CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Functional and evolutionary analysis of sugar transporters in Kluyveromyces marxianus
Authors	Varela, Javier A.
Publication date	2018
Original Citation	Varela, J. A. 2018. Functional and evolutionary analysis of sugar transporters in Kluyveromyces marxianus. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2018, Javier A. Varela. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-11-07 08:34:13
Item downloaded from	https://hdl.handle.net/10468/6566



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh

The National University of Ireland, Cork

University College Cork

School of Microbiology

Head of School: Professor Gerald F. Fitzgerald

Supervisor: Dr. John P. Morrissey



Functional and evolutionary analysis of sugar transporters in *Kluyveromyces marxianus*

Thesis presented by

Javier A. Varela, MSc.

for the degree of

Doctor of Philosophy

in

Microbiology

April 2018

Table of Contents

Table of contents i

Declaration ii

Acknowledgements iii

Abstract iv

Publications and conference contributions viii

Chapter 1 1

Applications of *Kluyveromyces marxianus* in biotechnology

Chapter 2 26

Development of methodologies for targeted mutagenesis in *Kluyveromyces marxianus*

Chapter 3 75

Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains

Chapter 4 118

Identification of new galactose transport systems in *Kluyveromyces marxianus*

Chapter 5 153

Ploidy variation in *Kluyveromyces marxianus* separates dairy and non-dairy strains

Chapter 6 200

General discussion and recommendations for future research

Appendix 219

Declaration

I, the undersigned Javier Alberto Varela Cabrera, declare that I have not obtained a degree from University College Cork, National University of Ireland, Cork or elsewhere on the basis of this Ph.D. thesis and that the results presented in this thesis were derived from experiments undertaken by myself. Most of the research presented in this thesis was carried out at University College Cork but some data and figures in peer-reviewed research publications presented as chapters in this thesis were generated in collaborations with other researchers from Delft Technical University, University College Dublin and Yamaguchi University. Specifically, the pUDP002 plasmid in Chapter 2 was provided by collaborators at Delft Technical University, mating and gene overexpression experiments in Chapter 3 were performed in Yamaguchi University, and analysis of genome sequence data in Chapter 5 was carried out in collaboration with University College Dublin. The contribution of collaborators to those chapters is indicated at each chapter.

Signature: _____ Date: _____

Javier A. Varela Cabrera

Acknowledgements

I gratefully acknowledge the financial support of the YEASTCELL ITN program, funded by the European Union FP7 Marie Curie Programme. I would also like to thank to all the partners within the network who contributed with ideas and suggestions to my research throughout the years.

I would like to sincerely thank my supervisor, Dr. John Morrissey, for providing me with the opportunity to pursue my PhD. in his research group. Thank you for your continuous help, patience and guidance over these last four years. I always felt that my ideas and thoughts about the project were well received, and I appreciate the fact that your door was always open whenever I needed to discuss them. I would also like to thank my supervisors within the YEASTCELL network, Prof. Eckhard Boles and Dr. Jan-Maarten Geertman. I also extend my gratitude to Prof. Ken Wolfe for contributing to my research and collaborating on different stages of the project.

I am also very grateful for the numerous people that supported me in one way or another during the PhD. To my lab colleagues Loughlin Gethins, Noemi Montini, Martina Puricelli, Valentina Sforza, Fabrizia Tittarelli and Arun Rajkumar. To my friends Felipe, Esteban, Javier, Rodrigo, Nacho, Bea, Ralph, Matthew and Victor and my YEASTCELL buddies David, Raúl and Nuri. Thanks all for your support, help and also thank you for the good times we spent together.

Finally, I would like to express my deepest gratitude to my parents for their commitment to my education, and to my siblings, a huge thank you for the support despite the distance.

Lisa no estuvo satisfecha con lo que escribí sobre ella en la primera versión de la sección de agradecimientos así que decidí escribirlo en español para que sobresalga. Muchas gracias por todo tu apoyo todos estos años, por los momentos buenos que pasamos (y que nos ayudaron a olvidar nuestros doctorados por un momento), y sobre todo por tu amor y compañía en los momentos más difíciles. Tu consejo siempre me ayudo a volver a poner los pies en la tierra y seguir caminando, gracias.

Abstract

Kluyveromyces marxianus is a yeast species traditionally found in fermented dairy products such as yogurt and kefir. Because of its association with these products, this yeast has a QPS/GRAS status that facilitates applications in the food industry. *K. marxianus* is also known for possessing remarkable physiological traits: the yeast is thermotolerant, has a rapid growth rate and is able to utilise a wide range of substrates. Since these traits are desirable from an industrial standpoint, *K. marxianus* has been exploited for several biotechnological applications including bioethanol, flavour molecule and organic acid production. Perhaps the best-known application involving this yeast is the production of ethanol from whey, a by-product of the dairy industry. The capacity of the yeast to utilise lactose, essential in this environment, was previously shown to be a variable trait with some strains being able to utilise the sugar better than others. To understand the genetic basis that accounts for these differences, a detailed analysis of lactose utilisation was performed. The *K. marxianus* genome was found to encode four copies of the lactose permease (*LAC12*) gene, which are present in both good and poor lactose utilising strains. To determine if these genes were responsible for the growth differences observed, the *LAC12* genes from a good and a poor lactose utiliser were cloned into a replicative plasmid and expressed in *S. cerevisiae*. Expression of one of these gene copies (*KmLAC12*) conferred growth on lactose, indicating that this is the main lactose transporter in *K. marxianus*. However, growth on lactose was only observed when expressing the *KmLAC12* gene from a good lactose utiliser, suggesting that good and poor utilisers carry a functional (*KmLAC12*-B) and non-functional (*KmLAC12-A*) version of the lactose permease gene, respectively. Overexpression of *KmLAC12-B* and not *KmLAC12*-A conferred growth on a poor lactose utiliser, supporting this idea. A comparison of KmLac12p sequences from 12 *K. marxianus* strains confirmed that poor and good lactose utilisers carry distinct versions of the lactose transporter. Through this analysis it was established that the KmLac12p-B and KmLac12p-A differ in 16 amino acid positions. Altogether, these results showed unequivocally that differences in the *KmLAC12* sequence are the main factor accounting for variability in lactose utilisation in *K. marxianus*.

Lactose utilisation is not the only variable trait in *K. marxianus*. In fact, several studies have shown that the yeast displays a high level of diversity in a range of phenotypes including thermotolerance, stress tolerance, glucose repression and sugar utilisation. While several *K. marxianus* genome sequences are currently available, a comparative analysis aiming to determine the level of genetic diversity between these sequences had not previously been performed. To investigate this aspect, the genomes of nine *K. marxianus* strains from different origins were sequenced and compared. Also, five previously published genomes were included in this analysis. Single Nucleotide Polymorphism (SNP) analyses revealed a relatively high level of genetic diversity in the species, with up to 3 % DNA sequence divergence between alleles. These data also showed that *K. marxianus* strains can be found in different ploidy states (i.e. haploid, diploid or triploid), a finding that was confirmed by flow cytometry experiments. Diploid and triploid strains showed large regions of loss of heterozygosity in several chromosomes. Also, sequence coverage analyses revealed differences in copy number of large regions of the genome, indicating that some of these isolates are partially aneuploid. Remarkably, ploidy was found to correlate with strain origin: all the strains isolated from dairy environments were diploid or triploid while non-dairy strains were haploid. A phylogenetic tree based on SNP data was constructed to further investigate the relationship between these isolates. The haplotype tree showed that dairy and non-dairy strains form two distinct clades. In addition, all of the dairy strains utilised lactose efficiently and were found to carry the *KmLAC12*-B gene, encoding the functional lactose transporter. The SNP data also allowed us to determine that the diploid strains in this study were hybrids between a dairy and non-dairy haplotype. On the other hand, triploid strains contained three copies of the dairy haplotype. The data obtained in this study reveals that *K. marxianus* displays a high level of genetic diversity and shows that there is a relationship between genetic diversity and the environment from where the strains were isolated.

The role of *KmLAC12*-B in lactose utilisation was clear, but the function of the rest of the *LAC12* genes was not known. Since the *LAC12* gene was previously shown to encode a galactose transporter in *K. lactis*, we tested whether any of the *LAC12* genes from *K. marxianus* encoded a similar function. Heterologous expression of these genes revealed that three out of the four copies encoded galactose transporters – no function was identified for *LAC12-3*. However, single or multiple disruptions of these genes in

K. marxianus did not abolish growth on galactose, indicating the presence of additional galactose transporters. A second galactose transport system, encoded by the *HGT1* gene, was previously described in *K. lactis*. To examine if *K. marxianus* genome encoded any orthologues of this gene, a bioinformatics analysis was carried out. Several tandem copies of the *HGT1* gene were identified in *K. marxianus* genome. Also, multiple tandem copies of the glucose transporters *KHT1* and *KHT2* were discovered through this analysis. The expansion of these sugar transporters was not found in other *Kluyveromyces* genomes examined, indicating that these gene expansions are specific to *K. marxianus.* Interestingly, copy number of the *KHT* and *HGT1* genes was variable among *K. marxianus* strain; the strains were found to encode either 6 or 5 copies of *KHT* and either 5 or 3 copies of *HGT1*. To assess the functionality of these genes, heterologous expression experiments were carried out in *S. cerevisiae* EBY.VW4000, a strain deleted for all hexose transporters. Expression of the *K. marxianus* transporters identified additional galactose transporters; 4 encoded by *HGT1* orthologues and 1 encoded by a *KHT* orthologue. *K. marxianus* strains carrying mutations in these genes were constructed and tested on galactose medium. Only a mutant carrying mutations in all eight of the identified galactose transporters showed a significant growth defect on 2 % galactose medium whereas a mutant in the *HGT1* locus and the three functional *LAC* genes failed to grow on 0.1 % galactose. These results demonstrate that galactose transport is highly redundant in *K. marxianus,* that different proteins appear to be involved in low or high-affinity transport and highlight important differences between the *K. lactis* and *K. marxianus* galactose transport systems.

In addition to studying the expansion of sugar transporter families in *K. marxianus*, molecular tools to genome engineer this yeast were also developed during this PhD. While some tools to genetically manipulate this yeast are currently available, there is a need to develop new techniques that allow simple and efficient gene targeting for construction of mutants and for allelic replacement. To tackle this issue, the CRISPR-Cas9 system was implemented in *K. marxianus*. To establish the methodology, a single-plasmid system, comprising Cas9 and an sgRNA, was constructed and successfully used to target the *ADE2* gene at high efficiency (> 96%) in this yeast. The system was also shown to function in *K. lactis,* where the *ADE2* gene was targeted at similar efficiency. Mutants were obtained in both yeasts both by non-homologous end-

joining (NHEJ) and homology-dependent repair (HDR) methods, though the capacity to carry out allelic replacement by homologous recombination using repair fragments was of variable efficiency in different strains. To improve gene targeting in *K. marxianus*, the *KU80* gene involved in the NHEJ DNA repair system was disrupted using the CRISPR-Cas9 system. Homologous recombination frequency was greatly enhanced in a *ku80* mutant, suggesting that the NHEJ had been inactivated. Finally, to further improve homologous recombination repair, the *RAD52* gene, involved in homologous recombination in *S. cerevisiae*, was expressed in *K. marxianus*. A strain expressing this gene showed a seven-fold increase in homologous recombination-mediated repair, albeit from a low base. These tools were used throughout the thesis to facilitate the functional analysis of sugar transporters. The results presented in this thesis contribute to better understanding the function and evolution of sugar transporters in *K. marxianus*. Also, the tools developed in this work will help to further develop this yeast for cell factory applications.

Publications and Conference contributions

Publications

A. Research Papers that form part of this thesis

1. **Varela JA**, Gethins L, Stanton C, Ross P, Morrissey JP. Applications of *Kluyveromyces marxianus* in biotechnology. In: Satyanarayana T, Kunze G, editors. Yeast diversity in human welfare. Singapore: Springer; 2017. p. 439e53
   This work is presented as Chapter 1 in the thesis

2. Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains. **Varela J.A.**, Montini N, Scully D, Van der Ploeg R, Oreb M, Boles E, Hirota J, Akada R, Hoshida H, Morrissey JP. *FEMS Yeast Res.* 2016. May 1;17(3). DOI: 10.1093/femsyr/fox021
   This work is presented as Chapter 3 in the thesis

3. Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. Juergens H., **Varela J.A**, Gorter de Vries A. R, Gast V.G.M., Gyurchev N.Y., Rajkumar A.S., Mans R., Pronk J.T, Morrissey JP, Daran J.G. *FEMS Yeast Res.* 2018. May 1;18(3). DOI:10.1093/femsyr/foy012
   Part of this work is presented as Chapter 2 in the thesis

4. Ploidy variation in *Kluyveromyces marxianus* separates dairy and non-dairy isolates. *Ortiz-Merino R., ***Varela J.A.,** Coughlan A., Hoshida H., da Silveira W., Wilde C., Kuijpers N., Geertman JM., Wolfe K., Morrissey JP. *Frontiers in genetics*. 2018. May 21;9:94. DOI: 10.3389/fgene.2018.00094
   * = Authors contributed equally to this work
   This work is presented as Chapter 5 in the thesis

B. Research Papers from research carried out as part of the YEASTCELL PhD programme but not included in the thesis

5. Development and Implementation of Multi Locus Sequence Typing to study diversity of the yeast *Kluyveromyces marxianus* in Italian Cheeses. Tittarelli F., **Varela J.**, Gethins L., Stanton C., Ross P., Suzzi G., Grazia L., Tofalo R., Morrissey JP. *Microbial Genomics.* 2017. Jan 18. DOI: 10.1099/mgen.0.000153

6. Transcriptional Response to Lactic Acid Stress in the Hybrid Yeast *Zygosaccharomyces parabailii*. Ortiz-Merino RA, Kuanyshev N, Byrne KP, **Varela** JA, **Morrissey** JP, Porro D, Wolfe KH, Branduardi P.Appl Environ Microbiol. 2018 Feb 14;84(5). DOI: 10.1128/AEM.02294-17.

Oral presentations

Varela, J.A. and Morrissey J.P. Towards a genome-scale metabolic model of *Kluyveromyces marxianus*: a key tool to improve its industry-relevant properties. *31st International Specialised Symposium on Yeast*. Vipava, Slovenia (9th-12th October 2014).

Varela J.A., Montini N., Scully D., Van der Ploeg R. and Morrissey, J.P. Genetic dissection of lactose assimilation in *Kluyveromyces marxianus*. *General Society of Microbiology Society Annual Conference,* Liverpool, UK. (21st-24th March 2016).

Varela, J.A., Ortiz R., Hoshida H., Wolfe K. and Morrissey J.P. Expansion of sugar transporters in the genome of the industrial yeast *Kluyveromyces marxianus*. *33rd International Specialised Symposium on Yeast*. Cork, Ireland (26th-29th June 2017).

Poster presentations

Varela J.A., Cliffe M., Claesson M. and Morrissey J.P. Towards a genome-scale metabolic model of *Kluyveromyces marxianus*: a key tool to improve its industry-relevant properties. *Irish Fungal Society meeting*. Galway, Ireland. (19th-20th June 2014).

Varela J.A., Wolfe K. and Morrissey, JP. Comparative genomic analysis reveals subtelomeric gene duplications and metabolic potential of the *Kluyveromyces* genus. 6th *Federation of European Microbiology Society meeting*. Maastricht, The Netherlands. (7th-11th June 2015).

Varela J.A., Montini N., Scully D., Van der Ploeg R. and Morrissey, J.P. Deciphering molecular elements involve in lactose assimilation in *K. marxianus*. *6th Conference on Physiology of Yeasts and Filamentous Fungi*. Lisbon, Portugal (11th-14th July 2016).

Chapter 1

Applications of *Kluyveromyces marxianus* in biotechnology

This chapter was published in its totality as a book chapter.

Varela JA, Gethins L, Stanton C, Ross P, Morrissey JP. Applications of *Kluyveromyces marxianus* in biotechnology. In: Satyanarayana T, Kunze G, editors. Yeast diversity in human welfare. Singapore: Springer; 2017. p. 439e53

Abstract

Kluyveromyces marxianus is a member of the *Saccharomycetales* yeast order and is used for a variety of commercial applications, most notably production of ethanol from food waste streams. Traits such as rapid growth rate, lactose utilization, good tolerance to inhibitory molecules and thermotolerance facilitate these applications. *K. marxianus* is frequently isolated from food and beverage environments and is especially associated with fermented dairy products such as kefir. This history of food association means that *K. marxianus* has GRAS/QPS status, thereby facilitating applications in the food sector. *K. marxianus* strains have the capacity to produce a range of volatile fragrance and flavour (F&F) molecules such as higher alcohols and acetate esters and there is consequent interest in exploiting high-producing strains as F&F cell factories. The availability of genome sequences and the development of molecular tools is facilitating further applications in *K. marxianus* as a novel yeast cell factory for biomolecule production. This chapter will provide an update on the genetics and biology of this yeast, and an overview of commercial applications. It will then focus on three specific areas: *K. marxianus* for bioethanol production; for the production of fragrance and flavours; and the future development of *K. marxianus* as a yeast cell factory.

1.1 Introduction

Kluyveromyces marxianus is a yeast traditionally associated with fermented milk products and decaying fruit (Fonseca *et al.* 2008, Lane and Morrissey 2010). These two habitats reflect important enzymatic traits that the species has, namely the capacity to transport and utilize lactose, and to degrade plant fructans to simple sugars. Interestingly, the much better-known yeast, *Saccharomyces cerevisiae* lacks both of these traits, which are encoded by the *LAC* genes, *LAC12* and *LAC4*, and the inulinase gene, *INU1*, respectively. *K. marxianus* has other distinctive characteristics, most importantly thermotolerance to temperatures between 44°C and 52°C, and a rapid growth rate that is twice that of *S. cerevisiae* on rich medium. These traits make this yeast quite attractive from a biotechnological perspective. The speed at which *K. marxianus* has emerged as an important yeast for biotechnology is apparent in a bibliometric survey (http://www.ncbi.nlm.nih.gov/pubmed) that shows that >50 % of the total number of publications on this yeast have come in the past 7 years (for *S. cerevisiae* in the same period, it is 25 %). There are now many publications detailing applications for enzyme production, for fermented foods and beverages, for biofuels, and for cell factory applications. One of the more esoteric applications that has been reported is tequila production which occurs when inulinases produced by wild *K. marxianus* strains break down plant fructans in agave, liberating sugars that can then be fermented to alcohol by *K. marxianus* and *S. cerevisiae*. Some studies have compared starter cultures for these species and identified additional benefits to the organoleptic profile of *K. marxianus*-fermented agave, most likely because of the production of fruity acetate esters (Lopez-Alvarez *et al.* 2012, Amaya-Delgado *et al.* 2013, Lopez *et al.* 2014). The taxonomic position of *K. marxianus* is clear (**Figure 1.1A**): it is a member of the order *Saccharomycetales*, and one of a genus of six species (Lachance 2007). Several but not all of the species in the genus are lac$^+$, but only *K. marxianus* is inulinase$^+$ and thermotolerant (44°C) and these are often considered diagnostic traits (Lachance 2011). Apart from these characteristics, the species is quite diverse and extensive physiological variation has been reported (Lane *et al.* 2011). It is also a polymorphic yeast and depending on the strain and the environmental conditions, may grow in yeast, pseudohyphal or hyphal forms (**Figure 1.1B**). For several years, the only genomic information in the public domain was a partial sequence from the Génolevures project (Llorente *et al.* 2000), but starting with the

publication of the strain CBS 6556 in 2012 (Jeong *et al.* 2012), several have been published, and several more are available in draft format (Gao *et al.* 2015, Lertwattanasakul *et al.* 2015), Morrissey, unpublished).

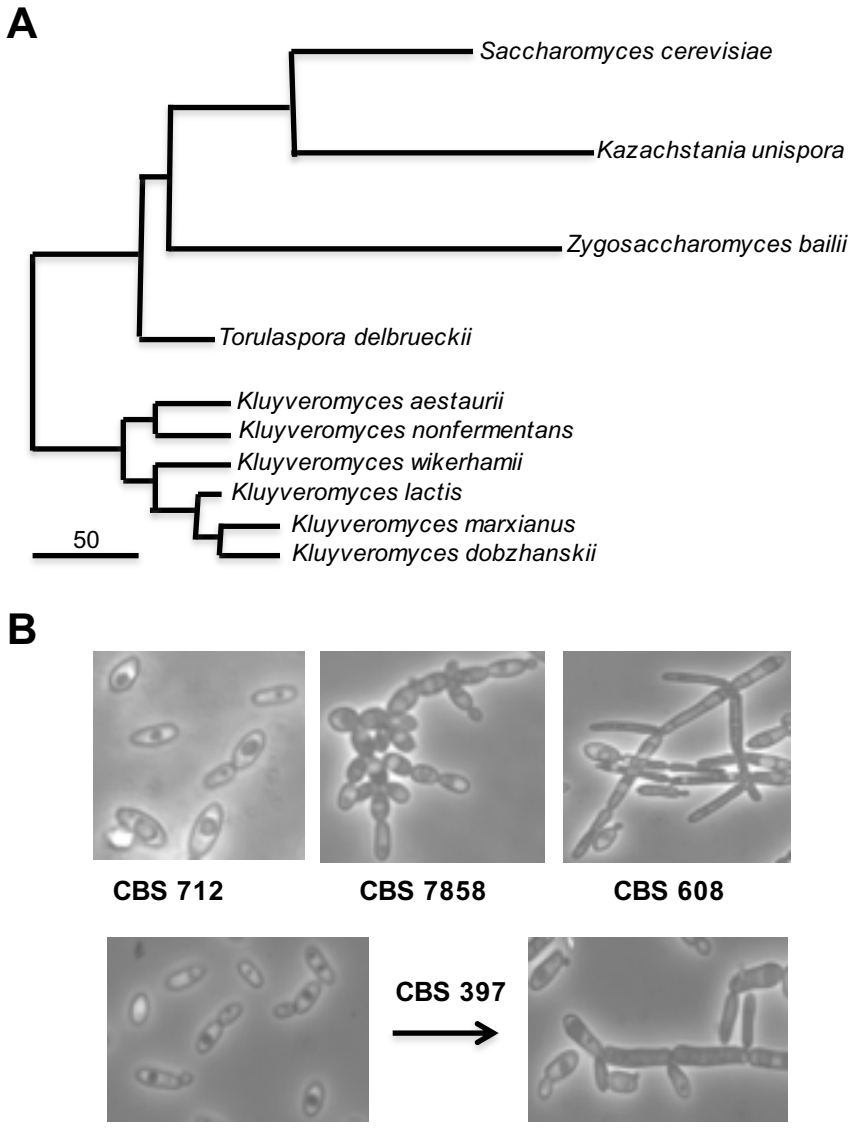


Figure 1.1. Phylogeny and morphology of *K. marxianus.*

A. The phylogeny of *K. marxianus* within the genus *Kluyveromyces* and the order *Saccharomycetales* based on 6 different loci is shown. The six species of *Kluyveromyces* and selected other species of industrial interest are included. Modified from Lachance *et al.,* 2007. **B**. Three different *K. marxianus* strains growing on identical medium show that this fungus can grow in yeast (CBS 712), pseudophyphae (CBS 7858) and hyphal forms (CBS 608). The capacity of some strains to switch morphology is shown by the growth of CBS 397 under two different environmental conditions (images courtesy of Niall Burke, UCC)

1.2 Advances made in the area

Genome sequence availability underpins much of the recent progress and advances with *K. marxianus*. Analysis of the best-annotated genome from the strain DMKU 3-1042 indicates that the genome is approximately 11 Mb, comprising 8 chromosomes ranging from 0.9 to 1.7 Mb in size (Lertwattanasakul *et al.* 2015). Interestingly, while there is strong synteny between genomes of sequenced strains, there is extensive rearrangement between *K. marxianus* and its closest relative *K. lactis* (Varela and Morrissey, unpublished). It is also worth noting that there are suggestions that some strains may have a different karyotype (Belloch *et al.* 1998, Fasoli *et al.* 2015). Comparative analyses showed that *K. marxianus* genome shares 1,552 genes with other 10 yeast species, while encoding 193 unique genes that might be involved in species-specific traits. Furthermore, a high number of sugar transporters, which might relate with the capacity of this yeast to utilize a wide variety of substrates, were identified. Along with these genome sequences, two transcriptome analyses have been published in which the global transcriptional response of *K. marxianus* to industrial-relevant conditions was evaluated. Comparing growth at different temperatures, it was found that at high temperatures *K. marxianus* switches its metabolism to cope with the production of reactive oxygen species (Lertwattanasakul *et al.* 2015). In a different study, the transcriptome of cells growing on inulin was determined (Gao *et al.* 2015). The authors found that despite being a Crabtree-negative yeast, the transcriptomic profile of *K. marxianus* strain Y179 showed up-regulation of genes associated with central carbon metabolism and ethanol production. While these studies represent the first attempts to obtain a global overview of *K. marxianus* genetic landscape, they also serve as platforms to assist in designing genetic manipulations in this yeast.

K. marxianus is considered a haploid species, though both haploid and diploid strains can be found in culture collections (Lane *et al.* 2011). The mating-type system resembles that of *K. lactis* and lacks the HO endonuclease found in *S. cerevisiae*. This is important as it means that stable haploid and diploid strains can be maintained as mating-type switching occurs at very low frequency. The potential for genetics in this yeast was demonstrated by Hoshida and colleagues, who generated auxotrophic mutants of a range of strains, successfully crossed these to create diploids, and subsequently sporulated these diploids to produce haploid spores (Yarimizu *et al.* 2013). It is also necessary to develop molecular tools for understanding the bases of

biological processes in yeast. In *K. marxianus*, molecular tools have heretofore been rather limited but the past 5 years have seen this change. The recent developments in molecular tools provide exciting opportunities to explore *K. marxianus* physiology for fundamental and applied research. While relevant and useful, the described tools need to be further improved to cope with the new requirements in the field. Ideally, tools must be universal and allow genetic modification of *K. marxianus* in a simple and versatile manner. In other words, it should be applicable to a great number of strains, regardless of their ploidy, providing fast and reliable results. Finally, it should allow performing more than one modification at the time, which could be particularly useful when introducing several genes from a metabolic pathway. Current tools are summarised in **Table 1.1**.

Table 1.1: Development of molecular tools in *K. marxianus*

Markers & Genetic integration	**Reference**
Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*	(Hoshida *et al.* 2014)
Deletion of a *KU80* homolog enhances homologous recombination in the thermotolerant yeast *Kluyveromyces marxianus*	(Choo *et al.* 2014)
Simultaneous integration of multiple genes into the *Kluyveromyces marxianus* chromosome	(Heo *et al.* 2013)
Identification of auxotrophic mutants of the yeast *Kluyveromyces marxianus* by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*	(Yarimizu *et al.* 2013)
Promoter-based gene assembly and simultaneous overexpression	(Chang *et al.* 2013)
Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*	(Abdel-Banat *et al.* 2010)
Sequential gene integration for the engineering of *Kluyveromyces marxianus*	(Pecota *et al.* 2007)
Application of the Cre-loxP system for multiple gene disruption in the yeast *Kluyveromyces marxianus*	(Ribeiro *et al*. 2007)
Promoters & Vectors	**Reference**
Characterization of *Saccharomyces cerevisiae* promoters for heterologous gene expression in *Kluyveromyces marxianus*	(Lee *et al*. 2013)
Evaluation of the tetracycline promoter system for regulated gene expression in *Kluyveromyces marxianus*	(Pecota *et al*. 2005)
Characterizing yeast promoters used in *Kluyveromyces marxianus*	(Yang *et al.* 2015)
Gene expression analysis using strains constructed by NHEJ-mediated one-step promoter cloning in the yeast *Kluyveromyces marxianus*	(Suzuki *et al.* 2015)
Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*	(Hoshida *et al*. 2014)

One essential tool is transformation and targeted genetic integration. The capacity to incorporate and remove DNA sequences from the genome has a significant importance for both fundamental and applied research. While *K. marxianus* can be transformed using routine protocols adapted from other yeasts, targeted genetic integration has proven challenging. Generally speaking, the fate of ectopic DNA transformed into a yeast cell will depend on the presence of nucleases and the efficiency of the DNA repair mechanisms available (Shrivastav *et al.* 2008). If a fragment with short flanking regions targeting a site in the genome is transformed it could be integrated at the desired spot by homologous recombination (HR) or could be incorporated at a random site through non-homologous end-joining (NHEJ). In *K. marxianus,* NHEJ is more efficient than HR, which leads to random DNA integrations. Some researchers have addressed this problem by increasing the length in the flanking regions to promote HR over NHEJ. In most cases the length used exceeds 1 kb, which is significantly more than what is used in *S. cerevisiae*, where targeted integrations can be achieved using 50 bp flanking regions. This length requirement is a hurdle that must be overcome. One approach to increase gene-targeting strategy is to eliminate the NHEJ system by deleting one of its components. Using long flanking regions, disruption of the *KU80* gene was achieved in the strain CBS 6556 (Choo *et al.* 2014). This gene encodes a subunit of the Ku heterodimer that binds to the DNA promoting random integration. In the constructed strain, HR was significantly improved from 13-70 % while no growth defects were found. Nonetheless, since the ku80 protein is part of the normal DNA repair apparatus, it would be preferable to work with wild-type strains. Recently, there has been a revisiting of earlier work on the links between HR/NHEJ and the cell cycle and it has been demonstrated that synchronizing cells and carrying out gene targeting during the S1 phase, greatly increases the likelihood of HR (Tsakraklides *et al.* 2015). This may have potential for *K. marxianus*. Although not yet published in this yeast, it is only a matter of time before CRISPR-Cas9 systems are developed and this will enhance the capacity to target genes to specific loci in the genome (Mans *et al.* 2015).

Heterologous expression requires gene cloning methods, promoters and plasmids. Although NHEJ is not desirable when performing gene targeting, it could be used as a tool for gene cloning (Hoshida *et al.,* 2014). It has been reported that a C-terminal-truncated non-functional *URA3* maker can be repaired by fusing the missing fragment

through NHEJ. Using this strategy, if a desired fragment is amplified using primers that complement the truncated *URA3*, then this sequence could be cloned by joining the maker through NHEJ. This technique allows cloning into shuttle plasmids and chromosomal integration into *K. marxianus* genome. A different system called "promoter-based gene assembly and simultaneous overexpression" (PGASO) was also developed for cloning multiple gene fragments (Chang *et al.* 2013). This method is particularly useful when simultaneous overexpression of multiple genes is required. Since ensuring controlled and stable gene expression is critical for many applications, promoters and plasmids are important elements to review. Yang *et al*., 2015, characterized the strength of 6 constitutive promoters from *S. cerevisiae* and *K. marxianus* at different temperatures. They showed that the *PGK* promoter is the strongest, regardless of the carbon source and temperatures used. Besides, they determined that the rest of the promoters had different levels of activity, important information to consider when tuning expression of a gene of interest. Another approach to evaluate *K. marxianus* promoters was established by Suzuki *et al.,* 2015. In this method a promoter sequence is fused to a reporter gene by NHEJ and then the joined fragments are randomly integrated into the genome. This methodology allowed the researchers to analyse 36 promoters in several culture conditions. In the case of plasmids there are a few cases where plasmids containing autonomously replicating sequences (ARS) have been described. For example, by constructing a genomic DNA library Hoshida *et al.,* 2014, identified a 2.8 kb ARS in *K. marxianus* (Hoshida *et al.* 2014). When dissecting this DNA fragment to determine the minimal sequence required to provide stable replication, the authors found that a 60 bp region was sufficient for ARS activity. A plasmid containing a centromere sequence was also built in this study.

1.3 Critical analysis

Advances in genomics and molecular tools pave the way for developing *K. marxianus* for cell factory applications. This moves the yeast beyond the realm of a component of fermented foods, or a starter culture, to a genuine industrial microbe capable of producing bulk or high value chemicals in large-scale industrial processes. The classic bulk chemical produced in yeast is ethanol and there is considerable interest in developing *K. marxianus* as a platform for production of biofuels, especially bioethanol. As previously mentioned, this yeast has traits of industrial relevance which could boost its use as an alternative platform to *S. cerevisiae*. However, there are also particular limitations, most importantly low ethanol yields relative to *S. cerevisiae* because of the bias towards respiratory metabolism. There is, however, a large variation in this trait between strains the potential to achieve acceptable yields that are close to the theoretical maximum (Lane *et al.* 2011, Rocha *et al.* 2011). Because the rate of production (ethanol productivity) is significantly lower in *K. marxianus* than *S. cerevisiae*, it is not likely to be a more efficient platform for conventional first-generation ethanol production from starch or sucrose. There is potential, however, for production of second or third generation bioethanol from other substrates such as cellulosic or lignocellulosic biomass, energy crops and waste streams. Some of these applications are possible because *K. marxianus* is capable of using sugars such as lactose and cellobiose, and through the production of extracellular enzymes (inulinase) capable of hydrolyzing plant fructans. Thermotolerance and production of these enzymes also contribute the potential for simultaneous saccharification and fermentation (SSF) where precursor hydrolysis and fermentation take place at the same production stage (Radecka *et al.* 2015). Multiple screening studies as well as exploitation of culture collections has resulted in the identification of lead strains for bioethanol production. In many cases, these strains have been developed further for applications using classical and GM methods.

The first success with *K. marxianus* for bioethanol production came with a process that first introduced on a commercial scale by the Carbery Group in Ireland in 1978 and subsequently adopted by dairies in the US and New Zealand (http://www.agmrc.org/media/cms/rr214_fb445fb28f0d6.pdf). This involved fermenting whey permeate to ethanol. Cheese manufacture generates large amounts of whey (approximately 9 litres per Kg cheese) that contains lactose (4-6 %), proteins,

and other nutrients. Traditionally, protein was recovered, and the resulting whey permeate treated as a waste stream but the need to comply with environmental standards and to achieve maximum commercial return has led to efforts to further valorise the whey permeate by making use of the lactose (Guimaraes *et al.* 2010). This sugar can be recovered as a dried powder or fermented to ethanol by microbes. The Carbery process involves a batch fermentation of whey permeate by a proprietary strain of *K. marxianus* that delivers a high yield of ethanol. Assimilation of lactose relies on the products of the *LAC12* gene, encoding a lactose permease, and *LAC4*, encoding a ß-galactosidase. Lac12p, which is a member of the MFS family of sugar transporters, transports lactose (and galactose), which is cleaved by Lac4p to the hexoses glucose and galactose. These are then metabolised via the standard glycolytic and Leloir pathways to pyruvate. Whether this pyruvate is then fermented to ethanol or directed into the TCA cycle for aerobic respiration is strain-dependent also depend on environmental parameters, for example oxygen availability (Lane *et al.* 2011, Rocha *et al.* 2011). There have been some efforts to optimise and refine this process and yield is influenced by pH, temperature and lactose concentration (Diniz *et al.* 2014). The second area where *K. marxianus* has been quite extensively studied for bioethanol is for production of ethanol from Jerusalem artichokes. This is a North American plant that is considered to have strong potential as an energy crop. The plant produces tubers that are rich in inulin, which may have applications as a functional food or as a substrate for bioethanol (Yang *et al.* 2015). The *K. marxianus* Inu1p protein is very effective in hydrolysing inulin to D-fructose monomers that are then transported into the cell by the Frt1p transporter and processed through the glycolytic pathway. This trait confers some *K. marxianus* strains with the capacity to ferment hydrolysed Jerusalem artichokes to ethanol in a consolidated bioprocess (CBP) where hydrolysis, saccharification and fermentation are carried out simultaneously (reviewed by Yang *et al*., 2015). Although under certain circumstances, high ethanol yields are obtained, these tend to be at lower inulin concentrations and under conditions not so amenable to large scale industrial processes. A recent transcriptome study addressed these issues and found that varying inulin concentrations and aeration levels dramatically shifted the transcription profile of the entire network of genes involved in ethanol fermentation (Gao *et al.* 2015). This illustrates the complexity of the process and the need to develop an improved understanding of the metabolic and regulatory

networks if it is to become feasible to transfer this process from proof of concept to commercial scale. Further improvements in both these processes are likely to involve selection of more robust strains, improving ethanol tolerance and anaerobic growth, and enhancing the rate of ethanol generation (productivity).

The future for a sustainable bioethanol industry is believed to lie in the production of ethanol from cellulosic material as it enables utilization of non-food crops or crop wastes. In brief, it first involves a saccharification step whereby the cellulosic material is hydrolysed with a combination of endo and exo enzymes that liberate glucose monomers from the cellulosic substrate. These enzymes are often fungal in origin and have optimum temperatures close to 50°C, whereas the subsequence fermentation of glucose to ethanol is typically carried out by *S. cerevisiae*, which requires temperatures close to 30°C for optimal growth and fermentation. There would be a considerable advantage if these steps could be combined, thereby allowing simultaneous saccharification and fermentation (SSF). The higher growth range of *K. marxianus* offers this potential and there have been a number of studies that have demonstrated this in SSF processes (for references, see (Hong *et al.* 2007). There has been considerable progress made with selection of strains with high activity and process modifications to establish the parameters that influence yield (Castro and Roberto 2014, Kang *et al.* 2014). More recently, the focus has shifted to consolidated bioprocesses, with the aim that the fermentative yeast could also produce the required enzymes. This could take advantage of the fact that at least some strains of *K. marxianus* can utilise cellobiose but it also requires the heterologous expression of other cellulolytic enzymes. The first successful demonstration of fermentation of cellulosic material to ethanol came with a study in 2010, in which an endoglucanase from *Trichoderma reesei* and a ß-glucosidase from *Aspergillus aculeatus* were expressed on the surface of *K. marxianus* NBRC1777 (Yanase *et al.* 2010). This strain was able to convert cellulosic ß-glucan to ethanol with a yield of 92 % of the theoretical maximum. Interestingly, although wild-type NBRC1777 produces ethanol optimally at 40°C, the recombinant strain had an optimum production temperature of 48°C, reflecting the higher activity of the final cellulolytic enzymes at this temperature. Cellulosic biomass derived from pre-treated straw is more complex than ß-glucans and requires a cocktail of enzymes to digest the material to monomers. For example, commercial enzyme preparations often use 5 or more fungal enzymes with

a range of activities. To address this challenge, Chang and colleagues expressed a cocktail of five cellulases as well as a transporter in strain KY3 to build a strain, designated KR7, that could successfully hydrolyse crystalline cellulose to produce ethanol (Chang *et al.* 2013).

Complete fermentation of lignocellulose biomass requires the capacity to ferment both the hexose (C6) monomers derived from cellulose/hemicellulose and the pentose (C5) sugars that comprise up to 50% of hemicellulose. Unlike *S. cerevisiae*, which is poorly able to assimilate pentose sugars, *K. marxianus* can effectively transport and metabolise the pentoses xylose and arabinose via the aldose reductase pathway. The enzymes encoded by the *XYL1*, *XYL2* and *XKS1* genes first reduce xylose to xylitol, then oxidise xylitol to xylulose, and finally phosphorylate xylulose to xylulose-5-P, which enters the pentose phosphate pathway (PPP). Although the conversion of xylose to xylulose is redox balanced overall, the two enzymes involved, xylose reductase (XR) and xylitol hydrolase (XH), use the cofactors NADPH and NAD, respectively. Under aerobic conditions, NADH is recycled to NAD^+, but in the absence of oxygen co-factor imbalance leads to the accumulation of xylitol and low ethanol yields. The consequence is that wild-type strains of *K. marxianus* generally do not efficiently ferment xylose and metabolic reprogramming is required if the yeast is to produce ethanol from pentoses. Initial studies selected the highest producing strains and assessed production parameters under various temperature and sugar regimes. This work demonstrated the potential of strains such as IMB4 (Suryawati *et al.* 2008), IIPE453 (Kumar *et al.* 2009) and DMKU3-1042 (Rodrussamee *et al.* 2011) but also highlighted the need for strain engineering if sufficient yields were to be achieved. The strategies to engineer *K. marxianus* to develop pentose fermenting strains have largely mirrored those adopted in *S. cerevisiae*, namely expression of heterologous xylose metabolizing enzymes and laboratory evolution to select higher yielding variants. One of the first studies replaced the *KmXYL1* gene in a derivative of *K. marxianus* NBRC 1777 with the *Schefferomyces stipitis* xylose reductase gene, which has a dual co-factor specificity (Zhang *et al.* 2013). This strain (YZB014) was capable of generating ethanol from xylose with a yield of 37% but the rate of production was low, and it also accumulated xylitol. A similar study that expressed *SsXYL1*, *SsXYL2* and *S. cerevisiae XKS1* in strain DMB1 generated a strain, DMB3-7, with that produced higher amounts of ethanol, though different methodologies preclude direct

comparisons of all parameters (Goshima *et al.* 2013, Goshima *et al.* 2013). More recently, this approach was extended to construct a strain that expressed the genes for *Neurospora crassa* xylose reductase (*NcXR*) and the *S. stipitis* xylitol hydrogenase (*SsXDH*) (Zhang *et al.* 2015, Zhang *et al.* 2015). The reason for using these genes is the higher activity of NcXR over ScXR and the specificity of SsXDH for $NADP^+$ (over NAD^+). That study also overexpressed a number of downstream *K. marxianus* genes from the PPP and ethanol fermentation pathways and built a strain (YZX088) that had ethanol yields (g/g) as high as 85%, with productivity that was also comparable to the best engineered *S. cerevisiae* strains. The alternative approach of expressing a xylose isomerase (XI) that directly converts xylose to xylulose with the use of co-factors has also been tried (Wang *et al.* 2013). The xylose isomerase gene, *XYLA*, from a fungus *Orpinomyces* was expressed in the background of strain NBRC 1777, replacing the endogenous *kmXYL1* and *kmXYL2* genes. The initial engineered strain did not effectively ferment xylose but following laboratory evolution a strain YRL005 that yielded 76% (g/g) of the theoretical ethanol was obtained. As was found in similar studies in *S. cerevisiae*, the evolved strain overexpressed multiple genes in the PPP. Productivity was still low relative to *S. cerevisiae* and the best pentose – fermenting *K. marxianus* strain constructed to date is strain YZX088. Unlike the engineered *S. cerevisiae* strains, all the *K. marxianus* strains described here are capable of fermentation at 42°C – 45°C.

Higher alcohols, also known as fusel alcohols, are alcohols produced from turnover of amino acids via the Ehrlich pathway (Hazelwood *et al.* 2008). While these higher alcohols are found in many fermented foods and beverages like wine, beer, bread of some yogurts, these alcohols are also widely produced synthetically as primary products as additives for food or indeed as fragrances for the perfume industry. However, increasing consumer preference for naturally derived products has created an ever-growing market niche for these higher alcohols to be of natural origin. Currently there are two methods for natural production, the first one is plant extraction and the second being microbial production. Plant extraction is main source of natural F&Fs, however it is an expensive process that can suffer from limited yield or unstable supply. Different bacteria and yeast can be used in microbial synthesis, but there are clear advantages to using a yeast like *K. marxianus,* which is non-pathogenic, food grade, and carries GRAS (US – Generally Regarded As Safe) and QPS (European

Food Safety Authority – Qualified Presumption of Safety) labelling. The Ehrlich pathway consists of three basic enzymatically catalysed reactions: first, transamination of an amino acid to a 2-oxo acid; second, decarboxylation to an aldehyde; and third, either reduction or oxidation to an alcohol or an acid respectively (Ehrlich 1907). Most knowledge of this pathway comes from studies on *S. cerevisiae,* but the core genes are conserved in *K. marxianus*. The transaminases encoded by *BAT1/BAT2* and *ARO8/ARO9* act on branched chain amino acids (leucine, isoleucine and valine) and aromatic amino acids (phenylalanine, tyrosine and tryptophan), respectively (Kispal *et al.* 1996, Iraqui *et al.* 1998). The most important decarboxylase is encoded by *ARO10* although, at least in *S. cerevisiae*, several other enzymes also have this activity. Expression of *ARO10* is strongly correlated with nitrogen source and is upregulated in the presence of branched-chain or aromatic amino acids (Vuralhan *et al.* 2003) but transcription of the transaminases does not seem to be linked with the presence of their preferred substrates (Boer *et al.* 2007). The fate of the aldehyde depends on the redox state of the cells. In *S. cerevisiae*, when glucose is limited in aerobic conditions, the aldehyde is oxidised to a fusel acid, whereas in anaerobic conditions, NADH is used to reduce the aldehyde to the corresponding alcohol such as 2-phenylethanol (phenylalanine); isoamyl alcohol (Leucine) or isobutanol (Valine) (Dickinson *et al.* 2003, Boer *et al.* 2007). There are multiple aldehyde dehydrogenases and alcohol dehydrogenases that may catalyse these reactions, and as found in a *K. marxianus* transcriptome analysis, genes encoding these enzymes are subject to complex regulation (Gao *et al.* 2015).

Presently, most work has focused on production of 2-Phenylethanol (2-PE) in K. marxianus (reviewed in (Morrissey *et al.* 2015). This flavour molecule has a rose-like smell and has a range of applications in the cosmetics and perfume industry to the food and beverage industry. The world annual production of 2-PE was estimated to be over 10,000 tons in 2010, the vast majority of which is produced from harmful carcinogens such as benzene and styrene (Clark 1990, Chen *et al.* 2011, Hua and Xu 2011). Although the amount of naturally produced 2-PE still pales in comparison to chemical synthesis the vast majority of it is produced by fermentation as plant extraction is not economically viable with prices upwards of $1,000/Kg compared to the $5/Kg price tag of synthetic production (Etschmann *et al.* 2002). This has created an interest in producing a naturally derived 2-PE at a more economically affordable price point.

Highest yields are obtained in what is essentially a bioconversion process: growth medium is supplemented with phenylalanine, the yeast coverts this to 2-phenylethanol, and the 2-PE is recovered using an *in-situ* product recovery (ISPR) system (to avoid toxicity issues). At least two companies (Puris, South Africa; Lasaffre, France) commercially produce 2-PE from *K. marxianus* but the precise details of the process remain proprietary knowledge. Higher alcohols can be further modified to esters by the addition of acyl group either acetyl coA (to yield an acetate ester) or acyl CoA (yielding an ethyl ester). These metabolites are major flavour contributors in all yeast based fermented foods and beverages and controlled production in cell factories would be very desirable. The enzyme responsible for catalysing this reaction is an alcohol acetyltransferase or AATase. In *S. cerevisiae* this activity is mostly shared between the two paralogs *ATF1* and *ATF2,* although as deletion of both of these genes does not completely abolish activity it appears that, as yet, unidentified genes also have a role to play (Verstrepen *et al.* 2003). There is very limited data available for *K. marxianus,* but one recent study found that production of different acetate esters was differentially regulated, implying that there may be more than one enzyme with aromatic acetyl transferase activity in *K. marxianus* (Gethins *et al.* 2015). This theory would be supported by some of the older *K. marxianus* literature (Kallel-Mhiri and Miclo 1993, Plata *et al.* 2003).

Alcohol acetyltransferases are also responsible for the synthesis of ethylacetate, which is formed by acetylation of ethanol. Ethyl acetate does have flavour characteristics, but the main commercial interest is as an industrial solvent. Because to its moderate polarity, ethyl acetate is an attractive industrial solvent for many applications from cleaning surfaces to extraction and chromatographic recovery of pharmaceuticals. In addition, ethyl acetate is easily degraded by bacteria, making disposal relatively straightforward. As ethyl acetate is derived from primary metabolism its synthesis is not easily increased by manipulation of carbon and nitrogen sources unlike the higher alcohol derived acetate esters. Interestingly, however it has been found that impairing the activity of the TCA cycle by limiting the availability of Iron (Fe) or Copper (Cu) dramatically increases ethyl acetate production (Loser *et al.* 2012, Urit *et al.* 2012). Also, since ethanol is the substrate for ethyl acetate production conditions that increase ethanol accumulation, most notably oxygenation, also favour the production of ethyl

acetate (Urit *et al.* 2013). The production of ethyl acetate in *K. marxianus* has been extensively reviewed recently (Loser *et al.* 2014, Loser *et al.* 2015).

1.4 Future perspectives

Oil-based chemistry processes are the most common means for synthetizing bulk and fine chemicals in our society. Although generally considered as feasible, these procedures are questionable from an environmental and economical point of view. Recently, researchers have focused on developing biotechnological processes that could serve as sustainable alternatives to chemical synthesis. Due to their robustness and versatility yeasts have become interesting candidates for cell factory applications and many examples are available in the literature (for a review see Borodina and Nielsen 2014). Although most of the work has focused on *Saccharomyces cerevisiae* there are many cases where other yeasts such as *K. marxianus* have been used. In developing a process, the starting point can either be a biodiversity screen to identify the best natural strain, or an engineering approach based on well-studied platform strains (Porro and Branduardi 2009). Increasingly, the trend is to pursue a hybrid strategy: identify natural strains with traits of interest and understand the genetic basis of the trait; reverse engineer the trait into a platform strain (inverse engineering); apply laboratory evolution to optimise the strain (Evolutionary engineering) (Crook and Alper 2012, Bachmann *et al.* 2015). In order to improve relevant aspects such as yield, substrate/product spectrum and stress tolerance, a combination of computational and genetic tools is required. First, a global view of cellular metabolism must be obtained. To accomplish this, genome-scale metabolic models, which are useful in predicting metabolic fluxes, can be constructed. Then, to obtain a more realistic overview of the system, different sets of data (e.g. transcriptomics, proteomics and metabolomics) must be integrated. Using this model, metabolic fluxes can be simulated using flux balance analysis and therefore targets for reprogramming cellular metabolism can be identified. These tools are particularly useful when trying to redirect metabolism towards a product of interest, identifying potential competing metabolic branches/pathways or optimizing cofactor usage and redox balance (Kavscek *et al.* 2015). Another key element required to utilize *K. marxianus* for cell factory applications is developing molecular tools that enable genome editing. These tools must be powerful enough to allow efficient targeted integration of DNA fragments in the genome. In *K. marxianus* these tools have not been fully developed but, as described earlier, considerable progress has been made. Future developments are also likely to involve synthetic biology. Two examples where this type of approach has

already been implemented in *K. marxianus* were the assembly of the cellulose enzyme cocktail discussed earlier (Chang *et al.* 2013) and the introduction of a pathway to synthesise hexanoic acid (Cheon *et al.* 2014). Significant tool development will be required to further progress synthetic biology in *K. marxianus.*

1.5 Conclusions

Microbial cell factories represent a feasible alternative to chemical synthesis, but more improvements are required to meet industrial standards. While yeasts are interesting candidates for many applications, the majority of the research efforts have focused in *S. cerevisiae,* which might not be suitable for all purposes. Lately, *K. marxianus* has arisen as a new potential host for cell factory applications. The development of new genetic and mathematical tools will be critical for engineering this yeast to obtain platform strains that can be exploited for different industrial applications. Although in the present scenario many technological aspects need to be addressed we can speculate about the future for *K. marxianus* cell factories. Using a systems biology approach, it will be possible to understand the complex relationships between metabolic components. By applying efficient molecular tools different genes/reactions could be removed from the system and other desired components could be integrated. As a result, a chassis strain, in which different metabolic pathways can be installed as independent circuits, could be obtained. Since modifying this strain might be simple, its construction would lead to new industrial applications for *K. marxianus*. *K. marxianus* is likely to develop as a platform, not to replace *S. cerevisiae,* but to provide opportunities to expand the range of biomolecules produced in yeast systems as a whole.

1.6 References

Abdel-Banat B.M., Nonklang S., Hoshida H. and Akada R. 2010. *Yeast* **27:** 29-39.

Amaya-Delgado L., Herrera-Lopez E.J., Arrizon J., Arellano-Plaza M. and Gschaedler A. 2013. *World J Microbiol Biotechnol* **29:** 875-881.

Bachmann H., Pronk J.T., Kleerebezem M. and Teusink B. 2015. *Curr Opin Biotechnol* **32:** 1-7.

Belloch C., Barrio E., Garcia M.D. and Querol A. 1998. *Yeast* **14:** 1341-1354.

Boer V.M., Tai S.L., Vuralhan Z., Arifin Y., Walsh M.C., Piper M.D.W., De Winde J.H., Pronk J.T. and Daran J.-M. 2007. *FEMS Yeast Research* **7:** 604-620.

Borodina I. and Nielsen J. 2014. *Biotechnol J* **9:** 609-620.

Castro R.C. and Roberto I.C. 2014. *Appl Biochem Biotechnol* **172:** 1553-1564.

Chang J.J., Ho F.J., Ho C.Y., Wu Y.C., Hou Y.H., Huang C.C., Shih M.C. and Li W.H. 2013. *Biotechnol Biofuels* **6:** 19.

Chen X.-M., Kobayashi H., Sakai M., Hirata H., Asai T., Ohnishi T., Baldermann S. and Watanabe N. 2011. *Journal of Plant Physiology* **168:** 88-95.

Cheon Y., Kim J.S., Park J.B.*, et al.* 2014. *J Biotechnol* **182-183:** 30-36.

Choo J.H., Han C., Kim J.Y. and Kang H.A. 2014. *Biotechnol Lett* **36:** 2059-2067.

Clark G.S. 1990. *Perfum Flavor* **15:** 37-44.

Crook N. and Alper H.S. 2012. Classical Strain Improvement. Engineering Complex Phenotypes in Industrial Strains,p.^pp. 1-33. John Wiley & Sons, Inc.

Dickinson J.R., Salgado L.E. and Hewlins M.J. 2003. *J Biol Chem* **278:** 8028-8034.

Diniz R.H., Rodrigues M.Q., Fietto L.G., Passos F.M. and Silveira W.B. 2014. *Biocatalysis and Agricultural Biotechnology* **3:** 111-117.

Ehrlich F. 1907. *Berichte der deutschen chemischen Gesellschaft* **40:** 1027-1047.

Etschmann M., Bluemke W., Sell D. and Schrader J. 2002. *Applied Microbiology and Biotechnology* **59:** 1-8.

Fasoli G., Tofalo R., Lanciotti R., Schirone M., Patrignani F., Perpetuini G., Grazia L., Corsetti A. and Suzzi G. 2015. *Int J Food Microbiol* **214:** 151-158.

Fonseca G.G., Heinzle E., Wittmann C. and Gombert A.K. 2008. *Appl Microbiol Biotechnol* **79:** 339-354.

Gao J., Yuan W., Li Y., Xiang R., Hou S., Zhong S. and Bai F. 2015. *Biotechnol Biofuels* **8:** 115.

Gethins L., Guneser O., Demirkol A., Rea M.C., Stanton C., Ross R.P., Yuceer Y. and Morrissey J.P. 2015. *Yeast* **32:** 67-76.

Goshima T., Tsuji M., Inoue H., Yano S., Hoshino T. and Matsushika A. 2013. *Biosci Biotechnol Biochem* **77:** 1505-1510.

Goshima T., Negi K., Tsuji M., Inoue H., Yano S., Hoshino T. and Matsushika A. 2013. *J Biosci Bioeng* **116:** 551-554.

Guimaraes P.M., Teixeira J.A. and Domingues L. 2010. *Biotechnol Adv* **28:** 375-384.

Hazelwood L.A., Daran J.M., van Maris A.J., Pronk J.T. and Dickinson J.R. 2008. *Appl Environ Microbiol* **74:** 2259-2266.

Heo P., Yang T.J., Chung S.C.*, et al.* 2013. *J Biotechnol* **167:** 323-325.

Hong J., Wang Y., Kumagai H. and Tamaki H. 2007. *J Biotechnol* **130:** 114-123.

Hoshida H., Murakami N., Suzuki A., Tamura R., Asakawa J., Abdel-Banat B.M., Nonklang S., Nakamura M. and Akada R. 2014. *Yeast* **31:** 29-46.

Hua D. and Xu P. 2011. *Biotechnol Adv* **29:** 654-660.

Iraqui I., Vissers S., Cartiaux M. and Urrestarazu A. 1998. *Mol Gen Genet* **257:** 238-248.

Jeong H., Lee D.-H., Kim S.H., Kim H.-J., Lee K., Song J.Y., Kim B.K., Sung B.H., Park J.C. and Sohn J.H. 2012. *Eukaryotic cell* **11:** 1584-1585.

Kallel-Mhiri H. and Miclo A. 1993. *FEMS Microbiology Letters* **111:** 207-212.

Kang Q., Appels L., Tan T. and Dewil R. 2014. *The Scientific World Journal* **2014**.

Kavscek M., Strazar M., Curk T., Natter K. and Petrovic U. 2015. *Microb Cell Fact* **14:** 94.

Kim J.S., Park J.B., Jang S.W. and Ha S.J. 2015. *Appl Biochem Biotechnol* **176:** 1975-1984.

Kim T.Y., Lee S.W. and Oh M.K. 2014. *Enzyme Microb Technol* **61-62:** 44-47.

Kispal G., Steiner H., Court D.A., Rolinski B. and Lill R. 1996. *J Biol Chem* **271:** 24458-24464.

Kumar S., Singh S.P., Mishra I.M. and Adhikari D.K. 2009. *Journal of industrial microbiology & biotechnology* **36:** 1483-1489.

Lachance M.-A. 2011. Kluyveromyces van der Walt. The Yeasts, Vol. 2 (Kurtzman CP, Fell JW & Boekhoet T, eds.), p.^pp. 471-482. Elsevier, Amsterdam.

Lachance M.A. 2007. *FEMS Yeast Res* **7:** 642-645.

Lane M.M. and Morrissey J.P. 2010. *Fungal Biology Reviews* **24:** 17-26.

Lane M.M., Burke N., Karreman R., Wolfe K.H., O'Byrne C.P. and Morrissey J.P. 2011. *Antonie Van Leeuwenhoek* **100:** 507-519.

Lee K.S., Kim J.S., Heo P.*, et al.* 2013. *Appl Microbiol Biotechnol* **97:** 2029-2041.

Lertwattanasakul N., Kosaka T., Hosoyama A.*, et al.* 2015. *Biotechnol Biofuels* **8:** 47.

Llorente B., Malpertuy A., Blandin G., Artiguenave F., Wincker P. and Dujon B. 2000. *FEBS Lett* **487:** 71-75.

Lopez C.L., Beaufort S., Brandam C. and Taillandier P. 2014. *World J Microbiol Biotechnol* **30:** 2223-2229.

Lopez-Alvarez A., Diaz-Perez A.L., Sosa-Aguirre C., Macias-Rodriguez L. and Campos-Garcia J. 2012. *J Biosci Bioeng* **113:** 614-618.

Loser C., Urit T. and Bley T. 2014. *Appl Microbiol Biotechnol* **98:** 5397-5415.

Loser C., Urit T., Keil P. and Bley T. 2015. *Appl Microbiol Biotechnol* **99:** 1131-1144.

Loser C., Urit T., Forster S., Stukert A. and Bley T. 2012. *Appl Microbiol Biotechnol* **96:** 685-696.

Mans R., van Rossum H.M., Wijsman M., Backx A., Kuijpers N.G., van den Broek M., Daran-Lapujade P., Pronk J.T., van Maris A.J. and Daran J.M. 2015. *FEMS Yeast Res* **15:** fov004.

Morrissey J.P., Etschmann M.M.W., Schrader J. and de Billerbeck G.M. 2015. *Yeast* **32:** 3-16.

Nonklang S., Ano A., Abdel-Banat B.M., Saito Y., Hoshida H. and Akada R. 2009. *Biosci Biotechnol Biochem* **73:** 1090-1095.

Pecota D.C. and Da Silva N.A. 2005. *Biotechnol Bioeng* **92:** 117-123.

Pecota D.C., Rajgarhia V. and Da Silva N.A. 2007. *J Biotechnol* **127:** 408-416.

Plata C., Millán C., Mauricio J.C. and Ortega J.M. 2003. *Food Microbiology* **20:** 217-224.

Porro D. and Branduardi P. 2009. *Microb Cell Fact* **8:** 51.

Radecka D., Mukherjee V., Mateo R.Q., Stojiljkovic M., Foulquie-Moreno M.R. and Thevelein J.M. 2015. *FEMS Yeast Res* **15:** fov053.

Ribeiro O., Gombert A.K., Teixeira J.A. and Domingues L. 2007. *J Biotechnol* **131:** 20-26.

Rocha S.N., Abrahao-Neto J. and Gombert A.K. 2011. *Antonie Van Leeuwenhoek* **100:** 619-630.

