Title: Multilocus analysis reveales large genetic diversity in *Kluyveromyces marxianus* strains isolated from Parmigiano Reggiano and Pecorino di Farindola cheeses

***Authors*:** Giuseppe Fasoli[1], Eladio Barrio[2,3], Rosanna Tofalo[1], Giovanna Suzzi[1] and Carmela Belloch[2]

[1]Faculty of BioSciences and Technology for Food, Agriculture and Environment, University of Teramo, Via C.R. Lerici 1, 64023, Mosciano S. Angelo, Italy

[2]Instituto de Agroquímica y Tecnología de Alimentos (IATA), Consejo Superior de Investigaciones Científicas (CSIC), Ave. Agustín Escardino 7, 46980 Paterna, Valencia, Spain.

[3]Department of Genetics, Universitat de València, C/ Dr. Moliner, 56100, Burjassot, Valencia, Spain.

Corresponding author: Carmela Belloch

belloch@iata.csic.es

Institute of Agrochemistry and Food Technology, Ave. Agustín Escardino 7, 46980 Paterna (Valencia), Spain.

Phone: + 34 963 90 0022 Ext. 2321

Fax: + 34 963 63 6301

Abstract

In the present study, we have analysed the genetic diversity in *Kluyveromyces marxianus* isolated from Parmigiano Reggiano and Pecorino di Farindola cheesemaking environment. Molecular typing methods inter-RTL fingerprint and mtDNA RFLPs, as well as, sequence diversity and heterozygosity in the intergenic region between KmSSB1 and KmRIO2 genes and analysis of the mating locus were applied to 54 *K. marxianus* strains. Inter-RTL fingerprint revealed a large degree of genetic heterogeneity and clustering allowed differentiation of *K. marxianus* strains from different geographical origins. In general, inter-LTR profiles were more discriminating than RFLPs of mtDNA; however our results also indicate that both techniques could be complementary unveiling different degrees of genetic diversity. Sequence analysis of the intergenic region between KmSSB1 and KmRIO2 genes revealed 26 variable positions in which a double peak could be observed in the sequence chromatogram. Further analysis revealed the presence of heterozygous strains in the *K. marxianus* population isolated from Parmigiano Reggiano. On the other hand, all strains isolated from Pecorino di Farindola were homozygous. Two very different groups of haplotypes could be observed as well as mixtures between them. Phylogenetic reconstruction divided *K. marxianus* dairy strains into two separate populations. A few heterozygous strains in an intermediate position between them could also be observed. Mating type locus analysis revealed a large population of diploid strains containing both *MATa* and *MAT*α alleles and few haploid strains, most of them presenting the *MAT*α allele. Different scenarios explaining the presence and maintaining of homozygous and heterozygous diploid strains as well as hybrids between them in the Parmigiano Reggiano *K. marxianus* population are proposed. A

principal component analysis ~~using all dataset~~ supported the large differences between *K. marxianus* isolated from Parmigiano Reggiano and Pecorino di Farindola.

Keywords

Kluyveromyces marxianus, Italian dairy, genetic diversity, inter-RTL, mtDNA RFLPs, IGS KmSSB1-KmRIO2 sequence, mating type

Introduction

The yeast *K. marxianus* and its sister species *Kluyveromyces lactis* are characterized by their unique ability to ferment lactose (Lachance, 2011). The study of lactose metabolism, among other reasons, has highlighted *K. lactis* as a model for non-conventional yeasts (Fukuhara et al., 2006, Dujon et al., 2004; Schaffrath and Breunig, 2000). On the other hand, *K. marxianus* has been found especially suitable for biotechnological applications, based on its broad substrate spectrum, high growth rates and thermotolerance (Fonseca et al. 2008; Lane and Morrissey, 2010). Among these, production of enzymes, bioingredients and aroma compounds, removal of lactose from food or bioremediation stand out for their industrial utilization (Fonseca et al., 2008). Additionally, its GRAS (Generally Regarded As Safe) and QPS (Qualified Presumption of Safety) status favours the interest of the dairy and beverage industry on *K. marxianus* (Lane and Morrissey, 2010).

Kluyveromyces marxianus is generally described as a lactose fermentative homothallic yeast, commonly isolated from dairy environments, though it can also be found in a variety of different habitats. Very limited knowledge is available on *K. marxianus* when compared with other yeasts of biotechnological interest such as *S. cerevisiae*. Several studies have pointed out the genetic heterogeneity of *K. marxianus*. Karyotyping of *K. marxianus* strains from different isolation sources and geographical origins revealed a rich intraspecific polymorphism (Belloch et al., 1998; Fasoli et al., 2015) which was in agreement with the wide haplotype diversity observed through RFLPs of mitochondrial DNA (mtDNA) (Belloch et al., 1997; Suzzi et al., 2000). Restriction analysis of the non-transcribed spacer (NTS) region of rDNA allowed reproducible discrimination at the intraspecific level on *K. marxianus* (Perez-Brito et al., 2007). Additional RAPD PCR fingerprinting studies were found useful to

differentiate at strain level (Tofalo et al., 2014; Lopandic et al., 2006; Suzzi et al., 2000); and the variability of the insertion of the long terminal repeat (LTR) retrotransposons Tkm1 was used for assessing genetic diversity within *K. marxianus* from diverse origins (Sohier et al., 2009). Analysis of the mating type ~~ploidy in the species~~ (*MAT* locus) supports the emerging view that stable haploid and diploid *K. marxianus* strains occur in natural environments (Lane et al., 2011). The ~~mating~~ MAT locus in *K. marxianus* seems to be organized as in *K. lactis* (Butler et al., 2004), and it is constituted by MATa1 and MATa2, and MATα1, MATα2 and MATα3 alleles. Lane et al. (2011) analysed the mating type in *K. marxianus* by PCR amplification of the *MAT* locus. They concluded that strains containing bands consistent with MATa and MATα genes were diploid and strains producing a single band consistent with MATa or MATα genes were haploid. In recent years, the complete genome sequences of several *K. marxianus* strains have been obtained unveiling gene encoding enzymes for several metabolic traits (Llorente et al., 2000; Jeong et al., 2012; Suzuki et al., 2014).

In this study, we have analysed genetic diversity, heterozygosity and mating type ~~ploidy~~ in *K. marxianus* strains isolated from Italian dairy sources. For this purpose, strains isolated from natural starters in the production of Parmigiano Reggiano cheese, cow whey, fermented milk as well as some strains isolated during the production and ripening of Pecorino di Farindola cheese, were analysed by inter-LTR PCR fingerprinting and mtDNA restriction analysis. The level of heterozygosity was evaluated by sequencing the intergenic spacer region between genes KmSSB1 and KmRIO2, and mating type was determined by PCR amplification of *MATa* and *MAT*α alleles.

2. Materials and methods

2.1 Yeast strains and growth conditions

Fifty four *K. marxianus* isolates from Italian dairy origin were examined in this work. Parmigiano Reggiano strains (LM) were isolated from natural whey starter cultures belonging to 35 different dairies located in geographically contiguous areas of Northern Italy (Emilia Romagna region, Italy). Pecorino di Farindola strains (M) were isolated along the cheesemaking process, strain FM09 was isolated from fermented milk and strain VG4 was isolated from cow whey. All yeast strains were conserved as culture stocks at -80 °C in 15% (v/v) glycerol.

2.2 DNA isolation and inter-LTR PCR fingerprinting

Yeast cells were cultured overnight on GPY medium (2% w/v glucose, 0.5% w/v peptone and 0.5% w/v yeast extract) at 25 °C. DNA was extracted as reported by Querol et al. (1992) and dilutions containing about 10 ng/μl prepared using a Nanodrop spectrophotomether ND-1000 (Thermo Fisher Scientific, USA). Inter-LTR PCR was carried out as described by Sohier et al. (2009) using KM1 (5'-GTTGGTATAATATCTGG-3') and KM2 (5'-TTCTAAGGTCCCTACTAC-3') primers. PCR products were separated on 2% (w/v) agarose gel in 1 × TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8) at 60 V for 3 h, stained with RedSafe (INtRON Biotech., Korea) and visualized under UV light. DNA fragment sizes were determined using a 100 bp DNA ladder (Life Technologies, USA). Reproducibility of the technique was verified using internal controls of few strains which were included in all PCR DNA amplifications and electrophoresis. *K. marxianus* CBS 834 type of *Saccharomyces kefir* was included as reference strain of dairy origin.

2.3 Mitochondrial DNA restriction analyses (mtDNA RFLPs)

MtDNA restriction analysis was performed as reported in Belloch et al. (1997). For this purpose, a total of 300 ng of DNA was digested with restriction endonuclease *Hin*fI (Roche, Switzerland) according to the manufacturer's instructions. Restriction

fragments were separated on 2% (w/v) agarose gel in 1 × TAE buffer at 25 V for 16 h, stained with GelRed™ (Biotium, USA) and visualized under UV light. DNA fragment sizes were estimated against a DNA marker consisting of lambda phage DNA digested with *Pst*I.

