Specific populations of the yeast Geotrichum candidum revealed by molecular typing

Noémie Jacques, Sandrine Mallet, Fatima Laaghouiti, Colin R. Tinsley, Serge Casaregola

Specific populations of the yeast *Geotrichum candidum* revealed by molecular typing

Noémie Jacques[#], Sandrine Mallet[#], Fatima Laaghouiti, Colin R. Tinsley and Serge Casaregola[*]

Micalis Institute, INRA, AgroParisTech, CIRM-Levures, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

[*]Corresponding author: serge.casaregola@grignon.inra.fr

Running title: Specific populations of *Geotrichum candidum*

Abbreviations: MLST, Multi Locus Sequence Typing, Sequence Type, ST; RAPD, Random Amplification of Polymorphic DNA; LTR, Long Terminal Repeat; CLP, Chromosome Length Polymorphism

The GenBank/EMBL/DDBJ accession numbers for the sequences determined in this study are: HF558450 for *URA1* gene, HF558451 for *URA3* gene, HF558452 for *NUP116* gene, HF558453 *SAPT4* (*YSP3)* gene, HF586414 for *PLB3* gene and HF558454 for the retro-transposon LTR. The *G. candidum* MLST database can be accessed online at http://pubmlst.org/geotrichum/

# These authors participated equally to the work

Abstract

Geotrichum candidum is an ubiquitous yeast, and an essential component in the production of many soft cheeses. We developed a Multi-Locus Sequence Typing (MLST) scheme with five retained loci (*NUP116*, *URA1*, *URA3*, *SAPT4*, *PLB3*) which were sufficiently divergent to distinguish 40 Sequence Types (STs) among the 67 *G. candidum* strains tested. Phylogenetic analyses defined five main clades; one clade was restricted to environmental isolates, three other clades included distinct environmental isolates and dairy strains, while the fifth clade comprised 34 strains (13 STs), among which all but two were isolated from milk, cheese or dairy environment. These findings suggest an adaptation to the dairy ecosystems by a group of specialized European *G. candidum* strains. In addition, we developed a PCR inter-LTR scheme, a fast and reproducible RAPD-like method for *G. candidum*, to type the closely related dairy strains, which could not be distinguished by MLST. Overall, our findings distinguished two types of dairy strains, one forming a homogeneous group with little genetic diversity, and the other more closely related to environmental isolates. Neither regional nor cheese specificity was observed in the dairy *G. candidum* strains analyzed. This present study sheds light on the genetic diversity of both dairy and environmental strains of *G. candidum* and thus extends previous characterizations that have focused on the cheese isolates of this species.

Key words: MLST, inter-LTR, *Geotrichium candidum*, dairy, population

Introduction

The dimorphic yeast *Geotrichum candidum* (teleomorph *Galactomyces candidus*) is commonly found in foodstuffs, either as part of their normal constitution or as a contaminant. It is present in raw milk (Desmasures *et al.*, 1997) and is marketed as a starter for cheese making, because of its proteolytic activities and its organoleptic properties. For these reasons, *G. candidum* is desirable on the surface of semi-hard cheeses and mould-ripened or smeared soft cheeses (Boutrou and Gueguen, 2005), and is of considerable importance in the cheese-making industry (Pottier *et al.*, 2008).

G. candidum displays substantial morphological diversity and a wide ecological distribution. The genetic basis of this variability has been investigated: work based on RAM-PCR and RAPD-PCR (Gente *et al.*, 2002; Gente *et al.*, 2006; Marcellino *et al.*, 2001) demonstrated a high degree of genetic diversity of *G. candidum,* to some extent correlated to its various ecological niches and morphological polymorphism. PFGE profiling of 13 strains revealed highly diverse chromosome numbers, and hence genome sizes. It was suggested that genome size correlated with strain morphology (Gente *et al.*, 2006). It was recently shown that *G. candidum* displays an unusual intra-specific polymorphism in the ribosomal DNA sequences used for species identification (Alper *et al.*, 2011).

Among various typing methods, Multi-Locus Sequence Typing (MLST) is based on comparisons of partial housekeeping gene sequences. It is accepted to be reliable and reproducible, and has the advantage of portability, via the creation of a common database which can be continuously enlarged and updated by independent laboratories (Bougnoux *et al.*, 2004; Taylor and Fisher, 2003). MLST is the most popular typing method for yeast and has been applied to more than 10 species, including *Candida albicans* (Bougnoux *et al.*, 2002), *Candida glabrata* (Dodgson *et al.*, 2003), *Candida tropicalis* (Tavanti *et al.*, 2005), *Candida dubliniensis* (McManus *et al.*, 2008), *Candida krusei* (Jacobsen *et al.*, 2007),

Saccharomyces cerevisiae (Ayoub *et al.*, 2006; Fay and Benavides, 2005; Munoz *et al.*, 2009) and *Cryptococcus gattii* (Feng *et al.*, 2008). The first MLST scheme for *G. candidum* was developed using six gene fragments covering 3009 nucleotides, defining 17 ST among 40 strains (Alper *et al.*, 2013). This work confirmed the intra-specific variability observed in *G. candidum* (Gente *et al.*, 2002; Gente *et al.*, 2006; Marcellino *et al.*, 2001).

Long Terminal Repeats (LTRs) are short transposon-associated sequences of around 300-400 bp widely repeated in some genomes, like those of the genus *Saccharomyces* and other Saccharomycotina yeasts (Bleykasten-Grosshans and Neuveglise, 2011; Neuveglise *et al.*, 2002), either as flanking parts of the cognate retrotransposon or as isolated elements resulting from deletion of the retrotransposon through recombination between the two flanking LTRs. The lattersolo elements are numerous in some genomes, and this feature has been used to differentiate isolates within a single species by amplifying the regions between LTRs, using primers in the conserved regions of LTRs (inter-LTR PCR). This approach has been applied to various species including *S. cerevisiae* (Legras and Karst, 2003; Ness *et al.*, 1993), *Debaryomyces hansenii* and *Kluyveromyces marxianus* (Sohier *et al.*, 2009).

In this study, we exploit the completed genome sequence of the *G. candidum* strain CLIB 918 (=ATCC 204307) (Morel *et al.*, 2015) to develop protocols for typing this species and to gain insights on the phylogenetic relationship between isolates.

Materials and methods

Yeast strains and growth conditions.

Isolates used in this study are listed in Table 1. “FM” strains were provided during the “Food Microbiomes” project (http://www.agence-nationale-recherche.fr/?Project=ANR-08-ALIA-0007) by producers of starter cultures and cheeses, and were given FM numbers to keep their origin confidential; UCMA strains were provided by the University of Caen (France) and CLIB strains by CIRM-Levures (Thiverval-Grignon, France; http://www6.inra.fr/cirm/Levures). Most of the *G. candidum* isolates were from dairy products and are involved in the ripening of various cheeses, some are from other habitats such as soil; these strains were isolated in various places in France and the rest of the world. Other *G. candidum* strains were from Centralbureau voor Schimmelcultures (Utrecht, the Netherlands), Museum National d'Histoire Naturelle, Laboratoire de Cryptogamie (Paris, France), VTT Technical Research Center (Finland), Kasetsart University (Bangkok, Thailand), DSMZ (Braunschweig, Germany), BCCM/MUCL (Louvain la Neuve, Belgium), NCYC (Norwich, United Kingdom) and NBRC (Chiba, Japan). Routinely, yeast strains were grown in YPD (Yeast Peptone Dextrose: yeast extract 10 g l^{-1}, bacto peptone 10 g l^{-1}, glucose 10 g l^{-1}) at 28°C with shaking. For solid medium, 14 g l^{-1} of agar was added to YPD.

DNA extraction.

