Genomics Data 9 (2016) 25–29

Contents lists available at ScienceDirect

Genomics Data

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

Genome annotation of a *Saccharomyces* sp. lager brewer's yeast

Patricia Marcela De León-Medina [a,b,c], Ramiro Elizondo-González [a,b,c], Luis Cástulo Damas-Buenrostro [c], Jan-Maarten Geertman [d], Marcel Van den Broek [e], Luis Jesús Galán-Wong [a], Rocío Ortiz-López [b], Benito Pereyra-Alférez [a,*]

[a] Instituto de Biotecnología, Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León. Pedro de Alba y Manuel L. Barragán S/N, Ciudad Universitaria, San Nicolás de los Garza, Nuevo León 66450, Mexico
[b] Centro de Investigación y Desarrollo en Ciencias de la Salud, Universidad Autónoma de Nuevo León, Avenida Carlos Canseco s/n esquina con Av. Gonzalitos, Mutualismo, Mitras Centro, 64460 Monterrey, Nuevo León, Mexico
[c] Laboratorio de Investigación y Desarrollo, Cervecería Cuauhtémoc Moctezuma S.A. de C.V., Alfonso Reyes Norte Col, Bella Vista, 2202 Monterrey, Nuevo León, Mexico
[d] Heineken Supply Chain, Global Research & Development, 2382 PH Zoeterwoude, The Netherlands
[e] Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands

ARTICLE INFO

Article history:
Received 25 October 2015
Received in revised form 13 February 2016
Accepted 19 May 2016
Available online 21 May 2016

Keywords:
Saccharomyces pastorianus
Lager beer
Lager yeast
Hybrid

ABSTRACT

The genome of lager brewer's yeast is a hybrid, with *Saccharomyces eubayanus* and *Saccharomyces cerevisiae* as sub-genomes. Due to their specific use in the beer industry, relatively little information is available. The genome of brewing yeast was sequenced and annotated in this study. We obtained a genome size of 22.7 Mbp that consisted of 133 scaffolds, with 65 scaffolds larger than 10 kbp. With respect to the annotation, 9939 genes were obtained, and when they were submitted to a local alignment, we found that 53.93% of these genes corresponded to *S. cerevisiae*, while another 42.86% originated from *S. eubayanus*. Our results confirm that our strain is a hybrid of at least two different genomes.

1. Introduction

The brewing process is probably the most ancient type of biotechnology. There is evidence that the production and consumption of beer began in Egypt in the Early Dynastic Period (5500–3100 BCE) [1]. The first fermentation process was termed "high" because yeast floats to the top of the tank at a temperature between 15 °C and 25 °C during production of ale beer with the yeast *Saccharomyces cerevisiae*. Lager yeast did not emerge until the 15th century. This yeast is capable of fermenting at a temperature lower than 10 °C and flocculates at the bottom of the tank. The fermentation is followed by a maturation process called "lagering" (Masschelein, 1986).

The *Saccharomyces sensu stricto* complex includes six parented species: *S. cerevisiae*, *Saccharomyces bayanus*, *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii*, *S. mikatae* and *S. paradoxus* [2]; however, it has been observed that lager-brewing yeast is a hybrid species of two combined genomes of *S. eubayanus* and *S. cerevisiae* [3–10]. This provides an important source of chromosomal rearrangements, leading to the gene number and the size of the complete genome [11–14]. It has been proposed and recently demonstrated that lager yeast is the product of two independent hybridization events that can be divided into two groups: Saaz and Frohberg, or group I and group II, respectively [15–19].

With the use of next generation sequencing (NGS) technologies, such as the Illumina Platform, 40,175 prokaryote and eukaryotes genomes have been reported, including 210 different strains of the *Saccharomyces* complex (http://www.ncbi.nlm.nih.gov/genome/browse/ - revised July 22, 2015).

To obtain a higher level of understanding of the sequenced organism, the data obtained from NGS has been assembled and annotated.

The annotation process consists of identifying the biological characteristics from sequences of the assembly. This can be performed through gene prediction and homologous sequence alignment [20–22].

