

Diploid genome assembly of *Kluyveromyces marxianus* NRRL Y-50883 (SLP1)

Carolina Gómez-Márquez ¹, Dania Sandoval-Nuñez ², Anne Gschaedler ², Teresa Romero-Gutiérrez ¹, Lorena Amaya-Delgado ^{2,*} and J. Alejandro Morales ^{1,*}

¹Departamento de Ciencias Computacionales, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara 44430, México, and ²Biocología Industrial, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco A.C, Zapopan 45019, México

*Corresponding author: Email: jalejandro.morales@academicos.udg.mx (J.A.M); lamaya@ciatej.mx (L.A-D)

Abstract

The yeast *Kluyveromyces marxianus* SLP1 has the potential for application in biotechnological processes because it can metabolize several sugars and produce high-value metabolites. *K. marxianus* SLP1 is a thermotolerant yeast isolated from the mezcal process, and it is tolerant to several cell growth inhibitors such as saponins, furan aldehydes, weak acids, and phenolics compounds. The genomic differences between dairy and nondairy strains related to *K. marxianus* variability are a focus of research attention, particularly the pathways leading this species toward polyploidy. We report the diploid genome assembly of *K. marxianus* SLP1 nonlactide strain into 32 contigs to reach a size of ~12 Mb (N50 = 1.3 Mb) and a ~39% GC content. Genome size is consistent with the *k*-mer frequency results. Genome annotation by Funannotate estimated 5000 genes in haplotype A and 4910 in haplotype B. The enriched annotated genes by ontology show differences between alleles in biological processes and cellular component. The analysis of variants related to DMKU3 and between haplotypes shows changes in *LAC12* and *INU1*, which we hypothesize can impact carbon source performance. This report presents the first polyploid *K. marxianus* strain recovered from nonlactic fermenting medium.

Keywords: *Kluyveromyces marxianus*; SLP1; non-*Saccharomyces* yeast; complete genome; gene annotation; nondairy yeast

Introduction

The yeast *Kluyveromyces marxianus* is emerging as a remarkable alternative for fermentation processes, metabolic engineering, omics analysis, and recombinant protein production, having several advantages over other yeasts. These include the ability to metabolize a wide variety of sugars, such as glucose, fructose, fructans, xylose, and lactose, which are abundant sugars in nature (i.e., agave juice, lignocellulosic biomass, whey, and molasses) (Fonseca et al. 2008; Lane and Morrissey 2010; Amaya-Delgado et al. 2013; Flores et al. 2013; Schabert et al. 2018).

K. marxianus is thermotolerant and shows rapid cellular growth even when cultured with growth-inhibitor molecules (Li et al. 2018; Flores-Cosío et al. 2019). Previous studies of this yeast highlighted the significant physiological variation amongst *K. marxianus* strains in terms of their fermentation characteristics and ability to use various substrates, suggesting genetic diversity within the species (Flores et al. 2013; Ortiz-Merino et al. 2018).

Our strain, *K. marxianus* NRRL Y-50883 (SLP1), was isolated from hostile fermentation musts during the mezcal fermentation process in San Luis Potosí, México, and belongs to the yeast collection of the Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco (CIATEJ, A.C). *K. marxianus* SLP1 is a strain with interesting characteristics because it produces ethanol from mixtures of glucose-xylose, and agave fructans and fructose, while producing several industry-desirable metabolites

such as fructanases, aromatic compounds, and higher alcohols (Flores et al. 2013; Flores-Cosío et al. 2018).

Compared to other yeasts, SLP1 produces a significant amount of ethanol, derived from different lignocellulosic hydrolysates and agave juice for tequila production (Flores et al. 2013; Sandoval-Nuñez et al. 2018; Alcazar-Valle et al. 2019). SLP1 can tolerate high concentrations of organic acids, furans, phenols, oxidant agents, and saponins, which are considered growth inhibitors, as well as high temperatures. This outstanding tolerance is related to the physiological ability of SLP1 to change the cell membrane composition to increase its robustness under unfavorable fermentation conditions (Arellano-Plaza et al. 2013; Alcázar et al. 2017; Mejía-Barajas et al. 2017, 2018; Flores-Cosío et al. 2019).

