This document is the accepted manuscript version of the following article:
Tsykun, T., Rellstab, C., Dutech, C., Sipos, G., & Prospero, S. (2017). Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus Armillaria cepistipes. Heredity. http://doi.org/10.1038/hdy.2017.48

Title: Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*

Full names and addresses of all authors:

Dr. Tetyana Tsykun, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

Dr. Christian Rellstab, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

Dr. Cyril Dutech, INRA, University of Bordeaux, UMR 1202 BIOGECO, F-33612 Cestas, France

Dr. György Sipos, Research Center for Forestry and Wood Science, University of West Hungary, Bajcsy-Zsilinszky u. 4, HU-9400 Sopron, Hungary

Dr. Simone Prospero, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

Corresponding author: Tetyana Tsykun, Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland, phone: +41447392249, fax: +41447392215, e-mail: tetyana.tsykun@wsl.ch

Running title: Fungal population structure from SSRs or SNPs

Word count for main text: 6966

Abstract

During the last years, simple sequence repeats (SSRs, also known as microsatellites) and single nucleotide polymorphisms (SNPs) have become the most popular molecular markers for describing neutral genetic variation in populations of a wide range of organisms. However, only a limited

number of studies has focused on comparing the performance of these two types of markers for describing the underlying genetic structure of wild populations. Moreover, none of these studies targeted fungi, the group of organisms with one of the most complex reproductive strategies. We evaluated the utility of SSRs and SNPs for inferring the neutral genetic structure of *Armillaria cepistipes* (basidiomycetes) at different spatial scales. For that, 407 samples were collected across a small (150 km^2) area in the Ukrainian Carpathians and a large (41,000 km^2) area in the Swiss Alps. All isolates were analyzed at 17 SSR loci distributed throughout the whole genome and at 24 SNP loci located in different single-copy conserved genes. The two markers showed different patterns of structure within the two spatial scales studied. The multi-allelic SSR markers seemed to be best suited for detecting genetic structure in indigenous fungal populations at a rather small spatial scale (radius of about 50-100 km). The pattern observed at SNP markers rather reflected ancient divergence of distant (about 1000 km) populations that in addition are separated by mountain ranges. Despite these differences, both marker types were suitable for detecting the weak genetic structure of the two *A. cepistipes* populations investigated.

Keywords: population genetic structure, genetic markers, genetic diversity, population differentiation, *Armillaria.*

INTRODUCTION

Investigating the population genetics of fungi may be a challenging task because of the complexity of their mating systems (e.g. Nieuwenhuis and

James, 2016). Besides sexual reproduction and clonal (asexual) spread, diploid-haploid mating resulting in the formation of a recombinant nucleus through complex events may also be observed. These different mating processes can occur in parallel and have contradictory impacts on traditional population genetic estimators, like heterozygosity (De Meeûs *et al.*, 2007). Potentially, this may lead to false inferences, e.g. revealing subdivisions in populations which are actually not subdivided or extensive genetic exchange when it is completely absent. In the recent past, in addition to traditional population genetic estimators, a range of clustering methods for probabilistically assigning individuals to genetic clusters have been developed (Pritchard *et al.*, 2000; Jombart *et al.*, 2010). Although these methods have been used with a variety of molecular markers, so far only a limited number of studies exist that have compared the particular outcome of depicting population genetic structure with different markers types, particularly in fungal study systems. On one hand, the different reproductive processes mentioned above act on the entire nuclear genome, therefore, we expect them to have a similar impact on different marker types. On the other hand, although these clustering methods are based on allele frequencies and do not directly include the characteristics (like mutation rate or ascertainment bias, see below) of the applied markers, we might expect differences in population structure inference when using them, because different marker types can exhibit different spectra of allele frequencies (see, e.g., Fischer at al. 2017).

During the last two decades, simple sequence repeats (SSRs, also known as microsatellites) and single nucleotide polymorphisms (SNPs) have become the most popular molecular markers for describing genetic variation in natural populations of a wide range of organisms with different biology. At the same time, an intense debate has arisen about the robustness and consistency of the results obtained with these two types of markers (Väli *et al.*, 2008; Coates *et al.*, 2009; Ljungqvist *et al.*, 2010; Guichoux *et al.*, 2011; Fischer *et al.*, 2017). Both SSRs and SNPs are abundant in the genome of most organisms and therefore potentially useful for detecting the population genetic structure and reconstructing the evolutionary history of species. However, because of different mutation rates and mechanisms, genome-wide distribution patterns, and biological functions, the use of SSRs and SNPs may lead to substantially different outcomes when testing specific hypotheses (Banke and McDonald, 2005; Coates *et al.*, 2009; Fischer *et al.*, 2017).

SSRs are generally abundant and polymorphic in non-expressed genomic regions and consequently considered to be selectively neutral. However, SSR loci can also occur in regions of chromosomes (e.g. telomeres or centromeres) involved in gene transcription, translation, chromatin organization, or recombination (Li *et al.*, 2002) and might therefore be under selection. Neutral SSRs evolve rapidly without vital consequences for the organisms. Due to replication slippage, SSR loci mutate from 10 to 100 thousand times more frequently per generation than single nucleotide substitutions occur (Guichoux *et al.*, 2011). Their high mutation rates and

assumed neutral evolution allow the accumulation of numerous population-specific (i.e. private) alleles, which are important for revealing hidden population structure. Their multi-allelic nature also results in a higher probability to detect heterozygosity than, for instance, an equal number of bi-allelic markers. However, the unusual high variability of SSRs in respect to other genomic regions might not necessarily reflect patterns of genome-wide genetic diversity (Väli *et al.*, 2008; Ljungqvist *et al.*, 2010; Fischer *et al.*, 2017). Moreover, the rapid mutation rates of SSRs may also confound signals of population structuring and divergence. For instance, the frequent forward and backward mutations of SSR loci can create identical alleles in populations that are unrelated or genetically isolated (i.e. homoplasy). This undesirable effect can be compensated by increasing the number of polymorphic SSR loci used, but the level of genetic differentiation of populations that diverged long time ago could still be underestimated (Estoup *et al.*, 2002).

In recent years, SNPs have started to replace SSRs in population genetic studies as well as in a wide range of other applications (Brumfield *et al.*, 2003; Guichoux *et al.*, 2011). SNPs occur twice as frequently in intergenic and non-coding regions of the genome than in coding regions (Zhao *et al.*, 2003). However, genome-wide association studies revealed that SNPs located in non-coding regions are often physically linked to functional or regulatory genomic sites, thus reflecting, for example, selection signatures (Kim *et al.*, 2007). Given that SNPs are mostly bi-allelic, traditional population genetic statistics can easily be applied to them but a higher number of loci

sufficiently polymorphic might be necessary to reach the same power as multi-allelic SSR loci (Guichoux *et al.*, 2011). The advent of next generation sequencing (NGS) techniques has considerably accelerated, simplified and automated genome-wide SNP detection and genotyping. However, considering that also a relatively low number of highly polymorphic SNPs can potentially give a similar genetic resolution as randomly chosen and multi-allelic SSRs (Kaiser *et al.*, 2016), an alternative strategy to genome-wide SNP screening might be targeting polymorphic sites in unlinked single copy genes, generally known to be conserved in the targeted phylum (Dutech *et al.*, 2016).

In the present study, we aimed at evaluating the use of SSRs and SNPs for inferences of neutral genetic population structure in the fungus *Armillaria cepistipes* (Basidiomycota, Physalacriaceae). Fungi of this genus belong to the most important and frequent component of the wood-decaying mycoflora in natural and managed forest ecosystems worldwide (Baumgartner *et al.*, 2011). Moreover, several *Armillaria* species can act as primary or secondary pathogens causing root rot on a wide range of tree and shrub species, thereby affecting crop and timber plantations. *Armillaria* has both sexual and asexual (vegetative) reproductive strategies. In addition, diploid-haploid mating was observed for this fungi *in vitro* (Carvalho *et al.*, 1995). This can create new genotypes in natural populations and potentially alter population structure. Due to vegetative propagation, genotypes (i.e. genets) of *Armillaria* may persist over centuries and even millennia in natural forests. Because of this high

longevity and the stable territoriality, *Armillaria* may influence forest structure and dynamics over several generations of host trees. The preferentially saprophytic *A. cepistipes* is widely distributed in Europe and frequently occurs in mountain forests from the Alps to the Carpathians (Heinzelmann *et al.*, 2012; Tsykun *et al.*, 2012). *Armillaria cepistipes* spreads locally by producing a dense network of vegetative rhizomophs in the soil, through which it can rapidly colonise new food bases (e.g. fresh stumps or wood debris). The fungus might also spread over longer distances with wind via basidiospores. These two different strategies of dispersal might have an impact on population genetic structure and different genetic markers may therefore be necessary to reveal it at different spatial scales.