2.4 Sequencing of the intergenic spacer region (IGS) between genes KmSSB1 and KmRIO2

The IGS region was amplified using primers KmSSB1 (5'-CAAATACCGCATATGAGATGTRTCTAAYTTCAT-3') and KmRIO2 (5`-GAACAAATGGTCAACCARGCYGARGA-3'). An additional internal primer (5'-CGCTTTTACATGTCTACGTTAT-3') was designed to obtain the whole sequence. PCR reactions were performed in 50 µl final volume containing 0.25 µl rTaq (5U) DNA polymerase, 4 µl dNTP mix (2.5 mM), 5 µl of 10X buffer (Takara Bio Inc., Japan), 3 µl $MgCl_2$ (1.5 mM) (Sigma, USA), 1 µl primer (50 pmol/µL) (Isogen Life Science, The Netherlands) and 7.5 µl solution containing around 100 ng of genomic DNA. Amplification was performed on a Mastercycler Pro (Eppendorf, Germany) with an initial denaturation at 95 °C for 5 min followed by 45 cycles consisting of 45 s at 94 °C, 35 s at 57, 1 min at 72 °C and a final extension of 10 min at 72 °C. Bands of PCR products were cut from the agarose gel and purified with High Pure PCR Product Purification Kit (Roche, Switzerland) following the manufacturer's instructions. Sequencing reactions were carried out using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, USA) in an Applied Biosystems (Model 310) automatic DNA sequencer. *K. marxianus* CECT 10585^T (CBS 712^T), type strain of the species, and CECT 1043 (CBS 608) were included as reference strains.

2.5 Mating type analysis

Primer design for determining the mating type in *K. marxianus* was carried out taking into account previous results by Butler et al. (2004) and Lane et al. (2011) as well as sequences deposited in GenBank. Primer pairs for amplification were SLA2 (5'-TATACATGGGATCATAAATC-3') (Lane et al., 2011) and MATa1D (5'-GGTTTGGCAGGAGTACAACTA-3') and MATα1D (5'-TGAAATCCAAAGCACCAACT-3'). PCR reactions were performed in 25 µl final volume containing 0.75 µl (5U) Expand Long Template PCR System (Roche, Switzerland), 5 µl 10X buffer 1 containing 1.75 mM $MgCl_2$, 3.5 µl dNTP mix (2.5 mM) (Takara Bio Inc., Japan), 0.5 µl each primer (50 pmol/µL) (Isogen Life Science, The Netherlands) and 2.5 µl solution containing around 200 ng of genomic DNA. Amplification was performed on a Mastercycler Pro (Eppendorf, Germany) with an initial denaturation at 94 °C for 2 min followed by 10 cycles consisting of 10 s at 94 °C, 30 s at 46 °C and 3 min at 68 °C and 35 cycles consisting of 10 s at 94 °C, 30 s at 50 °C and 3 min at 68 °C and a final extension of 10 min at 68 °C. The type strain of the closest relative *K. lactis* CECT 1961^T (CBS 683^T), was used as a negative control and, *K. marxianus* CECT 1043 (CBS 608) was used as a positive control (Lane et al., 2011). PCR products were separated on 1% (w/v) agarose gel in 1 × TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8) at 60 V for 3 h, stained with RedSafe (INtRON Biotech., Korea) and visualized under UV light. DNA fragments size was estimated against a DNA marker consisting of lambda phage DNA digested with *Pst*I.

2.7 Data analysis: dendrograms, heterozygosity analysis and phylogenetic reconstruction.

A similarity matrix was generated based on presence (1) and absence (0) of bands in the electrophoretic patterns of inter-LTRs and RFLPs of mtDNA (see DiB1). Bands ranging between 600 bp to over 2.0 Kb and 1.98 Kb to 11.5 Kb were

considered in the case of LTRs and *Hin*fI RFLPs of mtDNA, respectively. Clustering by the UPGMA (unweighted pair-group method using arithmetic averages) method (Sokal and Michener, 1958) was performed using the Jaccard Similarity Index. The multiple alignment of IGS sequences was done using MEGA 6. Sequence of strains CBS 712^T (CECT 10585^T) and CBS 608 (CECT 1043) were included in the alignment as references. To establish the degree of intraspecific diversity and heterozygosity within the sequences, the number of polymorphic sites with two or more variants was determined using the program DnaSP v5.10 (Librado and Rozas, 2009) which reconstructs haplotypes using the algorithm PHASE (Stephens et al., 2001). Haplotype reconstruction assigned a haplotype number for each unique haploid sequence. Allelic sequences from heterozygous strains were labelled with numbers 1 or 2 in the alignment. The Neighbor-Joining method (Saitou and Nei, 1987) in MEGA 6 was applied for phylogenetic tree reconstruction. Genetic distances were obtained using the absolute number of differences between the compared sequences. Bootstrap analysis was performed on 1000 replicates. DnaSP v5.10 was also used to determine nucleotide diversity (π) for ~~between~~ different *K. marxianus* populations. Principal Component Analysis was carried out using the programs in the NTSYS package version 2.2 (NTSYS Numerical Taxonomy and Multivariate Analysis System, Exeter Software, USA) merging results from presence and absence of bands in the electrophoretic patterns of inter-LTRs and RFLPs of mtDNA, variable positions in the KmSSB1-KmRIO2 IGS sequences and presence of mating type genes (see DiB 4).

3. Results

3.1 Typing of K. marxianus strains based on inter-RTL and mtDNA RFLP profiles

Inter-LTR fingerprinting revealed a high level of genetic diversity among *K. marxianus* strains. Up to 35 different electrophoretic patterns that differed in band

number, size and intensity were observed (Figure 1). Among them, pattern 7 displayed by strains LM15, LM142, LM154, LM161 and FM09 consisted of only one band of 850 bp. At the upper range in number of bands was pattern 10 with 13 bands, exhibited by strains LM116 and LM148, the top band over 2000 bp and the bottom one around 250 bp. Despite the large differences between the patterns a common band at 850 bp could be found in most patterns (23 patterns out of 35). To assess the level of diversity between the inter-LTR profiles a UPGMA dendrogram was generated. First, a presence/absence inter-LTR band matrix was compiled (see DiB1) and used to generate a similarity matrix (see DiB1). The high diversity observed between the inter-LTR profiles was supported by the low similarity value, 0.4 or lower, in 85% of the comparisons in the similarity matrix. The UPGMA dendrogram (Figure 1) revealed 26 strains grouped into 6 clusters and the remaining strains represented by unique inter-RTL profiles. Despite the large degree of genetic heterogeneity, Pecorino di Farindola cheese strains and Parmigiano Reggiano whey strains clustered in separated groups. Pattern 3 consisting of only three bands about 2000 bp, 1300 bp and 850 pb was present in 8 out of 13 strains from Pecorino di Farindola cheese. Similar patterns 4 and 34 were showed by Pecorino di Farindola cheese strains M131 and M135. The remaining three strains from the same origin exhibited pattern 5 which showed very few similarities with previous patterns 3, 4 and 34. Most of Parmigiano Reggiano strains appeared in clusters constituted by few strains or as independent branches. However, some strains clustered with an isolate from fermented milk (FM09) (pattern 7).

RFLPs of mitochondrial DNA obtained with *Hinf*I endonuclease showed 9 different patterns designated from A to I (Figure 2). All patterns except A displayed a doublet at the top of the agarose slab and shared several bands at different positions.

The mtDNA RFLP band presence/absence matrix (see DiB1) was used to generate a similarity matrix (see DiB1). The RFLP mtDNA profiles displayed a high level of similarity as demonstrated by the low percentage of values, less than 1%, showing similarity numbers lower than 0.4 (see DiB1). The UPGMA dendrogram generated from the RFLP mtDNA profiles revealed 6 clusters (Figure 2). The most common restriction pattern was I, shared by 34 strains, whereas patterns H, G and B were unique and represented by strains LM154, FM09 and LM148, respectively. UPGMA dendrogram showed no correlation between patterns/clusters of mtDNA RFLPs and the geographic origin of the strains. Moreover, strains isolated from Pecorino di Farindola shared patterns with strains isolated from Parmigiano Reggiano (patterns I, D and E).

Comparison of the number of different profiles generated by inter-LTR or restriction of mtDNA (Figures 1 and 2) indicated that the former method had a higher discriminating power than the latter. A dendrogram combining both type of patterns (Figure 3) showed that strains within inter-RTL clusters 5, 7, 9 and 10 appeared further separated by mtDNA RFLPs. UPGMA clustering at 80% similarity showed 22 *K. marxianus* strains grouped in clusters C1, C2, C3 and C4. Clusters C2 and C3 comprised all strains from Pecorino di Farindola cheese demonstrating that these *K. marxianus* strains are separated from the rest of strains. On the contrary, most strains isolated from Parmigiano Reggiano appear in isolated branches and few strains group in clusters (C1 and C4).

