1 **The complete mitochondrial genome of the conifer needle endophyte,**
2 ***Phialocephala scopiformis* DAOMC 229536 confirms evolutionary**
3 **division within the fungal *Phialocephala fortinii* s.l. – *Acephala***
4 ***appalanata* species complex.**

5 Brent M. ROBICHEAU^a, Alexander P. YOUNG^a, Kurt LABUTTI^b, Igor V.

6 GRIGORIEV^b, Allison K. WALKER^{a*}

7 ^aDepartment of Biology, Acadia University, 33 Westwood Ave, Wolfville, NS, B4P
8 2R6, Canada

9 ^bU.S. Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut
10 Creek, CA 94598, USA

11 *Corresponding author

12 Email: allison.walker@acadiu.ca Tel: 1 902 585 1333; Fax: 1 902 585 1059

13 **Abstract**

14 Despite the recent surge in mitochondrial (mt) genome sequencing, Kingdom
15 *Fungi* remains underrepresented with respect to mtDNA. We describe the 36,919 bp
16 mt genome of the conifer needle endophyte, *Phialocephala scopiformis* DAOMC
17 229536 (*Helotiales*, *Ascomycota*). This strain of *P. scopiformis* is of interest to the
18 Canadian forestry industry as it produces the potent anti-insectan compound
19 rugulosin. Mitochondrial sequence was obtained from whole genome shotgun
20 sequencing of *P. scopiformis* DAOMC 229536. Evolutionary comparison to the only
21 other published complete *Phialocephala* mt genome, *Phialocephala subalpina*,
22 indicates that the suite of common mt genes – *cox1-3* and *cob*, *nad1-6* and *4L*, *atp6*, 8
23 and 9, as well as *rrnL* and *rrnS* – has retained an identical gene order. *Nad4L* remains
24 one of the most conserved mitochondrial genes within the genus. Members of the
25 closely related *Phialocephala fortinii* s.l. – *Acephala appalanata* species complex
26 (PAC) share too much sequence similarity to properly resolve evolutionary branches
27 using ITS barcoding alone. Using *P. scopiformis* mt sequence as a reference outgroup
28 we determined ancestral gene states that help confirm clades within *Phialocephala*.
29 Our results show: (1) the complete mt genome of *P. scopiformis*, representing the 10th
30 complete mt genome for the diverse fungal order *Helotiales* (containing >3,800
31 species), and (2) how large-scale genomic patterns, such as mitochondrial gene order,
32 can be used to confirm evolutionary branches within fungal species complexes.

Key Words: endophyte, gene duplication, mitochondrial genome, rugulosin

Abbreviations: *atp*, ATP synthase subunit; *cob*, cytochrome *b*; *cox*, cytochrome *c* oxidase subunit; *ITS*, internal transcribed spacer region of ribosomal DNA; *PAC*, *Phialocephala fortinii* s.l. – *Acephala applanata* species complex; *PI*, percent identity; *ORF*, open reading frame; *OXPPOS*, oxidative phosphorylation; *rrnL*, ribosomal RNA large subunit; *rrnS*, ribosomal RNA small subunit; *trn*, transfer RNA; *rps3*, ribosomal protein S3

1.0 Introduction

Phialocephala scopiformis DAOMC 229536 was first isolated as a needle endophyte of white spruce, *Picea glauca* (Moench) Voss, in eastern Quebec (Sumarah & Miller 2009). This strain produces the potent anti-insectan secondary metabolite rugulosin (Calhoun et al. 1992). Rugulosin has a strong inhibitory effect on the growth rate of spruce budworm larvae, suppressing the population growth of this devastating pest in areas surrounding trees inoculated with the fungus (Miller 2011; Miller et al. 2008).

Kendrick (1961) described the genus *Phialocephala* as a group of species sharing morphologically similar conidiophores. Grünig et al. (2002) later showed the genus was polyphyletic, concluding that the similarity of conidiophores between species represents an example of convergent evolution. ITS rDNA sequencing placed *Phialocephala scopiformis* in a monophyletic clade with *Phialocephala compacta*, *Phialocephala dimorphospora*, and *Phialocephala fortinii* (Grünig et al. 2002). However, more recent evidence places *P. scopiformis* as a basal group to the aforementioned clade, which also includes members of the *Phialocephala fortinii* s.l. – *Acephala applanata* species complex (or PAC; see Jacobs et al. 2003 and Tanney et al. 2016). Phylogenetic placement based on nuclear ribosomal markers assigns PAC species to the diverse order *Helotiales* (Ascomycota). *Helotiales* comprises over 3,800

species including several economically important plant pathogens (Grünig et al. 2008). Members of the PAC are frequently recovered as dark septate endophytes (DSE) of conifer roots. Different *Phialocephala* species dominate the aboveground DSE community in these tree species (Grünig et al. 2008). *P. scopiformis* has been frequently isolated from healthy needles of multiple conifer species, indicating it is non-pathogenic, whereas *P. dimorphospora* and *P. fortinii* may range from mutualistic to pathogenic, with some strains causing soft-rot (Anagnost et al. 2000). In addition to its endophytic role in conifer needles, *Phialocephala scopiformis* is known as a saprotroph of decorticated softwood and Kowalski & Kehr (1995) first described it as a cambium endophyte from *Picea abies* in Germany; this fungus is not known from roots (Tanney et al. 2016).

