Mycologia, 100(4), 2008, pp. 577–589. DOI: 10.3852/07-042R1

Molecular diversity assessment of arctic and boreal *Agaricus* taxa

József Geml[1]
Gary A. Laursen
D. Lee Taylor

Institute of Arctic Biology, 311 Irving I Building, 902 N. Koyukuk Drive, P.O. Box 757000, University of Alaska at Fairbanks, Fairbanks, Alaska 99775

Abstract: We provide a phylogenetic diversity assessment study in genus *Agaricus* as part of our ongoing work to saturate ITS and LSU rDNA sequence diversity of soil-dwelling fungi in Alaska. Pairwise sequence similarity-based groupings and statistical parsimony analyses were applied to delimit operational taxonomic unit (OTU) and were compared to results of full phylogenetic analyses. Our results show that the proportion of section *Arvenses* taxa is particularly high in the boreal forest and hypo-arctic (low arctic) regions, whereas the genus is represented by section *Agaricus* in high arctic habitats. Furthermore our findings suggest that the commercially important *A. bisporus* occurs naturally in the boreal region of interior Alaska, substantially expanding the known northern limit of the species. Delimitations of OTU varied greatly with different methods. In general 95% similarity-based grouping proved to be the least sensitive method, often resulting in section- and subsection-level groups. The 95% connection-limit statistical parsimony separated far more groups. The 98% similarity-based groups and the 98% connection limit networks recognized respectively 11 and 13 OTU containing our specimens. The 98% connection limit statistical parsimony was the only method in which all recognized OTU consisted of members grouped by branches with significant (> .95) posterior probabilities, providing an independent support for the groups. Our results also point out that considerable additional efforts will be needed to elucidate the evolution of this diverse genus and to assess its phylogenetic diversity, given that most taxa in our analyses could not be placed convincingly within well characterized species using ITS/LSU data.

Key words: Alaska, fungi, Greenland, ITS, LSU, OTU, phylogenetics, rDNA, Svalbard

INTRODUCTION

It has been recognized for some time that single-locus DNA sequences can be used to discriminate among species. Because taxonomic expertise in many organism groups is sadly disappearing, while concerns over sustaining biodiversity are increasingly urgent, it has been argued that the generation of molecular databases of species-specific ''barcodes'' should be a high priority (Savolainen et al 2005). These databases then would allow the identification of samples that are too degraded for morphological identification and for rapid identification of specimens by a wider research community beyond specialists in that group. Microbial ecologists have long used similar methods, in particular for the identification of uncultured organisms in mixed environmental samples. Eumycota has roughly 80 000 described species, with estimates of the true numbers at 1.5 million (Hawksworth 1991) or perhaps much higher (Schadt et al 2003, O'Brien et al 2005, Lynch and Thorn 2006, Taylor et al 2007). While many fungi have macroscopic fruiting structures on which fungal taxonomy is built, molecular analyses are revealing unknown fungal diversity of two forms: (i) cryptic phylogenetic species within previously known morphological species and (ii) novel lineages known only from environmental samples, especially soil.

Although it has some limitations the nuclear ribosomal ITS region is arguably the best genomic region available at present for use as a barcode across the Eumycota. While the CO1 gene has been advocated strongly as a universal barcode region in eukaryotes (http://www.dnabarcodes.org) the mycological research community recently gained approval from the Consortium for the Barcode of Life (CBOL) for designating the ITS as the official barcode region for fungi (http://barcoding.si.edu/Fungi.html). However, to be used most effectively, issues such as standards for designating operational taxonomic units (OTUs) based on ITS sequences require further evaluation. One such method, the delimitation of OTUs based on an arbitrary cut-off value for pairwise sequence identity, has been used widely in diversity assessment studies (e.g. Floyd et al 2002, O'Brien et al 2005, Lynch and Thorn 2006, Higgins et al 2007). However, it is unclear how well such divergence cut-off based OTU diversity corresponds to species diversity. An alternative option is statistical parsimony analysis (Templeton et al 1992), which searches for discontinuities in sequence variation to detect puta-

Accepted for publication 22 April 2008.
[1] Corresponding author. E-mail: jgeml@iab.alaska.edu

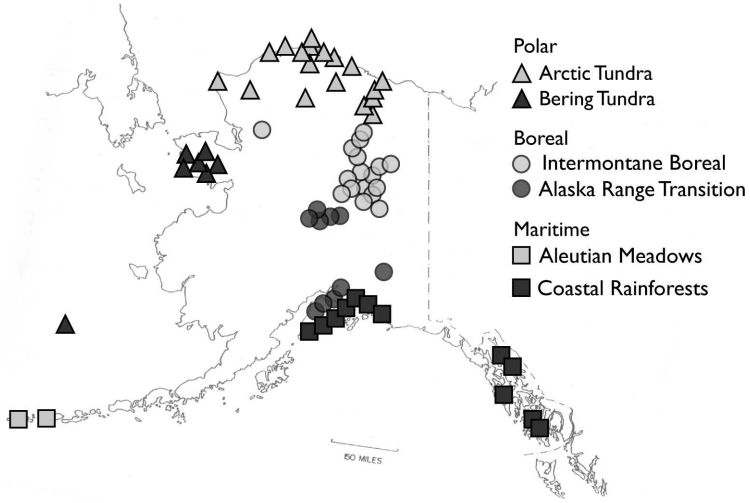

FIG. 1. Map of collecting localities representing different ecoregions of Alaska.

tive species boundaries. The method separates clusters of sequences into different networks if haplotypes are connected by relatively long branches. In practice, this method usually separates groups in approximate agreement with named species or phylogenetic species groups (e.g. Cardoso and Vogler 2005, Geml et al 2006, Templeton 2001). Because individual gene trees do not necessarily represent the species tree due to stochastic lineage sorting etc., the delimitation of putative species ideally should be based on sequence data from multiple unlinked loci to obtain a species tree from congruent parts of the individual gene trees (Taylor et al 2000). However, for large-scale biodiversity surveys, especially those including environmental clone libraries, a multigene approach currently is not feasible.

Due to its importance to the global climate system and position at the forefront of climate change, we have targeted boreal and arctic regions of Alaska for extensive surveys of fungal molecular biodiversity. We are generating ITS plus partial LSU sequences in a high throughput fashion from both curated sporocarp collections and soil PCR clone libraries. The soil sampling for clone-library construction was carried out at various plots of the Bonanza Creek LTER site in interior Alaska. The herbarium collections were gathered from across Alaska during 35 y of collecting. The collections span all three major ecoregions and six subregions in Alaska (FIG. 1, TABLE I), supplemented by materials from Svalbard, Nunavut, and subantartic islands of New Zealand and Australia. Here we evaluate OTU discrimination methods along with biodiversity patterns in a well known genus of fungi, *Agaricus*, with considerable economic importance.

