Cataloguing structural variations at the population level in Saccharomyces cerevisiae by generating a Reference genome Assembly Panel.

Samuel O'Donnell and Gilles Fischer
LCQB (IBPS), Paris

The advent of long-read sequencing has advanced the field of genomics by enhancing the ability to generate complete genome assemblies. This is currently allowing researchers to move the field of genomics from that of a single reference genome representing a species, to a Reference genome Assembly Panel (RAP) containing multiple reference quality genomes. The many benefits of an RAP include; containing a more pangenomic view of a species; reducing reference bias during reference guided analysis and greatly improving the ability to detect structural variations (SVs) compared to short read data. Additionally, long reads have enabled the phasing and assembling of haplotypes, meaning these RAPs may contain, for example, diploid genome assemblies.

The Phenovar project is a collaborative effort, that has focussed on creating the first rigorous RAP using the species *Saccharomyces cerevisiae* (ScRAP) which currently contains 140 genomes, 98 newly sequenced, and 18 newly assembled. Mentionable is that the ScRAP contains 41 genomes with 1 contig per chromosome, each assembled telomere to telomere, and therefore representing truly complete genomes. Furthermore, the ScRAP contains the first of both complete diploid genome assemblies and polyploid phased assemblies.

Although ~2500 *S. cerevisiae* strains have been sequenced by short-read technology, this still provides a poor view of SVs due to complications in repetitive regions and reference bias during mapping-based approaches. Using the ScRAP we have detected over 5500 non-redundant SVs of which 76% are novel compared to short read assemblies. This gain has been further improved in phased assemblies, which in diploids alone can increase the number of detected SVs by 60% as compared to haplotype collapsed assemblies. Lastly, the ScRAPs' and Long-read sequencing's resolution of genome structure has allowed us to detect a relationship between large SVs and aneuploidy.