{(g_{ik} \times g_{jk})} \times \log p_{kder} \right)$$

where p_{kanc} is allele frequency of ancestral allele, and p_{kder} is allele frequency of the derived allele (Guan et al., 2009).

3. Results

3.1. Individual-level whole genome sequencing of pinewood nematode

The whole genomes of 359 *B. xylophilus* male individuals were sequenced, consisting of 248 samples from 173 sites in South Korea, and 111 individuals of 2 inbred lines with their F_1, and N_2 progenies. The single nematode sequencing protocol was utilized (methods). Also, 42 individuals of *B. mucronatus* (18 and 24 individuals for Asian and European type, respectively) and 6 *B. doui* individuals were sequenced following the same protocol. Sequence reads of *B. xylophilus* samples with published data of 8 inbred lines from Japan and Portugal, were mapped to the Japanese Ka4C1 reference genome (Dayi et al., 2020). *B. mucronatus* and *B. doui* sequence reads were mapped to the Chinese *B. mucronatus* reference genome (Wu et al., 2020).

The average mapping rate of *B. xylophilus* sequence reads on the Japanese Ka4C1 reference genome was 0.69 (0.06-0.98), and the sequenced coverage of each sample ranges from 1.7× to 140.6× (35.0× on average) (Appendix 2). Individual-level sequencing was done well and even throughout the chromosomes (Figure 1). Sequenced reads from each individual contained bacterial DNA and gDNA from the *B. cinerea* genome (from 1.2 % to 87.3 %) due to insufficient washing, which leads to loss of nematode genome coverage. Nevertheless, it was possible to attain enough coverage for the following analysis. *B. mucronatus* European and *B. doui* individuals, due to considerable differences in their genomes to *B. mucronatus* Asian type reference genome, showed quite lower mapping rates than those of *B. mucronatus* Asian individuals.

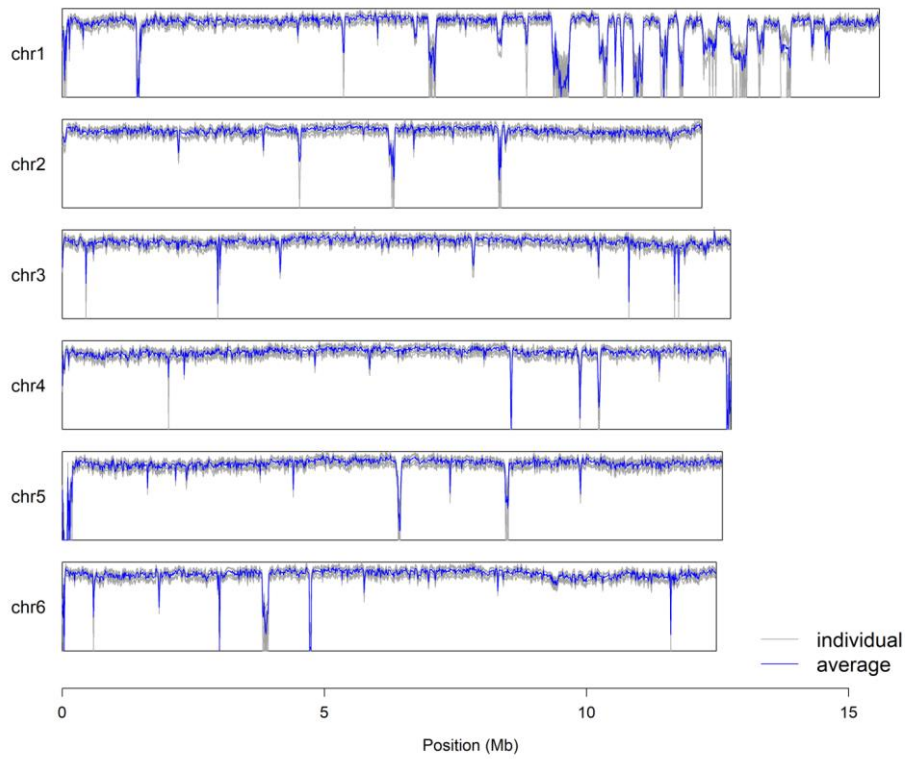


Figure 1. Mapped read depth plot of 7 *B. xylophilus* individuals. Samples with >100× coverage were selected and plotted. Log2 normalized.

3.2. Genome-wide variant identification and individual-level genotyping

3 types of diversity variant panels were produced; 367 individuals of all *B. xylophilus* samples, 359 Korean *B. xylophilus* samples, and 48 *B. mucronatus* and *B. doui* species samples (Table 1). Each variant panel contained 2,730,335, 1,189,294, and 5,124,329 high-quality variants, respectively. Filtered variants were sorted into 3 categories; bi-allelic SNPs, bi-allelic InDels, and multi-allelic variants. After genotype refinement and missingness filtering, a whole *B. xylophilus* variant panel with 2,078,778 bi-allelic SNPs was produced and used for the following analysis. Korean *B. xylophilus* variant panel, with 341,388 bi-allelic SNPs, lessened its SNP counts to almost 28% after the exclusion of monomorphic SNPs. The relative species panel, with 2 types of *B. mucronatus* and *B. doui* individuals, had more variants than the *B. xylophilus* panel because of the difference between the genomes of the two species, but most of the variants didn't make the missingness cutoff.

Table 1. The Number of produced variants for each diversity panel.

Variant panel	Whole *B. xylophilus*	Korean *B. xylophilus*	*B. mucronatus* and *B. doui*
# of samples	367	359	48
# of bi-allelic SNPs	2,158,749	947,855	3,892,368
# of bi-allelic InDels	395,642	191,350	649,794
# of multi-allelic variants	175,944	50,089	582,167
# of bi-allelic SNPs (finalized for analysis)	2,078,778	341,388	239,809

3.3. Genetic profile of *B. xylophilus* in South Korea and multiple introductions into the Korean Peninsula

To investigate the distribution of variants throughout the whole chromosomes of pinewood nematodes, the density of biallelic SNPs is counted with a non-overlapping 10kb window. Most bi-allelic SNPs were located in the arm of each chromosome, while the center of chromosomes showed a sparse distribution of SNPs (Figure 2). Since most genes of nematode are located at the center of the chromosome (*C. elegans* Sequencing Consortium, 1998; Rödelsperger et al., 2017), it is reasonable that neutral mutations on the chromosome arm remained while mutations on the center of the chromosome were selected out by purifying selection (Rödelsperger et al., 2014). Besides the overseas samples, variants that are polymorphic among Korean pinewood nematodes were distributed in restricted regions. 64 % of total biallelic SNPs are fixed in Korean samples, leading the SNP density to be much lower than with overseas samples, which implies the lower genetic diversity of the Korean lineages.

To examine the population structure of *B. xylophilus* in South Korea, we performed principal component analysis (PCA). Among 367 *B. xylophilus* individuals, All Korean samples are clustered on the same point of PCA (Figure 2). 4 inbred lines (C14, OKD1, Ka4C1, and T4) are separated from the rest of the inbred lines and all Korean samples. 54.2% of total variants used for PCA is explained by PC1, showing high divergence between Japanese inbred lines and Korean lineages. Excluding the most divergent 4 individuals, we classified the Korean samples into 5 distinct lineages, including 2 Korean inbred lines (Figure 3). Korean Lineage 01 (KL01) has 183 individuals sampled from 139 sites in South Korea. KL01 individuals are distributed in the entire area of South Korea (Figure 4.). LineF, sampled from Pohang, is pointed at the same point on the PCA plot with KL01 individuals. KL02 comprised 52 individuals from 32 sites and they were collected from the southern region of the Korean peninsula, including Busan, Jeolla-do, and Gyeongsang-do. From

only 1 site KL03 is found, and 9 individuals were sequenced. 4 individuals of KL04, a putative hybrid of KL01 and KL02, were found in 1 site of Busan, in which pinewood nematode was first reported (Yi et al., 1989). Finally, lineJ, sampled from Busan, was not clustered with the other Korean samples. Japanese inbred lines S10, and PT670 from the mainland of Portugal are clustered with KL02, while no other overseas samples were grouped with the other Korean lineages.

The relationships between 10 inbred lines and 4 Korean lineages were analyzed. F_{st} showed high genetic differences between 14 lineages (Table 2). F2 distance corroborated with the PCA result (Figure 5). C14 and OKD1 were opted out from the other lineages, then Ka4C1 and T4, serially. KL02 is clustered with S10 and MAD24 but without other Korean lineages.

To describe the relationships between the Korean lineages and overseas isolates, phylogenetic analysis was conducted using OrientAGraph v1.0 (Molloy et al., 2021). 10 inbred lines and 4 Korean lineages were included. 1,996,630 SNPs were used, and 0 to 4 migration edge was added to explain relationships that are not explained enough only with tree formation. For all trees, 4 Japanese inbred lines C14, OKD1, Ka4C1, and T4 formed the outgroup for the Korean lineages (Figure 6. The first two lineages, C14 and OKD1, are known as less virulent isolates than the rest (Filipiak, 2015). Also, it is reported that these two lineages don't share an endogenous viral element with the other inner clades, putatively derived from Nodavirus (Cotton et al., 2016).

The 4 Korean lineages are not monophyletic to overseas samples. Aside from KL01 and KL03, KL02 formed a sister group with KL04, Portuguese PT670, MAD24, and Japanese S10. The first migration edge indicates gene flow from KL01 to KL04, which implies KL04 is the hybrid of KL01 and KL02. The second and third migration edge explains the intermediate characteristics of PT670 and lineJ. PCA result showed that they lie in the midpoint of KL01 and KL02, while PT670 is slightly close to KL02 and lineJ to KL01. OrientAGraph result supports this, by adding gene flow edges from the other clades

to them. The last migration edge is placed from OKD1 to KL01, KL03, and lineF.

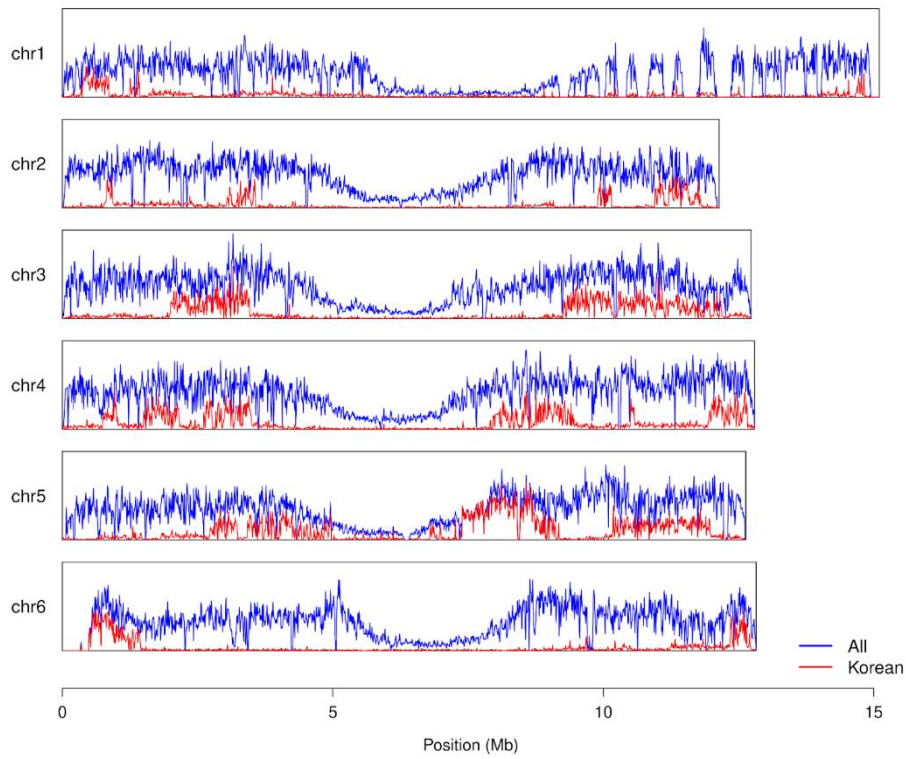


Figure 2. Distribution of bi-allelic SNPs throughout 6 chromosomes. The y-axis maximum value is 0.08, which means 800 SNPs in a 10kb window.

(A)

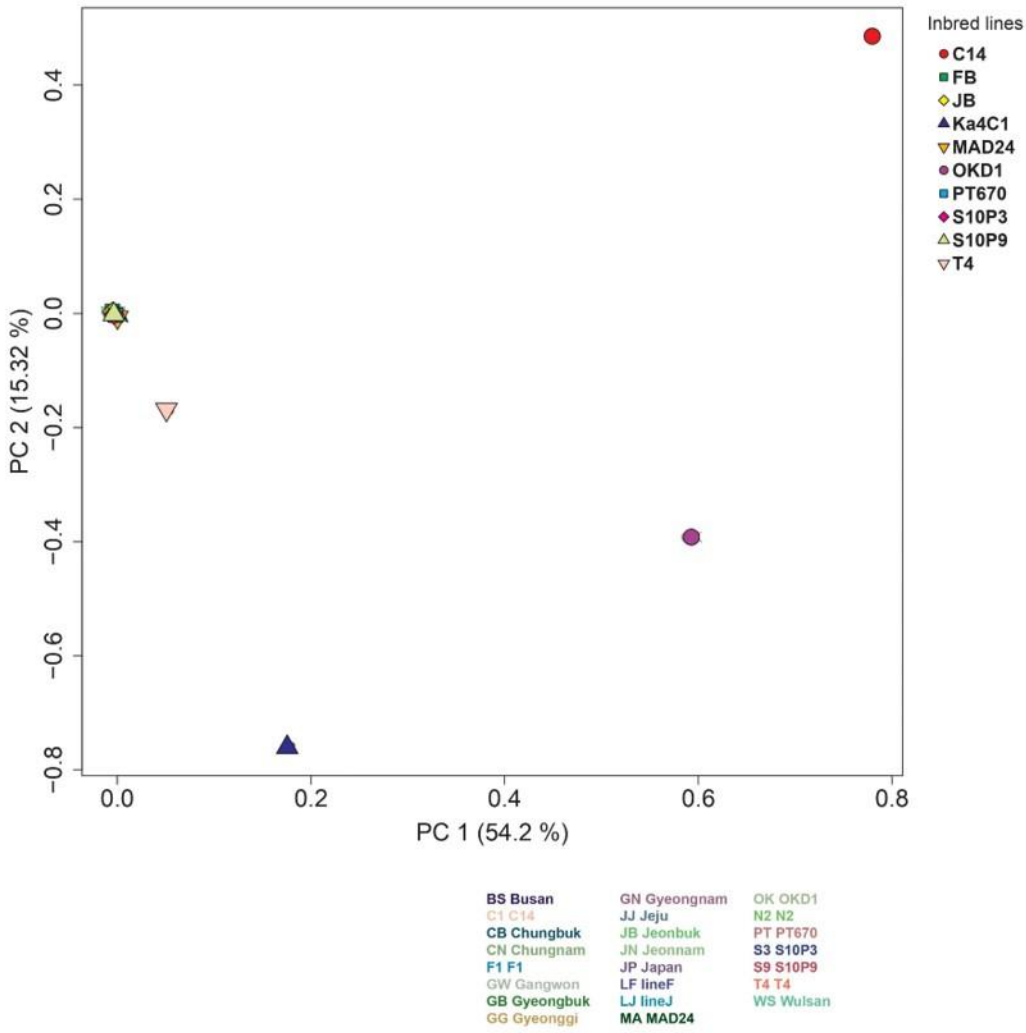


Figure 3. PCA plot of *B. xylophilus* samples. (A) whole 367 individuals (B) 363 individuals except 4 inbred lines (C14, OKD1, Ka4C1, and T4). These 4 samples were not included in the PCA calculation but instead projected on the PCA plot. The numbers in the bracket indicate the variation explained by each axis.

(B)

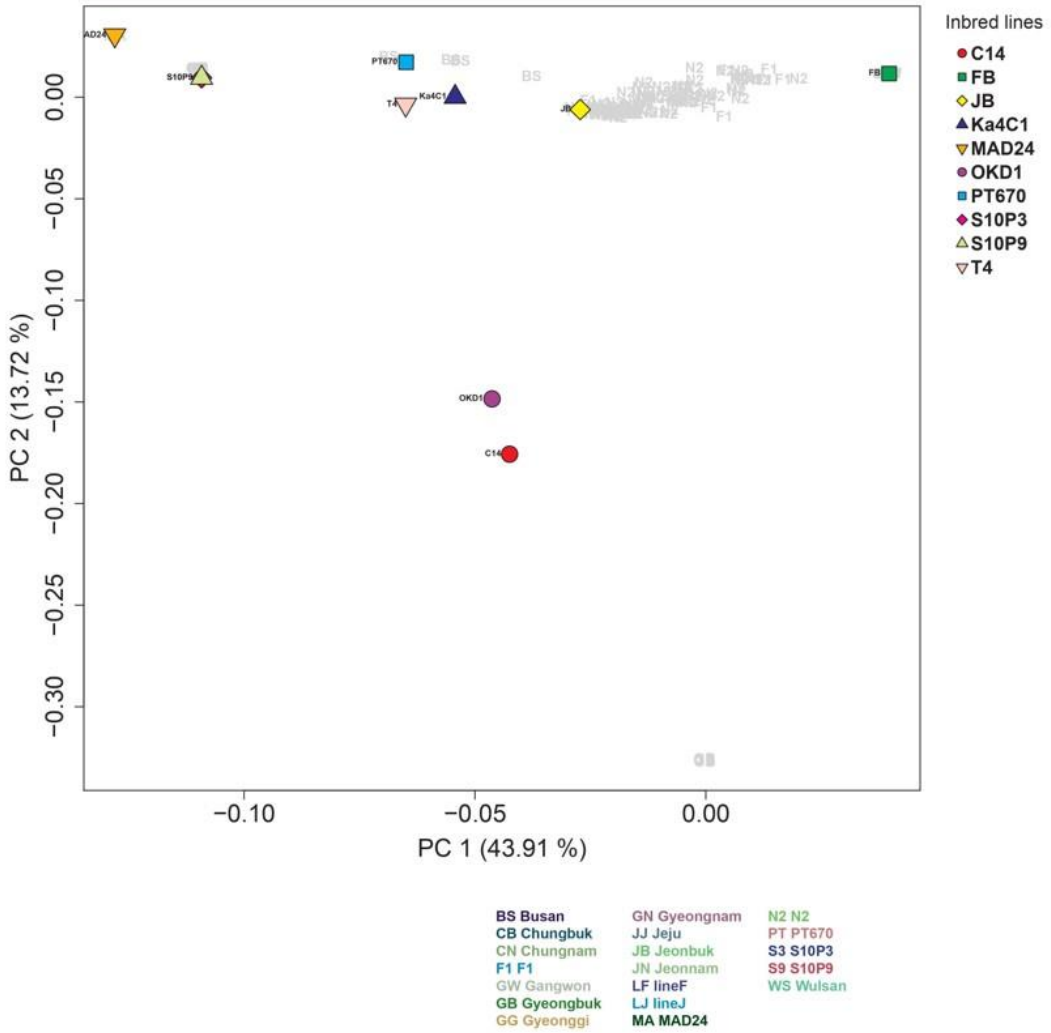


Figure 3. (continued)

