# SUPPLEMENTAL TABLES

## Table S1. General information of the *de novo S. pastorianus* assemblies used in this study

| Strain | Group | Assembly | Technology | Assembler | Scaffolding | Level | Seqs | N50 (bp) | Size (Mb) | %GC | Cov |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CBS 1513 (Walther *et al* 2014)[a] | 1 | GCA_000586595.1 | 454 GS FLX | gsAssembler/ Newbler 2.6 | Synteny | Scaffolds | 77 | 552,537 | 19.37 | 39.3 | 18x |
| CBS 1513 (Okuno *et al* 2016) | 1 | GCA_001515445.1 | Illumina Miseq | Platanus v.1.2.4 | Platanus v.1.2.4 | Contigs | 323 | 174,822 | 19.15 | 39.2 | 360x |
| CBS 1503 (Okuno *et al* 2016) | 1 | GCA_001515425.1 | Illumina Miseq | Platanus v.1.2.4 | Platanus v.1.2.4 | Contigs | 716 | 94,203 | 17.16 | 34.4 | 440x |
| CBS 1503 (Unpub)[a] | 1 | GCA_002557695.1 | 454 | Newbler 2.6 | - | Contigs | 720 | 88,533 | 16.41 | 39.3 | 18x |
| CBS 1538 (Okuno *et al* 2016)[a] | 1 | GCA_001515465.1 | Illumina Miseq | Platanus v.1.2.4 | Platanus v.1.2.4 | Contigs | 387 | 122,314 | 14.34 | 39.6 | 490x |
| WS 34/70 (Nakao *et al* 2009) | 2 | GCA_000182115.1 | Sanger | PCPA | - | Scaffolds | 3566 | 12,548 | 22.38 | 39.1 | 6x |
| WS 34/70 (Walther *et al* 2014) | 2 | GCA_000586535.1 | Illumina MiSeq v2 | CASAVA v. 1.8.2 | - | Contigs | 1358 | 63,323 | 22.96 | 39 | 18x |
| WS 34/70 (Okuno *et al* 2016)[a] | 2 | GCA_001515485.1 | Illumina Miseq | Platanus v.1.2.4 | Platanus v.1.2.4 | Contigs | 872 | 113,674 | 22.27 | 39.1 | 250x |
| CBS 1483 (van den Broek *et al* 2015)[a] | 2 | GCA_000805465.1 | Illumina HiSeq2000 | gsAssembler/ Newbler 2.6 | SSPACE 2.0 | Contigs | 878 | 53,101 | 21.92 | 39.1 | 52.2x |
| 790 (De León-Medina *et al* 2016)[a] | 2[b] | GCA_001640265.1 | FLX 454 Titanium/Illumina MiSeq | gsAssembler/ Newbler 2.6 | - | Contigs | 1098 | 85,203 | 22.34 | 39.1 | 37x |
| CCY48-91 (Dostálek *et al* 2013)[a] | 2 | GCA_000287895.1 | 454 FLX | gsAssembler/ Newbler 2.5p1 | - | Contigs | 2425 | 60,494 | 24.21 | 41.6 | 30x |
| M14 (Liu *et al* 2018)[a] | 2[b] | GCA_002375215.1 | Illumina and PacBio | SPAdes v. 3.7.1 | - | Contigs | 131 | 575,172 | 22.77 | 39.0 | 690x |

[a]Used in phylogenetic analysis.

[b]Lager yeast group inferred from this study.

**Table S2. BESs alignments' results with the *S. pastorianus* 790 assembly (De León-Medina *et al* 2016) using all the implemented tools in the developed pipeline, categorized by alignment type and expressed as percentages (%)**

| Alignment type | BOWTIE (Langmead *et al* 2009) | NUCMER (Kurtz *et al* 2004) | BOWTIE2 (Langmead and Salzberg 2012) | BWA (Li and Durbin 2009) | BLAST (Altschul *et al* 1990) | MEGABLAST (Altschul *et al* 1990) | BLAT (Kent 2002) |
|---|---|---|---|---|---|---|---|
| Paired-end | 8.24 | 14.09 | 16.73 | 16.78 | 16.78 | 15.58 | 16.78 |
| Opposite | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Positive | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Negative | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Unpaired-end | 28.07 | 56.69 | 66.83 | 72.68 | 72.68 | 73.88 | 72.68 |
| >100 Kb | 0.4 | 1.35 | 1.17 | 1.32 | 1.32 | 2.42 | 1.3 |
| <100 Kb | 27.67 | 55.34 | 65.66 | 71.35 | 71.35 | 71.45 | 71.38 |
| Singletons | 25.25 | 14.36 | 9.74 | 7.77 | 7.79 | 7.79 | 7.79 |
| BES Total | 61.56 | 85.14 | 93.31 | 97.23 | 97.25 | 97.25 | 97.25 |

**Table S3. Primers names, sequences, reference sequence localization and selected clones used for PCR amplification of gaps and homeologous translocations**

| Primer | Sequence (5'-3') | Localization | Selected clone |
|---|---|---|---|
| Gap1_Fw | TTCCACTCTCAGCTTATTGG | contig00086 | Plate_01B08 |
| Gap1_Rv | TGGGTTGATCTATTCACTCC | contig00042 | |
| Gap2_Fw | TCTACAGTCAAACCAGTAGC | contig00132 and scf17 | Plate_03C10 |
| Gap2_Rv | TCTAACCTCTCATCTTTAGGC | contig00066 and scf22 | |
| Gap3/ScXI-SeXI_Fw | TGTGATAGAACTGGTAGGAAACCGTT | contig00011 and ScXI | Plate_22C08 |
| Gap3/ScXI-SeXI_Rv | GATCTCCTCGGACCACACAC | contig00053 and SeXI | |
| ScXI_Fw | CTTACCAAGACTGAATGAGGATGACCG | *Sc*XI | Plate_03F06 |
| ScXI_Rv | GCGGATAGTTTGAGCTAATCTCTTCAG | *Sc*XI | |
| SeXI_Fw | GCTAGTGGGTAACCGCCTAGTG | *Se*XI | Plate_14B04 |
| SeXI_Rv | CAGCAACCTGTCATCTTCGTTCAGTC | *Se*XI | |
| SeXIII-ScXIII_Fw | CTTGTCTAGCGCTTCGGTA | *Se*XIII | Plate_14G10 |
| SeXIII-ScXIII_Rv | GGCCGATACGAGTGTCAAA | *Sc*XIII | |
| ScXIII_Fw | ATGGAATGACTGGGTTCGCAC | *Sc*XIII | Plate_12A05 |
| SeXIII-ScXIII_Rv | GGCCGATACGAGTGTCAAA | *Sc*XIII | |

**Table S4. BESs alignments' types with the *S. pastorianus* 790 assembly (De León-Medina *et al* 2016) using BOWTIE2, classified by pairs**

| Alignment type | Both BESs (%) | One BES (%) | Quality BESs (%) |
|---|---|---|---|
| Aligned[a] | 3439 (85.89) | 297 (7.42) | 3736 (93.31) |
| Paired-end | 670 (16.73) | 0 (0) | 670 (16.73) |
| Unpaired-end | 2676 (66.83) | 0 (0) | 2676 (66.83) |
| Single | 93 (2.32) | 297 (7.42) | 390 (9.74) |
| Not aligned | 147 (3.67) | 121 (3.02) | 268 (6.69) |
| Yeast positive[b] | 145 (3.62) | 14 (0.35) | 159 (3.97) |
| Vector positive[b,c] | 2 (0.05) | 107 (2.67) | 109 (2.72) |
| SUBTOTAL | 3586 (89.56%) | 418 (10.44) | 4004 (100) |

[a]Negative, positive, and opposite BESs are not shown because they are 0%.

