Supplementary Material

Interplay of chimeric mating-type loci impairs fertility rescue and accounts for intra-strain variability in *Zygosaccharomyces rouxii* interspecies hybrid ATCC42981

Melissa Bizzarri1, Stefano Cassanelli1\*, Laura Bartolini1, Leszek P. Pryszcz2, Michala Dušková3, Hana Sychrová3, Lisa Solieri1\*

**\* Correspondence:** Lisa Solieri: lisa.solieri@unimore.it

# Supplementary Tables

**Table S1. Primer sets used in the present study.** Primers named as old are from Watanabe et al. (2013), whereas primers named as new derive from Watanabe et al. (2017). DNA sequences complementary to the *kanMX* gene sequence and plasmid sequences are written in lower-case.

|  |  |  |
| --- | --- | --- |
| **Primer name** | **Sequence (5’ to 3’)** | **Corresponding Figure** |
| DIC1\_P\_F1 | GGTATTTGATGGGAGCAGCA | Figure 1A |
| 3\_old | TACTTACTGGATGAATCTTCTGTGA | Figure 1B |
| C\_old | TCAGTACCAGAAGTGGTCTTTGAAA | Figures 1A, 1B |
| MATα2cp2-893up-F | CGGTAACGACTGTATAGCTAAG | Figures 1C, S1 |
| SLA2\_P\_R3 | GGACAGTTGGGAGACACTGAA | Figure 1C |
| 3\_old | TACTTACTGGATGAATCTTCTGTGA | Figure 1D |
| SLA2\_P\_R2 | GGGAGACACTGAAGCGTTAGAT | Figure 1D |
| CHA1L\_P\_F1 | AGACAGCTACAAGGTGTTGTGA | Figure 1E |
| SLA2Tr\_R1 | AAAGTCCTATTCACGTGACGAA | Figure 1E |
| DIC1\_P\_F2 | GAGTGGTATGGTGAAGCTGTG | Figure 1F |
| SLA2\_uni\_R1 | ATATATCTTATCGAGACAGTGTATTTC | Figure 1F |
| 1\_old | GCTACTCCCTCATTAGAACATGAAA | Figure 1G |
| B\_old | TCTATTTCGTCCGTTTATCGTTGGT | Figure 1G |
| 6\_new | TGTATTGACCAGCTTCGTTTGA | Figure 1H |
| F\_new | ATGGACTACACGTACCACAA | Figure 1H |
| 301\_MATa1F7 | TGGATCTTAGACAGTGGTGTAAGG | Figure 3A |
| 301MATα1α2cp2R2 | TGTCATCCACATTGAAATCATCTC | Figure 3A |
| 301MATa1a2-cp2F1 | GTTTAGATGCCAGTGCTCTTCA | Figure 3B |
| 301MATa1a2-cp2R1 | GACACATTGCATTCTGTTAAACGT | Figure 3B |
| 301\_MATA1F3 | GTAGCTTCCACAAGGTCTTCAAG | Figure 3C |
| 301\_MATA1R4 | GCTGCTACAGCTTCCCTTTC | Figure 3C |
| 301\_MATA2F8 | AGCCAAGTGGGCGATTTA | Figure 3D |
| 301\_MATA2R2 | CATGTGTCTGCAATCACTTCAC | Figure 3D |
| MATA1\_008\_F1 | ATTCTCCAAATGATCTTCAGA | Figure 3E |
| MATA1\_008\_R1 | ATACCCATATTCTTACTTGAAGT | Figure 3E |
| MATα1/2cp2-kanMX-F-80nt | CATGTTTGAACGAGTGTTTTGTTCATTGGTTTGGAATAAACAGGTCTTCGACGTTTAGCCATGTCGAGGATTTAAACGTTTGACAttcgtacgctgcaggtcgac | Figure S1 |
| MATα1/2cp2-kanMX-R-80nt | CAACCGGTAAGTGTTCTTTCAATAAGTCAGTTGTGCAATGAAGTGGCAAGTCAGTTTTTAAGCAACACACCGCACGTACCGgcataggccactagtggatctg | Figure S1 |
| kanMX-R1 | CTCTGGCGCATCGGGC | Figure S1 |
| kanMX-F1 | CATTTGATGCTCGATGA | Figure S1 |
| MATα1cp2-374down-R | CCAAACTTTATGGATATGAGTTCTAGC | Figure S1 |
| AGA2\_backbone1\_rouxii\_F | CATGTACCACTGTACCCAGTAAG | Data not shown |
| AGA2\_backbone1\_rouxii\_R | ACCGTAGTAGTCCCGATTGA |
| STE2\_backbone15\_rouxii\_F | CCTATTGGCCTCGTCTGTTAAT |
| STE2\_backbone15\_rouxii\_R | TAGGCGGACAAGATATGAGGT |
| STE2\_backbone5\_no\_rouxii\_F | CCTATTGGCCTCGTCTGTTAAT |
| STE2\_backbone5\_no\_rouxii\_R | TAGGCGGACAAGATATGAGGT |
| STE6\_backbone1\_rouxii\_F | TAACACTACCAGTGGGTAA |
| STE6\_backbone1\_rouxii\_R | TCATAAGTGGACGTTTTGAAA |
| STE6\_backbone6\_no\_rouxii\_F | ATATCAAAATCGATGGCATGGA |
| STE6\_backbone6\_no\_rouxii\_R | AGCGGTTATTTTGTTGCCT |

