# **Supplementary information for**

Modular, synthetic chromosomes as new tools for large scale engineering of metabolism

Else-Jasmijn Hassinget al.

1 Department of Biotechnology, Delft University of Technology, van der Maasweg 9, 2627HZ Delft, The Netherlands

The construction of powerful cell factories requires intensive genetic engineering for the addition of new functionalities and the remodeling of native pathways and processes. The present study, using *Saccharomyces cerevisiae*, explores the potential of modular, specialized *de novo*-assembled neochromosomes (named NeoChrs) to facilitate extensive genome reprogramming. Linear and circular NeoChrs carrying 20 native and 21 heterologous genes were designed and constructed by *in vivo* assembly in yeast in two transformation steps leading to *de novo* production of anthocyanins, native to plants. Turned into exclusive expression platforms for heterologous and essential metabolic routes, the NeoChrs behaved like native chromosomes regarding mitotic and genetic stability, copy number, harmlessness for the host and editability by CRISPR/Cas9. This study paves the way for future microbial cell factories with modular genomes in which core metabolic networks, localized on satellite, specialized NeoChrs can be swapped for alternative configurations and serve as landing pads for the addition of functionalities.

## Supplementary Figure - Reactions from glucose to pelargonidin 3-O-glucoside

Yellow: glycolysis an ethanolic fermentation. Blue: pentose phosphate pathway. Brown: *E. coli* shikimate pathway. Green: plant anthocyanin pathway. The gene names encoding the enzymes involved in the indicated reactions are indicated in italics. Genes deleted in the present study are indicated in red and underlined. *Ec E. coli, At Arabidopsis thaliana, Rc Rhodobacter capsulatus, co* codon optimized. *fbr* feedback resistant, Glc glucose, Glc-6P glucose 6-phosphate, Fru-6p fructose-6-phosphate, Fru-1,6-BP fructose 1,6-biphosphate, GAP glyceraldehyde 3-phosphate, DHAP dihydroxyacetone, 1,3-BPG 1,3-biphosphoglycerate, 3-PG 3-phosphoglycerate,2-PG 2-phosphoglycerate, PEP phosphoenolpyruvate, Pyr pyruvate, AcAL acetaldehyde, EtOH ethanol, 6p-GLCN-lac 6-phosphogluconolactone, 6p-GLCN 6-phosphoglucono, RL5P ribulose 5-phosphate, R5P ribose 5-phosphate, S7P sedoheptulose 7-phosphate, XUL-5P xylulose 5-phosphate, GAP glyceraldehyde 3-phosphate, Ery-4P erythrose 4-phosphate, DAHP 3-deoxy-D-arabino-heptulosonate-7-P, DHQ 3-dehydroquinate, DHS 3-dehydroshikimate, SHIK shikimate, SHP shikimate 3-phosphate, EP3P 5-enolpyruvoyl-shikimate 3-phosphate, CHA chorismate, PPA prephenate, PPY phenylpyruvate, PAC phenylacetaldehyde 2PE 2-phenylethanol, PAA phenylacetic acid, PHE L-phenylalanine, *p*OHPPY *p-*hydroxyphenylpyruvate, *p*OHPAC *p*-hydroxyphenylacetaldehyde, *p*OH2PE, *p*-hydroxyphenylethanol, *p*OHPAA, *p*-hydroxyphenylacetic acid, TYR tyrosine, COUM coumaric acid, CIN cinnamic acid, COCOA coumaroyl-CoA, NARCC naringenin chalcone, PHLOR phloretic acid, NAR naringenin, DHK dihydrokaempferol, KAE kaempferol, K3G kaempferol 3-O-glucoside, LPE, leucopelargonidin, PEL pelargonidin, P3G pelargonidin 3-O-glucoside

## Supplementary Figure – Flow cytometric analysis of the test linear neochromosomes

Cells from shake flask cultures were analysed by FACS. The fluorescence is plotted on the y-axis and the forward scatter (FSC-A) on the x-axis. Negative control: CEN.PK113-7D. Positive controls: IMC111 (mRuby2), IMC112 (mTurquoise2) and IMF6 (mRuby2 and mTurquoise2). Gates for fluorescence of the two different fluorescent proteins were drawn based on the IMC111 and IMC112 controls. Approximately 10000 events are shown for each plot.