Given that the mitochondrial (mt) genome evolves more rapidly than the nuclear genome and has been shown to produce reliable phylogenetic trees, it has been established as an effective evolutionary marker in fungi (Wu et al. 2009, Duò et al. 2012, Youssar et al. 2013, Shen et al. 2015). Fungal mt genomes are frequently circular molecules consisting of 14 highly conserved protein coding genes related to oxidative phosphorylation (OXPHOS), one large (*rrnL*) and one small (*rrnS*) ribosomal RNA (rRNA) gene, and a set of transfer RNA (tRNA or *trn*) genes which can vary between species (Gray et al. 1999, Lavín et al. 2008). In fungal mt genomes, open reading frames (ORFs) exist and represent protein-coding genes with unknown functions (Wu et al. 2009, Ghikas et al. 2010). *Phialocephala subalpina*, the only other species within the genus *Phialocephala* to have a fully annotated mt genome, has four of these unidentified reading frames (also called URFs; Duò et al. 2012).

The fungal mitochondrial genome project was launched nearly two decades ago (Paquin et al. 1997). Since the beginning of this global initiative, 347 fungal mt

genomes have been deposited and made publicly available via NCBI's Organelle Genome database, 220 of which have been curated. However, there are only 9 sequenced mt genomes from the order *Helotiales* and just 14 total from the class *Leotiomyces* (*Ascomycota*) (Duò 2012, Youssar 2013). Given that very few mt genomes from the order *Helotiales* have been annotated, the *P. scopiformis* mt genome presented herein will serve as a basis for further exploration into the relationships among helotialean fungal mitochondrial genomes.

Here we describe the 36,919 bp mitochondrial genome of the rugulosin-producing conifer endophyte *Phialocephala scopiformis* DAOMC 229536 (*Helotiales*; *Ascomycota*). We provide a comparative analysis of the whole *Phialocephala subalpina* mt genome with all existing partial *Phialocephala* species mt genomes to explore evolutionary trends within the PAC. The objectives for completing this work include: (1) providing data and direction for the future development of strain-specific biomarkers; (2) using large scale genetic patterns, such as gene order, to infer evolutionary relationships between *P. scopiformis* and members of the PAC; and (3) documenting mitochondrial diversity within the large fungal order *Helotiales*. Additionally, we provide further evidence that *Phialocephala europaea*, *Phialocephala letzii*, and *Phialocephala turicensis* may be separated into their own clade within the PAC based on evolutionary events within the mitochondrial ORF2 gene; supporting trends noted in Duò et al. (2012).

2.0 Materials and Methods

2.1 Acquisition of *Phialocephala scopiformis* DAOMC 229536 mitochondrial genome sequence

The *Phialocephala scopiformis* DAOMC 229536 mt genome was sequenced

110 as part of an entire nuclear genome sequence project using the Illumina platform.
111 Total DNA was extracted using a modified Joint Genome Institute (JGI)
112 cetyltrimethylammonium bromide (CTAB) and genomic tip (Qiagen) protocol. DNA
113 was sheared and size selected to produce two 300 bp insert size fragments and 4 kb
114 Cre-lox long-mate-pair libraries, and sequenced on Illumina HiSeq in 2×150 bp and
115 2×100 bp reads, respectively. Stranded cDNA libraries were generated using the
116 Illumina TruSeq stranded RNA LT kit from mRNA fragmented and reverse
117 transcribed using random hexamers and SuperScript II reverse transcriptase
118 (Invitrogen) followed by second strand synthesis. The fragmented cDNA was treated
119 with end-pair, A-tailing, adapter ligation, and 10 cycles of PCR, and sequenced on
120 HiSeq in 2×150 bp reads format. Genomic data were filtered and assembled with
121 AllPathsLG vR49403 to produce an initial single 35.9 kb contig of the mitochondrion.
122 During initial assembly an $80\times$ coverage threshold was used for the fragment library,
123 and a $90\times$ coverage threshold was selected for the jumping library of long-mate-pair
124 data. The latter library aided in improving the scaffolding in the mt genome build. For
125 both assemblies an expected molecule size of 37 kb was assumed based on the known
126 size of sequence for *P. scopiformis*' closest relative in GenBank, *Phialocephala*
127 *subalpina* (>35kb; accession: NC_015789.1; Duò et al. 2012). The single contig of
128 the mitochondrial genome produced from the AllPathsLG run appeared circular with
129 1,461 links supporting circularity. Upon further inspection of the mitochondrial contig
130 a gap of ~1kb was detected due to the presence of a missing *nad3* gene that is
131 typically found in *Phialocephala* (Duò et al. 2012) and whose sequence was detected
132 by aligning *P. subalpina nad3* sequence to the *P. scopiformis* cDNA library using the
133 *Map to Reference* tool in Geneious v8.0.5 (Kearse et al. 2012). Primers *nad2*-forward
134 5' [TTT GCC TAT AGT ACG ATA AGT CAT GT] 3' and *cox2*-reverse 5' [CCT