3.2 Heterozygosity in K. marxianus

PCR amplification of the IGS region between genes KmSSB1 and KmRIO2 produced a single fragment of about 800 bp. The multiple alignment of 56 sequences covered 815 nucleotide positions and contained 9 gaps and an homopolymeric region

containing 11 to 18 adenines between positions 647 and 664 (see DiB2). Moreover, sequence chromatograms analysis revealed several positions in which two simultaneous peaks of different colours and similar height could be observed. This would indicate the presence of polymorphic positions showing both nucleotides simultaneously and therefore the presence of heterozygous strains in the analysed *K. marxianus*. The analysis of the sequences using DnaSP (Figure 4) revealed 29 polymorphic positions (26 double peaks and 3 nucleotidic substitutions), and separation of the sequences into their alleles allowed reconstruction of 19 haplotypes (including gaps). Most polymorphic sites were found isolated along the sequence; although a region comprising 7 polymorphic adjacent sites (640 to 646) could be observed (Figure 4, see DiB2). Most polymorphic positions showing a double peak appeared in more than one sequence, and only sites 521, 564 and 595 in sequence FM09 and site 637 in sequence LM154 were unique. Haplotype 1 represents the allele present in most homozygous strains, although similar haplotypes 2, 3 and 6 are also present in few homozygous strains. The remaining haplotypes 4 and 5 as well as 7 to 19 represent alleles found exclusively in heterozygous strains. Haplotypes 13 to 19 shared 12 polymorphic positions and displayed large differences respect to haplotype 1. Haplotypes 8 and 9 appeared to be similar to haplotype 1, but for the last four polymorphic positions which were similar to haplotype 13. Haplotypes 10 to 12 seemed to be similar to haplotype 13, but for the last polymorphic positions which were similar to haplotype 1. It is worth to note that all strains isolated from Pecorino di Farindola were homozygous in the IGS region studied showing haplotypes 1, 2 or 6. Heterozygosity was observed only in strains isolated from Parmigiano Reggiano (1SC-4, LM54, LM116, LM153, LM154, LM161 and LM167) and strain FM09 isolated from

298 fermented milk. Reference strain CECT 1043 (CBS 608) was also a heterozygous
299 strain.

300 The multiple alignment comprising homozygous as well as heterozygous strains
301 (see DiB2) was manually realigned in positions 629 and 630 in the case of alleles
302 LM54-2 and LM153-2, and from 638 to 646 for alleles 1SC4-2, LM116-2 and LM167-2.
303 The corrected multiple alignment revealed 21 polymorphic sites and 9 gap positions
304 and was used to reconstruct a Neighbor-Joining (NJ) phylogenetic tree (Figure 5). The
305 NJ tree shows *K. marxianus* Italian dairy strains divided into 4 groups supported by
306 bootstrap analysis. Group Km 1 was constituted by most homozygous stains and two
307 heterozygous strains, LM54 and LM153. This group was characterised by low
308 nucleotide diversity ($\pi = 0.00068$) and reduced population substructure. On the
309 contrary, group Km 2 enclosed 10 sequences constituted by alleles of heterozygous
310 strains FM09, CBS 608, LM154 and LM161 as well as sequence of strains LM15 and
311 CBS 712^{T}. This group showed a rich population structure as well as a high nucleotide
312 diversity in this group, π = 0.00331, which was 5 times higher than in Km 1. In a
313 intermediate position between groups Km 1 and Km 2 appeared two groups of
314 sequences constituted by alleles LM 167-1, LM116-1 and 1SC4-1 (haplotypes 8 and
315 9) and alleles LM167-2, LM116-2 and 1SC4-2 (haplotypes 10, 11 and 12).

316 Concerning the origin of the strains, all *K. marxianus* isolated from Pecorino di
317 Farindola seemed to be homozygous and clustered in Km 1, although strains M41,
318 M81 and M83 appeared in a separated branch within Km 1 in a similar way as in
319 figures 1 and 2. On the other hand, Parmigiano Reggiano *K. marxianus* were a mixture
320 of homozygous and heterozygous strains. Most homozygous strains were placed in
321 Km 1, whereas most heterozygous strains clustered in Km 2. Moreover, in the case of
322 Parmigiano Reggiano strains, correspondence between phylogenetic groups and inter-

RTLs or RFLPs mtDNA clusters was found solely in case of the heterozygous strains. Strains LM116 and 1SC4 grouped together in figures 5 and 2 (pattern C), whereas strains FM09, LM154 and LM161 clustered together in figures 5 and 1 (pattern 7). It is worth to note that most strains showing haplotypes other than haplotype 1 appeared in clusters at the bottom of the combined dendrogram (Figure 3).

3.3 Analysis of the mating type loci in K. marxianus

Amplification of SLA2 - MATa1 and SLA2 – MATα1 regions produced amplified products of size 2,6 Kb and 2,8 Kb, respectively. Amplification of *MATa1* gene indicated presence of MATa genes (MATa locus) and amplification of MATα1 gene indicated presence of MATα genes (MATα locus). Amplification of CECT 1043 (CBS 608) produced both bands confirming that this strain is diploid. No amplification product was obtained from *K. lactis* CECT 1961^T (CBS 683^T) used as negative control. Most of the Italian dairy *K. marxianus* strains ~~were diploid~~ showed both bands consistent with presence of mating types *MATa* and *MATα* (see DiB3) and evidencing the diploid state of these strains. The only exceptions were 13 strains, 9 ~~strains~~ containing the *MATα* locus and 4 ~~strains~~ containing the *MATa* locus (Figure 5) consistent with the haploid state. The percentage of haploid strains in Parmigiano Reggiano or Pecorino di Farindola *K. marxianus* populations was the same, about 23%. However, differences in the distribution of haploid strains in both *K. marxianus* populations, Parmigiano Regginao and Pedorino di Farindola, were found. Most haploid strains containing the MATα locus were ~~restricted to~~ found within the Parmigiano Reggiano isolates (9 *MATα* and 1 *MATa* strains), whereas only 3 haploid strains, 2 of them containing the *MATa* locus (M41 and M81) and strain M131 containing the *MATα* locus appeared within the Pecorino di Farindola strains. Moreover, haploid strains did not appear randomly distributed in the phylogenetic tree.

Group Km 1 included all haploid strains containing the *MAT*α locus, whereas group Km 2 contained only two haploid strains containing the *MATa* locus, LM15 and CBS 712^{T}. Location of haploid strains containing the *MATa* locus in the inter-RTL and mtDNA composite dendrogram (Figure 3) revealed that they appeared solely in groups C3 and C4.

Principal component analysis (Figure 6) was applied to ~~the entire dataset to~~ further analyse relationships between *K. marxianus* populations. The first component (50.5% variability) separated heterozygous and homozygous *K. marxianus*. This separation is supported by variable positions in KmSSB1-KmRIO2 IGS sequence but also by the typing techniques. Components 2 and 3, accumulate 8% and 6,4% of the variability, respectively. These two components separate the heterozygous strains which constitute a loose group except for LM154 and LM161. These strains together with LM15 cluster in C3 of the composite dendrogram as well as in Figure 5. Strains LM20 and LM148 appear in a separated position between the heterozygous and homozygous groups. These two strains appear close to the heterozygous strains on the basis of mtDNA RFLPs although LM20 is separated from them by inter-RTL fingerprint. The large population of homozygous *K. marxianus* is additionally separated by component 3 mostly supported by inter-RTL patterns. The separate position of heterozygous strains LM153 and LM54 and few homozygous strains is supported by variable positions in KmSSB1-KmRIO2 IGS sequence but also by mtDNA RFLPs.

4. Discussion

In this study, we have used inter-RTL fingerprinting, RFLPs of mtDNA, heterozygosity in the sequence of the internal gene spacer between KmSSB1 and KmRIOS2 and examination of the mating type locus to analyse the genetic diversity among 54 *K. marxianus* strains of Italian dairy origin.

373 Inter-RTL fingerprinting, developed by Sohier et al. (2009), is based on the
374 variability of the insertion of the LTR retrotransposon TKm1. In their study, Sohier et al.
375 (2009) revealed a rich genetic diversity within *K. marxianus* from French dairy
376 products. The percentage of unique patterns found by these authors (70%) and the
377 ones found in our study (65%) are very similar. Moreover, results from both studies
378 highlight the divergence between patterns in number and size of bands even between
379 strains from the same type of cheese. However, Sohier et al. (2009) could not
380 compare between *K. marxianus* strains from different geographical origins or different
381 types of cheeses; although they concluded that the high specificity of the patterns
382 could be used to associate strains to varieties of cheese. Our clustering results (Figure
383 1) show that *K. marxianus* strains from Pecorino di Farindola cluster separated from *K.*
384 *marxianus* from Parmigiano Reggiano, therefore confirming the conclusion of Sohier et
385 al. (2009). Moreover, based on their inter-RTL results for *D. hansenii* and previous
386 studies (Romano et al., 1996) Sohier et al. (2009) concluded that inter-LTR profiles are
387 more discriminating than RFLPs of mtDNA. Early studies had demonstrated the
388 discriminating power of RFLPs of mtDNA between *K. marxianus* strains from diverse
389 isolation sources (Belloch et al., 1997), however these authors could not compare
390 between strains from the same isolation source or same geographical origin. Our
391 results for *K. marxianus* strains of Italian dairy origin indicate that restriction of mtDNA
392 generates few different patterns which translate into higher values in the similarity
393 matrix than in the case of inter-LTR PCR. The lower discriminating level of RFLPs of
394 mtDNA did not allow separation of *K. marxianus* strains isolated from different cheeses
395 or geographical origins. However, our results indicate that both techniques could be
396 complementary unveiling different levels of genetic diversity as strains that clustered
397 together based on inter-LTR profiles could be separated by RFLPs of mtDNA.

Evidence of Ploidy in *K. marxianus* strains was tested by analysis of the mating loci type by PCR amplification of the *MAT* locus, as in previous studies (Lane et al., 2011). The mating MAT locus in *K. marxianus* seems to be organized as in *K. lactis* (Butler et al., 2004). The MAT locus in the latter is constituted by MATa1 and MATa2, and MATα1, MATα2 and MATα3 alleles. These authors concluded that strains producing PCR bands consistent with the presence of MATa and MATα genes were diploid and strains producing a single band consistent with MATa or MATα genes were haploid. Presence of MATa and MATα genes in *K. marxianus*, as proposed by Lane et al. (2011), consists of PCR amplification of a 10 Kb (MATα) and 6 Kb (MATa) regions spanning SLA2 to LAA1 genes which contained the whole mating loci. However, amplification of large DNA regions in *K. marxianus* dairy strains using the primers designed by Lane et al. (2011) was technically complicated, requiring very good quality DNA and very long PCR reactions. Moreover, amplification of both bands was not always possible or reproducible even using equivalent PCR reagents and the same amplification conditions used by these authors. In our study, a reduction in size of the expected PCR band by amplification of the regions spanning SLA2 and MATa1 or MATα1 alleles minimized technical difficulties and simplified the testing of the large number of strains investigated.