[b]Alignments performed with BLAST.

[c]Not reported in LIBGSS_039348.

**Table S5. Homeologous translocations in *S. pastorianus* 790 and other group-2 lager yeasts strains detected in this and other studies**

| Chromosomes (approximate loci) | 790 (This study) | 790 (Monerawela and Bond 2017) | WS 34/70 (Nakao *et al* 2009) | WS 34/70 (Monerawela and Bond 2017) | CBS 1483 (van den Broek *et al* 2015) | CBS 1483 (Monerawela and Bond 2017) |
|---|---|---|---|---|---|---|
| *Se*III-*Sc*III (*MAT* locus) | X | | X | X | X | |
| *Sc*VII-*Se*VII (YGL173 / *XRN1*) | X | X | X | X | X | X |
| *Sc*VII-*Se*VII (YGR285C / *ZUO1*) | X | X | | X | X | X |
| *Sc*X-*Se*X (YJR009C / *TDH2* (1)) | X | X | X | X | X | X |
| *Se*X-*Sc*X (YJR009C / *TDH2* (2)) | X | | X | X | | X |
| *Sc*XI-*Se*XI (YKL045W / *PRI2*) | X | X | X | X | | |
| *Sc*XII and *Se*XII (from YLR410W / *VIP1* to YLR413C-YLR414C) | | X | | | X | X |
| *Se*XIII-*Sc*XIII (YMR302C / *YME2*) | X | X | X | X | X | X |
| Mosaic *Se*XVI (YPL240C / *HSP82*) | X | X | X | X | X | X |
| Mosaic *Sc*XVI (YPR160W / *GPH1*) | X | X | X | X | X | X |
| Mosaic *Sc*XVI and Mosaic *Se*XVI (YPR191W / *QCR2*) | X | X | X | X | X | X |

**Table S6. Approximate chromosome copy number of *S. pastorianus* 790**

| Chromosome | Size | Approximate copies |
|---|---|---|
| **A. Chromosomes with homeologous pairs** | | |
| *Sc*I | 230,218 | 2 |
| *Se*I | 175,444 | 2 |
| *Sc*V | 576,874 | 3 |
| *Se*V | 588,913 | 2 |
| *Sc*VI | 270,161 | 1 |
| *Se*VI | 264,757 | 3 |
| *Sc*IX | 439,888 | 2.5 |
| *Se*IX | 401,362 | 2.5 |
| *Sc*XII | 1,078,177 | 3 |
| *Se*XII | 1,033,978 | 2 |
| *Sc*XIV | 784,333 | 2 |
| *Se*XIV | 768,019 | 2 |
| **B. Chromosomes without homeologous pairs** | | |
| *Sc*II | 813,184 | 3 |
| *Sc*IV | 1,531,933 | 2 |
| *Sc*VIII | 562,643 | 3 |
| *Sc*XV | 1,091,291 | 3 |
| *Se*II-*Se*IV | 1,274,804 | 3 |
| *Se*IV-*Se*II | 982,538 | 2 |
| *Se*VIII-*Se*XV | 813,827 | 2 |
| *Se*XV-*Se*VIII | 741,893 | 2 |
| **C. Chromosomes with chimeric pairs** | | |
| *Sc*III | 316,620 | 2 |
| *Se*III-*Sc*III | 305,615 | 2 |
| *Sc*VII | 1,090,940 | 2 |
| *Sc*VII-*Se*VII | 1,051,730 | 2 |
| *Sc*X | 745,751 | 1 |
| *Se*X | 747,934 | 1 |
| *Sc*X-*Se*X | 745,751 | 1 |
| *Se*X-*Sc*X | 747,934 | 1 |
| *Sc*XI | 666,816 | 2 |
| *Se*XI | 632,881 | 2 |
| *Sc*XI-*Se*XI | 666,816 | 1 |
| *Sc*XIII | 924,431 | 2 |
| *Se*XIII-*Sc*XIII | 953,685 | 3 |
| Mosaic *Sc*XVI | 948,066 | 3 |
| Mosaic *Se*XVI | 896,107 | 2 |
| Total | 55,790,387 | 74 |

## Table S7. Comparison of methods used to study the *S. pastorianus* genome and other genomes

| Method | Advantages | Disadvantages |
|---|---|---|
| Synteny with parental genomes (Walther *et al* 2014) | - Chromosome size contigs.<br>- No additional equipment or experiments required, besides initial whole genome sequencing and assembly. | - Loss of heterozygous sites information.<br>- Requires a high finishing level reference sequence. |
| Chromosome copy number quantification and chimeric chromosome detection using sequencing reads (Hewitt *et al* 2014; Walther *et al* 2014; van den Broek *et al* 2015) | - Assembly not required.<br>- Nucleotide resolution level of recombination sites (Hewitt *et al* 2014).<br>- No additional equipment or experiments required, besides whole genome sequencing. | - Requires a reference sequence.<br>- Sequencing libraries may not be publicly available.<br>- Cannot detect recombinations in chromosomes with equal copy number.<br>- Requires high sequencing depth for accurate chromosome quantification (Gorter de Vries *et al* 2017). |
| Whole genome short-reads sequencing and assembly (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; Gallone *et al* 2019) | - Generates important sequence assembly information. | - When applied to hybrid genomes, generates highly fragmented assemblies difficult to scrutinize (Pryszcz and Gabaldón 2016).<br>- Repeated sequences hiders the sequencing and assembly process (Schatz *et al* 2010; Alkan *et al* 2011).<br>- Homologous regions from different chromosomes collapse into one consensus (Pryszcz and Gabaldón 2016).<br>- Alternative alleles are not represented in the final assembly (Pryszcz and Gabaldón 2016).<br>- Sequencing libraries cannot be preserved for verifying experiments. |
| Scaffolding with short-insert mate pairs (van den Broek *et al* 2015; De León-Medina *et al* 2016; Okuno *et al* 2016) | - Generates more contiguos assemblies.<br>- No additional equipment or experiments required, besides whole genome sequencing and assembly. | - Sequencing libraries may not be publicly available.<br>- Requires specially paired (mate pairs) reads.<br>- Can introduce errors due to long repeated regions (Martin *et al* 2016).<br>- All alternative alleles are not represented in the final scaffolded assembly. See ratio parameter -a in Boetzer *et al* (2011).<br>- Cannot still generate complete chromosomes. |
| Bioinformatic finishing and phasing (Gordon and Green 2013; Hunt *et al* 2013; Schatz *et al* 2013; Pryszcz and Gabaldón 2016; Fay *et al* 2019) | - No additional equipment or experiments required, besides whole genome sequencing and assembly. | - *In silico* approach only; computational hypothesis that needs to be tested.<br>- Manual finishing is time consuming. |
| Whole genome long-reads sequencing and assembly (Liu *et al* 2018; Salazar *et al* 2019) | - Near chromosome-sized contigs after the assembly. | - Not widely available yet. |
| BACs library and BESs (This study) | - Independent and experimental alternative information to the whole genome sequencing process.<br>- Pairing information can be used for scaffolding.<br>- Long-insert information can span long repeated regions.<br>- Single allele information can be used to identify all alternative alleles and offers an improved detection of homeologous translocations.<br>- Clones are preserved and used for verification experiments (Ammiraju *et al* 2005), specially on heterozygous regions (Barrera-Saldaña *et al* 2017) (different gene alleles could be studied by this method). | - Additional experiments required.<br>- Clones with unrelated ligated insert may give ambiguous results.<br>- Centromeres and telomeres are not easily cloned.<br>- Requires a sequence reference.<br>- Sanger sequencing to obtain BESs can be laborious and expensive. |

