

© 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-NonCommercial-NoDerivatives 4.0 International Licence](https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited.

**How to Cite:**

Cintia Gómez-Muñoz, Luis Fernando García-Ortega, Javier Montalvo-Arredondo, Esmeralda Pérez-Ortega, Luis Cástulo Damas-Buenrostro, Lina Riego-Ruiz, Long-insert clone experimental evidence for assembly improvement and chimeric chromosomes detection in an allopolyploid beer yeast, *G3 Genes|Genomes|Genetics*, Volume 11, Issue 7, July 2021, jkab088, <https://doi.org/10.1093/g3journal/jkab088>

# Long-insert clone experimental evidence for assembly improvement and chimeric chromosomes detection in an allopolyploid beer yeast

Cintia Gómez-Muñoz,<sup>1</sup> Luis Fernando García-Ortega,<sup>1,2</sup> Javier Montalvo-Arredondo,<sup>1,3</sup> Esmeralda Pérez-Ortega,<sup>4</sup> Luis Cástulo Damas-Buenrostro,<sup>4</sup> and Lina Riego-Ruiz <sup>1,\*</sup>

<sup>1</sup>División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, A.C., San Luis Potosí 78216, Mexico

<sup>2</sup>Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del IPN, Irapuato 36824, Mexico

<sup>3</sup>Dirección General Académica, Universidad Autónoma Agraria Antonio Narro, Saltillo 25315, Mexico

<sup>4</sup>Cervecería Cuauhtémoc Moctezuma S.A. de C.V., Monterrey 64442, Mexico

\*Corresponding author: Camino a la Presa San José 2055, Col. Lomas 4a. sección C.P. 78216, San Luis Potosí, San Luis Potosí, Mexico. lina@ipicyt.edu.mx

## Abstract

Lager beer is made with the hybrid *Saccharomyces pastorianus*. Many publicly available *S. pastorianus* genome assemblies are highly fragmented due to the difficulties of assembling hybrid genomes, such as the presence of homeologous chromosomes from both parental types, and translocations between them. To improve the assembly of a previously sequenced lager yeast hybrid *Saccharomyces* sp. 790 and elucidate its genome structure, we proposed the use of alternative experimental evidence. We determined the phylogenetic position of *Saccharomyces* sp. 790 and established it as *S. pastorianus* 790. Then, we obtained from this yeast a bacterial artificial chromosome (BAC) genomic library with its BAC-end sequences (BESs). To analyze these data, we developed a pipeline (applicable to other assemblies) that classifies BES pairs alignments according to their orientation. For the case of *S. pastorianus* 790, paired-end BESs alignments validated parts of the assembly and unpaired-end ones suggested contig joins or misassemblies. Importantly, the BACs library was preserved and used for verification experiments. Unpaired-end alignments were used to upgrade the previous assembly and provided an improved detection of translocations. With this, we proposed a genome structure of *S. pastorianus* 790, which was similar to that of other lager yeasts; however, when we estimated chromosome copy number and experimentally measured its genome size, we discovered that one key difference is the outstanding *S. pastorianus* 790 ploidy level (allopolyploid). Altogether, our results show the value of combining bioinformatic analyses with experimental data such as long-insert clone information to improve a short-read assembly of a hybrid genome.

**Keywords:** lager; yeast; hybrid; genome; BAC; assembly; translocations

## Introduction

There are two main types of beers depending on the kind of yeast used for its production. Ale beer is elaborated with some strains of *Saccharomyces cerevisiae*, a well-studied organism (Goffeau *et al.* 1996; Engel *et al.* 2014), whereas lager beer is produced with *Saccharomyces pastorianus* (Martini and Kurtzman 1985; Martini and Martini 1987). Most large breweries utilize lager yeast, hence the importance to study this organism for industrial application. *S. pastorianus* is a hybrid between an *S. cerevisiae* (Engel *et al.* 2014; Goffeau *et al.* 1996) probably of ale origin, and *Saccharomyces eubayanus* (Baker *et al.* 2015; Libkind *et al.* 2011), another closely related wild yeast that has been isolated from multiple locations around the world (Libkind *et al.* 2011; Bing *et al.* 2014; Peris *et al.* 2014; Gayevskiy and Goddard 2016). All of these *S. eubayanus* isolates group into two principal populations potentially originating from Patagonia, one of which the other globally distributed subpopulations derive (Peris *et al.* 2014, 2016; Eizaguirre *et al.* 2018).

Even though the highest number of subpopulations, genetic diversity, and abundance have been found in South America (Eizaguirre *et al.* 2018; Langdon *et al.* 2020; Nespolo *et al.* 2020), based on higher genetic similarities and experimental observations, an Holarctic subpopulation seems to be the most likely parent of the lager beer yeast (Bing *et al.* 2014; Peris *et al.* 2016; Brouwers *et al.* 2019). A third type of beer, belonging to traditional Belgian beers, fermented with *S. cerevisiae* and *Saccharomyces kudriavzevii* hybrids can be considered (González *et al.* 2008; Gallone *et al.* 2019; Langdon *et al.* 2019). Furthermore, other hybrids, even multiple hybrids, have been isolated from different fermentations, such as wine and cider (Masneuf *et al.* 1998; González *et al.* 2006; Peris *et al.*, 2012), evidencing the important role of hybridization for evolution and environment adaptation (Gallone *et al.* 2019; Langdon *et al.* 2019).

The hybrid nature of *S. pastorianus* implies that its genetic content can be segregated in two parental sub-genomes. After the

Received: December 04, 2020. Accepted: March 12, 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-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

initial hybridization event, extensive recombination between the parental sub-genomes occurred. The *S. pastorianus* genome possesses homeologous chromosomes from both parental types and several chimeric chromosomes resulting from homeologous (interparental, Sc-Se or Se-Sc) and nonhomologous (intraparental, Sc-Sc or Se-Se) translocations (Tamai et al. 1998; Dunn and Sherlock 2008). Furthermore, yeasts belonging to the lager hybrid species can be divided into two groups, named group 1 or Saaz and group 2 or Froberg, that have different phenotype, geographical origin, and genetic elements distribution (Liti et al. 2005; Dunn and Sherlock 2008; Gibson et al. 2013). Yeasts from the distinct lager groups are thought to share one common origin due to the presence of conserved recombination points across several group-1 and group-2 strains (Hewitt et al. 2014). Although, the specific origin of each lager group has been highly debated, from hypotheses suggesting one hybridization and later divergence, two independent hybridization events, or more complex phylogenetic stories (Dunn and Sherlock 2008; Hewitt et al. 2014; Okuno et al. 2016; Monerawela and Bond 2017; Salazar et al. 2019). Nonetheless, the type and combination of chromosomes, or genome structure, varies greatly among lager strains, being more notorious between yeasts from different groups (Monerawela and Bond 2017; Gallone et al. 2019). Similarly, the total DNA content, or ploidy level, also fluctuates in lager yeasts due to changes in chromosomes copy number; however, it is generally acknowledged that group-1 yeasts are allotriploids, whereas group-2 are allotetraploids (Walther et al. 2014; van den Broek et al. 2015). The overall genome structure and chromosome copy number is thought to be unique for each lager yeast isolate and responsible for the observed differences in brewing-related traits (van den Broek et al. 2015). Therefore, knowing the genome structure of lager yeasts is important to characterize unstudied or newly isolated strains, to associate their genotype with their phenotype, and to serve as a background for laboratory experiments design, yeast breeding, and selection programs.

Massive parallel or short-reads sequencing allowed to study the genome of several *S. pastorianus* yeast strains (Nakao et al. 2009; Dostálek et al. 2013; Walther et al. 2014; van den Broek et al. 2015; De León-Medina et al. 2016; Okuno et al. 2016; Langdon et al. 2019). The initial sequencing results, or reads, have been aligned to parental-like genomes or genome sequence references to determine the genome structure and chromosome copy number of *S. pastorianus* strains (Hewitt et al. 2014; Walther et al. 2014; van den Broek et al. 2015), based on the principle that the average number of reads covering a nucleotide base, or depth, in a chromosome region is proportional to the number of copies of that region. Changes in sequencing depth across the chromosome indicate recombination sites. However, the detection of recombination sites requires that the involved chimeric chromosomes have a different number of copies; otherwise, they cannot be detected by this method. Furthermore, this process *per se* does not retrieve the chromosome sequence; for that, the reads have also been assembled into longer contiguous sequences or contigs. Nonetheless, most of the publicly available *S. pastorianus* assembly reports consist of hundreds, if not thousands, of contigs (NCBI Genome, <http://www.ncbi.nlm.nih.gov/genome>), and the N50 of the assemblies, a measure of the assemblies' contiguity, is usually smaller than the average size of a yeast chromosome. The above may be due to intrinsic limitations of the used sequencing and assembly technologies, as well as the inherent complexity of assembling hybrid genomes (Pryszcz and Gabaldón 2016). In more detail, short-reads produced by massive parallel sequencing cannot span long DNA repeats, and simultaneously, assembly algorithms fail in

such repeated regions (Schatz et al. 2010; Alkan et al. 2011). Moreover, highly similar intraparental sequences from nonrecombined and chimeric chromosomes are assembled into one consensus contig, and sequences flanking recombination points are not linked, producing fragmented assemblies (Pryszcz and Gabaldón 2016). Such assemblies contain small contigs difficult to scrutinize and use to design experiments.

One method used to order contigs in *S. pastorianus* assemblies is to utilize synteny, or whole genome alignments, with the parental genomes solely (Walther et al. 2014). Nonetheless, ordering contigs using a reference disregard the *S. pastorianus* heterozygosity. Another method used to upgrade lager yeast assemblies is scaffolding (van den Broek et al. 2015; Okuno et al. 2016). This process uses specially paired massive parallel reads, or mate pairs, separated by a fixed DNA length, or insert. Then the scaffolding orders and orients contigs to produce even longer assembled sequences called scaffolds. However, the mate pair sequencing libraries possess short inserts that may not span the aforementioned long repeated regions, are often not publicly available, and/or still may not produce assemblies at chromosome level. Most recently, long-reads sequencing technologies, such as Pacific Bioscience and Oxford Nanopore Technologies, have been used to sequence lager yeast genomes; but until now, only two assemblies of this kind of data are available in the public repositories (Liu et al. 2018; Salazar et al. 2019).