To understand the genomic differences of *K. marxianus* SLP1 and other strains of this species that can be isolated, for example, from dairy products, it is important to know the genetic features that provide thermotolerance, sugar transporter, and other traits that confer the most significant qualities of SLP1. Ortiz-Merino et al. (2018) showed divergence in *K. marxianus* UFS-Y2791 isolated from *Agave Americana* compared to *K. marxianus* isolated from dairy products. This opens the question as to what extent *K. marxianus* genomic variability is associated with nondairy strains.

Here, we report the whole-genome sequencing, assembly, and annotation results of *K. marxianus* SLP1, which allow the

Received: August 9, 2021. Accepted: September 21, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

identification of the genomic machinery available in this strain as a product of the adaptive response to a stress phenomenon and the evaluation of its impact on this mechanism through gene identification, enrichment by ontology, and variant classification. The results of this work allow us to study the dynamics of adaptation and provide knowledge on nonlactic *K. marxianus* that show diploidy.

Materials and methods

Yeast strain and library preparation

The yeast *K. marxianus* (NRRL Y-50883) was grown overnight at 30°C at 250 rpm in 50 mL of YPD medium to the early stationary phase before cells were harvested by centrifugation. Total genomic DNA was extracted using the Promega Wizard® Genomic DNA Purification Kit Technical Manual according to the manufacturer's instructions. DNA quality was verified using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA) and verifying its integrity in agarose gel.

PacBio sequencing

The sequencing was performed by Macrogen Inc, Korea. Whole-genome *de novo* sequencing was performed by PacBio using a SMRT Cell 8Pac V3 with DNA Polymerase Binding Kit P6 and DNA Sequencing Reagent 4.0 v2. The reads were processed by the parallel statistical analysis of the k-mer spectrum in sequencing Jellyfish data (Marçais and Kingsford 2011), which counts the occurrence of all 21-mer length fragments. Once the raw data were processed, they were moved to GenomeScope 2.0, a reference-free analyzer used to establish genome profiling (Ranallo-Benavidez et al. 2020). PacBio performed exploration with both tools, and generated outputs estimates for size, repetitiveness, and polyploid (through heterozygosity rate).

Genome assembly

K. marxianus SLP1 raw reads were assembled using Falcon software via Bioconda (<https://github.com/PacificBiosciences/pb-assembly>). The first Falcon module was run in three steps: (1) overlap detection and error correction of raw reads, (2) overlap detection of corrected reads, and (3) string graph assembly. Following, to extract haplotig fragments, we ran the Falcon-Unzip package with the previously generated data. The two main stages for the latter were: (1) identifying SNPs and assigning phases and (2) assembling the graph annotation with the phases. Falcon-Unzip was run with the parameters adjusted for yeast sequences. After, we generated two types of fasta files: primary contigs and haplotigs. Primary contigs are the longest continuous stretches of contiguously assembled sequences, and haplotigs are contigs from a specific haplotype (Chin et al. 2016).

Quality assembly

The assembly quality was assessed using the QUAST 5.1 (<https://github.com/ablab/quast>) algorithm. We calculated basic metrics (L50, N50, and number of sequences, amongst others) and included NG50 and L50 using a reference genome *K. marxianus* DMKU3 (*K. marxianus* GCA_001417885.1). The genome completeness was determined using Benchmarking Universal Single-Copy Orthologs (BUSCO) software v4.1.4 with the Saccharomycetes ortholog gene set from OrthoDB in terms of found single-copy orthologs in the genome by researching the absence or presence of highly conserved genes, which are reported as complete if the length is two standard deviations of BUSCO groups, duplicated if

a complete gene is found with more than one copy, fragmented if is partially recovered, and missing if is not recovered (Simão et al. 2015).

Variant identification and quality control

We also aligned the contigs with the reference genome to find structural variants with the MUMmer package nucmer. Sequences were aligned with MUMmer 3.23 using a reference genome (*K. marxianus* DMKU3-1043 with genome assembly identifier: GCA_001417885.1), and by calling the SNP positions, the output file was converted to a VCF file. Next, variant annotation was analyzed by SNPeff 5.0 using the public DMKU3-1042 genome annotation as a reference. In addition, we aligned the amino acid sequences of the SLP1 alleles to two genes using the InterProScan web service (<https://www.ebi.ac.uk/interpro/search/sequence/>): LAC12 with strain CBS 397 for its efficient lactose consumption, and *INU1* with strain DMKU3 for sequence similarity.

Genome annotation

We separated the *K. marxianus* SLP1 diploid genome into haplotypes to identify both sub-genomes. After the assembly, we assigned primary contigs as haplotype A. Then, we designed a Python (3.8) script that takes the coordinate data of each haplotig produced by Falcon, replacing the sequences in the primary contig to generate the haplotype B.

