Fungal Biology 123 (2019) 687–697

Contents lists available at ScienceDirect

Fungal Biology

journal homepage: www.elsevier.com/locate/funbio

Genomic analysis and lactose transporter expression in *Kluyveromyces marxianus* CCT 7735

Lílian Cardoso de Paiva [a], Raphael Hermano Santos Diniz [a, b], Pedro Marcus Pereira Vidigal [c], Tiago Antônio de Oliveira Mendes [d], Mateus Ferreira Santana [a], María-Esperanza Cerdán [e], María-Isabel González-Siso [e], Wendel Batista da Silveira [a, *]

[a] Laboratory of Microbial Physiology, Department of Microbiology, Universidade Federal de Viçosa, Viçosa, MG, Brazil
[b] Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, Campus Ouro Preto, CODACIB, Ouro Preto, MG, Brazil
[c] Center for Analysis of Biomolecules, Center for Biological and Health Sciences, Universidade Federal de Viçosa, Viçosa, MG, Brazil
[d] Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, MG, Brazil
[e] Universidade da Coruña, Exprela Research Group, Facultade de Ciencias and CICA (Centro de Investigacións Científicas Avanzadas), A Coruña, Spain

ARTICLE INFO

Article history:
Received 18 December 2018
Received in revised form
21 May 2019
Accepted 3 June 2019
Available online 15 June 2019

Corresponding Editor: Julian Rutherford

Keywords:
Fermentation
Genome
Hypoxia
Lac12
Metabolism
Real-time PCR

ABSTRACT

Kluyveromyces marxianus CCT 7735 has been used to produce ethanol, aromatic compounds, enzymes and heterologous proteins besides assimilates lactose as carbon source. Its genome has 10.7 Mb and encodes 4787 genes distributed in 8 nuclear chromosomes and one mitochondrial. Contrary to *Kluyveromyces lactis*, which has a unique *LAC12* gene (encodes lactose permease), *K. marxianus* possesses four. The presence of degenerated copies and Solo-LTRs related to retrotransposon TKM close to the *LAC12* genes in *K. marxianus* indicates ectopic recombinations. The Lac12 permeases of *K. marxianus* and *K. lactis* are conserved, however the conservation is higher between the copy of the left side of the chromosome three and the unique copy of *K. lactis*, indicating that this copy is the ancestor. The expression of the four *LAC12* genes occurred in aerobiosis and hypoxia. Notably, the high lactose consumption in hypoxia seems to be related to the high expression of the *LAC12* genes.

1. Introduction

Kluyveromyces marxianus is a homothallic, hemiascomycete yeast belonging to Saccharomycetaceae family, phylogenetically related to *Saccharomyces cerevisiae*. It is found in various habitats, like fruits, sewage, soil and dairy products, which results in a high metabolic diversity and a substantial polymorphism in their chromosomes.

K. marxianus is able to metabolize different carbohydrates such as galacturonic acid, inulin, xylose and arabinose (Kurtzman and Fell, 1998). It stands out for its ability to utilize lactose as the sole carbon source, a characteristic shared with other species of the genus *Kluyveromyces*, such as *Kluyveromyces lactis*, but found in only 2 % of the known species of yeast. Nevertheless, it is well-known that the efficiency of lactose utilization varies among the strains (Grba et al., 2002; Lane et al., 2011). Besides the ability to assimilate various sugars, *K. marxianus* has other characteristics desirable for biotechnological applications, such as respir-ofermentative metabolism, thermotolerance and high growth rate and secretory capacity (Fonseca et al., 2008; Groeneveld et al., 2009). Due to these traits, its biotechnological potential has been exploited for the production of ethanol, recombinant proteins, aromatic compounds and enzymes such as polygalacturonase, β-glucosidase and β-galactosidase (Lane and Morrissey, 2010).

The dairy yeast *K. lactis* has been used as a model for the study of the lactose–galactose metabolism (Fukuhara, 2006). In *K. lactis*, the permease (Lac12p) transports lactose from the medium into the cytosol, where this disaccharide is hydrolyzed by β-galactosidase

* Corresponding author. Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária-BIOAGRO, Universidade Federal de Viçosa, Av. P.H. Rolfs s/n, Campus Universitário, 36570-900, Viçosa, MG, Brazil. Fax: +55 31 3899 2573.
E-mail address: wendel.silveira@ufv.br (W.B. Silveira).

https://doi.org/10.1016/j.funbio.2019.06.004
1878-6146/

(Lac4p) into its monosaccharides glucose and galactose. The *LAC12* and *LAC4* genes are specific for lactose metabolism and they are co-regulated by the same bidirectional promoter with four Gal4p-binding sites upstream of the genes (Godecke et al., 1991; Rubio-Teixeira, 2005). On the other hand, information about the regulation of lactose–galactose regulon in *K. marxianus* is still scarce. Recently, analysis of some genomes of *K. marxianus* highlighted the fact that all studied strains possess four *LAC12* gene copies that are highly homologous (Lertwattanasakul et al., 2015; Varela et al., 2017). All copies are in subtelomeric chromosome regions. Notably, subtelomeric regions are subject to a high rate of duplication and recombination (Barton et al., 2008), however the nature of the *LAC12* duplication and recombination are unknown for instance.

The Lac12p is an integral membrane protein, member of the major facilitator superfamily and works via a proton-symport mechanism. It is saturable at high substrate concentrations and is also able to transport galactose with low affinity (Riley et al., 1987). It has been suggested that changes in 13 amino acids of Lac12p in *K. marxianus* strains are responsible for its great or poor capacity to transport lactose. Changes were found in transmembrane domains and cytoplasmic region alike, however, the distribution of these residues did not show any particular pattern regarding the potential protein structure (Varela et al., 2017).

The strain *K. marxianus* CCT 7735, previously designated as UFV-3, was isolated from a dairy industry because of its outstanding β-galactosidase production, and it showed potential for producing ethanol from medium content lactose, such as: whey permeate, synthetic medium or complex medium (Silveira et al., 2005; Diniz et al. 2012, 2014; Ferreira et al., 2015). The authors also showed that the specific lactose consumption is higher in hypoxia, indicating that oxygen concentration plays an important role in the lactose uptake. In addition, these authors observed that the specific ethanol production is also higher under hypoxic conditions, indicating a connection between transport and fermentative flow.

Compared to *K. lactis*, *K. marxianus* CCT 7735 displays a higher fermentative capacity, which is related to high expression of the *RAG6* gene (pyruvate decarboxylase), *LAC4* gene (β-galactosidase) and genes of the Leloir pathway under hypoxia and high lactose concentrations (Diniz et al., 2012). Nevertheless, it is still not clear whether the oxygen concentration regulates the expression of the lactose transport in yeasts of the *Kluyveromyces* genera.

The genome of *K. marxianus* CCT 7735 was sequenced and the transcriptome in ethanol stress conditions was recently published (Diniz et al., 2017). The draft genome was assembled in nine scaffolds representatives of *K. marxianus* CCT 7735 chromosomes and the predicted proteins presented high similarity to proteins of *K. marxianus* DMKU3-1042 and *K. lactis* CBS 2359 strains. To better understand the genetics features of *K. marxianus*, a comparative analysis of its genome (published in 2014, Silveira et al., 2014) with *K. lactis* and other genomic sequences of *K. marxianus* was performed. The main objective of this study was to evaluate the expression patterns of lactose permeases genes in both aerobiosis and hypoxia and analyze their polymorphisms.

2. Materials and methods

2.1. Strains, media and growth conditions

To evaluate the expression of lactose permease genes, the strains *K. lactis* CBS 2359 (=NRRL Y-1140) and *K. marxianus* CCT 7735 (=UFV-3) were used. *K. marxianus* CCT 7735 were deposited in the Tropical Culture Collection Tonsello André Foundation, Campinas, São Paulo, Brazil. The yeasts were maintained frozen in 20 % glycerol at −80 °C. The pre-culture was obtained by adding 1 mL of the culture stored to a volume of 20 mL of YNB (Yeast Nitrogen Base) medium without amino acids (Sigma-Aldrich, St. Louis, USA) and 2 % (w/v) of lactose in Erlenmeyer flasks of 125 mL. These pre-cultures were incubated in a rotary shaker (New Brunswick Scientific 25D, NJ, USA) at 37 °C with agitation of 200 rpm for 16 h. Posteriorly, the cultures were centrifuged at 3000 *g* for 5 min at 4 °C. The cell pellet was washed twice with peptone 0.01 % (w/v), before being added to a new culture medium.