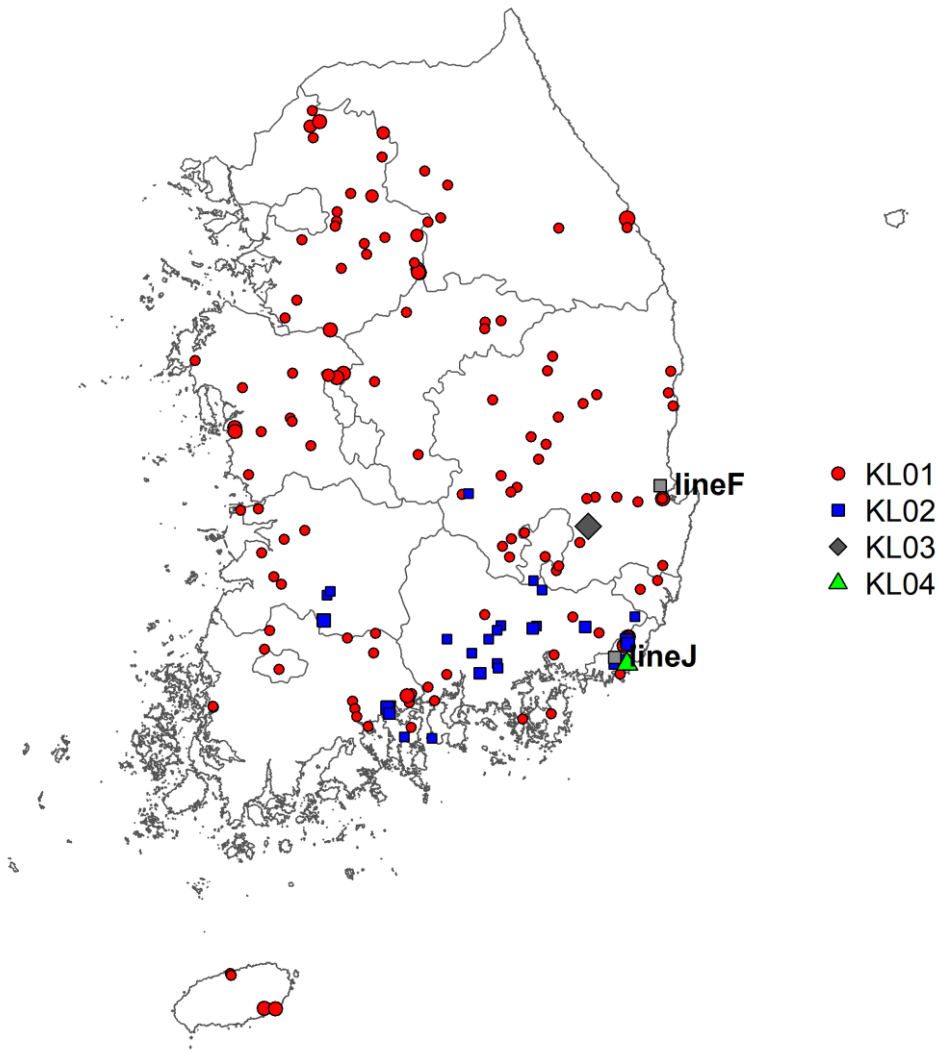


Figure 4. Sampling sites of *B. xylophilus* in South Korea. The figure size indicates the number of sequenced individuals from each site.

Table 2. F_{st} of each population set.

Set	F_{st}
All 14 lineages	0.9289
Except for 2 lineages (C14, OKD1)	0.9217
Except for the next 2 lineages (Ka4C1, T4)	0.6216
10 Korean lineages	0.5965

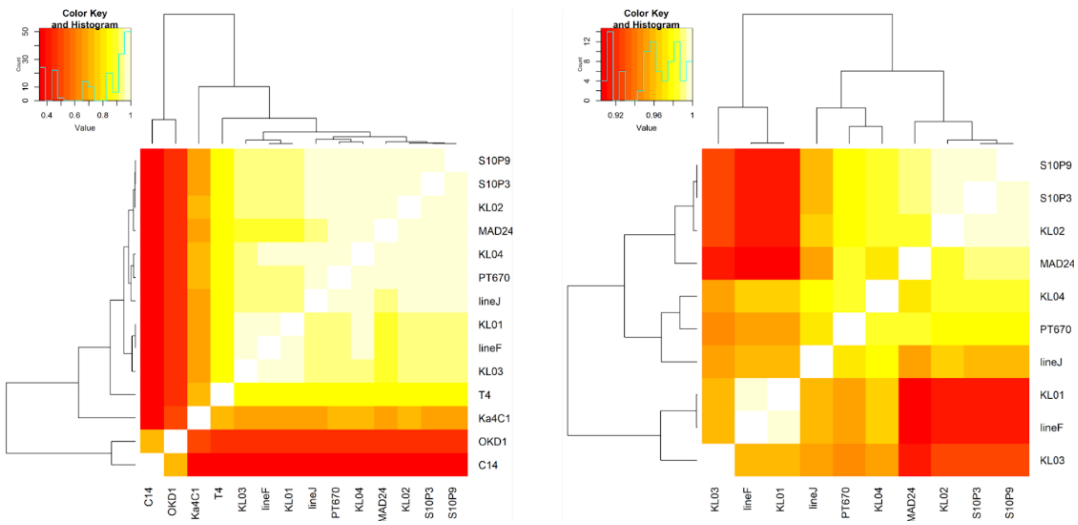
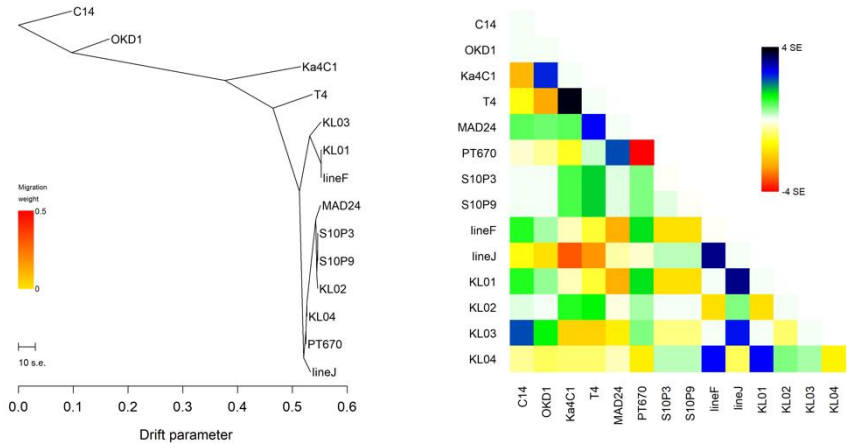


Figure 5. Heatmap plot of F_2 distance between lineages. (Left) All 14 lineages (right) and 10 lineages except the most divergent 4 lineages (C14, OKD1, T4, and Ka4C1).

(A)



(B)

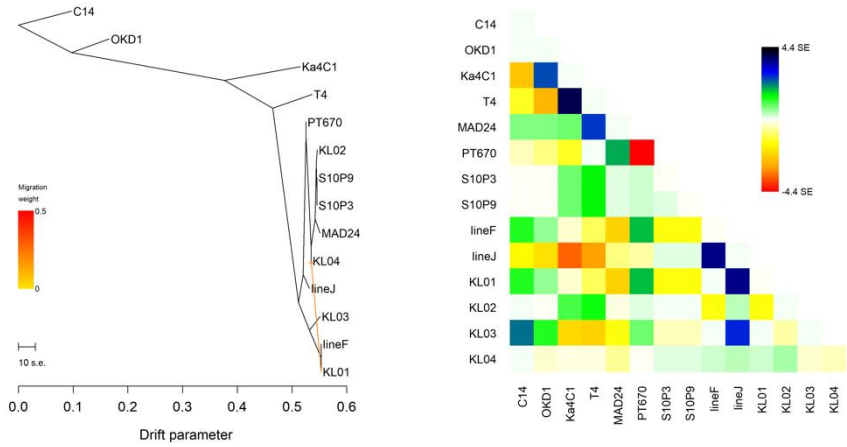
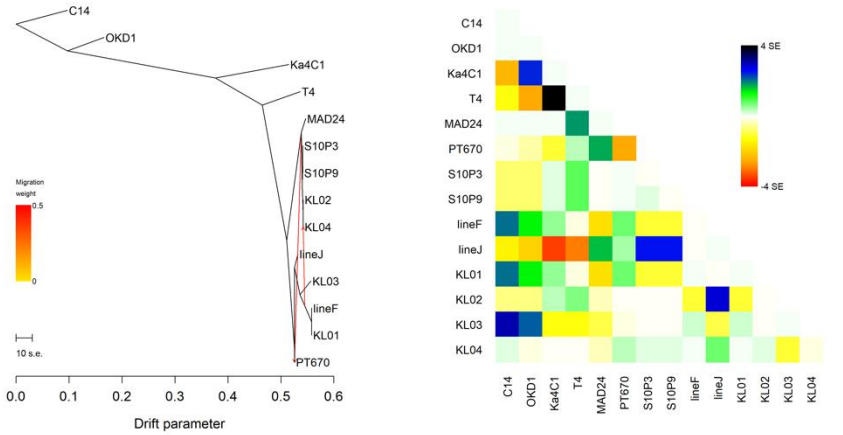
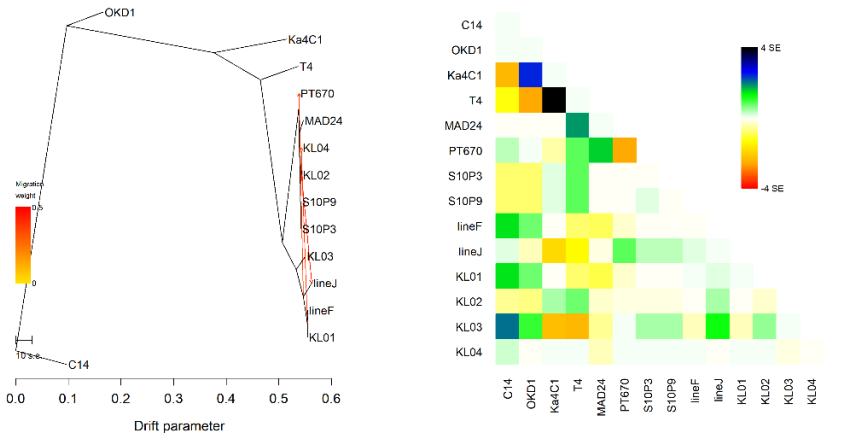


Figure 6. Phylogenetic tree constructed with 14 *B. xylophilus* lineages. C14 was set to root. 0 to 4 gene flow edges were added, respectively.

(C)



(D)



(E)

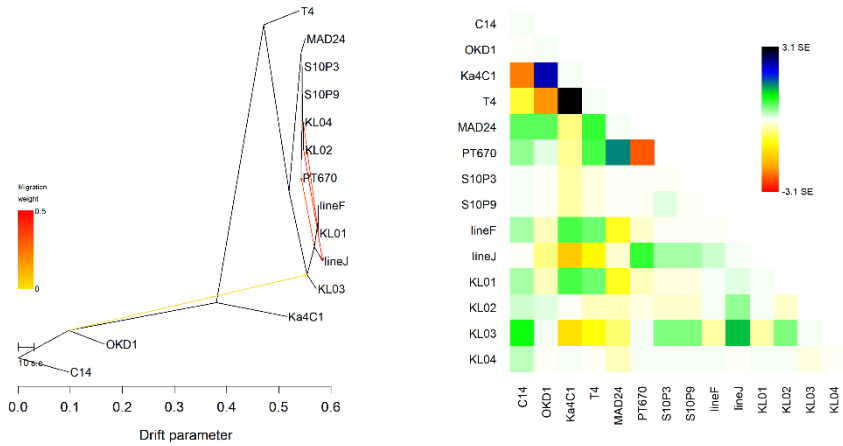


Figure 6. (continued)

3.4. Genetic diversity of Korean *B. xylophilus* lineages

To inspect the population structure of Korean *B. xylophilus* lineages, KL01, and KL02 samples were analyzed. For precise analysis,166 and 47 individuals with coverage > 10x were selected from KL01 and KL02, respectively. 1,838,858 SNPs were selected and re-labeled to ancestral and derived alleles using genotypes of 4 outgroup individuals. In each lineage, most variants were fixed (Table 3). The number of fixed alleles in each lineage was 1,773,835 (96.5%) and 1,812,165 (98.5%), respectively. 157,227 (8.6%) SNPs were fixed to different alleles between two lineages, and 66,408 (3.6%) SNPs were fixed in one lineage and polymorphic in the other lineage.

Using lineage-specifically polymorphic SNPs PCA was performed. For each lineage, some samples acted as outliers; GW12 in KL01, and JN12, JB3-1 in KL02 (Figure 7C,7D). For precise analysis, outlier individuals were excluded and PCA was performed again. In KL01, despite serial exclusions of outliers, no geographic subgroup was detected and individuals still clustered together (Figure 7). In KL02, Gyeongnam samples were slightly separated from the rest, with some overlapping individuals. The explained variance of each axis was lower in KL01 than in KL02.

Using the same dataset, the pairwise mismatch rate (PMR) between within-lineage individuals was calculated and visualized as a heatmap (Figure 8). KL01 individuals showed much more uniform genotypes compared to KL02, with average PMR values of 0.036 and 0.098 for KL01 and KL02, respectively. Individuals showing higher PMR values acted as outliers in PCA. In KL01, no geographic subpopulation was detected, although some pairs showed much closer similarities than other pairs. On the other hand, individuals in KL02 were divided into two subgroups; samples from Gyeongsang-namdo (Gyeongnam) and the others, although PMR values were much higher than those of KL01.

To detect subgroups in each lineage, a rare variant sharing rate was calculated. Among lineage-specific polymorphic SNPs, SNPs with minor allele frequency < 0.1 were selected. Visualized as

a heatmap, individuals in KL01 showed closer similarity with those from the same region (Figure 9). In the KL01 heatmap, about 45% of *B. xylophilus* individuals from 11 regions of South Korea, except Jeju Island, showed much closer similarities than the others. In KL02, unlike PMR and PCA results, the rare variant sharing rate showed the separation of Gyeongnam samples into more than 2 subgroups. This may be interpreted that Gyeongnam individuals shared common alleles rather than rare variants.

Lastly, genotype similarity was calculated, considering the rarity of each allele (Figure 10). The result indicated that outliers from each lineage, especially GW12 in KL01 and JB3, and JN12 in KL02, shared rare variants with themselves, which might let them pop out from the major cluster.

Table 3. The distribution of derived allele frequencies for KL01 and KL02. The square bracket means over/under #, and the round bracket means more / less than #.

KL01 / KL02	[0.00]	(0.00, 0.05]	(0.05, 0.10]	(0.10, 0.15]	(0.15, 0.20]	(0.20, 0.25]	(0.25, 0.30]	(0.30, 0.35]	(0.35, 0.40]	(0.40, 0.45]	(0.45, 0.50]	(0.50, 0.55]	(0.55, 0.60]	(0.60, 0.65]	(0.65, 0.70]	(0.70, 0.75]	(0.75, 0.80]	(0.80, 0.85]	(0.85, 0.90]	(0.80, 0.95]	(1.00, 1.00)	[1.00]
[0.00]	439293	6275	456	211	136	80	53	48	54	98	385	50	58	43	31	53	30	54	60	93	1647	79127
(0.00, 0.05]	28874	5891	272	8	6	6	0	0	0	7	6	4	3	3	4	3	3	0	1	6	23	2733
(0.05, 0.10]	247	122	136	28	15	1	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	18
(0.10, 0.15]	50	0	42	47	21	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	15
(0.15, 0.20]	20	1	11	26	23	19	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
(0.20, 0.25]	10	0	0	6	13	35	30	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(0.25, 0.30]	3	1	0	0	4	8	38	12	4	0	0	0	0	0	0	0	0	0	0	0	0	3
(0.30, 0.35]	3	4	0	0	0	2	6	44	20	3	2	0	0	0	0	0	0	0	0	0	0	2
(0.35, 0.40]	7	0	0	0	0	0	5	26	62	24	2	0	0	0	0	0	0	0	0	0	0	7
(0.40, 0.45]	23	1	0	0	3	0	0	3	17	67	34	0	0	0	0	0	0	0	0	0	0	4
(0.45, 0.50]	291	6	7	0	0	1	4	4	0	22	849	26	6	2	0	0	0	2	1	4	15	48
(0.50, 0.55]	61	0	0	0	0	0	0	0	0	0	245	147	22	0	3	0	0	0	0	0	3	45
(0.55, 0.60]	10	0	0	0	0	0	0	0	0	0	1	9	23	14	0	0	0	0	0	0	1	4
(0.60, 0.65]	8	0	0	0	0	0	0	0	0	0	0	1	17	39	14	0	0	0	0	0	0	3
(0.65, 0.70]	16	0	0	0	0	0	0	0	0	0	0	0	3	25	28	5	0	0	0	0	0	2
(0.70, 0.75]	17	2	0	0	0	0	0	0	0	0	0	1	3	5	16	16	9	1	1	2	0	1
(0.75, 0.80]	9	0	0	0	0	0	0	0	0	0	0	0	0	0	2	13	14	5	4	0	0	1
(0.80, 0.85]	15	0	0	0	0	0	0	0	0	0	7	0	0	0	0	9	16	27	9	1	2	2
(0.85, 0.90]	29	1	0	0	0	0	0	0	0	0	4	2	0	0	0	0	4	14	40	9	1	6
(0.90, 0.95]	43	0	0	0	0	0	0	0	0	0	6	0	0	6	0	0	0	2	34	71	60	33
(0.95, 1.00)	4844	16	2	0	0	0	0	2	0	4	17	0	2	0	4	1	1	1	9	139	3254	14839
[1.00]	77983	732	45	4	27	6	5	10	15	25	105	23	20	12	9	8	14	25	23	76	2937	1163159

(A)

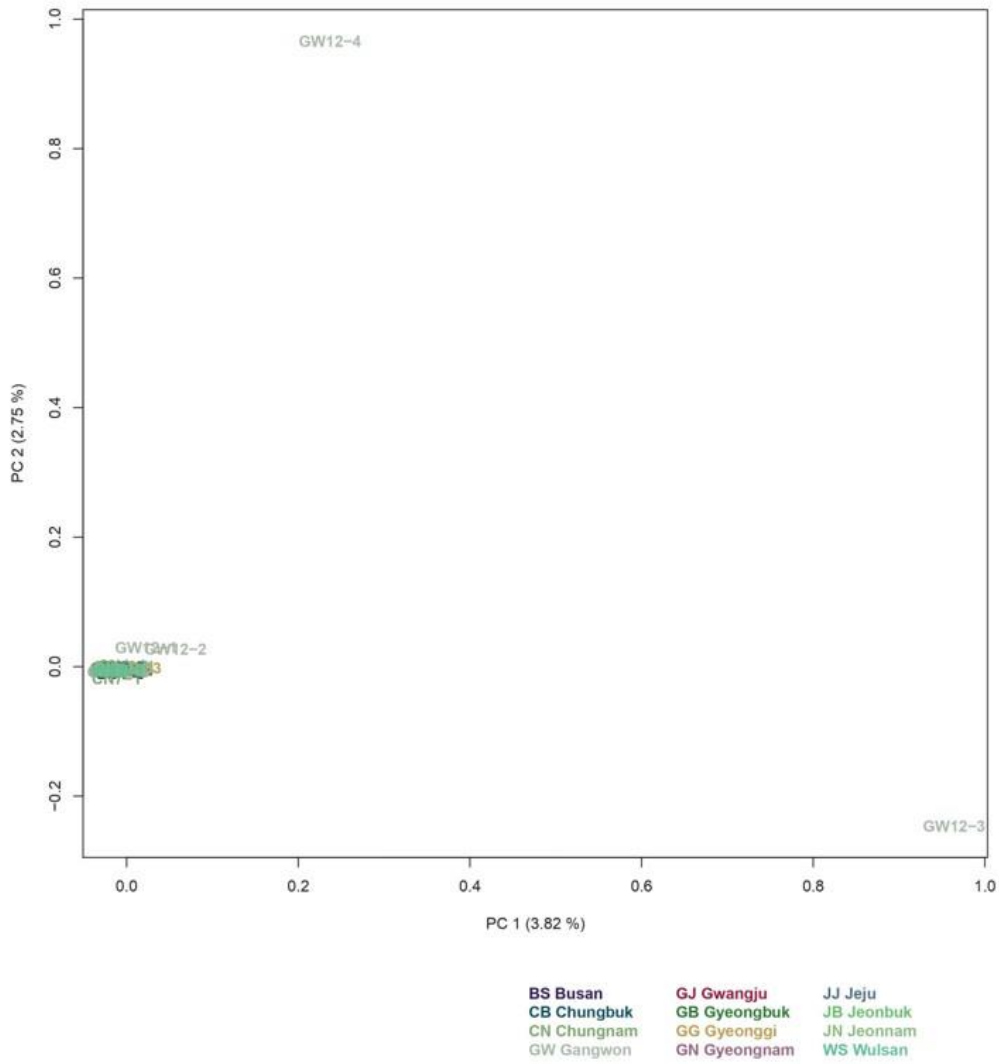


Figure 7. PCA plot of the Korean lineage individuals. The numbers in the bracket indicate the variation explained by each axis. (A) all KL01 individuals (B) all KL02 individuals (C) KL01 individuals without outliers (D) KL02 individuals without outliers.