Rodrussamee N., Lertwattanasakul N., Hirata K., Suprayogi, Limtong S., Kosaka T. and Yamada M. 2011. *Appl Microbiol Biotechnol* **90:** 1573-1586.

Shrivastav M., De Haro L.P. and Nickoloff J.A. 2008. *Cell research* **18:** 134-147.

Suryawati L., Wilkins M.R., Bellmer D.D., Huhnke R.L., Maness N.O. and Banat I.M. 2008. *Biotechnology and bioengineering* **101:** 894-902.

Suzuki A., Fujii H., Hoshida H. and Akada R. 2015. *FEMS Yeast Res* **15**.

Tsakraklides V., Brevnova E., Stephanopoulos G. and Shaw A.J. 2015. *PLoS One* **10:** e0133434.

Urit T., Stukert A., Bley T. and Loser C. 2012. *Appl Microbiol Biotechnol* **96:** 1313-1323.

Urit T., Manthey R., Bley T. and Löser C. 2013. *Engineering in Life Sciences* **13:** 247-260.

Verstrepen K.J., Derdelinckx G., Dufour J.-P., Winderickx J., Pretorius I.S., Thevelein J.M. and Delvaux F.R. 2003. *FEMS Yeast Research* **4:** 285-296.

Vuralhan Z., Morais M.A., Tai S.L., Piper M.D. and Pronk J.T. 2003. *Appl Environ Microbiol* **69:** 4534-4541.

Wang R., Li L., Zhang B., Gao X., Wang D. and Hong J. 2013. *J Ind Microbiol Biotechnol* **40:** 841-854.

Yanase S., Hasunuma T., Yamada R., Tanaka T., Ogino C., Fukuda H. and Kondo A. 2010. *Appl Microbiol Biotechnol* **88:** 381-388.

Yang C., Hu S., Zhu S., Wang D., Gao X. and Hong J. 2015. *World J Microbiol Biotechnol* **31:** 1641-1646.

Yarimizu T., Nonklang S., Nakamura J.*, et al.* 2013. *Yeast* **30:** 485-500.

Zhang B., Li L., Zhang J., Gao X., Wang D. and Hong J. 2013. *J Ind Microbiol Biotechnol* **40:** 305-316.

Zhang B., Zhang J., Wang D., Gao X., Sun L. and Hong J. 2015. *Data Brief* **5:** 179-186.

Zhang J., Zhang B., Wang D., Gao X., Sun L. and Hong J. 2015. *Metab Eng* **31:** 140-152.

Chapter 2

Development of methodologies for targeted mutagenesis in *Kluyveromyces marxianus*

Material from this chapter was published in the following paper. This includes Figure 2.1, Figure 2.3 and Figure 2.4

Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. Juergens H., **Varela J.A**, Gorter de Vries A. R, Gast V.G.M., Gyurchev N.Y., Rajkumar A.S., Mans R., Pronk J.T, Morrissey JP, Daran J.G. *FEMS Yeast Res.* 2018. May 1;18(3). DOI:10.1093/femsyr/foy012

Abstract

Kluyveromyces marxianus has recently emerged as an attractive organism for cell factory applications. Its high growth rate, thermotolerance and its capacity to utilise a wide arrange of substrates make it an ideal host for several biotechnological applications including bioethanol and bioflavour molecule production. Unfortunately, advances made in the area have been hampered by the lack of standardised molecular tools that allow precise modification of the yeast genome. In this study, the CRISPR-Cas9 system was implemented in *K. marxianus.* A Cas9/gRNA single plasmid expression system was developed and used to target the *ADE2* gene in *K. marxianus* where targeting efficiencies of 96% were obtained. The system was also successfully tested in *K. marxianus* sister species, *K. lactis*, where the *ADE2* locus was also targeted. The Cas9/gRNA plasmid was further modified by introducing a type IIS restriction enzyme cloning site that allows fast cloning of target sequences without compromising the functionality of the system. To enhance the frequency of homologous recombination-mediated repair in *K. marxianus*, the *KU80* gene was disrupted using the CRISPR-Cas9 system. The *ku80* mutant showed an increase in homologous repair suggesting that the non-homologous end-joining pathway, responsible for random DNA integration, had been inactivated in the strain. In order to further improve homologous recombination, the *RAD52* recombinase from *S. cerevisiae* was overexpressed in *K. marxianus*. A strain carrying the *KU80* mutation and overexpressing the *RAD52* gene showed a seven-fold increase in homologous recombination-mediated repair. The tools developed in this study provide a useful framework to genome engineer *K. marxianus*.

2.1 Introduction

Kluyveromyces lactis and *Kluyveromyces marxianus* are non-conventional yeast species commonly found in milk and fermented dairy products such as kefir and cheese (Gethins *et al*., 2016; Lopandic *et al*., 2006). Due to their association with these food products both yeasts hold a generally regarded as safe (GRAS) status that allows their use in food and pharmaceutical applications. While both species are closely related, *K. lactis* has received more attention and is considered a model for non-conventional yeasts (Rodicio *et al*., 2013). In contrast, *K. marxianus* has not been studied at the same extent but has been used in a number of biotechnological applications that include the production of ethanol from cheese whey, production of biomass and production of enzymes such as β-galactosidase and inulinase (Fonseca *et al*., 2008; Lane and Morrissey, 2010). *K. marxianus* also has other features of industrial interest that are not present in *K. lactis*. *K. marxianus* is thermotolerant, being able to grow up to temperatures of 50°C, it utilises a wide variety of substrates and has the fastest growth rate among eukaryotes with a doubling time of 45-70 minutes (Fonseca *et al*., 2008; Lane *et al*., 2011). The possibility of exploiting these traits and developing *K. marxianus* into a microbial cell-factory for diverse applications has motivated researchers to study its physiology in more detail. Several studies comparing different strains are available (Fonseca *et al*., 2007; Lane *et al*., 2011; Rocha *et al*., 2011) and numerous *K. marxianus* genomes have been sequenced (Inokuma *et al*., 2015; Jeong *et al*., 2012; Lertwattanasakul *et al*., 2015; Silveira *et al*., 2014; Suzuki *et al*., 2014). However, the limited availability of genetic tools to manipulate this yeast has stalled the advances made in the area. One of the main obstacles in developing efficient genetic tools in this yeast is the non-homologous end-joining (NHEJ) system (Yarimizu *et al*., 2013). NHEJ is a DNA repair mechanism that repairs double-strand breaks (DSBs) in the DNA without the need of a homologous DNA sequence (Daley *et al*., 2005). In the NHEJ system the heterodimer Ku70/Ku80 binds to the DNA ends and promotes binding of the Lig4/Lif1 complex (Palmbos *et al*., 2008, 2005). Lif1 stabilises the Lig4 DNA ligase that joins the ends of the two DNA molecules together (Ma *et al*., 2004). When a DSB occurs, the NHEJ system repairs the damage by reconnecting the ends of the two compatible DNA molecules, however if an additional linear DNA sequence is present, the system can ligate this molecule to one of the DBS ends creating an

imprecise repaired molecule (Daley *et al*., 2005). This phenomenon could lead to random integration of DNA fragments into the genome (Iiizumi *et al*., 2008). A DSB can also be repaired via homologous recombination (HR) when a homologous DNA sequence is present. Though this system could be used to target ectopic DNA to a desired location in the genome, in *K. marxianus* NHEJ is highly active and predominates over HR making gene targeting challenging (Abdel-Banat *et al*., 2009). It is worth taking into consideration that *K. lactis* also has an active NHEJ system, but in this case several genetic tools that ensure efficient gene targeting have been developed to solve the issue (Heinisch *et al*., 2010; Kooistra *et al*., 2004; Tsakraklides *et al*., 2015; Wésolowski-Louvel, 2011).

Current advances in molecular tool development in *K. marxianus* can be classified according to their dependence on the NHEJ and HR repair systems. The techniques that rely on the NHEJ repair system are mainly focused on randomly integrating genes into the yeast genome or using this repair system as gene cloning tool. For example, random gene integration has been used to complement *K. marxianus* auxotrophic mutants with *S. cerevisiae* genes (Nonklang *et al*., 2008; Yarimizu *et al*., 2013). The system has also been used to construct strains for biotechnological applications. A *K. marxianus* strain able to produce hexanoic acid has been constructed by randomly integrating the hexanoic acid biosynthesis pathway comprised of seven genes from bacteria and yeast (Cheon *et al*., 2014). Also, random integration of *S. cerevisiae* genes into *K. marxianus* genome has been performed to obtain a flocculent *K. marxianus* strain that can be used in high-temperature fermentations (Nonklang *et al*., 2009). The feasibility of using NHEJ as a DNA cloning tool has also been studied. A linearised plasmid carrying a truncated version of a *KanMX* and *URA3* selection markers can be joined by NHEJ to a DNA fragment containing flaking regions that functionally complement the markers (Hoshida *et al*., 2014). As cell survival depends on a NHEJ event taking place the cloning method is highly efficient. The same method has been used to construct promoter fusions by cloning promoter sequences next to a fluorescent reporter gene and integrating the resulting construct into a random position in the genome (Suzuki *et al*., 2015). Although these methods can be useful for several applications, additional methods that allow performing precise genetic modifications have also been developed. *K. marxianus* strains lacking NHEJ have been constructed by disrupting the *KU70* or

KU80 genes (Abdel-Banat *et al*., 2009; Choo *et al*., 2014). In these strains random gene integration was completely abolished, and HR-mediated repair was greatly enhanced with frequencies over 80 % for both *ku70* and *ku80* mutants. The size of the flanking regions used in these experiments was found to correlate inversely with transformation efficiency (only a few colonies were obtained when using short 50-60 bp homology arms). The Cre-*lox*P system has also been implemented in *K. marxianus* as a tool to disrupt multiple genes, however the efficiency of this tool depends on using long homology arms to ensure high HR frequency (Ribeiro *et al*., 2007). Overall, the different tools developed to date allow manipulating *K. marxianus* genetically but additional strategies to perform gene targeting at high efficiency are still needed. These tools should preferably work on different strains without the need of previous genetic modifications (i.e. generating *ku70/ku80* mutants) and should allow simple and reliable gene targeting regardless of strain ploidy. Recently it has been demonstrated that the introduction of a DSB in the targeted DNA locus can strongly facilitate genetic engineering, either by the introduction of mutations at the cut site or by stimulating the occurrence of HR-mediated DNA repair with co-transformed repair fragments (Jasin and Rothstein, 2013; Kuijpers *et al*., 2013).

Over the past five years, the CRISPR-Cas9 system has emerged as a powerful and versatile tool to engineer the genomes of a wide range of organisms (Hsu *et al*., 2014). In this system, the endonuclease CRISPR associated protein 9 (Cas9) binds a guide RNA molecule (gRNA) that targets a sequence-specific site in the genome (Jinek *et al*., 2012). The Cas9-gRNA complex then induces a DSB that is lethal unless repaired. Repair of DSBs typically occurs through NHEJ or HR, depending on the presence of a repair DNA fragment and the predominant DSB repair mechanism of the host cell (Shrivastav *et al*., 2008). The CRISPR-Cas9's ability to induce mutations in the target sequence has been widely exploited in the development of genetic tools for various non-conventional yeasts including *K. lactis* (Horwitz *et al*., 2015), *K. marxianus* (Löbs *et al*., 2017; Nambu-Nishida *et al*., 2017), *O. polymorpha* (Numamoto *et al*., 2017), *Pichia pastoris* (Weninger *et al*., 2016), *Scheffersomyces stipitis* (Cao *et al*., 2017) and *Yarrowia lipolytica* (Gao *et al*., 2016; Schwartz *et al*., 2016). However, the plasmid and Cas9/gRNA expression systems available today have generally only been designed for and tested in a single

yeast species, thus limiting the potential of each system to work on different yeasts with novel characteristics. Development of universal genetic tools is necessary to expand the range of species that can be used in biotechnological applications. Also, such tools could provide a framework for inter-species studies where a set of genetic modifications need to be transferred from one species to another. Here we have developed a wide-host-range CRISPR-Cas9 system for use across several yeast species including *K. marxianus* and *K. lactis.* The system was further refined and then used in combination with other tools to boost homologous recombination repair in *K. marxianus*.

2.2 Materials and methods

Strains and growth conditions

The *K. lactis* and *K. marxianus* strains used in this study are listed in **Table 2.1**. For cultivation under non-selective conditions, strains were grown in 500 mL shake flasks containing 100 mL YPD medium (10 g L^{-1} Bacto yeast extract, 20 g L^{-1} Bacto peptone, 20 g L^{-1} glucose, demineralized water), placed in a rotary shaker set to 30°C and 200 rpm. For antibiotic selection, YPD medium was supplemented with 200 µg mL^{-1} hygromycin B and 2% (w/v) agar. For *LAC4* gene targeting experiments cells were plated on YPGal agar medium (10 g L^{-1} Bacto yeast extract, 20 g L^{-1} Bacto peptone, 20 g L^{-1} galactose, 2% (w/v) agar, demineralized water) supplemented with 200 µg mL^{-1} hygromycin B and 40 µg mL^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). For homologous recombination efficiency experiments cells were plated on YPD medium supplemented with 200 µg mL^{-1} G418 or 50 µg mL^{-1} Nourseothricin sulfate (clonNAT) and then replica plated onto YPD medium containing 3 mg ml^{-1} 5-Fluoroorotic acid (5-FOA). Frozen stock cultures were prepared from exponentially growing shake-flask cultures by addition of 30% (v/v) glycerol, and aseptically stored in 1 mL aliquots at -80°C.

Table 2.1: *Kluyveromyces* strains used in this study. CBS 2359, CBS 5795 and CBS 397 were obtained from the CBS-KNAW fungal collection ([A] Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands). NBRC 1777 was obtained from the NBRC culture collection ([B] National Institute of Technology and Evaluation, Tokyo, Japan).

Species	Strain	Genotype	Origin
S. cerevisiae	S288C	*MATα SUC2 gal2 mal2 mel flo1 flo8-1 hap1 ho bio1 bio6*	CBS-KNAW[A]
K. lactis	CBS 2359	Wild type	CBS-KNAW[A]
K. marxianus	CBS 5795	Wildtype	CBS-KNAW[A]
	CBS 6556	Wildtype	CBS-KNAW[A]
	CBS 608	Wildtype	CBS-KNAW[A]
	NBRC 1777	Wildtype	NBRC[B]
	YBL001	NBRC1777 *ΔKU80*	This study
	YBL002	NBRC 1777 *ΔKU80 lac4::RAD52*	This study
	CBS 397	Wildtype	CBS-KNAW[A]

Construction of CRISPR-Cas9 cloning vectors

All plasmids used in this study are described in **Table 2.2**. For the construction of pUDP002, an intermediate plasmid containing the *Klebsiella pneumoniae hph* (Hyg^R) open reading frame (ORF), expressed from the *TEF1* promoter and terminator from the yeast *Ashbya gossypii* (*Eremothecium gossypii*), was constructed by Gibson assembly (Gibson *et al*., 2009) from plasmids pYTK079 (Lee *et al*., 2015) and pUD527: the *hph* ORF was amplified from pYTK079 using primers 9837 and 9838, and inserted into a plasmid backbone which was generated by PCR amplification of pUD527 using primers 9839 and 9840. Plasmid pUDP002 was constructed by Gibson assembly from five overlapping fragments, using synthetic homologous recombination sequences (Kuijpers *et al*., 2013): (i) the *AgTEF1p-hph-AgTEF1t* hygromycin resistance cassette was amplified from the intermediate plasmid using primers 9841 and 9842 and sequence-verified before further use; (ii) the *AaTEF1p-cas9*$^{D147Y\ P411T}$*-ScPHO5t* expression cassette (Bao *et al*., 2015; Gorter de Vries *et al*., 2017) was obtained by amplification of pUD423 using primers 3841 and 9393; (iii) the pangenomic yeast replication origin panARS(OPT) (Liachko and Dunham, 2014) was amplified from pUD530 using primers 9663 and 3856; (iv) the empty *ScTDH3p-ScCYC1t* gRNA expression cassette was obtained by digestion of pUD531 with SmaI; (v) the *Escherichia coli* pBR322 origin and *bla* gene for ampicillin selection were obtained by digestion of pUD532 with SmaI. The five fragments were gel-purified and 0.1 pmol of each fragment was assembled in a Gibson assembly reaction. An *E. coli* clone containing the correctly assembled plasmid (verified by digestion with PdmI) was stocked and pUDP002 (Addgene Plasmid #103872) was used as gRNA entry plasmid for subsequent plasmid construction. Plasmids pUDP002, pUDP025 and pUDP082 were deposited at Addgene (https://www.addgene.org/).

For constructing the pUCC001 plasmid, that allows cloning target sequences without the need of synthetic blocks, long overlapping oligonucleotides (HH-BsaI-HDV-F and HH-BsaI-HDV-R) encoding the hammer head ribozyme, a BsaI cloning site, the gRNA structural part and the HDV ribozyme were chemically synthesised and annealed in a thermocycler machine. In parallel, the pUDP002 plasmid was amplified by PCR using the pUDP002-F and pUDP002-R primers. The 205 bp DNA-duplex and the plasmid backbone were then assembled via Gibson assembly.

Table 2.2: Plasmids used in this study. Restriction enzyme sites are indicated in superscript and gRNA target sequences are indicated in subscript. SHRs represent specify synthetic homologous recombination sequences used for plasmid assembly. The addgene plasmid code (when relevant) is indicated next to the plasmid name between brackets. Plasmids are listed in order of appearance in the Materials and Methods section.

Name (Addgene Plasmid #)	Relevant characteristics	Origin
pYTK079	Template for *hph* (Hyg^R) open reading frame	(Lee *et al.* 2015)
pUD423	Template for *AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t* cassette	(Gorter de Vries *et al.* 2017)
pUD527	*ori kan*^{R} SHRA *AgTEF1p-amdS-AgTEF1t* SHRB	GeneArt™
pUD530	*ori kan*^{R} panARS(OPT)	GeneArt™
pUD531	*ori kan*^{R} SHRC *ScTDH3p*-$^{BsaI\ BsaI}$*-ScCYC1t* SHRI	GeneArt™
pUD532	*ori kan*^{R} SHRI *ori bla* SHRA	GeneArt™
pUD555	*ori amp*$^{R\ BsaI}$HH-gRNA$_{KlADE2}$-HDV^{BsaI}	GeneArt™
pUDP002 (#103872)	*ori amp*^{R} panARS(OPT) *AgTEF1p-hph-AgTEF1t ScTDH3p*$^{BsaI\ BsaI}$*ScCYC1t AaTEF1p-Spcas9*$^{D147Y\ P411T}$*-ScPHO5t*	This study
pUCC001	[pUDP002] HH-$^{BsaI\ BsaI}$-gRNAscafffold-HDV	This study
pUDP025 (#103874)	[pUDP002] gRNA$_{KlADE2}$	This study
pUDP082 (#103875)	[pUDP002] gRNA$_{KmADE2}$	This study
pUCC001-*KmADE2*	[pUCC001] gRNA$_{KmADE2}$	This study
pUCC001-*KmLAC4*	[pUCC001] gRNA$_{KmLAC4}$	This study
pUCC001-*KmKU80*	[pUCC001] gRNA$_{KmKU80}$	This study
pYTK078	Template for *clonNAT* ($clonNAT^R$) open reading frame	(Lee *et al.* 2015)
pYTK095	Template for *amp*^{R} and *colE1* for p*URA3*-clonNAT	(Lee *et al.* 2015)
p*URA3-KanMX*	*ori amp*^{R} *5'URA3 KanMX* 3'*URA3*	This study
p*URA3-NatMX*	*ori amp*^{R} *5'URA3 ClonNAT* 3'*URA3*	This study

Name (Addgene Plasmid #)	Relevant characteristics	Origin
pYTK001	Storage vector for cloning gene parts	(Lee *et al.* 2015)
pI1L	[pYTK001] *LAC4* 5′	This study
pI1Rbis	[pYTK001] *LAC4* 3′	This study
pT1	[pYTK001] *INU1p*	This study
pYTK002	Left-side spacer sequence for pI1P22-*RAD52* plasmid	(Lee *et al.* 2015)
pYTK030	Template for *GAL1p* for pI1P22-*RAD52* construction	(Lee *et al.* 2015)
pYTK067	Right-side spacer sequence for pI1P22-*RAD52* plasmid	(Lee *et al.* 2015)
pYTK077	Template for *KanMX* (G418^R) open reading frame	(Lee *et al.* 2015)
pYTK078	Template for *NatMX* (clonNAT^R) open reading frame	(Lee *et al.* 2015)
pYTK083	Template for *amp^R* and *colE1* for pI1P22-*RAD52* construction	(Lee *et al.* 2015)
pI1P22-*RAD52*	*ori amp^R 5'LAC4 ScGAL1p-ScRAD52-KmINU1t KanMX 3'LAC4*	This study
pIL75	*ori amp^R KanMX*	(Liachko and Dunham, 2014)

Design of gRNA target sequences and construction of CRISPR-gRNA plasmids

All 20 bp Cas9 gRNA target sequences used in this study are described in **Table 2.3**. For designing gRNA sequences for targeting the *ADE2* gene in *K. lactis* and *K. marxianus* approximately ten candidate target sequences were chosen from the first third of each targeted ORF, based on the presence of an NGG protospacer adjacent motif (PAM) site. Any target sequence with off-targets (defined as a sequence with NGG or NAG PAM and more than 15 nucleotides identical to the candidate sequence) was excluded, based on a blastn homology search against the respective yeast genome (https://blast.ncbi.nlm.nih.gov). The remaining target sequences were ranked based on AT content (‘AT score’; 0 being the lowest and 1 being the highest possible AT content) and secondary structure as predicted with the complete gRNA sequence, using minimum free energy prediction by *RNAfold* (Lorenz *et al.*, 2011) (‘RNA score’; 0 being the lowest and 1 being the highest possible number of unpaired target sequence nucleotides). Finally, the target sequences with the highest

cumulative AT and RNA score that did not exceed an AT score of 0.8 were chosen for use in this study. The gRNA expression plasmid pUDP025 containing the *ADE2* target sequence for *K. lactis* (Plasmid #103874) was assembled in a one-pot Golden Gate assembly (Engler *et al*., 2008) by BsaI digestion of pUDP002 and the donor plasmid pUD555, containing the respective gRNA sequence flanked by ribozymes and BsaI restriction sites. The Golden Gate assembly was carried out in a final volume of 20 µL, using T4 DNA ligase buffer (Thermo Fisher Scientific), 10 U BsaI (New England Biolabs), 1 U T4 DNA ligase (Thermo Fisher Scientific) and 10 ng of both pUDP002 and pUD555. For the construction of pUDP082 (Plasmid #103875), the partially overlapping primers Km-ade2-F and Km-ade2-R were annealed, extended and amplified by PCR to yield a 233 bp double-stranded DNA fragment containing the *KmADE2* gRNA target sequence flanked by ribozymes and BsaI sites, which was integrated into pUDP002 by Golden Gate assembly as described above. All constructed gRNA-harboring plasmids were verified by digestion with PdmI.

Design of the gRNA sequences targeting the *KU80*, *LAC4* and *URA3* genes in *K. marxianus* was carried out using the software sgRNA*cas9* (Xie *et al*., 2014). Complementary oligonucleotides containing the target sequences and specific overhangs were annealed, phosphorylated and cloned into pUCC001 by Golden Gate assembly, as described previously (Ng and Dean, 2017; Vyas *et al*., 2015). Presence of the target sequences in the pUCC001 plasmid was corroborated by colony PCR using the forward target primer and BSA-R.

Table 2.3: gRNA target sequences used in this study. Target sequences are shown including protospacer adjacent motif (PAM) sequence (underlined). Position in open reading frame (ORF) indicates the base pair after which the Cas9-mediated double-strand break is expected, out of the total length of the ORF.

Locus	Target sequence (5′-3′)	Position in ORF (bp)
KlADE2	TTTCAATACCTCAAGTGCTTCGG	508/1710
KmADE2	GCCCATTTTTCTGCGTATAGCGG	537/1710
KmLAC4	GACATCTCTTAGGACAAGTTGGG	1387/3078
KmKU80	CCTCCTCAGCCATCGCCAGCCGG	1271/1884
KmURA3	AGGTTCTTTCGTAACTTCCTCGG	412/803

Construction of *ADE2* repair fragments

The *ADE2* repair DNA fragments of *K. lactis* (962 bp) and *K marxianus* (959 bp) were constructed from strains CBS 2359 and CBS 5795 respectively, using primer sets 10723 & 10724 and 10725 & 10726 (*K. lactis*) and 10727 & 10728 and 10729 & 10730 (*K. marxianus*) to amplify the homology regions flanking *ADE2*. Both regions were then joined by overlap extension PCR using primer sets 10723 & 10726 and 10727 & 10730, in the case of *K. lactis* and *K. marxianus*, respectively. In both cases, the final repair fragment was gel-purified and further amplified to obtain quantities required for transformation, using primer sets 10723 & 10726 and 10727 & 10730 for *K. lactis* and *K. marxianus*, respectively. The primers pairs were designed to amplify the 480 bp (± 1 or 2 bp in some cases) upstream the ATG or the 480 bp terminator region downstream the stop codon of the interrupted/deleted ORF.

Construction of *URA3-KanMX* and *URA3-NatMX* cassettes

The p*URA3-KanMX* and p*URA3-NatMX* plasmids were constructed using the yeast toolkit as described by Lee *et al*., 2015. Briefly, 1 kb upstream and downstream regions of the *URA3* gene were amplified by PCR with the primers US_URA3-F/US_URA3-R and DS_URA3-F/DS_URA3-R. The products were purified and combined with the pYTK095 (containing the ampicillin resistance marker and colE1 for *E. coli* maintenance) and the pYTK077 or pYTK078 plasmids (containing the *KanMX* and *NatMX* resistance markers, respectively) in a Golden Gate assembly reaction. After assembly and transformation, white colonies were screened by colony PCR using primer pairs ASR_K001F/US_URA3-R and ASR_K011F/DS_URA3-R. To obtain the linear the fragments targeting the *URA3* gene, 10 ng of the p*URA3-KanMX* or p*URA3-NatMX* plasmid were used for a PCR reaction using the primers US_URA3-F and DS-URA3-R. The PCR product was quantified using a nanodrop machine.

Construction of a *lac4::RAD52 K. marxianus* strain

To construct the pI1P22-*RAD52* plasmid, sequences upstream and downstream of the gRNA targeting *LAC4* were respectively amplified from CBS 6556 genomic DNA by primers ASR_I001LF/I001LR and ASR_IL001RFbis/RRbis with Q5 DNA polymerase. These were subsequently cloned into the yeast toolkit plasmid pYTK001 by Golden Gate assembly using BsmBI following the protocol described

in (Lee *et al*., 2015) using 20 fmol of amplified part and plasmid to generate plasmids pI1L and pI1Rbis. Correct assembly was verified by colony PCR after transformation using primers ASR_K001F and ASR_K002R and sequence verification of the insert in plasmids extracted from PCR-positive colonies.

While pI1Rbis used BsaI overhangs suited for a Yeast Toolkit plasmid used for a yeast marker (part type 6), pI1L and the *GAL1* promoter downstream of it required new 3' and 5' BsaI overhangs, respectively. The *GAL1* promoter was amplified from pYTK030 with primers ASR_P022Fbis/R with Q5 DNA polymerase. 40 fmol of each plasmid (pYTK002, pI1L, p*RAD52*, pT1, pYTK067, pYTK077, and pI1Rbis), the amplified *GAL1* promoter and 20 fmol of pYTK083 were used in a Golden Gate assembly using BsaI. pT1 was created by amplifying the *INU1* promoter from CBS 6556 genomic DNA using primers ASR_T001F and ASR_T001Rc, and cloning that into pYTK001 by BsmBI Golden Gate assembly as for pI1L and pI1Rbis. Following transformation into *E. coli* and red/white selection, correct assembly of the *RAD52* plasmid was verified by colony PCR using primer pairs ASR_K001F/ASR_P0022R and ASR_T001F/ASR_M001R. To obtain the *LAC4-RAD52* as a linear fragment, 2 µg of pI1P22-*RAD52* were digested with NotI and transformed into YBL001.

Yeast transformations and molecular diagnosis of yeast mutants

K. lactis and *K. marxianus* strains were transformed using the LiAc/single-strand carrier DNA/polyethylene glycol method (Gietz and Schiestl, 2007). Overnight pre-cultures in YPD medium were used to inoculate a shake flask containing YPD medium to an initial OD_{660nm} of 0.5. Cultures were then incubated at 30˚C until an OD_{660nm} of 2.0 was reached, harvested and transformed with 200 ng of plasmid DNA and 300 or 1000 ng of repair DNA fragment in the case of *K. lactis* or *K. marxianus*, respectively. For homologous recombination experiments, cells were inoculated to an initial OD_{660nm} of 0.1 and harvested at OD_{660nm} 0.8. The cultures were then transformed with 2 µg of the *URA3-KanMX* or *URA3-NatMX* cassettes. After heat shock, cells were recovered in 1 mL YPD medium for 3 h and plated on selective YPD or YPGal medium, as mentioned in the text. For *ADE2* experiments plates were typically kept for 5 d at 30°C and then incubated at 4°C for 2 h before assessing the

percentage of red Ade$^-$ colonies. Plates were kept at the same temperature for 2 days for all the rest of experiments.

Molecular diagnosis of yeast mutants

For molecular analysis of *K. lactis* and *K. marxianus* mutants, colonies were grown overnight in YPD medium. Genomic DNA was extracted using the method described by (Singh and Weil, 2002), and used as a template for PCR reactions targeting the *ADE2* locus. Primer sets 10909 & 10910, and 10911 & 10912 were used for *K. lactis* and *K. marxianus*, respectively. For Sanger sequencing of putative NHEJ corrected mutants, PCR products were purified and then sequenced using primers 10737 and SeqADE2 for *K. lactis* and *K. marxianus*, respectively. Screening of *ku80* mutants was carried out following the same procedure and using primers Diag*KU80*-F and Diag*KU80*-R. Correct integration of the *RAD52* fragment into the *LAC4* gene was checked by PCR using the primers K11-F and I1DS-R.

Molecular biology techniques used to construct the pUDP002 system

PCR amplification with Phusion High Fidelity Polymerase (Thermo Fisher Scientific) or Q5 High-Fidelity DNA polymerase (New England Biolabs) was performed according to the manufacturer's instructions using PAGE-purified oligonucleotide primers (Sigma-Aldrich, St. Louis, MO). Diagnostic PCR was done using DreamTaq polymerase (Thermo Fisher Scientific) and desalted primers (Sigma-Aldrich). All primer sequences are shown in **Table 2.4**. DNA fragments obtained by PCR were separated by gel electrophoresis. Gel purification was carried out using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA). PCR purification was performed using either the GenElute™ PCR Clean-Up Kit (Sigma-Aldrich) or GeneJET PCR purification kit (Thermo Fisher Scientific). Gibson assembly was done using the NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, Ipswich, MA) according to manufacturer's recommendations. Restriction digest with PdmI and SmaI was performed using FastDigest enzymes (Thermo Fisher Scientific), according to manufacturer's instructions. *E. coli* strain XL1-blue was used for plasmid transformation, amplification and storage. Plasmids were isolated from *E. coli* with the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich).

Table 2.4: Primers used in this study

Name	Sequence (5' - 3')
pUDP002 and pUCC001 construction	
9837	ATACAGTTCTCACATCACATCCGAACATAAACAAGGATCCATGGGTAAAAAGCCTGAACTC
9838	ACAAGTTCTTGAAAACAAGAATCTTTTTATTGTCCTCGAGTTATTCCTTTGCCCTCGGAC
9839	CTCGAGGACAATAAAAAGATTCTTG
9840	GGATCCTTGTTTATGTTCGGATG
9841	ACTATATGTGAAGGCATGGC
9842	GTTGAACATTCTTAGGCTGG
3841	CACCTTTCGAGAGGACGATG
9393	TGCCGAACTTTCCCTGTATGAAGCGATCTGACCAATCCTTTGCCGTAGTTTCAACGTATGTTTTCATTTTTGCGATGCCAG
9663	CATACGTTGAAACTACGGCAAAGGATTGGTCAGATCGCTTCATACAGGGAAAGTTCGGCATCAACATCTTTGGATAATATCAGAATGAG
3856	CTAGCGTGTCCTCGCATAGTTC
HH-BsaI-HDV-F	CTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCGAGACCTGCGGAGGTCTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
HH-BsaI-HDV-F	GTCCCATTCGCCATGCCGAAGCATGTTGCCCAGCCGGCGCCAGCGAGGAGGCTGGGACCATGCCGGCCAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTG
pUDP002-F	TCGGACGAGCTTACTCGTTTCGTCCTCACGGACTCATCAGGTTTGTTTGTTTATGTGTGTTTATTC
pUDP002-R	GCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACCACAGGCCCCTTTTCCTTTG

Name	Sequence (5' - 3')
Km-ade2-F	GGTCTCGCAAAGTCAAGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCCCATTTTTCTGCGTATAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
Km-ade2-R	GGTCTCGCTGTGTCCCATTCGCCATGCCGAAGCATGTTGCCCAGCCGGCGCCAGCGAGGAGGCTGGGACCATGCCGGCCAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGC
gRNA sequences for pUCC001	
ADE2-F	CGTCGCCCATTTTTCTGCGTATAG
ADE2-R	AAACCTATACGCAGAAAAATGGGC
LAC4-F	CGTCGACATCTCTTAGGACAAGTT
LAC4-R	AAACAACTTGTCCTAAGAGATGTC
KU80-F	CGTCCCTCCTCAGCCATCGCCAGC
KU80-R	AAACGCTGGCGATGGCTGAGGAGG
URA3-F	CGTCAGGAAGTTACGAAAGAACCT
URA3-R	AAACAGGTTCTTTCGTAACTTCCT
Construction of *ADE2* repair fragments	
10723	GTAGTACCGACCTTATCCGTG
10724	GTTGTCTTAGTGAAGAAGGTGAAC
10725	TATATAATAACATCACGTTCACCTTCTTCACTAAGACAACAGCTGCCAAATTAGAAACTATCG
10726	TGTGCGTTGATATATGCCAAC
10727	ATCATAGACAGTCAGTTAGTTCCC

Name	Sequence (5' - 3')
10728	TTCTTTGGTCCATGATTAACAAGG
10729	ACTACAACAATATAAACCTTGTTAATCATGGACCAAAGAAGTATTCAACTACCTCCAACAAGAAG
10730	CAAATTTATGAAGTTTGTGCCATTTG
Construction of the p*URA3-KanMX* and p*URA3-NatMX* plasmids	
US_URA3-F	GCATCGTCTCATCGGTCTCACCCTCTGCTAGAAATCCAGACCTTAAGT
US_URA3-R	ATGCCGTCTCAGGTCTCATGTAGGGGCTAGCTGATGATACTTC
DS_URA3-F	GCATCGTCTCATCGGTCTCAGAGTCGAGTGTACTCGGATCAGAAG
DS_URA3-R	ATGCCGTCTCAGGTCTCATCGGCCTAGTTGAGTGCAATGCAG
ASR_K001F	TTTGCTGGCCTTTTGCTC
ASR_K011F	AGAAAGTAATATCATGCGTCAA
Construction of the *lac4::RAD52* strains	
ASR_I001LF	GCATCGTCTCATCGGTCTCAAACGTTCAGGCGCGCCGAAAGTGATTGAAGAACCCT
ASR_I001LR	ATGCCGTCTCAGGTCTCACCGCTTAAGCAATTGGATCCTACC
ASR_I001RFbis	GCATCGTCTCATCGGTCTCAGAGTTGCTTAATTAGCTTGTACATGG
ASR_I001RRbis	ATGCCGTCTCAGGTCTCATCGGCTGAGGCGCGCCGAAGGCCCATATTGAAGAC
ASR_P022Fbis	GCATCGTCTCATCGGTCTCAGCGGCCCCATTATCTTAGCCTAAAA

Name	Sequence (5' - 3')
ASR_P022R	ATGCCGTCTCAGGTCTCACATAGATCTTATAGTTTTTTCTCCTTGAC
ASR_T001F	GCATCGTCTCATCGGTCTCATAGGTCTGATCTGCTTACTTTACTAAC
ASR_T001Rc	ATGCCGTCTCAGGTCTCACAGCCACCTTTCTAAGGTTAGAATAGAAAT
ASR_M001R	ATGCCGTCTCAGGTCTCAACTCCAGTATAGCGACCAGCATTCA
Diagnosis of yeast mutants	
10737	AATTGCATCTCTTTGTGATGTC
SeqADE2	CTGCAACTGCTTGTTCAGCC
10909	TCTTCGTCGCCATTTATTGTTGAG
10910	CTATTGCGGTTCGCTCTTCC
10911	ATTCGCCGAATCTGACGTG
10912	TGGTGTGCAGACGGATAAAC
DiagKU80-F	GGGCAGACTTTTGTCCACTCTTG
DiagKU80-R	TGTGTCTAGCGATGAGCTGGAAAC
Diag-URA3-F	GGAAAGAGCAGCTGCTCATAGAAGT
Diag-URA3-R	CGGCAACAACTTCATCCACAGTTC
BSA-R	TACACGCGTTTGTACAGAAAAAAAAGAAAAATTTGA
I1DS-R	CTTAGTGGTTGTGAAGGTTT

2.3 Results

pUDP: a plasmid-based wide-host-range yeast CRISPR/Cas9 system

The pUDP system was designed to enable Cas9-mediated genome editing in different yeast species by simple transformation with a single plasmid. To this end, DNA parts encoding the *cas9* expression cassette, selection marker, gRNA expression cassette and plasmid replication origin were chosen to function in a wide range of yeast species including *K. lactis* and *K. marxianus* (**Figure 2.1**). The *cas9*$^{D147Y\ P411T}$ nuclease variant (Bao *et al*., 2015) was expressed under control of the *TEF1* promoter from *Arxula adeninivorans* (*Blastobotrys adeninivorans*), which enabled strong constitutive expression in various yeasts (Steinborn *et al*., 2006; Terentiev *et al*., 2004). Similarly, the *K. pneumoniae hph* gene encoding hygromycin B phosphotransferase, which conferred hygromycin resistance in a wide range of microorganisms, was expressed under control of the *Ashbya gossypii* (*Eremothecium gossypii*) *TEF1* promoter (Wach *et al*., 1994) that showed activity in a range of yeasts (Kim *et al*., 2010; Mazzoni *et al*., 2005). To avoid RNA modifications that interfere with biological function, transcription of gRNAs for genome editing is commonly controlled by RNA polymerase III promoters. By analogy with expression of gRNA in mammalian cells (Cong *et al*., 2013; Mali *et al*., 2013), the RNA polymerase III *SNR52* promoter was used in *S. cerevisiae* (DiCarlo *et al*., 2013; Mans *et al*., 2015), but accurate annotation and characterization of this promoter might not be available for other yeast species. Therefore, self-processing ribozyme-flanked gRNAs were expressed from an RNA polymerase II promoter (Gao and Zhao, 2014; Ryan *et al*., 2014). This concept has already been successfully applied in several different organisms (Gao *et al*., 2016; Gorter de Vries *et al*., 2017; Weninger *et al*., 2016). In this system, the gRNA is flanked by a hammerhead and hepatitis delta virus ribozyme at its 5′ and 3′ end, respectively, resulting in the precise release of mature gRNA after self-cleavage. To ensure sufficient gRNA transcription, the strong glycolytic *TDH3* promoter from *S. cerevisiae* was used. The pangenomic yeast origin of replication panARS(OPT) was chosen as the replication origin in pUDP002. This origin of replication, which was inspired by a *K. lactis* chromosomal ARS, has been shown to function in at least ten different species of ascomycetous yeasts, however its functionality has not been tested in *K. marxianus* (Liachko and Dunham, 2014). To check whether the replication origin could be used

in this yeast, transformation of the pIL75 plasmid carrying the panARS(OPT) replication origin was transformed into two *K. marxianus* type strains, CBS 6556 and CBS 608. Colonies were obtained at transformation efficiencies comparable to efficiencies reported in *K. lactis* (data not shown) (Liachko and Dunham, 2014). To test for plasmid stability the *K. marxianus* strains transformed with pIL75 were transferred to non-selective medium and then plated back onto solid medium containing antibiotic at different time points (**Figure 2.2**). A rapid decrease in the number of colonies resistant to the antibiotic was observed over a period of 24 hours, suggesting that the plasmid is loss rapidly in the absence of selective pressure. This finding indicates that the panARS(OPT) replication origin is functional in *K. marxianus* and its stability is dependent on selective pressure.

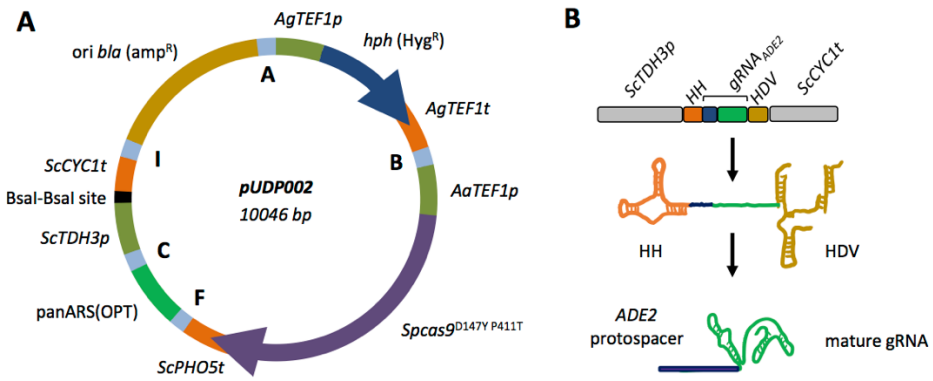


Figure 2.1: Components of the pUDP genome editing system.
A: Map of pUDP002 (Addgene plasmid #103872), a wide-host-range gRNA entry plasmid. pUDP002 is comprised of a *hph* (Hyg^R) selection marker cassette under control of the *TEF1* promoter from *Ashbya gossypii* conferring hygromycin resistance, *Spcas9*$^{D147Y\ P411T}$ under control of the *TEF1* promoter from *Arxula adeninivorans*, the pangenomic yeast replication origin panARS(OPT), a BsaI cloning site for entry of gRNA constructs, and a pBR322-derived *E. coli* origin and *bla* gene for ampicillin selection. A, B, F, C and I indicate 60 bp synthetic homologous recombination sequences used for Gibson assembly of the plasmid. **B:** Representation of the ribozyme-flanked gRNA expression cassette design. Guide RNAs (represented by gRNA$_{ADE2}$) were flanked on their 5' by a hammerhead ribozyme (HH represented in orange) and on their 3' by a hepatitis delta virus ribozyme (HDV represented in bronze). When integrated into pUDP002, this construct is under control of the RNA polymerase II promoter *TDH3* and the *CYC1* terminator from *S. cerevisiae*. Upon ribozyme self-cleavage, a mature gRNA comprised of the guiding protospacer (in blue) and the structural gRNA fragment (in green) is released.

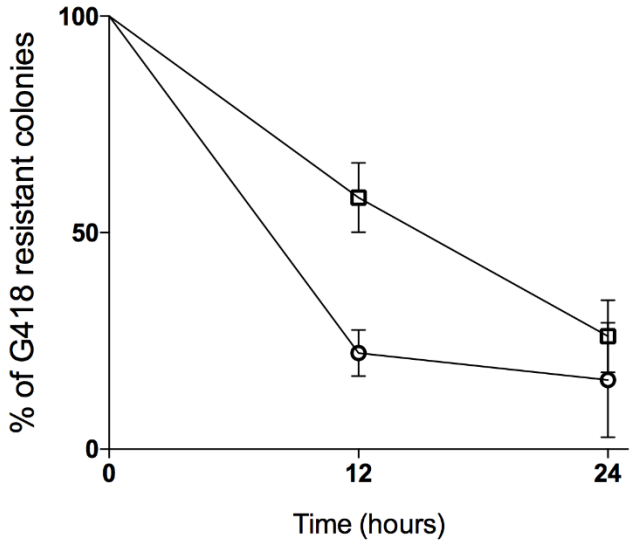


Figure 2.2: Stability of the pIL75 plasmid containing the panARS(OPT) replication origin in *K. marxianus*.
Functionality and stability of the panARS replication origin was tested in *K. marxianus.* CBS 6556 (open circles) and CBS 608 (open squares) transformed with pIL75 were grown in non-selective conditions and then plated on selective media at 12 and 24 hours. The percentage of G418 resistant colonies is shown for both strains.

To test the genome-editing efficiency of the pUDP system, the *ADE2* gene was targeted in *K. lactis* and *K. marxianus*. The *ADE2* gene encodes a phosphoribosylaminoimidazole carboxylase, also referred to as AIR carboxylase, involved in adenine biosynthesis. Besides causing adenine auxotrophy, loss-of-function mutations in *ADE2* results in a red-colour phenotype due to accumulation of the oxidized form of 5-amino imidazole ribonucleotide (AIR). This allows detection of *ADE2* targeting by simple visual inspection of transformation plates (Roman, 1956). Therefore, gRNAs were designed targeting *ADE2* in *K. lactis* and *K. marxianus* based on available genome data (**Table 2.3**), and gRNA-harbouring plasmids pUDP025 and pUDP082 for deletion of *ADE2* in *K. lactis* and *K. marxianus* respectively, were constructed.

Efficient CRISPR/Cas9 targeting enables gene deletion in *K. lactis* and *K. marxianus*

To test the effectiveness of the pUDP system in *K. lactis*, gRNA$_{KlADE2}$ was inserted into pUDP002 and the resulting plasmid pUDP025 was used to target *ADE2* in *K. lactis* CBS 2359. Transformation with pUDP025 without a repair DNA fragment yielded a total of 35 transformants of which 19 (54%) exhibited a red Ade$^-$ phenotype. In the presence of a 962 bp repair fragment which targeted the *ADE2* promoter and terminator (**Figure 2.3A**), the transformation of pUDP025 yielded 26 red transformants out of a total 27 (96%) (**Figure 2.3B**). In comparison, transformation of the backbone plasmid pUDP002, which does not express a gRNA, generated a number of transformants that was ca. 35-fold higher, none of which displayed a red phenotype. This difference in transformation efficiency already provided information about the quality of the gRNA and the editing. Unless repaired chromosomal DSB should be lethal, therefore this 35-fold drop in transformation efficiency would represent the fraction of transformants that did not successfully repair the CRISPR-Cas9 induced break and subsequently died from it. Out of the 26 Ade$^-$ transformants obtained after co-transformation with pUDP025 and repair fragment, 13 red colonies were randomly picked and analysed by diagnostic PCR, revealing two distinct groups. Nine colonies exhibited a normally sized *ADE2* locus while four colonies (31%) showed a PCR-fragment size compatible with complete deletion of the *KlADE2* ORF (**Figure 2.3C**). Subsequent Sanger sequencing of the amplified fragments derived from the nine clones exhibiting an *ADE2* wild-type size band identified the presence of indels at the targeted site. These indels reflected imperfect repair of the Cas9-induced DSB via NHEJ, resulting in introduction of loss of function mutations and disruption of the Cas9 target site (**Figure 2.3D**). These results demonstrated that the pUDP system could be used for efficient targeting of the *ADE2* gene in *K. lactis* CBS 2359, resulting in repair by either HR or NHEJ DNA repair mechanisms.

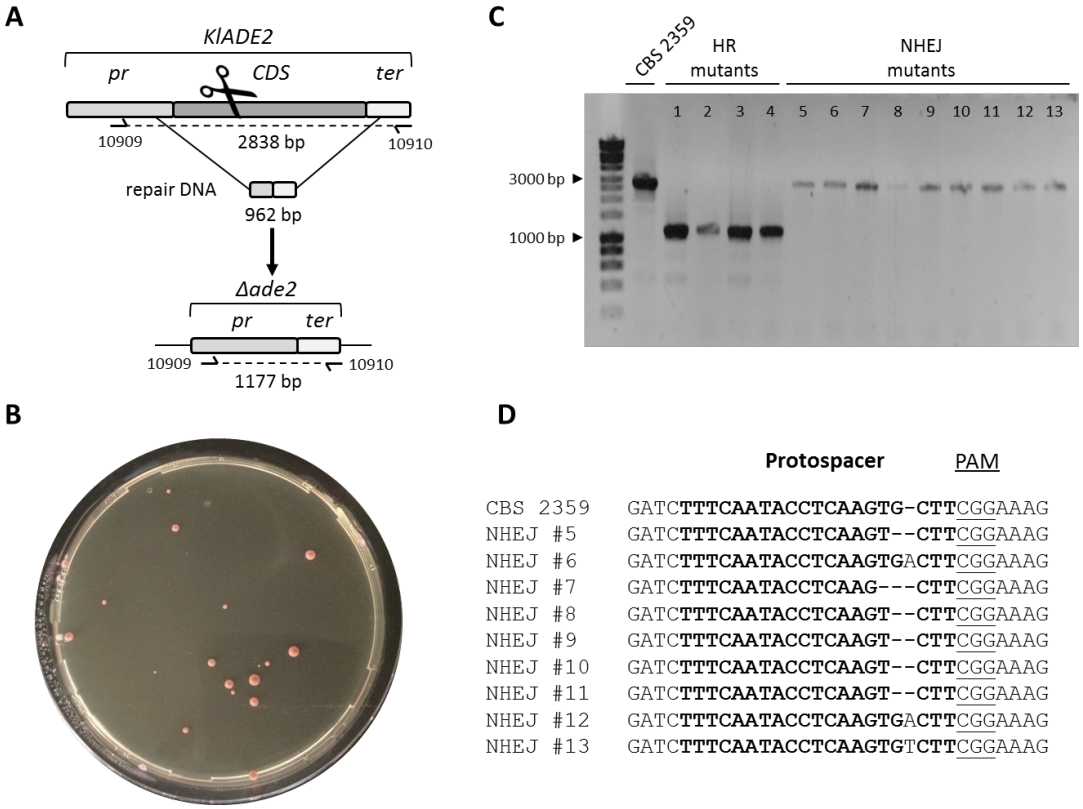


Figure 2.3: Efficient gRNA targeting in *K. lactis* CBS 2359 enables marker-free gene deletion. **A:** Schematic representation of *ADE2* editing upon transformation of CBS 2359 with pUDP025 (gRNA$_{KlADE2}$) and a repair DNA fragment. The primers for diagnostic PCR of transformants are indicated. **B:** Colony morphology of CBS 2359 upon transformation with pUDP025 and a marker-free 962 bp repair fragment. **C:** Diagnosis of 13 randomly picked red Ade$^-$ transformants of CBS 2359 upon transformation with pUDP025 and a 962 bp marker-free repair fragment. Four transformants (HR mutants 1-4) showed a PCR product of 1177 bp corresponding to the deleted allele. The control labelled CBS 2359 and nine transformants (NHEJ mutants 5-13) showed a PCR product of 2838 bp corresponding to the wild-type allele. **D:** Sanger sequencing results of purified PCR fragments from nine Ade$^-$ mutants (corresponding to mutants 5-13 in panel C) derived from the transformation of CBS 2359 with pUDP025 and repair fragment.

The pUDP system was similarly tested in *K. marxianus* by targeting *ADE2* in the haploid strain NBRC 1777 and the diploid strain CBS 397, using plasmid pUDP082 expressing gRNA$_{KmADE2}$. The transformation of *K. marxianus* NBRC 1777 with pUDP082, without and with a 959 bp repair DNA fragment resulted in 13 and 30 transformants, respectively. Of these, 13 out of 13 and 29 out of 30 transformants were red, indicating successful disruption of *ADE2* in both cases. To determine which repair mechanism contributed to the repair of the Cas9-induced DSB in the presence of the repair fragment, 13 randomly picked red transformants were subjected to diagnostic PCR. Similar to the results obtained in *K. lactis* two groups of transformants were identified. Three transformants (24%) showed a genotype

corresponding to the repair of the DSB with the repair fragment (**Figures 2.4A** and **2.4B**), while the remaining transformants exhibited a PCR result that was undistinguishable from that of the NBRC 1777 *ADE2* wild type locus. Sanger sequencing of the Cas9 target site again revealed the presence of indels, leading to nonsense mutations in the *ADE2* coding sequence (**Figure 2.4C**).

When the diploid strain CBS 397 was transformed with pUDP082 without repair DNA fragment, 106 out of 133 (80%) of the colonies were red, while 117 out of 143 (82%) colonies showed the red phenotype when a 959 bp repair fragment was provided. In comparison, a control transformation with pUDP002 yielded 262 transformants without any red phenotypes. From the transformation with pUDP082 and repair fragment, ten red colonies were randomly picked and analysed by diagnostic PCR, resulting in PCR products that were identical in size to the native *ADE2* locus in all cases. Sanger sequencing of the same mutants confirmed the link between the red-coloured phenotype and occurrence of small indels at the Cas9 cut site, which likely introduced loss of function mutations in *ADE2* (**Figure 2.4D**). Interestingly, Sanger sequencing also indicated that the editing occurred identically at both *KmADE2* alleles as clear and continuous sequence was observed over the cut site. These results demonstrated that the *ADE2* gene can be efficiently targeted by the pUDP system in both *K. marxianus* strains, with DNA repair mediated by HR or NHEJ mechanisms in the haploid strain NBRC 1777, and NHEJ in the diploid strain CBS 397.

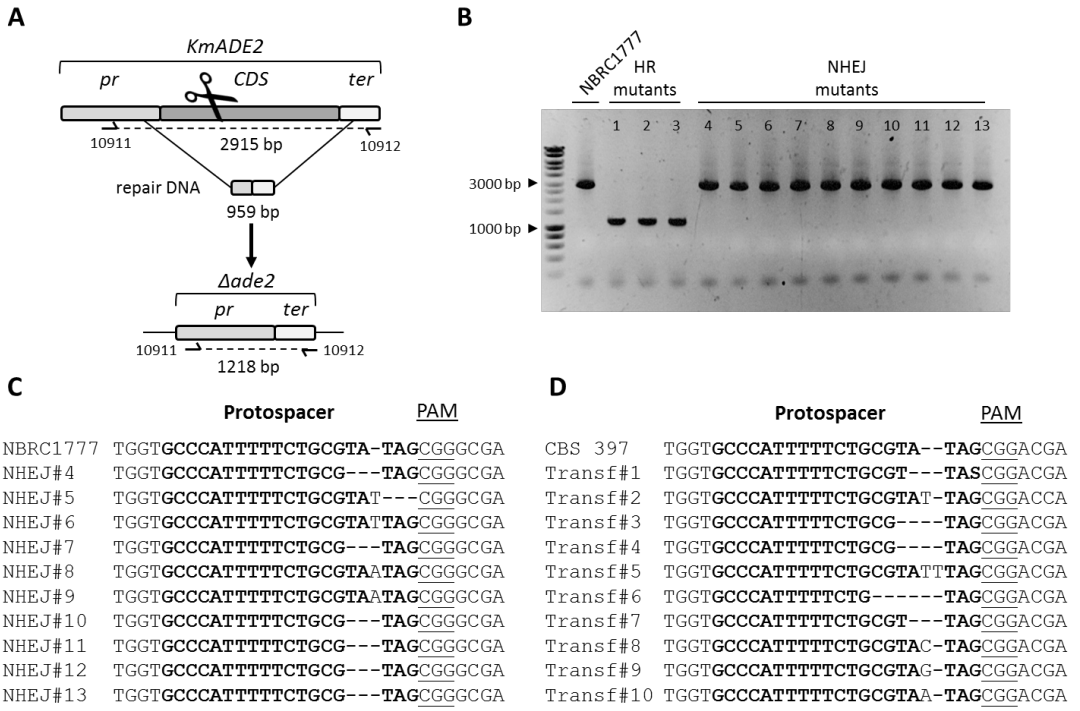


Figure 2.4: Efficient gRNA targeting enables marker-free gene deletion in haploid *K. marxianus* NBRC 1777 and gene disruption in diploid *K. marxianus* CBS 397.
A: Schematic representation of the *ADE2* editing upon transformation of NBRC 1777 with pUDP082 ($gRNA_{KmADE2}$) and a repair DNA fragment. The primers for diagnostic PCR of transformants are indicated. **B:** Diagnosis of 13 randomly picked red Ade^- transformants of NBRC 1777 upon transformation with pUDP082 and a 959 bp marker-free repair fragment. Three transformants (HR mutants 1-3) showed a PCR product of 1218 bp corresponding to the deleted allele. The control labelled NBRC 1777 and ten transformants (NHEJ mutants 4-13) showed a PCR product of 2915 bp corresponding to the wild-type allele. **C:** Sanger sequencing results of purified PCR fragments from ten Ade^- mutants (corresponding to mutants 4-13 in panel B) derived from the transformation of NBRC 1777 with pUDP082 and repair fragment. **D:** Sanger sequencing results of purified PCR fragments of ten randomly picked red Ade^- mutants derived from the transformation of CBS 397 with pUDP082 and repair fragment.

Modifications to the pUDP002 plasmid enable fast cloning of target sequences

The CRISPR-Cas9 system encoded in the pUDP002 plasmid can efficiently be used for gene targeting in *K. lactis* and *K. marxianus*, however cloning target sequences into the plasmid requires the use of synthetic DNA blocks or long overlapping DNA sequences. It has been shown that the 5′ end of the HH ribozyme present in the plasmid has to contain the reverse complement of the first 6 nucleotides of the target sequence cloned, which essentially involves modifying the HH 5′ sequence every time a new target sequence is cloned (Gao and Zhao, 2014). To test whether the CRISPR-Cas9 functionality is affected by omitting the HH 5′ sequence, the pUDP002 plasmid was modified by introducing a cloning site that allows fast and simple cloning of target sequences without additional modifications to the HH ribozyme (**Figure 2.5A**). The new plasmid, pUCC001, contains a cloning site comprised of a double recognition site for the type IIS restriction enzyme, BsaI. Since this enzyme cleaves the DNA outside its recognition site it can be used to generate specific overhangs at each end of the DNA sequence. A pair of 24 bp complementary DNA primers, encoding the target CRISPR sequence, can then be annealed to form a duplex with overhangs compatible with pUCC001 and cloned into the plasmid via Golden Gate assembly (Ng and Dean, 2017; Vyas *et al*., 2015). To test the system, the *KmADE2* target sequence was cloned into pUCC001 and transformed into CBS 397 without a repair fragment (**Figure 2.5B**). A total of 90 out of 99 colonies (91 %) showed the red phenotype indicating that the pUCC001 system can induce mutagenesis of the *ADE2* locus at frequencies comparable to the pUDP system. Interestingly, the colonies that did not exhibit the red phenotype grew faster than the *ADE2* mutants indicating that mutations in *ADE2* have consequences for cellular fitness in *K. marxianus*. Similar results were obtained when transforming NBRC 1777 (data not shown). The results obtained using the pUCC001 plasmid suggest that the system is functional under the conditions tested, despite lacking the reverse complement sequence at the 5′ end of the HH ribozyme.

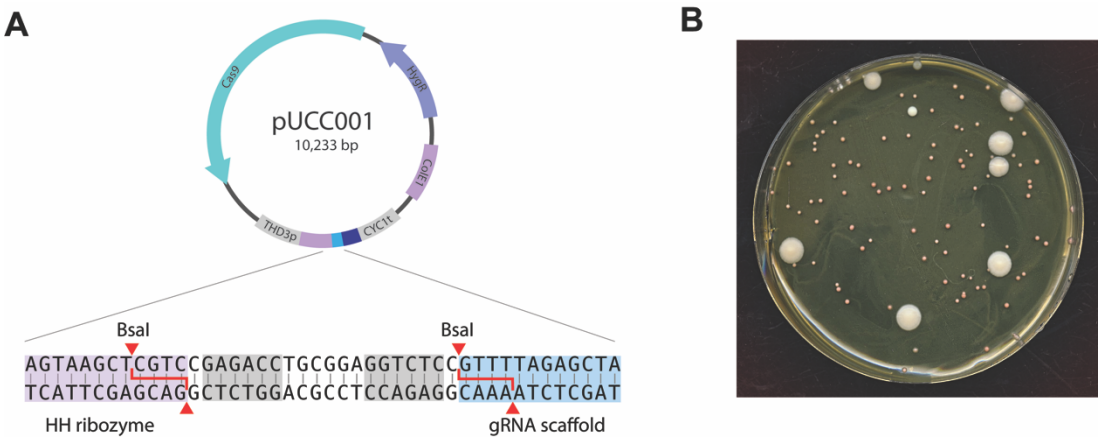


Figure 2.5: Cloning system in pUCC001 plasmid allows simple cloning of target sequences.
A: Map of the pUCC001 plasmid highlighting the double BsaI cloning site system. The BsaI recognition sites (grey shading) lie within the linker sequence between the HH ribozyme and the gRNA scaffold (purple and blue shading, respectively). Primer duplexes containing a target sequences and the 5′ CGTC 3′ and 5′ CAAA 3′ overhangs can be cloned readily into this position by golden gate assembly. Excluding the BsaI-BsaI cloning site all the components of the pUCC001 plasmid are identical to the ones in pUDP002. **B:** Efficient targeting of the *ADE2* gene in *K. marxianus* using pUCC001. The *KmADE2* target sequence was cloned into pUCC001 by Golden Gate assembly and transformed into *K. marxianus* CBS 397. Cells were plated onto YPD medium plates.