# SUPPLEMENTAL FIGURES

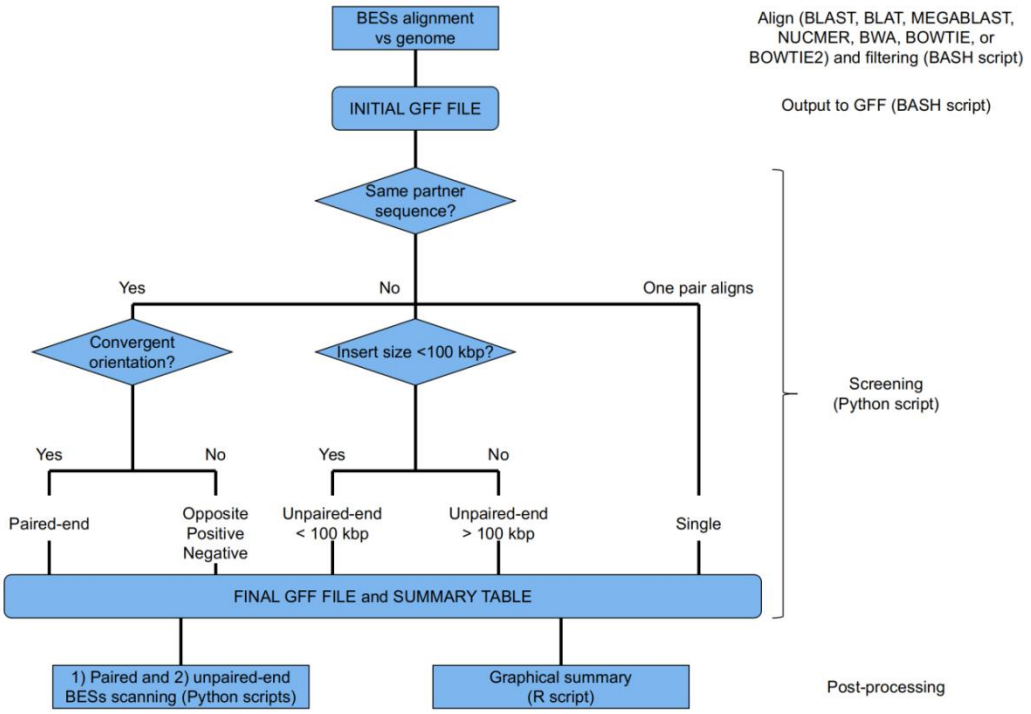



**Fig. S1. BESs comparison with *S. pastorianus* genome assemblies pipeline.** BESs can be aligned with the *S. pastorianus* genomes using one of the implemented alignment tools using a BASH script (bes_analysis.sh). The output alignment file is filtered to select either the longest or primary alignment (depending on the tool used) and then the file is converted to GFF format. This initial GFF file is screened with a Python script (screening.py) to classify BESs types. The results are a final GFF file, formatted to aid visual inspection in the Integrative Genomics Viewer (Thorvaldsdóttir *et al* 2013), and a summary table. Additionally, two Python scripts (sliding_paired.py and sliding_unpaired.py) to scan for 1) regions not spanned by paired-end BESs and 2) with consistent unpaired-end BESs alignments can be ran. Also, an R script to generate a graphical summary (graphic_summary.R) can be applied. All scripts are available at https://github.com/Lriego/BES-analysis.

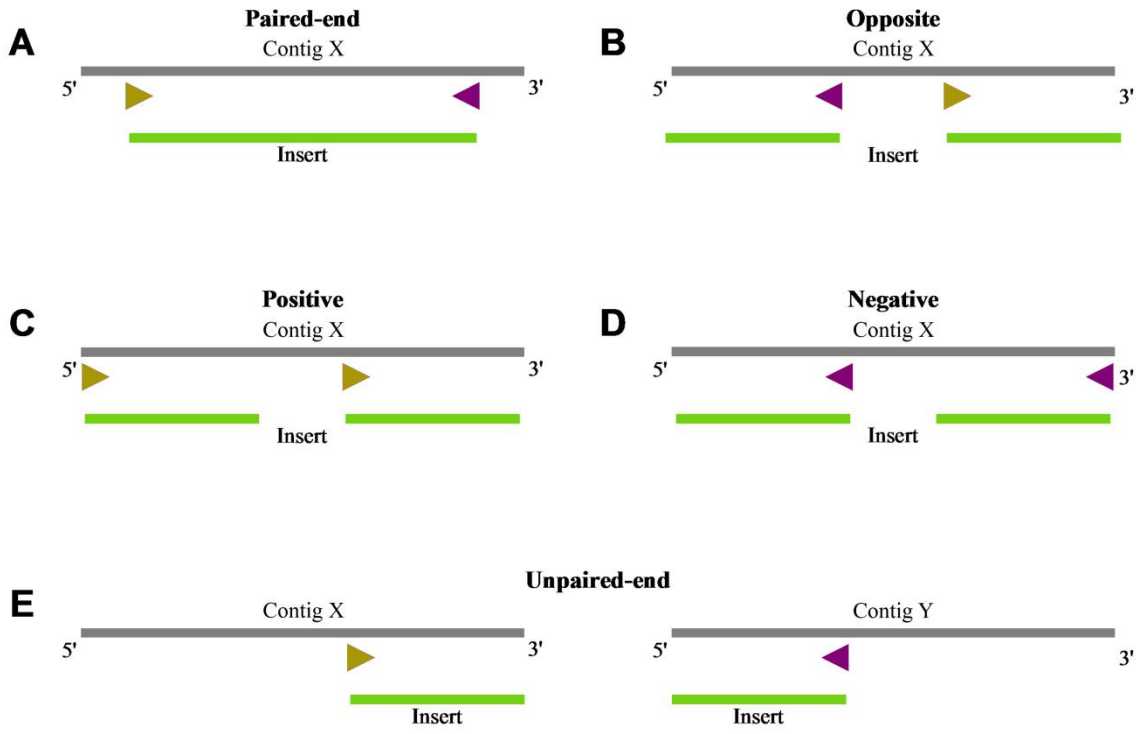



**Fig. S2. BESs alignments classification according to the orientation between pairs.** BESs pairs alignments can be (A) paired-end if the orientation according to the reference is convergent, (B) opposite when it is divergent, (C) positive, and (D) negative, when they are in the same direction, and (E) unpaired-end, when they align in different contigs. Grey: contigs; yellow arrows: positively aligned BESs; purple arrows: negatively aligned BESs; green: BAC insert.