Agaricus is a diverse, cosmopolitan genus with an estimated number of species exceeding 300 worldwide (Cappelli 1984, Bas 1991). The genus is characterized generally by a white, yellow or brown pileus, free lamellae that are pallid or pinkish when young, later becoming chocolate-brown, and dark brown, smooth basidiospores. Based on molecular evidence several secotioid taxa from genera *Endoptychum*, *Gyrophragmium* and *Longula* have been placed in genus *Agaricus* (Vellinga et al 2003, Geml 2004, Geml et al 2004). While separating genus *Agaricus* from other genera is relatively straightforward based on the morphological characters described above, identification and classification of several species has been controversial (Heinemann 1977, Cappelli 1984, Kerrigan 2005). This is due largely to the morphological, reproductive and ecological plasticity observed in numerous species (e.g. see Kerrigan et al 1999, Calvo-Bado et al 2000, Geml et al 2004, Callac and Guinberteau 2005, Kerrigan 2005). Efforts have been made to reconstruct the evolutionary history of the genus from molecular data (Bunyard et al 1996, Mitchell and Bresinsky 1999, Robison et al 2001, Challen et al 2003, Geml 2004, Geml et al 2004) and to detect and describe new species with combined molecular phylogenetics and morphological investigation (Callac and Guinberteau 2005, Kerrigan 2005). However, all studies included predominantly collections from temperate climates and the evolutionary history of *Agaricus* taxa in the northern high latitudes (i.e. arctic and boreal regions) remains largely unknown.

One species, *A. bisporus* (Lange) Imbach, merits particular attention because of its commercial importance and culinary value. Numerous wild populations of this species have been discovered and these represent valuable resources for commercial mush-

TABLE I. Ecoregions of Alaska represented by our collecting efforts. Classification is based on Nowacki et al (2001). Characteristic vegetations refer to low-elevation communities; at higher elevations, alpine tundra can be found in all ecological subregions

Ecoregion	Subregion	Characteristic vegetation	Sampled localities[1]	Number of collecting trips (1971–2006)
Polar	Arctic Tundra	*Betula*, *Salix* low shrubs; sedges; ericaceous heath, herbs, mosses and lichens; *Dryas* on dry ridges; shrubs absent from northernmost section (e.g. Barrow)	Atigun Pass, Barrow, Colville River, Dalton Highway (north of the Brooks Range), Driftwood, Franklin Bluffs, Ivotuk, Meade River, Prudhoe Bay, Toolik Lake LTER site, Umiat, Wainwright	33
	Bering Tundra	*Alnus*, *Betula*, *Salix* tall shrubs; sedges; ericaceous heath, herbs, and lichens; *Dryas* on dry ridges	Kotzebue, Seward Peninsula, St Lawrence and St Paul Islands	7
Boreal	Intermontane Boreal	Fire disturbance mosaic of *Picea glauca*, *P. mariana*, *Populus balsamifera*, *P. tremuloides*, and *Betula neoalaskana*	Bonanza Creek LTER site, Caribou-Poker Creek LTER site, Dalton Highway (south of the Brooks Range), Eagle Summit, Fairbanks and vicinity, Kobuk Valley National Park	52
	Alaska Range Transition	Similar to Intermontane Boreal, more precipitation	Anchorage and vicinity, Denali National Park and Preserve, Eagle River, Glennallen, Hatcher Pass, Parks, Seward and Sterling highways	17
Maritime	Aleutian Meadows	*Alnus*, *Betula*, *Salix* shrubs; *Dryas*, ericaceous heath, herbs, lichens, and grass communities	Adak and Amchitka Islands	2
	Coastal Rainforests	*Picea sitchensis*, *Tsuga heterophylla*, *T. mertensiana*	Columbia Glacier, Juneau, Kenai Peninsula, Prince of Wales Island, Sitka, Tongass National Forest, Valdez	7

[1] See map in FIG. 1.

room breeding programs. Several of these populations were characterized genetically and found to exhibit substantial genetic variation as opposed to the majority of the commercial cultivars (Royse and May 1982; Loftus et al 1988; Kerrigan and Ross 1989; Kerrigan 1990; Callac et al 1993, 2000; Kerrigan et al 1993, 1995, 1996, 1998; Xu et al 1997). Although most individuals of the species, as the name implies, have predominantly bisporic basidia and an amphithallic lifecycle (i.e. both secondarily homothallic and heterothallic), two tetrasporic varieties have been described respectively from the Mojave Desert and the Mediterranean with predominantly heterothallic and homothallic lifecycles (Callac et al 1993, 2000). Former genetic analyses of wild *A. bisporus* populations included specimens from Europe, Israel, Alberta and California (Challen et al 2003). Among other taxa, we have found *A. bisporus* at 65°N in Alaska, although the former northernmost record of wild populations in North America was from Alberta, Canada, at approximately 52°N (Kerrigan et al 1995). *A. bisporus* is sold widely in grocery stores in Alaska. We, therefore, sought to determine whether *A. bisporus* specimens in our collection were from

TABLE II. Species, isolate code and GenBank accession numbers of *Agaricus* spp. included in this study. Specimens with new sequence data are indicated in bold