The majority of *K. marxianus* strains examined by Lane et al. (2011) were haploid containing the *MATa* locus. In contrast, most of the Italian dairy *K. marxianus* in our study were diploid. The restricted number of dairy strains tested by Lane et al. (2011) and their diverse geographical origin prevents comparisons between both studies. However, the five dairy strains investigated by these authors were diploid or haploid containing the *MATa* locus. In our study, the percentage of haploid strains in Parmigiano Reggiano or Pecorino di Farindola *K. marxianus* populations appeared to

423 be the same. However, the ratio of haploid strains containing the MATα locus vs.
424 MATa locus was 8/1 in Parmigiano Reggiano strains and 1/2 in Pecorino di Farindola
425 strains.

426 Butler et al. (2004) and Lane et al. (2011) results indicate that mating locus
427 organization in *K. marxianus* is most likely as in *K. lactis*, lacking the HO gene and the
428 Ho site in MATα1. On the other hand, the presence of silent HMRa and HMLα loci was
429 confirmed in *K. marxianus* on the basis of CBS 608 whole genome sequence (Lane et
430 al., 2011). This was in agreement with early studies indicating that mating-type
431 switching in *K. lactis* occurs only in spores and in much lower frequency than in *S.*
432 *cerevisiae* (Herman & Roman, 1966).

433 The population of *K. marxianus* strains of Italian dairy origin analysed in our
434 study is mainly constituted by homozygous strains in the IGS region between genes
435 KmSSB1 and KmRIO2. Occurrence of a high number of homozygous strains is
436 favoured by haploselfing or autodiploidization as occurs in homothallic *S. cerevisiae*
437 (Magwene, 2014); however, this mechanism seems very unlikely in *K. marxianus*. The
438 most likely mechanism favouring the large percentage of diploid homozygous strains
439 observed in Km 1 may be automixis, or crossing between spores produced in the
440 same meiotic division of a diploid *K. marxianus* parental (Diezmann and Dietrich,
441 2009; Magwene et al., 2011). Moreover, automixis would also support the presence of
442 the reduced number of heterozygous strains, LM54 and LM153, found within Km 1.
443 The few polymorphic positions shown in haplotypes 2, 4, 5 and 7 could be explained
444 by few point mutations occurring during mitotic growth of homozygous ~~diploid~~ isolates
445 (Mortimer et al., 1994; Tsai et al., 2008). Another explanation for the presence of
446 closely related heterozygous alleles in Km 1 would be outcrossing of isolates within
447 closely related sub-populations as indicated by the low nucleotide diversity within Km

1. Crossing of two genetically identical or closely related compatible haploid parentals followed by asexual growth of the diploid progeny would support the large number of homozygous strains and also the low number of heterozygous isolates (Diezmann and Dietrich, 2009). The absence of haploid *MATa* strains in the largest homozygous population (haplotype 1) would make improbable the crossing of haploid parentals to produce homozygous diploid strains; however the presence of *MAT*α haploid strains in the dominant haplotype and *MATa* haploid strains in closely related haplotypes (haplotype 6) would also explain the low percentage of heterozygous alleles in Km 1. Successive automixing and outcrossing between closely related isolates would maintain the lower degree of heterozygosity and favour assimilation of heterozygous isolates in the homozygous *K. marxianus* population (Mortimer et al., 1994; Magwene, 2014).

In contrast, sequences constituting Km 2 are mostly alleles of heterozygous strains showing a large degree of nucleotide diversity, which is in agreement with the rich subpopulation structure found within this group in the phylogenetic tree. These results, as well as the absence of homozygous strains, can be explained by a high frequency of outcrossing in these strains. High levels of heterozygosity have been correlated with high levels of outcrossing in human-associated *S. cerevisiae* (Muller and McCusker, 2009; Magwene et al., 2011). Moreover, heterozygosity due to outcrossing increases the adaptive potential of natural populations as their progeny may exhibit hybrid vigour (Marsit and Dequin, 2015). Studies in human-associated *S. cerevisiae* revealed that crosses between genetically unrelated isolates increased the ability to grow at high temperatures among other traits (McCusker et al., 1994) and, in wine fermentations, hybrids present advantages over non-hybrids showing increased tolerance than the parents to various stresses during fermentation (Belloch et al.,

2008). Under these premises, phenotypic diversity of heterozygous strains would promote colonization of new niches and adaptation to new environments (Magwene, 2014). In our study, none of the alleles from the highly heterozygous strains constituting group Km 2 appears in any isolate pertaining to Km 1 indicating separation between both populations. Moreover, there are large differences in the patterns of sequence polymorphisms between populations Km 1 and Km 2. This would indicate that new strains appear regularly in the cheesemaking environment and outcrossing followed by asexual growth would maintain the high proportion of heterozygosity, as occurs in clinical *S. cerevisiae* (Muller and McCusker, 2009). Differences in the pattern of polymorphisms as well as the high nucleotide diversity in Km 2 population would be supported by the different genetic background of yeasts which find their way into the cheesemaking factory probably due to different milk batches or brine preparations.

Coexistence in the same environment of well established homozygous *K. marxianus* strains and newly introduced heterozygous *K. marxianus* strains would favour crossing between both groups of strains. In our study, two groups of sequences corresponding to both alleles of heterozygous strains 1SC4, LM116 and LM167 appear in an intermediate position between Km 1 and Km 2 isolates. The phylogenetic position of these heterozygous isolates, as well as their pattern of sequence polymorphisms with respect to haplotypes 1 and 13, indicate crossing between strains in Km 1 and Km 2 groups. Outcrossing between different yeast populations has been demonstrated repeatedly in *S. cerevisiae* as well as in other yeast species (Cogliati et al., 2012; Marsit and Dequin, 2015). Moreover, human association seems to have a large influence in outcrossing between yeast populations. In a recent study using comparative genomics, Cromie et al. (2013) observed two human-associated populations containing European and Asian alleles and situated between these two

498 groups in a principal coordinate analysis. Numerous studies on *S. cerevisiae* and other
499 yeast species have reported heterozygosity on microsatellites as well as coding genes
500 (Muller and McCusker, 2009; Diezmann and Dietrich, 2009; Cogliati et al., 2012);
501 however, there is also abundant heterozygosity in non-coding regions. Magwene
502 (2014) emphasized its potential effect on regulatory networks due to allele specific
503 gene expression or in protein-DNA interactions (Gagneur et al. 2009; Zheng et al.,
504 2010), as well as its role in cell to cell heterogeneity within clonal populations (Levy et
505 al., 2012).

506 In conclusion, the data generated in our study have highlighted large
507 differences between *K. marxianus* from Parmigiano Reggiano and Pecorino di
508 Farindola origin. Parmigiano Reggiano strains appear distributed in different groups,
509 showing a very rich population structure composed by a well established abundant
510 population of homozygous ~~diploid~~ strains and a small but highly heterozygous
511 population of probably newly acquired strains. Moreover, mating between Parmigiano
512 Reggiano strains must have happened as supported by the presence of hybrids
513 ~~phylogenetically situated between both populations~~. These mating events have
514 introduced new allele combinations in *K. marxianus* strains involved in the Parmigiano
515 Reggiano cheesemaking process. On the other hand, Pecorino di Farindola cheese
516 strains are genetically very similar and appear in closely related groups. These strains,
517 isolated from raw milk and several cheesemaking stages, showed differences in their
518 RAPD-PCR patterns as well as different ratios between free aminoacids and lactic
519 acid residual content in whey after fermentation (Tofalo et al., 2014). However, their
520 genetic similarity using inter-RTL fingerprinting, RFLPs of mtDNA as well as the low
521 level of polymorphisms in the IGS between KmSSB1 and KmRIO2 is in agreement
522 with a homogeneous population in which almost no new allele combinations are

introduced due to cheesemaking conditions or by inoculation of ~~diploid~~ homozygous strains well established in the cheesemaking factory.

Acknowledgements

CB thanks grants Consolider Ingenio Fun-C-Food, CSD2007-00063 and INIA RM2007-00010 for financial support.

References

Belloch, C., Barrio, E., Uruburu, F., Garcia, M.D., Querol, A., 1997. Characterisation of four species of the genus *Kluyveromyces* by mitochondrial DNA restriction analysis. Systematic and Applied Microbiology 20, 397-408.

Belloch, C., Barrio, E., García, M.M., Querol, A., 1998. Inter- and intraspecific chromosome pattern variation in the yeast genus *Kluyveromyces*. Yeast 14, 1341-1354.

Belloch, C., Orlic, S., Barrio, E., Querol, A., 2008. Fermentative stress adaptation of hybrids within the *Saccharomyces* sensu stricto complex. International Journal of Food Microbiology 122, 188-195.

Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C., Wolfe, K., 2004. Evolution of the MAT locus and its Ho endonuclease in yeast species. Proceedings of the National Academy of Science 101, 1632-1637.

