

SGRP Update

16 October 2007

Dear All,

The first stage of the Saccharomyces Genome Resequencing Project (<http://www.sanger.ac.uk/Teams/Team71/durbin/sgrp/>) is nearly finished and we plan on submitting an overview manuscript in January. We hope that others will respect our priority in publishing the primary global analyses as outlined below, but also that people will use our data for additional analyses either to be published at the same time or after. Thank you all for your patience.

Publication plans:

Timeline – draft in mid November, additions/corrections in mid December, submission in early January. We are planning on submitting to Nature.

Analyses included –

- Sequence assemblies and comparisons
 - S. paradoxus reference assembly
 - Alignment of S. cerevisiae and S. paradoxus references
 - SNPs and INDELS
 - Copy number variation
 - Potentially functional variation – pseudogenes etc.
- Sequences not in Reference (yeast, fungal, other)
- Population structures of S. paradoxus and S. cerevisiae
 - S. paradoxus has strong population structure
 - S. cerevisiae may be combination of ‘clean’ lineages and mosaic mixtures of these
- Global search for evidence of selection
 - Distribution of selection measures on genes
 - Selection on branch sites and CAI for strongly expressed genes
- Mosaic genomes and introgressions
- Lab strain comparisons
 - (W303, SK1, Y55 all >3X by ABI and 8-12X by Solexa)
 - S288c (original gal2 etc. from stock) compared to SGD reference
- Global phenotype analysis and comparison to phylogeny
 - General correlations
 - Specific pathways
- Evidence of human influence and what the influence might be
 - What does domestication mean?
 - Did human selection occur or did human activity select from available variation?
 - Are there feral S. cerevisiae?
 - Are there wild S. cerevisiae?

Future plans

Other analyses known to be done or being done:

- rDNA analysis – copy number variation and mosaicism
- tRNA and other small structural RNAs
- Transposable elements
- Sequence assembly/imputation
- LD structure and recombination maps

Analyses that should be done sometime (not exhaustive) –

- Assemble subtelomeric regions
- Regional variation in divergence/diversity
- ...(please add your ideas)

Invitations:

If anyone has done or plans on doing a global analysis with a tight clean result which you think should be included in the overview paper, please contact us (ed.louis@nottingham.ac.uk and rd@sanger.ac.uk) and we will consider its inclusion with associated authorship on the paper. The analysis would have to be complete by 14 December and you would have to be willing to have the details transparently displayed on the web pages associated with the project.

In addition to or alternatively, you may have an analysis you want to independently submit, either back to back or in another journal. We have already been approached by some journals concerning related papers and we would be happy to try to facilitate the submission of your work to one of these journals to coincide with the overview paper.

Yours sincerely,

Ed Louis, Institute of Genetics, University of Nottingham

Richard Durbin, The Wellcome Trust Sanger Institute