(B)

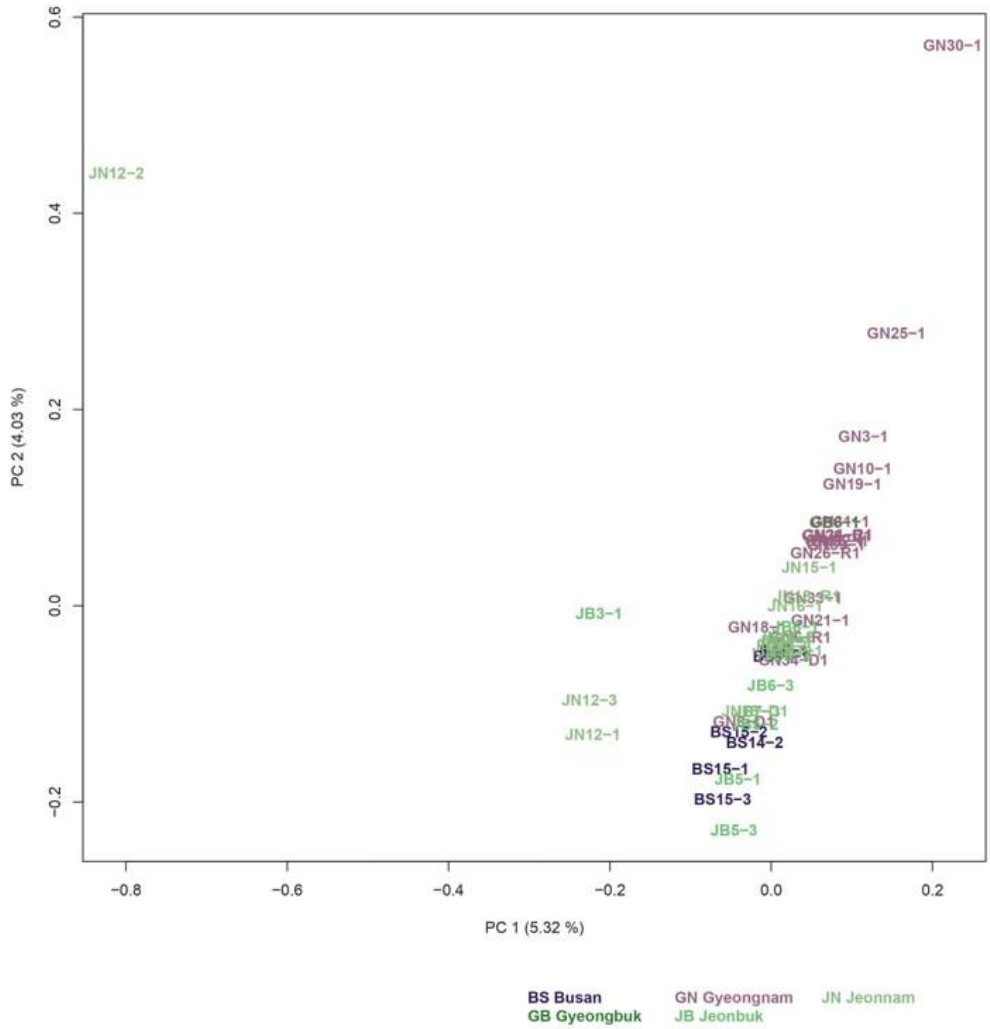


Figure 7. (continued)

(C)

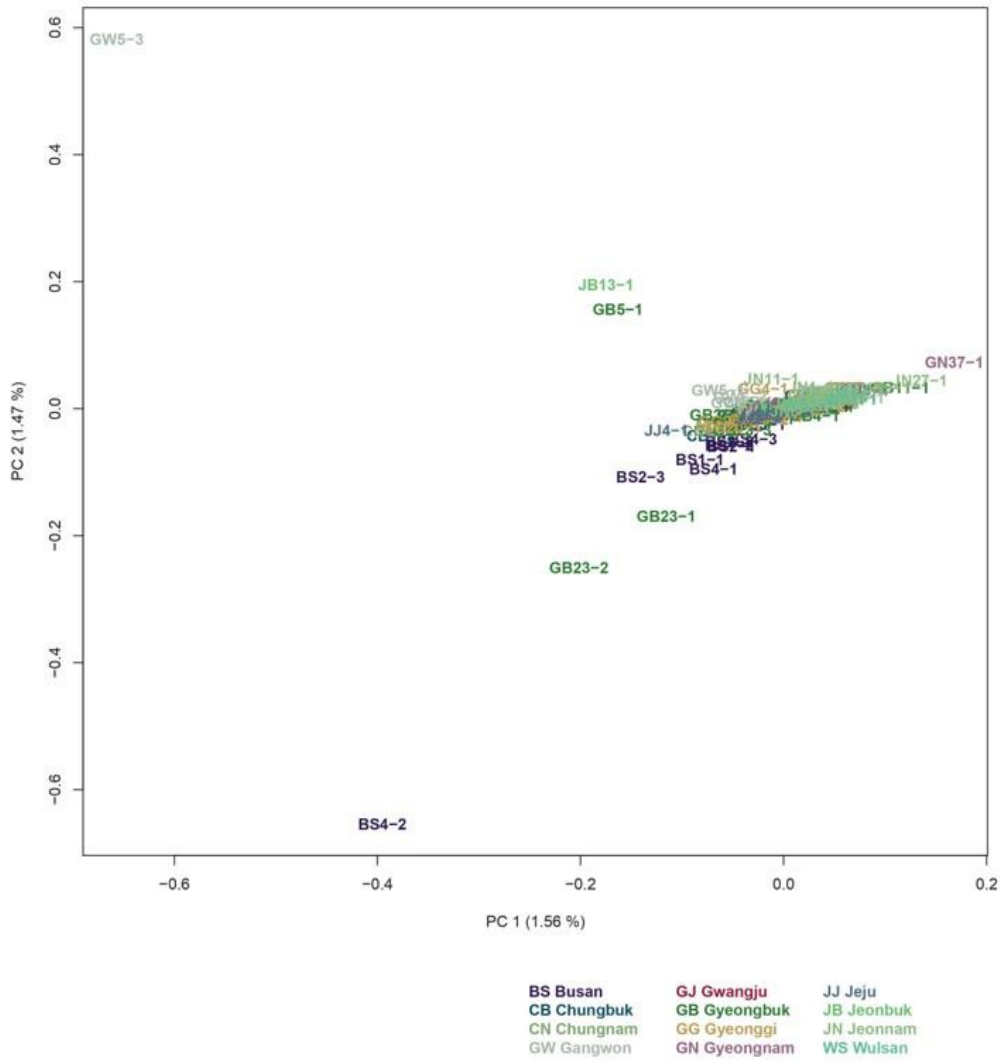


Figure 7. (continued)

(D)

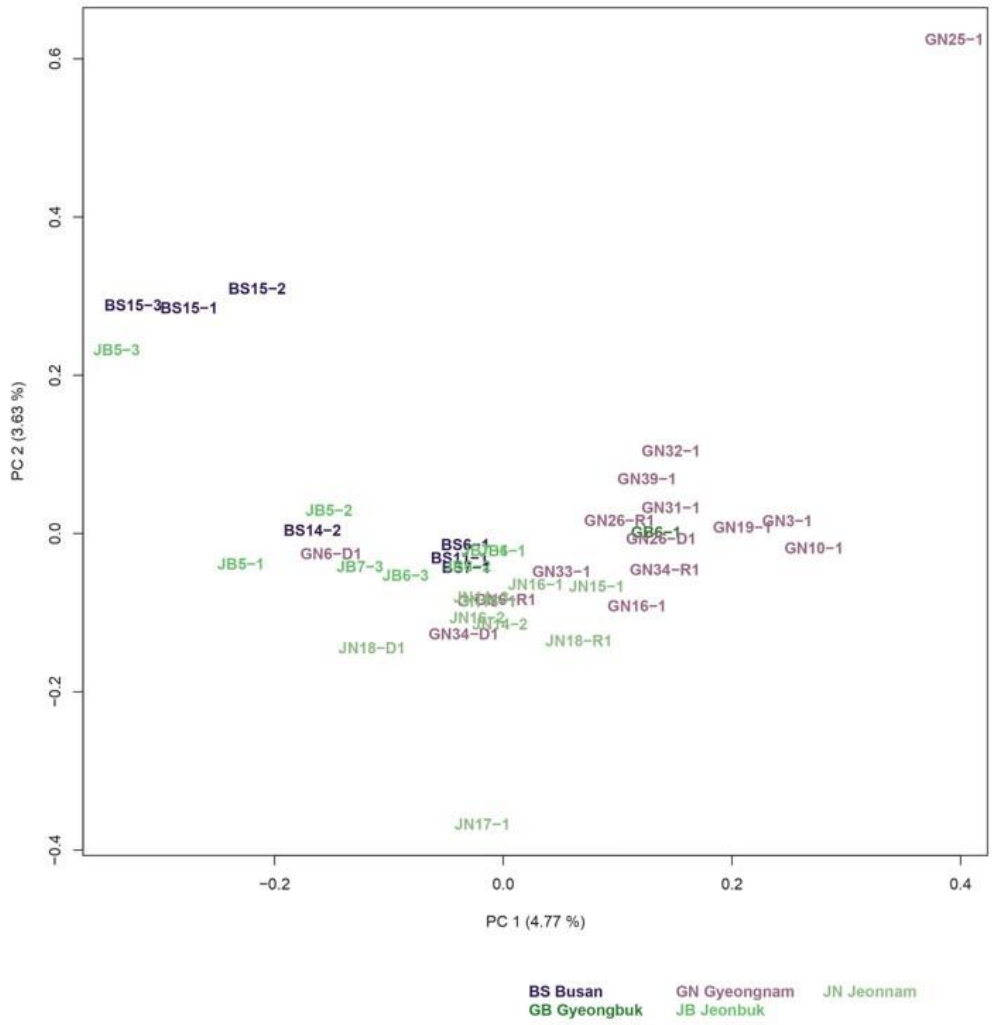


Figure 7. (continued)

(A)

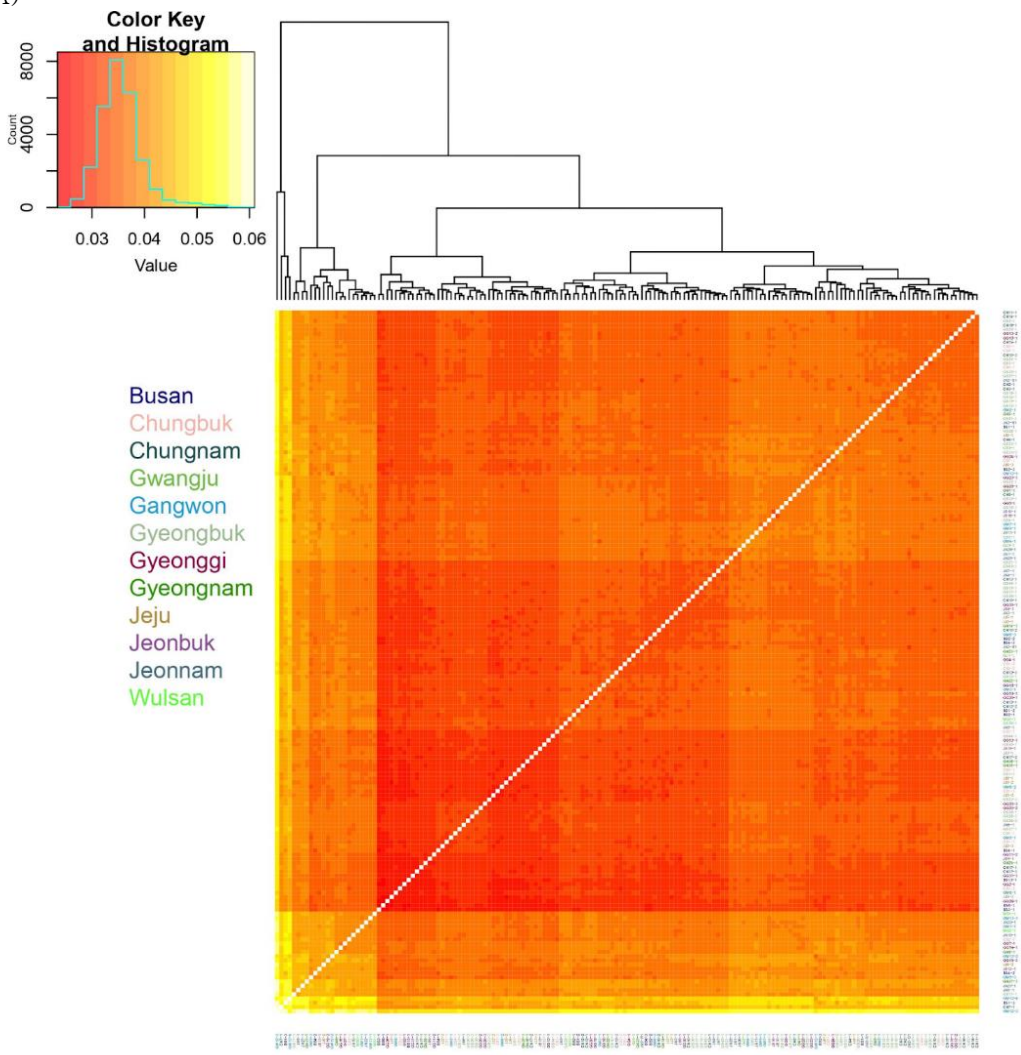


Figure 8. Heatmap of pairwise heatmap rate for the Korean lineage individuals. (A) KL01 (B) KL02.

(B)

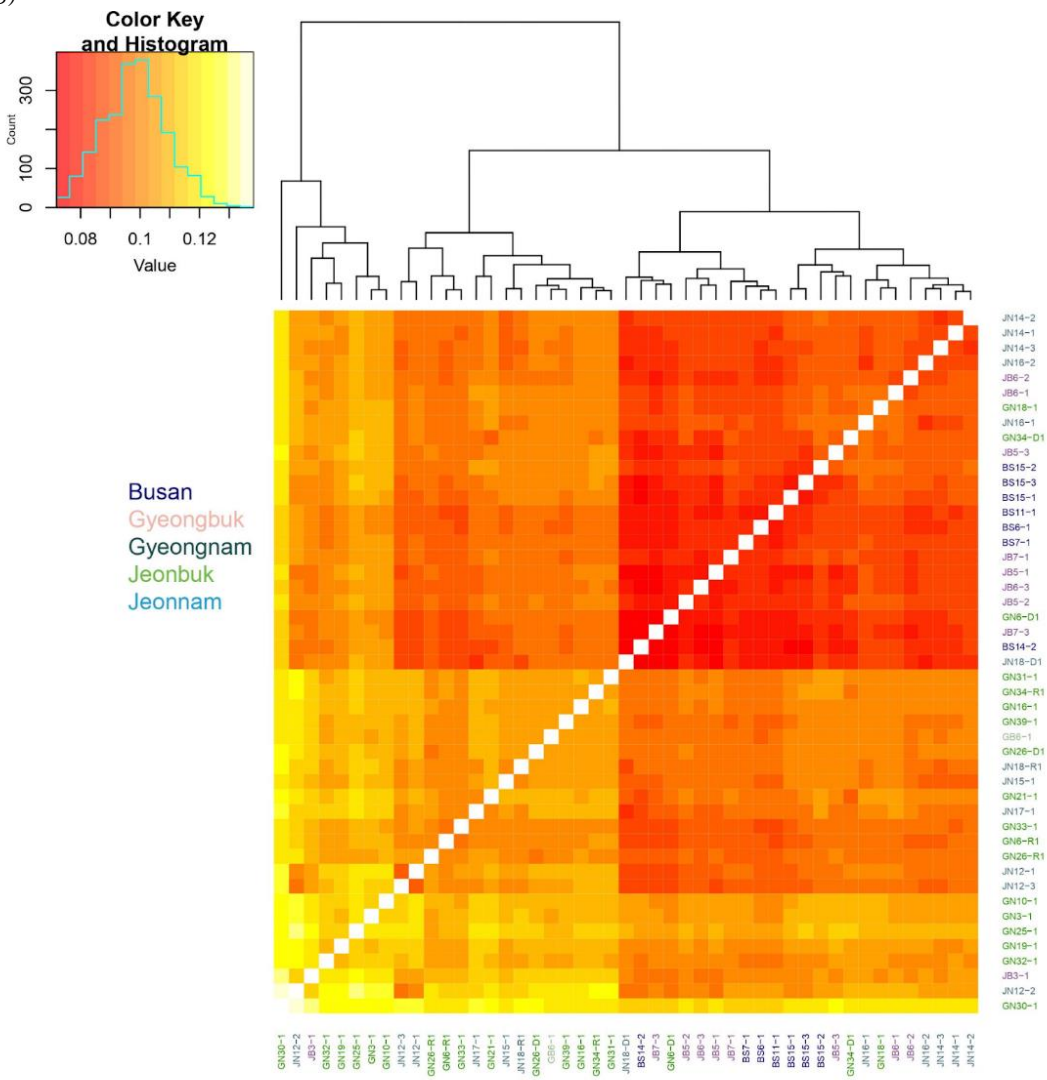


Figure 8. (continued)

