MOLECULAR ECOLOGY
RESOURCES

Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle

Journal:	*Molecular Ecology Resources*
Manuscript ID:	MER-09-0428.R1
Manuscript Type:	DNA Barcoding
Date Submitted by the Author:	
Complete List of Authors:	Roe, Amanda; University of Alberta, Biological Sciences Rice, Adrianne; University of Alberta, Biological Sciences Bromilow, Sean; University of Alberta, Biological Sciences Cooke, Janice; University of Alberta, Biological Sciences Sperling, Felix; University of Alberta, Biological Sciences
Keywords:	DNA Barcoding, Fungi, molecular diagnostics, Dendroctonus ponderosae, Ophiostomataceae, identification

Title: Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle

Authors: Amanda D Roe[1], Adrianne V Rice[1], Sean E Bromilow[1], Janice EK Cooke[1], and Felix AH Sperling[1].
[1] CW 405 – Department of Biological Sciences
University of Alberta
Edmonton, AB, CANADA
T6G 2E9

Corresponding Author:
Amanda D Roe
CW 405 – Dept. Biological Sciences
University of Alberta
Edmonton, AB, CANADA
T6G 2E9
amandaroe5@gmail.com
Fax: 780.492.9234

Keywords: DNA barcoding, *Dendroctonus ponderosae*, Ophiostomataceae, identification, molecular diagnostics, fungi

Running Title: Fungal DNA barcoding

Abstract:
There is strong community-wide interest in applying molecular techniques to fungal species delimitation and identification, but selection of a standardized region or regions of the genome has not been finalized. A single marker, the ribosomal DNA internal transcribed spacer region (ITS), has frequently been suggested as the standard for fungi. We used a group of closely related blue stain fungi associated with the mountain pine beetle (*Dendroctonus ponderosae* Hopkins) to examine the success of such single locus species identification, comparing ITS to four other nuclear markers. We demonstrate that single loci varied in their utility for identifying the six fungal species examined, while use of multiple loci was consistently successful. In a literature survey of 21 similar studies, individual loci were also highly variable in their ability to provide

consistent species identifications and were less successful than multilocus diagnostics. Accurate species identification is the essence of any molecular diagnostic system and this consideration should be central to locus selection. Moreover, our study and the literature survey demonstrate the value of using closely related species as the proving ground for developing a molecular identification system. We advocate use of a multilocus barcode approach that is similar to the practice employed by the plant barcode community, rather than reliance on a single locus.

Introduction

Accurate species identification and delimitation is vital to understanding our natural world. Errors can have far-reaching consequences, impacting biodiversity assessment, ecological studies, and management decisions (Bortolus 2008). Less than 10% of an estimated 1.5 million fungal species have been described (Hawksworth 2001), impeding our understanding of fungal diversity and ecology. Closely related fungi can differ in their pathological effects, toxicogenic products, beneficial attributes, and ecological niches (Bleiker & Six 2007; Degenkolb *et al.* 2007; Lee *et al.* 2006b). Morphological identification of such closely related fungi has remained elusive due to the scarcity and ambiguity of diagnostic characters. Furthermore, morphology-based identifications are impossible for many species since they cannot be cultured.

To counter difficulties in identification, use of DNA sequences has become increasingly popular for species delimitation and identification (e.g. Anderson & Cairney 2007; Giraud *et al.* 2008; Shenoy *et al.* 2007), allowing for the application of genealogical concordance phylogenetic species recognition (Taylor *et al.* 2000). Various diagnostic tools based on sequence variation have been developed for important groups of fungi (Druzhinina *et al.* 2005; Geiser *et al.* 2004; Kopchinskiy *et al.* 2005), but lack standardization. DNA barcoding (Hebert *et al.* 2003) is a currently common sequence-based approach that generally uses a single standard region of DNA to identify and recognize distinct animal species, but has received a wide range of responses among taxonomists (Janzen 2004; Sperling 2003; Tautz *et al.* 2002; Will *et al.* 2005)

Although there is now wide acceptance of the need for standardization of gene regions used in taxonomic studies (Caterino *et al.* 2000), it has proven more difficult to determine which particular gene regions and sequence length are most informative (Roe & Sperling 2007a, b; Rubinoff *et al.* 2006). Initially, a 658 bp segment of mitochondrial DNA from the cytochrome c oxidase gene (COI) was proposed as the single standard DNA barcoding region (Hebert *et al.* 2003). However, it has been demonstrated that COI is suboptimal for some groups of organisms, including fungi (Seifert 2009; Seifert *et al.* 2007) and plants (Chase *et al.* 2005). To develop a standardized molecular identification system like DNA barcoding in these groups, other genomic regions have been explored and suggested in place of COI (CBOL Plant Working Group *et al.*, 2009; Rossman 2007; Seifert 2009). Recently, application of DNA barcoding to fungi figured prominently in a special issue in Molecular Ecology Resources, including by Seifert (2009) who encouraged adoption of the internal transcribed spacer regions 1 and 2 (ITS) as the standard fungal DNA barcoding gene region.

Accurate species identification is the essence of any molecular diagnostic system (Erickson *et al.* 2008) and potential genomic regions should be tested rigorously prior to adoption in a standardized system. Using five independent gene regions (elongation factor 1 alpha, beta tubulin, actin, internal transcribed spacer 2 + 28S rDNA, and an anonymous nuclear locus), we examine the accuracy of each marker and combination of markers for species delimitation and identification in ophiostomatoid fungi associated with the mountain pine beetle (MPB; *Dendroctonus ponderosae* (Hopkins)). MPB is an eruptive forest pest,

attacking and feeding on pines across western North America. Western Canada is experiencing the largest outbreak on record, with over 7.8 million hectares of lodgepole pine (*Pinus contorta* Douglas var. *latifolia* Engelmann) and lodgepole x jack pine hybrid (*P. contorta* Douglas var. *latifolia* Engelmann x *P. banksiana* Lamb.) forest affected to date (British Columbia Ministry of Forests – www.for.gov.bc.ca/hfp/health/overview/2008table.htm, accessed Sept 01, 2009; Alberta Sustainable Resources Development – http://www.mpb.alberta.ca/Resources/regionalmaps.aspx, accessed Oct 16, 2009). MPB has several fungal symbionts (Lee *et al.* 2005, 2006a; Robinson 1962; Rumbold 1941; Six & Bentz 2003; Whitney & Farris 1970), and we focus on four closely related blue stain fungi (Ophiostomataceae): *Grosmannia clavigera* (Robinson-Jeffrey and Davidson) Zipfel, de Beer and Wingfield, *G. aurea* (Robinson-Jeffrey and Davidson) Zipfel, de Beer and Wingfield, *Leptographium longiclavatum* Lee, Kim and Breuil, and *L. terebrantis* Barras and Perry, collectively termed the *Grosmannia* clade (Zipfel *et al.* 2006). These fungi can be tree pathogens (Lee *et al.* 2006b; Owen *et al.* 1987; Reid *et al.* 1967; Rice *et al.* 2007; Stobel & Sugarwara 1986), aid in host suppression (Raffa & Berryman 1983), and provide nutrition to the beetle (Adams & Six 2007; Bentz & Six 2006; Six & Paine 1998). Species identification and delimitation in the *Grosmannia* clade is difficult due to interspecific overlap of morphological traits (Lim *et al.* 2004). Previous work on this group of fungi (Alamouti *et al.* 2009; Lee *et al.* 2005; Lim *et al.* 2004; Zipfel *et al.* 2006) has laid the foundation for an in

depth exploration of species identification and delimitation in MPB associated fungi using dense spatial and temporal sampling.

By comparing multiple independent loci we examined the contribution of independent character sets to the identification and delimitation of blue stain fungi associated with MPB. To relate our results to the larger body of ascomycete literature and assess the success of single- versus multilocus species identification, we surveyed studies that used multiple, independent loci to delimit closely related species. We then contrasted single and multilocus fungal species identification, examining the interaction between increased molecular sampling and successful species identification.

METHODS

Field Sampling

Sampling focused on pine stands in areas experiencing recent mountain pine beetle attacks in Alberta and eastern British Columbia (Fig. 1). Collections were made over two time periods: Jan-May 2007 (M001) and Sept 2007-May 2008 (M002). A 10 cm hole-saw and chisel were used to obtain pine discs 10 cm in diameter and approximately 3 cm thick, containing both xylem and bark with beetle infested phloem. Discs were placed in ziplock bags and transported to the laboratory on ice for processing.