Regarding the mentioned technologies limitations, we sought an alternative approach to validate and improve the available *Saccharomyces* sp. 790 genome assembly (De León-Medina et al. 2016), a previously sequenced lager yeast hybrid, and also to elucidate its genome structure. We envisioned that this alternative could be in the form of experimental data. Bacterial artificial chromosomes (BACs) (Shizuya et al. 1992) are an experimental resource consisting of circular and stable cloning vectors that can contain large DNA inserts (up to 300 kpb). The first bases from both ends of the BACs inserts, or BAC-end sequences (BESs), can be explored using universal primers anchored in the vector backbone and Sanger sequencing. BACs were frequently used to assist genome sequencing and assembly (Weber and Myers 1997) and could be applied to investigate an already sequenced and assembled hybrid genome; although, they have never been used for this particular purpose. BACs are also an alternative source of information, since they are created from the same organism but independently from the information used for genome assembly. Furthermore, the features of these vectors, such as the paired-end, long-insert, and single allele information, make them attractive to elucidate the genome structure of a hybrid organism, as we detail in this study. Therefore, we constructed a BAC library of *Saccharomyces* sp. 790 and obtained the BESs. As part of our efforts, we developed, curated, and made available a pipeline for analyzing the data. Afterward, we used the generated paired-end and long-insert information to validate and improve the available *Saccharomyces* sp. 790 assembly (De León-Medina et al. 2016), and using the single allele information, we elucidated the *Saccharomyces* sp. 790 genome structure. Finally, we observed a *Saccharomyces* sp. 790 genome feature that might differentiate it from other lager yeasts.

## Materials and methods

### Yeast strain and sequences files

The brewing yeast strain *Saccharomyces* sp. 790, specific genome sequences (scaffolds 17 and 22; File S1), and the short-reads sequencing library were obtained from Cervecería Cuauhtémoc

Moctezuma S.A. de C.V. (Monterrey, NL, Mexico). The reference sequence of *S. cerevisiae* S288c genome (Engel et al. 2014) was downloaded from the Saccharomyces Genome Database (<https://www.yeastgenome.org>). Genome sequence assemblies of *S. eubayanus* FM1318 (Baker et al. 2015), 10 *S. pastorianus* strains (Nakao et al. 2009; Dostálek et al. 2013; Walther et al. 2014; van den Broek et al. 2015; Okuno et al. 2016) (Supplementary Table S1), *Saccharomyces* sp. 790 (De León-Medina et al. 2016), and *Saccharomyces* sp. M14 (Liu et al. 2018) were retrieved from the NCBI Genome database (NCBI, Genome <http://www.ncbi.nlm.nih.gov/genome>). See Supplementary Table S1 for more information about the assemblies.

## Genome annotation and phylogenetic analysis

We annotated the genome assemblies of *S. pastorianus* strains CBS 1503, CBS 1513, CCY48-91, CBS 1483, CBS 1538, WS 34/70, *Saccharomyces* sp. M14, and *Saccharomyces* sp. 790 (Dostálek et al. 2013; Walther et al. 2014; van den Broek et al. 2015; De León-Medina et al. 2016; Okuno et al. 2016; Liu et al. 2018) (Supplementary Table S1). Genome annotation of the yeast assemblies was performed with MAKER2 (v2.31.10) (Holt and Yandell 2011). In brief, gene models were obtained with SNAP (v2013-11-29) (Korf 2004) and Augustus (v3.2.3) (Stanke et al. 2006) as *ab initio* gene predictors and with homology to the transcripts and proteins of *S. cerevisiae* S288c (Engel et al. 2014) and *S. eubayanus* FM1318 (Baker et al. 2015). For the functional annotation, we used *S. cerevisiae* S288c (Proteome ID: UP000002311) and *S. eubayanus* FM1318 (Proteome ID: UP000050240) proteomes (UniProt, <https://www.uniprot.org/>). Proteomes were combined in a single file, which means that both parental sequences are aligned competitively and the best hit of a gene corresponds to the most likely function of the gene, but also the most likely parental type. Orthologous genes' identification from each sub-genome was carried out using the OrthoMCL algorithm (v1.4) (Li et al. 2003) as implement in *get\_homologues.pl* (v23072018) (Contreras-Moreira and Vinuesa 2013) and with the gene information of *S. cerevisiae* S288c (Engel et al. 2014) and *S. eubayanus* FM1318 (Baker et al. 2015) as reference. This led to identify 316 and 1470 one-to-one orthologous groups for the *S. cerevisiae* and *S. eubayanus* sub-genomes, respectively. For each orthologous group, the nucleotide sequences among the eight lager strains and the respective parental sequences were aligned with MAFFT (v7.453) (Katoh and Standley 2013). A concatenated full-length data set containing all orthologous groups for each sub-genome was used for phylogenetic reconstruction. RaxML (v8.2.12) (Stamatakis, 2014) was used to build the maximum likelihood trees based on the GTRGAMMA model with 1000 rapid bootstraps. Both trees were visualized with *ggtree* (v4.11) (Yu et al. 2017).

## Library construction, characterization, and end-sequencing

The BAC genomic library and BESs of *Saccharomyces* sp. 790 were generated at Clemson University Genomics Institute (CUGI, Clemson, SC, USA). The library was built with *Hind*III, the pIndigoBAC536 vector, and *Escherichia coli* DH10B. Clones were randomly selected, purified, and characterized with NotI ( $n = 45$ ) (New England Biolabs, Ipswich, MA, USA) (Farrar and Donnison 2007). The insert was visualized in a 1% pulsed field gel electrophoresis agarose-0.5X TBE gel (Molecular Sigma Biology, St. Louis, MO, USA) in a CHEF Mapper XA (Bio-Rad, Hercules, CA, USA) at  $6.0 \text{ V cm}^{-1}$ , 120°, 5–15 s, lineal, 16 h, and 14°C. The estimated insert size was calculated by interpolation to a DNA ladder.

The formula  $W = NI/G$  was used to calculate the genomic coverage ( $W$ ), where  $N$  is the number of clones,  $I$  is the average insert size, and  $G$  is the genome size (the sum of the parental genomes sizes, i.e., 23.8 Mbp). All the clones were subjected to vector purification and sequencing from both ends with Sanger technology using universal primers T7 and M13.

## BESs processing, alignment pipeline, and other bioinformatic procedures

We designed a pipeline to analyze BESs information (Supplementary Figure S1; scripts available at <https://github.com/Lriegio/BES-analysis>). Input genome files preparation, BESs processing, and alignment were performed with a custom-made BASH script (*bes\_analysis.sh*) as follows: the raw BESs reads in AB1 format were converted to FASTQ with Seqret (vEMBOSS : 6.6.0.0) (Rice et al. 2000). Sequence's quality was assessed with FastQC (v0.11.5) (available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Sequences were trimmed and filtered with Trimmomatic (v0.36) (Bolger et al. 2014) with the following parameters: HEADCROP: 91 LEADING: 20 TRAILING: 20 SLIDINGWINDOW: 4:20 MINLEN: 35. The resulting sequences were converted to FASTA format with FASTX-Toolkit (v0.0.14) (available online at: [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Sequences from vectors with no insert (vector positive) and sequences from yeast DNA (yeast positive) were detected by aligning the BESs with the pIndigoBAC536 vector sequence, and the *S. cerevisiae* S288c (Engel et al. 2014), *S. eubayanus* FM1318 (Baker et al. 2015), and *Saccharomyces* sp. 790 (De León-Medina et al. 2016) assemblies using BLAST (v2.6.0+) (Altschul et al. 1990; Camacho et al. 2009) with default parameters and an e-value of  $1e^{-6}$ . BESs were deposited in the DDBJ/EMBL/GenBank database under the accession number LIBGSS\_039348. BESs with no insert were removed from the reported data set to comply with the database requirements, but the complete original data set was described and used in subsequent analyses. BESs nucleotide sequences can be located in the corresponding genome assembly using BLAST (v2.6.0+) (Altschul et al. 1990; Camacho et al. 2009), BLAT (v36x2) (Kent 2002), MEGA-BLAST (v2.6.0+) (Camacho et al. 2009), NUCMER (v3.1) (Kurtz et al. 2004), BOWTIE (v1.2.2) (Langmead et al. 2009), BOWTIE2 (v2.3.4.1) (Langmead and Salzberg 2012), and BWA (v0.7.17) (Li and Durbin 2009). All the alignment tools were tested by aligning the BESs with the *Saccharomyces* sp. genome assembly (De León-Medina et al. 2016) (Supplementary Table S2). A tool presenting intermediate results with respect to the total of aligned BESs, in this case BOWTIE2 (Langmead and Salzberg 2012), was selected to perform all BESs alignments presented in this work, unless otherwise specified. For any tool used, only one alignment per BES, either the longest or primary alignment, is selected by the BASH script. BESs alignment results were converted into GFF table format and screened with a Python script (*screening.py*). The Python script takes in the GFF file and detects BESs pairs that align in the same sequence. Then, the BESs alignments are classified according to the orientation between the pairs as paired-end, positive, negative, or opposite (Supplementary Figure S2). If only one BES pair aligns, it is classified as singleton. BESs pairs that align in different contigs are classified as unpaired-end (Supplementary Figure S2) and used for downstream analyses. Unpaired-end BESs alignments were categorized according to their hypothetical insert size (the added distance from each unpaired-end BES alignment site to the end of its aligning sequence). The screening Python script outputs a summary table, and a specially formatted GFF file that includes the BESs pairs classification, the insert size, the pair information



for unpaired-end alignments, and an alignment type color code to aid the visual inspection of the file in the Integrative Genomics Viewer (v2.4.14) (Thorvaldsdóttir et al. 2013). Two additional custom-made Python scripts (sliding\_paired.py and sliding\_unpaired.py) were designed and used to detect large regions (>60 kbp) not spanned by paired-end BESs alignments and regions enriched with unpaired-end BESs pairs pointing to the same contig (consistent). These scripts scan the formatted BESs alignments GFF file using a sliding window algorithm and output a BED file with the coordinates of the detected regions. A visual summary of the results can be obtained with an R script (graphic\_summary.R) from the summary table and the BESs scanning results. The integrated information of BESs alignments, BESs scanning results, and annotation were visualized in the Integrative Genomics Viewer (Thorvaldsdóttir et al. 2013) and figures were generated using the R package Gviz (v1.28.3) (Hahne and Ivanek 2016). Scaffolding of the *Saccharomyces* sp. 790 assembly (De León-Medina et al. 2016) was performed using SSPACE (v3.0) (Boetzer et al. 2011) and the BOWTIE2 BESs alignments results. To propose a genome structure of *Saccharomyces* sp. 790, the output scaffolds were ordered and joined into chromosomes using NUCMER (v3.1) (Kurtz et al. 2004) and the *S. cerevisiae* S288c (Engel et al. 2014) and *S. eubayanus* FM1318 (Baker et al. 2015) genomes as reference. The best mapped location of each query contig (show-tiling -p option) was chosen as best hit. After this process, BESs were realigned to the resulting sequences, and regions enriched with consistent unpaired-end BESs alignments were detected using our scanning script (sliding\_unpaired.py) at each of the following two steps. In the first step, the unpaired-end BESs scanning results were used to detect potential recombination sites in the newly joined chromosomes, then the closest gap toward the unpaired-end BESs alignments direction was selected, and the chromosomes were splitted in the gap position. In the second step, a final genome structure in FASTA format was generated with Biostrings (v2.52.0) (Pagès et al. 2019) using all of the possible homeologous chromosomes joins suggested by the unpaired-end BESs scanning results. The *Saccharomyces* sp. 790 genome structure in FASTA format was annotated identically as the other lager yeast assemblies and plotted to portray the *Saccharomyces* sp. 790 genome structure using genoPlotR (v0.8.9) (Guy et al. 2010).

