

# SCIENTIFIC REPORTS

OPEN

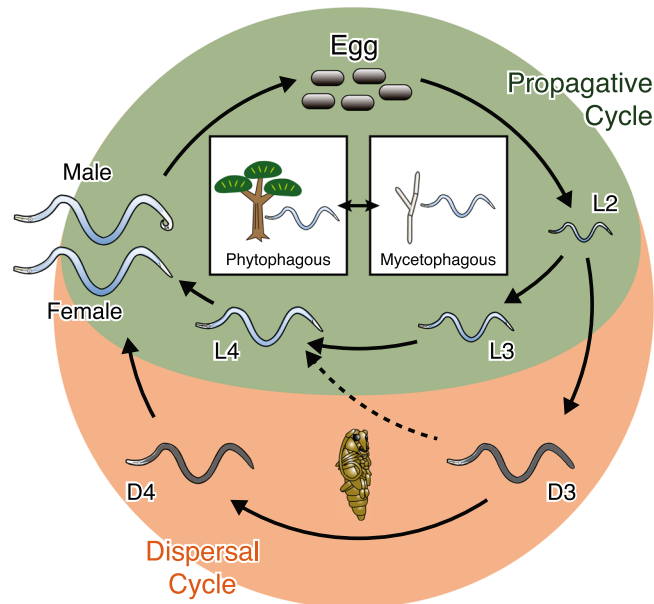
## Stage-specific transcriptome of *Bursaphelenchus xylophilus* reveals temporal regulation of effector genes and roles of the dauer-like stages in the lifecycle

Suguru E. Tanaka<sup>1</sup>, Mehmet Dayi<sup>2,3</sup>, Yasunobu Maeda<sup>2</sup>, Isheng J. Tsai<sup>4</sup>, Ryusei Tanaka<sup>2</sup>, Mark Bligh<sup>2</sup>, Yuko Takeuchi-Kaneko<sup>5</sup>, Kenji Fukuda<sup>1</sup>, Natsumi Kanzaki<sup>6</sup> & Taisei Kikuchi<sup>2</sup>

The pine wood nematode *Bursaphelenchus xylophilus* is the causal agent of pine wilt disease, one of the most devastating forest diseases in East Asian and West European countries. The lifecycle of *B. xylophilus* includes four propagative larval stages and gonochoristic adults which are involved in the pathogenicity, and two stages of dispersal larvae involved in the spread of the disease. To elucidate the ecological roles of each developmental stage in the pathogenic life cycle, we performed a comprehensive transcriptome analysis using RNA-seq generated from all developmental stages of *B. xylophilus* and compared transcriptomes between stages. We found more than 9000 genes are differentially expressed in at least one stage of the life cycle including genes involved in general nematode biology such as reproduction and moulting but also effector genes likely to be involved in parasitism. The dispersal-stage transcriptome revealed its analogy to *C. elegans* dauer and the distinct roles of the two larval stages from each other regarding survival and transmission. This study provides important insights and resources to understand *B. xylophilus* parasitic biology.

The pine wood nematode, *Bursaphelenchus xylophilus*, is the causal agent of Pine Wilt Disease (PWD), one of the most serious forest diseases in East Asian and West European countries<sup>1</sup>. This migratory endoparasitic nematode has a complex lifecycle which includes two distinct dietary phases (phytophagous and mycetophagous phases) and two developmental forms (propagative and dispersal forms) (Fig. 1). In the life cycle, the nematodes enter the host pine tree when the vector beetle (*Monochamus* spp.), which has a phoretic association with *B. xylophilus*, undergoes maturation feeding. On entering the tree the nematodes feed on plant epithelial cells around the cortical and xylem tissues<sup>2</sup>, spreading throughout the tree by migrating through the tissues (phytophagous phase), leading to wilting symptoms that result in the death of the tree within a year of infection. After the tree becomes symptomatic, the nematodes start feeding on fungi that invade the tree and multiply into millions (mycetophagous phase). In both phases, *B. xylophilus* develops from an egg to an adult via three propagative larvae (L2, L3 and L4. Nota bene: L1 moults to L2 inside the egg). However, when conditions become harsh, *B. xylophilus* enters into its dispersal phase in which L2 moults to an alternative L3 stage called the third stage dispersal juvenile (D3). Furthermore, in the presence of the vector beetle D3 moults to an alternative fourth stage larva called the fourth stage dispersal juvenile (D4). The D4 nematodes are transmitted by the vector to new healthy host trees to complete the pathogenic life cycle<sup>3,4</sup> (Fig. 1).

<sup>1</sup>Laboratory of Forest Botany, Graduate School of Agricultural and Life Sciences, the University of Tokyo, Tokyo, 113-8657, Japan. <sup>2</sup>Division of Parasitology, Faculty of Medicine, University of Miyazaki, Miyazaki, 889-1692, Japan. <sup>3</sup>Forestry Vocational School, Duzce University, 81620, Duzce, Turkey. <sup>4</sup>Biodiversity Research Center, Academia Sinica, Taipei, Taiwan. <sup>5</sup>Laboratory of Terrestrial Microbial Ecology, Graduate School of Agriculture, Kyoto University, Kyoto, 606-8502, Japan. <sup>6</sup>Kansai Research Center, Forestry and Forest Products Research Institute, Kyoto, 612-0855, Japan. Suguru E. Tanaka and Mehmet Dayi contributed equally. Correspondence and requests for materials should be addressed to T.K. (email: [taisei\\_kikuchi@med.miyazaki-u.ac.jp](mailto:taisei_kikuchi@med.miyazaki-u.ac.jp))



**Figure 1.** Life cycle of *B. xylophilus*. In favourable conditions *B. xylophilus* multiplies via a propagative cycle by feeding on plant cell (phytophagous phase) or fungal hyphae (mycetophagous phase). In unfavourable conditions, the nematode switches to the dispersal cycle in which the second-stage larvae (L2) moult to the dispersal third-stage juvenile (D3) and subsequently to the dispersal fourth-stage juvenile (D4) in a presence of the vector insect, which enters the vector beetle for dispersion.

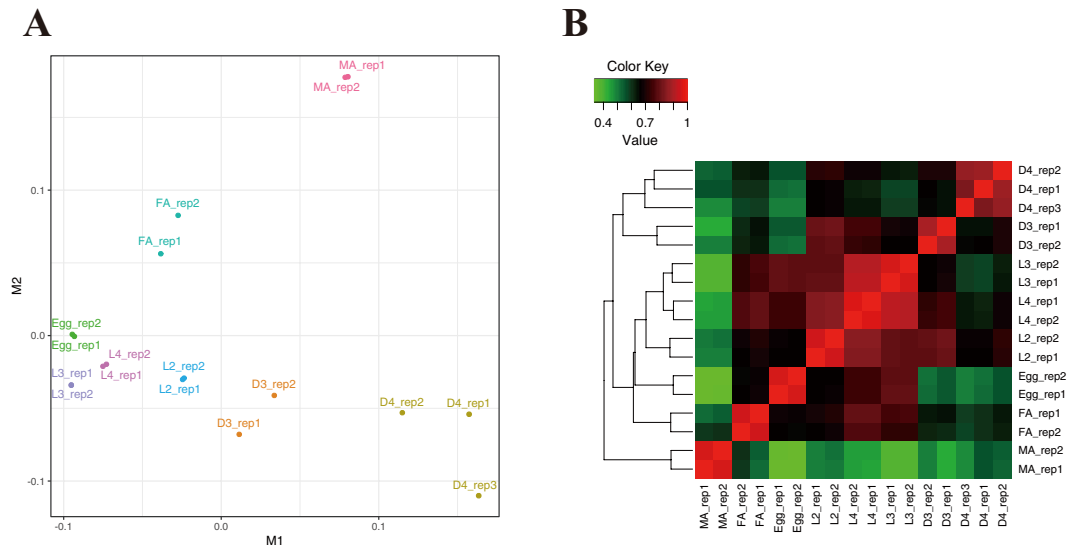
The dispersal stages of *B. xylophilus* are distinguishable by their morphology<sup>5–7</sup>, and thought to be analogous to the *C. elegans* arrested stage (dauer)<sup>8,9</sup>. However, whereas *C. elegans* dauer is an alternative form of its third stage larva, *B. xylophilus* has two developmental stages with dauer-like characteristics (D3 and D4). Interestingly, the later stage (D4) is morphologically more similar to the *C. elegans* third-stage dauer whose mouth is closed and body is covered with thick cuticle<sup>10</sup>, whilst D3 exhibit an active feeding behaviour, but also shows high stress tolerance and extended longevity<sup>11,12</sup>. Environmental conditions to form *B. xylophilus* D3 are similar to those of *C. elegans* dauer including scarcity of food and high nematode density<sup>10–12</sup>, whereas *B. xylophilus* D4 forms only with the vector beetle stimulus<sup>13,14</sup>. Because those stages are essential for the transmission of the pathogen, understanding molecular mechanisms underlying formations and maintenance of those stages are important for developing efficient disease control methods.

In contrast, the propagative cycle of *B. xylophilus* contains stages that are responsible for inducing the typical wilting symptom of the host tree. Like other plant parasitic nematodes, *B. xylophilus* produces and secretes a variety of proteins, called effectors, to manipulate their hosts in order to obtain the nutrients required for development and reproduction. Identified *B. xylophilus* effectors include cell wall degrading enzymes and peptidases for feeding and movement within the tree<sup>15–18</sup>, and anti-oxidant and detoxifying enzymes against host defences<sup>19,20</sup>. The regulation of transcription of effector proteins varies depending on the requirements of the life-cycle stage<sup>21</sup>. Recently two independent studies using RNA deep sequencing (RNA-seq) to compare the mycetophagous with phytophagous stages revealed that several effector candidate genes including those encoding peptidases, glycosyl hydrolases and detoxification proteins are differentially regulated during the infection<sup>20,22</sup>, suggesting transcriptional and physiological changes are important for the parasitism of *B. xylophilus*. However, those RNA-seq studies used mixed-stage nematodes and it remains unclear if each stage has any specific roles in the pathogenic life cycle.

