JMBFS

Journal of Microbiology, Biotechnology and Food Sciences

International peer-reviewed scientific online journal

Published by Faculty of Biotechnology and Food Sciences

REVISITING MOLECULAR CLONING TO SOLVE GENOME SEQUENCING PROJECT CONFLICTS

Hugo A. Barrera-Saldaña; Aarón Daniel Ramírez-Sánchez; Tiffany Editth Palacios-Tovar; Dionicio Aguirre-Treviño; Saúl Felipe Karr-de-León*

Address(es): Dr. Hugo A. Barrera-Saldaña,
Vitagénesis, SA de CV, Edificio Vitaxentrum, Laboratorio de Bioprocesos, Blvd. Puerta del Sol #1005, 64630, Monterrey, NL. Ph: +52 (818) 123-8249.

*Corresponding author: habarrera@gmail.com

doi: 10.15414/jmbfs.2017.6.5.1157-1160

ARTICLE INFO

Received 3. 1. 2017
Revised 25. 1. 2017
Accepted 26. 1. 2017
Published 3. 4. 2017

Short communication

OPEN ACCESS

ABSTRACT

In our laboratory, DNA sequencing by Sanger method is used as the "gold standard" for clinical diagnostics, microbe identification (bacteria and yeast, mainly) and genome characterization. In this research, we used it to characterize a conflicting locus in a *Saccharomyces cerevisiae* sequencing project. When sequenced, the resulting electropherogram of the analyzed locus showed a pattern indicating either sample contamination or allele variation. Molecular cloning was chosen as the most straight-forward strategy to solve the dilemma. The initial characterization of recombinant plasmids by restriction enzyme digestion confirmed the presence of two genomic sequences. Their Sanger sequencing revealed two alleles distinguishable by a total of 29 nucleotide differences (25 of which were SNPs). NCBI BLAST revealed that the conflicting locus covered an intergenic region and a coding sequence for a putative permease protein. The present study shows the utility of the classical molecular cloning technique to solve problems of modern genome projects.

Keywords: Allele segregation, molecular cloning, Sanger sequencing, troubleshooting

INTRODUCTION

Molecular cloning is a methodology conformed by a set of experiments that allow the asexual exponential copying of an isolated genomic region for posterior analyses and genetic manipulation **(Sambrook and Russell, 2001)**. The principle of this methodology relies on the use of plasmid vectors, restriction enzymes, ligases, calcium-induced bacterial transformation, and clonal propagation of recombined microorganisms (*Escherichia coli* and *Saccharomyces cerevisiae*, among others). This clonal propagation provides an additional value of this tool since allows segregating and thus the identification of coexisting highly similar gene sequences **(Barrera, Seeburg and Saunders, 1983)**. Molecular cloning was eventually substituted by the development of the Polymerase Chain Reaction (PCR) in which there is no need for a host cell, and exponential multiplication of DNA material can be done *in vitro* and in a much shorter time **(K. B. Mullis, *et al*. 1986)**.
Molecular cloning coupled to nucleotide sequencing has been commonly used to read genomes represented in clones of genomic libraries. In our case, the main objective was to achieve the resolution of a possible mix of genomes represented in a PCR product labeled Sc790Gap1, derived from the *S. cerevisiae* strain 790 genome, using molecular cloning to solve the putative overlapping sequences of at least two genetic variants.

MATERIAL AND METHODS

Conflicting amplicon

Gap1 is one of various genomic regions present in *S. cerevisiae* strain 790, amplified with *Gap1F* and *Gap1R* primers, troubling the genome assembly of this strain. Its sequences were obtained using the Applied Biosystems® (AB) Genetic Analyzer 3130 (AB, Foster city, CA, USA) while its analyses were attempted with various bioinformatics tools: Clone Manager 6 (Scientific Educational Software; NC, USA), SnapGene Viewer 2.6.2 (GSL Biotech LLC, IL, USA), BioEdit 7.2.5 (Biosciences, CA, USA) and MEGA 6.06 **(Tamura, *et al*., 2013)**.