We developed two sets of molecular markers for *A. cepistipes*; 24 SNPs from 24 single-copy genes and 17 SSRs partially derived from a draft genome assembly. We tested these markers on populations with different spatial scales from two European mountain ranges, i.e. a large-scaled population scattered over the Swiss Alps and a small-scaled population in two forest locations of the Ukrainian Carpathians (Figure 1). The questions we addressed were: (1) Are the two sets of developed markers sufficient for depicting population genetic structure? (2) Do SNP and SSR markers produce similar results regarding genetic differentiation and structure, and if not, (3) what are the possible reasons for that?

MATERIALS AND METHODS

Samples of *Armillaria cepistipes*

168 In total, 407 samples of *A. cepistipes* were used in this study, all originating
169 either from the Carpathians or the Alps (Figure 1). In the Carpathians,
170 samples were collected across 150 km^2 in two different protected forests as
171 described in Tsykun *et al.* (2012), i.e. 121 samples within 40 plots in a
172 virgin pure beech (*Fagus sylvatica*) forest and 160 samples in 35 plots in a
173 natural mixed forest with conifers (*Picea abies, Abies alba, Acer*
174 *pseudoplatanus, Fagus sylvatica*). These two Carpathian subpopulations
175 (forests) were about 50 km apart (Figure 1). The pure beech forest is
176 located between 400 and 1150 m a.s.l. The annual average temperature is
177 +7°C and the annual average precipitation sum is 948 mm. The natural
178 conifer and mixed forest lies between 700 and 1550 m a.s.l. The territory
179 has average temperatures from 0 to +7°C and the average precipitation
180 sum varies between 1000 and 1500 mm. In both Carpathian forests,
181 *Armillaria* was systematically sampled on a 1.5 x 1.5 km square grid
182 (Tsykun *et al.*, 2012). In each intersecting point of the grid, a 20 x 25 m
183 (500 m^2) plot was established and rhizomorphs were sampled from the soil
184 and from the root collar of trees at all four corners of the rectangle. In
185 Switzerland, an individual-based sampling approach was adopted as
186 described in Heinzelmann et al (2012). Here, the *A. cepistipes* samples
187 were collected in managed and unmanaged forests in an area of about 41
188 000 km^2 (Figure 1) covering all altitudinal zones, both north (i.e. 62 isolates
189 from the subpopulation Plateau in Heinzelmann *et al.* (2012)) and within or
190 south of the Alps (64 isolates from the subpopulations Southern, Western
191 Inner, Eastern Inner Alps in Heinzelmann *et al.* (2012)). These four

192 investigated subpopulations are named as follows: "Beech" and
193 "Mixed/conifer" (for the subpopulations of the Carpathians), and "North" and
194 "South" (for the subpopulations of the Alps).

195 DNA was extracted as described in Heinzelmann *et al.* (2012) and Tsykun
196 *et al.* (2012). All 407 isolates were genotyped with SSRs and SNPs.

197 **Screening and genotyping of SSRs and SNPs**

198 Briefly, 32 single copy protein-coding gene sequences of 12 *A. cepistipes*
199 samples were sequenced and screened for the presence of SNPs (Dutech
200 *et al.*, 2016). Based on the polymorphism detected in these sequences, 131
201 primer pairs amplifying each one SNP were initially designed and
202 dispatched in multiplexes with MASSARRAY ASSAY DESIGN 4.1 for
203 Sequenom screening. These multiplexes were tested with 48 samples from
204 the Ukrainian Carpathians and 47 samples from the Swiss Alps. SNPs in
205 genes that showed evidence to be present in multiple copies in the genome
206 of *A. cepistipes*, monomorphic loci (minor allele frequency < 5%), and loci
207 that were not amplified in the majority of the screened samples were
208 discarded. After considering the compatibility of the primers pairs, one
209 multiplex of 31 SNPs located in 24 single copy protein-coding genes was
210 selected for final genotyping (for more details, see Table S1.1 and Figure
211 S1.1, Appendix 1). SNP genotyping was performed using the medium
212 throughput MassARRAY iPLEX genotyping assay from Sequenom (San
213 Diego, CA, USA). Raw genotypic data were analyzed with TYPER 4.0
214 (Agena Biosciences). For 32 of the 407 DNA isolates genotyping was
215 repeated.

A total of 17 SSR loci were used in this study (Table S1.2, Appendix 1). Six of them (Arm02, Arm05, Arm09, Arm11, Arm15, Arm16) were previously specifically developed for *A. cepistipes* by Prospero *et al.* (2010). Two loci (Am109 and Am111) were initially developed for *A. mellea* (Baumgartner *et al.*, 2009) but successfully applied in *A. cepistipes* (Heinzelmann *et al.*, 2012). The remaining nine loci (AC18, AC31, AC37, AC12, AC16, AC04, AC38, AC22, AC34) were developed in the present study (for fragment sequences of the selected SSR loci and for annotations see Table S1.2, Appendix 1). An *A. cepistipes* draft genome assembly (G Sipos, unpublished) was screened for simple sequence repeat patterns with UNIPRO UGENE v1.170 (Okonechnikov *et al.*, 2012). Search parameters included a nucleotide repeat size of two to six and a minimum tandem length of eight repeats. Among the SSRs detected, 24 loci scattered over 14 major assembled scaffolds were selected. Di-, tri-, tetra-, and penta-nucleotide repeats were chosen. SSR loci that might be involved in chromosomal organization and recombination were avoided by ignoring tandem repeats close to telomere regions and repeats in a dense cluster, which might indicate centromere sites (Li et al. 2002).

For each of the 24 selected SSR loci, specific PCR reverse and forward primers were designed using UNIPRO UGENE v1.170 fulfilling the following criteria: GC content of 50-60%, start with G or C, primer size of 20-27 bp, similar annealing temperature (60 ± 3°C) of forward and reverse primer, and a resulting fragment size of 150-400 bp. Initially, each SSR locus was amplified using a forward primer with an M13pa

240 (CACGACGTTGTAAAACGAC) tail, the designed reverse primer, and a
241 universal fluorescent dye-labeled M13pa (Boutin-Ganache *et al.*, 2001).
242 Each primer pair was tested separately with four isolates from the Ukrainian
243 Carpathians and three isolates from the Swiss Alps. The PCR program was
244 set as follows: 95°C (5 min), then 28 cycles at 95°C (30 s) / 60°C (90 s) /
245 72°C (30 s), followed by 8 cycles 95°C (30 s) / 55°C (45 s) / 72°C (30 s),
246 and a final extension at 60°C for 30 min. PCR products were initially scored
247 for successful amplification on 1.2% agarose gels, and subsequently run on
248 an ABI 3130 Genetic Analyzer using GeneScan LIZ500 (Applied
249 Biosystems, CA, USA) as internal size standard. For each locus that was
250 successfully amplified with the M13pa labeled primer, the designed forward
251 primer was then labeled at the 5'-end with a fluorescent dye (6FAM-blue,
252 NED-yellow, VIC-green, PET-red, Applied Biosystems, CA, USA). Four
253 multiplexes including each three to four primer pairs were subsequently
254 developed (Table S1.2, Appendix 1). The final PCR was run as follows:
255 95°C (15 min), followed by 30 cycles at 94°C (30 s) / 60°C (90 s) / 72°C (30
256 s), and a final extension at 60°C for 30 min and 72°C for 30 min. The
257 amplified fragments were run on an ABI DNA Analyzer (Applied
258 Biosystems, CA, USA). Alleles were scored with GENEMAPPER 3.7 (Applied
259 Biosystems).

260 **Data analysis**

261 In order to assess and compare the utility of the two marker types, only the
262 isolates with an almost complete SSR and SNP dataset (i.e. not more than
263 5% missing data in both marker types) were included in the analyses. To

remove a putative clonal effect on the genetic structure, only one representative of each multi-locus genotype (MLG) per sampling plot was considered for genetic analysis for SNPs and SSRs separately.

In each population, number (Na) and evenness (E) of alleles per SSR locus, and observed and expected heterozygosity (H_{obs}, H_{exp}) for SSR and SNP loci were estimated using ARLEQUIN 3.5.2.1 (Excoffier *et al.*, 2009) and the R-package POPPR (Kamvar *et al.*, 2014). Allelic richness (Ar) per SSR locus was estimated using FSTAT 2.9.3.2 (Goudet, 2002) and rarefied to the population with the lowest sample size (i.e. the North Alpine subpopulation with 62 isolates). SSR and SNP loci were tested for significant departure from Hardy-Weinberg equilibrium (HWE) by conducting a Fisher's exact test with a Markov chain algorithm (1,000,000 chain steps, 100,000 de-memorization steps) in ARLEQUIN 3.5.2.1. Pair-wise linkage disequilibrium (LD) between loci was tested with the log-likelihood ratio using a Markov chain algorithm (default parameters) as implemented in the web version of GENEPOP 4.2 (Rousset, 2008). The statistical significance (for HWE and LD) was inferred using 1000 permutations and sequential Bonferroni correction with α=0.05. Multilocus linkage disequilibrium was evaluated based on the *P* values from one-sided permutation tests with the R-package POPPR for the indexes of association I_A and $rbar_D$ (Agapow and Burt, 2001). Genetic differentiation among populations was assessed by calculating pairwise F_{ST}-values (Weir and Cockerham, 1984) and corresponding *P* values (α=0.05) with ARLEQUIN 3.5.2.1.