**Table S2**. **Inventory of *MTL* cassettes in ATCC42981\_R draft genome derived from DBG2OLC and MaSuRCA *de novo* assemblies.** DBG2OLC scaffolds (accession numbers: UEMZ01000001.1-UEMZ01000033.1) were derived from ATCC42981\_R BioProject PRJEB26771. Grey shadow indicates *MTL* cassettes found in both the assemblies. JCM66020 *MTL* cassettes were described according to the nomenclature reported by Watanabe et al. (2017). Briefly, numbers from 1 to 6 indicate 5’ *MTL*-flanking genes *DIC1*T, *CHA1*LT, *CHA1*T, *DIC1*P, *CHA1*LP, and *CHA1*P, respectively. Capital letters A to F indicate 3’ *MTL*-flanking genes *SLA2*T, ZYRO0F18634gT, ZYRO0C18392gT, *SLA2*P, ZYRO0F18634gP and ZYRO0C18392gP, respectively. Abbreviation: r.c., reverse complement.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cassettes** | **Bizzarri *et al*. (2016)** | **JCM22060** | **PCR** | ***In silico* analysis** | **NBRC110957 (Accession number)** |
| **YαT** |  |  |  |  |  |
| *DIC1*P-*MTL*αT-ZYRO0F18634gT | - | 4B | + | DBG2OLC, MaSuRCA | *CHA1*P-*MTL*αP-ZYRO0C18392gP (BDGX01000045) |
| *CHA1*LT-*MTL*αT-ZYRO0F18634gT | + | - | + | DBG2OLC | *CHA1*LT-*MTL*αT-ZYRO0F18634gT (BDGX01000025) |
| **YαP** |  |  |  |  |  |
| *DIC1*T-*MTL*αP-*SLA2*P | + | 1D | + | DBG2OLC,MaSuRCA | *DIC1*T-*MTL***a**P-*SLA2*P (BDGX01000009) |
| *CHA1*LT-*MTL*αP-*SLA2*P | + | 2D | + | DBG2OLC,MaSuRCA | - |
| *CHA1*LP-*MTL*αP-ZYRO0F18634gP | - | 5E | + | DBG2OLC,MaSuRCA | *CHA1*LP-*MTL*αP-ZYRO0F18634gP (BDGX01000013) |
| *DIC1*T-*MTL*αP-*SLA2***N** | - | - | - | DG2OLC | *-* |
| **Ya** |  |  |  |  |  |
| *DIC1***N**-*MTL***aN**-*SLA2*T | + (partially) | - | + | DBG2OLC | *DIC1*P-*MTL***a**T-ZYRO0C18392gT (BDGX01000035) |
| *CHA1*T-*MTL***a**T-ZYRO0C18392gT | + | 3C | + | DBG2OLC,MaSuRCA | *CHA1*T-*MAT***a**P-*SLA2*P (BDGX01000021) |
| *CHA1*P-*MTL***a**P-ZYRO0C18392gP | - | 6F | + | MaSuRCA | - |