### Gaps and homeologous translocations amplification

Briefly, primers flanking potential contig joins and homeologous translocations were designed using the appropriate sequences (Supplementary Table S3). Library clones spanning regions of interest were selected to be used as template (Supplementary Table S3). BACs were purified (Farrar and Donnison 2007), DNA was amplified by PCR using standard conditions, and the resulting fragments were visualized in an electrophoresis gel.

### Flow cytometry analysis to measure total DNA content

DNA content of *S. cerevisiae* CLA1 (Avendaño et al. 1997), *S. eubayanus* FM1318 (Baker et al. 2015), *S. pastorianus* strains CBS 1513 (Walther et al. 2014) and CBS 1483 (van den Broek et al. 2015), and the study yeast was measured. *S. cerevisiae* CLA1 and *S. eubayanus* FM1318 cells were grown in casamino acids medium (yeast nitrogen base 1.7 gL<sup>-1</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 5 gL<sup>-1</sup>, casamino acids 6 gL<sup>-1</sup>, and dextrose 20 gL<sup>-1</sup>), whereas the lager yeasts cells were grown in YPD media (yeast extract 10 gL<sup>-1</sup>, peptone 20 gL<sup>-1</sup>, and dextrose 20 gL<sup>-1</sup>) to avoid nutritional stress. The Haase and Reed (2002) protocol was used to stain yeasts' DNA with SYTOX Green

(Thermo Fisher Scientific, Waltham, MA, USA). Samples were analyzed in the Molecular Biology Division's (IPICYT) flow cytometer BD FACSCalibur (Becton Dickinson & Company, Franklin Lakes, NJ, USA). Fluorescence was captured in the FL1-A channel and 5000 events in the assigned window were registered. The fluorescence geometric means of G0/1 and G2 cell cycle phases were calculated in the FlowJo X software (v10) (Becton Dickinson & Company, Franklin Lakes, NJ, USA). A linear regression was employed to determine the DNA content of the *Saccharomyces* sp. 790 strain, considering the fluorescence geometric means and the reported genome size of the control strains (Engel et al. 2014; Walther et al. 2014; Baker et al. 2015; van den Broek et al. 2015). Fluorescence intensity in FLA-1 channel against the number of events (fluorescence histogram) was plotted using the R packages flowCore (v1.50.0) (Hahne et al. 2009) and ggcyto (v1.12.0) (Van et al. 2018).

### Bioinformatic estimation of *Saccharomyces* sp. 790 chromosome copy number

A *Saccharomyces* sp. 790 short-reads sequencing library was aligned with the parental-like genomes of *S. cerevisiae* S288c (Engel et al. 2014) and *S. eubayanus* FM1318 (Baker et al. 2015) using BOWTIE2 (v2.3.4.1) (Langmead and Salzberg 2012). The output file in SAM format was processed in two separate ways. First, to estimate the haploid distribution, the SAM file was converted to BAM using Samtools (v1.7) (Li et al. 2009) and from BAM to BED using Bedtools (v2.26.0) (Quinlan 2014). Next, the sequencing depth histograms per parental chromosome were obtained using Bedtools genomcov (v2.26.0) (Quinlan 2014). Second, to calculate the sequencing depth mean across each parental chromosome, sequencing depth per base was calculated using Samtools (v1.7) (Li et al. 2009). Results of the first and second analysis were processed with a bespoke R script. The script establishes the smallest distribution in the chromosome sequencing depth histograms as the haploid distribution and calculates the sequencing depth per base mean in 1000bp windows. At last, the formula  $C = \log_2(W/H)$  was used to obtain the window copy number (C), where W is the window depth mean and H is the haploid distribution. It was established that the logarithm base two equal to zero is one copy. All windows copy number for each chromosome were plotted using the R graphics package (v3.6.3).

### Data availability

All lager yeast assemblies used in this study are publicly available (see Supplementary Table S1). BESs generated in this study were deposited in the DDBJ/EMBL/GenBank database under the accession number LIBGSS\_039348. Curated code to analyze the LIBGSS\_039348 is available at <https://github.com/Lriego/BES-analysis>. Not curated code and raw data are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables. Supplementary material available at figshare: <https://doi.org/10.25387/g3.14156912>. Supplementary File S1 contains the potential nonhomologous translocation (scaffolds 17 and 22) in FASTA format. File S2 is the BED file of consistent unpaired-end BESs alignments in *S. pastorianus* 790. Supplementary File S3 is the BED file of consistent unpaired-end BESs alignments in the combined parental genomes of *S. cerevisiae* S288c and *S. eubayanus* FM1318 with at least three supporting clones. File S4 is the BED file of consistent interparental unpaired-end BESs alignments in the combined parental genomes of *S. cerevisiae* S288c and *S. eubayanus* FM1318 with at least two supporting clones. The lager yeast strain 790 as well as the BAC library are property of Cervecería Cuauhtémoc Moctezuma S.A. de C.V.

## Results

### Saccharomyces sp. 790 is a group-2 lager yeast

*Saccharomyces* sp. 790 was established as a hybrid between *S. cerevisiae* and *S. eubayanus* (De León-Medina et al. 2016); nonetheless, its phylogenetic position with respect to other hybrids of this type is unknown. To answer the former question and to provide clarity of the source DNA used for the library construction, we performed a phylogenetic analysis using the orthologous genes from each sub-genome of *Saccharomyces* sp. 790 strain (De León-Medina et al. 2016) and seven lager beer-related strains (Dostálek et al. 2013; Walther et al. 2014; van den Broek et al. 2015; Okuno et al. 2016; Liu et al. 2018) (Supplementary Table S1). In both *S. cerevisiae* and *S. eubayanus* orthologous genes trees, all group-1 *S. pastorianus* strains (Supplementary Table S1) grouped together, and strains classified as group-2 (Supplementary Table S1) were in close distance with each other (Figure 1; Supplementary Figure S3). The phylogenetic position of *Saccharomyces* sp. M14 was unknown, but in our phylogenetic reconstruction this yeast always clustered with other group-2 yeasts, such as *S. pastorianus* CBS 1483. *Saccharomyces* sp. 790 also clustered with other group-2 *S. pastorianus* strains (Figure 1; Supplementary Figure S3). Based on these results, we classified *Saccharomyces* sp. 790 as *S. pastorianus* 790 of group-2.

### A redundant, long-insert, and paired-end experimental and sequence resource for studying the *S. pastorianus* 790 genome

To generate resources for studying the *S. pastorianus* 790 genome, we constructed from this yeast a BAC genomic library comprised of 2304 clones. The BAC library was preserved and could be used for verification experiments. The average estimated genomic coverage, or redundancy, was 9.0X, the measured clone insert proportion was 100% (n = 45) (Supplementary Figure S4), and the insert size of the library was  $93.4 \pm 31.7$  kbp (Table 1). The observed large insert size could be a key feature to study the *S. pastorianus* 790 genome at large scale, since it can span conflicting regions of DNA such as repeated sequences or translocation points. We obtained the BESs of the library (4608 total BESs), we filtered the resulting sequences and 4004 passed the quality parameters. From the

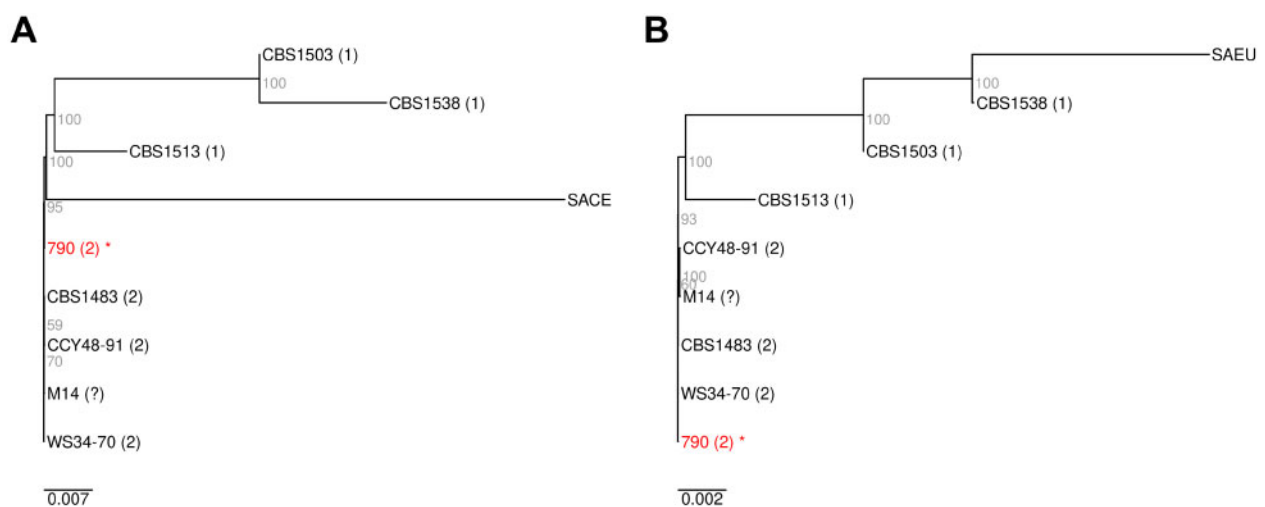
**Table 1** Genomic library characterization and end-sequencing summary

BAC-based genomic library	
Clones	2304
Genomic coverage	9.0X
Clone insert proportion	100%
	n = 45
Average estimated insert size	$93.4 \pm 31.7$ kbp
	n = 45
BESs	
Total BESs	4608
Quality BESs	4004
Paired BESs	3586
Single BESs	418
Average sequence length	$710.44 \pm 212.24$ bp
Sequenced bases	2,844,613 bp

passing BESs, 3586 were paired (both pairs) and 418 were single (one pair) (Table 1 and Supplementary Table S4). Furthermore, the average sequence length was  $710.44 \pm 212.24$  bp and the total sequenced bases were 2,844,613 bp (Table 1). This sequence information is the link between the BAC library experimental data and the bioinformatic genome information. To detect sequences from vectors with no insert (vector positive), we aligned the 4004 quality BESs with the vector sequence using BLAST (Altschul et al. 1990; Camacho et al. 2009). Of these, 109 aligned exclusively to the vector, which indicated sequences from BACs with no insert, whereas 3895 did not align (Supplementary Table S4). BESs are available in the DDBJ/EMBL/GenBank database under the accession number LIBGSS\_039348. BESs with only vector sequences were excluded for a total of 3895 BESs in the reported data set.

