Chapter 6

Materials and Methods

This chapter provides further details of the materials and methods used in this work. Many of the methods used are described elsewhere in Chapters 2-4. To avoid repetition this chapter contains only additional materials and methods used during this Thesis.

6.1 Growth Medium

6.1.1 *Escherichia coli* Growth Media

LB

LB mix (FM) 200g

NaOH (10M) a few drops

H_2O up to 8L

LB-Amp

LB mix (FM) 200g

NaOH (10M) a few drops

800μl Ampicillin (50mg/ml)

H_2O up to 8L

LB-Amp-Agar

LB mix 125g

NaOH (10M) a few drops

500μl Ampicillin (50mg/ml)

H_2O up to 5L

Agar per 1L bottle 14g

6.1.2 *Saccharomyces cerevisiae* Growth Media

YPD/ rich medium

Yeast extract 10g

Peptone 20g

H_2O up to 1000ml

pH 5.4-5.7

10x Glucose (20% w/v solution) final conc. 2%

YPD-Agar

YPD medium

Agar 2%

Water-agar Made by autoclaving 500ml bottle filled with H_2O 300ml Agar 8g

YNB (10X) Final concentration is 0.17%. Made by dissolving 8.5g in 500ml water. Filter sterilised and stored at 4°C.

Ammonium sulphate (100X) Final concentration is 5g/L. Made by dissolving 5g in 500ml H_2O and sterilising. Stored at 4° C.

Monosodium glutamate (MSG; 100X) Final concentration is 1g/L. Prepared by dissolving 50g in 500ml and filter sterilise. Stored at 4° C.

Amino acids Mixture (25X)

L-Arginine 1.25g (f.c.: 50mg/L)

L-Aspartate 2.00g (f.c.: 80mg/L)

L-Isoleucine 1.25g (f.c.: 50mg/L)

L-Methionine 0.5g (f.c.: 20mg/L)

L-Phenylalanine 1.25g (f.c.: 50mg/L)

L-Threonine 2.5g (f.c.: 100mg/L)

L-Tyrosine 1.25g (f.c.: 50mg/L)

L-Valine 3.5g (f.c.: 140mg/L)

Prepared by covering the powdered amino acids with 20ml ethanol ON at RT, then dissolved by adding 980ml H_2O. Stored at 4° C.

Amino acid bases (100X) (Adenine, Histidine, Leucine, Lysine, Tryptophan, Uracil) Final concentration is 100mg/L. Prepared by dissolving 5g in 500ml and filter sterilisation (Uracil is sterilised with ethanol). Stored at 4° C.

SD (Synthetic-dropout)

10X YNB (DIFCO) Solution 40ml

25X Amino acid Mixture

16ml 100X MSG or Ammonium sulphate 4ml [1]

10X Glucose 40ml

100X Adenine 4ml

[1] SD plates containing G418 use MSG

100X Histidine 4ml

100X Tryptophan 4ml

100X Uracil 4ml

100X Leucine 4ml

100X Lysine 4ml

Other chemicals (G418, Thialysine) as needed

H_2O up to 400ml

This media is filter sterilised. Bases (e.g. uracil, histidine) are omitted according to experimental requirements to generate the required auxotrophic marker selection. Glucose can be substituted with other sugars as needed. SD-Agar is made by substituting the water with a bottle of melted water-agar, followed by pouring into petri dishes.

FOA medium This is used to counter-select the *URA3* marker. *URA3*$^+$ cells die on FOA medium, while Ura$^-$ cells survive. The solution is added (after filter sterilisation once FOA has dissolved) to a sterile bottle of 200ml H_2O and 8g agar.

10X YNB (DIFCO) Solution 40ml

100X Ammonium Sulphate 4ml

25X Amino acid Mixture 16ml

100X Histidine 4ml

100X Tryptophan 4ml

100X Uracil 2ml

100X Leucine 4ml

100X Lysine 4ml

100X Adenine 4ml

100X Lysine 4ml

FOA 400mg

H_2O up to 200ml

VB medium This is used to starve yeast cells to induce them to undergo meiosis/sporulation.

NaAC anhydrous 8.2g

KCl 1.9g

$MgSO_4$ 0.35g

NaCl 1.2g

Agar 15g

H_2O up to 1L

6.2 Other solutions

Most solutions were prepared by the staff of the Gurdon Institute as follows.