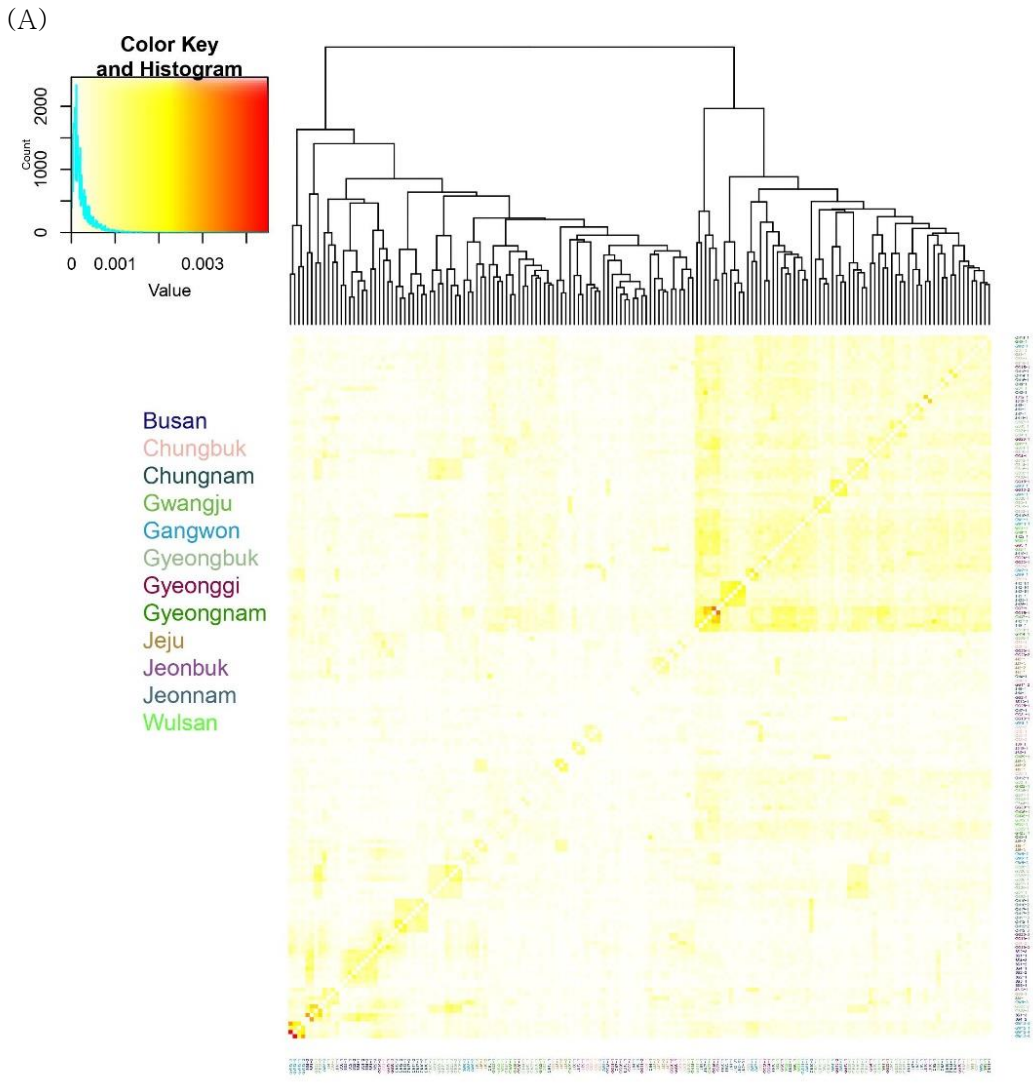


Figure 9. Heatmap of rare variant sharing rate for the Korean lineage individuals. (A) KL01 (B) KL02.

(B)

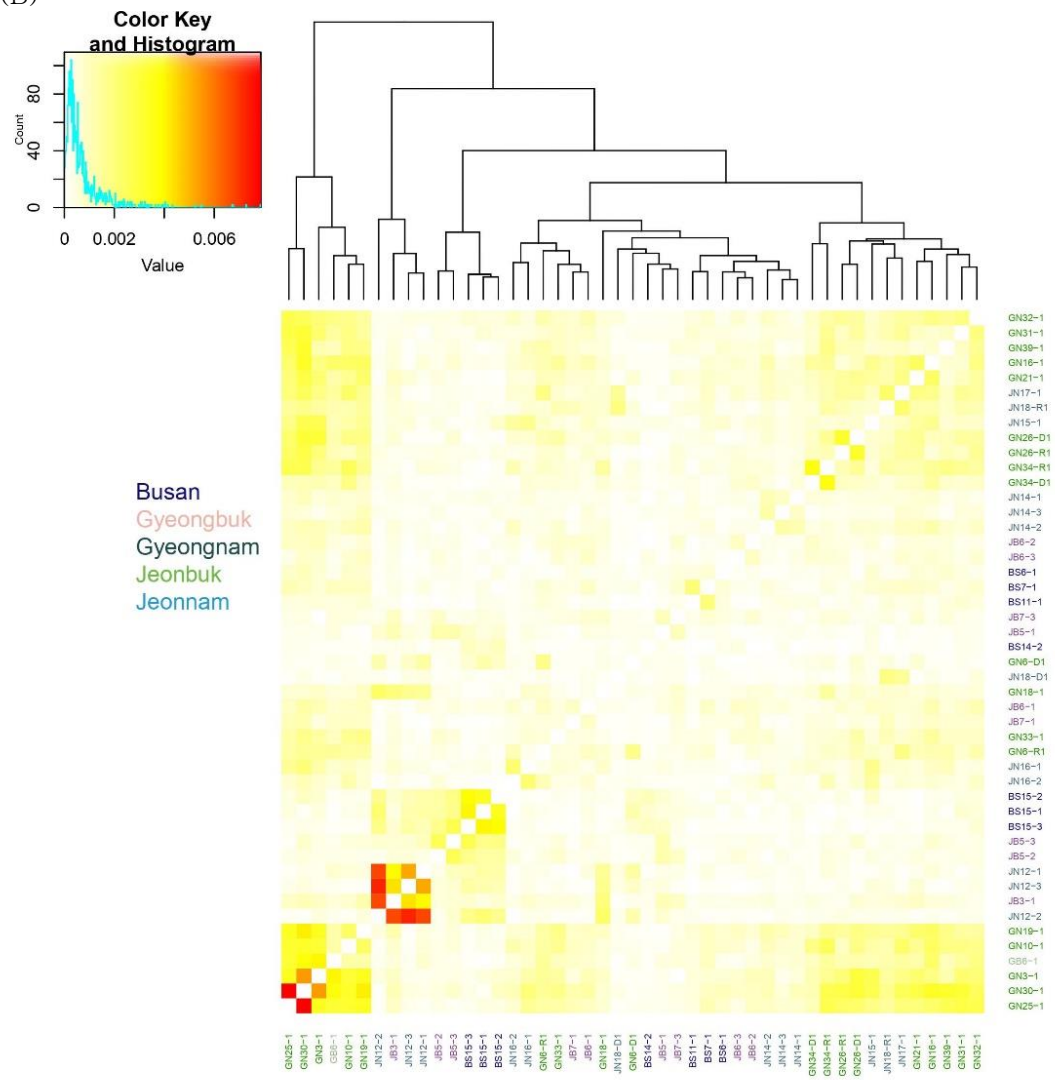


Figure 9. (continued)

(A)

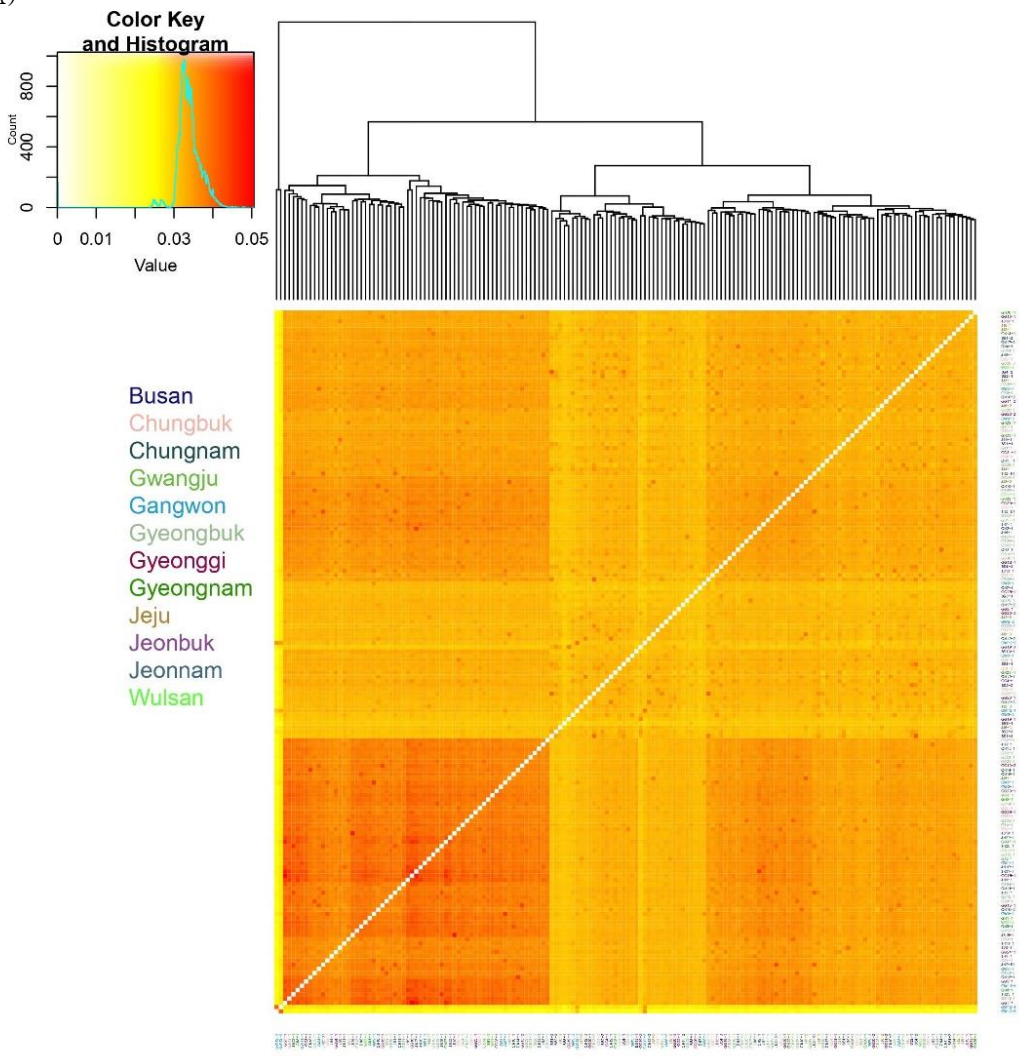


Figure 10. Heatmap of genotype similarity rate for the Korean lineage individuals. (A) KL01 (B) KL02.

(B)

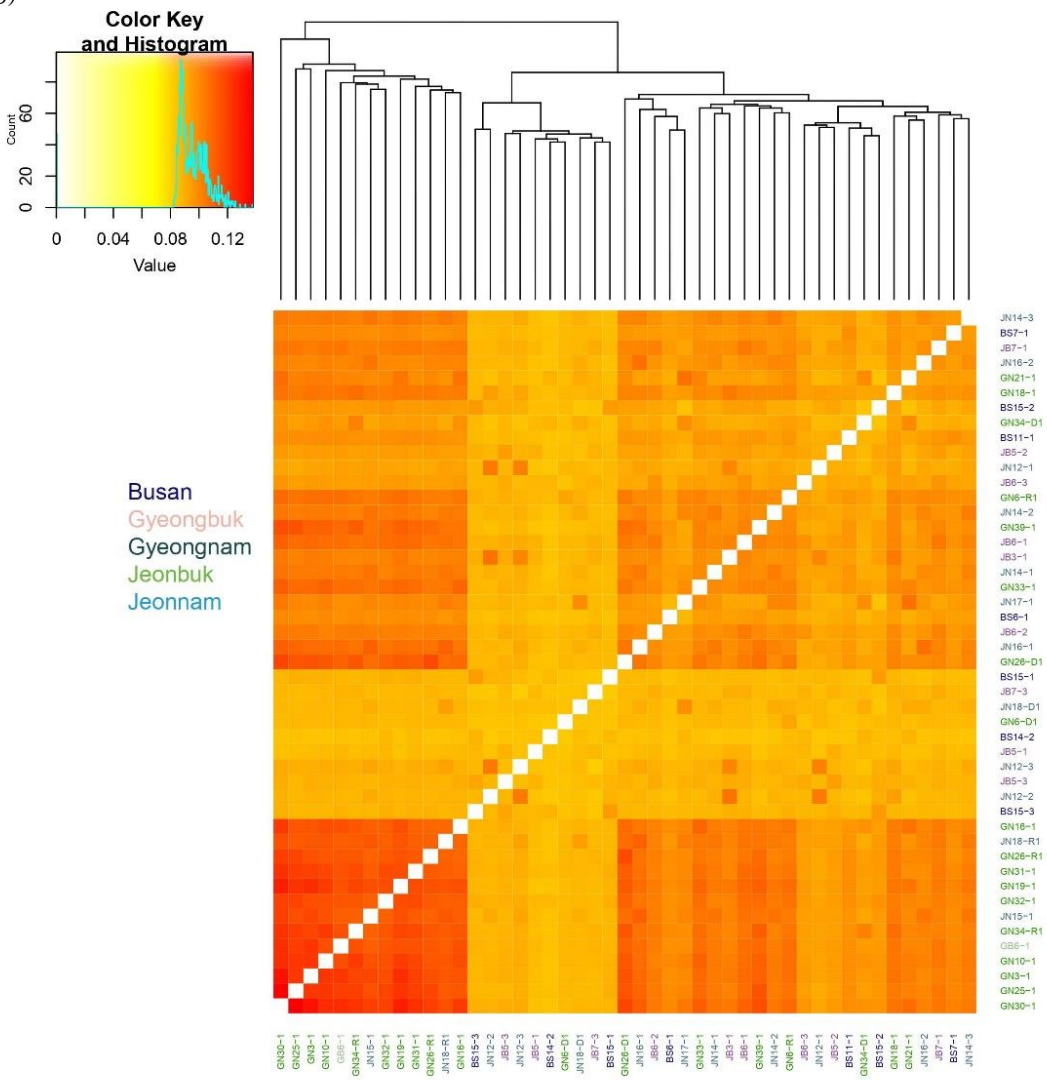


Figure 10. (continued)

3.5. Comparison of genetic diversity between *B. xylophilus* and *B. mucronatus*

To compare the genetic characteristics of *B. xylophilus* and its related species *B. mucronatus*, we calculated the heterozygosity of 281 collected *Bursaphelenchus* samples with >5x coverage (Figure 11). 4 KL04 individuals, the putative hybrid of KL01 and KL02 lineage, inbred line samples, and progenies of two Korean inbred line samples were excluded. 237 *B. xylophilus* individuals showed low and even heterozygosity (0.0041 on average). Interestingly, all 9 KL03 individuals showed relatively higher heterozygosity (0.0047 – 0.0050) than the other *B. xylophilus* individuals. Two *B. mucronatus* lineages (East Asian and European types) showed a much wider range of heterozygosity, while Asian types showed lower values than those European types. The heterozygosity of European type individuals may result from the reference bias, which overestimates the heterozygosity of a distant population due to missed alignment of reads containing alternative alleles.

PCA was performed using *B. mucronatus* samples. East Asian types and European types showed distinguished from each other, with quite a variation within species (Figure. 12). Unlike *B. xylophilus*, even *B. mucronatus* individuals captured from the same site didn't clade to each other (Figure 13). The Asian type samples from each site showed variation between themselves, while European types from 7 sites were clustered into 4 subgroups.

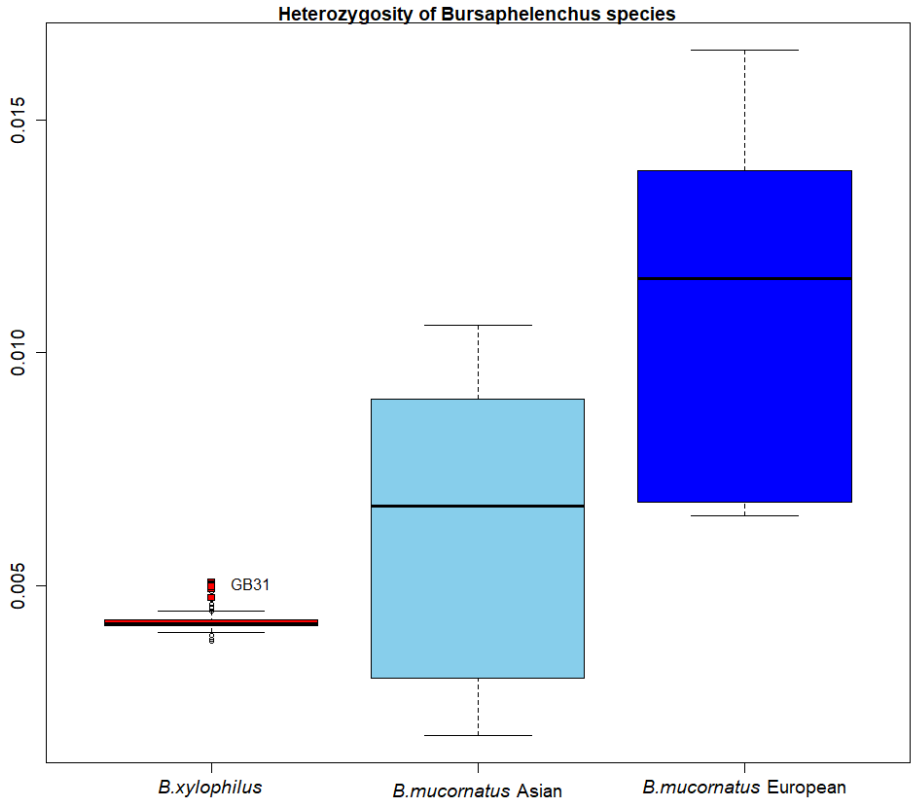


Figure 11. Boxplot of heterozygosity of each species. The red dots indicate the heterozygosity of GB31 individuals.