In this study, we aimed to obtain comprehensive transcriptional profiles during development of *B. xylophilus* using stage-specific RNA-seq data. To achieve accurate analyses, we produced an improved genome assembly and gene models using newly obtained genome/transcriptome sequencing data<sup>23</sup>. Then, we performed RNA-seq comparisons between developmental stages to understand stage-dependent gene expression pattern. In particular, we generated transcriptome data for the dispersal (dauer-like) stages for the first time as well as the propagative stages and achieved a comprehensive gene expression profile.

## Results

**Reference Genome Improvement.** We generated a new genome assembly using sequencing data from multiple Illumina libraries (see Materials and Methods). The new assembly is 75.9 Mb in length, which is 1.3 Mb larger than the previously published version (v1.2), composed of 501 scaffolds with the largest scaffold of 4.1 Mb (Table S1). Mapping ratio of Illumina short reads generated using *B. xylophilus* genomic DNA to the new reference genome was 99.1% whereas 98.3% reads were mapped to the previous version, suggesting a higher genome coverage by the new assembly. We have predicted gene models on the new assembly using RNA-seq alignments as hints (please see Methods). The new gene set (v2.0) contains 16,346 gene models, which is ~2,000 smaller than the



**Figure 2.** Multi-dimensional scaling plot (A) and heat map of clustered correlation matrix (B) of all RNA-seq samples. L2, L3 and L4: propagative second, third and fourth stage, respectively, MA: male, FA: female, D3 and D4: dispersal third and fourth stage, respectively.

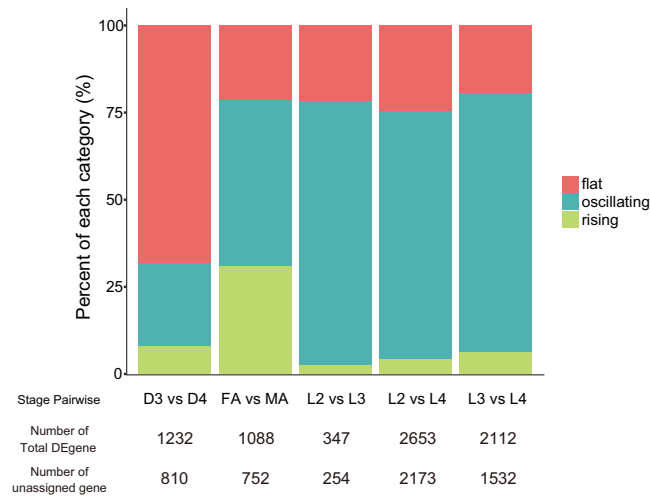
previous version (v1.2) but having a longer average gene length with more exons per gene (Table S1). Gene model comparison revealed that the new gene set has a high number of full-length genes, some of which are truncated or fragmented in the previous version, e.g., a new gene model BXYJ\_1053000.1 (1107 amino acids) corresponds to three old gene models which are located next to each other on a scaffold (BUX\_s00543.22, BUX\_s00543.23 and BUX\_s00543.24) and matched *C. elegans prp-6* gene (968 amino acids) (Fig. S1).

**RNA preparation and sequencing.** RNA was extracted from each nematode developmental stage from the egg to the adult (egg, L2, L3, L4, adult female and adult male). RNA samples were also prepared from the two dauer-like stages (D3 and D4). At least two biological replicates of at least 20 million mRNA-seq reads in each stage was produced (except one D4 sample (D4\_rep1) with ~10 million reads) (Table S2). Aligning the RNA-seq reads to the reference genome exhibited a high mapping ratio for most of the samples (87.3% to 92.9%), suggesting high quality RNA-seq data. D4 samples were the exceptions having relatively low mapping ratios (43.3% to 75.4%) (Table S2) partly due to mRNA contamination from the insects from which the nematodes were isolated, and some nematode rRNA contamination (Table S2). Overall, in each stage, 12301–14699 genes had >5 mapped reads; adult female was the largest (14699 genes having >5 mapped reads) and D4 the smallest (12301 genes having >5 mapped reads) (Fig. S2).

Gene expression values (read count and FPKM (Fragments Per Kilobase of transcript per Million mapped reads)) were calculated for each gene model and used for further comparisons (Table S2). Multidimensional scaling (MDS) analysis of FPKM showed that replicates from the same stages are similar to each other (Fig. 2A). The propagative stages (L2, L3 and L4) and egg stage were positioned closely to each other whereas adult female, adult male, D3 and D4 samples were distantly located from all other samples with relatively high deviations within the replicates (except adult male). Consistently, high correlation was observed between propagative larval stages (L2, L3 and L4) with the highest Pearson's correlation coefficient of 0.92 between L3 and L4 (Fig. 2B). Correlations between the adult male and other stages were relatively low (0.26 to 0.43). Two dispersal stages (D3 and D4) formed a heat map cluster, however D3 showed higher correlation values with L2 than D4 (Fig. 2B). Both in the MDS plot and the heat map, the egg stage is located between the larval stage cluster (L2, L3 and L4) and adult female. This is probably because the gravid adult females contain many eggs inside the body and the egg samples contained a range of developmental stages from the early immediate post-delivery stage to later stages containing L1 larvae.

**Changes in gene expression during development.** Comparing gene expression values between all possible stage pairs identified a total of 9612 genes that were differentially expressed in any pairs with a minimum fold change of 2 and a FDR of <0.001 (Table S3). The number of DE genes was smallest at L3-L4 comparison (347) and largest at egg-male comparison (4327) (Table S3).

Pairwise comparisons between propagative larvae (L2, L3 and L4) showed smaller numbers (347 to 1232) of DE genes than other pairwise combinations (Table S3). Those DE genes were enriched in GO (Gene Ontology) terms of structural constituent of cuticle (GO:0042302), collagen (GO:0005581) and lipid binding (GO:0008289) (Table S4) and contain many orthologues of the *C. elegans* “oscillating genes”, which have oscillatory expressions within a developmental stage and mainly involve moulting functions, constituting 19.9–26.0% of the total DE genes (65.7–73.2% excluding ‘unclassified’) (Fig. 3 and Table S5). In contrast, non-DE genes in the analysis were mainly composed (45.3% of the total, 51.3% excluding ‘unclassified’) of orthologues of *C. elegans* ‘flat genes’



**Figure 3.** Proportions of differentially expressed genes assigned to “flat”, “rising” and “oscillating” genes in pair-wise comparisons. The “flat” category is higher in the two dauer-like stage comparison (D3-D4) and “rising” has higher proportions in adult female and male comparison (FA-MA) than the other pairs, whereas the “oscillating” gene category are dominant in comparisons between larvae (L2-L3, L2-L4 and L3-L4).

(Table S5). Other than oscillating genes, genes having GO terms of proteolysis (GO:0006508) are up-regulated in L3 and L4 compared to L2 (Fig. 3 and Table S4).

Expression patterns of the 9612 DE genes along the development can be grouped into 10 clusters (C1 to C10) using the k-means method (Fig. 4) with cluster sizes ranging from 285 to 1820 genes. These clusters can be further classified roughly into sex-biased (C3, C6 and C10 clusters), dauer-biased (C1, C2, C5, C7 and C8 clusters) and others (C4 and C9 clusters) (Fig. 4).

Cluster C4 genes showed high expression in the egg stage and relatively flat in other stages, representing genes involved in later embryo development (Fig. 4). Enriched GO terms in these genes contain many categories related to cell cycle (GO:0007049), suggesting higher cell division activity in eggs than other stages (Table S6D).

