

# Single nucleotide polymorphisms in the ectomycorrhizal mushroom *Tricholoma matsutake*

Jianping Xu,<sup>1,2</sup> Hong Guo<sup>2</sup> and Zhu-Liang Yang<sup>1</sup>

## Correspondence

Jianping Xu

jpxu@mcmaster.ca

fungi@mail.kib.ac.cn

<sup>1</sup>Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, People's Republic of China

<sup>2</sup>Department of Biology, McMaster University, Hamilton, ON, Canada

Single nucleotide polymorphisms (SNPs) are becoming increasingly popular markers for studying a variety of biological phenomena. This paper describes the development and analysis of a set of SNP markers for the basidiomycete fungus *Tricholoma matsutake*. *T. matsutake* is a gourmet mycorrhizal mushroom primarily associated with pine forests. However, little is known about its genetics and genomic variation, including SNP variation. To identify and analyse SNPs in *T. matsutake*, a genomic library was constructed and >72 000 nt were analysed from >200 random clones. Primers from 20 sequenced fragments were then designed to amplify and sequence >10 000 bp sequences from the original strain, from which the genomic library was constructed, as well as another strain from >350 km away; both strains were from south-western China. These two strains had similar intra-strain SNP frequencies (1.104 and 1.278 % per nucleotide, respectively). The combined analysis revealed that 14 of the 20 examined fragments contained SNPs, ranging from two to 47 per fragment, and yielding a total of 178 SNPs out of the 10 428 sequenced nucleotides (an SNP frequency of 1.707 %). Among the 178 SNPs, one site had three alternative nucleotides, while the remaining 177 had two each, with 148 transitions and 29 transversions, resulting in a combined transition to transversion ratio of 5.1. In addition, the haplotype phases of all SNPs within individual fragments for both strains were determined. Phylogenetic analyses of these haplotypes revealed three kinds of haplotype relationship, including haplotype sharing both within and between strains. Furthermore, a subset of the SNPs detectable by restriction enzyme digests was screened for its distribution among 31 additional wild strains from five distinct locations in south-western China. The implications of these SNPs and haplotypes for our understanding of the genetics, population history, ecology and evolution of this important mushroom species are discussed.

Received 28 December 2006

Revised 8 March 2007

Accepted 28 March 2007

## INTRODUCTION

Molecular markers are essential tools to study a variety of biological properties and processes, such as recombination, population structure, and genetic relatedness among individuals. Among these markers, single nucleotide polymorphisms (SNPs) are among the fastest-developing categories in biomedical and biological research. This is mainly because SNPs are the most frequently observed differences between DNA sequences obtained from different individuals, or between alleles from within an individual in diploid or higher ploidy organisms. In addition, SNPs have several properties, such as a relatively low mutation rate and the ease of scoring and data sharing, that make them highly desirable for a variety of biological analyses (Brumfield *et al.*, 2003; Xu, 2006b). So far, most

such studies have focused on model organisms, and relatively little is known about the patterns and distributions of SNPs in non-model organisms, including most fungi. Our objective in this study was to develop and analyse a set of SNP markers for the economically and ecologically important fungus *Tricholoma matsutake*.

*T. matsutake* is an ectomycorrhizal basidiomycete that produces the economically important edible mushrooms commonly known as 'matsutake'. It has a monokaryon–dikaryon alternating life cycle typical of most basidiomycete mushrooms. Specifically, haploid monokaryotic basidiospores are released from mature fruiting bodies and germinate to form monokaryotic mycelia. When monokaryotic mycelia with different mating types meet, they can mate and form dikaryotic mycelia. Under appropriate conditions, the dikaryotic mycelia can form mature mushrooms and release basidiospores to complete the life cycle (Tominaga, 1978).

Abbreviations: IGS, intergenic spacer; ITS, internal transcribed spacer; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

*T. matsutake* is predominantly associated with pine forests in the northern hemisphere but has been found to be associated with oaks in south-western China (Nakayama & Nakanishi, 2004). It is widely distributed in Europe and Asia. Closely related species that produce fruiting bodies similar to *T. matsutake* exist in North Africa, North and Central America, as well as Europe and Asia (Wang *et al.*, 1997; Redhead, 1997). The commercial importance of *T. matsutake* has been mostly driven by the Japanese market, in which 1 kg of top-quality fresh matsutake may be sold for as much as US \$2000. However, in the past century in Japan, because of deforestation and infestation by the pinewood nematode (*Bursaphelenchus xylophilus*), populations of the matsutake mushroom host plant *Pinus densiflora* have declined rapidly (Gill *et al.*, 2000; Wang *et al.*, 1997). As a result, the annual harvest of matsutake in Japan has been declining steadily, and is much lower than in the early twentieth century. Artificial cultivation has not been developed for this mushroom, therefore, to satisfy its domestic demand, Japan imports about 3000 tons of these mushrooms annually, mostly from Pacific North America, Korea and China (Wang *et al.*, 1997). The commercial demands from Japan have subsequently placed serious pressures on natural populations of this species elsewhere, such as in south-western China, from where up to a third of the Japanese imports come. Consequently, understanding and conserving the natural populations of this mushroom have attracted significant attention in recent years from mushroom collectors, government agencies, non-governmental agencies, conservation groups, and environmental biologists.

Several molecular markers have been used/developed to help understand natural populations of *T. matsutake*. These include simple sequence repeats (SSRs) or microsatellites (Lian *et al.*, 2003), retrotransposon-like repetitive elements (Murata & Yamada, 2000; Murata *et al.*, 2001), and DNA sequences at the internal transcribed spacer (ITS) and the intergenic spacer (IGS) regions of the rRNA gene clusters. The ITS and, to a lesser extent, the IGS have been used to analyse potential species boundaries in the matsutake species complex, which includes *Tricholoma magnivelare*, *Tricholoma caligatum*, and *T. matsutake*, as well as the potential geographical relationships among strains and populations within *T. matsutake* (e.g. Chapela & Garbelotto, 2004). However, due to the relatively conservative nature of the ITS regions, few variable nucleotide sites have been found among strains and populations within *T. matsutake*. To overcome the limited variation observed in the ITS regions, six microsatellite markers have recently been developed using a method called 'inter simple sequence repeat (ISSR)-suppression PCR'. Among 10 clones screened in an ISSR-suppression PCR library, four were not variable, and six had alleles ranging from two to nine in number, in a sample of 41 fruiting bodies from various areas in Japan (Lian *et al.*, 2003). The use of five of these six microsatellite markers has allowed the identification of genetic individuals and the analysis of relationships among

isolates obtained from above and below ground in a *P. densiflora* forest in Japan (Lian *et al.*, 2006). A third type of marker for *T. matsutake*, two highly repetitive elements named marY1 and marY2N, have been discovered and used to type strains (Murata & Yamada 2000; Murata *et al.*, 2001). These two markers have allowed effective discrimination among genotypes, even for strains isolated in very close proximity (e.g. in the same 'shiro') (Murata *et al.*, 2005).