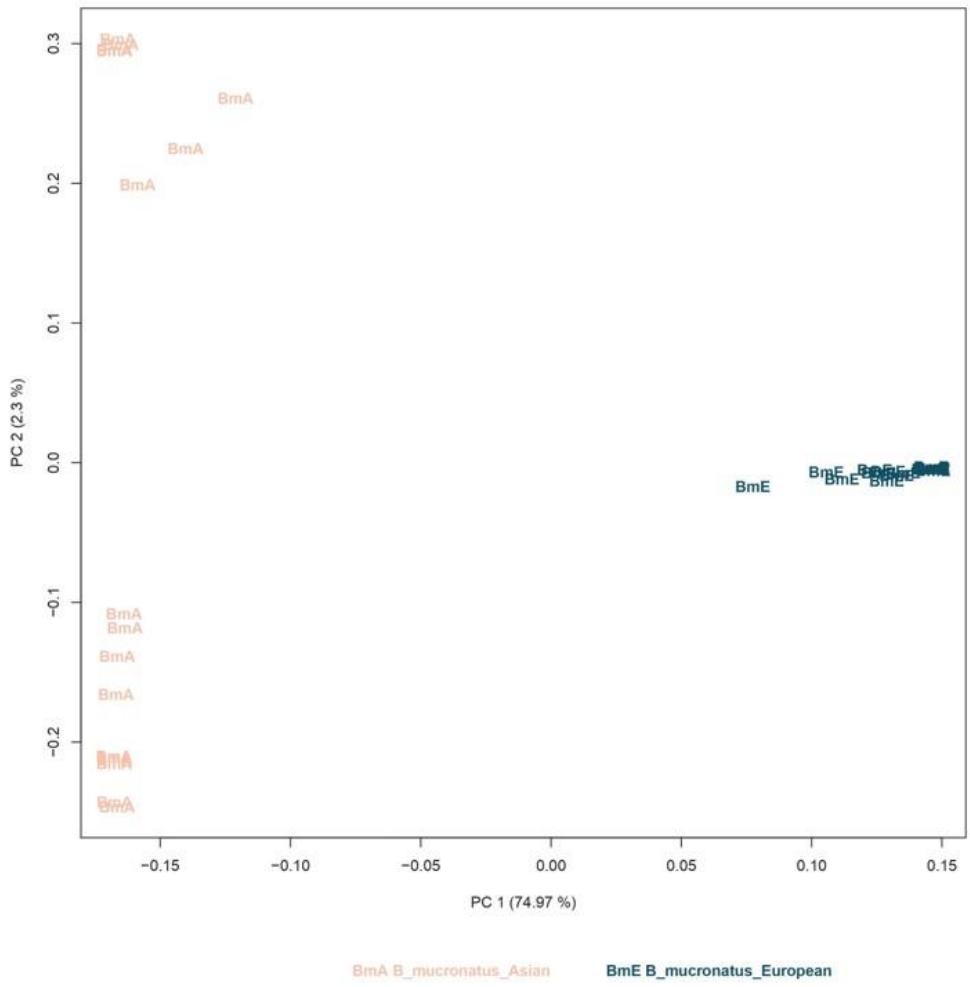


Figure 12. PCA plot of *B. mucronatus* samples. The numbers in the bracket indicate the variation explained by each axis.

(A)

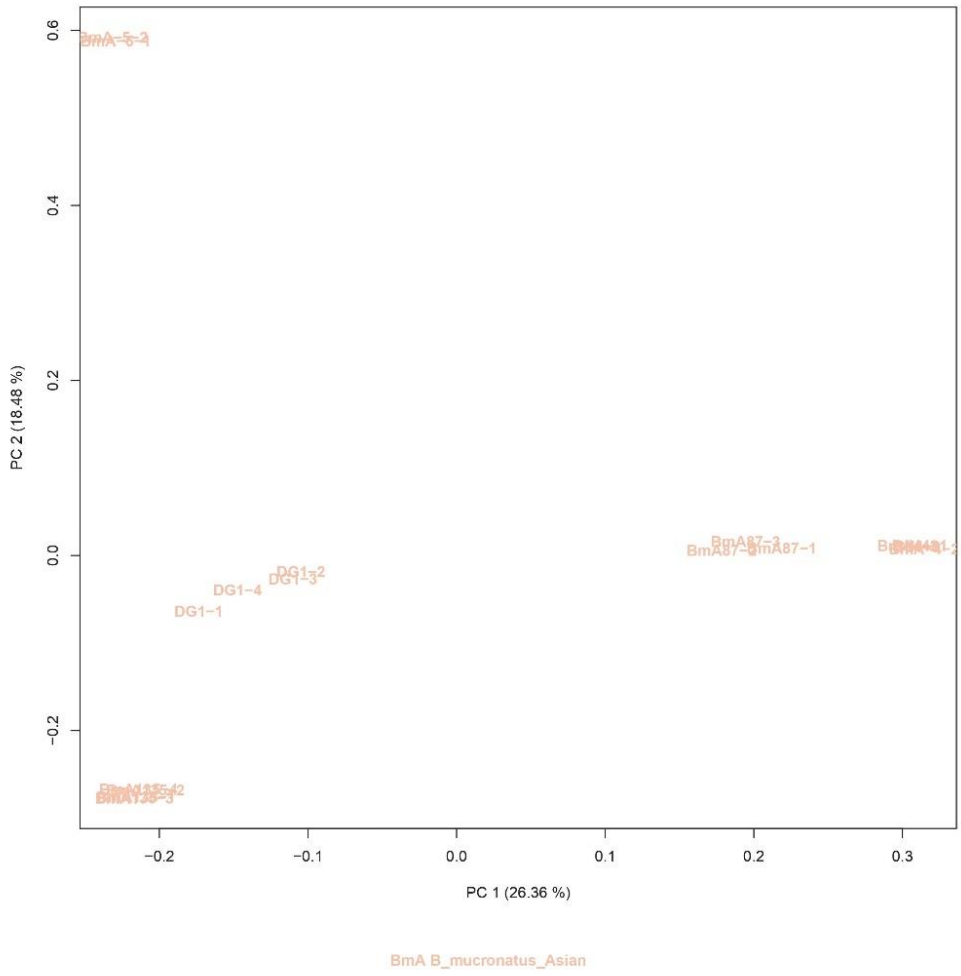


Figure 13. PCA plot of *B. mucronatus* individuals in each type. The numbers in the bracket indicate the variation explained by each axis. (A) Asian types (B) European types.

(B)

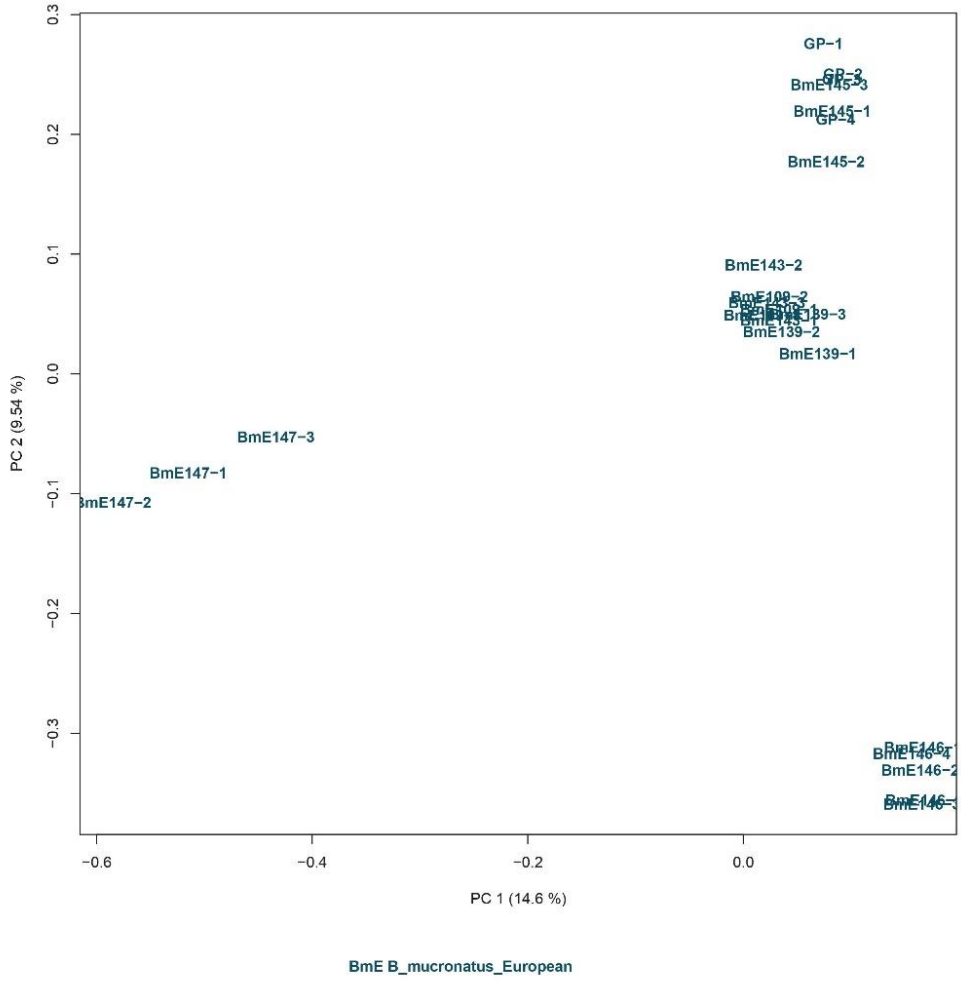


Figure 13. (continued)

4. Discussion

Through the novel individual-level nematode sequencing protocol, whole-genome data of *Bursaphelenchus* individuals was successfully produced. This protocol does not require producing and maintaining the artificial inbred lines which consumes a lot of time and manpower, and provides fine-resolution markers of pinewood nematode. It was available to identify the genetic characteristics of *B. xylophilus* in South Korea, especially genetic heterozygosity, and genetic diversity of nematode individuals collected from the same isolate, which has been impossible to look at in detail so far. Also, it was possible to see the entire view of *B. xylophilus* in South Korea, with the distribution of each lineage. *B. xylophilus* samples collected from all over the Korean Peninsula showed distinct differences between 5 lineages, and only one lineage among them, KL02, was close to overseas *B. xylophilus* samples S10, and MAD24. Individuals in each lineage showed uniform genetic profiles, regardless of the actual physical distance between sampling sites. It is intriguing how their genetic profiles are maintained. During the dispersal by vector beetles, transferred nematode populations went through severe and serial genetic drift. And after inoculation to host trees, there would be hardly any mixture event with other populations. Yet, *B. xylophilus* showed almost clonal genetic similarity within lineages. Further analysis is required to explain how this uniformity was maintained.

Since the first report of the introduction of *B. xylophilus* into South Korea was in 1988 (Yi et al., 1989), it was unclear how many times these infectious pathogens have been transported. Considering definite reasons, it is reasonable that there were multiple introduction events of *B. xylophilus* into South Korea. First, the distinct differences between KL01 and KL02 disagree with the single introduction hypothesis. Genetic similarity within lineages was sustained during the dispersal, whereas the F_{st} of the Korean population was quite high. Also, no populations with intermediate

profiles between KL01 and KL02 were found except in only one site, which is clear to be a hybrid between the two lineages. It is unconvincing that the ancestral *B. xylophilus* population of the Korean lineages, with high genetic diversity, diverged after introduction into South Korea, and then lost its heterozygosity. Rather, it is much easier and reasonable to explain that multiple lineages were introduced independently. Thus, the two Korean lineages KL01 and KL02 might be introduced and dispersed individually, and then KL04 might be produced.

The chances of producing a hybrid of two lineages are very low. Considering its procedure, two vector beetles carrying different lineages should lay the eggs in the same host tree. From 173 sampled sites, only one isolate was a hybrid. This supports the low possibility of forming a hybrid in filed.

It is still unsettled which lineage was introduced into South Korea in advance. If KL01 was the former invader, for three decades they would be moved north from Busan, and then KL02 was introduced recently. This hypothesis corroborates the distribution pattern of each lineage, as KL01 is dispersed all over the Korean Peninsula, while KL02 is located only in the southern part. On the contrary, if KL02 was the former invader and KL01 was the latter, there must be a phenotypic disparity between the two lineages. This scenario might be plausible when considering the temperature, which is the key element to their growth (Panesar et al., 1994; Dwinell, 1997). The southern part of the Korean Peninsula has a moderate and warm climate, compared to the central region. And there was a report about cold tolerance variation between *B. xylophilus* isolates sampled from different sites (Li et al., 2022).

The introduction route of *B. xylophilus* into South Korea was analyzed. Considering the distribution in South Korea and the existence of Japanese inbred line S10 (Shinya et al., 2012; Tanaka et al., 2017; Ekino et al., 2018), KL02 seems to be introduced from Japan, through the port in Busan and go north. Yet, the relationships between KL02 and other overseas isolates are unclear. It is confident that Portuguese *B. xylophilus* originated from East Asia

(Burgermeister et al., 1999; Metge et al., 2006; Vieira et al., 2007; Fonseca et al., 2012), but which country, Japan or South Korea, or another East Asian country, was the direct origin is uncertain. Also, the origins of KL01, KL03, and lineJ remained unknown. Additional sources are required to explain the origin of the rest Korean lineages.

To discover the dispersal routes of *B. xylophilus* in South Korea, subpopulation detection was conducted with various statistical methods. A geographic variant-sharing pattern was shown, but it was not enough to track the spreading path, due to the almost clonal genetic profile of individuals in each lineage, regardless of the geographic distance between sampling sites. The outliers from each Korean lineage might be the key to tracking the dispersal route. 3 individuals of JN12 and 1 individual of JB3 from KL02, for example, acted as an outlier in PCA analysis and share the rare variants with themselves (Figure. 7B, Figure. 9B). The distance between the two sampling sites is about 76 kilometers, but the genetic similarity seems clear.

Compared to its relative species, *B. xylophilus* showed lower heterozygosity and a uniform genetic profile. Two subspecies of *B. mucronatus* were differentiated from each other, and genetic diversity within the subspecies was high. It is unclear whether there are several sublineages in each type of *B. mucronatus*. A lower level of intraspecific diversity than *B. mucronatus* was already reported (Pereira et al., 2013), even in a small subset of the genome. Interestingly, only *B. xylophilus* sustained its genetic profile, while *B. mucronatus* also undergo a severe genetic drift process during its dispersal by vector beetles. The maintainability of genetic property might be associated with the pathogenicity of *B. xylophilus* (2013; Selman et al., 2013).

In conclusion, using the novel sequencing protocol allowing individual-level genome sequencing, the multiple invasion scenario of *B. xylophilus* in South Korea was suggested, and excessive homozygosity of its genome with the genetic similarity between within-lineage individuals regardless of geographic distance was discovered. Also, by comparing its non-pathogenic relative species,

the uniformity of genetic property was proposed to be the cause of the pathogenicity of *B. xylophilus*. This study will help to identify the source of the pathogenicity of *B. xylophilus* and prevent the worldwide spreading of the invasive species.

References

Abelleira, A., & Navas, A. (2011). Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhrer, 1934 (Nickle, 1970) in Spain. *Nematology*, *13*(6), 755-757.

Akbulut, S., Yüksel, B., Serin, M., Baysal, I., & Erdem, M. (2007). Pathogenicity of *Bursaphelenchus mucronatus* in pine seedlings under greenhouse conditions. *Turkish Journal of Agriculture and Forestry*, *31*(3), 169-173.

Bakke, A., Anderson, R. V., & Kvamme, T. (1991). Pathogenicity of the nematodes *Bursaphelenchus xylophilus* and *B. mucronatus* to Pinus sylvestris seedlings: a greenhouse test. *Scandinavian Journal of Forest Research*, *6*(1-4), 407-412.

Burgermeister W et al (1999) First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. Nematology 1:727–734

Blunt, T. D., Jacobi, W. R., Appel, J. A., Tisserat, N., & Todd, T. C. (2014). First report of pine wilt in Colorado, USA. *Plant health progress*, *15*(3), 103-104.

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *The American Journal of Human Genetics*, *98*(1), 116-126.

C. elegans Sequencing Consortium*. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science, 282(5396), 2012-2018.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-

015.

Chen, G., Yuan, A., Shriner, D., Tekola-Ayele, F., Zhou, J., Bentley, A. R., ... & Rotimi, C. N. (2015). An improved FST estimator. *Plos one*, *10*(8), e0135368.

Cheng, X. Y., Cheng, F. X., Xu, R. M., & Xie, B. Y. (2008). Genetic variation in the invasive process of *Bursaphelenchus xylophilus* (Aphelenchida: *Aphelenchoididae*) and its possible spread routes in China. *Heredity*, *100*(4), 356-365.

Cotton, J. A., Steinbiss, S., Yokoi, T., Tsai, I. J., & Kikuchi, T. (2016). An expressed, endogenous Nodavirus-like element captured by a retrotransposon in the genome of the plant parasitic nematode *Bursaphelenchus xylophilus*. *Scientific reports*, *6*(1), 1-8.

Dayi, M., Sun, S., Maeda, Y., Tanaka, R., Yoshida, A., Tsai, I. J., & Kikuchi, T. (2020). Nearly complete genome assembly of the pinewood nematode *Bursaphelenchus xylophilus* strain Ka4C1. *Microbiology resource announcements*, *9*(42), e01002-20.

Ding, X., Guo, Y., Ye, J., Wu, X., Lin, S., Chen, F., ... & Li, D. W. (2022). Population differentiation and epidemic tracking of *Bursaphelenchus xylophilus* in China based on chromosome-level assembly and whole-genome sequencing data. *Pest Management Science*, *78*(3), 1213-1226.

Dwinell, L. D. (1997). The pinewood nematode: regulation and mitigation. Annual review of Phytopathology, 35(1), 153-166.

Ekino, T., Tanaka, S. E., Kanzaki, N., & Takeuchi-Kaneko, Y. (2018). Tolerance to oxidative stress of inbred strains of the pine wood nematode, *Bursaphelenchus xylophilus*, differing in terms of virulence. *Nematology*, *20*(6), 539-546.

Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what?. *Molecular ecology*, *19*(19), 4113-4130.

Figueiredo, J., Simões, M. J., Gomes, P., Barroso, C., Pinho, D., Conceição, L., ... & Egas, C. (2013). Assessment of the geographic origins of pinewood nematode isolates via single nucleotide polymorphism in effector genes. *PLoS One*, *8*(12), e83542.

Filipiak, A. (2015). Pathogenicity of selected isolates of the quarantine pinewood nematode *Bursaphelenchus xylophilus* to Scots pine (Pinus sylvestris L.). *Journal of Plant Protection Research*, *55*(4).

Filipiak, A., Malewski, T., Matczyńska, E., & Tomalak, M. (2021). Molecular variation among virulent and avirulent strains of the quarantine nematode *Bursaphelenchus xylophilus*. *Molecular Genetics and Genomics*, *296*(2), 259-269.

Fonseca, L., Cardoso, J. M. S., Lopes, A., Pestana, M., Abreu, F., Nunes, N., ... & Abrantes, I. (2012). The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia*, *49*(2), 96-103.

Guan, W., Liang, L., Boehnke, M., & Abecasis, G. R. (2009). Genotype-based matching to correct for population stratification in large-scale case-control genetic association studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, *33*(6), 508-517.

Han, H., Chung, Y. J., & Shin, S. C. (2008). First report of pine wilt disease on *Pinus koraiensis* in Korea. *Plant disease*, *92*(8), 1251-1251.