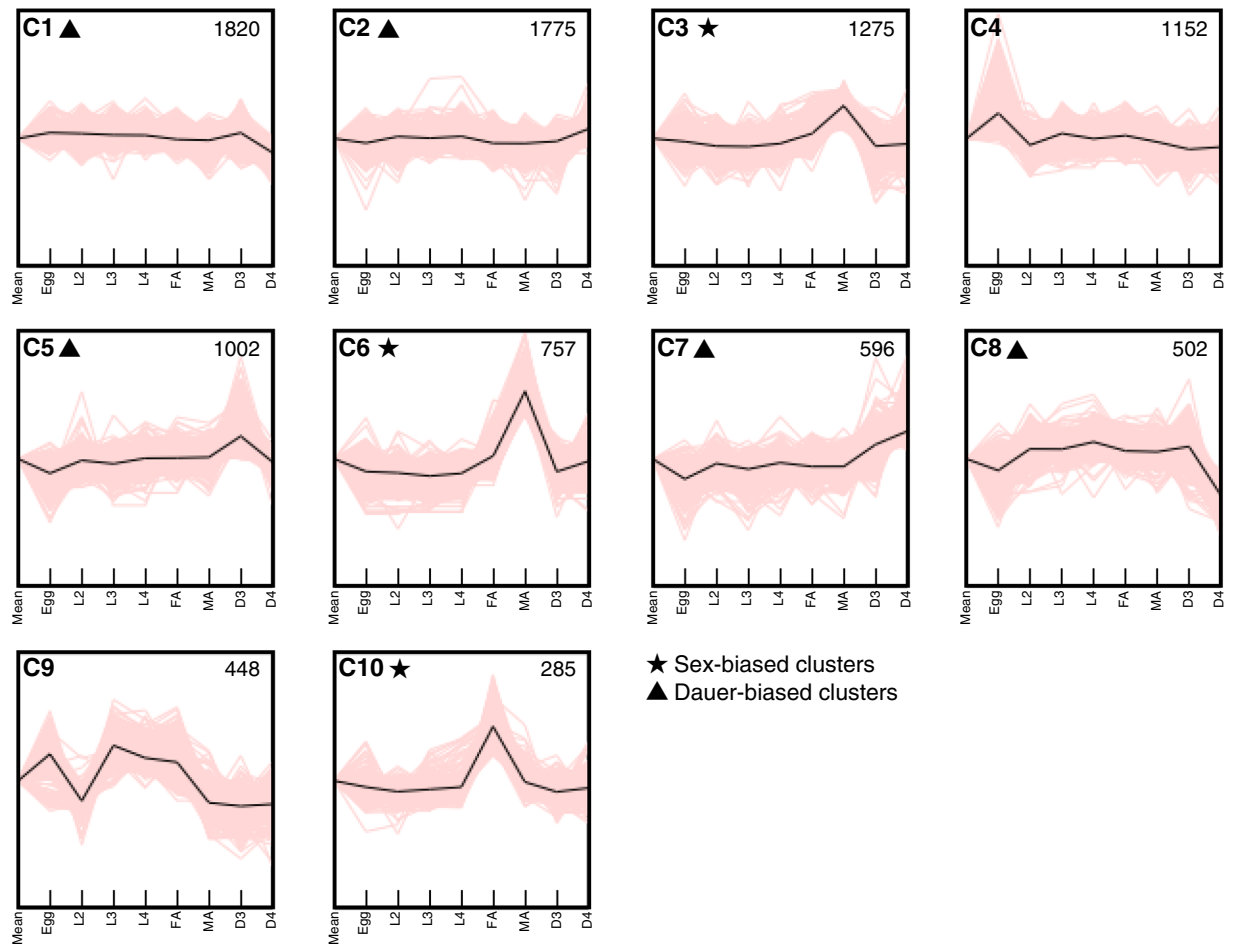
Expression values of cluster C9 genes are high in egg, L3, L4 and adult female, and low in L2, D3, D4 and adult male stages (Fig. 4). This cluster is enriched with GO terms related to the moulting cycle (GO:0042303). In addition, 166 out of 448 genes in this cluster are orthologues of *C. elegans* ‘oscillating’ genes which show pervasive and phase-locked oscillations of gene expression in propagative larvae, suggesting the cluster C9 genes are changeable depending on developmental stages (Table S6I).

**Sex-biased and germline-enriched transcriptome.** Two clusters (C3 and C6) exhibit up-regulated gene expression in the male stage and contained 1275 and 757 genes, respectively. The C6 cluster genes showed a sharper peak in the male stage than C3 genes but both are enriched in similar GO terms including protein kinase activity (GO:0004672), protein phosphorylation (GO:0006468), dephosphorylation (GO:0016311), and ATP binding (GO:0005524) (Fig. 4 and Table S6C,F). In contrast, C10 cluster genes showed high expression in the female stage and are enriched in GO terms involved in proteolysis (carboxy (GO:0004180), cysteine-type (GO:0008234) and serine-type (GO:0008236) peptidases) and structural constituent of cuticle (GO:0042302) (Table S6J).

A direct comparison of the adult female and adult male stages identified 2653 DE genes (Table S3). Within those, 968 and 1685 genes are female-biased and male-biased, respectively. We then directly compared adult female and adult male stage gene expression data to elucidate up-regulated expressions in germline and somatic tissues using the strategy described in Choi *et al.*<sup>24</sup>. Using these criteria (please see Methods), a total of 1977 and 1550 genes were identified as germline-enriched and somatic-enriched genes, respectively (Table S7). GO term categories including DNA binding (GO:0003677), regulation of macromolecule biosynthetic process (GO:0010556) and regulation of RNA biosynthetic process (GO:2001141) are highly represented in germline-enriched genes, whereas genes implicated in proteolysis (GO:0070011), peptidase activity (GO:0008233) and extracellular region (GO:0005576) are more frequently found in the somatic-enriched gene set (Table S7).

**Dispersal stage gene expressions.** As contrasted with the dauer stage of *C. elegans*, which is an alternative third stage larva, *B. xylophilus* has two developmental stages with dauer-like characteristics (D3 and D4) (Fig. 1). Heatmap (Fig. 2B) and MDS plot of expression data (Fig. 2A) suggested the two dauer-like stages are transcriptionally very distinct from the propagative stages.

Cluster C1 and C8 genes showed down-regulation at D4 (Fig. 4). C1 contained a large number of genes (1820 genes) and is enriched in a wide variety of GO terms involved in general biological processes (Table S6A,H), suggesting a low metabolic activity of D4 stage. C8 (502 genes) is enriched in GO terms of proteolysis, especially in aspartic-type (GO:0004190 and GO:0070001) and cysteine-type (GO:0008234) peptidases, which are expanded in the *B. xylophilus* genome and involved in feeding behaviour of the nematode<sup>19,23</sup>, reflecting the D4



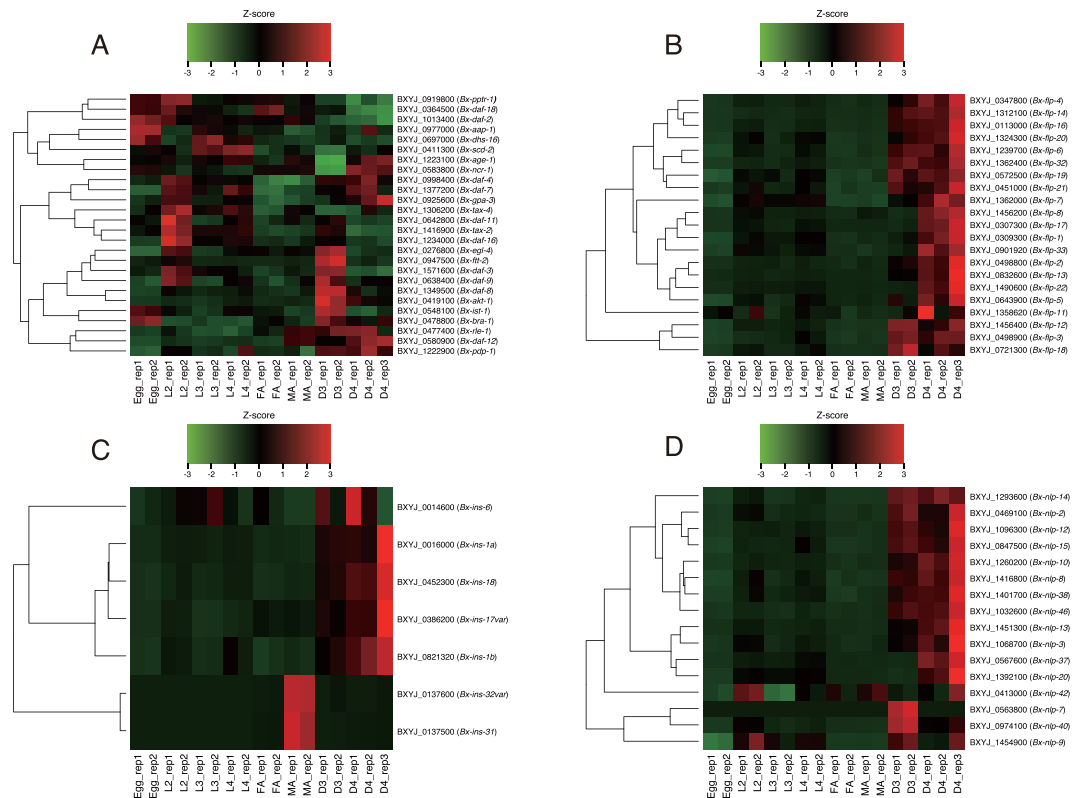
**Figure 4.** Clustering analysis of genes differentially expressed between developmental stages. A total of 9612 genes are clustered into 10 (C1 to C10) based on expression dynamics by K-means clustering. The upper-left alphanumeric values designate the number of the cluster. Sex-biased and dauer-biased clusters were indicated by stars and triangles, respectively. The upper-right values in each cluster shows cluster size (number of genes). Black lines indicate representative transcriptional expressions.

characteristics of the covered mouth and lack of feeding behaviour. This corresponds to the observation of low expression of C8 genes in the egg stage (Fig. 4).

In contrast, the C2 cluster (1775 genes) is up-regulated at D4 and enriched in signal transduction-related GO terms including potassium channel activity (GO:0015079), neuropeptide receptor activity (GO:0008188), G-protein coupled receptor activity (GO:0004930), and integral component of membrane (GO:0016021) (Table S6B). The C7 cluster (596 genes), up-regulated at both D3 and D4, is enriched in “neuropeptide signalling pathway” GO term (GO:0007218) (Table S6G). The C5 cluster showed higher expression at D3 than D4 (Fig. 4). GO enrichment analysis revealed C5 genes are enriched in the categories involved in peptide metabolic processes (GO:0006518) with more anabolic terms including biosynthetic process (GO:0009058), peptide biosynthetic process (GO:0043043) and amide biosynthetic process (GO:0043604) (Table S6E). This suggests D3 is a stage that readjusts its metabolism to accumulate internal energy sources. Additionally, a high representation of the term oxidoreductase activity (GO:0016491) in the C5 cluster may indicate a higher oxidative stress tolerance in D3 than other stages.