Investigating the neutral genetic structure and demographic history of populations implies the use of loci whose population-specific allele frequencies are not affected by natural selection. Although SSR loci are commonly considered as neutral makers, recent studies suggest that this is not always the case (e.g. Li *et al.*, 2002). The SNPs used in this study originated from putative functional genes (Dutech *et al.*, 2016). Thus, allele frequencies at these loci might be driven by natural selection instead of reflecting neutral population processes. Therefore, the SSR and SNP loci used in this study were tested for deviation from neutral patterns using two statistical approaches as implemented in ARLEQUIN 3.5.2.1 (Excoffier *et al.*, 2009) and BAYESCAN v2.1 (Foll and Gaggiotti, 2008). The tests were conducted for the Carpathian and Alpine populations, ignoring subpopulations, and separately within the mountain ranges for the two subpopulations each. Both approaches assume a non-hierarchical finite island migration model.

The neutral genetic structure in the four subpopulations was investigated using two different approaches. First, a multivariate clustering method (discriminant analysis of principle components, DAPC) was applied that identifies synthetic variables and builds discriminant functions maximizing the variation between groups but minimizing it within groups (Jombart *et al.*, 2010). In a first step, genetic data from both markers sets separately were transformed into principal components (PCs) and the optimal number of PCs was assessed with cross-validation (Jombart *et al.*, 2010). Thereafter, we predefined geographic groups (i.e., the four subpopulations), following

311 the approach previously applied in a population genetic study of the
312 oomycete pathogen *Phytophthora infestans* (Wang *et al.*, 2017). The DAPC
313 then probabilistically assigned individuals to one of the four pre-defined
314 groups. The DAPC was performed with the R package ADEGENET (Jombart
315 *et al.*, 2010).

316 We also performed a Bayesian model-based cluster analysis with
317 STRUCTURE 2.3.4 that uses allele frequencies at each locus to
318 probabilistically assign individuals to genetic clusters. We used sampling
319 locations of the subpopulations as prior geographic information
320 (LOCPRIOR), the admixture ancestral model with correlated allele
321 frequencies, and indication of null alleles (RECESSIVEALLELES settings)
322 in locus AC18. Analyses were run with 200,000 burn-in iterations followed
323 by the same number of iterations for Markov chain Monte Carlo (MCMC) in
324 ten independent runs for each number of clusters (K) from one to 20. The
325 most likely number of clusters (K) was determined by (a) considering the
326 maximal mean and small standard deviation of the posterior probability of K
327 among runs (Pritchard *et al.*, 2000) using STRUCTURE HARVESTER (Earl
328 and vonHoldt, 2011) and (b) looking at the alterations of individual
329 assignment probabilities with increasing K (i.e. whether additional clusters
330 were reflected by whole individuals or whether they rather resulted in
331 individuals represented by several clusters). Average assignment
332 probabilities of MLGs to the clusters were computed with CLUMPP 1.1.2
333 (Jakobsson and Rosenberg, 2007) using the Greedy algorithm for K≥10

and visualized using DISTRUCT 1.1 (Rosenberg, 2004) and R graphic functions.

RESULTS

SNP and SSR marker sets

A total of 117 (89%) of the 131 detected SNP loci in the 24 single copy protein-coding genes were successfully amplified in the majority (78%) of the samples. Among these, 31 were selected for genotyping (Table S2.1, Appendix 2) based on the criteria described in the Materials and Methods section. In the selected set of loci, genotyping was successful in 97% of the samples and repeated genotyping of 32 samples revealed less than 1% difference in allele calling. All 31 SNP loci were bi-allelic, and in 29 of them the minor allele frequency was above 5%.

Based on the selection criteria given in the Materials and Methods section, a total of 281 SSR loci were found across the draft genome assembly (total length of about 75 Mbp; Münsterkötter *et al.*, 2015). Twelve out of the 24 loci, for which specific primers with the M13pa modification at the 5'-end were designed, successfully amplified a PCR product of the expected size. However, three of them were either monomorphic or did not yield a PCR product in the three final multiplexes using the labeled forward primers, and were consequently not used for further analyses. Hence, in combination with the eight previously developed markers (Baumgartner *et al.*, 2009; Prospero *et al.*, 2010), the samples were genotyped at totally 17 polymorphic SSR loci scattered all over the genome (Figure S1.1, Appendix

1). All SSR loci revealed a considerable level of polymorphism with a frequency of the most common allele lower than 95% (Table S2.1, Appendix 2). Missing data for both sets of markers were observed in less than 5% of the samples across loci and populations, except for the SSR locus AC18 which revealed 15% missing data in the Carpathian and almost 40% in the Alpine populations. Multilocus genotypes with missing data in loci other than AC18 were excluded from further analyses.

In addition to the high percentage of missing data, SSR locus AC18 also revealed a significant deficiency of heterozygotes and a statistically consistent deviation from Hardy-Weinberg equilibrium in all four subpopulations (Table S2.1, Appendix 2). This is strong evidence for the presence of null alleles. However, since null allele frequencies at this locus differ between Carpathian and Alpine populations and thus can be used for the STRUCTURE analysis (RECESSIVEALLELES settings), we kept it for further tests.

The linkage disequilibrium test showed that a significant correlation between allele frequencies among all subpopulations was only present at SNPs which are located in the same gene. Thus, for population structure analyses, only one random SNP locus per gene was considered, resulting in a total of 24 SNPs. Non-random association of allele frequencies was also observed for other pairs of SNP and SSR loci in specific subpopulations, but the statistical support was not consistent across loci and subpopulations (Table S2.2, Appendix 2). Noteworthy, these rarely pair-wise correlated loci were located in different scaffolds across the

genome (Figure S1.1, Appendix 1). The two multilocus estimators of non-random association between alleles (I_A and $rbar_D$) were low across all subpopulations and markers (Table 1). However, in the two Carpathian subpopulations, non-random associations in SSR loci were supported by significant *P* values.

BAYESCAN revealed no F_{ST} outliers in both SSRs and SNPs (Figures S3.1B and S3.2B, Appendix 3), suggesting that all loci were selectively neutral in the studied populations and subpopulations. In contrast, ARLEQUIN showed locus-specific F_{ST} and heterozygosity at the SSR locus AC31 (95% confidence interval) when analyzing the Carpathian subpopulations separately (Figure S3.3A, Appendix 3), which might indicate that natural selection acted on this locus in Carpathian *A. cepistipes*. Similarly, SNP loci FG848_7 and FG894_7 were identified as F_{ST} outliers (99% confidence interval) with ARLEQUIN in the two Alpine subpopulations (Figure S3.6A, Appendix 3). In the overall analysis, assuming a finite island migration scenario between the Carpathian and Alpine populations, ARLEQUIN revealed no SSR locus being under selection, whereas the two SNP loci MS481_16 and FG730_11 showed evidence of being under purifying selection (Figure S3.2A, Appendix 3). Because these analyses revealed no consistent F_{ST} outliers among statistical approaches and population sets, all loci were considered as neutral and kept for further analysis.

Genetic diversity

Among the 407 isolates genotyped with 17 SSRs, 359 exhibited a unique combination of alleles, whereas only 278 unique SNP MLGs were detected

405 (Table 1). All 17 SSR loci were polymorphic in the investigated *A.*
406 *cepistipes* populations and harbored a total of 135 alleles. Eighteen alleles
407 were specific for the Carpathian population, whereas only 11 alleles were
408 detected in the Alpine population. Within the Carpathian population, each
409 subpopulation harbored five private alleles. The two Alpine subpopulations
410 exhibited five (North) and three (South) private alleles. Mean allelic
411 richness (Ar) across SSR loci was similar in all four subpopulations (Table
412 2), but varied considerably among loci (Table S2.1, Appendix 2). Two of the
413 31 SNP loci were polymorphic in only the Carpathian population (Table
414 S2.1, Appendix 2). Just as for Ar, evenness (E) of alleles and expected
415 heterozygosity (H_{exp}) across SSR and SNP loci showed only little variation
416 among subpopulations (Table 2), but high variation among loci (Table S2.1,
417 Appendix 2). Within subpopulations, SSR loci exhibited higher H_{exp} than
418 SNP loci (Table 2).

419 **Population structure**

420 Pairwise F_{ST} values between subpopulations were low for both types of
421 molecular markers (across SSR loci: from -0.01 to 0.08; across SNP loci:
422 from -0.01 to 0.17), and even between the two substantially distant (> 1000
423 km apart) Alpine and Carpathian populations (SSR-F_{ST} = 0.025, SNP-F_{ST} =
424 0.026). Despite the generally low values of F_{ST} (resulting in $P < 0.05$), those
425 computed between geographically distant subpopulations, e.g. Alpine
426 South and Carpathian Beech, were about two (SNPs) to ten (SSRs) times
427 higher than those between subpopulations within a mountain range (Table
428 3). An extremely low and statistically non-significant F_{ST} value was obtained

429 with SNPs markers between the two subpopulations of the Carpathians
430 (Beech and Mixed/conifer) which were sampled at small spatial scale. In
431 contrast, SSR markers revealed a non-significant F_{ST} value between the
432 two Alpine subpopulations which were randomly sampled at a large scale
433 (Table 3).