### BESs alignment pipeline benchmarking with several *S. pastorianus* assemblies

To investigate our specific type of data, we designed a pipeline for analyzing BESs information. The scripts and pipeline were curated and published for other potential users (<https://github.com/Lriego/BES-analysis>). To test the proper functioning of our



**Figure 1** Phylogenetic relationships of *Saccharomyces* sp. 790 and seven lager beer-related yeast strains. All yeast assemblies were annotated with MAKER2 pipeline (Holt and Yandell 2011) and the orthologous genes from each sub-genome segregated and used to construct specific parental sub-genome trees. (A) Maximum likelihood tree of the *S. cerevisiae* parental sub-genome genes. (B) Maximum likelihood tree of the *S. eubayanus* parental sub-genome genes. In both trees, *Saccharomyces* sp. 790 clustered closer to other group-2 lager yeast strains; therefore, we classified it as *S. pastorianus* 790 of group-2. Numbers at the nodes are the percent of bootstrap replicates that support it (only numbers higher than 50% are shown). The scale on the bottom left represents genetic distances in substitutions per nucleotide. Numbers in parenthesis are the lager group reported for that yeast (see Supplementary Table S1 for references information). SACE: *S. cerevisiae*; SAEU: *S. eubayanus*. \* Study yeast.

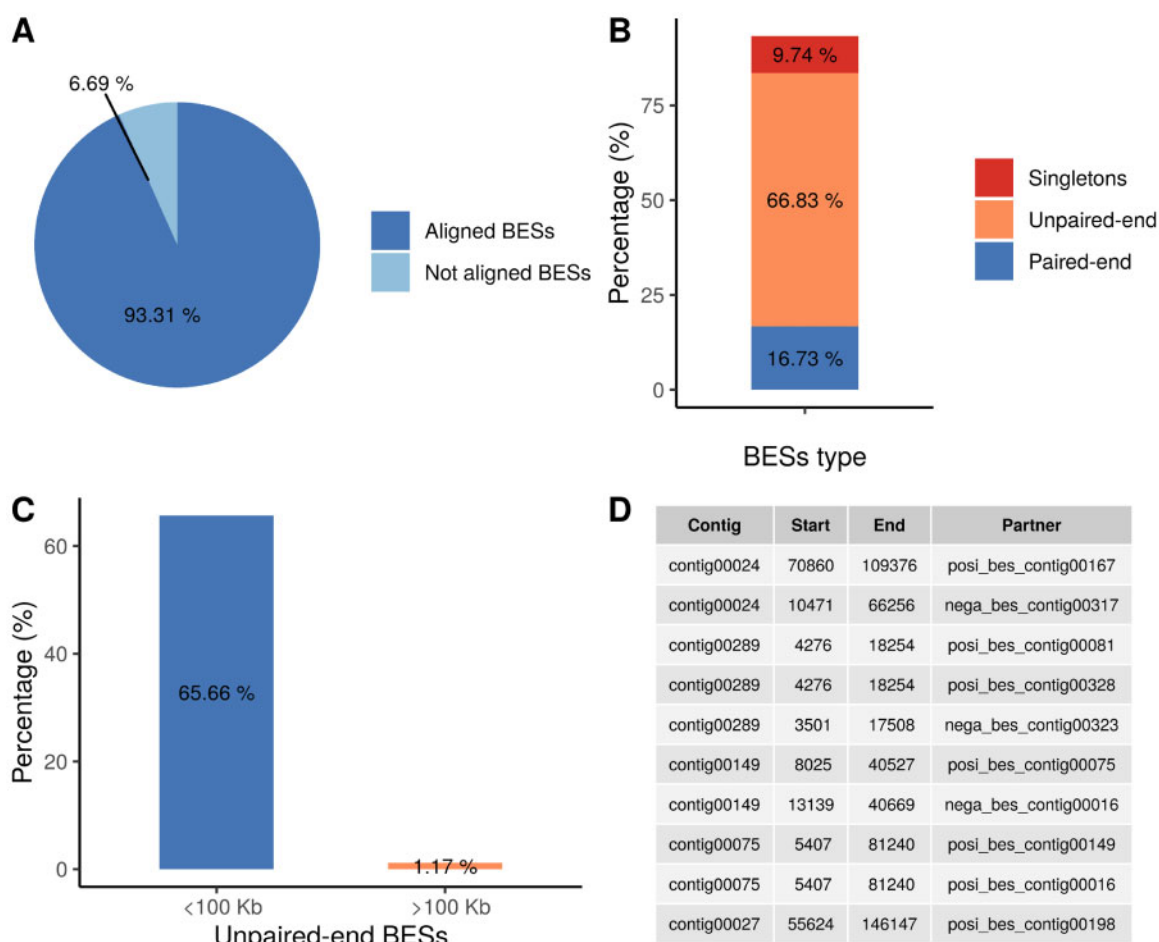
pipeline, we aligned using BOWTIE2 (Langmead and Salzberg 2012) the BESs collection with several lager yeast assemblies of different origin, sequencing, and assembly technologies (Supplementary Table S1). These were the *S. pastorianus* 790 assembly (De León-Medina et al. 2016), 10 *S. pastorianus* assemblies (Nakao et al. 2009; Dostálek et al. 2013; Walther et al. 2014; van den Broek et al. 2015; Okuno et al. 2016) (Supplementary Table S1), and the *Saccharomyces* sp. strain M14 assembly (Liu et al. 2018). The percentage of total aligned BESs of group-1 yeast assemblies ranged from 60.74% to 80.74% (Table 2). The range for group-2 yeast assemblies was 88.06–97.0% (Table 2). The difference in the total percent of aligned BESs among the lager groups could be due to a higher nucleotide identity of the group-2 yeast sequences with the same group yeast BESs data set. Alternatively, it could be due to the larger genome size of group-2 yeasts and, as consequence, absent sequences in group-1 yeasts. The percentage of total aligned BESs did not correlate with the N50 of the assemblies ( $R=0.12$ ,  $P>0.05$ ), which suggest that the BESs can be aligned to any *S. pastorianus* assembly, independently of the assemblies' contiguity, and that the alignment results may not be biased by this assembly's parameter. Oppositely, the percentage of paired-end BESs varied largely among assemblies but correlated with their N50 ( $R=0.96$ ,  $P<0.05$ ), which indicates that the number of this type of alignments depends on the assembly's completeness. The percentage of the rest of the alignment types, such as unpaired-end alignments, may also be related to the N50 of the assemblies; but it could also be due to genomic differences between the strains. Therefore, the *S. pastorianus* 790 BESs alignments with other lager yeast assemblies only exemplify the type of results that could be obtained with BESs data, and information from strain specific BAC libraries should be used for correct interpretation. Finally, simplistic graphics of the alignments results can be obtained (Figure 2), which are intended to provide the potential users a rapid overview of the pipeline results. Similarly, a specially formatted file can be used to aid the visual inspection of the alignments results (Figure 3). Altogether, the pipeline works properly with any given *S. pastorianus* assembly, regardless of their sequencing and assembly technology, and BESs can be aligned with the assemblies independently of their contiguity.

### BESs alignments validate and improve the *S. pastorianus* 790 assembly

To assess the BAC library information potential, we analyzed in detail the BESs alignments with the available *S. pastorianus* 790 assembly (De León-Medina et al. 2016). We observed that 93.31% of the BESs could be located in the 790 strain assembly, whereas 6.69% did not align (Figure 2A and Supplementary Table S4). Of the aligning BESs, 16.73% were paired, 66.83% were unpaired-end, and 9.74% were singletons (Table 3 and Figure 2B). We did not find BESs alignments in positive, negative, or opposite orientations. Only paired-end alignments are considered in agreement with biological data and, therefore, validated the covered regions (Figure 3B). Unpaired-end BESs alignments indicate potential assembly improvements, such as contig joins (Figure 3C); however, these contig joins can only be made if the hypothetical insert size is close to the one observed experimentally. Otherwise, unpaired-end BESs alignments may indicate discrepancies in the experimental data compared to the assembly, such as assembly errors, or something else. Therefore, we categorized the unpaired-end alignments according to their hypothetical insert size in greater or less than 100 kbp (the estimated rounded insert size

**Table 2** BESs alignments results with the different *S. pastorianus* strain assemblies, categorized by alignment type and expressed as percentages

Alignment type	Group 1										Group 2				
	CBS 1538 (Okuno et al. 2016)	CBS 1503 (Unpub)	CBS 1503 (Okuno et al. 2016)	CBS 1513 (Walther et al. 2014)	CBS 1513 (Okuno et al. 2016)	WS34-70 (Walther et al. 2014)	WS34-70 (Okuno et al. 2016)	WS34-70 (Nakao et al. 2009)	CBS 1483 (van den Broek et al. 2015)	CCY48-91 (Dostálek et al. 2013)	M14 (Liu et al. 2018)	790 (De León-Medina et al. 2016)			
Paired-end	17.68	12.84	18.93	57.89	30.17	23.28	24.53	0.35	6.89	9.34	68.03	16.73			
Opposite	0	0	0	0.65	0	0	0	0	0	0	0.95	0			
Positive	0	0	0	0	0	0.1	0	0	0	0	0	0			
Negative	0	0	0	0.1	0.05	0	0	0	0	0	0	0			
Unpaired-end	33.37	42.81	41.56	11.54	39.36	56.09	57.54	74.63	75.32	73.78	20.13	66.83			
>100 kbp	1.45	1.2	0.9	1.35	2.37	1.37	1.85	0.02	0.4	1.0	6.47	1.17			
<100 kbp	31.92	41.61	40.66	10.19	36.99	54.72	55.69	74.6	74.93	72.78	13.66	65.66			
Singletons	9.69	10.59	10.81	10.56	10.29	11.31	9.74	13.09	10.71	9.52	7.89	7.74			
BES Total	60.74	66.23	71.3	80.74	79.87	90.78	91.81	88.06	92.93	92.63	97.0	93.31			



**Figure 2** BESs alignments with the *S. pastorianus* 790 De León-Medina et al. (2016) assembly. (A) Percentage of aligned BESs. (B) Percentage of the different BESs alignments types. Negative, positive, and opposite BESs bars are not shown because they are 0%. (C) Percentage of unpaired-end BESs according to the insert size. Unpaired-end BESs alignments <100 kbp can be used to join sequences, whereas unpaired-end alignments >100 kbp indicate contradictory information (misassemblies or homeologous translocations). (D) Regions with consistent unpaired-end BESs aligned. Only the first 10 regions of the file resulting from the unpaired-end BESs scanning script are shown.