Fungal Culturing

Fungal cultures were obtained, following Rice & Langor (2009), directly from adult and larval MPB individuals or from wood samples taken from beetle galleries. Live individuals were allowed to crawl on malt extract agar (MEA) for up to 24h, while dead individuals were wiped across the surface of the media. Fungi were allowed to grow for 2-5 days at room temperature (RT) and ambient light before subculturing onto MEA. Subcultures were incubated at RT for 2-4 weeks then scored for cultural and microscopic traits and compared to previously authenticated strains by A. Rice and grouped according to colony morphology ('morphotype'). Representative cultures were deposited in the University of Alberta Microfungus Collection and Herbarium (UAMH), with unique identifiers listed in Appendix 1 and images of strains deposited on MorphoBank project number P327 (www.morphobank.org).

DNA extraction, PCR, and sequence analysis

Cultures were selected at random for multilocus typing within each collecting locality for each morphotype using a random number generator. Cultures with unusual morphology were also selected for typing. Single spore isolates (SSI) were grown for each selected strain.

Fungal tissue (~25 mg) was harvested for DNA extraction from SSI grown on 1.5% MEA covered with a cellophane membrane and ground in liquid nitrogen. DNA was extracted using a CTAB protocol modified from Chang *et al.* (1993). CTAB buffer (700 µL) was added to each sample, followed by 10 uL of Proteinase K (600 mAU/mL, Qiagen). Samples were incubated at 65˚C for 1 hr,

159 then cooled to room temperature (RT) before adding 600 µL 24:1 chloroform:
160 isoamyl alcohol. Tubes were mixed vigorously then centrifuged for 5 min (all
161 centrifugation steps performed at 17000 x g and RT). The aqueous supernatant
162 was mixed with 600 µL isopropanol, and chilled at -20˚C for at least 2 hours.
163 Samples were centrifuged for 15 min, supernatants were discarded, then 500 µL
164 95% ethanol (v/v) added to the pellet, which was vortexed briefly and centrifuged
165 for 3 min. This process was repeated with 500 µL 70% ethanol (v/v). Pellets
166 were re-suspended in 100 µL Milli-Q water (Millipore, Billerica MA) with gentle
167 agitation.

168 Five gene regions were amplified: actin, elongation factor 1 alpha (EF1a),
169 beta tubulin (Btub), an anonymous nuclear locus (UFM), and ITS2 (partial 5.8S +
170 internal transcribed spacer 2 + partial 28S). Primers used to amplify the above
171 regions as well as their corresponding PCR profiles are listed in Table 1. Initially,
172 amplification of the ITS region with universal primers resulted in non-specific
173 amplification as well as impediments to sequencing due to poly-A repeats in the
174 ITS1 region. In fact, the UFM locus was obtained as a nonspecific product while
175 trying to amplify ITS2 using primers ITS3+LR3 (Lim *et al.* 2004; Vilgalys &
176 Gonzalez 1990; White *et al.* 1990). ITS2 was eventually specifically amplified
177 using primers ITS3+TW13 (Table 1). The non-ITS identity of the UFM locus was
178 confirmed by comparison with available ITS sequence and alignment with the *G.*
179 *clavigera* genome sequence (whole genome shotgun sequence, GenBank
180 ACXQ01001232.1, contig 29.6, 42669-43146 bp, Diguistini *et al.* 2009). Locus-
181 specific internal primers were then designed to amplify the UFM region (Table 1).

PCR reactions for all regions were performed in 25 µL final volume. Each reaction contained 50 ng DNA, 0.1 µM of each primer, 1X modified Thermopol buffer (M0273S: New England BioLabs, Pickering ON) with BSA fraction V added to the buffer to give a final concentration of 1 ng/µL BSA, 1.25 U Taq DNA Polymerase (New England BioLabs, Pickering ON), 200 µM each dNTP, 1% DMSO (v/v) and 4 mM $MgCl_2$. In the case of ITS amplification, 2 mM $MgSO_4$ was used in place of 4 mM $MgCl_2$.

PCR products were purified in 10 µL reactions using an EXO-SAP reaction with Exonuclease I and Shrimp Alkaline Phosphatase (70073Z and 70092Y; USB Corporation, Cleveland OH). Purified PCR products were sequenced with ABI Big Dye Terminator v 3.1 cycle sequencing reaction kit (Applied Biosystems, Foster City CA) and purified using an ethanol/EDTA/sodium acetate precipitation protocol (ABI Big Dye Terminator v3.1 cycle sequencing kit protocol manual, Applied Biosystems). Sequencing was performed on an Applied Biosystems 3730 DNA Analyzer. Initially, all reactions were sequenced bi-directionally to identify common haplotypes. Following the initial survey, reactions were sequenced uni-directionally to streamline data collection, with all new haplotypes confirmed by bi-directional sequencing. Sequence data and associated chromatograms for unique multilocus haplotypes were submitted to GenBank and the NCBI Trace Archive under accession numbers GU370130-GU370344.

Phylogenetic Analyses

Parsimony Haplotype Networks

Sequence data were initially aligned in Sequencher 4.8 (Gene Codes, Ann Arbor, MI) with manual adjustments made by eye. Representative GenBank sequences from previous studies of closely related members of the *Grosmannia* clade (*G. clavigera, G. aurea, G. robusta, L. longiclavatum, L. terebrantis,* and *L. pyrinum*) were included (Appendix 2). These representatives helped to confirm morphotype identifications and to identify species not diagnosed during morphotyping. Unique haplotypes were determined for each individual locus and a concatenated multilocus data set, and haplotype networks were calculated using TCS 1.21 (Clement *et al.* 2000), which infers a haplotype network using a statistical parsimony framework (Templeton 1998). Gaps were treated as missing data and the connection limit was set to 95%.

Neighbor-Joining

Following sequence alignment, a neighbor-joining tree was produced using PAUP* v. 4.0b10 for each gene region and the concatenated multilocus data set. Two additional taxa were used as outgroups (*L. koreanum* Kim and Kim, and *L. yunnanensis* Zhou, Jacobs, Wingfield, and Morelet). Uncorrected “p” distances were used to generate pairwise distances and missing data (including gaps) were distributed proportionally to unambiguous changes.

Maximum Likelihood

Following sequence alignment, a maximum likelihood (ML) tree was estimated for each gene region and the multilocus data set. The ML trees were calculated using a maximum likelihood framework as implemented in RaxML v. 7.0.4 (Stamatakis 2006) hosted by CIPRES portal v. 1.0 (Cyberinfrastructure for Phylogenetic Research - www.phylo.org/portal/Home.do, accessed June 30, 2009). RAxML, or "randomized axelerated maximum likelihood" for high performance computing, uses a simultaneous optimized ML tree-space search method and rapid bootstrapping algorithm (Stamatakis *et al.* 2008) that allows a full ML analysis in a single run. Matrices and associated ML trees for each individual locus and the multilocus data set were submitted to TreeBASE (www.treebase.org; accession number SN4846).

Multilocus Power Analysis

To determine whether species identification is improved by the addition of data for multiple genes, we compared the congruence of species limits in single locus data sets to successively larger multilocus data sets. These comparisons were conducted on three data sets that consisted of several closely related individuals which had sequence data for five independent loci: the present study (four species), *Mycosphaerella* (five species, Crous *et al.* 2004), and *Cercospora* (three species, Groenewald *et al.* 2005). For each study, we produced NJ trees for each individual locus, as well as all combinations of two, three, four, and five loci. The average proportion of species identification success was calculated for

each combination. Congruence with species limits was scored as in the literature survey below.

Literature Survey of Multilocus Species Identification

To relate our study to broader patterns of single locus versus multilocus species identification in the mycological literature, we used the literature survey approach described by Roe and Sperling (2007b). We selected fungal studies that examined at least two independent loci and sampled at least 5 individuals from two closely related species. For inclusion, studies had to provide results from single locus analyses (e.g. phylogenetic or distance trees) either in the publication or in TreeBASE. To summarize each study, loci were grouped into different character classes (ribosomal, mtDNA, autosomal, and unknown), and their haplotype fixation and phylogenetic/cluster congruence were compared to the multilocus species delimitations preferred by the authors. Haplotype fixation was defined as either fixed (species do not share haplotypes) or shared (species share haplotypes). Congruence with species limits was scored based on the type of analysis used. Phylogenetic results (parsimony, likelihood, Bayesian) were scored as reciprocally monophyletic (RM), paraphyletic (PA), or polyphyletic (PO). Distance or cluster-based methods (neighbor-joining) were scored as congruent (CO) or non-congruent (NC) with species limits.