To elucidate the mechanism underlying the function of pUCC001 a RNA secondary structure prediction was performed comparing the *ADE2* gRNA molecules encoded by pUDP002 and pUCC001 (**Figure 2.6**). According to this analysis the two systems give rise to slightly different gRNA structures. In the *ADE2* gRNA from pUDP002 the first 6 nucleotides of the *ADE2* target sequence anneal with the 5′ end of the HH ribozyme, as expected (**Figure 2.6A**). Also, the HDV sequence anneals with the gRNA scaffold sequence forming a linker structure that separates the HDV ribozyme from the HH and *ADE2* elements. On the contrary, in the pUCC001 system the *ADE2* target sequence partially anneals with the HDV ribozyme while the HH structure shows slight differences when compared to the structure encoded by pUDP002 (**Figure 2.6B**). These are simply computational predictions and experiments are required to establish the exact structure of the gRNA molecule encoded by the pUCC001 plasmid and the mechanism underlying its functionality. However, as shown below the system was successfully used to disrupt other genes in *K. marxianus* (also see Chapter 4).

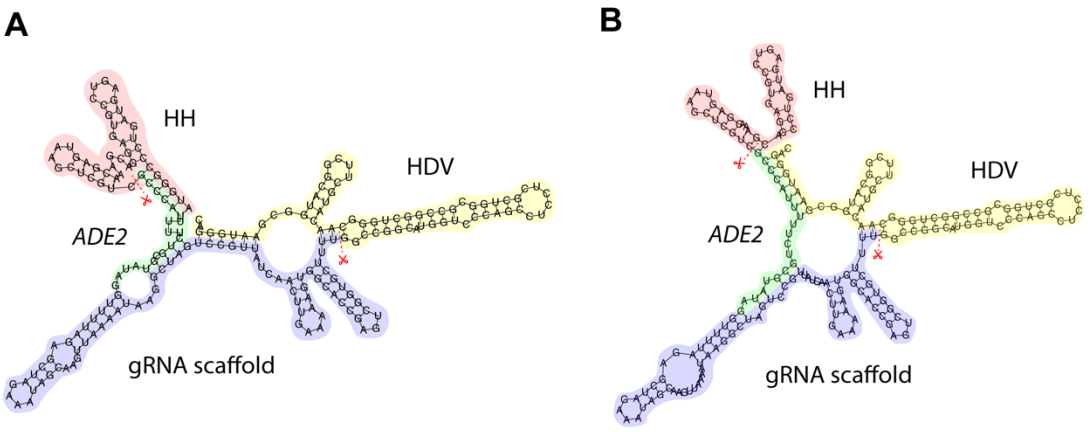


Figure 2.6: Secondary structure prediction of gRNA molecules encoded by pUDP002-*ADE2* and pUCC001-*ADE2*.
A: Secondary structure of *ADE2* gRNA molecule encoded by pUDP002-*ADE2*. **B:** Predicted secondary structure of *ADE2* gRNA molecule encoded by pUCC001-*ADE2*. The different components of the pre-processed gRNA molecules are shown in different colours; yellow; HDV ribozyme; blue; gRNA scaffold; green; *ADE2* target sequence; red; HH ribozyme. Cleavage sites for HH and HDV are indicated with dotted red lines. The prediction was performed using the software RNA*fold*.

Deletion of *KU80* and overexpression of *ScRAD52* improve HR in *K. marxianus*

While the CRISPR-Cas9 system is functional in *K. marxianus*, the NHEJ system present in the yeast presents a major limitation when performing targeted integrations in the genome. To inactivate this system, the *KU80* gene was targeted using the CRISPR-Cas9 system. In this case, the pUCC001 system was used as an expression system for the *KU80* gRNA and NBRC 1777 was chosen as a platform strain. After transforming NBRC 1777 with the pUCC001-*KU80* plasmid 86 colonies were obtained. DNA was extracted from three of these transformants and used for PCR reactions targeting the *KU80* gene (**Figure 2.7A**). To genotype these strains, the PCR products were purified, sequenced and compared against the wild-type *KU80* sequence (**Figure 2.7B**). One of these transformants (*KU80-3*) did not contain any mutations in the targeted site, however the remaining two contained distinct mutations adjacent to the protospacer sequence. The *KU80*-1 and the *KU80*-2 strains had a two-nucleotide deletion and a single nucleotide insertion, respectively. Both frame-shift mutations result in premature stop codons in the Ku80p sequence that render the gene non-functional. As the double deletion in the *KU80*-1 mutant is less likely to revert than the mutation in *KU80*-2, the former strain was stocked and used for further experiments. The strain was renamed YBL001.

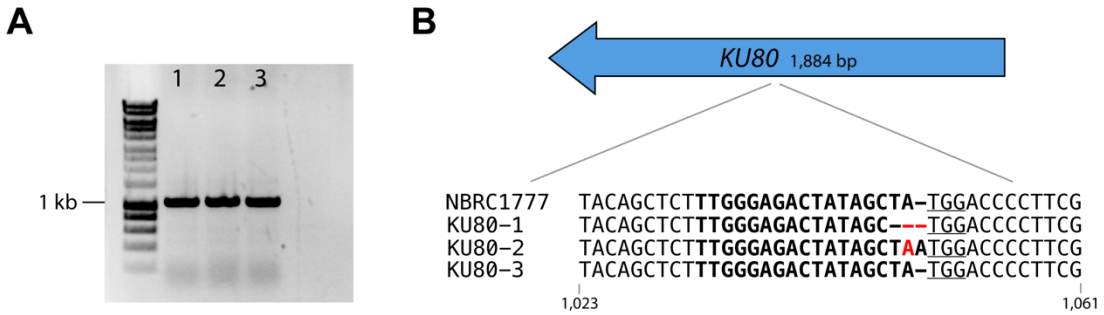


Figure 2.7: Mutagenesis of the *KU80* gene in *K. marxianus*.
The pUCC001-*KmKU80* plasmid, containing the target sequence for the *KU80* gene, was used to obtain a NHEJ mutant in *K. marxianus* NBRC 1777. **A:** Gel electrophoresis showing *KU80* PCR products obtained from three colonies after transformation with pUCC001-*KmKU80*. Samples were later purified and sequenced. DNA molecular marker: HyperLadder 1. **B:** DNA alignment of transformants against the wildtype *KU80* sequence. Mutations are shown in red. The protospacer and PAM sequences are shown in bold and underlined styles, respectively.

To test whether the YBL001 mutant lacked NHEJ, and if the disruption of *KU80* resulted in an increase in HR-mediated repair frequency, a plasmid containing a cassette targeting the *URA3* gene in *K. marxianus* was constructed. The assembly of this plasmid was performed using standardised genetic parts from the yeast toolkit, a genetic engineering platform that contains multiple modules that can be readily assembled using the Golden Gate assembly method (Lee *et al*., 2015). The technique relies on using the BsaI enzyme to create specific 4 bp overhang sequences in every part to be assembled. Since only sequences containing complementary overhangs can be ligated, the use of different overhangs allows for directional cloning of multiple parts in a single Golden Gate reaction. While the platform was developed for *S. cerevisiae*, compatibility of some parts (e.g. antibiotic markers and some promoters) allows its use in *K. marxianus*.

To construct the p*URA3-KanMX* plasmid, 1kb regions flaking the *URA3* gene were amplified with primers containing the BsaI site and assembled via Golden Gate assembly with the pYTK077 and pYTK095 plasmids, containing the *KanMX* marker (conferring resistance to G418) and the *E. coli* maintenance elements, respectively (**Figure 2.8A**). The resulting plasmid was amplified by PCR to obtain a linear fragment containing the flaking *URA3* regions and the *KanMX* marker.

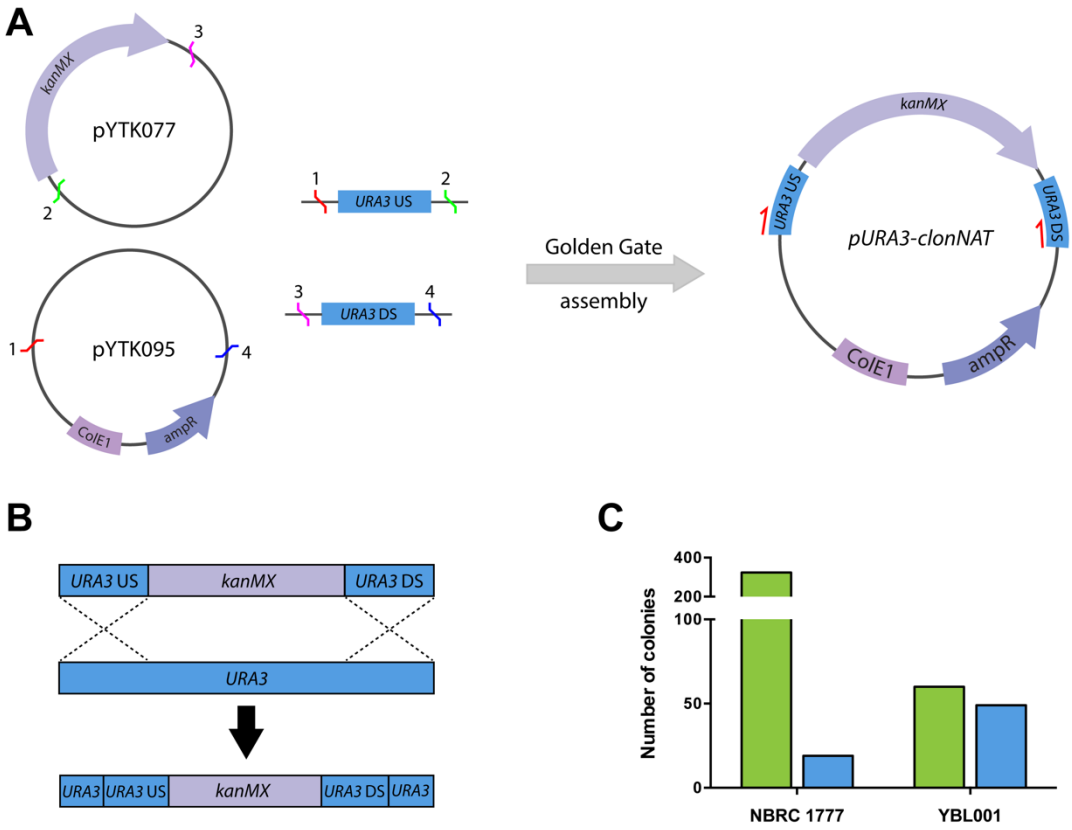


Figure 2.8: *KU80* mutation impairs NHEJ in *K. marxianus*.
A: Construction of the p*URA3-KanMX* plasmid using the yeast toolkit method. The *URA3* upstream and downstream fragments were PCR amplified with primers containing BsaI recognition sites and specific overhangs. The fragments were mixed with the pYTK077 and pYTK095 plasmids in a one-pot Golden Gate reaction to obtain the p*URA3-KanMX* plasmid. Specific overhangs sequences, represented by coloured lines and numbers flanking the parts, allow directional cloning of the different elements. Primers used to amplify p*URA3-KanMX* and obtain the *URA3-KanMX* fragment are shown as red arrows. **B:** Diagram showing the integration of the *URA3-KanMX* cassette into the *URA3* locus. Correct incorporation of the cassette results in uracil auxotrophy. **C:** Histogram displaying the number of colonies obtained when transforming NBRC1777 and YBL001 with the *URA3-KanMX* cassette. NBRC1777 and YBL001 were transformed with 2 µg of the cassette and selected on G418 medium. Colonies were counted after two days and then replicated plated onto 5-FOA medium to score for uracil auxotrophy. Green and blue bars represent the number of G418 resistant colonies and -ura colonies, respectively.

The *URA3-KanMX* cassette can be used to study both HR and NHEJ-mediated repair in *K. marxianus*. Random integration of the cassette into the genome by NHEJ results in G418 resistance and uracil prototrophy. On the contrary, correct integration of this fragment into the *URA3* locus by HR results in transformants that are resistant to G418 but still uracil auxotrophs (**Figure 2.8B**). In order to establish whether the NHEJ system was active in the *ku80* mutant YBL001, the *URA3-KanMX* cassette was transformed in this strain and the wild-type strain NBRC 1777 (**Figure 2.8C**). The strains were selected on medium containing G418 and then transferred to 5-FOA

plates, which only allow the growth of uracil auxotrophs (to screen for *URA3* disruption). A total of 324 colonies were obtained when transforming the *URA3-KanMX* cassette into NBRC 1777. Screening of these colonies on 5-FOA plates revealed that only 6% of the colonies (19/324) had the cassette integrated in the *URA3* locus. In comparison, although the number of G418 resistant colonies in YBL001 was considerably lower (60), the targeting efficiency in this case reached 82%, which suggested that the NHEJ activity in this strain had been greatly impaired. The HR frequency in YBL001 is comparable with frequencies reported by Choo *et al*., 2014, using a *K. marxianus* CBS 6556 *ku80* mutant and similar experimental conditions. In that study, the authors reported a targeting efficiency of 70% when targeting the *URA3* gene with a linear DNA cassette containing 1 kb homology arms. Altogether, these results suggest that NHEJ is inactive in YBL001. Because a single flanking region length was tested in this experiment, the impact of flanking region length in HR-mediated repair in YBL001 is not known. Nonetheless, it has been demonstrated that flanking region length and transformation efficiency correlate inversely in a *ku80* mutant (Choo *et al*., 2014). This could limit some applications that require integrating multiple genes into the genome simultaneously. One strategy that could solve this issue is overexpressing DNA recombinases involved in HR repair from other organisms. This approach has been successfully used to boost HR in human cells where a 37-fold increase in gene targeting was observed when overexpressing the *RAD52* gene from *S. cerevisiae* (Di Primio *et al*., 2005; Kalvala *et al*., 2010). To test if the frequency of gene targeting in *K. marxianus* could be enhanced, a similar approach was followed. First, the *LAC4* gene was chosen as a potential insertion site for *RAD52* because mutations in this gene do not affect cell fitness (only when lactose is used as carbon source) (Sheetz and Dickson, 1980). Also, colonies carrying loss-of-function mutations in *LAC4* can be easily identified in galactose-X-gal plates – mutant colonies are white while wild-type colonies are blue (Ribeiro *et al*., 2007). The *LAC4* gene was targeted using the CRISPR-Cas9 system resulting in a high number of colonies with 133 out of 223 being white (60% efficiency) (**Figure 2.9**). Colony size of both blue and white colonies was comparable suggesting that mutations in *LAC4* do not impart cellular fitness and that the gene can be used as an insertion site for *RAD52*.



Figure 2.9: Mutagenesis of the *LAC4* using the CRISPR-Cas9 system.
The *LAC4* gene, selected as a potential insertion site for the *RAD52* gene, was targeted using the CRISPR-Cas9 system. NBRC 1777 was transformed with the pUCC001-*LAC4* plasmid and then plated onto YPGal medium supplemented with x-gal.

A plasmid containing the *RAD52* gene from *S. cerevisiae* and the *KanMX* marker flanked by 1 kb homology arms that allow integration into the *LAC4* locus was constructed using the yeast toolkit method. The *GAL1* promoter from *S. cerevisiae* was chosen to express *RAD52* since it can be induced by the presence of galactose and has been previously used in *K. marxianus* for gene expression experiments (Almeida *et al.*, 2003; West *et al.*, 1984). The pI1P22-*RAD52* plasmid containing the *LAC4-RAD52* cassette was digested and transformed into YBL001 (**Figure 2.10A**). Seven white colonies were obtained after transformation. These were checked for insertion of the cassette into the *LAC4* locus by PCR using an internal and external primer annealing the *LAC4* DS and *LAC4* genomic DNA, respectively. A 1 kb PCR product was obtained in all cases confirming the correct integration of the cassette in the transformants (**Figure 2.10B**). One of these transformants was stocked and renamed YBL002.

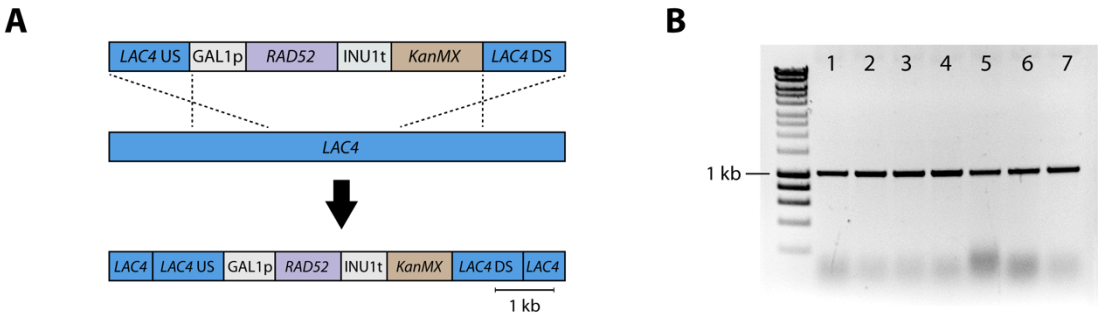


Figure 2.10: Construction of a *K. marxianus* strain expressing the *RAD52* recombinase from *S. cerevisiae*.
A: Scheme showing the double recombination event leading to the integration of the *RAD52* cassette into *K. marxianus* genome. A linear fragment containing the *RAD52* gene from *S. cerevisiae*, the *KanMX* marker and *LAC4* flaking sequences was obtained by digesting the pI1P22-*RAD52* plasmid with NotI and then transformed into *K. marxianus*. Integration of the fragment into the genome results into disruption of the *LAC4*, which can be observed in YPGal X-gal plates. Screening of white colonies was carried out by PCR using primers that bind to the *LAC4* DS and *LAC4* elements, the predicted PCR product is 1kb as shown in the figure. **B:** Screening of white colonies by PCR. DNA from seven white colonies was extracted and used for PCR reactions. DNA molecular marker: HyperLadder 1.

To test the effect of *RAD52* overexpression in HR-mediated repair the wild-type strain NBRC 1777, the *ku80* mutant YBL001 and the *lac4::RAD52* strain, YBL002, were transformed with a linear cassette targeting the *URA3* gene. Since the YBL002 strain contains the *KanMX* marker integrated into the *LAC4* locus, a new plasmid containing the *NatMX* marker (encoding resistance to Nourseothricin or ClonNat) was constructed to score for HR-mediated repair. The cassette containing the marker and the *URA3* flanking regions was amplified and transformed into the different strains, as previously described (**Figure 2.11**). In order to establish if the *GAL1* promoter controlling *RAD52* expression was responsive to galactose induction, the YBL002 strain was grown in glucose (YPD) and galactose (YPGal) media. The different strains were plated on selective medium containing ClonNat and then transferred to 5-FOA plates to calculate the number of *ura3* mutant colonies. The frequencies obtained for NBRC 1777 matched previous results with 7% of the colonies containing the cassette inserted in the *URA3* locus. On the other hand, only one colony per plate was obtained when transforming YBL001 with the *NatMX* cassette but these colonies were found to be uracil auxotrophs. This result contrasts with the number of colonies obtained when transforming YBL001 with the *KanMX* cassette, which suggests that antibiotic marker preference influences integration of the cassette into the genome.

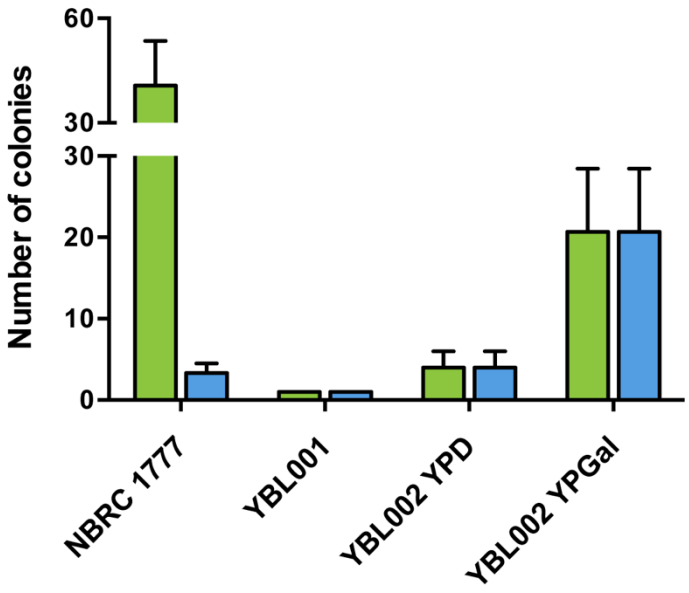


Figure 2.11: *RAD52* expression in *K. marxianus* increases HR-mediated repair.
NBRC 1777, YBL001 and YBL002 were transformed with a linear DNA fragment targeting the *URA3* locus. For transformations, the strains were grown on YPD with the exception of YBL002, which was grown on YPD and YPGal. Cells were transformed with 2 μg of the cassette and then plated on YPD or YPGal medium containing ClonNat. The transformants were then transferred to 5-FOA medium to screen for uracil auxotrophy. Green and blue bars represent the number of G418 resistant colonies and -ura colonies, respectively. The experiments were performed in triplicate with error bars showing SD.

The number of colonies obtained for the YBL002 strain grown in YPD was comparable with the amount of ura- colonies obtained when transforming NBRC 1777. However, an increase in the number of colonies was observed in the YBL002 strain grown in galactose medium. All of these colonies were able to grow on 5-FOA medium indicating that the *NatMX* marker had been integrated in the *URA3* locus. These results suggest that overexpression of *RAD52* has a positive effect in HR-mediated repair in *K. marxianus*. The number of ura- colonies obtained in this strain was seven-fold greater than the number observed in the wild-type strain NBRC 1777. To evaluate the effect of double-strand breaks in HR repair in YBL002, a pUCC001 plasmid targeting the *URA3* was constructed. The plasmid itself was able to induce mutagenesis of the *URA3* locus, however no differences in HR-mediated repair frequencies were observed when this plasmid was transformed in combination with the *NatMX* cassette (data not shown). Interestingly, the overall number of colonies obtained in the two experiments performed to assess HR-mediated repair varied depending on the selection marker used (i.e. fewer colonies were obtained when using the *NatMX* marker). This phenomenon could be due to either a possible

secondary structure formed by the *NatMX* cassette that difficult genome integration or poor translation efficiency of the gene in *K. marxianus*. Regardless of the case, these results should be taken into account if future experiments involving this genetic marker are to be performed.

2.4 Discussion

Molecular tools are crucial for developing *K. marxianus* for cell factories applications. In this study, the CRISPR-Cas9 system was implemented in *K. marxianus* allowing genetic manipulation of this yeast be performed. The pUDP system developed here permitted disruption of the *ADE2* gene at high frequencies in haploid and diploid *K. marxianus* strains. Also, the system was used in *K. lactis, K. marxianus* closest relative. In both yeasts, NHEJ and HR-mediated mutations were obtained, demonstrating the versatility of the CRISPR-Cas9 system. The pUDP plasmid was refined by the introduction of a type IIS restriction enzyme site that allows cloning of target sequences easily while minimising costs associated with chemical synthesis of DNA blocks. It was demonstrated that this modification does not impair the functionality of the system in the conditions tested. However, the mechanism underlying the function of the system remains unknown. It is possible that the lack of the complementary sequence at the 5′ end of the HH ribozyme disables its self-cleavage activity resulting in a gRNA molecule formed by the HH ribozyme, gRNA scaffold and protospacer sequence (**Figure 2.12A**). Under this assumption, Cas9 activity would not be affected by the presence of the ribozyme. On the contrary, another possibility is that the 5′ end of the ribozyme weakly anneals with the target sequence and this interaction is enough for HH self-cleavage activity (**Figure 2.12B**). This implies that the functionality of the system could depend on the target sequence used, which would require careful design of target sequences. However, as shown in Chapter 4, the pUCC001 system was successfully used to disrupt other six genes in *K. marxianus*. In that case, mutants were readily obtained without considering RNA secondary structures when designing the target sequences. Also, no sequence patterns could be observed among the target sequences used throughout this study meaning that either the system does not depend on the target sequence used or that its functionality only requires minor interactions between the HH ribozyme and the target sequence. The effect of modifying the HH ribozyme in a CRISPR system has been investigated to some extent in *P. pastoris* (Weninger *et al*., 2016). In that study, the authors placed a random 6 bp linker sequence between the HH ribozyme and target sequence. This linker sequence anneals with the HH ribozyme allowing self-cleavage but also forms part of the mature gRNA molecule once the HH and HVD ribozymes have been processed.

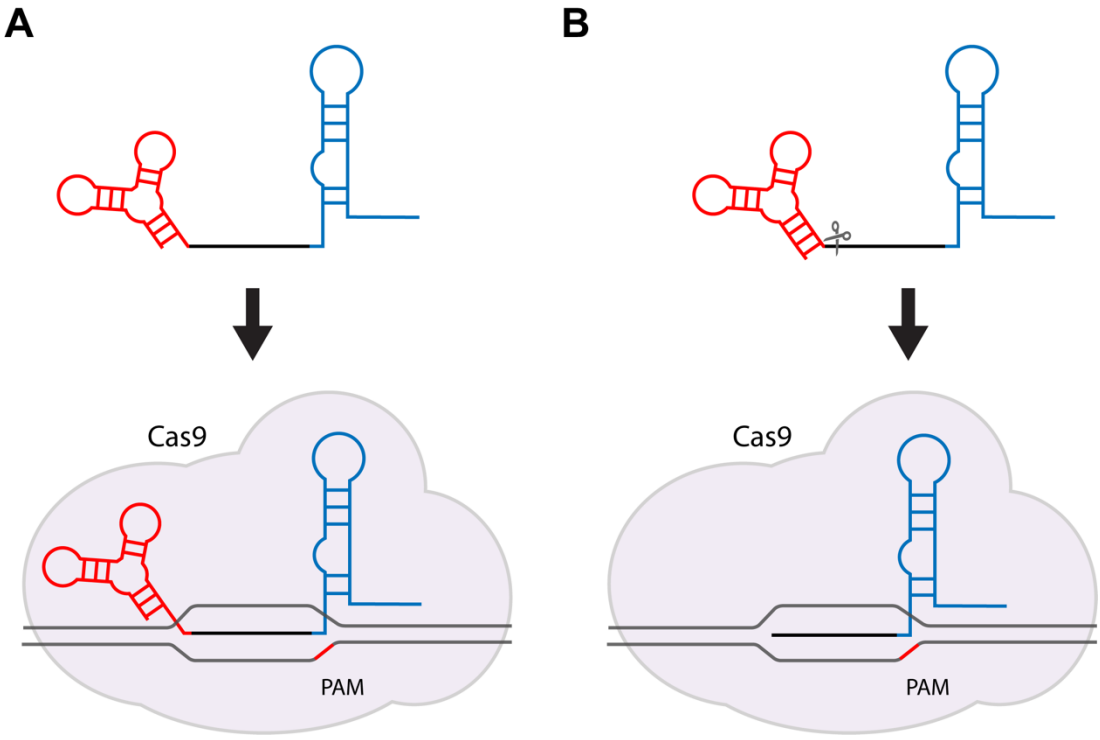


Figure 2.12: Possible mechanism behind pUCC001 functionality.

A: The gRNA molecule is expressed and the HDV ribozyme processed. However, the HH ribozyme (red) cannot be processed and remains attached to the rest of the molecule. The mature gRNA binds to Cas9 and catalyses the formation of DSBs at the target site despite containing the intact HH ribozyme. **B:** The lack of the 5′ sequence does not affect the activity of the ribozyme which is capable of self-cleavage. This produces a mature gRNA molecule comprised by the protospacer and gRNA scaffold (blue) sequences only.

The system was evaluated using three, previously tested, protospacer sequences targeting the glycerol kinase gene, *GUT1*. High mutagenesis efficiency was observed when using two of these targets. However, the system was not functional when using the third protospacer sequence, likely due to an interaction between the protospacer and the linker sequences that inactivates the gRNA molecule (Weninger *et al.*, 2016). Although the system used in that study is different than pUCC001, where the 6 bp linker sequence is absent, the data suggest that modifications in the HH ribozyme could dramatically affect CRISPR-Cas9 function. In order to test if HH self-cleavage is not required for Cas9 activity and to check for sequence-dependent self-cleavage effects, a RT-PCR experiment could be performed. The information obtained by this approach, in combination with additional modifications (different targets, other HH modifications) can be used to re-engineer the system maximising both functionality and cloning simplicity.

Other CRISPR-Cas9 systems are also available for *K. marxianus* (Löbs *et al*., 2017; Nambu-Nishida *et al*., 2017). Nevertheless, these systems rely on using RNA polymerase III promoters for gRNA expression, which limits cross-species applications. A comparison between these systems and pUDP/pUCC could be useful in further refining the CRISPR-Cas9 system in *K. marxianus*. At the same time, it would be interesting to test if the two RNA polymerase III-based systems are functional in other *Kluyveromyces* species such as *K. lactis*. This would open new possibilities for developing new cross-platform tools including other *Kluyveromyces* species.

The CRISPR-Cas9 system was also used to disrupt the *KU80* gene in *K. marxianus*. Disruption of *KU80* resulted in 82% efficiency in gene targeting, which contrasts with the 6% obtained in the wild-type strain. While this data indicates that NHEJ inactivation improves the frequency of HR-mediated repair, the experiment was only performed using 1 kb flanking regions, which excludes the effect of flanking region size in HR-mediated repair. Previous studies have reported that transformation efficiency correlates inversely with flaking region size in a *K. marxianus ku80* strain (Choo *et al*., 2014). However, it has not been established whether this phenotype is universal across the species and further experiments are required to test if this is the case in NBRC 1777.

Despite being useful for genetic engineering experiments, loss-of-function mutations in *KU80* could be detrimental and affect cellular fitness (Banerjee *et al*., 2006). *KU80* plays an important role in telomere maintenance and elevated rates of subtelomeric recombination have been observed in a *K. lactis ku80* strain (Carter *et al*., 2007). Nevertheless, mutations in *NEJ1*, another component of the NHEJ system, do not seem to cause the same effect (Carter *et al*., 2007). Both *K. lactis* and *K. marxianus nej1* strains have been constructed to successfully increase the rate of HR-mediated repair (Kooistra *et al*., 2004; Nambu-Nishida *et al*., 2017). This strategy should be considered in the future when engineering *K. marxianus* strains for industrial applications.

To further enhance the frequency of HR-mediated repair, a *K. marxianus* strain overexpressing the *RAD52* recombinase from *S. cerevisiae*, was constructed. The strain showed a seven-fold increase in HR-mediated repair when compared to the

wild-type strain NBRC 1777. Though these results are promising, a seven-fold increase contrasts with other studies where a 37-fold increase in HR-mediated repair has been observed (Di Primio *et al*., 2005; Kalvala *et al*., 2010). A possible explanation for these results is that another gene, other than *RAD52*, is poorly active in *K. marxianus*, limiting HR repair. In this scenario, overexpression of *RAD52* increases HR repair to a certain level but does not fully compensate the low activity of other components of the HR machinery. Heterologous expression of other genes involved in HR could help in establishing if this is the case in *K. marxianus*.

Altogether, the results obtained in this study show that the CRISPR-Cas9 system can be efficiently used to genome engineer *K. marxianus*. The expression system developed can also be used in *K. lactis*, opening new possibilities for comparative studies involving both yeasts. Homologous recombination-mediated repair was optimised in *K. marxianus* by inactivating the NHEJ system and expressing the *RAD52* recombinase from *S. cerevisiae*. The tools developed here constitute significant progress in obtaining a comprehensive toolbox to genome engineer *K. marxianus* to enhance fundamental knowledge and for biotechnological applications.

2.5 References

Abdel-Banat, B.M.A., Nonklang, S., Hoshida, H., Akada, R., 2009. Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. Yeast 27, 29–39. doi:10.1002/yea.1729

Almeida, C., Queirós, O., Wheals, A., Teixeira, J., Moradas-Ferreira, P., 2003. Acquisition of flocculation phenotype by *Kluyveromyces marxianus* when overexpressing *GAP1* gene encoding an isoform of glyceraldehyde-3-phosphate dehydrogenase. J. Microbiol. Methods 55, 433–440. doi:10.1016/S0167-7012(03)00189-1

Banerjee, S., Smith, S., Myung, K., 2006. Suppression of gross chromosomal rearrangements by yKu70-yKu80 heterodimer through DNA damage checkpoints. Proc. Natl. Acad. Sci. 103, 1816–1821. doi:10.1073/pnas.0504063102

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., Zhao, H., 2015. Homology-Integrated CRISPR–Cas (HI-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*. ACS Synth. Biol. 4, 585–594. doi:10.1021/sb500255k

Cao, M., Gao, M., Lopez-Garcia, C.L., Wu, Y., Seetharam, A.S., Severin, A.J., Shao, Z., 2017. Centromeric DNA Facilitates Nonconventional Yeast Genetic Engineering. ACS Synth. Biol. 6, 1545–1553. doi:10.1021/acssynbio.7b00046

Carter, S.D., Iyer, S., Xu, J., McEachern, M.J., Aström, S.U., 2007. The role of nonhomologous end-joining components in telomere metabolism in *Kluyveromyces lactis*. Genetics 175, 1035–45. doi:10.1534/genetics.106.067447

Cheon, Y., Kim, J.-S., Park, J.-B., Heo, P., Lim, J.H., Jung, G.Y., Seo, J.-H., Park, J.H., Koo, H.M., Cho, K.M., Park, J.-B., Ha, S.-J., Kweon, D.-H., 2014. A biosynthetic pathway for hexanoic acid production in *Kluyveromyces marxianus*. J. Biotechnol. 182–183, 30–36. doi:10.1016/j.jbiotec.2014.04.010

Choo, J.H., Han, C., Kim, J.-Y., Kang, H.A., 2014. Deletion of a *KU80* homolog enhances homologous recombination in the thermotolerant yeast *Kluyveromyces marxianus*. Biotechnol. Lett. 36, 2059–2067. doi:10.1007/s10529-014-1576-4

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 339, 819–823. doi:10.1126/science.1231143

Daley, J.M., Palmbos, P.L., Wu, D., Wilson, T.E., 2005. Nonhomologous End Joining in Yeast. Annu. Rev. Genet. 39, 431–451. doi:10.1146/annurev.genet.39.073003.113340

Di Primio, C., Galli, A., Cervelli, T., Zoppè, M., Rainaldi, G., 2005. Potentiation of gene targeting in human cells by expression of *Saccharomyces cerevisiae* Rad52. Nucleic Acids Res. 33, 4639–48. doi:10.1093/nar/gki778

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., Church, G.M., 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 41, 4336–4343. doi:10.1093/nar/gkt135

Engler, C., Kandzia, R., Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS One 3, e3647. doi:10.1371/journal.pone.0003647

Fonseca, G.G., Gombert, A.K., Heinzle, E., Wittmann, C., 2007. Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. FEMS Yeast Res. 7, 422–435. doi:10.1111/j.1567-1364.2006.00192.x

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. doi:10.1007/s00253-008-1458-6

Gao, S., Tong, Y., Wen, Z., Zhu, L., Ge, M., Chen, D., Jiang, Y., Yang, S., 2016. Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. J. Ind. Microbiol. Biotechnol. 43, 1085–1093. doi:10.1007/s10295-016-1789-8

Gao, Y., Zhao, Y., 2014. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J. Integr. Plant Biol. 56, 343–9. doi:10.1111/jipb.12152

Gethins, L., Rea, M.C., Stanton, C., Ross, R.P., Kilcawley, K., O'Sullivan, M.,

Crotty, S., Morrissey, J.P., 2016. Acquisition of the yeast *Kluyveromyces marxianus* from unpasteurised milk by a kefir grain enhances kefir quality. FEMS Microbiol. Lett. 363, fnw165. doi:10.1093/femsle/fnw165

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. doi:10.1038/nmeth.1318

Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31–34. doi:10.1038/nprot.2007.13

Gorter de Vries, A.R., de Groot, P.A., van den Broek, M., Daran, J.-M.G., 2017. CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. Microb. Cell Fact. 16, 222. doi:10.1186/s12934-017-0835-1

Heinisch, J.J., Buchwald, U., Gottschlich, A., Heppeler, N., Rodicio, R., 2010. A tool kit for molecular genetics of *Kluyveromyces lactis* comprising a congenic strain series and a set of versatile vectors. FEMS Yeast Res. 10, 333–342. doi:10.1111/j.1567-1364.2009.00604.x

Horwitz, A.A., Walter, J.M., Schubert, M.G., Kung, S.H., Hawkins, K., Platt, D.M., Hernday, A.D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S.S., Newman, J.D., 2015. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Syst. 1, 88–96. doi:10.1016/j.cels.2015.02.001

Hoshida, H., Murakami, N., Suzuki, A., Tamura, R., Asakawa, J., Abdel-Banat, B.M.A., Nonklang, S., Nakamura, M., Akada, R., 2014. Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*. Yeast 31, 29–46. doi:10.1002/yea.2993

Hsu, P.D., Lander, E.S., Zhang, F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157, 1262–1278. doi:10.1016/j.cell.2014.05.010

Iiizumi, S., Kurosawa, A., So, S., Ishii, Y., Chikaraishi, Y., Ishii, A., Koyama, H., Adachi, N., 2008. Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting. Nucleic Acids

Res. 36, 6333–42. doi:10.1093/nar/gkn649

Inokuma, K., Ishii, J., Hara, K.Y., Mochizuki, M., Hasunuma, T., Kondo, A., 2015. Complete Genome Sequence of *Kluyveromyces marxianus* NBRC1777, a Nonconventional Thermotolerant Yeast. Genome Announc. 3, e00389-15. doi:10.1128/genomeA.00389-15

Jasin, M., Rothstein, R., 2013. Repair of Strand Breaks by Homologous Recombination. Cold Spring Harb. Perspect. Biol. 5, a012740–a012740. doi:10.1101/cshperspect.a012740

Jeong, H., Lee, D.-H., Kim, S.H., Kim, H.-J., Lee, K., Song, J.Y., Kim, B.K., Sung, B.H., Park, J.C., Sohn, J.H., Koo, H.M., Kim, J.F., 2012. Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var. marxianus KCTC 17555. Eukaryot. Cell 11, 1584–5. doi:10.1128/EC.00260-12

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 337, 816–821. doi:10.1126/science.1225829

Kalvala, A., Rainaldi, G., Di Primio, C., Liverani, V., Falaschi, A., Galli, A., 2010. Enhancement of gene targeting in human cells by intranuclear permeation of the *Saccharomyces cerevisiae* Rad52 protein. Nucleic Acids Res. 38, e149–e149. doi:10.1093/nar/gkq486

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., Han, S., Jeffery, L., Baek, S.-T., Lee, H., Shim, Y.S., Lee, M., Kim, L., Heo, K.-S., Noh, E.J., Lee, A.-R., Jang, Y.-J., Chung, K.-S., Choi, S.-J., Park, J.-Y., Park, Y., Kim, H.M., Park, S.-K., Park, H.-J., Kang, E.-J., Kim, H.B., Kang, H.-S., Park, H.-M., Kim, K., Song, K., Song, K. Bin, Nurse, P., Hoe, K.-L., 2010. Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. Nat. Biotechnol. 28, 617–623. doi:10.1038/nbt.1628

Kooistra, R., Hooykaas, P.J.J., Steensma, H.Y., 2004. Efficient gene targeting in *Kluyveromyces lactis*. Yeast 21, 781–792. doi:10.1002/yea.1131

Kuijpers, N.G.A., Chroumpi, S., Vos, T., Solis-Escalante, D., Bosman, L., Pronk, J.T., Daran, J.-M., Daran-Lapujade, P., 2013. One-step assembly and targeted

integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. FEMS Yeast Res. 13, 769–81. doi:10.1111/1567-1364.12087

Kuijpers, N.G., Solis-Escalante, D., Bosman, L., van den Broek, M., Pronk, J.T., Daran, J.-M., Daran-Lapujade, P., 2013. A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. Microb. Cell Fact. 12, 47. doi:10.1186/1475-2859-12-47

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*. Antonie Van Leeuwenhoek 100, 507–19. doi:10.1007/s10482-011-9606-x

Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. Fungal Biol. Rev. 24, 17–26. doi:10.1016/j.fbr.2010.01.001

Lee, M.E., DeLoache, W.C., Cervantes, B., Dueber, J.E., 2015. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth. Biol. 4, 975–986. doi:10.1021/sb500366v

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Suprayogi, Tsuchikane, K., Limtong, S., Fujita, N., Yamada, M., 2015. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. Biotechnol. Biofuels 8, 47. doi:10.1186/s13068-015-0227-x

Liachko, I., Dunham, M.J., 2014. An autonomously replicating sequence for use in a wide range of budding yeasts. FEMS Yeast Res. 14, 364–367. doi:10.1111/1567-1364.12123

Löbs, A.-K., Engel, R., Schwartz, C., Flores, A., Wheeldon, I., 2017. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol. Biofuels 10, 164. doi:10.1186/s13068-017-0854-5

Lopandic, K., Zelger, S., Bánszky, L.K., Eliskases-Lechner, F., Prillinger, H., 2006.

Identification of yeasts associated with milk products using traditional and molecular techniques. Food Microbiol. 23, 341–350. doi:10.1016/J.FM.2005.05.001

Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. Algorithms Mol. Biol. 6, 26. doi:10.1186/1748-7188-6-26

Ma, Y., Lu, H., Tippin, B., Goodman, M.F., Shimazaki, N., Koiwai, O., Hsieh, C.-L., Schwarz, K., Lieber, M.R., 2004. A Biochemically Defined System for Mammalian Nonhomologous DNA End Joining. Mol. Cell 16, 701–713. doi:10.1016/j.molcel.2004.11.017

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013. RNA-Guided Human Genome Engineering via Cas9. Science. 339, 823–826. doi:10.1126/science.1232033

Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G.A., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., Daran, J.-M.G., 2015. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. 15. doi:10.1093/femsyr/fov004

Mazzoni, C., Serafini, A., Falcone, C., 2005. The inactivation of KlNOT4, a *Kluyveromyces lactis* gene encoding a component of the CCR4-NOT complex, reveals new regulatory functions. Genetics 170, 1023–32. doi:10.1534/genetics.105.041863

Nambu-Nishida, Y., Nishida, K., Hasunuma, T., Kondo, A., 2017. Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. Sci. Rep. 7, 8993. doi:10.1038/s41598-017-08356-5

Ng, H., Dean, N., 2017. Dramatic Improvement of CRISPR/Cas9 Editing in *Candida albicans* by Increased Single Guide RNA Expression. mSphere 2, e00385-16. doi:10.1128/mSphere.00385-16

Nonklang, S., Abdel-Banat, B.M.A., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., Yamada, M., Akada, R., 2008. High-Temperature Ethanol

Fermentation and Transformation with Linear DNA in the Thermotolerant Yeast *Kluyveromyces marxianus* DMKU3-1042. Appl. Environ. Microbiol. 74, 7514–7521. doi:10.1128/AEM.01854-08

Nonklang, S., Ano, A., Abdel-Banat, B.M.A., Saito, Y., Hoshida, H., Akada, R., 2009. Construction of flocculent *Kluyveromyces marxianus* strains suitable for high-temperature ethanol fermentation. Biosci. Biotechnol. Biochem. 73, 1090–5.

Numamoto, M., Maekawa, H., Kaneko, Y., 2017. Efficient genome editing by CRISPR/Cas9 with a tRNA-sgRNA fusion in the methylotrophic yeast *Ogataea polymorpha*. J. Biosci. Bioeng. 124, 487–492. doi:10.1016/j.jbiosc.2017.06.001

Palmbos, P.L., Daley, J.M., Wilson, T.E., 2005. Mutations of the Yku80 C Terminus and Xrs2 FHA Domain Specifically Block Yeast Nonhomologous End Joining. Mol. Cell. Biol. 25, 10782–10790. doi:10.1128/MCB.25.24.10782-10790.2005

Palmbos, P.L., Wu, D., Daley, J.M., Wilson, T.E., 2008. Recruitment of *Saccharomyces cerevisiae* Dnl4-Lif1 complex to a double-strand break requires interactions with Yku80 and the Xrs2 FHA domain. Genetics 180, 1809–19. doi:10.1534/genetics.108.095539

Ribeiro, O., Gombert, A.K., Teixeira, J.A., Domingues, L., 2007. Application of the Cre-*loxP* system for multiple gene disruption in the yeast *Kluyveromyces marxianus*. J. Biotechnol. 131, 20–26. doi:10.1016/j.jbiotec.2007.05.027

Rocha, S.N., Abrahão-Neto, J., Gombert, A.K., 2011. Physiological diversity within the *Kluyveromyces marxianus* species. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 100, 619–630. doi:10.1007/s10482-011-9617-7

Rodicio, R., Heinisch, J.J., Heinisch, J.J., 2013. Yeast on the milky way: genetics, physiology and biotechnology of *Kluyveromyces lactis*. doi:10.1002/yea.2954

Roman, H., 1956. A system selective for mutations affecting the synthesis of adenine in yeast. C. R. Trav. Lab. Carlsb. 299–314.

Ryan, O.W., Skerker, J.M., Maurer, M.J., Li, X., Tsai, J.C., Poddar, S., Lee, M.E., DeLoache, W., Dueber, J.E., Arkin, A.P., Cate, J.H.D., 2014. Selection of chromosomal DNA libraries using a multiplex CRISPR system. Elife 3. doi:10.7554/eLife.03703

Schwartz, C.M., Hussain, M.S., Blenner, M., Wheeldon, I., 2016. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. ACS Synth. Biol. 5, 356–359. doi:10.1021/acssynbio.5b00162

Sheetz, R.M., Dickson, R.C., 1980. Mutations affecting synthesis of beta-galactosidase activity in the yeast *Kluyveromyces lactis*. Genetics 95, 877–90.

Shrivastav, M., De Haro, L.P., Nickoloff, J.A., 2008. Regulation of DNA double-strand break repair pathway choice. Cell Res. 18, 134–147.

Silveira, W.B., Diniz, R.H.S., Cerdan, M.E., Gonzalez-Siso, M.I., Souza, R. d. A., Vidigal, P.M.P., Brustolini, O.J.B., de Almeida Prata, E.R.B., Medeiros, A.C., Paiva, L.C., Nascimento, M., Ferreira, E.G., dos Santos, V.C., Braganca, C.R.S., Fernandes, T.A.R., Colombo, L.T., Passos, F.M.L., 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, e01136-14-e01136-14. doi:10.1128/genomeA.01136-14

Singh, M. V, Weil, P.A., 2002. A method for plasmid purification directly from yeast. Anal. Biochem. 307, 13–7.

Steinborn, G., Böer, E., Scholz, A., Tag, K., Kunze, G., Gellissen, G., 2006. Application of a wide-range yeast vector (CoMed) system to recombinant protein production in dimorphic *Arxula adeninivorans*, methylotrophic *Hansenula polymorpha* and other yeasts. Microb. Cell Fact. 5, 33. doi:10.1186/1475-2859-5-33

Suzuki, A., Fujii, H., Hoshida, H., Akada, R., 2015. Gene expression analysis using strains constructed by NHEJ-mediated one-step promoter cloning in the yeast *Kluyveromyces marxianus*. FEMS Yeast Res. 15, fov059.

Suzuki, T., Hoshino, T., Matsushika, A., 2014. Draft Genome Sequence of *Kluyveromyces marxianus* Strain DMB1, Isolated from Sugarcane Bagasse Hydrolysate. Genome Announc. 2, e00733-14-e00733-14. doi:10.1128/genomeA.00733-14

Terentiev, Y., Pico, A.H., Böer, E., Wartmann, T., Klabunde, J., Breuer, U., Babel, W., Suckow, M., Gellissen, G., Kunze, G., 2004. A wide-range integrative yeast

expression vector system based on *Arxula adeninivorans*-derived elements. J. Ind. Microbiol. Biotechnol. 31, 223–228.

Tsakraklides, V., Brevnova, E., Stephanopoulos, G., Shaw, A.J., 2015. Improved Gene Targeting through Cell Cycle Synchronization. PLoS One 10, e0133434. doi:10.1371/journal.pone.0133434

Vyas, V.K., Barrasa, M.I., Fink, G.R., 2015. A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. Sci. Adv. 1, e1500248–e1500248. doi:10.1126/sciadv.1500248

Wach, A., Brachat, A., Pöhlmann, R., Philippsen, P., 1994. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. Yeast 10, 1793–808.

Weninger, A., Hatzl, A.-M., Schmid, C., Vogl, T., Glieder, A., 2016. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. J. Biotechnol. 235, 139–149. doi:10.1016/j.jbiotec.2016.03.027

Wésolowski-Louvel, M., 2011. An efficient method to optimize *Kluyveromyces lactis* gene targeting. FEMS Yeast Res. 11, 509–513.

West, R.W., Yocum, R.R., Ptashne, M., 1984. *Saccharomyces cerevisiae GAL1-GAL10* divergent promoter region: location and function of the upstream activating sequence UASG. Mol. Cell. Biol. 4, 2467–78. doi:10.1128/MCB.4.11.2467

Xie, S., Shen, B., Zhang, C., Huang, X., Zhang, Y., 2014. sgRNAcas9: A Software Package for Designing CRISPR sgRNA and Evaluating Potential Off-Target Cleavage Sites. PLoS One 9, e100448. doi:10.1371/journal.pone.0100448

Yarimizu, T., Nonklang, S., Nakamura, J., Tokuda, S., Nakagawa, T., Lorreungsil, S., Sutthikhumpha, S., Pukahuta, C., Kitagawa, T., Nakamura, M., Cha-aim, K., Limtong, S., Hoshida, H., Akada, R., 2013. Identification of auxotrophic mutants of the yeast *Kluyveromyces marxianus* by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*. Yeast 30, 485–500. doi:10.1002/yea.2985

Chapter 3

Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains

This chapter was published in its totality in the following paper:

Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains. **Varela J.A.**, Montini N, Scully D, Van der Ploeg R, Oreb M, Boles E, Hirota J, Akada R, Hoshida H, Morrissey JP. *FEMS Yeast Res.* 2016. May 1;17(3). DOI: 10.1093/femsyr/fox021

Abstract

Kluyveromyces marxianus is a safe yeast used in the food and biotechnology sectors. One of the important traits that sets it apart from the familiar yeasts, *Saccharomyces cerevisiae*, is its capacity to grow using lactose as a carbon source. Like in its close relative, *Kluyveromyces lactis*, this requires lactose transport via a permease and intracellular hydrolysis of the disaccharide. Given the importance of the trait, it was intriguing that most, but not all, strains of *K. marxianus* are reported to consume lactose efficiently. In this study, primarily through heterologous expression in *S. cerevisiae* and *K. marxianus*, it was established that a single gene, *LAC12*, is responsible for lactose uptake in *K. marxianus*. Strains that failed to transport lactose showed variation in 13 amino acids in Lac12p, rendering the protein non-functional for lactose transport. Genome analysis showed that the *LAC12* gene is present in 4 copies in the sub-telomeric regions of three different chromosomes but only the ancestral *LAC12* gene encodes a functional lactose transporter. Other copies of *LAC12* may be non-functional or have alternative substrates. The analysis raises some interesting questions regarding the evolution of sugar transporters in *K. marxianus*. The *LAC12* genes could have duplicated and functionally diverged to transport other sugars. Also, it is not clear whether the *KmLAC12* versions identified in this study represent an adaptation of some strains to dairy or two distinct steps in the evolution of lactose utilisation in *K. marxianus.*

3.1 Introduction

Kluyveromyces marxianus is a yeast commonly isolated from food, fruit and plant material. It has frequently been isolated from fermented dairy products, such as cheese, yoghurt and kefir, where it is associated with production of pleasant flavours and other traits (Morrissey *et al*., 2015). It is closely related to *Kluyveromyces lactis*, itself considered to be a milk or dairy yeast (Lachance, 2007; Rodicio *et al*., 2013). Because of its history of association with food, *K. marxianus,* like *K. lactis*, has European Food Safety Authority (http://www.efsa.europa.eu/) Qualified Presumption of Safety (QPS) status and is a Generally Regarded as Safe (GRAS) organism. Despite the close relationship between *K. marxianus* and *K. lactis*, there are important phenotypic differences that have particular implications when considering the yeast for biotechnological applications (Fonseca *et al*., 2008; Lane and Morrissey, 2010). For instance, *K. marxianus* encodes an inulinase enzyme that gives the yeast the capacity to hydrolyse complex plant fructans (Arrizon *et al*., 2012, 2011), and it also has the capacity to grow at temperatures of 44 °C and above, both traits absent in *K. lactis* (Nonklang *et al*., 2009). *K. marxianus* is a respire-fermentative yeast and some strains are also reasonably good ethanol producers. Consequentially, there have been many studies seeking to exploit the beneficial traits of *K. marxianus* for bioethanol production. For example, ethanol has been produced from different substrates such as glucose (Castro and Roberto, 2014), lignocellulosic material (Goshima *et al*., 2013) and Jerusalem artichoke (Hu *et al*., 2012).

The capacity of *K. marxianus* and *K. lactis* to utilise lactose is one of the phenotypes that distinguishes them from *Saccharomyces cerevisiae,* which lacks this feature. Important biotechnology applications derive from this: namely, production of the enzyme β-galactosidase for the food industry, and remediation of whey by lactose removal (Rubio-Texeira, 2006). Whey is the waste material remaining from cheese manufacture and was long considered a problem until it was realised that whey permeate (deproteinised whey), which contains up to 5% residual lactose, can serve as a cheap substrate for industrial applications. As a result, *K. marxianus* is now used commercially for production of ethanol and yeast biomass from whey permeate (reviewed by Guimarães *et al*., 2010). These are, however, low value bulk products and the interest in yeast biotechnology is now to engineer strains to produce higher value products like flavours/aromas or low molecular weight bioactives (Morrissey *et*

al., 2015). An ideal scenario from a biotechnology perspective is to be able to grow a yeast on a cheap substrate like whey permeate to produce high-value products and so the capacity to transport and metabolise lactose is very important.

Lactose utilisation in yeast involves transport of the sugar into the cell, intracellular hydrolysis of lactose into glucose and galactose, which are then metabolised through the glycolytic and Leloir pathways. The genetics of lactose assimilation have been intensively studied in *K. lactis*, where it is shown that the *LAC12 – LAC4* gene pair encodes a lactose permease and a β-galactosidase enzyme, respectively (Godecke *et al*., 1991). Lac12p is a member of the major facilitator superfamily (MFS) and works via a proton-symport mechanism. It is saturable at high substrate concentrations and is also reported to transport galactose. Therefore, it is considered a high-affinity lactose transporter and low-affinity galactose transporter in *K. lactis* (Riley *et al*., 1987). *K. lactis* also has at least one other transporter capable of transporting galactose but Lac12p is the only lactose transporter described (Baruffini *et al*., 2006). Notably, *Kluyveromyces* spp. lack the *GAL2* gene, which encodes the main galactose transporter in *S. cerevisiae* (Tschopp *et al*., 1986). *LAC12* and *LAC4* are divergently transcribed from a common promoter where key regulatory elements are located. This 2.8 kb *LAC12-LAC4* intergenic region contains several binding sites for Lac9p, which is the orthologue of the well-studied *S. cerevisiae* transcriptional activator, Gal4p (Godecke *et al*., 1991), as well as binding sites for the transcriptional repressor Mig1p. *LAC12* and *LAC4* are therefore strongly induced by lactose or galactose in the growth medium but potentially repressed by the presence of glucose.

The same genes, *LAC12* and *LAC4*, are present in *K. marxianus* and these are believed to account for the capacity of this yeast to utilise lactose. Several studies, however, suggest that efficient lactose-utilisation is not universally found in *K. marxianus*. For instance, differences in biomass production by *K. marxianus* strains grown in whey permeate have been reported (Grba *et al*., 2002). These findings are further supported by evidence suggesting that some *K. marxianus* strains lack lactose-proton symport activity and do not ferment lactose (Carvalho-Silva and Spencer-Martins, 1990). More recently, studies under defined media conditions reported that some strains either grew very slowly, or not at all, when provided with lactose as the sole carbon source (Lane *et al*., 2011; Rocha *et al*., 2011). Publication of *K. marxianus* genome sequences has highlighted some interesting questions regarding lactose utilisation in *K. marxianus*.

For example, it was reported that strain DMKU3-1042 encodes 4 genes that are highly homologous to the *LAC12* gene (Lertwattanasakul *et al*., 2015).

Given the importance of Lac12p for lactose transport and biotechnological applications of *K. marxianus*, we were interested in determining the reason for the phenotypic variability among a number of strains for this trait. The availability of draft and completed genome sequences enabled us to build a more complete view of the potential lactose transport genes in this yeast. Through a series of phenotypic, biochemical and genetic approaches, it was possible to determine that there is a single functional lactose transporter in *K. marxianus* and, in some strains, this protein no longer transports lactose into the cell. Apart from immediate relevance when engineering strains for applications, the work opens some interesting questions regarding the evolution of yeast sugar transporters.

3.2 Materials and methods

Strains and culture conditions

The yeast strains used in study are listed in **Table 3.1**. Strains were typically cultured in YP medium (10 g/L yeast extract, 20 g/L bactopeptone) supplemented with 20 g/L glucose, lactose or raffinose. To determine growth kinetics and sugar transport, CBS 397 and CBS 6556 were grown on minimal media (MM) (Fonseca *et al*., 2007). For β - galactosidase measurements, cells were grown to mid exponential phase (A_{600} 2) on YPD and YPLac media. The *K. lactis* strain CBS 2359 was used as a control in this experiment. Enzymatic activity was measured as described by Ribeiro *et al*., 2007. *S. cerevisiae* EBY.VW4000 and BY4741 were used for heterologous expression and cellular localization experiments, respectively. EBY.VW4000 was grown in synthetic complete (SC) medium (1.7 g/L yeast nitrogen base, 5 g/L ammonium sulphate plus synthetic complete drop-out lacking L-leucine and uracil) (Formedium, Norfolk, United Kingdom). The growth medium was supplemented with 380 mg/L L-leucine and 76 mg/L uracil (final concentration) when required. Maltose or lactose was added as carbon source to a final concentration of 20 g/L. *S. cerevisiae* BY4741 was typically grown in YP medium supplemented with 20 g/L glucose or galactose and 200 µg/mL G418, when required for plasmid selection. Yeast strains were grown at 30 °C with 200 rpm agitation in a New Brunswick Innova 40/40R shaker (Eppendorf, Hamburg, Germany). Cloning experiments were performed using *Escherichia coli* DH5α. *E. coli* was routinely cultured in LB medium (5 g/L yeast extract, 10 g/L bactopeptone, 10 g/L NaCl) supplemented with 100 µg/mL ampicillin when required.