In this study, we described the discovery of 178 SNP sites in two individuals of *T. matsutake*. These SNPs were identified through analysis of a shotgun genomic library, and comparison between sequences from within and between two strains isolated from ~350 km apart. Furthermore, these SNPs were sorted into haplotypes for each of the analysed fragments. A selected subset of the SNPs detectable by restriction enzyme digests was screened for its distribution among 31 wild mushrooms from five distinct locations in south-western China. We discuss the implications of these SNPs and haplotypes in the inference of the population genetic and ecological patterns of this species.

## METHODS

**Strains.** A total of 33 fruiting bodies were analysed in this study. These strains were collected from seven distinct geographical locations in south-western China: five locations in Yunnan province and one each in Tibet and Sichuan. Among these 33 strains, we primarily focused on two: HKAS48691 and HKAS48901. These two fruiting bodies were collected in Yunnan Province. Strain HKAS48691 was collected in a mixed forest dominated by oaks on the snow-capped mountain of Yulong (~3500 m above sea level), situated 20 km north-west of the City of Lijiang, Yunnan. Strain HKAS48901 was collected in a mixed pine forest in Nanhua County in Chuxiong Prefecture, ~350 km south-east of Lijiang. Both strains were collected in August 2005, dried and deposited at the Herbarium of Cryptogams, Kunming Institute of Botany, Chinese Academy of Sciences. The other 31 strains were collected in the summer and fall of 2006. They were identified as belonging to *T. matsutake*, based on morphological characteristics of the fruiting bodies and sequences of the ITS regions.

**DNA isolation.** For each of the two strains, ~0.01 g tissue from the cap of the dried mushroom was ground into a fine powder using a blue micropipette tip in a 1.5 ml microcentrifuge tube. The remaining steps essentially followed those in the protocol described by Xu *et al.* (2000a) for extracting DNA from live yeasts, with slight modifications. Briefly, the ground mushroom tissue was suspended in 0.5 ml cell-wall-digesting buffer [1 M sorbitol, 1%  $\beta$ -mercaptoethanol, 0.25 mg lysing enzyme ml<sup>-1</sup> (Sigma), 0.25 mg lyticase ml<sup>-1</sup> (Sigma)], and incubated at 37 °C for 2 h. The cell-wall-less materials were collected by centrifugation at 5000 g for 5 min, and incubated at 65 °C in 0.5 ml lysing buffer (50 mM EDTA, 1% SDS, 0.1  $\mu$ g RNase A ml<sup>-1</sup>) for 30 min to dissolve cytoplasmic membranes and lipid materials. One-quarter volume of 7.5 M ammonium acetate (~125  $\mu$ l) and 0.5 ml chloroform/isoamyl alcohol (24:1) solution were then added. The mixture was vortexed and centrifuged at 13 000 g for 10 min. The supernatant was transferred to a new 1.5 ml microcentrifuge tube and 0.5 ml ice-cold 100% isopropanol was added. The tube contents were mixed gently by inversion and centrifuged at 13 000 g for 2 min to precipitate the DNA. The

supernatant fluid was discarded and the DNA was washed with 70 % ethanol, air-dried, and suspended in 100 µl Tris/EDTA (TE) buffer. The DNA was then cleaned using the commercial GeneClean III kit Bio 101 (Qbiogene), following the supplier's instructions. The final DNA was suspended in 50 µl TE buffer and stored at -20 °C. By this method, 0.01 g dried mushroom tissue yielded ~5 µg genomic DNA.

**Genomic library construction.** To construct a random shotgun genomic library, the genomic DNA extracted from strain HKAS48691 was first sheared using sonication to yield the majority of DNA fragments in the range 0.5–1.5 kb. The sheared DNA fragments were then ligated into the pPCR-Script Amp SK(+) cloning vector and transformed into XL10-Gold Kan ultracompetent *Escherichia coli* cells, following the manufacturer's instructions. Both the vector and the competent cells were included in the PCR-Script Amp Cloning kit purchased from Stratagene. Aliquots of the transformed cells were spread-plated onto Luria–Bertani (LB)/ampicillin agar medium containing X-Gal and IPTG. Plates were incubated at 37 °C for 17–20 h to distinguish transformed cells with and without inserts.

**Sequencing cloned fragments.** Primers T3 and T7 matching the flanking sequences of the cloning site region on the pPCR-Script Amp SK(+) vector were used for PCR to amplify cloned DNA inserts. Each PCR reaction contained ~10<sup>7</sup>–10<sup>5</sup> cells, 0.5 U Taq DNA polymerase, 1 µM each primer, and 200 µM of each of the four dNTPs in a total volume of 30 µl. The following PCR conditions were used for all amplifications: 4 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 55 °C, 45 s at 72 °C, and finally, 7 min at 72 °C. After confirmation of the PCR products by agarose gel electrophoresis, the fragments were cleaned using the DiaMed PCR cleanup kit according to the manufacturer's instructions. The purified PCR products were then sequenced using an Applied BioSystems Prism 3100 automated sequencer with dRhodamine-labelled terminators (PE Applied BioSystems), following the manufacturer's instructions.

**Sequence analysis for SNP identification.** To analyse the sequences obtained above, flanking vector sequences were first trimmed and the remaining nucleotides were then compared, using the BLASTN algorithm, with sequences in GenBank to determine their potential matches with published sequences, and the possibility of contamination due to either impurity of the original material or laboratory artefacts. A subset of the sequences was then selected and used for the design of PCR primers for direct amplification and sequencing of DNA from strains HKAS48691 and HKAS48901. PCR primers were designed using in-house software at the MoBix Laboratory at McMaster University. The protocols for PCR, PCR product confirmation and cleaning, and DNA sequencing followed those described above. Sequencing was done in both directions using both the forward and reverse primers, and the obtained sequences from the two strains were used to identify SNPs.

**Haplotype inference.** Fruiting bodies of *T. matsutake* develop from dikaryotic mycelia, and such mycelia are the products of mating between genetically different and compatible monokaryons; therefore, each cell in the fruiting body is a functional diploid and has two alleles at each locus. If the two alleles were the same, there should be no double peak in the sequencing chromatograph, and a clear sequence with no ambiguous site should be obtained. On the other hand, if the two alleles at an individual locus within a fruiting body have different sequences, direct sequencing of the PCR product from such a fruiting body would result in a composite chromatograph, with certain sites containing two nucleotides. In strain HKAS48691, in which the original genomic library was constructed, one sequence (also called a haplotype) is known from the shotgun genomic library; therefore, the other haplotype can be directly inferred by comparing the composite diploid sequence with that of the genomic library (Xu, 2006a).