Pairwise comparison of D3 with L2, L3 and L4 identified 1036, 1792 and 1472 genes differentially expressed, respectively (Table S3). Genes up-regulated in D3 are enriched in stress response-related GO terms, including oxidoreductase activity (GO:0016491) and response to stress (GO:0006950) as well as proteolysis (GO:0006508) (Table S4). Interestingly many genes with cellulase activity (GO:0008810) are up-regulated in D3 relative to L2. Highly presented GO terms in genes down-regulated in D3 include structural constituent of cuticle (GO:0042302), proteolysis activity (GO:0006508) and lipid binding (GO:0008289) (Table S4).

Pairwise comparisons of D4 with the propagative stages identified larger numbers of DE genes than the D3 comparisons (2567, 3121 and 2618 related to L2, L3 and L4 respectively) (Table S3). Among them, numbers of D4 up-regulated genes are 1170, 1297 and 933 and GO terms related with protein kinase activity (GO:0004672), and transmembrane signalling receptor activity (GO:0004888) was significantly enriched in D4 relative to L4. GO terms enriched in genes down-regulated in D4 include structural constituent of cuticle (GO:0042302), proteolysis



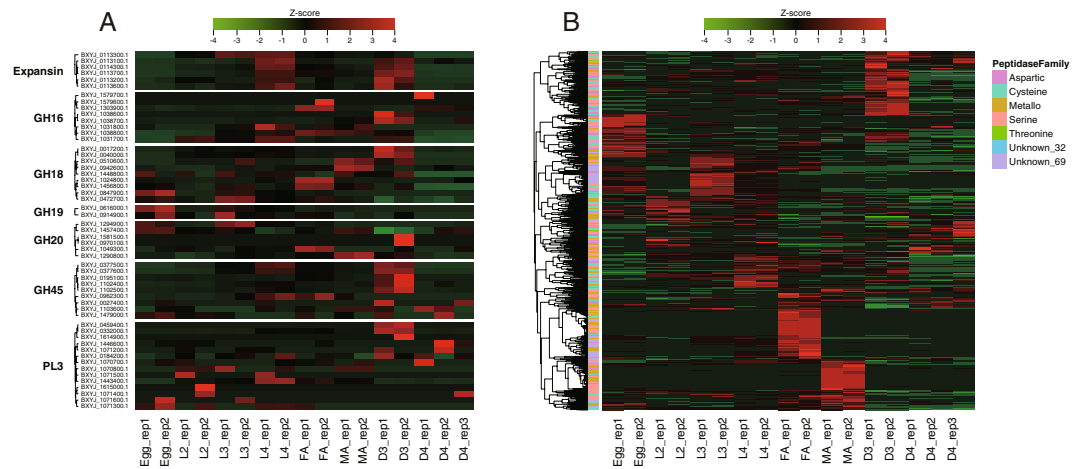
**Figure 5.** Gene expression patterns of dauer-related genes and neuropeptide genes. Heatmaps represent expression dynamics of *B. xylophilus* orthologues of (A) *C. elegans* dauer genes and neuropeptide (B) flp, (C) ins and (D) nlp genes. Rows represent genes “*B. xylophilus* gene ID (*C. elegans* gene name or *B. xylophilus* gene name)” and columns represent RNA-seq samples. The intensity of each colour denotes the standardized ratio between each value and the average expression of each gene across all samples. Genes showing null FPKM in all stages were removed from the plot.

activity (GO:0006508) and lipid binding (GO:0008289). In particular, the GO term of aspartic-type endopeptidase activity (GO:0004190) enriched as a down-regulated function in all the three pairwise comparisons, as well as down-regulated cellulase activity (GO:0008810) relative to L4, suggests again the low feeding activity of D4 nematodes.

A direct comparison between D3 and D4 identified 1081 genes and 1033 genes up-regulated in D3 and D4, respectively (Table S4). Genes highly expressed in D3 are enriched in GO terms of proteolysis (aspartic-type endopeptidase activity (GO:0004190) and cysteine-type peptidase activity (GO:0008234)) and oxidation-reduction process (GO:0055114), whereas genes highly expressed in D4 are enriched in G-protein coupled receptor signaling pathway (GO:0007186). These results indicate substantial biological differences between the two dauer-like stages; D3 has high feeding/metabolic activity and higher stress-tolerances whereas D4 has low metabolic activity but increased and/or distinct signal pathway activity.

**Expression of genes involved in dauer pathways.** To see whether nematode dauer formation/maintenance pathways are conserved in *B. xylophilus*, we compared gene expressions of *B. xylophilus* orthologues of *C. elegans* dauer genes. Four signalling pathways have major roles in dauer formation of *C. elegans*: cGMP, IIS, TGF- $\beta$ , and DA/DAF-12 pathways<sup>25–28</sup>. *B. xylophilus* has most orthologues of *C. elegans* genes involved in these four pathways<sup>23,29</sup>.

Many *B. xylophilus* orthologues of *C. elegans* dauer genes that are up-regulated in the dauer relative to other developmental stages (Fig. S3) also exhibit high expression values in D3 and/or D4 stages (Fig. 5). Four major clusters were observed in the heatmap showing the expression pattern of the *B. xylophilus* orthologues (Fig. 5). Most orthologues of genes up-regulated in *C. elegans* dauer are included in two clusters showing high expression both in D3 and D4, particularly D3. The third cluster contained two orthologues showing high expression mainly in D4, and both *C. elegans* genes (*ncr-1* and *age-1*) showed high expression in the dauer. Orthologues in the fourth clusters have high expressions in the egg and/or L2 and relatively low in D3 and D4. Those orthologues in *C. elegans* mostly show a similar pattern (i.e. high in the egg and/or L1, and relatively low in the dauer). These results suggest that nematode dauer pathways are largely conserved between *C. elegans* and *B. xylophilus*, and *B. xylophilus* D3 and D4 use different sets of *C. elegans* dauer genes from each other to form and maintain their special developmental stages. *Bx-daf-7* (BXJ\_1377200), a single-copy orthologue of *C. elegans daf-7* which encodes a key ligand of the TGF- $\beta$  signaling pathway<sup>30</sup>, showed a similar expression pattern to *C. elegans daf-7*, showing high expressions in the early larval stage and the dauer stages, suggesting a conserved role in the two species although diverged



**Figure 6.** Expression dynamics of *B. xylophilus* effector genes along the development. **(A)** Heatmap of 54 genes encoding cell-wall degrading enzymes (expansin, GH45 cellulases, PL3 polysaccharide lyases, GH16 beta-1,3-glucanases, and chitin-degrading enzymes (GH18, GH19, and GH20)). **(B)** Heatmap of 771 genes encoding peptidases belonging to aspartic (91 genes), cysteine (139 genes), metallo (223 genes), serine (171 genes), threonine (10 genes), and unknown\_69 (134 genes) families. Genes showing null FPKM in all stages were removed from the heatmap.

functions have been suggested in other nematodes<sup>31</sup>. There are, however, some exceptions which show a different expression pattern between *C. elegans* and *B. xylophilus*, including *tax-4*, *rle-1*, *daf-8* and *daf-16*, indicating that unique gene regulations are also present in *B. xylophilus*. Notably there are seven genes showing a null FPKM value in all stages of *B. xylophilus* (not shown in the heatmap), including BXYJ\_0789700, BXYJ\_1393800, and BXYJ\_1599300, whose *C. elegans* homologues, *daf-38* and *pdk-1*, are active in the dauer stage.

**Expression of neuropeptide genes.** Neuropeptides are short peptides acting as transmitters and neuromodulators, which can control the activity of neurons and modulate behaviour in specific way<sup>32</sup>. The *C. elegans* genome encodes 31 FMRFamide-like peptides (*flp* genes), 40 insulin-related peptides (*ins* genes), and 47 neuropeptide-like proteins (*nlp* genes), and they, especially *flp* genes, are significantly up-regulated in the dauer entry<sup>33</sup>. We identified 22 *flp*, 7 *ins* and 17 *nlp* genes in the *B. xylophilus* genome<sup>23</sup>. Among them 21 *flp*, 7 *ins*, and 16 *nlp* genes were expressed in at least one stage. Interestingly, we found almost all *flp* and *nlp* genes were up-regulated in D3 and/or D4 (Fig. 5B,D). An exception is two *nlp* genes (*Bx-nlp-9* and *-42*) showing high expression in developmental stages other than D3 or D4 stage. Most of *flp* and *nlp* genes showed higher expression levels in D4 than in D3, whereas *Bx-nlp-7*, *nlp-40*, *flp-3*, *flp-12* and *flp-18* showed higher expressions in D3 than in D4. The two *ins* genes (*Bx-ins-31* and *-32var*) were mainly up-regulated in adult male and the other five were up-regulated in D3 and/or D4 (Fig. 5C). In *C. elegans*, *ins-17* and *ins-18* likely work as antagonists of Daf-2 pathway, where over-expression of those *ins* genes leads to dauer arrest<sup>34,35</sup>. We found *Bx-ins-17* and *-18*, which show high expression in dauer-like stages, shares conserved cysteine bond structure with *C. elegans* though *Bx-ins-18* does not have the conserved PPG motif<sup>36</sup>, indicating conserved roles of those *ins* genes and the pathway in the two species. These results altogether suggest that those neuropeptides are tightly incorporated in the dauer biology of *B. xylophilus* as is suggested in other parasitic nematodes<sup>33</sup>.