Han, H., Chung, Y. J., & Shin, S. C. (2008). Molecular biological characterization of *Bursaphelenchus xylophilus* and *Bursaphelenchus mucronatus* in Korea. *Forest Science and Technology*, *4*(2), 45-50.

Jaeho Woo. (2022). High-quality Genome Assembly and Comparative Genomic Analysis of Korean Pinewood Nematodes / 우재호 [electronic Resource].

Jones, J. T., Moens, M., Mota, M., Li, H., & Kikuchi, T. (2008). *Bursaphelenchus xylophilus*: opportunities in comparative genomics and molecular host-parasite interactions. *Molecular plant pathology*, *9*(3), 357-368.

Jung, J., Han, H., Ryu, S. H., & Kim, W. (2010). Amplified fragment length polymorphism analysis and genetic variation of the pinewood nematode *Bursaphelenchus xylophilus* in South Korea. *Animal Cells and Systems*, *14*(1), 31-36.

Jung, J., Han, H., Ryu, S. H., & Kim, W. (2010). Microsatellite variation in the pinewood nematode, *Bursaphelenchus xylophilus* (Steiner and Buhrer) Nickle in South Korea. *Genes & Genomics*, *32*(2), 151-158.

Kanzaki, N., & Giblin-Davis, R. M. (2018). Diversity and plant pathogenicity of *Bursaphelenchus* and related nematodes in relation to their vector bionomics. *Current Forestry Reports*, *4*(2), 85-100.

Kim, B.-N., Kim, J.H., Ahn, J.-Y., Kim, S., Cho, B.-K., Kim, Y.-H., and Min, J. (2020). A short review of the pinewood nematode, *Bursaphelenchus xylophilus*. Toxicology and Environmental Health Sciences 12, 297-304.

Kikuchi, T., Cotton, J. A., Dalzell, J. J., Hasegawa, K., Kanzaki, N., McVeigh, P., ... & Berriman, M. (2011). Genomic insights into the

origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS pathogens*, *7*(9), e1002219.

KIYOHARA, T., & TOKUSHIGE, Y. (1971). Inoculation experiments of a nematode, *Bursaphelenchus sp.*, onto pine trees. Journal of the Japanese Forestry Society, 53(7), 210-218.

Kiyohara, T., & Bolla, R. I. (1990). Pathogenic variability among populations of the pinewood nematode, *Bursaphelenchus xylophilus*. Forest Science, 36(4), 1061-1076.

Lee, D. S., Nam, Y., Choi, W. I., & Park, Y. S. (2017). Environmental factors influencing on the occurrence of pine wilt disease in Korea. *Korean Journal of Ecology and Environment*, *50*(4), 374-380.

Lee, I. H., Han, H., Koh, Y. H., Kim, I. S., Lee, S. W., & Shim, D. (2019). Comparative transcriptome analysis of *Pinus densiflora* following inoculation with pathogenic (*Bursaphelenchus xylophilus*) or non-pathogenic nematodes (*B. thailandae*). *Scientific reports*, *9*(1), 1-11.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Li, Y., Meng, F., Deng, X., Wang, X., Feng, Y., Zhang, W., ... & Zhang, X. (2019). Comparative transcriptome analysis of the pinewood nematode *Bursaphelenchus xylophilus* reveals the molecular mechanism underlying its defense response to host-derived α-pinene. *International journal of molecular sciences*, *20*(4), 911.

Li, Y. L., Fan, C. J., Jiang, X. H., Tian, X. Y., & Han, Z. M. (2021). *Bursaphelenchus xylophilus*: an important pathogenic factor of pine

wilt disease and its relationship with *Bursaphelenchus mucronatus*. *Plant Disease*, *105*(10), 3055-3062.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *bioinformatics*, *25*(14), 1754-1760.

Li, Z., Tao, J., & Zong, S. (2022). Cold Tolerance in Pinewood Nematode *Bursaphelenchus xylophilus* Promoted Multiple Invasion Events in Mid-Temperate Zone of China. *Forests*, *13*(7), 1100.

Li, M., Li, H., Sheng, R. C., Sun, H., Sun, S. H., & Chen, F. M. (2020). The first record of Monochamus saltuarius (Coleoptera; Cerambycidae) as vector of *Bursaphelenchus xylophilus* and its new potential hosts in China. *Insects*, *11*(9), 636.

Mamiya, Y., & Kiyohara, T. (1972). Description of *Bursaphelenchus lignicolus* n. sp.(Nematoda: Aphelenchoididae) from pine wood and histopathology of nematode-infested trees. *Nematologica*, *18*(1), 120-124.

Mamiya, Y. (1988). History of pine wilt disease in Japan. *Journal of nematology*, *20*(2), 219.

Mallez, S., Castagnone, C., Espada, M., Vieira, P., Eisenback, J. D., Mota, M., ... & Castagnone-Sereno, P. (2013). First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci. *PLoS One*, *8*(3), e59165.

Mallez, S., Castagnone, C., Espada, M., Vieira, P., Eisenback, J. D., Harrell, M., ... & Guillemaud, T. (2015). Worldwide invasion routes of the pinewood nematode: What can we infer from population genetics analyses?. *Biological invasions*, *17*(4), 1199-1213.

Mallez, S., Castagnone, C., Lombaert, E., Castagnone-Sereno, P., &

Guillemaud, T. (2021). Inference of the worldwide invasion routes of the pinewood nematode *Bursaphelenchus xylophilus* using approximate Bayesian computation analysis. *Peer Community Journal*, *1*.

Matsunaga, K., Akiba, M., Kanzaki, N., & Togashi, K. (2019). A simple method for distinguishing *Bursaphelenchus xylophilus*, *B. mucronatus mucronatus*, and *B. m. kolymensis* (Nematoda: Aphelenchoididae) by polymerase chain reaction with specific primers designed based on cytochrome oxidase subunit I genes. Applied entomology and zoology, 54(1), 147-153.

Metge, K., & Bürgermeister, W. (2006). Intraspecific variation in isolates of *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) revealed by ISSR and RAPD fingerprints. *Journal of Plant Diseases and Protection*, *113*(6), 275-282.

Molloy, E. K., Durvasula, A., & Sankararaman, S. (2021). Advancing admixture graph estimation via maximum likelihood network orientation. *Bioinformatics*, *37*(Supplement_1), i142-i150.

Naves, P. M., Camacho, S., De Sousa, E. M., & Quartau, J. A. (2007). Transmission of the pine wood nematode *Bursaphelenchus xylophilus* through feeding activity of *Monochamus galloprovincialis* (Col., Cerambycidae). *Journal of Applied Entomology*, *131*(1), 21-25.

Nunes da Silva, M., Solla, A., Sampedro, L., Zas, R., & Vasconcelos, M. W. (2015). Susceptibility to the pinewood nematode (PWN) of four pine species involved in potential range expansion across Europe. *Tree physiology*, *35*(9), 987-999.

Palomares-Rius, J. E., Tsai, I. J., Karim, N., Akiba, M., Kato, T., Maruyama, H., ... & Kikuchi, T. (2015). Genome-wide variation in the pinewood nematode *Bursaphelenchus xylophilus* and its

relationship with pathogenic traits. *BMC genomics*, *16*(1), 1-13.

Panesar, T. S., Peet, F. G., Sutherland, J. R., & Sahota, T. S. (1994). Effects of temperature, relative humidity and time on survival of pinewood nematodes in wood chips. European journal of forest pathology, 24(5), 287-299.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, *2*(12), e190.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., ... & Reich, D. (2012). Ancient admixture in human history. Genetics, 192(3), 1065-1093.

Pereira, F., Moreira, C., Fonseca, L., van Asch, B., Mota, M., Abrantes, I., & Amorim, A. (2013). New insights into the phylogeny and worldwide dispersion of two closely related nematode species, *Bursaphelenchus xylophilus* and *Bursaphelenchus mucronatus*. *PLoS One*, *8*(2), e56288.

Robertson, L., Arcos, S. C., Escuer, M., Merino, R. S., Esparrago, G., Kanzaki, N., Aikawa, T., Maehara, N., & Matsumoto, K. (2008). *Bursaphelenchus doui* Braasch, Gu, Burgermeister & Zhang, 2005 (Aphelenchida: Parasitaphelenchidae), an associate of *Monochamus subfasciatus* Bates (Coleoptera: Cerambycidae) and *Pinus densiflora* Sieb. & Zucc. *Nematology*, *10*(1), 69-78.

Robinet, C., Roques, A., Pan, H., Fang, G., Ye, J., Zhang, Y., & Sun, J. (2009). Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One*, *4*(2), e4646.

Rödelsperger, C., Neher, R. A., Weller, A. M., Eberhardt, G., Witte, H., Mayer, W. E., ... & Sommer, R. J. (2014). Characterization of genetic diversity in the nematode *Pristionchus pacificus* from population-scale resequencing data. Genetics, 196(4), 1153-1165.

Rödelsperger, C., Meyer, J. M., Prabh, N., Lanz, C., Bemm, F., & Sommer, R. J. (2017). Single-molecule sequencing reveals the chromosome-scale genomic architecture of the nematode model organism *Pristionchus pacificus*. Cell reports, 21(3), 834-844.

Selman, M., Sak, B., Kváč, M., Farinelli, L., Weiss, L. M., & Corradi, N. (2013). Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*. *Eukaryotic Cell*, *12*(4), 496-502.

Shin, H., Lee, H., Woo, K. S., Noh, E. W., Koo, Y. B., & Lee, K. J. (2009). Identification of genes upregulated by pinewood nematode inoculation in Japanese red pine. *Tree physiology*, *29*(3), 411-421.

Sultana, T., Kim, J., Lee, S. H., Han, H., Kim, S., Min, G. S., ... & Park, J. K. (2013). Comparative analysis of complete mitochondrial genome sequences confirms independent origins of plant-parasitic nematodes. *BMC evolutionary biology*, *13*(1), 1-17.

Sultana, T., Han, H., & Park, J. K. (2013). Comparison of complete mitochondrial genomes of pine wilt *nematode Bursaphelenchus xylophilu*s and *Bursaphelenchus mucronatus* (Nematoda: Aphelenchoidea) and development of a molecular tool for species identification. Gene, 520(1), 39-46.

Ste-Croix, D. T., St-Marseille, A. F. G., Lord, E., Bélanger, R. R., Brodeur, J., & Mimee, B. (2021). Genomic profiling of virulence in the soybean cyst nematode using single-nematode sequencing. *Phytopathology®*, *111*(1), 137-148.

Tanaka, S. E., Aikawa, T., Takeuchi-Kaneko, Y., Fukuda, K., & Kanzaki, N. (2017). Artificial induction of third-stage dispersal juveniles of *Bursaphelenchus xylophilus* using newly established inbred lines. *PloS one*, *12*(10), e0187127.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*, *43*(1), 11-10.

Valadas, V., Laranjo, M., Mota, M., & Oliveira, S. (2013). Molecular characterization of Portuguese populations of the pinewood nematode *Bursaphelenchus xylophilus* using cytochrome b and cellulase genes. *Journal of Helminthology*, *87*(4), 457-466.

Valadas, V., Laranjo, M., Barbosa, P., Espada, M., Mota, M., & Oliveira, S. (2012). The pine wood nematode, *Bursaphelenchus xylophilus*, in Portugal: possible introductions and spread routes of a serious biological invasion revealed by molecular methods. *Nematology*, *14*(8), 899-911.

Vicente, C. S., Nascimento, F., Espada, M., Barbosa, P., Mota, M., Glick, B. R., & Oliveira, S. (2012). Characterization of bacteria associated with pinewood nematode *Bursaphelenchus xylophilus*.

Vicente, C. S., Nascimento, F. X., Espada, M., Barbosa, P., Hasegawa, K., Mota, M., & Oliveira, S. (2013). Characterization of bacterial communities associated with the pine sawyer beetle *Monochamus galloprovincialis*, the insect vector of the pinewood nematode *Bursaphelenchus xylophilus*. *FEMS microbiology letters*, *347*(2), 130-139.

Vieira, P., Burgermeister, W., Mota, M., Metge, K., & Silva, G. (2007). Lack of genetic variation of *Bursaphelenchus xylophilus* in Portugal revealed by RAPD-PCR analyses. *Journal of Nematology*, *39*(2), 118.

Vincent, B., Altemayer, V., Roux-Morabito, G., Naves, P., Sousa, E.,

& Lieutier, F. (2008). Competitive interaction between *Bursaphelenchus xylophilus* and the closely related species *Bursaphelenchus mucronatus*. Nematology, 10(2), 219-230.

Wang, X. Z., Wang, X., Zhang, S. F., Zhang, Y. L., Cao, Y. F., & Wang, L. F. (2022). The Effect of Pine Wood Nematode *Bursaphelenchus xylophilus* (Steiner and Buhrer) Nickle on Intestinal Bacterial Community of Insect Vector *Monochamus saltuarius* (Coleoptera: Cerambycidae). *Forests*, *13*(10), 1673.

Wu, S., Gao, S., Wang, S., Meng, J., Wickham, J., Luo, S., ... & Sun, J. (2020). A reference genome of *Bursaphelenchus mucronatus* provides new resources for revealing its displacement by pinewood nematode. *Genes*, *11*(5), 570.

Yi, C. K., Byun, B. H., Park, J. D., Yang, S. I., & Chang, K. H. (1989). First finding of the pine wood nematode*, Bursaphelenchus xylophilus* (Steiner et Buhrer) Nickle and its insect vector in Korea. *Research Reports of the Forestry Research Institute (Seoul)*, (38), 141-149.

Zas, R., Moreira, X., Ramos, M., Lima, M. R., Nunes da Silva, M., Solla, A., ... & Sampedro, L. (2015). Intraspecific variation of anatomical and chemical defensive traits in Maritime pine (*Pinus pinaster*) as factors in susceptibility to the pinewood nematode (*Bursaphelenchus xylophilus*). *Trees*, *29*(3), 663-673.

Zhang, W., Yu, H., Lv, Y., Bushley, K. E., Wickham, J. D., Gao, S., ... & Sun, J. (2020). Gene family expansion of pinewood nematode to detoxify its host defense chemicals. *Molecular ecology*, *29*(5), 940-955.

Zhao, L. L., Wei, W., Kang, L., & Sun, J. H. (2007). Chemotaxis of the pinewood nematode, *Bursaphelenchus xylophilus*, to volatiles associated with host pine, Pinus massoniana, and its vector

Monochamus alternatus. *Journal of chemical ecology*, *33*(6), 1207-1216.

Zhao, L., Mota, M., Vieira, P., Butcher, R. A., & Sun, J. (2014). Interspecific communication between pinewood nematode, its insect vector, and associated microbes. *Trends in parasitology*, *30*(6), 299-308.

Zhuo, K., Luo, M., Cui, R. Q., & Liao, J. L. (2011). A multiplex one-step PCR method for the simultaneous identification of *Bursaphelenchus xylophilus*, *B. mucronatus* and *B. doui*–three species within the xylophilus group. *Forest Pathology*, *41*(1), 66-69.

Zhou, Z., Sakaue, D., Wu, B., & Hogetsu, T. (2007). Genetic structure of populations of the pinewood nematode *Bursaphelenchus xylophilus*, the pathogen of pine wilt disease, between and within pine forests. *Phytopathology*, *97*(3), 304-310.

Zhou, L. F., Chen, F. M., Xie, L. Y., Pan, H. Y., & Ye, J. R. (2017). Genetic diversity of pine-parasitic nematodes *Bursaphelenchus xylophilus* and Bursaphelenchus mucronatus in China. *Forest Pathology*, *47*(4), e12334.

국문 초록

국내 소나무재선충의 기원 및 유전적 다양성 분석

소나무재선충(*Bursaphelenchus xylophilus*)은 침입한 국가의 산림에 막대한 피해를 끼치는 해충으로서, 소나무마름병의 원인으로 지목된 이래 전세계적으로 많은 연구가 진행되었다. 한국에서는 1988년 부산에서 처음으로 보고되었으며, 유입된 이후 약 30년 동안 한반도 전역으로 퍼져나갔으나 정확한 유래와 전파 경로 파악 및 유입 횟수 등에 대한 연구는 아직까지 명확한 진척을 보이지 못하였다. 소나무재선충의 생활사 및 상호작용하는 생물종이 밝혀짐에 따라 이것이 소나무재선충의 병원성에 미치는 영향에 대한 연구가 많이 이루어졌으며, 소나무재선충의 기원 및 전파 경로에 대한 연구 또한 수행되었다. 그러나 집단 분석의 경우 개체로부터 충분한 양의 유전 물질을 확보하는 것이 어려워 대부분의 연구에서는 소규모 마커 유전자를 활용하거나 실험실에서 제작된 동계교배라인, 혹은 대량의 선충으로부터 일괄적으로 유전 물질을 추출하는 등 여러 우회책을 통하여 진행되었다는 한계점이 존재하였다. 본 연구에서는 개체 단위 총유전체 서열 결정 기법을 새로이 개발하였으며, 이를 이용하여 국내 유래 소나무재선충 359개체와 두 종의 근연종 48개체의 유전체를 시퀀싱하였다. 국내 173개 지역에서 채집된 248개체의 소나무재선충 및 두 종의 한국 동계교배라인과 이들의 자손 세대 111개체, 그리고 해외에서 출판된 동계교배라인 8개체의 유전자형 자료를 바탕으로 고품질의 변이 패널을 제작하였다. 또한 생산한 자료를 바탕으로 국내 소나무재선충의 유전적 특성 및 집단 구조, 기원에 대한 분석을 수행하였다. 그 결과 국내에서 채집된 소나무재선충을 5개의 계통으로 구분할 수 있었으며, 이들 계통이 국내에 최소 2회 이상 독립적으로 유입되었음을 유추할 수 있었다. 또한 이중 한 계통은 일본 및 포르투갈에서 보고된 소나무재선충과 가까운 계통임을 확인하였다. 각 계통은 분포 지역에서 큰

차이를 보였으며, 국내에 가장 많이 퍼진 2개 계통은 채집 지역에 관계없이 거의 동일한 유전자형을 나타내었다. 이로 인해 지역별 하위 소집단의 흔적은 발견하였으나, 정확한 전파 경로 추적은 불가능하였다. 한편, 소나무재선충 가까운 계통 중 숙주에 병원성을 나타내지 않는 두 종류의 어리소나무재선충(*B. mucronatus*) 42개체의 유전체 자료를 생산하고 이를 소나무재선충 자료와 비교함으로써 소나무재선충의 유전체는 비병원성의 근연종에 비해 낮은 이형접합도 및 유전적 다양성을 보임을 확인하였다. 소나무재선충의 낮은 유전적 다양성 및 전파 과정 중 유전적 특성이 유지되는 특성이 소나무재선충의 병원성과 관계가 있을 것으로 추론하였다. 본 연구를 통해 소나무재선충의 전세계적인 전파 경로 파악과 소나무재선충의 병원성을 유발하는 유전적 특성에 대한 향후 분석에 큰 도움이 될 것으로 기대된다.

Appendices

Appendix 1. Summary of *B. xylophilus* and *B. mucronatus* samples used in this study.