# Supplementary Figures

****

**Figure S1. Validation of *MAT*αPdeletion in ATCC42981\_R.** Panel (**A**) outlines the diagnostic PCR strategy used to verify the integration of *loxP*-*kanMX*-*loxP* disruption cassette in the *DIC1*T-*MAT*αP-*SLA*2P locus of ATCC42981\_R genome. Full-length, 5’ and 3’-PCRs are shown. Flanking genes from T and P-subgenomes are marked with T and P superscripts, respectively. *MAT*α1 and *MAT*α2 genes from P-subgenome are indicated as αP1 and αP2. In panel (**B**) the *loxP*-*kanMX*-*loxP* module was integrated at the *DIC1*T-*MTL*αP-*SLA2*P cassette in four ATCC42981\_R clones, as demonstrated by size difference in full-length PCR products between *kanMX* (2,164 bp) and *MAT*αP (2,889 bp) cassettes. The correct orientation of *loxP*-*kanMX*-*loxP* module was confirmed by the5’ and 3’ PCRs. Numbers from 1 to 5 indicate *MAT*αPΔ clone\_6, \_65, \_74, and \_177, and ATCC42981\_R, respectively (Table 1), while M represents molecular weight marker GeneRuler 100 bp or 1 Kb Plus DNA Ladders (Thermo Scientific, Waltham, MA). Abbreviations: wt, wild type; HR, homologous recombination; *kanMX*, kanamycin resistance gene; NTC, no template control.



**Figure S2.** **Multiple sequence alignment of *SLA2* genes from DBG2OLC, MaSuRCA assemblies and Sanger sequences.** Alignment was carried out using Clustal Omega. Formatting of aligned sequences was done in Jalview alignment viewer. Residues in the alignment follow the default Clustal colour scheme of Jalview.



**Figure S3. Sequence alignment highlighting the 27 bp indel in X regions downstream the *DIC1N* and *DIC1T* gene variants.** X region from CBS732T was used as reference. Sequences were retrieved from Sanger sequencing and aligned using Clustal Omega. Formatting of aligned sequences was done in Jalview alignment viewer. Residues in the alignment follow the default Clustal colour scheme of Jalview.



**Figure S4.** **Cartoon illustration of main differences between MaSuRCA and DBG2OLC assembliesin *MTL*α loci–containing scaffolds**. In panel (**A**) black horizontal bar represents size scale with tick marks for every 100 Kbp. MaSuRCA and DBG2OLC scaffolds are represented as green and orange rectangles, respectively. Synteny representation around *MTL*α loci omits X and Z regions for brevity. T and P variants are depicted as filled and dot arrows and marked with T and P superscripts, respectively. Panel (**B**) details collinear sets of homologous regions between the jcf180000000243 (r.c) MaSuRCA scaffold and either UEMZ01000013.1 or UEMZ01000003.1 (r.c) DBG2OLC scaffolds. Gene size and distance are not in scale. The yellow rectangle represents the starting point of scaffold synteny. Abbreviation: r.c., reverse complement.

****

**Figure S5. Cartoon detailing gene order and synteny blocks around *MTL* loci in ATCC42981\_R**. Rectangles represent regions in ATCC42981\_R scaffolds that are collinear with a region of the reference strain NBRC1130T ancestral genome. Blue and light blue colours correspond to regions that in CBS732T genome are on chromosomes C and F, respectively. Solid rectangles and rectangles with diagonals represent T- and P-sequences, respectively. Scaffold (sc) numbers referred to the DBG2OLC genome assembly deposited in European Nucleotide Archive under accession number PRJEB26771 (Bizzarri et al., 2018); for simplicity the last number of ENA code marked each scaffold (*i.e*. UEMZ01000028.1 in short sc28). Genes from T and P-subgenomes are marked with T and P superscripts, respectively, while *DIC1* and *MAT***a**2 new variants with N superscript. Scaffolds are not in scale.