## Supplementary Figure - Separation of the test linear neochromosomes on pulsed-field electrophoresis

**A.** 1) IMX1338: control strain without neochromosome. 2) IMF6: control strain with 100 kb in plug linearized neochromosome NeoChr1. 3) IMF23: strain with 100 kb in plug linearized NeoChr12. 4) NeoChr10.10: correct size. 5)NeoChr10.13: correct size. 6) NeoChr10.16: wrong size. 7) NeoChr11.7: no visible neochromosome. 8) NeoChr11.8: wrong size. 9) NeoChr11.12: wrong size. 10) NeoChr11.19: correct size. 11)NeoChr11.22: wrong size. 12) NeoChr11.26: no visible neochromosome. 13) Size ladder

**B.** 1) IMX1338: control strain without neochromosome. 2) IMF6: control strain with 100 kb in plug linearized neochromosome (NeoChr1). 3) NeoChr10.30: no visible neochromosome. 4) NeoChr10.47: correct size. 5)NeoChr10.54: correct size. 6) NeoChr10.60: no visible neochromosome. 7) NeoChr10.62: correct size. 8) NeoChr10.67: correct size. 9) NeoChr10.69: correct size. 10) NeoChr11.29: no visible neochromosome. 11) NeoChr11.34: wrong size. 12) Size ladder

## Supplementary Figure - Sequencing results of test linear neochromosomes

*in silico* fragment configuration in NeoChr10 / NeoChr11 as well as *in vivo* fragment configuration of the neochromosome transformants as measured by long-read nanopore sequencing. The fragments of the *in silico* design which are present in the neochromosome transformants are connected by the same colored line. A dotted line indicates an area which is inverted. Fragments are color coded as follows: blue represents a correctly assembled fragment; red represents a missing fragment; green represents a duplicated fragment; and yellow indicates an inverted fragment.

A) NeoChr10.13 and NeoChr10.47 are missing an internal part of chunk 7A. B) NeoChr10.69 is missing 3 chunks: 8A, 15B and 4D. C) NeoChr10.10 is missing 4 chunks: 7A, 15C, 16B and 4D. D) NeoChr10.67 is missing 6 chunks: 7A, 19D, 16B, 4B, 4D and 9D , E) NeoChr10.62 is missing 4 chunks: 15B, 17B, 19A and 17D. In addition, a region containing the chunks 4A, 4B, 4C and 4D is duplicated , F) Neochr10.54 has a large inversion from 8D until 19A, from this region 2 chunks are missing: 15C and 18B. This region is link to a region spanning from 19A (which is thus duplicated) until the right telomere. From this region chunk 17D is missing, G) NeoChr11.19 contains several duplicated and inverted areas, from one area chunk 16A is missing.

## Supplementary Figure - NeoChr copy number estimation based on fluorescence

mRuby2 and mTurquoise2 fluorescence was measured by flow cytometry. CEN.PK113-7D with no fluorescent markers was used as negative control. IMX2224 and IMX2226 with a single copy of *mRuby2* and *mTurquoise2* integrated in the genome, respectively, were used as positive controls. All strains showed a fluorescence corresponding to the expected NeoChr. copy number.

Supplementary Figure 6 - NeoChr copy number estimation based on sequencing

IMF22 and IMF48 were analyzed by long-read Nanopore sequencing and IMF23, IMF41, IMF42 and IFM47 by short-read Miseq sequencing. Plots on the left represent the copy number of native chromosomes, while plots on the right show the NeoChrs copy number

## Supplementary Figure 7 - Flow cytometric analysis of (linear) NeoChr25 and (circular) NeoChr26 designed for anthocyanin production

Cells from shake flask cultures were analyzed by FACS. The fluorescence is plotted on the y-axis and the FSC-A on the x-axis. Negative control: CEN.PK113-7D. Positive controls: IMX2224 (mRuby2), IMX2226 (mTurquoise2). Gates for fluorescence of the two different fluorescent proteins were drawn based on the IMX2224 and IMX2226 controls. Approximately 10000 or 100000 events are shown for each plot.