RESULTS

Morphotyping

A total of 5063 strains were isolated, morphotyped, and preserved for future examination from 45 stands in Alberta and eastern British Columbia (Fig. 1). Strains were initially identified as three common MPB fungal associates: *G. clavigera* (n=2162), *L. longiclavatum* (n=675), and *Ophiostoma montium* (Rumbold) von Arx (n=1991) based on diagnostic characters from original species descriptions (Lee *et al.* 2005; Robinson-Jeffery & Davidson 1968; Rumbold 1941), as well as via comparisons with previously authenticated strains. Some strains appeared to be morphologically intermediate between *G. clavigera* and *L. longiclavatum* and were identified as 'intermediate' (n=235). A subset of morphotyped strains was selected for multilocus typing.

Phylogenetic Analyses

For the remainder of this study, we focused on strains belonging to the *Grosmannia* clade (*G. clavigera*, *L. longiclavatum*, and intermediate strains). *Ophiostoma montium* was clearly genetically divergent from the *Grosmannia* clade, although not always morphologically distinct, and phylogenetic analysis for this species will be reported elsewhere. In total, 350 SSI from the *Grosmannia* clade were selected for multilocus characterization (see Appendix 1 for a complete list of all cultures and collecting localities). In addition to these strains, sequence data were included for six previously published representative species within the *Grosmannia* clade and two outgroup species (Appendix 2).

Five loci (actin, EF1a, Btub, UFM, and ITS2) were amplified for all isolates. Unique haplotypes for each locus were selected for phylogenetic

295 analysis, treating insertions and deletions as missing data. For each locus and
296 the combined multilocus data set, parsimony haplotype networks (Figs. 2 and 3),
297 maximum likelihood (ML) and neighbor-joining (NJ) trees (Figs. 4 and 5) were
298 produced. A summary of the phylogenetic data and ML model parameters for
299 each data set is presented in Table 2.
300
301 Single Locus Analyses
302 Generally, similar patterns of tree topology and species limits were found
303 for each locus, with the exception of ITS2. ITS2 had several haplotypes that
304 were shared between species (*L. terebrantis, L. pyrinum*, *G. robusta,* and *G.*
305 *aurea,* Fig. 2), while no haplotypes were shared between species for the
306 remaining loci (actin, EF1a, Btub, and UFM). Monophyly of individual species
307 was well supported where species were represented by multiple haplotypes, with
308 the exception of species sharing ITS2 haplotypes. There was no evidence of
309 paralogous ITS copies or interspecific hybridization, suggesting that shared ITS
310 haplotypes resulted from incomplete lineage sorting. *Leptographium terebrantis*
311 was paraphyletic in actin, EF1a, and Btub, and *G. clavigera* was paraphyletic in
312 UFM. While monophyly of individual species was generally well-supported,
313 internal nodes within and between species received little bootstrap support.
314 Monophyly of the *Grosmannia* clade was well supported for each locus (Fig. 4),
315 although this could not be assessed for UFM because sequences were
316 unavailable for outgroup taxa.
317

318 Multilocus Analyses

319 Minimal topological conflict existed among supported clades in each
320 single-locus analysis; accordingly, the five data sets were combined into a single
321 concatenated data set. No haplotypes were shared between species (Fig. 3),
322 similar to the individual loci, although contrasting with the results obtained for
323 ITS2 (Fig. 2). *G. clavigera*, *L. longiclavatum, L. terebrantis*, and *G. aurea* were all
324 represented by multiple haplotypes and were monophyletic, with strong support
325 for all except *L. terebrantis*. As with the single locus analyses, internal nodes
326 within and between species received little bootstrap support. Monophyly of the
327 *Grosmannia* clade was also strongly supported by the concatenated data set
328 (Fig. 5).

329

330 **Multilocus Power Analysis**

331 Using three multilocus data sets, including the one from our study, we
332 examined how species identification success changed with the inclusion of
333 additional loci. Based on single loci, average proportion of identification success
334 ranged from 0.4 to 0.7 (40 – 70%) for *Grosmannia, Mycosphaerella,* and
335 *Cercospora* species. The proportion of successful identification increased as
336 additional loci were added (Fig. 6). Eighty percent success was attained in the
337 *Grosmannia* data set using as few as two loci, while four loci were needed for
338 both *Mycosphaerella* and *Cercospora*. Furthermore, the addition of more loci
339 resulted in an increase in identification success to 100% in both the *Grosmannia*

and *Mycosphaerella* with three and five loci respectively, while *Cercospora* never surpassed 80% success, even with the entire five locus data set.

Literature Survey

We obtained multilocus data for 22 closely related species pairs of ascomycetes from 21 studies (Appendix 2). Species pairs occurred in six orders: Ophiosomatales (two pairs), Hypocreales (eight pairs), Capnodiales (five pairs), Calosphaeriales (one pair), Eurotiales (two pairs), Botryosphaeriales (three pairs), and one unplaced Sordariomycetidae genus. This was not an exhaustive review but we believe that it represents the overall patterns in the literature. Results of this survey are summarized in Table 3 and presented in greater detail in Appendix 3. Gene regions examined in the studies fall into 5 major categories: ribosomal (ITS1, 5.8S, ITS2, 28S, intergenic spacer region), mtDNA (mitochondrial small subunit rDNA, adenosine triphosphatase 6), autosomal (actin, beta tubulin, elongation factor 1 alpha, calmodulin, histone H3, RNA polymerase II subunit, chit18-5), anonymous (UFM, BotF15), and amplified fragment length polymorphisms.

A total of 77 loci were examined across the studies examined, with two to six loci per species pair (Table 3, Appendix 3). Excluding amplified fragment length polymorphisms, 60 loci showed fixed haplotype differences between species pairs, while 16 loci had shared or partially shared haplotypes between species. All loci with shared or partially shared haplotypes (16 loci) were polyphyletic (PO) or non-congruent (NC). Of the loci with fixed haplotype

differences, the majority of haplotypes were reciprocally monophyletic (RM – 28 loci) or congruent (C – 22 loci) with species limits. The remaining fixed loci (10 loci) were paraphyletic (PA) or non-congruent (NC). Of the surveyed studies, the majority of loci fall into two locus types, ribosomal (20 loci) and autosomal loci (52 loci). Proportion of haplotype fixation was higher in autosomal loci (46 loci - 88%), compared to rDNA loci (11 loci - 55%) (Table 3, Appendix 3). Of the loci with fixed haplotype differences, levels of reciprocal monophyly and congruence were similar between autosomal loci (87%) and rDNA loci (82%).

DISCUSSION

Accurate identification and delimitation of closely related species should be the proving ground for any diagnostic tool designed to elucidate biodiversity (Hollingsworth 2008; Roe & Sperling 2007b; Sperling 2003). Boundaries between closely related species are often indistinct, lacking clear limits and diagnostic characters (Bickford *et al.* 2007; Sites & Marshall 2004), due to processes such as introgressive hybridization and lineage sorting (Funk & Omland 2003). These evolutionary phenomena can lead to discordance between species limits and diagnostic traits, which are especially problematic - and undetectable - when a single character set is used. Currently many molecular identification approaches, such as DNA barcoding, rely on a single standardized fragment of DNA for species identification, delimitation, and discovery (Hebert *et al.* 2003; Rubinoff *et al.* 2006).

Single locus delimitation and identification

The ideal barcode identification system recognizes distinct species, has universal primers, and reliably produces high quality sequence (Erickson *et al.* 2008). The fungal community is currently working toward a standardized molecular identification system to improve the identification, delimitation, and discovery of cryptic fungal biodiversity. For example, the All Fungi Barcode Initiative (FBI) (Rossman 2007) has promoted the adoption of ITS as the sole fungal barcode region, a recommendation also followed by Seifert (2009). ITS, in many respects, appears to fit the criteria of an ideal molecular marker. Given its successful usage in a range of fungal groups (e.g. Druzhinina *et al.* 2005; Feau *et al.* 2009; Iwen *et al.* 2002; Koljalg *et al.* 2005; Le Gac *et al.* 2007; Seifert *et al.* 2007; Shenoy *et al.* 2007; Summerbell *et al.* 2007), universal primers (White *et al.* 1990), manageable sequence length, high copy number, and phylogenetic utility (Coleman 2003; Kohn 1992), selecting ITS as the molecular marker for fungi initially seems reasonable.

In addition to ITS, other loci have been suggested as potential molecular diagnostic markers. Mitochondrial genes, such as COI, work well for certain groups (Seifert *et al.* 2007; Vialle *et al.* 2009). Since a database of COI sequence already exists for animals (Ratnasingham & Hebert 2007), there is a strong advantage to using the same region to avoid a 'Tower of Babel' (Caterino *et al.* 2000). Single/low copy nuclear genes have also been suggested (Geiser *et al.* 2004; Lieckfeldt & Seifert 2000). These types of loci can have a number of advantages, including easier alignment, fewer indels, more variable 3rd codon

positions, and known copy number (Bruns & Shefferson 2004). Of these, variability is the most important advantage, as indicated in our literature survey.