**Temporal regulation of effector genes.** Plant parasitic nematodes secrete molecules affecting the host tissues called effectors<sup>37–39</sup>. Cell-wall degrading enzymes (CWDEs) are secreted by *B. xylophilus* and used to feed and migrate within the host tissues<sup>23</sup>. Expression patterns of 62 genes encoding CWDEs in *B. xylophilus* are shown in Fig. 6A, including those working on plant cell wall (GH45 cellulases, PL3 pectate lyases and expansins) and those on fungal cell wall (GH16 beta-1,3-glucanases and chitin degrading enzymes (GH18, GH19, and GH20)). Many genes both for plant and fungal CWDEs showed high expression in D3, suggesting the highly active feeding behaviour and metabolism of the dauer-like stage (please note that a sealed mouth, a typical morphological character of *C. elegans* dauers, is seen only in D4, not in D3). In particular, all highly-expressed GH45 cellulases (FPKM > 20 in any sample) showed the highest expression in D3. Additionally, D3 expansin expression is highest among the stages. It is also noteworthy that expansins and GH45 cellulases showed high FPKM values in L4 (Fig. 6A). In pectate lyases (PL3), there are three genes showing high FPKM values. Two of them (BXYJ\_0459400.1 and BXYJ\_0332000.1) are specifically upregulated in D3. The other PL3 gene (BXYJ\_1443400.1) appears constantly expressed in all stages except eggs with the highest FPKM value in L4. In chitin degrading enzymes, GH19 genes showed high expression in eggs. Two other families (GH18 and GH20) also have a few genes that show high expression in egg, suggesting chitinase roles in egg formation or hatching. The other GH18 genes generally showed constant expression across the stages (except D4) with upregulations at D3, adult female and adult male (two genes each) (Fig. 6A). Five genes of GH16 beta-1,3-glucanases, which are involved in fungal cell wall degradation, have high expression levels at L4, adult female, and D3. Among them, two genes (BXYJ\_1038600 and BXYJ\_1038700) are specifically upregulated in D3 whereas the other three FPKM

values are relatively constant across the stages. The D4 stage showed very low expression levels for all the enzymes with only a small number of exceptions including genes encoding chitinases, a PL3 and an expansin.

Nematode peptidases, which hydrolyse polypeptides or proteins, participate in a wide range of molecular, biological and cellular processes such as digestion of host proteins, moulting and embryonic development of the egg<sup>40</sup>. The *B. xylophilus* genome contains 808 peptidase genes, representing the highest gene number among characterised nematode genomes (Table S8)<sup>23</sup>, which are composed of aspartic (106 genes), metallo (230 genes), cysteine (142 genes), serine (170 genes), threonine (13 genes), unknown\_32 (8 genes) and unknown\_69 (136 genes). Gene expression pattern of the peptidases is visualised in Figs 6B and S4. Intriguingly, most peptidases specifically upregulate at only one developmental stage comprising clear stage-specific clusters in the heatmap. The clusters have a similar size to each other except L2- and D4-specific clusters which show fewer genes. Each stage-specific cluster contains various types of peptidase (i.e. aspartic, cysteine, metallo, serine, threonine peptidase, unknown\_32 and unknown\_69 peptidases). The number of genes expressed in each life stage of *B. xylophilus* varies depending on peptidase type. The female-specific cluster has more highly expressed genes of unknown\_69 type and the D3 cluster has more highly expressed genes of aspartic, cysteine, serine and threonine types. In contrast, the number of highly expressed genes of metallo and unknown\_32 type peptidase families is increased in L3- and male-specific clusters (Figs 6B and S4).

## Discussion

In the present study, we performed a comprehensive stage-specific transcriptome analysis of *B. xylophilus* and identified genes differentially expressed between developmental stages.

*B. xylophilus* is a migratory endoparasite which multiplies within the host via multiple generations unlike other major plant parasitic nematodes, including root-knot and cyst nematodes, which are sedentary parasites whose developmental stages have clear stage-specific roles in their lifecycles<sup>3,41</sup>. This indicates that all propagative stages of *B. xylophilus* share the same niche in the host and therefore, previous studies paid little attention to stage differences within the propagative life cycle. When investigating the pathogenicity of *B. xylophilus*, many studies used mixed-stage nematodes<sup>2,11,42–44</sup> though differences between the dispersal and propagative stages have been more extensively studied<sup>7,45–47</sup>. Recently, two independent transcriptome studies investigated gene expression of *B. xylophilus* in the host plants and identified ‘effectors’ and genes important for parasitism<sup>20,22</sup>. However, those studies also used mixed-stage nematodes and did not reveal roles in the pathogenic life cycle or particular biological characteristics of each developmental stage.

In this study, we found that many genes are differentially regulated in *B. xylophilus* even between the propagative stages. For example, genes involved in germline development in *C. elegans* were up-regulated in the adult stages of *B. xylophilus*: genes with GO terms of protein kinase and phosphatase in males, and those with GO terms of oocyte maturation in females. Genes involved in DNA replication and pharynx development were up-regulated in eggs. In addition, genes involved in moulting such as collagen genes are likely to have oscillating expression patterns within each larval stage as seen in *C. elegans*<sup>48</sup>. These results suggest that roles and regulation of genes involved in basic biological processes are well conserved between *B. xylophilus* and *C. elegans* though the two species belong to different evolutionary clades from each other<sup>21</sup>.

Intriguingly we also found that genes encoding effectors involved in parasitism are differentially expressed between stages. In particular, genes encoding CWDEs were highly expressed in D3 and L4 relative to other propagative stages indicating their highly active feeding status. This may reflect a high demand for nutrients by D3 to facilitate lipid granule accumulation in the body for long-term survival, or by L4 to develop the germlines. It also indicates those stages may have particular roles in the pathogenesis of the nematode by destroying plant tissues.

Genes encoding peptidases, which are highly expanded in the *B. xylophilus* genome<sup>23</sup>, were also differentially expressed between stages. Within a peptidase family, *B. xylophilus* is likely to use different sets of genes from the multiple gene copies depending on the stage. It is known that peptidase families have a diverse range of biological roles such as moulting, development, food digestion, and parasitism in nematodes. Our results suggest that each gene in an expanded peptidase family has distinct roles or distinct regulation in *B. xylophilus*, representing a clear example of gene family evolution by gene duplication and functional divergence. Stage-specific regulation of genes in an expanded gene family were also observed in other parasitic nematodes such as SCP/TAPS gene families in animal parasitic *Strongyloides* species<sup>49</sup> and *Necator* species<sup>50</sup>, astacin-like metallopeptidase in *Strongyloides* species<sup>49</sup>, the cysteine peptidase gene family in *Schistosoma mansoni*<sup>51</sup>, ShKT-domain containing proteins in *Teladorsagia circumcincta*<sup>52,53</sup>, and metallopeptidases in *Globodera pallida*<sup>54</sup>.

Another example of genome evolution via gene duplication and functional divergence are the chitinase gene families. *B. xylophilus* has a higher number of GH18 chitinase genes than other obligate plant parasites<sup>23</sup> and those genes are likely to have different roles in *B. xylophilus*. Expression of nine GH18 chitinase genes in *B. xylophilus* showed variable patterns (Fig. 6A). This result is mostly consistent with a previous study<sup>55</sup>, where they cloned seven chitinase genes in *B. xylophilus* designated *Bx-Chi-1* to *Bx-Chi-7* and performed functional analysis. They grouped *B. xylophilus* chitinase genes into four types: egg-specific involved in egg hatching (*Bx-chi-7*), female spermatheca-specific involved in reproduction in females (*Bx-chi-1*), male-specific involved in sperm metabolism (*Bx-chi-3, 4, 5* and *6*), and pharyngeal gland-specific involved in fungal feeding (*Bx-chi-2*). In this study, we found that two genes including BXYJ\_1024800.1 and BXYJ\_1456800.1 were highly expressed in the adult female stage and one gene (BXYJ\_0847900.1) was observed to be highly expressed in eggs. Three chitinase genes including *Bx-chi-3* and *4* (BXYJ\_1448800.1 and BXYJ\_0510600.1) were mainly up-regulated in adult males. Two genes, including *Bx-chi-2* (BXYJ\_0040000.1), were highly expressed in larval stages with the highest expression at D3, likely involved in fungal feeding. These results support the idea that each of multiple GH18 chitinases in the genome has a distinct role in its lifecycle<sup>55</sup>, from general nematode biology including reproduction and development to a specific fungal feeding character of *B. xylophilus*.



*C. elegans* dauer is a specialised stage favouring survival and dispersal whose characteristics include tolerance to starvation and high stress<sup>56</sup>. Parasitic nematodes also have the same type of arrested stage, called “infective juveniles (or infective larvae)” which are normally an alternative 3<sup>rd</sup> stage larva and can survive harsh conditions while searching for a new host<sup>57</sup>. Unlike many other nematodes, *B. xylophilus* possesses two developmental stages with dauer-like characteristics in its life cycle (D3 and D4). D3 emerges in the host tree when the host is dying (harsh conditions). D4 larvae develop from D3 and are dispersed by the vector beetle. We found that the two dauer-like stages have very different gene expressions not only from other propagative stages, but also between the two stages. Many CWDEs and peptidases involved in “food degradation” were highly up-regulated in D3 relative to the propagative stages, whereas D4 showed a low metabolic-state gene expression pattern like the *C. elegans* dauer<sup>58</sup>. D3 doesn't have a typical morphological character of dauers - a sealed buccal opening, indicating D3's active feeding state, whereas D4 has a sealed mouth and degenerate digestive systems, showing no-feeding behavior<sup>7</sup>. In this sense, D3 is more like *C. elegans* pre-dauer (or L2d) which is the alternative 2<sup>nd</sup> stage larva *C. elegans* forms under harsh conditions before progressing onto dauer arrest<sup>59</sup>. However, the stress tolerance and longevity of D3 are similar to those of the *C. elegans* dauer. Genes involved in stress responses, including those encoding oxidoreductase and deubiquitinating enzymes, were more active in D3 than D4, which reflects the natural life cycle where D3 survives in the host tree after emergence from autumn till spring, suggesting its long life, which is another typical character of the *C. elegans* dauer and infective larvae of other parasitic nematodes. In contrast, *B. xylophilus* D4 occurs only at the presence of the vector beetle and is likely specialized for dispersion. Genes involved in signalling pathways were up-regulated in D4. This is probably because D4 needs more receptors and signal mediators to successfully recognize the insect vector. These results suggest that *C. elegans* dauer abilities and roles are apportioned between the two *B. xylophilus* dauer-like stages to achieve the unique *B. xylophilus* life cycle.