* Corresponding author at: Pedro de Alba y Manuel L. Barragán S/N, Ciudad Universitaria, San Nicolás de los Garza, Nuevo León 66450, Mexico.
E-mail addresses: bpereyra@gmail.com, benitopereyraal@uanl.edu.mx (B. Pereyra-Alférez).

http://dx.doi.org/10.1016/j.gdata.2016.05.009

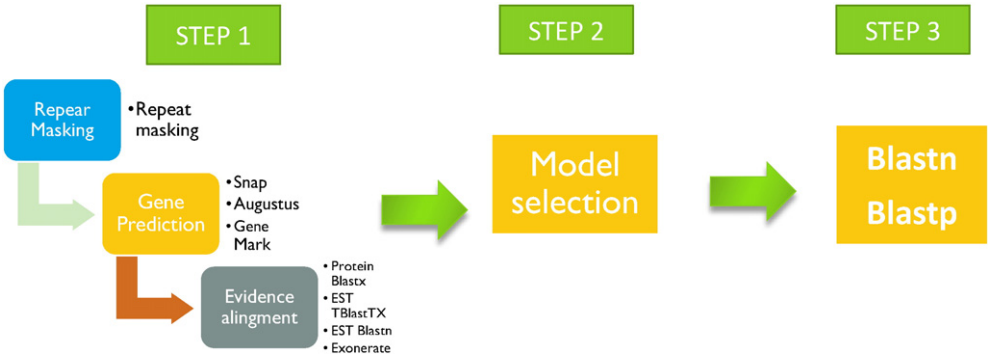

Fig. 1. Maker's annotation pipeline. In step 1, the repetitive sequences are identified and masked; proteins and ESTs are aligned; and using this information, a gene prediction model is made and the structural annotation is completed. Those models are then submitted to a local alignment in Blast for functional annotation.

Here, we present the analysis of the lager yeast genome *Saccharomyces* sp. strain 790 and its comparison with *S. eubayanus* and *S. cerevisiae* S288c. This study provides information about the genome structure of *Saccharomyces* sp. strain 790.

2. Materials and methods

2.1. Strains and sequences

The brewing yeast *Saccharomyces* sp. strain 790 and a reference sequence of 76 scaffolds from *S. eubayanus* were obtained from the yeast collection of Cervecería Cuauhtémoc Moctezuma S.A. de C.V. The *S. cerevisiae* S288c reference genome sequence was retrieved from the yeast genome database (www.yeastgenome.org).

2.2. Sequencing and genome assembly

The brewing yeast genome was sequenced using the FLX 454 Titanium (Roche) and MiSeq (Illumina) massive sequencing platforms according to the manufacturer's protocols. We obtained 0.8 million reads from FLX 454 Titanium (454 Life Sciences, Branfort, CT) with an average size of 400 bp; 6 million pair-end reads from Illumina (Illumina, San Diego, CA) with an average size of 150 bp; 5 million mate-pair reads from Illumina with an insert size of 350 bp and a size of 101 × 2 bp; and 11.7 million mate-pair reads from Illumina with an insert of 8 kb and a size of 51 × 2 bp. Approximately 454 Illumina pair-end reads were assembled with a Newbler DeNovo Assembler (Roche). Contigs were then processed using SSPACE 1.0 software (Boetzer et al. 2011). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LSMH00000000. The version described in this paper is version LSMH01000000.

The sequencing quality data were analyzed with FastQC 0.10.1 software with a value ≥ Q30 [23].

Table 1
Assembly of *Saccharomyces* sp. 790.

Assembled reads	17,034,361
Depth	~70×
Estimated genome size (bp)	22,741,276
Number of scaffolds	133
Average scaffold size (bp)	170,987
Largest scaffold (bp)	1,404,408
N50	568,800
'Ns'	399,699

Likewise, alignments were made against reference sequences (*S. cerevisiae* S288c and *S. eubayanus*) with the MUMmer 3.23 software package [24].

2.3. Genome annotation

The bioinformatics analysis for the annotation was performed with MAKER v2.31.8 software (University of Utah). MAKER is an integrative tool that yields a putative position of the genes. Fig. 1 depicts the annotation steps [22].

Table 2
Size range among scaffolds.