434 In the overall STRUCTURE analysis using all four subpopulations, SSRs
435 revealed a larger number of genetic clusters than SNPs (Figure 2). For the
436 SSR data, the log-likelihood increased constantly up to 20 clusters (Figure
437 2A). However, the standard deviations of the log-likelihood values
438 increased simultaneously and became particularly high for some $K \geq 5$, like
439 $K = 9$ and $K = 12$. Up to $K = 20$, increasing K led to individuals assigned to
440 new clusters with high probability and revealed a complex structure within
441 the small-scaled Carpathian subpopulations (Figure 2B, Figure S4.1,
442 Appendix 4). For the SNP data, the log-likelihood was highest at $K = 2$ and
443 high standard deviations of log-likelihood values suggested unstable results
444 for $K > 3$ (Figure 2A). Unlike in the SSR dataset, increasing $K > 3$ did not
445 assign fungal MLGs to new clusters and did not suggest a more complex
446 genetic structure within any of the studied subpopulations. However, $K = 3$
447 supports the existence of two large-scaled Alpine subpopulations and one
448 Carpathian population (Figure 2B, Figure S4.2, Appendix 4). Additional
449 STRUCTURE analyses using only the two Alpine subpopulations (Figure
450 S4.4, Appendix 4) showed the highest log-likelihood at $K = 2$ and confirmed
451 the assignment pattern. Considering the two methods of K determination
452 described above (highest log-likelihood and sequential cluster assignment

453 of the MLGs) and in order to avoid an overestimation of the number of
454 genetic clusters in the studied populations, we assumed K = 4 as
455 reasonable to describe the genetic structure with the SSR markers and K =
456 3 as the most likely number of genetic clusters with the SNP markers.

457 The four genetic clusters identified using SSRs split the 359 MLGs
458 according to their geographic origin (Figure 2B, SSRs, K=4). All MLGs of
459 the two Alpine subpopulations belonged to the same (green) cluster,
460 whereas those from the two Carpathian subpopulations were mainly
461 attributed to one (blue) of the three other clusters. However, in these two
462 latter subpopulations a significant mixture of MLGs that belonged to the
463 remaining two clusters (yellow and orange) was also observed. The
464 STRUCTURE analyses revealed an almost complete absence of MLGs with
465 admixed genetic origin in the large and randomly sampled population from
466 the Alps. In contrast, such admixed MLGs were frequent in the two
467 subpopulations of the Carpathians sampled across a smaller spatial scale,
468 especially in the mixed and conifer forests. Most admixed MLGs from the
469 orange cluster were present in this specific subpopulation (Figure 2B,
470 SSRs, K=4).

471 The STRUCTURE analysis with the SNP data for K = 3 (Figure 2B, SNPs,
472 K=3) revealed that the Carpathian and Alpine populations were clearly
473 separated, with the first population including mainly MLGs from one cluster
474 (green, membership probability of 60-80%) and the latter from two clusters
475 (blue and yellow). Noteworthy, the pattern of subdivision within populations
476 was the opposite as the one observed with SSRs. The small-scaled

477 subpopulation of the Carpathian Mountains was homogeneous, whereas in
478 the large Alpine population the two subpopulations (North and South) were
479 clearly differentiated (Figure 2B, SNPs, K=3). This signal of differentiation
480 between the two Alpine subpopulations substantially disappeared when the
481 five SNPs (loci: MS481_16, FG730_11, FG848_7, FG894_7, FG524_2)
482 that showed deviation from neutral patterns (Figure S3.2, Figure S3.6,
483 Appendix 3) were excluded from the cluster analyses (Figure S4.3,
484 Appendix 4).

485 Cluster assignments with both types of loci were also examined with a
486 discriminant analysis of principal components (DAPC, Figure 3). Based on
487 lowest root mean squared error and highest mean of successful
488 reassignments with 1000 replicates (cross-validation), 90 (of 110) and 20
489 (of 22) computed principal components were retained in the discriminant
490 analysis using SSRs and SNPs, respectively. Three discriminant functions
491 were built for each analysis. In both types of markers, no strong association
492 was detected between MLGs assignments and their geographic origin. The
493 individual posterior probabilities of assignment to a predefined geographic
494 group were low for both the SSR and SNP data in all studied
495 subpopulations. However, the proportion of MLGs successfully assigned
496 (with posterior probability >95%) to the predefined geographic
497 subpopulation differed between the genetic markers (SSRs 35% of studied
498 MLGs; SNPs 0.36%, i.e. only one MLG).

499 Overall, the DAPC clustering agreed with the one of STRUCTURE (Figures 2
500 and 3). In both SSRs and SNPs, the first discriminant separated the MLGs

of the Carpathians and the Alps (Figure 3). With the SSR data, the two subpopulations of the Carpathians were further discriminated along the second axis, whereas the individuals from the two Alpine subpopulations largely overlapped (Figure 3). The opposite situation was observed in the SNP data, where the least overlap was observed between the two Alpine subpopulations. One noteworthy difference between the STRUCTURE and DAPC clustering could be observed with the SSR data (Figures 2 and 3). While STRUCTURE revealed a complex structure within the Carpathian subpopulations, the DAPC identified prominently differences between Beech and Mixed/conifer forests.

DISCUSSION

In this study, we aimed to compare the utility of SNP and SSR markers for investigating the neutral genetic population structure of the basidiomycete *A. cepistipes* at different spatial scales. Analyzing the population structure of such an organism implies also addressing several other issues, such as the contribution of different reproduction modes to its spread at large and small geographic scales, the connectivity among populations in a heterogeneous environment, and their demographic history. Our analysis revealed differences in the information that the two types of markers give about genetic structure within large and small geographic scales based on the example of two populations from mountain forests in Europe. Noteworthy, SSRs provided a higher resolution at a smaller geographic scale under a systematic sampling (Carpathian population), whereas SNPs

were able to differentiate the two subpopulations which were randomly sampled across a large area in the Alps.

Both types of markers revealed the presence of repeatedly occurring genotypes in the investigated populations. However, a higher number of MLGs was detected when using the 17 SSRs (359 MLGs in 407 isolates) than when using the 24 SNPs (278 MLGs). This result confirms that multi-allelic SSR markers have a higher discrimination power than bi-allelic SNP markers (Guichoux *et al.*, 2011). High levels of genotypic diversity are generally expected in populations of fungi that mainly reproduce sexually. However, previous population genetic studies conducted in natural and managed forests in Europe showed that this is not always the case in *Armillaria* species (Prospero *et al.*, 2008). Because of the spread via vegetative rhizomorphs, *Armillaria* species may produce large genets that occupy a forest area of several hectares (Bendel *et al.*, 2006). Thus, the presence of such genets can reduce genotypic diversity. Moreover, mating of closely related haploids produced by spatially distributed clones via basidiospores during sexual reproduction will influence population structure and consequently change statistical estimators, (e.g. I_A and $rbar_D$) even after the clone-correction procedure.

Significant differences between SNPs and SSRs were observed with respect to heterozygosity (paired t-test, n=4, $p<0.001$), but not to F_{ST} (paired t-test, n=4, $p=0.30$). As expected due to their multi-allelic nature and usually higher level of polymorphism, SSR loci exhibited a significantly higher heterozygosity than bi-allelic SNP loci. However, locus-specific

548 values showed a wide range, possibly because of uneven allelic richness.
549 Previous studies argued that both high and low numbers of alleles at SSR
550 loci may affect the accuracy of heterozygosity estimates and consequently
551 of population-specific fixation indexes (Wang, 2015; Fischer *et al.*, 2017). In
552 our study, SSRs were selected with an emphasis on different nucleotide
553 number and GC content in the tandem repeats. Therefore, the
554 ascertainment bias due to the selection of genomic fragments with
555 exclusively high levels of polymorphism should be minor. For this reason,
556 allelic richness and heterozygosity estimators vary considerably among
557 loci.