To determine the kinetic parameters of growth, the cells were cultured in 500 mL (aerobiosis) or 250 mL (hypoxia) Erlenmeyer flasks filled with 150 mL of the culture. The cultures were incubated at 30 °C in a rotary shaker at 200 rpm (aerobiosis) or 50 rpm (hypoxia). In hypoxic condition, the culture medium was purged with nitrogen gas (99.9 %, v/v purity) prior to cell inoculation. The cultures were initiated at an optical density 600 nm (OD_{600}) between 0.08 and 0.1. Growth of the cells was measured every hour in a spectrophotometer (Multiskan™ Go, Thermo Scientific, MA, USA), until the stationary phase. The growth rate was determined by linear regression of the plot of ln OD_{600} unit *versus* time (h) in the exponential growth phase.

2.2. Analytical methods

Samples were collected at regular intervals (hourly) in order to analyze both lactose consumption and ethanol production. The culture supernatants were collected by centrifugation at 10 000 *g* for 10 min. Supernatant was passed through a 0.13 mm syringe filter (0.22 μm, Merck Milipore Co., Germany) prior to the injection into the HPLC (CTO-20A, Shimadzu, Japan). Ethanol and lactose concentrations were analyzed by HPLC by injecting 10 μL into an ion exchange column (Aminex HPX-87H 300 × 7.8 mm, 9 μm, Bio-Rad, Munich, Germany) with 5 mmol H_2SO_4 as eluent at 0.6 mL/min, 60 °C and detection by refraction index (RID-20A, Shimadzu, Japan). Quantification of all substances was performed by calibration and verification with external standards.

2.3. Determination of the dry weight and fermentative parameters

Dry weight of biomass was determined in samples of 5 mL each by centrifugation at 10 000 *g* for 5 min at 4 °C. The wet pellets were washed twice and then dried for 24 h at 105 °C and then weighed. A suspension of 1 mL was used to dilute the cells: 2×10^{-1}, 3×10^{-1}, 4×10^{-1}, 5×10^{-1} and 6×10^{-1}. The dry mass (gdw) was calculated by a linear regression of the plot of the OD_{600} *versus* dry weight.

The determination of the ethanol yield ($Y_{P/X}$ [g g^{-1}]) was obtained by angular coefficient (slope) of the plot cell dry weight (gdw) *versus* ethanol concentration. Cell biomass yield ($Y_{X/S}$ [g g^{-1}]) was determined by slope from a linear regression of the plot gdw *versus* lactose concentrations (g L^{-1}). All linear regression was performed at exponential growth phase. The specific ethanol production rate (q_P, [g g^{-1} h^{-1}]) was calculated as follows: $q_P = \mu * Y_{P/X}$, where μ = specific growth rate (h^{-1}). Whereas the specific lactose consumption rate (q_S, [g g^{-1} h^{-1}]) was determined as follows: $q_S = \mu / Y_{X/S}$.

2.4. Gene prediction and functional annotation

The genes of *K. marxianus* CCT 7735 (Silveira et al., 2014) were predicted using Augustus version 2.5.5 (Stanke and Waack, 2003; Stanke et al., 2006). Gene prediction was performed by using the trained dataset of *K. lactis*, which is available in the Augustus default installation. All proteins encoded by predicted genes were functionally annotated through similarity searches using BLAST version 2.6.0 (Altschul et al., 1990) and HMMER version v3.1b2 (Finn et al., 2011). In these searches, protein sequences were aligned to the yeast reference proteomes available at UniProt

database. Additionally, the predicted proteins were manually classified in fourteen functional classes selected at Gene Ontology (GO) consortium database (www.geneontology.org/), by analyzing the results of similarity searches.

2.5. Comparative genomic analysis

In order to gain insights regarding the genetics features of *K. marxianus* CCT 7735, its genome – GenBank database under the accession numbers CP009303 to CP009311 – and predicted proteins were compared those of *K. lactis* and other *K. marxianus* strains which are available at NCBI Assembly database (https://www.ncbi.nlm.nih.gov/assembly). The genome sequences of *K. lactis* CBS 2359 strain (accession code: GCA_000002515) and *K. marxianus* strains DMKU3-1042 (GCA_001417885) and NBRC 1777 (GCA_001417835) were downloaded and analyzed using BLAST version 2.6.0, OrhtoVenn (Wang et al., 2015), SynChro (Drillon et al., 2014) and Mauve (Darling et al., 2004).

To evaluate the degree of conservation of genes and proteins of *K. marxianus* strains and *K. lactis*, the gene set of each genome was aligned by a bidirectional approach using BLASTN tool and the orthologous proteins were clustered using OrthoVenn. To analyze the synteny of genomes and identify conserved blocks of genes, the chromosomes sequences were aligned using Mauve and the syntenic blocks between *K. marxianus* and *K. lactis* were identified using SynChro.

2.6. Identification, classification and analysis of transposable elements (TEs)

The identification and classification of retrotransposon elements in the regions close to the loci of lactose permease genes of *K. marxianus* CCT 7735 genome were performed using the LTR-Finder (http://tlife.fudan.edu.cn/ltr_finder/) (Xu and Wang, 2007) and Censor programs (Kohany et al., 2006). In addition, the analysis of open reading frames in coding regions of each identified TE were carried out using ORF-finder tool (http://www.ncbi.nlm.nih.gov/projects/gorf/). The sequences found were classified as complete elements and degenerate sequences. Complete elements contain sequence similarity with proteins related to transposition machinery and terminal repeats conserved. Degenerate sequences contain sequence identity with TKM retrotransposon, however lack structural features or protein coding sequences related to transposition.

2.7. Evidence of ectopic recombination

Evidence of ectopic recombination in the TEs flanking the duplicate genes of the lactose metabolism were analyzed using RDP (Recombination Detection Program) (Martin and Rybicki, 2000), Geneconv (Padidam et al., 1999), Bootscan (Martin et al., 2005), Maximum Chi Square (Smith, 1992) Chimaera (Posada and Crandall, 2001), Sister Scan (Gibbs et al., 2000), and 3Seq (Boni et al., 2007) implemented in RDP version 3.0 (Martin et al., 2010). TEs belonging to each family were aligned using Mega version 6.06 (Tamura et al., 2013). Alignments were analyzed using standard configurations for the different methods implemented in RDP software with a 0.005 cutoff for the Bonferroni-corrected p-value. Only ectopic recombination events detected by at least four of the methods used were considered reliable.

2.8. Structural modeling of lactose permease

The prediction of three-dimensional structure of the Lac12 protein was based on homology modeling from multiple templates using Phyre2 (Protein Homology/analogy Recognition Engine version 2.0) (Kelley et al., 2015). The structure in PDB format was visualized with Pymol version 2.0 (Schrödinger, LLC). The quality of obtained models was measured by evaluation of geometry and stereochemistry, energy distribution and other characteristics of the 3D models, using RAMPAGE (Ramachandran chart) (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php), and PROSA-Web (Wiederstein and Sippl, 2007). The correction of the molecular dynamics was done through the YASARA Energy Minimization Server (Krieger and Vriend, 2014). TMHMM Server v.2.2 (Krogh et al., 2001) was used for prediction of the transmembrane helixes of the proteins, the common regions were aligned by Clustal Omega web server (Sievers and Higgins, 2014) and a selection of the exposed parts of potential interaction with lactose was chosen and aligned by WEblogo 2.8.2 (Crooks et al., 2004).

2.9. Total RNA extraction and cDNA synthesis

The total RNA was isolated using miRneasy Mini Kit (Qiagen, Netherlands), as suggested by Tesorero et al. (2013) and its quality was evaluated by agarose gel electrophoresis 1 % (w/v). Subsequently, the concentration of total RNA was determined by spectrophotometer, $\lambda = 260$ nm and A_{260}/A_{280} ratio calculated. Then, 1 μg of RNA was treated with DNAse I free RNAse (BioLabs, NE, USA) for complementary DNA strand synthesis. Posteriorly, the Polymerase Chain Reaction (PCR) was performed to check the RNA contamination with genomic DNA. A primer pair for the gene encoding glyceraldehyde 3P dehydrogenase was used in PCR. In the synthesis of 20 μL of cDNA, 0.5 μL of treated RNA was used in the reverse transcription reaction IMPROM-II (Promega, WI, USA) containing random primers as described in the manufacturer's manual. The amplification products generated were analyzed by agarose gel electrophoresis 1.0 % (w/v). The positive control of reverse transcription reaction was performed with chromosomal DNA, while the negative control was carried out with the total RNA treated with DNase.