Unlike haplotype inferences for these loci in strain HKAS48691, in which one allele for each locus was already available in the shotgun library, those in strain HKAS48901 were not available. To help identify the haplotypes, loci in strain HKAS48901 that showed heterozygosity (double peaks on the chromatograph) were subjected to further experimentation. Specifically, PCR products from such loci were cloned following the protocol described above, and a random colony was picked for each locus for further PCR and sequencing, to obtain one haplotype for this locus in strain HKAS48901. The other haplotype was then directly inferred by comparing the known haplotype sequence with the composite diploid sequence (Xu, 2006a). To confirm the inferred haplotypes in both strains, we further sequenced four to six additional random clones for four of the 20 DNA fragments.

**Relationships among allelic sequences.** After haplotype inferences, the loci with all four sequences available from the two strains were analysed for their phylogenetic relationships. The phylogenetic analyses were performed using the software PAUP4.0b10 (Swofford, 2004). The maximum parsimonious trees were obtained using heuristic searches and the tree-bisection reconnection (TBR) branch, swapping with 100 starting trees obtained by a random sequential addition of sequences. Maximum-parsimony analyses were conducted for each gene fragment that contained all four sequences. Mid-point rooting was used for all phylogenetic trees.

**Population survey of selected SNPs.** To examine the potential usefulness of the SNPs for genotype determination of natural *T. matsutake* strains, a subset of the SNPs detectable by restriction enzyme digests was screened for 33 fruiting bodies of *T. matsutake* collected from various locations in south-western China. To achieve this, we first checked the restriction recognition maps of all 14 fragments containing SNPs, using the online web server of New England Biolabs. Differences in restriction maps among haplotypes for each locus were used as the basis for enzyme selection. Here, for ease of scoring on agarose gels, we selected enzymes that had only one or no recognition site among the alternative haplotypes. The final chosen combinations of gene fragment–restriction enzyme were based on enzyme availability. These enzymes were then used to digest PCR products from the 33 wild strains. PCR, restriction digestion, agarose gel electrophoresis and scoring followed those in Xu *et al.* (1999).

## RESULTS

### BLAST searches

Among the colonies containing inserts, we amplified and sequenced a total of 221 random cloned fragments. These 221 DNA fragments ranged from 43 to >1000 bp, with a total of 73 065 bp of sequence obtained. Based on results from BLAST searches, these fragments were grouped into the following four categories.

The first category contained six fragments with a total of 2451 bp of sequence. These six fragments had significant matches to non-fungal DNA in GenBank. Five fragments had best matches with those from three bacterial species: the common non-pathogenic saprophyte *Pseudomonas fluorescens* (three fragments), the plant pathogen *Pseudomonas syringae* pv. *syringae* (one fragment), and the psychrotolerant anaerobe *Rhodospirillum rubrum* (one fragment). The sixth fragment in this category had a significant match to human DNA. These sequences likely

**Table 1.** Primers used to screen for SNPs and haplotypes within and between strains of *T. matsutake*

Fragment name	Primer sequence (5'→3')	Temperature (°C)	Expected size (bp)
TmRC1	Forward: CCCACAATAAGATTTTCATACC	48	425
	Reverse: TCTTGTTGTTTCAGGCGGTA	53	
TmRC2	Forward: TCTTGAAAATCTTTGAAAATCC	51	737
	Reverse: TGAGGACAAATGCCTTTGCA	56	
TmRC3	Forward: CACACAAATATACAAAATGACAAACT	51	530
	Reverse: AAGCTGAGGCTGAGCATCTTAA	55	
TmRC4	Forward: ATTACAAGGACAGTCAGCTCCAT	57	455
	Reverse: GCTTACAGAGTCATTTGAGTCACAC	53	
TmRC5	Forward: GACAACATCAAAACTATTCTTTGC	51	494
	Reverse: CACACAGAATCACCGCTAAGT	51	
TmRC6	Forward: CTATATACGCTGCACAGGGGA	54	390
	Reverse: AGAGGTTGAAAACACCTCGAA	53	
TmRC7	Forward: CGTGCCATGTCACATGTGTTT	56	380
	Reverse: TCTTGTTTTCTCCTCCCTCCCT	55	
TmRC8	Forward: TTTGCCCTTCAAACCTGGAA	56	432
	Reverse: AGCGTCGAGGCATCAAACT	56	
TmRC9	Forward: TCCCAATCCCAGAGCATCTA	54	412
	Reverse: TTTCAGACCCAAAAAGGTCC	53	
TmRC10	Forward: GGCATCGCCCGGCCTATC	62	380
	Reverse: CGACCGACAGCCCCGCCTC	66	
TmRC11	Forward: AAATCCCATGTTGTGGCTCCT	57	426
	Reverse: ACGATTGGCGATGTCAAAC	54	
TmRC12	Forward: CCCTCTACCTTTGCAGAAATGAC	56	928
	Reverse: GGGTAGGTTTACATATTTCTGCT	56	
TmRC13	Forward: AACTCTTCAATAGCGCTCAGG	53	448
	Reverse: TAACCGAGCTGCATGTGTTT	53	
TmRC14	Forward: CCGATGAAAGCTTCCACTTTG	56	727
	Reverse: GCAGGACCCTCTATCTGGATT	53	
TmRC15	Forward: TTCCATGTCGTCAAACCTCTTC	53	519
	Reverse: GACTCAACATAAATCTCCAATCCA	54	
TmRC16	Forward: TGCATTGGGCACTATTCCTA	53	406
	Reverse: TCCCCTCATTACATTACACA	53	
TmRC17	Forward: AACGTCAACAGCAATGGCAT	55	777
	Reverse: TCGTCAGAGGAAGGTCAAACA	54	
TmRC18	Forward: CGACAGTCAATGACCCTGAAT	53	694
	Reverse: CCAAGCCCTCGATCATGACTA	56	
TmRC19	Forward: CATCTACTAATGAAATAACAATGC	47	469
	Reverse: CAAGTAACAGAACTAGTTGAAGAC	46	
TmRC20	Forward: TTTGATTTGGGGTTTGAGCC	56	386
	Reverse: TTACTGTGACTAGAACATTTTCGG	52	

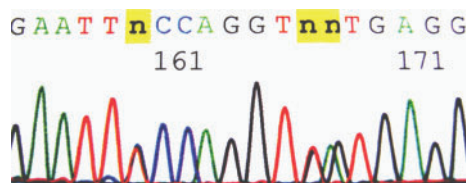
came from contamination during collecting and/or sample processing, and were not included in further analysis.

The second category contained one 573 bp fragment that had a significant match to the mitochondrial DNA sequence of another basidiomycete, *Crinipellis perniciosa*, which belongs to the same fungal group as *T. matsutake*, the family *Tricholomataceae*.