DNA preparation

The yeast strain was harvested from an axenic culture used for beer production, and its total DNA was isolated as previously described by **Harju, *et al.*, (2004)**.

PCR amplification

PCR reaction was prepared using *Gap* primers (*Gap1F*: 5'-TTTACCATGAGCGCAACAGC-3' and *Gap1R:* 5'-AAAAAGCAGAACGACGCACC-3') at 0.6 µM, 100 ng of DNA sample, 11 µl of Master Mix (2X) from Promega (Madison, WI, USA), and nuclease free water (NF H_2O) to complete 15 µl. All amplification reactions were hot started at 95°C for 5 min, followed by 30 cycles of denaturation at 95°C for 60s, annealing at 58°C for 60s, and extension at 72°C for 120s. After a final extension step at 72°C for 3 min, the amplification product was stored at 4°C.

Sequencing of amplicons

The PCR products were treated with Shrimp Alkaline Phosphatase (SAP) and Exonuclease I (EXO I) **(**Affymetrix, Santa Clara, CA, USA**)** enzymes during 15 minutes at 37°C, followed by 15 minutes at 80°C. 1 µl of treated product was added to a PCR sequencing master mix with 2 µl of Big Dye **(**Applied Biosystems, Foster city, CA, USA**)**, 0.5 pmol of primer and pure sterile water up to 6 µl. The resulting PCR product was purified again with the SAM/X-terminator Kit **(**Applied Biosystems**)**. The supernatant was loaded into the AB Genetic Analyzer 3130 for sequencing. Finally, raw data was processed using the Sequencing Analysis Software v5.3.1 **(**Applied Biosystems**)** to obtain the electropherograms.

Molecular cloning

The PCR product was cloned into a TOPO XL plasmid (Invitrogen, Carlsbad, CA) using the T/A cloning approach **(Sambrook and Russell, 2001)**. Bacteria were transformed using heat shock and a rapid screening of the clones was performed through colony PCR using *Gap* primers **(Sambrook and Russell, 2001)**. Lastly, a restriction analysis of the PCR products was realized using *TaqI* enzyme **(**New England Biolabs, Ipswich, MA**)** during one hour at 65°C in order to select the clones with a differential pattern.

Plasmid screening and characterization

Selected colonies were picked up with sterile toothpicks, inoculated in 3 mL of LB (Luria-Bertani) broth and incubated overnight at 37°C. Biomass was collected by microfugation and plasmids were isolated essentially as described by **Birboim and Doly, (1979)**. Insert presence and orientation were established by using restriction mapping and positive clones were sequenced using Sanger sequencing **(Smith, *et al*., 1986)**.

Sequence analyses

Sequences were aligned against consensus sequences found in GenBank using Bioinformatic software MEGA **(Tamura, *et al*., 2013)**.

RESULTS AND DISCUSSION

Sanger sequencing

The product obtained from the PCR reaction was bidirectionally read using *Gap1F* and *Gap1R* primers (according to the PCR assay, an expected product of approximately 1750 bp was seen). A clear reading of the first ~200 bp was obtained from the *Gap1F* primer; whereas approximately 400 bp were obtained using *Gap1R* primer. Beyond these lengths, electropherogram anomalies complicated its interpretation. Apparently, the artifact detected was due to two overlapping readings, and troubleshooting advice suggested that this could be due either to contamination or heteroploidy. Nevertheless, seeking to discard possible errors or artifacts, we designed new nested primers (*IntergapF* and *IntergapR*) located closer to the troubling region (Figure 1). Again, the same undecipherable readings were obtained (data not shown). Therefore, we decided to test the heterozygosity hypothesis and for this, molecular cloning was chosen as the most suitable method to segregate alleles in order to characterize them separately.