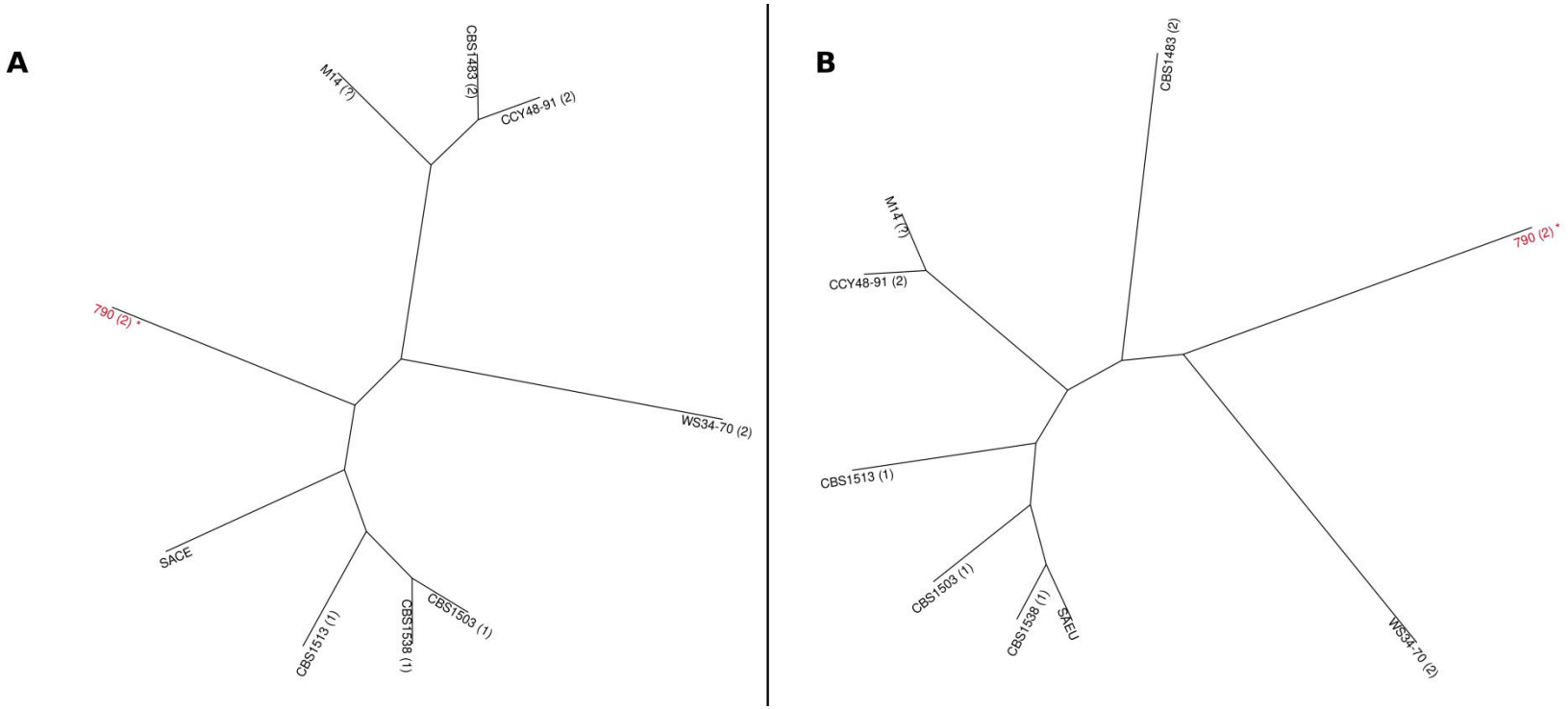



**Fig. S3. Unrooted representation of the phylogenetic relationships of *Saccharomyces* sp. 790 and seven lager beer-related yeast strains (without distances).** All yeast assemblies were annotated with MAKER2 pipeline (Holt and Yandell 2011) and the orthologous genes from each sub-genome segregated 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 in parentheses are the lager group reported for that yeast (see Table S1 for references information). SACE: *S. cerevisiae*; SAEU: *S. eubayanus*. ***** Study yeast.