The third category contained 26 fragments with a total of 10 230 bp of sequence. Fragments in this category all had significant matches to published sequences from *T. matsutake*. This category contained three groups. The first group had 14 fragments with a total of 4706 bp. Each of the

14 had a significant match to part of the repetitive element marY1. Their sequence identities to the published marY1 sequence ranged from 88 to 99 %. A complete copy of marY1 has been cloned and sequenced by Murata & Yamada (2000). Their analysed marY1 is 6.0 kb long, and has been identified as belonging to the gypsy-type retrotransposons. The second group in this category contained 11 fragments with a total of 5358 bp. Each of the 11 had a significant match to part of another repetitive element, marY2N. Their sequence identities to the published marY2N sequence ranged from 89 to 97 %. MarY2N is a member of the long intersperse nuclear elements (LINES) (Murata *et al.*, 2001). The remaining fragment in this





**Fig. 1.** A representative chromatogram showing heterozygosity at nucleotide positions 159, 166 and 167 of fragment TmRC4 for strain HKAS48691. These three positions on the chromatogram correspond to nucleotides 185, 192 and 193, respectively, of TmRC4 in Fig. 2.

group was 166 bp long. This fragment had 100 % sequence identity to parts of ITS1 and the 5.8S rRNA gene from *T. matsutake*.

The fourth group contained the remaining 188 fragments with a total of 59 551 nt. Fragments in this group had no obvious sequence identity to any DNA in GenBank. These fragments likely represented novel sequences in *T. matsutake*.

### Identification of SNPs

Based on the obtained sequences from the genomic library, we designed PCR primers based on the sequences of 20 fragments from the fourth category above, but with sequences greater than 400 bp. This length of cut-off was chosen to maximize the sequence length obtained. The PCR primers for all 20 fragments are presented in Table 1. They were expected to amplify >10 000 bp sequences from each of the two strains HKAS48691 and HKAS48901. As expected, all 20 primer pairs successfully amplified fragments of expected sizes using the genomic DNA from strain HKAS48691. However, in our initial screening, only 14 of the 20 primer pairs resulted in PCR products for the second strain, HKAS48901. The obtained composite diplotype sequences were compared to each other and to those in the genomic library of strain HKAS48691, for inferences of SNPs. A representative chromatograph containing heterozygous sites is shown in Fig. 1. Below, we briefly summarize the results of the SNP analyses.

For strain HKAS48691, we obtained a total of 10 415 nt among the 20 analysed fragments. Based on our analysis, eight of the 20 fragments were homozygous with no evidence of heterozygosity within this strain, while 12 fragments had SNPs ranging from 1 to 33, with a mean of ~10 SNPs per fragment (Fig. 2). Overall, within strain HKAS48691, there were a total of 115 SNPs for the analysed 10 415 bp, yielding an average SNP frequency of 1.104 % (115/10 415) per nucleotide.

For strain HKAS48901, we were able to obtain sequences from only 14 of the 20 analysed fragments, using the designed primers. The six primer pairs that failed to amplify

from strain HKAS48901 were TmRC2, TmRC9, TmRC10, TmRC13, TmRC17 and TmRC20. Our analyses identified that six of the 14 amplified fragments were homozygous with no evidence of heterozygosity within this strain, while the eight remaining fragments (TmRC1, TmRC3, TmRC4, TmRC12, TmRC14, TmRC16, TmRC18 and TmRC19) had SNPs ranging from one to 27, with a mean of ~11 SNPs per fragment (Fig. 2). Overall, within strain HKAS48901, there were 93 SNPs for the analysed 7275 bp sequences, yielding an average SNP frequency of 1.278 % (93/7275) per nucleotide site, a rate similar to that within strain HKAS48691.

When sequences from both strains were compared, 14 of the 20 fragments contained SNPs. Of the six fragments without any SNP, four (TmRC5, TmRC6, TmRC7 and TmRC11) were obtained from both strains, but were shown to be identical and homozygous within and between the two strains. Among the 14 fragments containing SNPs, the comparisons of the two strains yielded an additional 44 SNPs not seen within HKAS48691 or HKAS48901. In total, we identified 178 SNPs, representing an SNP frequency of 1.707 % per nucleotide site. In addition to the SNPs, the comparisons between these two strains identified a 13 bp indel in fragment TmRC12. This 13 bp sequence was absent in both alleles in strain HKAS48691, but present in both alleles in strain HKAS48901. All the identified SNPs and the indel are presented in Fig. 2.

Among the 178 SNPs identified here, one site had three alternative nucleotides, and 177 had two each with 148 transitions and 29 transversions. The overall ratio of transition to transversion is 5.1. This combined ratio is very similar to that obtained from within the two strains (data not shown).

### Haplotype inferences and relationships

Among the 12 DNA fragments that showed heterozygosity within strain HKAS48691, we were able to infer unambiguously both haplotypes for each strain. Similarly, for all eight loci that showed heterozygosity in strain HKAS48901, we were able to obtain their respective haplotypes through cloning PCR products, sequencing representative clones, and comparing the clone sequences with the composite diplotype sequences. Sequencing three to four additional clones confirmed all inferred haplotypes and SNPs for both strains.

In total, we obtained all four allelic sequences from the two strains for the following 14 DNA fragments: TmRC1, TmRC3, TmRC4, TmRC5, TmRC6, TmRC7, TmRC8, TmRC11, TmRC12, TmRC14, TmRC15, TmRC16, TmRC18 and TmRC19. Among these 14 fragments, four were not variable among the four alleles from the two strains (TmRC5, TmRC6, TmRC7 and TmRC11), and these four were not subjected to further phylogenetic analysis. Of the remaining 10 fragments, the relationships among the four that were allelic for each fragment were all

Fragment	TmRC1	TmRC2	TmRC3	TmRC4
Position	12	35	11233	011111122
	41	22	89168	912789977
	33	22	18990	891352314
HKAS48691-1 <sup>1</sup>	GG	GG	TTTGG	TCTCTGTT
HKAS48691-2	AG	TA	CCCGG	CTTTCACC
HKAS48901-1	GA	?? <sup>2</sup>	TTTGA	TCTCTGTT
HKAS48901-2	GG	??	CCCAA	CTTTCACC

Fragment	TmRC8	TmRC9	TmRC10
Position	122222333411111111222333333		12
	03556968901345677990080123779		70
	38065205706590128470689417238		39
HKAS48691-1	GAATACTGCGGATGACTCTTAGAACTGT		TT
HKAS48691-2	GAATACTGCGAGCAGGCTCCGAGGGCAC		CC
HKAS48901-1	ACGCGGGTTA????????????????		??
HKAS48901-2	ACGCGGGTTA????????????????		??