2.10. Real-time PCR analysis

The analysis of the expression of lactose permease genes was performed by Real Time PCR. Primers were designed by Primer3-Web version 4.0.0 (http://primer3.sourceforge.net) and are shown in Table 1. The qPCR was prepared with the SYBR Green PCR Master (Invitrogen, CA, USA). PCR reaction contented: 1 μL of cDNA, 12.5 μL of the mixture SYBR® Green (dye SYBR® Green I, $MgSO_4$, dNTPs, and Taq polymerase), 1 μL of each primer, and 4.5 μL of DEPC water were added. The efficiency of the primers was previously checked by PCR. The amplification was performed in the equipment of Bio-Rad (Real-Time CFX96 System) using the following program: 95 °C for 10 min, then 40 cycles (95 °C for 30 s and 60 °C for 1 min) followed by a gradual denaturation for elaboration of the melting curve, with increment of 1 °C per minute until the temperature of 95 °C. The fluorescence intensity of each transcript, TC, was determined from the linear increase in fluorescence above the baseline fluorescence (background). The relative expression was determined using the relative quantification method $2^{-\Delta\Delta CT}$ (Winer et al., 1999; Livak and Schmittgen, 2001), in which the expression of the constitutive gene encoding glyceraldehyde 3-P dehydrogenase is evaluated as an endogenous control. To calculate the efficiency of each primer pair, a standard curve using a series of cDNA dilutions was constructed. This experiment was carried out in three replicates.

Table 1
Primers used in qPCR to amplify lactose permease genes of *Kluyveromyces marxianus*.

Target (Reference)	Primer name (Sequence 5′-3′)	Amplicon size
LAC12 Chr2 (this work)	Chr2.1Lac12L (GCAGACACTTCTAGCTTGCC) Chr2.1Lac12R (TCGTCCTGCTTCTTATCTGGA)	110 nt
LAC12 Chr3.c1 (this work)	Chr3.c1Lac12L (ATCGGGATCACAAGGAAGCC) Chr2.c1Lac12R (GCACTTCCTCTCTAGCGTCC)	100 nt
LAC12 Chr3.c2 (this work)	Chr3.c2Lac12L (TGGGAGCAGTAGAACATAAGGAG) Chr3.c2Lac12R (TTCGTTGGTGTCTTCTGCGT)	119 nt
LAC12 Chr8 (this work)	Chr8.1Lac12L (ACAGATCCAGTATTTGCCGCA) Chr8.1Lac12R (TCTTCAATCGTCATGGCATCC)	117 nt
All *LAC12* (this work)	ALLLac12L (GAATGTTTTGATGAATGGTGT) ALLLac12R (CCTTGGGTTTGGAGCCTCAA)	113 nt
GAPDH endogenous (Wu et al., 2012)	G3PR (TCTCTGAAGGTAAGTTGAAGGACG) G3PF (TCTCTGAAGGTAAGTTGAAGGACG)	120 nt

2.11. Phylogenetic tree

The phylogenetic tree was computed using the MEGA 6.0.6 program (Tamura et al., 2013), using the statistical method of maximum likelihood and Jones–Taylor–Thornton (JTT) model.

2.12. Statistical analysis

All experiments were performed in triplicates. The statistical significance of the physiological parameters was determined by Student's *t* test, and the normality test was determined by Shapiro–Wilk test. A p value $\leq$ 0.05 was considered to be significant.

3. Results and discussion

3.1. Functional analysis of coding sequences

Table 2 summarize the characteristics of three *Kluyveromyces* genomes: *K. lactis*, *K. marxianus* CCT 7735 and *K. marxianus* DMKU3-1042. Based on similarity analysis, the proteins encoded by *K. marxianus* CCT 7735 genes were classified into 14 functional groups based on gene ontology (GO) terms (Fig. 1A and Table S1). In total, 814 (17 %) of 4774 predicted proteins were annotated as "hypothetical proteins" and proteins without defined function. These proteins were included in the first group. The "metabolic processes" group was formed by 860 (18 %) proteins involved in central carbon metabolism as well as other enzymes that participate in other catabolic or anabolic reactions.

Taking into account that some strains of *K. marxianus* are one of the few yeasts capable of assimilating lactose as the sole carbon and energy source, we proposed a specific group for the "lactose metabolism". This group is formed by 28 proteins including the lactose permease, β-galactosidase and enzymes of the Leloir pathway. Notably, we found four copies of the gene encoding lactose permease (*LAC12*) in *K. marxianus* CCT 7735 genome; one in the chromosome 2, two in the chromosome 3 and one in chromosome 8. One of the copies of the chromosome 3 is located physically close to the gene encoding β-galactosidase (*LAC4*). It should be pointed out that *K. lactis* has only one copy of the *LAC12* gene and it is also close to the *LAC4* gene. The β-galactosidase enzyme is responsible of hydrolyzing lactose into glucose and galactose; these monomers are metabolized by the glycolytic and the Leloir pathways respectively. Three genes, encoding the enzymes of the Leloir pathway, were identified on chromosome 2 of *K. marxianus* CCT 7735. These genes are organized in sequentially in the chromosome 2. The first gene encodes a galactose-1-phosphate uridyltransferase enzyme (*GAL7*). The second gene encodes a uridine diphosphoglucose 4-epimerase (*GAL10*), while the third one encodes the enzyme galactokinase (*GAL1*), which plays both catalytic and regulatory roles.

Another important feature of *K. marxianus* CCT 7735 is its fermentative capacity (Diniz et al., 2012). Thus, we retrieved from the genome sequence a fourth subset of 25 genes associated with ethanol metabolism ("alcohol metabolic process"). Among them, those encoding alcohol dehydrogenases, aldehyde dehydrogenases and pyruvate decarboxylase.

The fifth group includes 296 proteins related to "transcription", such as RNA polymerases and transcription factors. In the "cell division" group, 418 proteins related to cell division such as chromosome segregation, budding and sporulation, were included. The 651 proteins involved in two-component systems are in the "signaling" group. The 213 genes encoding proteins related to cellular transport processes ("proteins, sugars and ions") were put in the eight group ("transport"). The group "DNA metabolic process" was formed by 148 proteins involved in the process of DNA replication, DNA repair and transposition ("transposon activity"). A total of 572 proteins with structural function, as tubulin, actin, myosin and cell wall synthesis, formed the group "structural molecule activity". The group "protein metabolic process" is constituted by 143 proteins involved in the translation process, such as, those encoding ribosomal components and elongation factors.

A specific group for "response to stress" was created and it includes 84 heat shock proteins and other proteins related to oxidation–reduction processes. The ubiquitin, proteases and other components were included in the "protein catabolic process" group (group 13). Finally, we also proposed a "mitochondrial derivate" group related to division and segregation in the mitochondria, including 129 proteins.

It is important to point out that there is no evidence of specialized chromosomes for *K. marxianus*, containing a high frequency of genes with related functions. Our proposed classification

Table 2
General characteristics of three *Kluyveromyces* genomes.

Species	Genome size (Mb)	Average GC content (%)	Total CDS	Source
K. marxianus CCT 7735	10.70	39.11	4787	Silveira et al. (2014)
K. marxianus DMKU3-1042	10.97	40.14	4952	Lertwattanasakul et al., 2015
K. lactis	10.60	38.70	5329	Dujon et al. (2004)

GC: guanine + cytosine.
CDS: coding DNA sequence.