Size level	Number of scaffolds	Bp	Whole genome proportion (%)	Cumulative % of the whole genome
0:999	47	31,168	0.137	0.137
1 k:9999	21	44,610	0.196	0.333
10 k: 99,999	18	1,063,816	4.678	5.011
100 k 199, 999	11	1,608,982	7.075	12.086
200 k: 299,999	6	1,528,906	6.723	18.809
300 k: 399,999	6	2,092,805	9.203	28.01
400 k: 499, 999	4	1,700,637	7.478	35.490
500 k: 599, 999	7	3,759,004	16.529	52.020
600 k: 699,999	5	3,213,208	14.129	66.149
700 k: 799,999	1	712, 480	3.133	69.282
800 k: 899,999	1	806,158	3.545	72.827
900 k: 999,999	5	4,775,094	20.997	93.824
1;404,408	1	1,404,408	6.176	100
Total	133	22,741,276		

Table 3
Sequenced genome assembly level of *Saccharomyces* species.

Organism	Finished genome	Number of chromosomes	Number of scaffolds	Number of contigs	Size (Mpb)
S. cerevisiae	✓	16	17	–	12.16
S. kudriavzevii	✗	16	2054	–	11.19
S. pastorianus	✗	–	–	2425	24.21
S. paradoxus	✗	16	–	832	11.87
S. mikatae	✗	16	–	1648	11.47
S. bayanus	✗	16	–	586	11.87
S. boulardii	✗	16	48	–	11.64
S. arboricola	✓	16	35	–	11.62
S. uvarum	✗	–	–	3985	11.60
S. carlsbergensis	✗	29	77	–	19.37
S. cerevisiae - S. kudriavzevii	✗	–	60	419	23.37
S. pastorianus —S. weihenstephan 34/70	✗	–	–	1358	22.96
790	✗	**~32**	**133**	–	**22.74**