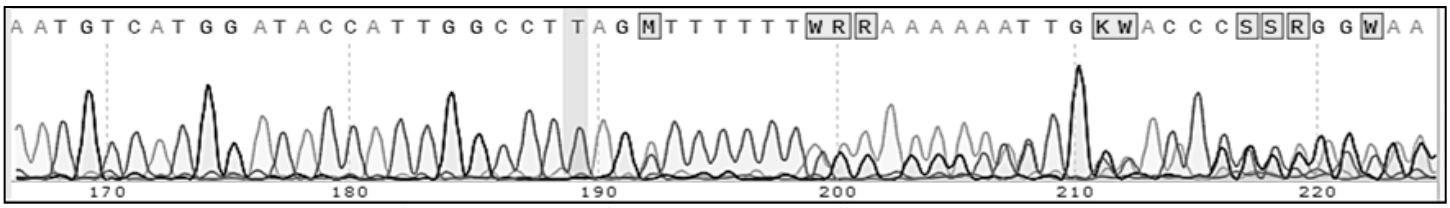

Figure 1 Electropherograms showing region of reading overlaps. The figure shows two clearly different overlapping readings following the T homopolymer. The upper sequence belongs to the forward end reading using *Gap1F* primer (5'>3'), while the lower one corresponds to the reverse end obtained with *Gap1R* primer (3'>5').

Molecular cloning and resolving sequencing

The original PCR product was inserted into a plasmid vector and then used for transforming *E. coli* cells to generate clones and continue with their identification, independent propagation and characterization of the cloned insert. The presence of the alleles in the transformed candidate clones was confirmed by restriction analysis using *Taq*I restriction enzyme of products obtained by colony PCR. Two different patterns predicted *in silico* were evidenced in the agarose gel image (Figure 2) and their insert orientation was established by double digestion with *Bgl*II and *Kpn*I enzymes (Figure 3).

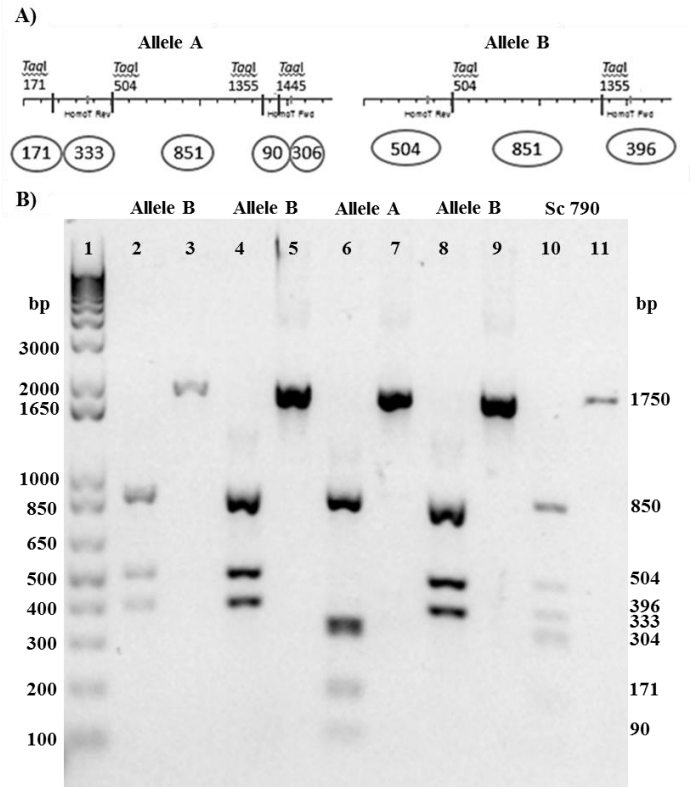

Figure 2 Characterization of clones by PCR and restriction analyses. (A) *In silico* *Taq*I restriction maps of possible PCR products obtained from recombinant plasmids using *Gap1F* and *Gap1R* primers. (B) Agarose gel electrophoresis of actual digestions with *Taq*I of the PCR products. Lane 1 shows the DNA base pair (bp) ladder. Lanes 2, 4 and 8 correspond to the size pattern identified as "Allele B". Lane 6 shows the pattern characterizing "Allele A". Lane 10 shows the restriction of the PCR product obtained directly from *S. cerevisiae* 790 (GCF_000146045.2), and lanes 3, 5, 7, 9, and 11 correspond to PCR products from the recombinant clones analyzed.