Gene prediction and annotation were conducted in both haplotypes with Funannotate pipeline v. 1.7.0 (Palmer 2016), which includes softmasking with RepeatMasking and RepeatModeler tools, an *ab initio* gene predictor (Augustus, snap, glimmerHMM, codingQuarry, and GenMark), and functional annotation of the protein-coding genes. Then, we assessed the ontology enrichment of genes categorized by the PANTHER web server (<http://pantherdb.org/>) into the three main ontologies: molecular function, biological process, and cellular component. An NJ phylogenetic tree was generated with CLC Sequence Viewer 8.0 to establish taxonomic distance using the *K. marxianus* SLP1 fatty acid synthase subunit beta (*FAS1*) gene. The *FAS1* gene is present during different stress responses, considered a wide-range fermentative yeasts taxonomic marker (Roscini et al. 2020) and was proven to be essential in the adaptive evolution of *K. marxianus*.

Results and discussion

Sequencing and size estimation

Sequencing was performed on the PacBio RS System using Single Molecule Real Time (SMRT), generating 82,360 reads with an average length 10,384 bp for a total length of >855 Mb with N50 = 16,868 b.

Pacbio reads showed a high heterozygosity rate, consistent with our previous analysis of *K. marxianus* SLP1 k-mer frequency in sequencing data produced by the Illumina platform (data not shown). Figure 1 shows the k-mer profile results from GenomeScope. There are two major peaks between ~30 and 55. The first peak is higher than the second; this is an indicator of high heterozygosity. This evidence is consistent with the findings of a diploid genome (Ranallo-Benavidez et al. 2020). The size estimation obtained by GenomeScope is 11.9 Mb, with a uniqueness of 88.33% and heterozygosity of 3.27%. After k-mer profile analysis, 70× coverage of the estimated 11.9 Mb *K. marxianus* SLP1 genome was calculated.

Kluyveromyces marxianus assembly and quality measures

The *K. marxianus* SLP1 diploid genome was assembled into 32 primary contigs covering 12.04 Mb with N50 = 1.3 Mb and L50 = 4,

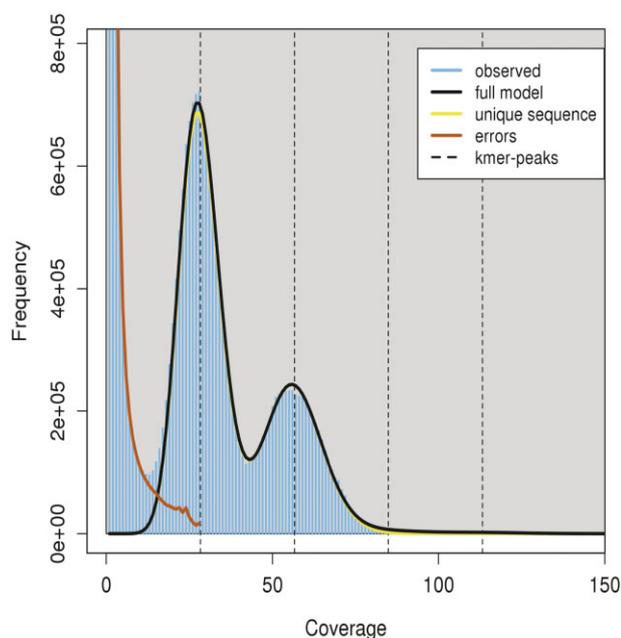


Figure 1 k-mer frequency of PacBio raw reads. Plot obtained by GenomeScope using $k = 21$. The fit of the GenomeScope model (black line) and the k-mer frequencies observed (blue area).

with the same values for NG50 and LG50. We also identified 38 haplotigs covering 955,365bp with N50 = 368,823bp, NG50 = 361,944 bp, and L50 = 9, with the same value for LG50, which means that our contigs are same length and quantity in terms of genome average in comparison to *K. marxianus* DMKU3. The smallest fragment in the primary contigs is 20,533 and 3040 bp in haplotigs as we show in Table 1.