The possession of two dauer-like stages of *B. xylophilus* is uncommon in nematodes. Even within the genus, only a few species have this two-step dauer formation<sup>60</sup>. This atypical dauer formation seem adaptive to its life history, *i.e.*, the worms propagate in the dead/dying wood, and transmitted by longhorn beetle species. The vector longhorn beetles, *Monochamus* spp. (and their relatives associated with other *Bursaphelenchus* spp.) are univoltine<sup>61</sup>, and can carry the nematodes only in early summer season, *i.e.*, *B. xylophilus* which is highly specialized to its vector insects has to survive from autumn to next early summer under harsh conditions.

Several signalling pathways including cGMP<sup>62</sup>, insulin-like<sup>26</sup>, TGF- $\beta$ <sup>25</sup> and DA/DAF-12<sup>27</sup> control *C. elegans* dauer formation. We found that orthologues of genes involved in those pathways as well as genes encoding neuropeptides were also present in *B. xylophilus*. Among them, genes upregulated in *C. elegans* dauer showed high expression values in D3 and/or D4 in *B. xylophilus*. In particular, *B. xylophilus* orthologue of *daf-12* (BXYJ\_0580900), a key regulator of DA/DAF-12 pathways in *C. elegans*<sup>25</sup>, is up-regulated in D4 of *B. xylophilus*, suggesting the pathways are conserved in *B. xylophilus* as in some other nematodes<sup>63–65</sup>. Orthologues of *egl-4* (BXYJ\_0276800) in cGMP pathway associated with receiving chemical signals, and *ist-1* (BXYJ\_0518100 and BXYJ\_0548100) and *akt-1* (BXYJ\_0419100) in IIS pathway associated with transmission of signals to downstream pathways showed high expression pattern in D3 as in the *C. elegans* dauer, but not in D4. This again indicated that D3 and D4 of *B. xylophilus* are similar to *C. elegans* dauer but have distinct characteristics from each other. However, we should also note that post-translational regulations should play important roles in those pathways and may be different between species. It may be noteworthy that expression of some dauer pathway orthologues including orthologue of *daf-38* were not detected in any stages of *B. xylophilus* in this study. *C. elegans* *daf-38* encodes a G-protein-coupled receptor which binds an ascaroside, dauer-inducing pheromone<sup>66</sup>, indicating diverged environmental signal receptions between these two species although ascarosides are an evolutionarily conserved family of nematode pheromones and ascaroside-like substances with heat-tolerant and water-soluble characteristics seem to be used by *B. xylophilus* to induce D3 formation<sup>12</sup>.

FMRFamide-like peptide genes (*flp*) are specifically upregulated during dauer entry in *C. elegans* and are shown to be involved in the dauer specific nictation behaviour and CO<sub>2</sub> attraction<sup>33</sup>. This seems to be conserved in many parasitic nematodes, whose infective juveniles show similar behaviours<sup>33</sup>. We observed upregulation of *flp* genes in D3 and D4 of *B. xylophilus*. This is interesting because strong nictation behaviour was seen only in D4 in *B. xylophilus* though both D3 and D4 are attracted by CO<sub>2</sub><sup>4</sup>. The *flp* genes that were differentially expressed between D3 and D4 may be responsible to those differences.

Finally, the results and data from this study provide a foundation for further investigation of *B. xylophilus* biology and the development of novel pathogen control strategies. *B. xylophilus* transcriptome data and unique life cycle also provide a unique opportunity to perform comparative studies on many conserved and not-conserved processes for which the free-living *C. elegans* remains the pre-eminent model.

## Methods

**Genome reference.** Illumina reads from 600 bp paired-end and 3-kb mate-pair libraries, generated using DNA extracted from the *B. xylophilus* Ka4C1 isogenic line<sup>23</sup>, were assembled using the MaSuRCA assembler 2.2.1<sup>67</sup> with parameters (filter 60X- with cgwErrorRate = 0.2). The resulting assembly was further improved by the Gapfiller<sup>68</sup> and Image<sup>69</sup> to fill gaps in the initial assembly using Illumina pair-end reads. After base corrections using ICORN<sup>70</sup>, haplotypes within the assembly were collapsed using Haplomerger<sup>71</sup>.

For the gene prediction, Augustus (v. 3.0.1)<sup>72</sup> was trained for *B. xylophilus* based on a training set of about 500 confident genes, manually curated in Artemis<sup>73</sup> using aligned RNA-seq data and transfers of the previous version of gene models (v1.2)<sup>23</sup>. RNA-seq alignments to the genome using TopHat v2.0.11<sup>74</sup> were used to produce intron hints. Cufflinks (v2.0.1)<sup>75</sup> was used to assemble transcript fragments, and this information was converted into exon hints. The trained versions of Augustus were run using all the hints for that species as input to generate v2.0 gene models. Functional annotations and Gene Ontology (GO) term assignments were performed using Pfam scan (v27.0)<sup>76</sup>, Interproscan<sup>77</sup> and Blast2Go v2<sup>78</sup>.

**RNA sequencing.** Nematodes (*B. xylophilus* strain Ka4C1) were cultured on a fungal mat of *Botrytis cinerea* on PDA (39 g/L Potato dextrose agar, Nissui) plate for 4 days and eggs were collected as described in Iwahori and Futai<sup>79</sup>. Collected eggs were incubated in distilled water at 25 °C for 28 h to let them hatch and age-synchronised. Age-synchronised L2s were then transferred to a 1/4 PDA (4% agar) plate with *B. cinerea* inoculated a priori and cultured at 25 °C. L2, L3, and L4 nematodes were harvested at 3 h, 20 h, and 38 h post nematode inoculation, respectively. Adult males and females were obtained from 52 h culture by hand-picking using a needle. D3 nematodes were collected from two-month old cultures (1/5 PDA supplemented with 3% glycerol). Nematodes of D4 stage were produced and isolated using the artificial pupal chamber method<sup>18,80</sup>. Nematodes were frozen in 150 µL TRI reagent (Life Technology) in liquid nitrogen and homogenised using Biomasher (TaKaRa). This step was repeated >6 times till all nematode bodies were disrupted. Total RNA was then extracted according to TRI reagent standard procedures (Life Technology).

Total RNA samples were qualified using the Bioanalyzer 2100 (Agilent Technology, Inc.). Only samples with an RNA integrity value (RIN) greater than 7.0 were used for library construction. One hundred ng of total RNA was used to construct an Illumina sequencing library using the TruSeq RNA-seq Sample Prep kit according to the manufacturer's recommended protocols (Illumina, San Diego, USA). The libraries were sequenced for 101-bp paired-ends on an Illumina HiSeq2000 sequencer using the standard protocol (Illumina). RNA-seq experiments were conducted in triplicate for D4 and in duplicates for the other stages.

**Differential gene expression analysis.** RNA-seq reads were mapped against the *B. xylophilus* genome reference (v2.0) using Tophat v.2.0<sup>74</sup> under options (fr-unstranded, minimum intron length of 15, maximum intron length of 50,000, microexon search; mate inner distance of 100, mate standard deviation of 40). Host RNA contaminations were checked by mapping RNA-seq reads to the *M. alternatus* reference genome (T. Kikuchi, unpublished) using Tophat v.2.0 with aforementioned options. Mapped read count of each gene was calculated using HTSeq with options (-s no, -a 10, -m union)<sup>81</sup> and differential expression analyses were performed using EdgeR v3.2.4<sup>82</sup>. A transcript was identified as differentially expressed in a pairwise comparison if the following criteria were met: false discovery rate (FDR) ≤ 0.001 and fold change ≥ 2.0. FPKM values were calculated using Cufflinks packages v2.2.1<sup>75</sup> and used to generate for MDS plot using "cmdscale" and ggplot2<sup>83</sup> and for correlation coefficient heatmap using "cor" and "heatmap.3" functions implemented in R (version 3.1.3)<sup>84</sup>. Gene expression heatmaps were generated by converting the FPKM values to z-scores with clustering per gene based on Ward's hierarchical agglomerative clustering method (Euclidean distance measure; Ward.2 criterion) using ComplexHeatmap R Bioconductor package<sup>85</sup>.

We used the method described by Choi *et al.*<sup>24</sup> to elucidate somatic-enriched and germline-enriched genes defined as:

$$\begin{aligned} \text{somatic - enriched} &= (\text{genes upregulated in adult female relative to eggs}) \\ \text{germline - enriched} &= (\text{genes upregulated in eggs relative to male}) \text{ or} \\ &(\text{genes upregulated in eggs and/or female relative to other stages}) \\ &\text{-(somatic - enriched genes).} \end{aligned}$$

Temporal gene expression patterns were clustered by the k-means method using the STEM program<sup>86</sup>. The optimal cluster number  $k = 10$  was estimated by the gap statistic method<sup>87</sup> using R (v3.3.3)<sup>84</sup> and the factoextra package. Gene ontology term enrichment analyses were performed using the Goseq (v1.4.0)<sup>88</sup> package implemented in R program (v3.1.3)<sup>84</sup> or the function implemented in STEM<sup>86</sup>.