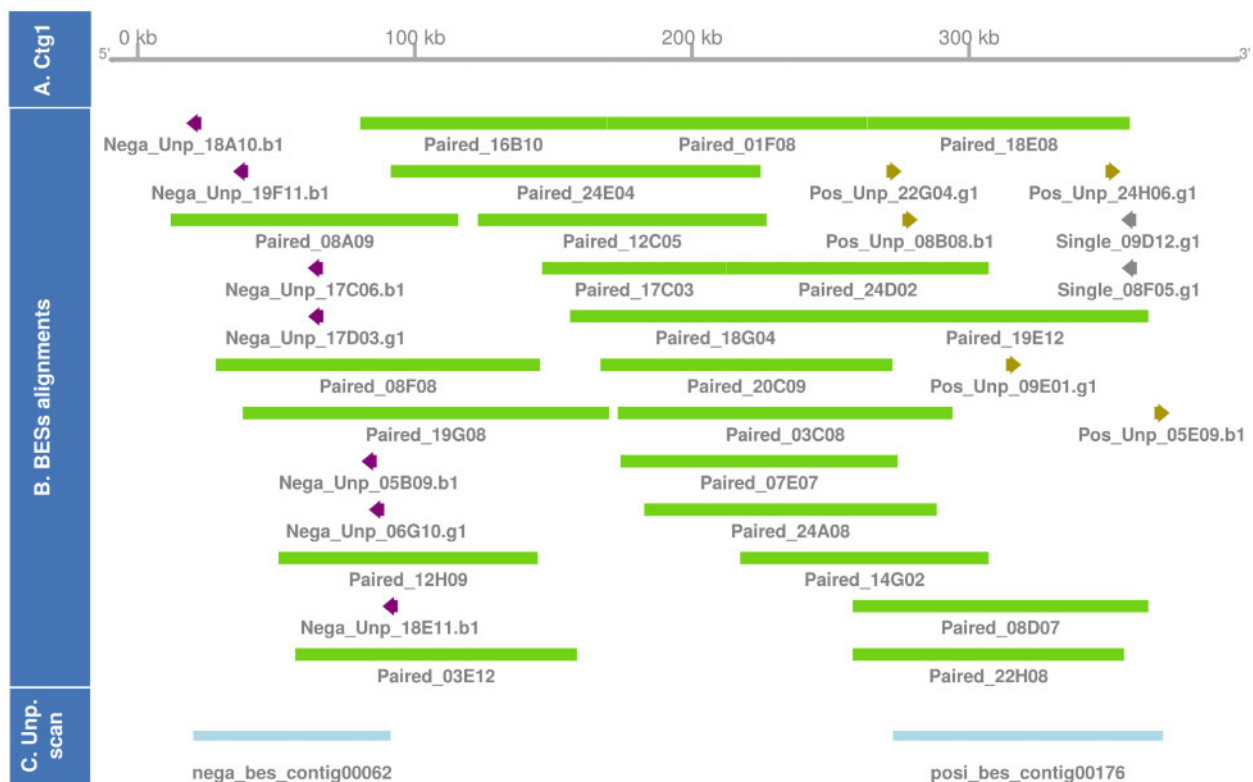
of the BAC library). We observed that 1.17% unpaired-end BESs had an insert size >100 kbp, and as mentioned, they point differences between the BESs and the assembly (or, to a lesser extent, presence of BESs from clones with unrelated ligated inserts). Whereas the remaining majority of the unpaired-end BESs (65.66%) had an insert size of <100 kbp, suggesting contig joins (Figure 2C). Accordingly, we found 334 regions with consistent unpaired-end BESs alignments (Figure 2D and Supplementary File S2). To proof that unpaired-end alignments with short insert size (<100 kbp) could be used to improve the assembly, we designed primers flanking three possible joins (Supplementary Table S3) and amplified the region using the DNA obtained from spanning clones as template. All three regions were successfully amplified, indicating that the information contained in the BACs agrees with the bioinformatic inferences (Figure 4). Overall, BESs can be used to validate and improve the available assembly, and BACs are a reliable source of information.

### BESs substantially improve the *S. pastorianus* 790 assembly by scaffolding

Our previous analysis suggested that many contigs could be joined with the BESs. To test whether the long-insert information of the BAC library sequences could be used to improve the *S. pastorianus* 790 available assembly (De León-Medina et al. 2016), we used the BESs

alignments with BOWTIE2 to scaffold the contigs using SSPACE (Boetzer et al. 2011). The number of sequences went from 1098 contigs in the original assembly to 1001 scaffolds in the scaffolded version. The former seemed like a modest improvement; however, more than a half of the new assembly (523 scaffolds) consisted of scaffolds smaller than 1000 bp that were already present in the initial assembly. Nonetheless, the N50 statistic improved from 85,203 bp (with 75 contigs) in the first assembly to 271,957 bp (with 28 contigs) in the last one. The later N50 is larger than the smallest chromosome of *S. cerevisiae* S288c (Engel et al. 2014), suggesting that the largest scaffolds in the current assembly are chromosome-sized. We realigned the BESs to the scaffolded sequences and we observed that the percentage of paired-end BESs alignments increased from the original 16.73% to 45.05%; whereas the percentage of unpaired-end BESs alignments decreased from 66.83% to 38.41% (Table 3), indicating a better agreement of the biological data with the new scaffolded assembly. Notwithstanding, when scanning for consistent unpaired-end BESs alignments, 147 regions still harbored potential contig joins. The former could be due to unresolved unions originated from repeated regions of assorted kinds, as well as homeologous translocations. Nevertheless, the use of BESs information in combination with scaffolding produces a substantially improved assembly with respect to the initial one.





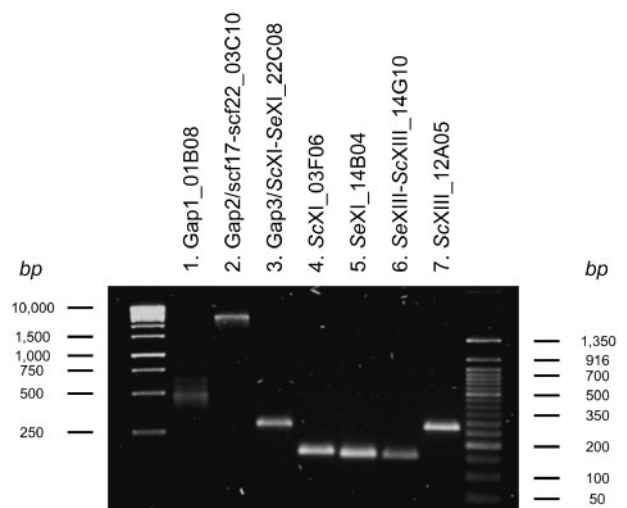
**Figure 3** BESs alignments visualization with the *S. pastorianus* 790 De León-Medina et al. (2016) assembly. (A) Contig 1. (B) BESs alignments. (C) Unpaired-end BESs alignments scanning results. Most of the BESs alignments in contig 1 were paired-end (represented with their simulated insert as thick lines with “Paired” prefix). This type of alignments validates the assembled region. We also observed unpaired-end BESs aligned positively (right arrows with “Pos\_Unp” prefix), unpaired-end BESs aligned negatively (left arrows with “Nega\_Unp” prefix), and single (left arrows with “Single” prefix). The unpaired-end BESs scanning script detected two regions with consistent unpaired-end BESs representing possible contig joins to contig00062 and contig00176 (thin lines with names nega\_bes\_contig00062 and posi\_bes\_contig00176).

**Table 3** BESs alignments results with the original De León-Medina et al. (2016) assembly and improved assemblies’ versions, categorized by alignment type and expressed as percentages (%)

Alignment type	Original	Scaffolded	Joined	Splitted
Paired-end	16.73	45.05	65.53	62.89
Opposite	0	0	0	0
Positive	0	0	0.25	0.25
Negative	0	0	0	0
Unpaired-end	66.83	38.41	3.79	6.39
>100 kbp	1.17	4.42	2.52	1.12
<100 kbp	65.66	33.99	1.27	5.27
Singletons	9.74	9.79	14.54	14.56
BES total	93.31	93.25	84.11	84.09

### BESs indicate the presence of homeologous translocations but not new nonhomologous translocations in the *S. pastorianus* 790 genome

Previously, De León-Medina et al. (2016) observed nonhomologous translocations in the *S. pastorianus* 790 assembly, whereas homeologous translocations were not searched. To systematically detect nonhomologous and homeologous translocations present in the BESs collection, we assigned the BESs to a parental sub-genome by aligning them with the combined parental-like genomes of *S. cerevisiae* S288c (Engel et al. 2014) and *S. eubayanus* FM1318 (Baker et al. 2015). BESs that align in different sequences (unpaired-end) within the same parental genome are thought to harbor nonhomologous translocations, but if such BESs align in different parental genomes, then they are thought to harbor



**Figure 4** Gaps and homeologous translocations amplification. Vectors were purified from clones spanning the potential contig join or homeologous translocation. Primers flanking the join were designed and used for PCR. Three contig joins (lanes 1–3) were successfully amplified. Likewise, homeologous translocations ScXI-SeXI (lane 3) and SeXIII-ScXIII (lane 6) were detected. ScXI (lane 4) and SeXI (lane 5) alleles, and a ScXIII (lane 7) allele, were also identified by PCR. Primers spanning Gap2 (lane 2) were also used to confirm the union between scaffold 17 and scaffold 22 of an *S. pastorianus* 790 assembly improved by scaffolding (De León-Medina et al. 2016). Primers used to amplify the union of Gap3 were also flanking the chimeric allele ScXI-SeXI (lane 3). Clone names used are followed by the underscore sign of each lane name.

homeologous translocations. The results showed that most of the BESs alignments (79.72%) in the combined parental genomes were paired-end (Supplementary Figure S5B) and all the parental chromosomes were covered by this type of alignments (except for chromosomes III and VII of the *S. eubayanus* genome that were partially covered), which suggest that *S. pastorianus* 790 possesses at least the same chromosomes of both parental yeasts. The unpaired-end BESs alignments were 5.19% (Supplementary Figure S5B) and, as established, indicate potential translocations in the *S. pastorianus* 790 genome. Therefore, this type of alignments were scanned for regions with consistent BESs. Of the found regions, 14 occurred between chromosomes of different parental type, which translated to seven potential recombinations between homeologous chromosomes III, VII, X, XI, XIII, and XVI in two regions (Supplementary File S3). Four additional regions, which corresponded to two potential homeologous translocations, one in a second position of chromosomes X, and another in a third position of chromosomes XVI, were identified with a lower threshold (File S4) and included to the list because we considered BESs a reliable source of information, and because they were near to genes YJR009C/TDH2 and YPR191W/QCR2, other previously reported recombination sites in group-2 lager yeasts (Nakao et al. 2009; van den Broek et al. 2015; Monerawela and Bond 2017). In total, nine potential homeologous translocations were identified with the *S. pastorianus* 790 BESs and the parental genomes as reference (Supplementary Figure S5D). To validate these observations, two possible homeologous translocations in *S. pastorianus* 790, between chromosomes ScXI and SeXI, and chromosomes SeXIII and ScXIII, were confirmed by PCR amplification using the BACs DNA as template (Figure 4). Opposed to what De León-Medina et al. (2016) described, we did not find nonhomologous translocations in the unpaired-end alignments (File S3). However, *S. eubayanus* possesses two nonhomologous translocations between chromosomes II and IV, and between chromosomes VIII and XV, which were already present in this parental yeast prior to the hybridization event (Baker et al. 2015). These translocations are present in the Baker et al. (2015) assembly and did not produce unpaired-end BESs alignments, the kind of alignments we used for searching translocations. The above means that the translocations of chromosomes II and IV and VIII and XV are also present in *S. pastorianus* 790. To confirm that other nonhomologous translocations detected by De León-Medina et al. (2016) were inaccurate, we aligned the BESs to a potential nonhomologous translocation present in scaffold 17 of an *S. pastorianus* 790 assembly improved by scaffolding with short-insert sequencing libraries (De León-Medina et al. 2016) (Supplementary File S1). We observed that the region (near loci YBL074C and YER092W) was not covered by paired-end BESs alignments, which suggest that this region is not sustained by the BESs experimental evidence (Supplementary Figure S6). The region was also flanked by unpaired-end alignments pointing out a union between scaffold 17 and scaffold 22. Using BACs DNA as template, we confirmed the union of scaffold 17 and scaffold 22 by PCR (Figure 4). Furthermore, the region had a gap (Supplementary Figure S6 and Supplementary File S1) possibly generated during the De León-Medina et al. (2016) scaffolding, a procedure that has been reported introduces errors (Martin et al. 2016). The former findings indicated that the nonhomologous translocation in the De León-Medina et al. (2016) scaffold 17 is a misassembly. In summary, the *S. pastorianus* 790 genome structure, with its chromosomes homeologous to both parental types, is represented in the BAC library; moreover, the BESs information allows the detection