Region	Isolate	No. of sample	Lineage	Latitude	Longitude
Busan	BS1	3	KL01	35.29	129.12
Busan	BS2	3	KL01	35.29	129.11
Busan	BS3	1	KL01	35.29	129.11
Busan	BS4	3	KL01	35.24	129.09
Busan	BS6	1	KL02	35.14	129.03
Busan	BS7	1	KL02	35.17	129.10
Busan	BS8	1	KL01	35.23	129.14
Busan	BS11	1	KL02	35.15	129.11
Busan	BS12	4	KL04	35.15	129.11
Busan	BS13	1	KL01	35.09	129.07
Busan	BS14	3	KL02	35.27	129.11
Busan	BS15	3	KL02	35.25	129.12
Chungbuk	CB1	3	KL01	36.68	127.26
Chungbuk	CB2	3	KL01	36.66	127.22
Chungbuk	CB3	2	KL01	36.67	127.17
Chungbuk	CB4	1	KL01	36.25	127.75
Chungbuk	CB5	1	KL01	36.64	127.47
Chungbuk	CB6	1	KL01	36.95	128.19
Chungbuk	CB7	1	KL01	36.92	128.19
Chungbuk	CB8	1	KL01	36.96	128.29
Chungbuk	CB9	1	KL01	37.01	127.68
Chungnam	CN2	1	KL01	36.61	126.61
Chungnam	CN3	1	KL01	36.30	127.05
Chungnam	CN4	1	KL01	36.30	127.05
Chungnam	CN5	1	KL01	36.30	127.05
Chungnam	CN7	1	KL01	36.68	126.93
Chungnam	CN8	1	KL01	36.37	126.73
Chungnam	CN10	3	KL01	36.40	126.56
Chungnam	CN11	1	KL01	36.91	127.18
Chungnam	CN12	3	KL01	36.91	127.18
Chungnam	CN13	1	KL01	36.15	126.65
Chungnam	CN14	1	KL01	36.44	126.92
Chungnam	CN15	1	KL01	36.43	126.93
Chungnam	CN16	1	KL01	36.75	126.30
Chungnam	CN17	3	KL01	36.38	126.56
Gyeongbuk	GB1	1	KL01	36.00	129.19
Gyeongbuk	GB2	1	KL01	35.71	128.35
Gyeongbuk	GB3	1	KL01	35.84	128.45

Appendix 1. (continued)

Region	Isolate	No.of sample	Lineage	Latitude	Longitude
Gyeongbuk	GB4	1	KL01	35.77	128.30
Gyeongbuk	GB5	1	KL01	36.04	128.04
Gyeongbuk	GB6	1	KL02	36.05	128.08
Gyeongbuk	GB9	1	KL01	36.45	128.67
Gyeongbuk	GB10	1	KL01	36.57	128.92
Gyeongbuk	GB11	1	KL01	36.52	128.83
Gyeongbuk	GB12	1	KL01	36.58	129.38
Gyeongbuk	GB14	1	KL01	36.51	129.41
Gyeongbuk	GB15	1	KL01	36.03	129.05
Gyeongbuk	GB16	1	KL01	36.03	128.91
Gyeongbuk	GB17	1	KL01	36.02	128.85
Gyeongbuk	GB18	1	KL01	35.64	128.65
Gyeongbuk	GB19	1	KL01	35.71	128.58
Gyeongbuk	GB20	1	KL01	35.66	128.67
Gyeongbuk	GB21	1	KL01	36.08	128.40
Gyeongbuk	GB22	1	KL01	36.05	128.36
Gyeongbuk	GB23	3	KL01	36.02	129.34
Gyeongbuk	GB26	3	KL01	36.02	129.34
Gyeongbuk	GB28	1	KL01	36.02	129.34
Gyeongbuk	GB29	1	KL01	36.02	129.34
Gyeongbuk	GB31	9	KL03	35.87	128.86
Gyeongbuk	GB32	1	KL01	35.81	128.36
Gyeongbuk	GB35	1	KL01	36.31	128.59
Gyeongbuk	GB36	1	KL01	36.35	128.49
Gyeongbuk	GB37	1	KL01	36.14	128.29
Gyeongbuk	GB38	1	KL01	36.14	128.29
Gyeongbuk	GB40	1	KL01	36.14	128.29
Gyeongbuk	GB44	1	KL01	36.23	128.54
Gyeongbuk	GB45	1	KL01	36.23	128.54
Gyeongbuk	GB49	1	KL01	36.70	128.60
Gyeongbuk	GB50	1	KL01	36.54	128.24
Gyeongbuk	GB51	1	KL01	36.77	128.63
Gyeongbuk	GB52	1	KL01	36.69	129.40
Gyeongbuk	GB56	1	KL01	35.79	128.81
Gyeongbuk	GB57	1	KL01	36.31	128.59
Gyeonggi	GG2	2	KL01	37.95	127.52
Gyeonggi	GG4	1	KL01	37.37	127.40
Gyeonggi	GG5	1	KL01	37.46	127.21
Gyeonggi	GG7	1	KL01	37.63	127.31
Gyeonggi	GG8	1	KL01	37.93	127.07
Gyeonggi	GG11	2	KL01	37.62	127.45
Gyeonggi	GG13	2	KL01	37.41	127.74

Appendix 1. (continued)

Region	Isolate	No. of sample	Lineage	Latitude	Longitude
Gyeonggi	GG15	1	KL01	37.27	127.73
Gyeonggi	GG16	1	KL01	37.40	127.54
Gyeonggi	GG18	1	KL01	38.07	127.06
Gyeonggi	GG19	2	KL01	37.99	127.05
Gyeonggi	GG20	3	KL01	38.01	127.11
Gyeonggi	GG23	1	KL01	37.24	127.25
Gyeonggi	GG24	1	KL01	37.39	127.00
Gyeonggi	GG25	1	KL01	37.31	127.42
Gyeonggi	GG26	1	KL01	36.97	126.89
Gyeonggi	GG30	1	KL01	37.49	127.22
Gyeonggi	GG31	1	KL01	37.54	127.23
Gyeonggi	GG32	1	KL01	37.07	126.96
Gyeonggi	GG33	1	KL01	37.83	127.52
Gwangju	GJ1	1	KL01	35.22	126.75
Gwangju	GJ2	1	KL01	35.12	126.85
Gyeongnam	GN1	1	KL01	34.88	128.62
Gyeongnam	GN3	1	KL02	35.12	128.27
Gyeongnam	GN5	1	KL01	35.31	128.93
Gyeongnam	GN6	2	KL02	35.34	128.84
Gyeongnam	GN8	1	KL01	34.75	127.84
Gyeongnam	GN10	1	KL02	34.75	127.84
Gyeongnam	GN14	1	KL01	35.39	128.76
Gyeongnam	GN16	1	KL02	35.28	127.94
Gyeongnam	GN18	1	KL02	35.40	129.16
Gyeongnam	GN19	1	KL02	35.59	128.50
Gyeongnam	GN21	1	KL02	35.54	128.56
Gyeongnam	GN22	1	KL01	35.19	128.64
Gyeongnam	GN23	1	KL01	34.85	128.43
Gyeongnam	GN25	1	KL02	35.35	128.52
Gyeongnam	GN26	2	KL02	35.33	128.50
Gyeongnam	GN29	1	KL01	35.41	128.19
Gyeongnam	GN30	1	KL02	35.15	128.27
Gyeongnam	GN31	1	KL02	35.35	128.29
Gyeongnam	GN32	1	KL02	35.32	128.27
Gyeongnam	GN33	1	KL02	35.20	128.10
Gyeongnam	GN34	2	KL02	35.10	128.15
Gyeongnam	GN35	1	KL01	35.09	127.94
Gyeongnam	GN36	1	KL01	35.02	127.82
Gyeongnam	GN37	1	KL01	34.95	127.86
Gyeongnam	GN39	1	KL02	35.28	128.21

Appendix 1. (continued)

Region	Isolate	No. of sample	Lineage	Latitude	Longitude
Gangwon	GW1	1	KL01	37.75	127.80
Gangwon	GW2	1	KL01	37.45	128.67
Gangwon	GW3	1	KL01	37.48	127.82
Gangwon	GW4	1	KL01	37.50	127.90
Gangwon	GW5	3	KL01	37.23	127.75
Gangwon	GW6	3	KL01	37.22	127.75
Gangwon	GW7	1	KL01	37.45	129.11
Gangwon	GW8	1	KL01	37.45	129.11
Gangwon	GW9	1	KL01	37.45	129.11
Gangwon	GW10	3	KL01	37.21	127.76
Gangwon	GW11	1	KL01	37.68	127.95
Gangwon	GW12	4	KL01	37.50	129.12
Jeonbuk	JB2	1	KL02	35.51	127.16
Jeonbuk	JB3	1	KL02	35.53	127.18
Jeonbuk	JB5	3	KL02	35.38	127.14
Jeonbuk	JB6	3	KL02	35.37	127.14
Jeonbuk	JB7	3	KL02	35.37	127.14
Jeonbuk	JB8	1	KL01	35.97	126.71
Jeonbuk	JB9	1	KL01	35.96	126.60
Jeonbuk	JB10	1	KL01	35.80	126.88
Jeonbuk	JB13	1	KL01	35.73	126.73
Jeonbuk	JB14	1	KL01	35.60	126.81
Jeonbuk	JB15	1	KL01	35.57	126.86
Jeonbuk	JB16	1	KL01	35.85	127.01
Jeju	JJ2	3	KL01	33.33	126.75
Jeju	JJ3	1	KL01	33.51	126.53
Jeju	JJ4	1	KL01	33.50	126.54
Jeju	JJ5	3	KL01	33.32	126.82
Jeju	JJ6	3	KL01	33.32	126.82
Jeonnam	JN1	1	KL01	34.99	127.71
Jeonnam	JN2	3	KL01	34.98	127.68
Jeonnam	JN3	1	KL01	35.20	127.46
Jeonnam	JN4	1	KL01	35.31	127.47
Jeonnam	JN5	1	KL01	34.92	126.42
Jeonnam	JN6	1	KL01	34.92	126.42
Jeonnam	JN7	1	KL01	34.87	127.35
Jeonnam	JN9	1	KL01	34.81	127.43
Jeonnam	JN10	1	KL01	34.95	127.32
Jeonnam	JN11	1	KL01	34.91	127.34
Jeonnam	JN12	3	KL02	34.91	127.56
Jeonnam	JN14	3	KL02	34.92	127.55
Jeonnam	JN15	1	KL02	34.91	127.55

Appendix 1. (continued)

Region	Isolate	No. of sample	Lineage	Latitude	Longitude
Jeonnam	JN16	3	KL02	34.92	127.55
Jeonnam	JN17	1	KL02	34.76	127.66
Jeonnam	JN18	2	KL02	34.88	127.56
Jeonnam	JN21	1	KL01	34.81	127.71
Jeonnam	JN22	1	KL01	35.32	126.78
Jeonnam	JN25	1	KL01	34.94	127.70
Jeonnam	JN26	1	KL01	34.94	127.70
Jeonnam	JN27	1	KL01	35.28	127.29
Wulsan	WS1	1	KL01	35.54	129.20
Wulsan	WS2	1	KL01	35.59	129.31
Wulsan	WS3	1	KL01	35.67	129.35
Unknown	BmA87	3	BmA		
Unknown	BmE145	3	BmE		
Gyeongbuk	BmA135	4	BmA	36.00	128.40
Gyeongnam	BmA136	2	BmA	35.02	128.73
Gyeonggi	BmE109	3	BmE	37.75	127.18
Gangwon	BmE139	3	BmE	37.38	128.66
Gyeonggi	BmE143	3	BmE	37.49	127.49
Gyeonggi	BmE146	5	BmE	38.25	127.94
Seoul	BmE147	3	BmE	37.60	126.93
Gyeonggi	GP1	4	BmE	37.83	127.51
Daegu	DG1	4	BmA	35.83	128.57

Appendix 2. Summary of sequenced *B. xylophilus* and *B. mucronatus* samples.

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
BS1-1	61287616	10164487	0.17	13.56	63.01
BS1-2	37802712	11026042	0.29	14.89	71.38
BS1-3	59039066	11873484	0.20	14.18	50.26
BS2-1	30276224	14538540	0.48	20.29	37.72
BS2-2	31904622	13116801	0.41	18.15	43.11
BS2-3	66949834	12380786	0.18	15.21	34.53
BS3-1	10658672	9996202	0.94	15.15	5.03
BS4-1	23019066	11875922	0.52	17.04	32.88
BS4-2	34650956	11447317	0.33	16.01	42.96
BS4-3	31688606	14830232	0.47	20.61	32.61
BS6-1	13278906	12697777	0.96	19.72	5.10
BS7-1	23859920	22718800	0.95	35.36	5.00
BS8-1	7881378	7466882	0.95	11.77	3.62
BS11-1	13184068	12349571	0.94	18.49	4.09
BS12-1	39408666	37254403	0.95	52.68	6.67
BS12-2	21406140	20162452	0.94	29.51	6.03
BS12-3	22162148	20860613	0.94	30.69	5.93
BS12-4	39693054	36489938	0.92	56.54	5.84
BS13-1	12793820	12119987	0.95	19.23	3.71
BS14-1	29241692	5692727	0.19	8.27	27.44
BS14-2	21371932	8258172	0.39	12.41	49.67
BS14-3	21109478	4349040	0.20	6.26	26.79
BS15-1	53445784	9510300	0.18	12.85	35.43
BS15-2	37487142	10996642	0.29	15.20	36.78
BS15-3	62055650	10291146	0.17	14.12	44.22
CB1-1	22907762	13137785	0.57	18.53	33.52
CB1-2	46013596	12162653	0.26	16.37	29.72
CB1-3	25747462	12312523	0.48	17.91	21.59
CB2-1	13567890	12984569	0.96	18.80	9.84
CB2-2	16261794	15482011	0.95	21.25	18.89
CB2-3	69191962	66341474	0.96	97.10	8.59

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
CB3-1	31219510	28060435	0.90	43.94	2.12
CB3-2	37519744	17032701	0.45	25.65	42.67
CB4-1	55197712	15471925	0.28	22.15	33.74
CB5-1	69079232	53581047	0.78	78.43	27.37
CB6-1	49791896	3259595	0.07	3.81	78.95
CB7-1	68284798	63592648	0.93	91.73	10.59
CB8-1	70584742	46747144	0.66	67.55	31.30
CB9-1	61798996	57891200	0.94	84.22	11.10
CN2-1	78299436	75153565	0.96	108.23	10.78
CN3-1	44805538	33992728	0.76	49.66	17.88
CN4-1	68006302	14745866	0.22	20.24	62.71
CN5-1	61642598	5780848	0.09	6.87	69.70
CN7-1	57551896	8915044	0.15	11.19	57.77
CN8-1	62503152	58561536	0.94	88.24	6.09
CN10-1	15411772	14729104	0.96	20.98	11.32
CN10-2	15810806	14703407	0.93	19.83	17.80
CN10-3	16682876	15731653	0.94	22.61	10.55
CN11-1	68912342	65260461	0.95	100.78	5.98
CN12-1	29658518	12498882	0.42	17.52	24.91
CN12-2	33150820	12410585	0.37	17.34	25.54
CN12-3	35876880	9695472	0.27	13.70	28.55
CN13-1	46377582	44174350	0.95	64.72	8.70
CN14-1	68807030	65431784	0.95	96.67	8.61
CN15-1	42126548	39939525	0.95	61.62	5.63
CN16-1	50037432	47898293	0.96	70.88	10.09
CN17-1	18818268	16958646	0.90	25.41	5.65
CN17-2	18951672	7624484	0.40	10.98	58.47
CN17-3	21323966	11206375	0.53	17.02	41.73
GB1-1	36050912	15075163	0.42	22.70	57.02
GB2-1	39331876	37410815	0.95	56.09	7.44
GB3-1	45202910	42918126	0.95	65.68	6.44
GB4-1	51459056	39999542	0.78	59.25	21.84
GB5-1	52297338	11449233	0.22	16.18	67.02

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
GB6-1	55084278	43930985	0.80	66.78	18.85
GB9-1	61638168	58522056	0.95	82.23	11.35
GB10-1	84042524	22650090	0.27	30.91	53.95
GB11-1	95989494	88750061	0.92	131.62	6.43
GB12-1	52025272	40876583	0.79	60.17	16.26
GB14-1	36834968	34358372	0.93	51.23	5.80
GB15-1	62345534	59356042	0.95	88.88	7.36
GB16-1	29110390	27153578	0.93	39.61	4.88
GB17-1	21044376	19614240	0.93	28.99	4.84
GB18-1	42622936	39833231	0.93	53.11	11.71
GB19-1	17810728	16531940	0.93	24.32	5.92
GB20-1	18528192	17427395	0.94	24.74	7.04
GB21-1	21347522	20132512	0.94	29.19	5.97
GB22-1	23344280	21897146	0.94	31.48	6.78
GB23-1	38368230	12082682	0.31	16.68	54.01
GB23-2	72866632	14138280	0.19	18.06	46.94
GB23-3	31511858	12159503	0.39	17.21	42.26
GB26-1	47703822	11249397	0.24	14.46	58.41
GB26-2	25913170	12448652	0.48	17.42	51.16
GB26-3	29736782	10604209	0.36	15.10	32.40
GB28-1	35384400	32659091	0.92	46.38	8.35
GB29-1	23919232	22312184	0.93	32.36	6.97
GB31-B1	30468124	27901249	0.92	41.78	5.19
GB31-B2	19139232	18130681	0.95	25.23	6.33
GB31-B3	21504042	19940712	0.93	27.60	6.23
GB31-B4	22443784	21159115	0.94	29.01	9.32
GB31-D1	31862596	6580588	0.21	9.25	64.92
GB31-R1	40047414	37448350	0.94	55.65	6.423
GB31-R2	19660940	18809908	0.96	27.15	6.45
GB31-R3	19609274	18696893	0.95	27.66	5.99
GB31-R4	22870060	21559061	0.94	31.72	5.54
GB32-1	27353348	23944447	0.88	34.05	10.25
GB35-1	30537936	28749987	0.94	40.82	6.98
GB36-1	18895292	17241815	0.91	24.82	7.38