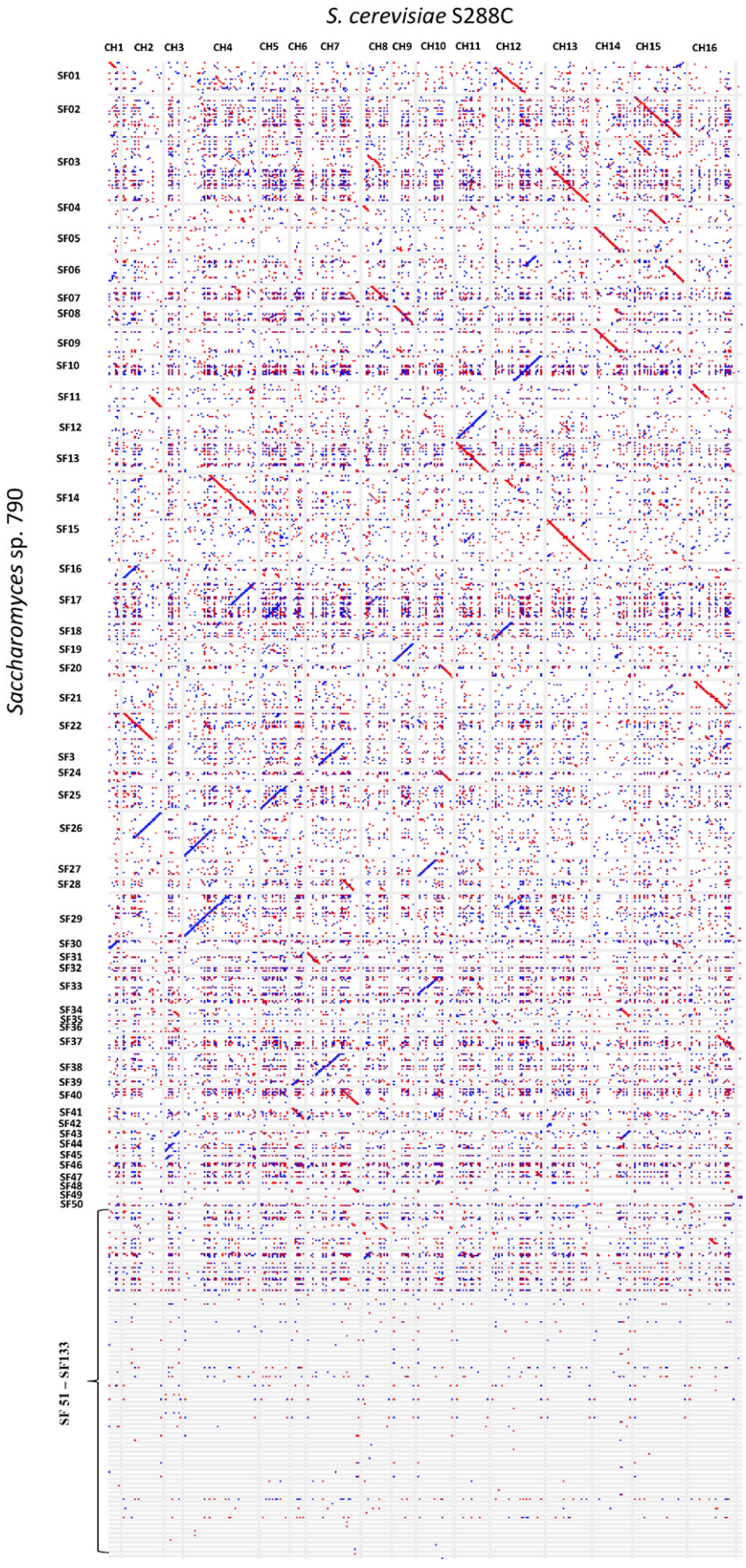

Fig. 2. Alignment dot plot of *Saccharomyces* sp. 790 versus *S. cerevisiae* S288C.

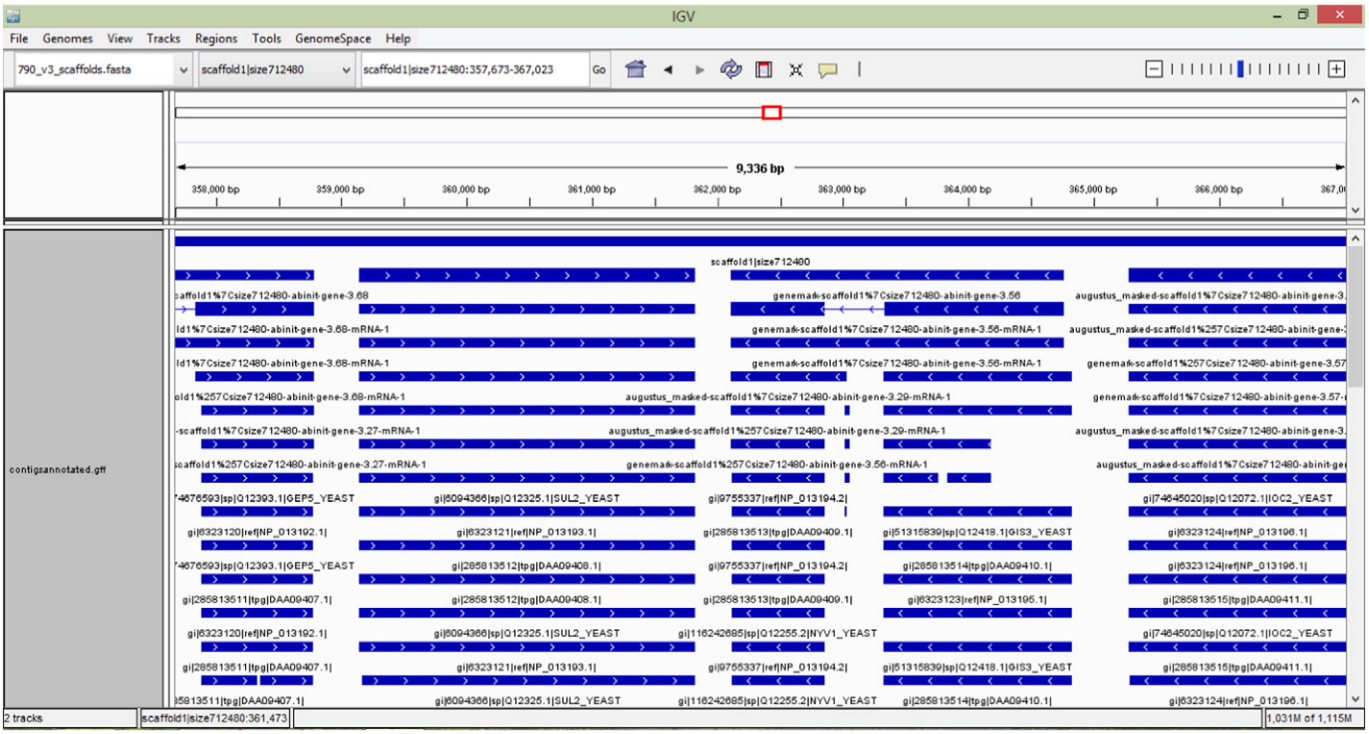

Fig. 3. Visualization of annotations with IGV Tools v2.3 (http://www.broadinstitute.org/software/igv/home). Scaffold 01 alignments and predictions from the Maker's annotation.