558 In *Armillaria* populations, heterozygosity may also be strongly influenced by
559 the mixed mating system. Clonal reproduction by rhizomorphs maintains
560 heterozygosity (De Meeûs *et al.*, 2007), whereas sexual reproduction
561 increases it. In *Armillaria* species, spore release is intense but spore
562 dispersal seems to be spatially limited (Travadon *et al.*, 2012; Dutech *et al.*,
563 2017). This may lead to inbreeding processes like mating between closely
564 related haplotypes and plasmogamy of haploid spores (or mycelium) with
565 their diploid parents, which may both reduce heterozygosity. As inbreeding
566 and outbreeding processes can occur simultaneously in *Armillaria*
567 populations, heterozygosity may not accurately explain demographic
568 processes (e.g. gene flow between populations or a Wahlund effect due to
569 population subdivision) in these fungi, regardless whether SNPs or SSRs
570 are used. Nonetheless, in the large- and small-scaled *A. cepistipes*
571 populations, we observed a heterozygote deficit at most loci for both types

572 of genetic markers, suggesting a predominance of inbreeding processes.
573 However, the high abundance of rhizomorphs (Tsykun *et al.*, 2012) and
574 presence of repeated MLGs within different localities, also the low but
575 significant indexes of multilocus association (I_A and $rbar_D$) in the
576 Carpathians suggest that clonal reproduction also might influence
577 demographic processes in this population. Therefore, we can assume that
578 this population is driven by inbreeding processes along with clonal
579 spreading via rhizomorphs. However, long-distance spore dispersal cannot
580 be excluded and is indirectly supported in our study by the low number of
581 private alleles, the lack of a strong structure between subpopulations within
582 mountain ranges (regardless of sampling design or spatial scale) and the
583 low differentiation between the geographically distant Carpathian and
584 Alpine populations (see below).

585 Pairwise F_{ST} values between the studied fungal populations and
586 subpopulations, even between geographically distant ones (like Alpine vs.
587 Carpathian), were low (0.001 – 0.036) with both types of markers. It is
588 known that F_{ST} is very sensitive to the level of within-population variation,
589 resulting in suspiciously low values in SSR studies and a consequent
590 underestimation of the level of population divergence (Brumfield *et al.*, 2003).
591 However, low pairwise F_{ST} values in *Armillaria* (Giraud *et al.*, 2008;
592 Baumgartner *et al.*, 2010; Heinzelmann *et al.*, 2012), as well as in other
593 fungi (Giraud *et al.*, 2008), are rather the rule than the exception. This
594 suggests low overall population differentiation due to extensive gene flow
595 among populations. Although such high gene flow may be present between

596 the two subpopulations of the rather continuous Carpathian primeval
597 forests, it is less realistic between the two Alpine subpopulations and
598 between the Alpine and Carpathian populations, which are separated by
599 the Alps and a large geographic distance, respectively. An alternative
600 explanation for the low F_{ST} values may be a common glacial refugium of the
601 *A. cepistipes* populations, possibly coupled with a relatively slow population
602 divergence and homoplasy events. Recently, several authors (e.g. Jost,
603 2008; Meirmans and Hedrick, 2011) have criticized the use of F_{ST} as a
604 measure of population differentiation. Since this estimator seems to be
605 negatively correlated with the number of alleles per locus, F_{ST} tends to have
606 values towards zero in populations with high allelic richness and thus
607 underestimates the actual divergence between populations (Jost, 2008). In
608 our study, low F_{ST} values were not only detected with the multi-allelic SSR
609 markers, but also with the bi-allelic SNPs. In a comparative study by
610 Fischer *et al.* (2017), substantially higher F_{ST} values were obtained with a
611 limited number of SSRs than with genome-wide SNPs. The authors
612 emphasized that pairwise F_{ST} calculated from SSRs must be used with
613 caution. In particular, one should not rely on absolute values because it can
614 reflect rather the highly polymorphic nature of the markers than a real
615 whole-genome differentiation of populations. In our study, we used a limited
616 number of SNPs, which were selected because they exhibited a sufficient
617 level of polymorphism. Thus, this specific set of SNPs might induce an
618 ascertainment bias and show a higher population differentiation than
619 genome-wide SNPs that contain many low-frequency alleles. Regardless of

620 the overall low absolute F_{ST} values, for both types of markers, the genetic
621 differentiation between distant populations (Carpathian vs. Alpine) was
622 substantially higher than the one between subpopulations of the same
623 mountain region. Using SNPs, the highest differentiation was observed
624 between the most distant subpopulations that are separated by a high
625 mountain range, i.e. between both Carpathian subpopulations and the
626 subpopulation of the South Alps. In contrast, with the SSR data, the F_{ST}
627 value for the subpopulations sampled at a small spatial scale, i.e. Beech
628 and Mixed/conifer subpopulations of Carpathians, was higher (and
629 significantly different from zero) than the one from the SNP data.

630 Overall, the two clustering methods (DAPC and STRUCTURE) used for
631 investigating population genetic structure produced consistent results. The
632 geographically distant populations (i.e. Carpathian and Alpine) showed a
633 clear separation with both types of markers. SSRs and SNPs, however,
634 gave different signals within the two populations sampled at different spatial
635 scales. SSRs exhibited a considerable admixture of clusters in the two
636 geographically close and systematically sampled subpopulations of the
637 Carpathian forests in the STRUCTURE analysis, suggesting the same
638 ancestral origin and/or possible gene flow among populations. The DAPC
639 analysis, however, was able to define genetic components that differ
640 between the Carpathian subpopulations, suggesting a weak subdivision.
641 Because in this region the landscape barriers for spore dispersal are
642 relatively weak (e.g. low mountain relief and small distance between the
643 studied forests), genetic exchange between these subpopulations is a

644 realistic scenario. However, the clustering with SSRs showed also that
645 fungal populations sampled within a small-scaled area might have a
646 complex genetic structure. The mainly monocultural beech forests of the
647 Carpathians seem to harbor a more homogenous *A. cepistipes* population
648 which were resolved into only two genetic clusters in the STRUCTURE
649 analysis, whereas the mixed and pure conifer forests contain a more
650 diverse *A. cepistipes* population (three clusters). This result, however, is
651 different using SNPs with both DAPC and STRUCTURE, evidencing only one
652 single genetic cluster across both spatially close subpopulations of the
653 Carpathians. It is important to note that these SNPs were initially selected
654 from housekeeping genes present in the genomes of five fungal species
655 other than *Armillaria* (Dutech *et al.*, 2016). Therefore, SNPs in such
656 conserved genes may rather reflect long-term divergence among
657 populations than recent processes. Apparently, the two Carpathian
658 subpopulations have not yet diverged enough to reveal nucleotide
659 differences in the genes considered.

660 In contrast to using SSRs, the two large-scaled subpopulations that are
661 separated by a high mountain range (North and South of Alpine population)
662 were assigned to two different clusters using SNPs. The two SNP loci that
663 were mainly responsible for this discrimination were also significant outliers
664 in ARLEQUIN (but not in BAYESCAN). This suggests that the Alpine mountain
665 range left its traces on the long-term divergence of the northern and
666 southern *A. cepistipes* subpopulations. The presence of only one genetic
667 cluster in the large-scaled Alpine population based on SSRs might be at

668 least partially due to the particular sampling design applied. The two Alpine
669 subpopulations mainly share the same alleles at all 17 SSR loci. Thus, a
670 random sampling of distant individuals at a large spatial scale may not
671 accurately reveal local population allele frequencies to infer subpopulation
672 structure with SSRs. In contrast, scattered sampling at large scale did not
673 affect the discrimination power of SNPs. This is most likely because
674 differences among geographically distant populations in SNP loci were
675 fixed along an evolutionary time scale, making it easier to detect population
676 specific allele frequencies even with a scattered random sampling. Our
677 results are in agreement with those of a study on the global migration
678 patterns of the pathogenic crop fungus *Mycosphaererella graminicola*
679 (Banke and McDonald, 2005). The authors found that SSRs were sensitive
680 to detect recent (50-150 years) migration events between North and South
681 American populations, whereas protein-coding sequences loci were not.
682 Based on these and on our results, we conclude that SSRs have a higher
683 resolution for genetically and spatially close populations (as in the
684 Carpathian subpopulations) with extensive gene flow. In contrast, SNPs in
685 housekeeping genes seem to be more appropriate for phylogeographic
686 large-scale studies (as in the Alpine subpopulations). However, this
687 conclusion should be treated with caution, because our study design does
688 not allow disentangling the possible confound effects of sampling scale and
689 population-specific demography on marker performance.

690 In summary, the present study revealed differences on inferences of
691 population genetic structure of the fungus *Armillaria cepistipes* at different

692 spatial scales when using two different marker types (SSRs and SNPs).
693 SSRs were found to be better suited for detecting structure in populations
694 at a small spatial scale with a systematic and continuous sampling design
695 (as shown in the example of the Carpathian population). The patterns
696 observed in the SNP markers rather reflect ancient divergence of distant
697 and naturally separated populations, being less sensitive to sampling
698 design (as shown in the example of the Alpine population). A full factorial
699 sampling design and a higher genomic resolution would help to strengthen
700 the reliability of the obtained results. Nevertheless, both marker types were
701 suitable for detecting weak genetic structure of the two fungal populations
702 considered.