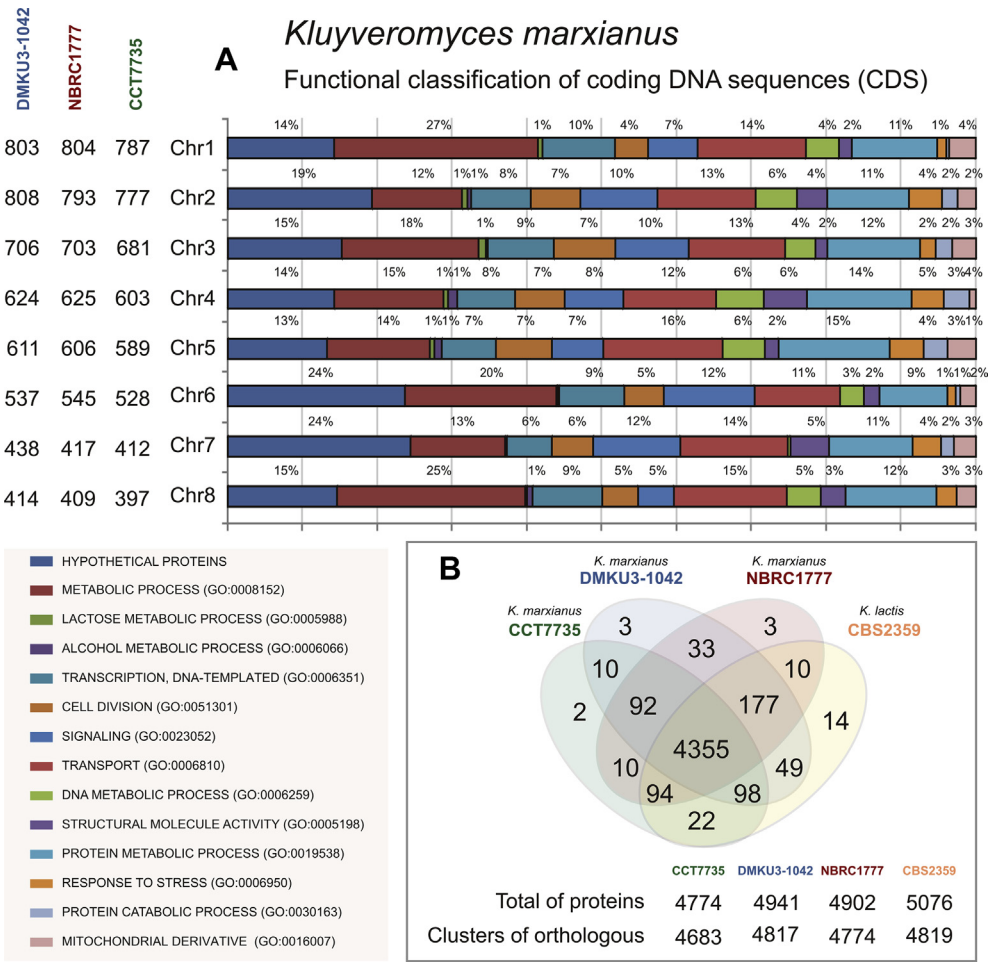

Fig. 1. An overview of *K. marxianus* CCT 7735 genome. (A) Functional classification of predicted Coding DNA Sequences (CDS) of *K. marxianus* CCT 7735. The numbers in left correspond to the CDS counting in each chromosome of *K. marxianus* DMKU3-1042, NBRC1177, and CCT 7735, for comparison purposes. The stacked bar chart shows the frequencies of functional classes in which the CDS of *K. marxianus* CCT 7735 were classified in each chromosome. (B) Clustering of orthologous proteins of *K. marxianus* CCT 7735, *K. marxianus* DMKU3-1042, *K. marxianus* NBRC 1777 and *K. lactis* CBS 2359. The Venn diagram plotted by OrthoVenn shows shared orthologous protein clusters among the *Kluyveromyces* strains. The number of proteins of each *Kluyveromyces* strain and the number of orthologous groups in which they were included are listed below.

showed that the eight chromosomes contains proportionally a similar number of genes involved in each functional group. Groups “lactose metabolism” and “alcohol metabolic process” are found in lowest frequencies, once their proteins are very specific in comparison with the other groups. It is noteworthy that *K. marxianus*, in contrast to *S. cerevisiae*, is a pre-whole genome duplication (WGD) yeast, which explains the low redundancy of genes. Consistent with that, Dujon et al. (2004) showed that in *K. lactis*, which is also pre-WGD yeast, the frequency of blocks of genes involved in the same function is low.

In order to better understand the metabolic diversity among *K. marxianus* strains, we clustered the orthologous proteins of *K. marxianus* CCT 7735, *K. marxianus* DMKU3-1042 and *K. marxianus* NBRC1777 using Orthovenn (Fig. 1B). We also compared them with the well-studied *K. lactis* CBS 2359 strain. The comparison among proteins of three *K. marxianus* strains and *K. lactis* CBS 2359 showed high conservation among them. These yeasts have a similar number of proteins, which were clustered in a similar number of orthologous groups. Most of proteins were clustered in 4335 groups that are shared by all analyzed yeasts.

Six out of the ninety-two orthologous groups shared exclusively among *K. marxianus* strains are related to sugar transport. In *K. lactis* CBS 2359, there is also a group including a sugar transport protein among its fourteen exclusive clusters. It is noteworthy that *K. marxianus* CCT 7735 has two exclusive clusters of orthologous proteins, one of them includes an ABC transporter and the other a hypothetical protein. *K. marxianus* DMKU3-1042 showed three exclusive clusters, one of them includes a transposon protein and the others are hypothetical proteins. Likewise, *K. marxianus* DMKU3-1042, *K. marxianus* NBRC1777 also presents 3 exclusive clusters, including a nucleolin, a 3-dehydrosphinganine reductase and a hypothetical protein. *K. marxianus* CCT 7735 and *K. lactis* share 22 unique clusters of orthologous proteins, that include nine ribosomal proteins, one nuclear pore complex protein Nup62, one vesicle-associated protein and eleven hypothetical proteins. These unique cluster comparisons do not reveal a clear difference in terms of metabolic diversity among *K. marxianus* strains. Therefore, the differences may be related to a polymorphism between the strains/species once that most of proteins were included in orthologous groups shared by all yeasts.

The study of the conservation of synteny between species allows to trace the events which have shaped the genome evolution. When applied to yeasts, synteny analysis revealed a considerable extent of genome reorganization (Dujon et al., 2004). In order to gain insights about genome organization in *K. marxianus*, we compared the genes set of the six chromosomes of *K. lactis* CBS 2359 with those of the eight chromosomes of *K. marxianus* DMKU3-1042, selected as reference due its complete genome, using

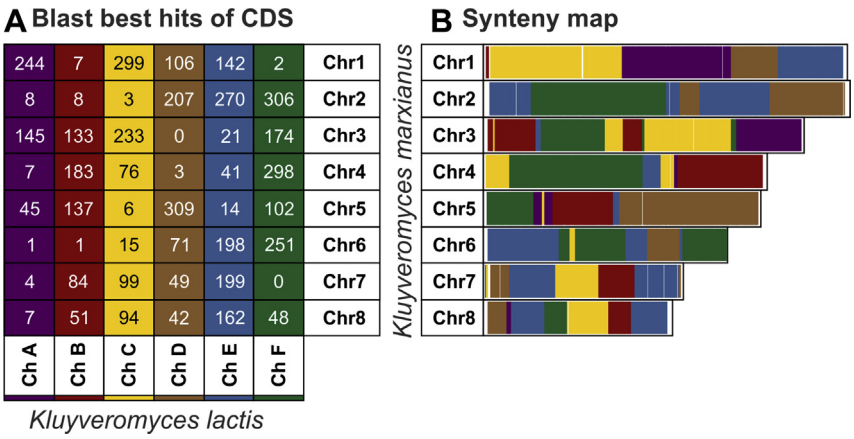

Fig. 2. Genome comparison between *K. marxianus* and *K. lactis*. (A) Comparison of Coding DNA sequences (CDS) repertory of *K. marxianus* DMKU3-1042 (selected as reference) and *K. lactis* CBS 2359 through similarity searches. The set of CDS of each chromosomes of *K. marxianus* (chromosomes 1 to 8) and *K. lactis* chromosomes (A to F) were compared using BLASTn. (B) Synteny map reconstructed by Synchro. The horizontal bars correspond to the eight chromosomes of *K. marxianus* and the colored segments corresponds to segments of the six chromosomes of *K. lactis*. The synteny map suggest that ancestral yeasts underwent a chromosomal shuffling which resulted in the speciation event between *K. marxianus* and *K. lactis*.

similarity searches (Fig. 2). Due to the phylogenetic proximity between the two species, most of *K. marxianus* genes have homologs in *K. lactis*. The analysis of best-hits in similarity searches provided some insights about recombination events between these two yeasts, suggesting that chromosome rearrangement with exchange of gene blocks occurred in the course of evolution.