3. Results and discussion

The read assembly yielded 133 scaffolds with a ~70× depth and a N50 of 568,800 bp, suggesting a complete genome size of ~22.7 Mbp (Table 1), similar to previous reports of other lager beer yeasts [11,15,25,26]. Approximately 65/133 scaffolds had a size >10 kbp, which represents 99.667% of the assembled genome (Table 2). Table 3 shows a comparison of the assembly level of the sequenced genomes of the *Saccharomyces* species (as of August 2015). It also shows a similar genome size compared to other brewing yeasts [11,26]. The alignments against the reference genome, *S. cerevisiae* S288C, assigned scaffolds to each of its 16 chromosomes, and some scaffolds covered different portions of more than one chromosome; for example, scaffold01 (SF01) aligns with two chromosomes: a small portion with chromosome 1 and with chromosome 12. Scaffold17 (SF17) aligns with chromosomes 4 and 5 (Fig. 2).

The estimated size matches the previous and known information; this is due to the presence of 16 chromosomes of the *S. cerevisiae* subgenome and 16 of *S. eubayanus*. This suggests an overall estimation of 32 chromosomes without considering ploidy. Likewise, its size is close to the sum of the aforementioned genomes (~12 Mbp each). This observation is consistent with the previously reported data by Nakao et al. (2009), Borneman et al. (2011) and Walther et al. (2014), who reported the sequence and assembly of the lager brewing yeast genomes *Saccharomyces carlsbergensis* (78 scaffolds, 29 chromosomes with a 19.5 Mbp length), and *Saccharomyces pastorianus* Weihenstephan 34/70 (985 scaffolds, ~29 chromosomes, and 22.9 Mbp).

The annotation yielded 9939 CDS and a gff file with their locations in the scaffolds of the assembly (Fig. 3). The protein and transcript sequences, were subjected to a local alignment with the Blast tool [27] against a local database using the sequence of *S. cerevisiae* S288C as a reference. The transcripts were considered to be genes because previous reports showed that only approximately 5% of the yeast genome contains introns [28,29].

The scaffolds were classified using the results obtained from Blastn according to the mean identity percentage in all of the genes contained in the same scaffold, as follows (Table 4):

%Id >99.0% and E value $<10^{-6}$ = scaffold belongs to *S. cerevisiae*.

%Id <90.0% and E value $<10^{-6}$ = scaffold does not belong to *S. cerevisiae*.

99.0% > %Id >90.0% and E value $<10^{-6}$ = hybrid scaffold.

Our identity criterion was validated by subjecting the gene sequences from *S. cerevisiae* S288C to a local alignment against *S. eubayanus*, and we found that the %Id between these strains was <90% (Supplementary Table S1) and the average size of the CDS was 1550 bp. Approximately 96.8% of the genome was annotated; 53.93% corresponded to *S. cerevisiae*, 42.86% were non-*cerevisiae* and 3.20% remained un-annotated Fig. 4 (Supplementary Table S2).

4. Conclusions

From the findings in this work, it can be concluded that *Saccharomyces* sp. 790 is a hybrid between *S. cerevisiae* and *S. eubayanus*. Its nuclear genome consists of approximately 32 chromosomes, 16 of which correspond to the *S. cerevisiae* genome and 16 to the *S. eubayanus* genome,