CAA GGT TTA GGT GTA TCA CAA] 3' were designed using Primer3 (Untergasser et al. 2012) for the region suspected to house *nad3* (based on conservation of gene order). The forward primer is designed to sit in *nad2* and the reverse primer in *cox2*, which were predicted to flank *nad3*. The remaining 1,042 bp of sequence containing *nad3* was obtained using successful Long Range-PCR amplification with the aforementioned primers. Reactions were carried out using the volume recommendations stated in the protocol for a Phusion® High-Fidelity DNA polymerase kit (New England BioLabs Inc.; Catalogue No. M0530S). Thermocycler parameters were as follows: initiation at 98 °C for 30 sec, 33 cycles of denaturation at 98 °C for 10 sec, annealing at 63 °C for 30 sec, and extension at 72°C for 1 min 45 sec, with a final extension at 72 °C for 3 min. We then sequenced the amplicon in the forward and reverse direction using Sanger sequencing at McGill University's Genome Innovation Centre (Montréal, QC). We also aligned the final mitochondrial build to the scaffolds of the nuclear genome using the *Map to Reference* tool in Geneious v8.0.5 (Kearse et al. 2012) to ensure that any potential mitochondrial fragment(s) present in nuclear DNA was not present in our mitochondrial genome build. The prior alignment test did not resolve any matches. Lastly, we remapped initial sequence read datasets with AllPathsLG vR49403 to our final mitochondrial assembly that included *nad3*. The total aligned coverage from both read libraries was 722× coverage for the final complete mitochondrial genome (175× for fragment library + 547× for long-mate-pair library) and 575× coverage for the specific *nad3* region that was added to the initial assembly (48× for fragment library + 527× for long-mate-pair library).

Phialocephala scopiformis DAOMC 229536 nuclear, cDNA, and mitochondrial sequence is referenced under the NCBI GenBank accession number

LKNI00000000 (Walker et al. 2016) and is also available at the JGI MycoCosm portal (Grigoriev et al. 2014).

2.2 *Phialocephala scopiformis* DAOMC 229536 mitochondrial genome annotation

Genes were initially identified using the MITOS web server (Bernt et al. 2013), and further confirmed by: (1) the presence of open reading frames (using a mold and protozoan mitochondrial genetic code and the ORF Finder web server (Wheeler et al. 2003), (2) cross-comparison to the *Phialocephala subalpina* mt genome build (accession: NC_015789.1; Duò et al. 2012), (3) alignment to fungal sequences in BLAST (taxid: 4751; Altschul et al. 1990; Camacho et al. 2009), and (4) alignment to sequence within the *P. scopiformis* cDNA library using the *Map to reference* tool in Geneious v8.0.5 (Kearse et al. 2012). Transfer RNAs were predicted based on the gene sequences' expected ability to fold into stem and leaf secondary structures using the programs tRNAscan-SE v1.21 (Lowe & Eddy 1997) and ARWEN v1.2 (Laslett & Canbäck 2008). The start and ends of ribosomal RNA genes were identified based on sequence homology to *P. subalpina* mtDNA using both Clustal Omega (Li et al. 2015) and BLAST (Altschul et al. 1990; Camacho et al. 2009) alignment methods (i.e., both a multiple sequence alignment and a local alignment approach). The putative identification of a *coxI* intron in *P. scopiformis* as mobile DNA was inferred by local alignment of the intron in BLAST to fungal sequence (Altschul et al. 1990; Camacho et al. 2009). We further tested if our final mitochondrial genome annotation was correct by running the unannotated sequence through the webserver MFannot (Beck and Lang 2010).

2.3 Gene conservation analysis

All individual nucleotide gene sequences of *Phialocephala scopiformis* and *Phialocephala subalpina* were independently aligned using Clustal Omega v1.2.1 (Li

et al. 2015). Individual gene alignments were imported into Jalview v15.0 (Waterhouse et al. 2009) and converted to the *.aln file format prior to being imported to MEGA v6 (Tamura et al. 2013) and pairwise distances were computed using the p-distance model. In calculating p-distances all three codon positions were included and pairwise deletion was used to account for gaps and missing data-points. This process was repeated after converting nucleotide sequences to amino acid sequences using EMBOSS Transeq v6.6.0 (Li et al. 2015). Total evolutionary distances were also computed using the p-distance model on concatenated gene sequences. To accomplish this, each nucleotide gene alignment was trimmed to a conserved initial and terminal position and then sequences for multiple genes were concatenated in their proper orientation starting with *coxI* using MEGA v6 (Tamura et al. 2013); total p-distance was then calculated for all three codon positions using pairwise deletion for gaps. This process was repeated for a calculation of total amino acid sequence divergence (measured as p-distance). To determine the conservation of gene order a dot-plot was generated using the software JDotter v1.0 (Brodie et al. 2004) with a sliding window of 100bp/100bp.

2.4 Phylogenetic analysis and PAC species identification test

Partial mitochondrial genome sequences for additional *Phialocephala* species were retrieved from GenBank (Benson et al. 2005; Table 1). A mitochondrial genome architecture diagram was produced in Geneious v8.0.5 (Kearse et al. 2012). The phylogenetic tree used to test the hypothesis that mitochondrial evolutionary events match the division of clades within the PAC was produced using ORF2 nucleotide sequences obtained herein and from sequences described in Table 1. The dendrogram was constructed in MEGA v6 (Tamura et al. 2013) using the neighbour-joining tree reconstruction method, 1,199 nucleotides representing 1st+2nd+3rd codon positions,

and 500 bootstrap replications. The Tamura 3-parameter model with a discrete Gamma distribution (+G) with five rate categories was selected as the best-fit nucleotide substitution model for ORF2 sequences based on calculated Bayesian Information Criterion (BIC) and corrected Akaike Information Criterion (AIC) values determined in MEGA v6 (Tamura et al. 2013)). Alignment data for ORF2 sequences are available in Appendix B and were aligned in MEGA v6 using the ClustalW algorithm option (Tamura et al. 2013).