Cultures grown in 3 ml YPD medium overnight at 28°C were centrifuged at 2,500 *g* for 3 min, and the cell pellets were washed in 750 µl 50 mM EDTA. Each cell pellet was resuspended in 200 µl lysis buffer (1 % SDS, 2 % triton, 100 mM NaCl, 50 mM EDTA, 50mM Tris, pH=8), 200 µl chloroform/phenol (pH=8) and 300 mg glass beads and mechanically shaken by vortexing for 4 min. Then, 200 µl TE buffer was added and the

samples centrifuged for 5 min at 12,000 *g*. The aqueous phase was transferred to a new tube, and two chloroform extractions were carried out. DNA was precipitated with an equal volume of 100% ethanol and centrifuged for 4 min at 12,000*g*. The pellet was rinsed with 400 µl of 70 % ethanol, dried at room temperature for 15 min, resuspended in 50 µl TE buffer, and incubated with 2 µl RNase (10 mg ml^{-1}) for 30 min at 37°C. The DNA concentration was quantified on a 0.8% agarose gel (wt/vol) with 1X TAE electrophoresis buffer containing 0.2 mg ml^{-1} ethidium bromide, run at 100 V in a SUB-CELL GT electrophoresis system (Bio-Rad) for 60 min.

PCR amplification.

The primers used in this study, based on the complete genome sequence of *G. candidum* CLIB 918 (Morel *et al.*, 2015) were designed using Primer3 (http://fokker.wi.mit.edu/primer3). For inter-LTR PCR fingerprinting, sequences with similarities to a *G. candidum* retrotransposon LTR identified by Morel *et al.*, (2015) were aligned using MAFFT7 (Katoh and Standley, 2013). Primers were designed in the conserved regions of the aligned sequences. The primers used were GC_LTR3 (5'-TCTTTCCTTTAATTAATCAA) and GC_LTR5 (5'-TGAAGTGAACCAACATAA). For MLST, the primers used were NUP116_for (5'-ACCGCTACAACTGGATTTGG), NUP116_rev (5'-GAGACCTGTTTGAGGGCTTG), PLB3_for (5'-AAGAATATCTGGGATCTTTC), PLB3_rev (5'-TGAAGAAGAAGTACCAAGAA), SAPT4_for (5'-ATCATTAACACCCCGGCATA), SAPT4_rev (5'-GTGTCACCAAGCAGAGCAAA), URA1_for (5'-CAAGCCAATTGTGCTGAGAA), URA1_rev (5'-GGTGTCGTAGGGCAGTTGAT), URA3_for (5'-GCCAAAAAGACCAACCTGTG), URA3_rev (5'-CCTCATCCATACGGTTCTGC).

For MLST, between 25 and 50 ng of genomic DNA was amplified in a 50 µl PCR reaction mix containing 0.8 mM dNTP mixture, 0.1 µM forward and reverse primers in the manufacturer's recommended buffer and 1 U TaKaRa Ex *Taq*. Reactions were run on a 2720 thermal cycler (Applied Biosystems) as follows: 5 min at 94°C followed by 30 cycles of 30 sec at 94°C, 40 s at temperatures between 35 and 50°C and 30 sec at 72°C, with a final extension step of 7 min at 72 °C.

For inter-LTR PCR fingerprinting, 2 µl of DNA preparation (containing 20 ng) was added to a 23 µl PCR reaction mix containing 0.8 mM dNTP mixture, 0.8 µM GC1_LTR and GC2_LTR primers in the manufacturer's recommended buffer and 0.625 U TaKaRa Ex *Taq*. PCR conditions were as follows: 94°C for 5 min, 30 cycles of 94°C for 30 s, 37°C for 40 s and 72°C for 2 min, followed by a final extension at 72°C for 7 min. PCR products were visualized on a 1.2% agarose gel (wt/vol) 15 cm in length with 1X TAE electrophoresis buffer containing 0.2 mg ml^{-1} ethidium bromide, run at 100 V in a SUB-CELL GT electrophoresis system (Bio-Rad) for 90 min. The molecular marker was Gene ruler DNA Ladder Mix (Thermo Scientific). Acquisition of data by the Bionumerics software 6.5 (Applied Maths, Belgium) was as described in (Sohier *et al.*, 2009). Briefly, electrophoresis patterns were normalized to compensate for migration bias, non-linear background was subtracted from the patterns and comparison was based on the rolling disk principle. Inter-LTR variability not being appropriate for phylogenetic analysis, and the object being only to discriminate strains within an MLST sequence type, interpretation of the Bionumerics output was by visual observation ; no clustering algorithm was performed.

Synonymous and non-synonymous base ratio

Codon by codon analysis was performed using MEGA7 (Kumar *et al.*, 2016). Pairwise peptide comparisons were conducted with MEGA7 (codon based Z-test for selection) and

with the package *Dnasp* (Librado and Rozas, 2009). The likelihood analysis was performed using the REL algorithm of the HyPhy package implemented in *DataMonkey* (Pond and Frost, 2005).

DNA sequence determination and phylogenetic analysis

PCR fragments were sequenced on both strands by Eurofins MWG Operon (Ebersberg, Germany). Sequences were assembled with the phred/phrap/consed package (Ewing and Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998) and analyzed with BLAST implemented at NCBI (http://www.ncbi.nlm.nih.gov). Sequence alignments were generated using MAFFT7 (Katoh and Standley, 2013). Phylogenetic trees were constructed with the Neighbor-Joining program implemented in MEGA6 (Tamura *et al.*, 2013). Phylogenetic trees were visualized with NJplot (Perriere and Gouy, 1996). Bionumerics software 6.5 (Applied Maths, Belgium) was used for the acquisition of inter-LTR profiles and the integration of MLST, inter-LTR and phenotypic data.

Phylogenetic analysis with *Structure*

The MLST data were used to predict the population structure on the basis of the clustering program *Structure* (Pritchard *et al.*, 2000). Analyses were performed with burn-in of 80,000 cycles and 80,000 rounds of calculation. In accordance with the authors' recommendations, the best estimation of the number of populations, K, was taken as the value of K for which the difference in log(probability of data) between K and K+1 diminished abruptly (*ie.*, the log probability ceased to increase rapidly). The population structure was thus calculated with K = 5 populations. For all other parameters, the recommended default values were used.

Phylogenetic analysis using *Splits tree*

191 The MLST data were used to predict genetic exchanges with the program *Splits tree* (Huson
192 and Bryant, 2006). Recommended default values were used.
193

Results

Design of a MLST scheme for *G. candidum*.

A literature survey was undertaken to identify genes which have been used for MLST schemes in previous studies on yeasts (Bougnoux *et al.*, 2002; Dodgson *et al.*, 2003; Jacobsen *et al.*, 2007; McManus *et al.*, 2008; Munoz *et al.*, 2009; Tavanti *et al.*, 2005). The sequences of 28 *G. candidum* orthologs of such genes were extracted from the genome of CLIB 918. PCR primers were designed to amplify fragments of around 500 bp and were tested on 67 strains, which were considered to be haploid on the basis of their mating type characteristics (Morel *et al.*, 2015). The study population included 34 *G. candidum* strains isolated from cheese, 4 from milk, 1 from the dairy environment, 1 from cow's udder, 3 dairy product contaminants and 17 isolates from diverse origins (fruit, human faeces, industrial plant, industrial waste…). The latter group contained three *G. silvicola* strains, now reassigned to the *G. candidum* species. A preliminary analysis was performed on closely related dairy strains from France to select the most divergent markers. Five markers were selected as showing the greatest sequence divergence between strains: *URA1, URA3, SAPT4* (*YSP3* in *S. cerevisiae*), *NUP116* and *PLB3.* Data for the five MLST markers are presented in Table 2. The PCR fragments analyzed were between 393 bp for *PLB3* and 501 bp for *SAPT4,* yielding a concatenated sequence length of 2254 bp which contained 90 polymorphic sites.

Table 2 summarizes information concerning the diversity at each locus. The number of polymorphic sites, only consisting in substitutions, varied between 11 for *URA1* and 28 for *PLB3*. The number of allelic profiles between loci was five for *SAPT4* and nine for *PLB3*, *URA1, URA*3 and *NUP116*. The average divergence for the alleles ranges from 0.2 % for *SAPT4* to 4.7 % for *PLB3* (Supplementary Table 1).

Analysis using MEGA7 showed evidence of positive selection in 4 of the 14 variable codons of the *NUP116* sequence, in 2 of the 28 codons of *PLB3,* in 2 of 21 of *SAPT4*, in 1 of 11 of

URA1 and in 1 of 16 of *URA3*, though without good statistical support (P ≈ 0.7). Pairwise peptide comparisons showed a positive overall dN - dS for the sequences of *NUP116* and of *PLB3* in strain MUCL8652 when compared with about half of the other strains (P=0.16). Analysis with the package DnaSP gave values of dN/dS always smaller than unity, excepting cases where dS was zero. In the case of *NUP116* over 90% of the values of dN/dS were smaller than 0.25; for the other four genes, over 90% were less than 0.15. A likelihood approach showed 2 positively selected sites (Bayes factors > 100) in the sequence of *SAPT4* and none in the other proteins. Considering the feeble evidence from the codon by codon analysis and the Z-test, the absence of evidence from the analysis in Dnasp and the prediction of 2/21 codons subject to selection in *SAPT4* by the likelihood analysis, we may conclude that the large majority of sites used in the analysis are not subject to positive selection.