703 **ACKNOWLEDGEMENTS**

704 We are grateful to E. Guichoux and A. Delcamp (Genome Transcriptome
705 Facility of INRA, Bordeaux), E. Jung (Phytopathology, WSL), and the team
706 of the Genetic Diversity Center (GDC, ETH Zürich) for laboratory
707 assistance and expert help. We also thank R. Holderegger for useful
708 discussions while analyzing and interpreting results, and three anonymous
709 referees for the critical comments on the manuscript. This work was
710 financially supported by the Swiss Federal Research Institute WSL, by the
711 program "Trees4Future", a project co-funded by the European Union
712 Seventh Framework FP7 under grant agreement n° 284181, and by the
713 Swiss Government Excellence Scholarships for Foreign Scholars and
714 Artists for the Academic Year 2014-2015 of the State Secretariat for
715 Education, Research and Innovation (SERI) to T. Tsykun. The *A. cepistipes*

716 genome project was founded by the European Union in the frame of the
717 Széchenyi 2020 Programme (GINOP-2.3.2-15-2016-00052) and by the
718 WSL to G. Sipos.

719 **DATA ARCHIVING**

720 Genotype data have been submitted to Dryad:

721 **CONFLICT OF INTEREST**

722 The authors declare no conflict of interest.

723 **REFERENCES**

724 Agapow P-M, Burt A (2001). Indices of multilocus linkage disequilibrium.
725 *Molecular Ecology Notes* **1:** 101-102.

726 Banke S, McDonald B (2005). Migration patterns among global populations of the
727 pathogenic fungus *Mycosphaerella graminicola*. *Molecular Ecology* **14:** 1881-1896.

728 Baumgartner K, Coetzee MPA, Hoffmeister D (2011). Secrets of the subterranean
729 pathosystem of *Armillaria*. *Mol Plant Pathol* **12:** 515-534.

730 Baumgartner K, Grubisha LC, Fujiyoshi P, Garbelotto M, Bergemann SE (2009).
731 Microsatellite markers for the diploid basidiomycete fungus *Armillaria mellea*.
732 *Mol Ecol Resour* **9:** 943-946.

733 Baumgartner K, Travadon R, Bruhn J, Bergemann SE (2010). Contrasting patterns
734 of genetic diversity and population structure of *Armillaria mellea sensu stricto* in
735 the Eastern and Western United States. *Phytopathology* **100:** 708-718.

736 Bendel M, Kienast F, Rigling D (2006). Genetic population structure of three
737 *Armillaria* species at the landscape scale: a case study from Swiss Pinus mugo
738 forests. *Mycol Res* **110:** 705-712.

739 Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001). M13-tailed
740 primers improve the readability and usability of microsatellite analyses performed
741 with two different allele-sizing methods. *Biotechniques* **31:** 24-26, 28.

742 Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003). The utility of single
743 nucleotide polymorphisms in inferences of population history. *Trends in Ecology &*
744 *Evolution* **18:** 249-256.

745 Carvalho DB, Smith M, Anderson J (1995). Genetic exchange between diploid and
746 haploid mycelia of *Armillaria gallica*. *Mycol Res* **99:** 641-647.

747 Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD *et al.*
748 (2009). Comparative performance of single nucleotide polymorphism and
749 microsatellite markers for population genetic analysis. *Journal of Heredity* **100:**
750 556-564.

De Meeûs T, McCoy KD, Prugnolle F, Chevillon C, Durand P, Hurtrez-Bousses S *et al.* (2007). Population genetics and molecular epidemiology or how to "débusquer la bête". *Infection, Genetics and Evolution* **7:** 308-332.

Dutech C, Labbé F, Capdevielle X, Lung-Escarmant B (2017). Genetic analysis reveals efficient sexual spore dispersal at a fine spatial scale in *Armillaria ostoyae*, the causal agent of root-rot disease in conifers. *Fungal Biol*: 10.1016/j.funbio.2017.03.001.

Dutech C, Prospero S, Heinzelmann R, Fabreguettes O, Feau N (2016). Rapid identification of polymorphic sequences in non-model fungal species: the PHYLORPH method tested in Armillaria species. *Forest Pathol* **46:** 298-308.

Earl DA, vonHoldt BM (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359-361.

Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular ecology* **11:** 1591-1604.

Excoffier L, Hofer T, Foll M (2009). Detecting loci under selection in a hierarchically structured population. *Heredity* **103:** 285-298.

769 Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK *et al.*
770 (2017). Estimating genomic diversity and population differentiation – an empirical
771 comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC*
772 *Genomics* **18:** 69.

773 Foll M, Gaggiotti O (2008). A Genome-Scan Method to Identify Selected Loci
774 Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective.
775 *Genetics* **180:** 977-993.

776 Giraud T, Enjalbert J, Fournier E, Delmotte F, Dutech C (2008). Population genetics
777 of fungal diseases of plants. *Parasite* **15:** 449-454.

778 Goudet J (2002). FSTAT v 2.9.3.2, A Program to Estimate and Test Gene Diversities
779 and Fixation Indices. *http://www2unilch/popgen/softwares/fstathtm*.

780 Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O *et al.* (2011).
781 Current trends in microsatellite genotyping. *Mol Ecol Resour* **11:** 591-611.

782 Heinzelmann R, Rigling D, Prospero S (2012). Population genetics of the wood-
783 rotting basidiomycete *Armillaria cepistipes* in a fragmented forest landscape.
784 *Fungal Biol* **116:** 985-994.

785 Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation
786 program for dealing with label switching and multimodality in analysis of
787 population structure. *Bioinformatics* **23:** 1801-1806.

788 Jombart T, Devillard S, Balloux F (2010). Discriminant analysis of principal
789 components: a new method for the analysis of genetically structured populations.
790 *BMC Genet* **11:** 94.

791 Jost L (2008). GST and its relatives do not measure differentiation. *Molecular*
792 *ecology* **17:** 4015-4026.

793 Kaiser SA, Taylor SA, Chen N, Sillett TS, Bondra ER, Webster MS (2016). A
794 comparative assessment of SNP and microsatellite markers for assigning
795 parentage in a socially monogamous bird. *Mol Ecol Resour***:** 10.1111/1755-
796 0998.12589.

797 Kamvar ZN, Tabima JF, Grunwald NJ (2014). Poppr: an R package for genetic
798 analysis of populations with clonal, partially clonal, and/or sexual reproduction.
799 *PeerJ* **2:** e281.

800 Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S *et al.* (2007).
801 Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature genetics*
802 **39:** 1151-1155.

803 Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002). Microsatellites: genomic
804 distribution, putative functions and mutational mechanisms: a review. *Molecular*
805 *Ecology* **11:** 2453-2465.

806 Ljungqvist M, Åkesson M, Hansson B (2010). Do microsatellites reflect genome-
807 wide genetic diversity in natural populations? A comment on. *Molecular ecology*
808 **19:** 851-855.

809 Meirmans PG, Hedrick PW (2011). Assessing population structure: FST and related
810 measures. *Mol Ecol Resour* **11:** 5-18.

811 Münsterkötter M, Walter M, Güldener U, Sipos G. (2015). *28th Fungal Genetics*
812 *Conference, Vol. (Suppl): Abstract*. Fungal Genetics Reports: Pacific Grove, CA, pp
813 237.

814 Nieuwenhuis BPS, James TY (2016). The frequency of sex in fungi. *Philosophical*
815 *Transactions of the Royal Society B: Biological Sciences* **371**.

816 Okonechnikov K, Golosova O, Fursov M (2012). Unipro UGENE: a unified
817 bioinformatics toolkit. *Bioinformatics* **28:** 1166-1167.

818 Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure
819 using multilocus genotype data. *Genetics* **155:** 945-959.

820 Prospero S, Jung E, Tsykun T, Rigling D (2010). Eight microsatellite markers for
821 *Armillaria cepistipes* and their transferability to other *Armillaria* species. *Eur J*
822 *Plant Pathol* **127:** 165-170.

823 Prospero S, LUNG-ESCARMANT B, Dutech C (2008). Genetic structure of an
824 expanding *Armillaria* root rot fungus (*Armillaria ostoya*e) population in a managed
825 pine forest in southwestern France. *Molecular ecology* **17:** 3366-3378.

826 Rosenberg NA (2004). distruct: a program for the graphical display of population
827 structure. *Molecular Ecology Notes* **4:** 137-138.

828 Rousset F (2008). genepop'007: a complete re-implementation of the genepop
829 software for Windows and Linux. *Mol Ecol Resour* **8:** 103-106.

830 Travadon R, Smith ME, Fujiyoshi P, Douhan GW, Rizzo DM, Baumgartner K (2012).
831 Inferring dispersal patterns of the generalist root fungus *Armillaria mellea*. *New*
832 *Phytol* **193:** 959-969.

833 Tsykun T, Rigling D, Nikolaychuk V, Prospero S (2012). Diversity and ecology of
834 *Armillaria* species in virgin forests in the Ukrainian Carpathians. *Mycol Prog* **11:**
835 403-414.

836 Väli Ü, Einarsson A, Waits L, Ellegren H (2008). To what extent do microsatellite
837 markers reflect genome-wide genetic diversity in natural populations? *Molecular*
838 *Ecology* **17:** 3808-3817.