Fragment	TmRC12
Position	0012222233333444444455555556666677788888899
	79701347803799333344601467788223688166567789901
	01740067068667013579275642389396723723670654853
HKAS48691-1	AAGAATTCTATTACATTAGTCTCCCTCATTCAGTCATTTCTGTTAT
HKAS48691-2	AAGAATTCTATTACATTAGTCTCCCTCATTCAGTCATTTCTGTTAT
HKAS48901-1	GGACGCCACCCCTGTGCCGACTCTTTGTGCCTGTACTGCCCAACCGC
HKAS48901-2	GGACGCCACCCCTGCCTGCGACTCCTCGCGTTCGCGCCGCCAATCGC

Fragment	TmRC14	TmRC15
Position	001111222222333333344444445556667	22
	7811581133348355788923666784590020	14
	9314555615799689425630234679035685	53
HKAS48691-1	GTTAACGCGCCAGCAACGTTGCGATGCGGTGAC	TA
HKAS48691-2	ACCGGTCAAATTGATGCTTCCATGGGATGACAGT	GG
HKAS48901-1	GCCAACCGAGCCAGCATCGTTGACGGCGGCGAC	GG
HKAS48901-2	ACCGGTCAAATTGATGCTTCCATGGGATAACAGT	GG

Fragment	TmRC16	TmRC17
Position	000011111122333334	145
	889902457944045690	230
	070118998489213465	319
HKAS48691-1	ATCAGCTTATGTAGTTT	CTC
HKAS48691-2	GATGTATCATCAGACCC	TCT
HKAS48901-1	ATCGTACTTATGTAGCTT	???
HKAS48901-2	GACGTATCATCAGACTC	???

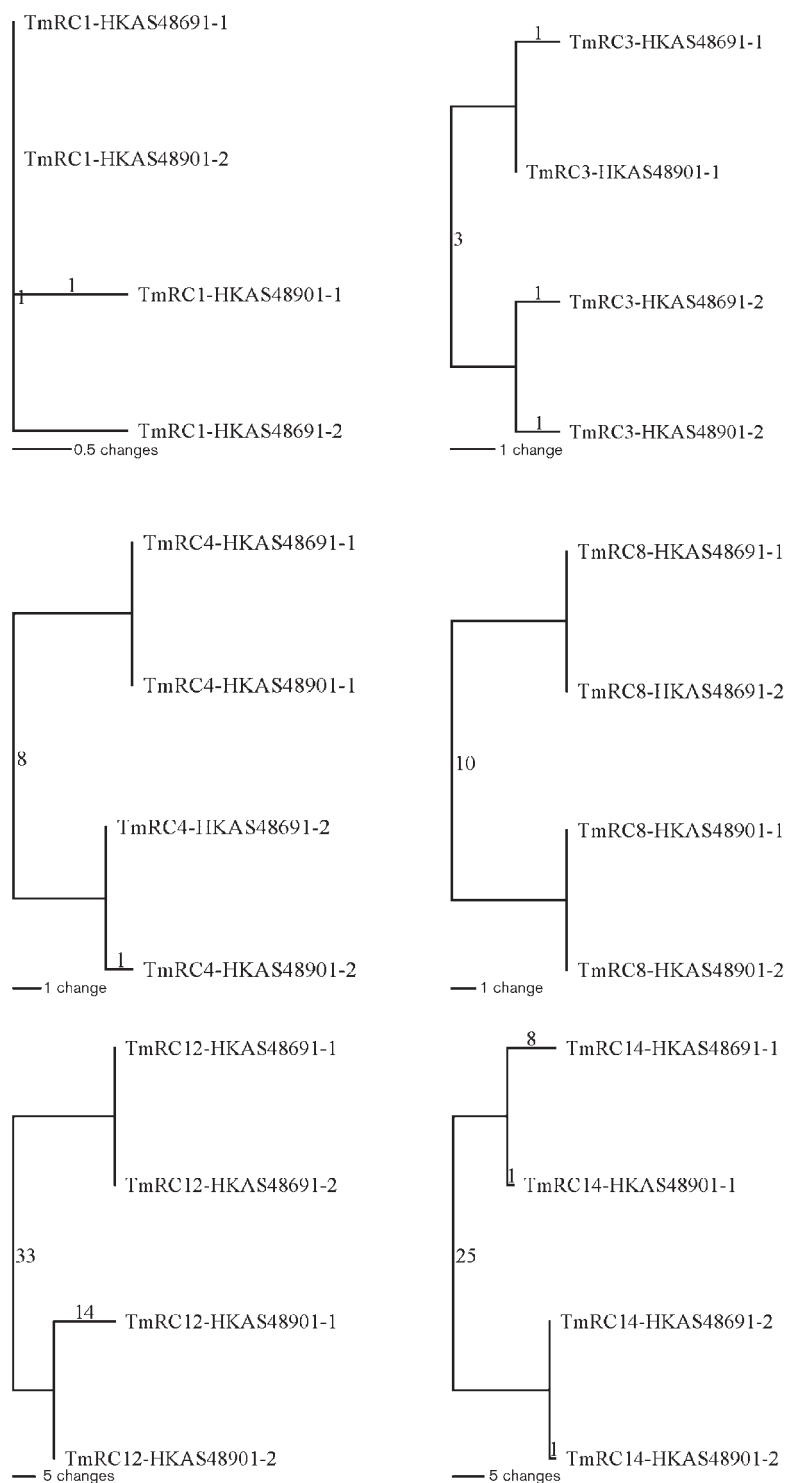
Fragment	TmRC18	TmRC19
Position	1111	1111111122222222344
	0012	322356677002233335511
	5909	478347847470113473612
HKAS48691-1	CTTT	TGGGTTGTTGTCCACAGGGAA
HKAS48691-2	TCTC	GTTACCACCTCTGCTGACATG
HKAS48901-1	CTTT	TGGGTTGTTGTCCACAGGAAA
HKAS48901-2	TCCC	GTTACCACCGCTGCTAGGAAA

**Fig. 2.** SNPs and haplotypes identified in this study. For each panel, the top row indicates the random clone fragment. The second row contains three digits, indicating the nucleotide position of each SNP in the sequenced fragment, reading from top to bottom. In all cases, the first base of the forward primer for amplifying each fragment is referred to as position 1. The third and bottom row includes the SNP information for each allelic sequence. Each allelic sequence identifier contains the strain name (HKAS48691 or HKAS48901) followed by an arbitrary number (1 or 2) to represent the allelic sequences within each of the two strains for each fragment. Please note that the allelic sequence identifiers among the DNA fragments are independent of each other within each of the two strains. ?, missing data.

analysed. The phylogenetic analyses identified that for each of the ten fragments, there was only one most parsimonious tree, and the consistency indices were all 1, with no evidence of homoplasy. The respective 10 maximum-parsimonious trees are presented in Fig. 3, and the relationships among the allelic sequences are briefly summarized below.