Geographical and industrial influence on strain grouping.

Forty sequence types (ST) were defined amongst the 67 strains, based on the possession of identical concatenated sequences. Overall, 31 strains were the sole member of their ST, while the other 36 strains were distributed among the remaining 9 STs: ST 21 and ST 23 each contained two strains, ST 14 and ST 19 contained three strains, ST 7, ST 18, ST 29 and ST 30 contained five strains and ST 22 contained six strains (Table 3). The pairwise divergence between STs (averaged over the concatenated sequences) was from 0.04 % to 2.57 % and average divergence for all STs was 0.96 %.

Phylogenetic trees were constructed by neighbor joining following alignment of the sequences using MAFFT7 (Figures 1 and 2). The tree constructed with ST type sequences, shown in Figure 1, differentiated five main clades. Two well-separated clades 1 and 2 included mainly environmental isolates and dairy strains. Clade 1 contained three STs represented by a single isolate and one (ST 7) which contained five strains. Clade 2 contained

five single-isolate STs. A third clade also heterogeneous with regard to the origin of the strains contained 14 single-strain STs and one (ST 14) composed of three strains. Clade 4 contained three single-strain STs and consisted of two strains isolated from soil and one from fruit. Finally, in contrast to the previous clades, clade 5 contained only strains involved in cheese making (either from cheese made with raw milk or from starters used in the cheese industry), with the exception of two isolates (FM 268 and FM 269, in ST 18 and ST 27 respectively) which were isolated from human stools. It may be reasonable to propose that these latter isolates have a cheese origin, considering their position in the tree and their geographical origin. Isolates assigned to clade 5 were less divergent; seven of the thirteen STs contained from two to six strains.

The slovakian strain CBS 11176 was classified among cheese isolates of clade 5. This strain was first described as *Geotrichum bryndzea* (Sulo *et al*., 2009), but it was recently shown to belong to *Gal. candidus*, the teleomorph of *G. candidum* (Groenewald *et al.*, 2012); the position of this strain in the phylogenetic tree shown in Figure 2 confirmed the recent taxonomic change.

Population analysis

Since *G. candidum* is a sexually reproducing species, we performed analyses using the programs *Structure* and *Splits tree*. *Structure* clusters the isolates based on their respective genetic profiles which are considered in terms of the relative contributions from a predefined number of hypothetical ancestral genomes. *Splits tree* calculates a phylogenetic tree which permits the visualization of genetic exchanges in addition to mutations.

Preliminary analysis of the data with *Structure* suggested that the data were best explained under the hypothesis of five ancestral genotypes. The results of clustering on this basis (Figure 3a) distinguish the previously defined clades 1 and 5, and suggest that the clade 2

contains hybrid organisms with genetic contributions originating in clades 1 and 3. Individual gene trees for each of the markers also indicated genetic exchanges between isolates (Supplementary Figure 1). As suggested by the above Neighbor-Joining analysis (Figures 1 and 2), the clade 3 is composed of two subgroups, while clade 4 branches from this group. Clades 1 and 5 contained almost exclusively dairy strains, or strains isolated from human faeces, presumably as a consequence of having eaten dairy products, while clades 2, 3 and 4 contained a mixture of dairy and environmental strains. *Splits tree* gave similar results (Figure 3b), separating the strains into groups which, with few exceptions, corresponded to the clades previously defined. The analysis separated clades 1 and 5 and illustrated the derivation of clades 2 and 4 from the clade 3. Again, the distinction of clades 1 and 5, of dairy origin, from the other groups, of which environmental strains constituted about half the isolates, suggests that dairy strains have developed separately from non-dairy strains.

Inter-LTR PCR distinguishes most of the *G. candidum* dairy strains.

Although MLST was very informative about the divergence of environmental strains and dairy strains, it was not clear whether the large ST groups were due to clonality or limits to the discrimination of the MLST markers. We therefore developed an additional typing method more sensitive to genetic changes occurring at short time-scales.

Sequence analysis of the complete genome of *G. candidum* CLIB 918 revealed the presence of a Long Terminal Repeat (LTR)-retrotransposon likely related to the yeast Ty families (Morel *et al*., 2015). The corresponding 385 bp long LTR sequences found in the genome of CLIB 918 were aligned with MAFFT7 and oligonucleotide primers were designed on the basis of conserved regions (Supplementary Figure 2). These were used to amplify genomic DNA from the 36 strains which were in multi-strain STs (listed in Table 3). Various primer pairs and various PCR conditions were tested; those leading to the most discriminating results

(GC_LTR3 and GC_LTR5) were retained. In the subsequent analysis all the inter-LTR PCRs were repeated at least twice as previously described (Sohier *et al.*, 2009).

Results of electrophoresis are shown in Figure 3 and Supplementary figure 3. The inter-LTR profiles were compared to the MLST tree obtained previously to highlight the variability of strains within the STs, now observable with inter-LTR analysis. The amplified bands ranged from 200 to 4000 bp and the profiles of the various strains differed in number, size and relative intensity of fragments. In the majority of cases, strains within the multi-strain ST were distinguished by inter-LTR analysis. Thus, 28 strains were uniquely discriminated by a combination of MLST and inter-LTR, while two strains of ST 29 (CLIB 1239 and CLIB 1241) shared the same inter-LTR profiles, as did FM 76 and FM 77 in ST 18, and CLIB 1244, CLIB 1245, CLIB 1246, CLIB1247 and CLIB 1253 in ST 30. The two strains of ST 29 originated from the Franche-Comté region and the five strains in ST 30 were isolated from the Haute-Savoie; their classification into the same clusters is therefore coherent. In conclusion, the combination of MLST and inter-LTR allowed precise discrimination between cheese strains.

As described previously (Marcellino *et al.*, 2001), we observed a limited, but reproducible, phenotypic diversity in terms of carbon source assimilation (galactose, lactose and ribose; Figure 3). The capacity to utilize ribose is rarely found in *G. candidum*. We found that six of the 37 isolates could utilize ribose. They were mainly associated with ST 18 (4 of 5 isolates tested). Only two, unrelated strains were lactose-negative. Galactose assimilation was absent from ST 30 and from two thirds of the isolates belonging to the sister clade ST 22, these being all cheese strains from the region Haute-Savoie region.

Discussion

We report the development of tools for investigating the evolution of *G. candidum* and for studying industrial strains used in cheese production. The MLST analysis we describe here involving five single copy genes confirms the high genetic diversity reported for *G. candidum*. Previous typing studies based on RAM-PCR, RAPD and Chromosome Length Polymorphism (CLP) techniques suggest substantial diversity within the species (Marcellino *et al*., 2001; Gente *et al*., 2002; Gente *et al*., 2006), which was again evidenced by the variability of the ribosomal DNA sequences observed by (Alper *et al.*, 2011). Our results demonstrate a large degree of genetic variation in the five genes considered (between 2.4 and 7.1 % of polymorphic sites, this being higher than the 0.9 to 3.2 % found in the study of (Alper *et al.*, 2013).

In this work, *G. candidum* strains could be readily differentiated on the basis of the partial sequences of five genes: 40 different STs were obtained for 67 isolates based on 90 polymorphic sites within the 2254 bp. This compares favorably with the 17 STs distinguished in the MSLT scheme of Alper *et al*. (2013) using 58 polymorphic sites in 6 loci (3009 bp),our choice of candidate genes being facilitated by the availability of the whole genome sequence of *G. candidum*. In addition, the analysis was performed on a larger number of isolates from a wide range of geographical or ecological origins, including the environment, industrial plant or food contaminant, which permitted a more general view of the phylogenetic diversity of the species.