EDTA (0.5M, pH 8.0)

EDTA (Fisher) 372.2g

NAOH pellets 100g

10M NaOH to pH

H_2O up to 2L

Sodium Acetate (3M pH 5.2)

Sodium Acetate (anhydrous, Fisher) 492.18g

Glacial Acetic Acid ~200ml (enough for pH 5.2)

H_2O up to 2L

Sodium Chloride (5M)

NaCl (Fisher) 584.4g

H_2O up to 2L

Sodium dodecyl sulphate, SDS (20%)

SDS (Melford) 800g

H_2O up to 4L

TAE (50X)

Tris 1210g

Glacial acetic acid 285.5ml

EDTA 0.5M pH8.0

H_2O up to 5L

TBE (10X)

Tris (Melford) 540g

Orthoboric Acid (Fisher) 275g

EDTA (0.5M pH 8.0) 200ml

H_2O up to 5L

TE (pH 8.0)

1M Tris pH8.o

EDTA 0.5M pH8.0

H_2O up to 2L

Tris (1M, pH 6.8)

Tris (Melford) 242.2g

Conc. HCl to pH ~160ml

H_2O up to 2L

Tris (1M, pH 7.4)

Tris (Melford) 242.2g

Conc. HCl to pH ~146ml

H_2O up to 2L

Tris (1M, pH 7.5)

Tris (Melford) 242.2g

Conc. HCl to pH ~142ml

H_2O up to 2L

Tris (1M, pH 8.0)

Tris (Melford) 242.2g

Conc. HCl to pH ~96ml

H_2O up to 2L

Tris (1M, pH 8.8)

Tris (Melford) 242.2g

Conc. HCl to pH ~36ml

H_2O up to 2L

6.3 Microbial Strains

6.3.1 *Escherichia coli* strains

One Shot® TOP10 F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG

This chemically competent strain for plasmid construction was purchased from Invitrogen (Cat# C404010).

XL1-Blue recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacI^{q}ZΔM15 Tn10 (Tetr)]

This chemically competent strain used for plasmid construction was made in-house.

MAX Efficiency® Stbl2™ F- mcrA Δ(mcrBC-hsdRMS-mrr) recA1 endA1lon gyrA96 thi supE44 relA1 λ- Δ(lac-proAB)

This chemically competent strain was used for plasmid construction with unstable inserts and was purchased from Invitrogen (Cat# 10268019).

6.3.2 *Saccharomyces cerevisiae* strains

Name	Genotype	Reference
K699	MATa ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3 can1-100	Kim Nasmyth
K700	MATα ade2-1 trp1-1 leu2-3,112 his3-11,15 ura3 can1-100	Kim Nasmyth
YMH8	(K699) pol2::URA3-POL2	This work
YMH9	(YMH8)(K700)	This work
YMH10	(K699) pol3::URA3-pol3-S483N	This work
YMH11	(YMH10)(K700)	This work
YMH12	(K699) pol2::URA3-pol2-L439V	This work
YMH13	(YMH12)(K700)	This work
YMH14	(K699) pol2::URA3-pol2-V426L	This work
YMH15	(YMH14)(K700)	This work
YMH16	(K699) pol2::URA3-pol2-S312F	This work
YMH17	(YMH16)(K700)	This work
YMH18	(K699) pol2::URA3-pol2-P301R	This work
YMH19	(YMH18)(K700)	This work
YMH20	(K699) pol2::URA3-pol2-D290V	This work
YMH21	(YMH20)(K700)	This work
YMH22	(K699) pol2::URA3-pol2-M459K	This work
YMH23	(YMH22)(K700)	This work
YMH24	(K699) pol2::URA3-pol2-Q468R	This work
YMH25	(YMH24)(K700)	This work