Despite their various positive attributes, each marker system has potential negative issues and no single marker is perfect. For example, low variability in ITS has resulted in failure to separate closely related species (Appendix 2, Bruns & Shefferson 2004; Crouch *et al.* 2009; Landis & Gargas 2007; Lieckfeldt & Seifert 2000; Varga *et al.* 2007), which is where identifications matter most (Hollingsworth 2008; Sperling 2003). As well, ITS may have multiple paralogous copies (O'Donnell & Cigelnik 1997; Simon & Weiss 2008), leading to inaccurate identifications, or variation in copy number (Pukkila & Skrzynia 1993; Vilgalys & Gonzalez 1990) , preventing its use as a quantitative marker. Problems with COI have also been well documented, including problems with multiple paralogous copies (Gilmore *et al.* 2009), low species resolutions, mobile introns, and limited historical usage (for a detailed discussion see Seifert 2009). Single/low copy nuclear genes, while more variable, experience rapid nucleotide saturation and may be more difficult to amplify in degraded material (Bruns & Shefferson 2004). Development of universal primer regions is important for standardization of a molecular identification system, which can be more problematic in nuclear genes. Universality problems can also occur with ITS (Taylor & McCormick 2008), as we saw in our study. Non-orthologous and inconsistent amplification of universal regions may cause missing or inaccurate sequence data, leading to errors in species identification.

Multilocus species delimitation and identification

Our data, and to an even greater extent our survey of the literature, clearly demonstrate the limitations of relying on a single locus for identifying and inferring species limits among closely related fungal species. If we had relied on a single locus in this study, such as ITS, we would have underestimated fungal species diversity (Table 3, Appendix 3). This is similar to the findings of Seifert et al. (2007) where beta tubulin, a single-copy nuclear locus, provided diagnostic barcodes for 80% of the *Penicillium* species examined, outperforming both ITS (25%) and COI (66%). Such failure to separate closely related species occurred in every type of gene class, with nuclear genes also being susceptible to incongruence. However, the inclusion of additional loci in a multilocus framework helps to minimize the risk of inaccurate species identification, increasing the proportion of successful species identifications relative to single locus data sets. The number of additional loci needed to achieve a high proportion of successful identifications varied between groups.

Demonstration of the improvement in species delimitation and identification that can be obtained with multilocus data sets has prompted recommendations to move towards an integrative multilocus approach (Roe & Sperling 2007b; Rubinoff *et al.* 2006). A recent example of a newly developed multilocus diagnostic system is the 2-locus plant barcode proposed by the CBOL Plant Working Group *et al.* (2009). Rigorous testing of chloroplast gene regions in this study demonstrated that two or three regions are optimal, balancing universality, resolution, and consistency. Using the plant barcode as a benchmark, we

strongly support the adoption of a similar multilocus fungal identification system. Other studies with fungi have also demonstrated that multiple loci provide the most accurate species identification and delimitation (Balajee *et al.* 2007; Dettman *et al.* 2006; Petit & Excoffier 2009; Summerbell *et al.* 2005). Despite this, much of the fungal literature relies on a single locus (often ITS) and does not include additional characters to independently evaluate lineage diversity. It is difficult to know when additional characters are needed to clarify species limits since species-level variability is not predictable (Nilsson *et al.* 2008). This unpredictability makes it nearly impossible to reliably include additional characters using only a targeted approach to counter single locus invariability (Rossman 2007; Seifert 2009).

Conclusions

Any method that aims to achieve accurate species identification and delimitation should be capable of delimiting closely related species and identifying gene tree versus species tree discordance (Fazekas *et al.* 2008; Funk & Omland 2003; Lahaye *et al.* 2008; Olmstead 1995). Examination of multiple, independent character sets in an integrative taxonomic framework (Roe & Sperling 2007b; Rubinoff *et al.* 2006) can minimize the risk of inaccurate species delimitation by inferring species boundaries through congruence among multiple character sets. With multiple character sets, a genealogical concordance phylogenetic species concept (Taylor *et al.* 2000) could provide a means to detect discordance between gene trees and species trees in an identification

framework. Discordance can be particularly problematic for closely related species as their lineages may have not had sufficient time to coalesce, compared to species with longer, independent, evolutionary histories. Ultimately, the development of a molecular identification system like DNA barcoding is a compromise between universality, resolution, efficiency, and consistency (Erickson *et al.* 2008; Vialle *et al.* 2009). Fungal biodiversity vastly exceeds that of many other taxa, so extensive testing will be needed to evaluate potential combinations of loci.

Acknowledgements

We wish to thank all the field and laboratory technicians who obtained and processed the fungal samples. We especially thank Dr. Patricia Crane, who developed protocols for and cultured the single spore isolates used in this study. This study is dedicated to Pat. We are grateful to Richard Hamelin, Clement Tsui, Sepideh Alamouti, Colette Breuil, and David Coltman for supportive collaboration and advice throughout this project. We are also grateful for the helpful comments and suggestions provided by two anonymous reviewers. This research was carried out as part of The TRIA Project (www.thetriaproject.ca). Funding for this research has been provided by the Government of Alberta through Genome Alberta, Alberta Advanced Education and Technology, and Alberta Sustainable Resources Development.

References

Adams AS, Six DL (2007) Temporal variation in mycophagy and prevalence of fungi associated with developmental stages of *Dendroctonus ponderosae* (Coleoptera : Curculionidae). *Environmental Entomology* **36**, 64-72.

Alamouti SM, Tsui CK, Breuil C (2009) Multigene phylogeny of filamentous ambrosia fungi associated with ambrosia and bark beetles. *Mycological Research* **113**, 822-835.

Anderson IC, Cairney JW (2007) Ectomycorrhizal fungi: exploring the mycelial frontier. *FEMS Microbiology Review* **31**, 388-406.

Balajee SA, Houbraken J, Verweij PE*, et al.* (2007) *Aspergillus* species identification in the clinical setting. *Studies in Mycology* **59**, 39-46.

Bentz BJ, Six DL (2006) Ergosterol content of fungi associated with *Dendroctonus ponderosae* and *Dendroctonus rufipennis* (Coleoptera : Curculionidae, Scolytinae). *Annals of the Entomological Society of America* **99**, 189-194.

Bickford D, Lohman DJ, Sodhi NS*, et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22**, 148-155.

Bleiker KP, Six DL (2007) Dietary benefits of fungal associates to an eruptive herbivore: potential implications of multiple associates on host population dynamics. *Environmental Entomolology* **36**, 1384-1396.

Bortolus A (2008) Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio* **37**, 114-118.

Bruns TD, Shefferson RP (2004) Evolutionary studies of ectomycorrhizal fungi: recent advance and future directions. *Canadian Journal of Botany* **82**, 1122-1132.

Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: a thriving Tower of Babel. *Annual Review of Entomology* **45**, 1-54.

CBOL Plant Working Group, Hollingsworth PM, Forrest LL*, et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12794-12797.

Chang S, Puryear J, Cairney JW (1993) A simple method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**, 113-116.

Chase MW, Salamin N, Wilkinson M*, et al.* (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1889-1895.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics* **19**, 370-375.

Crouch JA, Clarke BB, Hillman BI (2009) What is the value of ITS sequence data in *Colletotrichum* systematics and species diagnosis? A case study using the falcate-spored graminicolous *Colletotrichum* group. *Mycologia* **101**, 648-656.

Crous PW, Groenewald JZ, Pongpanich K*, et al.* (2004) Cryptic speciation and host specificity among *Mycosphaerella* spp. occurring on Australian

Acacia species grown as exotics in the tropics. *Studies in Mycology* **50**, 457-469.

Degenkolb T, Kirschbaum J, Bruckner H (2007) New sequences, constituents, and producers of peptaibiotics: an updated review. *Chemistry & Biodiversity* **4**, 1052-1067.

Dettman JR, Jacobson DJ, Taylor JW (2006) Multilocus sequence data reveal extensive phylogenetic species diversity within the *Neurospora discreta* complex. *Mycologia* **98**, 436-446.

Diguistini S, Liao NY, Platt D*, et al.* (2009) De novo genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biology* **10**, R94.91-12.

Druzhinina IS, Kopchinskiy AG, Komon M*, et al.* (2005) An oligonucleotide barcode for species identification in *Trichoderma* and *Hypocrea*. *Fungal Genetics and Biology* **42**, 813-828.

Erickson DL, Spouge J, Resch A, Weigt LA, Kress WJ (2008) DNA barcoding in land plants: developing standard to quantify and maximize success. *Taxon* **57**, 1304-1316.