Among these 10 DNA fragments, two showed two distinct haplotypes each (TmRC8 and TmRC15), four showed three distinct haplotypes each (TmRC1, TmRC4, TmRC12

and TmRC18), and the remaining four had four haplotypes each (TmRC3, TmRC14, TmRC16 and TmRC19) (Fig. 3). Allelic relationships for the 10 fragments can be classified into three types. In the first type, the two alleles from within a strain were more similar to each other than either was to alleles from the other strain. Fragments TmRC8 and TmRC12 showed this type of relationship (Fig. 3). In the second type, one allele from each of the two strains was more similar to that from the other strain than either was to the second allele from the same strain. Seven fragments (TmRC1, TmRC3, TmRC4, TmRC14, TmRC16, TmRC18

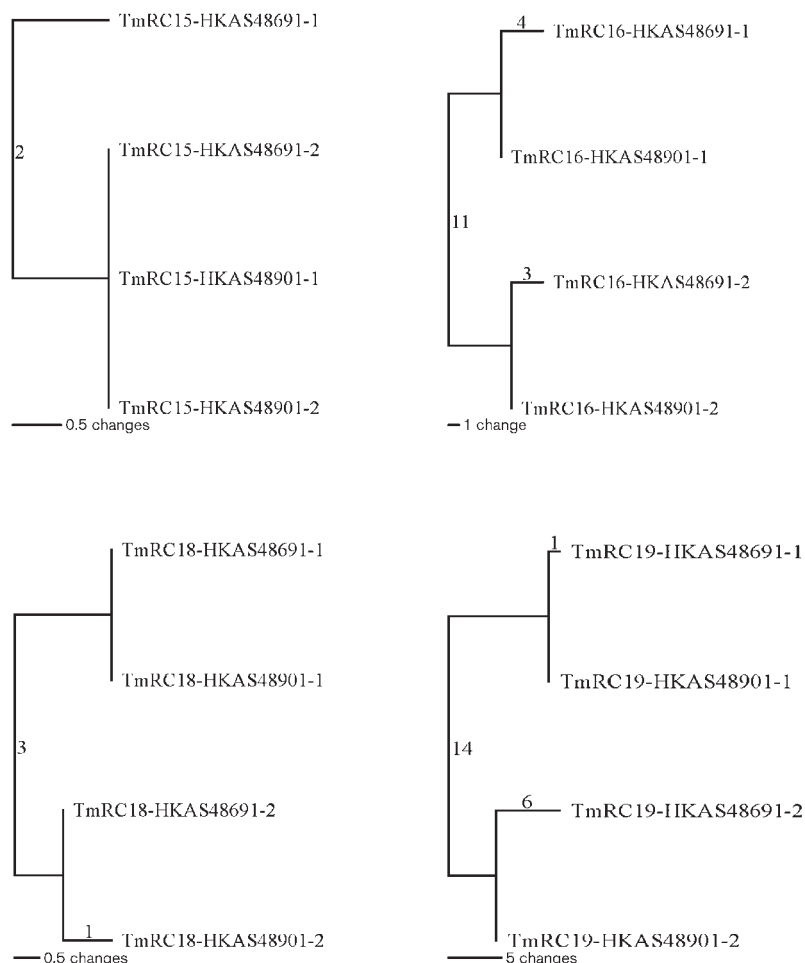


and TmRC19) showed this type of relationship (Fig. 3). The third type included only fragment TmRC15. Within this fragment, strain HKAS48901 was homozygous for one haplotype, and strain HKAS48691 was heterozygous, with one of the two haplotypes identical to that of strain HKAS48691 (Fig. 3). Among the distinct haplotypes, the distances (i.e. the number of steps) among them varied

widely (1–47), depending on the fragment and haplotype pair analysed (Figs. 2 and 3).

### SNP survey among natural strains

The potential usefulness of a subset of the discovered SNPs for future population studies was examined in a sample of



**Fig. 3.** Ten maximum-parsimonious trees for the four haplotypes in two strains of *T. matsutake*, one tree for each of the 10 DNA fragments. The consistency index for each tree is 1. In each tree, the haplotype identifiers included the analysed DNA fragment name (TmRC1, TmRC3..., and TmRC19), followed by the strain name (HKAS48691 or HKAS48901). The last digit (1 or 2) refers to the arbitrarily assigned haplotype number within each of the two strains. Within each of the 10 trees, the branch lengths are proportional to the amount of sequence divergence among the haplotypes for each DNA fragment. The number of nucleotide changes for each branch is shown.

33 wild mushroom strains. In all, 11 SNPs in seven DNA fragments predicted to be detectable by restriction enzyme digests were screened by PCR, enzyme digestion and agarose gel electrophoresis (Table 2). Using these seven primer pairs, we successfully amplified fragments of expected sizes from all 33 strains, including the two initial strains used to identify SNPs. Analysis by subsequent restriction enzyme digestion and gel electrophoresis revealed variable nucleotide frequencies at the SNP sites. For example, of the 11 SNPs the following six showed highly balanced allele frequencies: position 185 of TmRC4, positions 476 and 359 of TmRC14, position 123 of TmRC17, and positions 220 and 207 of TmRC19. In contrast, the remaining five SNPs had highly skewed allele frequencies. In addition, in this screening, we were able to amplify two fragments (TmRC9 and TmRC17) that previously failed to amplify from strain HKAS48901.

## DISCUSSION

In this study, we identified a large number of SNPs from the economically and ecologically important mushroom *T. matsutake*. These SNPs were distributed among 14 random

DNA fragments, and these fragments varied in their numbers of SNPs, from 1 to >40. Our observed SNP frequency here [1.1–1.7 % depending on the strain(s) analysed] was comparable to those observed in several other fungal species. For example, within the sequenced diploid strain SC5314 of the human pathogenic yeast *Candida albicans*, there is an SNP frequency of ~1 % per nucleotide (Jones *et al.*, 2004). Similarly, a survey of seven genomic loci (3013 nt in total) for 84 natural strains of the model yeast *Saccharomyces cerevisiae* from Asia has identified a total of 62 SNPs, yielding an SNP frequency of 2.05 % per nucleotide (Ayoub *et al.*, 2006). The SNP frequencies are higher in populations of several opportunistic pathogenic yeasts, such as *Candida parapsilosis* (~3.4 %; Fundyga *et al.*, 2004), and the species complexes of *Candida guilliermondii* (~6.3 %; Lan & Xu, 2006) and *Cryptococcus neoformans* (~20 % per nucleotide; Xu *et al.*, 2000b). However, for *C. parapsilosis*, *C. guilliermondii* and *Crypt. neoformans*, the population samples contain strains from distinct cryptic species and/or species (Fundyga *et al.*, 2004; Lan & Xu, 2006; Xu *et al.*, 2000b). The observed SNP frequency in *T. matsutake* was likely an underestimate of the true SNP frequency in this species. This is because sequences from only two strains were analysed for SNP discovery, and both strains were from the



**Table 2.** Confirmation and distribution of selected SNPs detectable by restriction endonuclease digests for 33 strains from seven regions in south-western China

These geographical regions are separated by high mountains (all >2000 m above sea level) and great rivers.