Twenty-eight of the strains studied here have previously been studied by RAPD (Marcellino *et al.*, 2001). The previous RAPD results and our findings are generally coherent, although there were some differences. The three strains, CLIB 1267 (GC129 in Marcellino *et al*., 2001), CLIB 1258 (=GC101) and CLIB 1237 (=GC37), isolated from milk used respectively for manufacturing Camembert (Normandy), St Nectaire (Auvergne) and Chaource cheeses

(Champagne-Ardenne), were previously found by (Marcellino *et al.*, 2001) to be closely related to other cheese strains. However, we found that they are well separated from other cheese strains and closely related to environmental strains (Figure 2). Another difference concerns the yeasts that we assign to ST30 (composed of isolates from ripening cheese from Annecy). The isolate CLIB 1253 (=GC90) was suggested to be different from the other strains that we classified as ST 30 (Marcellino *et al.*, 2001). We clearly show here that the five strains belonging to ST 30 are indistinguishable on the basis of MLST, and inter-LTR profile. In addition they have identical phenotypic characteristics (this work), and we have previously shown that they carry the same mating type (Morel *et al*., 2015). These strains were isolated from two different cheeses, Reblochon and Tomme de Savoie, made in different dairy facilities. Our results therefore indicate that these are clones of the same strain present at more than one particular site and in more than one cheese, suggesting that *G. candidum* isolates are not region or cheese specific.

The 67 isolates studied here were classified in 40 STs. A total of 31 isolates were typed by MLST alone as they all display a unique ST. Though all of the environmental isolates could be distinguished, there were a number of dairy isolates (eight isolates in clades 1, 2 or 3 and 28 isolates belonging to the dairy clade 5) that were not typed by this approach. We were able to improve strain differentiation by developing inter-LTR PCR for *G. candidum*, which successfully differentiated all but nine of the strains that were not separated by MLST. Inter-LTR PCR is simple, reliable and fast, as demonstrated previously for other species (Ness *et al*., 1993; Legras and Karst, 2003; Sohier *et al*., 2009) and may therefore be useful for strain differentiation, for example, in such applications as analysis of yeast population dynamics during cheese ripening.

It has been reported that *G. candidum* is highly variable as assessed by chromosome separation using PFGE (Gente *et al.*, 2006). However, CLP very often results from ectopic

368 recombination between similar sequences, such as transposons (Zolan, 1995; Casaregola et
369 al., 1998; Rachidi et al., 1999). Accordingly, we observed a very large number of various
370 kinds of transposons in the genome of *G. candidum* (Morel *et al.*, 2015). The discrepancy
371 between MLST and RAPD may be inherent to the techniques used because methods based on
372 the amplification of repeats do not measure genetic diversity directly, but rather, like PFGE,
373 indicate chromosomal differences. It has been demonstrated, for example, that the species
374 *Schizosaccharomyces pombe* displays substantial CLP which is not associated with either
375 sequence or phenotypic variability (Brown *et al.*, 2011). In this study, strains which could not
376 be separated on the basis of sequence divergence, *i.e.* MLST, displayed little phenotypic
377 diversity; they are therefore presumably closely related. The use of inter-LTR PCR permitted
378 a fine-scale discrimination of these related isolates within a single ST.
379 Our MLST analysis differentiated a clade containing dairy isolates from those containing
380 environmental isolates mixed with dairy isolates, in agreement with the suggestion that *G.*
381 *candidum* species could be separated into two groups, on the basis of rRNA variability and
382 genomic sequence diversity (Alper *et al.*, 2013). The observation of these authors concerns
383 dairy yeasts, and may correspond to the two types of dairy yeasts that we evidenced (clade 5
384 vs. clades 2, 3 and 4). In summary, our work has brought to light a population of dairy yeasts
385 which is clearly distinct from the clades containing a mixture of dairy and environmental
386 strains. This is reminiscent of population genomics findings for *S. cerevisiae* (Liti et al., 2009;
387 Schacherer et al., 2009), in which specialized isolates formed separated clades. *G. candidum*
388 clade 5 is reminiscent of the *S. cerevisiae* European wine group in which strains are closely
389 related and specialized in wine making.
390 In addition to shedding light on the population structure and probable evolution of the species
391 *G. candidum*, our work provides information and methods with potential to facilitate genetic
392 improvement and studies of genetic diversity within this species. For example, it provides a

framework which will make it easier to correlate genetic divergence with industrially important characteristics of *G. candidum*, such as aroma production, morphology and growth characteristics.

Acknowledgements

This work received funding from the *Agence National pour la Recherche* grant "Food Microbiomes" (ANR-08-ALIA-007-02). This publication made use of the *Geotrichum* MLST website (http://pubmlst.org/geotrichum/) developed by Keith Jolley and sited at the University of Oxford (Jolley and Maiden, 2010). The development of this site has been funded by the Wellcome Trust. We thank Christelle Louis-Mondésir for expert technical assistance. We are grateful to Prof. Savitree Limtong (Kasetsart University, Thailand) for providing us with the isolate NT12. We are grateful to Guillaume Morel (supported by a CIFRE fellowship with CNIEL) for the extraction from the genome sequence of CLIB 918 of the genes of interest. We also thank the *Centre National Interprofessionnel de l'Economie Laitière* (CNIEL) and the *Syndicat Professionnel des Producteurs d'Auxiliaires pour l'Industrie Laitière* (SPPAIL) for providing us with dairy strains.

References

Alper I, Frenette M, Labrie S. 2011. Ribosomal DNA polymorphisms in the yeast Geotrichum candidum. *Fungal Biol* **115:** 1259-69.

Alper I, Frenette M, Labrie S. 2013. Genetic diversity of dairy Geotrichum candidum strains revealed by multilocus sequence typing. *Appl Microbiol Biotechnol* **97:** 5907-20.

Ayoub MJ, Legras JL, Saliba R, *et al.* 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.

Bleykasten-Grosshans C, Neuveglise C. 2011. Transposable elements in yeasts. *C R Biol* **334:** 679-86.

Bougnoux ME, Aanensen DM, Morand S, *et al.* 2004. Multilocus sequence typing of Candida albicans: strategies, data exchange and applications. *Infect Genet Evol* **4:** 243-52.

Bougnoux ME, Morand S, d'Enfert C. 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of Candida albicans. *J Clin Microbiol* **40:** 1290-7.

Boutrou R, Gueguen M. 2005. Interests in Geotrichum candidum for cheese technology. *Int J Food Microbiol* **102:** 1-20.

Brown WR, Liti G, Rosa C, *et al.* 2011. A Geographically Diverse Collection of Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. *G3 (Bethesda)* **1:** 615-26.

Desmasures N, Bazin F, Gueguen M. 1997. Microbiological composition of raw milk from selected farms in the Camembert region of Normandy. *J Appl Microbiol* **83:** 53-8.

Dodgson AR, Pujol C, Denning DW, *et al.* 2003. Multilocus sequence typing of Candida glabrata reveals geographically enriched clades. *J Clin Microbiol* **41:** 5709-17.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186-94.

Ewing B, Hillier L, Wendl MC, *et al.* 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8:** 175-85.

Fay JC, Benavides JA. 2005. Evidence for domesticated and wild populations of Saccharomyces cerevisiae. *PLoS Genet* **1:** 66-71.

Feng X, Yao Z, Ren D, *et al.* 2008. Genotype and mating type analysis of Cryptococcus neoformans and Cryptococcus gattii isolates from China that mainly originated from non-HIV-infected patients. *FEMS Yeast Res* **8:** 930-8.

Gente S, Desmasures N, Panoff JM, *et al.* 2002. Genetic diversity among Geotrichum candidum strains from various substrates studied using RAM and RAPD-PCR. *J Appl Microbiol* **92:** 491-501.

Gente S, Sohier D, Coton E, *et al.* 2006. Identification of Geotrichum candidum at the species and strain level: proposal for a standardized protocol. *J Ind Microbiol Biotechnol* **33:** 1019-31.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* **8:** 195-202.

Groenewald M, Coutinho T, Smith MT, *et al.* 2012. Species reassignment of Geotrichum bryndzae, Geotrichum phurueaensis, Geotrichum silvicola and Geotrichum vulgare based on phylogenetic analyses and mating compatibility. *Int J Syst Evol Microbiol* **62:** 3072-80.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23:** 254-67.

Jacobsen MD, Gow NA, Maiden MC, *et al.* 2007. Strain typing and determination of population structure of Candida krusei by multilocus sequence typing. *J Clin Microbiol* **45:** 317-23.

Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11:** 595.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30:** 772-80.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33:** 1870-4.
Larkin MA, Blackshields G, Brown NP, *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.
Legras JL, Karst F. 2003. Optimisation of interdelta analysis for Saccharomyces cerevisiae strain characterisation. *FEMS Microbiol Lett* **221:** 249-55.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-2.
Marcellino N, Beuvier E, Grappin R, *et al.* 2001. Diversity of Geotrichum candidum strains isolated from traditional cheesemaking fabrications in France. *Appl Environ Microbiol* **67:** 4752-9.
McManus BA, Coleman DC, Moran G, *et al.* 2008. Multilocus sequence typing reveals that the population structure of Candida dubliniensis is significantly less divergent than that of Candida albicans. *J Clin Microbiol* **46:** 652-64.
Morel G, Sterck L, Swennen D, *et al.* 2015. Differential gene retention as an evolutionary mechanism to generate biodiversity and adaptation in yeasts. *Sci Rep* **5:** 11571.
Munoz R, Gomez A, Robles V, *et al.* 2009. Multilocus sequence typing of oenological Saccharomyces cerevisiae strains. *Food Microbiol* **26:** 841-6.
Ness F, Lavallée F, Dubourdieu D, *et al.* 1993. Identification of yeast strains using the polymerase chain reaction. *Journal Sci Food Agri.* **62:** 69-94.
Neuveglise C, Feldmann H, Bon E, *et al.* 2002. Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. *Genome Res* **12:** 930-43.
Perriere G, Gouy M. 1996. WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78:** 364-9.
Pond SL, Frost SD. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21:** 2531-3.
Pottier I, Gente S, Vernoux JP, *et al.* 2008. Safety assessment of dairy microorganisms: Geotrichum candidum. *Int J Food Microbiol* **126:** 327-32.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945-59.
Sohier D, Dizes AS, Thuault D, *et al.* 2009. Important genetic diversity revealed by inter-LTR PCR fingerprinting of Kluyveromyces marxianus and Debaryomyces hansenii strains from French traditional cheeses. *Dairy Science & Technology* **89:** 569-581.
Tamura K, Stecher G, Peterson D, *et al.* 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30:** 2725-9.
Tavanti A, Davidson AD, Johnson EM, *et al.* 2005. Multilocus sequence typing for differentiation of strains of Candida tropicalis. *J Clin Microbiol* **43:** 5593-600.
Taylor JW, Fisher MC. 2003. Fungal multilocus sequence typing--it's not just for bacteria. *Curr Opin Microbiol* **6:** 351-6.

Table 1 List of strains studied

Species	Isolate	Other names	Origin	Source
Geotrichum bryndzae[a]	CLIB 1366^T	CBS 11176^T	Slovakia	Bryndza cheese
Geotrichum candidum	CLIB 1368^T	CBS 615.84^T	Ile de France	Brie cheese
//	CLIB 918	ATCC 204307	Normandy	Pont l'Evêque cheese
//	CLIB 1235	GC12	Normandy	Camembert cheese
//	CLIB 1236	GC 21	Normandy	Goat's cheese
//	CLIB 1237	GC37	Normandy	Camembert cheese
//	CLIB 1239	GC43	Franche-Comté	Mont d'Or cheese
//	CLIB 1240	GC44	Haute-Savoie	Reblochon cheese
//	CLIB 1241	GC45	Franche-Comté	Mont d'Or cheese
//	CLIB 1242	GC42	Haute-Savoie	Reblochon cheese
//	CLIB 1243	GC51	Haute-Savoie	Reblochon cheese
//	CLIB 1244	GC59	Haute-Savoie	Tomme de Savoie cheese
//	CLIB 1245	GC60	Haute-Savoie	Reblochon cheese
//	CLIB 1246	GC63	Haute-Savoie	Reblochon cheese
//	CLIB 1247	GC64	Haute-Savoie	Tomme de Savoie cheese
//	CLIB 1248	GC74	Haute-Savoie	Reblochon cheese
//	CLIB 1251	GC79	Burgondy	Epoisses cheese
//	CLIB 1252	GC84	Burgondy	Epoisses cheese
//	CLIB 1253	GC90	Haute-Savoie	Reblochon cheese
//	CLIB 1255	GC96	Haute-Savoie	Reblochon cheese
//	CLIB 1256	GC97	Haute-Savoie	Reblochon cheese
//	CLIB 1257	GC100	Auvergne	Saint Nectaire cheese
//	CLIB 1258	GC101	Auvergne	Saint Nectaire cheese
//	CLIB 1260	GC105	Auvergne	Saint Nectaire cheese
//	CLIB 1262	GC110	Auvergne	Saint Nectaire cheese
//	CLIB 1263	GC120	Auvergne	Saint Nectaire cheese
//	CLIB 1267	GC129	Champagne-Ardenne	Chaource cheese
//	CLIB 1270	GC146	Auvergne	Saint Nectaire cheese
//	CLIB 1274	GC164	Haute-Savoie	Reblochon cheese
//	CLIB 1283	UCMA 103	Normandy	Pont l'Evêque cheese
//	CLIB 1285	UCMA 3936	Normandy	Dairy
//	CLIB 1361	CBS 182.33	Italy	Yoghurt
	CLIB 1415	LCP 51.590	Spain	Sand
//	FM 03		unknown	Cheese contaminant
//	FM 115		unknown	unknown
//	FM 122		unknown	unknown
//	FM 125		unknown	unknown
//	FM 127		unknown	unknown
//	FM 128		unknown	unknown
//	FM 136		unknown	unknown
//	FM 212		France	Corn silage

//	FM 214		Normandy	Cow's milk
//	FM 260		Normandy	Cow's udder
//	FM 268		Normandy	Stools
//	FM 269		Normandy	Stools
//	FM 270		Normandy	Stools
//	FM 29		Auvergne	Goat's cheese
//	FM 30		Auvergne	Goat's cheese
//	FM 31		Auvergne	Goat's cheese
//	FM 34		Auvergne	Goat's cheese
//	FM 76		Normandy	Cow's milk
//	FM 77		unknown	unknown
//	CBS 117138		Italy	Municipal compost
//	CBS 11628		South Africa	Soil
//	CBS 476.83		Senegal	Soil
//	CBS 557.83		Egypt	Fruit
//	DSM 10452		Germany	Sauerkraut
//	DSM 13629		United Kingdom	Polyurethane
//	MUCL 11539		United Kingdom	Polluted water
//	MUCL 14462		United States	Squash, Curcubita
//	MUCL 881		Belgium	Milk
//	MUCL 8652		Belgium	Soaked hay
//	NBRC 5368		unknown	Butter
//	NCYC 49		unknown	Milk
Geotrichum silvicola[a]	CLIB 1378^T	CBS 9194^T	Brazil	Insect
//		NT 12	Thailand	Rain forest
//		VTTC 4559	Sweden	Malting system

[a] previous species denomination. GC, strains described in Marcelino et al. (2001).

Table 2 Characteristics of the five loci studied

Locus	Size of examined amplicon (bp)	Sequence divergence (%)	Number of polymorphic sites	Number of allelic profile per locus
NUP116	425	3.29	14	9
URA1	465	2.36	11	9
URA3	470	3.40	16	9
SAPT4	501	4.19	21	5
PLB3	393	7.12	28	9

Table 3 MLST profiles for each strains of this study and distribution of isolates in STs