Cogliati, M., Barchiesi, F., Spreghini, E., Tortorano, A.M., 2012. Heterozygosis and pathogenicity of *Cryptococcus neoformans* AD-hybrid isolates. Mycopathologia 173, 347-357.

Cromie, G.A., Hyma, K.E., Ludlow, C.L., Garmendia-Torres, C., Gilbert, T.L., May, P., Huang, A.A., Dudley, A.M., Fay, J.C., 2013. Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. Genes, Genomes, Genetics 3, 2163-2171.

548 Diezmann, S., Dietrich, F.S., 2009. *Saccharomyces cerevisiae*: population divergence
549 and resistance to oxidative stress in clinical, domesticated and wild isolates. PLoS
550 One 4, e5317.
551 Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De
552 Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M.,
553 Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne,
554 E., Bleykasten, C., Boisramé, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar,
555 A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F.,
556 Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire,
557 M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-
558 Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A.,
559 Swennen, D., Tekaia, F., Wésolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-
560 Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B.,
561 Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., Souciet, J.L., 2004.
562 Genome evolution in yeasts. Nature 430, 35-44.
563 Fasoli, G., Tofalo, R., Lanciotti, R., Schirone, M., Patrignani, F., Perpetuini, G., Grazia,
564 L., Corsetti, A., Suzzi, G., 2015. Chromosome arrangement, differentiation of growth
565 kinetics and volatile molecule profiles in *Kluyveromyces marxianus* strains from Italian
566 cheeses. International Journal of Food Microbiology 214, 151-158.
567 Fonseca, G.G., Heinzle, E., Wittmann, C., Gombert, A.K., 2008. The yeast
568 *Kluyveromyces marxianus* and its biotechnological potential. Applied Microbiology and
569 Biotechnology 79, 339-354.
570 Fukuhara, H., 2006. *Kluyveromyces lactis* – a retrospective. FEMS Yeast Research 6,
571 323-324.

572 Gagneur, J., Sinha, H., Perocchi, F., Bourgon, R., Steinmetz, H.W., 2009. Genome-
573 wide allele- and strand-specific expression profiling. Molecular Systems Biology 5,
574 274.
575 Herman, A., Roman, H.L., 1966. Allele specific determinants of homothallism in
576 *Saccharomyces lactis*. Genetics 53, 727-740.
577 Jeong, H., Lee, D.H., Kim, S.H., Kim, H.J., Lee, K., Song, J.Y., Kim, B.K., Sung, B.H.,
578 Park, J.C., Sohn, J.H., Koo, H.M., Kim, J.F., 2012. Genome sequence of the
579 thermotolerant yeasts *Kluyveromyces marxianus* var. *marxianus* KCTC 17555.
580 Eukaryotic Cell 11, 1584-1585.
581 Lachance, M.A., 2011. *Kluyveromyces* van der Walt (1971). In: Kurtzman C.P., Fell,
582 J.W., Boekhout, T. (Eds.), The Yeasts, a Taxonomic Study, vol. 2. Elsevier B.V., New
583 York, pp. 471-481.
584 Lane, M.M., Burke, N., Karreman, R., Wolfe, K.H., O'Byrne, C.P., Morrissey, J.P.,
585 2011. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*.
586 Antonie van Leeuwenhoek 100, 507-519.
587 Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: a yeast emerging from
588 its sister's shadow. Fungal Biology Reviews 24, 17–26.
589 Levy, S.F., Ziv, N., Siegal, M.L., 2012. Bet hedging in yeast by heterogeneous, age-
590 correlated expression of a stress protectant. PLoS Biology 10, e1001325.
591 Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of
592 DNA polymorphism data. Bioinformatics 25, 1451-1452.
593 Llorente, B., Malpertuy, A., Blandin, G., Artiguenave, F., Wincker, P., Dujon, B., 2000.
594 Genomic exploration of the hemiascomycetous yeasts: *Kluyveromyces marxianus* var.
595 *marxianus*. FEBS Letters, 487, 71–75.

596 Lopandic, K., Zelger, S., Bánszky, L.K., Eliskases-Lechner, F., Prillinger, H., 2006.
597 Identification of yeasts associated with milk products using traditional and molecular
598 techniques. Food Microbiology 23, 341-350.
599 Magwene, P.M., 2014. Revisiting Mortimer's genome renewal hypothesis:
600 heterozygosity, homothallism, and the potential for adaptation in yeast. In: Landry,
601 C.R., Aubin-Horth, N. (Eds.) Ecological Genomics: Ecology and Evolution of Genes
602 and Genomes. Springer, The Netherlands, pp. 37-48.
603 Magwene, P.M., Kayikçi, Ö., Granek, J.A., Reininga, J.M., Scholl, Z., Murray, D., 2011.
604 Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution
605 in *Saccharomyces cerevisiae*. Proceedings of the National Academy of Science 108,
606 1987-1992.
607 Marsit, S., Dequin, S., 2015. Diversity and adaptative evolution of *Saccharomyces*
608 wine yeast: a review. FEMS Yeast Research 15, fov067
609 McCusker, J.H., Clemons, K.V., Stevens, D.A., Davis, R.W., 1994. Genetic
610 characterization of pathogenic *Saccharomyces cerevisiae* isolates. Genetics 136,
611 1261-1269.
612 Mortimer, R.K., Romano, P., Suzzi, G., Polsinelli, M., 1994. Genome renewal: a new
613 phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae*
614 derived from natural fermentation of grape musts. Yeasts 10, 1543-1552.
615 Muller, L.A.H., McCusker, J.H., 2009. Microsatellite analysis of genetic diversity
616 among clinical and non-clinical *Saccharomyces cerevisiae* isolates suggest
617 heterozygote advantage in clinical environments. Molecular Ecology 18, 2779-2786.
618 Pérez-Brito, D., Tapia-Tussell, R., Quijano-Ramayo, A., Larqué-Saavedra, A., Lappe,
619 P., 2007. Molecular characterization of *Kluyveromyces marxianus* strains isolated from
620 *Agave fourcroydes* (Lem.) in Yucatan, Mexico. Molecular Biotechnology 37, 181–186.

621 Querol, A., Barrio, E., Ramon, D., 1992. A comparative study of different methods of
622 yeast strains characterization. Systematic and Applied Microbiology 15, 439-446.
623 Romano, A., Casaregola, S., Torre, P., Gaillardin, C., 1996. Use of RAPD and
624 mitochondrial DNA RFLP for typing of *Candida zeylanoides* and *Debaryomyces*
625 *hansenii* yeast strains isolated from cheese. Systematic and Applied Microbiology 19,
626 255-264.
627 Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for
628 reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406-425.
629 Schaffrath, R., Breunig, K.D., 2000. Genetics and molecular physiology of the yeast
630 *Kluyveromyces lactis*. Fungal Genetics and Biology 30, 173-190
631 Sohier, D., Le Dizes, A.S., Thuault, D., Neuveglise, C., Coton, E., Casaregola, S.,
632 2009. Important genetic diversity revealed by inter-LTR PCR fingerprinting of
633 *Kluyveromyces marxianus* and *Debaryomyces hansenii* strains from French traditional
634 cheeses. Dairy Science Technology 89, 569-581.
635 Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic
636 relationships. University of Kansas Science Bulletin 28, 1409-1438.
637 Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype
638 reconstruction from population data. American Journal of Human Genetics 68, 978-
639 989.
640 Suzuki, T., Hoshino, T., Matsushika, A., 2014. Draft genome sequence of
641 *Kluyveromyces marxianus* strain DMB1, isolated from sugarcane bagasse
642 hydrolysate. Genome Announcement 2, e00733-14.
643 Suzzi, G., Lombardi, A., Lanorte, M.T., Caruso, M., Andrighetto, C., Gardini, F., 2000.
644 Phenotypic and genotypic diversity of yeasts isolated from water-buffalo Mozzarella
645 cheese. Journal of Applied Microbiology 88, 177-123.

646 Tsai, I.J., Bensasson, D., Burt, A., Koufopanou, V., 2008. Population genomics of the
647 wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. Proceedings of the
648 National Academy of Science 105, 4957-4962.
649 Tofalo, R., Fasoli, G., Schirone, M., Perpetuini, G., Pepe, A., Corsetti, A., Suzzi, G.,
650 2014. The predominance, biodiversity and biotechnological properties of
651 *Kluyveromyces marxianus* in the production of Pecorino di farindola cheese.
652 International Journal of Food Microbiology 187, 41-49.
653 Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M., Snyder, M., 2010. Genetic
654 analysis of variation intranscription factor binding in yeast. Nature 464, 1187-1191.
655

656 **Figure captions**

657 **Figure 1.-** UPGMA dendrogram based on inter-LTR patterns clustering *K. marxianus*
658 strains of Italian dairy origin. Strain name, pattern number and profile are given. The
659 vertical bar indicates the 80% similarity cutt-off.

660 **Figure 2.-** UPGMA dendrogram based on mtDNA RFLPs patterns clustering *K.*
661 *marxianus* strains of Italian dairy origin. Strain name, pattern number and profile are
662 given.

663 **Figure 3.-** Composite dendrogram based on inter-LTR and mtDNA RFLPs patterns
664 clustering *K. marxianus* strains of Italian dairy origin. Strain name and pattern number
665 are given.

666 **Figure 4.-** Haplotypes reconstructed for each IGS (KmSSB1 and KmRIO2) sequence
667 type observed in the *K. marxianus* Italian dairy population. Sequence polymorphisms
668 are indicated by their positions in the sequence. Gaps were included in the analysis.