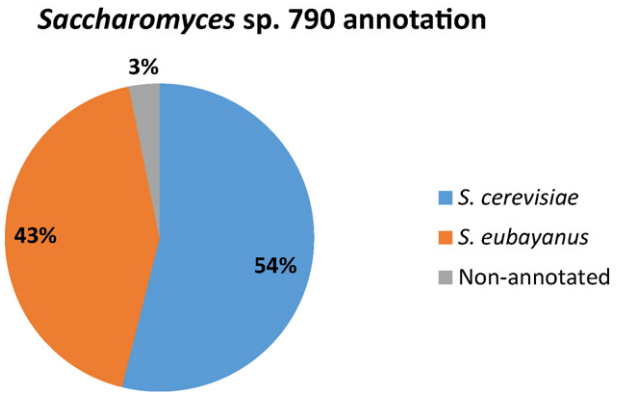

Fig. 4. Schematic representation of *Saccharomyces sp.* 790 annotations. *S. cerevisiae*, *S. eubayanus* and non-annotated proportions are represented in blue, orange and gray, respectively.

without considering ploidy. A total of 133 scaffolds were obtained in the last version of the assembly. Nine scaffolds presented continuous translocations (scaffolds 1, 4, 6, 23 for the *S. cerevisiae* sub-genome and 26, 11, 17, 22 and 32 for the *S. eubayanus* sub-genome), which indicate homologous recombination events. One scaffold presented a possible recombination event (scaffold 3). Data on the chromosome number and size, as well as the number of scaffolds obtained, are consistent with previous reports on lager yeast [6,15,16]. The next step is to improve the assembly with physical mapping techniques.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.05.009.

Acknowledgements

This work was supported, in part, by the Consejo Nacional de Ciencia y Tecnología (CONACYT) of México Grant number: 212773. P. M. De León-Medina and R. Elizondo-González were recipients of a fellowship from CONACYT.

