4-3-2022

Baker's yeast breeding and engineering to improve biologics manufacturing

Chris Finnis
Phenotypeca Ltd, United Kingdom, chris@phenotypeca.com

Les Evans
Phenotypeca Ltd, United Kingdom

Ed Louis
Phenotypeca Ltd, United Kingdom

Anne Henstra
Phenotypeca Ltd, United Kingdom

Baker's Yeast Breeding and Engineering to Improve Biologics Manufacturing

Chris Finnis, Les Evans, Anne Henstra, David Harvey, Charlotte Smith, Alex Pritchard, Ed Louis

Contact: chris@phenotypeca.com Website: www.phenotypeca.com

Abstract

The regulatory-friendly baker's yeast ***Saccharomyces cerevisiae*** has a bright future for cost reduction and improved access to medicines, especially for lower and middle-income countries. It has been used safely for biopharmaceutical manufacturing since 1987, although this work only used a limited number of closely related strains. These were improved by chemical mutagenesis and rational genome engineering for first-generation platforms, producing biopharmaceuticals currently worth over $40 billion/year in market value.

Phenotypeca has developed a next-generation biologics production platform using genetically diverse baker's yeast from different geographical and ecological niches[1]. Four parent strains with around 100,000 segregating functional SNPs were engineered for protease reduction, chaperone overexpression and other modifications known to improve product yield and quality during cGMP manufacturing of recombinant products. Advanced multi-generational breeding[2] produced libraries of approximately a billion genetically distinct progeny for the intracellular production or secretion of a range of recombinant proteins. These populations contain industrially stable, exchangeable, high copy number expression plasmids and consist of only yeast DNA plus a codon-optimised synthetic gene of interest. Results for mCherry and mCherry-tagged proteins, including SARS-CoV-2 antigens, demonstrated that progeny populations contained phenotypes better than the best parent and worse than the worst parent. High throughput screening and selection workflows using flow-cytometry and robotics systems have been developed to select individuals with improved properties for robust manufacturing processes.

Introduction

Production Strains: Since the original *S. cerevisiae* S288c genome sequence was published in 1996, it has been has revealed that strains used to develop first generation baker's yeast expression platforms were closely related to S288c and W303, with mosaic genomes derived from the Wine/European lineage (**Figure 1**)[1, 3]. Phenotypeca has similarly engineered strain collections of this type (**YPT** strains) with modifications known to improve product yield and quality during cGMP biologics manufacturing.

Additionally, four parent strains, isolated from diverse geographical and ecological niches which contain around 100,000 independently segregating naturally selected genomic SNPs, were chosen as the basis for the generation of novel strains for recombinant protein production. Haploid derivatives of these strains lacking the natural 2-micron plasmid were prepared with auxotrophic markers for breeding and expression plasmid maintenance. Derivatives were also produced with gene disruptions and insertions to increase yield and quality during recombinant protein production, e.g. to control proteolysis, glycosylation, protein folding and mRNA levels.

All production strains are designed for cGMP biologics manufacturing and simple regulatory approval, and are devoid of antibiotic resistance markers, containing only *S. cerevisiae* DNA plus a synthetic gene codon-optimised for the recombinant product of interest.