839 Wang J (2015). Does GST underestimate genetic differentiation from marker data?
840 *Molecular ecology* **24:** 3546-3558.

841 Wang J, Fernández-Pavía SP, Larsen MM, Garay-Serrano E, Gregorio-Cipriano R,
842 Rodríguez-Alvarado G *et al.* (2017). High levels of diversity and population
843 structure in the potato late blight pathogen at the Mexico centre of origin.
844 *Molecular Ecology* **26:** 1091-1107.

845 Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of
846 population structure. *Evolution***:** 1358-1370.

847 Zhao Z, Fu Y-X, Hewett-Emmett D, Boerwinkle E (2003). Investigating single
848 nucleotide polymorphism (SNP) density in the human genome and its implications
849 for molecular evolution. *Gene* **312:** 207-213.

850

851

852 **Table 1** Multilocus genotype summary inferred from 17 SSR and 24 SNP markers
853 in two populations of *Armillaria cepistipes*. Ind – number of genotyped individuals;
854 MLGs – number of multilocus genotypes; I_A and $rbar_D$ – indexes of association; $p.I_A$
855 and $p.rbar_D$ – corresponding *P* values from one-sided permutation tests for indexes
856 of association

17 SSR loci							
Populations:	**Subpopulations:**	Ind	MLGs	I_A	$p.I_A$	$rbar_D$	$p.bar_D$
Carpathian	Beech	160	130	0.121	0.022	0.008	0.020
	Mixed/conifer	121	103	0.221	0.001	0.014	0.001
Alpine	North	62	62	0.064	0.190	0.004	0.189
	South	64	64	-0.039	0.669	-0.002	0.669
Overall		407	359	0.039	0.122	0.003	0.119
24 SNP loci							
Populations:	**Subpopulations:**	Ind	MLGs	I_A	$p.I_A$	$rbar_D$	$p.bar_D$
Carpathian	Beech	160	87	-0.113	0.985	-0.005	0.985
	Mixed/conifer	121	65	-0.047	0.762	-0.002	0.762
Alpine	North	62	62	-0.006	0.520	0.000	0.520
	South	64	64	0.062	0.175	0.003	0.173
Overall		407	278	-0.027	0.799	-0.001	0.799

857

858 **Table 2** Population specific summary statistics inferred from 17 SSR and 24 SNP
859 loci in populations of *Armillaria cepistipes*. N – average number of alleles; Ar –
860 average allelic richness; E – average evenness of alleles; H_{exp} – average expected
861 heterozygosity; H_{obs} – average observed heterozygosity; ± – standard deviation

17 SSR loci						
Populations:	**Subpopulations:**	**N**	**Ar**	**E**	**H_{exp}**	**Hobs**
Carpathian	Beech	7 ± 2	6 ± 2	0.64 ± 0.20	0.53 ± 0.26	0.46 ± 0.24
	Mixed/conifer	6 ± 3	6 ± 2	0.64 ± 0.16	0.53 ± 0.26	0.45 ± 0.23
Alpine	North	6 ± 2	5 ± 2	0.67 ± 0.17	0.55 ± 0.24	0.47 ± 0.24
	South	6 ± 2	5 ± 2	0.66 ± 0.16	0.52 ± 0.24	0.46 ± 0.23
24 SNP loci						
Populations:	**Subpopulations:**	**N**	**Ar**	**E**	**H_{exp}**	**Hobs**
Carpathians	Beech	2	-	0.76 ± 0.21	0.30 ± 0.16	0.33 ± 0.18
	Mixed/conifer	2	-	0.76 ± 0.21	0.30 ± 0.17	0.31 ± 0.17
Alpine	North	2	-	0.77 ± 0.17	0.30 ± 0.14	0.34 ± 0.15
	South	2	-	0.73 ± 0.19	0.28 ± 0.15	0.29 ± 0.14

862

863 **Table 3** Comparison of pairwise F_{ST} inferred from SSR and SNP multilocus
864 genotypes from two Carpathian and two Alpine subpopulations of *Armillaria*
865 *cepistipes*. * *P value* < 0.05

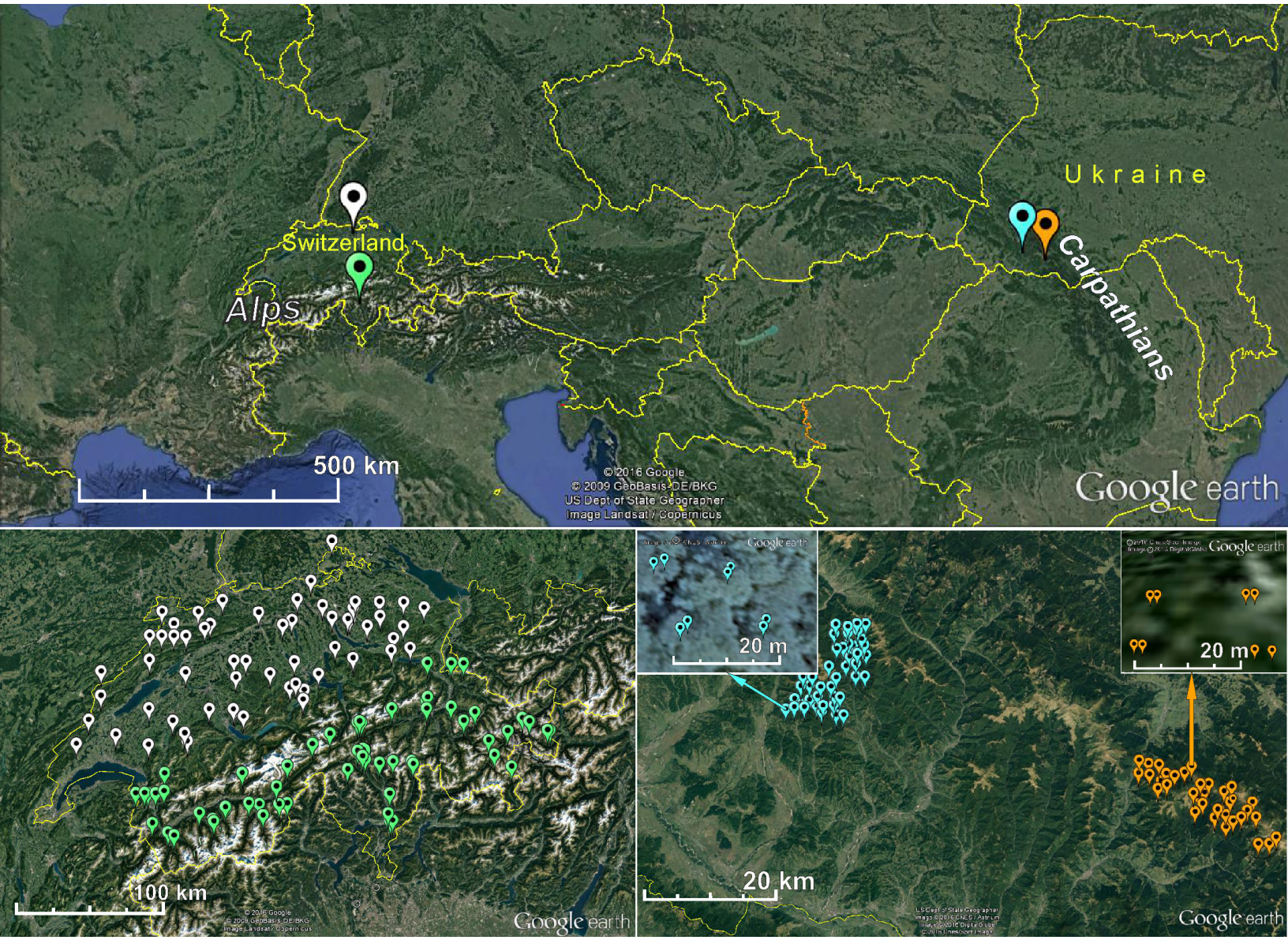
Loci:		17 SSRs			24 SNPs		
Populations:		Carpathian:		Alpine:	Carpathian:		Alpine:
		Beech	Mixed/ conifer	North	Beech	Mixed/ conifer	North
Carpathian:	Mixed/ conifer	*0.004	-	-	-0.001	-	-
Alpine:	North	*0.029	*0.026	-	*0.026	*0.024	-
	South	*0.028	*0.026	0.002	*0.036	*0.036	*0.019