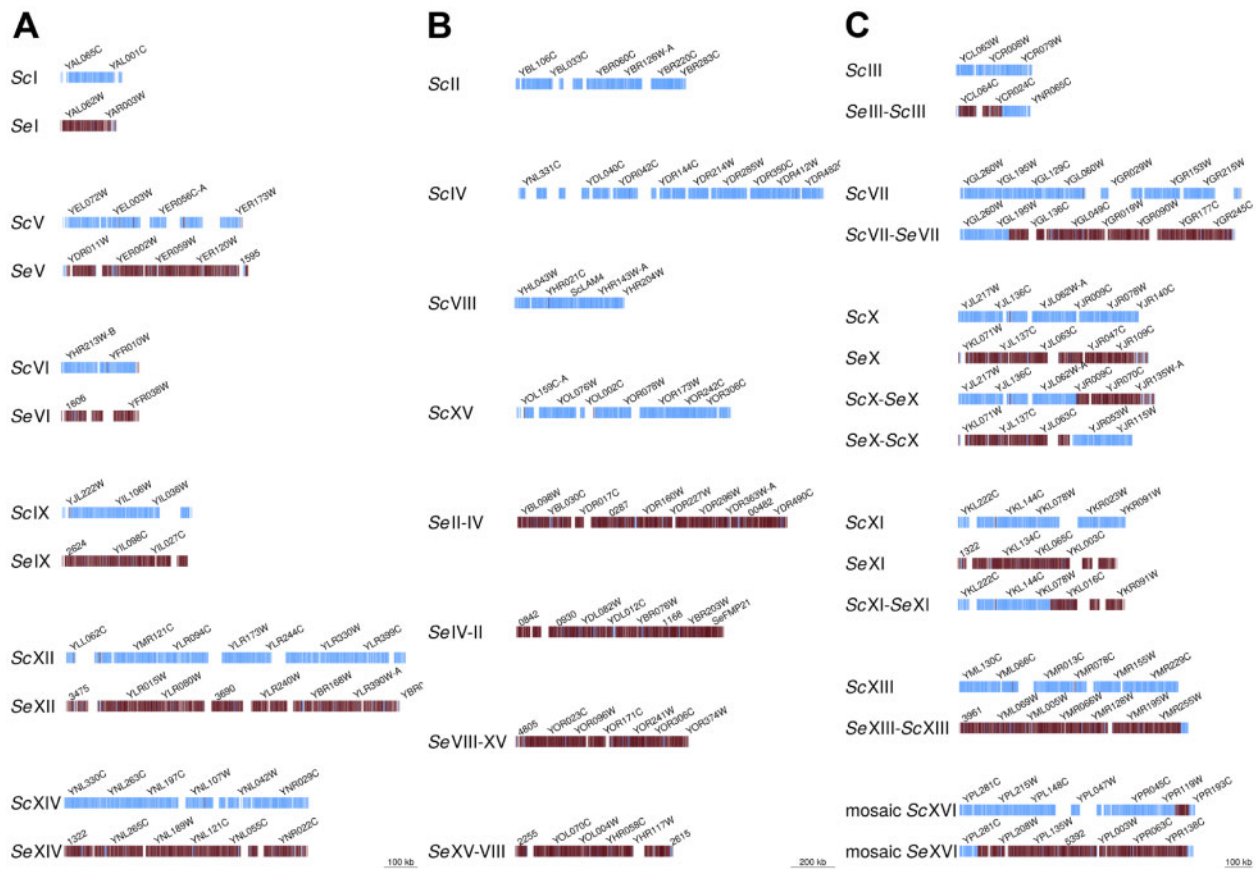
of homeologous translocations between the chromosomes and also misassemblies.

### ***Saccharomyces pastorianus* 790 genome structure is similar to that of other group-2 yeasts**

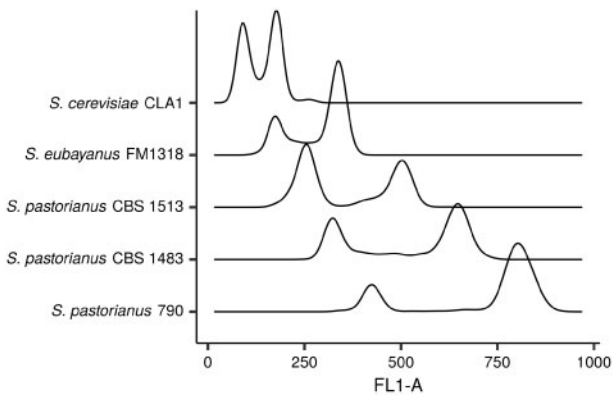
With the prior information, specifically the BES-supported scaffolded assembly version, and the parental chromosomes and homeologous translocations detected with the BESs, we proposed a hypothesis of the genome structure of *S. pastorianus* 790 based on the following observations. We used the parental-like genomes of *S. cerevisiae* S288c (Engel et al. 2014) and *S. eubayanus* FM1318 (Baker et al. 2015) to order and join the scaffolds into chromosomes using NUCMER (Kurtz et al. 2004). Given that ordering the scaffolds using another species reference may bias the result, we considered the output as a potential genome structure and not as a new assembly. We realigned the BESs with this genome structure, detected the site of potential homeologous translocations using our unpaired-end BESs scanning script (Supplementary Figure S7), splitted the sequences, and generated alternative chromosomes. We annotated the new genome structure and assigned the sequences to a parental type using the observed gene information. With the former, we could reconstruct 35 chromosome-sized pseudo-molecules. Of these, 12 were formed by sequences mainly containing genes of one parental type and that were present in both allelic versions (Sc and Se). These were chromosomes ScI, SeI, ScV, SeV, ScVI, SeVI, ScIX, SeIX, ScXII, SeXII, ScXIV, and SeXIV (Figure 5A). Eight chromosomes almost exclusively contained genes assigned to one parental type and were present in only one parental allelic version. These were chromosomes ScII, ScIV, ScVIII, and ScXV from the *S. cerevisiae* sub-genome and SeII-SeIV, SeIV-SeII, SeVIII-SeXV, and SeXV-SeVIII from the *S. eubayanus* sub-genome (Figure 5B). Fifteen sequences formed chromosome groups present in one or both parental versions plus one or two chimeric versions (i.e., contained long stretches of genes assigned to one or the other parental type, and hence harbored homeologous translocations). These were the chromosome groups ScIII and SeIII-ScIII; ScVII and ScVII-SeVII; ScX, SeX, ScX-SeX, and SeX-ScX; ScXI, SeXI, and ScXI-SeXI; ScXIII and SeXIII-ScXIII; and mosaic ScXVI and mosaic SeXVI (Figure 5C and Supplementary Table S5). Monerawela and Bond (2017) reported for *S. pastorianus* 790 a second translocation in the YGR285C/ZUO1 gene of chromosomes VII. We could locate one ZUO1 gene in the sub-telomeric region of each chromosome ScVII and ScVII-SeVII. When we inspected the BESs alignments near the ZUO1 gene of chromosome ScVII, we observed a small region (approximately of 10,000 bp) with two unpaired-end BESs alignments pointing to chromosome ScVII-SeVII (Supplementary Figure S8); therefore, we also considered this translocation in the genome structure. The *S. pastorianus* 790 proposed structure is highly similar to that of other lager yeasts such as *S. pastorianus* WS 34/70 (Supplementary Table S5) (Nakao et al. 2009; Monerawela and Bond 2017), which suggest that potential differences between *S. pastorianus* 790 and other yeast of the same type may be due to other genome features instead of the genome structure exclusively. Overall, it was possible to propose a genome structure of *S. pastorianus* 790 exploiting the information contained in the BESs.

### **High copy chromosome number and ploidy levels in *S. pastorianus* 790**

The fact that the genome structure of our study yeast was similar to that of other group-2 lager strains prompted us to question whether there were other differentiating features in the *S. pastorianus* 790 genome. It is thought that group-1 lager yeast strains



**Figure 5** Hypothetical genome structure of *S. pastorianus* 790. (A) Chromosomes present in both allelic versions (Sc and Se). (B) Chromosomes present in only one allelic version (Sc or Se). (C) Chromosome groups present in one or two parental version plus one or two chimeric versions (Sc-Se or Se-Sc). The genome structure was proposed with the combined information of the BES-supported scaffolded assembly version, the parental genomes as reference, and the detected homeologous chromosomes and homeologous translocations. The final assembly version was annotated and the genes were assigned to a parental type. Each vertical line represents a gene. Genes were shaded according to their parental type (light for *S. cerevisiae* and dark for *S. eubayanus*). The scale on the bottom right of each panel represents the distance in kilobase pairs (kb).



**Figure 6** Fluorescence histograms in channel FL1-A of different yeast cells stained with SYTOX Green. In each of the analyzed yeasts, two fluorescence peaks, which correspond to the G0/1 and G2 of the cell cycle phases, were observed. Fluorescence is directly proportional to the DNA content. The yeast with the smallest DNA content was the haploid *S. cerevisiae* CLA1 (Avendaño et al. 1997; Engel et al. 2014), followed by the diploid *S. eubayanus* FM1318 (Baker et al. 2015), the allotriploid *S. pastorianus* CBS 1513 (Walther et al. 2014), the allotetraploid *S. pastorianus* CBS 1483 (van den Broek et al. 2015), and lastly the allotriploid *S. pastorianus* 790.

are generally allotriploids, whereas, group-2 yeasts are thought to be allotetraploids (Walther et al. 2014; van den Broek et al. 2015). We measured the DNA content of *S. pastorianus* 790 and

other yeasts with known ploidy levels. We observed that the DNA content of *S. pastorianus* 790 was of 60.1 Mbp, which is approximately five times larger than the *S. cerevisiae* S288c genome (Engel et al. 2014) (Figure 6), and suggesting that *S. pastorianus* 790 is an allotriploid. To further investigate this, we took one short-reads sequencing library from a previous whole genome sequencing of *S. pastorianus* 790 (De León-Medina et al. 2016), and we obtained the approximate chromosomes copy number using the parental chromosomes as reference. We observed that the overall number of chromosome copies ranged from one to three (Supplementary Table S6; Supplementary Figures S9 and S10), and for the specific case of chromosomes with homeologous pairs the chromosomes sum ranged from four to five (Supplementary Table S6). We also observed copy number changes in all the regions flanking the homeologous translocations detected using the BESs. These copy number changes suggest recombination sites and validate our previous inferences. The only exceptions from the former observation were chromosomes X, which exhibited no copy number changes across the sequence and an equal chromosome copy number. Finally, when taking chromosome sizes and their estimated copy number, we obtained a total size of 55.79 Mbp (Supplementary Table S6), which is close to the genome size we determined experimentally (Figure 6). All of these results favor the hypothesis that *S. pastorianus* 790 is an allotriploid.