Strains	Markers					STs
	NUP116	***PLB3***	***SAPT4***	***URA1***	***URA3***	
NT 12	6	2	2	5	4	1
CLIB 1267	2	2	4	1	1	2
CLIB 1237	2	2	2	2	1	3
LCP 51.590	2	2	5	2	3	4
CLIB 1235	1	2	1	2	1	5
CBS 182.33	2	2	2	1	1	6
CLIB 1236	2	2	1	2	1	7
CLIB 1251	2	2	1	2	1	7
CLIB 1252	2	2	1	2	1	7
CLIB 1263	2	2	1	2	1	7
FM 125	2	2	1	2	1	7
CBS 9194	4	3	3	3	1	8
FM 122	3	7	2	2	1	9
CLIB 1283	2	1	2	1	1	10
FM 270	2	1	2	4	3	11
FM 03	5	4	2	2	1	12
FM 212	2	4	2	2	1	13
FM 136	1	4	2	2	1	14
CLIB 1258	1	4	2	2	1	14
CBS 117138	1	4	2	2	1	14
VTTC 4559	1	4	2	2	5	15
CLIB 1274	2	5	1	2	1	16
FM 260	3	4	1	2	3	17
FM 214	2	1	1	2	3	18
FM 268	2	1	1	2	3	18
FM 34	2	1	1	2	3	18
FM 76	2	1	1	2	3	18
FM 77	2	1	1	2	3	18
CBS 11176	2	1	1	2	1	19
CLIB 1248	2	1	1	2	1	19
CLIB 1260	2	1	1	2	1	19
CBS 615.84	3	1	3	2	1	20
CLIB 1257	1	1	1	2	1	21
FM 128	1	1	1	2	1	21
CLIB 1240	1	1	1	1	1	22
CLIB 1242	1	1	1	1	1	22
CLIB 1243	1	1	1	1	1	22
CLIB 1255	1	1	1	1	1	22
FM 127	1	1	1	1	1	22
MUCL 881	1	1	1	1	1	22
FM 29	1	1	1	1	3	23
FM 31	1	1	1	1	3	23

FM 30	1	6	1	1	3	24
CLIB 1285	2	1	1	4	1	25
FM 115	3	6	5	1	1	26
FM 269	2	1	1	1	3	27
CLIB 918	2	1	1	1	2	28
CLIB 1239	2	1	1	1	1	29
CLIB 1241	2	1	1	1	1	29
CLIB 1256	2	1	1	1	1	29
CLIB 1262	2	1	1	1	1	29
CLIB 1270	2	1	1	1	1	29
CLIB 1244	5	1	1	1	1	30
CLIB 1245	5	1	1	1	1	30
CLIB 1246	5	1	1	1	1	30
CLIB 1247	5	1	1	1	1	30
CLIB 1253	5	1	1	1	1	30
DSM 10452	2	2	1	2	6	31
MUCL 11539	2	4	2	2	6	32
CBS 11628	7	7	2	6	7	33
CBS 476.83	8	7	2	6	8	34
CBS 557.83	7	8	2	9	8	35
NBRC 5368	9	4	2	7	1	36
MUCL 8652	3	9	2	1	1	37
DSM 13629	2	6	2	7	1	38
MUCL 14462	2	3	5	8	9	39
NCYC 49	2	2	1	2	2	40

1

2

Figure legends

Figure 1: Neighbor-Joining tree of the 40 STs generated from the analysis of the concatenation of the five selected sequences in 67 *G. candidum* isolates. Bootstrap values (1000 repetitions) are indicated at the main nodes. The evolutionary distances were computed using the Kimura 2-parameter method and are expressed as number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). The analysis involved 40 nucleotide sequences with a total of 2254 positions in the final dataset. Evolutionary analyses were conducted using MEGA6 (Tamura *et al*., 2011). Numbers in brackets indicate the number of isolates with the same ST. Bar, 0.001 substitutions per site.

Figure 2: Neighbor-Joining tree of 67 *G. candidum* isolates based on the concatenated sequences. Details of the calculations are as for figure 1.. The analysis involved 67 nucleotide sequences with a total of 2254 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.*, 2011). Strain numbers in bold are from dairy origin. Strains with boxed numbers were isolated from stools. Strains with numbers in italic are of unknown origin. Bar, 0.001 substitutions per site.

Figure 3: Population structure in connection with the phylogenetic analysis of 67 strains of the 67 *G. candidum* strains. a) Clustering of the strain genotypes by *Structure*. For each of the strains, presented on the abscissa, the estimated contribution (ordinate) to its genotype from each of five hypothetical ancestral genotypes is represented as rectangles of different colours. The origins of each strain are shown above the histogram: “D” milk or dairy, “E” environment, “F” non-dairy food processing, “H” human stools, “n” not known. b) Phylogenetic analysis of the yeasts using *Splits tree*. Strain names are colour-coded as for the histogram labels in a) (D, E, F, H and n): blue, milk or diary; orange, environmental strains; green, non-dairy food processing; pink, human stools; grey, origin unknown. Groups 1 to 5

are highlighted for comparison with the results of the Neighbor-Joining phylogenetic tree of Figure 2, and of the analysis with Structure.

Figure 4: Composite-analysis of the 36 isolates which were in non-unique STs. The MLST dendrogram was generated as described in the Materials and Methods and is shown on the left. The inter-LTR profile of each strain was super-imposed on the MLST tree. Variability of the inter-MLST profile must therefore be observed within each represented ST. The ability to assimilate galactose, lactose and ribose is indicated by a black rectangle. The geographical origin and the substrate of isolation of the strains and the ST to which the strains belong are shown at the right.

Figure 1

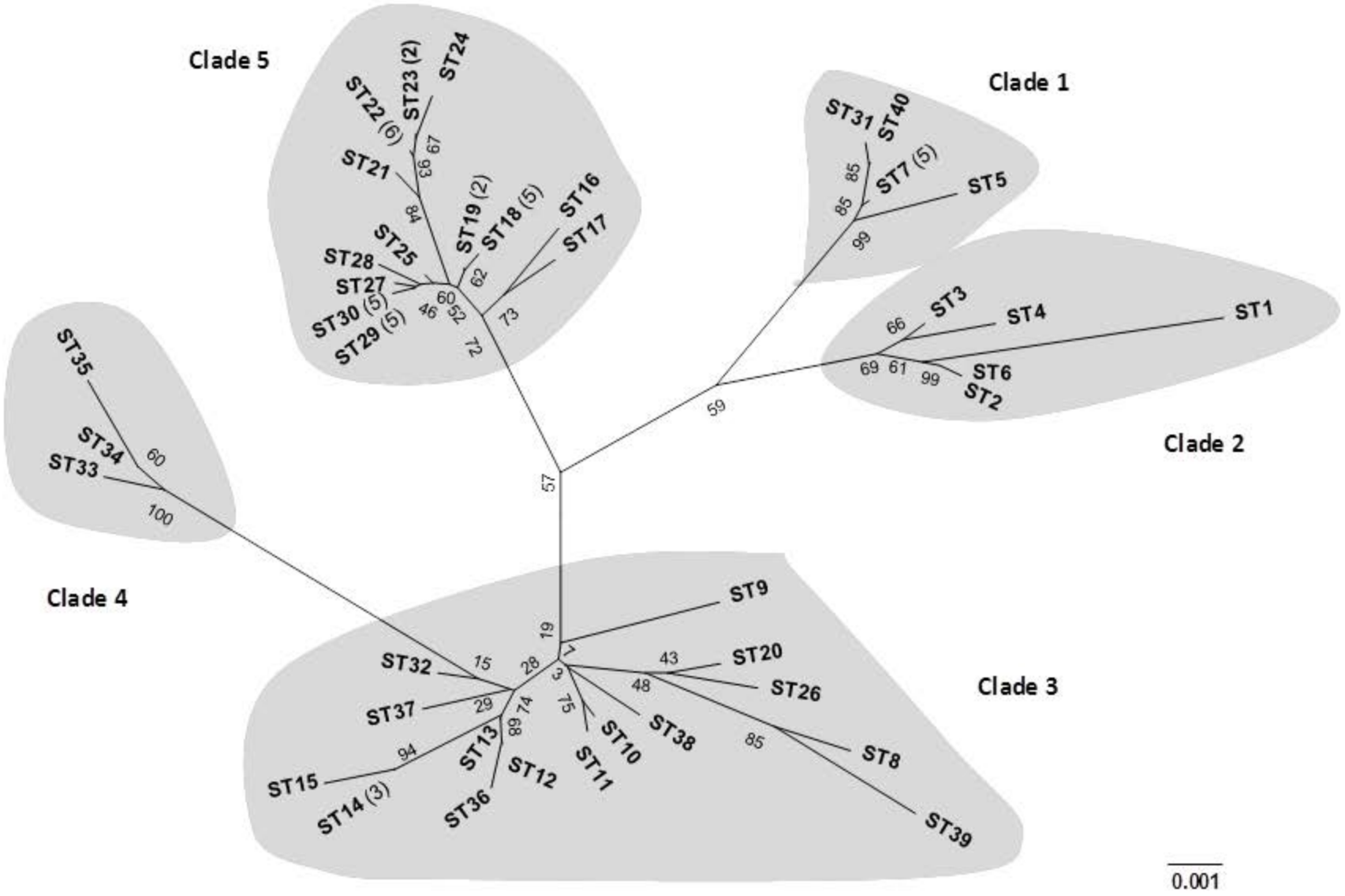

Clade 5
Clade 1
Clade 2
Clade 4
Clade 3
ST24
ST23 (2)
ST22 (6)
ST21
ST19 (2)
ST18 (5)
ST16
ST17
ST25
ST28
ST27
ST30 (5)
ST29 (5)
ST31
ST40
ST7 (5)
ST5
ST3
ST4
ST1
ST6
ST2
ST35
ST34
ST33
ST9
ST20
ST26
ST32
ST37
ST38
ST8
ST13
ST10
ST15
ST12
ST11
ST14 (3)
ST36
ST39
0.001