The chromosome 1 genes of *K. marxianus* share homologs mainly with the genes of chromosomes A, C and E of *K. lactis* (Fig. 2A). The genes of chromosome 2 with the chromosomes D, E and F. The genes chromosome 3 share homologous with those located in chromosomes A, B, C and E. For the chromosome 4, the highest number of homologs is on chromosomes B and E. Approximately, half of the genes on chromosome 5 of *K. marxianus* have homologs on chromosome D of *K. lactis*. Chromosome 6 shares a higher number of homologs with chromosomes D and E. Chromosomes 7 and 8 have a similar pattern of homologs distribution with *K. lactis*, sharing the highest number of homologs with chromosome E.

The synteny map (Fig. 2B) agreed with the patterns of homologs distribution observed in similarity searches, suggesting that ancestral yeasts underwent a chromosomal shuffling which resulted in the speciation event between *K. marxianus* and *K. lactis*. Gene clusters and proteins are conserved but their arrangement is different in the genomes of these two *Kluyveromyces* species, which may be associated with the physiological differences in response to environmental shifts that these species experienced.

3.2. Transposable elements analysis

The gain and loss of specific genes may be critical for the functional differentiation between species and evolution. The acquisition of three further copies of lactose permease in *K. marxianus* compared to *K. lactis* raises the question of their origin. These copies are in subtelomeric regions, which are hotspots of gene evolution and functional divergence (Anderson et al., 2015). Gene duplication events could be products of DNA repair mechanisms, replication, and due to ectopic recombination among transposable elements of the same family (Chia and Goldenfeld, 2011). To address this issue, we analyzed repetitive DNA sequences close to lactose permease genes in *K. marxianus* CCT 7735 genome. Five copies of retrotransposons were identified (Table 3). These elements were aligned against sequence of transposable elements deposited in RepBase (http://www.girinst.org/repbase/update/index.html) via CENSOR program (Kohany et al., 2006). The copies found belong to retrotransposon TKM (Neuvéglise et al., 2002). Table 3 shows complete retrotransposon copies in chromosome 3 and 8 and one degenerated element in the chromosome 8. The retrotransposon TKM is 5994 nucleotides in length and has two ORFs (open reading frames) related to the gag (capsid-related retrotransposon protein) and pol (polyprotein) regions containing protease, integrase, reverse transcriptase and RNase H (Neuvéglise et al., 2002). Furthermore, in the chromosome 2 were identified two Solo-LTRs (long terminal repeats) belonging to TKM retrotransposon (Neuvéglise et al., 2002) (Table 3). The presence of degenerated copies and Solo-LTRs related to retrotransposon TKM indicates possible ectopic recombinations. Furthermore, evidences of ectopic recombination in the TEs close to the *LAC12* gene were analyzed using RDP software ($p < 0.005$), the results (Table S2) showed four recombination events between the retrotransposons located in the chromosomes 3 and 8.

Table 3
Genome localization of the identified transposable elements (TEs).

Name	Chromosome	Distance to *LAC12*	Status
TKM1	Chromosome 3	22 184 pb upstream	Complete
TKM2	Chromosome 3	15 000 pb upstream	Complete
TKM3	Chromosome 3	7815 pb upstream	Complete
TKM4	Chromosome 8	10 567 pb upstream	Complete
TKM5	Chromosome 8	8342 pb upstream	Degenerated
Solo-TKM1	Chromosome 2	5921 pb downstream	LTR solo
Solo-TKM2	Chromosome 2	3078 pb downstream	LTR solo

Therefore, the characterization of TEs close to *LAC12* region suggests that they have been responsible of increasing the copy number of the *LAC12* gene. The extra *LAC12* copies may be related to the higher lactose uptake in some *K. marxianus* strains than in others.

3.3. Lactose metabolism analysis

The major common feature between *K. lactis* and *K. marxianus* strains isolated from dairy environments is the capacity to assimilate lactose as the sole carbon source, which is absent in *S. cerevisiae* (Lane and Morrissey, 2010). Regarding lactose–galactose genes, *K. marxianus* CCT 7735 has, in contrast to *K. lactis*, four copies of the *LAC12* gene (Table S3) and two copies of

Table 4
Physiological parameters of *K. marxianus* CCT 7735 and *K. lactis* CBS 2359 cultured under aerobiosis and hypoxia, in minimal medium (YNB) containing lactose (20 g/L) as the sole carbon source.

Parameters	*K. marxianus* CCT 7735		*K. lactis* CBS 2359	
	Aerobiosis	Hypoxia	Aerobiosis	Hypoxia
Growth rate ([μ] h^{-1})	$0.41^{a*} \pm 0.00$	$0.20^{b} \pm 0.01$	$0.30^{a} \pm 0.01$	$0.15^{b} \pm 0.01$
$Y_{X/S}$ (g g^{-1})	$0.18^{a} \pm 0.02$	$0.06^{b} \pm 0.01$	$0.17^{a} \pm 0.03$	$0.06^{b} \pm 0.01$
$Y_{P/X}$ (g g^{-1})	$2.10^{b} \pm 0.16$	$5.11^{a} \pm 0.24$	$1.88^{b} \pm 0.16$	$3.47^{a} \pm 0.28$
q_S (g g^{-1} h^{-1})	$2.26^{b} \pm 0.20$	$2.81^{a} \pm 0.28$	$1.80^{b} \pm 0.26$	$2.74^{a} \pm 0.50$
q_P (g g^{-1} h^{-1})	$0.86^{b} \pm 0.07$	$1.00^{a} \pm 0.03$	$0.57^{a} \pm 0.06$	$0.53^{a} \pm 0.06$

*Equal letters mean that there is no statistical difference between treatments
± standard deviation.
$Y_{X/S}$ = Cell biomass yield.
$Y_{P/X}$ = ethanol yield per cell mass.
q_S = specific lactose consumption rate.
q_P = specific ethanol production rate.

the *GAL5* gene. As aforementioned, *K. lactis* has only one copy of both *LAC12* and *GAL5* genes. It is noteworthy that other *K. marxianus* strains have the same number of copies of these genes (Lertwattanasakul et al., 2015; Varela et al., 2017).

Although the capacity of metabolizing lactose is a remarkable trait of *K. marxianus*, several studies have pointed out that this trait is variable between strains (Grba et al., 2002; Lane et al., 2011; Rocha et al., 2011). Varela et al. (2017) evaluated fifteen *K. marxianus* strains in order to detect the degree of variability between them. They observed a clear bimodal distribution because six strains (CBS 397, CBS 608, NCYC 179, CBS 5795, CBS 6432 and CBS 1555) grew and consumed a high amount of lactose. On the other hand, nine strains (CBS 6556, CBS 5670, CBS 4857, CBS 712, CBS 7894, CBS 7858, CBS 1596, CBS 2233 and CBS 4354) grew poorly and showed a low capacity of lactose uptake. The strain used in this study, that is, *K. marxianus* CCT 7735 grew well on lactose with maximum growth rate of 0.41 h^{-1}, which is expected since it was isolated from a dairy industry (Silveira et al., 2005).

We hypothesize that this may be related to a higher lactose uptake. To address that, we evaluate the physiological parameters of both *K. marxianus* CCT 7735 and *K. lactis* CBS 2359, which is considered a model organism for lactose metabolism studies, grown on minimal medium containing lactose as the sole carbon source under aerobiosis and hypoxia (Table 4).

K. marxianus strains display higher growth rates (μ); indeed *K. marxianus* CCT 7735 grew faster than *K. lactis* CBS 2359 under both aerobic and hypoxic conditions (Table 4). In hypoxia, we observed in both yeasts that the growth rate and cell biomass ($Y_{X/S}$) decreased. In agreement with these results, the ethanol yield per cell mass ($Y_{P/X}$) increased (Table 4). *K. marxianus* CCT 7735 stood out in terms of ethanol production, since its $Y_{P/X}$ and q_P parameters were higher than in *K. lactis* CBS 2359 (Table 4). As shown in Table 4, both specific lactose consumption rate (q_s) and qp in *K. marxianus* CCT 7735 were higher in hypoxia than in aerobiosis, highlighting its great capacity to convert lactose into ethanol.