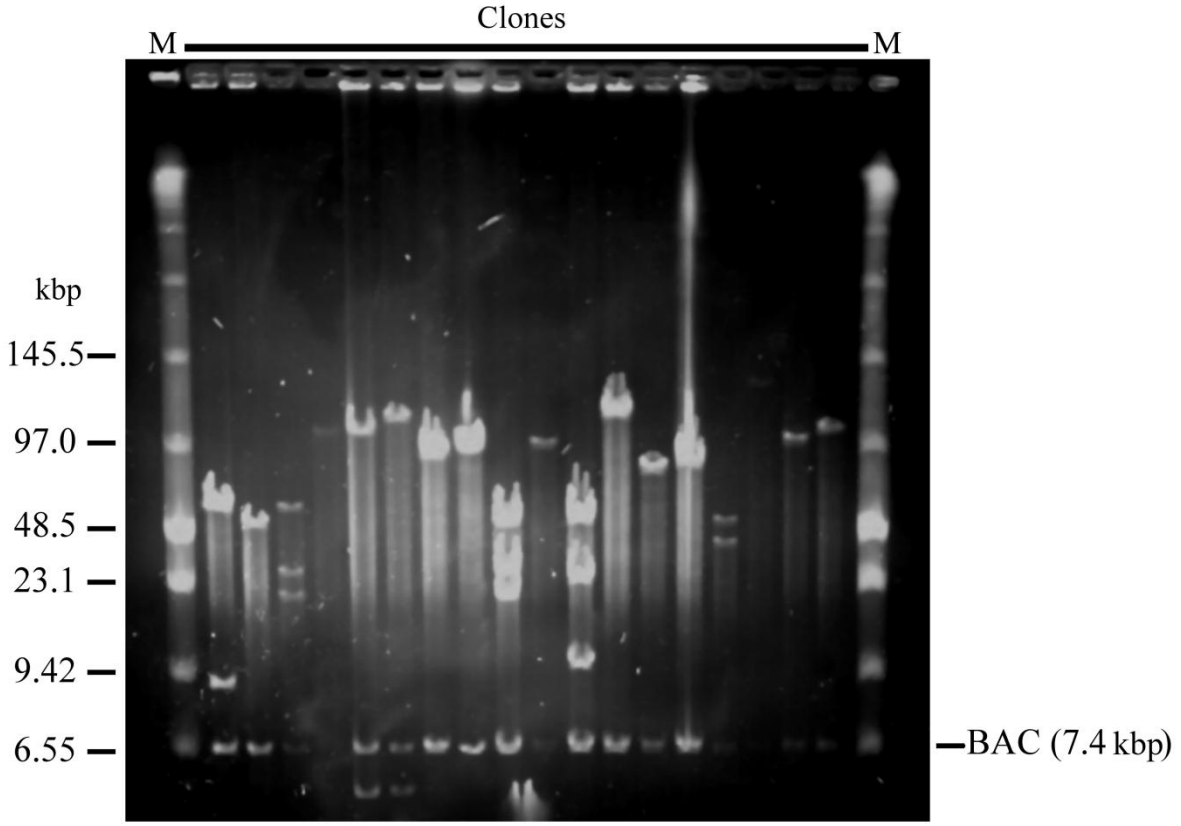
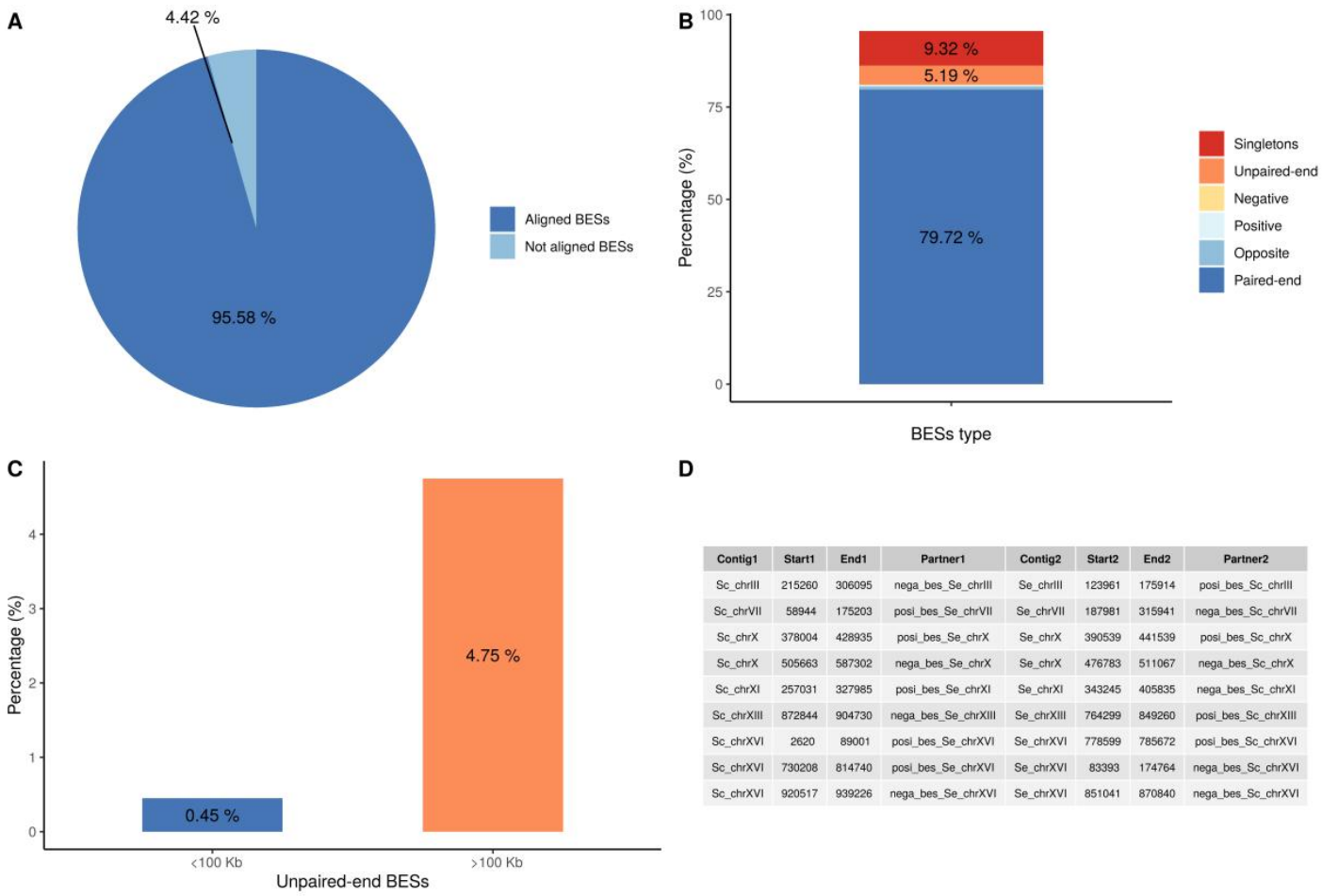



**Fig. S4. *Not*I characterization of the BAC genomic library.** Clones were randomly selected and digested with *Not*I. Insert release was characterized with pulsed field gel electrophoresis. The common band at 7.4 kbp in all the samples is the BAC backbone. First and last lanes (M) are the molecular weight marker.




| Contig1 | Start1 | End1 | Partner1 | Contig2 | Start2 | End2 | Partner2 |
|---|---|---|---|---|---|---|---|
| Sc_chrIII | 215260 | 306095 | nega_bes_Se_chrIII | Se_chrIII | 123961 | 175914 | posi_bes_Sc_chrIII |
| Sc_chrVII | 58944 | 175203 | posi_bes_Se_chrVII | Se_chrVII | 187981 | 315941 | nega_bes_Sc_chrVII |
| Sc_chrX | 378004 | 428935 | posi_bes_Se_chrX | Se_chrX | 390539 | 441539 | posi_bes_Sc_chrX |
| Sc_chrX | 505663 | 587302 | nega_bes_Se_chrX | Se_chrX | 476783 | 511067 | nega_bes_Sc_chrX |
| Sc_chrXI | 257031 | 327985 | posi_bes_Se_chrXI | Se_chrXI | 343245 | 405835 | nega_bes_Sc_chrXI |
| Sc_chrXIII | 872844 | 904730 | nega_bes_Se_chrXIII | Se_chrXIII | 764299 | 849260 | posi_bes_Sc_chrXIII |
| Sc_chrXVI | 2620 | 89001 | posi_bes_Se_chrXVI | Se_chrXVI | 778599 | 785672 | posi_bes_Sc_chrXVI |
| Sc_chrXVI | 730208 | 814740 | posi_bes_Se_chrXVI | Se_chrXVI | 83393 | 174764 | nega_bes_Sc_chrXVI |
| Sc_chrXVI | 920517 | 939226 | nega_bes_Se_chrXVI | Se_chrXVI | 851041 | 870840 | nega_bes_Sc_chrXVI |

**Fig. S5. BESs alignments with the combined parental-like genomes of *S. cerevisiae* S288c (Engel *et al* 2014) and *S. eubayanus* FM1318 (Baker *et al* 2015) as reference.** (A) Percentage of aligned BESs. (B) Percentage of the different BESs alignments types. Negative, positive, and opposite BESs bars are not observed because they are less than 1%. (C) Percentage of unpaired-end BES according to the insert size. (D) Regions with consistent inter-parental unpaired-end BESs aligned (File S4) (graphic_summary.R script was modified to show all resulting regions).