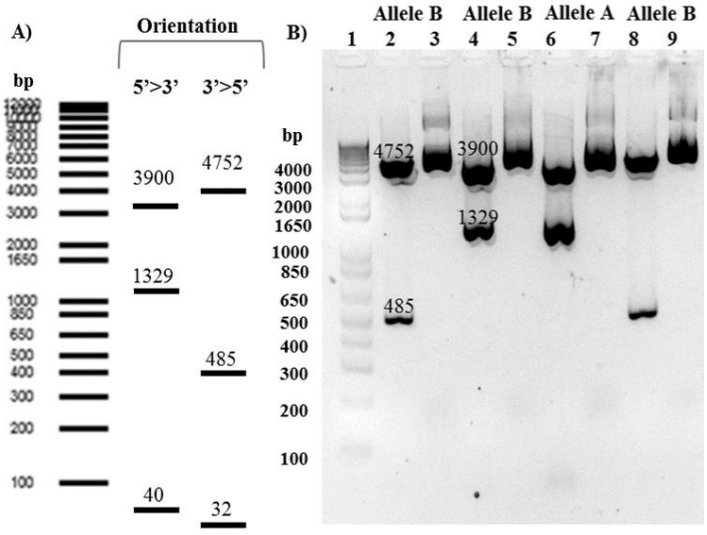

Figure 3 Strategy to determine the insert orientation in the plasmid. (A) *In silico* prediction of the two possible digestion patterns using *Bgl*II y *Kpn*I enzymes to determine the orientation of the cloned amplicon inserted into the plasmid. (B) Agarose gel electrophoresis showing the experimental results from the digestion assay. Lane 1 shows the DNA bp ladder. Lanes 2 and 8 show the pattern for insertion in the 3'>5'orientation, whereas lanes 4 and 6 present the opposite orientation pattern. Lanes 3, 5, 7, and 9 correspond to undigested plasmid DNA.

The cloning segregation (Fig. 3B, clones in lanes 4 and 6) permitted the sequencing of the two different alleles (resulting in the reading of sequences of approximately 850 nucleotide bases of good quality from each end), detecting four indel positions throughout T homopolymers, which were probably the primary cause of electropherograms overlapping. In addition, 25 single nucleotide polymorphisms (SNPs) were identified (Figure 4).