669 **Figure 5.-** Neighbour-joining tree based on total number of differences between pairs
670 of IGS (KmSSB1 and KmRIO2) sequences. Percentage of bootstrap values are given
671 on the nodes. Strains labelled “a” or “α" are haploid strains with mating type locus
672 MATa or MATα respectively. Strains not labelled represent diploid strains.

673 **Figure 6.-** Principal component analysis 3D plot showing relatedness among *K.*
674 *marxianus* strains and their distribution in populations. Strains from different dairy
675 sources are represented by the following symbols: ● Parmigiano Reggiano, ● Pecorino
676 di Farindola, ●other sources.

677

Table 1. List of strains used in this study.

Strain number	**Isolation source**	**Culture collection**
LM3, LM6, LM9, LM14, LM15, LM17, LM20, LM28, LM30, LM32, LM36, LM38, LM42, LM44, LM47, LM48, LM50, LM54, LM72, LM92, LM96, LM114, LM116, LM127, LM133, LM136, LM139, LM141, LM142, LM148, LM153, LM154, LM161, LM167, LM169, LM174, 6M2, 1SC4, K326	Parmigiano Reggiano natural whey starter culture	*Department of Agricultural and Food Sciences, University of Bologna*
M12, M14, M38, M41, M48, M68, M81, M83, M123, M131, M135, M166, M169	Pecorino di Farindola cheese	*Faculty of BioSciences and Technology for Food, Agriculture, and Environment, University of Teramo*
VG4	cow milk whey	
FM09	fermented milk	

Figure 1

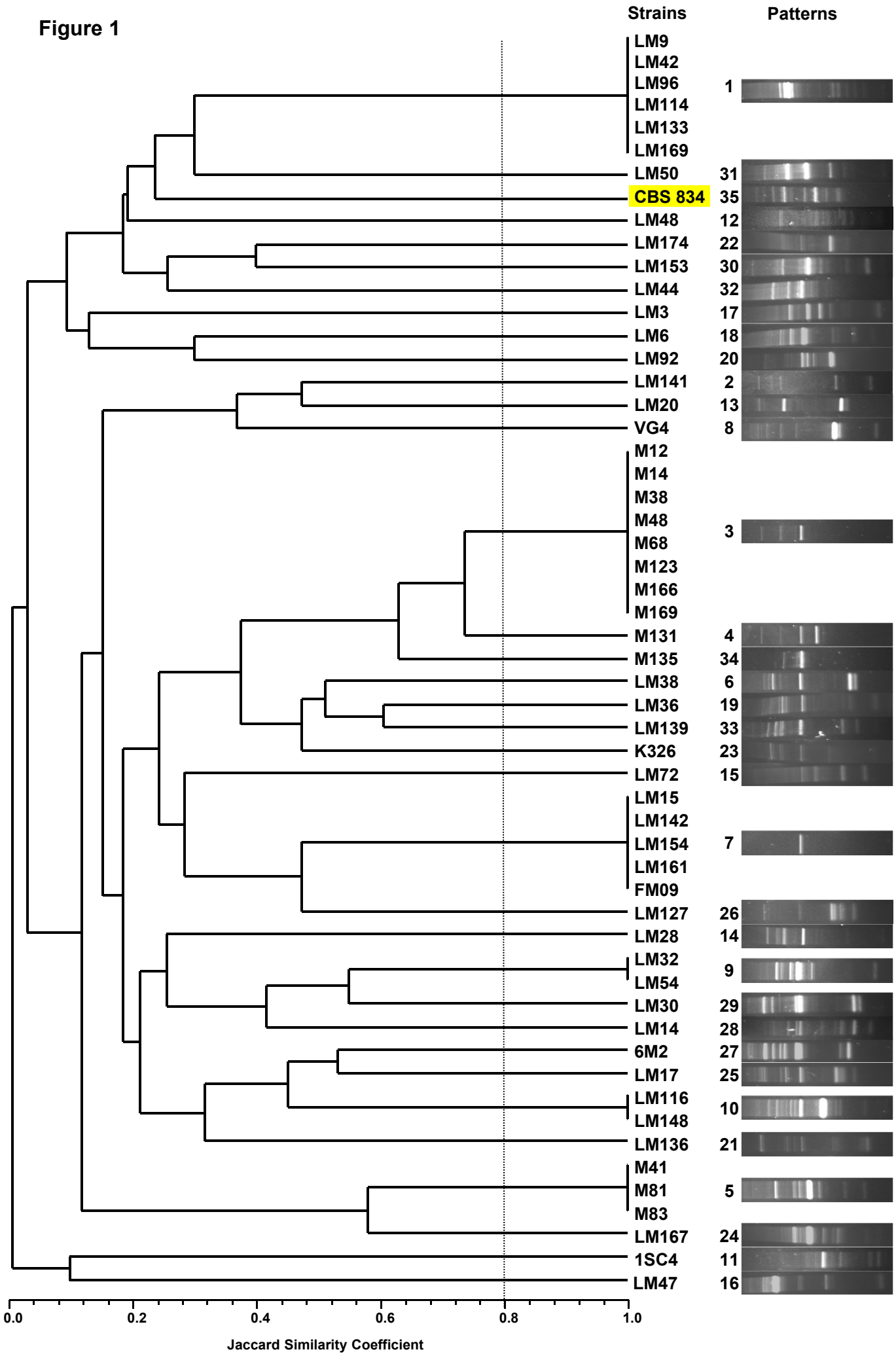

Strains
Patterns
LM9
LM42
LM96
LM114
LM133
LM169
1
LM50
31
CBS 834
35
LM48
12
LM174
22
LM153
30
LM44
32
LM3
17
LM6
18
LM92
20
LM141
2
LM20
13
VG4
8
M12
M14
M38
M48
M68
M123
M166
M169
3
M131
4
M135
34
LM38
6
LM36
19
LM139
33
K326
23
LM72
15
LM15
LM142
LM154
LM161
FM09
7
LM127
26
LM28
14
LM32
LM54
9
LM30
29
LM14
28
6M2
27
LM17
25
LM116
LM148
10
LM136
21
M41
M81
M83
5
LM167
24
1SC4
11
LM47
16
0.0
0.2
0.4
0.6
0.8
1.0
Jaccard Similarity Coefficient