## Discussion

The use of BACs and its end-sequences validates and improves the available *S. pastorianus* 790 assembly. Furthermore, BACs can be used to study the genome structure of this yeast. The generated library and sequences represent an independent experimental alternative that contain several levels of information, namely paired-end, long-insert (around 100 kbp), and single allele information, that are particularly advantageous compared to other strategies to study this hybrid genome with a high ploidy level (Supplementary Table S7).

BESs, as the Hi-C technique (Marie-Nelly et al. 2014), are an independent experimental resource. This is an advantage compared to other options to produce better assemblies that include bioinformatic finishing (Gordon and Green 2013; Hunt et al. 2013; Schatz et al. 2013) or different assembly strategies such as phasing (Pryszcz and Gabaldón 2016; Fay et al. 2019), because these latter options rely on *in silico* analyses exclusively and use the same source of information generated for the assembly. Instead, our strategy combines newly generated data with bioinformatic analyses, which provides an additional source of information to confirm observations. Furthermore, BACs of the end-sequences are preserved and used later for experimental verification (Ammiraju et al. 2005), as we showed by PCR amplification of the suggested joins and homeologous translocations, and to facilitate experiments on heterozygous regions (Barrera-Saldaña et al. 2017). The former pose an additional advantage compared to short-reads sequencing libraries that are discarded after the process (Supplementary Table S7).

The BESs pairing information can be used to improve the *S. pastorianus* 790 assemblies in two ways. First, unpaired-end BESs alignments <100 kbp can be used to join contigs by scaffolding with SSPACE (Boetzer et al. 2011) and substantially improve the N50 of the assembly (with scaffolds the size of chromosomes). Second, when the unpaired-end BESs alignments hypothetical insert is larger than expected (>100 kbp), these alignments detect misassemblies. The use of BESs could be relevant because despite the great availability of lager yeast assemblies (Nakao et al. 2009; Dostálek et al. 2013; Walther et al. 2014; van den Broek et al. 2015; De León-Medina et al. 2016; Okuno et al. 2016; Liu et al. 2018), many of them have a small N50, probably due to the complexity of sequencing and assembling hybrid genomes. To improve them, specific BESs collections could be generated and compared with their assemblies, independently of the N50, given that the total of aligned BESs does not correlate with this parameter. Whereas paired-end BESs alignments could validate regions and be seen as a measurement of an assemblies' completeness, since the number of BESs in this orientation correlates with the N50.

The long-insert BESs information can be used to span long repeated regions during scaffolding. Scaffolding (Boetzer et al. 2011) is an approach that has been used before to improve lager yeast assemblies (van den Broek et al. 2015; De León-Medina et al. 2016; Okuno et al. 2016); however, it is a bioinformatic-only approach that can introduce joining errors (Martin et al. 2016). This is indeed what we observed in the misassembly of scaffold 17 of the De León-Medina et al. (2016) assembly, which was probably generated during their scaffolding and due to the inability of the short-insert libraries (around tens or less kilobases) to span repeated regions. Supporting this, the homologous region of scaffold 17 in *S. cerevisiae* (near loci YBL074C and YER092W) is a repeated region containing genes derived from the whole genome duplication event (Engel et al. 2014) that may have caused the joining error. Unlike mate pairs sequencing libraries commonly used for

scaffolding, BESs have an insert length that can easily surpass these repeated regions (Supplementary Table S7).

Most importantly, the single allele information contained in the BESs was key to detect chimeric chromosomes in the *S. pastorianus* 790 genome. Many assembly algorithms retrieve one consensus sequence and do not allow to easily retrieve alternative alleles (Supplementary Table S7). Detection of chimeric chromosomes and misassemblies had to be performed in a blind automatic process (Hunt et al. 2013; Pryszcz and Gabaldón 2016) or manual in a time-consuming way not specially designed for hybrids and chimeric chromosomes (i.e., without a clear mechanism to deal with regions with paired-end and unpaired-end reads simultaneously aligned) (Gordon and Green 2013; Schatz et al. 2013). The pipeline presented in this work combines both, automatic and manual approaches, by automating the searches for our specific type of data, suggesting all possible joins or alleles, and facilitating the visual inspection of regions with misassemblies and homeologous translocations for the user to decide. Furthermore, the resulting files (in BED and GFF format) can be used to guide analyses with other more commonly used tools such as Biostrings (Pagès et al. 2019). One possible disadvantage could be the high number of chimeric regions or joins to inspect whether our approach is applied to other highly heterozygous genomes. Notwithstanding, it was possible for us to propose a genome structure of *S. pastorianus* 790 exploiting the single allele information contained in the BESs, which was not possible with the assembly and scaffolding solely. Another possible disadvantage of our approach could be the reduced throughput, higher cost, and time required for Sanger sequencing compared to massive parallel sequencing (Supplementary Table S7). Nevertheless, new techniques to generate BESs using short-reads sequencing have been developed (Wei et al. 2017; Yang et al. 2020), which circumvent the downsides of Sanger sequencing and make the application of BESs and approaches such as the one described in this work more feasible.

The lager genome structure has been widely studied with special focus on chimeric chromosomes (Nakao et al. 2009; Walther et al. 2014; van den Broek et al. 2015; Monerawela and Bond 2017). All the chimeric chromosomes we detected in our studied yeast have been already reported in other lager strains (Nakao et al. 2009; van den Broek et al. 2015; Monerawela and Bond 2017). For instance, the *S. pastorianus* 790 genome structure is very similar to that of *S. pastorianus* WS 34/70 (Supplementary Table S5) (Nakao et al. 2009; Monerawela and Bond 2017). Of the 36 reported chromosomes of WS 34/70 strain, we could identify 35 equivalent sequences. We did not find a chromosome with the nonhomologous translocation ScV-ScXI or any other nonhomologous translocations in the *S. cerevisiae* sub-genome. Similarly, almost all the homeologous translocations we observed in *S. pastorianus* 790 have been detected previously in this same yeast by other researchers (Monerawela and Bond 2017), through a strategy that uses sequence similarity among known recombination sites. Nonetheless, we did not detect an introgression in chromosomes XII probably because it is small and did not have BESs aligned that would give unpaired-end alignments. Contrarily, we observed two additional translocations in chromosomes SeIII-ScIII and SeX-ScX that were not detected by the Monerawela and Bond (2017) method, probably because the recombination site was not present in the highly fragmented assembly of De León-Medina et al. (2016). Therefore, the use of BESs provides an improved detection of homeologous translocations (Supplementary Table S5). We noticed that three homeologous translocations detected with a lower threshold (YGR285C/ZUO1, YJR009C/TDH2, and



YPR191W/QCR2) were near centromeric or telomeric regions. This lower sequence representation could have occurred due to the difficulty of cloning these particular chromosomal regions. Another technique to detect homeologous translocations is to align the whole genome sequencing reads to the parental chromosomes (Hewitt et al. 2014), which we used to estimate the chromosome copy number and that in part helped us to corroborate our proposed genome structure. Nevertheless, translocations involving chromosomes with an equal copy number, such as the case of chromosomes ScX-SeX and SeX-ScX, cannot be observed by this method and became evident only when studying the unpaired-end BESs alignments, which represent an advantage of the BESs over the reads alignment strategy (Supplementary Table S7).

More intriguing was the *S. pastorianus* 790 ploidy level compared to other lager yeasts. The genome size of 790 strain, established as an allopolyploid, is the highest determined experimentally for this type of yeasts. This is interesting because group-2 strains formed a more compact group in the phylogenetic tree, which could be an indication of lower genetic variability of yeasts belonging to this lager group. Likewise, the genome structure of *S. pastorianus* 790 and other group-2 yeasts (Supplementary Table S5) appeared to be highly conserved, i.e., they shared many recombination sites. Hence, the key differences between *S. pastorianus* 790 and other group-2 strains may rely on the copy number of chromosomes and genes. Indeed, deviating chromosome copy number has been associated with specific phenotypes (Bolat et al. 2008; van den Broek et al. 2015). The former could be explained by a gene dose effect or by single nucleotide differences in the coding and/or regulatory sequences of the alleles. In the last case, BACs could become handy to study dissimilarities between the alternative copies, since they originated from individual DNA molecules. The specific phenotype resulting from the extraordinary *S. pastorianus* 790 ploidy level is yet to be determined. Nonetheless, these conjectures may give direction to future experiments. For example, genome diversity of lager yeast strains is thought to be limited due to an evolutionary bottleneck that occurred during the industrialization of lager beer production (Gallone et al. 2019). However, extra-variation and innovation in the lager beer industry could be achieved by screening for different ploidy levels yeasts in available collections or by artificially increasing the number of chromosome copies by genetic modification, physical or chemical treatments (Gorter de Vries et al. 2017, 2020).

Beyond the proposed genome structure and peculiarities of *S. pastorianus* 790, we would like to highlight the importance of the use of BESs, with their paired-end, long-insert, and single allele information, to improve short-reads assemblies and detect chromosome structural variants in a hybrid genome. Altogether, our results show the value of combining bioinformatic analyses with experimental information. Our strategy represents an alternative that can be generalized to study other hybrid genomes, especially if obtaining the most complete assembly of the study organism is in mind. The elucidated genome structure and ploidy levels will allow us to design future experiments, to understand their phenotypic implications and possible industrial applications.

## Acknowledgments

The authors gratefully acknowledge the computing time granted by the IPICYT Supercomputing National Center for Education & Research (CNS-IPICYT), grant TKII-R2020-CGM.

## Funding

This project was funded by “Programa de Estímulos a la Investigación, de Desarrollo o de Innovación Tecnológica” grant 231412-2016 of the Consejo Nacional de Ciencia y Tecnología (CONACYT), Mexico. During the development of this project, C.G-M. and J.M-A. received a doctoral scholarship, and L.F.G-O. a postdoctoral scholarship, all from CONACYT.