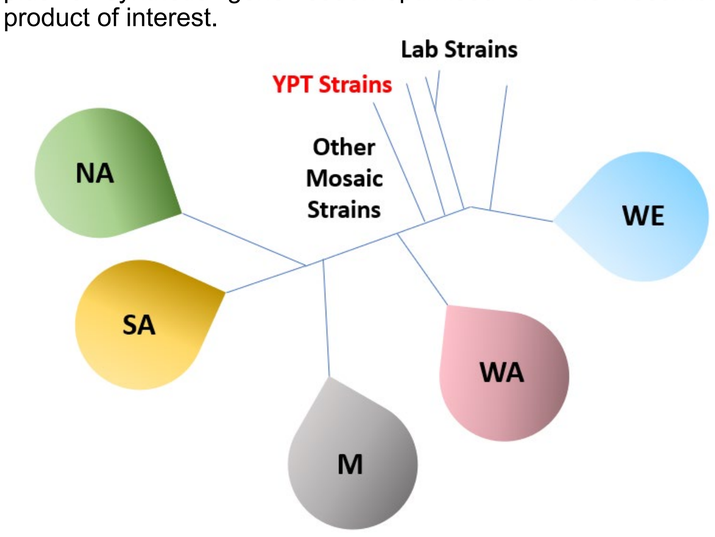

Figure 1: Phylogenetic tree diagram from the *Saccharomyces* Genome Resequencing Project showing clean lineages in relation to traditional laboratory strains[1]. A total of 235,127 SNPs, 14,051 indels and 38 new ORFs were identified among the five clean lineages and mosaics strains. One individual representative from each of the North American (**NA**), Sake (**SA**), West African (**WA**) and Wine/European (**WE**) clean lineages was selected for recombinant protein production and breeding. The Malaysian (**M**) lineage was not used. All strains were compliant with the Nagoya Protocol (UN Convention on Biological Diversity).

References

1. Liti, G. *et al*. Population genomics of domestic and wild yeasts. Nature 458, 337–41 (2009).

2. Cubillos, F. A. *et al*. High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. Genetics 195, 1141–55 (2013).

3. Goffeau, A. *et al*. Life with 6000 genes. Science 274, 546–567 (1996).

4. Lan, J. *et al*. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215-220 (2020).

Expression Plasmids: High-copy number episomal expression plasmids based on the whole 2-micron plasmid were designed for stable maintenance during industrial scale production. Plasmid copy number is enhanced by complementary genome engineering in the parental strains to elevate the transcript levels for the recombinant product. Promoter, leader and terminator sequences are selected from libraries for each recombinant protein of interest. Recombinant protein production is repressed during breeding and activated afterwards for selection of strains with improved phenotypes from the diverse progeny generated. Fluorescent protein tags, e.g. mCherry, were used to facilitate selection during flow cytometry. Plasmids expressing mCherry-tagged proteins can be exchanged for the final plasmids without the mCherry-tag by growth in non-selective media and re-transformation with the desired final expression plasmid. Similarly, strains improved for one recombinant product can easily be tested for new products.

Multi-Generational Breeding

Pair-wise crosses with parents of opposite mating-types were performed for all four parents, with one parent strain transformed to leucine prototrophy with an expression plasmid. Diploids were sporulated and spores pooled for multi-generational crosses to produce populations of approximately a billion progeny (**Figure 2**). Auxotrophic markers *ura3* and/or *lys2*, were used for haploid/diploid selection. Final populations were induced for recombinant protein expression and screened for product yield and quality.

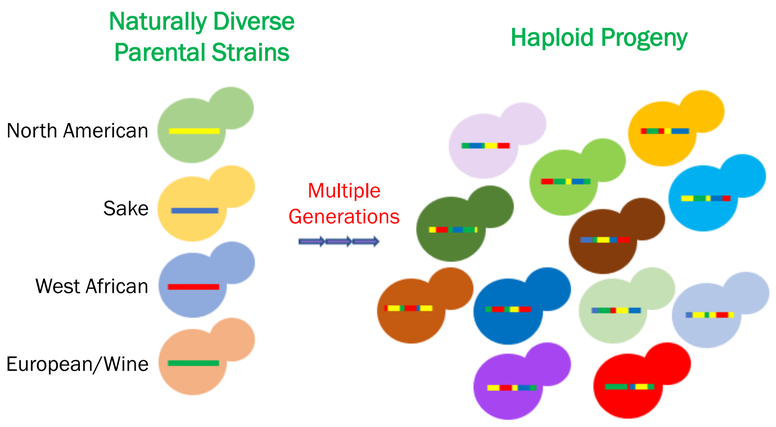

Figure 2: Four diverse parental yeast strains are bred to generate diverse populations for phenotypic screening.

Typical workflows (**Figure 3**) use flow cytometry followed by individual strain phenotyping for feasibility studies (PhenoStart™), strain optimisation (PhenoDev™) and process development (PhenoFull™).