The completeness of the genome as determined by BUSCO software is shown in Figure 2, where haplotypes A and B have 14 and 61 genes, respectively, appearing as duplicated copies due to the divergence between haplotypes. Conversely, 60% of the genes are complete and fragmented, while 40% are missing. Missing genes may be due to either no significant matches, or the BUSCO matches scored below the range of scores for the BUSCO profile. The missing genes could also be due to the assembler used (Falcon) and its limitation with mitochondrial sequences as reported by Hamlin et al. (2019). The assembly completeness in

Table 1 *K. marxianus* SLP1 summary assembly. Statistics of primary contigs and haplotigs. NG50 and LG50 were calculated using *K. marxianus* DMKU3 as the reference.

	Primary contigs	Haplotigs
Sequences	32	38
GC %	39.91	39.13
N50	1,343,707 bp	368,823 bp
NG50	1,343,707 bp	361,944 bp
L50	4	9
LG50	4	9
Assembly length	12,044,577 bp	955,365 bp

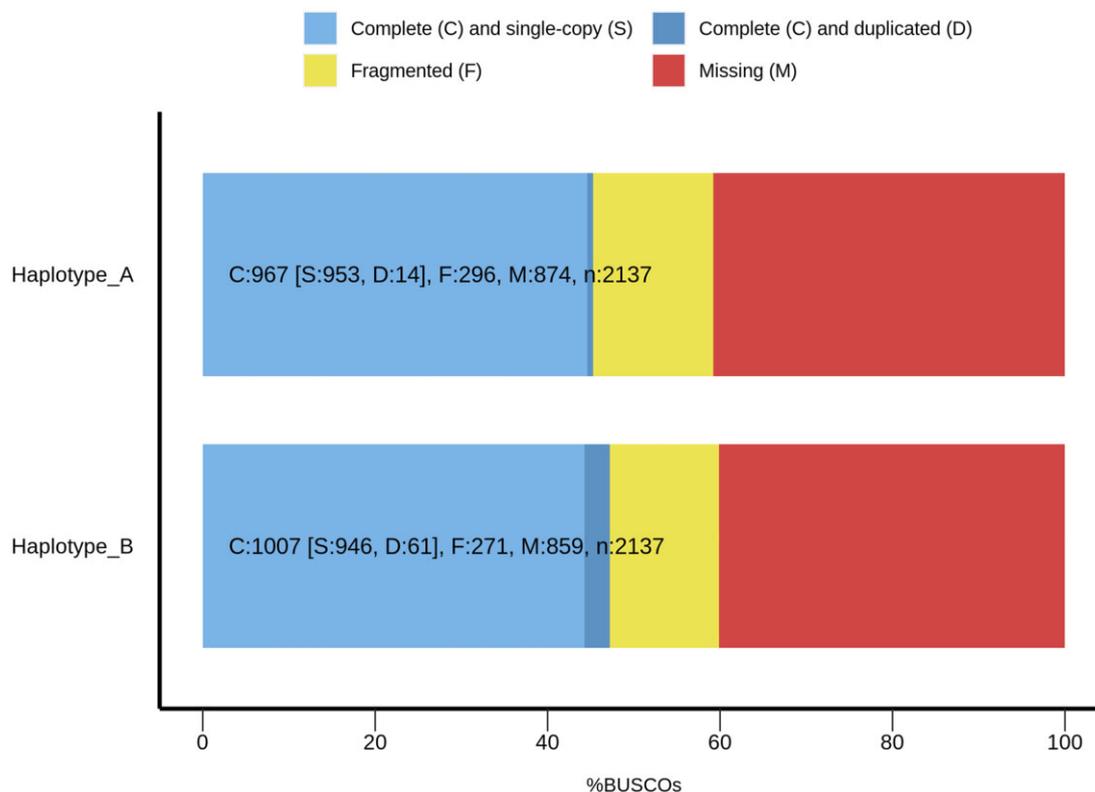


Figure 2 BUSCO summary of *K. marxianus* SLP1 haplotypes. Haplotype A and haplotype B of *K. marxianus* SLP1 with Saccharomycetes orthology data from orthoDB represented by the number of universal single-copy reference orthologues (BUSCOs). The blue and dark blue areas represent the number of complete and single-copy genes, and complete and duplicated genes, respectively; yellow and red indicate the number of fragmented and missing genes, respectively; and n is the number of genes used.