Fazekas AJ, Burgess KS, Kesanakurti PR*, et al.* (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* **3**, e2802.

Feau N, Vialle A, Allaire M*, et al.* (2009) Fungal pathogen (mis-) identifications: a case study with DNA barcodes on *Melampsora* rusts of aspen and white poplar. *Mycological Research* **113**, 713-724.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics* **34**, 397-423.

Geiser DM, Jiménez-Gasco MdM, Kang S*, et al.* (2004) FUSARIUM-ID v. 1.0: A DNA sequence database for identifying *Fusarium*. *European Journal of Plant Pathology* **110**, 473-479.

Gilmore S, Gräfenhan T, Louis-Seize G, Seifert KA (2009) Multiple copies of cytochrome oxidase 1 in species of the fungal genus *Fursarium*. *Molecular Ecology Resources* **9**, 90-98.

Giraud T, Refregier G, Le Gac M, de Vienne DM, Hood ME (2008) Speciation in fungi. *Fungal Genetics and Biology* **45**, 791-802.

Groenewald M, Groenewald JZ, Crous PW (2005) Distinct species exist within the *Cercospora apii* morphotype. *Phytopathology* **95**, 951-959.

Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited*. *Mycological Research* **105**, 1422-1432.

Hebert PD, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceeding of the Royal Society B: Biological Sciences* **270 Suppl 1**, S96-99.

Hollingsworth PM (2008) DNA barcoding plants in biodiversity hot spots: progress and outstanding questions. *Heredity* **101**, 1-2.

Iwen P, Hinrichs S, Rupp M (2002) Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. *Medical Mycology* **40**, 87-109.

Janzen DH (2004) Now is the time. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 731-732.

Kohn LM (1992) Developing new characters for fungal systematics - an experimental approach for determining the rank of resolution. *Mycologia* **84**, 139-153.

Koljalg U, Larsson KH, Abarenkov K*, et al.* (2005) UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* **166**, 1063-1068.

Kopchinskiy A, Komon M, Kubicek CP, Druzhinina IS (2005) TrichoBLAST: a multilocus database for *Trichoderma* and *Hypocrea* identifications. *Mycological Research* **109**, 658-660.

Lahaye R, van der Bank M, Bogarin D*, et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2923-2928.

Landis FC, Gargas A (2007) Using ITS2 secondary structure to create species-specific oligonucleotide probes for fungi. *Mycologia* **99**, 681-692.

Le Gac M, Hood ME, Fournier E, Giraud T (2007) Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. *Evolution* **61**, 15-26.

Lee S, Kim JJ, Breuil C (2005) *Leptographium longiclavatum* sp nov., a new species associated with the mountain pine beetle, *Dendroctonus ponderosae*. *Mycological Research* **109**, 1162-1170.

Lee S, Kim JJ, Breuil C (2006a) Diversity of fungi associated with the mountain pine beetle, *Dendroctonus ponderosae* and infested lodgepole pines in British Columbia. *Fungal Diversity* **22**, 91-105.

Lee S, Kim JJ, Breuil C (2006b) Pathogenicity of *Leptographium longiclavatum* associated with *Dendroctonus ponderosae* to *Pinus contorta*. *Canadian Journal of Forest Research* **36**, 2864-2872.

Lieckfeldt E, Seifert KA (2000) An evaluation of the use of ITS sequences in the taxonomy of the Hypocreales. *Studies in Mycology* **45**, 35-44.

Lim YW, Alamouti SM, Kim JJ, Lee S, Breuil C (2004) Multigene phylogenies of *Ophiostoma clavigerum* and closely related species from bark beetle-attacked *Pinus* in North America. *FEMS Microbiology Letters* **237**, 89-96.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH (2008) Intraspecific its variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary bioinformatics online* **4**, 193-201.

O'Donnell K, Cigelnik E (1997) Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Molecular Phylogenetics and Evolution* **7**, 103-116.

Olmstead RG (1995) Species concepts and pleisiomorphic species. *Systematic Botany* **20**, 623-630.

Owen DR, Lindahl KQJ, Wood DL, Parmeter JRJ (1987) Pathogenicity of fungi isolated from *Dendroctonus valens, D. brevicomis,* and *D. ponderosae* to ponderosae pine seedlings. *Phytopathology* **77**, 631-636.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution* **24**, 386-393.

Pukkila PJ, Skrzynia C (1993) Frequent changes in the number of reiterated ribosomal RNA genes throughout the life cycle of the basidiomycete *Coprinus cinereus*. *Genetics* **133**, 203-211.

Raffa KF, Berryman AA (1983) Physiological aspects of lodgepole pine wound response to a fungal symbiont of the mountain pine beetle. *Canadian Entomologist* **115**, 723-734.

Ratnasingham S, Hebert PD (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol Ecol Notes* **7**, 355-364.

Reid RW, Whitney HS, Watson JA (1967) Reactions of lodgepole pine to attack by *Dendroctonus ponderosae* Hopkins and blue stain fungi. *Canadian Journal of Botany* **40**, 609-614.

Rice A, Langor D (2009) Mountain pine beetle-associated blue-stain fungi in lodgepole x jack pine hybrids near Grande Prairie, Alberta (Canada). *Forest Pathology* **39**, 323-334.

Rice A, Thormann M, Langor D (2007) Virulence of, and interactions among, mountain pine beetle associated blue-stain fungi on two pine species and their hybrids in Alberta. *Canadian Journal of Botany* **85**, 316-323.

Robinson RC (1962) Blue stain fungi in lodgepole pine (*Pinus contorta* Dougl. var. *latifolia* Engelm.) infested by the mountain pine beetle (*Dendroctonus monticolae* Hopk.). *Canadian Journal of Botany* **40**, 609-614.

Robinson-Jeffery RC, Davidson RW (1968) Three new *Europhium* species wih *Verticicladiella* imperfect states on blue-stained pine. *Canadian Journal of Botany* **46**, 1523-1527.

Roe AD, Sperling FAH (2007a) Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution* **44**, 325-345.

Roe AD, Sperling FAH (2007b) Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach. *Molecular Ecology* **16**, 3617-3633.

Rossman A (2007) Report of the planning workshop for All Fungi DNA barcoding. *Inoculum* **58**, 1-5.

Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *Journal of Heredity* **97**, 581-594.

Rumbold CT (1941) A sapstaining fungus *Ceratostomella montium* n. sp. and some yeasts associated with two species of *Dendroctonus*. *Journal of Agricultural Research* **62**, 589-601.

Seifert KA (2009) Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* **9**, 83-89.

Seifert KA, Samson RA, Dewaard JR*, et al.* (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case.

Proceedings of the National Academy of Sciences of the United States of America **104**, 3901-3906.

Shenoy BD, Jeewon R, Hyde KD (2007) Impact of DNA sequence-data on the taxonomy of anamorphic fungi. *Fungal Diversity* **26**, 1-54.

Simon UK, Weiss M (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Biology and Evolution* **25**, 2251-2254.

Sites JW, Marshall JC (2004) Operational criteria for delimiting species. *Annual Review of Ecology Evolution and Systematics* **35**, 199-227.

Six DL, Bentz BJ (2003) Fungi associated with the North American spruce beetle, *Dendroctonus rufipennis*. *Canadian Journal of Forest Research* **33**, 1815-1820.

Six DL, Paine TD (1998) Effects of mycangial fungi and host tree species on progeny survival and emergence of *Dendroctonus ponderosae* (Coleoptera : Scolytidae). *Environmental Entomology* **27**, 1393-1401.

Sperling FAH (2003) DNA barcoding: *deus ex machina*. *News of the Lepidopterists' Society* **45**, 1-3.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology* **57**, 758-771.

Stobel GA, Sugarwara F (1986) The pathogenicity of *Ceratocystis montia* to lodgepole pine. *Canadian Journal of Botany* **64**, 113-116.

Summerbell RC, Levesque CA, Seifert KA*, et al.* (2005) Microcoding: the second step in DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1897-1903.

Summerbell RC, Moore MK, Starink-Willemse M, Van Iperen A (2007) ITS barcodes for *Trichophyton tonsurans* and *T. equinum*. *Medical Mycology* **45**, 193-200.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2002) DNA points the way ahead in taxonomy. *Nature* **418**, 479.

Taylor DL, McCormick MK (2008) Internal transcribed spacer primers and sequences for improved characterization of basidiomycetous orchid mycorrhizas. *New Phytologist* **177**, 1020-1033.

Taylor JW, Jacobson DJ, Kroken S*, et al.* (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* **31**, 21-32.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**, 381-397.