Figure 2

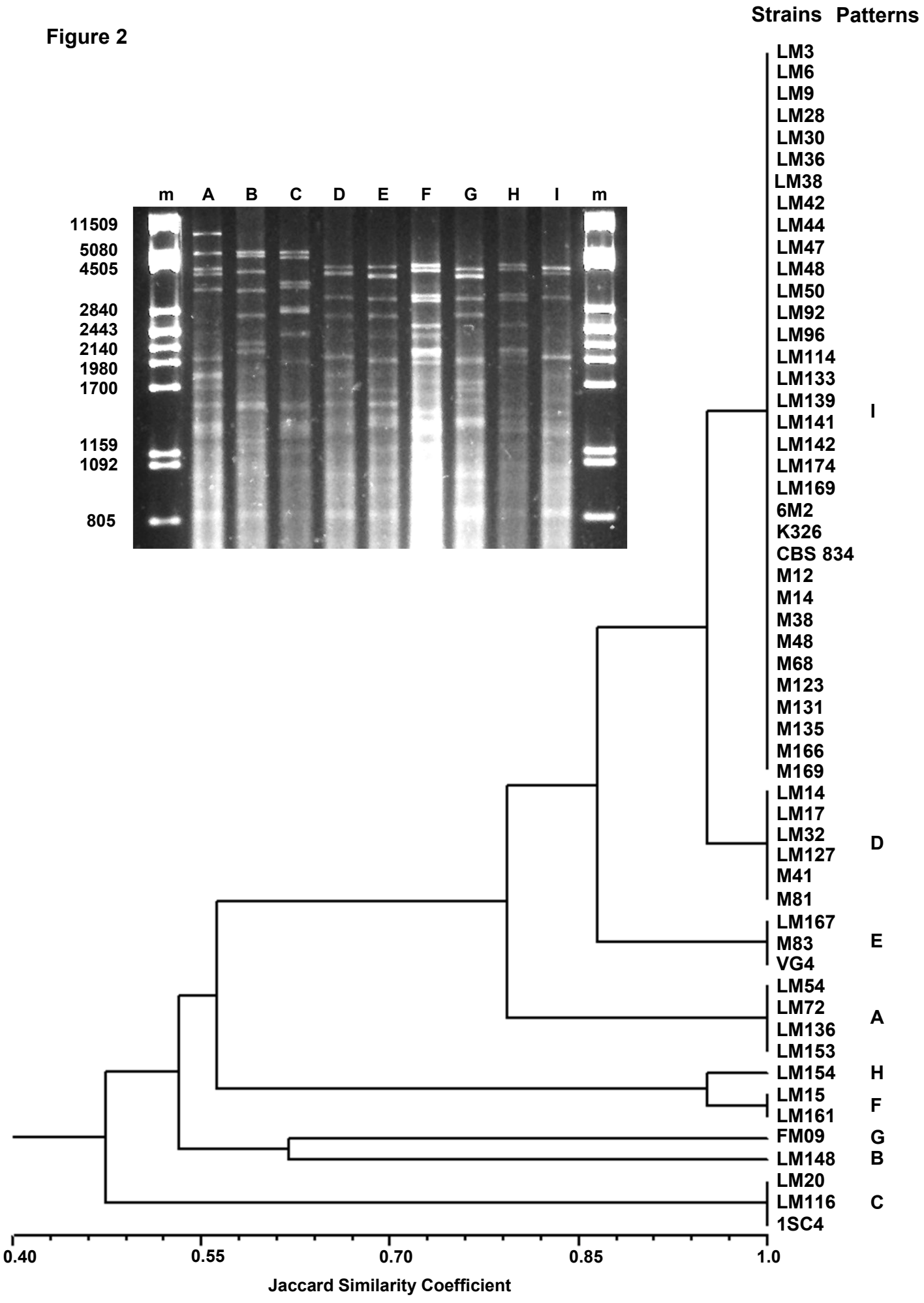

Strains Patterns
m A B C D E F G H I m
11509
5080
4505
2840
2443
2140
1980
1700
1159
1092
805
LM3
LM6
LM9
LM28
LM30
LM36
LM38
LM42
LM44
LM47
LM48
LM50
LM92
LM96
LM114
LM133
LM139
LM141
LM142
LM174
LM169
6M2
K326
CBS 834
M12
M14
M38
M48
M68
M123
M131
M135
M166
M169
LM14
LM17
LM32
LM127
M41
M81
LM167
M83
VG4
LM54
LM72
LM136
LM153
LM154
LM15
LM161
FM09
LM148
LM20
LM116
1SC4
I
D
E
A
H
F
G
B
C
0.40
0.55
0.70
0.85
1.0
Jaccard Similarity Coefficient

Figure 3

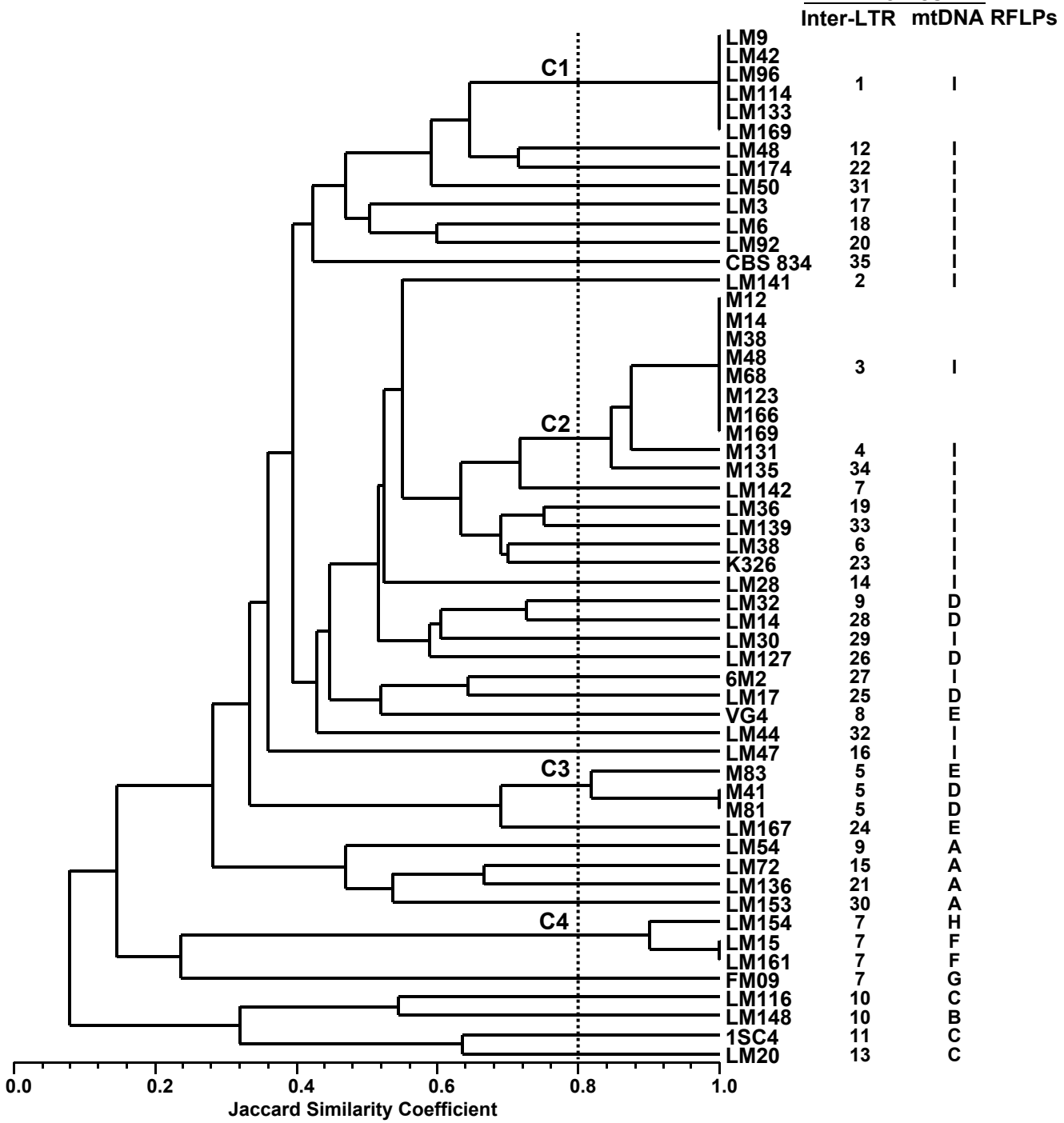

Profiles
Inter-LTR mtDNA RFLPs
C1
C2
C3
C4
LM9 1 I
LM42
LM96
LM114
LM133
LM169
LM48 12 I
LM174 22 I
LM50 31 I
LM3 17 I
LM6 18 I
LM92 20 I
CBS 834 35 I
LM141 2 I
M12
M14
M38
M48
M68 3 I
M123
M166
M169
M131 4 I
M135 34 I
LM142 7 I
LM36 19 I
LM139 33 I
LM38 6 I
K326 23 I
LM28 14 I
LM32 9 D
LM14 28 D
LM30 29 I
LM127 26 D
6M2 27 I
LM17 25 D
VG4 8 E
LM44 32 I
LM47 16 I
M83 5 E
M41 5 D
M81 5 D
LM167 24 E
LM54 9 A
LM72 15 A
LM136 21 A
LM153 30 A
LM154 7 H
LM15 7 F
LM161 7 F
FM09 7 G
LM116 10 C
LM148 10 B
1SC4 11 C
LM20 13 C
0.0
0.2
0.4
0.6
0.8
1.0
Jaccard Similarity Coefficient

Figure 4

Haplotype	96	160	341	487	494	495	516	520	521	564	580	587	595	598	629*	630	637	639	641	642	643	644	645	646	647	658*	659*	660*	661*	662*	663*	664*	668	725	729	757	762	Strains
1	C	G	A	C	T	G	T	G	G	C	T	T	A	C	A	T	-	A	C	T	A	C	A	C	A	A	A	A	A	A	-	-	T	G	G	T	C	Homozygous strains
2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	LM54-1, LM36, LM72, LM136, M48
3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	C	.	.	.	.	LM47
4	.	A[1]	.	.	.	.	.	.	.	.	.	.	.	.	-	A[1]	T	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	LM54-2
5	.	A[1]	.	.	.	.	.	.	.	.	.	.	.	.	-	A[1]	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	-	C	.	.	.	.	LM153-2
6	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	C	.	.	.	.	LM 148, M41, M81, M83, VG4
7	.	.	.	.	.	T[1]	.	.	.	.	.	.	.	.	-	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	-	C	.	.	.	.	LM153-1
8	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	C	A[1]	C[1]	A[1]	T[1]	LM167-1
9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	.	A[1]	C[1]	A[1]	T[1]	1SC4-1, LM116-1
10	T[1]	A[1]	.	T[1]	G[1]	T[1]	C[1]	A[1]	.	.	.	C[1]	.	A[1]	.	.	.	T[1]	A[1]	C[1]	T[1]	A[1]	T[1]	A[1]	C[1]	.	.	.	-	-	-	-	.	.	.	.	.	1SC4-2
11	T[1]	A[1]	.	T[1]	G[1]	T	C[1]	A[1]	.	.	.	C[1]	.	A[1]	.	.	.	T[1]	A[1]	C[1]	T[1]	A[1]	T[1]	A[1]	C[1]	.	.	.	-	-	-	-	C	.	.	.	.	LM167-2
12	T[1]	.	.	T[1]	G[1]	T[1]	C[1]	A[1]	.	.	.	C[1]	.	A[1]	.	.	.	T[1]	A[1]	C[1]	T[1]	A[1]	T[1]	A[1]	C[1]	.	.	.	-	-	-	-	.	.	.	.	.	LM116-2
13	T	.	.	T	G	T	C	A	.	.	.	C	.	A[2]	.	.	T	.	.	.	.	T	.	.	.	.	.	-	-	-	-	-	.	A	C	A	T	CBS 712[T], FM09-1
14	T	A[1]	.	T	G	T	C	A	T[1]	T[1]	.	C	G[1]	.	.	.	T	.	.	.	.	T	.	.	.	.	.	-	-	-	-	-	.	A	C	A	T	FM09-2
15	T	.	C	T	G	T	C	A	.	.	.	C	.	.	.	.	C	.	.	.	.	T	.	.	.	-	-	-	-	-	-	-	.	A	C	A	T	CBS 608-1
16	T	A[1]	C	T	G	T	C	A	.	.	.	C	.	.	.	.	C	.	.	.	.	T	.	.	.	-	-	-	-	-	-	-	.	A	C	A	T	CBS 608-2
17	T	A[1]	C	T	G	T	C	A	.	.	G	C	.	A[1]	.	.	T	.	.	.	.	T	.	.	.	.	-	-	-	-	-	-	.	A	C	A	T	LM154-2
18	T	A[1]	C	T	G	T	C	A	.	.	G	C	.	.	.	.	C	.	.	.	.	T	.	.	.	.	-	-	-	-	-	-	.	A	.	A	T	LM161-2
19	T	.	C	T	G	T	C	A	.	.	G	C	.	.	.	.	C[3]	.	.	.	.	T	.	.	.	.	-	-	-	-	-	-	.	A	C[4]	A	T	LM161-1, LM154-1, LM15