Name	Genotype	Reference
YMH26	(K699) pol2::URA3-pol2-A480V	This work
YMH27	(YMH26)(K700)	This work
YMH28	(K699) pol2::URA3-pol2-4	This work
YMH29	(YMH28)(K700)	This work
YMH30	(K699) pol3::URA3-POL3	This work
YMH31	(K700)pol3::URA3-POL3	This work
YMH32	(K699) pol3::URA3-pol3-01	This work
YMH33	(K700) pol3::URA3-pol3-01	This work
YMH34	(K699) pol3::URA3-pol3-R316C	This work
YMH35	(K700) pol3::URA3-pol3-R316C	This work
YMH36	(K699) pol3::URA3-pol3-P332L	This work
YMH37	(K700) pol3::URA3-pol3-P332L	This work
YMH38	(K699) pol3::URA3-pol3-S375R	This work
YMH39	(K700) pol3::URA3-pol3-S375R	This work
YMH40	(K699) pol3::URA3-pol3-V397M	This work
YMH41	(K700) pol3::URA3-pol3-V397M	This work
YMH42	(K699) rev3::KanMX	This work
YMH43	(K699) pol2::URA3-pol2-A480V rev3::KanMX	This work
YMH44	(K699) pol2::URA3-pol2-P301R rev3::KanMX	This work
YMH46	(K699) pol2::URA3-pol2-S312F rev3::KanMX	This work
YMH52	(K700) pol2::URA3-pol2-S312F	This work
YMH53	(K700) pol2::URA3-pol2-P301R	This work
YMH54	(K699) msh2::KanMX	This work
YMH56	(K699) pol2::URA3-pol2-A480V msh2::KanMX	This work
YMH58	(K699) pol2::URA3-pol2-D290A-E292A-A480V	This work
YMH60	(K699) pol2::URA3-pol2-D290A-E292A-M459K	This work
YMH62	(K699) pol2::URA3-pol2-D290A-E292A-P301R	This work
YMH64	(K699) pol2::URA3-pol2-D290A-E292A-S312F	This work
YMH66	(K700) pol2::URA3-pol2-A480V	This work
YMH67	(K700) pol2::URA3-pol2-M459K	This work
YMH68	(YMH30)(K700)	This work
YMH69	(YMH36)(K700)	This work
YMH70	(YMH38)(K700)	This work

Name	Genotype	Reference
YMH71	(K700) pol3::URA3-pol3-01	This work
YMH72	(YMH34)(K700)	This work
YMH73-75	(K699) pol2::URA3-pol2-A480V msh2::KanMx	This work
YMH78	(K699)pol3::URA3-pol3-S483N(K700)pol2::URA3-pol2-P301R	This work
YMH81	(K699)(K700)pol2::URA3-pol2-P301R/pol2::URA3-pol2-P301R	This work
YMH82	(K699)(K700)pol2::URA3-pol2-M459K/POL2 msh2::KanMx/MSH2	This work
YMH83	(K699)(K700)pol2::URA3-pol2-S312F/POL2 msh2::KanMx/MSH2	This work
YMH84	(K699)(K700)pol2::URA3-pol2-P301R/POL2 msh2::KanMx/MSH2	This work
NOY408-1b	MATa ade2-1 ura3-1 his3-11 trp1-1 leu2-3,112 can1-100	Nogi et al. 1991
YSI101	(NOY408-1b) except fob1Δ::LEU2 (~150 rDNA copies)	[866]
YSI102	(NOY408-1b) except ~20 rDNA copies	[866]
YSI103	(NOY408-1b) except ~40 rDNA copies	[866]
YSI104	(NOY408-1b) except ~80 rDNA copies	[866]
YSI105	(NOY408-1b) except ~110 rDNA copies	[866]

6.4 Oligonucleotides

Oligonucleotides to generate POL2 mutants by site directed mutagenesis

pol2_S312F_SDM_fw
TAGATCAAATAATGATGATT**TTT**TATATGATCGATGGGGAAGG
pol2_S312F_SDM_rv
CCTTCCCCATCGATCATATA**AAA**AATCATCATTATTTGATCTA
pol2_V426L_SDM_fw
ACATGGATTGTTTCCGTTGG**CTG**AAGCGTGATTCTTATTTACC
pol2_V426L_SDM_rv
GGTAAATAAGAATCACGCTT**CAG**CCAACGGAAACAATCCATGT
pol2_P301R_SDM_fw
CGAAGCCGCCTTTAAAATTC**CGG**GATTCCGCCGTAGATCAAAT
pol2_P301R_SDM_rv
ATTTGATCTACGGCGGAATC**CCG**GAATTTTAAAGGCGGCTTCG
pol2_D290V_SDM_fw
ACCCTGTGGTAATGGCATTT**GTT**ATAGAAACCACGAAGCCGCC
pol2_D290V_SDM_rv