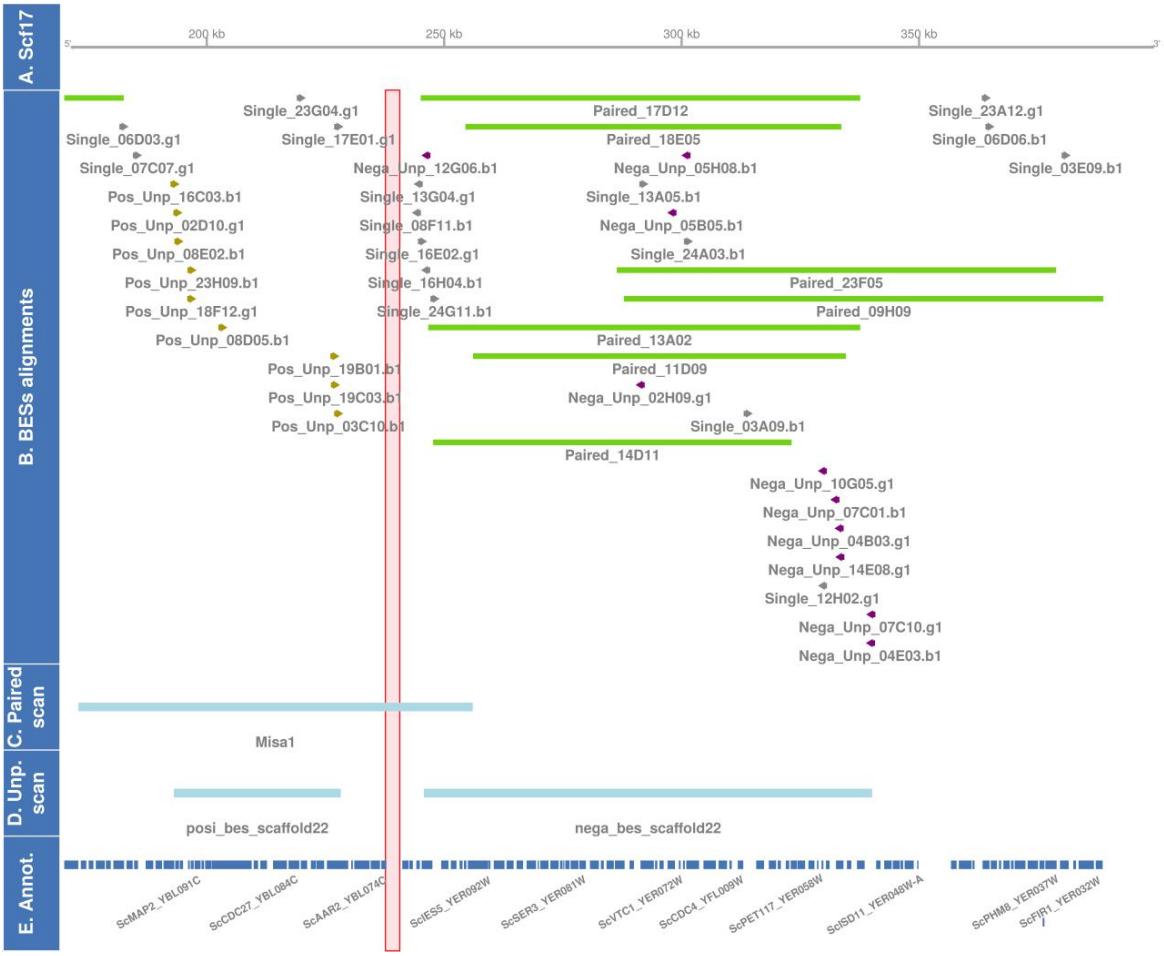



**Fig. S6. Misassembly in an *S. pastorianus* 790 assembly improved by scaffolding with short insert libraries (De León-Medina *et al* 2016).** (A) Scaffold 17. (B) BESs alignments. (C) Paired-end BESs alignments scanning results. (D) Unpaired-end BESs alignments scanning results. (E) Genome annotation. The visualized region presented BESs aligned as paired-end (represented with their simulated insert as green lines, 'Paired' prefix), unpaired-end aligned positively (yellow arrows, 'Pos_Unp' prefix), unpaired-end aligned negatively (purple arrows, 'Neg_Unp' prefix), and single (gray arrows, 'Single' prefix). The paired-end BESs scanning script successfully detected a region (Misa1) not sustained by this type of alignments. The unpaired-end BESs scanning script detected two regions with consistent unpaired-end BESs pointing to scaffold 22 (posi_bes_scaffold22 and nega_bes_scaffold22). A sub-region in between regions posi_bes_scaffold22 and nega_bes_scaffold22 had a portion of unknown sequence or a gap (red vertical bar). The inspected region was associated with a false non-homologous translocation in the *S. cerevisiae* sub-genome (note change of chromosome B to E and the 'Sc' prefix in the systematic gene name of the annotation).