Based on our results, we hypothesize that the observed difference between *K. marxianus* and *K. lactis* in terms of lactose uptake is due to the expression level of the *LAC12* genes. Thus, we also analyzed the expression of the four *LAC12* genes by real time PCR in order to evaluate which of them are expressed in *K. marxianus* CCT 7735 in both aerobiosis and hypoxia (Fig. 3).

In aerobiosis, the expression of the *LAC12* genes found in the left arm of chromosome 3 (Chr3.c1), chromosome 2 (Chr2) and chromosome 8 (Chr8) was similar. The copy located in the right arm of the chromosome 3 (Chr3.c2) had a lower expression (0.98). These results suggest that the high lactose consumption rate in *K. marxianus* CCT 7735 is, in part, explained by the expression of the three copies aforementioned. These results are in contrast to the data previously published by Varela et al. (2017). These authors observed that the expression of the Chr3.c1 copy was higher than the other *LAC12* genes, while the copies Chr3.c2 and Chr8 had only basal expression, and the copy Chr2 had some degree of lactose induction but the level of expression was low compared to Chr3.c1. They also suggested, based on phenotypic, biochemical and genetic data, that the main transporter responsible for lactose uptake in *K. marxianus* CBS 397 and *K. marxianus* CBS 6556 is the *LAC12* gene located on the left arm of chromosome 3, transcribed divergently from the *LAC4* gene. Oppositely, our results suggest that in *K. marxianus* CCT 7735, contrary to *K. marxianus* CBS 397 and *K. marxianus* CBS 6556, three permeases are significantly expressed. Silveira et al. (2019) evaluated the lactose transport in *K. marxianus* CCT 7735 in aerobiosis, which was characterized by maximal velocity (V_{max}) of 3.67 ± 0.30 mmol h^{-1} g^{-1} (dry weight) and a half-saturation constant (K_S) of 1.67 ± 0.41 mmol l^{-1}. These authors indicated that two different lactose transporters or two sequence variants of the same transporter are functional in *K. marxianus* CCT 7735.

Under hypoxic conditions, we observed a higher expression of the *LAC12* genes compared to aerobiosis. The expression of the copies Chr3.c1 and Chr2 were four and three-fold higher, respectively. These results agreed well with the results published by Diniz

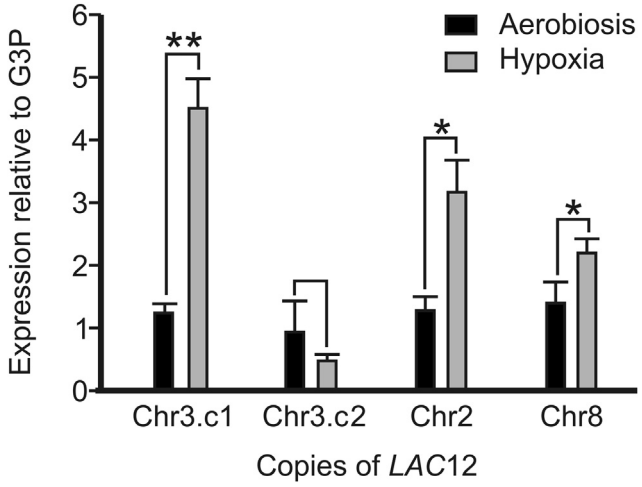

Fig. 3. Relative *LAC12* genes expression in *K. marxianus* CCT 7735 by quantitative real time PCR (qRT-PCR). *LAC12* genes expression were normalized to endogenous glyceraldehyde 3-P dehydrogenase (GAPDH) gene. *K. marxianus* CCT 7735 was cultured in YNB medium containing lactose as the sole carbon source under aerobic and hypoxic conditions.

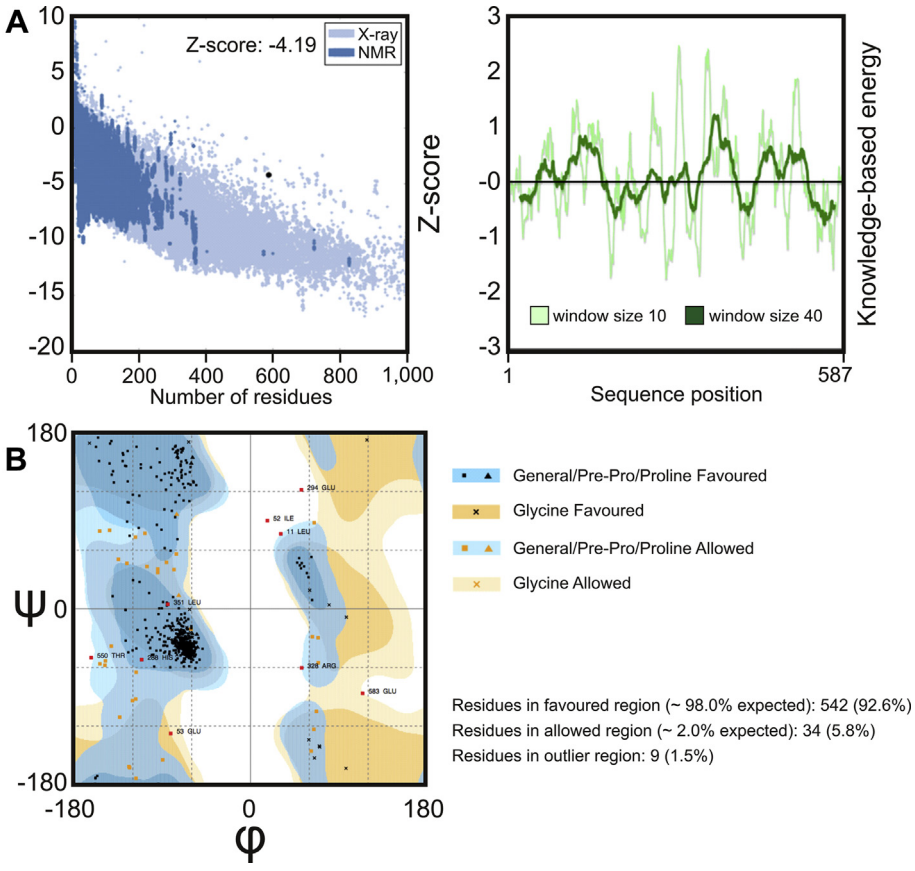

Fig. 4. Model to the protein biologically feasible for Lac12p copy Chr3.c1 of *K. marxianus*. (A) Results of PROSA and (B) Ramachandram graph.

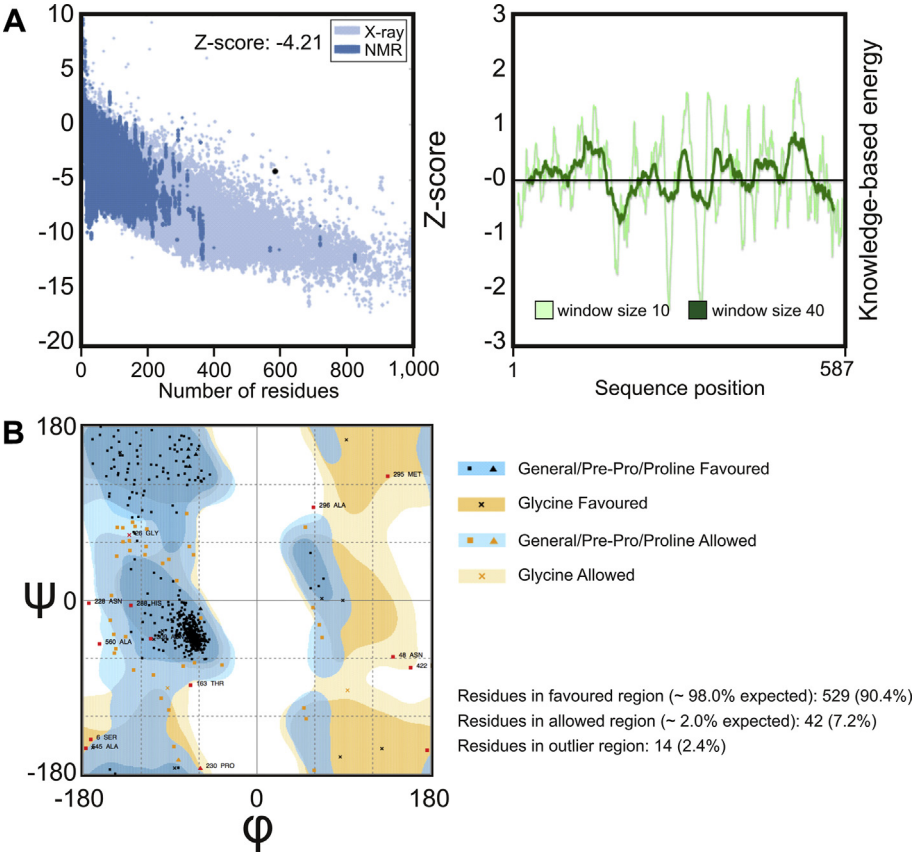

Fig. 5. Model to the protein biologically feasible for Lac12p of *K. lactis*. (A) Results of PROSA and (B) Ramachandram graph.