Table 3.1: Yeast strains used in this study

Species	Strain	Relevant genotype	Source	Origin
K. marxianus	CBS 397	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
K. marxianus	CBS 608	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
K. marxianus	CBS 5795	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
K. marxianus	CBS 6432	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Not specified
K. marxianus	CBS 1555	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
K. marxianus	CBS 6556	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dough
K. marxianus	CBS 5670	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Winery
K. marxianus	CBS 4857	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
K. marxianus	CBS 712	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Not specified
K. marxianus	CBS 7894	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Not specified
K. marxianus	CBS 7858	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Food
K. marxianus	CBS 4354	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
K. marxianus	CBS 2233	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
K. marxianus	CBS 1596	Wild-type strain	CBS-KNAW Fungal Biodiversity Centre	Not specified
K. marxianus	NCYC 179	Wild-type strain	National Collection of Yeast Cultures	Dairy

Species	Strain	Relevant genotype	Source	Origin
K. marxianus	NCYC 1429	Wild-type strain	National Collection of Yeast Cultures	Not specified
K. marxianus	DMKU 3 - 1042	Wild-type strain	NITE Biological Resource Center	Bioethanol plant
K. marxianus	RAK 3605	Δ*URA3*	NITE Biological Resource Center	Lab strain
K. marxianus	RAK 4072	Δ*URA5*	NITE Biological Resource Center	Lab strain
K. lactis	CBS 2359	Wild type strain	CBS-KNAW Fungal Biodiversity Centre	Dairy
S. cerevisiae	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EUROSCARF	Lab strain
S. cerevisiae	EBY.VW4000	EBY.VW1000 stl1v::loxP agt1v: :loxP ydl247wv: :loxP yjr160cv: :loxP	Boles group, Goethe-Universität Frankfurt, Germany	Lab strain

Analytic methods for sugar detection

Lactose concentration was determined by high-performance liquid chromatograph (HPLC). An Agilent 1200 HPLC system (Agilent Technologies, CA, USA) equipped with a refractive index detector and a REZEX 8 μL 8 % H^+ organic acid column (300 x 7.8 mm) (Phenomenex, CA, USA) was used to separate and detect the analytes. The column was eluted with 0.01 N H_2SO_4 at a flow rate of 0.6 mL/min. The temperature of the column was maintained at 65 °C. Lactose consumption was calculated by subtracting the lactose concentration at the end and the beginning of the experiment.

Sugar uptake assays to measure uptake of lactose and glucose were performed as previously described (Farwick *et al*., 2014). *K. marxianus* CBS 397 and CBS 6556 were grown on MM supplemented with glucose or lactose, as appropriate, and collected at an A_{600} of 2, which corresponds to mid exponential phase. Cells were centrifuged and then washed twice in cold uptake buffer (100 mM potassium phosphate, pH 6.5). Cell aliquots were equilibrated at 30 °C and then mixed with 100 mM radiolabelled glucose or lactose (American Radiolabeled Chemicals, MO, USA), for 15 seconds. Radioactivity was measured in a Wallac 1409 liquid scintillation counter. Counts from the unfiltered sample, referred as total counts, and from non-specific binding, were also determined. Intracellular sugar concentration was established by subtracting the non-specific binding counts to the filter counts and dividing this value by the total counts. Finally, uptake velocity was calculated by dividing the intracellular sugar concentration by the dry weight and multiplying this value by the time of the assay. All experiments were carried out as 3 independent replicates.

Bioinformatic analysis

The unannotated genome sequence from the *K. marxianus* strain CBS 6556, also known as KCTC 17555, is publicly available (Jeong *et al*., 2012). Scaffold sequences (assembly GCA_000299195.2) were retrieved from the NCBI database and then submitted to the YGAP server (Yeast Genome Annotation Pipeline) for gene prediction (Proux-Wéra *et al*., 2012). Predicted proteins were then annotated using Blast2GO v 2.7.2 (Conesa *et al*., 2005). The genome of the *K. marxianus* strain DMKU3-1042, used for high-temperature ethanol fermentations, was used for genome comparisons (Lertwattanasakul *et al*., 2015; Nonklang *et al*., 2008). Clustering of the

full set of genes of *K. marxianus* CBS 6556 and *K. marxianus* DMKU3-1042 (assembly GCA_001417885.1) was computed using OrthoMCL v 5 (Li *et al*., 2003).

To investigate sequence variation in the *LAC12* genes, the fragments of interest were amplified by PCR, sequenced and then multiple sequence alignments were created using AliView (Larsson, 2014). Maximum likelihood trees of the Lac12p sequences were computed using MEGA v 6.0.6 (Tamura *et al*., 2013). The CBS 6556 and CBS 397 *LAC12* sequences can be found in the EBI site under the ID numbers LT708111-LT708118

Functional expression of *LAC12* genes

All the plasmids used in this study are listed in **Table 3.2**. All yeast transformations were performed using the standard Lithium Acetate procedure (Gietz and Schiestl, 2007). Plasmids for functional expression of *LAC12* genes in *S. cerevisiae* were constructed using the pGREG-505 backbone (Jansen *et al*., 2005). The *GAL1* promoter in pGREG-505 was replaced with the constitutive *TEF1* promoter. The *TEF1* sequence was amplified from BY4741, digested with NotI/SacI and cloned into the similarly cut pGREG-505 using T4 DNA ligase (New England Biolabs Inc., MA, USA) to yield plasmid pGREG-505-TEF1. The *LAC12* genes were cloned into pGREG505-TEF1 by *in vivo* recombination in *S. cerevisiae*. The four *LAC12* open reading frames were amplified by PCR from CBS 397 and CBS 6556 using Phusion Polymerase (New England Biolabs Inc., MA, USA). Then, 1 µL of each product was mixed with 200 ng of *SalI*-digested pGREG505-TEF1 and transformed into *S. cerevisiae* EBY.VW4000**.** Cells were plated on SC media lacking leucine. Colonies were screened by replica plating in –His media and colony PCR. Plasmids were recovered from yeast cultures and transformed into *E. coli* as described elsewhere (Singh and Weil, 2002).

Table 3.2: Plasmids used in this study

Plasmid	**Source**
pGREG-505	EUROSCARF
pGREG-506	EUROSCARF
pGREG-574	EUROSCARF
pGREG-506 - TEF1 - *LAC4*	This study
pGREG-505 - TEF1	This study
pGREG-505 - TEF1 - *KmLAC12* - CBS 6556	This study
pGREG-505 - TEF1 - *LAC12-2* - CBS 6556	This study
pGREG-505 - TEF1 - *LAC12-3* - CBS 6556	This study
pGREG-505 - TEF1 - *LAC12-4* - CBS 6556	This study
pGREG-505 - TEF1 - *KmLAC12* - CBS 397	This study
pGREG-505 - TEF1 - *LAC12-2* - CBS 397	This study
pGREG-505 - TEF1 - *LAC12-3* - CBS 397	This study
pGREG-505 - TEF1 - *LAC12-4* - CBS 397	This study
pGREG-574 - GFP::*KmLAC12* - CBS 6556	This study
pGREG-574 - GFP::*KmLAC12* - CBS 397	This study
p11256	Hoshida *et al.,* 2014
p11256 - *KmLAC12* - DMKU3-1042	This study
p11256 - *LAC12-2* - DMKU3 - 1042	This study
p11256 - *LAC12-3* - DMKU3 - 1042	This study
p11256 - *LAC12-4* - DMKU3 - 1042	This study
p11256 - *KmLAC12* - NCYC 1429	This study
p11256 - *LAC12-2* - NCYC 1429	This study
p11256 - *LAC12-3* - NCYC1429	This study
p11256 - *LAC12-4* - NCYC1429	This study

To express the β-galactosidase gene required for lactose hydrolysis, the pGREG506-TEF1-*LAC4* was constructed by recombinational cloning. Briefly, the *TEF1* promoter from *S. cerevisiae*, the *LAC4* gene from *K. marxianus* CBS 397 and the pGREG506 plasmid, containing the uracil maker, were PCR amplified using primers designed for *in vivo* assembly (**Table 3.3**). As mentioned previously, 1 µL of each product was transformed into *S. cerevisiae* EBY.VW4000 and the resulting plasmid was recovered. Finally, this plasmid was transformed into the EBY.VW4000 strain containing the different *LAC12* plasmids. Cells were plated in SC media lacking uracil and leucine.

The *LAC12* PCR products were also cloned into the pGREG574 plasmid to obtain in frame N-terminal GPF fusions. In this case, the *S. cerevisiae* BY4741 strain was used. *S. cerevisiae* BY4741 strains transformed with the pGREG574 – *LAC12* plasmids were grown in inducing YPGal medium to an A_{600} of 2. Cells were collected and washed twice with sterile water. Suspensions of 10 µL were then spread into glass slides. Samples were visualized using a Zeiss LSM 5 laser scanning confocal microscope. For detection of GFP fluorescence, samples were excited with laser light at 488 nm and emission light acquired using a bandpass emission filter of 505-550 nm. Samples were acquired and analysed using Zen 2008 SP2 software.

Table 3.3: Primers used in this study

Identification of variations in the *LAC12* sequences

Primer	Sequence	Source
KmLAC12-A-F	TCTGTGTGTTGTCTTATAATTAT	This study
KmLAC12-A-R	ATGAGTTCTATCGCCATTTAAG	This study
KmLAC12-B-F	AAACCCTCTAGAAATGTTAGATG	This study
KmLAC12-B-R	CTTTTTGCTGCAATTAGGAGAAG	This study
LAC12-2-A-F	ACTGATTATTACTCAACAAAAGT	This study
LAC12-2-A-R	ATGTTCCCTGTTACCATTGAAA	This study
LAC12-2-B-F	TTGGAGGTCTTAGACCTAAGA	This study
LAC12-2-B-R	TTAATTCACTGATCATATCCCA	This study
LAC12-3-A-F	TTCATCTAAAATATATGGGGGCC	This study
LAC12-3-A-R	GTGGTCTCTGTTACCCCCAAAA	This study
LAC12-3-B-F	TGCTCCACTAGAAGTTTTGAACC	This study
LAC12-3-2-R	TTCTCTTGCAAAAAAAAACGGAG	This study

LAC12-4-A-F	TTAATACTGCAAACTCTCAAAAA	This study
LAC12-4-A-R	GCGATGACCAATCCTAACCTG	This study
LAC12-4-B-F	CGGCATCAATATGTCCGCTCCGT	This study
LAC12-4-B-R	ATTTACATACTTTGATAGACGA	This study

Construction of pGREG plasmids

Primer	Sequence	Source
TEF1-SacI-F	ATGCGAGCTCCACACACCATAGCTTCAAA	This study
TEF1-NotI-R	ATGCGCGGCCGCTTTGTAATTAAAACTTAGATTAGATTGC	This study
KmLAC12-1-F	gaattcgatatcaagcttatcgataccgtcgATGGCAGATCATTCGAGCAGCTCA	This study
KmLAC12-1-R	gcgtgacataactaattacatgactcgaggtcgacTAAACAGATTCTGCCTCTGAAAGCTT	This study
LAC12-2-ORF-F	gaattcgatatcaagcttatcgataccgtcgATGGCAGATAAAATAAGTTCGAT	This study
LAC12-2-ORF-R	gcgtgacataactaattacatgactcgaggtcgacTCAAACTGATAGTGAATCAACCTCTT	This study
LAC12-3-ORF-F	gaattcgatatcaagcttatcgataccgtcgATGGCAGACACTTCTAGCTTGC	This study
LAC12-3-ORF-R	gcgtgacataactaattacatgactcgaggtcgacGACAGTCTCAAACGTACTCCTT	This study
LAC12-4-Short-ORF-F	gaattcgatatcaagcttatcgataccgtcgATGGACTGGAAAGGAAGAAAACCA	This study

LAC12-4-Full-ORF-F	gaattcgatatcaagcttatcgataccgtcgATGACAGAAAAGTGGAGTGAATCA	This study
LAC12-4-ORF-R	gcgtgacataactaattacatgactcgaggtcgacTTAAACTGATTCCAAATCGGAGACT	This study
TEF1_506-F	agctggagctcgtttaaacgCCACACACCATAGCTTCAAAATG	This study
TEF1_506-R	ggcaagacatCTTAGATTAGATTGCTATGCTTTCTTTC	This study
LAC4-F	ctaatctaagATGTCTTGCCTTATTCCTG	This study
LAC4-R	attacatgactcgaggtcgaTTATTCAAAAGCGAGATCAAAC	This study
pGREG506-F	TCGACCTCGAGTCATGTAATTAG	This study
pGREG506-R	CGTTTAAACGAGCTCCAG	This study

Primers used for overexpressing the *LAC12* copies in *K. marxianus*

Primer	Sequence	Source
KmLAC12+1[*1]	ATGGCAGATCATTCGAGCAGCTCAT	This study
LAC12 - 2+1[*1]	ATGACAGAAAAGTGGAGTGAATCAT	This study
LAC12 - 3+1[*1]	ATGGCAGACACTTCTAGCTTGCCTT	This study
LAC12 - 4+1[*1]	ATGGACTGGAAAGGAAGAAAACCAG	This study
3CG9-*KmLAC12*c[*1]	cccgggcccTTAAACAGATTCTGCCTCTGAAAGC	This study

3CG9-*LAC12 - 2*c[*1]	cccgggcccTCAAACTGATAGTGAATCAACCTCT	This study
3CG9-*LAC12 - 3*c[*1]	cccgggcccTCAAACGTACTCCTTTATTTTGATT	This study
3CG9-*LAC12 - 4*c[*1]	cccgggcccTTAAACTGATTCCAAATCGGAGACT	This study
KmURA3+772786TGA+3CG9	gcttacttgaagagaTGAcccgggccc	This study
URA3+771c	ttcccagcctgcttttctgtaacgt	This study
KmCDC19-1c	TGCGTACGGTTTTTTTTGTGTGAGTG	This study

Primers used in qPCR experiments

Primer	Sequence	Source
KmLAC12-Q-F	GATCAAGCATTCTTGGCTCA	This study
KmLAC12-Q-R	CATGATCAGCATCGCTCTTT	This study
LAC12-2-Q-F	GATTGGATGTTTCGGTGTTGT	This study
LAC12-2-Q-R	CAGCTGAATTGGCGATTGTA	This study
LAC12-3-Q-F	AAGTGTCGGCATGAAGTCTG	This study
LAC12-3-Q-R	CCTGCTTGATAACGTGCTGT	This study
LAC12-4-Q-F	GCTACAATTGCTTGCTCTGC	This study

LAC12-4-Q-R	CGCAGCAACAATAGATCCAC	This study
LAC4-Q-F	CCGTCATGAACCTCTTCCTT	This study
LAC4-Q-R	ACCAAGGTGGCTTTATCTGG	This study
ALG9-Q-F	CCATCTCAGGATCCCTCTTC	This study
ALG9-Q-R	GCATTCCAGCGAATAGTTGA	This study
ACT1-Q-F	GGCTGAACGTGGTTACTCCT	This study
ACT1-Q-R	AGAAGCGGTTTGCATTTCTT	This study

For expression in *K. marxianus*, the *KmLAC12* genes were cloned into YCp11256 (**Fig. 3.1**) under the *KmCDC19* promoter by using functional marker selection (Hoshida *et al*., 2014). The vector DNA was linearized and amplified with primers *Km*CDC19-1c and *URA3*+771c. *KmLAC12* genes from NCYC 1429 and DMKU3-1042 were amplified with primers shown in **Table 3.3**. The *URA3* C-terminal sequence was added to the resulting *KmLAC12* gene fragments by PCR using each +1 primer and *KmURA3*+772786TGA+3CG9. The linearized vector DNA and one of *KmLAC12* gene fragments were simply mixed and used for transformation of RAK3605, a *ura3-1* derivative of DMKU3-1042 (Nonklang *et al*., 2008). *In vivo* assembly using the NHEJ repair mechanism joined the fragments end to end, recreating an intact *URA3* gene and allowing selection of transformants on SC media lacking uracil. Correct cloning of each *KmLAC12* gene was confirmed by PCR and sequencing.

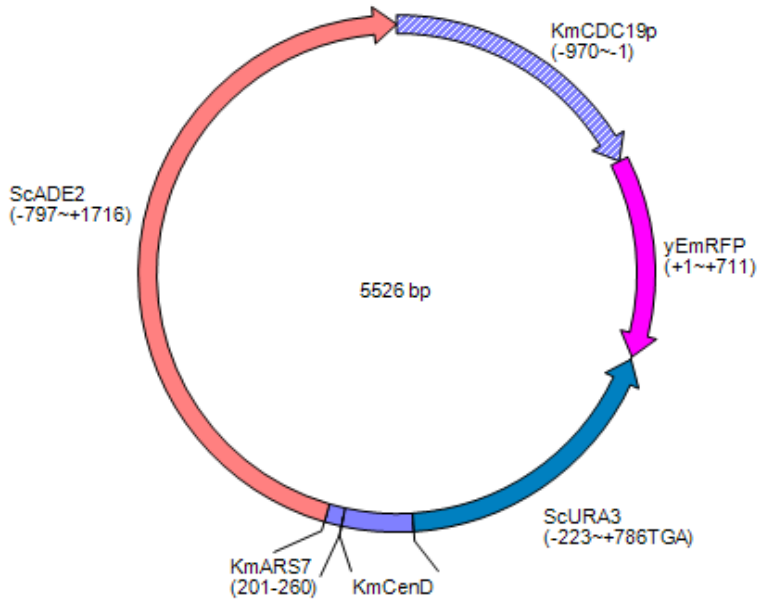


Figure 3.1: Plasmid YCp11256 use for overexpression of the *LAC12* genes. The *LAC12* genes were inserted between KmCDC19p and ScURA3 with substitution of yEmRFP. As result, the expression of the *LAC12* genes was driven by the *KmCDC19* promoter. The *KmCDC19p* sequence was obtained from *K. marxianus* DMKU3-1042. Sc*ADE2* and Sc*URA3* were obtained from conventional laboratory strains of *S. cerevisiae*. The plasmid replication elements, *KmARS7* and *KmCenD* were described by Hoshida *et al*., 2014.

Gene expression analysis

K. marxianus strains were grown in YPD, YPLac and YRaff to an A_{600} of 2. Cells were then recovered by centrifugation and RNA was extracted using the Ambion RiboPure RNA Purification Kit for Yeasts (Ambion, Austin, TX, USA). The extracted RNA was quantified using a Nanodrop spectrophotometer. cDNA was synthesized from 800 ng of RNA, using the ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs Inc., MA, USA). cDNA samples were diluted by adding 30 µL of nuclease-free water, resulting in a final volume of 50 µL per cDNA sample. Aliquots were then diluted 1/10 and stored at -20 °C. Real-time quantitative PCR (qPCR) was used to measure gene expression. cDNA samples were amplified by using LightCycler 480 SYBR Green I Master qPCR and ran in a LightCycler 480 real-time PCR machine. Primers were design to amplify the *ACT1, ALG9*, *LAC4,* and *LAC12* genes (**Table 3.3**). Standard curves were performed to check for amplification efficiency with all primers showing > 90 % efficiency. Additionally, melting curves were carried out to check reaction specificity and to ensure that each primer pair produced a single amplicon. *ALG9* (coding for alpha-1,2-mannosyltransferase) and *ACT1* (coding for actin) were used as reference genes, as both these genes are commonly used reference genes involved in cell processes, unrelated to lactose metabolism (Teste *et al*. 2009). To compare expression of the *LAC12* genes in control media varied between strains, the data was normalised against *ACT1* expression. Data was further calculated as $2^{-\Delta Ct}$ as described by (Pfaffl, 2001). *ALG9* was used as an internal control to ensure that the *ACT1* expression was constant in all samples.

Genetic crosses and spore analysis

RAK 3605 and RAK 4072 (DMKU3-1042 *ura3-1*, and NCYC1429 *ura5-2*, respectively, were crossed on 2% glucose medium and incubated at 30ºC for one day (Yarimizu *et al*., 2013). Yeast cells were replica-plated on a MM plate and incubated for two days to obtain zygotes. For sporulation, zygotes were suspended in 4 mL sterilized water at a cell concentration A_{600} 0.1. After two-day incubation, cells were spread on MM supplemented with uracil and 1 mg/mL 5-Fluoroorotic acid. One hundred ninety-two colonies were randomly selected and used for growth tests in YPLac medium.

3.3 Results

K. marxianus strains can be classified according to their lactose consumption capacities

Although the capacity to utilise lactose is considered to be a typical trait of *K. marxianus*, several studies have found this trait to be variable between strains (Carvalho-Silva and Spencer-Martins, 1990; Grba *et al*., 2002; Lane *et al*., 2011; Rocha *et al*., 2011). In order to establish the degree of variability, lactose utilisation was evaluated in a set of 15 *K. marxianus* strains derived from culture collections and originating from different niches (**Table 3.1**). Strains were grown on YPLac and growth was assessed after 11h (**Fig 3.2**, open bars). In the same experiment, the amount of lactose consumed after 11h of growth was determined (**Fig 3.2**, solid bars). In both cases, there was a very clear bimodal distribution with 6 strains (CBS 397, CBS 608, NCYC 179, CBS 5795, CBS 6432 and CBS 1555) growing to a higher density (~A_{600} 1.8) and consuming a higher amount of lactose and 9 strains (CBS 6556, CBS 5670, CBS 4857, CBS 712, CBS 7894, CBS 7858, CBS 1596, CBS 2233 and CBS 4354) growing poorly (A_{600} < 1) and consuming low amounts of lactose. The precise overlap between growth and lactose consumption in YPLac suggests that strains can be classified as either poor or good consumers of lactose based on growth patterns.

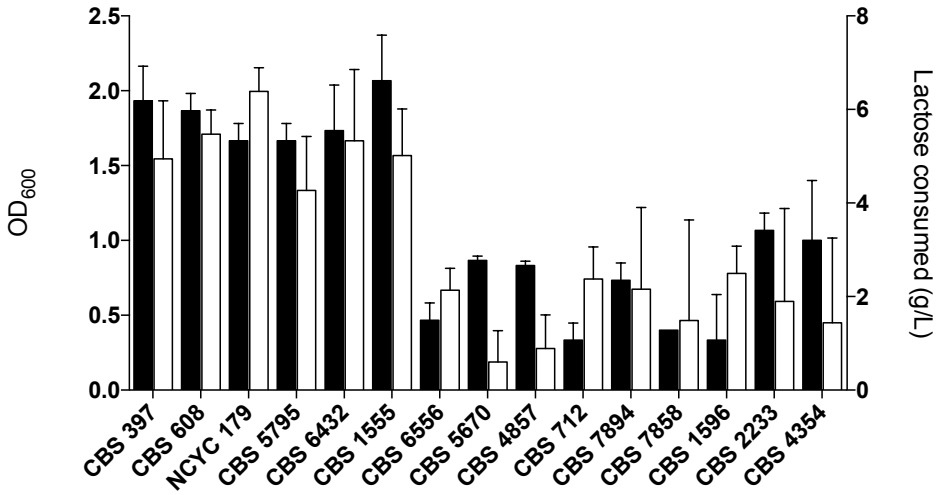


Figure 3.2: Lactose consumption is a variable trait among *K. marxianus* strains. A set of 15 strains was grown in YPLac media. Growth (open bars) and lactose consumed (solid bars) were evaluated after 11 hours. Lactose consumption was calculated as the difference between the initial and final lactose concentration in the medium. Experiments were performed in triplicates with error bars showing SD.

CBS 397 transport lactose more efficiently than CBS 6556

To study this trait in more detail, two strains, CBS397 and CBS 6556, which are among the most studied in the field, were selected as representatives of the good and poor lactose consumers (Etschmann *et al*., 2003; Fonseca *et al*., 2013, 2007; Jeong *et al*., 2012; Wang and Bajpai, 1997). First, to test whether the differences observed were exclusive to lactose consumption, the yeasts were grown in MM glucose (**Fig. 3.3A**). Both strains grew equally well when grown in this media reaching stationary phase after 11-12 hours. Sugar consumption was also monitored showing that CBS 397 and CBS 6556 utilised all the glucose in the media. In contrast, when the strains were cultured in MM lactose (**Fig. 3.3B**), CBS 397 grew at a similar rate as when using glucose as a carbon source, but CBS 6556 exhibited a dramatic reduction in growth rate. Lactose consumption analysis showed that CBS 397 utilised lactose efficiently whereas CBS 6556 barely consumed any sugar. Since there was some growth of CBS 6556 on MM with 2% lactose with little apparent use of lactose, growth on the lower concentration of 0.2% lactose was assessed (**Fig. 3.3C**). Although overall yields (as measured by A_{600}) were lower, a similar pattern of growth on 2% lactose was observed. In this case, it was possible to observe low levels of lactose utilisation by CBS 6556 indicating that the strain can utilise lactose, but at a much lower rate than CBS 397. In control experiments with MM lacking any sugar, neither strain could grow, thereby confirming that all growth observed relied on lactose as the carbon source (data not shown). As a key step in lactose utilisation is the efficient hydrolysis of this sugar into glucose and galactose, β - galactosidase activity was measured in the strains to test for possible differences. Overall, low activity was detected when the strains were grown in glucose and high levels when grown in lactose (**Fig 3.3D**). Interestingly, CBS 6556 showed higher levels of activity then CBS 397 but lower than the *K. lactis* strain CBS 2359 suggesting that differences in β - galactosidase are not accountable for the growth defect exhibited by CBS 6556. Next, to establish whether the variation in lactose consumption was due to differences in lactose transport efficiency, sugar uptake assays were conducted using radiolabelled substrates (**Fig. 3.3E**). Glucose and lactose were assayed separately using saturating concentrations and maximum velocities (Vmax) were calculated accordingly. The strains did not show significant differences in glucose transport (**Fig. 3.3E**, black bars) but the Vmax for lactose transport was 10 times higher in CBS 397 than in CBS 6556 (**Fig. 3.3E**,

white bars). These results indicate that the phenotypic differences observed between the strains can be explained by changes in lactose transport efficiency.

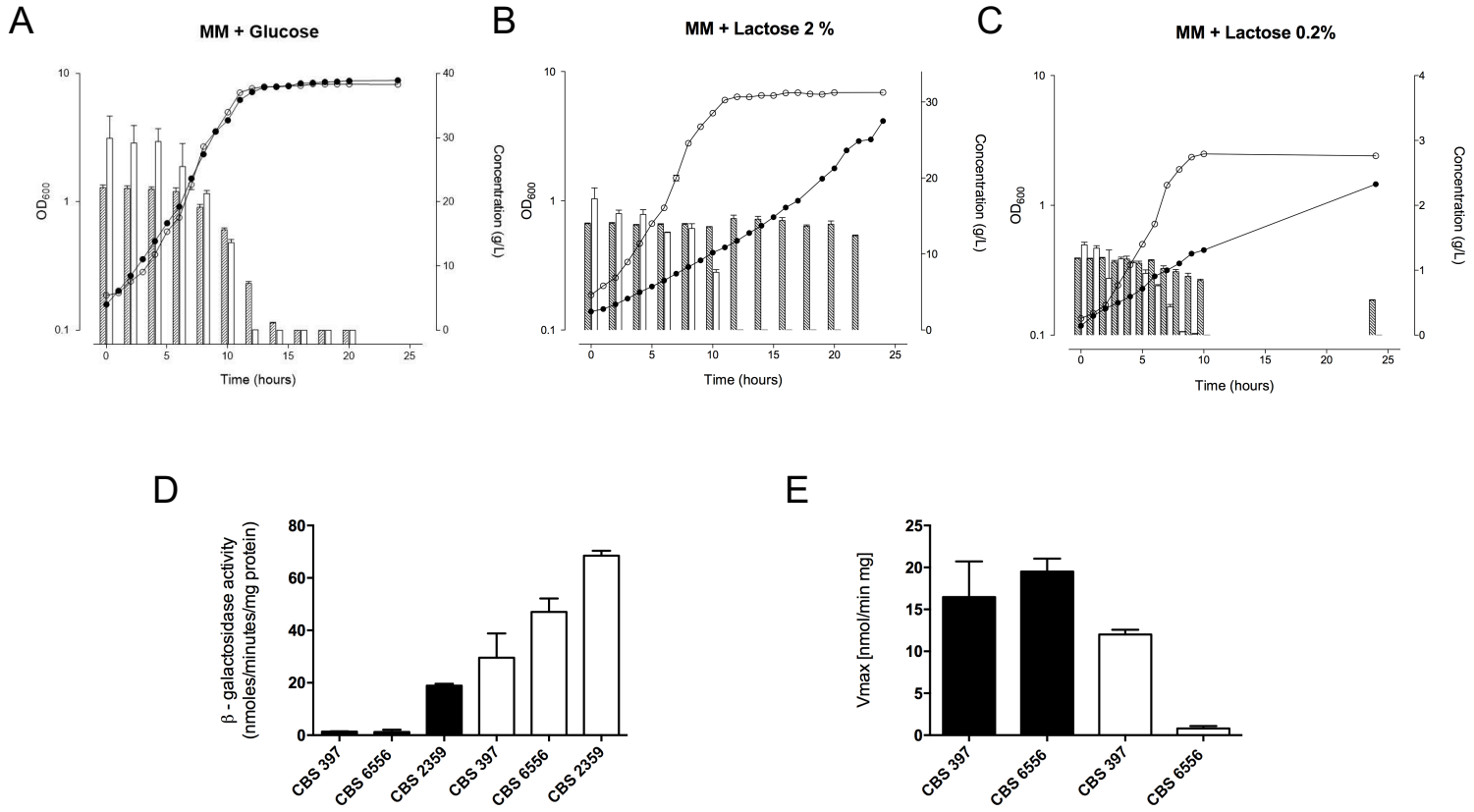


Figure 3.3: Growth, consumption and transport of glucose and lactose by *K. marxianus* strains. CBS 397 (open circles) and CBS 6556 (solid circles) were grown in triplicate on MM medium supplemented with 2 % glucose (**A**), 2 % lactose (**B**) or 0.2 % lactose (**C**). A_{600} was monitored every hour and supernatants were analysed by HPLC. Sugar content in the medium is shown as white (CBS 397) and grey bars (CBS 6556). (**D**) β - galactosidase assays were performed in *K. marxianus* CBS 397 and CBS 6556 and in *K. lactis* CBS 2359. Cells were grown in YPD (black bars) and YPLac (white bars) to mid-exponential phase at which point the assay was carried out. (**E**) Sugar uptake assays were performed on cells pre-grown on 2 % glucose or lactose to an A_{600} of 2. Radiolabelled glucose (black bars) or lactose (white bars) was added and the kinetics of uptake measured. Histograms show the average ± S.D. of 3 independent experiments.

K. marxianus strains carry four copies of *LAC12*

In the sister species, *K. lactis*, lactose is transported by Lac12p, therefore one possible explanation for the differences could have been deletion or inactivation of the CBS 6556 *LAC12* gene. To investigate this possibility, the published genome sequence of CBS 6556 was interrogated using tBLASTn and the *K. lactis* Lac12p. This identified 3 homologous sequences, all of which were automatically annotated as *LAC12* by Blast2GO.

As Lertwattanasakul *et al*., 2015 showed the presence of 4 putative *LAC12* genes in the DMKU3-1042 genome, the CBS 6556 and DMKU3-1042 genomes were compared. Three of the *LAC12* genes were present in both strains but one *LAC12* gene, located in the subtelomeric region of chromosome 8, appeared to be absent in CBS 6556. Comparative genome analysis, however, revealed that a 50 kb section is missing from the CBS 6556 assembly and PCR confirmed the presence of the 4th copy of *LAC12* in CBS 6556 (data not shown). Thus, both CBS 6556 and DMKU3-1042 carry 4 very similar copies of the *LAC12* gene in virtually identical genomic contexts (**Fig. 3.4**). One of the *LAC12* genes, located in the subtelomeric region of chromosome 3, shares the *K. lactis LAC12-LAC4* locus topology and is also the closest in sequence to *K. lactis LAC12* and is therefore designated *LAC12* (*KmLAC12).* The homologous sequences, designated *LAC12-2, LAC12-3 and LAC12-4,* are found on the opposite extreme of chromosome 3 (to *KmLAC12*), on chromosome 2, and on chromosome 8, respectively. Interestingly, in CBS 6556, *LAC12-2* also displays the *K. lactis LAC12-LAC4* locus organisation except that the *LAC4* sequence at this locus has in-frame stop codons, indicating that it is a non-functional pseudogene. In DMKU3-1042, the *LAC4* pseudogene is absent and *LAC12-2* is flanked at the 5' end by a putative oxidoreductase. With that exception, the genomic context of the *LAC12* genes is similar in both CBS 6556 and DMKU3-1042.

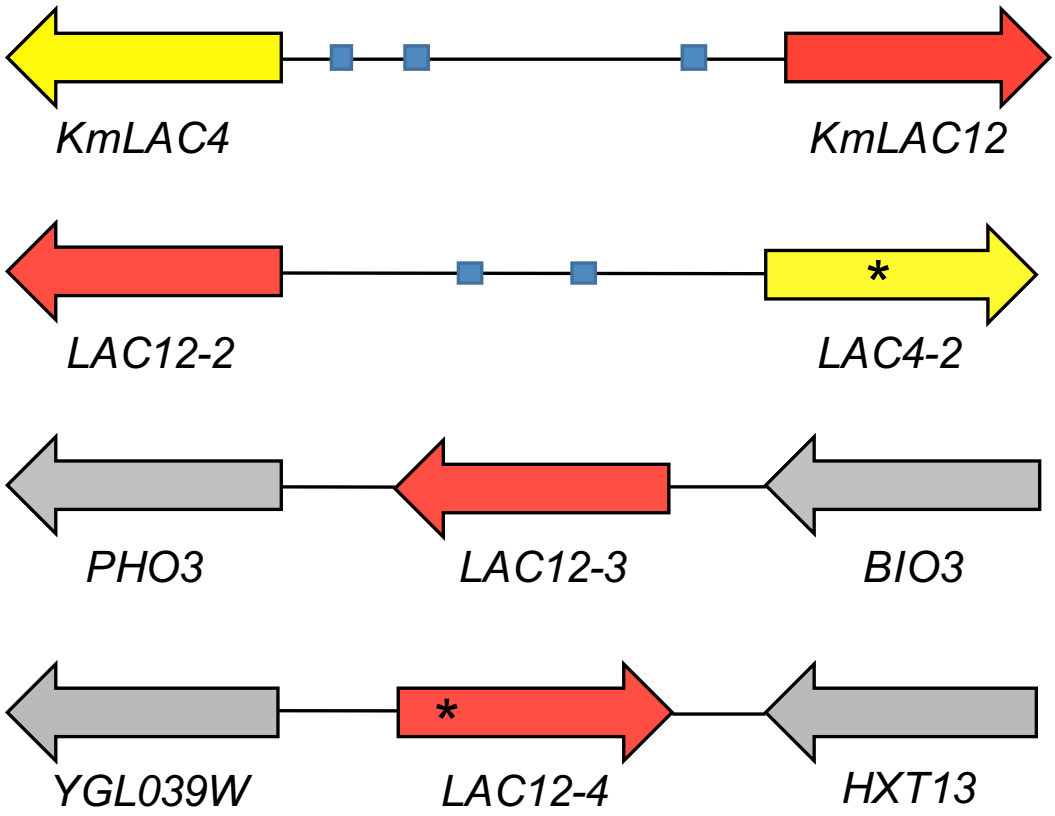


Figure 3.4: *K. marxianus* genome contains multiple copies of the lactose permease gene. The figure represents the CBS 6556 genome and was constructed with the information obtained from the CBS 6556 and DKMU 3 - 1042 genomes. Adjacent genes, shown in grey, are named according to their *S. cerevisiae* orthologues. *KmLAC12* and *KmLAC4* refer to the ancestral lactose permease and β-galactosidase genes, respectively. The *LAC4-2* and *LAC12-4* pseudo-genes are shown as arrows containing asterisks that indicate mutations that interrupt the reading frame. Blue squares represent putative Gal4p binding sites. Distances between genes are not drawn to scale.

More extensive analysis in other *K. marxianus* strains revealed that all tested strains carry the 4 *LAC12* genes but in some (CBS 6556 and DMKU3-1042), the *LAC12-4* gene is a pseudogene whereas in others, the sequence appears to be capable of encoding a functional protein (**Fig. 3.5**). Analysis by PCR and sequencing established that CBS 397 carries the four *LAC12* genes, including a putative functional *LAC12-4* gene. The presence of the *LAC4* pseudo gene, next to *LAC12-2*, was also confirmed in this strain.

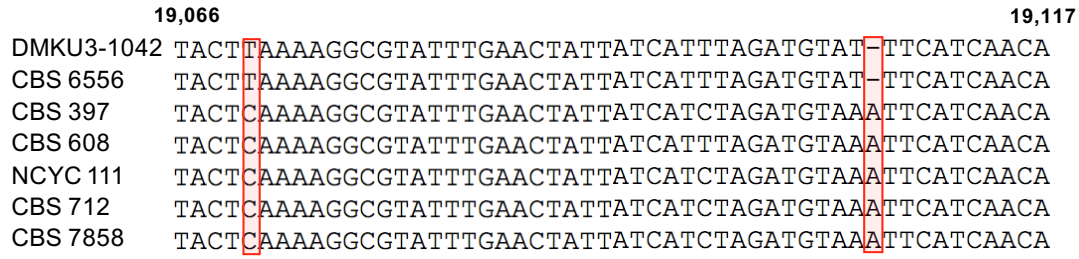


Figure 3.5: Presence of premature stop codons in the *LAC12-4* gene of some *K. marxianus* strains. The *LAC12-4* gene was sequenced and compared across all strains shown. Strains DMKU3-1042 and CBS 6556 contain mutations that independently cause premature stop in the sequence (red boxes). Numbers at the top of the figure correspond to coordinates of the *LAC12-4* gene in chromosome 8 of the DMKU 3 – 1042 genome.

A single gene is responsible for the differences in lactose consumption

The presence of multiple *LAC12* copies suggested that the lactose consumption differences between *K. marxianus* strains could be caused by the activity of multiple genes. To test whether this was the case or whether a single gene was responsible for the phenotypic variation, an analysis of a genetic cross between representatives of good and poor lactose consumers was carried out. Because specific markers were required for this analysis, auxotrophic derivatives of DMKU3–1042, which displays the CBS 6556 poor lactose consumption phenotype, and NCYC 1429, which represents the CBS 397 phenotype (**Fig. 3.6**), were used. For this experiment, the two strains were mated, the diploid sporulated, and random spore analysis was performed, scoring for the lactose utilisation phenotype (**Fig. 3.7**). The diploid had the lactose-utilisation (NCYC 1429) phenotype and all progeny showed either one of the parental phenotypes and the absence of intermediate phenotypes. These genetic data indicate that a single dominant gene is responsible for the differentiation of phenotypes.

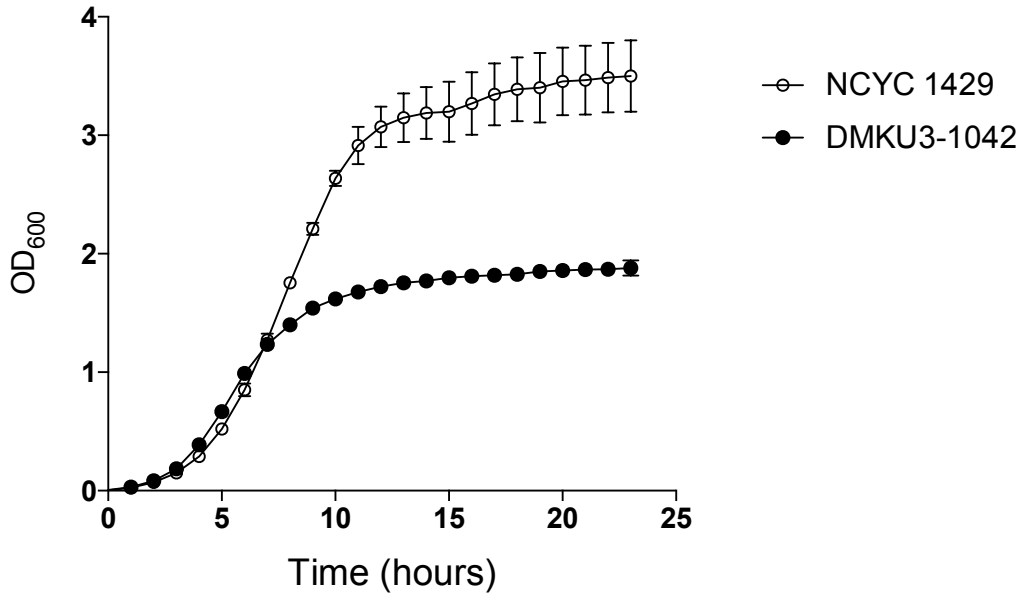


Figure 3.6: Growth of lactose medium by NCYC 1429 and DMKU3-1042 strains. Cells were grown in triplicates in a 96-well plate containing YPLac medium. Growth was monitored hourly. NCYC 1429 and DMKU3 – 1042 are shown in open and closed circles, respectively.

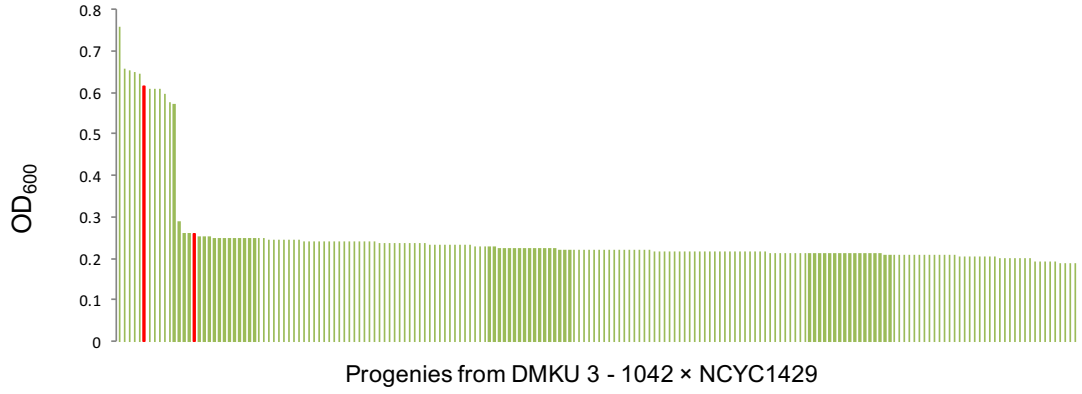


Figure 3.7: Growth on lactose by DMKU 3 -1042 x NCYC 1429 progenies. 190 random progenies were selected and grown in 96-well plates containing YPLac. Growth was measured after 24 hours using a multi-well spectrophotometer. Blue arrows indicate the parental strains DMKU 3 – 1042 and NCYC 1429.

KmLAC12 is induced by lactose in both good and poor lactose-utilising strains

Since it is known in *K. lactis* that expression of the *LAC12* and *LAC4* genes is strongly induced by lactose, RTqPCR analyses were performed in strains CBS 397 and CBS 6556 (**Fig. 3.8**). Strains were grown to mid-exponential phase on YP with 2% glucose, raffinose or lactose as the carbon source and expression compared to the *ACT1* gene assessed by RTqPCR. There were marked differences between the *LAC12* genes, with *KmLAC12* expressed to a much higher level than the other *LAC12* genes. There was also strong induction on lactose compared to raffinose or glucose, possibly as a consequence of the presence of putative Gal4p binding sites in the *LAC12-LAC4* intergenic region (**Fig. 3.4**). *LAC12-2, LAC12-3* and *LAC12-4* all showed lactose induction to some degree, but the level of expression was low compared to *KmLAC12*. For example, in the case of *LAC12-2* and *LAC12-4*, induced expression was only at the level of basal expression of *KmLAC12*.

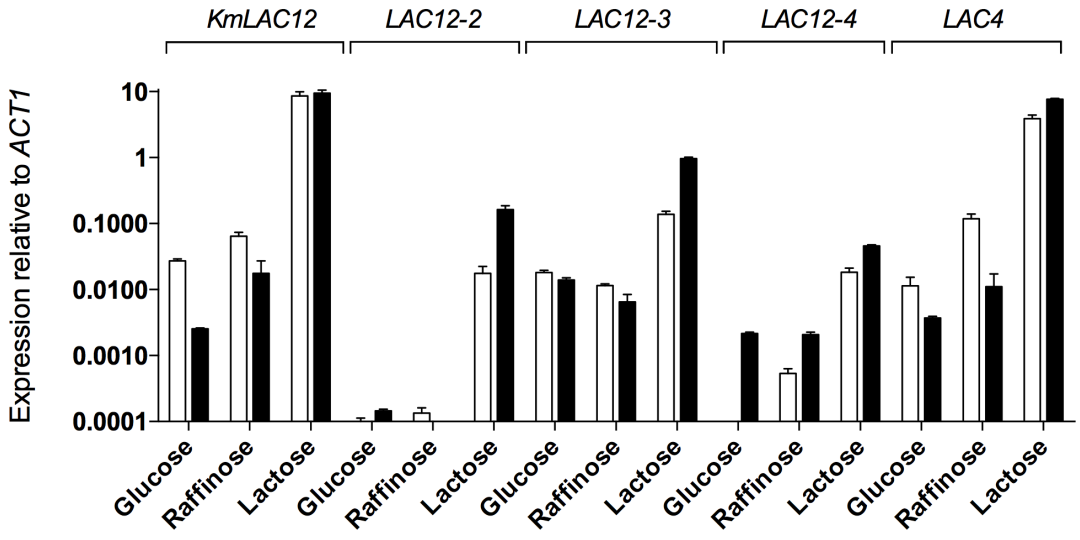


Figure 3.8: Gene expression analysis of the *LAC12* gene copies. CBS 397 (black bars) and CBS 6556 (white bars) were grown in glucose, raffinose and lactose media. Then, RNA extraction and cDNA synthesis were carried out. Relative quantification of gene expression was performed by quantitative PCR. The values shown correspond to delta Ct calculations obtained when comparing against the expression of internal control *ACT1*. Histograms show the average ± S.D. of 3 independent experiments.

It was significant that there was no substantial difference in expression patterns between CBS 6556 (open bars) and CBS 397 (solid bars), suggesting that expression differences do not explain the lactose utilisation phenotype. Expression of *LAC4* is also strongly induced but comparable in both strains, consistent with earlier data showing that differences in β-galactosidase activity are not responsible for the different phenotypes of the good and poor lactose utilising strains (**Fig. 3.3D**)

Functional differences in the *Km*Lac12p protein are responsible for the variable lactose-utilisation phenotype.

Since neither comparative genomics nor expression analysis identified any differences that would account for why lactose transport is efficient in CBS 397, NCYC 1429 and other strains, but poor in CBS 6556, DMKU-1042 and other strains, heterologous expression of the *LAC12* genes was carried out. First, each of the 4 genes from each strain was expressed in *S. cerevisiae* to assess lactose transport (**Fig. 3.9A**). The host strain, *S. cerevisiae* EBY.VW4000 lacks hexose transporters and was used to avoid growth on hexoses which might arise from impurities in the lactose or as a consequence of unspecific lactose hydrolysis. However, this strain can grow on maltose using the Mal2p maltose transporter (Wieczorke *et al*., 1999). *S. cerevisiae* strains expressing the different *LAC12* genes (and the β-galactosidase gene, *LAC4)* were washed, diluted and plated onto different plates with either maltose or lactose as the carbon source. All strains grew on the maltose control plate, but when growth in lactose media was examined, it was found that only the CBS 397 version of *KmLAC12* conferred growth. This indicates that, of the 4 *LAC12* genes in CBS 397, only *KmLAC12* encodes a functional lactose transporter. Furthermore, it demonstrates that the same protein from CBS 6556 is not able to transport lactose. To rule out a problem of expression or protein localisation of the CBS 6556 version of *Km*Lac12p, an in-frame fusion of CBS 6556 *Km*Lac12p to GFP was made. Fluorescence microscopy showed that the fusion protein correctly localised to the cell membrane (**Fig. 3.10**), suggesting that CBS 397 and CBS 6556 genuinely have different variants of *KmLAC12*.

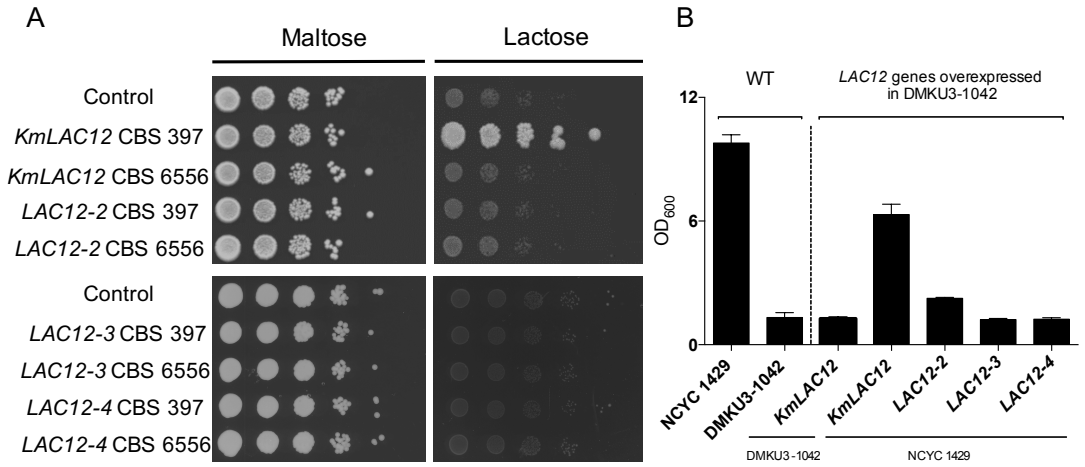


Figure 3.9: Functional analysis of the *LAC12* genes in *S. cerevisiae* EBY.VW4000 and *K. marxianus* DMKU3-1042. A. *S. cerevisiae* EBY.VW4000 strains expressing the different *LAC12* copies and the *LAC4* gene were grown in SC maltose media lacking leucine and uracil to an A_{600} of 2. Cells were collected by centrifugation and then washed twice with sterile water. Cultures were serially diluted (10-fold) and then inoculated into different carbon sources using a replicator. Plates were incubated for a period of 3-5 days at 30 °C. **B.** Overexpression of the *LAC12* genes in DMKU3–1042 *ura3-1* (RAK 3605). The strains expressing the *LAC12* genes were grown in YPLac medium. A_{600} was measured after 24 hours. The wild-type strains NCYC 1429 and DMKU3-1042 were used as controls for efficient and inefficient lactose utilisation, respectively. Experiments were performed in triplicates with error bars showing SD.

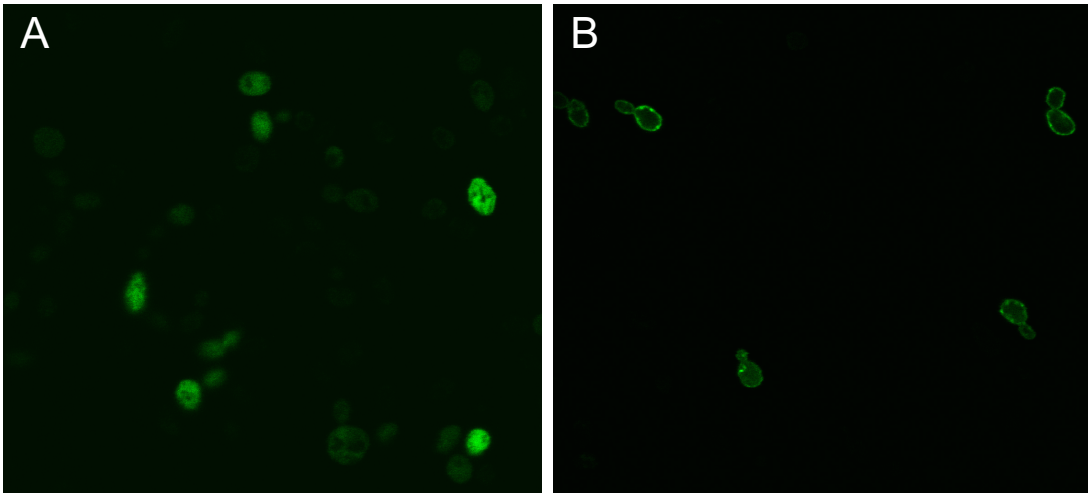


Figure 3.10: Subcellular location of the *Km*Lac12p - CBS 6556 transporter in *S. cerevisiae*. N-terminal GFP fusions were constructed for all the *LAC12* genes. Plasmids were transformed into *S. cerevisiae* BY4741, grown in presence of galactose, collected and visualized using a Zeiss LSM 5 laser scanning confocal microscope with a Plan-Apochromat 63x/1.40 Oil DIC M27 lens. **A.** Control strain expressing the GFP gene. **B.** Cells expressing the *KmLAC12::GFP* gene from CBS 6556. Similar results were obtained for *KmLAC12*, *LAC12-2* and *LAC12-4* – 397.

This data suggested that expression of the *KmLAC12* allele from a good lactose-utilising strain in a poor lactose-utilising strain would be sufficient to change the strain phenotype. To test this hypothesis, the two versions of the gene were expressed in DMKU3-1042, a poor utiliser (**Fig. 3.9B**). In this case, expression of the NCYC 1429 *KmLAC12* gene (identical in sequence to CBS 397) conferred growth on lactose, whereas expression of the DMKU3-1042 gene (identical in sequence to CBS 6556) had no effect. Furthermore, none of other three *LAC12* genes (*LAC12-2, LAC12-3* or *LAC12-4*) cloned from NCYC 1429 conferred growth, consistent with the conclusion that in good lactose utilisers like NCYC 1429 and CBS 397, only *KmLAC12* encodes a lactose transporter.

Changes in amino acid sequence are responsible for lactose consumption variability

The amino acid sequence of the CBS 397 / NCYC 1429, the CBS 6556 / DMKU3-1042, and the *K. lactis* CBS 2359 KmLac12p were compared to identify the changes that could give rise to the functional variation. This revealed 19 variant amino acids across the transporter (**Fig. 3.11**). To add to the resolution, the same gene was fully sequenced in the 13 strains shown in **Figure 3.2**. With the exception of CBS 1596, in which there was a deletion rendering *KmLAC12* a pseudogene, full protein sequences were obtained for all strains. Multiple sequence alignment (**Fig 3.12**) showed that there were two sequence variants of *Km*Lac12p. These are differentiated by changes in 13 amino acids (**Fig 3.11**, red shading). The distribution of these residues did not seem to show any particular pattern regarding the potential protein structure: changes were found in transmembrane domains and cytoplasmic regions alike. The *K. lactis* strain CBS 2359, known to be a good lactose consumer, also shared the CBS 397 amino acid residues that define the good lactose consumption allele.

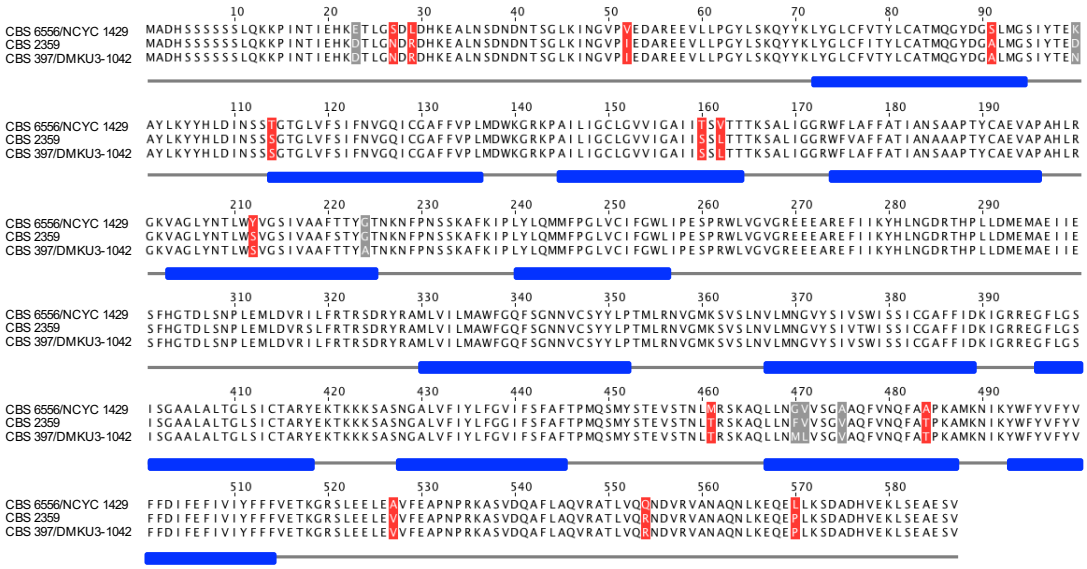


Figure 3.11: Amino acid sequence alignment of Lac12p sequences from CBS 6556, CBS 397 and *K. lactis* CBS 2359. Differences between the sequences are represented by coloured residues; red-labelled residues represent amino acids conserved in efficient or non-efficient lactose-consuming strains whereas grey-labelled positions show changes between the CBS 397 and CBS 6556 sequences that are not necessarily conserved in other efficient / non-efficient strains. Blue rectangles represent transmembrane domains as predicted by InterProScan. The alignment was computed using MUSCLE.

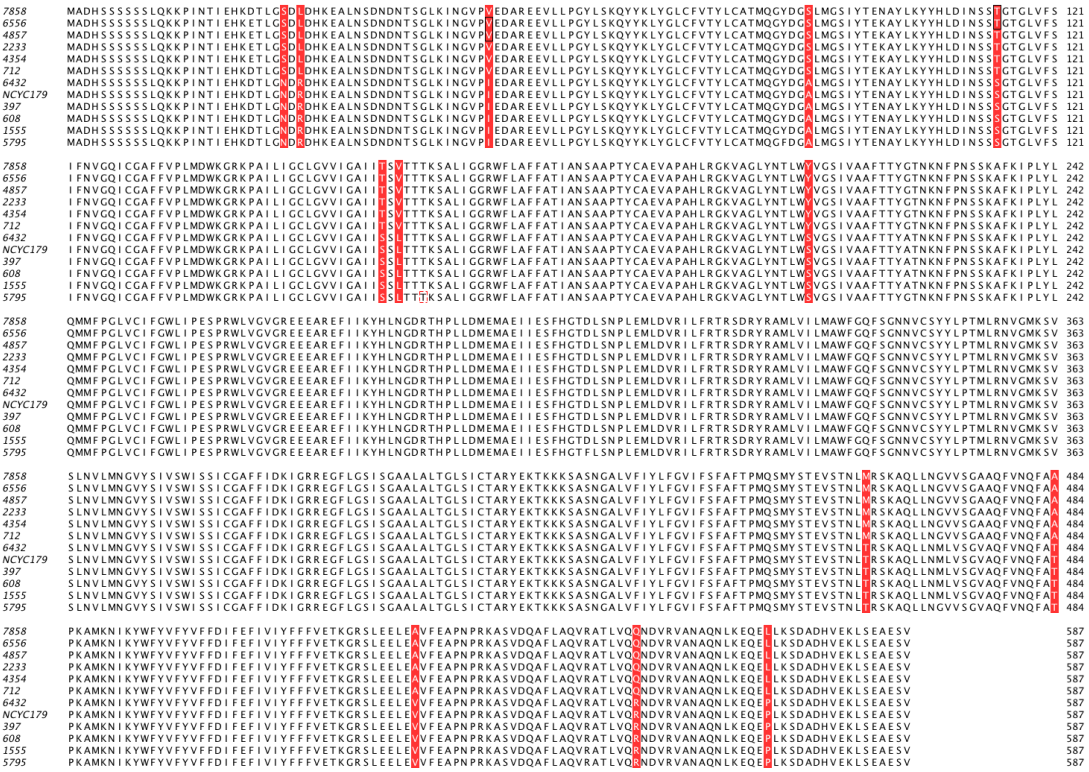


Figure 3.12: Multiple sequence alignment of *KmLAC12* sequences from *K. marxianus* strains. The *KmLAC12* gene was sequenced in 12 strains. Full-length amino acid sequences were obtained and used to generate a multiple sequence alignment in AliView. Strain names are shown at left. Residues that differ in good and poor utilisers are labelled in red.

To gain an insight into the evolution of the *LAC12* genes, protein sequences from CBS 6556 and CBS 397 were used to generate a phylogenetic tree (**Fig. 3.13**). As shown, the *Km*Lac12p sequences form a distinctive cluster that includes the *K. lactis* Lac12p transporter. Within this group the CBS 397 protein is more similar to the *K. lactis* sequence than to the CBS 6556 one. The remainder of the Lac12p sequences are located in another branch of the tree. Here, the Lac12-2p and Lac12-4p transporters are more closely related to each other than to Lac12-3p. The putative low-affinity galactose transporter *Km*Hgt1p (Baruffini *et al*., 2006) forms a distinct branch and is more related to the *S. cerevisiae* hexose transporters Hxt1p (glucose) and Gal2p (galactose).