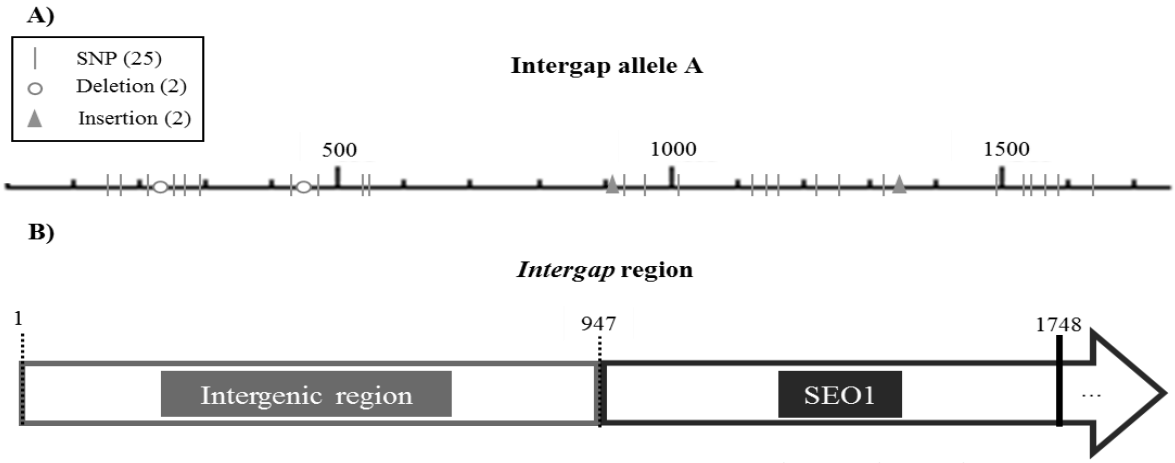

Figure 4 Genetic elements contained in the *Gap1* region. (A) Schematic representation of sequence features of allele A with respect to allele B sequences. *Vertical bars* indicate SNPs (25), *circles* indicate deletions and *triangles* represent insertions. (B) The alleles sequences cover an intergenic region (1-946 bp) and the starting coding region (947-1748 bp) of an open reading frame (a permease protein called SEO1) for a total length of 1748 bp.

Having solved the *Gap* region sequencing problem, its correct genome assembly and annotation was achieved. It turned out that the locus codes for a putative permease (SEO1, NM_001178208) and an intergenic region (Figure 5). The genome variants show a 98.51% similarity, and their origin was traced back to strains EC1118 **(Novo, *et al.*, 2009)** and YJM1463 **(Strope, *et al.*,2015)** of *Saccharomyces* using the GenBank database **(Benson, *et al.*, 2005)** (Table 1).