Species	ID number	GenBank accession number		Origin, reference
		ITS	LSU	
Agaricus sp.	**GAL2619**	EF460376	EF460401	Arctic tundra; Spitsbergen, Svalbard, Norway
	GAL2673	EF460375	EF460400	Arctic tundra; Spitsbergen, Svalbard, Norway
	GAL3083	EF460374	EF460399	Arctic tundra; Dalton Hwy: Mile 352.5, North Slope, Alaska, USA
	GAL5812	EF460364	EF460389	Arctic tundra; Serpentine Hot Springs, Seward Peninsula, Alaska, USA
	GAL6020	EF460362	EF460387	Boreal forest; Quartz Creek, Seward Peninsula, Alaska, USA
	GAL8621	EF460372	EF460397	Open boreal forest; Bear Creek, Seward Peninsula, Alaska, USA
	GAL8696	EF460368	EF460393	Arctic tundra; Teller Highway, Mile 43, Seward Peninsula, Alaska, USA
	GAL12153	EU244584	—	Maritime meadow; Adak, Aleutian Islands, Alaska, USA
	GAL12215	EU244585	—	Maritime meadow; Amchitka, Aleutian Islands, Alaska, USA
	GAL13796	EF460370	EF460395	Alpine tundra; Eagle Summit, Interior, Alaska, USA
	GAL13951	EF460373	EF460398	Boreal forest; Steese Highway, Interior, Alaska, USA
	GAL15446	EF460366	EF460391	Boreal forest; North Pole, Interior, Alaska, USA
	GAL15786	EF460365	EF460390	Boreal forest; Bonanza Creek LTER site, Interior, Alaska, USA
	GAL16010	EF460367	EF460392	Boreal forest; Fairbanks, Interior, Alaska, USA
	GAL16056	EF460361	EF460386	Boreal forest; Fairbanks, Interior, Alaska, USA
	GAL16070	EF460363	EF460388	Boreal forest; UAF Ski Trails, Interior, Alaska, USA
	GAL16645	EF460369	EF460394	Boreal forest; Creamer's Field, Interior, Alaska, USA
	GAL16893	EF460371	EF460396	Boreal forest; Creamer's Field, Interior, Alaska, USA
	GAL17591	EF460355	EF460380	Boreal forest; UAF Ski Trails, Interior, Alaska, USA
	GAL17637	EF460356	EF460381	Fairbanks, Interior, Alaska, USA
	GAL17717	EF460353	EF460378	Lawn; Fairbanks, Interior, Alaska, USA
	GAL17755	EF460352	EF460377	Boreal forest; UAF Ski Trails, Interior, Alaska, USA
	GAL18055	EF460354	EF460379	Boreal forest; UAF Ski Trails, Interior, Alaska, USA
	GAL18862	EF460359	EF460384	Open coastal forest; Homer, Kenai Peninsula, Alaska, USA
	GAL18885	EF460358	EF460383	Open coastal forest; Homer, Kenai Peninsula, Alaska, USA
	GAL18889	EF460360	EF460385	Open coastal forest; Homer, Kenai Peninsula, Alaska, USA
	GAL18890	EF460357	EF460382	Open coastal forest; Homer, Kenai Peninsula, Alaska, USA
	TKN10 3241L4	EU244586	EU244587	Boreal forest, UAF Arboretum, Interior, Alaska, USA
A. abruptibulbus Peck	—	AY484673	AY484673	Geml et al (2004)
A. albolutescens Zeller	—	AY484675	AY484675	Geml et al (2004)
A. aristocratus Gulden	**O73061**	EU567323	—	Arctic tundra; Greenland
A. arvensis (Schäffer) Fries	—	AY484691	AY484691	Geml et al (2004)
A. augustus Fries	—	AY484672	AY484672	Geml et al (2004)
A. bernardii (Quélet apud Cooke & Quélet) Saccardo	—	AY484678	AY484678	Geml et al (2004)

TABLE II. Continued

Species	ID number	GenBank accession number ITS	GenBank accession number LSU	Origin, reference
A. bisporus (Lange) Imbach	—	AY484692	AY484692	Geml et al (2004)
A. bitorquis (Quélet) Saccardo	—	AY484695	AY484695	Geml et al (2004)
A. blazei Murrill	—	AY484697	AY484697	Geml et al (2004)
A. californicus Peck	—	AY484679	AY484679	Geml et al (2004)
A. campestris (Linnaeus) Fries 1	—	U85307	U85273	Johnson and Vilgalys (1999)
A. campestris 2		DQ182533	—	Kerrigan and Challen unpubl
A. campestris 3		DQ486682	DQ110871	Curtis et al (AFTOL) unpubl
A. cupreobrunneus (Schäffer & Steer) Möller 1	—	AY484680	AY484680	Geml et al (2004)
A. cupreobrunneus 2		DQ182532	—	Kerrigan and Challen unpubl
A. devoniensis Orton	—	AJ418755	AF059225	Challen et al (2003), Mitchell and Bresinsky (1999)
A. diminutivus Peck	—	AY484681	AY484681	Geml et al (2004)
A. excellens (Möller) Möller	—	AY484682	AY484682	Geml et al (2004)
A. fissuratus (Möller) Möller	—	AY484683	AY484683	Geml et al (2004)
A. fuscofibrillosus (Möller) Pilát	—	AY484684	AY484684	Geml et al (2004)
A. fuscovelatus Kerrigan	—	AY484677	AY484677	Geml et al (2004)
A. hondensis Murrill	—	AY484685	AY484685	Geml et al (2004)
A. inapertus Vellinga	—	AF482834	AF482878	Vellinga et al (2003)
A. langei (Möller) Möller	—	AY484699	AY484699	Geml et al (2004)
A. lilaceps Zeller	—	AY484676	AY484676	Geml et al (2004)
A. macrocarpus (Möller) Möller	—	AY484686	AY484686	Geml et al (2004)
A. macrosporus (Möller et Schäffer) Pilát	—	AY484687	AY484687	Geml et al (2004)
A. nivescens (Möller) Möller	—	AY484670	AY484670	Geml et al (2004)
A. pocillator Murrill	—	U85308	AF041542	Johnson and Vilgalys (1999), Hopple and Vilgalys (1999)
A. semotus Fries	—	AJ133390	AF059224	Huang unpubl, Mitchell and Bresinsky (1999)
A. subfloccosus (Lange) Pilát	—	AY484698	AY484698	Geml et al (2004)
A. subperonatus (Lange) Singer	—	AF432889	AF059216	Challen et al (2003), Mitchell and Bresinsky (1999)
A. subrutilescens (Kauffman) Hotson et Stuntz	—	AY484688	AY484688	Geml et al (2004)
A. xanthoderma Genevier	—	AY484689	AY484689	Geml et al (2004)

natural populations and whether they were genetically different from other wild populations and/or cultivars using molecular phylogeographic methods.

MATERIALS AND METHODS

Generation of molecular data.—Eighty-nine specimens were collected from the high arctic (Svalbard), low arctic and boreal (Alaska) regions. Sporocarps were deposited in Gary A. Laursen's personal mycological herbarium at the University of Alaska at Fairbanks (UAF). Because the majority of the specimens were collected in interior Alaska and fell into the morphological section *Arvenses*, only a subset of these were chosen to reduce redundancy. Twenty-seven specimens, representing the morphological groups and geographic areas of origin among the collections, were selected for molecular work (TABLE II). One additional specimen, *A. aristocratus* (O73061) from the Mycological Herbarium at the Natural History Museum of the University of Oslo, also was sampled and included in the analyses. DNA was extracted from small samples of dried specimens with the E-Z 96® Fungal DNA Kit (Omega Bio-tek Inc., Doraville, Georgia). In addition, 65 000 sequences of ITS/LSU clones derived from soil DNA extracts from various boreal forest types (Taylor et al unpubl) were subject to BLAST analyses to identity *Agaricus*-affiliated sequences and thereby evaluate the diversity and abundance of *Agaricus* in the sampled soils. To our surprise, only one such clone (TKN10 3241L4) was found in a *Picea mariana* stand (TABLE II). ITS and LSU sequences of *Agaricus* species, representing the known major phylogenetic clades in the genus, were downloaded from GenBank (TABLE II). In preliminary analyses, all specimens, except *A. bisporus*, fell in two major clades, I and II *sensu* Geml et al (2004) that approximately correspond respectively to sections *Arvenses* Hein and *Agaricus* Hein. We, therefore, downloaded additional nonredundant ITS sequences from GenBank to represent the known molecular phylogenetic diversity of these clades. Homologous sequences of *Chlorophyllum molybdites* (U85309, U85303) were used to root all trees. In addition, ITS sequences (TABLE III) of *A. bisporus* were downloaded from GenBank for the intraspecific phylogeographic analyses. For this analysis, we only included full ITS sequences with no IUPAC ambiguity codes to avoid creating ambiguous connections among haplotypes.