GGCGGCTTCGTGGTTTCTAT**AAC**AAATGCCATTACCACAGGGT
pol2_M459K_SDM_fw
TTGAACTGGATCCCGAATTA**AAG**ACGCCGTATGCATTTGAAAA
pol2_M459K_SDM_rv
TTTTCAAATGCATACGGCGT**CTT**TAATTCGGGATCCAGTTCAA
pol2_Q468R_SDM_fw
CGTATGCATTTGAAAAGCCA**CGG**CACCTTTCCGAATATTCTGT
pol2_Q468R_SDM_rv
ACAGAATATTCGGAAAGGTG**CCG**TGGCTTTTCAAATGCATACG
pol2_A480V_SDM_fw
ATTCTGTTTCCGATGCAGTC**GTT**ACGTATTACCTTTACATGAA
pol2_A480V_SDM_rv
TTCATGTAAAGGTAATACGT**AAC**GACTGCATCGGAAACAGAAT
Pol2Faiat-SDM-fw
CCCTGTGGTAATGGCATTT**GCT**ATA**GCA**ACCACGAAGCCGCCTTTAAA
Pol2Faiat-SDM-rv
TTTAAAGGCGGCTTCGGGT**TGC**TAT**AGC**AAATGCCATTACCACAGGG

Oligonucleotides to generate POL3 mutants by site directed mutagenesis

pol3_FAIAC_SDM_fw
TGCGTATCATGTCCTTT**GAT**ATC**GAG**TGTGCTGGTAGGATTGG
pol3_FAIAC_SDM_rv
CCAATCCTACCAGCACA**CTC**GAT**ATC**AAAGGACATGATACGCA
pol3_V397M_SDM_fw
TCATCAAAGTTGATCCTGAT**ATG**ATCATTGGTTATAATACTAC
pol3_V397M_SDM_rv
GTAGTATTATAACCAATGAT**CAT**ATCAGGATCAACTTTGATGA
pol3_R316C_SDM_fw
GGTCTCATACAGCTCCATTG**TGT**ATCATGTCCTTTGATATCGA
pol3_R316C_SDM_rv
TCGATATCAAAGGACATGAT**ACA**CAATGGAGCTGTATGAGACC
pol3_S483N_SDM_fw
CCTACACGTTGAATGCAGTC**AAT**GCGCACTTTTTAGGTGAACA
pol3_S483N_SDM_rv
TGTTCACCTAAAAAGTGCGC**ATT**GACTGCATTCAACGTGTAGG

pol3_P332L_SDM_fw
CTGGTAGGATTGGCGTCTTT**CTG**GAACCTGAATACGATCCCGT
pol3_P332L_SDM_rv
ACGGGATCGTATTCAGGTTC**CAG**AAAGACGCCAATCCTACCAG
pol3_S375R_SDM_fw
TAACAGGTTCAATGATTTTT**CGC**CACGCCACTGAAGAGGAAAT
pol3_S375R_SDM_rv
ATTTCCTCTTCAGTGGCGTG**GCG**AAAAATCATTGAACCTGTTA

Oligonucleotides to check polymerase mutant generation

Sc-pol3_out_fw GAAGAGCATGACCTGTCATCATTC
pRS306_out_rv GACCATGATTACGCCAAGCTCG
pol3_sq1 TACCAAAAGGAAAGTATTCG
pol3_sq2 GTCATCCAAATTGCCAACGT
pol3_sq3 ACTACAAATTTTGATATCCC
Pol2_promoter_rv: 5'GATCCATATTGCACACCAGAGCTGTT
pRS306_fw: 5'GGCGGACAGGTATCCGGTAAG

Oligonucleotides to delete MSH2

MSH2-F1 TTATCTGCTGACCTAACATCAAAATCCTCAGATTAAAAGT CGGATCCCCGGGTTAATTAA
MSH2-R1 TATCTATCGATTCTCACTTAAGATGTCGTTGTAATATTAA GAATTCGAGCTCGTTTAAAC
MSH2.3 TAAAGCCAATGAATTGGACG
MSH2.4 TTTCCAGTGGTCTAGAGACC