**Orthologue analyses.** Based on stage specific RNA-seq analyses<sup>48</sup>, *C. elegans* genes were categorized into three groups; 'flat' showing stable expression in propagative juveniles, 'rising' showing increasing expression along the nematode's growth and 'oscillating' showing pervasive and phase-locked oscillations of gene expression in developing juveniles. The oscillatory pattern demonstrates an eight-hour periodic cosine-curve wave. Therefore, when comparing gene expression between stages, those genes may lead to false positives<sup>43</sup>. Orthologous relationships between *B. xylophilus* and *C. elegans* genes were identified using OrthoMCL version 5<sup>89</sup>. The orthologue information (orthofamily) was used to classify *B. xylophilus* genes into the three expression pattern categories ('flat', 'rising' and 'oscillation') according to Hendirks *et al.*<sup>48</sup>.

**Identification of effector candidate genes.** Genes encoding cell wall degrading enzymes and expansins in the v2.0 genome were identified using dbCAN<sup>90</sup> and BLASTP<sup>91</sup>. Peptidases were identified using MEROPS v.11.0<sup>92</sup> with HmmerWeb v.2.19 search engine<sup>93</sup> (with cut-off e-value 0.01). Peptidases were classified into aspartic, cysteine, metallo, serine, threonine, unknown\_32 (collagenase) and unknown\_69 (self-processing peptidase) types. Non-peptidase homologues and peptidase inhibitors were not considered in this study.

**Identification of dauer genes and neuropeptide genes.** OrthoMCL results (please see above) were used to identify orthologous genes of *C. elegans* dauer genes<sup>29</sup> and neuropeptide genes in *B. xylophilus*. The orthologous relationships were confirmed using the orthologue gene trees in the Wormbase Parasites<sup>94</sup>. Gene expressions of *C. elegans* genes in each developmental stage were derived from Wormbase WS220 (<http://www.wormbase.org>; expression page)<sup>95,96</sup>.

### Data Availability

The genome assembly/annotation and RNA sequence data has been deposited to NCBI/EMBL/DDBJ under BioProject accession ID PRJDB7519 and PRJDB3458, respectively.

## References

- Jones, J. T. *et al.* Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology* **14**, 946–961 (2013).
- Ichihara, Y., Fukuda, K. & Suzuki, K. Early symptom development and histological changes associated with migration of *Bursaphelenchus xylophilus* in seedling tissues of *Pinus thunbergii*. *Plant Disease* **84**, 675–680 (2000).
- Jones, J. T., Moens, M., Mota, M., Li, H. & Kikuchi, T. *Bursaphelenchus xylophilus*: opportunities in comparative genomics and molecular host–parasite interactions. *Molecular plant pathology* **9**, 357–368 (2008).
- Futai, K. Pine wood nematode, *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* **51**, 61–83 (2013).
- Mamiya, Y. Pine wood nematode, *Bursaphelenchus lignicolus* Mamiya and Kiyohara, as a causal agent of pine wilting disease. *Review of Plant Protection Research* **5**, 46–60 (1972).
- Ishibashi, N. & Kondo, E. The developmental stages of nematodes for survival, with some discussions on the term “infective” and “dauer”. *Japanese Journal of Nematology* **4**, 1–10 (1974).
- Mamiya, Y. The life history of the pine wood nematode, *Bursaphelenchus lignicolus*. *Japanese Journal of Nematology* **5**, 16–25 (1975).
- Mamiya, Y. & Enda, N. Transmission of *Bursaphelenchus lignicolus* (Nematoda: Aphelenchoididae) by *Monochamus alternatus* (Coleoptera: Cerambycidae). *Nematologica* **18**, 159–162 (1972).
- Cassada, R. C. & Russell, R. L. The dauerlarva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. *Developmental biology* **46**, 326–342 (1975).
- Ishibashi, N. & Kondo, E. Occurrence and Survival of the Dispersal Forms of Pine Wood Nematode, *Bursaphelenchus lignicolus* MAMIYA and KIYOHARA: Ecological significance of dormancy in plant parasitic nematodes. *V. Applied Entomology and Zoology* **12**, 293–302 (1977).
- Kiyohara, T. & Suzuki, K. Seasonal change in population of *Bursaphelenchus lignicolus* in the wood of Japanese black pine. *Trans Jpn For Soc* **86**, 296–298 (1975).
- Tanaka, S. E., Aikawa, T., Takeuchi-Kaneko, Y., Fukuda, K. & Kanzaki, N. Artificial induction of third-stage dispersal juveniles of *Bursaphelenchus xylophilus* using newly established inbred lines. *PLoS one* **12**, e0187127, <https://doi.org/10.1371/journal.pone.0187127> (2017).
- Warren, J. E. & Linit, M. Effect of *Monochamus carolinensis* on the life history of the pinewood nematode, *Bursaphelenchus xylophilus*. *Journal of nematology* **25**, 703 (1993).
- Zhao, L. *et al.* Chemical signals synchronize the life cycles of a plant-parasitic nematode and its vector beetle. *Current Biology* **23**, 2038–2043 (2013).
- Kikuchi, T., Jones, J. T., Aikawa, T., Kosaka, H. & Ogura, N. A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode *Bursaphelenchus xylophilus*. *FEBS letters* **572**, 201–205 (2004).
- Kikuchi, T., Shibuya, H. & Jones, J. T. Molecular and biochemical characterization of an endo- $\beta$ -1, 3-glucanase from the pinewood nematode *Bursaphelenchus xylophilus* acquired by horizontal gene transfer from bacteria. *Biochemical Journal* **389**, 117–125 (2005).
- Kikuchi, T., Shibuya, H., Aikawa, T. & Jones, J. T. Cloning and characterization of pectate lyases expressed in the esophageal gland of the pine wood nematode *Bursaphelenchus xylophilus*. *Molecular plant-microbe interactions* **19**, 280–287 (2006).
- Kikuchi, T. *et al.* Expressed sequence tag (EST) analysis of the pine wood nematode *Bursaphelenchus xylophilus* and *B. mucronatus*. *Molecular and biochemical parasitology* **155**, 9–17 (2007).
- Shinya, R. *et al.* Secretome analysis of the pine wood nematode *Bursaphelenchus xylophilus* reveals the tangled roots of parasitism and its potential for molecular mimicry. *PLoS one* **8**, e67377 (2013).
- Espada, M. *et al.* Identification and characterization of parasitism genes from the pinewood nematode *Bursaphelenchus xylophilus* reveals a multilayered detoxification strategy. *Molecular plant pathology* **17**, 286–295 (2016).
- Kikuchi, T., Eves-van den Akker, S. & Jones, J. T. Genome evolution of plant-parasitic nematodes. *Annual review of phytopathology* **55**, 333–354 (2017).
- Tsai, I. J. *et al.* Transcriptional and morphological changes in the transition from mycetophagous to phytophagous phase in the plant-parasitic nematode *Bursaphelenchus xylophilus*. *Molecular plant pathology* **17**, 77–83 (2016).
- Kikuchi, T. *et al.* Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS pathogens* **7**, e1002219 (2011).
- Choi, Y. J. *et al.* A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, *Brugia malayi*. *PLoS Negl Trop Dis* **5**, e1409, <https://doi.org/10.1371/journal.pntd.0001409> (2011).
- Thomas, J. H., Birnby, D. A. & Vowles, J. J. Evidence for parallel processing of sensory information controlling dauer formation in *Caenorhabditis elegans*. *Genetics* **134**, 1105–1117 (1993).
- Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994 (1997).
- Antebi, A., Yeh, W.-H., Tait, D., Hedgecock, E. M. & Riddle, D. L. *daf-12* encodes a nuclear receptor that regulates the dauer diapause and developmental age in *C. elegans*. *Genes & development* **14**, 1512–1527 (2000).
- Fielenbach, N. & Antebi, A. C. *elegans* dauer formation and the molecular basis of plasticity. *Genes & development* **22**, 2149–2165 (2008).
- Gilbert, A., Curran, D. M., Harvey, S. C. & Wasmuth, J. D. Expanding the view on the evolution of the nematode dauer signalling pathways: refinement through gene gain and pathway co-option. *BMC genomics* **17**, 476 (2016).
- Nolan, K. M., Sarafi-Reinach, T. R., Horne, J. G., Saffer, A. M. & Sengupta, P. The DAF-7 TGF- $\beta$  signaling pathway regulates chemosensory receptor gene expression in *C. elegans*. *Genes Dev* **16**, 3061–3073, <https://doi.org/10.1101/gad.1027702> (2002).
- Crook, M., Thompson, F. J., Grant, W. N. & Viney, M. E. *daf-7* and the development of *Strongyloides ratti* and *Parastrongyloides trichosuri*. *Mol Biochem Parasitol* **139**, 213–223, <https://doi.org/10.1016/j.molbiopara.2004.11.010> (2005).
- Li, C. & Kim, K. Neuropeptides. *WormBook*, 1–36, <https://doi.org/10.1895/wormbook.1.142.1> (2008).
- Lee, J. S. *et al.* FMRFamide-like peptides expand the behavioral repertoire of a densely connected nervous system. *Proc Natl Acad Sci USA* **114**, E10726–E10735, <https://doi.org/10.1073/pnas.1710374114> (2017).
- Murphy, C. T. & Hu, P. J. Insulin/insulin-like growth factor signaling in *C. elegans*. *WormBook: the online review of C. elegans biology*, 1–43, <https://doi.org/10.1895/wormbook.1.164.1> (2013).
- Zheng, S. *et al.* A functional study of all 40 *Caenorhabditis elegans* insulin-like peptides. *The Journal of biological chemistry* **293**, 16912–16922, <https://doi.org/10.1074/jbc.RA118.004542> (2018).
- Pierce, S. B. *et al.* Regulation of DAF-2 receptor signaling by human insulin and *ins-1*, a member of the unusually large and diverse *C. elegans* insulin gene family. *Genes Dev* **15**, 672–686, <https://doi.org/10.1101/gad.867301> (2001).
- Smant, G. *et al.* Endogenous cellulases in animals: isolation of  $\beta$ -1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences* **95**, 4906–4911 (1998).
- Bellaïf, S. *et al.* Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS pathogens* **4**, e1000192 (2008).
- Wang, J. *et al.* Dual roles for the variable domain in protein trafficking and host-specific recognition of Heterodera glycines CLE effector proteins. *New Phytologist* **187**, 1003–1017 (2010).
- Malagón, D., Benítez, R., Kašný, M. & Adroher, F. J. Peptidases in parasitic nematodes: a review. *Parasites: ecology, diseases and management*. Nova Science Publishers, Hauppauge, 61–102 (2013).
- Davis, E. L. *et al.* Nematode parasitism genes. *Annual Review of Phytopathology* **38**, 365–396 (2000).