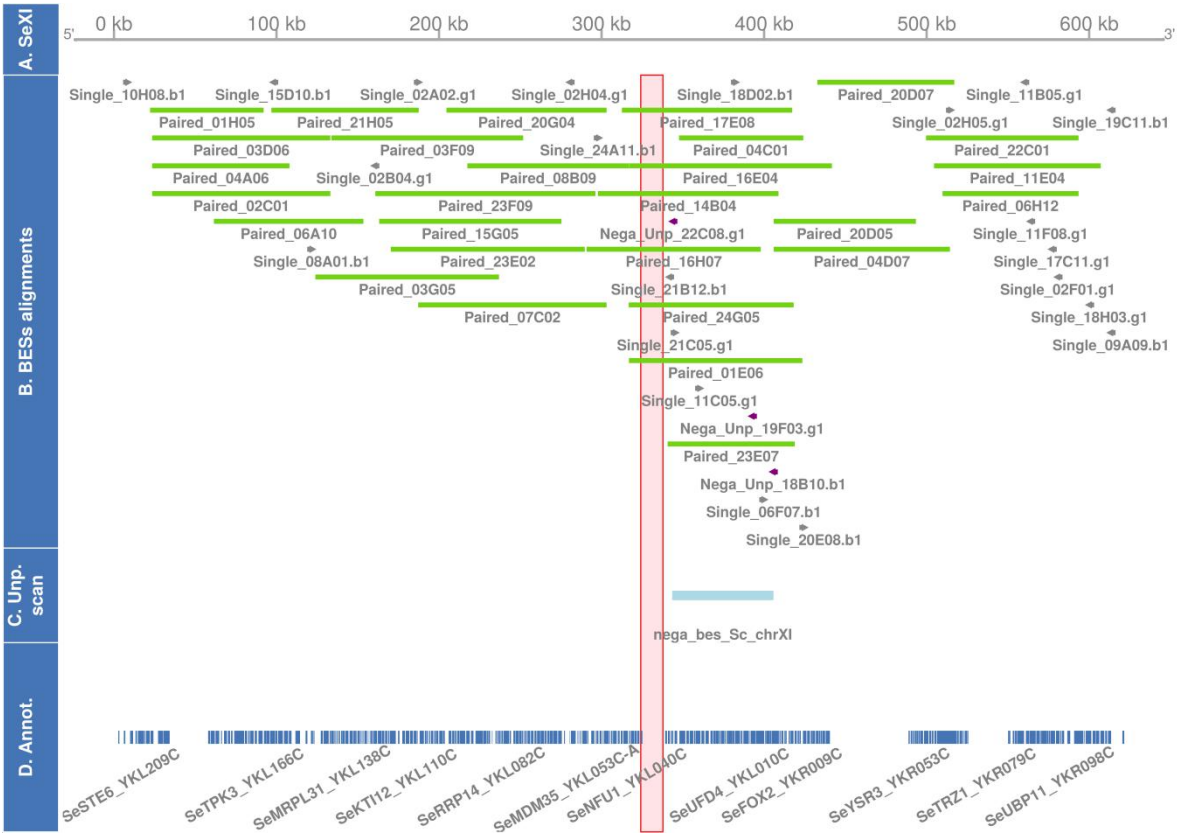



**Fig. S7. Homeologous translocation detection in the joined *S. pastorianus* 790 assembly.** (A) Chromosome *Se*XI. (B) BESs alignments. (C) Unpaired-end BESs alignments scanning results. D Genome annotation. Chromosome *Se*XI presented paired-end BESs alignments (represented with their simulated insert as green lines, 'Paired' prefix), unpaired-end BESs aligned negatively (purple arrows, 'Neg_Unp' prefix), and single BESs (gray arrows, 'Single' prefix). The entire chromosome was covered by paired-end BESs alignments; nonetheless, the unpaired-end BESs scanning script detected a region with consistent unpaired-end BESs pointing a join to chromosome *Sc*XI (nega_bes_Sc_chrXI) and indicating an homeologous translocation. The closest gap towards the detected join (red vertical bar) was used to split the sequence. The same procedure was use with the rest of the chromosomes presenting homeologous translocations.

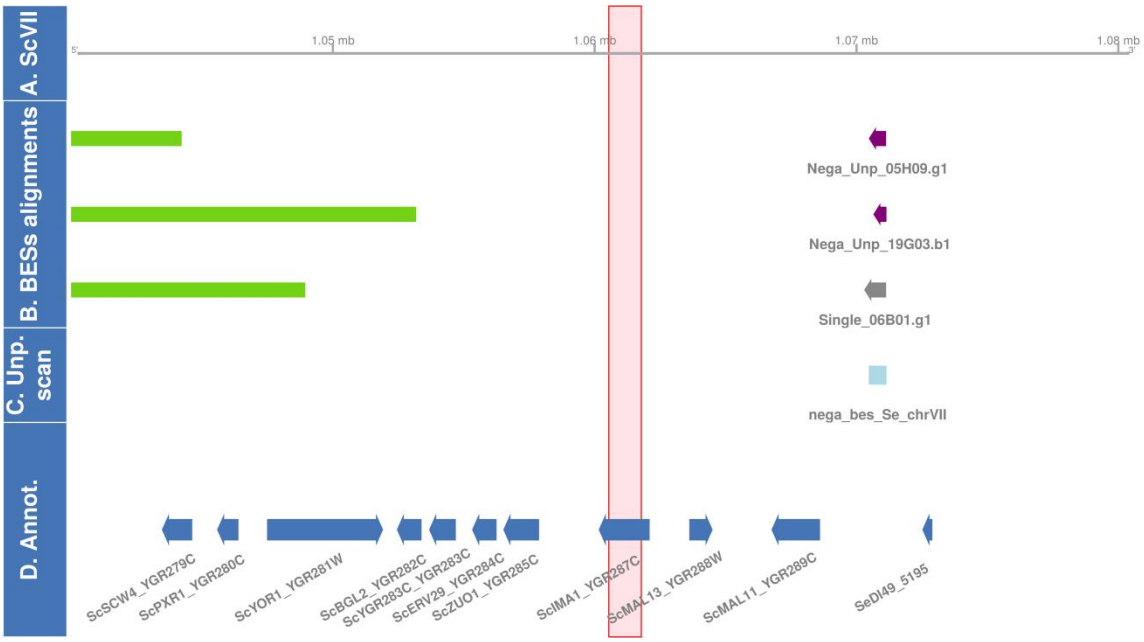



**Fig. S8. Unpaired-end BESs alignments near the YGR285C/*ZUO1* gene in the *S. pastorianus* 790 joined assembly.** (A) Chromosome *Sc*VII. (B) BESs alignments. (C) Unpaired-end BESs alignments scanning results. (D) Genome annotation. The visualized region presented BESs aligned as paired-end (represented with their simulated insert as green lines), unpaired-end BESs aligned negatively (purple arrows, 'Neg_Unp' prefix), and single BESs (gray arrows, 'Single' prefix). The unpaired-end BESs scanning results showed that the two unpaired-end BESs aligning in the region pointed to chromosome *Se*VII (nega_bes_Se_chrVII), suggesting an homeologous translocation. The chromosome was splitted in the next detected gap (red vertical bar) towards the direction of the unpaired-end BESs alignments and the homeologous translocation was included to the genome structure of *S. pastorianus* 790.