## Supplementary Figure 8 - Separation of (linear) NeoChr25 transformants on pulsed-field electrophoresis.

Pulsed-field electrophoresis was used to estimate the size of NeoChr25 in several yeast transformants. 1) Size ladder. 2) NeoChr25.4: correct size. 3) NeoChr25.15: no visible neochromosome. 4) NeoChr25.19: no visible neochromosome. 5) NeoChr25.23: no visible neochromosome. 6) NeoChr25.24: no visible neochromosome. 7) NeoChr25.25: correct size. 8) NeoChr25.47: correct size. 9) NeoChr25.53: correct size. 10) NeoChr25.56: correct size. 11) NeoChr25.57: no visible neochromosome. 12) NeoChr25.73: no visible neochromosome. 13) NeoChr25.75: correct size. 14) IMF22: positive control. 15) Size ladder.

Supplementary Figure 9 – Duplication and inversion of four plant genes in linear NeoChr25 and circular Neochr26.

An unexpected recombination was observed upon integration of the genes encoding the anthocyanin production pathway in the linear and circular NeoChrs. A) Schematic representation of the *in silico* design for the integration of the anthocyanin pathway in the circular NeoChr26 of IMF40 resulting in IMF41 and the linear NeoChr25 of IMF34 resulting in IMF42. B) Schematic representation of the genetic organization observed in IMF41 and IMF42. The last four genes in the anthocyanin pathway (*coAtF3H*, *coGhDFR*, *coAtANS* and *coAt3GT*) were duplicated and inversed, and *ARS106* was absent. The dashed boxes illustrate the recombination events that occurred on the left and right flank of this duplicated region. For the left flank, there was probably an exonuclease and subsequent Non-Homologous End Joining (NHEJ) event between the two SHR CJ, since there was no homology between the inverted and non-inverted sequences. In the sequenced IMF41 strain (circular) 57 bp of SHR CJ was retained and in the sequenced IMF42 (linear) 51 bp of SHR CJ was retained. For the right flank, in the IMF41 strain (circular) the first 649 bp showed exact homology to *pSeFBA1*, while the last 414 bp showed exact homology to *pScFBA1* (100% homology overlap of 7 bp). In the sequenced IMF42 strain (linear) the first 29 bp showed exact homology to *pSeFBA1* and the last 710 bp showed exact homology to *pSceFBA1* (overlap of 100% homology is 24 bp).

Supplementary Figure 10 – schematic representation of *coAtANS* mutation in strains IMF41, IMF42, IMF44 and IMF47

**A)** The original coAtANS has a length of 1071 bp and encodes for an enzyme consisting of 356 amino acids. **B)** In strains IMF41, IMF42, IMF44 and IMF47, 21 nucleotides of non-homologous DNA (indicated in grey) together with 25 of the first 26 nucleotides of the *coAtANS* gene (indicated in dark green) were inserted right after the 26th nucleotide. **C)** This insertion resulted in a total insertion of 46 nucleotides disrupting the original ORF. However, this also resulted in a new ORF starting from the 284th nucleotide. **D)** The new ORF of the truncated coAtANS has a length of 834 bp and encodes for an enzyme consisting of 277 amino acids.

## Supplementary Figure 11 - Substrates and products profiles during aerobic batch cultivation in bioreactors of IMF41, IMF42 and IMF48.

**A)** IMF41 (Cir, 1x *coAtCHS3*), **B)** IMF42 (Lin, 1x *coAtCHS3*), and **C)** IMF48 (Lin, 9x *coAtCHS3, coAtANS*), were grown at 30°C in aerobic batch cultures in bioreactors, in chemically defined medium with 20 g L-1 glucose as sole carbon source (SMD). Biological duplicates were performed and are shown in two columns as #1 and #2.