Table 2 *K. marxianus* SLP1 haplotype annotation summary

	Haplotype	
	A	B
Genes	5000	4910
CDS	4732	4646
tRNA	268	264

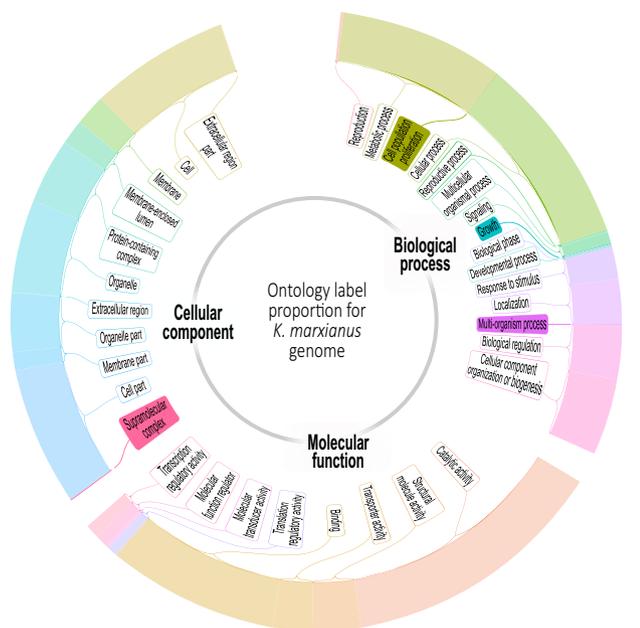


Figure 3 Ontology enrichment. The bar graphs surrounding the plot circumference are represented by the relative frequency of genes in each ontology term. Nonfilled boxes are ontologies present in both haplotypes while color-filled boxes represent functions present in only one haplotype.

Table 3 Variant annotation summary. High, frameshift variant; low, synonymous variant; moderate, missense variant; modifier, up- and downstream gene variant.

Chromosome	Gene	Coordinate	High	Low	Moderate	Modifier
1	INU1	1,092,964 to 1,098,595	11	5	1	198

the gene content term can be used to measure the quality of the assembly, but it is for known genes that are incorporated in databases, and high-quality genomes may reflect the organism’s evolutionary history rather than an incomplete assembly.

Kluyveromyces marxianus annotation

Funannotate pipeline reported 4732 predicted CDS in haplotype A and 4646 in haplotype B. The summary annotation is shown in Table 2, which lists the number of genes, CDS, and tRNAs. Each haplotype was annotated separately; the results show 1429 gene IDs in haplotype A and 1341 gene IDs in haplotype B. The remaining CDS are reported as hypothetical proteins, so we decided to work only with CDS with known function.

The enrichment by ontologies in annotated genes in Table 2 is presented by genes in each haplotype, as we described in the Methods section. Figure 3 shows the ontologies present in our yeast strain where some functions are different between haplotypes. Such gained functions could be the consequence of the characteristic variability generated during sexual reproduction that occurs -in the case of *K. marxianus*- under stressful conditions. The genome contains 187,219 SNPs, 36,674 INS, and 70,565 DEL in haplotype A, and 210,342 SNPs, 41,416 INS, and 68,323 DEL in haplotype B, as reported by Snpeff. In addition, we report the effect of variants at the *INU1* gene because its influence on fructan assimilation is a trait of this strain. Table 3 shows 198 variants located in the up- and down-stream regions have an impact as modifiers. However, there are variants have high, low, and moderate impacts. Those with high impact are frameshift variants caused by INDEL, those with moderate impact are missense variants, and those with low impact are synonymous variants. The phylogenetic tree (Supplementary Figure S1) shows the distances between *Saccharomyces cerevisiae*, *K. marxianus*, *Kluyveromyces lactis*, and other related species.

Relevance

This is the first report of a diploid *K. marxianus* living in a fermenting agave must medium. All other polyploid *K. marxianus* have been isolated from lactic mediums, and Ortiz-Merino et al. (2018) demonstrated that *LAC12* (functional and nonfunctional) gene alleles are associated with their chromosome set number. Mo et al. (2019) showed that *K. marxianus* *FIM1*, a haploid yeast isolated from nonlactic environments, was highly stable when grown in ethanol-rich medium; our report underlies the need to further study the mechanisms of adaptive polyploidy evolution in fructan-rich growing-medium *K. marxianus*. Such evidence