Figure 2

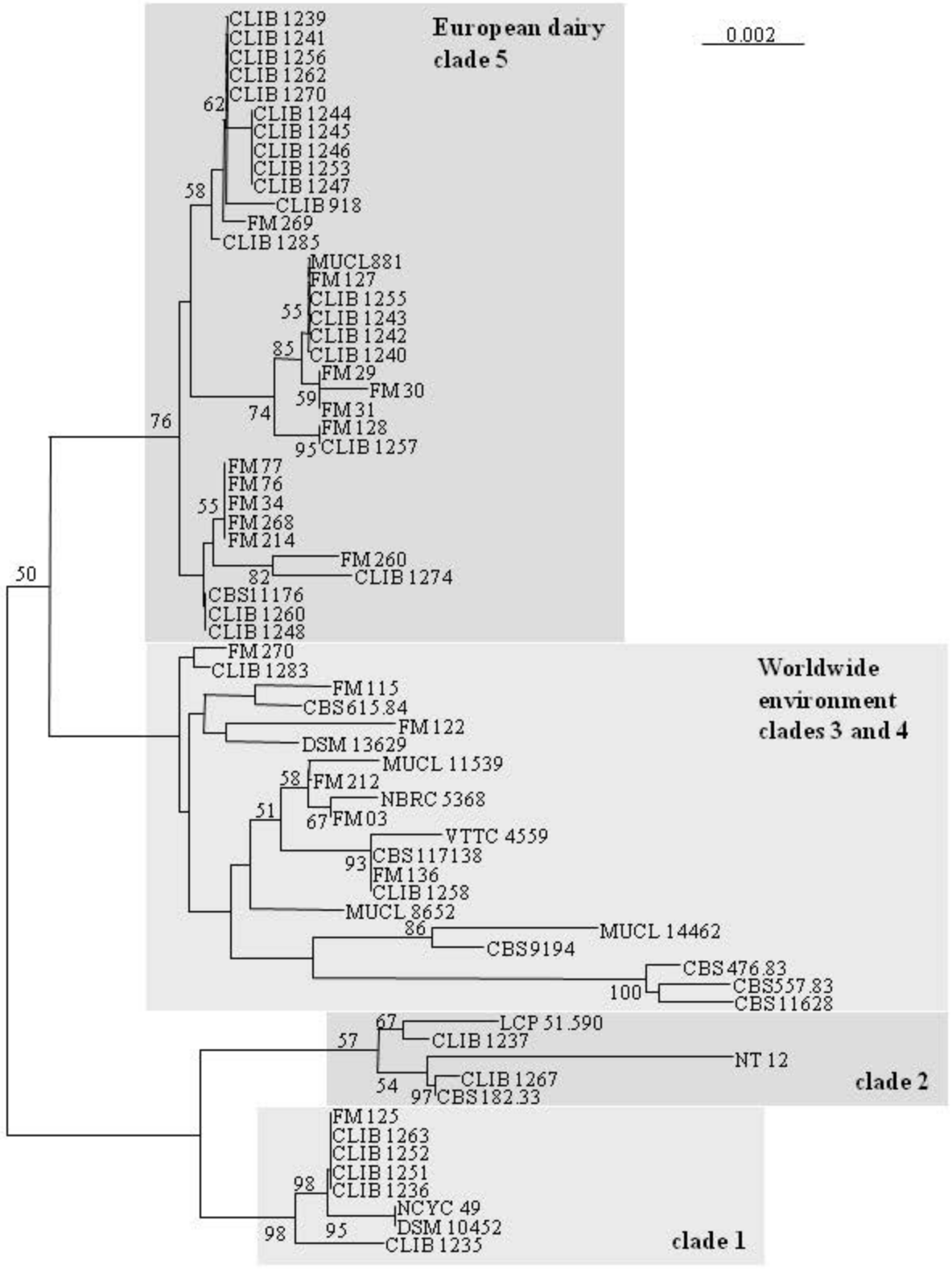

European dairy clade 5
0.002
CLIB 1239
CLIB 1241
CLIB 1256
CLIB 1262
CLIB 1270
CLIB 1244
CLIB 1245
CLIB 1246
CLIB 1253
CLIB 1247
CLIB 918
FM 269
CLIB 1285
MUCL 881
FM 127
CLIB 1255
CLIB 1243
CLIB 1242
CLIB 1240
FM 29
FM 30
FM 31
FM 128
CLIB 1257
FM 77
FM 76
FM 34
FM 268
FM 214
FM 260
CLIB 1274
CBS 11176
CLIB 1260
CLIB 1248
Worldwide environment clades 3 and 4
FM 270
CLIB 1283
FM 115
CBS 615.84
FM 122
DSM 13629
MUCL 11539
FM 212
NBRC 5368
FM 03
VTTC 4559
CBS 117138
FM 136
CLIB 1258
MUCL 8652
MUCL 14462
CBS 9194
CBS 476.83
CBS 557.83
CBS 11628
LCP 51.590
CLIB 1237
NT 12
CLIB 1267
CBS 182.33
clade 2
FM 125
CLIB 1263
CLIB 1252
CLIB 1251
CLIB 1236
NCYC 49
DSM 10452
CLIB 1235
clade 1