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
GB37-1	39800400	31796948	0.80	49.40	18.08
GB38-1	49410150	13267869	0.27	18.09	30.08
GB40-1	40522814	3346103	0.08	4.02	71.15
GB44-1	12494892	11706335	0.97	16.88	6.65
GB45-1	18330618	16851752	0.92	24.21	8.17
GB47-1	17260942	169960	0.01	0.03	37.39
GB49-1	14447668	11739375	0.81	16.96	16.28
GB50-1	51941244	6997737	0.13	9.10	48.05
GB51-1	22436000	20386790	0.91	26.73	12.42
GB52-1	13498934	12857159	0.95	19.11	6.28
GB55-1	58036	48978	0.84	0.08	9.99
GB56-1	26831568	3163466	0.12	3.87	87.34
GB57-1	60447240	17965635	0.30	24.39	58.31
GG11-1	15361292	1554083	0.10	1.84	31.56
GG11-2	37030328	34683963	0.94	50.02	6.92
GG13-1	16122760	7977266	0.49	11.43	32.60
GG13-2	31947614	30315670	0.95	42.74	9.27
GG15-1	50390446	47624570	0.95	71.82	9.39
GG16-1	62270576	58988257	0.95	83.17	11.24
GG18-1	24988106	23703400	0.95	31.70	10.84
GG19-1	43034044	8631438	0.20	11.58	70.67
GG19-2	12957810	320342	0.02	0.11	75.59
GG19-3	80065530	8871660	0.11	10.58	58.49
GG20-1	26599370	12477420	0.47	17.28	54.10
GG20-2	24332490	12314795	0.51	16.95	52.86
GG20-3	39325044	9430921	0.24	12.82	34.30
GG2-1	13061868	12215502	0.94	17.26	8.19
GG2-2	19888172	4404073	0.22	6.06	64.56
GG23-1	35410962	33163281	0.94	45.39	7.65
GG24-1	45167524	42322075	0.94	60.65	5.91
GG25-1	38011986	35879984	0.94	50.11	7.56
GG26-1	10124014	8129796	0.80	12.41	13.61
GG30-1	13099740	12050585	0.92	17.97	5.25
GG31-1	12911878	11838848	0.92	17.75	3.88

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
GG32-1	35980126	4014820	0.11	5.02	86.62
GG33-1	14135520	5406385	0.38	7.86	36.28
GG4-1	54099964	31426497	0.58	45.44	43.66
GG5-1	50336560	47966397	0.95	70.17	10.31
GG7-1	60235294	56450795	0.94	85.56	7.36
GG8-1	72999918	6064919	0.08	6.65	84.05
GJ1-1	35389034	26382465	0.75	36.80	25.20
GJ2-1	28774854	27060937	0.94	37.49	7.38
GN10-1	44021330	41641432	0.95	61.14	8.17
GN1-1	56277188	54088442	0.96	77.39	11.96
GN14-1	55066880	18856869	0.34	26.15	27.45
GN16-1	61413398	55154491	0.90	84.56	11.32
GN18-1	43254372	23760783	0.55	35.34	47.72
GN19-1	62009054	59568008	0.96	88.11	8.93
GN21-1	69974510	66469432	0.95	99.12	6.99
GN22-1	68389368	37824597	0.55	55.79	45.88
GN23-1	27831156	25764972	0.93	37.51	5.65
GN25-1	33400068	31084938	0.93	44.43	6.42
GN26-D1	29942250	27178126	0.91	40.19	5.53
GN26-R1	35513924	33120491	0.93	49.43	7.00
GN29-1	12623718	11927242	0.94	18.56	4.03
GN30-1	80858994	77383087	0.96	108.92	11.06
GN3-1	54598280	49290390	0.90	72.48	14.03
GN31-1	27781910	26076647	0.94	38.05	6.09
GN32-1	27973854	26309225	0.94	37.22	9.88
GN33-1	38311644	35157552	0.92	52.12	9.64
GN34-D1	26086454	23857022	0.91	35.58	5.08
GN34-R1	36348894	33941963	0.93	49.99	5.83
GN35-1	14151250	13151371	0.93	20.14	3.94
GN36-1	12834566	12092466	0.94	16.83	8.02
GN37-1	30011826	27720244	0.92	36.96	8.98
GN39-1	14860962	13321773	0.90	19.73	3.29
GN5-1	61466718	58535697	0.95	85.73	8.56
GN6-D1	44402380	8940672	0.20	12.06	28.51

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
GN6-R1	27437870	25530202	0.93	38.02	5.67
GN8-1	66228320	63886045	0.96	93.51	11.42
GW1-1	71089098	61844208	0.87	89.25	16.57
GW2-1	54694046	52349982	0.96	79.09	7.80
GW3-1	45220740	42759991	0.95	61.84	8.74
GW4-1	58015206	55282949	0.95	79.45	10.26
GW5-1	39385326	9637470	0.24	13.41	53.75
GW5-2	85402856	6401194	0.07	6.44	56.00
GW5-3	51368880	12407703	0.24	15.23	56.66
GW6-1	90138518	7162367	0.08	7.89	33.97
GW6-2	20653170	10071098	0.49	13.97	25.09
GW6-3	21685800	10176188	0.47	13.72	41.03
GW7-1	54748220	51820563	0.95	75.45	10.44
GW8-1	18112030	7493918	0.41	10.74	22.12
GW9-1	38823794	36879651	0.95	52.62	10.16
GW10-1	20896064	6355281	0.30	9.42	25.51
GW10-2	29406566	4181291	0.14	5.77	32.00
GW10-3	23228608	4011529	0.17	5.82	21.93
GW11-1	39517946	34189265	0.87	48.09	9.92
GW12-1	60697718	14991701	0.25	19.59	50.00
GW12-2	53995406	16061711	0.30	20.07	46.66
GW12-3	86785800	14312531	0.16	11.90	55.23
GW12-4	50792614	17879449	0.35	16.75	52.98
JB2-1	50831194	6618757	0.13	7.80	85.80
JB3-1	62188304	23811797	0.38	31.22	66.55
JB5-1	44885932	9614361	0.21	13.32	78.26
JB5-2	28818496	10911865	0.38	15.83	61.48
JB5-3	42049912	11117395	0.26	15.04	74.56
JB6-1	16740936	15376635	0.92	22.46	13.17
JB6-2	17853856	16790319	0.94	24.62	9.29
JB6-3	16674148	15548737	0.93	22.58	10.77
JB7-1	24318644	22013169	0.91	33.79	2.77
JB7-2	30337730	4678529	0.15	6.86	55.18
JB7-3	47510474	9705128	0.20	14.44	70.55

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
JB8-1	15171590	14244598	0.94	21.53	5.37
JB9-1	14776838	10320945	0.70	16.01	25.66
JB10-1	63437302	45492466	0.72	67.44	29.19
JB13-1	57435866	17167670	0.30	22.15	65.49
JB14-1	15120236	1338575	0.09	1.62	31.85
JB15-1	17857518	16956785	0.95	25.17	7.35
JB16-1	50355574	47250975	0.94	70.09	8.41
JJ2-1	34095364	12939096	0.38	18.12	50.85
JJ2-2	21014988	11811742	0.56	16.92	42.07
JJ2-3	22953960	14640110	0.64	20.47	30.18
JJ3-1	47995432	17669945	0.37	25.78	24.29
JJ4-1	56211228	21159174	0.38	28.18	31.91
JJ5-1	27664872	25212549	0.91	38.07	4.24
JJ5-2	28628282	19166099	0.67	29.26	14.93
JJ5-3	28664056	16987066	0.59	26.44	34.89
JJ6-1	26045428	12401228	0.48	17.83	46.51
JJ6-2	42001788	10159733	0.24	13.98	48.38
JJ6-3	53403584	11035445	0.21	14.43	36.14
JN1-1	35917470	27087733	0.75	37.79	26.47
JN2-B1	32797284	30775848	0.94	46.16	5.84
JN2-D1	34170008	31554864	0.92	46.37	5.54
JN2-R1	31900650	28909169	0.91	43.05	10.67
JN3-1	22015678	20593659	0.94	30.84	6.39
JN4-1	14401152	13260036	0.92	20.29	4.62
JN5-1	34118850	32118383	0.94	44.50	7.98
JN6-1	15624958	14375218	0.92	21.45	8.41
JN7-1	20074128	18659371	0.93	27.73	6.18
JN9-1	1.01E+08	94452479	0.94	133.92	10.52
JN10-1	62042508	59025678	0.95	84.98	14.01
JN11-1	49221090	19259824	0.39	25.29	62.40
JN12-1	34275326	11168729	0.36	15.08	47.87
JN12-2	56776834	10558118	0.19	12.63	52.55
JN12-3	41215714	9117263	0.22	12.14	48.64
JN14-1	15791610	14896891	0.94	22.26	6.90

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
JN14-2	20423098	19393155	0.95	27.86	9.69
JN14-3	15376752	14579738	0.95	21.36	8.53
JN15-1	42661058	40472831	0.95	60.54	8.86
JN16-1	44559432	19829097	0.45	29.97	42.13
JN16-2	47848378	16501214	0.35	25.31	50.62
JN16-3	35989884	3921532	0.11	5.63	47.79
JN17-1	39445066	28896827	0.73	38.97	27.44
JN18-D1	26378626	12073854	0.46	17.98	36.90
JN18-R1	31414506	29736540	0.95	43.16	8.94
JN21-1	61145772	3982965	0.07	4.17	74.19
JN22-1	31039024	29040161	0.94	40.93	6.86
JN25-1	26975646	25572475	0.95	35.82	10.63
JN26-1	55784220	53473596	0.96	78.25	9.93
JN27-1	60418212	57008248	0.94	80.35	10.62
WS1-1	37819406	35689490	0.94	49.96	7.37
WS2-1	30184060	28280138	0.94	40.20	6.73
WS3-1	32509972	30790231	0.95	44.17	6.82
0gen-1	13148086	10945010	0.83	16.11	3.04
0gen-2	15543176	8394853	0.54	12.12	34.00
3gen-1	12077006	9694237	0.80	13.91	2.97
3gen-2	21791324	170787	0.01	0.13	27.54
8gen-1	9514474	7907674	0.83	12.10	2.43
8gen-2	11639532	9734418	0.84	14.56	2.82
FB	46497944	41811273	0.90	61.24	2.73
FJ_12-1	18185298	16076079	0.88	23.26	6.47
FJ_12-2	23860696	21439096	0.90	32.29	4.59
FJ_5-1	19193978	17146424	0.89	24.79	6.74
FJ_5-2	14838688	13079785	0.88	19.06	5.21
Japan	58326120	57524142	0.99	67.22	0.53
JB	43185518	39639932	0.92	58.75	1.24
M1	23171124	19654473	0.85	27.20	6.41
M2	22702942	20235875	0.89	29.45	4.91
M3	21627010	18459934	0.85	26.28	3.72

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
P2	20812038	20166533	0.97	24.53	3.92
P2U	21544328	20888552	0.97	27.22	4.91
fHa-8-1	10359382	9059439	0.87	14.47	2.43
fHa-8-2	7456112	6688981	0.90	10.77	3.52
fHa-8-3	8423338	7640300	0.91	12.34	3.16
fHa-8-4	9842834	8656776	0.88	14.10	2.51
fHa-8-5	8693954	7698791	0.89	12.23	2.97
fHa-8-6	11447694	10372774	0.91	16.72	2.66
fHa-8-7	7671676	6956910	0.91	11.22	3.32
fHa-8-9	8163178	7388124	0.91	11.45	5.80
fHa-8-10	10153328	9225750	0.91	14.82	4.23
fHa-8-11	6818360	6228484	0.91	9.82	5.94
fHa-8-12	8387524	7668297	0.91	11.70	6.38
fHa-8-13	11553274	10480966	0.91	16.44	6.09
fHa-8-14	10263186	9318480	0.91	14.84	4.04
fHa-8-15	12330390	11279184	0.91	17.76	4.79
fHa-12-1	8891292	7820106	0.88	12.60	1.54
fHa-12-2	14012552	13429559	0.96	20.64	7.81
fHa-12-3	9880016	8769680	0.89	14.00	2.36
fHa-12-4	7540886	6741625	0.89	10.71	3.86
fHa-12-5	9438600	8569661	0.91	13.69	2.68
fHa-12-6	9329806	8546103	0.92	13.63	3.58
fHa-12-7	12074970	11108591	0.92	17.72	3.61
fHb-9-1	12851618	11824043	0.92	18.43	5.26
fHb-9-2	11992686	10731440	0.89	17.01	5.81
fHb-9-3	12068398	10921082	0.90	17.66	5.01
fHb-9-4	10919854	9696080	0.89	15.39	6.14
fHb-9-5	7792858	7161649	0.92	11.42	4.43
fHb-9-6	10064188	9136390	0.91	14.75	3.54
fHb-9-7	10222244	9314117	0.91	14.55	4.14
fHb-9-9	15233392	14029408	0.92	22.16	6.35
fHb-9-10	13957646	13122691	0.94	20.93	5.89
fHb-9-11	7516700	7000386	0.93	11.16	5.01
fHb-9-12	12055890	11104346	0.92	17.98	3.97

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
fHb-9-13	15153690	14056680	0.93	22.30	4.69
fHb-9-14	10037382	9563676	0.95	14.80	7.39
fHb-9-15	11715822	11047170	0.94	17.52	6.14
fHb-12-1	10752160	9541974	0.89	15.20	6.58
fHb-12-2	9004952	8170595	0.91	13.28	4.56
fHb-12-3	7700814	6931548	0.90	11.16	3.50
fHb-12-4	8691310	7902576	0.91	12.71	3.49
fHb-12-5	14323554	13495541	0.94	21.14	6.42
fHb-12-6	9983228	9084365	0.91	14.76	3.77
fHb-12-7	10829526	9924244	0.92	15.97	4.22
mHa-12-1	9105544	8476364	0.93	13.74	3.57
mHa-12-2	7666634	6875029	0.90	10.95	3.10
mHa-12-3	8452914	7715784	0.91	12.74	3.05
mHa-12-4	8419926	7304640	0.87	11.97	2.69
mHa-12-5	10221528	9571681	0.94	15.21	5.03
mHa-12-6	11736642	10737063	0.91	17.46	2.53
mHa-12-7	10485306	9251997	0.88	14.45	3.75
mHa-12-9	13953314	12759683	0.91	20.23	4.72
mHa-12-10	8977252	8174178	0.91	12.84	3.76
mHa-12-11	14105994	13009089	0.92	20.27	5.63
mHa-12-12	13546128	12402535	0.92	19.64	4.83
mHa-12-13	10122034	124027	0.01	0.02	56.45
mHa-12-14	3802096	3477057	0.91	5.46	5.62
mHa-12-15	14782304	13852168	0.94	21.57	5.46
mHa-15-1	9726388	8968388	0.92	14.34	4.29
mHa-15-2	9388002	8402530	0.90	13.78	2.31
mHa-15-3	9286284	8300454	0.89	13.24	4.00
mHa-15-4	12528890	11550468	0.92	18.81	3.27
mHa-15-5	11166960	9948730	0.89	15.91	3.07
mHa-15-6	12672404	11383114	0.90	18.33	3.27
mHa-15-7	9693128	8645552	0.89	13.57	4.77
mHb-1-1	8114520	7465306	0.92	11.77	5.27
mHb-1-2	14336776	13259603	0.92	21.01	4.88

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
mHb-1-3	14718166	13481137	0.92	21.20	4.32
mHb-1-4	11223456	10274509	0.92	16.60	3.99
mHb-1-5	8025202	7437361	0.97	11.31	6.87
mHb-1-6	9630300	9031776	0.94	14.01	7.53
mHb-1-7	9829098	9072311	0.92	14.00	7.88
mHb-1-8	13360148	12712968	0.95	18.77	9.50
mHb-6-1	12422376	11437488	0.92	18.09	4.61
mHb-6-2	13390628	12131412	0.91	19.26	4.95
mHb-6-3	12445726	11224503	0.90	17.75	5.29
mHb-6-4	11554074	10351312	0.90	15.95	6.22
mHb-6-5	11611118	10549315	0.91	16.52	5.31
mHb-6-6	11097332	10010199	0.90	16.03	3.84
mHb-6-7	12224818	11115758	0.91	17.27	4.27
mHb-6-8	27260924	25857790	0.95	37.39	10.21
mHb-16-1	12291164	10976440	0.89	17.31	4.38
mHb-16-2	8123694	7538581	0.93	11.61	6.16
mHb-16-3	13394368	12534874	0.94	19.61	6.40
mHb-16-4	10532728	9620074	0.91	14.98	6.79
mHb-16-5	12708642	11513737	0.91	18.24	4.37
mHb-16-6	14335682	13073249	0.91	20.69	4.30
mHb-16-7	9701042	8921402	0.92	13.96	5.43
mHb-16-8	14961928	14281271	0.95	19.79	14.24
NTC1	22949778	19645401	0.86	27.80	4.72

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
Bd4-1	21361536	19847361	0.93	26.83	6.11
Bd4-2	25890506	23908106	0.92	33.06	4.80
Bd4-3	23424970	21642410	0.92	29.68	4.79
BmA135-1	33696710	32164714	0.95	44.54	4.62
BmA135-2	37515440	35137514	0.94	47.66	3.68
BmA135-3	24689152	23377731	0.95	32.59	3.95
BmA135-4	29401794	27841351	0.95	38.48	4.46
BmA-4-1	1.04E+08	12115776	0.12	13.91	45.24
BmA-4-2	90129316	14379318	0.16	16.62	30.90
BmA-5-1	34989240	33374534	0.95	42.06	8.70
BmA-5-2	34039348	31920821	0.94	34.60	7.32
BmA87-1	59014386	18044762	0.31	16.24	9.95
BmA87-2	19213960	16193369	0.84	18.85	19.57
BmA87-3	44213334	26922065	0.61	10.96	45.53
BmE109-1	27428786	18590035	0.68	19.16	20.06
BmE109-2	29717928	17263301	0.58	17.54	25.76
BmE109-3	26166278	17846707	0.68	17.57	17.60
BmE139-1	19457306	10474528	0.54	2.99	40.49
BmE139-2	41409594	26066123	0.63	13.32	13.90
BmE139-3	23097166	13127924	0.57	3.62	34.04
BmE143-1	24368580	8782813	0.36	2.83	48.44
BmE143-2	42105018	22920730	0.54	15.00	36.95
BmE143-3	35079788	17718966	0.51	15.47	35.39
BmE145-1	23438934	17998353	0.77	20.16	5.95
BmE145-2	34445192	26063048	0.76	29.92	4.37
BmE146-1	36638448	28085973	0.77	32.22	4.84
BmE146-2	23451630	17317111	0.74	19.33	4.41
BmE146-3	22226034	16303722	0.73	18.01	4.34
BmE146-4	12513958	9967327	0.80	10.91	8.55
BmE146-5	26332566	20919848	0.79	21.06	14.66
BmE147-1	28169642	20769820	0.74	22.88	4.55
BmE147-2	21499142	16106265	0.75	18.22	4.72
BmE147-3	22345832	16468183	0.74	18.22	5.09

Appendix 2. (continued)

Sample ID	Number of total reads	Number of mapped reads	% of mapped reads	Coverage	% of contamination
DG1-1	27160670	25780038	0.95	33.45	7.93
DG1-2	12894578	10102622	0.78	14.59	4.28
DG1-3	14532056	9878310	0.68	14.00	3.16
DG1-4	22411234	29002000	1.29	40.72	5.12
GP-1	55652652	41406321	0.74	43.46	7.50
GP-2	29942110	23075308	0.77	26.00	6.50
GP-3	21393634	15950883	0.75	17.25	6.61
GP-4	44974362	32322094	0.72	35.25	5.37
Bd-1	36471548	7295349	0.20	2.78	6.18
Bd-2	25635652	4946737	0.19	2.02	5.69
Bd-3	37981302	8047670	0.21	3.13	8.25
Bd-4	27669648	5851225	0.21	2.26	6.63
Bd-5	26249152	5241649	0.20	2.02	8.30
Bd-6	24953370	4715789	0.19	1.90	6.85