To test the utility of mitochondrial genes for species-level identification, ITS, *atp9*, and ORF2 sequences (described in Table 1) were locally aligned in a pairwise fashion using BLAST (Altschul et al. 1990; Camacho et al. 2009). The percent identity (PI) values generated were then plotted as frequency distributions to show the PI values (for each gene) for both intraspecific and within-PAC comparisons. The frequency density plot was produced in R v3.2.1 (R Core Team 2015) using the ggplot package (Wickham 2009).

We used a Z-test of Selection in MEGA v6 to test whether ORF2 sequences are under selection or if they are under neutral evolution, i.e. drifting (Tamura et al. 2013). This analysis was completed using the Nei-Gojobori (Jukes-Cantor) substitution model using pairwise deletion to account for missing data-points/gaps (Tamura et al. 2013). As a more stringent model we selected the modified Nei-Gojobori (Jukes-Cantor) with a transition/transversion bias of 1.63. This bias value was estimating in MEGA v6 under the assumption that the best substitution model was the Tamura 3-parameter + G model, which was originally determine above for use in our phylogenetic tree reconstruction (Tamura et al. 2013). Variance was estimated using 1000 bootstrap replicates and the final dataset had a total of 397 positions.

3.0 Results and Discussion

3.1 Mitochondrial genome content

The complete mitochondrial genome of *Phialocephala scopiformis* DAOMC 229536 is 36,919 bp with a circular topology and a mean GC content of 27.1%. It codes for the typical 14 conserved protein-coding genes responsible for oxidative phosphorylation (OXPHOS) and ATP synthesis in mitochondria: cytochrome c oxidase subunits I, II and III (*coxI-3*), NADH dehydrogenase subunits 1-6 and 4L (*nad1-6* and *nad4L*), ATP synthase subunits 6, 8 and 9 (*atp6*, *atp8* and *atp9*) and cytochrome b (*cob*; Fig 1). The mt genome also codes for the standard large and small subunit ribosomal RNA (rRNA) genes (*rrnL* and *rrnS*, respectively), which are involved in translational processes within the mitochondria (Fig 1). When compared to sequences from its close relative, *Phialocephala subalpina*, all of the above OXPHOS and rRNA genes are conserved between the two genomes (compare Fig 1 to Fig 2A). Additionally, a conserved ribosomal protein S3 (*rps3*) is also present (Fig 1 and Fig 2A). *Phialocephala scopiformis*, however, does lack both GIY-YIG endonuclease and N-Acetyl-transferase genes identified by Duò et al. (2012) in *P. subalpina* (Fig 1 and Fig 2A). The *P. scopiformis* mt genome also contains 25 tRNA (*trn*) genes that encode tRNAs for all 20 essential amino acids (with three copies of *trnM* and two copies of *trnR*, *trnL* and *trnS* (Fig 1)). Transfer RNAs were confirmed based on the sequences' ability to fold into characteristic cloverleaf structures (see supplementary data in Appendix A). Furthermore, *Phialocephala scopiformis* DAOMC 229536 possesses an unidentified reading frame, known as ORF2, which was identified in the mt genome of *P. subalpina* and in partial mt sequence from other PAC species (i.e., *A. applanata*, *P. europaea*, *P. fortinii*, *P. helvetica*, *P. letzii*, *P. turicensis* and *P. uotilensis*, [Fig 1; Duò et al. 2012]). All of the protein-coding rRNA

and *rps3* genes identified above in *P. scopiformis* were confirmed by cross-comparison of ORFs to *P. subalpina* mt sequence, and also by local alignment to both fungal sequences using NCBI BLAST and transcriptome data for *P. scopiformis* (i.e., its cDNA sequence library). Identical gene predictions for protein coding, rRNA, and tRNA sequences were resolved by MITOS (Bernt et al. 2013) and MFannot (Beck and Lang 2010) webserver. As noted in Section 2.1, the initial mitochondrial genome build lacked *nad3*. We are not sure why this gap was present in our initial build, however, because the complete mt genome sequence for the close relative *Phialocephala subalpina* was available on GenBank, we could compare the *P. subalpina* mt genome (with its *nad3* sequence) to the cDNA library generated for *P. scopiformis*, which allowed us to detect the gap. It is important to use published mt genomes to examine any atypical properties in new mt genome builds.

A marked difference between *P. scopiformis* and *P. subalpina* is the presence of a 2,138 bp intron in the *cox1* gene of *P. scopiformis* (Fig 1). This was confirmed using cDNA sequence for *cox1*. Despite the paucity of sequence data available for comparison with other *Helotiales*, the presence of an intron in *cox1* is not unexpected. Seifert et al. (2007) indicated introns within *cox1* are a common feature in fungi, with intron sizes ranging from 134 bp to over 3 kb. *P. scopiformis* is a basal evolutionary group to *P. subalpina* and other PAC individuals (Jacobs et al. 2003 and Tanney et al. 2016), therefore, the acquisition of the intron in the *cox1* gene of *P. scopiformis* must either: (1) have arisen after separation of the two groups to account for the lack of an intron in the *cox1* gene of *P. subalpina*, or (2) the intron may be composed of a mobile sequence which could have been excised from mtDNA within the *P. subalpina* lineage after speciation. Again the lack of sequence data for *Helotiales* makes testing the above hypotheses difficult, however some data are available which