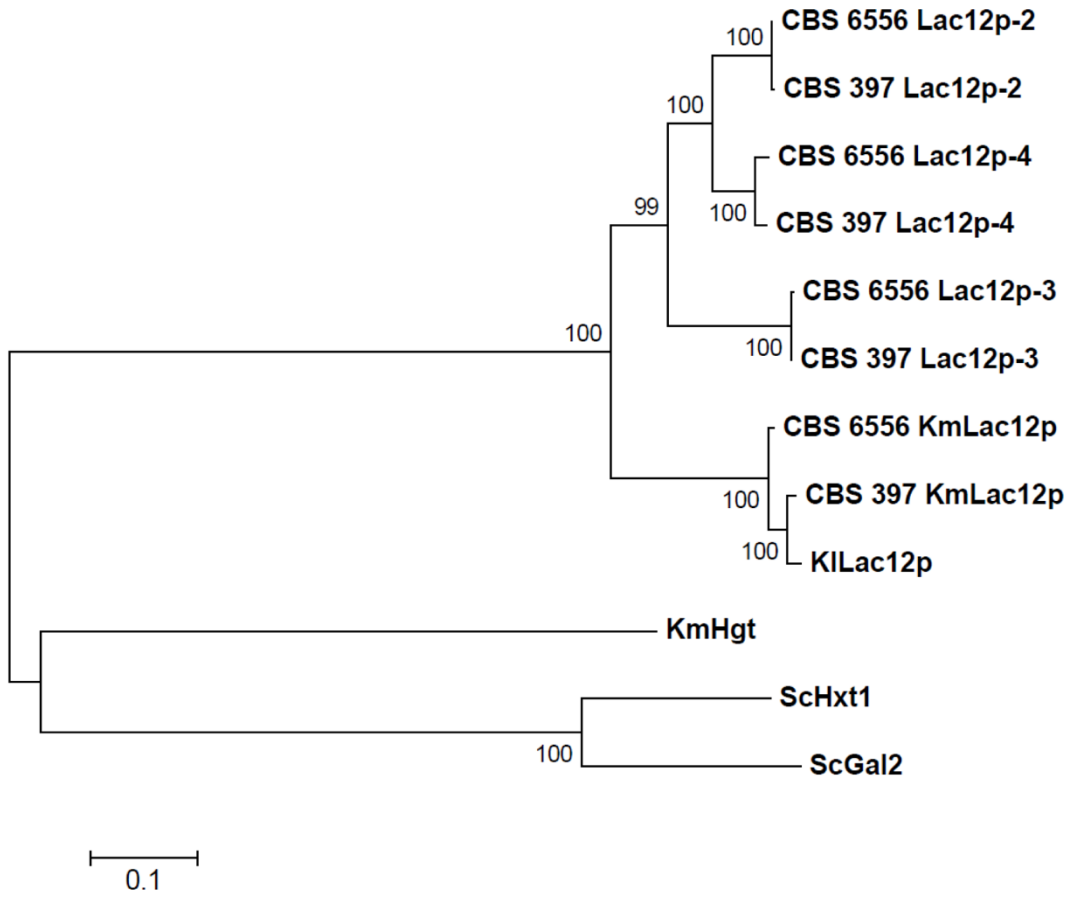


Figure 3.13: Phylogeny of lactose transporters from *K.* marxianus. Maximum-likelihood phylogenetic tree generated by comparing amino acid sequences of lactose permeases and hexose transporters. KlLac12p: lactose permease from *K. lactis*, KmHgt1: putative galactose transporter from *K. marxianus*. ScHxt1 and ScGal2: low-affinity glucose transporter and galactose transporter from *S. cerevisiae.* The tree was created using MEGA. Numbers represent % bootstrap values based on 1000 samplings.

3.4. Discussion

Transport and metabolism of lactose is required for many biotechnological applications of *K. marxianus,* so it is important to understand the basis of this trait. The basic mechanism of transport via Lac12p and hydrolysis by the intracellular Lac4p β-galactosidase in *Kluyveromyces* species has been known for some time from studies with *K. lactis* (Godecke *et al*., 1991). The presence of these genes is conserved in *K. marxianus* so the variable growth of different strains on lactose-containing medium was a conundrum. Further uncertainty came from the observation from genome sequencing projects that *K. marxianus* carries several genes that are homologous to *LAC12* and therefore, are also potential lactose transporters (Lertwattanasakul *et al*., 2015). In this study, it is unequivocally demonstrated that the main transporter responsible for uptake of lactose by *K. marxianus* is the *KmLAC12* gene that is located on the left arm of chromosome 3, transcribed divergently from the *LAC4* gene. As is the case with *K. lactis,* the intergenic region between these genes has several putative binding sites for the Gal4p transcription factor and both *LAC12* and *LAC4* are very strongly induced by lactose and show some evidence of glucose repression. The *LAC12* and *LAC4* nucleotide sequences show a high degree of conservation with those genes from *K. lactis* and it is clear that this *LAC12-LAC4* locus shares a common evolutionary origin in *K. lactis* and *K. marxianus* and is most likely ancestral to the speciation.

Since comparison with *K. lactis* indicates that the *K. marxianus* locus on the left art of chromosome 3 is ancestral, it follows that the other copies of *LAC12* that are present in *K. marxianus* arose by gene duplication events. The location of all 4 homologous *LAC12* genes in sub-telomeric chromosome regions immediately suggests the mechanism of duplication. Yeast subtelomeric regions have been shown to be subject to a high rate of duplication and recombination (Barton *et al*., 2008). Indeed, in *K. lactis* 7 out of 12 subtelomeres contain a duplicated 9 kb region that contains several duplicated genes (Fairhead and Dujon, 2006). Our analysis of other chromosomes in *K. marxianus* found that several chromosomes have duplicated regions (Varela and Morrissey, unpublished data). In the case of chromosome 3, the *LAC12-LAC4* locus became duplicated on the right sub-telomeric region of chromosome 3 (*LAC12-2 – LAC4).* From comparison of sequenced *K. marxianus* strains and analysis of some strains by targeted PCR, it is seen that no strains retained a functional *LAC4* gene at

this locus – in all cases it has either acquired deletion/frameshift mutations or it is completely absent, presumably lost by a deletion event. In contrast, the *LAC12-2* gene retains the coding capacity for an intact protein, albeit one that has lost the capacity to transport lactose. Phylogenetic analysis indicates that *LAC12-3* and *LAC12-4* probably arose from duplications of *LAC12-2* (**Figure 3.13**) so it is not surprising that these proteins also do not encode functional lactose transporters. In some strains (e.g. CBS 6556 and DMKU3-1042), *LAC12*-4 is a pseudogene, whereas in others (e.g. CBS 397, NCYC 1429) it still has the potential to encode a functional protein. The variation seen between strains suggests that diversification in the sub-telomeric regions is an ongoing process. More broadly, sub-telomeric regions have been shown to be hot spots of gene evolution and functional divergence (Anderson *et al*., 2015). Rapid evolution of subtelomeric gene families has been shown in the *MAL* genes, required for maltose consumption in *S. cerevisiae* (Brown *et al*., 2010). Moreover, it was recently demonstrated that certain duplicated hexose transporter family members in *S. cerevisiae* (Hxt13 and Hxt15-Hxt17; all encoded in subtelomeric regions) have lost their ability to transport hexoses and evolved as hexitol transporters (Jordan *et al*., 2016).

As well as establishing that *KmLAC12* encodes the main *K. marxianus* lactose transporter, the work reported in this study determined that the reason why CBS 6556, DMKU3-1042 and other strains fail to grow efficiently on lactose is due to variation in that gene. More specifically, the *KmLAC12* allele in those strains has sufficiently diverged from the ancestral sequence to no longer encode a lactose transporter. In one poor lactose-utilising strain tested (CBS 1596), there was a deletion rendering the gene non-functional, but in all other cases, it is still possibly encoding an intact protein. Analysis of the amino acid sequence of 12 strains determined that there are 13 amino acid variations that distinguish a functional / non-functional lactose transporter. It remains to be established whether any single amino acid change is sufficient for this effect, or whether a combination of some or all of the 13 changes is required to abolish lactose transport. It has been previously shown that lactose permease specificity can be modified in *E. coli* (Varela *et al*., 1997). For example, it has been shown that chemical modification of a single cysteine residue in the lactose permease LacY could change the transporter into a galactose specific transporter (Guan *et al*., 2002). Although some of the studied amino acid changes in *E. coli* could potentially be

applied to the *K. marxianus* transporters, the *Km*Lac12p and LacY are only 33 % identical making the comparison difficult.

Comparative analysis of the amino acid sequence of each of the Lac12p sequences between CBS 397 and CBS 6556 is intriguing. There is very little variation between Lac12-2p (99 % identity) or Lac12-3p (100 % identity), but more variation between Lac12-4p (97 % identity) and *Km*Lac12p (97 % identity). In contrast, the Lac12p from CBS 397 and *K. lactis* are 99 % identical. One could speculate that the CBS 6556 *KmLAC12* gene has undergone selection to transport a sugar other than lactose. In this regard, it is notable that unlike CBS 397 (yoghurt), both CBS 6556 (pozol) and DMKU 3-1042 (bioethanol) were isolated from non-dairy environments and would not necessarily have a requirement to transport lactose. Since the substrate from which many of the strains used in this study were isolated is not known, it is not possible to extrapolate this observation. Nor is it known what this alternative substrate would be, but our analysis indicates that it is not galactose (Varela and Morrissey, unpublished). If this suggestion is correct, however, it seems peculiar that the β - galactosidase activity remains in yeast strains that have lost the capacity to utilise lactose. Thus, there are certainly still unanswered questions. It must also be considered that proteins in this family can also transport other organic and inorganic small molecules, so there is also the possibility of a non-sugar substrate. In any case, the phenotypic variability observed across strains indicates that lactose utilisation is not universal in *K. marxianus.* Although the original classification of the species assumed this trait to be present in all strains, our data supports the idea that sequence-based methods are more robust than phenotypic assays for classifying yeasts.

Much of what is known about sugar transporters in eukaryotes has come from research with *S. cerevisiae*. In particular, that yeast has an impressive repertoire of high and low affinity glucose sensors and transporters that play a key role in the ecology and physiology of the organism (Ozcan and Johnston, 1999). Furthermore, research on maltose transport in brewing strains have been a model for how selection can give rise to gene families (Jespersen *et al*., 1999). There has been a perception that domestication had created some unique evolutionary pressures for *S. cerevisiae* and other yeasts will not show similar complexity. The findings of this study and others (Lertwattanasakul *et al*., 2015) indicate that this, in fact, is not the case. *K. marxianus* has a large family of related sugar transporters and decipering their regulatory and

metabolic networks is a key challenge to aid in the exploitation of *K. marxianus* as a cell factory yeast for biotechnology.

3.5 Acknowledgments

J. V. is supported by the YEASTCELL Marie-Curie ITN project (REA Grant No. 606795) under the EU's Seventh Framework Programme for Research (FP7). Technical assistance of Dan Walsh with HPLC analysis, and Paddy O' Reilly with Microbiology work, is greatly acknowledged. This study was supported in part by the Advanced Low Carbon Technology R&D Program (JST, Japan).

3.6 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, 907–919. doi:10.1534/genetics.115.177451

Arrizon, J., Morel, S., Gschaedler, A., Monsan, P., 2012. Fructanase and fructosyltransferase activity of non-Saccharomyces yeasts isolated from fermenting musts of Mezcal. Bioresour. Technol. 110, 560–565. doi:10.1016/j.biortech.2012.01.112

Arrizon, J., Morel, S., Gschaedler, A., Monsan, P., 2011. Purification and substrate specificities of a fructanase from *Kluyveromyces marxianus* isolated from the fermentation process of Mezcal. Bioresour. Technol. 102, 3298–3303. doi:10.1016/j.biortech.2010.10.071

Barton, A.B., Pekosz, M.R., Kurvathi, R.S., Kaback, D.B., 2008. Meiotic recombination at the ends of chromosomes in *Saccharomyces cerevisiae*. Genetics 179, 1221–35. doi:10.1534/genetics.107.083493

Baruffini, E., Goffrini, P., Donnini, C., Lodi, T., 2006. Galactose transport in *Kluyveromyces lactis*: major role of the glucose permease Hgt1. FEMS Yeast Res. 6, 1235–42. doi:10.1111/j.1567-1364.2006.00107.x

Brown, C.A., Murray, A.W., Verstrepen, K.J., 2010. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr. Biol. 20, 895–903. doi:10.1016/j.cub.2010.04.027

Carvalho-Silva, M., Spencer-Martins, I., 1990. Modes of lactose uptake in the yeast species *Kluyveromyces marxianus*. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 57, 77–81. doi:10.1007/BF00403158

Castro, R.C.A., Roberto, I.C., 2014. Selection of a Thermotolerant *Kluyveromyces marxianus* Strain with Potential Application for Cellulosic Ethanol Production by Simultaneous Saccharification and Fermentation. Appl. Biochem. Biotechnol. 172, 1553–1564. doi:10.1007/s12010-013-0612-5

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.

doi:10.1093/bioinformatics/bti610

Etschmann, M.M., Sell, D., Schrader, J., 2003. Screening of yeasts for the production of the aroma compound 2-phenylethanol in a molasses-based medium. Biotechnol. Lett. 25, 531–6.

Fairhead, C., Dujon, B., 2006. Structure of *Kluyveromyces lactis* subtelomeres: duplications and gene content. FEMS Yeast Res. 6, 428–441. doi:10.1111/j.1567-1364.2006.00033.x

Farwick, A., Bruder, S., Schadeweg, V., Oreb, M., Boles, E., 2014. Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. Proc. Natl. Acad. Sci. U. S. A. 111, 5159–64. doi:10.1073/pnas.1323464111

Fonseca, G.G., de Carvalho, N.M.B., Gombert, A.K., 2013. Growth of the yeast *Kluyveromyces marxianus* CBS 6556 on different sugar combinations as sole carbon and energy source. Appl. Microbiol. Biotechnol. 97, 5055–5067. doi:10.1007/s00253-013-4748-6

Fonseca, G.G., Gombert, A.K., Heinzle, E., Wittmann, C., 2007. Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. FEMS Yeast Res. 7, 422–35. doi:10.1111/j.1567-1364.2006.00192.x

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. doi:10.1007/s00253-008-1458-6

Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31–34. doi:10.1038/nprot.2007.13

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, 5351–5358.

Goshima, T., Tsuji, M., Inoue, H., Yano, S., Hoshino, T., Matsushika, A., 2013. Bioethanol Production from Lignocellulosic Biomass by a Novel *Kluyveromyces marxianus* Strain. Biosci. Biotechnol. Biochem. 77, 1505–1510.

doi:10.1271/bbb.130173

Grba, S., Stehlik-Tomas, V., Stanzer, D., Vahcic, N., Skrlin, A., 2002. Selection of yeast strain *Kluyveromyces marxianus* for alcohol and biomass production on whey. Chem. Biochem. Eng. Q. 16, 13–16.

Guan, L., Sahin-Tó th, M., Ronald Kaback, H., 2002. Changing the lactose permease of *Escherichia coli* into a galactose-specific symporter.

Guimrães, P.M.R., Teixeira, J.A., Domingues, L., 2010. Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. Biotechnol. Adv. 28, 375–384. doi:10.1016/j.biotechadv.2010.02.002

Hoshida, H., Murakami, N., Suzuki, A., Tamura, R., Asakawa, J., Abdel-Banat, B.M.A., Nonklang, S., Nakamura, M., Akada, R., 2014. Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast *Kluyveromyces marxianus*. Yeast 31, 29–46. doi:10.1002/yea.2993

Hu, N., Yuan, B., Sun, J., Wang, S.-A., Li, F.-L., 2012. Thermotolerant Kluyveromyces marxianus and *Saccharomyces cerevisiae* strains representing potentials for bioethanol production from Jerusalem artichoke by consolidated bioprocessing. Appl. Microbiol. Biotechnol. 95, 1359–1368. doi:10.1007/s00253-012-4240-8

Jansen, G., Wu, C., Schade, B., Thomas, D.Y., Whiteway, M., 2005. Drag&Drop cloning in yeast. Gene 344, 43–51. doi:10.1016/j.gene.2004.10.016

Jeong, H., Lee, D.-H., Kim, S.H., Kim, H.-J., Lee, K., Song, J.Y., Kim, B.K., Sung, B.H., Park, J.C., Sohn, J.H., Koo, H.M., Kim, J.F., 2012. Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var. marxianus KCTC 17555. Eukaryot. Cell 11, 1584–5. doi:10.1128/EC.00260-12

Jespersen, L., Cesar, L.B., Meaden, P.G., Jakobsen, M., 1999. Multiple alpha-glucoside transporter genes in brewer's yeast. Appl. Environ. Microbiol. 65, 450–6.

Lachance, M.-A., 2007. Current status of *Kluyveromyces* systematics. FEMS Yeast Res. 7, 642–645. doi:10.1111/j.1567-1364.2006.00197.x

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. Antonie Van Leeuwenhoek 100, 507–19. doi:10.1007/s10482-011-9606-x

Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. Fungal Biol. Rev. 24, 17–26. doi:10.1016/j.fbr.2010.01.001

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–8. doi:10.1093/bioinformatics/btu531

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Suprayogi, Tsuchikane, K., Limtong, S., Fujita, N., Yamada, M., 2015. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. Biotechnol. Biofuels 8, 47. doi:10.1186/s13068-015-0227-x

Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–89. doi:10.1101/gr.1224503

Morrissey, J.P., Etschmann, M.M.W., Schrader, J., Billerbeck, G.M., 2015. Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. Yeast 32, 3–16. doi:10.1002/YEA.3054

Nonklang, S., Abdel-Banat, B.M.A., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., Yamada, M., Akada, R., 2008. High-Temperature Ethanol Fermentation and Transformation with Linear DNA in the Thermotolerant Yeast *Kluyveromyces marxianus* DMKU3-1042. Appl. Environ. Microbiol. 74, 7514–7521. doi:10.1128/AEM.01854-08

Nonklang, S., Ano, A., Abdel-Banat, B.M.A., Saito, Y., Hoshida, H., Akada, R., 2009. Construction of flocculent *Kluyveromyces marxianus* strains suitable for high-temperature ethanol fermentation. Biosci. Biotechnol. Biochem. 73, 1090–5.

Ozcan, S., Johnston, M., 1999. Function and regulation of yeast hexose transporters. Microbiol. Mol. Biol. Rev. 63, 554–69.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45.

Proux-Wéra, E., Armisén, D., Byrne, K.P., Wolfe, K.H., 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny

approach. BMC Bioinformatics 13, 237. doi:10.1186/1471-2105-13-237

Ribeiro, O., Gombert, A.K., Teixeira, J.A., Domingues, L., 2007. Application of the Cre-loxP system for multiple gene disruption in the yeast *Kluyveromyces marxianus*. J. Biotechnol. 131, 20–26. doi:10.1016/j.jbiotec.2007.05.027

Riley, M.I., Sreekrishna, K., Bhairi, S., Dickson, R.C., 1987. Isolation and characterization of mutants of *Kluyveromyces lactis* defective in lactose transport. Mol. Gen. Genet. 208, 145–51.

Rocha, S.N., Abrahão-Neto, J., Gombert, A.K., 2011. Physiological diversity within the *Kluyveromyces marxianus* species. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 100, 619–630. doi:10.1007/s10482-011-9617-7

Rodicio, R., Heinisch, J.J., Heinisch, J.J., 2013. Yeast on the milky way: genetics, physiology and biotechnology of *Kluyveromyces lactis*. doi:10.1002/yea.2954

Rubio-Texeira, M., 2006. Endless versatility in the biotechnological applications of *Kluyveromyces* LAC genes. Biotechnol. Adv. 24, 212–225. doi:10.1016/j.biotechadv.2005.10.001

Singh, M. V, Weil, P.A., 2002. A method for plasmid purification directly from yeast. Anal. Biochem. 307, 13–7.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729. doi:10.1093/molbev/mst197

Tschopp, J.F., Emr, S.D., Field, C., Schekman, R., 1986. GAL2 Codes for a Membrane-Bound Subunit of the Galactose Permease in *Saccharomyces cerevisiae.* J. Bacteriol. 166, 313–318.

Varela, M.F., Brooker, R.J., Wilson, T.H., 1997. Lactose carrier mutants of *Escherichia coli* with changes in sugar recognition (lactose versus melibiose). J. Bacteriol. 179, 5570–3.

Wang, C.-J., Bajpai, R.K., 1997. A mathematical model of ethanol fermentation from cheese whey. Appl. Biochem. Biotechnol. 63–65, 495–509. doi:10.1007/BF02920449

Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C.P., Boles, E.,

1999. Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. FEBS Lett. 464, 123–8.

Yarimizu, T., Nonklang, S., Nakamura, J., Tokuda, S., Nakagawa, T., Lorreungsil, S., Sutthikhumpha, S., Pukahuta, C., Kitagawa, T., Nakamura, M., Cha-aim, K., Limtong, S., Hoshida, H., Akada, R., 2013. Identification of auxotrophic mutants of the yeast *Kluyveromyces marxianus* by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*. Yeast 30, 485–500. doi:10.1002/yea.2985

Chapter 4

Identification of new galactose transport systems in *Kluyveromyces marxianus*

In preparation for submission to FEMS Yeast Research

Abstract

K. marxianus is an attractive yeast for multiple biotechnological applications including bioethanol and flavour molecule production. Despite the great interest on this yeast, several fundamental aspects related to *K. marxianus* physiology remain poorly understood. For example, sugar transport has not been well studied with only a few studies performed, most of them based on general characterisations of transport systems present in the yeast. Recently, it was showed that *K. marxianus* contains four homologues of the lactose permease gene, *LAC12*, with only one of them encoding a functional lactose transporter. The function of the remaining transporters was not determined but here, we show that three out of the four *LAC12* genes encode galactose transporters. Disruption of these genes in *K. marxianus* did not abolish galactose growth, suggesting the existence of additional galactose transporters in this yeast, with the *HGT1* gene being a strong candidate since it encodes a galactose transporter in *K. lactis*. A whole-genome bioinformatics analysis was performed to identify candidate sugar transporters in *K. marxianus* and, surprisingly, it was found that *K. marxianus* carries multiple homologues of the *HGT* gene and of *KHT1* gene, previously reported to encode a low-affinity glucose transporter. Heterologous expression of these genes revealed the presence of five additional galactose transporters in *K. marxianus*; four encoded by *HGT*-like genes, and a fifth encoded by a *KHT*-like gene. To explore the physiological role of these genes, mutant strains lacking these genes were constructed using the CRISPR-Cas9 system. Multiple combinations of mutants were made but only a strain mutated in all eight demonstrated galactose transport genes was unable to grow in 2 % galactose, confirming that these genes are the main galactose transporters in *K. marxianus* and that there is extensive redundancy for galactose transport. On the other hand, mutations of the *LAC* and *HGT* genes abolished growth on 0.1 % galactose medium, suggesting that these genes encode high-affinity galactose transporters. The findings of this chapter reveal that *K. marxianus* has several galactose transporters that are not found in other *Kluyveromyces* species, further illustrating the distinct genetic and phenotypic status of this yeast.

4.1 Introduction

Kluyveromyces marxianus is a yeast traditionally found in fermented dairy products such as cheese and kefir (Coloretti *et al*., 2017; Gethins *et al*., 2016; Tittarelli *et al*., 2018). Because of its connection with these products, *K. marxianus* has been granted a QPS (EU) or GRAS (FDA) status that regards this yeast as safe to use in food applications (Fonseca *et al*., 2008). *K. marxianus* is also found during the production of other non-dairy products such as chocolate and tequila/mezcal (Ho *et al*., 2014; Lappe-Oliveras *et al*., 2008). In order to establish and often colonise these environments the yeast employs physiological traits that allow efficient substrate utilisation. This involves the production of specific enzymes that degrade complex substrates and transporters that mediate the uptake of monomers into the cell. For example, Cocoa bean fermentation, an essential step in chocolate production, is generally carried out by a consortium of indigenous species that includes *K. marxianus* (De Vuyst and Weckx, 2016). In this environment the pectinase enzyme produced by *K. marxianus* degrades the pectin present in the cocoa pulp allowing the establishment of other yeast species required in the fermentation process (Schwan and Wheals, 2004). *K. marxianus* is also present as part of the natural flora of the *Agave* plant, used to produce the alcoholic beverages tequila and mezcal (Escalante-Minakata *et al*., 2008). Degradation of complex fructans to simple sugars by the enzyme inulinase is essential in the establishment of *K. marxianus* in this niche. Besides utilising various substrates and producing relevant enzymes *K. marxianus* also has other interesting physiological traits, uncommon to most yeasts (Lane and Morrissey, 2010). For example, it is able to grow at high temperatures (up to 50 °C), exhibits the fastest growth rate among eukaryotes and naturally produces significant amounts of flavour molecules such as esters and fusel alcohols (Morrissey *et al*., 2015). All these characteristics make this yeast an attractive host for different biotechnological purposes with some commercial applications available to date. Some of these include the production of bioethanol from cheese whey and production of the flavour molecule 2-phenylethanol (Fonseca *et al*., 2008; Morrissey *et al*., 2015; Varela *et al*., 2017a).

The prospect of further developing *K. marxianus* for cell-factory applications has motivated researchers to sequence the genome of several strains (Inokuma *et al*., 2015; Jeong *et al*., 2012; Lertwattanasakul *et al*., 2015; Ortiz-Merino *et al*., 2018;

Suzuki *et al*., 2014). A comparison of these genomes showed that *K. marxianus* displays a high level of genome variation at a structural and sequence level. Divergence in DNA sequence was found to be relativity high with up to 3% divergence between some isolates. Also, different strains were found to exist in haploid, diploid and triploid states (Ortiz-Merino *et al*., 2018). This high degree of variation between genomes is not surprising considering previous observations of phenotypic diversity across isolates (Lane *et al*., 2011). Lactose utilisation, for example, is a variable trait with some strains displaying good and poor growth on lactose (Carvalho-Silva and Spencer-Martins, 1990; Lane *et al*., 2011) This phenomenon is due to polymorphisms in the *LAC12* gene, which encodes a lactose permease that mediates lactose uptake into the cell. Strains that display good and poor lactose consumption carry functional (*LAC12*-B) and non-functional *(LAC12*-A) versions of the lactose transporter, respectively (Varela *et al*., 2017b). The sequence differences between the two haplotypes translate into 14 amino acid amino acid changes in the protein sequence that are sufficient to alter the function of the transporter (Ortiz-Merino *et al*., 2018). Unbiased analyses of genome sequences have shown that isolates from dairy environments are diploid or triploid and carry the *LAC12*-B haplotype while non-dairy strains are haploid and contain the *LAC12*-A haplotype that encodes the non-functional lactose transporter (Ortiz-Merino *et al*., 2018). Although this analysis was performed in a relatively small number of genomes it raises interesting questions regarding the evolution of lactose utilisation in *K. marxianus*.

Information on the transport of other key sugars, such as glucose and galactose, is limited in *K. marxianus*. Early experiments established the existence of three carbon-repressible proton-sugar transporters that mediate glucose/galactose, fructose and lactose uptake (De Bruijne *et al*., 1988). Induction of the different systems was shown to depend on the carbon source used to grow the cells. In yeast cells grown on lactose, the activity of a high-affinity proton-glucose transporter was detected whereas a low-affinity transporter was present in cells grown in glucose. Galactose was only taken up by cells grown in lactose and not by glucose-grown cells which indicates that the galactose transport system requires induction by lactose or presumably galactose (Gasnier, 1987). Unfortunately, these studies were focused on a single strain and were not followed up by studies aiming to identify the molecular

determinants encoding these transport systems. However, much more research on the subject has been done on *K. lactis*, *K. marxianus'* closest relative. Glucose uptake has been well studied in this yeast where also a low and high-affinity uptake system was described (Billard *et al*., 1996; Wésolowski-Louvel and Goffrini, 1992). In most strains, the low-affinity system is encoded by two tandemly-arrayed genes, *KHT1* and *KHT2* but some strains carry a single gene, *RAG1*, that arose from a recombination event between *KHT1* and *KHT2* (Weirich *et al*., 1997). The high affinity transport system is encoded by *HGT1*, which was found to exist as a single copy (Billard *et al*., 1996). Expression of this gene was found to be constitutive while expression of the *KHT* genes is glucose inducible (Milkowski *et al*., 2001). Interestingly, a strain carrying null mutations for both of the uptake systems is still able to grow on glucose, which points out to the existence of another glucose transporter not yet identified. Galactose uptake was also investigated in this yeast where the *LAC12* gene was found to play a role as a low-affinity galactose transporter (Riley *et al*., 1987). In fact, strains lacking *LAC12* were greatly impaired in galactose uptake, which led to the idea that galactose was exclusively transported via *LAC12*. Conversely, further studies on the topic showed that the high-affinity glucose carrier, *HGT1,* also acts as a high-affinity galactose transporter in *K. lactis* (Baruffini *et al*., 2006). Disruption of both *LAC12* and *HGT1* abolished galactose growth completely, ruling out the existence of other galactose transporters. It was later concluded that *LAC12* and *HGT1* encode a low-affinity and high-affinity galactose system, respectively.

Much of what is known from *K. marxianus* biology has been inferred from *K. lactis* studies. However, there are substantial differences between *K. marxianus* and *K. lactis* that suggest that the two species are more different than originally thought (Lane and Morrissey, 2010). Besides being phenotypically distinct (*K. marxianus* is more thermotolerant and has a higher growth rate), differences in genome structure and lactose utilisation variability point out to additional differences that further separate the two species as genetically and phenotypically unique. One of these differences is the number of *LAC12* genes present in each species. *K. lactis* contains a single copy of *LAC12* whereas *K. marxianus* has four copies of the gene (Varela *et al*., 2017b). Heterologous expression of the additional *K. marxianus LAC12* genes in *S. cerevisiae* (a naturally lactose negative yeast) showed that none of them encode a

functional lactose transporter however, it is possible that some of these genes are involved in the uptake of other sugars. Studying gene function in *K. marxianus* has proven laborious in past due to the lack of adequate molecular tools. But the recent implementation of the CRISPR-Cas9 system in *K. marxianus* allows performing genetic modifications readily thus opening new opportunities to study the molecular basis of sugar uptake (Juergens *et al*., 2018; Löbs *et al*., 2017; Nambu-Nishida *et al*., 2017).

Here, we have investigated the role of the *LAC12* and *HGT1* genes in galactose utilisation in *K. marxianus*. Our findings show an expansion in the number of genes that encode the *KHT* and *HGT* transport systems in this species Functional characterisation revealed a high level of functional overlap between the *HGT* and *LAC12* genes, with several of them encoding galactose transporters. This work highlights the existence of a species-specific gene expansion and raises fascinating questions about the evolution of galactose and lactose utilisation in *K. marxianus*.

4.2 Materials and methods

Strains and growth conditions

K. marxianus CBS6556 and NBRC1777 were obtained from the CBS-KNAW and the NITE Biological Research Centre culture collection, respectively. *S. cerevisiae* EBY.VW4000 was kindly provided by Dr Eckhard Boles, Goethe University Frankfurt, Germany. *K. marxianus* strains were typically grown in YPD medium (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose). When used in transformation experiments NBRC1777 was plated onto YPD plates containing 200 µg/mL of Hygromycin B (Sigma-Aldrich, MI., USA). Mutants were grown on mineral media (MM) supplemented with 20 g/L and 0.1 g/L glucose, galactose, mannose, fructose or sucrose (Fonseca *et al*., 2007). *S. cerevisiae* EBY.VW4000, used for heterologous expression of sugar transporters (Wieczorke *et al*., 1999), was maintained in synthetic complete (SC) medium (1.7 g/L yeast nitrogen base, 5 g/L ammonium plus synthetic complete drop-out lacking uracil and or L-leucine) (Formedium, Norfolk, United Kingdom) supplemented with 20 g/L maltose. For functional analysis the EBY.VW4000 strains expressing the different transporters were grown on SC medium supplemented with glucose, galactose, mannose or fructose. All the sugars were used at a final concentration of 20 g/L, 10 g/L or 1 g/L in the medium, as mentioned in the text. Yeast strains were grown at 30°C with 200 rpm agitation unless otherwise mentioned. *E. coli* DH5α was used for cloning purposes. The strain was maintained in LB medium (5 g/L yeast extract, 10 g/L bactopeptone, 10 g/L NaCl) supplemented with 100 µg/mL ampicillin when required.

Identification of potential sugar transporters

A genome-wide identification of sugar transporters was performed in the *K. marxianus* CBS6556 and *K. lactis* CBS2359 genomes. The predicted proteomes from the two species were independently compared against the TransportDB 2.0 database to identify possible candidates. After filtering out hits with low identity (< 35 %) and coverage (< 80 %) the resulting sequences were submitted to the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) for the prediction of transmembrane domains. All sequences containing 2 or more transmembrane domains were considered as potential transporters. In order to classify these transporters, their sequences were aligned against the PFAM database (Finn *et al*.,

2014) using HMMER v 3.0 (Finn *et al*., 2011). Sequences belonging to the MFS superfamily were extracted and further analysed. Amino acid sequences were aligned using MUSCLE 3.8 (Edgar, 2004) and then used to construct a maximum-likelihood tree using PhyML 3.1 (Guindon *et al*., 2010). Tree visualisation was carried out using the ETE 3 toolkit (Huerta-Cepas *et al*., 2016).

To compare the structure of the *KHT* and *HGT* loci across species, the publically available *Kluyveromyces* genomes were aligned and visualised using progressiveMauve (Darling *et al*., 2010). *K. marxianus* genomes from different strains (CBS6556, NBRC1777, UFV-3, CBS397, CBS4857 and DMKU3) were also analysed following the same procedure.

Heterologous expression of sugar transporters

Plasmids and primers used in this study are listed in **Table 4.1** and **Table 4.2**, respectively. The plasmid p426 was used to clone and express the *KHT* and *HGT* genes in *S. cerevisiae*. The p426 backbone, *ScTEF1* promoter, *ScCYC1* terminator and the E03650 gene from *K. marxianus* CBS6556 were amplified by PCR with primers containing overlapping regions. PCR products were then purified and assembled using a Gibson assembly kit (New England Biolabs Inc., MA, USA). The resulting plasmid, which expresses the E03650 gene under the control of the *ScTEF1* promoter, was used as a backbone for cloning the rest of the transporters. The primers p426-TEF-F and p426-CYC-R, were used to PCR amplify the p426-E03650 plasmid from the *ScTEF1* promoter to the *ScCYC1* terminator (excluding the E03650 gene). Then, purified PCR products from the rest of the CBS6556 transporters were individually combined with the purified p426 backbone and cloned by Gibson assembly. The *KHT* genes 50343 and 50344, absent in CBS6556, were PCR amplified from NBRC1777 and cloned following the same procedure. All the PCR reactions were performed using DNA Phusion Polymerase (New England Biolabs Inc., MA, USA). Plasmids were finally transformed into EBY.VW4000 using the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl, 2007). Expression of the *LAC12* genes was also tested using plasmids previously described (Varela *et al*., 2017b). After transformations, the cells transformed with the p426-*KHT/HGT* and *LAC12* plasmids were plated onto SC media lacking uracil and leucine, respectively.

Table 4.1: Plasmids used in this study

Plasmid	**Source**
p426	Addgene
p426-TEF1-E03650	This study
p426-TEF1-E03660	This study
p426-TEF1-E03670	This study
p426-TEF1-E03680	This study
p426-TEF1-E03690	This study
p426-TEF1-50344	This study
p426-TEF1-50343	This study
p426-TEF1-A02920	This study
p426-TEF1-A02930	This study
p426-TEF1-A02940	This study
p426-TEF1-A02950	This study
p426-TEF1-A02960	This study
p426-TEF1-E00380	This study
pGREG-505 - TEF1	Varela *et al.* 2017
pGREG-505 - TEF1 - *KmLAC12* - CBS 6556	Varela *et al.* 2017
pGREG-505 - TEF1 - *KmLAC12* - CBS 397	Varela *et al.* 2017
pGREG-505 - TEF1 - *LAC12-2* - CBS 397	Varela *et al.* 2017
pGREG-505 - TEF1 - *LAC12-3* - CBS 397	Varela *et al.* 2017
pUDP002	Juergens *et al.* 2018
pUCC001	This study
pUCC001-*KmLAC12*	This study
pUCC001-*LAC12-2*	This study
pUCC001-*LAC12-4*	This study
pUCC001-E03670	This study
pUCC001-*HGT*	This study

Table 4.2: Primers used in this study. The section of the primer that anneals to the target sequence is shown in uppercase. Sequences in lower case are homologous to the plasmid where the sequences were cloned.

Heterologous expression of *KHT* and *HGT* genes

Primer	Sequence
p426-F	TTCGCCAGCTGGCGTAATAG
p426-R	GGGAGAGGCGGTTTGCGTATTG
ScTEF1-F	gcccaatacgcaaaccgcctctcccCATAGCTTCAAAATGTTTCTACTC
ScTEF1-R	cagcttcggacatCTTAGATTAGATTGCTATGCTTTC
ScCYC1-F	tatgatgaaataaTCATGTAATTAGTTATGTCACGC
ScCYC1-R	cttcgctattacgccagctggcgaaGCAAATTAAAGCCTTCGAGC
E03650-F	aatctaatctaagATGTCCGAAGCTGCTGGTTTAC
E03650-R	actaattacatgaTTATTTCATCATAGCCTTGTACCATG
p426-TEF-F	CTTAGATTAGATTGCTATGCTTTC
p426-CYC-R	TCATGTAATTAGTTATGTCACGC
E03660-F	gcatagcaatctaatctaagATGTCTGAAGCTGCCGCTG
E03660-F	tgacataactaattacatgaTTAATGCTTCATCATGGCCTTG
E03670-F	gcatagcaatctaatctaagATGTCTGAAGAAGCTGCATTACAG
E03670-R	tgacataactaattacatgaTTAGGACGTCATGCGCTTG
E03680-F	gcatagcaatctaatctaagATGTCCAATCAATTAACGG
E03680-R	tgacataactaattacatgaTTAGTTCTTCTTGAAGGACATG
E03690-F	gcatagcaatctaatctaagATGTCTAATCAATTGACGGC
E03690-R	tgacataactaattacatgaTTATTTCAAAGAAATCCTCTTG
50343-F	gcatagcaatctaatctaagATGTCTGAAGCTGCTGCTGATTTACA

50343-R	tgacataactaattacatgATTAATGCTTCATCATGGCCTTGTACCA
50344-F	gcatagcaatctaatctaagATGTCTGAAGCTGCCGCTGA
50344-R	gcgtgacataactaattacatgaCTCGAGGTCGACTTAATGCTTCAACAAAGCCTTGTACCAT
A02920-F	gcatagcaatctaatctaagATGACTTTAAAAGATAAACTATTGCTCC
A02920-R	tgacataactaattacatgaCTAGACCGAGCTGCTGCTATTAG
A02930-F	gcatagcaatctaatctaagATGTCATTCTTAGACAAGAAAAC
A02930-R	tgacataactaattacatgaTCAAACATTGTCATTCCTTAC
A02940-F	gcatagcaatctaatctaagATGTCATTTAAAGACAAGTTTTC
A02940-R	tgacataactaattacatgaTTAAACGCTGTTACCAGAG
A02950-F	gcatagcaatctaatctaagATGAAACAATTCGCTACG
A02950-R	tgacataactaattacatgaTTATACGCGAGAGTCGTTC
A02960-F	gcatagcaatctaatctaagATGTCATTGAAAGACAAGATTTTG
A02960-R	tgacataactaattacatgaTTAGTTAGAGTTTGAGTTTGAGTTG

Construction of pUCC001

Primer	Sequence
pUDP002-F	tcggacgagcttactcgtttcgtcctcacggactcatcagGTTTGTTTGTTTATGTGTGTTTATTC
pUDP002-R	gcgccggctgggcaacatgcttcggcatggcgaatgggacCACAGGCCCCTTTTCCTTTG
	CTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCGAGACCTGCGGAGGTCTCCGTTTTAGAGCTAG
HH-BSA-HDV-F	AAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
	GTCCCATTCGCCATGCCGAAGCATGTTGCCCAGCCGGCGCCAGCGAGGAGGCTGGGACCATGCCGGCCAA
HH-BSA-HDV-R	AAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTG

Construction of CRISPR plasmids

Primer	Sequence
*Km*LAC12-gRNA-F	CGTCTCGTCTCTTGGATTTCTTCA
*Km*LAC12-gRNA-R	AAACTGAAGAAATCCAAGAGACGA
LAC12-2-gRNA-F	CGTCCGCTAAAGCCGCGCCTAGAG
LAC12-2-gRNA-R	AAACCTCTAGGCGCGGCTTTAGCG
LAC12-4-gRNA-F	CGTCGTGCTGCATTAGCATTATCA
LAC12-4-gRNA-R	AAACTGATAATGCTAATGCAGCAC
E03670-gRNA-F	CGTCGTACGTCTTGGTACCGTAGT
E03670-gRNA-R	AAACACTACGGTACCAAGACGTAC
HGT-gRNA-F	CGTCCGCTCCAACCTGGGGTAT
HGT-gRNA-R	AAACATACCCCAGGTTGGAGCG
BSA-R	TACACGCGTTTGTACAGAAAAAAAAGAAAAATTTGA

Diagnostic primers to check for CRISPR-*Cas9*-induced mutagenesis

Primer	Sequence
Diag*KmLAC12*-F	TTTGGTTGGTTAATCCCAGAATCTCCAAG
Diag*KmLAC12*-R	GGATGCCTTTCTTGGGTTTGGAG
Diag*LAC12-2*-F	CGGGTCTCGTCTGTATTTTCGGT
Diag*LAC12-2*-R	TAGGCTGCATAGGAGTAAATGCAAACG
Diag*LAC12-4*-F	TTGTCGTCTCATCTGTGGTCACG
Diag*LAC12-4*-R	TTCATAGCTCTGGGTGCGGC
DiagE03670-F	GAAGTTGACAATACCCAAGACGATGGA
DiagE0367*0*-R	CTCTGGTTTGGGTGTTGGTGGT

Diag*HGT*-F	CTCATCTTGGATAAAATTTGCTTG
Diag*HGT*-R	ATTAAAAGGGAGAGAGGGTGG

Construction of *K. marxianus* mutants

The CRISPR-Cas9 system was used to disrupt the *LAC12* and *HGT* genes in *K. marxianus* NBRC1777. The pUDP002 plasmid, containing the *Cas9* gene and the elements required to express the guide RNA (gRNA) molecule in *K. marxianus* (Juergens *et al*., 2018), was modified to simplify cloning of target sequences. Long overlapping oligonucleotides encoding the hammerhead ribozyme, a BsaI cloning site, the gRNA structural part and the HDV ribozyme were chemically synthesised (Integrated DNA Technologies, IA, USA) and annealed in a thermocycler machine. In parallel, the pUDP002 plasmid was amplified by PCR using the pUDP002-F and pUDP002-R primers. The 205 bp DNA-duplex and the plasmid backbone were assembled via Gibson assembly to yield the pUCC001 plasmid. Design of the gRNA sequences targeting the genes of interest was carried out using the software sgRNAcas9 (Xie *et al*., 2014). Complementary oligonucleotides containing the gRNA sequences and specific overhangs (5' CGTC 3' and 5' AAAC 3') were annealed, phosphorylated and cloned into pUCC001 by Golden Gate assembly, as described previously (Ng and Dean, 2017; Vyas *et al*., 2015). Presence of the target sequences in the pUCC001 plasmid was corroborated by colony PCR using the forward target primer and BSA-R. Plasmids were transformed into NBRC1777 using the LiAc/SS carrier DNA/PEG method and transformants checked by PCR and DNA sequencing. In order to cure the pUCC001 plasmid, the mutant strains were grown on YPD broth overnight without selection and diluted 1/10 into fresh YPD next morning. The procedure was repeated twice and then the strains were grown on YPD medium supplemented with Hygromycin B to corroborate plasmid loss. Construction of additional mutations was achieved by transforming plasmids sequentially into a plasmid-curated mutant. Primers used for cloning the gRNA molecules and checking transformants are listed in **Table 4.2**. The genotype of all the *K. marxianus* mutants constructed in this study is shown in **Table 4.3**.

Table 4.3: Genotype of *K. marxianus* mutants constructed in this study. Mutation coordinates refer to the position where the mutations are located in the NBRC1777 genome. Single or double nucleotide deletions are represented by a (-) symbol followed by the nucleotide/s deleted.

Strain	**Genotype**	**Mutation coordinates**
NBRC1777 Δ*LAC*	*Kmlac12 lac12-2 lac12-4*	*KmLAC12:*CHR III 16,869..-T
		LAC12-2: CHR III 1,557,231..-T
		LAC12-4: CHR VIII 12,834..-GT
NBRC1777 Δ*E03670*	ΔE03670	E03670: CHR V 782,125..-T
NBRC1777 Δ*HGT*	Δ*HGT*	CHR I Δ1,131,464-1,145,439
NBRC1777 Δ*LAC* Δ*E03670*	*Kmlac12 lac12-2 lac12-4* E03670	*LAC12* genes: same as Δ*LAC*
		E03670: CHR V 782,125..-T
NBRC1777 Δ*LAC* Δ*HGT*	*Kmlac12 lac12-2 lac12-4* Δ*HGT*	*LAC12* genes: same as Δ*LAC*
		HGT: CHR I Δ1,131,464-1,145,439
NBRC1777 Δ*HGT* ΔE03670	Δ*HGT* ΔE03670	*HGT:* CHR I Δ1,131,464-1,145,439
		E03670: CHR V 782,125..-T
NBRC1777 Δ*LAC* Δ*HGT*	*Kmlac12 lac12-2 lac12-4* Δ*HGT*	
ΔE03670	ΔE03670	*LAC12* genes: same as Δ*LAC*
		HGT: CHR I Δ1,131,464-1,145,439
		E03670: CHR V 782,125..-T

4.3 Results

Extra *LAC12* copies encode galactose transporters

It has been recently shown that *K. marxianus* contains four copies of the lactose permease gene with only *KmLAC12* encoding a functional lactose transporter in some strains (Varela *et al*., 2017b). Because the *K. lactis* Lac12p protein is known to transport lactose and galactose (Riley *et al*., 1987), it was hypothesised that the *K. marxianus LAC12* copies could encode functional galactose transporters. To test this idea, the *LAC12* copies were expressed in *S. cerevisiae* EBY.VW4000 (**Figure 4.1A**), a yeast strain that has been subjected to systematic disruption of all known hexose transporters (Wieczorke *et al*., 1999). The *S. cerevisiae* strains expressing the *KmLAC12*, *LAC12*-2 and *LAC12*-4 grew in 2 % galactose medium possibly indicating that these genes encode transporters with low-affinity galactose transport activity. It was also noted that the strain carrying the *KmLAC12* gene from *K. marxianus* CBS397, previously shown to encode a functional lactose transporter, grew better than the strain expressing the *KmLAC12* gene from CBS6556, a non-functional lactose transporter.

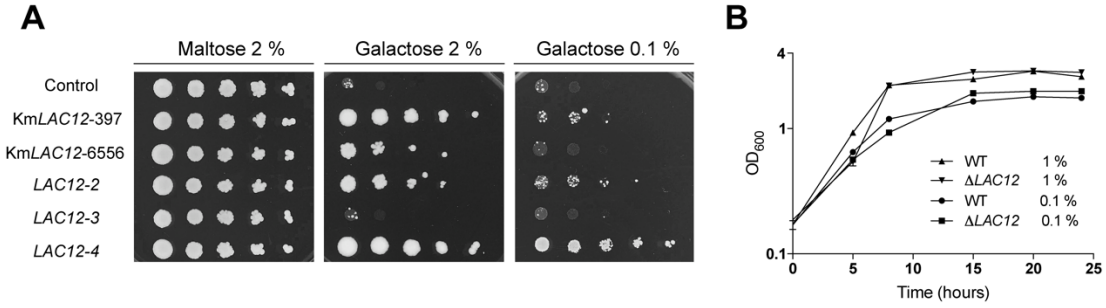


Figure 4.1: Role of the *LAC12* genes in galactose transport. **A.** Heterologous expression of the *K. marxianus LAC12* genes in *S. cerevisiae* EBY.VW4000. Yeast strains were grown on SC maltose to an OD_{600} of 2 then washed, diluted serially and spotted onto SC plates supplemented with maltose, galactose or glucose, as indicated. Plates were incubated for 5 days at 30°C. The *LAC12-2* and *LAC12-4* genes were obtained from *K. marxianus* CBS397, as described by (Varela *et al*., 2017b). **B.** Growth phenotype of *K. marxianus* Δ*LAC* mutant carrying null mutations for the *KmLAC12*, *LAC12-2* and *LAC12-4* genes. NBRC1777 and the Δ*LAC* strain were grown overnight on MM supplemented with 1 % or 0.1 % galactose and then transferred to fresh medium containing the same amount of sugar to an OD_{600} of 0.1. Growth was recorded every 6 hours using a spectrophotometer. The experiment was conducted in triplicates.

In order to check whether the *LAC12* genes encoded high-affinity galactose transporters, the different strains were also grown on galactose 0.1 %. In this case, the strain expressing the *LAC12*-4 copy exhibited strong growth while the *KmLAC12*-397 and *LAC12*-2 strains only grew weakly and the *LAC12*-3 not at all. Since these results suggested that *KmLAC12*, *LAC12*-2 and *LAC12*-4 encode galactose transporters, we then tested if a *K. marxianus* strain carrying non-functional versions of these genes could still utilise galactose as sole carbon source. The previously published CRISPR-Cas9 system (Juergens *et al*., 2018) was modified to allow fast cloning of target sequences (**Figure 4.2**) and used to sequentially target the three genes in NBRC1777. The triple mutant (NBRC1777 *ΔLAC*) and the wild-type strain were grown in 1 % and 0.1 % galactose media but no major differences in growth were observed (**Figure 4.1B**). Altogether, these findings point out to the existence of an additional galactose transporter(s) in *K. marxianus*. An idea compatible with previous studies in *K. lactis* where the *HGT1* gene was shown to encode a high-affinity galactose transporter (Baruffini *et al*., 2006).

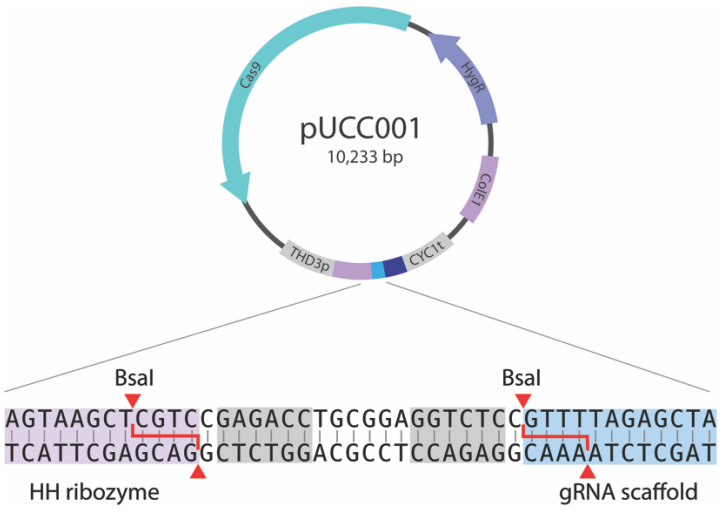


Figure 4.2: Map of the CRISPR-Cas9 plasmid used in this study.
The pUCC001 plasmid was obtained by introducing a double-restriction site for the type IIS restriction enzyme, BsaI (grey shading) into the pUDP002 plasmid (Juergens *et al*., 2018). The desired target sequence can be chemically synthesized as a pair of complementary DNA oligos containing specific overhangs compatible with pUCC001. This oligo duplex can then be cloned between the HH ribozyme and the gRNA scaffold sequences (purple and blue shading, respectively) by Golden Gate assembly. Figure modified from Figure 2.5, Chapter 2.

KHT and *HGT* sugar transporter genes are expanded in *K. marxianus*

In order to identify other possible galactose transporters in *K. marxianus*, a genome-wide annotation of membrane transporters was performed. Predicted proteins from *K. marxianus* and *K. lactis* were annotated and categorised into protein families. Protein sequences from both species belonging to the major facilitator superfamily (MFS) were compared through a phylogenetic tree revealing an expansion in *K. marxianus* sugar transporters (data not shown). To gain a better understanding of the relationship between these sequences a maximum-likelihood phylogenetic tree was constructed including the Kht1/2p, Hgt1p and Lac12p amino acid sequences from *K. lactis* and *K. marxianus* (**Figure 4.3**).

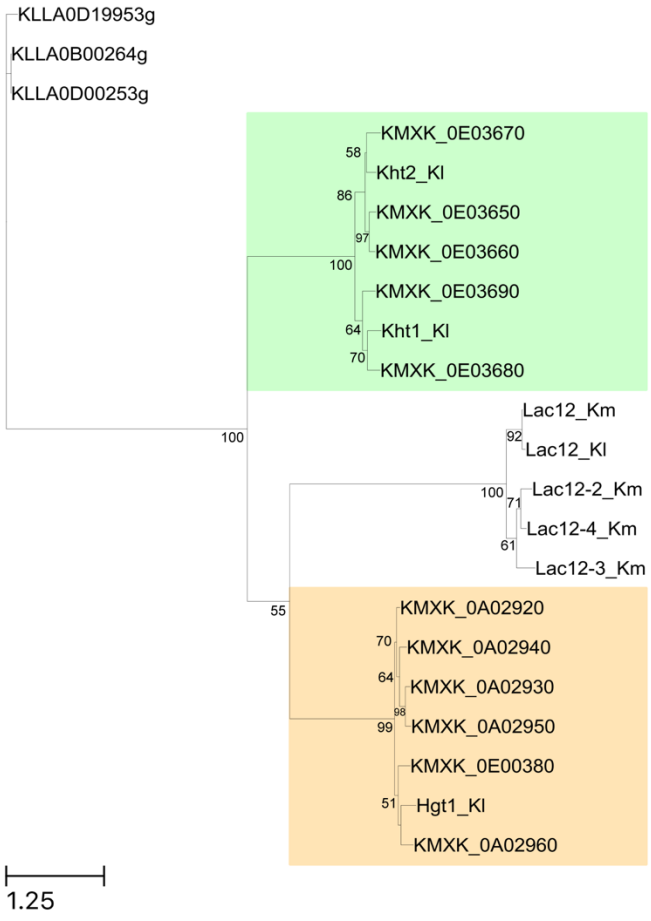


Figure 4.3: Phylogenetic tree of *K. marxianus* sugar transporters. Sequences corresponding to the Kht, Hgt and Lac12 proteins were aligned using MUSCLE and then used to construct a maximum-likelihood tree using PhyML with a bootstrapping value set to 1000. Branches with bootstrapping values over 50 are shown. The Kht and Hgt sequences are highlighted in green and brown boxes, respectively. The *K. lactis* maltose permease genes KLLA0D19953g, KLLA0B00264g and KLLA0D00253g were used as outgroups for the tree construction.

As previously described, the *K. lactis* genome was found to contain a single Lac12p transporter while *K. marxianus* genome contained four copies of this transporter (Varela *et al*., 2017b). Also, an amplification in the number of Kht1/2p and Hgt1p sugar transporters was also found in this species. Five copies of the putative low-affinity glucose transporter Kht1/2p and six copies of the putative high-affinity glucose transporter Hgt1p were identified in this analysis, suggesting that *LAC12* is not the only gene family that has undergone expansion in this yeast. The *K. marxianus* Kht1/2p transporters are 75 % identical when compared to each other. Two of these sequences, E03690 and E03680, are orthologues of the *K. lactis* Kht1p gene, while E03650, E03660 and E03670 are orthologous to Kht2p (**Table 4.4**). On the other hand, the Hgt1p sequences show an overall identity of 78 %. All of these protein sequences show a high resemblance to the *K. lactis* Hgt1p transporter (**Table 4.5**).

Table 4.4: Identity matrix of the Kht transporter sequences from *K. marxianus* CBS6556 and *K. lactis* CBS2359. The matrix was obtained by comparing the corresponding protein sequences using the MUSCLE server (https://www.ebi.ac.uk/Tools/msa/muscle/). The Kht1 and Kht2 homologues in *K. marxianus* are shown in red and blue, respectively.

	Kht1p_*Kl*	E03690	E03680	Kht2p_*Kl*	E03650	E03660	E03670
Kht1p_*Kl*	100	71.63	78.27	78.99	79.45	74.98	75.81
E03690	78.27	100	80.58	75.38	77.15	76.91	73.41
E03680	78.99	80.58	100	77.07	78.3	81.26	75.7
Kht2p_*Kl*	79.45	75.38	77.07	100	83.19	83.54	78.12
E03650	74.98	77.15	78.3	83.19	100	87.95	78.19
E03660	75.81	76.91	81.26	83.54	87.95	100	79.55
E03670	71.63	73.41	75.7	78.12	78.19	79.55	100

Table 4.5: Identity matrix of the putative Hgt transporters from *K. marxianus* CBS6556 and *K. lactis* CBS2359. Matrix constructed from MUSCLE protein sequence alignment.

	Hgt1p_*Kl*	A02920	A02930	A02950	A02940	A02960
Hgt1p_*Kl*	100	76.6	77.52	76.48	78.08	79.29
A02920	76.6	100	78.28	78.06	80.67	81
A02930	77.52	78.28	100	90.27	84.39	84.14
A02950	76.48	78.06	90.27	100	84.85	84.98
A02940	78.08	80.67	84.39	84.85	100	88.2
A02960	79.29	81	84.14	84.98	88.2	100

An inspection of the genomic context of these genes revealed that the *KHT1/2* and *HGT1* genes are present as tandem gene copies with the exception of E00380 that is not present in the same location as the rest of the *HGT1* genes. In order to test whether the expansion in these gene families could be found in other *K. marxianus* strains, the *KHT* and *HGT* loci were compared across a set of publicly available *K. marxianus* genomes and other *Kluyveromyces* genomes (*K. lactis*, *K. aestuarii*, *K. wickerhamii* and *K. dobzhanskii*) (**Figure 4.4**). Expansion of the *KHT* and *HGT* genes was found in all the *K. marxianus* genomes analysed. In contrast, the rest of the *Kluyveromyces* genomes contain two *KHT* genes and a unique *HGT* gene. Interestingly, copy number was found to vary across *K. marxianus* strains. While the *K. marxianus* strains CBS6556, DMKU3 and CBS397 carry five copies of the *KHT* genes, the strains NBRC1777, UFV3 and CBS4857 encode six copies of the transporter (**Figure 4.4A**). An alignment of the NBRC1777 and CBS6556 sequences showed that the 5' and 3' sections of the E03660 sequence are identical to KMAR_50344 and KMAR_50345, respectively, suggesting that E03660 arose by a recombination event between two *KHT* genes (**Figure 4.5**). A similar scenario was found in the *HGT1* locus where some strains carry five (NBRC1777, CBS6556, CBS397 and CBS4857) or three (DMKU3 and UFV3) copies of the gene (**Figure 4.4B**). A comparison of the NBRC1777 and DMKU3 sequences suggested that this reduction in copy number occurred as a result of a recombination event between the second and fourth *HGT* gene, leading to a loss of the third gene (**Figure 4.6**). Altogether, these results suggest that the *KHT* and *HGT* gene expansions are specific to *K. marxianus*. Also, differences in copy number between strains likely indicate that these gene clusters have a recent origin.

Expression of genes involved in galactose metabolism, including *LAC12* and *HGT1*, is activated by the Gal4p/Lac9p transcriptional factor in *K. lactis* (Godecke *et al*., 1991). This activation system is conserved in *K. marxianus* with several Galp4 binding sites present in the *LAC12-LAC4* promoter (Varela *et al*., 2017b). Interestingly, no Gal4p sites were found in the promoters of the *HGT* genes from *K. marxianus*. However, two Gal4p sites were identified in the promoter of the *KHT* gene E03670, indicating that this gene is potentially regulated by the Gal4p protein.



Figure 4.4: Organisation of the *KHT* and *HGT* genes in *K. marxianus* genomes.
A. Genomic context of the *KHT* genes in *K. marxianus* strains NBRC1777, UFV3 and CBS4857 (top), CBS6556, DMKU3 and CBS397 (middle) and rest of *Kluyveromyces* genomes, represented by *K. lactis* (bottom). **B.** Position of the *HGT* genes in the CBS6556, NBRC1777, CBS397 and CBS4857 genomes (top), DMKU3 and UFV3 (middle) and *K. lactis* (bottom). Sequence conservation across the different regions is represented by grey areas connecting the genes in different genomes.