**Conflicts of interest:** The authors declare that there is no conflict of interest.

## Literature cited

- Alkan C, Sajjadian S, Eichler EE. 2011. Limitations of next-generation genome sequence assembly. *Nat Methods*. 8:61–65.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.
- Ammiraju JSS, Yu Y, Luo M, Kudrna D, Kim H, et al. 2005. Random sheared fosmid library as a new genomic tool to accelerate complete finishing of rice (*Oryza sativa* spp. Nipponbare) genome sequence: sequencing of gap-specific fosmid clones uncovers new euchromatic portions of the genome. *Theor Appl Genet*. 111: 1596–1607.
- Avedaño A, Deluna A, Olivera H, Valenzuela L, Gonzalez A. 1997. GDH3 encodes a glutamate dehydrogenase isozyme a previously unrecognized route for glutamate biosynthesis in *Saccharomyces cerevisiae*. *J Bacteriol*. 179:5594–5597.
- Baker E, Wang B, Bellora N, Peris D, Hulfacher AB, et al. 2015. The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol Biol Evol*. 32:2818–2831.
- Barrera-Saldaña HA, Ramírez-Sánchez AD, Palacios-Tovar TE, Aguirre-Treviño D, Karr-de-León SF. 2017. Revisiting molecular cloning to solve genome sequencing project conflicts. *J. Microbiol. Biotechnol. Food Sci*. 7:1157–1160.
- Bing J, Han P-J, Liu W-Q, Wang Q-M, Bai F-Y. 2014. Evidence for a Far East Asian origin of lager beer yeast. *Curr Biol*. 24:R380–R381.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 27:578–579.
- Bolat I, Walsh MC, Turtoi M. 2008. Isolation and characterization of two new lager yeast strains from the WS34/70 population. *Roum. Biotechnol. Lett*. 6:62–73.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
- Brouwers N, Brickwedde A, Gorter de Vries AR, van den Broek M, Weening SM, et al. 2019. The genome sequences of Himalayan *Saccharomyces eubayanus* revealed genetic markers explaining heterotic maltotriose consumption by hybrid *Saccharomyces pastorianus*. *Appl. Environ. Microbiol*. 85:e01516–e01519.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.
- Contreras-Moreira B, Vinuesa P. 2013. GET\_HOMOLOGUES a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol*. 79:7696–7701.
- De León-Medina PM, Elizondo-González R, Damas-Buenrostro LC, Geertman J-M, Van den Broek M, et al. 2016. Genome annotation of a *Saccharomyces* sp. lager brewer’s yeast. *Genom Data*. 9:25–29.
- Dostálek P, Kvasnička J, Cejnar R, Vohanka J, Mokrejš M, et al. 2013. Sequencing the genome of bottom brewer’s yeast. *Kvasný Průmysl*. 59:313–316.

- Dunn B, Sherlock G. 2008. Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res.* 18:1610–1623.
- Eizaguirre JI, Peris D, Rodríguez ME, Lopes CA, Ríos PDL, et al. 2018. Phylogeography of the wild Lager-brewing ancestor (*Saccharomyces eubayanus*) in Patagonia. *Environ Microbiol.* 20:3732–3743.
- Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, et al. 2014. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3 (Bethesda)*. 4:389–398.
- Farrar K, Donnison IS. 2007. Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. *Nat Protoc.* 2:1661–1674.
- Fay JC, Liu P, Ong GT, Dunham MJ, Cromie GA, et al. 2019. A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. *PLoS Biol.* 17:e3000147.
- Gallone B, Steensels J, Mertens S, Dzialo MC, Gordon JL, et al. 2019. Interspecific hybridization facilitates niche adaptation in beer yeast. *Nat Ecol Evol.* 3:1562–1575.
- Gayevskiy V, Goddard MR. 2016. *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ Microbiol.* 18:1137–1147.
- Gibson BR, Storgårds E, Krogerus K, Vidgren V. 2013. Comparative physiology and fermentation performance of Saaz and Froberg lager yeast strains and the parental species *Saccharomyces eubayanus*. *Yeast.* 30:255–266.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, et al. 1996. Life with 6000 Genes. *Science.* 274:546–567.
- González SS, Barrio E, Gafner J, Querol A. 2006. Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res.* 6:1221–1234.
- González SS, Barrio E, Querol A. 2008. Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. *Appl Environ Microbiol.* 74:2314–2320.
- Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. *Bioinformatics.* 29:2936–2937.
- Gorter de Vries AR, Knibbe E, van Roosmalen R, van den Broek M, de la Torre Cortés P, et al. 2020. Improving industrially relevant phenotypic traits by engineering chromosome copy number in *Saccharomyces pastorianus*. *Front Genet.* 11:518.
- Gorter de Vries AR, Pronk JT, Daran J-MG. 2017. Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts. *Appl. Environ. Microbiol.* 83:e03206–e03216.
- Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics.* 26:2334–2335.
- Haase SB, Reed SI. 2002. Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle.* 1:117–121.
- Hahne F, Ivanek R. 2016. Visualizing genomic data using Gviz and bioconductor. In: E Mathé, S. Davis *Statistical Genomics: Methods and Protocols*. New York: Humana Press. p. 335–351.
- Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, et al. 2009. flowCore: a bioconductor package for high throughput flow cytometry. *BMC Bioinform.* 10:106.
- Hewitt SK, Donaldson IJ, Lovell SC, Delneri D. 2014. Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS One.* 9:e92203.
- Holt C, Yandell M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* 12:491
- Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, et al. 2013. REAPR: a universal tool for genome assembly evaluation. *Genome Biol.* 14:R47.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software Version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kent WJ. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* 12:656–664.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics.* 5:59.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.
- Langdon QK, Peris D, Baker EP, Opulente DA, Nguyen H-V, et al. 2019. Fermentation innovation through complex hybridization of wild and domesticated yeasts. *Nat Ecol Evol.* 3:1576–1586.
- Langdon QK, Peris D, Eizaguirre JI, Opulente DA, Buh KV, et al. 2020. Postglacial migration shaped the genomic diversity and global distribution of the wild ancestor of lager-brewing hybrids. *PLoS Genet.* 16:e1008680.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 25:1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, 1000 Genome Project Data Processing Subgroup, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics.* 25:2078–2079.
- Li L, Stoekert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, et al. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci U S A.* 108:14539–14544.
- Liti G, Peruffo A, James SA, Roberts IN, Louis EJ. 2005. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. *Yeast.* 22:177–192.
- Liu C, Niu C, Zhao Y, Tian Y, Wang J, et al. 2018. Genome analysis of the yeast M14 an industrial brewing yeast strain widely used in China. *J Am Soc Brew Chem.* 76:223–235.
- Marie-Nelly H, Marbouty M, Courmac A, Flot J-F, Liti G, et al. 2014. High-quality genome (re)assembly using chromosomal contact data. *Nat Commun.* 5:5695.
- Martin G, Baurens F-C, Droc G, Rouard M, Cenci A, et al. 2016. Improvement of the banana “*Musa acuminata*” reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genom.* 17:243
- Martini AV, Kurtzman CP. 1985. Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces Sensu Stricto*. *Int. J. Syst. Bacteriol.* 35:508–511.
- Martini AV, Martini A. 1987. Three newly delimited species of *Saccharomyces sensu stricto*. *Antonie Van Leeuwenhoek.* 53:77–84.
- Masneuf I, Hansen J, Groth C, Piskur J, Dubourdieu D. 1998. New hybrids between *Saccharomyces Sensu Stricto* yeast species found among wine and cider production strains. *Appl Environ Microbiol.* 64:3887–3892.
- Monerawela C, Bond U. 2017. Recombination sites on hybrid chromosomes in *Saccharomyces pastorianus* share common sequence motifs and define a complex evolutionary relationship between group I and II lager yeasts. *FEMS Yeast Res.* 17:fox047.
- Nakao Y, Kanamori T, Itoh T, Kodama Y, Rainieri S, et al. 2009. Genome sequence of the lager brewing yeast an interspecies hybrid. *DNA Res.* 16:115–129.

- Nespolo RF, Villarroel CA, Oporto CI, Tapia SM, Vega-Macaya F, et al. 2020. An Out-of-patagonia migration explains the worldwide diversity and distribution of *Saccharomyces eubayanus* lineages. *PLoS Genet.* 16:e1008777.
- Okuno M, Kajitani R, Ryusui R, Morimoto H, Kodama Y, et al. 2016. Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res.* 23:67–80.
- Pagès H, Aboyou P, Gentleman R, DebRoy S. 2019. Biostrings: Efficient manipulation of biological strings. R package version 2.52.0.
- Peris D, Belloch C, Lopandić K, Álvarez-Pérez JM, Querol A, et al. 2012. The molecular characterization of new types of *Saccharomyces cerevisiae* × *S. kudriavzevii* hybrid yeasts unveils a high genetic diversity. *Yeast.* 29:81–91.
- Peris D, Langdon QK, Moriarty RV, Sylvester K, Bontrager M, et al. 2016. Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. *PLoS Genet.* 12:e1006155.
- Peris D, Sylvester K, Libkind D, Gonçalves P, Sampaio JP, et al. 2014. Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol Ecol.* 23:2031–2045.
- Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44:e113.
- Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinform.* 47:11.12.1–11.12.34.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16:276–277.
- Salazar AN, Gorter de Vries AR, van den Broek M, Brouwers N, de la T, Cortés, et al. 2019. Chromosome level assembly and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization. *BMC Genomics.* 20:916.
- Schatz MC, Delcher AL, Salzberg SL. 2010. Assembly of large genomes using second-generation sequencing. *Genome Res.* 20:1165–1173.
- Schatz MC, Phillippy AM, Sommer DD, Delcher AL, Puiu D, et al. 2013. Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. *Brief Bioinform.* 14:213–224.
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, et al. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A.* 89:8794–8797.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30:1312–1313.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, et al. 2006. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34:W435–W439.
- Tamai Y, Momma T, Yoshimoto H, Kaneko Y. 1998. Co-existence of two types of chromosome in the bottom fermenting yeast *Saccharomyces pastorianus*. *Yeast.* 14:923–933.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 14:178–192.
- van den Broek M, Bolat I, Nijkamp JF, Ramos E, Luttk MAH, et al. 2015. Chromosomal copy number variation in *Saccharomyces pastorianus* is evidence for extensive genome dynamics in industrial lager brewing strains. *Appl Environ Microbiol.* 81:6253–6267.
- Van P, Jiang W, Gottardo R, Finak G. 2018. ggCyto: next generation open-source visualization software for cytometry. *Bioinformatics.* 34:3951–3953.
- Walther A, Hesselbart A, Wendland J. 2014. Genome sequence of *Saccharomyces carlsbergensis* the world's first pure culture lager yeast. *G3 (Bethesda).* 4:783–793.
- Weber JL, Myers EW. 1997. Human whole-genome shotgun sequencing. *Genome Res.* 7:401–409.
- Wei X, Xu Z, Wang G, Hou J, Ma X, et al. 2017. pBACode: a random-barcode-based high-throughput approach for BAC paired-end sequencing and physical clone mapping. *Nucleic Acids Res.* 45:e52.
- Yang X, Yang Y, Ling J, Guan J, Guo X, et al. 2020. A high-throughput BAC end analysis protocol (BAC-anchor) for profiling genome assembly and physical mapping. *Plant Biotechnol J.* 18:364–372.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 8:28–36.

Communicating editor G. W. Brown