Figure 3

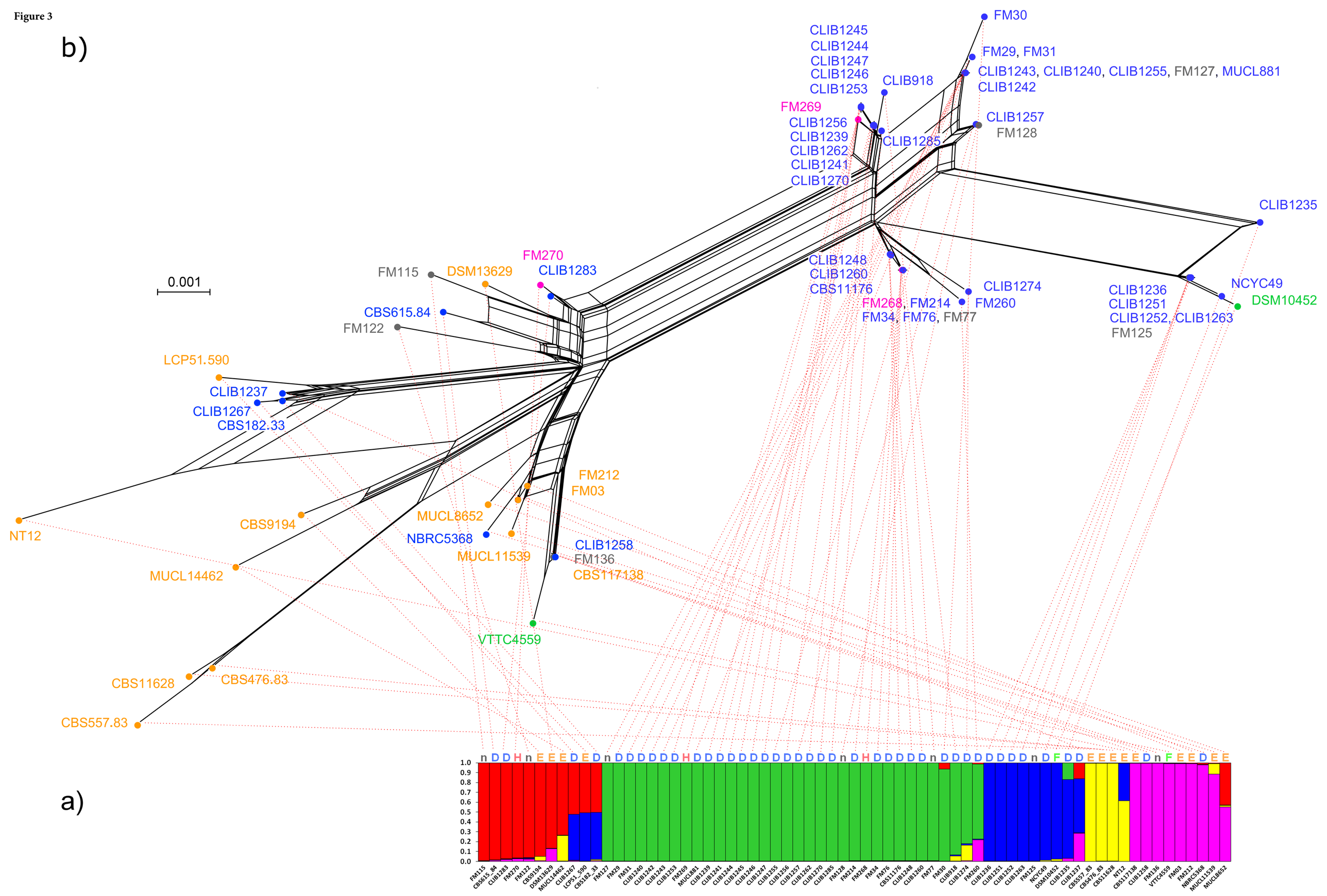
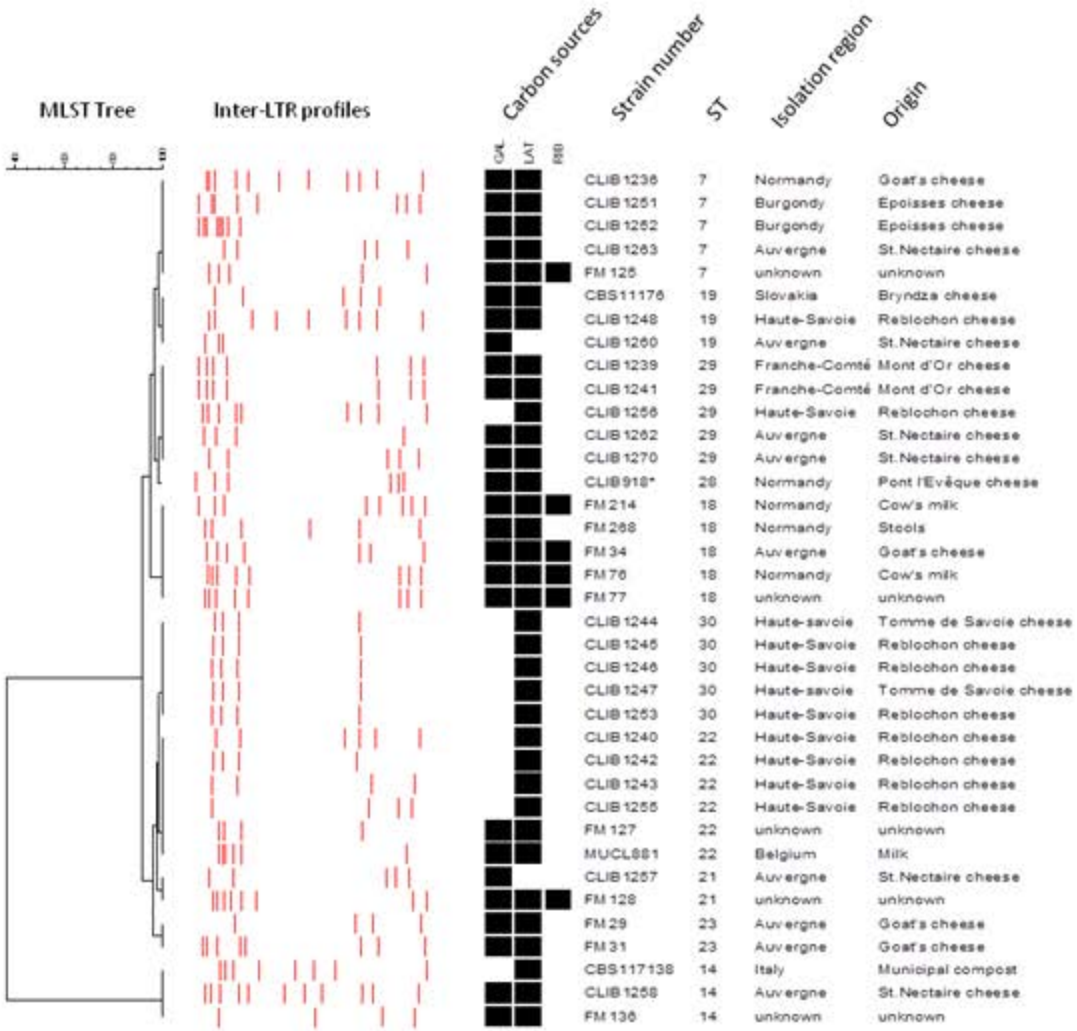

b)
0.001
FM30
CLIB1245
CLIB1244
CLIB1247
CLIB1246
CLIB1253
FM29, FM31
CLIB1243, CLIB1240, CLIB1255, FM127, MUCL881
CLIB1242
CLIB918
FM269
CLIB1256
CLIB1239
CLIB1262
CLIB1241
CLIB1270
CLIB1257
FM128
CLIB1285
CLIB1235
FM270
CLIB1283
FM115
DSM13629
CLIB1248
CLIB1260
CBS11176
CLIB1274
FM268, FM214
FM260
FM34, FM76, FM77
CLIB1236
CLIB1251
CLIB1252, CLIB1263
FM125
NCYC49
DSM10452
CBS615.84
FM122
LCP51.590
CLIB1237
CLIB1267
CBS182.33
FM212
FM03
CBS9194
MUCL8652
NBRC5368
MUCL11539
CLIB1258
FM136
CBS117138
NT12
MUCL14462
VTTC4559
CBS11628
CBS476.83
CBS557.83
a)

Figure 4


MLST Tree
Inter-LTR profiles
Carbon sources
GAL
LAT
RIB
Strain number
ST
Isolation region
Origin
CLIB1236 7 Normandy Goat's cheese
CLIB1251 7 Burgondy Epoisses cheese
CLIB1252 7 Burgondy Epoisses cheese
CLIB1263 7 Auvergne St.Nectaire cheese
FM125 7 unknown unknown
CBS11176 19 Slovakia Bryndza cheese
CLIB1248 19 Haute-Savoie Reblochon cheese
CLIB1260 19 Auvergne St.Nectaire cheese
CLIB1239 29 Franche-Comté Mont d'Or cheese
CLIB1241 29 Franche-Comté Mont d'Or cheese
CLIB1256 29 Haute-Savoie Reblochon cheese
CLIB1262 29 Auvergne St.Nectaire cheese
CLIB1270 29 Auvergne St.Nectaire cheese
CLIB918* 29 Normandy Pont l'Evêque cheese
FM214 18 Normandy Cow's milk
FM268 18 Normandy Stools
FM34 18 Auvergne Goat's cheese
FM76 18 Normandy Cow's milk
FM77 18 unknown unknown
CLIB1244 30 Haute-savoie Tomme de Savoie cheese
CLIB1245 30 Haute-Savoie Reblochon cheese
CLIB1246 30 Haute-Savoie Reblochon cheese
CLIB1247 30 Haute-savoie Tomme de Savoie cheese
CLIB1253 30 Haute-Savoie Reblochon cheese
CLIB1240 22 Haute-Savoie Reblochon cheese
CLIB1242 22 Haute-Savoie Reblochon cheese
CLIB1243 22 Haute-Savoie Reblochon cheese
CLIB1255 22 Haute-Savoie Reblochon cheese
FM127 22 unknown unknown
MUCL881 22 Belgium Milk
CLIB1257 21 Auvergne St.Nectaire cheese
FM128 21 unknown unknown
FM29 23 Auvergne Goat's cheese
FM31 23 Auvergne Goat's cheese
CBS117138 14 Italy Municipal compost
CLIB1258 14 Auvergne St.Nectaire cheese
FM136 14 unknown unknown