Varga J, Frisvad JC, Samson RA (2007) Polyphasic taxonomy of *Aspergillus* section Candidi based on molecular, morphological and physiological data. *Studies in Mycology* **59**, 75-88.

Vialle A, Feau N, Allaire M*, et al.* (2009) Evaluation of mitochondrial genes as DNA barcode for Basidiomycota. *Molecular Ecology Resources* **9**, 99-113.

Vilgalys D, Gonzalez D (1990) Organization of ribosomal DNA in the basidiomycete *Thanatephorus praticola*. *Current Genetics* **18**, 277-280.

White TJ, Bruns TD, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA gene for phylogenetics. In: *PCR Protocols: a Guide to Methods and Applications* (eds. Innis MA, Gelfand DH, Sninsky JJ, White TJ), p. 482. Academic Press, San Diego, CA.

Whitney HS, Farris SH (1970) Maxillary mycangium in the mountain pine beetle. *Science* **167**, 54-55.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* **54**, 844-851.

Zipfel RD, de Beer ZW, Jacobs K, Wingfield BD, Wingfield MJ (2006) Multi-gene phylogenies define *Ceratocystiopsis* and *Grosmannia* distinct from *Ophiostoma*. *Studies in Mycology* **55**, 75-97.

Table 1: Primers used in this study.

Gene	Direction	Sequence (5'-3')	Source
actin			
LepactF	F	TACGTCGGTGACGAGGC	Lim *et al*. 2004
LepactF_2	F	ACAGTCCAAGCGTGGTATCC	This study
LepactR	R	CAATGATCTTGACCTTCAT	Lim *et al*. 2004
EF1a			
EF3E	F	GTCGTYATCGGCCACGTCGA	Lim *et al*. 2004
TEF1-rev	R	GCCATCCTTGGAGATACCAGC	Samuels *et al*. 2002
Btub			
T1	F	AACATGCGTGAGATTGTAAGT	O'Donnell and Cigelnik 1997
BT12	R	GTTGTCAATGCAGAAGGTCTCG	Kim *et al*. 2003
BT12_internal	R	CCTCMGTGTAGTGMCCCTTG	This study
ITS2			
ITS3	F	GCATCGATGAAGAACGCAGC	White *et al.* 1990
TW13	R	GGTCCGTGTTTCAAGACG	White *et al.* 1990
Anon. Locus			
UFM1_F	F	AGTATGTGCCCCGAATCAG	This study
UFM1_R	R	ATCTGTTGCCGATGCTTG	This study

Table 2: Locus information for single-locus analyses and concatenated multilocus maximum likelihood analysis. Values calculated for the multilocus partitioned ML analysis are shown in brackets for each gene where values differ from single locus analyses. Ingroup character information includes representative sequence data obtained from GenBank.

	actin	EF1a	Btub	UFM	ITS2	Combined
# haplotypes*	15	15	11	10	8	40
# sites	765	630	1045	393	887	3720
# constant char.*	735	615	1007	379	880	3616
# variable char. (uninformative)*	9	2	13	2	3	30
# pars. inform. char.*	21	13	25	12	4	74
% informative*	2.75%	2.06%	2.39%	3.05%	0.34%	1.99%
ML Model	GTR+Γ	GTR+Γ	GTR+Γ	GTR+Γ	GTR+Γ	GTR+Γ
-ln	-1453.2358	-1264.0232	-2020.8777	-638.2299	-1315.2930	-6947.2673
∝ gaps/missing	0.003862	0.1160	0.1580	0.0000	0.03908	0.09674
Base freq.						partitioned
A	0.2079	0.2179	0.2111	0.2203	0.2124	-
C	0.3039	0.3036	0.2790	0.3022	0.2974	-
G	0.2721	0.2267	0.2783	0.2960	0.3187	-
T	0.2162	0.2518	0.2316	0.1814	0.1714	-
Rate Matrix						partitioned
A-C	0.0721 (0.8423)	0.3959 (0.4898)	0.8292 (1.2951)	0.000017	0.4284 (0.2233)	-
A-G	0.8200 (2.5714)	1.2255 (1.8538)	1.9061 (2.8402)	4.1214 (7.1599)	0.1973 (0.1053)	-
A-T	2.6728 (0.000017)	1.5994 (3.1258)	0.5950 (0.8616)	0.000017	0.000017	-
C-G	0.3688 (0.4131)	0.2826 (0.4170)	0.4735 (0.7996)	1.4752 (3.1950)	0.000017	-
C-T	5.3126 (4.2967)	4.2956 (7.2423)	4.3915 (6.2634)	5.6200 (9.4605)	2.4492 (2.1376)	-
G-T	1.0000	1.0000	1.0000	1.0000	1.0000	-
Γ (alpha)	0.02000	0.9497 (0.2318)	0.1820 (0.2457)	0.0200	29.5889 (0.0200)	partitioned

* ingroup only

Table 3: Literature survey of multilocus data for pairs of closely related ascomycete fungi. Total number of loci (TOTAL) with fixed haplotype differences (FIXED), shared haplotypes (SHARED), or with a combination of fixed and shared haplotypes (PARTIAL) between a fungal species pair. Relationships between the species is classified as follows: for phylogenetic analyses RM = reciprocally monophyletic, M = monophyletic, PA = paraphyletic, and PO = polyphyletic; for distance analyses C = congruent and NC = not congruent.

LOCATION	TOTAL	FIXED	SHARED	PARTIAL	RM	PA	PO	C	NC
rDNA	20	11	5	4	4	1	5	5	5
mtDNA	2	2	0	0	0	2	0	0	0
Autosomal	52	46	1	5	23	6	0	17	6
Anonymous	2	1	1	0	1	0	1	0	0
AFLP	1	-	-	-	-	-	-	1	0

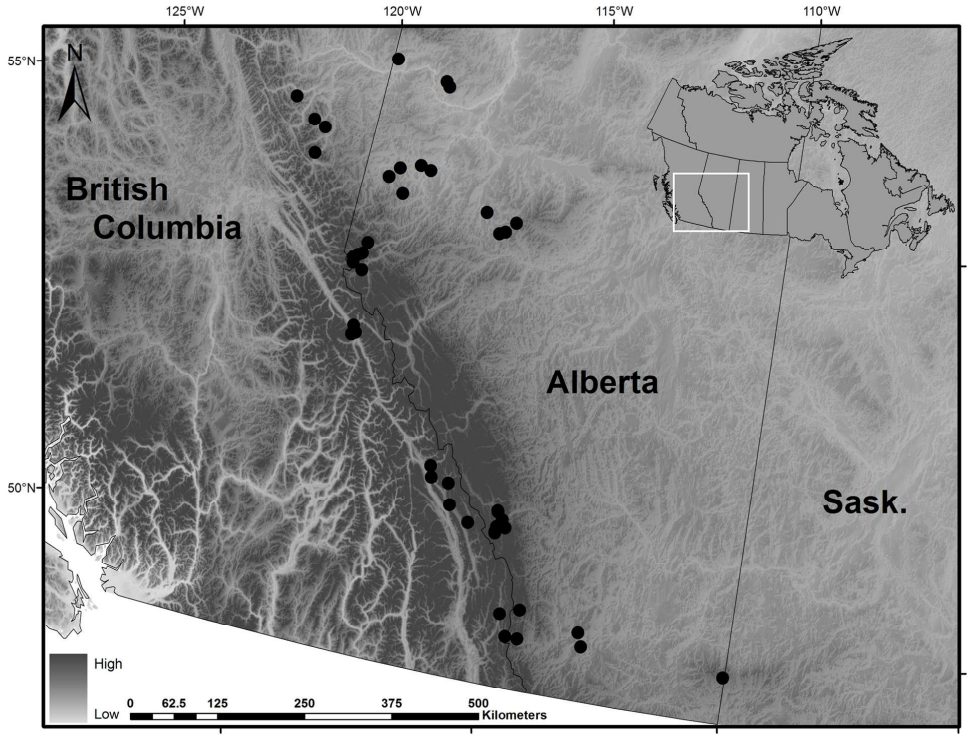

Figure 1: Map of collecting localities surveyed for MPB fungal associates.
168x127mm (600 x 600 DPI)