References

[1] I.S. Hornsey, A History of Beer and Brewing. 2003.
[2] E.S. Naumova, S.A. Bulat, N.V. Mironenko, G.I. Naumov, Differentiation of six sibling species in the *Saccharomyces sensu* stricto complex by multilocus enzyme electrophoresis and UP-PCR analysis, Antonie van Leeuwenhoek. Int. J. Gen. Mol. Microbiol. 83 (2003) 155–166.
[3] T. Nilsson-Tillgren, C. Gjermansen, M.C. Kielland-Brandt, J.G.L. Petersen, S. Holmberg, Genetic differences between *Saccharomyces carlsbergensis* and *S. cerevisiae*. Analysis of chromosome III by single chromosome transfer. Carlsberg Res. Commun. 46 (1981) 65–76, http://dx.doi.org/10.1007/BF02906199.
[4] A.V. Martini, C.P. Kurtzman, Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces sensu* stricto. Int. J. Syst. Bacteriol. 35 (1985) 508–511, http://dx.doi.org/10.1099/00207713-35-4-508.
[5] A. Vaughan-Martini, A. Martini, Facts, myths and legends on the prime industrial microorganism. J. Ind. Microbiol. 14 (1995) 514–522, http://dx.doi.org/10.1007/BF01573967.
[6] S. Casaregola, H.V. Nguyen, G. Lapathitis, A. Kotyk, C. Gaillardin, Analysis of the constitution of the beer yeast genome by PCR, sequencing and subtelomeric sequence hybridization. Int. J. Syst. Evol. Microbiol. 51 (2001) 1607–1618 (http://www.ncbi.nlm.nih.gov/pubmed/11491364).
[7] E.S. Naumova, G.I. Naumov, I. Masneuf-Pomarède, M. Aigle, D. Dubourdieu, Molecular genetic study of introgression between *Saccharomyces bayanus* and *S. cerevisiae*. Yeast 22 (2005) 1099–1115, http://dx.doi.org/10.1002/yea.1298.
[8] S. Rainieri, C. Zambonelli, Y. Kaneko, *Saccharomyces sensu* stricto: systematics, genetic diversity and evolution. J. Biosci. Bioeng. 96 (2003) 1–9 (http://www.ncbi.nlm.nih.gov/pubmed/16233475).
[9] S. Rainieri, Y. Kodama, Y. Kaneko, K. Mikata, Y. Nakao, T. Ashikari, Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. Appl. Environ. Microbiol. 72 (2006) 3968–3974, http://dx.doi.org/10.1128/AEM.02769-05.
[10] H.-V. Nguyen, J.-L. Legras, C. Neuvéglise, C. Gaillardin, Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of saccharomyces bayanus strains NBRC1948 and CBS380T. PLoS ONE 6 (2011), e25821http://dx.doi.org/10.1371/journal.pone.0025821.
[11] Y. Nakao, T. Kanamori, T. Itoh, Y. Kodama, S. Rainieri, N. Nakamura, et al., Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res. 16 (2009) 115–129, http://dx.doi.org/10.1093/dnares/dsp003.
[12] A. Querol, U. Bond, The complex and dynamic genomes of industrial yeasts. FEMS Microbiol. Lett. 293 (2009) 1–10, http://dx.doi.org/10.1111/j.1574-6968.2008.01480.x.
[13] K. Krogerus, F. Magalhães, V. Vidgren, B. Gibson, New lager yeast strains generated by interspecific hybridization (2015)http://dx.doi.org/10.1007/s10295-015-1597-6.
[14] M. Hebly, A. Brickwedde, I. Bolat, M.R.M. Driessen, E.A.F. de Hulster, M. van den Broek, et al., *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. FEMS Yeast Res. 15 (2015) fov005, http://dx.doi.org/10.1093/femsyr/fov005.
[15] B. Dunn, G. Sherlock, Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. Genome Res. 18 (2008) 1610–1623, http://dx.doi.org/10.1101/gr.076075.108.
[16] J. Usher, U. Bond, Recombination between homoeologous chromosomes of lager yeasts leads to loss of function of the hybrid GPH1 gene. Appl. Environ. Microbiol. 75 (2009) 4573–4579, http://dx.doi.org/10.1128/AEM.00351-09.
[17] J.C. Wright, D. Sugden, S. Francis-McIntyre, I. Riba-Garcia, S.J. Gaskell, I.V. Grigoriev, et al., Exploiting proteomic data for genome annotation and gene model validation in Aspergillus niger. BMC Genomics. 10 (2009) 61, http://dx.doi.org/10.1186/1471-2164-10-61.
[18] D. Libkind, C.T. Hittinger, E. Valério, C. Gonçalves, J. Dover, M. Johnston, et al., Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 14539–14544, http://dx.doi.org/10.1073/pnas.1105430108.
[19] B.R. Gibson, E. Storgårds, K. Krogerus, V. Vidgren, Comparative Physiology and Fermentation Performance of Saaz and Frohberg Lager Yeast Strains and the Parental Species *Saccharomyces eubayanus*. 2013 255–266, http://dx.doi.org/10.1002/yea.
[20] B.L. Cantarel, I. Korf, S.M.C. Robb, G. Parra, E. Ross, B. Moore, et al., MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18 (2008) 188–196, http://dx.doi.org/10.1101/gr.6743907.
[21] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 12 (2011) 491, http://dx.doi.org/10.1186/1471-2105-12-491.
[22] M. Yandell, D. Ence, A beginner's guide to eukaryotic genome annotation. Nat. Rev. Genet. 13 (2012) 329–342, http://dx.doi.org/10.1038/nrg3174.
[23] S. Andrews, FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/2010.
[24] A.L. Delcher, S.L. Salzberg, A.M. Phillippy, Using MUMmer to identify similar regions in large sequence sets. Curr. Protoc. Bioinformatics. Chapter 10 (2003) Unit 10.3.
[25] A.R. Borneman, B.A. Desany, D. Riches, J.P. Affourtit, A.H. Forgan, I.S. Pretorius, et al., Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of Saccharomyces cerevisiae. PLoS Genet. 7 (2011), e1001287http://dx.doi.org/10.1371/journal.pgen.1001287.
[26] A. Walther, A. Hesselbart, J. Wendland, Genome sequence of Saccharomyces carlsbergensis, the world's first pure culture lager yeast. G3 (Bethesda) 4 (2014) 783–793, http://dx.doi.org/10.1534/g3.113.010090.
[27] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, et al., BLAST+: architecture and applications. BMC Bioinformatics. 10 (2009) 421, http://dx.doi.org/10.1186/1471-2105-10-421.
[28] M. Spingola, L. Grate, D. Haussler, M. Ares, Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA 5 (1999) 221–234http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1369754&tool=pmcentrez&rendertype=abstract.
[29] J. Parenteau, M. Durand, S. Veronneau, A.E.-A. Lacombe, G. Morin, V. Guerin, Deletion of many yeast introns reveals a minority of genes that require splicing for function. Mol. Biol. Cell 19 (2008) 1932–1941, http://dx.doi.org/10.1091/mbc.E07.