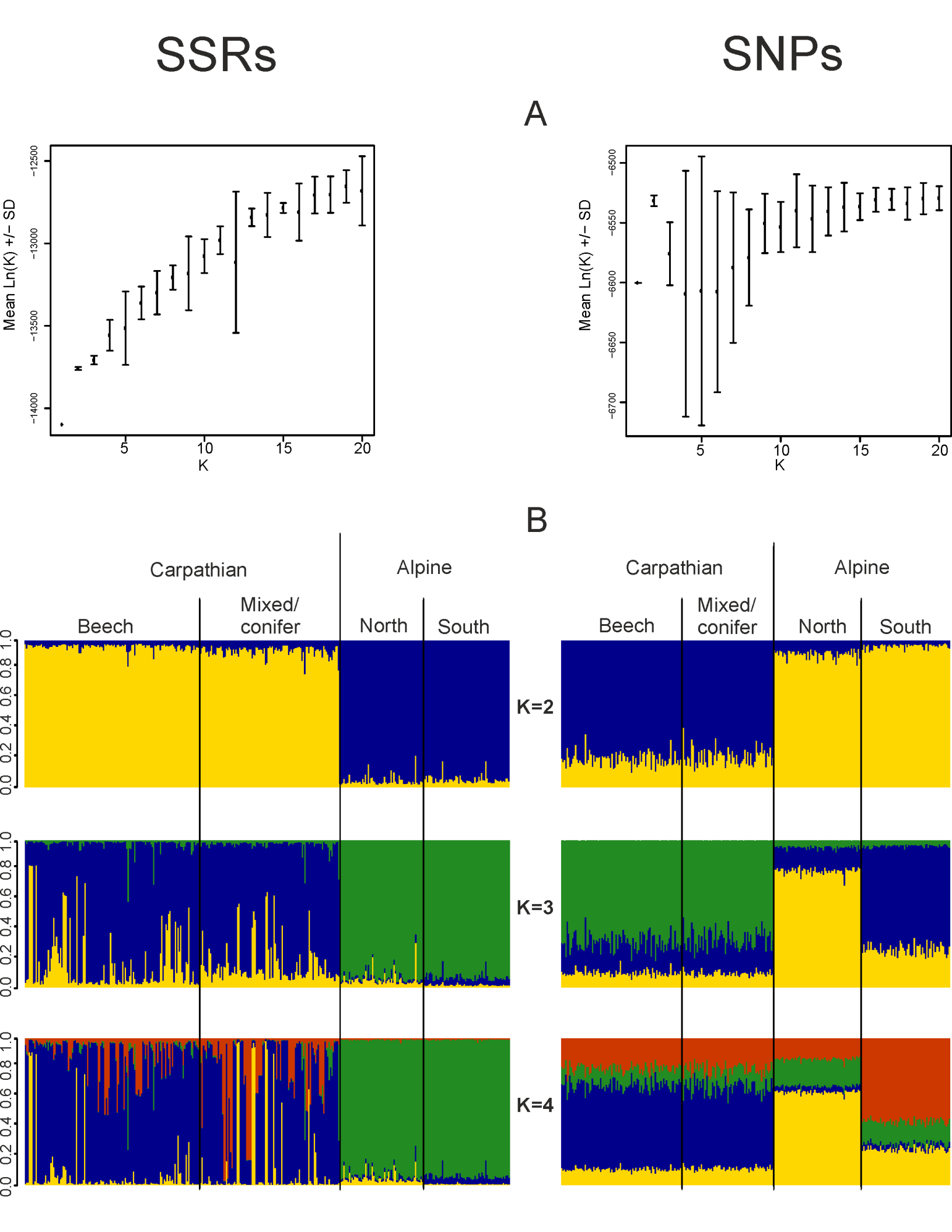
866

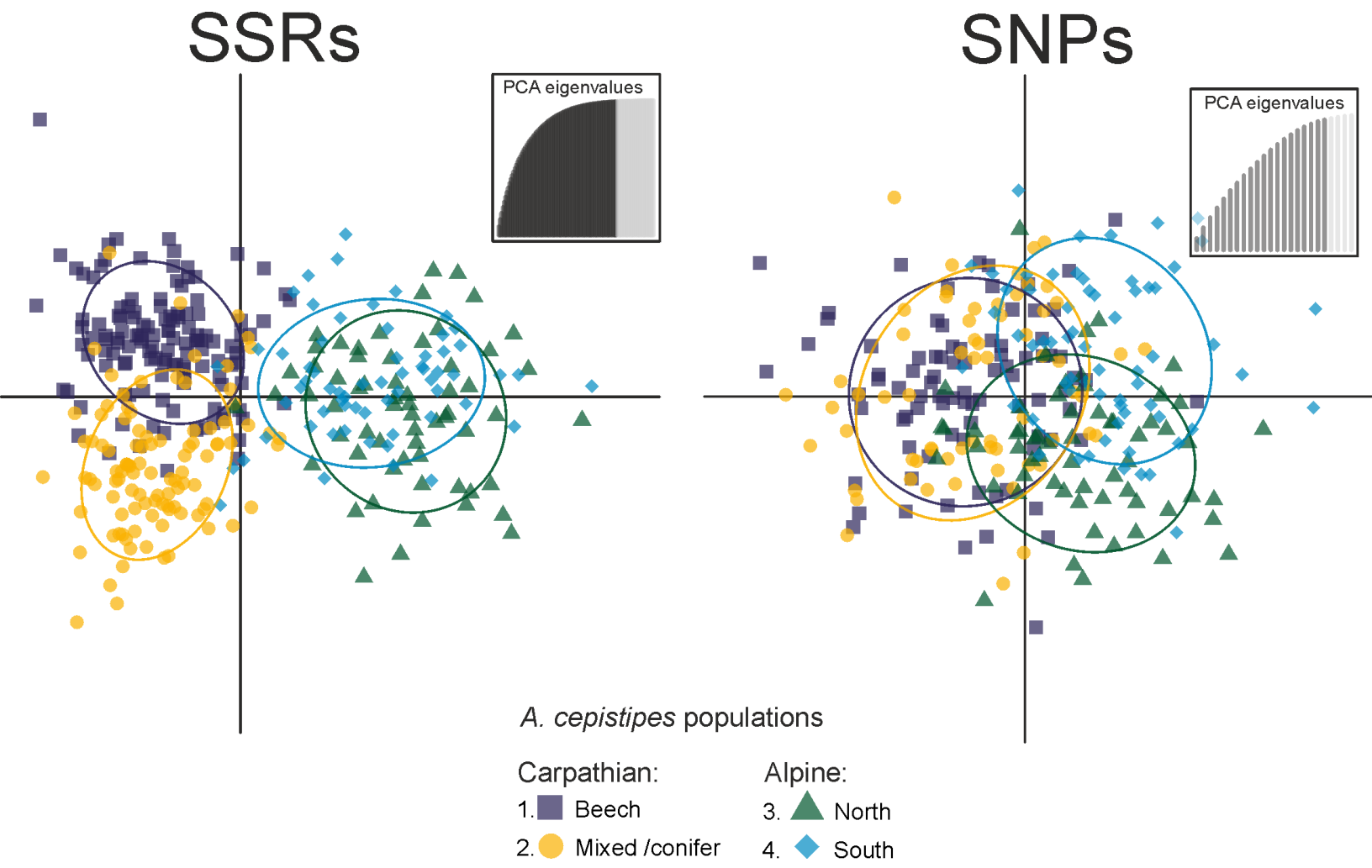
Figure 1 Geographic origin of the *Armillaria cepistipes* populations sampled in the Alps (Switzerland) and in the Carpathian mountains (Ukraine). In the lower left part of the figure, the Alpine population is shown, with the white drops indicating isolates of the North subpopulation and the green drops isolates of the South subpopulation. In the lower right part, the Carpathian population is illustrated, with the blue drops representing isolates sampled within plots (a zoomed example is shown) of the Beech subpopulation and the orange drops isolates within plots (a zoomed example is shown) of the Mixed/conifer subpopulation.

Figure 2 STRUCTURE results using 17 SSR loci and 24 SNP loci in four geographic subpopulations of *Armillaria cepistipes*. (A) Scatterplots with mean log-likelihood values (± standard deviation) for different numbers of clusters (K). (B) Barplots representing the average estimated membership probability (y-axis) of an individual to belong to a specific cluster (indicated by specific color).

Figure 3 Discriminant analysis of principle components (DAPC) computed with 17 SSRs and 24 SNPs in four subpopulations of *Armillaria cepistipes*. Scatterplots represent the distribution of individuals (symbols) along the axes build with the first two discriminant functions. The cross-validated principal components used in the analysis is shown in dark color in the bar plots on the top right of each scatterplot.


Switzerland
Alps
Ukraine
Carpathians
500 km
© 2016 Google
© 2009 GeoBasis-DE/BKG
US Dept of State Geographer
Image Landsat / Copernicus
Google earth
100 km
20 m
20 m
20 km
Google earth
Google earth
Google earth


SSRs
SNPs
A
Mean Ln(K) +/− SD
−12500
−13000
−13500
−14000
5
10
15
20
K
Mean Ln(K) +/− SD
−6500
−6550
−6600
−6650
−6700
5
10
15
20
K
B
Carpathian
Alpine
Beech
Mixed/ conifer
North
South
K=2
K=3
K=4
0.0 0.2 0.4 0.6 0.8 1.0


SSRs
PCA eigenvalues
SNPs
PCA eigenvalues
A. cepistipes populations
Carpathian:
1. Beech
2. Mixed /conifer
Alpine:
3. North
4. South