support the second mobile sequence hypothesis. When the *cox1* intron of *P. scopiformis* was aligned using BLAST to the NCBI nucleotide collection (nr/nt) for fungal taxa (taxid: 4751), the highest returned match was for a 975 bp alignment hit with 80% identity to an LAGLIDADG endonuclease (i.e., a mobile homing endonuclease; Taylor et al. (2012)) in *Sclerotinia borealis* mitochondrial sequence (accession code: KJ434027). Since the *cox1* intron in *P. scopiformis* shares a high degree of similarity to a homing endonuclease, there is support that this intron originated in the *P. scopiformis* lineage as a mobile element. However, it is presently impossible to infer whether the intron was [gained and then lost within the PAC lineage, or alternatively, never present in the PAC lineage but gained in *P. scopiformis* after a split between *P. scopiformis* and the PAC lineage. Either scenario is possible given the](#) intron's likely status as a mobile element. In either case such evolutionary events are not uncommon for fungal mitochondrial genomes (Hamari et al. 2001, Hamari et al. 2002, Juhász et al. 2004, Seifert et al. 2007).

3.2 Mitochondrial genome organization and sequence conservation

Like *P. subalpina*, all *P. scopiformis* protein and rRNA genes are transcribed in the same direction (Fig 1 and Fig 2A). Additionally, with the exception of *atp9* [which will be discussed further below; see subsection 3.3 Phylogenetic relationships], all protein and rRNA genes occur in the same order (Fig 2B). The presence of tRNA genes also appears to be concentrated between *rrnS* and *nad2* within both mt genomes (Fig 1 and Fig 2A). The only exception to this is in the region between *cox1* and *nad4*, where in *P. subalpina* a single *trn* gene punctuates the spaces between each of the five protein genes in the region (i.e., four *trn* genes total; Fig 2A).

In *P. scopiformis* this same region only contains two *trn* genes (*trnC* and *trnR* (*UCU*)) between ORF2 and *nadI* (Fig 1).

APPROXIMATE LOCATION OF FIGURES 1 AND 2

Regarding sequence conservation between *P. scopiformis* and *P. subalpina*, the two genomes have an overall nucleotide divergence level of 10% (p-distance for nucleotides = 0.10 ± 0.002 SE) and an overall amino acid divergence level of 6.5% (p-distance for amino acids = 0.065 ± 0.003 SE). NADH dehydrogenase related genes appear least conserved between the two taxa (when observing both nucleotide and amino acid divergences), while cytochrome c oxidase and ATP synthase related genes appear most conserved (Fig 3). Of note is that *nad4L* does not seem to follow the general trends in nucleotide divergence, being the most conserved despite its sister *nad* subunit sequences being least conserved (Fig 3). The two most divergent genes between the two mitochondrial genomes are *rps3* and ORF2; *rps3* and ORF2 are the only genes where the amino acid sequences differ more from its *P. subalpina* counterpart than do the nucleotide sequences (Fig 3).

APPROXIMATE LOCATION OF FIGURE 3

3.3 Phylogenetic relationships

Duò et al. 2012 obtained partial mt sequences of seven PAC species (in addition to *P. subalpina*) to determine if the ORF2 gene is a common feature of *Phialocephala* mt genomes. In amplifying the mitochondrial genomic neighbourhood of ORF2, the region between *coxI* and *nadI*, Duò et al. 2012 determined that the presence of ORF2 is a conserved feature of the PAC. The authors additionally found duplicated *trnC* and *atp9* genes between *coxI* and *nadI* (see mitochondrial genome architecture presented in Fig 4), yet the authors were unable to infer an ancestral state to either of the duplications due to the lack of an informative outgroup. We have

extended the analysis of the region between *cox1* and *nad1* to include *P. scopiformis*, and in particular, have used *P. scopiformis* as an outgroup to identify ancestral *trnC* and *atp9* genes. With an ancestral state for *trnC* and *atp9* genes in mind, we also reconfirmed that the presence/absence of these duplications matches taxonomic divisions in the evolution of the PAC when ORF2 sequences are used for phylogenetic reconstruction. Based on the synteny analysis presented in Fig 4, it is evident that when *P. scopiformis* is used as the basal outgroup [made possible due to its lack of both duplicated genes and its well established position as a basal clade to PAC individuals (Jacobs et al. 2003; Tanney et al. 2016)], the *atp9* gene prior to the conserved ORF2 gene and the *trnC* gene following ORF2 represent the ancestral states of each gene. Mitochondrial gene order thus indicates that the ancestor to the PAC likely obtained an *atp9* duplication prior to divergence of the species complex (see lineage 1 in Fig 4), and that within the PAC, *P. europaea*, *P. letzii* and *P. turicensis* are most closely related based on shared acquisition of a second *trnC* between *cox1* and the ancestral *atp9* (see lineage 2 in Fig 4). When these evolutionary events are compared to neighbour-joining trees reconstructed from ORF2 sequences the same two groupings arise, i.e., PAC species versus *P. scopiformis* (see position (1) in Fig 5) and *P. europaea*, *P. letzii* and *P. turicensis* falling within a monophyletic group (see position (2) in Fig 5). This finding further supports the dendrogram presented by Duò et al. (2012).