Row 1) ■ CDW ( g L-1), ○ Glucose (mM), ● EtOH (mM), ⯆ PYR (mM), △ Glyc (mM)

Row 2) ■ PPY (mM), ⯅ COUM (mM), ● Phlor (mM), ⯆ *p*OH2PE (mM)

Row 3) ■ DHK (mM), ⯅ PEL (mM), ⯀ KEA (mM) **○** K3G (mM)⯆ P3G (µM)

## Supplementary Figure 12 – Detection and quantification of pelargonidin and pelargonidin 3-O-glucoside by LC-MS/MS

**A)** Extracted ion chromatogram for the pelargonidin 3-0-glucoside (P3G) mass peak with the composition C21H21O10+ and the m/z of 433.1. Data shown for the cell pellet extract of IMF48 duplicate #1 (Table 2), grown in aerobic bioreactor (sample, upper trace), for a blank injection (trace in the middle) analysed just before the sample and for a synthetic P3G standard shown in the lower trace (Pelargonidin 3-O-glucoside chloride, Sigma Aldrich, Cat No PHL89753).

**B)** The mass spectra show the accurate mass of P3G observed in the sample (upper mass spectrum) and the standard (lower mass spectrum). No corresponding P3G peak was observed for the blank injection (spectrum in the middle) analysed before the sample.

**C)** Extracted ion chromatogram of the pelargonidin (PEL) fragment with the composition C15H11O5+, and a m/z of 271.06 Da. The corresponding fragment was observed in the sample (upper mass spectrum) and the standard (lower mass spectrum). No corresponding PEL fragment peak was observed for the blank injection analysed just before the sample. (Pelargonidin chloride, Sigma Aldrich, Cat No PHL80084).

**D)** The spectra show the accurate mass of the PEL major fragment with the composition C15H11O5+ and a m/z of 271.06 Da, as observed for the sample (upper spectrum) and the standard (lower spectrum). No corresponding fragment mass peak was observed for the blank injection (spectrum in the middle), which was performed just before the sample.

**E)** The table summarised the chemical compositions of P3G and the major fragment of pelargonidin (PEL) (loss of the sugar unit), the resulting theoretical m/z values, the sobered m/z values and the mass deviations (ppm). The observed mass deviations for standard and sample peaks were <5 ppm compared to their theoretical m/z values.

## Supplementary Table - Promoter-gene-terminator combinations in the NeoChrs.

Promoters, genes or terminators originate from *S. cerevisiae* unless indicated by: *Ec*= *Escherichia coli*, *At*= *Arabidopsis thaliana*, *Rc= Rhodobacter capsulatus*, *Gh=* *Gerbera hybrida* *Se*= *Saccharomyces eubayanus*, *Sk*= *Saccharomyces kudriavzevii*, *co= codon optimized.* Watermarked *S. cerevisiae* genes1 are indicated with an \*.

## Supplementary Table - Sequence fidelity of NeoChrs

Mutation identified in the neochromosomes as compared to the *in silico* design and with the most relevant parental strain. The \* indicates mutations which are the same in two separate transformations and therefore probably resulting from the template DNA and not during the *in vivo* assembly. Non-synonymous mutations are indicated in bold.

## Supplementary Table - Amino acid substitution in native genome of NeoChr strains

Amino acid substitutions identified in the genome of the constructed strains as compared to most relevant parental strain.

## Supplementary Table - Extracellular concentration of aromatic compounds produced by engineered *S. cerevisiae* strains in shake flask cultures

Determination of the intermediates of the anthocyanin pathway in *S. cerevisiae* strains IMF41 (Cir NeoChr, 1X *coAtCHS3*), IMF42 (Lin NeoChr, 1X *coAtCHS3*), IMF47 (Lin NeoChr, 9X *coAtCHS3*) and IMF48 (Lin NeoChr, 9X *coAtCHS3* repaired *coAtANS*), grown in aerobic shake flask batch cultures on glucose (20 g L-1) and urea. The data represents the average ± mean deviation of independent biological triplicates. Intermediates of the anthocyanin pathway coumaroyl-CoA, naringenin-chalcone, and leucopelargonidin were not measured. \* Indicates statistical significance when comparing IMF47 or IMF48 to IMF42, and # when comparing IMF48 to IMF47 (Student *t*-test, two-tailed, homoscedastic, *p*-value threshold 0.05).