PCR and DNA sequencing.—The entire ITS and partial LSU regions were PCR amplified in reaction mixtures containing 1.75 μL ultrapure water (Invitrogen), 1 μL 10× Herculase PCR buffer (Stratagene), 0.05 μL 100 mM dNTP mixture, 25 mM of each dNTP (Applied Biosystems), 0.2 μL Herculase DNA polymerase (Stratagene), 2 μL of 1 μM forward primer, ITS1F (Gardes and Bruns 1993) and reverse primer TW13 (White et al 1990) and 3 μL of template DNA at a concentration of 0.1 ng/μL. PCR reactions were performed with this temperature program: 95 C/2 min, 34 cycles of 95 C/0.5 min, 54 C/1 min, 72 C/2 min and 72 C/10 min. The concentration of the amplification products was determined with Picogreen (Molecular Probes). The amplification products were normalized to a concentration of 4 ng/μL and sequenced with the Applied Biosystems (ABI) BigDye v. 3.1 terminator kit and an ABI 3730xl automated capillary DNA sequencer (Applied Biosystems, Foster City, California). Because the amplification products were 1300+ bp long, we used two internal primers for cycle sequencing, ITS4 and CTB6 (White et al 1990), in addition to the primers used in the PCR reactions. All PCR and sequencing work was done at the MIT Broad Institute as part of our large-scale diversity assessment project.

Phylogenetic analyses.—Sequence data obtained for both strands of each locus were edited and assembled for each isolate with Aligner v. 1.3.4 (CodonCode Corp., Dedham, Massachusetts). Sequence alignments were initiated with Clustal W (Higgins et al 1991) and corrected manually. Phylogenetic analyses were conducted with the maximum-likelihood (ML) and maximum-parsimony (MP) method in PAUP* 4b10 (Swofford 2002) and Bayesian analysis in MrBayes 3.0 (Huelsenbeck and Ronquist 2001). The best-fit evolutionary model was determined by comparing different evolutionary models with varying values of base frequencies, substitution types, α-parameter of the γ-distribution of variable sites and proportion of invariable sites via the Akaike information criterion (AIC) with PAUP* and MrModeltest 2.2 (Nylander 2004). ML and MP analyses were carried out with the heuristic search option using the tree bisection and reconnection (TBR) algorithm. The stability of clades was tested with the MP bootstrap (Felsenstein 1985) with fast stepwise addition and 1000 replications. In Bayesian phylogenetic analyses, 200 000 generations were run in four chains. The chains were sampled every 100th generation. When the likelihood scores of trees sampled approached similar values, they were considered to have converged. In each run, trees after this convergence point were used to compute a majority rule consensus tree.

Delimitation of OTU.—Separate ITS alignments were made for Clades I and II for subsequent species delimitation analyses. Pairwise sequence similarity-based groupings were conducted by Cap3 (Huang and Madan 1999) with 95% and 98% cut-off values. (Note that Cap3 uses a single-linkage clustering algorithm, meaning that sequences less than 95% similar, for example, will be grouped together if an intermediate sequence greater than 95% similar to both is found.) Statistical parsimony analyses were carried out with TCS v.1.21 (Clement et al 2000) using only individuals with complete sequence information. The data were partitioned into independent networks of haplotypes connected by changes that are nonhomoplastic with 95% and 98% probabilities (Templeton 2001). Pairwise intragroup and intergroup divergences for 95% connection limit networks were calculated with PAUP* using p-distances and analyzed with JMP 3.2.6 (SAS Institute Inc.).

Intraspecific ITS diversity and phylogeography of A. bisporus.—Sequence alignments were initiated with Clustal W and corrected manually. Haplotype networks were generated by TCS v.1.21 (including indels) and were used to define a series of nested clades that in turn were used to perform random, two-way contingency permutation analysis to

TABLE III. Identity, frequency and GenBank accession numbers of *A. bisporus* ITS haplotype sequences

Haplotype	GenBank accession number	Isolate code	Geographic origin
I	EF460354	GAL18055	Alaska, USA
II	AJ133385	ATCC 62488	Commercial strain, Europe
	AF465404	Bs261	France
	AF465403	Bs261-150	France
	DQ404388	RWK1885	Denmark
III (*var. eurotetrasporus*)	AF465400	Bs423	France
	AF465399	Bs514	Greece
IV	AJ133380	ATCC 62471	Commercial strain, Europe
V	EF460355	GAL17591	Alaska, USA
VI (*var. burnettii*)	AF465401	JB	California, USA
VII	AJ133382	CCRC 36012	Commercial strain, Europe
	AF465402	PS-1	Commercial/wild hybrid, Europe
	AJ884646	5341	Ukraine
	AJ884645	459	Ukraine
	AJ884644	HAI0235	UK
	AY484692	10071	Hungary
	AJ131132	JFM-BY	Commercial strain, Europe
	AJ131131	TFM-BT70903	Commercial strain, Europe
	AJ131130	JFM-BW	Commercial strain, Europe
	AF432885	RWK1737	Ontario, Canada
	AF432883	KP1	Massachusetts, USA
	AJ409228	D1	Spain
	AJ409229	F1	France
	AJ409227	C3	Spain
	AJ409226	C2	Spain
	AJ409225	C1	Spain
	AJ409224	B2	Spain
	AJ409223	B1	Spain
	AJ409222	A4	Spain
	AJ409221	A3	Spain
	AJ409220	A2	Spain
VIII	AJ133383	ATCC 62473	Commercial strain, Europe
	AJ133381	CCRC 36011	Commercial strain, Europe
	AJ133378	ATCC 62455	Commercial strain, Europe
IX	AJ133384	ATCC 62469	Commercial strain, Europe
X	AJ133377	CCRC 36001	Commercial strain, Europe
XI	AF188033	KME59003	Commercial strain, Europe
XII	AF188035	KME75005	Commercial strain, Europe
XIII	AJ133379	ATCC 62457	Commercial strain, Europe

detect any association between geographic distribution and genetic variation (Templeton 1998). The nested clade information, sample size for each haplotype and geographic location of each clade were entered into the software package GeoDis v.2.0 (Posada et al 2000). GenBank sequences from commercial cultivars were considered to have a European origin, based on Fritsche (1991). GeoDis was used to calculate clade distance (D_c) and nested clade distance (D_n) and to test them for significance at an α level of 0.05 using a permutation technique with 10 000 resampling replicates (Posada et al 2000). D_c was calculated as the average distance of all individuals in Clade X from the geographic center of that clade, while D_n was the average distance of individuals in Clade X from the geographic center of clades of the next highest nesting level. Where significant D_c and/or D_n values were detected a set of criteria was used to detect the effects of contemporary (e.g. gene flow) versus historical (e.g. allopatric fragmentation and range expansion) processes (Posada et al 2000, Templeton 1998).