* gap positions included in the analysis

[1] positions/strains where a double peak was observed

[2] double peak was observed on FM09-1

[3] double peak was observed on LM154-1

[4] double peak was observed on LM161-1

Figure 5

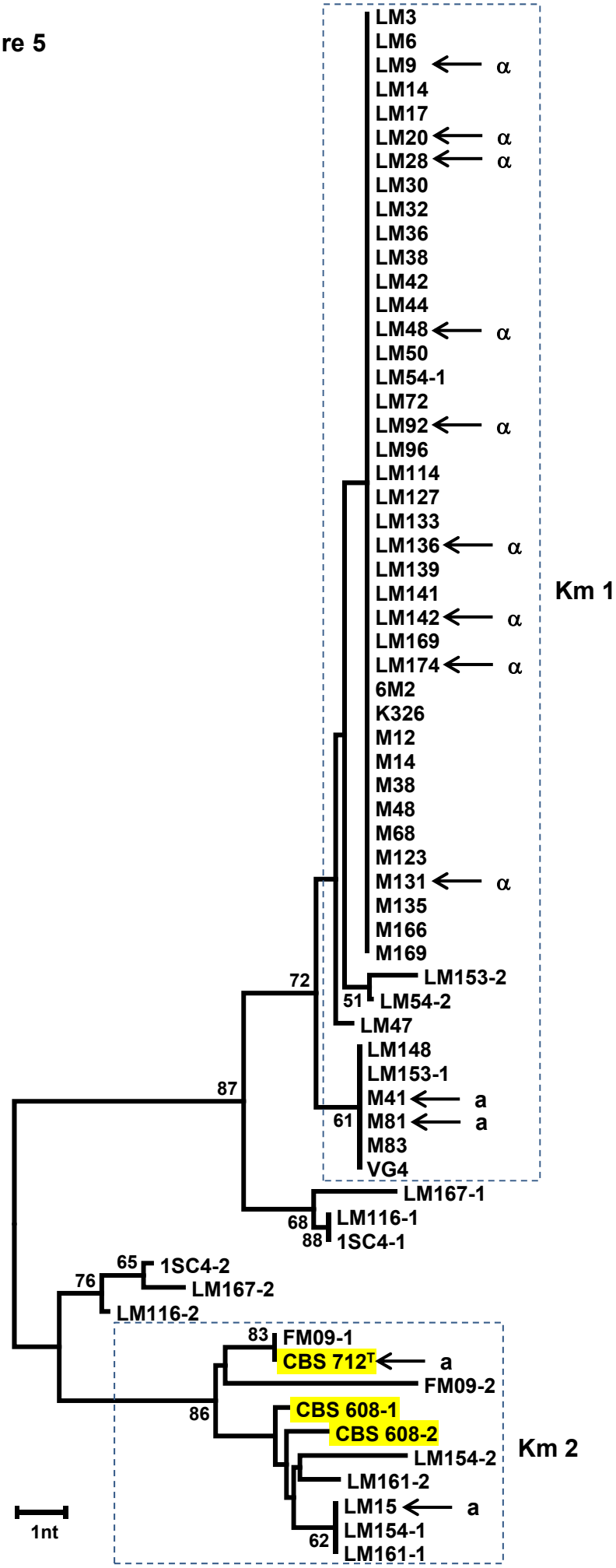
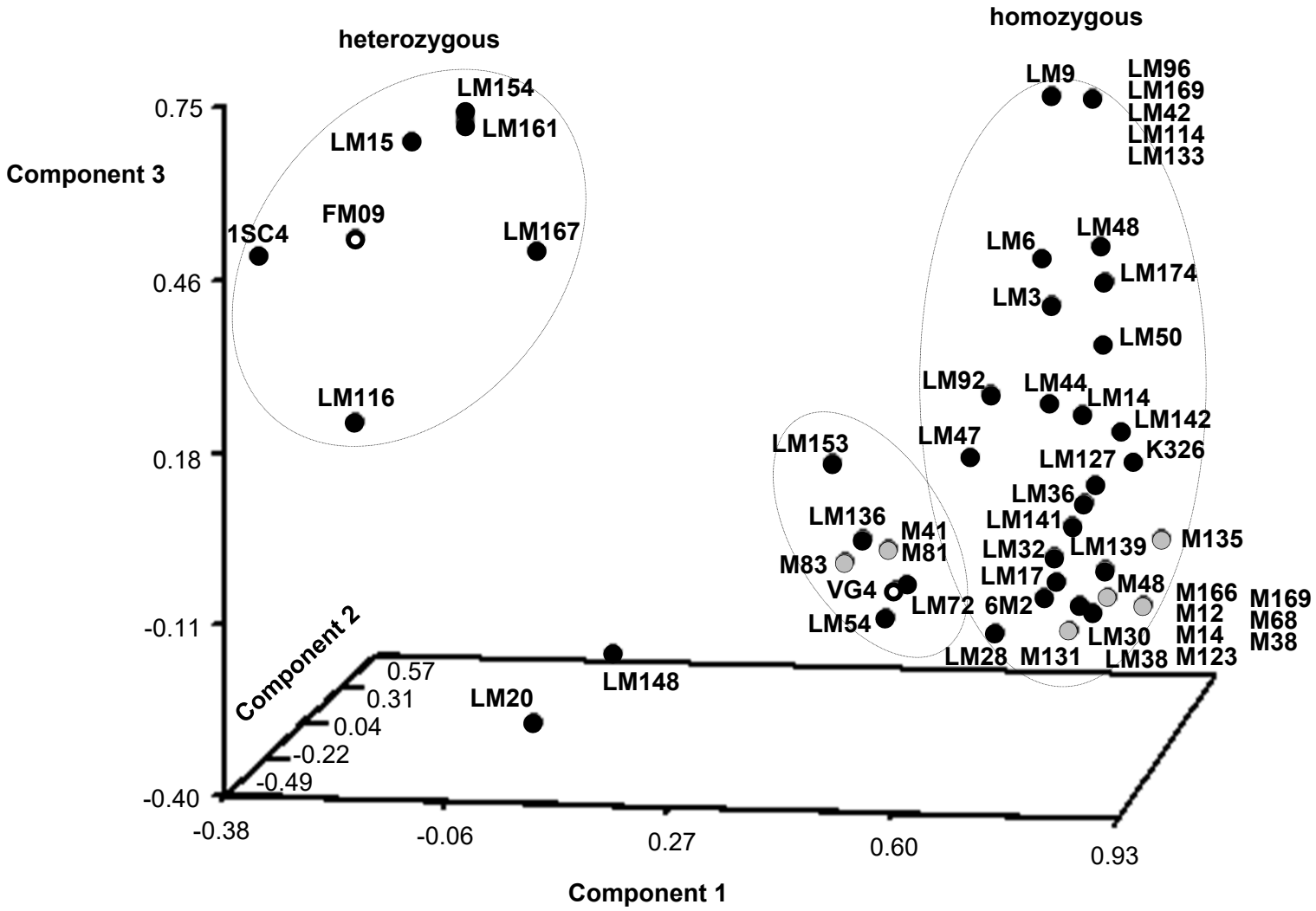

LM3
LM6
LM9 α
LM14
LM17
LM20 α
LM28 α
LM30
LM32
LM36
LM38
LM42
LM44
LM48 α
LM50
LM54-1
LM72
LM92 α
LM96
LM114
LM127
LM133
LM136 α
LM139
LM141
LM142 α
LM169
LM174 α
6M2
K326
M12
M14
M38
M48
M68
M123
M131 α
M135
M166
M169
Km 1
72
LM153-2
51
LM54-2
LM47
LM148
LM153-1
87
M41 a
61
M81 a
M83
VG4
LM167-1
68
LM116-1
88
1SC4-1
65
1SC4-2
76
LM167-2
LM116-2
83
FM09-1
CBS 712T a
FM09-2
86
CBS 608-1
CBS 608-2
LM154-2
LM161-2
LM15 a
62
LM154-1
LM161-1
Km 2
1nt

Figure 6


heterozygous
homozygous
LM154
LM161
LM15
FM09
1SC4
LM167
LM116
LM9
LM96
LM169
LM42
LM114
LM133
LM6
LM48
LM174
LM3
LM50
LM92
LM44
LM14
LM142
LM47
LM153
LM127
K326
LM36
LM136
LM141
M41
M81
M83
LM32
LM139
M135
LM17
M48
VG4
LM72
6M2
M166
M169
LM54
M12
M68
LM30
M14
M38
LM28
M131
LM38
M123
LM148
LM20
0.75
0.46
0.18
-0.11
-0.40
Component 3
Component 2
0.57
0.31
0.04
-0.22
-0.49
-0.38
-0.06
0.27
0.60
0.93
Component 1