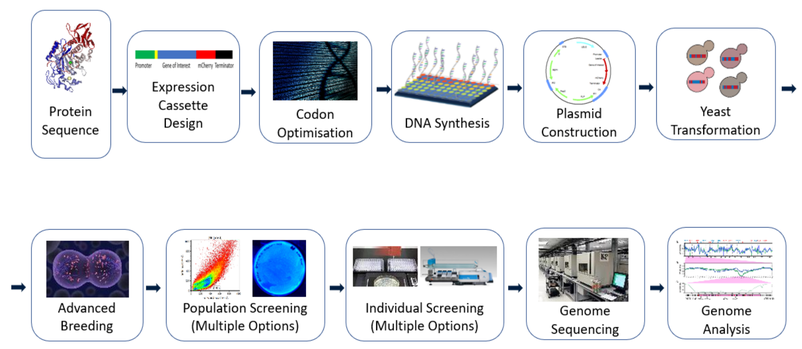

Figure 3: PhenoStart™ feasibility study and PhenoDev™ strain optimisation workflows.

Results

Intracellular Products: The mCherry fluorophore was used to compare phenotypic diversity for recombinant protein production between parental and progeny strains.

Figure 4 (right) shows the phenotypic diversity for intracellular mCherry expression after breeding, with a wide range of visible differences in colony colour. Typically, progeny better than the best parent, and worse than the worst parent, are generated by breeding.

Figure 5 shows a Box and Whisker Plot for mCherry fluorescence from 24 colonies from each clean lineage parent and 192 colonies for 3rd generation progeny. Colonies were randomly picked and grown for 5-days in 0.5mL selective minimal media (30ºC, 200rpm, 48-well microtitre plates with humidity control). mCherry signal was measured with Tecan Infinite M1000 PRO plate reader. Eight progeny (4.2%) had higher, and two (1.0%) had lower mCherry levels than the highest and lowest parent colonies, respectively. The parental average (broken red line) was comparable to the progeny average.

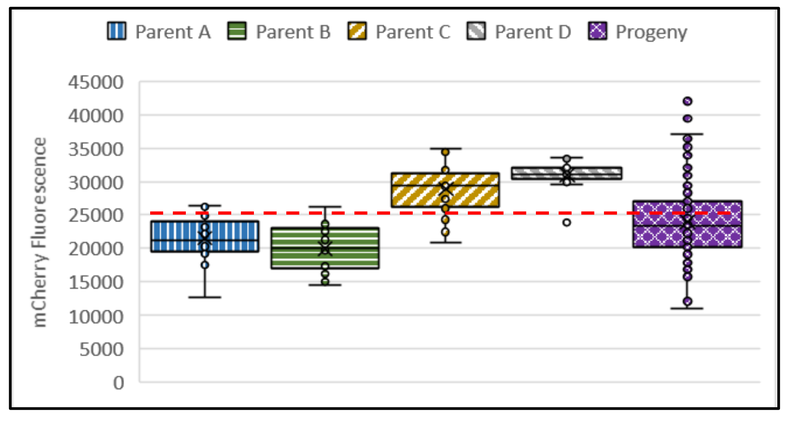

Figure 5: Comparison of *S. cerevisiae* parental and progeny strains for intracellular mCherry production.

Secreted Products: A range of SARS-CoV-2 spike protein fragments were secreted with C-terminal mCherry tags using the invertase leader sequence. **Figure 6** shows flow cytometry results for cell sorting of 4th generation progeny secreting a SARS-CoV-2 Spike protein fragment (RBD 344-604) fusion. Microtitre plate cultures for individuals selected for high mCherry signal by flow cytometry were visibly red compared to individuals with low mCherry signal.

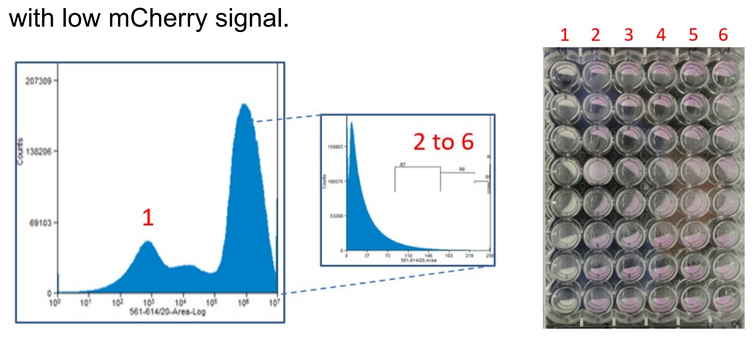

Figure 6: Single cells were sorted for mCherry fluorescence using a Beckman Coulter Astrios EQ cell sorter for expression analysis.