RESULTS

Phylogenetic analyses.—The combined ITS+LSU dataset of the 64 sequences consisted of 1414 characters, including gaps. Of these 99 characters were excluded due to incomplete data on the ends (1–44, 1360–1414). The Hasegawa-Kishino-Yano model (Hasegawa et al 1985) with calculated proportion of invariable sites (I = 0.4763) and estimated α-parameter (α = 0.4388) of γ-distribution characterizing among-site

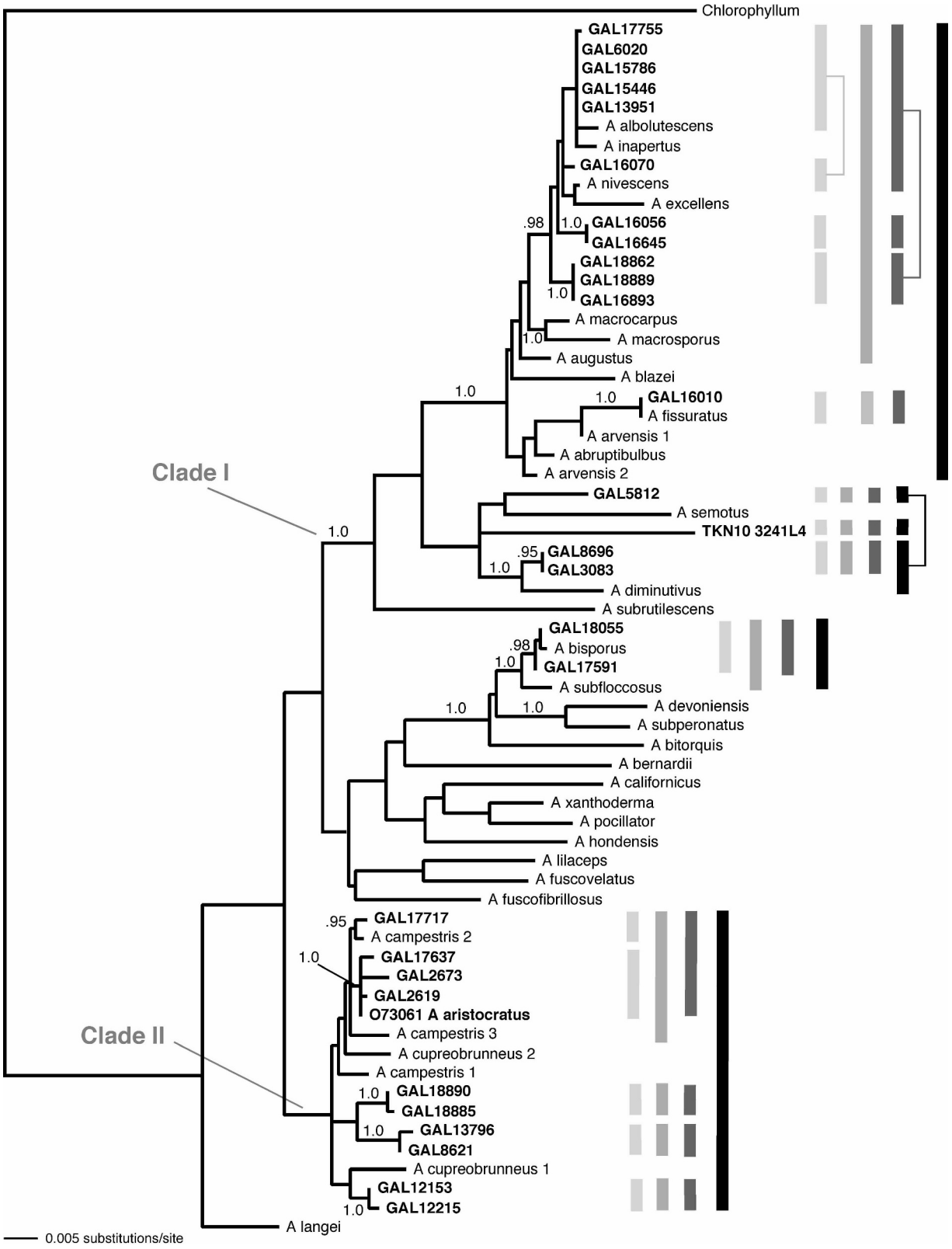

FIG. 2. The ITS/LSU maximum-likelihood phylogram (−lnL = 6168.5773) showing the phylogenetic spread of boreal and arctic *Agaricus* specimens (GAL and TKN numbers) sampled in this study. Bayesian posterior probability values are shown for the branches. OTUs delimited by 98% parsimony connection limit, 95% parsimony connection limit, 98% similarity cut-off, and 95% similarity cut-off are indicated, respectively, by light gray (98P), gray (95P), dark gray (98S) and black (95S) bars. OTU

variation (HKY+I+G) was selected as the best-fit evolutionary model. The base frequencies were freqA = 0.2600, freqC = 0.2085, freqG = 0.2408, freqT = 0.2907. The transition/transversion ratio was ti/tv = 3.2122. In Bayesian analysis, the consensus trees were computed from 1350 trees after discarding the first 651 trees as burn-in. The ML phylogram with the highest likelihood score is shown (FIG. 2).

The arctic and boreal *Agaricus* taxa sampled in our project fell into three major phylogenetic clades *sensu* Geml et al (2004): Clade I (section *Arvenses* Hein.), Clade II (section *Agaricus* Hein.) and Clade IV (section *Duploannulatae* Wasser ex Wasser). While Clade IV was represented only by a single species (*A. bisporus*: GAL17591, GAL18055) we found several taxa in Clades I and II.