Strain	Geographical region	DNA fragment/position of SNP/restriction enzyme										
		TmRC4	TmRC9	TmRC12		TmRC14		TmR-C17	TmR-C18	TmRC19		
		185	206	762	91	476	359	123	110	220	207	204
		<i>EcoRI</i>	<i>FokI</i>	<i>RsaI</i>	<i>TaqI</i>	<i>FokI</i>	<i>BanII</i>	<i>TaqI</i>	<i>AluI</i>	<i>RsaI</i>	<i>AluI</i>	<i>DdeI</i>
HKAS48691	Lijiang, YN*	1/2†	1/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	1/2
HKAS48901	Chuxiong, YN	1/2	2/2	1/2	1/1	1/1	1/2	2/2	1/2	2/2	1/2	1/2
SL2-1	Shangri-la, YN	1/1	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	1/1
SL2-2	Shangri-la, YN	1/1	2/2	2/2	2/2	1/2	1/2	2/2	2/2	1/2	1/2	1/1
SL2-3	Shangri-la, YN	1/1	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	1/1
SL2-4	Shangri-la, YN	1/1	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/1	2/2
SL3-1	Shangri-la, YN	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	1/1
SL4-1	Shangri-la, YN	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	1/1
SL4-4	Shangri-la, YN	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/1	2/2
SL4-5	Shangri-la, YN	1/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	1/1
SL7-3	Shangri-la, YN	1/1	2/2	2/2	2/2	1/2	1/2	2/2	1/2	2/2	1/2	2/2
SL10-1	Shangri-la, YN	1/1	2/2	2/2	2/2	1/2	1/1	1/2	2/2	2/2	1/2	2/2
HQ2-1	Dali, YN	2/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	2/2	1/1	2/2
HQ3-1	Dali, YN	1/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	2/2
HQ4-1	Dali, YN	1/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	2/2
HQ5-1	Dali, YN	1/1	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	2/2
HQ6-1	Dali, YN	1/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	2/2	1/2	2/2
HQ7-1	Dali, YN	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/1	2/2
AL1-1	Jingdong, YN	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	2/2
AL1-2	Jingdong, YN	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	2/2
AL1-3	Jingdong, YN	1/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	2/2
AL1-4	Jingdong, YN	1/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	2/2	1/2	2/2
AL1-5	Jingdong, YN	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/1	2/2
AL2-1	Jingdong, YN	1/1	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/1	2/2
AL2-2	Jingdong, YN	1/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	1/2	1/2	2/2
XZ1	Tibet	1/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	2/2
XZ2	Tibet	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	2/2
SC2	Sichuan	1/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	2/2
SC3	Sichuan	1/1	2/2	2/2	2/2	1/2	1/1	2/2	2/2	2/2	1/2	2/2
SC4	Sichuan	1/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	2/2	1/2	2/2
SC6	Sichuan	1/1	2/2	2/2	2/2	1/2	1/2	2/2	2/2	2/2	1/2	2/2
SC7	Sichuan	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	1/2	2/2
SC11	Sichuan	1/1	2/2	2/2	2/2	1/2	1/2	2/2	1/2	2/2	1/2	2/2

\*YN, Yunnan province, south-western China.

†1/2, heterozygous; 1/1, homozygous with both alleles uncut; 2/2, homozygous with both alleles cut.

same general geographical area in south-western China. Analysing more strains should yield additional SNPs within the sequenced DNA fragments. Nevertheless, all the above-mentioned fungi have SNP frequencies much higher than that in the human genome, in which SNPs are observed approximately once every 250 bp (Miller *et al.*, 2005).

Interestingly, there is significant variation in the transition to transversion ratios among the above-mentioned species. In *T. matsutake*, our observed transition to transversion ratio was ~5:1. In *C. albicans*, *Crypt. neoformans* and *C.*

*guilliermondii*, the ratios are ~3:1. However, in *C. parapsilosis*, the ratio is ~1.4:1, similar to that in humans, in whom the transition to transversion ratio for genes in the human nuclear genome is ~1. The mechanism(s) responsible for differences in the nucleotide substitution patterns among these species and their potential evolutionary significance are worthy of further investigation.

The analyses of our small shotgun library identified a large number of gene fragments with sequences highly similar to two repetitive elements, marY1 and marY2N, identified in

earlier studies. Our results indicated that these two types of repetitive elements existed in our library at a frequency of ~10 % (25/221 fragments). In terms of nucleotides, the percentage was ~13.9 % (10 064/73 065 bp, including sequences of all 221 sequenced clones in the library). Both estimates suggest that these two repetitive elements are widespread in the genome of *T. matsutake*. Indeed, the results from real-time PCR analyses of the copy numbers of marY1 and marY2N suggest that these two elements make up ~5.5 % of the *T. matsutake* genomes of several strains isolated from Japan, Korea and China (Murata & Babasaki, 2005).

The diverse relationships among allelic sequences from within and between strains, as well as the identified SNPs, can help us address a variety of ecological and evolutionary questions. One such question relates to the reproductive mode of the species in nature. Specifically, although basidiospores are highly abundant in fruiting bodies of *T. matsutake*, their contribution to natural genetic variation within and between populations of this species has not been determined conclusively. One reason might be that basidiospores from *T. matsutake* have not been successfully germinated in the laboratory, and as a result, there has been no report of mating in this species. In addition, vegetative genetic individuals (i.e. 'shiro') can persist in nature (Lian *et al.*, 2006; Murata *et al.*, 2005), which suggests the potential existence of ancient asexual clones in this species. However, the recent analysis of gene and genotype frequencies at four microsatellite loci for one population of *T. matsutake* in Japan has identified that all four loci are in Hardy–Weinberg equilibrium (Lian *et al.*, 2006), consistent with the hypothesis of sexual reproduction and genetic exchange in natural populations of this species. Our data here are also consistent with this hypothesis. Specifically, the distribution of SNPs and the relationships among allelic sequences from within and between the two strains are highly variable among the gene fragments analysed here. Such a high variability is inconsistent with the notion that individuals in this species are ancient asexuals, in which SNPs should have accumulated at a similar rate across the entire genome (Welch & Meselson, 2000). Furthermore, several patterns of allelic relationships have been found, including allele sharing, both within and between individuals. In ancient asexual organisms, such variabilities among loci should not be seen. The existence of similarly divergent haplotypes within an individual strain across many loci has been used as evidence for ancient asexuality in bdelloid rotifers (Welch & Meselson, 2000).