Oligonucleotides to delete REV3

REV3-F1 ATACAAAACTACAAGTTGTGGCGAAATAAAATGTTTGGAA CGGATCCCCGGGTTAATTAA
REV3-R1 ATAACTACTCATCATTTTGCGAGACATATCTGTGTCTAGA GAATTCGAGCTCGTTTAAAC
REV3.3 ACTGTTTAGAGAAAAGAAGC
REV3.4 AATGTGTGGGGAACTTATACG

Oligonucleotides to check the integration of polymerase mutant constructs into MEFs

EHom1_1fw GCTTGGGTGATGATGTTGGCTCCTGTAAA
EHom1_1rv CCGCGCTGTTCTCCTCTTCCTCATCTC
EHom2_1fw GCGGCATGGACGAGCTGTACAAGTGATTA
EHom2_1rv CCAGGACCTGCGGTAGTGGAAAGAGAAA
D1Hom1_1fw AGAGAATTGCTGAGAAAGGGGAGTGAGACA
D1Hom1_1rv CCGCGCTGTTCTCCTCTTCCTCATCTC
D1Hom2_1fw CCGCGATAATATGAGCCTGAAGGAGACCGT
D1Hom2_1rv TGGGTGGAGAAGGGCATCAGGAAGGAC

Oligonucleotides to generate the *POLE* P286R mutation in human cells

sgRNA-1 5'-ACCG-ATCTGGTCTGTCTCAGCATC-3'
and 5'-AAAC-GATGCTGAGACAGACCAGAT-3'
sgRNA-2 5'-ACCG-TCGATGGCCAGGTGAGCAGG-3'
and 5'-AAAC-CCTGCTCACCTGGCCATCGA-3'
ssODN (all mutations in lower case) CAAGGTCCCCATCCCAGGAGCTTACTTCCCAGAAG-gCACCTGCTCACCTGGCCAT CGATCATGTAGGAAATCATCATAATCTGGTCTGTCTCAGCAT-CAcGAAAtTTGAG GGGCAGTTTGGTCGTCTCAATGTCAAATGCCAAAACCACAGGGTC-CTGTGGGGA CAAAATAAGCATAAAGCCAAGCTCTAAACTCCCCA

6.5 Solutions

TE (1X) Tris-HCl pH 7.4 10mM
EDTA 1mM

TAE(1X) Tris-Acetate pH 8.0 40mM
EDTA 10mM

Gel Loading Dye, Purple(6X) Purchased from NEB (Cat# B7024S).

HyperLadder™ 1kb Purchased from BIOLINE (Cat# BIO-33026).

6.6 Protocols

DNA restriction DNA is digested with appropriate restriction enzymes according to specification of the supplier (New England Biolabs).

DNA ligation Fragments of DNA were ligated using the Quick Ligation™ Kit according to the specifications of the supplier (New England Biolabs).

Agarose gel electrophoresis DNA to be run on the gel is mixed with 1/6 Volume of 6X Gel Loading Dye and loaded onto an agarose gel (0.6-2%) containing 5μg/ml Ethidium bromide. A molecular marker (Hyperladder 1kb) is also loaded for size measurements. The gel is run in 1X TAE buffer and DNA is visualised under UV-light (260nm).

Plasmid extraction from *Escherichia coli* Plasmids were extracted from *E. coli* grown overnight in the appropriate culture medium using the QIAprep Spin Miniprep Kit (QIAGEN) as directed.

DNA extraction from agarose gels After gel electrophoresis a small slice of agarose, containing the DNA to be purified, is excised from the gel, weighed and the DNA is extracted using the QIAquick® Gel Extraction Kit (QIAGEN). A small aliquot is run on an agarose gel to assess the quality and efficiency of purification.

DNA precipitation 1/16 volume of KAc 3M pH 5.0 and 1 volume of Isopropanol (propan-2ol) are added to the DNA solution. Samples are spun for 10' at top centrifuge speed at RT and the supernatant is discarded. The pellet is washed with 1ml of 70% (-20°C) EtOH and the pellet is dried. The pellet is resuspended in 10-30μl TE buffer or water.