APPROXIMATE LOCATION OF FIGURES 4 AND 5

With existing GenBank sequences, we can now begin testing which mitochondrial gene sequences may best resolve species-level identifications within the genus *Phialocephala* as compared to standard nuclear ITS rDNA barcoding

sequence (the accepted DNA barcode region for fungi; Schoch et al. 2012). In this regard, we calculated the percent identities (PI) of interspecific *Phialocephala* groupings for ITS, functional *atp9*, and ORF2. We limited our analyses to these three genes, as these loci are most well-represented in GenBank for *Phialocephala*. Although *rps3* might also be well suited as a future marker within *Phialocephala* due to its high degree of sequence divergence between *P. subalpina* and *P. scopiformis*, there are presently not enough published *rps3* sequences in GenBank to test its suitability as a genetic marker (only *P. subalpina* and *P. scopiformis* are currently available). When the frequency distributions of BLAST percent-identity (PI) scores across species were plotted for ITS, ATP9, and ORF2 the data were bimodal (Fig 6). PAC individuals versus *P. scopiformis* comparisons represent the left-most peaks in the bimodal distributions and within-PAC comparisons represent right-most peaks in the bimodal distributions (Fig 6); this is because within-PAC PI scores are quite similar and closer to 100% while PAC versus *P. scopiformis* PI scores are well beneath 97% (downwards of 75% PI). Based on the spread of the PI scores, mitochondrial ORF2 is currently the better option for both interspecific and within-PAC species identification; this is because the two peaks for ORF2 both fall beneath the conventional 97% PI barcoding threshold for fungi (see spread of data and 97% PI threshold represented by dashed line in Fig 6; for references to 97% threshold see for example Walker et al. 2008 and Bjorbækmo et al. 2010). Using the same logic, both functional mitochondrial *atp9* and nuclear ITS sequences are inadequate for identification of species within PAC (see overlapping maxima for ITS and *atp9* distributions under ‘within PAC comparisons’ header in Fig 6). The nuclear ITS region is better than mitochondrial *atp9* at resolving identification between PAC members and *P. scopiformis*; however, ITS is not as effective as a *Phialocephala*

species-level marker as mitochondrial ORF2 sequence (see spread of peaks under ‘PAC individuals versus *Phialocephala scopiformis* comparisons’ header in Fig 6). Additionally, despite ORF2 presently having no known function, a Z-test of Selection indicated strong statistical support, at $\alpha = 0.01$, for rejecting a null hypothesis of neutrality in favour of an alternate hypothesis supporting purifying selection for ORF2 sequences. For $H_a = d_N \neq d_S$ [departure from neutrality], p-value < 0.01 . For $H_a = d_N < d_S$ [purifying selection], p-value < 0.01 . For $H_a = d_N > d_S$ [positive selection], p-value > 0.01 . Whereby, d_N = nonsynonymous substitutions/nonsynonymous sites, and d_S = synonymous substitutions/synonymous sites (Tamura et al. 2013). The same conclusions were drawn (due to the same returned p-values) when assuming a more complex model (the modified Nei-Gojobori (Jukes-Cantor) model/method) with a fixed transition/transversion bias of 1.63. This bias was calculated specifically for ORF2 sequences using its best-predicted model of nucleotide substitution, T92+G (Tamura et al. 2013).

APPROXIMATE LOCATION OF FIGURE 6

3.4 Conclusion

We reveal conserved mitochondrial gene order among currently sequenced members of the genus *Phialocephala* including members of the PAC, as well as the presence of a 2,138 bp intron in the *cox1* gene of the *P. scopiformis* DAOMC 229536 mt genome. The genus *Phialocephala* serves as a strong case study of how mitochondrial evolutionary events can help resolve evolutionary branches within a fungal species complex. Lastly, our comparative mitochondrial genomics study indicates that mitochondrial genes may be more fruitful in species-level identification for fungal groups in cases where ITS sequence alone lacks resolving power. This last

point is of interest as it implies if more mitochondrial genomes were obtained for strains of *Phialocephala*, strain-specific biomarkers could be developed.

4.0 Acknowledgements

The assistance of J.D. Miller, S. Frasz and B. Green (Carleton University); K. Seifert, E. Ponomareva, H.D.T. Nguyen, M. Balcerzak, T. Ouellet, T. Barasubiye, G. Louis-Seize, and the Microbiology Molecular Technologies Laboratory (Agriculture and Agri-Food Canada, Eastern Cereal and Oilseed Research Centre, Ottawa); and M. Nolan (JGI) is gratefully acknowledged. We also thank J. Tanney, H. Nguyen, S. Redhead and J.D. Miller for comments on an earlier version of the manuscript. This work was supported by NSERC CRDPJ 421782-11 to J. D. Miller and K. A. Seifert, an EYC internship to BMR and AKW from ECO Canada, and by the 1000 Fungal Genomes Project. The initial mt genome assembly was produced by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

5.0 References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.
- Anagnost SE, Richard ME, Hanna RB, 2000. Utilization of soft-rot cavity orientation for the determination of microfibril angle. Part I. *Wood and fiber science: journal of the Society of Wood Science and Technology* **1**: 32.
- Beck N, Lang BF, 2010. MFannot, organelle genome annotation webserver. Université de Montréal, QC, Canada. <http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL, 2005. GenBank. *Nucleic Acids Research* **33**: D34–D38.

- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF, 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**: 313–319.
- Bjorbækmo MFM, Carlsen T, Brysting A, Vrålstad T, Høiland K, Ugland KI, Geml J, Schumacher T, Kauserud H, 2010. High diversity of root associated fungi in both alpine and arctic *Dryas octopetala*. *BMC Plant Biology* **10**: 244.
- Brodie R, Roper RL, Upton C, 2004. JDotter: a Java interface to multiple dotplots generated by dotter. *Bioinformatics* **20**: 279–281.
- Calhoun LA, Findlay JA, Miller DJ, Whitney NJ, 1992. Metabolites toxic to spruce budworm from balsam fir needle endophytes. *Mycological Research* **96**: 281–286.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL, 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Duò A, Bruggmann R, Zoller S, Bernt M, Grünig CR, 2012. Mitochondrial genome evolution in species belonging to the *Phialocephala fortinii* s.l. - *Acephala applanata* species complex. *BMC Genomics* **13**: 166.
- Ghikas DV, Kouvelis VN, Typas MA, 2010. Phylogenetic and biogeographic implications inferred by mitochondrial intergenic region analyses and ITS1-5.8S-ITS2 of the entomopathogenic fungi *Beauveria bassiana* and *B. brongniartii*. *BMC Microbiology* **10**: 174.
- Gray MW, Burger G, Lang BF, 1999. Mitochondrial evolution. *Science* **283**: 1476–1481.
- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I, 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*. **42**: D699–704.
- Grünig CR, Sieber TN, Rogers SO, Holdenrieder O, 2002. Genetic variability among strains of *Phialocephala fortinii* and phylogenetic analysis of the genus *Phialocephala* based on rDNA ITS sequence comparisons. *Canadian Journal of Botany* **80**: 1239–1249.
- Grünig CR, Duò A, Sieber TN, Holdenrieder O, 2008. Assignment of species rank to six reproductively isolated cryptic species of the *Phialocephala fortinii* s.l.-*Acephala applanata* species complex. *Mycologia* **100**: 47–67.
- Hamari Z, Juhász Á, Gácsér A, Kucsera J, Pfeiffer I, Kevei F, 2001. Intron mobility results in rearrangement in mitochondrial DNAs of heterokaryon incompatible *Aspergillus japonicus* strains after protoplast fusion. *Fungal Genetics and Biology* **33**: 83–95.

- Hamari Z, Juhász Á, Kevei F, 2002. Role of mobile introns in mitochondrial genome diversity of fungi. *Acta Microbiologica et Immunologica Hungarica* **49**: 331–335.
- Jacobs A, Coetzee MPA, Wingfield BD, Jacobs K, Wingfield MJ, 2003. Phylogenetic relationships among *Phialocephala* species and other ascomycetes. *Mycologia* **95**: 637–645.
- Juhász Á, Ládai M, Gácsér A, Kucsera J, Pfeiffer I, Kevei F, Hamari Z, 2004. Mitochondrial DNA organisation of the mtDNA type 2b of *Aspergillus tubingensis* compared to the *Aspergillus niger* mtDNA type 1a. *FEMS Microbiology Letters* **241**: 119–126.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.
- Kendrick WB, 1961. The *Leptographium* complex. *Phialocephala* gen. nov. *Canadian Journal of Botany* **39**: 1079–1085.
- Kowalski T, Kehr RD, 1995. Two new species of *Phialocephala* occurring on *Picea* and *Alnus*. *Canadian Journal of Botany* **73**: 26–32.
- Laslett D, Canbäck B, 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**: 172–175.
- Lavín JL, Oguiza JA, Ramírez L, Pisabarro AG, 2008. Comparative genomics of the oxidative phosphorylation system in fungi. *Fungal Genetics and Biology* **45**: 1248–1256.
- Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R, 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research* **43**: W580–W584.
- Lowe TM, Eddy SR, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**: 955–964.
- Miller JD, 2011. Foliar endophytes of spruce species found in the Acadian forest: basis and potential for improving the tolerance of the forest to spruce budworm. In: Frank CA, Pirttilä AM (eds) *Endophytes of forest trees: biology and applications*. Springer, New York. pp. 237–249.
- Miller JD, Sumarah MW, Adams GW, 2008. Effect of a rugulosin-producing endophyte in *Picea glauca* on *Choristoneura fumiferana*. *J Chemical Ecology* **34**: 362–368.

- Paquin B, Laforest MJ, Forget L, Roewer I, Wang Z, Longcore J, Lang BF, 1997. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. *Current Genetics* **31**: 380–395.
- R Core Team, 2015. R: a language and environment for statistical computing. Vienna. <http://www.R-project.org/>.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium, 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* **109**: 6241–6246.
- Seifert KA, Samson RA, deWaard JR, Houbraken J, Lévesque CA, Moncalvo J-M, Louis-Seize G, Hebert PDN, 2007. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Sciences* **104**: 3901–3906.
- Shen X-Y, Li T, Chen S, Fan L, Gao J, Hou C-L, 2015. Characterization and phylogenetic analysis of the mitochondrial genome of *Shiraia bambusicola* reveals special features in the order of Pleosporales. *PLOS One* **10**: e0116466.
- Sumarah MW, Miller JD, 2009. Anti-insect metabolites from foliar fungal endophytes of conifer trees. *Nat Prod Commun* **4**: 1497–1504.
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S, 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725–2729.
- Tanney JB, Douglas B, Seifert KA, 2016. Sexual and asexual states of some endophytic *Phialocephala* species of *Picea*. *Mycologia* **108**: 255–280.
- Taylor GK, Petrucci LH, Lambert AR, Baxter SK, Jarjour J, Stoddard BL, 2012. LAHEDES: the LAGLIDADG homing endonuclease database and engineering server. *Nucleic Acids Research* **40**: W110–W116.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG, 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**: e115–e115.
- Walker JF, Jr OKM, Horton JL, 2008. Seasonal dynamics of ectomycorrhizal fungus assemblages on oak seedlings in the southeastern Appalachian Mountains. *Mycorrhiza* **18**: 123–132.
- Walker AK, Frasz SL, Seifert KA, Miller JD, Mondo SJ, LaButti K, Lipzen A, Dockter RB, Kennedy M, Grigoriev IV, Spatafora JW, 2016. Full genome of *Phialocephala scopiformis* DAOMC 229536, a fungal endophyte of spruce producing the potent anti-insectan compound rugulosin. *Genome Announcements* **4**: e01768–15.

- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ, 2009. Jalview
Version 2—a multiple sequence alignment editor and analysis workbench.
Bioinformatics **25**: 1189–1191.
- Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler
GD, Schriml LM, Sequeira E, Tatusova TA, et al., 2003. Database resources
of the National Center for Biotechnology. *Nucleic Acids Research* **31**: 28–33.
- Wickham H, 2009. ggplot2: elegant graphics for data analysis. Springer-Verlag, New
York. <http://ggplot2.org>.
- Wu Y, Yang J, Yang F, Liu T, Leng W, Chu Y, Jin Q, 2009. Recent dermatophyte
divergence revealed by comparative and phylogenetic analysis of
mitochondrial genomes. *BMC Genomics* **10**: 238.
- Youssar L, Grüning BA, Günther S, Hüttel W, 2013. Characterization and
phylogenetic analysis of the mitochondrial genome of *Glarea lozoyensis*
indicates high diversity within the order Helotiales. *PLOS One* **8**: e74792.

Figure Captions

Fig 1 – Complete mitochondrial genome of *Phialocephala scopiformis* DAOMC 229536. Blue boxes indicate protein-coding genes, turquoise boxes indicate ribosomal DNA genes and red boxes indicate tRNA genes. The direction of arrows indicates transcriptional directionality.

Fig 2 – Conservation of *Phialocephala scopiformis* DAOMC 229536 mitochondrial gene position compared to *Phialocephala subalpina*. (A) gene order of protein coding and rRNA genes for *Phialocephala subalpina* [modified from Duò et al. 2012]. (B) Dotplot of *Phialocephala scopiformis* gene order versus *Phialocephala subalpina* gene order. A diagonal line indicates matching sequence identity; sliding window size = 100bp/100bp.

Fig 3 – Sequence conservation of *Phialocephala* mitochondrial genes. P-distances for both nucleotide and amino acid sequences of *Phialocephala scopiformis* compared to *Phialocephala subalpina* are shown. Pairwise deletion was used to account for gaps and missing datapoints.

Fig 4 – Genome organization of region between *cox1* and *nad1*, comparing *Phialocephala scopiformis* DAOMC 229536 and eight members of the PAC. Evolutionary event ‘1’ represents the duplication of *atp9* and evolutionary event ‘2’ represents the duplication of *trnC*. The lineages in the dendrogram have been collapsed to a polytomy and branch lengths differ only so that the two evolutionary events inferred by the mitochondrial genome architecture may be more clearly displayed.

Fig 5 – Phylogenetic reconstruction using available mitochondrial ORF2 sequences for the genus *Phialocephala*. Tree was produced in MEGA v6 using a Neighbour-joining statistical method and 500 bootstrap replications.

Fig 6 – Frequency distribution of BLAST alignment scores for interspecific comparisons of *Phialocephala* species. Distributions are bimodal with peaks under the blue header representing distribution of percent-identity (PI) scores between *Phialocephala scopiformis* and PAC individuals, and peaks under grey header indicating distribution of PI scores exclusively among PAC sequences.

Table 1. Sequences of duplicated *atp9* and *trnC* mitochondrial genes used in this study, as well as nuclear ITS reference sequences for comparison of *atp9* and ORF2 for *Phialocephala* species-level identification.

Organism	Strain ^a	Sequence ^b [GenBank Accession]
<i>Acephala applanata</i>	T1_50_2	<i>cox1-nad1</i> [JN091449]
	CBS 109321	ITS [NR_119482]
<i>Phialocephala europaea</i>	3_117_2	<i>cox1-nad1</i> [JN091445]
	117-2	ITS [AY347404]
<i>Phialocephala fortinii</i>	7_6_7v	<i>cox1-nad1</i> [JN091448]
	FG2V1	ITS [FN678808]
<i>Phialocephala helvetica</i>	4_123_4	<i>cox1-nad1</i> [JN091446]
	93-315	ITS [AY078142]
<i>Phialocephala letzii</i>	2_120_3	<i>cox1-nad1</i> [JN091444]
	188J14	ITS [KU516548]
<i>Phialocephala subalpina</i>	70-1	Complete mt [JN031566]
	PFO-1	ITS [EF446148]
<i>Phialocephala turicensis</i>	1_124_1	<i>cox1-nad1</i> [JN091443]
<i>Phialocephala uotilensis</i>	5_134_3	<i>cox1-nad1</i> [JN091447]
<i>Phialocephala scopiformis</i>	DAOMC 229536	Complete mt [LKNI00000000]
	CBS 468.94	ITS [NR_119460]

Atp9, ATP synthase subunit 9; *trnC*, transfer RNA cysteine; ITS, internal transcribed spacer; ORF2, open reading frame 2; *cox1*, cytochrome c oxidase subunit 1; *nad1*, NADH dehydrogenase subunit 1; mt, mitochondria.

^a Isolate identifier is provided if GenBank reference lacks strain identifier.

^b *Cox1-nad1* regions contain sequence for mitochondrial ORF2 gene; no GenBank ITS barcode data is presently available for *Phialocephala turicensis* or *Phialocephala uotilensis*.

671
672

673
674

675
676

677
678

679