## Supplementary Table 5 - Physiological characterization of anthocyanin-producing strains grown in bioreactors

A) The specific growth rate (µ) and the yield (Y) of biomass (X) and ethanol (ETOH) on glucose (S)

B) The overall yield (Y) of glycerol (GLYC), pyruvate (PYR), coumaric acid (COUM), phloretic acid (PHLOR) and dihydrokaempferol (DHK) on glucose and ethanol (S) during aerobic bioreactor batch cultivation of IMF41 (Cir NeoChr, 1x *coAtCHS3*), IMF42 (Lin NeoChr, 1x *coAtCHS3*),) and IMF48 (Lin NeoChr, 9x *coAtCHS3*, repaired *coAtANS*).

## Supplementary Table 6 - *S. cerevisiae* strains used in this study

Strains that were short-read or long-read sequenced in this study are marked with a \**.* SHRs are differently annotated than in Kuijpers *et al.* 6. SHRs are annotated in bold subscript between de genetic fragments that they join together.

## Supplementary Table 7 - Neochromosome configurations

SHRs are differently annotated than in Kuijpers *et al.*. SHRs are annotated in subscript between the genetic fragments that they join together.

Supplementary Table 8 – Plasmids used in this study, 8A gRNA plasmids, 8B in-house golden gate part plasmids, 8C Part plasmids subcloned by GeneArt in entry vector pUD565, 8D Part plasmids ordered from GeneArt and subcloned in house in entry vector pUD565, 8E Part plasmids made in house by PCR, 8F in house expression plasmids described in other studies, 8G Expression plasmids constructed in this study, 8H Expression plasmids made by Gibson assembly in this study, 8I Other plasmids.

## Supplementary Table 9 - pROS/pMEL gRNA primers. gRNA sequence is underlined.

## Supplementary Table 10 - Primers to check correct construction of gRNA plasmids

## Supplementary Table 11 - Primers to make golden gate part plasmids and expression plasmids with Gibson assembly

## Table 12 - Diagnostic primers to check golden gate part plasmids

## Supplementary Table 13 - Diagnostic primers to check golden gate and Gibson assembly expression plasmids, for PCR and Sanger sequencing

## Supplementary Table 15 - List of primers for amplifying NeoChr10 and NeoChr11 chromosome parts

## Supplementary Table 16 - Primers for of amdSYM deletion

## Supplementary Table 17 - Primers for deletion of GND2, NQM1, SOL4 and TKL2

## Supplementary Table 18 - Primers for deletion of ura3, his3 and SpHIS5

## Supplementary Table 19 - Primers for deletion of ARO10

## Supplementary Table 20 - List of NeoChr25 (linear) and NeoChr26 (circular) chromosome parts

## Supplementary Table 21 - List of primers for amplifying NeoChr25 and NeoChr26 chromosome parts

## Supplementary Table 22 - Primers for glycolysis deletion

## Supplementary Table 23 - Primers to repair *RKI1* mutation in IMF32

## Supplementary Table 24 - Primers for deletion of native *ZWF1, GND1, SOL3, RKI1, TAL1, TKL1* and *RPE1* ORFs

## Supplementary Table 25 - Parts of the “basic design” of the anthocyanin pathway

## Supplementary Table 26 - List of primers for amplifying the fragments of the “basic design” of the anthocyanin pathway and diagnosing integration

## Supplementary Table 27 - List of primers for amplifying the fragments of the “elaborate design” of the anthocyanin pathway with several copies of the chalcone synthase and diagnostic PCR.

## Supplementary Table 28 List of primers for amplifying the correct *CoAtANS* transcriptional unit and the diagnostic primers used to confirm correct integration.

# Supplementary Methods

## Supplementary Methods 1: Strains, growth medium and maintenance.

## Supplementary Methods 2: Molecular biology techniques

## Supplementary Methods 3: Detailed construction of the host strain IMX2770

## Supplementary Methods 4: MinION long-read sequencing

## Supplementary Methods 5: Analysis of aromatics

### HPLC analysis of aromatic compounds up until naringenin

### Mass spectrometric analysis of anthocyanin pathway compounds