42. Odani, K., Sasaki, S., Yamamoto, N., Nishiyama, Y. & Tamura, H. Differences in dispersal and multiplication of two associated nematodes, *Bursaphelenchus xylophilus* and *Bursaphelenchus mucronatus* in pine seedlings in relation to the pine wilt disease development. *Journal of the Japanese Forestry Society* **67**, 398–403 (1985).
43. Fukushige, H. & Futai, K. Seasonal changes in *Bursaphelenchus xylophilus* populations and occurrence of fungi in *Pinus thunbergii* trees inoculated with the nematode. *Japanese Journal of nematology* **17**, 8–16 (1987).
44. Shinya, R., Morisaka, H., Takeuchi, Y., Ueda, M. & Futai, K. Comparison of the surface coat proteins of the pine wood nematode appeared during host pine infection and *in vitro* culture by a proteomic approach. *Phytopathology* **100**, 1289–1297 (2010).
45. Mamiya, Y. Pathology of the pine wilt disease caused by *Bursaphelenchus xylophilus*. *Annual review of Phytopathology* **21**, 201–220 (1983).
46. Kondo, E. & Ishibashi, N. Ultrastructural differences between the propagative and dispersal forms in pine wood nematode, *Bursaphelenchus lignicolus*, with reference to the survival. *Applied entomology and zoology* **13**, 1–11 (1978).
47. Forge, T. & Sutherland, J. Population dynamics of the pine wood nematode, *Bursaphelenchus xylophilus*, in excised branch segments of western North American conifers. *Fundamental and Applied Nematology* **19**, 349–356 (1996).
48. Hendriks, G.-J., Gaidatzis, D., Aeschmann, F. & Grobthans, H. Extensive oscillatory gene expression during *C. elegans* larval development. *Molecular cell* **53**, 380–392 (2014).
49. Hunt, V. L. *et al.* The genomic basis of parasitism in the Strongyloides clade of nematodes. *Nature genetics* **48**, 299–307, <https://doi.org/10.1038/ng.3495> (2016).
50. Tang, Y. T. *et al.* Genome of the human hookworm *Necator americanus*. *Nature genetics* **46**, 261–269, <https://doi.org/10.1038/ng.2875> (2014).
51. Chalmers, I. W. *et al.* Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC genomics* **9**, 89 (2008).
52. Rangaraju, S. *et al.* Potassium channel modulation by a toxin domain in matrix metalloprotease 23. *Journal of biological chemistry* **285**, 9124–9136 (2010).
53. McNeilly, T. N. *et al.* Niche-specific gene expression in a parasitic nematode; increased expression of immunomodulators in *Teladorsagia circumcincta* larvae derived from host mucosa. *Scientific reports* **7**, 7214 (2017).
54. Cotton, J. A. *et al.* The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome biology* **15**, R43 (2014).
55. Ju, Y., Wang, X., Guan, T., Peng, D. & Li, H. Versatile glycoside hydrolase family 18 chitinases for fungi ingestion and reproduction in the pinewood nematode *Bursaphelenchus xylophilus*. *International journal for parasitology* **46**, 819–828 (2016).
56. Klass, M. & Hirsh, D. Non-ageing developmental variant of *Caenorhabditis elegans*. *Nature* **260**, 523 (1976).
57. Crook, M. The dauer hypothesis and the evolution of parasitism: 20years on and still going strong. *International journal for parasitology* **44**, 1–8 (2014).
58. Harvey, S. C., Barker, G. L., Shorto, A. & Viney, M. E. Natural variation in gene expression in the early development of dauer larvae of *Caenorhabditis elegans*. *BMC genomics* **10**, 325 (2009).
59. Golden, J. W. & Riddle, D. L. The *Caenorhabditis elegans* dauer larva: developmental effects of pheromone, food, and temperature. *Developmental biology* **102**, 368–378 (1984).
60. Kanzaki, N. & Giblin-Davis, R. M. Diversity and plant pathogenicity of *Bursaphelenchus* and related nematodes in relation to their vector bionomics. *Current Forestry Reports* **4**, 85–100 (2018).
61. Obayashi, N., Obayashi, N. & Niisato, T. *Longicorn beetles of Japan*. (Tokai University Press, 2007).
62. Coburn, C. M. & Bargmann, C. I. A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans*. *Neuron* **17**, 695–706 (1996).
63. Ogawa, A., Streit, A., Antebi, A. & Sommer, R. J. A conserved endocrine mechanism controls the formation of dauer and infective larvae in nematodes. *Current Biology* **19**, 67–71 (2009).
64. Wang, Z. *et al.* Identification of the nuclear receptor DAF-12 as a therapeutic target in parasitic nematodes. *Proceedings of the National Academy of Sciences* **106**, 9138–9143 (2009).
65. Bai, X. *et al.* A lover and a fighter: the genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora*. *PLoS one* **8**, e69618 (2013).
66. Park, D. *et al.* Interaction of structure-specific and promiscuous G-protein-coupled receptors mediates small-molecule signaling in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **109**, 9917–9922 (2012).
67. Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013).
68. Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome biology* **13**, R56 (2012).
69. Tsai, I. J., Otto, T. D. & Berriman, M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome biology* **11**, R41 (2010).
70. Otto, T. D., Sanders, M., Berriman, M. & Newbold, C. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* **26**, 1704–1707, <https://doi.org/10.1093/bioinformatics/btq269> (2010).
71. Huang, S. *et al.* HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. *Genome Res* **22**, 1581–1588, <https://doi.org/10.1101/gr.133652.111> (2012).
72. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* **33**, W465–467, <https://doi.org/10.1093/nar/gki458> (2005).
73. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
74. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, <https://doi.org/10.1186/gb-2013-14-4-r36> (2013).
75. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562 (2012).
76. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* **44**, D279–D285 (2015).
77. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
78. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
79. Iwahori, H. & Futai, K. A simple method for disinfection of pine wood nematode eggs using their adhesiveness. *Japanese Journal of Nematology* **15**, 64–64 (1985).
80. Aikawa, T. & Togashi, K. An effect of inoculum quantity of *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) on the nematode load of *Monochamus alternatus* (Coleoptera: Cerambycidae) in laboratory. *Nematological Research (Japanese Journal of Nematology)* **27**, 14–21 (1997).
81. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, <https://doi.org/10.1093/bioinformatics/btu638> (2015).
82. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
83. Goff, L., Trapnell, C. & Kelley, D. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. *R package version 2* (2013).

84. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, 2010).
85. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849, <https://doi.org/10.1093/bioinformatics/btw313> (2016).
86. Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC bioinformatics* **7**, 191 (2006).
87. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 411–423 (2001).
88. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology* **11**, R14 (2010).
89. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* **13**, 2178–2189 (2003).
90. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **40**, W445–W451 (2012).
91. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410, [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
92. Rawlings, N. D., Barrett, A. J. & Bateman, A. Using the MEROPS database for proteolytic enzymes and their inhibitors and substrates. *Current protocols in bioinformatics*, 1.25. 21–21.25. 33 (2014).
93. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic acids research* **43**, W30–W38 (2015).
94. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol Biochem Parasitol* **215**, 2–10, <https://doi.org/10.1016/j.molbiopara.2016.11.005> (2017).
95. Li, J. J., Huang, H., Bickel, P. J. & Brenner, S. E. Comparison of *D. melanogaster* and *C. elegans* developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res* **24**, 1086–1101, <https://doi.org/10.1101/gr.170100.113> (2014).
96. Rogers, A. *et al.* WormBase 2007. *Nucleic Acids Res* **36**, D612–617, <https://doi.org/10.1093/nar/gkm975> (2008).

## Acknowledgements

The authors thank members of the labs in Miyazaki, Tokyo and Kyoto University for their valuable discussion and supports.

## Author Contributions

Y.T.K., K.F., N.K. and T.K. designed the study. S.E.T. and R.T. prepared materials. S.E.T. performed experiments. S.E.T., M.D., Y.M., I.J.T. and T.K. performed data analysis. S.E.T., M.D., M.B. and T.K. wrote the manuscript text. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-42570-7>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019