Delimitation of OTU.—In Clade I, Alaskan specimens grouped in five independent 95% connection-limit parsimony networks (FIG. 2): (A) GAL6020, GAL13951, GAL15446, GAL15786, GAL17755, *A. albolutescens*, *A. inapertus*, GAL16893, GAL18862, GAL18889, GAL16056, GAL16645, GAL16070, *A. nivescens*, *A. augustus*, *A. macrocarpus*, *A. macrosporus*, and *A. excellens*; (B) GAL16010 and *A. fissuratus*; (C) GAL5812; (D) the clone TKN10 3241L4; (E) GAL3083 and GAL8696. The mean pairwise intragroup and intergroup p-distances with standard deviation were respectively 1.22 ± 0.71% and 4.95 ± 2.05%. When compared with *t*-test the means were significantly different ($P < 0.0001$). At 98% connection limit, the first network was subdivided further into multiple OTUs (discussing only those including arctic or boreal specimens): (A/1) GAL6020, GAL13951, GAL15446, GAL15786, GAL17755, *A. albolutescens*, GAL16070, *A. nivescens*; (A/2) GAL16893, GAL18862, GAL18889; (A/3) GAL16056, GAL16645 (FIG. 2). The 98% similarity-based OTUs were identical to those obtained by 98% parsimony connection limit, except that A/1, A/2 and *A. inapertus* were merged in a 98% similarity group. The 95% similarity-based OTU approximately corresponded to subsection-level groups *Arvenses* and *Minores* (Heinemann 1977).

In Clade II, northern high-latitude specimens were included in four independent 95% connection-limit networks: (F) GAL2619, GAL2673, GAL17637, GAL17717, O73061, *A. campestris* 2, and *A. campestris* 3; (G) GAL18885 and GAL18890; (H) GAL8621 and GAL13796; and (I) GAL12153 and GAL12215. The mean pairwise intragroup and intergroup p-distances were respectively 1.08 ± 0.57% and 2.07 ± 0.47%, which were significantly different ($P < 0.0001$). At 98% connection limit the first network was subdivided further into two OTUs: (F/1) GAL2619, GAL2673, GAL17637, O73061; (F/2) GAL17717, *A. campestris* 2. With respect to our samples the 98% similarity-based OTU were similar to those obtained by 98% parsimony connection limit, except that F/1 and F/2 were merged in a 98% similarity group. The single 95% similarity-based OTU corresponded to the entire section-level clade, *Agaricus* (Heinemann 1977).

Intraspecific ITS diversity and phylogeography of A. bisporus.—The full ITS dataset of the 39 sequences consisted of 564 characters, including gaps. The nested haplotype network comprised 13 zero-step clades (haplotypes), four one-step clades and the entire network (FIG. 3). Both Alaskan haplotypes were unique. Of these, Haplotype I was connected to three other haplotypes, each one step away: Haplotype VI (found in *A. b. var. burnettii*), Haplotype II (found in European wild populations) and Haplotype VII (found in European cultivars and wild populations). The other Alaskan haplotype, V, was connected to only one haplotype, II. The null hypothesis of no association between genotype and geographic origin could be rejected at $P = 0.05$ level in only Clade 1-1 (tip), where significantly large D_n value was observed, $P = 0.002$. Based on the inference key of Templeton (1998) the significant statistical association between haplotype and geography was due to restricted gene flow with isolation by distance.

DISCUSSION

Based on our results, a moderate amount of phylogenetic diversity was found among arctic and boreal *Agaricus* taxa. The majority of specimens in the boreal forests, both in the number of sporocarps and in the number of taxa, belong to section *Arvenses*. Section *Arvenses* also is represented in the low arctic tundra, particularly with taxa grouping with species in subsection *Minores*. Based on our previous molecular analyses section *Arvenses* (with representatives of both subsections *Arvenses* and *Minores*) is the dominant infrageneric group within *Agaricus* in other high-latitude habitats as well, such as the subantarctic islands of New Zealand (Geml et al 2007). Therefore, several taxa in section *Arvenses* seem particularly well

←

delimitations were based only on ITS sequences. Bars connected by lines indicate OTUs that are not monophyletic on the ML tree. Only OTUs containing arctic or boreal specimens are indicated.

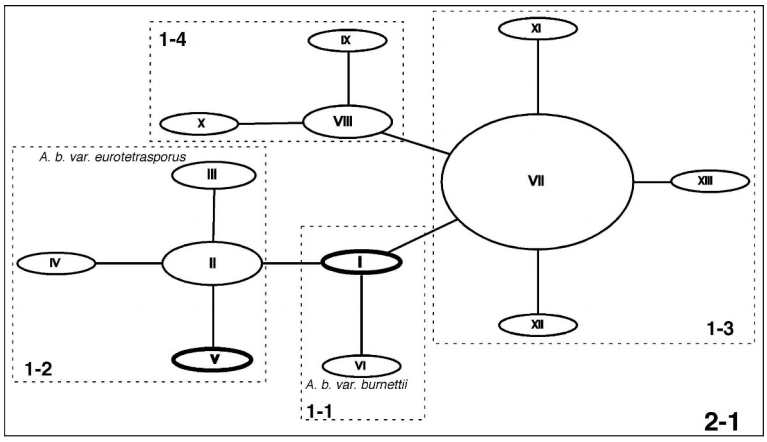

FIG. 3. Maximum-parsimony haplotype network constructed based on *A. bisporus* ITS sequences at 95% connection limit. Gaps were scored as ''new state''. Roman numbers indicate sampled haplotypes. Size of circles refers to the observed haplotype frequency. Haplotypes in bold have been found in Alaska.

adapted to harsh climates as opposed to groups (e.g. *Xanthodermatei*, *Sanguinolenti*) that have not been found in high-latitude ecosystems despite more than 30 y of collecting (GAL).

In addition to section *Arvenses*, the other well represented group is section *Agaricus*, containing *A. campestris*, *A. cupreobrunneus* and allies (Heinemann 1977). We found at least four, possibly five, putative high-latitude taxa in this group, at least three of which are unidentified and seem different from either *A. campestris* or *A. cupreobrunneus*, the only two species in this group with previously available ITS and LSU sequences. An Alaskan (GAL17637) and other arctic specimens from Svalbard (GAL2619, GAL2673) and Greenland (O73061) grouped together to form an OTU that appear to be *A. aristocratus*, based on the identification of O73061. The four *Agaricus* species reported from Svalbard are *A. aristocratus*, *A. macrosporus*, *A. comptulus* and *A.* cf. *crocodilinus*, according to the checklist of Agaricales of Svalbard by Gulden and Torkelsen (1996). Additional studies are needed to clarify the taxonomic status of these specimens. In general, high-latitude members of section *Agaricus* seem to have similarities with related low-latitude taxa in that they prefer open habitats (either tundra or meadow) to forest, as opposed to the taxa in section *Arvenses* detailed above.

Although a limited amount of ITS polymorphism was found in *A. bisporus*, both haplotypes from Alaskan populations were unique. Furthermore, their sequences were more similar to wild populations (both in Europe and North America) than to cultivars. These findings suggest that *A. bisporus* occurs naturally in the boreal region of interior Alaska, substantially expanding the known northern limit of the species. The haplotype network indicated little phylogeographic structure in the species. The null hypothesis of no geographical association of haplotypes could not be rejected in most clades, suggesting panmixia, extensive dispersal or inadequate geographical sampling, possibilities that cannot be distinguished with this method. While our geographic sampling admittedly is limited, being a widely cultivated agricultural crop provides many possible human-mediated dispersal scenarios over great distances and *A. bisporus* cultivars already have been shown to be able to escape cultivation and interbreed with local wild populations, for example in California (Kerrigan et al 1998). On the other hand, the statistically significant results obtained in one clade, containing only North American isolates, suggests restricted gene flow due to isolation by distance for at least some wild populations. Nonetheless more exhaustive geographic sampling is needed to elucidate the phylogeographic origin of the Alaskan *A. bisporus* populations.