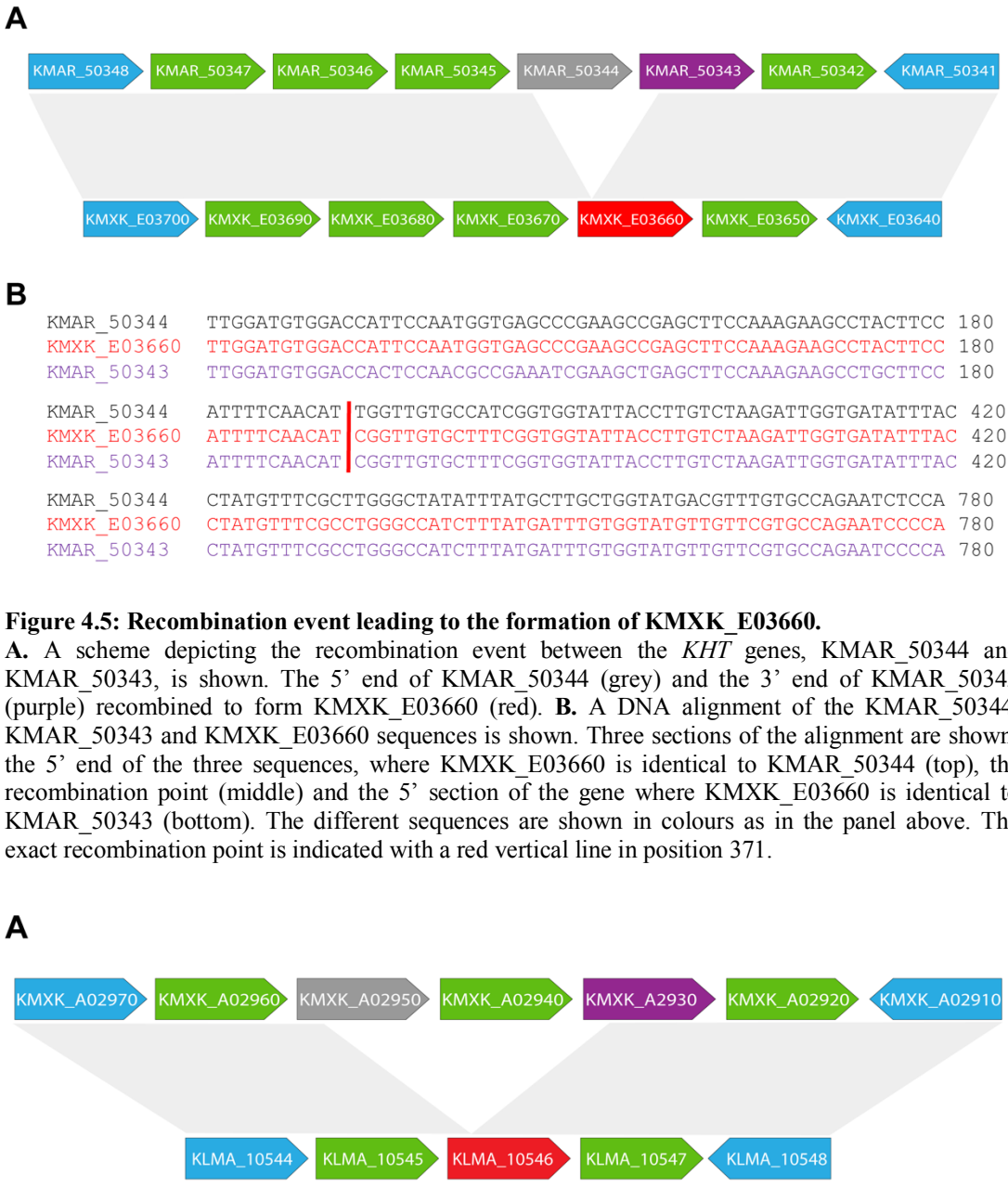


Figure 4.5: Recombination event leading to the formation of KMXK_E03660.
A. A scheme depicting the recombination event between the *KHT* genes, KMAR_50344 and KMAR_50343, is shown. The 5' end of KMAR_50344 (grey) and the 3' end of KMAR_50343 (purple) recombined to form KMXK_E03660 (red). **B.** A DNA alignment of the KMAR_50344, KMAR_50343 and KMXK_E03660 sequences is shown. Three sections of the alignment are shown; the 5' end of the three sequences, where KMXK_E03660 is identical to KMAR_50344 (top), the recombination point (middle) and the 5' section of the gene where KMXK_E03660 is identical to KMAR_50343 (bottom). The different sequences are shown in colours as in the panel above. The exact recombination point is indicated with a red vertical line in position 371.

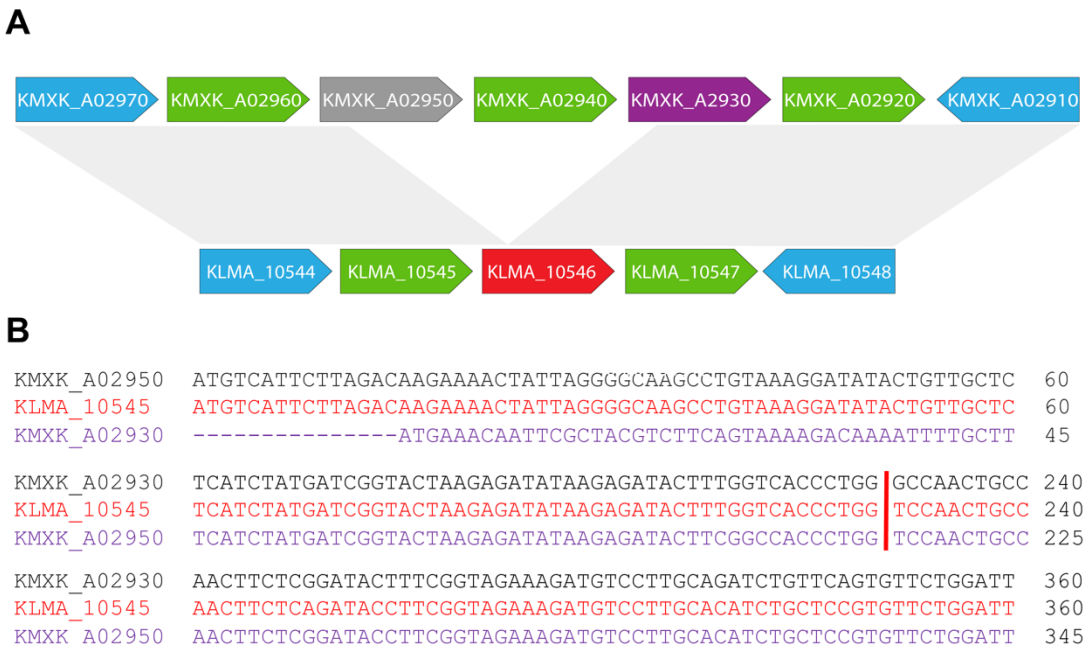


Figure 4.6: Recombination event leading to the formation of KLMA_10546.
A. A scheme depicting the recombination event between KMXK_A02950 and KMXK_A02930 is shown. The 5' end of KMXK_A02950 (grey) and the 3' end of KMXK_A02930 (purple) recombined to form KLMA_10546 (red). **B.** A DNA alignment of these sequences is shown as in Figure 4.5. The exact recombination point is indicated with a red vertical line in position 230.

KHT and *HGT* sugar transporter encode additional galactose transporters

To functionally characterise these genes and to test whether they encoded galactose transporters, the *K. marxianus* CBS6556 genes were cloned and expressed in *S. cerevisiae* EBY.VW4000 (**Figure 4.7**). All the *S. cerevisiae* strains expressing *KHT* genes grew on media containing high and low glucose concentrations, however only the strain expressing the E03670 gene grew on galactose 2 % medium (**Figure 4.7A**). On the other hand, no growth was observed on galactose 0.1 % medium, probably indicating that none of these genes encoded high-affinity galactose transporters. The KMAR_50344 and KMAR_50345 genes that presumably recombined to form E03660 in CBS6556 were also tested. In this case, expression of both genes conferred growth on glucose suggesting that they encode glucose transporters (**Figure 4.8**). Strains expressing the *HGT* genes displayed a different growth pattern (**Figure 4.7B**). Expression of three of these genes (A02960, A02940 and A02920) conferred growth on galactose 2 % medium, while expression of A02950 conferred weak growth on galactose 0.1 %. Growth on glucose media was only observed in the strain expressing the A02960 gene showing that the *KHT* and *HGT* genes are functionally distinct. Heterologous expression of the KLMA_10546 gene (arisen by recombination between KMXK_A02950 and KMXK_A02930) was not performed because the DMKU3 strain was not available when the experiments were carried out. Nevertheless, these results obtained here confirm the existence of additional galactose transport systems in *K. marxianus*.

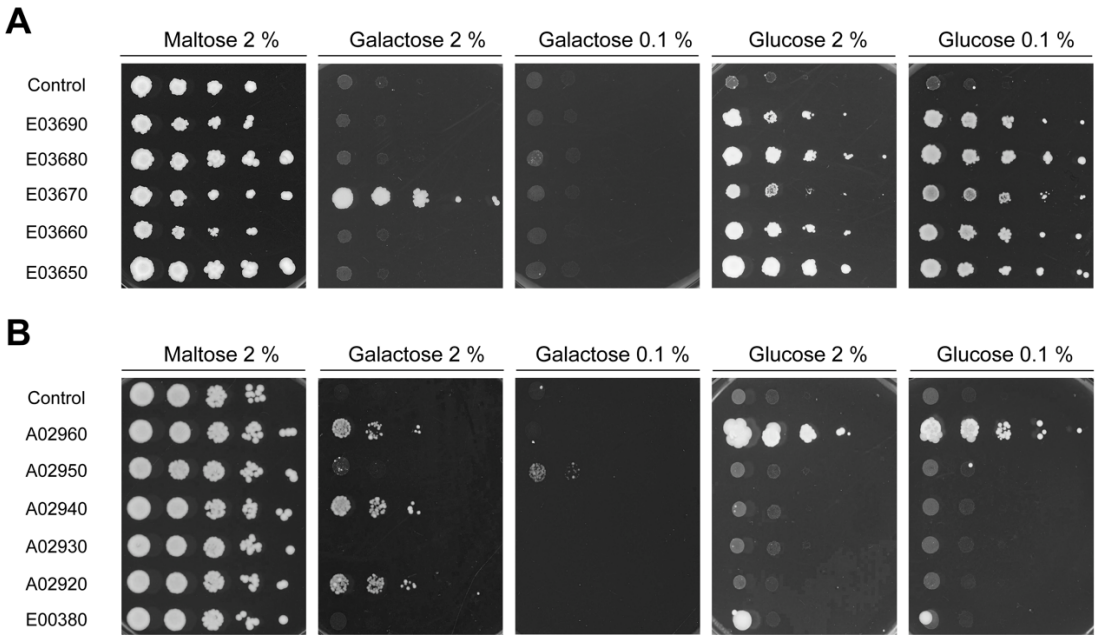


Figure 4.7: Functional analysis of the *KHT* and *HGT* genes in *S. cerevisiae* EBY.VW4000.
S. cerevisiae strains expressing the *KHT* and *HGT* genes from CBS6556 were grown as previously described and spotted onto SC plates containing maltose, galactose, or glucose. **A.** *KHT* genes. **B.** *HGT* genes.

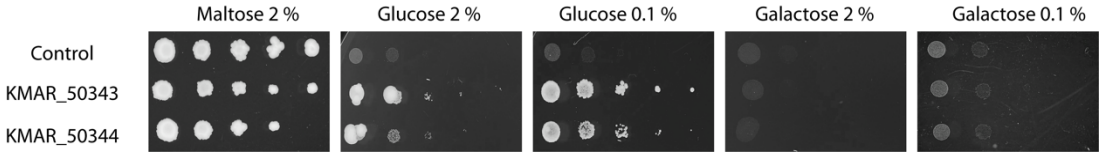


Figure 4.8: Functional analysis of the KMAR_50343 and KMAR_50344 genes in *S. cerevisiae* EBY.VW4000.
S. cerevisiae strains expressing the two genes from NBRC1777 were grown as previously described and spotted onto SC plates containing maltose, galactose, or glucose.

The data obtained by overexpressing the *HGT* and *KHT* genes in *S. cerevisiae* does not necessarily correlate with their function in *K. marxianus*. To characterise the physiological role of these transporters, the CRISPR-Cas9 was used to mutate the E03670 gene and the *HGT* genes in *K. marxianus*. To simplify this experiment, a simultaneous mutagenesis of all the *HGT* genes was attempted by targeting a common sequence present in all these genes. Efficient targeting of these genes by the CRISPR-Cas9 should result in five double-strand breaks in *K. marxianus* genome. The repair of these breaks only relies on the endogenous DNA repair machinery, which should result in indel mutations in all genes. However, since these genes

display high sequence similarity, repair through a recombination event could take place, resulting in a deletion of the whole region containing the *HGT* genes. Transformation of the CRISPR-Cas9 plasmid targeting the *HGT* genes into *K. marxianus* NBRC1777 resulted in a low number of colonies (46). DNA was extracted from six of these colonies and PCR reactions were performed to check for a full deletion of the *HGT* region (**Figure 4.9A**). Amplification of the *HGT* region in the six candidates, but not in the wild-type strain, resulted in the expected 2 kb product, confirming the deletion of the *HGT* region (**Figure 4.9B**). Sequencing analysis in one of these candidates (NBRC1777 *ΔHGT*) revealed that the recombination event took place between the A02960 and A02920 genes, resulting in a hybrid sequence that contains several stop codons rendering the sequence of the single remaining gene non-functional.

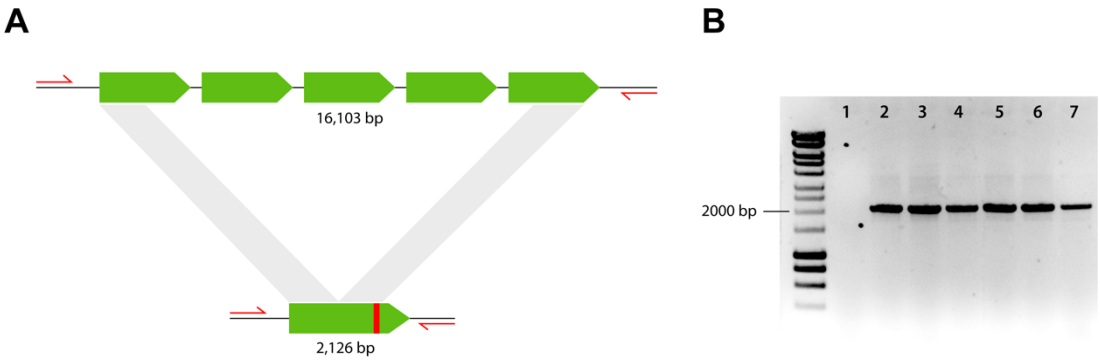


Figure 4.9: Construction of an *HGT* null mutant in *K. marxianus* NBRC1777. The CRISPR-*Cas9* system was used to induce mutagenesis of the *HGT* genes by targeting the 5' GCTCCAACCTGGGGTAT 3' sequence present in all *HGT* genes with a single gRNA. **A.** Scheme showing the repair event occurred upon targeting the *HGT* genes with the CRISPR-*Cas9* system. Recombination between the two ends of the *HGT* loci is shown by grey shading. Stop codon in the recombinant *HGT* is displayed in red. Primers used to screen the mutant colonies are represented by red arrows. **B.** Screening of *HGT* mutant colonies. Transformants were screen by PCR using the Diag*HGT* primers. Lane 1, NBRC1777 wild-type. Lanes 2-7, NBRC1777 strains transformed with the pUCC001-*HGT* plasmid. DNA ladder Hyperladder I.

To test the effect of the different mutations in galactose transport, strains carrying combinations of the *LAC, HGT* and E03670 mutations were constructed and grown on galactose and glucose media (**Figure 4.10**). None of the mutants showed growth defects when grown on glucose media, indicating that these mutations did not affect glucose transport. However, a different pattern was observed in galactose media where some mutants displayed a significant decrease in growth. Strains carrying single mutations (i.e. *ΔLAC*, *Δ*E03670 or *ΔHGT*) did not show growth differences when compared to the wild-type strain, but a combination of these mutations affected growth on galactose. On 2 % galactose, only the strain carrying a mutation in all identified galactose transporter genes (*ΔLAC Δ*E03670 *ΔHGT*) showed a significant decrease in growth, indicating that these genes are required for galactose transport in *K. marxianus*. On the other hand, the *ΔLAC ΔHGT* and *ΔLAC Δ*E03670 *ΔHGT* strains failed to grow on 0.1 % galactose. The *ΔHGT Δ*E03670 mutant showed a decrease in growth when compared to the wild-type strain but grew better than *ΔLAC ΔHGT* and *ΔLAC Δ*E03670 *ΔHGT* on this medium.

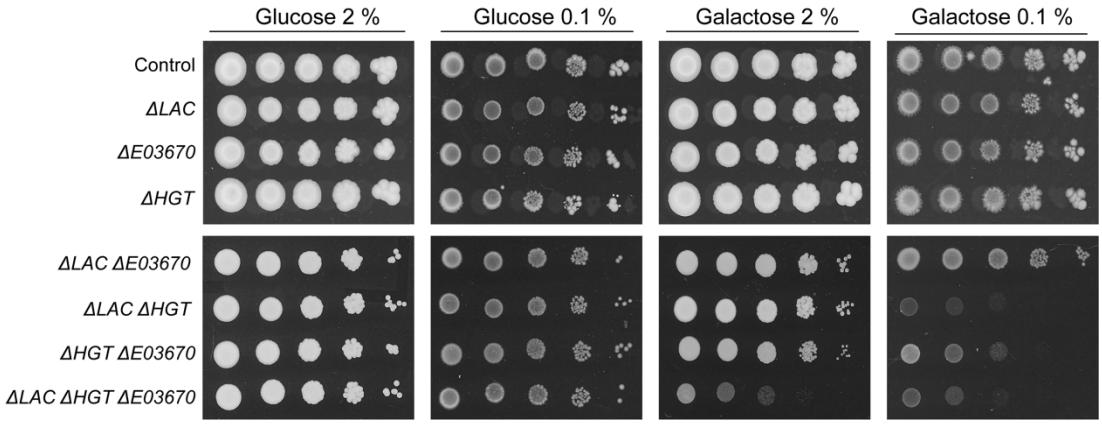


Figure 4.10: Phenotype of *K. marxianus* strains carrying mutations in galactose transport genes. The strains were grown on MM supplemented with 20 g/L to an OD_{600} of 2 then washed, diluted serially, and spotted onto MM plates supplemented with galactose or glucose, as indicated. The *ΔLAC* strain contains mutations in the *KmLAC12*, *LAC12*-2 and *LAC12*-4 genes, the *Δ*E03670 strain contains a single nucleotide deletion in the E0670 gene and *ΔHGT* carries a deletion of the whole region containing the *HGT* genes. The precise genotype of these mutants is shown in Table 4.3. The wild-type strain NBRC 1777 was used as a control in this experiment.

4.4 Discussion

Uniqueness of galactose transport in *K. marxianus*

The genetics of galactose transport has been well studied in *K. lactis* but only a few studies on this topic were previously performed with *K. marxianus*. The evidence presented in this work revealed that *K. marxianus* contains eight galactose transporters: three encoded by *LAC12* genes, four encoded by *HGT* genes and one encoded by a *KHT2* orthologue. Mutant analyses revealed that disruption of these genes results in a growth defect on galactose media, indicating that these genes are the main galactose transporters in *K. marxianus* (**Figure 4.10**). There may have been a slight amount of residual growth when the mutant lacking these transporters was cultured in 2 % galactose, possibly due to slow galactose uptake by another transporter(s) but more detailed growth and kinetics studies would be required to establish if such a transporter actually exists. A different scenario was observed in galactose 0.1 % where a mutation in the *LAC* and *HGT* genes completely abolished growth. Interestingly, this phenotype was observed in the *ΔLAC ΔHGT* mutant and not in the *ΔHGT Δ*E03670 strain, showing that at least one of the *LAC* and *HGT*-like genes encode high-affinity galactose transporters. Expression of these genes in *S. cerevisiae* showed that *LAC12-2, LAC12*-4 and A02950 conferred growth on 0.1 % galactose medium, supporting this finding and indicating that these genes are the main high-affinity galactose transporters in *K. marxianus* (**Figure 4.1**; **Figure 4.7**). These data also suggest that among these genes *LAC12*-4 is the main galactose transporter, but mutant data supports A02950 being more important. Gene expression experiments could contribute to better understanding the relevance of these genes in galactose transport. The evidence presented in this study also suggests that E03670 encodes a low-affinity transporter. Deletion of this gene in *K. marxianus* had no effect on growth on 0.1 % galactose. Also, heterologous expression of E03670 in *S. cerevisiae* did not confer growth on 0.1 % galactose, indicating the transporter is not able to uptake galactose when the sugar is present in low concentrations.

The results presented in this study indicate that *K. marxianus* has a unique genetic makeup for sugar transport, different from that in *K. lactis*. Functionally, the *HGT1* gene encodes a glucose and galactose transporter in *K. lactis* (Baruffini *et al*., 2006; Billard *et al*., 1996). However, heterologous expression experiments showed that none of the Hgt transporters from *K. marxianus* can transport glucose at detectable levels (**Figure 4.7**). Also, galactose transport by the *KHT2* homologue, E03670 is intriguing considering that the *KHT* genes in *K. lactis* are thought to encode glucose transporters (Milkowski *et al*., 2001). It is possible that *KHT2* encodes a galactose transporter in *K. lactis*, but this idea has not been tested. A previous study revealed that Lac12p and Hgt1p are the only galactose transporters in *K. lactis*, but those experiments were performed in a strain containing *RAG1*, a gene that arose by a recombination event between *KHT1* and *KHT2* (Weirich *et al*., 1997). Another aspect that seems to differ between the two species is gene regulation. The *HGT1* promoter in *K. lactis* contains several putative Gal4p activation sites (Baruffini *et al*., 2006), but none of these sites were identified in the *HGT* promoters from *K. marxianus*. This could indicate that *HGT* expression in *K. marxianus* is Gal4p-independent and may require the activity of another transcriptional factor. In contrast, two putative Gal4p binding sites were found in the E03670 promoter, indicative of galactose regulation. Altogether, the differences between hexose transport in these two species suggest that *K. marxianus* has diverged functionally from the rest of the *Kluyveromyces* species. Expansion of the *KHT* and *HGT* genes could have been essential in this process, permitting the evolution of several galactose transporters.

Importance of sugar transporter expansion

Previous studies showed that the *LAC12* gene is present in four copies in *K. marxianus* (Varela *et al*., 2017b). Our data demonstrate that the *KHT* and *HGT* genes are also present in multiple copies in this yeast (**Figure 4.3**). Also, *KHT* and *HGT* copy number was variable between strains, indicating that evolution at these loci is recent and ongoing. Studies carried out on bacterial and eukaryotic genomes established that recent gene duplications are usually involved in membrane and secreted proteins, supporting this idea (Kondrashov *et al*., 2002). Moreover, differences in copy number could be explained by the elimination of genes with redundant functions in some species (Mendonça *et al*., 2011). In fact, the *KHT* genes KMAR_50343 and KMAR_50344 (that recombined to form E03660) were found to be functionally equivalent to E03660 (**Figure 4.4**). A similar scenario could explain *HGT* copy number differences among strains. However, in this case, recombination of A02950 and A02930 led to the loss of a gene (A02940), encoding a low-affinity galactose transporter. More experiments are required to establish if the gene arisen from this recombination event (KLMA_10546 in DMKU3) has a similar function to other *HGT* genes. However, these data point out to an ongoing evolution process in both the *HGT* and *KHT* loci.

Among the *Kluyveromyces* genomes analysed, expansion in *HGT* and *KHT* was only observed in *K. marxianus*, which could indicate that these genes play a species-specific role in this yeast. *K. marxianus* is thermotolerant and displays a higher growth rate when compared to other *Kluyveromyces* species (Lane and Morrissey, 2010). Perhaps the expansion of these transporters contributed to the development of some of these functions in this yeast. It is known that stress is an important factor contributing to gene duplication and evolution of new functions (Kondrashov, 2012). Glucose limitation was shown to trigger duplication of the hexose transporters in *S. cerevisiae* (Brown *et al*., 1998). In that study, three tandem copies of a hybrid transporter formed by recombination between the *HXT6* and *HXT7* genes were found in an evolved strain. Further experiments showed that this strain outcompetes the parental strain in low glucose competition experiments. Similarly, duplication in *HGT* and *KHT* could have initially evolved in response to nutrient deficiency. *K. marxianus* is traditionally associated with fermented dairy products but it can also be found on plant sources (Suzuki *et al*., 2014; Zhong *et al*., 2016). Duplication of sugar

transporters could have allowed rapid sugar uptake and may have contributed to outcompete other microorganisms present in these environments.

4.5 References

Baruffini, E., Goffrini, P., Donnini, C., Lodi, T., 2006. Galactose transport in *Kluyveromyces lactis*: major role of the glucose permease Hgt1. doi:10.1111/j.1567-1364.2006.00107.x

Billard, P., Ménart, S., Blaisonneau, J., Bolotin-Fukuhara, M., Fukuhara, H., Wésolowski-Louvel, M., 1996. Glucose uptake in *Kluyveromyces lactis*: role of the HGT1 gene in glucose transport. J. Bacteriol. 178, 5860–5866.

Brown, C.J., Todd, K.M., Rosenzweig, R.F., 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol. Biol. Evol. 15, 931–942. doi:10.1093/oxfordjournals.molbev.a026009

Carvalho-Silva, M., Spencer-Martins, I., 1990. Modes of lactose uptake in the yeast species *Kluyveromyces marxianus*. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 57, 77–81. doi:10.1007/BF00403158

Coloretti, F., Chiavari, C., Luise, D., Tofalo, R., Fasoli, G., Suzzi, G., Grazia, L., 2017. Detection and identification of yeasts in natural whey starter for Parmigiano Reggiano cheese-making. Int. Dairy J. 66, 13–17. doi:10.1016/J.IDAIRYJ.2016.10.013

Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS One 5, e11147. doi:10.1371/journal.pone.0011147

De Bruijne, A.W., Schuddemat, J., Van den Broek, P.J.A., Van Steveninck, J., 1988. Regulation of sugar transport systems of *Kluyveromyces marxianus*: the role of carbohydrates and their catabolism. BBA - Biomembr. 939, 569–576. doi:10.1016/0005-2736(88)90104-6

De Vuyst, L., Weckx, S., 2016. The cocoa bean fermentation process: from ecosystem analysis to starter culture development. J. Appl. Microbiol. 121, 5–17. doi:10.1111/jam.13045

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. doi:10.1093/nar/gkh340

Escalante-Minakata, P., Blaschek, H.P., Barba de la Rosa, A.P., Santos, L., De León-Rodríguez, A., 2008. Identification of yeast and bacteria involved in the mezcal fermentation of *Agave salmiana*. Lett. Appl. Microbiol. 46, 626–630. doi:10.1111/j.1472-765X.2008.02359.x

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, D222–D230. doi:10.1093/nar/gkt1223

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37. doi:10.1093/nar/gkr367

Fonseca, G.G., Gombert, A.K., Heinzle, E., Wittmann, C., 2007. Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. FEMS Yeast Res. 7, 422–435. doi:10.1111/j.1567-1364.2006.00192.x

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. doi:10.1007/s00253-008-1458-6

Gasnier, B., 1987. Characterization of low- and high-affinity glucose transports in the yeast *Kluyveromyces marxianus*. BBA - Biomembr. 903, 425–433. doi:10.1016/0005-2736(87)90049-6

Gethins, L., Rea, M.C., Stanton, C., Ross, R.P., Kilcawley, K., O'Sullivan, M., Crotty, S., Morrissey, J.P., 2016. Acquisition of the yeast *Kluyveromyces marxianus* from unpasteurised milk by a kefir grain enhances kefir quality. FEMS Microbiol. Lett. 363, fnw165. doi:10.1093/femsle/fnw165

Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31–34. doi:10.1038/nprot.2007.13

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, 5351–5358.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. 59, 307–321. doi:10.1093/sysbio/syq010

Ho, V.T.T., Zhao, J., Fleet, G., 2014. Yeasts are essential for cocoa bean fermentation. Int. J. Food Microbiol. 174, 72–87. doi:10.1016/j.ijfoodmicro.2013.12.014

Huerta-Cepas, J., Serra, F., Bork, P., 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. 33, 1635–8. doi:10.1093/molbev/msw046

Inokuma, K., Ishii, J., Hara, K.Y., Mochizuki, M., Hasunuma, T., Kondo, A., 2015. Complete Genome Sequence of *Kluyveromyces marxianus* NBRC1777, a Nonconventional Thermotolerant Yeast. Genome Announc. 3, e00389-15. doi:10.1128/genomeA.00389-15

Jeong, H., Lee, D.-H., Kim, S.H., Kim, H.-J., Lee, K., Song, J.Y., Kim, B.K., Sung, B.H., Park, J.C., Sohn, J.H., Koo, H.M., Kim, J.F., 2012. Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var. marxianus KCTC 17555. Eukaryot. Cell 11, 1584–5. doi:10.1128/EC.00260-12

Juergens, H., Varela, J.A., de Vries, A.R.G., Perli, T., Gast, V.J.M., Gyurchev, N.Y., Rajkumar, A.S., Mans, R., Pronk, J.T., Morrissey, J.P., Daran, J.-M.G., 2018. Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. FEMS Yeast Res. doi:10.1093/femsyr/foy012

Kondrashov, F.A., 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc. R. Soc. B Biol. Sci. 279, 5048–5057. doi:10.1098/rspb.2012.1108

Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I., Koonin, E. V, 2002. Selection in the evolution of gene duplications. Genome Biol. 3, RESEARCH0008.

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*. Antonie Van Leeuwenhoek 100, 507–19. doi:10.1007/s10482-011-9606-x

Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. Fungal Biol. Rev. 24, 17–26. doi:10.1016/j.fbr.2010.01.001

Lappe-Oliveras, P., Moreno-Terrazas, R., Arrizán-Gaviño, J., Herrera-Suárez, T., García-Mendoza, A., Gschaedler-Mathis, A., 2008. Yeasts associated with the production of Mexican alcoholic nondistilled and distilled *Agave* beverages. FEMS Yeast Res. 8, 1037–1052. doi:10.1111/j.1567-1364.2008.00430.x

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Suprayogi, Tsuchikane, K., Limtong, S., Fujita, N., Yamada, M., 2015. Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. Biotechnol. Biofuels 8, 47. doi:10.1186/s13068-015-0227-x

Löbs, A.-K., Engel, R., Schwartz, C., Flores, A., Wheeldon, I., 2017. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol. Biofuels 10, 164. doi:10.1186/s13068-017-0854-5

Mendonça, A.G., Alves, R.J., Pereira-Leal, J.B., 2011. Loss of Genetic Redundancy in Reductive Genome Evolution. PLoS Comput. Biol. 7, e1001082. doi:10.1371/journal.pcbi.1001082

Milkowski, C., Krampe, S., Weirich, J., 2001. Feedback regulation of glucose transporter gene transcription in *Kluyveromyces lactis* by glucose uptake. J. Bacteriol. 183, 5223–5229. doi:10.1128/JB.183.18.5223

Morrissey, J.P., Etschmann, M.M.W., Schrader, J., Billerbeck, G.M., 2015. Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. Yeast 32, 3–16. doi:10.1002/YEA.3054

Nambu-Nishida, Y., Nishida, K., Hasunuma, T., Kondo, A., 2017. Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. Sci. Rep. 7, 8993. doi:10.1038/s41598-017-08356-5

Ng, H., Dean, N., 2017. Dramatic Improvement of CRISPR/Cas9 Editing in *Candida albicans* by Increased Single Guide RNA Expression. mSphere 2, e00385-16. doi:10.1128/mSphere.00385-16

Ortiz-Merino, R.A., Varela, J.A., Coughlan, A.Y., Hoshida, H., da Silveira, W.B., Wilde, C., Kuijpers, N.G.A., Geertman, J.-M., Wolfe, K.H., Morrissey, J.P., 2018. Ploidy Variation in *Kluyveromyces marxianus* Separates Dairy and Non-dairy Isolates. Front. Genet. 9, 94. doi:10.3389/fgene.2018.00094

Riley, M.I., Sreekrishna, K., Bhairi, S., Dickson, R.C., 1987. Isolation and characterization of mutants of *Kluyveromyces lactis* defective in lactose transport. MGG Mol. Gen. Genet. 208, 145–151. doi:10.1007/BF00330435

Schwan, R.F., Wheals, A.E., 2004. The Microbiology of Cocoa Fermentation and its Role in Chocolate Quality. Crit. Rev. Food Sci. Nutr. 44, 205–221. doi:10.1080/10408690490464104

Suzuki, T., Hoshino, T., Matsushika, A., 2014. Draft Genome Sequence of *Kluyveromyces marxianus* Strain DMB1, Isolated from Sugarcane Bagasse Hydrolysate. Genome Announc. 2, e00733-14-e00733-14. doi:10.1128/genomeA.00733-14

Tittarelli, F., Varela, J.A., Gethins, L., Stanton, C., Ross, R.P., Suzzi, G., Grazia, L., Tofalo, R., Morrissey, J.P., 2018. Development and implementation of multilocus sequence typing to study the diversity of the yeast *Kluyveromyces marxianus* in Italian cheeses. Microb. Genomics. doi:10.1099/mgen.0.000153

Varela, J.A., Gethins, L., Stanton, C., Ross, P., Morrissey, J.P., 2017a. Applications of *Kluyveromyces marxianus* in Biotechnology, in: Yeast Diversity in Human Welfare. Springer Singapore, Singapore, pp. 439–453. doi:10.1007/978-981-10-2621-8_17

Varela, J.A., Montini, N., Scully, D., Van der Ploeg, R., Oreb, M., Boles, E., Hirota, J., Akada, R., Hoshida, H., Morrissey, J.P., 2017b. Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains. FEMS Yeast Res. 17. doi:10.1093/femsyr/fox021

Vyas, V.K., Barrasa, M.I., Fink, G.R., 2015. A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. Sci. Adv. 1, e1500248–e1500248. doi:10.1126/sciadv.1500248

Weirich, J., Goffrini, P., Kuger, P., Ferrero, I., Breunig, K.D., 1997. Influence of mutations in hexose-transporter genes on glucose repression in *Kluyveromyces lactis*. Eur. J. Biochem. 249, 248–57. doi:10.1111/j.1432-1033.1997.t01-1-00248.x

Wésolowski-Louvel, M., Goffrini, P., 1992. Glucose transport in the yeast *Kluyveromyces lactis*. Mol. Gen. 101, 89–96. doi:10.1007/BF00587565

Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C.P., Boles, E., 1999. Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. FEBS Lett. 464, 123–8.

Xie, S., Shen, B., Zhang, C., Huang, X., Zhang, Y., 2014. sgRNAcas9: A Software Package for Designing CRISPR sgRNA and Evaluating Potential Off-Target Cleavage Sites. PLoS One 9, e100448. doi:10.1371/journal.pone.0100448

Zhong, S., Liu, H., Zhang, H., Han, T., Jia, H., Xie, Y., Zhong, S., Liu, H., Zhang, H., Han, T., Jia, H., Xie, Y., 2016. Effects of *Kluyveromyces marxianus* Isolated from Tibetan Mushrooms on the Plasma Lipids, Egg Cholesterol Level, Egg Quality and Intestinal Health of Laying Hens. Rev. Bras. Ciência Avícola 18, 261–268. doi:10.1590/1806-9061-2015-0070

Chapter 5

Ploidy variation in *Kluyveromyces marxianus* separates dairy and non-dairy isolates

This chapter was published in its totality in the following paper:

Ploidy variation in *Kluyveromyces marxianus* separates dairy and non-dairy isolates. Ortiz-Merino R., **Varela J.A.,** Coughlan A., Hoshida H., da Silveira W., Wilde C., Kuijpers N., Geertman JM., Wolfe K., Morrissey JP. *Frontiers in genetics*. 2018. May 21;9:94. DOI: 10.3389/fgene.2018.00094

Ortiz-Merino R performed read mapping of *K. marxianus* genome data and SNP analysis in this study.

Ortiz-Merino R and Varela J.A drafted the manuscript.

Abstract

Kluyveromyces marxianus is traditionally associated with fermented dairy products but can also be isolated from diverse non-dairy environments. Because of thermotolerance, rapid growth and other traits, many different strains are being developed for food and industrial applications but there is, as yet, little understanding of the genetic diversity or population genetics of this species. *K. marxianus* shows a high level of phenotypic variation but the only phenotype that has been clearly linked to a genetic polymorphism is lactose utilisation, which is controlled by variation in the *LAC12* gene. The genomes of several strains have been sequenced in recent years and, in this study, we sequenced a further 9 strains from different origins. Analysis of the Single Nucleotide Polymorphisms (SNPs) in 14 strains was carried out to examine genome structure and genetic diversity. SNP diversity in *K. marxianus* is relatively high, with up to 3% DNA sequence divergence between alleles. It was found that the isolates include haploid, diploid and triploid strains, as shown by both SNP analysis and flow cytometry. Diploids and triploids contain long genomic tracts showing loss of heterozygosity. All 6 isolates from dairy environments were diploid or triploid, whereas 6 out 7 isolates from non-dairy environment were haploid. This also correlated with the presence of functional *LAC12* alleles only in dairy haplotypes. The diploids were hybrids between a non-dairy and a dairy haplotype, whereas triploids included 3 copies of a dairy haplotype.

5.1 Introduction

The yeast *Kluyveromyces marxianus* is best-known because of its frequent association with traditional dairy products such as kefir and cheese (Lachance, 2011; Gethins *et al*., 2016; Coloretti *et al*., 2017). This association with fermented dairy beverages, a consequence of its capacity to use the milk sugar lactose as a carbon source, has led to inclusion of *K. marxianus* on GRAS (FDA) and QPS (EU) lists of safe micro-organism for use in foods (Lane and Morrissey, 2010; EFSA-BIOHAZ *et al*., 2017). The yeast is also regularly isolated from non-dairy environments (e.g. decaying fruit) and is part of the natural flora involved in production of *Agave*-based alcoholic beverages such as tequila and mezcal (Lappe-Oliveras *et al*., 2008; Verdugo Valdez *et al*., 2011). In the latter case, the production of enzymes that degrade plant fructans to simpler sugars (inulinases) undoubtedly contributes to its growth in this environment (Arrizon *et al*., 2012). The capacity of *K. marxianus* to utilise a broad array of sugars also creates potential for biotechnological applications (Fonseca *et al*., 2008; Lane and Morrissey, 2010), which is illustrated by the many studies exploring potential for bioethanol production from diverse substrates such as whey permeate, crop plants and lignocellulosic biomass (Nonklang *et al*., 2009; Guimaraes *et al*., 2010; Wu *et al*., 2016; Kobayashi *et al*., 2017). This yeast is used commercially for production of the flavour molecule 2-phenylethanol, and there is considerable interest in development of *K. marxianus* as a cell factory for production of other bioflavours (Morrissey *et al*., 2015). *K. marxianus* is also distinguished by thermotolerance (Lane *et al*., 2011), and the fastest reported growth rate of any eukaryote (Groeneveld *et al*., 2009). Recent years have seen increasing interest in new application such as production of biomolecules (Hughes *et al*., 2017; Lin *et al*., 2017), biocatalysis (Oliveira *et al*., 2017; Wang *et al*., 2017) and heterologous protein production (Gombert *et al*., 2016; Lee *et al*., 2017).

One of the interesting aspects of the nascent development of *K. marxianus* as an important yeast for biotechnology is the wide variety of strains that are being used, both for research and for application. This contrasts with the traditional yeast, *S. cerevisiae,* where, until recently, there was a very strong focus on a relatively narrow set of model strains. While giving access to the broad diversity that exists within any species, the non-reliance on model strains also creates challenges since findings with one isolate are not automatically transferrable to other isolates. This is illustrated

well by studies that demonstrate wide variance in tolerance to different external stresses (Lane *et al*., 2011; Rocha *et al*., 2011). Indeed, it has emerged that even a trait such as lactose utilisation, long considered one of the defining characteristics of *K. marxianus*, is not universal, and many strains exhibit very poor growth on lactose, a phenotype that was shown to be due to polymorphisms in the *LAC12* gene, which encodes a permease responsible for transport of lactose into the cell (Varela *et al*., 2017). Although recent studies on sugar transport and physiology are starting to address the deficit (Fonseca *et al*., 2013; Signori *et al*., 2014; Beniwal *et al*., 2017; Dias *et al*., 2017; Diniz *et al*., 2017), it is true to say that a lot of the underlying knowledge about the biology of *K. marxianus* is based on inference of similarity with its sister species, *Kluyveromyces lactis*, which was developed as a model for studying lactose-positive yeasts since the 1960′s (Fukuhara, 2006). The genome of *K. lactis* was sequenced more than a decade ago (Souciet *et al*., 2000), with a more recent functional reannotation and genome-scale model that provides a deeper understanding of the core metabolism of this species (Dias *et al*., 2012; Dias *et al*., 2014). Notwithstanding the utility of a related species for comparison, the many metabolic and physiological differences between *K. lactis* and *K. marxianus* necessitate independent studies of *K. marxianus* to provide the comprehensive understanding of its genetics and physiology that will underpin future developments in fundamental biology and biotechnology.

Genomic and transcriptomic studies have started to shed light on *K. marxianus* and a growing number of genome sequences of *K. marxianus* strains are now available (Jeong *et al*., 2012; Silveira *et al*., 2014; Inokuma *et al*., 2015; Lertwattanasakul *et al*., 2015; Quarella *et al*., 2016). As yet, however, there has not been a systematic comparison of the sequenced *K. marxianus* genomes, nor a comparison to the single *K. lactis* genome that is in the public domain. In contrast to *K. lactis*, whose genome comprises 6 chromosomes, several studies have reported that *K. marxianus* has a full complement of 8 chromosomes, with many areas of local synteny between the species. There is strong conservation of the mating type locus (Lane *et al*., 2011) and thus *K. marxianus* could be expected to be capable of mating type switching and mating in a manner similar to *K. lactis* (Barsoum *et al*., 2010; Rajaei *et al*., 2014). Based on information to date, however, there does appear to be a fundamental

difference in life-cycles. Studies of natural isolates of *K. lactis* suggest that this yeast is primarily a haploid (haplontic) species. Mating is induced by depletion of nitrogen or phosphate in the environment, and zygotes formed by mating usually sporulate immediately (although diploids can be maintained in the lab, e.g. by selection for auxotrophic markers) (Schaffrath and Breunig, 2000; Zonneveld and Steensma, 2003; Booth *et al*., 2010; Rodicio and Heinisch, 2013). In contrast, analysis of the mating-type locus of natural and culture collection *K. marxianus* isolates identified both haploid and diploid strains (Lane *et al*., 2011; Fasoli *et al*., 2016).

To put the phenotypic diversity of *K. marxianus* into context, it is important to characterise its genomic diversity and to assess the population structure of the species. There have been some pre-whole genome sequence studies that addressed this question using different methods. Pulsed-field gel electrophoresis studies suggested that there were variable numbers of chromosomes in *K. marxianus* strains (Belloch *et al*., 1998; Fasoli *et al*., 2015), a finding not in accordance with genome sequence data, which has consistently indicated 8 chromosomes. Mitochondrial DNA haplotypes and variation at some genomic loci was used to try to determine population structure in a collection from Italian cheeses (Fasoli *et al*., 2016). That particular study identified variations in population structure and proposed the occurrence of homozygous and heterozygous strains. A Multi Locus Sequence Typing (MLST) method was developed to further explore the diversity in that collection and in this case, the analysis was extended to other strains that are sequenced or available in culture collections (Tittarelli, 2018). MLST analysis did not identify distinct sub-populations but while the method was very diagnostic for strain identification, the surprisingly high level of heterozygosity in diploid strains reduced resolution to a level too low for population-type analysis. Analysis of population structure in diploid yeasts is challenging and, in many cases, has relied on SNPs identified in genome sequences derived from haploids or from completely homozygous diploids (made by self-mating of single spore derivatives) (Liti *et al*., 2009; Schacherer *et al*., 2009; Strope *et al*., 2015). In highly heterozygous species, this method may not generate an accurate view of the relationships among haplotypes or among strains.

In this study, we set out to explore the genomic diversity of *K. marxianus* by analysing whole-genome data from 14 strains isolated from different sources. Some

of these strains had been previously sequenced and published, whereas others were sequenced for this study. To take heterozygosity into account, raw sequence reads were used to allow analysis of single nucleotide polymorphisms between strains. The results indicate a high degree of variation among isolates, in both ploidy and heterozygosity, and show a correlation between ploidy and environmental niche. Our work raises important questions about the life cycle of *K. marxianus* and emphasizes the need to take ploidy and heterozygosity into account when considering using *K. marxianus* for biotechnological purposes.

5.2 Materials and methods

Yeast strains, growth and phenotypic analysis

The 14 *K. marxianus* strains analysed in this study are listed in **Table 5.1**. Two strains (DMKU3-1042 and UFS-Y2791) were not available for phenotypic assessment but the remaining 12 strains were obtained from the sources indicated in **Table 5.1** and were routinely cultured at 30 °C in YPD medium (10 g/L yeast extract, 20 g/L bactopeptone, 20 g/L glucose). For lactose utilisation tests, yeast strains were first grown overnight in 5 mL minimal media (MM) supplemented with 2% glucose (Fonseca *et al*., 2007). Cells from the overnight cultures were harvested by centrifugation, washed twice with 5 mL of water and used to inoculate MM supplemented with 2% lactose to an OD_{600} of 0.1. These cultures were incubated for 15 hours, when the final OD_{600} was determined. Lactose concentration was determined by HPLC at 0 and 15h and used to calculate lactose consumption as previously described (Varela *et al*., 2017). Experiments were performed in triplicate with error bars showing standard deviation.

Flow cytometry

DNA content was determined by flow cytometry using SYTOX green (Thermo-Fisher) as previously described (Haase and Reed, 2002). Yeast strains were grown in YPD at 30° C with 200 rpm agitation in a New Brunswick Innova 40/40R orbital shaker (Eppendorf, Hamburg, Germany). Cultures were harvested by centrifugation and resuspended in 1 mL sterile water. Cells were then washed, resuspended in 400 µL sterile water and fixed by adding 950 µL 100% ethanol. The suspensions were incubated overnight at 4° C, then centrifuged and washed in 50 mM sodium citrate (pH 7.2). The cells were resuspended in 500 µL RNAse A solution (0.25 mg/mL RNAse A, 50 mM sodium citrate pH 7.2) and incubated for 1 h at 37° C. Then, 100 µL of 20 mg/mL Proteinase K was added to each sample and the tubes were incubated at 50° C for 2 h. Finally, 500 µL of SYTOX Green solution (4 µM SYTOX Green, 50 mM sodium citrate pH 7.2) was added to each tube. Samples were analysed using a BD FACSCelesta system (BD Biosciences, CA, USA) and the data was processed using FlowJo software v10 (BD Biosciences, CA, USA).

Table 5.1: Sources of *K. marxianus* strains and genomes analysed in this study

Strain	Synonym	Country	Sample source	Strain source	Reference for genome sequence	Source of Illumina FASTQ data	Accession numbers for Illumina data
L01		Unknown	Dairy	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRR6450386
L02		Unknown	Dairy	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRR6450387
L03		Unknown	Dairy	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRR6450384
L04		Unknown	Baking	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRR6450385
L05		Unknown	Distillery	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRR6450382
CBS397*		Netherlands	Yoghurt	Westerdijk Institute, Netherlands	This study	University College Cork (J.P. Morrissey)	SRR6450383
NBRC0272		Unknown	Miso	Biological Resource Center, NITE (NBRC), Japan	This study	Yamaguchi University (H. Hoshida)	SRR6450380
NBRC0288	DSM4906	Unknown	Unknown	Biological Resource Center, NITE (NBRC), Japan	This study	Yamaguchi University (H. Hoshida)	SRR6450381
NBRC0617	ATCC8622	Denmark	Yoghurt	Biological Resource Center, NITE (NBRC), Japan	This study	Yamaguchi University	SRR6450388

						(H. Hoshida)	
Strain	**Synonym**	**Country**	**Sample source**	**Strain source**	**Reference for genome sequence**	**Source of Illumina FASTQ data**	**Accession numbers for Illumina data**
NBRC1777		Japan	Soil	Biological Resource Center, NITE (NBRC), Japan	Inokuma *et al*. (2015)**	Yamaguchi University (H. Hoshida)	SRR6450389
CBS6556	KCTC17555, ATCC26548	Mexico	Pozol	Westerdijk Institute, Netherlands	Jeong *et al*. (2012)	Yonsei University (J. F. Kim)	SRR6660862
UFV-3	CCT7735	Brazil	Dairy	Universidade Federal de Viçosa, Brazil	Silveira *et al*. (2014)	BIOAGRO, Brazil (F. M. L. Passos)	SRR6660864
DMKU3-1042†		Thailand	Soil	Strain not obtained	Lertwattanasakul *et al*. (2015)	Yamaguchi University (H. Hoshida)	SRR6450379
UFS-Y2791		South Africa	*Agave americana* juice	Strain not obtained	Schabort *et al*. (2016)	Univ. of the Free State (Du Toit. W. P. Schabort)	SRR6660863

* CBS397 is the type strain of *Kluyveromyces fragilis.*

**The reference genome sequence of NBRC1777 (Inokuma *et al*., 2015) was based on an assembly of Pacific Biosciences and Ion Torrent data, but Dr. A. Kondo generously provided with Illumina FASTQ data files from the same strain.

† Genome sequence obtained from a *ura3* mutant generated by UV mutagenesis.

Genome data, sequencing, and read mapping

The genomes of five strains (NBRC1777, CBS6556, UFV-3, DMKU3-1042 and UFS-Y2791) had previously been published and some of the authors kindly made the source Illumina FASTQ data available for this analysis. The 11 strains including accession numbers in **Table 5.1** were sequenced in this study. All strains were sequenced on Illumina HiSeq 2000 or 2500 instruments after Truseq genomic library preparation. Details of the sequencing data type and coverage for all 14 strains, including those sequenced elsewhere, are summarised in **Table 5.2.** We performed quality control checks for all libraries with FastQC v. 0.10.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads for strain CBS6556 were trimmed using skewer v. 0.2.2 (Jiang *et al*., 2014) with parameters -q 20 -m pe -l 70. For BLAST analyses, *de novo* assemblies of each newly sequenced genome were made using SPAdes 3.5.0 (Bankevich *et al*., 2012).

The NBRC1777 genome sequence was selected as a reference because of the high quality of its assembly into 8 chromosomes, and because our analysis confirmed that the haploid nature of this strain (Inokuma *et al*., 2015). Sequencing libraries from all strains were aligned to the NBRC1777 reference using the Burrows-Wheeler Aligner (BWA) v. 0.7.9a-r786 (Li and Durbin, 2009) with default parameters. The BWA "mem" alignment algorithm was used for libraries with read length ≥ 100 bp, the "aln" alignment algorithm for libraries with read length < 100 bp, and the alignment modes were set to "samse" and "sampe" for single-end and paired-end data respectively. Samtools v. 0.1.19-44428cd (Li *et al*., 2009) was used to remove unmapped reads from the BWA output files and to generate indexes for downstream steps. Picard tools v. 2.0.1 (http://broadinstitute.github.io/picard) function AddOrReplaceReadGroups was used to add identifiers to the BAM files, followed by MarkDuplicates to mark and discard PCR duplicates. Indel realignment and coverage calculation were performed using the RealignerTargetCreator, IndelRealigner, and DepthOfCoverage tools from the Genome Analysis Tool Kit (GATK) v. 3.5-0-g36282e4 (Van der Auwera *et al*., 2013). Mean coverage was calculated omitting a 19 kb region on chromosome 5 that contains the array encoding the rRNA genes. Coverage plots were obtained by calculating the average in 10-kb windows. Segment means were calculated using the R Bioconductor package DNAcopy v 1.50.1 (DOI: 10.18129/B9.bioc.DNAcopy).

Table 5.2: Summary of Illumina sequencing strategies and coverage for 14 strains used in SNP analysis.

Strain	Mean coverage (x)	Pairedness *	Read length	Million reads	BWA algorithm
L01	47	SE	50	16.6	aln samse
L02	47	SE	50	16.7	aln samse
L03	45	SE	50	16.7	aln samse
L04	45	SE	50	15.6	aln samse
L05	47	SE	50	16.2	aln samse
CBS397	141	PE	126	15.1	mem
NBRC0272	160	PE	100	20.4	mem
NBRC0288	212	PE	100	25.6	mem
NBRC0617	35	SE	50	11.6	aln samse
NBRC1777	110	PE	100	12.8	mem
CBS6556	623	PE	70-150	56.9	mem
UFV-3	359	PE	90	50.3	mem
DMKU3-1042	341	PE	100	44.6	mem
UFS-Y2791	50	PE	75-100	17.1	mem

* SE, single-end. PE, paired-end.

Variant calling

'Variable sites' were defined as the set of sites in the genome that contain a non-reference base, hereafter called 'variants', in at least one of the 14 strains. Variant calling in the 14 strains was done using the GATK tool HaplotypeCaller in DISCOVERY and GVCF modes, requiring a minimum quality score of 20. The output files were then used for multi-sample analysis using the GenotypeGVCF tool and a custom Perl script was used to remove all variants that had low genotype quality (GQ < 20), or had low approximate read depth (below 10% of the mean coverage for the sample excluding the rDNA locus and telomeric regions). For every remaining variable site, the output from GATK enabled us to calculate the empirical allele frequencies of the reference base (designated *fA*) and the variant (designated *fB* and referred as alternative allele frequencies), which sum to 1. Empirical allele frequencies were calculated at each variable site on each strain by dividing the allelic depth of the variant by the approximate read depth observed at that site (AD and DP fields in GATK's HaplotypeCaller). Each variant remaining after the filtering steps and having an $fB > 0.15$ is considered to be Single Nucleotide Polymorphisms (SNPs). Accordingly, variable sites were only used for further analysis when having a SNP. If $fA \geq 0.85$ the strain was called homozygous for the reference base (*AA*) at this variable site, and if $fB \geq 0.85$ it was called homozygous for the alternative base (*BB*). Sites with intermediate allele frequencies ($0.15 < fA < 0.85$) were called heterozygous (*AB*). Using these thresholds, a small number of SNPs were called in some strains later shown to be haploid. These apparent SNPs clustered in sub-telomeric regions known to contain repetitive DNA and are likely to be technical artefacts due to misaligning of reads to different repeats or to copy number variants. Note that these calls were only made for variable sites; the genomes also contain a much larger number of invariant sites that are considered identical among all strains and are therefore *AA*.

Nucleotide diversity (π) and average SNP density were calculated with VariScan v 2.0.3 (Hutter *et al*., 2006) using a non-overlapping window size of 1 kb along all chromosomes. In the case of heterozygous variants, only the variant with highest *fB* was used. Sites considered for the analysis of the 14 strains were required to show variation in a minimum number of 4 strains. SnpEff v 4.3s (Cingolani *et al*., 2012)

was used to produce summary statistics and annotate the SNPs using the public NBRC1777 genome annotation as a reference (Inokuma *et al*., 2015).

Phylogenetic analysis

Because the data consisted of a mixture of strains with different ploidies, we developed a custom method for phylogenetic analysis of haplotypes. This method is based on a window approach similar to our previous development for an interspecies hybrid (Schroder *et al*., 2016). Homozygosity and heterozygosity were first assessed in 1 kb windows of the genome of each diploid strain. For each 1 kb window in each strain, the total numbers of variable sites that were called with each genotype (#*AA*, #*BB* and #*AB*) in the window were calculated. The whole window was then classified as either heterozygous for the two haplotypes if #*AB* ≥ 3, or homozygous otherwise. Homozygous windows were then classified as either homozygous for the alternative haplotype if #*BB* ≥ 9, or homozygous for the reference haplotype otherwise. The cut-off values of 3 and 9 were chosen based on analysis of the distributions of window frequencies, using strain L01 as a test case (**Fig. 5.1**). Each 1 kb window of the genome was only used if it was heterozygous in all 5 diploid strains, and the regions with aberrant allele frequencies in NBRC0272 were excluded (chromosome 6). Concatenated nucleotide sequences of the shared heterozygous windows were extracted from the NBRC1777 genome and used to infer the sequences of alleles in each strain, depending on its ploidy.

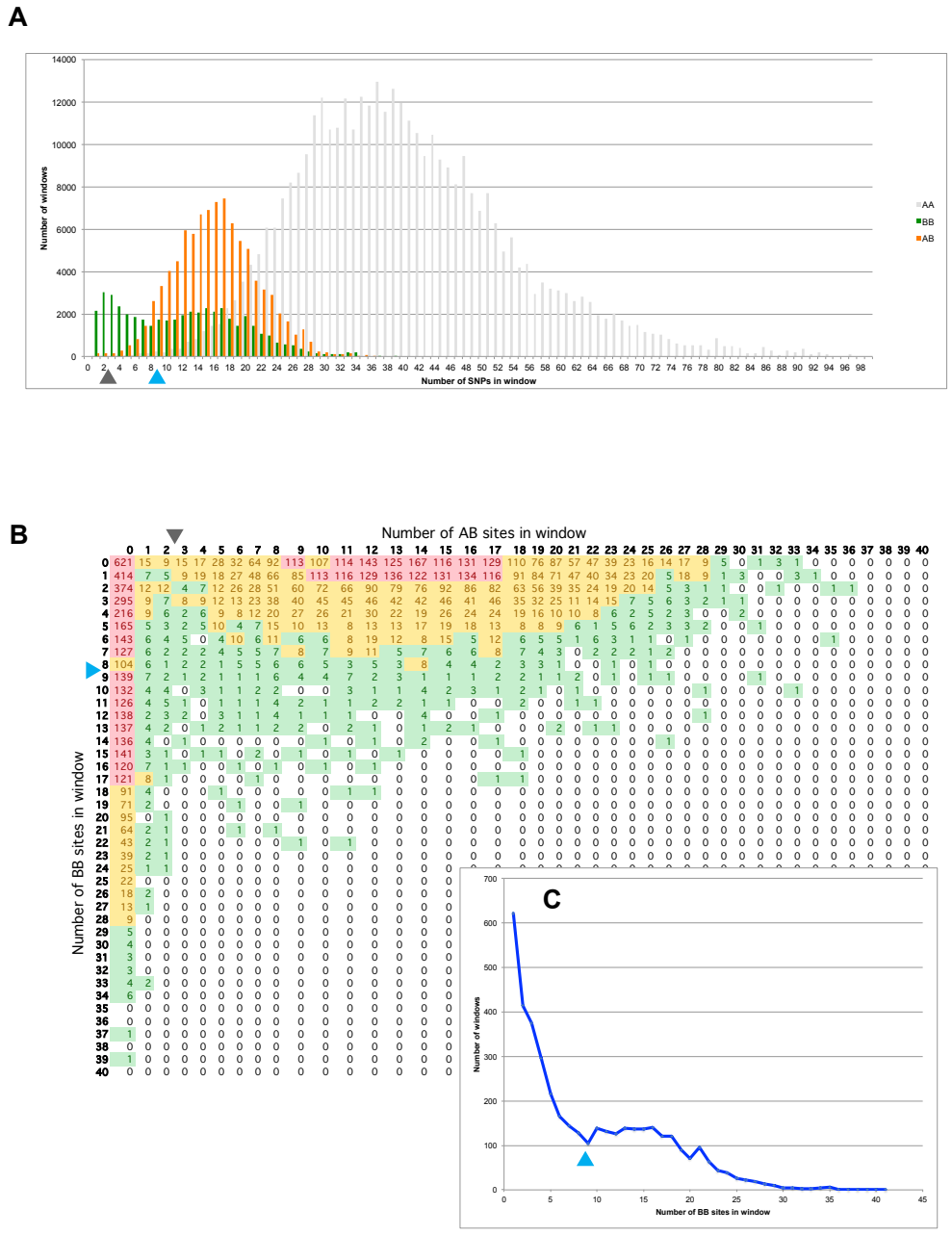


Figure 5.1: Method of classifying 1-kb genomic windows into haplotypes. The cutoff values chosen are marked by triangles in all panels (windows with ≥ 3 *AB* SNPs were classified as *AB* heterozygous windows; windows with ≥ 9 *BB* SNPs were classified as *BB* heterozygous windows; other windows were classified as *AA* homozygous windows). The data plotted are for the diploid strain L01. **(A)** Histogram of numbers of SNPs of each type, in all 1-kb windows in the genome. **(B)** Heatmap showing the distribution of numbers of *AB* and *BB* SNPs per window. Cells in the matrix show numbers of windows. The cutoff values were chosen to coincide with minima on the two axes. **(C)** Distribution of numbers of *BB* SNPs in windows that have zero *AB* SNPs.