PCR (Polymerase chain reaction) PCR uses DNA as a template to amplify a target DNA fragment. Two oligonucleotides, flanking the fragment, acting as primers for the polymerase are required. The DNA polymerases used are Taq (qiagen 201203), Phusion (NEB #M0530L), The reaction mix contains:

Template DNA 25-100ng (depending on whether it is plasmid or genomic)
Oligonucleotides 20pmol each
10X DNA polymerase buffer 5μl
dNTPs (2mM each) 5μl
DNA polymerase 2units

dH_2O up to 50μl

Reactions are carried out in cycler machines from and consist of the following steps:

1| First denaturation 2' @94°C

2| Denaturation 1' @94°C

3| Annealing 1' @Tm-5°C

4| Extension 1' per kb of target fragment size + 2' @72°C

5| Repeat steps 2-4 for 25-30 cycles

6| Final extension 10' @72°C

The Tm is the lower melting temperature of the two oligonucleotides. All parameters can be adjusted depending on the DNA template, the purpose of the PCR and the DNA polymerase. For instance, for a yeast colony PCR (a diagnostic PCR where whole yeast cells are added to the reaction mix skipping the DNA extraction step), Step 1 should be increased to 7' to allow breaking of the cells and liberation of genomic DNA.

Site-directed mutagenesis Performed using Agilent Technologies QuickChange Lightning Kit (#210519-5) according to the manufacturers' instructions. Primers are designed according to the manufacturers' instructions (see Chapter 6.4 and 6.4).

***Escherichia coli* transformation** Chemically competent cells are transformed with DNA according to the manufacturers' protocols.

***Saccharomyces cerevisiae* transformation** The strain to be transformed is grown up in 50μl of the appropriate medium until the culture has reached a concentration between $5x10^6$ and $1x10^7$ cells/ml. The cells are pelleted and washed with 25ml of sterile water. Cells are re-suspended in 500μl of water of which 100μl are used for a transformation. Cells are pelleted again and resuspended in 360μl transformation mix (33% PEG-4000, 0.1M LiAc, 0.27mg/ml salmon-sperm DNA) and an appropriate amount of transforming DNA is added. The suspension is incubated at 42°C for 5' (plasmid transformation) - 40' (a transformation requiring an integration event). Cells are pelleted and washed with sterile water, resuspended in 200μl water and plated on selective medium. Should the selection require some time for gene expression (for instance resistance to G418) cells are suspended in rich medium and grown for 2hours at 30°C before plating.

***Saccharomyces cerevisiae* ONE-STEP gene deletion and tagging[900]** To generate a transformation cassette that features the selectable marker flanked by two regions of homology

suitable oligonucleotides are designed and ordered. The transformation cassette is amplified by PCR Mix preparation:

5μl F1 Oligonucleotide

5μl R1 Oligonucleotide

50μl 2mM dNTPs

50μl 10x Taq/Dynazime Buffer

5μl l pFA6 template plasmid (1:20 QIA)

382.5μl l H2O

2.5μl l Taq/Dynazime

The solution is mixed and 100\mu l are aliquoted in each tube.

Program:

2' @94°C

1' @94°C

1' @45°C

4' @72°C - 5 cycles

1' @94°C

1' @52°C

4' @72°C - 30 cycles

10' @72°C

The PCR product is purified with Gel Cleanup Kit (Eppendorf)/Gel Cleanup System (Promega) without band extraction and resuspended in 30\mu l H2O. The yield is checked on a gel and 1-2μg (usually 5-6μl) is transformed into yeast cells using standard transformation protocols. Plates are replicated at least once and at least 8 single colonies are isolated to check integration of the cassette. Deletion is checked by colony PCR (and subsequently perhaps by Western blotting): a small amount of cells is placed in a PCR tube. The following mix is prepared and 50μl of it is aliquoted into each PCR tube:

1μl FOR Oligonucleotide

1μl REV Oligonucleotide

5μl 2mM dNTPs

5μl 10X Taq/Dynazime Buffer

7.5μl H2O

0.5μl Taq/Dynazime

The PCR is run with the following programme:

7' @94°C

30" @94°C

30" @50/42°C (ca 5° below lower melting temperature)
4' @72°C - 45 cycles
10' @72°C
PCR products are checked on an agarose gel.

***Saccharomyces cerevisiae* gDNA extraction** Collection - Cells were collected by pelleting 50ml of yeast culture (107 cells/ml, 3000rpm, 2min). Cells were washed with 1ml 0.9Msorbitol 0.1M EDTA, the supernatant is discarded and cells are frozen for storage.