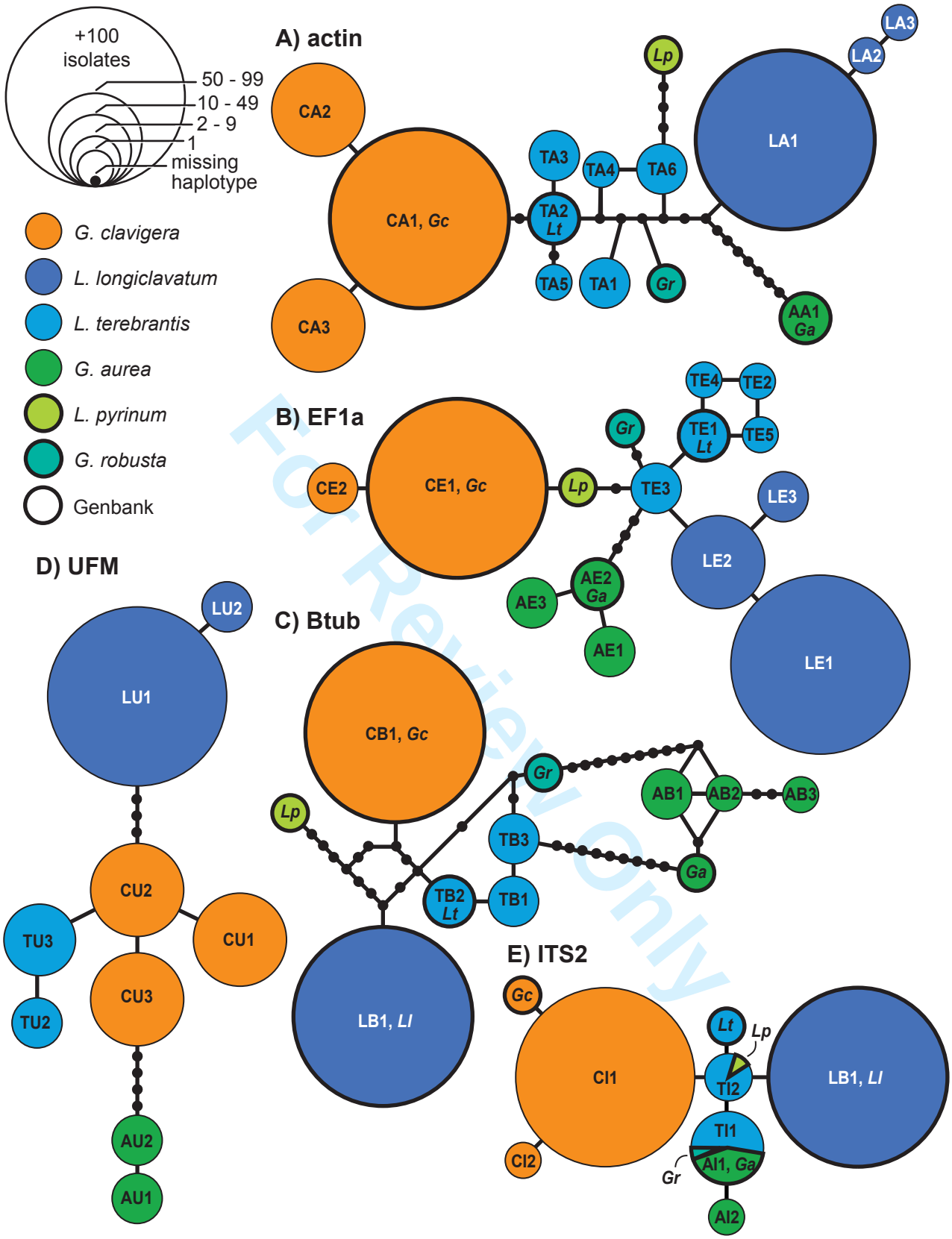

Figure 2: Parsimony haplotype networks for five independent loci of six closely related members of the *Grosmannia* clade. Coloured circles represent individual haplotypes, with circle size proportional to the number of isolates. Lines linking haplotypes represent single nucleotide differences. Black circles indicate missing haplotypes. Haplotypes with thick black borders are identical to representative GenBank sequences.
A) actin; B) elongation factor 1 alpha; C) beta tubulin; D) anonymous locus; E) internal transcribed spacer + partial 28S.

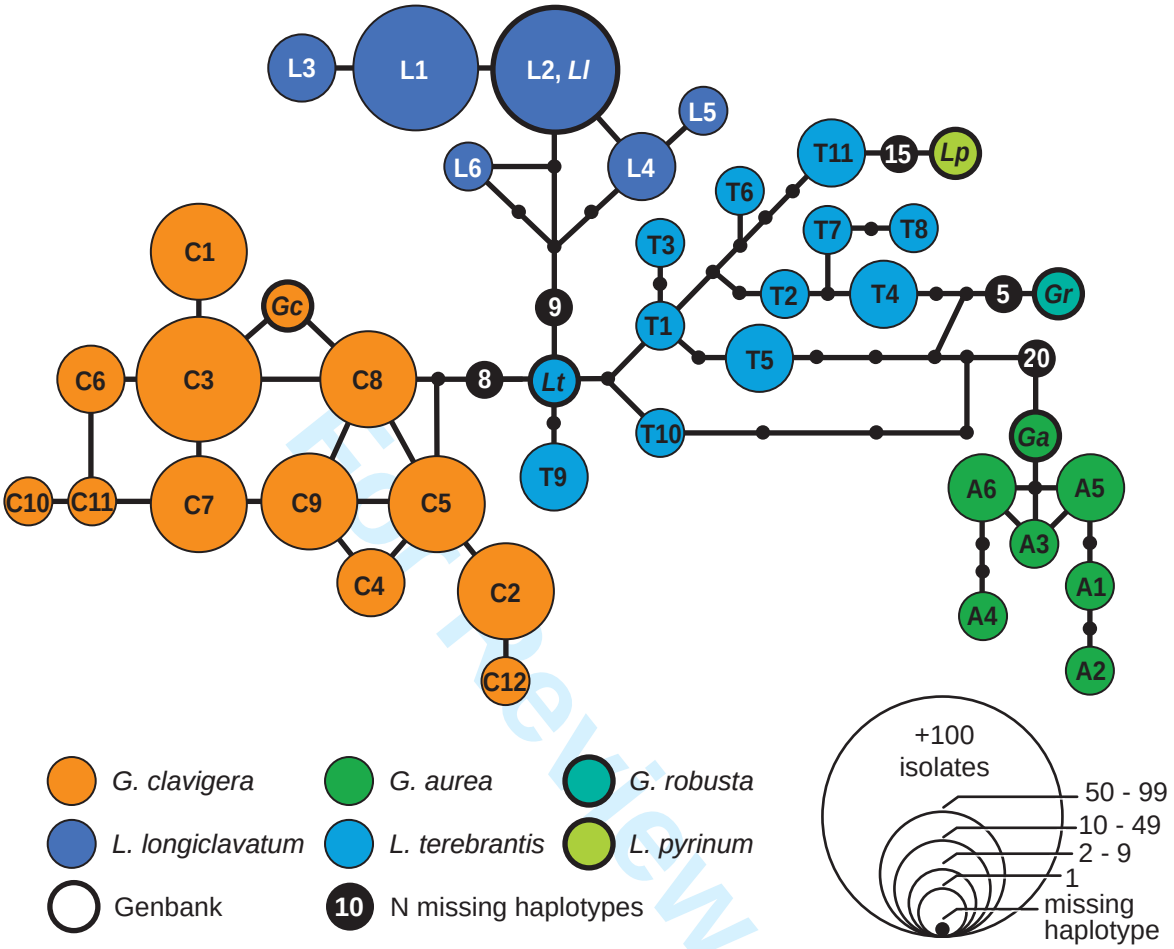

Figure 3: Parsimony haplotype network for the concatenated multilocus data set of six closely related members of the *Grosmannia* clade. See Fig. 2 caption for description of haplotype network.