For haploid strains, the variant was used to replace the reference base at the corresponding position of the variable site in the concatenated sequence. For diploid strains, two different putative A and B alleles were first generated and used as a template for base replacement depending on the type of variable site. For homozygous alternative (*BB*) sites, the variant was used for replacement in both A and B alleles. For heterozygous (*AB*) variable sites, the variant was only used for replacement in the B allele. For triploid strains, putative alleles 1, 2, and 3 were first generated and used as a template for base replacement depending on the types of variable sites and their allele frequencies. For homozygous alternative (*BB*) sites, the variant was used for replacement in all three putative alleles. For heterozygous (*AB*) sites, the variant was used for replacement in both alleles 2 and 3 if $fB > 0.6$, or for replacement only in allele 3 if $fB < 0.4$. The phylogenetic tree of the inferred and concatenated haplotype sequences was generated using PhyML v. 3.1 (Guindon *et al*., 2010) selecting for the best of NNI and SPR methods, using 5 random starts, and with empirical estimation of base frequencies and proportions of invariable sites (parameters: --search BEST --rand_start --n_rand_starts 5 -f e –v e). The tree was visualized using FigTree v. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Data Availability

The Illumina sequences generated and analysed for this study can be found in the NCBI Sequence Read Archive under the accession number SRP128575, strain-specific accessions are provided in **Table 5.1**.

5.3 Results

K. marxianus displays a high level of genomic variation

The genome sequences of 14 strains of *K. marxianus* were analysed for Single Nucleotide Polymorphisms (SNPs) to determine the extent of variability in this species. The strains selected for analysis included the 5 strains with published whole genome sequences (at the time of the study) and 9 other strains from different collections (**Table 5.1**). The 5 sequenced strains (NBRC1777, CBS6556, UFV3, DMKU3-1042 and UFS-Y2791) have been phenotypically analysed to different extents by a number of research teams and are of interest for biotechnological applications (Nonklang *et al*., 2008; Fonseca *et al*., 2013; Costa *et al*., 2014; Schabort du *et al*., 2016; Nambu-Nishida *et al*., 2017). This is also the case for the previously unsequenced strains CBS397, NBRC0272, NBRC0288 and NBRC0617, which were obtained from national culture collections (Lane *et al*., 2011; Foukis *et al*., 2012; Yarimizu *et al*., 2013). The additional 5 strains (L01 to L05) are from the in-house culture collection of the Lallemand company and there are no published data available for these strains. The original source of isolation is known for 13/14 strains and is almost evenly divided between dairy and non-dairy environments. The genome sequences of the 9 previously un-sequenced strains were obtained as described in methods and thus all 14 genomes could be analysed and compared. It should be noted that although all 14 strains were sequenced using Illumina technology, this was performed by different laboratories using a diversity of sequencing strategies and thus the depth of coverage is quite variable (**Table 5.2**).

We used the genome sequence of strain NBRC1777, which was assembled into 8 complete chromosomes using Pacific Biosciences technology, as the reference for SNP analysis (Inokuma *et al*., 2015). GATK software was used to identify sequence variants present in the Illumina reads from all 14 strains relative to this reference. Variable sites were defined as the set of sites in the genome that contain a non-reference base in at least one of the 14 strains. For each variable site in each strain, the empirical allele frequency in the Illumina reads from that strain was calculated (see Methods). Depending on frequency, the variant was classified as homozygous SNP ($fB \geq 0.85$), or heterozygous SNP ($0.15 < fB < 0.85$). Variants appearing at frequencies below 0.15 were assumed to be due to sequencing errors and were

ignored. Only 249 SNPs were identified in the Illumina data from the reference NBRC1777, confirming that the reference sequence is accurate, and this strain is haploid. All other strains contained more than 30,000 SNPs relative to the reference (**Table 5.3**). The highest number of SNPs is in strain UFS-Y2791 (Schabort du *et al*., 2016) and corresponds to 3.0% nucleotide sequence divergence in the 10.9 Mb genome. From the numbers of heterozygous and homozygous (non-reference) SNPs, the other strains fall into three groups: one group with low numbers of heterozygous SNPs (< 2000), one group with more heterozygous than homozygous SNPs, and one group with fewer heterozygous than homozygous SNPs. Below, we show that these three groups are haploid, diploid, and triploid strains, respectively. In total, 667,472 variable sites, comprising 597,466 SNPs, and 70,006 indels were found. Indels were not analysed further, and, after filtering (see methods) a subset of 571,339 SNPs was retained for analysis. SNP diversity in *K. marxianus* is relatively high, with average pairwise difference between strains (π) of 12×10^{-3}. The average density of SNPs in *K. marxianus* is 7.6 SNPs/kb in coding regions and 11.1 SNPs/kb in intergenic regions (**Table 5.4**). Of the 359,354 variants located within the coding regions of genes, 71.6% were predicted to be silent, 28.1% missense, and 0.2% (745) nonsense mutations, when compared against the NBRC1777 annotation (Inokuma *et al*., 2015).

Table 5.3: SNPs identified in 14 *K. marxianus strains.*

Strain	**Heterozygous** SNPs *	**Homozygous** SNPs †	**Total** SNPs
Haploid strains			
NBRC1777 (reference strain)	248	1	249
L05	611	33326	33937
L04	726	35138	35864
CBS6556	1439	40922	42361
DMKU3-1042	1647	39648	41295
UFS-Y2791	714	325190	325904
Diploid strains			
L02	115648	42561	158209
L01	96202	49325	145527
CBS397	110132	52415	162547
NBRC0288	96304	60554	156858
NBRC0272	147241	40085	187326
Triploid strains			
NBRC0617	29010	145016	174026
L03	26388	145856	172244
UFV-3	27347	177365	204712

* Number of sites at which a variant (non-reference base) was present, at a frequency in the reads between 0.15 and 0.85

† Number of sites at which a variant (non-reference base) was present, at a frequency ≥ 0.85 in the reads.

Table 5.4: SNP density

					Coding		Intergenic	
Strain	**pi**	**theta**	**SNPs**	**SNPs/kb**	**SNPs**	**SNPs/kb**	**SNPs**	**SNPs/kb**
CBS397	0.009	0.009	101645	9.368	68065	6.835	33580	10.15870019
CBS6556	0.004	0.004	41329	3.809	28748	2.887	12581	3.806033566
DMKU3-1042	0.004	0.004	40078	3.694	22607	2.270	17471	5.285367811
L01	0.008	0.008	87996	8.110	55584	5.582	32412	9.805354101
L02	0.008	0.008	88900	8.194	63530	6.380	25370	7.674991779
L03	0.014	0.014	154446	14.235	106919	10.737	47527	14.37797928
L04	0.003	0.003	35359	3.259	22444	2.254	12915	3.907076028
L05	0.003	0.003	33521	3.089	20501	2.059	13020	3.938840874
NBRC0272	0.010	0.010	105198	9.696	58213	5.846	46985	14.21401217
NBRC0288	0.010	0.010	108751	10.023	72786	7.309	35965	10.88021598
NBRC0617	0.014	0.014	156518	14.426	109982	11.045	46536	14.07817964
UFS-Y2791	0.030	0.030	325510	30.001	227782	22.874	97728	29.56490329
UFV-3	0.017	0.017	189331	17.450	131438	13.199	57893	17.51392586
Average				**10.412**		**7.637**		**11.16966004**

K. marxianus isolates show different ploidy states

The distribution of allele frequencies for each strain was assessed using a graphical method similar to that used in a recent study in *S. cerevisiae* (Zhu *et al*., 2016). Histograms (**Fig. 5.2A**), of the distribution of allele frequencies at all the variable sites in the genome created three sets of strains that corresponded to those identified on the basis of the numbers of heterozygous and homozygous SNPs (**Table 5.3**). The five strains with the highest numbers of heterozygous SNPs in (L02, L01, CBS397, NBRC0288 and NBRC0272) all show a symmetrical peak of allele frequencies centred on 0.5, suggesting that they are diploid. The three strains with high numbers of both homozygous and heterozygous SNPs (NBRC0617, L03 and UFV-3) all show bimodal distributions, with peaks at 0.33 and 0.66 for the frequency of the variant, suggesting that they are triploid. In contrast, in the 6 strains UFS-Y2791, DMKU3-1042, L05, L04, CBS6556 and NBRC1777, only a low number of sites were designated as heterozygous, and these sites show little pattern in their frequency distributions. These data are most consistent with a haploid genome.

To provide an independent measurement of ploidy, DNA content in each strain was measured by flow cytometry (**Fig. 5.2C**). Each analysed strain shows a bimodal distribution of DNA content, corresponding to the G1 and G2 phases of the cell cycle. For the haploid strains, the DNA content is 1*n* (in G1 phase) and 2*n* (in G2 phase), where *n* is the DNA content of the haploid genome. For the diploids, it is 2*n* and 4*n*, and for the triploids it is 3*n* and 6*n*. The patterns observed were consistent with the designation based on allele frequencies and confirmed that this set of strains was comprised of 6 haploid, 5 diploid and 3 triploid strains.

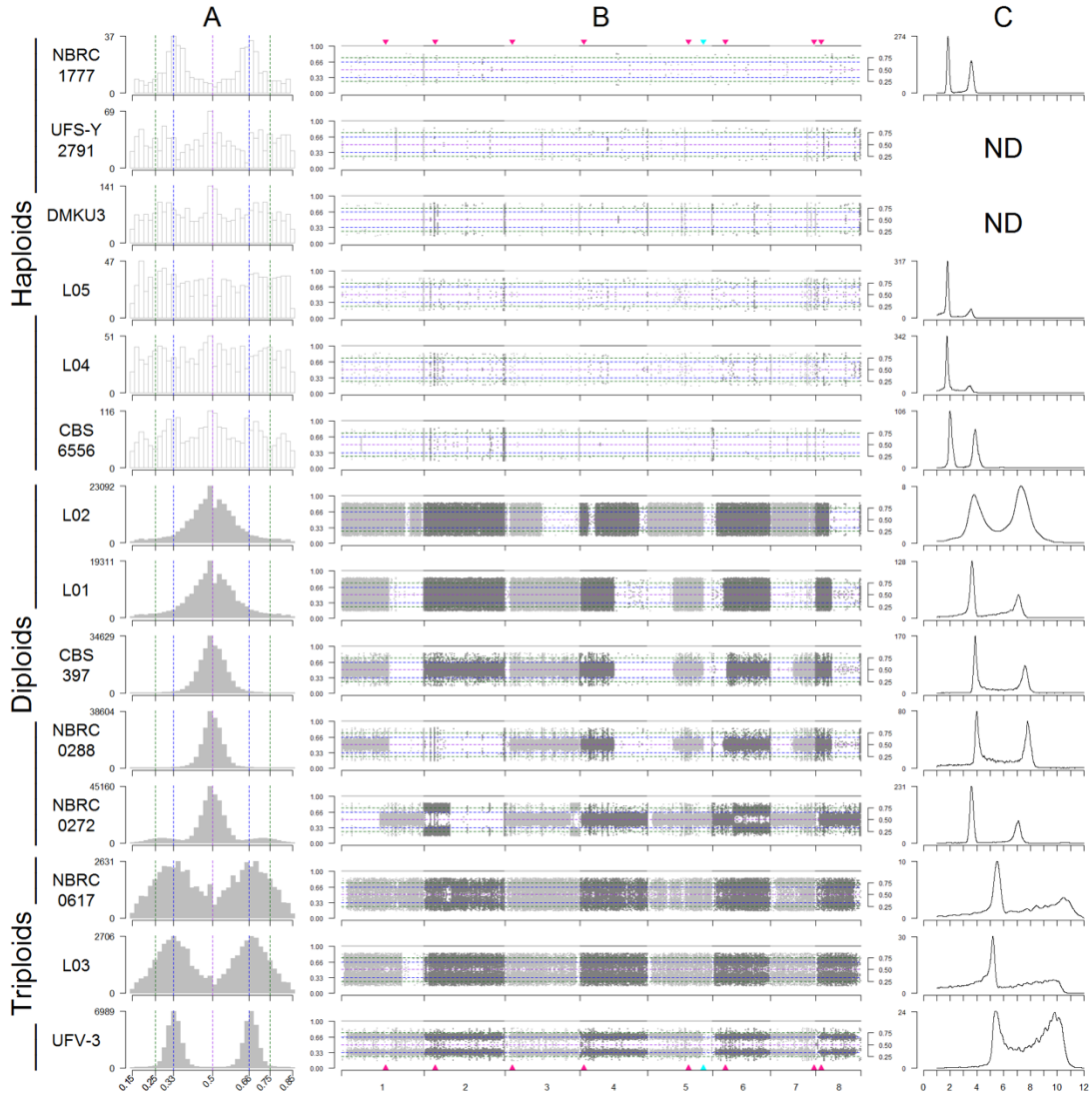


Figure 5.2: Variable ploidy in *Kluyveromyces marxianus* strains. Strain names are shown on the left. **(A)** Histograms of the alternative allele frequencies of variant (non-reference) bases, for SNPs designated as heterozygous (sites with alternative allele frequencies *fB* between 0.15 and 0.85). Histograms are coloured grey if at least 10% of the SNPs in a strain are heterozygous. Dashed vertical lines mark frequencies of 0.5 (purple), 0.33/0.66 (blue) and 0.25/0.75 (green). Bin sizes are 2% intervals. **(B)** Plots of alternative allele frequencies along the 8 chromosomes, for each strain. Horizontal dashed lines mark frequencies as in panel A. Light and dark grey points indicate SNPs on different chromosomes. Red triangles mark the locations of centromeres, and the blue triangle marks the ribosomal DNA locus. Allele frequencies ≥ 0.85 are shown as 1. Alternative allele frequencies ≤ 0.15 are not shown. **(C)** Flow cytometry of DNA content. The Y-axis shows numbers of cells, and the X-axis shows SYTOX Green fluorescence signal intensity (arbitrary units) which is proportional to DNA content. Flow cytometry was not carried out for UFS-Y2791 and DMKU3-1042.

Common patterns of loss of heterozygosity in diploid strains

For each SNP, its allele frequency versus its chromosomal location in the reference assembly was plotted to determine whether heterozygosity was uniformly distributed (**Fig. 5.2B**). For haploid strains, the low level of variation does not show any particular pattern, whereas in all 5 diploid strains, large regions of the genome with loss of heterozygosity (LOH) are apparent. In heterozygous regions of the genome, allele frequency should be distributed about 0.5 but there are regions where no variability is seen. These predominantly white areas in the plots indicate stretches of chromosome that are homozygous in that strain (**Fig. 5.2B**). For example, strain L02 is heterozygous through most of its genome but shows homozygosity on the right half of chromosome 3 and over most of chromosome 8. In the diploids, most chromosomes are heterozygous over at least some of their length, but chromosome 2 in NBRC0288 and chromosome 7 in L01 are essentially completely homozygous. Three diploid strains L01, CBS397 and NBRC0288 exhibit almost identical extents of partial LOH on chromosomes 1, 3 (left end), 4, 5, 6, and 8 (**Fig. 5.2B**), but different patterns on chromosomes 2 and 7. On chromosome 6, the region of LOH is slightly larger in CBS397 than in L01 and NBRC0288 (it crosses the centromere only in CBS397). Strain L02 shares two LOH boundaries with this group of three strains, on chromosome 5 (the boundary occurs at the rDNA locus) and chromosome 3 (left end). Only one region of LOH is visible in our triploids, on chromosome 1 in strain L03. This LOH region occupies approximately 4% of the L03 genome, whereas in the diploid strains the LOH regions total 25-51% of the genome by length.

Copy number variation and partial aneuploidy in multiple strains

Read coverage in 10 kb windows was determined across the genome of each strain to investigate whether any of the strains displayed aneuploidy (**Fig. 5.3**). In this analysis, a value of zero indicates no variation between expected and actual numbers of reads, whereas higher or lower numbers could indicate DNA duplications or deletions. Although the SNP and flow cytometry results (**Fig. 5.2**) indicated that there are three groups of strains with genomes that are primarily haploid, diploid and triploid, there is also evidence in multiple strains of partial aneuploidy, segmental duplications, or deletions that alter the copy number of some parts of the genome.

These possible aneuploidies do not correspond to the regions of LOH described in **Fig. 5.2**, indicating that these are distinct phenomena.

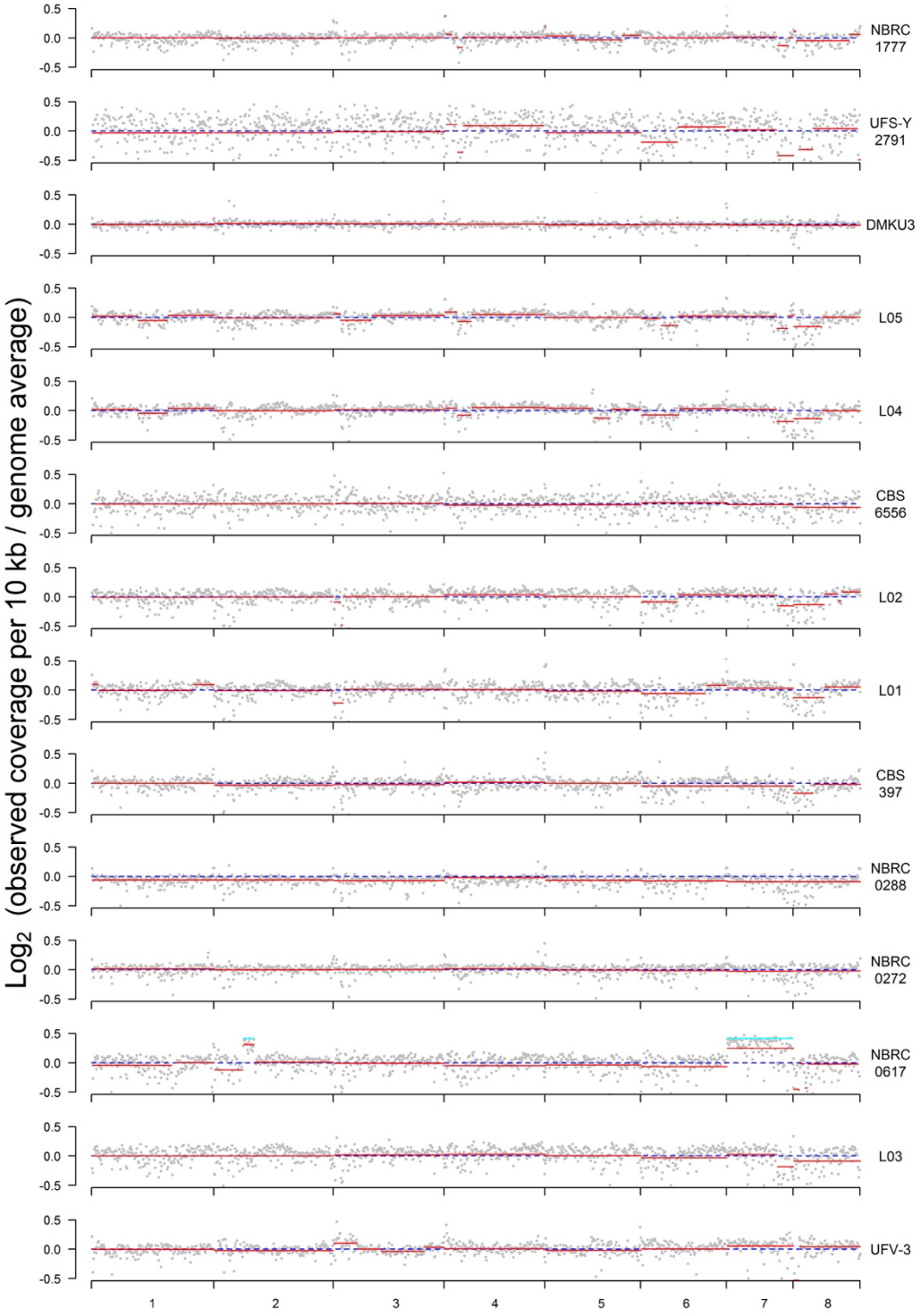


Figure 5.3: Plots of sequence coverage in each strain. The Y-axis is $\log_2$ of the ratio between the observed and expected coverage, for 10-kb windows through the genome; a value of zero (dashed blue line) indicates no difference. Expected coverage is based on the average in the whole genome. Red lines show the segmental means for consecutive 10-kb windows calculated using the Bioconductor package DNAcopy. Cyan lines for NBRC0617 indicate the value expected for the 1.33-fold increase in coverage that would result from a fourth copy of a region in a triploid.

The clearest case of aneuploidy is in strain NBRC0617, which has an extra copy of chromosome 7 (**Fig. 5.3**). The analysis of ratios shows that reads from chromosome 7 are present at 1.18x the expected frequency ($\log_2$ value 0.249) in this strain. Since NBRC0617 is primarily triploid, a fourth copy of a chromosome should increase the read coverage on that chromosome by approximately 4/3 = 1.33-fold (cyan line in **Fig. 5.3**) relative to the genome average, and many of the 10-kb windows in chromosome 7 are not significantly different from this value (**Fig. 5.4A**). Furthermore, the distribution of allele frequency values for SNPs on chromosome 7 of NBRC0617 shows a peak at 0.25 (**Fig. 5.4B**), which is consistent with the presence of four copies of the chromosome, and contrasts with the peaks at 0.33 and 0.66 that are seen when the whole genome of NBRC0617 is considered (**Fig. 5.2A**). NBRC0617 also shows increased copy number of a circa 150 kb segment of chromosome 2 (coordinates 420 – 570 kb) (**Fig. 5.3**). Within this segment, allele frequencies of 0.25 and 0.75 are visible (**Fig. 5.4B**), indicating that it is present in 4 copies. With the current data, we are unable to determine the precise structure of the chromosomal rearrangement that increased the segment′s copy number. The region immediately to the left of it (0–420 kb) may be present at a reduced copy number.

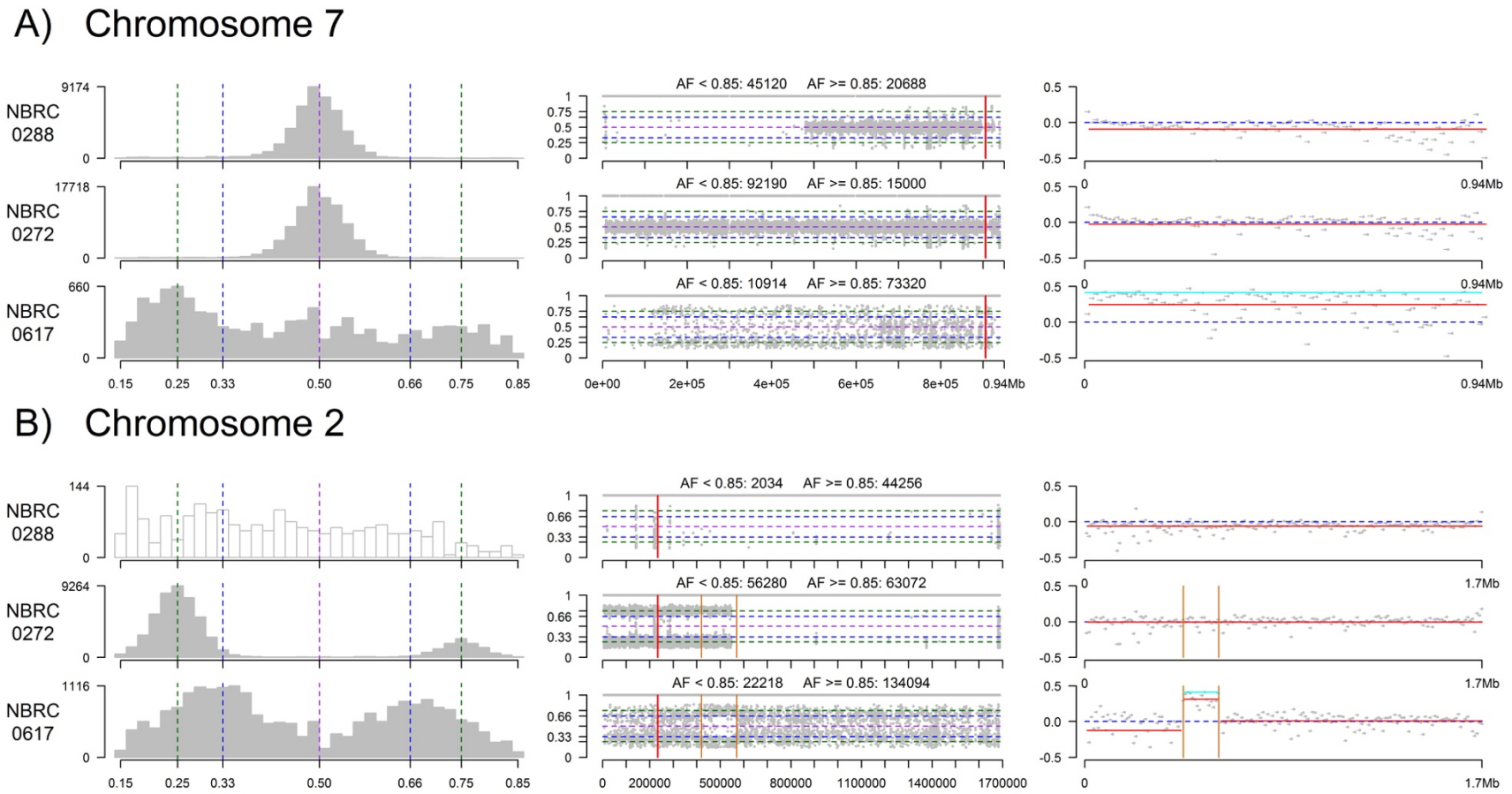


Figure 5.4: Allele frequencies and sequence coverage in three strains on **(A)** chromosome 7 and **(B)** chromosome 2. The *left* and *middle panels* show distributions of allele frequency on each chromosome, as in Figure 5.2. The centromere is marked by a vertical red line. The *right panels* show $\log_2$ ratios between observed and expected sequence coverage, as in Figure 5.3. Cyan lines for NBRC0617 indicate the value expected for the 1.33-fold increase in coverage that would result from a fourth copy of a region in a triploid. The vertical orange lines in B mark the 150 kb region from with increased coverage in NBRC0617 (coordinates 420 to 570 kb).

Some intriguing possible examples of copy number variation are seen in strain NBRC0272, which was designated as a diploid based on its overall allele frequency and flow cytometry patterns. Examination of allele frequencies indicates that there are three genomic regions (in chromosomes 2, 3 and 6) present at higher copy numbers (**Fig. 5.2B**). This can be seen in more detail in the allele frequency plots of those chromosomes for this strain (**Figure 5.5)**. Chromosomes 2 (left end) and 3 (right end) contain SNPs with 0.25/0.75 allele frequencies, indicating the presence of four copies. Chromosome 6 (right end) contains SNPs with allele frequency peaks close to 0.33/0.66, indicating three copies, and contrasting with the peaks at 0.5 on the left part of this chromosome. Since the flow cytometry (**Fig. 5.2C**) shows that the total DNA content of NBRC0272 is close to diploid, it is concluded that the extra copies of these three chromosomal regions must be the result of segmental duplications and not extra copies of the whole chromosome. Puzzlingly, however, these putative segmental duplications were not apparent in the coverage plot of NBRC0272 (**Fig. 5.3**).

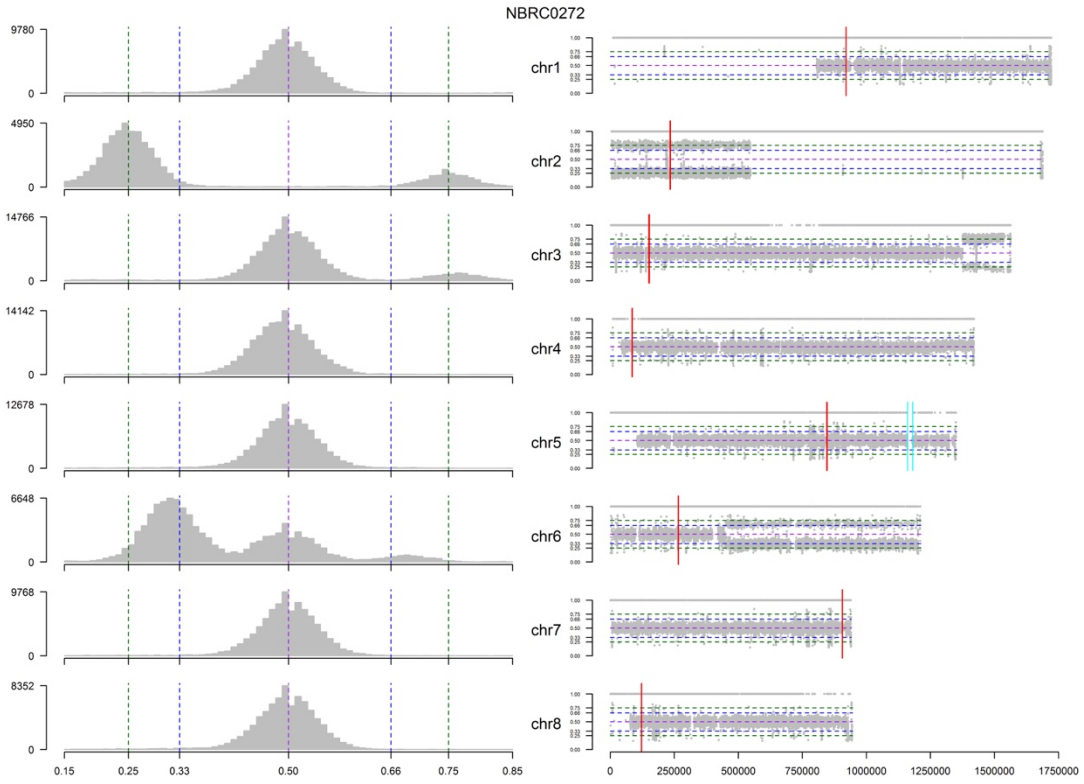


Figure 5.5: Allele frequencies on each chromosome of NBRC0272. Details are as in Figure 5.2A and B. In the plots on the right, red vertical lines mark centromeres and blue lines mark the rDNA array.

Phylogenetic analysis separates dairy from non-dairy haplotypes

The presence of strains with different ploidy in the dataset presents a problem for phylogenetic analysis. In studies on diploid eukaryotes such as mammals, the standard approaches for constructing phylogenies of individuals from SNP data either exclude all heterozygous sites, or randomly choose one of the alleles at these sites (Lischer *et al*., 2014). In our preliminary analyses of the *K. marxianus* data, it was noticed that in the diploid strains, one allele was often very similar to the NBRC1777 reference sequence, but the other allele was considerably different. We were therefore motivated to construct a phylogenetic tree of the *K. marxianus* strains that kept the alleles separate - namely, a tree of haplotypes rather than a tree of strains.

To make a tree of haplotypes, we used a method previously developed to investigate the pathogenic yeast *Candida orthopsilosis*, which is an interspecies hybrid (Schroder *et al*., 2016). In each of the 5 diploid *K. marxianus* strains, each region of the genome was classified as either heterozygous, homozygous for the 'A' haplotype (the haplotype more similar to the NBRC1777 reference), or homozygous for the 'B' haplotype (the haplotype less similar to the reference) (**Fig. 5.6**; see Methods for details). Approximately 18% of the genome was heterozygous in all 5 diploid strains, and we then extracted the sequences of the 'A' and 'B' haplotypes from these regions in each diploid. The aberrant (trisomic or tetrasomic) regions of the NBRC0272 genome were excluded from this dataset. A similar process was used to estimate the sequences of the three haplotypes present in each of the three triploid strains in these regions (see Methods). In the phylogenetic tree of haplotypes then generated, each haploid strain appears once, each diploid strain appears twice, and each triploid strain appears three times (**Fig. 5.7**). Three clades are evident, but Clade 3 contains only the haploid strain UFS-Y2791. Despite the high divergence of UFS-Y2791 from the other strains, a phylogenetic tree using *K. lactis* and *Lachancea thermotolerans* as outgroups confirms that it is indeed a strain of *K. marxianus* (**Fig. 5.7**, inset). All other *K. marxianus* haplotypes lie in Clades 1 and 2. Clade 1 contains all the haploid strains except the outlier UFS-Y2791, and the 'A' haplotypes of each of the five diploid strains. Clade 2 contains the 'B' haplotypes of the diploid strains, and all three haplotypes of the triploid strains.

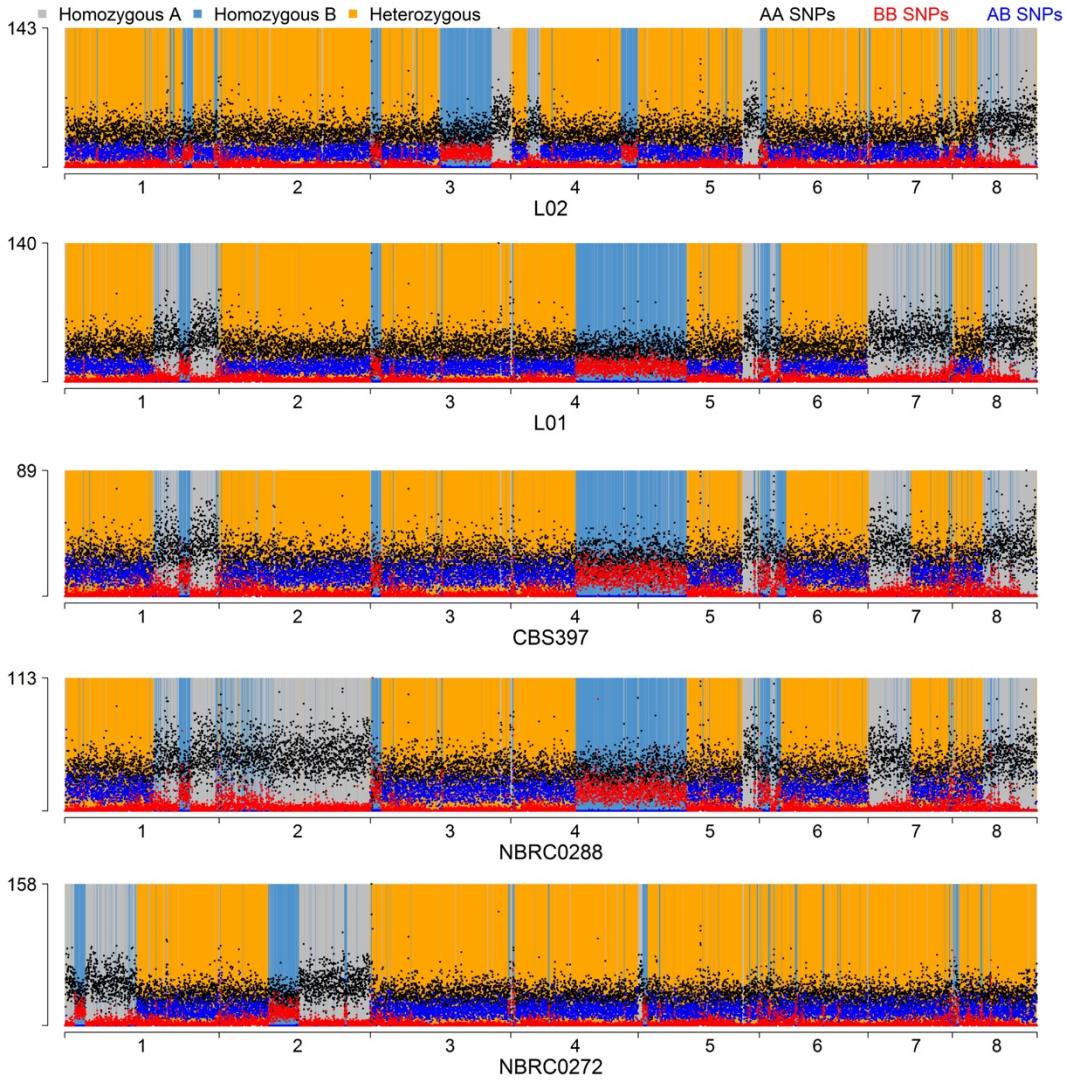


Figure 5.6: Haplotype assignment in diploid strains. For each 1-kb window through the genome, dots show the number of variable sites in the window with genotypes *AA* (black dots), *BB* (red dots), or *AB* (blue dots). Each window was then classified as either heterozygous (yellow background), homozygous for the *B* haplotype (blue background), or homozygous for the reference *A* haplotype (grey background) following rules as described in Methods and Figure 5.1. The green bars along the bottom axis show the regions (18%) that were heterozygous in all diploids and used for the haplotype phylogeny analysis.

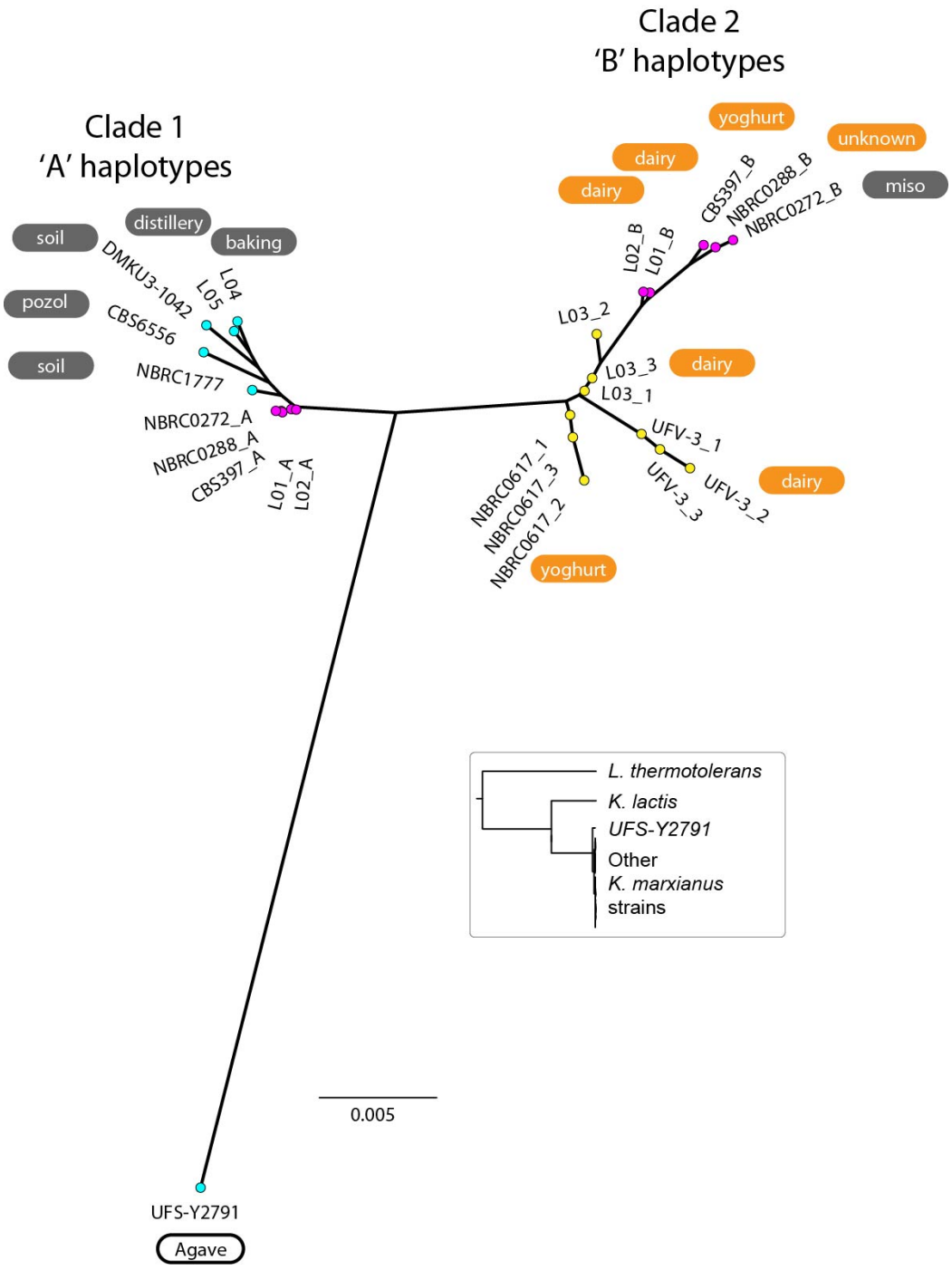


Figure 5.7: Phylogenetic tree of *K. marxianus* haplotypes. Coloured dots indicate the ploidy of strains as haploid (cyan; 1 haplotype per strain), diploid (magenta; 2 haplotypes per strain) or triploid (yellow; 3 haplotypes per strain). The environmental source of each strain is shown (for diploid strains, the source is only labelled on the 'B' haplotype). Orange backgrounds indicate Lac+ strains, grey backgrounds indicate Lac- strains, and white backgrounds indicate Lac phenotypes not tested. The tree was constructed from SNP data from regions that are heterozygous in diploids, totalling 18% of the genome. Inset, phylogenetic tree constructed from DNA sequences of *GCN1* (an arbitrarily chosen large gene), confirming that UFS-Y2791 is a strain of *K. marxianus*.

Lactose consumption phenotypes and *LAC12* genotypes

When the environments from which the strains were isolated are considered, an unexpected relationship is apparent between environment, ploidy and clade (**Fig. 5.7**). The 6 strains from 'dairy' environments (and strain NBRC0288, from an unknown source) are all either diploid or triploid whereas, with the exception of strain NBRC0272 (isolated from miso), all the strains from 'non-dairy' environments are haploid (**Table 5.1**). Since the use of lactose as a sugar source is considered important for growth of dairy yeasts, the capacity of the strains to grow on lactose was assessed to determine whether a similar pattern would emerge (**Fig. 5.8**). Indeed, none of the tested haploid strains used lactose, whereas all the diploid and triploid strains were Lac+, again with the exception of NBRC0272, which is diploid but Lac-. Although DMKU3-1042 was not available for this study, our previous work has demonstrated that it is Lac- (Varela *et al*., 2017). We also previously established that the variable ability of *K. marxianus* strains to consume lactose is explained by polymorphism of a single gene, the lactose transporter *LAC12*, and that functional and non-functional (in terms of lactose transport) alleles of this gene differ by 13 key amino acid substitutions (Varela *et al*., 2017).

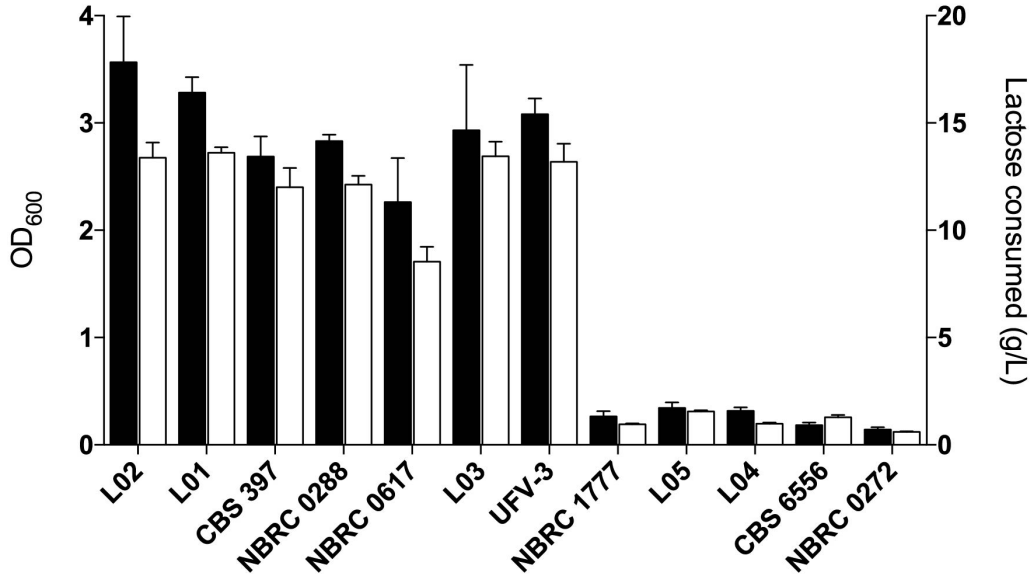


Figure 5.8: Growth and lactose consumption by *K. marxianus* strains. Growth in MM + 2 % lactose and lactose consumption are shown in black and white bars, respectively. Lactose consumption was calculated by subtracting the final and initial lactose concentrations in the medium.

BLAST searches against *de novo* assemblies of the genomes showed that the key polymorphisms in all the Lac+ strains in **Fig. 5.8** match were an exact match to the functional *LAC12*+ allele, except for NBRC0617 which matched in 11 positions. Similarly, all the Lac- strains exactly matched the non-functional *LAC12* allele, except for NBRC0272, which diverged at a single amino acid (**Fig. 5.9**). None of the diploid strains is a *LAC12*+/- heterozygote, due to LOH at the left end of chromosome 3 where *LAC12* is located (the gene is only 15 kb from the telomere). Thus, although all five diploids are AB heterozygous for most of chromosome 3, the four Lac+ diploids are 'BB' homozygous and the Lac- diploid strain NBRC0272 is 'AA' homozygous in this region (**Fig. 5.4**). In addition to the polymorphisms associated with a non-functional allele, the *LAC12* gene in NBRC0272 contains an internal stop codon (**Fig. 5.9**).

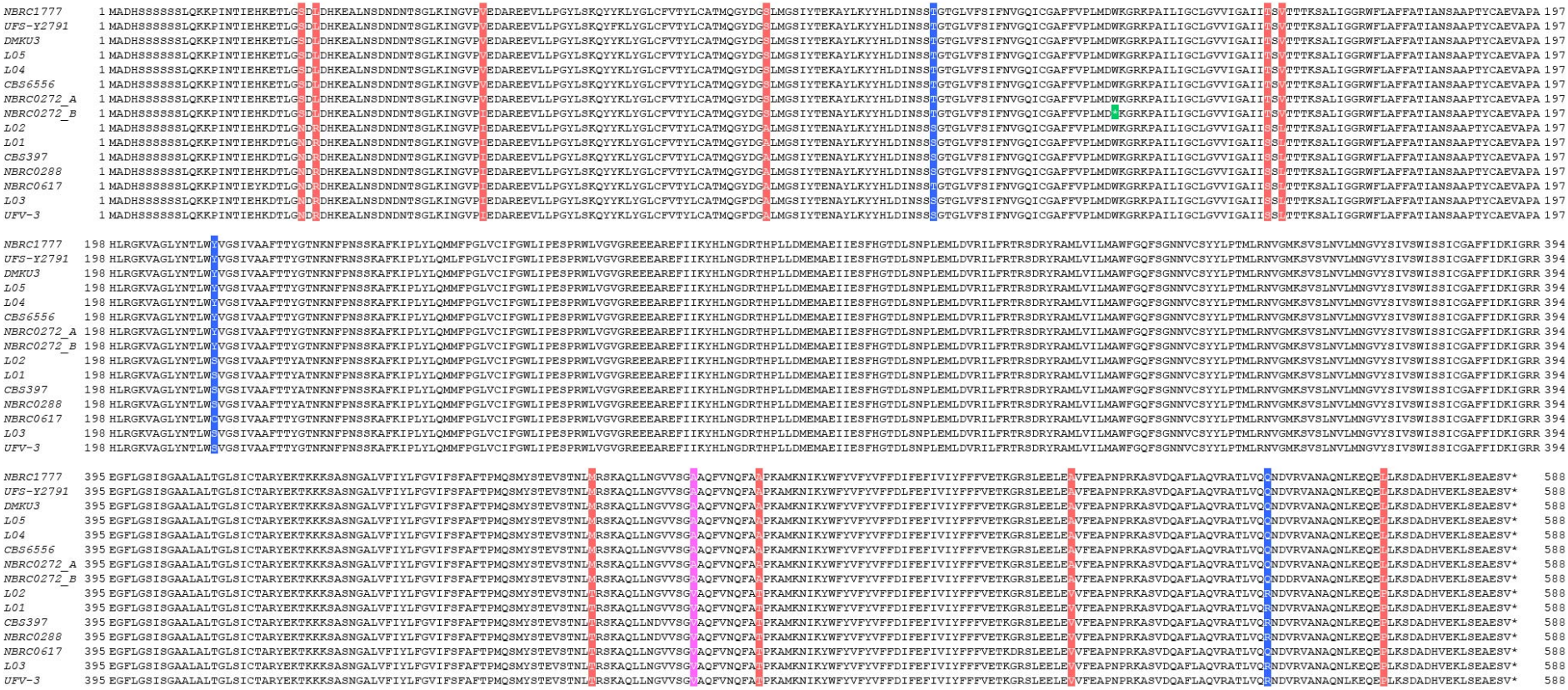


Figure 5.9: Multiple alignment of Lac12p sequences from the strains used in this study. Amino acid sequences shown were derived from the haplotypes of each strain. Haplotypes from diploid and triploid strains are denoted by letters and numbers, respectively. The residues marked in colour are those associated with either functional or non-functional alleles. The positions marked in blue were previously part of the set that distinguished alleles, but these are not conserved in all haplotype B Lac12p. Those in red are those that still distinguish based on functionality and an additional differentiating AA at position 475 (pink shading) that was overlooked in the previous study is also marked. Thus, based on sequences currently available, there are 11 differentiating amino acids. The stop codon in the NBRC0272_B sequence at position 139 is indicated with a green asterisk.

5.4 Discussion

Ploidy in *K. marxianus* distinguishes dairy and non-dairy strains

Although this study relied on a relatively small set of strains (14), it delivered some remarkable insights into the life-cycle of *K. marxianus*. Our previous report that natural isolates of this yeast can be either haploid or diploid (Lane *et al*., 2011) was confirmed and then extended by the discovery of three triploid strains. The sample size is too small to draw statistical conclusions, but it can be said that haploids, diploids, and triploids were present at roughly equal frequencies (43%, 36% and 21%, respectively) in this set. This appears to contrast with *K. lactis*, which is considered to be haploid, though it must be borne in mind that there are as yet no published population level studies with that yeast. Variable ploidy is not uncommon in yeasts; for example, among 144 mainly clinical isolates of *S. cerevisiae*, the basal ploidy levels (*i.e.* ignoring aneuploid chromosomes) were haploid (11%), diploid (57%), triploid (16%), and tetraploid (16%) (Zhu *et al*., 2016).

The most striking finding was that all the isolates from a dairy environment were either diploid or triploid, whereas non-dairy isolates were haploid. Furthermore, it was possible to distinguish two genomic haplotypes, described here as "A" and "B" that mapped 13/14 strains into distinct clades (Clades 1 and 2). The 14th strain, UFS-Y2791 may represent a third clade. All the dairy isolates contained at least one of the B haplotype genomes, suggesting that this is a dairy-niche associated genome. In the case of the three triploid strains, the B genome was represented three times, whereas the diploid strains contained one A haplotype genome and one B haplotype genome. The sequence divergence (approximately 2%) between the 'A' and 'B' haplotypes indicates that the diploid dairy strains were probably formed by mating between haploid representatives of Clades 1 and 2, as opposed to any other mechanism of ploidy change. Nevertheless, we were unable to examine *MAT* locus genotypes because the sequence assemblies are too fragmented in the *MAT/HML/HMR* regions. The phylogeny of the haplotypes indicates that the diploid dairy strains were formed by at least two independent matings between parents from the A and B clades, and that the A parents in these matings were very closely related to each other (much more so than the B parents). Triploids may have arisen by self-mating of B-haplotype (clade 2) strains, with one scenario being mating of a BB diploid

with a B haploid to form a triploid; other routes to a triploid are also possible. It is implicit in these scenarios that B-haplotype haploid strains should also exist, though none were found in the current study. It is also notable that in our recent study developing an MLST method for *K. marxianus,* all 57 strains that were listed as coming from (6 different) dairy environments were heterozygous and therefore presumably diploid (Tittarelli, 2018). In fact, in that study, only 13/83 strains were homozygous in the regions included in the MLST. It is noted that one well-studied strain in the literature is *K. marxianus* CBS397 and this study and those of Fasoli *et al* (2016) and Titarelli *et al* (2018) show that this strain is diploid (Fasoli *et al*., 2016; Tittarelli, 2018), which contrasts with what appears to be a previous erroneous suggestion based on long range PCR of the *MAT* locus that it was haploid (Lane *et al*., 2011).

The data suggest that the B-haplotype is a dairy-associated genome. It was gratifying, therefore, to identify one locus in this haplotype that confers a growth advantage in milk, the *LAC12* gene. Our previous work identified positions in the Lac12p where the functional protein had one particular amino acid and the non-functional protein a different one (Varela *et al*., 2017). In six of the strains with a B-haplotype genome, there was an exact match to this functional sequence and in the seventh (NBRC0617), there was a match in 11 positions (**Fig. 5.9**). Since all these strains grew on lactose as a sole sugar source, this now allows us to propose that the number of amino acid positions that distinguish a functional and non-functional lactose-transporting Lac12p protein can be refined to these eleven amino acids though confirmation would require functional tests. The lac- strain with the B-haplotype genome was NBRC0272, which is homozygous for the non-functional *LAC12* allele. This strain was isolated from a non-dairy environment (miso) and the most likely explanation is that it arose like the other diploids as a hybrid between an A and a B strain but since lactose transport was not required in its niche, it was possible for it to lose the B-haplotype *LAC12* allele through a LOH event at the left end of chromosome 3 whereas the other diploids lost the non-functional A-haplotype *LAC12* allele via a similar LOH event (**Fig. 5.7**).

The situation in *K. marxianus* seems to resemble a pattern seen in *Saccharomyces* and *Zygosaccharomyces* species, where strains used in industrial processes or isolated from industrial environments are often polyploids or interspecies hybrids, whereas 'natural' isolates (e.g. from non-anthropogenic environments) tend to be haploid or homozygous diploid (Hittinger, 2013; Suh *et al*., 2013; Wendland, 2014; Ortiz-Merino *et al*., 2017). This pattern is thought to reflect selection towards stress tolerance in the industrial environment, but towards maintenance of the ability to mate and sporulate in natural environments. If this is also the case with *K. marxianus*, it could be expected that the AB diploids display enhanced stress tolerance, at least over B-haplotype haploid strains in a dairy environment. Experiments to date have not succeeded in identifying any correlations between ploidy and stress tolerance (data not shown), but more studies that also include B-haplotype haploids are required to further address this question. As mentioned, B-haplotype strains have not yet been positively identified but there are LAC+ candidates worth investigating, for example, *K. marxianus* NCYC1424, shown to be homozygous by Tittarelli *et al* (Tittarelli, 2018), and *K. marxianus* NCYC1429, which appears to be haploid based on genetic crossing (Varela *et al*., 2017).

Sequence Diversity, aneuploidy and LOH in *K. marxianus*

One of the aims of this study was to assess if the wide phenotypic diversity that has been observed in *K. marxianus* was reflected in its genome diversity. The large number of SNPs (>500k) observed in the set of *K. marxianus* strains used in our study shows relatively high SNP diversity. We found an average pairwise difference (π) of 12×10^{-3}, which is comparable with reported values from other yeasts; for example, 4×10^{-3} in *S. cerevisiae*, 12×10^{-3} in *S. uvarum*, and 17×10^{-3} in *L. kluyveri* (Peter and Schacherer, 2016).

Previous studies with *S. cerevisiae* isolates showed variable ploidy (from 1 to 4 copies of the genome), aneuploidy (unequal copy numbers of different chromosomes), or variation in the copy number of segments of chromosomes (Hose *et al*., 2015; Strope *et al*., 2015; Zhu *et al*., 2016). Similar to S. cerevisiae but unlike *L. kluyveri* (Friedrich *et al*., 2015), all these phenomena were observed in this study of *K. marxianus*. The most unambiguous example of aneuploidy in *K. marxianus* is the presence of an extra copy of chromosome 7 in NBRC0617, but, as described in

the results, there are multiple other likely cases of aneuploidies or copy number variation that would need to be investigated in more detail.

There are also quite extensive regions of loss of heterozygosity in the diploid strains (25-51% by genome length) but not in the triploids. LOH arises when the genome homogenises in a region and it is expected to be rarer in triploid strains, because it will only be apparent if all three copies of a genomic region were homogenised. The shared patterns of LOH in different diploid *K. marxianus* isolates was unexpected. This observation could indicate that these strains are closely related and are mitotic descendants of a recent diploid common ancestor that had already lost heterozygosity in the shared regions. Alternatively, it could indicate that this species mostly reproduces by mitosis and rarely goes through meiosis and sporulation, at least in the dairy environment. Nonetheless, the divergence between the *K. marxianus* clades is low enough that it would not be expected to cause problems in meiosis. In *S. paradoxus*, crosses between strains with sequence divergence of up to 4.6% can still produce viable gametes, as long as the genomes are collinear (Liti *et al*., 2006). It should also be considered that dairy is probably not the original niche for *K. marxianus* and the strains that we studied were most likely selected during a fermentation process. This could have specifically selected hybrids between A and B clade strains and may also promote LOH. Other than the preservation of the functional lactose transporting allele of *LAC12* (B-haplotype), there was not an obvious preference for either the A-haplotype or the B-haplotype during LOH events (**Fig. 5.6**). Because lactose utilisation confers a benefit during growth in milk, one can speculate that the LOH of the region containing the functional *LAC12* gene is an adaptive response. It is not possible to say, however, whether or not other selective pressures played a role in determining the overall patterns of LOH.

Implications for biotechnology

This study focused on a small set of strains of biotechnological interest and therefore may not be fully representative of the species diversity. Indeed, one strain (UFS-Y2791) was far more diverse than the others, suggesting that there is further diversity to be accessed. Given that UFS-Y2791was isolated from agave juice (in South Africa), it will be interesting to see whether strains associated with tequila/mezcal fermentation (also from agave) in Mexico show any relationship to

this strain. The divergence between strains used, either deliberately or traditionally, in the food biotechnology sector is very significant in comparison to those isolated from "natural" environments. Perhaps the natural state of *K. marxianus* is haploid like its sister *K. lactis,* and diploids only arise after biotechnological selection. It is possible that diploids will have advantages, though, other than for lactose utilisation, these are not yet apparent. Haploid strains are much easier to engineer and manipulate so, for most biotechnological applications, it may be preferable to choose Clade 1 (A haplotype) strains. Nonetheless, divergent alleles in Clade 2 (B haplotype) may also be functionally important (for example *LAC12*) so this will still need to be considered in future studies.

5.5 Acknowledgements

RAOM and JAV were supported by the YEASTCELL Marie Curie ITN project which received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° 606795. RAOM was also partially supported by CONACyT, Mexico (fellowship number 440667). AYC was supported by Science Foundation Ireland (13/IA/1910). This study was supported in part by the Adaptable and the Advanced Low Carbon Technology R&D Program (JST, Japan). We thank Kevin Byrne for computational support and colleagues listed in Table 5.1 for providing raw Illumina source files of genome sequences.

5.6 References

Arrizon, J., Morel, S., Gschaedler, A., and Monsan, P. (2012). Fructanase and fructosyltransferase activity of non-Saccharomyces yeasts isolated from fermenting musts of Mezcal. *Bioresour Technol* 110**,** 560-565.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al*. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5)**,** 455-477.

Barsoum, E., Martinez, P., and Astrom, S.U. (2010). Alpha3, a transposable element that promotes host sexual reproduction. *Genes Dev* 24(1)**,** 33-44.

Belloch, C., Barrio, E., Garcia, M.D., and Querol, A. (1998). Inter- and intraspecific chromosome pattern variation in the yeast genus *Kluyveromyces*. *Yeast* 14(15)**,** 1341-1354.