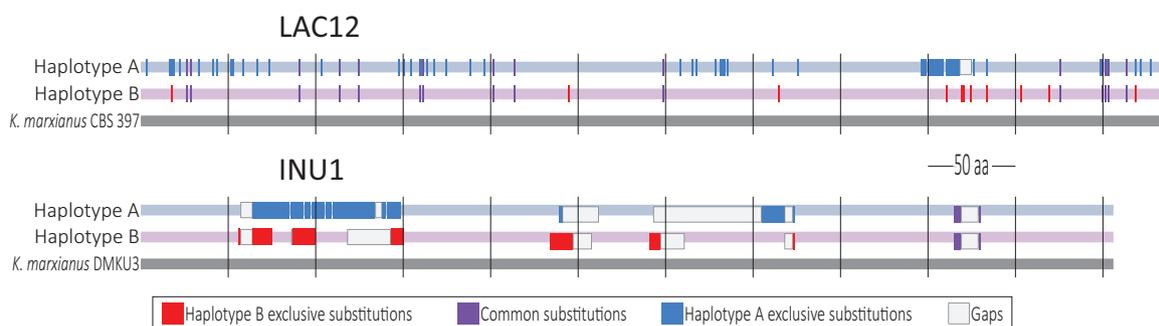


Figure 4 Alignment scheme of *LAC12* and *INU1* genes. *LAC12* amino acid sequence alignment representation from *K. marxianus* CBS 397 to SLP1 haplotypes. *INU1* amino acid sequence alignment representation from *K. marxianus* DMKU3 to SLP1 haplotypes. Common substitutions are represented by purple lines, red lines indicate haplotype B exclusive substitutions, and blue lines indicate haplotype A exclusive substitutions. Gaps in the alignment are indicated by white boxes.

opens the question if using other carbon-sugar sources is related to the ploidy phenomenon. The ability of *K. marxianus* SLP1 to consume a variety of substrates and to cope with hostile environments suggests a sexual recombination event that allowed functional gene gain in this strain. Figure 4 depicts the alignment representation of the LAC12 gene from *K. marxianus* CBS 397 to *K. marxianus* SLP1 haplotypes. We observe that the sequence of the haplotype A contains the largest number of exclusive substitutions, whereas haplotype B shows the highest homology with the LAC12 reference gene. However, we observe several exclusive substitutions and gaps in both haplotypes compared to the *INU1* gene from *K. marxianus* DMKU3.

Data availability

Raw reads are available in the National Center for Biotechnology Information (NCBI) with SRA accession number SRR14000474; the genome assembly generated in this study was deposited under BioProject ID PRJNA714250.

Supplementary material is available at G3 online.

Funding

Financial support is kindly acknowledged from the Energy Sustainability Fund 2014 (CONACYT-SENER, Mexico) Grants 245750. Sandoval-Nuñez D. and Gómez-Márquez C. acknowledge financial support from CONACYT in the form of Ph.D. scholarships 487000 and 443922, respectively.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