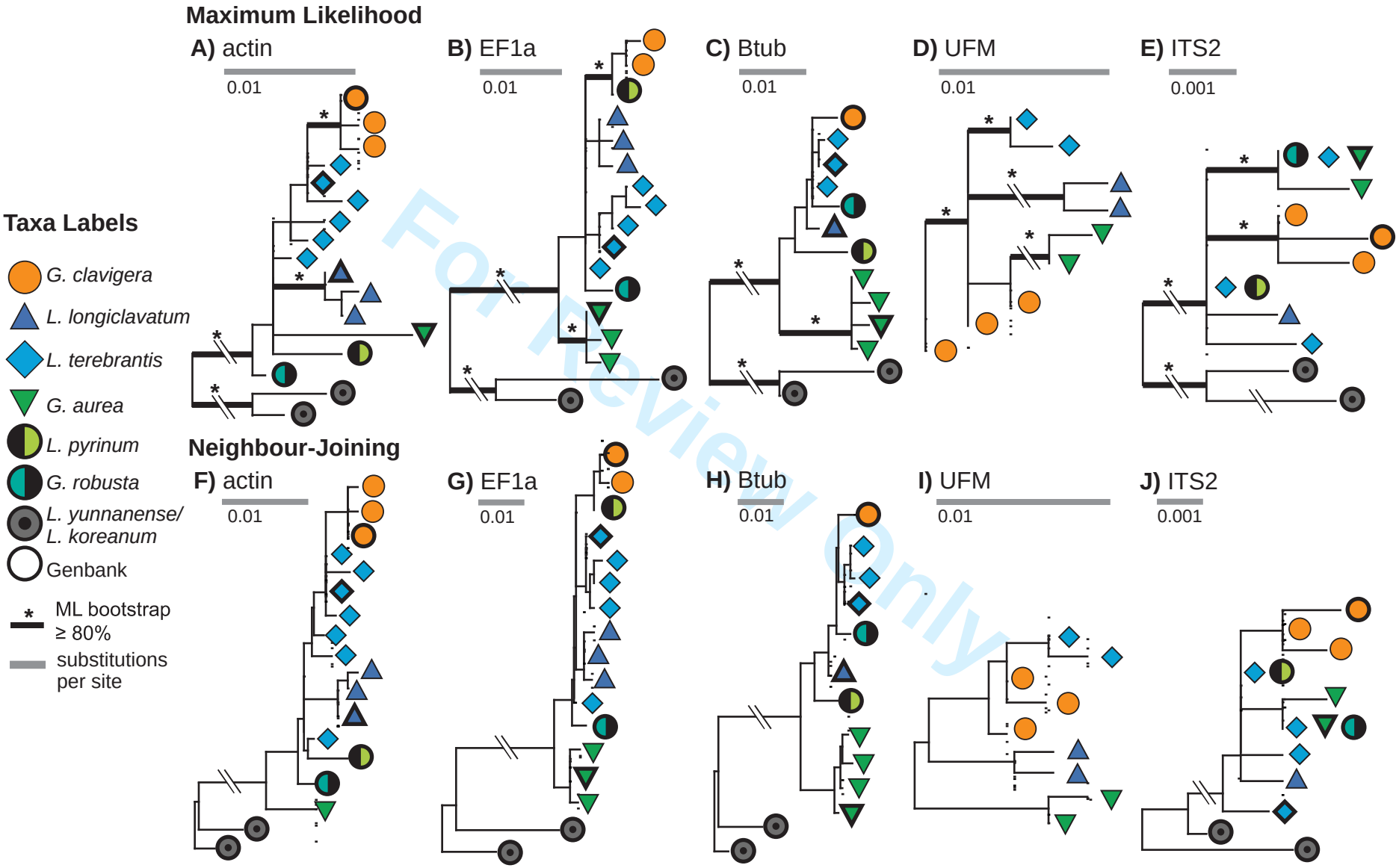

Figure 4: Maximum likelihood (ML) and neighbour-joining (NJ) trees for five independent loci of six closely related members of the *Grosmannia* clade.

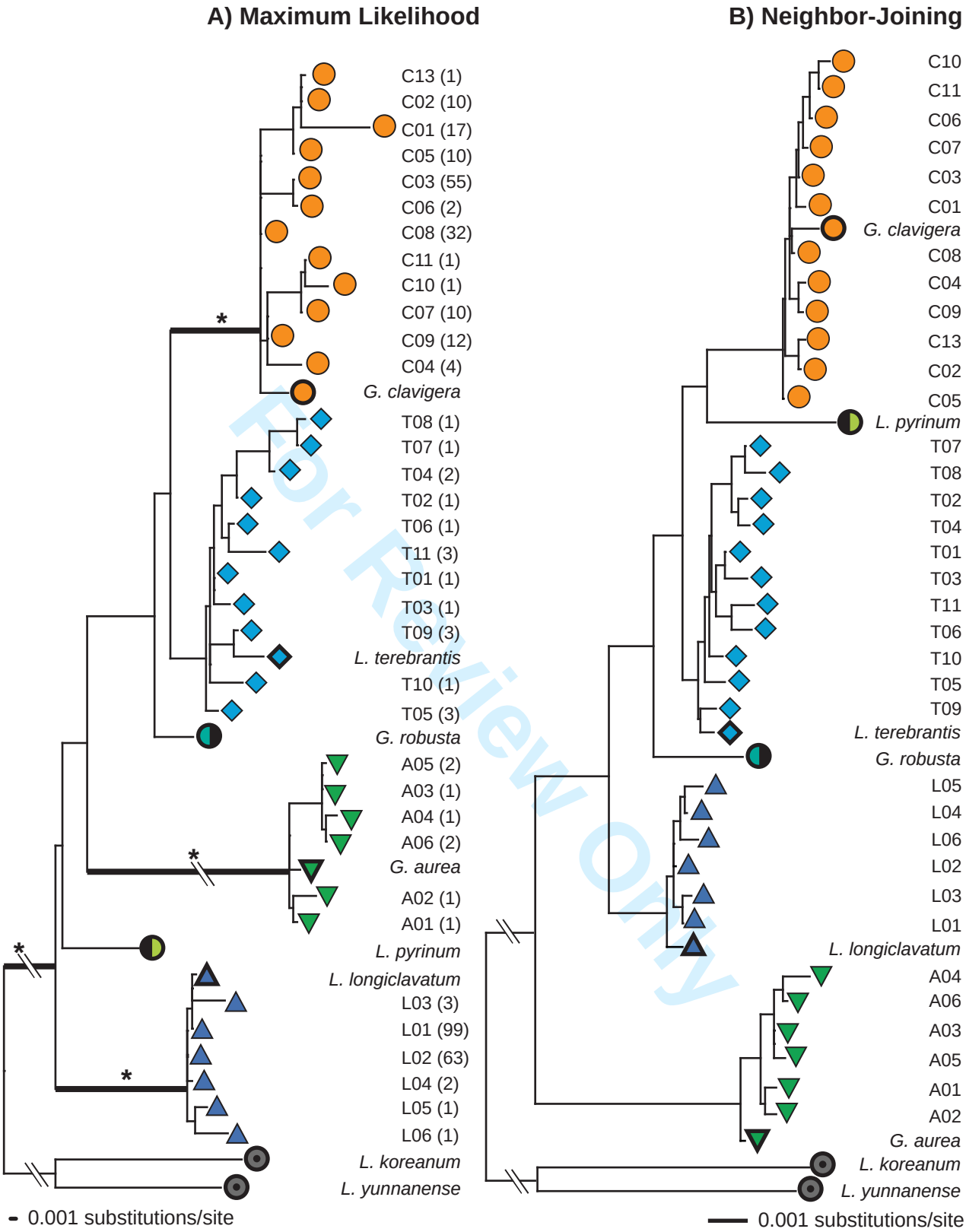

Figure 5: Maximum likelihood and neighbour-joining trees for the concatenated multilocus data set. A) Maximum likelihood phylogram. B) Neighbour-joining tree. Legend as in Fig. 4.

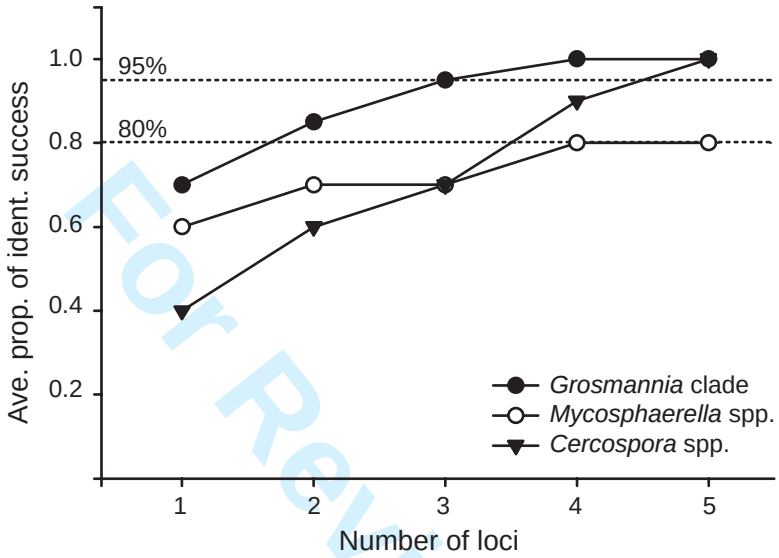

Figure 6: Changes in species identification success with the addition of loci. Proportion of NJ haplotype clusters congruent with published species limits was used as a measure of accurate species identification. All possible locus combinations were examined and averaged for each locus category. Benchmark levels of identification success (80% and 95%) are indicated. Data for this analysis were obtained from the current study (*Grosmannia* clade), Crous et al. (2004) (*Mycosphaerella* spp.), and Groenewald et al. (2005) (*Cercospora* spp.)