Beniwal, A., Saini, P., Kokkiligadda, A., and Vij, S. (2017). Physiological growth and galactose utilization by dairy yeast *Kluyveromyces marxianus* in mixed sugars and whey during fermentation. *3 Biotech* 7(5)**,** 349.

Booth, L.N., Tuch, B.B., and Johnson, A.D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* 468(7326)**,** 959-963. doi: 10.1038/nature09560.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., *et al*. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6(2)**,** 80-92. doi: 10.4161/fly.19695.

Coloretti, F., Chiavari, C., Luise, D., Tofalo, R., Fasoli, G., Suzzi, G., *et al*. (2017). Detection and identification of yeasts in natural whey starter for Parmigiano Reggiano cheese-making. *International Dairy Journal* 66(Supplement C)**,** 13-17. doi: https://doi.org/10.1016/j.idairyj.2016.10.013.

Costa, D.A., de Souza, C.J., Costa, P.S., Rodrigues, M.Q., dos Santos, A.F., Lopes, M.R., *et al*. (2014). Physiological characterization of thermotolerant yeast for cellulosic ethanol production. *Appl Microbiol Biotechnol* 98(8)**,** 3829-3840. doi: 10.1007/s00253-014-5580-3.

Dias, O., Basso, T.O., Rocha, I., Ferreira, E.C., and Gombert, A.K. (2017). Quantitative physiology and elemental composition of *Kluyveromyces lactis* CBS 2359 during growth on glucose at different specific growth rates. *Antonie Van Leeuwenhoek*. doi: 10.1007/s10482-017-0940-5.

Dias, O., Gombert, A.K., Ferreira, E.C., and Rocha, I. (2012). Genome-wide metabolic (re-) annotation of *Kluyveromyces lactis*. *BMC Genomics* 13**,** 517. doi: 10.1186/1471-2164-13-517.

Dias, O., Pereira, R., Gombert, A.K., Ferreira, E.C., and Rocha, I. (2014). iOD907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*. *Biotechnol J* 9(6)**,** 776-790. doi: 10.1002/biot.201300242.

Diniz, R.H.S., Villada, J.C., Alvim, M.C.T., Vidigal, P.M.P., Vieira, N.M., Lamas-Maceiras, M., *et al*. (2017). Transcriptome analysis of the thermotolerant yeast *Kluyveromyces marxianus* CCT 7735 under ethanol stress. *Appl Microbiol Biotechnol* 101(18)**,** 6969-6980. doi: 10.1007/s00253-017-8432-0.

EFSA-BIOHAZ, Hazards), E.P.o.B., Ricci, A., Allende, A., Bolton, D., Chemaly, M., *et al*. (2017). Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 5: suitability of taxonomic units notified to EFSA until September 2016. *EFSA Journal* 15(3)**,** e04663-n/a. doi: 10.2903/j.efsa.2017.4663.

Fasoli, G., Barrio, E., Tofalo, R., Suzzi, G., and Belloch, C. (2016). Multilocus analysis reveals large genetic diversity in *Kluyveromyces marxianus* strains isolated from Parmigiano Reggiano and Pecorino di Farindola cheeses. *Int J Food Microbiol* 233**,** 1-10. doi: 10.1016/j.ijfoodmicro.2016.05.028.

Fasoli, G., Tofalo, R., Lanciotti, R., Schirone, M., Patrignani, F., Perpetuini, G., *et al*. (2015). Chromosome arrangement, differentiation of growth kinetics and volatile molecule profiles in *Kluyveromyces marxianus* strains from Italian cheeses. *Int J Food Microbiol* 214**,** 151-158. doi: 10.1016/j.ijfoodmicro.2015.08.001.

Fonseca, G.G., de Carvalho, N.M., and Gombert, A.K. (2013). Growth of the yeast *Kluyveromyces marxianus* CBS 6556 on different sugar combinations as sole carbon and energy source. *Appl Microbiol Biotechnol* 97(11)**,** 5055-5067. doi: 10.1007/s00253-013-4748-6.

Fonseca, G.G., Gombert, A.K., Heinzle, E., and Wittmann, C. (2007). Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. *FEMS Yeast Res* 7(3)**,** 422-435. doi: 10.1111/j.1567-1364.2006.00192.x.

Fonseca, G.G., Heinzle, E., Wittmann, C., and Gombert, A.K. (2008). The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl Microbiol Biotechnol* 79(3)**,** 339-354. doi: 10.1007/s00253-008-1458-6.

Foukis, A., Stergiou, P.Y., Theodorou, L.G., Papagianni, M., and Papamichael, E.M. (2012). Purification, kinetic characterization and properties of a novel thermo-tolerant extracellular protease from *Kluyveromyces marxianus* IFO 0288 with potential biotechnological interest. *Bioresour Technol* 123**,** 214-220. doi: 10.1016/j.biortech.2012.06.090.

Friedrich, A., Jung, P., Reisser, C., Fischer, G., and Schacherer, J. (2015). Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol Biol Evol* 32(1)**,** 184-192. doi: 10.1093/molbev/msu295.

Fukuhara, H. (2006). *Kluyveromyces lactis*- a retrospective. *FEMS Yeast Res* 6(3)**,** 323-324. doi: 10.1111/j.1567-1364.2005.00012.x.

Gethins, L., Rea, M.C., Stanton, C., Ross, R.P., Kilcawley, K., O'Sullivan, M., *et al*. (2016). Acquisition of the yeast *Kluyveromyces marxianus* from unpasteurised milk by a kefir grain enhances kefir quality. *FEMS Microbiol Lett* 363(16). doi: 10.1093/femsle/fnw165.

Gombert, A.K., Madeira, J.V., Jr., Cerdan, M.E., and Gonzalez-Siso, M.I. (2016). *Kluyveromyces marxianus* as a host for heterologous protein synthesis. *Appl Microbiol Biotechnol* 100(14)**,** 6193-6208. doi: 10.1007/s00253-016-7645-y.

Groeneveld, P., Stouthamer, A.H., and Westerhoff, H.V. (2009). Super life--how and why 'cell selection' leads to the fastest-growing eukaryote. *FEBS J* 276(1)**,** 254-270. doi: 10.1111/j.1742-4658.2008.06778.x.

Guimaraes, P.M., Teixeira, J.A., and Domingues, L. (2010). Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. *Biotechnol Adv* 28(3)**,** 375-384. doi: 10.1016/j.biotechadv.2010.02.002.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3)**,** 307-321. doi: 10.1093/sysbio/syq010.

Haase, S.B., and Reed, S.I. (2002). Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* 1(2)**,** 132-136.

Hittinger, C.T. (2013). Saccharomyces diversity and evolution: a budding model genus. *Trends Genet* 29(5)**,** 309-317. doi: 10.1016/j.tig.2013.01.002.

Hose, J., Yong, C.M., Sardi, M., Wang, Z., Newton, M.A., and Gasch, A.P. (2015). Dosage compensation can buffer copy-number variation in wild yeast. *Elife* 4. doi: 10.7554/eLife.05462.

Hughes, S.R., Qureshi, N., Lopez-Nunez, J.C., Jones, M.A., Jarodsky, J.M., Galindo-Leva, L.A., *et al*. (2017). Utilization of inulin-containing waste in industrial fermentations to produce biofuels and bio-based chemicals. *World J Microbiol Biotechnol* 33(4)**,** 78. doi: 10.1007/s11274-017-2241-6.

Hutter, S., Vilella, A.J., and Rozas, J. (2006). Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics* 7**,** 409.

Inokuma, K., Ishii, J., Hara, K.Y., Mochizuki, M., Hasunuma, T., and Kondo, A. (2015). Complete Genome Sequence of *Kluyveromyces marxianus* NBRC1777, a Nonconventional Thermotolerant Yeast. *Genome Announc* 3(2). doi: 10.1128/genomeA.00389-15.

Jeong, H., Lee, D.H., Kim, S.H., Kim, H.J., Lee, K., Song, J.Y., *et al*. (2012). Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var. marxianus KCTC 17555. *Eukaryot Cell* 11(12)**,** 1584-1585. doi: 10.1128/EC.00260-12.

Jiang, H., Lei, R., Ding, S.W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15**,** 182. doi: 10.1186/1471-2105-15-182.

Kobayashi, Y., Sahara, T., Suzuki, T., Kamachi, S., Matsushika, A., Hoshino, T., *et al*. (2017). Genetic improvement of xylose metabolism by enhancing the expression of pentose phosphate pathway genes in *Saccharomyces cerevisiae*

IR-2 for high-temperature ethanol production. *J Ind Microbiol Biotechnol* 44(6), 879-891. doi: 10.1007/s10295-017-1912-5.

Lachance, M.-A. (2011). "Chapter 35 - *Kluyveromyces* van der Walt (1971) A2 - Kurtzman, Cletus P," in *The Yeasts (Fifth Edition),* eds. J.W. Fell & T. Boekhout. (London: Elsevier), 471-481.

Lane, M.M., Burke, N., Karreman, R., Wolfe, K.H., O'Byrne, C.P., and Morrissey, J.P. (2011). Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. *Antonie Van Leeuwenhoek* 100(4), 507-519.

Lane, M.M., and Morrissey, J.P. (2010). *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. *Fungal Biology Reviews* 24(1), 17-26. doi: https://doi.org/10.1016/j.fbr.2010.01.001.

Lappe-Oliveras, P., Moreno-Terrazas, R., Arrizon-Gavino, J., Herrera-Suarez, T., Garcia-Mendoza, A., and Gschaedler-Mathis, A. (2008). Yeasts associated with the production of Mexican alcoholic nondistilled and distilled Agave beverages. *FEMS Yeast Res* 8(7), 1037-1052.

Lee, J.W., In, J.H., Park, J.B., Shin, J., Park, J.H., Sung, B.H., *et al*. (2017). Co-expression of two heterologous lactate dehydrogenases genes in *Kluyveromyces marxianus* for l-lactic acid production. *J Biotechnol* 241, 81-86. doi: 10.1016/j.jbiotec.2016.11.015.

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., *et al*. (2015). Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. *Biotechnol Biofuels* 8, 47. doi: 10.1186/s13068-015-0227-x.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14), 1754-1760. doi: 10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al*. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16), 2078-2079. doi: 10.1093/bioinformatics/btp352.

Lin, Y.J., Chang, J.J., Lin, H.Y., Thia, C., Kao, Y.Y., Huang, C.C., *et al*. (2017). Metabolic engineering a yeast to produce astaxanthin. *Bioresour Technol* 245(Pt A)**,** 899-905. doi: 10.1016/j.biortech.2017.07.116.

Lischer, H.E., Excoffier, L., and Heckel, G. (2014). Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of microtus voles. *Mol Biol Evol* 31(4)**,** 817-831.

Liti, G., Barton, D.B., and Louis, E.J. (2006). Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174(2)**,** 839-850.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., *et al*. (2009). Population genomics of domestic and wild yeasts. *Nature* 458(7236)**,** 337-341. doi: 10.1038/nature07743.

Morrissey, J.P., Etschmann, M.M., Schrader, J., and de Billerbeck, G.M. (2015). Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. *Yeast* 32(1)**,** 3-16. doi: 10.1002/yea.3054.

Nambu-Nishida, Y., Nishida, K., Hasunuma, T., and Kondo, A. (2017). Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. *Sci Rep* 7(1)**,** 8993. doi: 10.1038/s41598-017-08356-5.

Nonklang, S., Abdel-Banat, B.M., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., *et al*. (2008). High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast *Kluyveromyces marxianus* DMKU3-1042. *Appl Environ Microbiol* 74(24)**,** 7514-7521. doi: 10.1128/AEM.01854-08.

Nonklang, S., Ano, A., Abdel-Banat, B.M., Saito, Y., Hoshida, H., and Akada, R. (2009). Construction of flocculent *Kluyveromyces marxianus* strains suitable for high-temperature ethanol fermentation. *Biosci Biotechnol Biochem* 73(5)**,** 1090-1095.

Oliveira, S.S.S., Bello, M.L., Rodrigues, C.R., Azevedo, P.L., Ramos, M., Aquino-Neto, F.R., *et al*. (2017). Asymmetric bioreduction of beta-ketoesters derivatives by *Kluyveromyces marxianus*: influence of molecular structure on

the conversion and enantiomeric excess. *An Acad Bras Cienc* 89(3)**,** 1403-1415. doi: 10.1590/0001-3765201720170118.

Ortiz-Merino, R.A., Kuanyshev, N., Braun-Galleani, S., Byrne, K.P., Porro, D., Branduardi, P., *et al*. (2017). Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol* 15(5)**,** e2002128. doi: 10.1371/journal.pbio.2002128.

Peter, J., and Schacherer, J. (2016). Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* 33(3)**,** 73-81. doi: 10.1002/yea.3142.

Quarella, S., Lovrovich, P., Scalabrin, S., Campedelli, I., Backovic, A., Gatto, V., *et al*. (2016). Draft Genome Sequence of the Probiotic Yeast *Kluyveromyces marxianus* fragilis B0399. *Genome Announc* 4(5).

Rajaei, N., Chiruvella, K.K., Lin, F., and Astrom, S.U. (2014). Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. *Proc Natl Acad Sci U S A* 111(43)**,** 15491-15496. doi: 10.1073/pnas.1406027111.

Rocha, S.N., Abrahao-Neto, J., and Gombert, A.K. (2011). Physiological diversity within the *Kluyveromyces marxianus* species [corrected]. *Antonie Van Leeuwenhoek* 100(4)**,** 619-630. doi: 10.1007/s10482-011-9617-7.

Rodicio, R., and Heinisch, J.J. (2013). Yeast on the milky way: genetics, physiology and biotechnology of *Kluyveromyces lactis*. *Yeast* 30(5)**,** 165-177. doi: 10.1002/yea.2954.

Schabort du, T.W., Letebele, P.K., Steyn, L., Kilian, S.G., and du Preez, J.C. (2016). Differential RNA-seq, Multi-Network Analysis and Metabolic Regulation Analysis of *Kluyveromyces marxianus* Reveals a Compartmentalised Response to Xylose. *PLoS One* 11(6)**,** e0156242.

Schacherer, J., Shapiro, J.A., Ruderfer, D.M., and Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458(7236)**,** 342-345.

Schaffrath, R., and Breunig, K.D. (2000). Genetics and molecular physiology of the yeast *Kluyveromyces lactis*. *Fungal Genet Biol* 30(3)**,** 173-190.

Schroder, M.S., Martinez de San Vicente, K., Prandini, T.H., Hammel, S., Higgins, D.G., Bagagli, E., *et al*. (2016). Multiple Origins of the Pathogenic Yeast *Candida orthopsilosis* by Separate Hybridizations between Two Parental Species. *PLoS Genet* 12(11)**,** e1006404. doi: 10.1371/journal.pgen.1006404.

Signori, L., Passolunghi, S., Ruohonen, L., Porro, D., and Branduardi, P. (2014). Effect of oxygenation and temperature on glucose-xylose fermentation in *Kluyveromyces marxianus* CBS712 strain. *Microb Cell Fact* 13(1)**,** 51.

Silveira, W.B., Diniz, R.H., Cerdan, M.E., Gonzalez-Siso, M.I., Souza Rde, A., Vidigal, P.M., *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).

Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., *et al*. (2000). Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett* 487(1)**,** 3-12.

Strope, P.K., Skelly, D.A., Kozmin, S.G., Mahadevan, G., Stone, E.A., Magwene, P.M., *et al*. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res* 25(5)**,** 762-774.

Suh, S.O., Gujjari, P., Beres, C., Beck, B., and Zhou, J. (2013). Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int J Syst Evol Microbiol* 63(Pt 5)**,** 1922-1929. doi: 10.1099/ijs.0.048058-0.

Tittarelli, F; Varela., J; Gethins, C; Stanton, C; Ross, P; Suzzi, G; Grazia, L; Tofalo, R; Morrissey, JP (2018). Development and Implementation of Multi Locus Sequence Typing to study diversity of the yeast *Kluyveromyces marxianu*s in Italian Cheeses. *Microbial Genomics* doi: 10.1099/mgen.0.000153

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., *et al*. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43**,** 11 10 11-33. doi: 10.1002/0471250953.bi1110s43.

Varela, J.A., Montini, N., Scully, D., Van der Ploeg, R., Oreb, M., Boles, E., *et al*. (2017). Polymorphisms in the *LAC12* gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains. *FEMS Yeast Res* 17(3). doi: 10.1093/femsyr/fox021.

Verdugo Valdez, A., Segura Garcia, L., Kirchmayr, M., Ramirez Rodriguez, P., Gonzalez Esquinca, A., Coria, R., *et al*. (2011). Yeast communities associated with artisanal mezcal fermentations from Agave salmiana. *Antonie Van Leeuwenhoek* 100(4)**,** 497-506. doi: 10.1007/s10482-011-9605-y.

Wang, Y.J., Ying, B.B., Shen, W., Zheng, R.C., and Zheng, Y.G. (2017). Rational design of *Kluyveromyces marxianus* ZJB14056 aldo-keto reductase KmAKR to enhance diastereoselectivity and activity. *Enzyme Microb Technol* 107**,** 32-40. doi: 10.1016/j.enzmictec.2017.07.012.

Wendland, J. (2014). Lager yeast comes of age. *Eukaryot Cell* 13(10)**,** 1256-1265. doi: 10.1128/EC.00134-14.

Wu, W.H., Hung, W.C., Lo, K.Y., Chen, Y.H., Wan, H.P., and Cheng, K.C. (2016). Bioethanol production from taro waste using thermo-tolerant yeast *Kluyveromyces marxianus* K21. *Bioresour Technol* 201**,** 27-32.

Yarimizu, T., Nonklang, S., Nakamura, J., Tokuda, S., Nakagawa, T., Lorreungsil, S., *et al*. (2013). Identification of auxotrophic mutants of the yeast Kluyveromyces marxianus by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*. *Yeast* 30(12)**,** 485-500. doi: 10.1002/yea.2985.

Zhu, Y.O., Sherlock, G., and Petrov, D.A. (2016). Whole Genome Analysis of 132 Clinical *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation. *G3 (Bethesda)* 6(8)**,** 2421-2434. doi: 10.1534/g3.116.029397.

Zonneveld, B.J.M., and Steensma, H.Y. (2003). "Mating, Sporulation and Tetrad Analysis in *Kluyveromyces lactis*," in *Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols,* eds. K. Wolf, K. Breunig & G. Barth. (Berlin, Heidelberg: Springer Berlin Heidelberg), 151-154.

Chapter 6

General discussion and recommendations for future research

6.1 Overview of the findings in this thesis

The overall objective of this thesis was to study the function and evolution of *K. marxianus* sugar transporters. This topic was chosen due to two main reasons: the subject has been poorly studied in *K. marxianus* (much of what is known has been inferred from *K. lactis* studies) and because obtaining a deeper understanding of sugar transport could lead to better substrate usage in strains used for different biotechnological applications. To study sugar transport in this yeast, a combination of biochemical methods, genomics and molecular techniques was used.

First, lactose utilisation variability was studied in *K. marxianus* (Chapter 3). The molecular basis of this trait has been mainly studied in *K. lactis* where the *LAC4-LAC12* gene pair is the genetic determinant involved in lactose utilisation. In this thesis, it was shown that *K. marxianus* contains additional elements involved in lactose transport that had not been identified previously. Unlike its sister species, *K. marxianus* encodes four copies of the *LAC12* gene with only one of them (*KmLAC12*) encoding a functional lactose transporter. Variation in lactose utilisation among *K. marxianus* strains correlated with changes in the *KmLAC12* sequence. Good and poor lactose utilising strains carry a functional (*LAC12*-B) and non-functional (*LAC12*-A) lactose transporter, respectively. The two sequences differ in 14 amino acid positions (Chapter 3 & 5). These differences are sufficient to separate both sequences functionally.

Although the *LAC12*-A gene does not encode a lactose permease, it was hypothesised that it, and the additional *LAC12* genes in *K. marxianus,* might encode galactose transporters (Chapter 4). This idea was based on the fact that the Lac12p transporter from *K. lactis* was previously reported to transport both lactose and galactose. Heterologous expression of the *LAC12* genes established that three out of four of these genes encode galactose transporters when expressed in *S. cerevisiae.* However, a *K. marxianus* strain carrying null-mutations of these genes could still grow on galactose, indicating the presence of additional galactose transport systems. To identify these transporters, a genome-wide annotation of membrane transporters was performed showing an expansion of the hexose transporter genes *KHT* and *HGT*. Expression of these genes in *S. cerevisiae* revealed the presence of four additional galactose transporters. To establish their role in galactose transport, *K. marxianus* mutant strains

were constructed using the CRISPR-Cas9 system (Chapter 2). Growth tests of these strains showed that only the mutant lacking all eight of the identified galactose transporters failed to grow on galactose. These results indicate an important degree of redundancy in galactose transport in *K. marxianus*.

K. marxianus is known for being a physiologically diverse species (Lane *et al*., 2011) but the degree of genomic diversity of the species has not been established to date. To obtain an insight into the diversity of the species, a comparative analysis of 14 *K. marxianus* genomes was performed (Chapter 5). Analysis of these genome sequences revealed a high degree of genetic diversity among isolates. Also, the strains were found in different ploidy states (haploid, diploid or triploid), with some of them displaying partial aneuploidy in some regions of the genome. Interestingly, the analysis of these genomes revealed that dairy and non-dairy strains have distinct ploidy patterns: dairy isolates were diploid or triploid, while non-dairy isolates were haploid. When compared using a phylogenetic haplotype tree, the strains formed three clades; a dairy clade (B), and two non-dairy clades (A and C). Interestingly, clade C only contains the strain UFS-Y2791, which was isolated from an *Agave* plant. Since this strain was the only one isolated from a plant source, it is possible to speculate that other strains isolated from similar environments would form part of the same clade. Also, including strains isolated from other niches (e.g. decaying fruit or plant material) could help expanding this phylogenetic tree, thus contributing to obtaining a better understanding of the genome diversity in the species.

Molecular tools to genome-engineer *K. marxianus* were also developed in this thesis (Chapter 2). A single-plasmid CRISPR-Cas9 expression system, functional in *K. lactis* and *K. marxianus*, was constructed and used to successfully target the *ADE2* gene. The system was then modified to allow fast cloning of target sequences and used to target several genes, including the *KHT* and *HGT* galactose transporters (Chapter 4). To enhance gene targeting, the *KU80* gene, involved in the non-homologous end-joining DNA repair system, was disrupted, resulting in a significant increase in homologous repair. Finally, to further increase the efficiency of homologous-recombination mediated repair, the *ScRAD52* was integrated into *K. marxianus* genome. Expression of this gene resulted in a seven-fold increase in homologous-recombination mediated repair.

6.2 Evolution of lactose utilisation in *K. marxianus*

One of the most interesting aspects related to lactose utilisation in *K. marxianus* is understanding how and when this trait evolved in the yeast. To explain this phenomenon, it is hypothesized that lactose utilisation evolved from cellobiose utilisation. The various points that support this idea are presented in this section.

Cellobiose is a disaccharide with structural resemblance to lactose. In fact, the two sugars only differ structurally in the position of a chiral carbon (Kurtzman and Fell, 2000). While other sugars, such as maltose and sucrose, are also similar to lactose, *K. marxianus* does not utilise maltose and degrades sucrose extracellularly (Rouwenhorst *et al*., 1990). Cellobiose can be utilised by *K. marxianus* (Barron *et al*., 1994). Also, it was demonstrated that the *LAC12* transporter from *K. lactis* (identical in sequence to *LAC12-B*) is able to confer growth on cellobiose when expressed in combination with a β-glucosidase (Sadie *et al*., 2011). Similar results have been obtained in *E. coli*, where the LacY transporter was found to transport cellobiose, and in *S. cerevisiae* where a strain engineered to ferment cellobiose was also able to ferment lactose (Liu *et al*., 2016; Sekar *et al*., 2012). Altogether, these findings show that the cellobiose and lactose utilisation systems overlap functionally and in some cases lactose permeases can transport cellobiose and some β-glucosidases can hydrolyse lactose and cellobiose (Akiyama *et al*., 2001; Hassan *et al*., 2015). Taking this into consideration, the *LAC12* gene could have originally encoded a cellobiose transporter that later evolved into a lactose transporter (**Figure 6.1**). Initially, the early *K. marxianus* ancestor may not have been able to utilise lactose but could have transported and hydrolysed cellobiose by the activity of *LAC12-A* and the β-glucosidase gene (*BGL*), which is located next to the *LAC12-LAC4* gene pair (**Figure 6.1**, number 1). While *LAC12-A* is considered a non-functional lactose transporter, strains carrying this gene grow slowly on lactose suggesting that the sugar gets transported into the cell at a very low rate. Partial transport of lactose into the cell could have played a role in the evolution of the β-galactosidase gene, *LAC4* (**Figure 6.1**, number 2). The origin of this gene is not clear but it is possible that either evolved from a gene encoding an enzyme with a different substrate or was acquainted by a horizontal gene transfer event. If the gene originated in *K. marxianus*, weak transport of lactose into the cell by Lac12p-A could have played a role in its evolution. Lactose is toxic to *K. lactis lac4* mutants but

this species carries the *LAC12-B* gene encoding the efficient lactose transporter (Lodi and Donnini, 2005). Weak lactose transport by Lac12p-A could have contributed in maintaining non-toxic intracellular lactose levels that may have acted as selective pressure for *LAC4* evolution. According to this model, the acquisition of a functional β-galactosidase allowed lactose utilisation for the first time, thus becoming a critical step in the evolution of lactose utilisation in *K. marxianus*. The extra *LAC12* copies may have also arisen at this point by duplication of the *LAC12-LAC4* pair and later duplication of *LAC12* into chromosomes II and VIII (Chapter 3).

The yeast carrying the *LAC12-A* gene transports lactose inefficiently but permanent exposure to lactose could have acted as selective pressure for a more efficient lactose transporter, resulting in mutations leading to *LAC12-B* (**Figure 6.1**, number 3). The fact that this haplotype was only found in dairy strains, supports this theory (Chapter 5). It is worth mentioning that even though *K. marxianus* strains can be classified according to the *LAC12* variant they carry, other non-functional variants of the gene could exist. For example, the strain UFS-Y2791 has a distinct *LAC12* haplotype that is closer to A than B in sequence (Chapter 5). Further experiments are required to determine the presence of additional haplotypes and their role in the evolution of lactose utilisation.

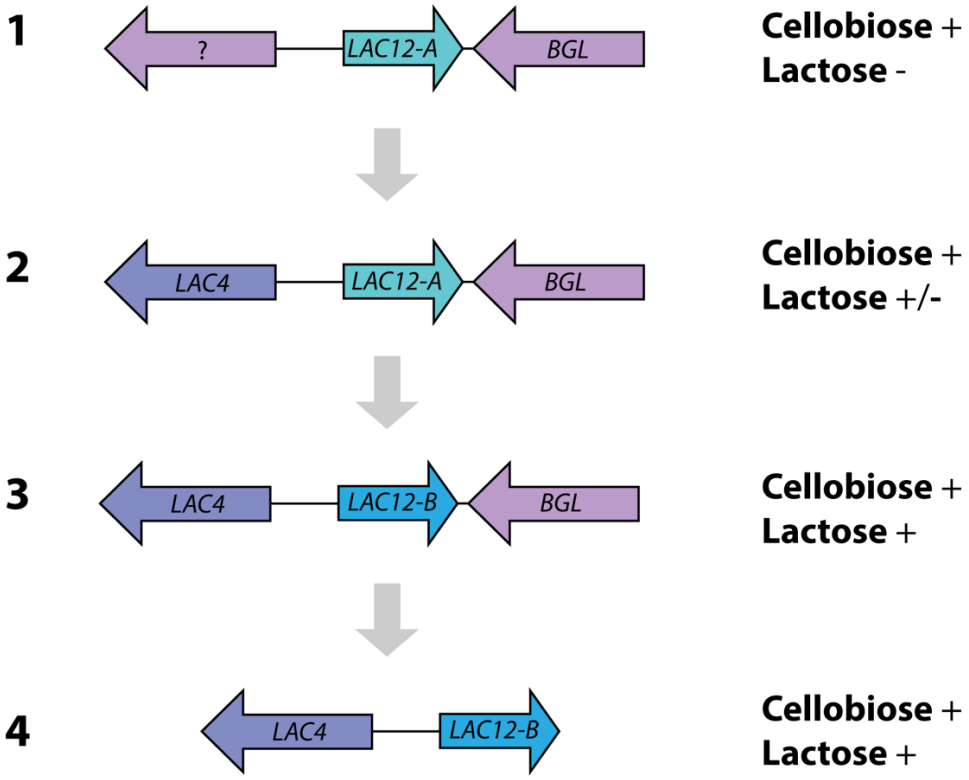


Figure 6.1: Evolution of lactose utilisation genes in *K. marxianus*.
1: the *LAC12* gene may have originally encoded a cellobiose transporter. The absence of β-galactosidase activity impeded lactose utilisation. 2: partial exposure to lactose resulted in the evolution of *LAC4* by a neo-functionalisation process. Acquisition of this new gene enabled weak lactose utilisation. 3: further exposure to lactose acted as selective pressure resulting in the evolution of the *LAC12-A* into *LAC12-B*. 4: the *LAC4-LAC12* gene pair was acquired by *K. lactis* by an introgression process, enabling efficient lactose utilisation in this yeast species.

6.3 Evolution of lactose utilisation in other *Kluyveromcyces* species

Another interesting aspect about the evolution of lactose utilisation is understanding how other *Kluyveromyces* species acquired this trait. The *K. lactis* variant *K. lactis* var. *drosophilarum* does not have the *LAC12-LAC4* genes and is unable to utilise lactose (Naumov *et al*., 2006). However, these genes are present in *K. lactis* and it was proposed that they were acquired in this species by a horizontal gene transfer event (Naumov, 2008). Unfortunately, the *K. lactis* var. *drosophilarum* genome is not available to date but a comparison between the *K. marxianus* and *K. lactis* genomes provides useful information regarding the origin of the *LAC* genes in *K. lactis*. The *LAC12-LAC4* gene pair is located in a subtelomeric region in both *K. lactis* and *K. marxianus*. A comparison of these regions shows that the first 15 kb (containing *LAC4* and *LAC12*) from chromosome II in *K. lactis* and chromosome III in *K. marxianus* haplotype-B are 99.8 % identical, which suggests that *K. lactis* acquired this DNA

segment from *K. marxianus* (**Figure 6.1**, number 4). A comparison of the organisation of the *LAC12-LAC4* gene pair in other *Kluyveromyces* species supports this idea (**Figure 6.2**). While *K. marxianus* and *K. lactis* have a similar gene organisation (all the genes up to the *LAC12* sequence are highly similar in both species), the genes have a different position in *K. aestuarii* and *K. wickerhamii*. Instead of sharing a common promoter and facing opposite directions, the *LAC12* and *LAC4* genes in these two species are positioned one after the other. Due to the poor quality of the genome sequences available for these yeasts, it is difficult to establish the exact position of the *LAC12-LAC4* genes. However, the differences in gene orientation and position raise questions about the evolution of lactose utilisation in the *Kluyveromyces* genus. A possible answer is that the genes required to transport and hydrolyse this sugar evolved in *K. marxianus* and another *Kluyveromyces* species at the same time. Nevertheless, is also possible that *K. lactis* loss *LAC12-LAC4* and these genes were later acquired from *K. marxianus.* More analyses, including detailed phylogenetic analysis of the *LAC4-LAC12* sequences and genome sequencing of other *Kluyveromyces* species, are required to shed a light on the evolution of lactose utilisation in the genus.

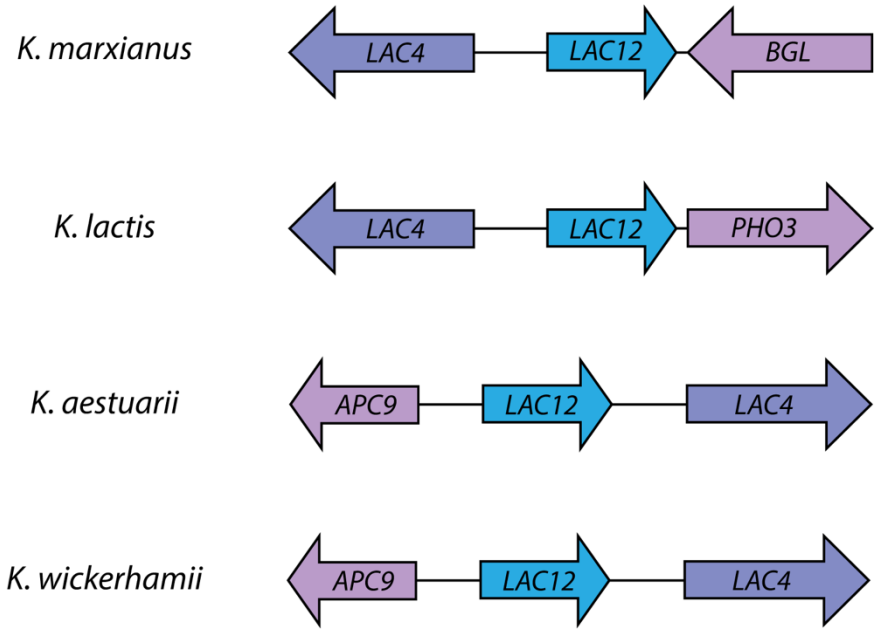


Figure 6.2: Genomic context of the *LAC12* and *LAC4* genes in *Kluyveromyces* genomes. The organisation of the *LAC* genes in the different genomes is shown. The genes are located in the subtelomeric region of chromosome II and III in *K. lactis* and *K. marxianus*, respectively. In *K. marxianus* the genes are next to the β-glucosidase whereas they are located next to the acid phosphatase gene, *PHO3,* in *K. lactis.* Positions of these genes are conserved between *K. aestuarii* and *K. wickerhamii*. Both genes are absent in *K. dobzhanskii* genome.

6.4 Gene regulation in *LAC12-A* and *LAC12-B K. marxianus* strains

The regulatory elements controlling *LAC12* and *LAC4* expression are conserved in *K. lactis* and *K. marxianus*. The main transcriptional factor interacting with the *LAC12-LAC4* promoter is Lac9p (orthologous to Gal4p in *S. cerevisiae*) which binds to the 5′ CGG-N_{11}-CCG 3′ consensus sequence to activate gene expression (Godecke *et al*., 1991). The *K. lactis* promoter contains four of these sites and it was shown that deletion of even one of these sites has a negative effect on gene expression (Godecke *et al*., 1991). In this study, it was shown that CBS6556, a *LAC12-A* haplotype strain, contains three Gal4p activation sites (Chapter 3). Nevertheless, new data obtained by sequencing the genome of the B-haplotype strain, CBS 397, has revealed that this strain has four potential Gal4p binding sites as in *K. lactis* (**Figure 6.3A**). Further comparisons between A and B genomes established that this pattern is conserved: A and B haplotype strains contain three or four Gal4p sites, respectively. Interestingly, RT-qPCR analyses showed that expression of *LAC12* and *LAC4* in lactose is comparable between CBS397 and CBS6556 (**Figure 6.3B**). Since the direct inducer of the Gal regulon is galactose and not lactose, expression analyses were also performed using this sugar as a carbon source. Preliminary results show that expression of *LAC12* and *LAC4* in these conditions is 100-fold lower when compared to expression in lactose (**Figure 6.3C**). Expression of *GAL1*, encoding a galactokinase, whose expression is activated by galactose was also measured to rule out a possible galactose activation defect on CBS6556. Nevertheless, expression of this gene was comparable between strain in both conditions, indicating that the differences observed are specific to the *LAC12-LAC4* promoter in CBS6556.

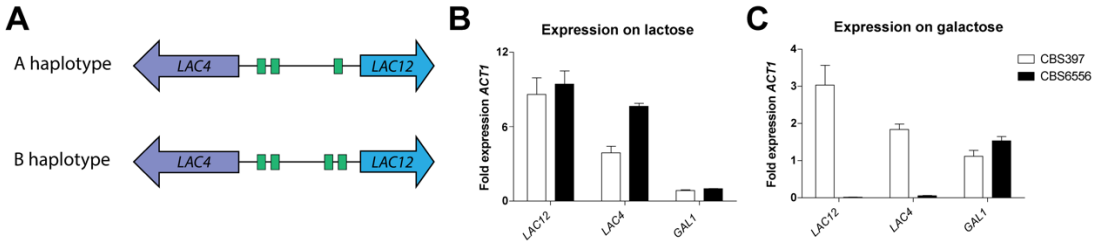


Figure 6.3: Promotor structure an expression patterns in haplotype A and B strains. A: Structure of the *LAC12-LAC4* genes in haplotype A and B strains. The Gal4p sites are shown in green. B: Expression of the *LAC12*, *LAC4* and *GAL1* genes on lactose. C: Expression of the *LAC12*, *LAC4* and *GAL1* genes on galactose. CBS397 and CBS6556 data are shown in white and black bars, respectively. Data shown in panel B corresponds to results presented in Chapter 3. The data from panel C was obtained using the same methodology described in Chapter 3 but was not published.

These findings could indicate that lactose activates gene expression by an alternative Gal4p-independent mechanism. Hence, the overall levels of expression observed in lactose could be the result of two simultaneous induction mechanisms: weak Gal4p activation (by galactose after lactose hydrolysis) and an unknown induction system activated by lactose. A possible explanation for this phenomenon could be related to activation of a cellobiose induction mechanism. As previously mentioned, lactose utilisation could have evolved from cellobiose utilisation. If this is the case, then some cellobiose-responsive elements controlling the expression of *LAC12* (able to transport cellobiose) must exist in *K. marxianus*. Because of structural similarities, lactose could mimic cellobiose and activate this system. Since the *LAC12-LAC4* promoter is bidirectional this activation could translate into high expression of both genes in lactose.

The interplay between cellobiose and lactose regulation systems has been reported in the fungus *Trichoderma reesei*. In this organism, production of cellulase and hemicellulose enzymes, which degrade cellulose into other sugars such as cellobiose, is induced by lactose (Li *et al*., 2017). In fact, lactose is used as an inducer for producing these enzymes in industrial scale. A null mutant in a predicted lactose transporter did not exhibit lactose-dependent activation of these genes, suggesting that lactose uptake is necessary for cellulase induction (Ivanova *et al*., 2013). Since the fungus is unlikely to encounter lactose in its natural habitat (decaying wood) it was proposed that the sugar induces expression of these enzymes by mimicking a plant substance available in its natural niche, possibly cellobiose.

More experiments are required to evaluate if a similar system is responsible for the induction of *LAC12-LAC4* in *K. marxianus*. One approach to explore this possibility is to construct a *gal4* mutant in a haplotype A strain. If lactose induction of *LAC12* and *LAC4* expression is observed in this strain, that would suggest that lactose is activating an alternative regulatory system. Also, it would be interesting to evaluate if cellobiose induces expression of the *LAC* genes in *K. marxianus*. This experiment could further support the idea of secondary regulation mechanism governing *LAC12-LAC4* expression. The *K. marxianus* genome does not contain homologues of any of the known fungal cellobiose regulators but another six genes (excluding *GAL4*) encode Gal4p-like transcriptional factors, that could play a role in cellobiose metabolism and activation of *LAC12* expression. To test this idea, mutants could be

constructed using the CRISPR-Cas9 system and then tested for *LAC12* expression and growth on different carbon sources. In addition, it would be of interest to evaluate if the other *LAC12* genes from *K. marxianus* encode cellobiose transporters. The physiological role of these genes was not determined in this study, but it is possible that they could transport other disaccharides such as cellobiose. Co-expression of these transporters and a β-glucosidase in *S. cerevisiae* could be performed to test this idea.

6.5 Genome-engineering tool development

K. marxianus is an attractive organism for several biotechnological applications but further developments in the area require precise molecular tools to manipulate this yeast. Traditionally, *K. marxianus* has been considered hard to genome engineer because of possessing an active non-homologous end-joining recombination system, which makes gene targeting difficult. To overcome this issue, the CRISPR-Cas9 system was implemented in this yeast allowing the disruption of several genes in a targeted fashion (Chapter 2). While this system was being tested, two additional studies on the topic were published. One of these studies aimed to develop a comprehensive molecular toolbox in *K. marxianus* based on the deaminase-mediated targeted point mutagenesis technology (Target-AID) (Nambu-Nishida *et al*., 2017). In this technique a hybrid protein comprising a nuclease-deficient version of Cas9 fused to an activation-induced cytidine deaminase (AID) is used to introduce point mutations in the target sequence. Recognition of the target sequence is mediated by Cas9 (in combination with an adequate gRNA molecule) and the introduction of point mutations in the target region is mediated by the AID enzyme, which converts cytosine to uracil (Nishida *et al*., 2016). The DNA replication machinery treats the uracil base as thymine resulting in a C/G to A/T mutation. The authors used this system to induce mutagenesis of several genes including the *NEJ1* and *LIG4* genes, involved in the NHEJ system. Also, targeting the 5′ and 3′ of the *URA3* gene with two gRNAs molecules at the same time was used to integrate three overlapping DNA fragments into the genome without using a selection marker (Nambu-Nishida *et al*., 2017). A second study aimed to use the CRISPR-Cas9 system to study ethyl acetate production in *K. marxianus* CBS 6556 (Löbs *et al*., 2017). In this case, the system was used to

introduce point mutations in the genes of interest, therefore no additional modifications were introduced in the host strain.

While these systems are substantially different, both rely on using *K. marxianus* RNA polymerase III promoters to express the gRNA molecule; the *SNR52* and the *RPR1*-$tRNA^{gly}$ promoters were used in the Target-AID and ethyl acetate production studies, respectively. Choice of promoter was shown to affect targeting efficiency. A comparison of five *K. marxianus* promoters performed by Löbs *et al*., 2017, showed that a hybrid promoter composed of the *RPR1* and $tRNA^{gly}$ sequences gives the best results when targeting the *XYL2* gene (60 % efficiency), encoding the xylitol dehydrogenase gene. On the contrary, the *SNR52* promoter was the weakest with a targeting efficiency of 10 %. Expression of the gRNA molecule is the most critical element determining Cas9 activity (Hsu *et al*., 2013). A further comparison between different promoters, including a ribozyme-based RNA polymerase II system, has been performed in the yeast *Yarrowia lipolytica* (Schwartz *et al*., 2016). In that study, higher targeting efficiencies were observed when using RNA polymerase III promoters, however, only one RNA polymerase II promoter was tested. Conducting a similar analysis, including several RNA pol II and III promoters, would be useful in optimising the CRISPR-Cas9 system in *K. marxianus*. Also, since *K. marxianus* has been shown to display a high degree of physiological diversity it would be of interest to evaluate these systems on different isolates. Another aspect that has not been considered in any these studies is multiplexing. Targeting multiple genes simultaneously with CRISPR-Cas9 has been successfully shown in *S. cerevisiae* (Jakočiūnas *et al*., 2015; Mans *et al*., 2015). While there are different gRNA delivery systems the most convenient methods involve expressing several gRNA molecules from a single plasmid (Mans *et al*., 2015). Nevertheless, alternative strategies to express and process several gRNA molecules are also available. One of these multiplexing techniques relies on using the Csy4 endoribonuclease from *Pseudomonas aeruginosa* to generate up to 4 gRNA molecules from a single transcript (Ferreira *et al*., 2018). This protein binds to a RNA stem-loop in the structural gRNA sequence and cleaves the RNA molecule releasing the gRNA molecules sequentially. Unfortunately, since the nuclease has to be expressed from a plasmid or the genome, the approach could be impractical in some cases. Lately, Cpf1, a new type of CRISPR-Cas nuclease, has been implemented as a tool to genome-engineer *S. cerevisiae* (Swiat

et al., 2017; Verwaal *et al*., 2018). Besides being an alternative to Cas9, Cpf1 contains an endoribonuclease domain that processes the CRISPR-RNA transcript. This means that a single CRISPR transcript containing several gRNAs can be processed without an additional RNase, thus facilitating multiplexing applications. Although this technology has only been recently developed it could potentially be applied to other yeast species such as *K. marxianus*. The implementation of these and other tools would further contribute to obtaining a universal molecular toolbox in this yeast.

6.6 Biotechnological potential of *K. marxianus*

Because of its physiological traits, *K. marxianus* has become an important candidate for several cell factory applications. The lack of physiological information together with the scarce availability of genetic tools to modify the yeast have traditionally limited the range of applications mainly to bioethanol and enzyme production (Fonseca *et al*., 2008). However, recent studies focused on strain physiology and genetic tool development have opened possibilities for new cell factory applications including bioflavour and organic acid production (Bae *et al*., 2018; Morrissey *et al*., 2015; Zhang *et al*., 2017). For example, a *K. marxianus* strain has been developed to produce lactic acid, an important precursor for the cosmetic and pharmaceutical industries (Bae *et al*., 2018). In comparison to other production systems that rely on using glucose or xylose as substrates (Skory, 2003; Turner *et al*., 2015), the engineered *K. marxianus* strain can synthetize lactic acid from Jerusalem artichoke tuber powder, an inulin-rich substance derived from the Jerusalem artichoke plant. The use of this renewable crop as substrate together with *K. marxianus* favourable characteristics (rapid growth, thermotolerance and stress resistance) could lead to the implementation of a cost-effective industrial process for lactic acid biosynthesis. Production of other chemicals such as pyruvate and astaxanthin has been recently demonstrated in *K. marxianus* (Lin *et al*., 2017; Zhang *et al*., 2017). Also, a patent for the production of cannabinoids in *S. cerevisiae* and *K. marxianus* has been published recently (U.S. Patent No. 14795816). Because *K. marxianus* can be used for high-temperature fermentations, ethanol production has received great attention with several studies investigating different aspects of the process such as stress tolerance (Gao *et al*., 2017a, 2017b), alternative substrates (Buenrostro-Figueroa *et al*., 2017; Chang *et al*.,

2018; Hughes *et al*., 2017; Islam *et al*., 2018; Pilap *et al*., 2018) and cofactor balance (Kobayashi *et al*., 2017), among others. Nevertheless, more studies are required to optimise other characteristics that could be important in an industrial setup. For example, anaerobic growth and ethanol tolerance are essential in ensuring a good fermentation performance. In fact, ethanol toxicity affects *S. cerevisiae* cell viability in industrial-like fermentations (Bai *et al*., 2004). Since *K. marxianus* is less tolerant to ethanol than *S. cerevisiae*, developing ethanol-tolerant strains would be critical in positioning *K. marxianus* as an alternative platform for industrial ethanol production. Similarly, oxygen depletion, often present in industrial fermentations, is one of the main constraints in xylose utilisation in *K. marxianus* (Signori *et al*., 2014). Future studies focused on this topic could further contribute in developing *K. marxianus* for efficient ethanol production.

Overall, the metabolic engineering strategies used in other yeasts such as *S. cerevisiae* can be transferred to *K. marxianus*. However, since the two species are physiologically distinct, the success of such of approach could be limited. For this reason, the development of new biotechnological applications in *K. marxianus* (and improvement of existing ones) will most likely require combining several approaches (e.g. strain characterisation, genetic tools, omics analysis, metabolic modelling and laboratory evolution). Some of these tools are partially available; some genetic tools have been recently developed (Chapter 2, Löbs *et al*., 2017; Nambu-Nishida *et al*., 2017), a model of its central metabolism is available (Pentjuss *et al*., 2017) and several transcriptome analyses have been published (Diniz *et al*., 2017; Gao *et al*., 2015; Lertwattanasakul *et al*., 2015; Schabort *et al*., 2018). While all of these studies collectively represent an important progress in the field, more efforts are required to establish a framework for the creation of new chassis strains. One of the aspects that could further developed is the implementation of an omics − metabolic modelling approach. The benefits of undertaking such an analysis are two-fold: a deep understanding of metabolic fluxes would be gained, and the information obtained could be used as a framework to design and perform metabolic engineering experiments in *K. marxianus*.

6.7 References

Akiyama, K., Takase, M., Horikoshi, K., Okonogi, S., 2001. Production of galactooligosaccharides from lactose using a β-glucosidase from Thermus sp. Z-1. Biosci. Biotechnol. Biochem. 65, 438–441. doi:10.1271/bbb.65.438

Bae, J.-H., Kim, H.-J., Kim, M.-J., Sung, B.H., Jeon, J.-H., Kim, H.-S., Jin, Y.-S., Kweon, D.-H., Sohn, J.-H., 2018. Direct fermentation of Jerusalem artichoke tuber powder for production of L-lactic acid and D-lactic acid by metabolically engineered *Kluyveromyces marxianus*. J. Biotechnol. 266, 27–33. doi:10.1016/j.jbiotec.2017.12.001

Bai, F., Chen, L., Zhang, Z., Anderson, W., Moo-Young, M., 2004. Continuous ethanol production and evaluation of yeast cell lysis and viability loss under very high gravity medium conditions. J. Biotechnol. 110, 287–293. doi:10.1016/J.JBIOTEC.2004.01.017

Barron, N., Marchant, R., McHale, L., McHale, A.P., 1994. Growth of a thermotolerant ethanol-producing strain of *Kluyveromyces marxianus* on cellobiose containing media. Biotechnol. Lett. 16, 625–630. doi:10.1007/BF00128612

Buenrostro-Figueroa, J., Tafolla-Arellano, J.C., Flores-Gallegos, A.C., Rodríguez-Herrera, R., De la garza-Toledo, H., Aguilar, C.N., 2017. Native yeasts for alternative utilization of overripe mango pulp for ethanol production. Rev. Argent. Microbiol. doi:10.1016/j.ram.2016.04.010

Chang, J. J., Lin, Y. J., Lay, C. H., Thia, C., Wu, Y.-C., Hou, Y. H., Huang, C. C., Li, W.-H., 2018. Constructing a cellulosic yeast host with an efficient cellulase cocktail. Biotechnol. Bioeng. 115, 751–761. doi:10.1002/bit.26507

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., da Silveira, W.B., 2017. Transcriptome analysis of the thermotolerant yeast *Kluyveromyces marxianus* CCT 7735 under ethanol stress. Appl. Microbiol. Biotechnol. 101, 6969–6980. doi:10.1007/s00253-017-8432-0

Ferreira, R., Skrekas, C., Nielsen, J., David, F., 2018. Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation Using Csy4 in *Saccharomyces cerevisiae*. doi:10.1021/acssynbio.7b00259

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. doi:10.1007/s00253-008-1458-6

Gao, J., Feng, H., Yuan, W., Li, Y., Hou, S., Zhong, S., Bai, F., 2017a. Enhanced fermentative performance under stresses of multiple lignocellulose-derived inhibitors by overexpression of a typical 2-Cys peroxiredoxin from *Kluyveromyces marxianus*. Biotechnol. Biofuels 10, 79. doi:10.1186/s13068-017-0766-4

Gao, J., Yuan, W., Li, Y., Bai, F., Jiang, Y., 2017b. Synergistic effect of thioredoxin and its reductase from *Kluyveromyces marxianus* on enhanced tolerance to multiple lignocellulose-derived inhibitors. Microb. Cell Fact. 16, 181. doi:10.1186/s12934-017-0795-5

Gao, J., Yuan, W., Li, Y., Xiang, R., Hou, S., Zhong, S., Bai, F., 2015. Transcriptional analysis of *Kluyveromyces marxianus* for ethanol production from inulin using consolidated bioprocessing technology. Biotechnol. Biofuels 8, 115. doi:10.1186/s13068-015-0295-y

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, 5351–5358.

Hassan, N., Nguyen, T.-H., Intanon, M., Kori, L.D., Patel, B.K.C., Haltrich, D., Divne, C., Tan, T.C., 2015. Biochemical and structural characterization of a thermostable β-glucosidase from *Halothermothrix orenii* for galacto-oligosaccharide synthesis. Appl. Microbiol. Biotechnol. 99, 1731–1744. doi:10.1007/s00253-014-6015-x

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat.

Biotechnol. 31, 827–832. doi:10.1038/nbt.2647

Hughes, S.R., Qureshi, N., López-Núñez, J.C., Jones, M.A., Jarodsky, J.M., Galindo-Leva, L.Á., Lindquist, M.R., 2017. Utilization of inulin-containing waste in industrial fermentations to produce biofuels and bio-based chemicals. World J. Microbiol. Biotechnol. 33, 78. doi:10.1007/s11274-017-2241-6

Islam, S.M.M., Elliott, J.R., Ju, L.-K., 2018. Minimization of fermentation inhibitor generation by carbon dioxide-water based pretreatment and enzyme hydrolysis of guayule biomass. Bioresour. Technol. 251, 84–92. doi:10.1016/j.biortech.2017.12.032

Ivanova, C., Bååth, J.A., Seiboth, B., Kubicek, C.P., 2013. Systems Analysis of Lactose Metabolism in *Trichoderma reesei* Identifies a Lactose Permease That Is Essential for Cellulase Induction. PLoS One 8, e62631. doi:10.1371/journal.pone.0062631

Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S.J., Kristensen, M., Pedersen, L.E., Jensen, M.K., Keasling, J.D., 2015. Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. Metab. Eng. 28, 213–222. doi:10.1016/J.YMBEN.2015.01.008

Kobayashi, Y., Sahara, T., Suzuki, T., Kamachi, S., Matsushika, A., Hoshino, T., Ohgiya, S., Kamagata, Y., Fujimori, K.E., 2017. Genetic improvement of xylose metabolism by enhancing the expression of pentose phosphate pathway genes in *Saccharomyces cerevisiae* IR-2 for high-temperature ethanol production. J. Ind. Microbiol. Biotechnol. 44, 879–891. doi:10.1007/s10295-017-1912-5

Kurtzman, C.P., Fell, J.W., 2000. The yeasts : a taxonomic study. Elsevier.

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*. Antonie Van Leeuwenhoek 100, 507–19. doi:10.1007/s10482-011-9606-x

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Suprayogi, Tsuchikane, K., Limtong, S., Fujita, N., Yamada, M., 2015. Genetic basis of the highly efficient yeast

Kluyveromyces marxianus: complete genome sequence and transcriptome analyses. Biotechnol. Biofuels 8, 47. doi:10.1186/s13068-015-0227-x

Li, C., Lin, F., Zhou, L., Qin, L., Li, B., Zhou, Z., Jin, M., Chen, Z., 2017. Cellulase hyper-production by *Trichoderma reesei* mutant SEU-7 on lactose. Biotechnol. Biofuels 10, 228. doi:10.1186/s13068-017-0915-9

Lin, Y.-J., Chang, J.-J., Lin, H.-Y., Thia, C., Kao, Y.-Y., Huang, C.-C., Li, W.-H., 2017. Metabolic engineering a yeast to produce astaxanthin. Bioresour. Technol. 245, 899–905. doi:10.1016/j.biortech.2017.07.116

Liu, J.-J., Zhang, G.-C., Oh, E.J., Pathanibul, P., Turner, T.L., Jin, Y.-S., 2016. Lactose fermentation by engineered *Saccharomyces cerevisiae* capable of fermenting cellobiose. J. Biotechnol. 234, 99–104. doi:10.1016/j.jbiotec.2016.07.018

Löbs, A.-K., Engel, R., Schwartz, C., Flores, A., Wheeldon, I., 2017a. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol. Biofuels 10, 164. doi:10.1186/s13068-017-0854-5

Lodi, T., Donnini, C., 2005. Lactose-induced cell death of β -galactosidase mutants in. FEMS Yeast Res. 5, 727–734. doi:10.1016/j.femsyr.2005.01.005

Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G.A., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., Daran, J.-M.G., 2015. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. 15. doi:10.1093/femsyr/fov004

Morrissey, J.P., Etschmann, M.M.W., Schrader, J., Billerbeck, G.M., 2015. Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. Yeast 32, 3–16. doi:10.1002/YEA.3054

Nambu-Nishida, Y., Nishida, K., Hasunuma, T., Kondo, A., 2017. Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. Sci. Rep. 7, 8993. doi:10.1038/s41598-017-08356-5

Naumov, G.I., 2008. Identification of the lactose *LAC* gene superfamilies in *Kluyveromyces yeast*. Dokl. Biochem. Biophys. 420, 158–60.

Naumov, G.I., Naumova, E.S., Barrio, E., Querol, A., 2006. Genetic and molecular study of inability of the yeast *Kluyveromyces lactis* var drosophilarum to ferment lactose. Mikrobiologiia 75, 299–304.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., Shimatani, Z., Kondo, A., 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science. 353, aaf8729-aaf8729. doi:10.1126/science.aaf8729

Pentjuss, A., Stalidzans, E., Liepins, J., Kokina, A., Martynova, J., Zikmanis, P., Mozga, I., Scherbaka, R., Hartman, H., Poolman, M.G., Fell, D.A., Vigants, A., 2017. Model-based biotechnological potential analysis of *Kluyveromyces marxianus* central metabolism. J. Ind. Microbiol. Biotechnol. 44, 1177–1190. doi:10.1007/s10295-017-1946-8

Pilap, W., Thanonkeo, S., Klanrit, P., Thanonkeo, P., 2018. The potential of the newly isolated thermotolerant *Kluyveromyces marxianus* for high-temperature ethanol production using sweet sorghum juice. 3 Biotech 8, 126. doi:10.1007/s13205-018-1161-y

Poulos, J., Farnia, A., 2016. Production of cannabinoids in yeast. U.S. Patent No. 14,795,816. Washington, DC: U.S. Patent and Trademark Office.

Rouwenhorst, R.J., Hensing, M., Verbakel, J., Scheffers, W.A., van Duken, J.P., 1990. Structure and properties of the extracellular inulinase of *Kluyveromyces marxianus* CBS 6556. Appl. Environ. Microbiol. 56, 3337–45.

Sadie, C.J., Rose, S.H., den Haan, R., van Zyl, W.H., 2011. Co-expression of a cellobiose phosphorylase and lactose permease enables intracellular cellobiose utilisation by *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 90, 1373–1380. doi:10.1007/s00253-011-3164-z

Schabort, D.T.W.P., Kilian, S.G., du Preez, J.C., 2018. Gene regulation in *Kluyveromyces marxianus* in the context of chromosomes. PLoS One 13, e0190913. doi:10.1371/journal.pone.0190913

Schwartz, C.M., Hussain, M.S., Blenner, M., Wheeldon, I., 2016. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. ACS Synth. Biol. 5, 356–359. doi:10.1021/acssynbio.5b00162

Sekar, R., Shin, H.-D., Chen, R., 2012. Engineering *Escherichia coli* cells for cellobiose assimilation through a phosphorolytic mechanism. Appl. Environ. Microbiol. 78, 1611–4. doi:10.1128/AEM.06693-11

Signori, L., Passolunghi, S., Ruohonen, L., Porro, D., Branduardi, P., 2014. Effect of oxygenation and temperature on glucose-xylose fermentation in *Kluyveromyces marxianus* CBS712 strain. Microb. Cell Fact. 13, 51. doi:10.1186/1475-2859-13-51

Skory, C.D., 2003. Lactic acid production by *Saccharomyces cerevisiae* expressing a Rhizopus oryzae lactate dehydrogenase gene. J. Ind. Microbiol. Biotechnol. 30, 22–27. doi:10.1007/s10295-002-0004-2

Swiat, M.A., Dashko, S., den Ridder, M., Wijsman, M., van der Oost, J., Daran, J.-M., Daran-Lapujade, P., 2017. FnCpf1: a novel and efficient genome editing tool for *Saccharomyces cerevisiae*. Nucleic Acids Res. 45, 12585–12598. doi:10.1093/nar/gkx1007

Turner, T.L., Zhang, G.-C., Kim, S.R., Subramaniam, V., Steffen, D., Skory, C.D., Jang, J.Y., Yu, B.J., Jin, Y.-S., 2015. Lactic acid production from xylose by engineered *Saccharomyces cerevisiae* without PDC or ADH deletion. Appl. Microbiol. Biotechnol. 99, 8023–33. doi:10.1007/s00253-015-6701-3

Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S., Roubos, J.A., 2018. CRISPR/Cpf1 enables fast and simple genome editing of *Saccharomyces cerevisiae*. doi:10.1002/yea.3278

Zhang, B., Zhu, Y., Zhang, J., Wang, D., Sun, L., Hong, J., 2017. Engineered *Kluyveromyces marxianus* for pyruvate production at elevated temperature with simultaneous consumption of xylose and glucose. Bioresour. Technol. 224, 553–562. doi:10.1016/j.biortech.2016.11.110