- Alcázar M, Kind T, Gschaedler A, Silveria M, Arrizon J, et al. 2017. Effect of steroidal saponins from Agave on the polysaccharide cell wall composition of *Saccharomyces cerevisiae* and *Kluyveromyces marxianus*. *LWT*. 77:430–439. doi:10.1016/j.lwt.2016.11.048.
- Alcazar-Valle M, Gschaedler A, Gutierrez-Pulido H, Arana-Sanchez A, Arellano-Plaza M. 2019. Fermentative capabilities of native yeast strains grown on juices from different Agave species used for tequila and mezcal production. *Braz J Microbiol*. 50:379–388. doi:10.1007/s42770-019-00049-7.
- Amaya-Delgado L, Herrera-López EJ, Arrizon J, Arellano-Plaza M, Gschaedler A. 2013. Performance evaluation of *Pichia kluyveri*, *Kluyveromyces marxianus* and *Saccharomyces cerevisiae* in industrial tequila fermentation. *World J Microbiol Biotechnol*. 29:875–881. doi:10.1007/s11274-012-1242-8.
- Arellano-Plaza M, Gschaedler-Mathis A, Noriega-Cisneros R, Clemente-Guerrero M, Manzo-Ávalos S, et al. 2013. Respiratory capacity of the *Kluyveromyces marxianus* yeast isolated from the mezcal process during oxidative stress. *World J Microbiol Biotechnol*. 29:1279–1287. doi:10.1007/s11274-013-1291-
- Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 13:1050–1054. doi:10.1038/nmeth.4035.
- Flores JA, Gschaedler A, Amaya-Delgado L, Herrera-López EJ, Arellano M, et al. 2013. Simultaneous saccharification and fermentation of *Agave tequilana* fructans by *Kluyveromyces marxianus* yeasts for bioethanol and tequila production. *Bioresour Technol*. 146:267–273. doi:10.1016/j.biortech.2013.07.078.
- Flores-Cosio G, Arellano-Plaza M, Gschaedler A, Amaya-Delgado L, Industrial Biotechnology, CIATEJ 2018. Physiological response to furan derivatives stress by *Kluyveromyces marxianus* SLP1 in ethanol production. *Rev Mex Ing Quim*. 17:189–202. doi:10.24275/uam/izt/dcbi/revmexingquim/2018v17n1/Flores.
- Flores-Cosío G, Herrera-López EJ, Arellano-Plaza M, Gschaedler A, Sanchez A, et al. 2019. Dielectric property measurements as a method to determine the physiological state of *Kluyveromyces marxianus* and *Saccharomyces cerevisiae* stressed with furan aldehydes. *Appl Microbiol Biotechnol*. 103:9633–9642. doi:10.1007/s00253-019-10152-2.
- Fonseca GG, Heinzle E, Wittmann C, Gombert AK. 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl Microbiol Biotechnol*. 79:339–354. doi:10.1007/s00253-008-1458-6.
- Hamlin JAP, Dias GB, Bergman CM, Bensasson D. 2019. Phased diploid genome assemblies for three strains of *Candida albicans* from oak trees. *G3 (Bethesda)*. 9:3547–3554. doi:10.1534/g3.119.400486.
- Lane MM, Morrissey JP. 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.
- Li P, Fu X, Li S, Zhang L. 2018. Engineering TATA-binding protein Spt15 to improve ethanol tolerance and production in *Kluyveromyces marxianus*. *Biotechnol Biofuels*. 11:1–13. doi:10.1186/s13068-018-1206-9.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 27:764–770. doi:10.1093/bioinformatics/btr011.
- Mejía-Barajas JA, Montoya-Pérez R, Salgado-Garciglia R, Aguilera-Aguirre L, Cortés-Rojo C, et al. 2017. Oxidative stress and antioxidant response in a thermotolerant yeast. *Braz J Microbiol*. 48:326–332. doi:10.1016/j.bjm.2016.11.005.
- Mejía-Barajas J, Montoya-Pérez R, Manzo-Avalos S, Cortés-Rojo C, Riveros-Rosas H, et al. 2018. Fatty acid addition and thermotolerance of *Kluyveromyces marxianus*. *FEMS Microbiol Lett*. 365:fny043. doi:10.1093/femsle/fny043.
- Mo W, Wang M, Zhan R, Yu Y, He Y, et al. 2019. *Kluyveromyces marxianus* developing ethanol tolerance during adaptive evolution with significant improvements of multiple pathways. *Biotechnol Biofuels*. 12:1–16. doi:10.1186/s13068-019-1393-z.
- Ortiz-Merino RA, Varela JA, Coughlan AY, Hoshida H, da Silveira WB, et al. 2018. Ploidy variation in *Kluyveromyces marxianus* separates dairy and non-dairy isolates. *Front Genet*. 9:94. doi:10.3389/fgene.2018.00094.
- Palmer JM. 2016. Funannotate: A Genome Annotation and Lightweight Comparative Genomics Pipeline for Eukaryotic Genomes. <https://github.com/nextgenusfs/funannotate>
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 11:1432. doi:10.1038/s41467-020-14998-3
- Roscini L, Conti A, Pierantoni DC, Robert V, Corte L, et al. 2020. Do metabolomics and taxonomic barcode markers tell the same story about the evolution of *saccharomyces sensu stricto* complex in fermentative environments? *Microorganisms*. 8:1242–1218. doi:10.3390/microorganisms8081242.
- Sandoval-Nuñez D, Arellano-Plaza M, Gschaedler A, Arrizon J, Amaya-Delgado L. 2018. A comparative study of lignocellulosic

- ethanol productivities by *Kluyveromyces marxianus* and *Saccharomyces cerevisiae*. *Clean Techn Environ Policy*. 20: 1491–1499. doi:10.1007/s10098-017-1470-6.
- Schabert DTWP, Kilian SG, Du Preez JC. 2018. Gene regulation in *Kluyveromyces marxianus* in the context of chromosomes. *PLoS One*. 13:e0190913. doi:10.1371/journal.pone.019.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM, et al. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31: 3210–3212. doi:10.1093/bioinformatics/btv351.

Communicating editor: G. Brown