Table 5
Results of the mean deviation of the alignment between the structures of the lactose permease copies. The mean deviations are the RMSD (Root-Mean-Square Deviation) values, which represent the displacement among the atoms of two protein structures aligned using the Pymol version 2.0 software.

	Chr2	Chr3 copy1	Chr3 copy 2	Chr8
K. lactis	0.935	1.046	0.986	0.919
Chr8	1.067	1.202	0.737	–
Chr3 copy 2	0.972	1.148	–	–
Chr3 copy1	0.925	–	–	–

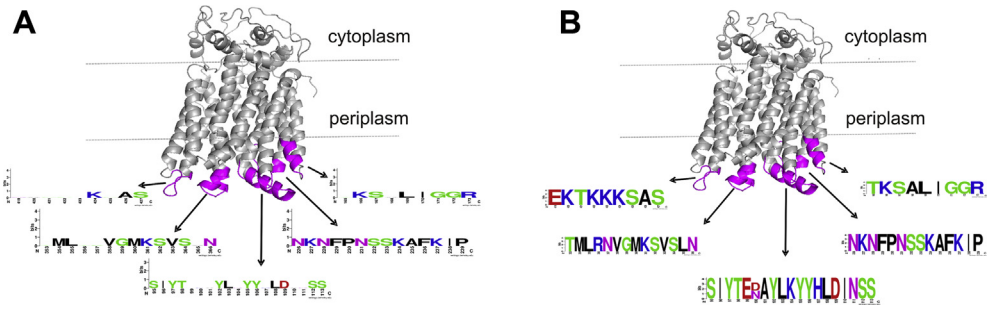

Fig. 6. Model structure of Lac12 protein by *K. marxianus* and *K. lactis*. (A) Homology model structure of Lac12 protein and Weblogo plots comparing the extracellular regions of the four copies of *K. marxianus* CCT 7735. (B) Homology model structure of Lac12 protein and Weblogo plots comparing the extracellular regions of Chr3.c1 of *K. marxianus* CCT 7735 and *K. lactis* CBS 2359. Extracellular regions were highlighted in magenta. Amino acid conservation of extracellular region is showed as Weblogo plots.

et al. (2012). These authors showed that the expression of the genes encoding β-galactosidase, pyruvate decarboxylase and most of the enzymes of the Leloir pathway (galactokinase, galactose-1-phosphate uridyltransferase, and epimerase) were also upregulated in *K. marxianus* CCT 7735 under hypoxic conditions. Beyond expression, they also observed enhanced activities of β-galactosidase and pyruvate decarboxylase. Therefore, the upregulation of lactose transporters and of those enzymes along with the increase of lactose uptake and enzymatic activities are likely related to the high fermentative metabolism of *K. marxianus* CCT 7735 under hypoxia.

3.4. Structural modeling of lactose permease

The high lactose consumption rate in *K. marxianus* CCT 7735 might also be associated with sequence variation among the copies of lactose permease protein, which influences both protein structure and activity. In order to detect structural differences among the lactose permeases of *K. marxianus* CCT 7735 and *K. lactis* CBS 2359, we predicted structural models using a homology modeling approach (Figs. S1–S5). The models of the proteins encoded by the four *LAC12* genes of *K. marxianus* and the single gene of *K. lactis* were based on twenty templates, with 99–100 % of confidence and identity of 23 to 8 % (Table S3). The predicted structures are basically composed by α-helixes. The validated quality by PROSA and the Ramachandram graph can be seen in Figs. 4 and 5 and in Supplementary Figs. S6–S8. The mean deviations of the alignment of the proteins were made by PyMOL and can be seen in Table 5.

In order to evaluate differences between the permeases, as well as in lactose transport, we selected the peptides outside of the membrane using TMHMM server 2.0 (Krogh et al., 2001). They were located in the structure and aligned. Amino acid conservation between the models was identified by Weblogo plots. The four proteins of *K. marxianus* CCT 7735, and the single one of *K. lactis* CBS 2359 (Fig. 6A and B) were compared. The peptides of lactose permease which are exposed to the extracellular environment are those that most likely interact with lactose. In the Weblogo plots, the sizes of the letters are related with the conservation of the amino acids; on the other hand, absence of letter means that there is no conservation. Our results show that the conservation is higher between the copy one of the chromosome 3 of *K. marxianus* and the protein of *K. lactis*, than among the four copies of *K. marxianus*. These results might suggest that the lactose permease encoded by the gene located in the chromosome 3 could have kinetic parameters more similar to those of the transporter of *K. lactis*.

Indeed, the first copy of the chromosome 3 of *K. marxianus* is closer to the *K. lactis* copy than to the other copies (Fig. 7). The phylogenetic tree also shows that the copies are very similar between strains, therefore, the differences in lactose uptake between them can be related to a small difference in their sequences. Varela et al. (2017) presented evidences that only the Lac12p encoded by the gene located in the chromosome 3 is functional. It is noteworthy that these authors also verified that the expression of this gene was higher. Since the expression of this gene in *K. marxianus* CCT 7735 was similar to other two copies, further studies should

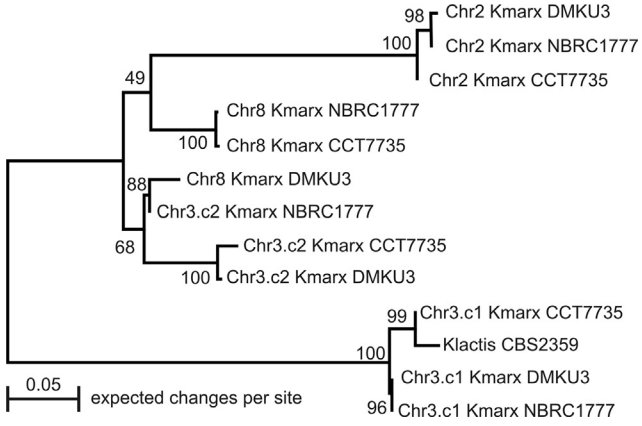

Fig. 7. Phylogenetic tree of lactose permease. The majority-rule consensus tree of amino acid sequences of the lactose permease copies in genomes of *K. marxianus* DMKU3-1042, *K. marxianus* NBRC 1777, *K. marxianus* CCT 7735, and *K. lactis* CBS 2359 was obtained by Maximum Likelihood (ML). Beside each node, the bootstrap values (expressed as percentage) calculated by MEGA are shown.

evaluate whether the permeases encoded by these genes play an important role in this strain.

The nucleotide sequence of *LAC12* present in the left arm of chromosome three is probably ancestral to the speciation because of its high degree of conservation with the gene from *K. lactis*, and the other three copies probably arose by gene duplication events.

4. Conclusions

The comparative genomic analysis presented expands our understanding about genome organization and its evolution in *K. marxianus*, which was shaped by chromosome shuffling and gene duplication events when compared with *K. lactis*. In addition, the results obtained by combination of genomic, physiological and structural analyses provide important information about *LAC12* gene organization and regulation in this yeast. Under hypoxia, a higher expression of the lactose permease genes is associated with the increase of both specific lactose consumption and specific ethanol production rates. These data interestingly show that lactose uptake is a promising target in order to optimize ethanol production by *K. marxianus* CCT 7735.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Besides, this study was supported by the Brazilian Agencies Foundation, Minas Gerais Research Foundation (FAPEMIG) as well as National Council for Scientific and Technological Development (CNPq). The work carried out at Universidade da Coruña was cofunded from Xunta de Galicia (Consolidación D.O.G. 10-10-2012. Contract no. 2012/118 and D.O.G 12-20-2016 Contract no ED431C-2016-012) cofinanced by FEDER. The authors thank the Center for Analysis of Biomolecules of Universidade Federal de Viçosa for the software used in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.funbio.2019.06.004.