Delimitations of OTUs varied greatly with different methods. In general, 95% similarity-based grouping proved to be the least sensitive method, often resulting in section- and subsection-level groups. The 95% connection-limit statistical parsimony separated far more groups. The 98% similarity-based groups and the 98% connection limit networks recognized respectively 11 and 13 OTUs containing our specimens. It is worth noting that the 98% connection limit statistical parsimony was the only method in which all recognized OTU consisted of members grouped by branches with significant ($>$ .95) posterior probabilities, providing independent support for the groups. Twelve of the 13 such OTUs were monophyletic, while one (A/1) was paraphyletic in the ML tree (FIG. 2). OTUs, particularly at 95%

similarity cut-off or connection limit, sometimes included multiple known species (e.g. *A. inapertus*, *A. augustus*, *A. nivescens* and *A. albolutescens*). There are many possible explanations for this observation. First, in general, ITS divergence values between pairs of many species in *Agaricus* are relatively low and described species can differ from each other by as few as 2–3 polymorphic sites in ITS (less than 0.5% divergence), for example, *A. moelleri* vs. *A. xanthodermus*, and *A. xanthodermulus* vs. *A. laskibarii* (Callac and Guinberteau 2005). This also implies that no universal similarity cut-off value for OTU delimitation can be used for all fungal groups and such delimitations should take into account what is known about the evolutionary rate and age of the group of interest. Second, some specimens with previously published sequences simply might have been misidentified. This is particularly plausible in *Arvenses*, where many described species have similar morphologies. Multilocus molecular studies involving type specimens are needed to clarify the systematics of that group.

With the estimated millions of species of fungi on Earth, little realistic hope is held for their comprehensive, in-depth taxonomic treatment with traditional methods. Large-scale DNA-based diversity-assessment projects have immense potential to help fill the enormous gaps in our current knowledge of fungal diversity. Our paper is the first to assess the phylogenetic diversity of *Agaricus* in northern high latitudes. Our findings point to an interesting ecological trend within the genus (i.e. that some of the major infrageneric evolutionary groups found commonly in temperate regions seem entirely absent in high latitudes). The proportion of section *Arvenses* taxa is particularly high in the boreal forest and low arctic regions, whereas the genus is represented only by section *Agaricus* in high arctic habitats. Our results also point out that considerable additional efforts will be needed to elucidate the evolution of this genus and to assess its phylogenetic diversity, given that most of the taxa in our analyses could not be convincingly identified with ITS/LSU data. These taxa of unknown identity might represent newly discovered species, the formal description of which is beyond the scope of this paper. Based on the recognized OTUs, at least 11–13 *Agaricus* taxa are in Alaska, of which several are undescribed or at least previously unsequenced. The actual diversity might be substantially higher because the soil clone sequence did not match any of the sporocarp sequences. We also have known taxa that were not expected to occur in Alaska (e.g. *A. bisporus*) and most OTUs described here are distributed over hundreds of kilometers, suggesting they are not rare or accidental colonists but members of large populations. While we have not completely resolved the ITS OTU delimitation problem, our results illustrate two points. First, if other lineages evolve similarly to *Agaricus*, then most environmental sequence studies, which typically use 90–97% identity for OTU delimitation (Floyd et al 2002, O'Brien et al 2005, Lynch and Thorn 2006, Higgins et al 2007, Taylor et al 2007), likely have lumped a number of distinct phylogenetic species. Second, parsimony network approaches appear somewhat more promising than arbitrary percent identity boundaries because the parsimony networks with a 98% connection limit distinguished expected groups in *Agaricus* and also have a stronger theoretical rationale related to thresholds in levels of gene flow.

ACKNOWLEDGMENTS

This research was supported by the NSF (grant No. 0333308) and a UAF Presidential International Polar Year Postdoctoral Fellowship. J. Geml thanks Christian Brochmann, Gro Gulden and Katriina Bendiksen for their hospitality when sampling arctic fungi at the Natural History Museum, University of Oslo. The authors thank Harris C. Nusbaum (Sequence and Analysis Program, Broad Institute) for the large-scale sequencing efforts related to the metagenomics project.

LITERATURE CITED

Bas C. 1991. A short introduction to the ecology, taxonomy and nomenclature of the genus *Agaricus*. In: van Griensven LJLD, ed. Genetics and breeding of *Agaricus*. Proceedings of the 1st International Seminar on Mushroom Science, Pudoc Wageningen. p 21–24.

Bunyard BA, Nicholson MS, Royse DJ. 1996. Phylogeny of the genus *Agaricus* inferred from restriction analysis of enzymatically amplified ribosomal DNA. Fungal Genet Biol 20:243–253.

Callac P, Billette C, Imbernon M, Kerrigan RW. 1993. Morphological, genetic, and interfertility analyses reveal a novel, tetrasporic variety of *Agaricus bisporus* from the Sonoran Desert of California. Mycologia 85: 835–851.

———, Guinberteau J. 2005. Morphological and molecular characterization of two novel species of *Agaricus* section *Xanthodermatei*. Mycologia 97:416–424.

———, Imbernon M, Guinberteau J, Pirobe L, Granit S, Olivier JM, Theochari I. 2000. Discovery of a wild Mediterranean population of *Agaricus bisporus*, and its usefullness for breeding work. Mush Sci 15:245–252.

Calvo-Bado L, Noble R, Challen MP, Dobrovin-Pennington A, Elliott TJ. 2000. Sexuality and genetic identity in the *Agaricus* section *Arvenses*. Appl Environ Microbiol 66: 728–734.

Cappelli A. 1984. *Agaricus* L.: Fr., Fungi Europaei. Vol. 1. Saronno, Italia: Libreria editrice Biella Giovanna. 560 p.

Cardoso A, Vogler AP. 2005. DNA taxonomy, phylogeny and Pleistocene diversification of the *Cicindela hybrida*

species group (Coleoptera: Cicindelidae). Mol Ecol 14: 353–3546.

Challen MP, Kerrigan RW, Callac P. 2003. A phylogenetic reconstruction and emendation of *Agaricus* section *Duploannulatae*. Mycologia 95:61–73.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. Mol Ecol 9: 1657–1659.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Floyd R, Abebe E, Papert A, Blaxter M. 2002. Molecular barcodes for soil nematode identification. Mol Ecol 11: 839–850.