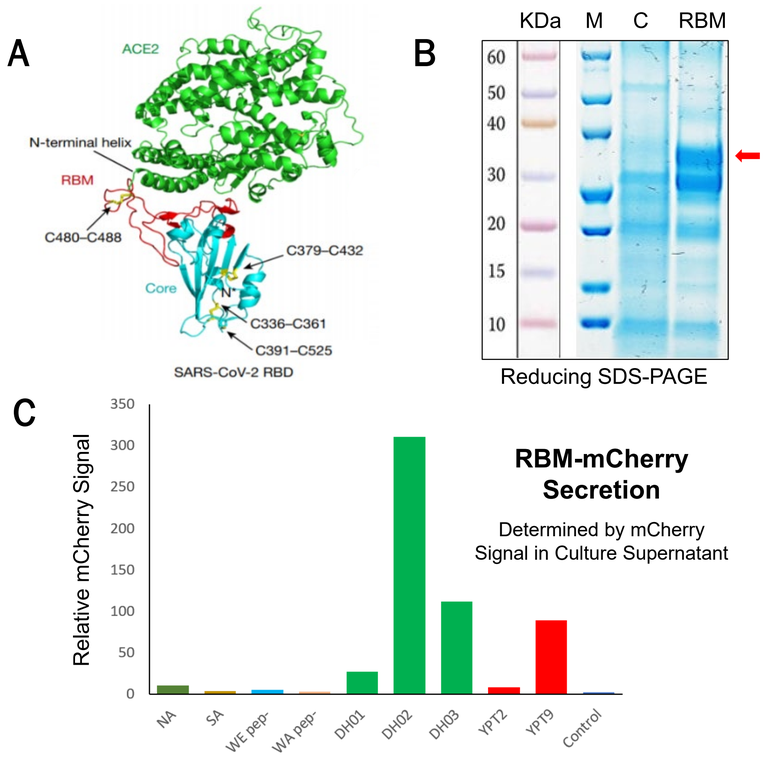

Figure 7 (above) shows expression analysis for the ACE2 receptor binding motif fragment (RBM 438-505)[4] of SARS-CoV-2 Spike protein (**7A**) with a C-terminal mCherry-tag. SDS-PAGE analysis indicated a substantial increase in secreted product yield after breeding (RBM ⬅) compared to the control (C) grown in a BioLector Pro mini-bioreactor system (mp2-labs). A proteolytic degradation product is apparent below the expected size for RBM-mCherry (**7B**). Significant yield improvements after breeding was also evident for mCherry levels from the whole culture and supernatants for progeny strains producing RBM-mCherry (DH01 to DH03). Rational engineering of progenitor laboratory strain YPT2 was performed to disrupt proteinase A (*pep4*) and an aspartic protease (*yps1*), along with engineering to over-express protein disulphide isomerase (*PDI1*) and increase plasmid copy number. This also gave a large increase in product yield (from strain YPT9) (**7C**).

Conclusion

- Yeast breeding generated progeny phenotypes better than the best parent and worse than the worst parent.
- Combining yeast breeding with rational genome engineering offers a powerful strategy for improving recombinant protein production from *S. cerevisiae.*
- Phenotypic improvement of production strains promises significant cost reductions for biologics production.

Acknowledgements: We gratefully acknowledge Innovate UK and the Future Vaccine Manufacturing Research Hub (Vax-Hub) for research funding and thank collaborators Dr Aadil El-Turabi and Professor Dame Sarah Gilbert at the University of Oxford's Jenner Institute.

www.phenotypeca.com | info@phenotypeca.com

Office Address: Phenotypeca Ltd, BioCity Nottingham, Pennyfoot Street, Nottingham, NG1 1GF

Research Facilities: Synthetic Biology Research Centre, Biodiscovery Institute, University of Nottingham, University Boulevard, Nottingham, United Kingdom, NG7 2RD