Our study also identified the distribution of 11 SNPs detectable by restriction digestion and agarose gel electrophoresis. These SNPs are easy to detect and score. Therefore, they should be ideal for genotyping large population samples. Among the 11 SNPs screened, the six with balanced allele frequencies will be especially useful for strain typing. In combination with markers developed previously, the analysis of additional strains using the SNP markers developed here should allow further elucidation of gene flow and population structure in this important basidiomycete.

## ACKNOWLEDGEMENTS

We thank Yanchun Li for her technical assistance. This study was supported by grants from the National Natural Science Foundation of China (no. 30525002), the Innovation Program of Kunming Institute of Botany, the Chinese Academy of Sciences (KSCX2-YW-G-025), Genome Canada, and the Natural Science and Engineering Research Council (NSERC) of Canada.

## REFERENCES

- Ayoub, M.-J., Legras, J.-L., Saliba, R. & Gaillardin, C. (2006). Application of multi locus sequence typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon. *J Appl Microbiol* **100**, 699–711.
- Brumfield, R. T., Beerli, P., Nickerson, D. A. & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inference of population history. *Trends Ecol Evol* **18**, 249–256.
- Chapela, I. H. & Garbelotto, M. (2004). Phylogeography and evolution in matsutake and close allies inferred by analyses of ITS sequences and AFLPs. *Mycologia* **96**, 730–741.
- Fundyga, R. E., Kuykendall, R. J., Lee-Yang, W. & Lott, T. J. (2004). Evidence for aneuploidy and recombination in the human commensal yeast *Candida parapsilosis*. *Infect Genet Evol* **4**, 37–43.
- Gill, W. M., Guerin-Laguet, A., Lapeyrie, F. & Suzuki, K. (2000). Matsutake – morphological evidence of ectomycorrhizal formation between *Tricholoma matsutake* and host roots in a pure *Pinus densiflora* forest stand. *New Phytol* **147**, 381–388.
- Jones, T., Federspiel, N., Chibana, H., Dungan, J., Kalman, S., Magee, B. B., Newport, G., Thorstenson, R., Agabian, N. & other authors (2004). The diploid genome sequence of *Candida albicans*. *Proc Natl Acad Sci U S A* **101**, 7329–7334.
- Lan, L. & Xu, J. (2006). Multiple gene genealogical analyses suggest divergence and recent clonal dispersal in the opportunistic human pathogen *Candida guilliermondii*. *Microbiology* **152**, 1539–1549.
- Lian, C. L., Hogetsu, T., Matsushita, N., Guerin-Laguet, A., Suzuki, K. & Yamada, A. (2003). Development of microsatellite markers from an ectomycorrhizal fungus, *Tricholoma matsutake*, by an ISSR-suppression-PCR method. *Mycorrhiza* **13**, 27–31.
- Lian, C. L., Narimatsu, M., Nara, K. & Hogetsu, T. (2006). *Tricholoma matsutake* in a natural *Pinus densiflora* forest: correspondence between above- and below-ground genets, association with multiple host trees and alteration of existing ectomycorrhizal communities. *New Phytol* **171**, 825–836.
- Miller, R. D., Philips, M. S., Jo, I., Donaldson, M. A., Studebaker, J. F., Addleman, N., Alfisi, S. V., Ankener, W. M., Bhatti, H. A. & other authors (2005). High-density single-nucleotide polymorphism maps of the human genome. *Genomics* **86**, 117–126.
- Murata, H. & Babasaki, K. (2005). Intra- and inter-specific variations in the copy number of two types of retrotransposons from the ectomycorrhizal basidiomycete *Tricholoma matsutake*. *Mycorrhiza* **15**, 381–386.
- Murata, H. & Yamada, A. (2000). marY1, a member of the gypsy group of long terminal repeat retroelements from the ectomycorrhizal basidiomycete *Tricholoma matsutake*. *Appl Environ Microbiol* **66**, 3642–3645.
- Murata, H., Miyazaki, Y. & Yamada, A. (2001). marY2N, a LINE-like non-long terminal repeat (non-LTR) retroelement from the ectomycorrhizal homobasidiomycete *Tricholoma matsutake*. *Biosci Biotechnol Biochem* **65**, 2301–2305.

- Murata, H., Babasaki, K. & Yamada, A. (2005).** Highly polymorphic DNA markers to specify strains of the ectomycorrhizal basidiomycete *Tricholoma matsutake* based on  $\sigma$ marY1, the long terminal repeat of gypsy-type retroelement marY1. *Mycorrhiza* **15**, 179–186.
- Nakayama, K. & Nakanishi, J. (2004).** *Tricholoma matsutake* in Xichang, Sichuan and Chuxioug, Yunnan, China. In *Abstracts of the 48th Annual Meeting of the Mycological Society of Japan*, p. 45.
- Redhead, S. A. (1997).** The pine mushroom industry in Canada and the United States: why it exists and where it is going. In *Mycology in Sustainable Development*, pp. 15–54. Edited by M. A. Palm & I. H. Chapela. Boone, NC: Parkway Publishers.
- Swofford, D. L. (2004).** PAUP\*: *Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sunderland, MA: Sinaur Associates.
- Tominaga, Y. (1978).** *Tricholoma matsutake*. In *The Biology and Cultivation of Edible Mushrooms*, pp. 683–697. Edited by S. T. Chang & W. A. Hayes. London: Academic Press.
- Wang, Y., Hall, I. R. & Evans, L. A. (1997).** Ectomycorrhizal fungi with edible fruiting bodies. I. *Tricholoma matsutake* and related fungi. *Econ Bot* **51**, 311–327.
- Welch, D. M. & Meselson, M. (2000).** Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**, 1211–1215.
- Xu, J. (2006a).** Extracting haplotypes from diploid organisms. *Curr Issues Mol Biol* **8**, 113–122.
- Xu, J. (2006b).** Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* **15**, 1713–1731.
- Xu, J., Mitchell, T. G. & Vilgalys, R. (1999).** PCR-restriction fragment length polymorphism (RFLP) analyses reveal both extensive clonality and local genetic differences in *Candida albicans*. *Mol Ecol* **8**, 59–73.
- Xu, J., Ramos, A. R., Vilgalys, R. & Mitchell, T. G. (2000a).** Clonal and spontaneous origins of fluconazole resistance in *Candida albicans*. *J Clin Microbiol* **38**, 1214–1220.
- Xu, J., Vilgalys, R. & Mitchell, T. G. (2000b).** Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol Ecol* **9**, 1471–1481.
- 

Edited by: B. A. Horwitz