Fritsche G. 1991. A personal view on mushroom breeding from 1957–1991. Genetics and breeding of *Agaricus*. In: van Griensven LJLD, ed. Genetics and breeding of Agaricus. Proceedings of the 1st International Seminar on Mushroom Science, Pudoc Wageningen. p 3–20.

Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity of basidiomycetes: application to the identification of mycorrhizae and rusts. Mol Ecol 2:113–118.

Geml J. 2004. Evolution in action: molecular evidence for recent emergence of secotioid genera *Endoptychum*, *Gyrophragmium* and *Longula* from *Agaricus* ancestors. Acta Microbiol Immunol Hung 51:97–108.

———, Geiser DM, Royse DJ. 2004. Molecular evolution of *Agaricus* species based on ITS and LSU rDNA sequences. Mycol Progress 3:157–176.

———, Laursen GA, O'Neill K, Nusbaum HC, Taylor DL. 2006. Beringian origins and cryptic speciation events in the fly Agaric (*Amanita muscaria*). Mol Ecol 15:225–239.

———, ———, Nusbaum HC, Taylor DL. 2007. Two new species of *Agaricus* from the Subantartic. Mycotaxon 100:193–208.

Gulden G, Torkelsen AE. 1996. A catalogue of Svalbard, plants, fungi, algae and cyanobacteria 3. Fungi I. Basidiomycota: Agaricales, Gasteromycetes, Aphyllophorales, Exobasidiales, Dacrymycetales and Tremellales. Norsk Polarinst Skr 198:173–206.

Hasegawa M, Kishino H, Yano TA. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 21:160–174.

Hawksworth DL. 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. Mycol Res 95:641–655.

Heinemann P. 1977. Essai d'une clé de détermination des genres *Agaricus* et *Micropsalliota*. Sydowia 30:6–37.

Higgins DG, Bleasby AJ, Fuchs A. 1991. Clustal W: improved software for multiple sequence alignment. CABIOS 8: 189–191.

Higgins KL, Arnold AE, Miadlikowska J, Sarvate SD, Lutzoni F. 2007. Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. Mol Phyl Evol 42:543–555.

Hopple JS, Vilgalys R. 1999. Phylogenetic relationships in the mushroom genus *Coprinus* and dark-spored allies based on sequence data from the nuclear gene coding for the large ribosomal subunit RNA: divergent domains, outgroups, and monophyly. Mol Phyl Evol 13:1–19.

Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. Genome Res 9:868–877.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.

Johnson J, Vilgalys R. 1999. Phylogenetic relationships within *Lepiota sensu lato* inferred from nuclear and mitochondrial rDNA sequences. Mycologia 91:443–458.

Kerrigan RW. 1990. Evidence of genetic divergence in two populations of *Agaricus bisporus*. Mycol Res 94:721–733.

———. 2005. *Agaricus subrufescens*, a cultivated edible and medicinal mushroom, and its synonyms. Mycologia 97: 12–24.

———, Billette C, Callac P, Velcko AJ. 1996. A Summary of allelic diversity and geographical distribution at six allozyme loci of *Agaricus bisporus*. In: Royse DJ, ed. Proceedings of the 2nd International Conference on Mushroom Biology and Mushroom Products. Pennsylvania State Univ. p 25–35.

———, Callac P, Xu J, Noble R. 1999. Population and phylogenetic structure within the *Agaricus subfloccosus* complex. Mycol Res 103:1515–1523.

———, Carvalho DB, Horgen PA, Anderson JB. 1995. Indigenous and introduced populations of *Agaricus bisporus*, the cultivated mushroom, in eastern and western Canada: implications for population biology, resource management, and conservation of genetic diversity. Can J Bot 73:1925–1938.

———, ———, ———, ———. 1998. The indigenous coastal Californian population of the mushroom *Agaricus bisporus*, a cultivated species may be at risk of extinction. Mol Ecol 7:35–45.

———, Horgen PA, Anderson JB. 1993. The California population of *Agaricus bisporus* comprises at least two ancestral elements. Syst Bot 18:123–136.

———, Ross IK. 1989. Allozymes of a wild *Agaricus bisporus* population: new alleles, new genotypes. Mycologia 81: 433–443.

Loftus MG, Moore D, Elliott TJ. 1988. DNA polymorphisms in commercial and wild strains of the cultivated mushroom, *Agaricus bisporus*. Theor Appl Genet 76: 712–718.

Lynch MDJ, Thorn RG. 2006. Diversity of Basidiomycetes in Michigan agricultural soils. Appl Environ Microbiol 72: 7050–7056.

Mitchell AD, Bresinsky A. 1999. Phylogenetic relationships of *Agaricus* species based on ITS-2 and 28S ribosomal DNA sequences. Mycologia 91:811–819.

Nowacki G, Spencer P, Fleming M, Brock T, Jorgenson T. 2001. Ecoregions of Alaska. U.S. Geological Survey Open-File Report 02-297 (map).

Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala University.

O'Brien H, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R. 2005. Fungal community analysis by large-scale se-

quencing of environmental samples. Appl Environ Microbiol 71:5544–5550.

Posada D, Crandall KA, Templeton AR. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Mol Ecol 9:487–488.

Robison MM, Chiang B, Horgen PA. 2001. A phylogeny of the genus *Agaricus* based on mitochondrial *atp6* sequences. Mycologia 93:30–37.

Royse DJ, May B. 1982. Use of isozyme variation to identify classes of *Agaricus brunnescens*. Mycologia 74:93–102.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Phil Trans R Soc B 360:1805–1811.

Schadt CW, Martin AP, Lipson DA, Schmidt SK. 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. Science 301:1359–1361.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). 4.0b4a. Sunderland, Massachusetts: Sinauer Associates.

Taylor DL, Herriott IC, Long J, O'Neill K. 2007. TOPO-TA is A-OK: a test of phylogenetic bias in fungal environmental clone library construction. Environ Microbiol 9:1329–1334.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC. 2000. Phylogenetic species recognition and species concepts in fungi. Fung Genet Biol 31:21–32.

Templeton AR. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol Ecol 7:381–397.

———. 2001. Using phylogeographic analyses of gene trees to test species status and processes. Mol Ecol 10:779–791.

———, Crandall KA, Sing CF. 1992. Cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III. Cladogram estimation. Genetics 132:619–633.

Vellinga EC, de Kok RPJ, Bruns TD. 2003. Phylogeny and taxonomy of *Macrolepiota* (Agaricaceae). Mycologia 95: 442–456.

White TJ, Bruns T, Lee S, Taylor JW. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. PCR protocols: a guide to methods and applications. New York: Academic Press. p 315–322.

Xu J, Kerrigan RW, Callac P, Horgen PA, Anderson JB. 1997. Genetic structure of natural populations of *Agaricus bisporus*, the commercial button mushroom. J Hered 88:482–488.