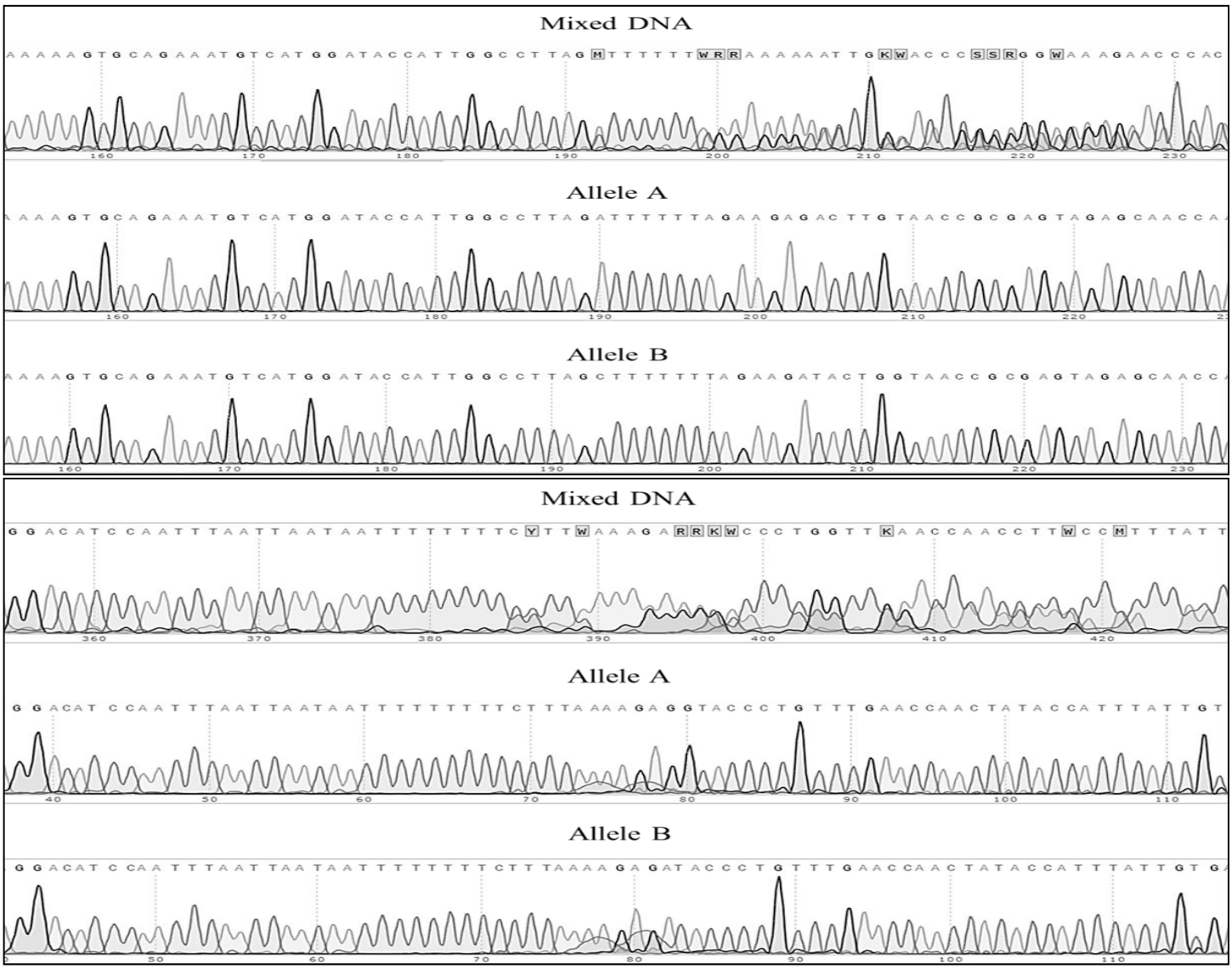

Figure 5 Electropherograms showing the GAP region resolved. The figure shows clearly the resolution of the two different sequences. (A) Forward sequences 5'>3': There is an indel in the T homopolymer, thus the mixed DNA had the two lectures overlapped. (B) Reverse sequences 3'>5': Results were similar to the Forward sequences, showing an indel in the homopolymer.

Table 1 Comparison of analyzed alleles with those of *S. cerevisiae* strains.

Strain	Allele A (%)	Allele B (%)
S. cerevisiae S288c	98.51	97.76
S. cerevisiae EC1118	99.14	99.37
S. cerevisiae YJM248	99.08	99.43
S. cerevisiae YJM270	98.97	99.43
S. cerevisiae YJM1078	99.08	99.43
S. cerevisiae YJM1252	99.08	99.43
S. cerevisiae YJM1450	99.08	99.31
S. cerevisiae YJM1463	98.97	99.54

CONCLUSION

In conclusion, our research shows that using a classical method, such as molecular cloning, problems of complex genomic sequencing projects can be solved in a simple way.

Acknowledgments: This research was supported by Cervecería Cuauhtémoc Moctezuma, S.A. de C.V. and CONACyT "Pharma research support" National Laboratory (#124140).

REFERENCES

Sambrook, Joseph and David W. Russell. (2001). Molecular Cloning: A Laboratory Manual 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Barrera-Saldaña, H.A., Seeburg, P.H. and G.F. Saunders. (1983). Two structurally different genes produce the same secreted human placental lactogen hormone. *J Biol Chem.* 258(6):3787-93.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology* 51: 263-273. https://doi.org/10.1101/SQB.1986.051.01.032

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and S. Kumar. (2013). Mega6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol*. 30: 2725-2729. https://doi.org/10.1093/molbev/mst197

Harju, S., Fedosyuk, H., Peterson, K. R. (2004). Rapid isolation of yeast genomic DNA: Bust n' Grab. *BMC Biotechnol.* 4(8), 1-6. https://doi.org/10.1186/1472-6750-4-8

Bimboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic acids research*, 7(6), 1513-1523.

Smith, L. M., Sanders J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., Hood, L. E. (1986). Fluorescent detection in automated DNA sequence analysis. *Nature.* 321(6071), 674-9. https://doi.org/10.1038/321674a0

Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J.-L., Wincker, P., Casaregola, S., Dequin, S. (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc Natl Acad sci U S A.* 106(38), 16333-16338. https://doi.org/10.1073/pnas.0904673106

Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome res.* 25(5), 762-74. https://doi.org/10.1101/gr.185538.114

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L. (2005). GenBank. *Nucleic acids research*, 33(suppl 1), D34-D38. https://doi.org/10.1093/nar/gki063