References

Anderson, M.Z., Wigen, L.J., Burrack, L.S., Berman, J., 2015. Real-time evolution of a subtelomeric gene family in *Candida albicans*. Genetics 200 (3), 907–919.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Barton, A.B., Pekosz, M.R., Kurvathi, R.S., Kaback1e, D.B., 2008. Meiotic recombination at the ends of chromosomes in *Saccharomyces cerevisiae*. Genetics 179, 1221–1235.
Boni, M.F., Posada, D., Feldman, M.W., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047.
Chia, N., Goldenfeld, N., 2011. Dynamics of gene duplication and transposons in microbial genomes following a sudden environmental change. Mob. Genet. Elem. 1 (3), 221–224.
Crooks, G.E., Hon, G., Chandoni, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14 (6), 1188–1190.
Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14 (7), 1394–1403.
Diniz, R.H.S., Silveira, W.B., Fieto, L.G., Passos, F.M.L., 2012. The high fermentative metabolism of *Kluyveromyces marxianus* UFV-3 relies on the increased expression of key lactose metabolic enzymes. Anton. Leeuwenhoek 101 (3), 541–550.
Diniz, R.H.S., Rodrigues, M.Q.R.B., Fietto, L.G., Passos, F.M.L., Silveira, W.B., 2014. Optimizing and validating the production of ethanol from cheese whey permeate by *Kluyveromyces marxianus* UFV-3. Biocatal. Agric. Biotechnol. 3 (2), 111–117.
Diniz, R.H.S., Villada, J.C., Alvim, M.C.T., Vidigal, P.M.P., Vieira, N.M., Lamas-Maceiras, M., Cerdán, M.E., González-Siso, M.I., Lahtvee, P.J., Silveira, W.B., 2017. Transcriptome analysis of the thermotolerant yeast *Kluyveromyces marxianus* CCT 7735 under ethanol stress. Appl. Microbiol. Biotechnol. 101, 6969–6980.
Drillon, G., Carbone, A., Fischer, G., 2014. SynChro: a fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. PLoS One 9 (3), e92621.
Dujon, B., Sherman, D., Fischer, G., et al., 2004. Genome evolution in yeasts. Nature 430, 35–44.
Ferreira, P.G., Silveira, F., Santos, R., Genier, H., Diniz, R.H.S., Ribeiro-Junior, J., Fietto, L.G., Passos, F.M.L., Silveira, W.B., 2015. Optimizing ethanol production by thermotolerant *Kluyveromyces marxianus* CCT 7735 in a mixture of sugarcane bagasse cellulosic biomass and ricotta whey. Food Sci. Biotechnol. 24, 1421–1427.
Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.
Fonseca, G.G., Heinzle, E., Wittmann, C., Gombert, A.K., 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. Appl. Microbiol. Biotechnol. 79, 339–354.
Fukuhara, H., 2006. *Kluyveromyces lactis* – a retrospective. FEMS Yeast Res. 6, 323–324.
Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signal in recombination sequences. Bioinformatics 16, 573–582.
Godecke, A., Zachariae, W., Arvanitidis, A., Breunig, K.D., 1991. Coregulation of the *Kluyveromyces lactis* lactose permease and beta-galactosidase genes is achieved by interaction of multiple *LAC9* binding sites in a 2.6 kbp divergent promoter. Nucleic Acids Res. 19 (19), 5351–5358.
Grba, S., Stehlik-Tomas, V., Stanzer, D., Vahčić, N., Škrlin, A., 2002. Selection of yeast strain *Kluyveromyces marxianus* for alcohol and biomass production on whey. Chem. Biochem. Eng. Q. 16, 13–16.
Groeneveld, P., Stouthamer, A.H., Westerhoff, H.V., 2009. Super life-how and why "cell selection" leads to the fastest-growing eukaryote. FEBS J. 276 (1), 254–270.
Kelley, L., Mezulis, S., Yates, C., Wass, M., Sternberg, M., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858.
Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and censor. BMC Bioinform. 25 (7), 474.
Krieger, E., Vriend, G., 2014. YASARA View – molecular graphics for all devices- from smartphones to workstations. Bioinformatics 30 (20), 2981–2982.
Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305 (3), 567–580.
Kurtzman, C.P., Fell, J.W., 1998. The Yeasts: a Taxonomic Study, fourth ed. Elsevier, Amsterdam.
Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: a yeast emerging from its sister's shadow. Fungal Biol. Rev. 24 (1–2), 17–26.
Lane, M.M., Burke, N., Karreman, R., Wolfe, K.H., O'Byrne, C.P., Morrissey, J.P., 2011. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. Anton. Leeuwenhoek 100, 507–519.
Lertwattanasakul, N., Kosaka, T., Hosoyama, A., et al., 2015. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. Biotechnol. Biofuels 8, 47.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)). Methods 25, 402–408.
Martin, D., Rybicki, E.P., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.
Martin, D.P., Posada, D., Crandall, K.A., Willianmson, C., 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retrovir. 21, 98–102.
Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.
Neuvéglise, C., Feldmann, H., Bon, E., Gaillardin, C., Casaregola, S., 2002. Genomic evolution of the long terminal repeat retrotransposons in Hemiascomycetous yeast. Genome Res. 12, 930–943.
Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–224.
Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U. S. A. 98, 13757–13762.
Riley, M.I., Sreekrishna, K., Bhairi, S., Dickson, R.C., 1987. Isolation and characterization of mutants of *Kluyveromyces lactis* defective in lactose transport. Mol. Genet. Genom. 208, 145–151.
Rocha, S.N., Abrahão-Neto, J., Gombert, A.K., 2011. Physiological diversity within the *Kluyveromyces marxianus* species. Anton. Leeuwenhoek 100, 619–630.
Rubio-Teixeira, M., 2005. A comparative analysis of the *GAL* genetic switch between not-so distant cousins – *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. FEMS Yeast Res. 5, 1115–1128.
Sievers, F., Higgins, D.G., 2014. Clustal omega, accurate alignment of very large numbers of sequences. Methods Mol. Biol. 1079, 105–116.
Silveira, F.A., Diniz, R.H.S., Sampaio, G.M.S., Brandão, R.L., Silveira, W.B., Castro, I.M., 2019. Sugar transport systems in *Kluyveromyces marxianus* CCT 7735. Anton. Leeuwenhoek 112 (2), 211–223.
Silveira, W.B., Passos, F.J.V., Mantovani, H.C., Passos, F.M.L., 2005. Ethanol production from cheese whey permeate by *Kluyveromyces marxianus* UFV-3: a flux analysis of oxido-reductive metabolism as a function of lactose concentration and oxygen levels. Enzym. Microb. Technol. 36 (7), 930–936.

Silveira, W.B., Diniz, R.H.S., Cerdán, M.E., González-Siso, M.I., et al., 2014. Genomic sequence of the yeast *Kluyveromyces marxianus* CCT 7735 (UFV-3), a highly lactose-fermenting yeast isolated from the Brazilian dairy industry. Genome Announc. 2 (6) e01136-14.

Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Stanke, M., Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19, 215–225.

Stanke, M., Schoffmann, O., Morgenstern, B., Waack, S., 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinform. 7, 62.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., 2013. Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30 (12), 2725–2729.

Tesorero, R.A., Yu, N., Wright, J.O., Svencionis, J.P., Cheng, Q., Kim, J.H., Cho, K.H., 2013. Novel regulatory small RNAs in *Streptococcus pyogenes*. PLoS One 8 (6), e64021.

Varela, J.A., Montini, N., Scully, D., Ploeg, R., Oreb, M., Boles, E., Hirota, J., Akada, R., Hoshida, H., Morrissey, J.P., 2017. Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains. FEMS Yeast Res. 17.

Wang, Y., Coleman-Derr, D., Chen, G., Gu, Y.Q., 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 43 (W1), W78–W84.

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 407–410.

Winer, J., Jung, C.K., Shackel, I., Williams, P.M., 1999. Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes *in vitro*. Anal. Biochem. 270, 41–49.

Wu, Y., Wu, M., He, G., Zhang, X., Li, W., Gao, Y., Li, Z., Wang, Z., Zhang, C., 2012. Glyceraldehyde-3-phosphate dehydrogenase: a universal internal control for Western blots in prokaryotic and eukaryotic cells. Anal. Biochem. 423 (1), 15–22.

Xu, Z., Wang, H., 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35, W265–W268.