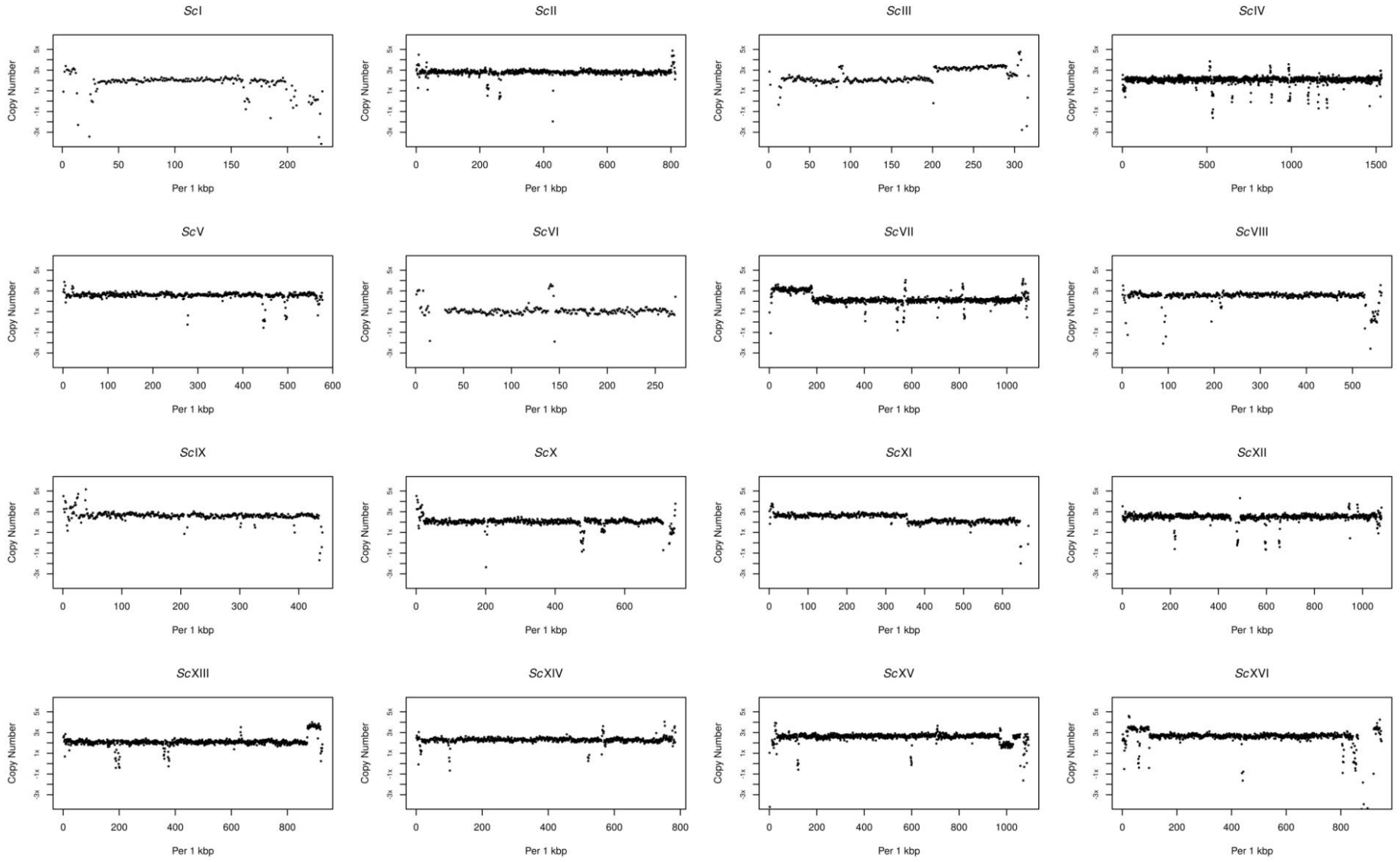



**Fig. S9. *S. pastorianus* 790 chromosome copy number of the *S. cerevisiae* sub-genome.** The estimated copy number of the 1000 bp windows of each of the chromosomes of the *S. cerevisiae* sub-genome was plotted.

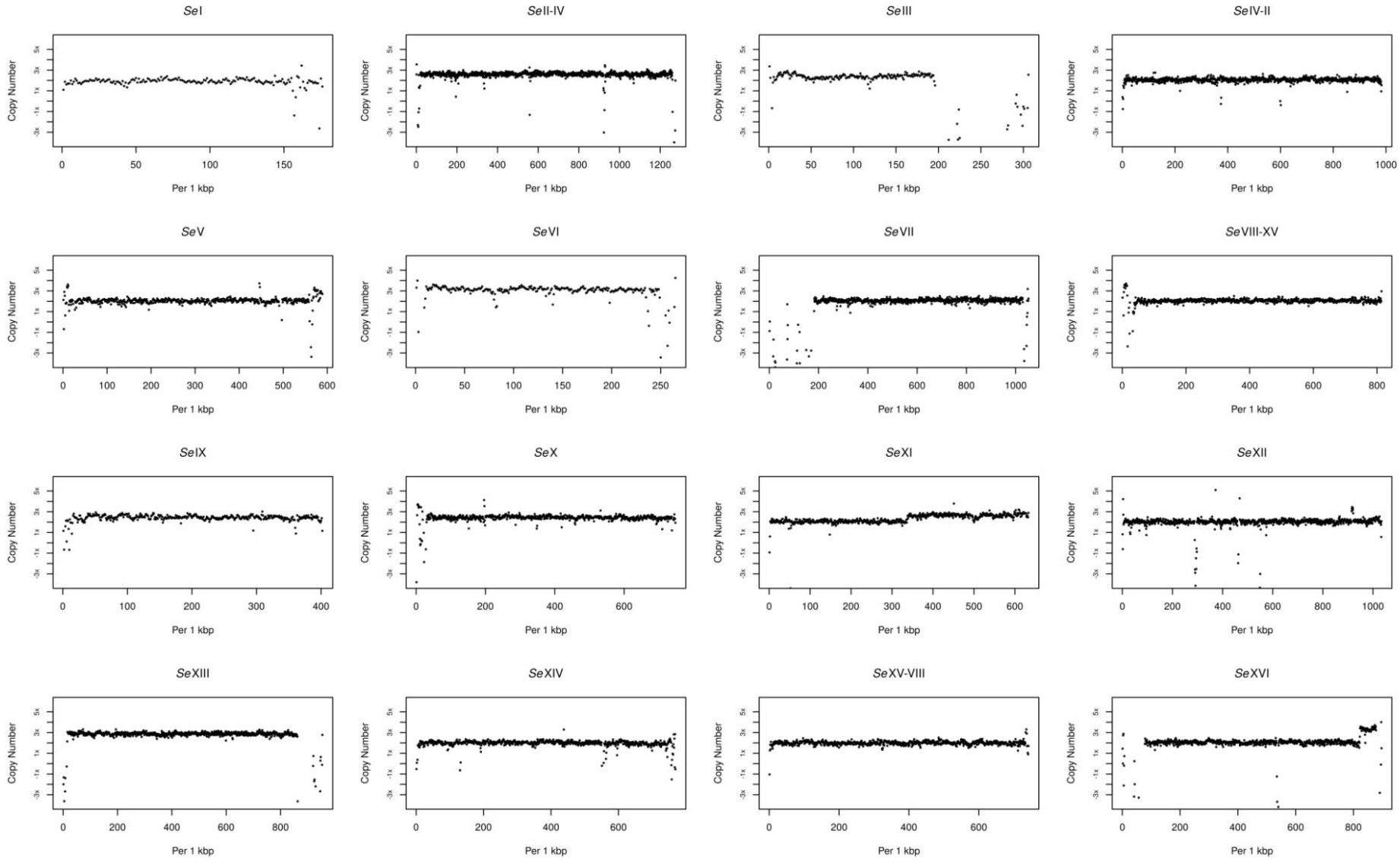



**Fig. S10. *S. pastorianus* 790 chromosome copy number of the *S. eubayanus* sub-genome.** The estimated copy number of the 1000 bp windows of each of the chromosomes of the *S. eubayanus* sub-genome was plotted.