Extraction - Cells are resuspended in 400μl 0.9M sorbitol 0.1M EDTA 14mM β-mercaptoethanol with 100μl of 4-5mg/ml zymoliase and incubated at 37°C 30-45 minutes, 850rpm shaking. Cells are centrifuged for 30" at 13,000rpm, the supernatant is removed and cells are resusupended in 400μl of 1Z TE (pH8) with 90μl of the following freshly prepared solution: 1.5 ml of EDTA ph8.5 + 0.6 ml TRIS base 2M + 0.6 ml SDS 10%. The solution with the cells is gently mixed and incubated for 30min at 65°C, shaking 850rpm. 80μl Potassium Acetate 5M is added and cells are incubated 60min on ice. Cells are spun 15min at 13,000rpm at 4°C, the supernatant is decanted into a new tube, 500-1000μl 100% ethanol (EtOH) kept at -20°C is added and the liquids are mixed by inverting. Samples are left 30min at -80°C or at -20°C overnight to precipitate the DNA. Tubes are centrifuged 5min at 13,000rpm at 4°C, the supernatant is discarded and the pellet is washed with 1ml of chilled 70% EtOH (centrifuged 5min, 13,000rpm). The supernatant is removed, the DNA pellet is allowed to dry and the DNA is resuspended in 500μl 1X TE. Once resuspended, 5μl RNAseA is added and incubated for 30min at 37°C. Green phenol/chloroform tubes are pulsed down and the DNA solution is added. 500μl of phenol/chloroform is added, the solutions are mixed by vortexing and the tubes are centrifuged for 5min at maximum speed. The layer of liquid above the gel phase is moved to fresh tubes, 0.5ml isopropanol is added and the liquids are mixed by inverting. Samples are centrifuged 15min at 13,000rpm, the supernatant is discarded and the pellet is washed with 1ml 70% EtOH and the DNA is allowed to dry, then resuspended in 50μl 1X TE.

***Saccharomyces cerevisiae* high-throughput gDNA extraction** Cells were grown in 2ml 96-well plates 1.5ml YPD at 30°C shaking for 48 hours. Then, plates were spun down at 4000rpm for 5' and the supernatant removed. Cells were resuspended in 500μl of:

22.5ml 2M sorbitol
10ml 0.5M EDTA
50μl 14mM \beta -mercaptoethanol

5ml RNase A (stock 10mg/ml)

12.5ml H2O

200 – 250mg zymoliase

and incubated for 2hours at 37°C with shaking, followed by spinning down and removal of the supernatant. Cells were resuspended in 200μl of:

16ml ATL buffer (qiagen)

2ml proteinase k (qiagen)

2ml RNAse A (stock 10mg/ml)

and incubated at 56°C with shaking for 24 hours. The plate was placed on the robot (CAS1820 by Corbett Robotics) which carried out the following steps in a 96-well format.

1 | 400μl of buffer AL mixed 50:50 (qiagen 19075) with 100% ethanol was added to samples using fiter tips (qiagen 990610).

2 | After mixing the total volume (600μl) was loaded onto a capture plate (qiagen 950901) and vacuum applied at 70kPa for 2'30".

3 | The capture plate was washed twice with 600μl of buffer DXW (qiagen 950154) and once with 600μl of buffer DWF (qiagen 950163).

4 | The vaccum was applied at 30kPa to remove remaining liquid.

5 | 100μl of buffer E (qiagen 950172) was added, incubated for 30" and vaccum applied for 5' at 50kPa to elute samples into the elution plate (qiagen 990602).

The eluted samples were then transfered into a 96-well plate for sequencing.

Mating *Saccharomyces cerevisiae* There are two different options for mating two haploid cells to generate a diploid.

1 | Two small quantaties of yeast cells are mixed on a YPD plate and incubated at 30°C. The next day a small quantity of yeast is suspended in 50μl H_2O and a drop of 30μl is placed on a new YPD plate and the plate is tilted to spread the cells thinly. Under the dissection microscope roughly 10 diploids are identified (which at this stage appear in a "dumbbell" shape) and placed to an empty space on the plate.

2| Small amounts of strains to be mated are inoculated in 5ml YPAD and incubated static overnight at 30C. Cells are resuspended and diluted 1:2000 in ddH_2O. 100μl are plated on a YPAD plate and incubated overnight at 30°C. The next morning approximately 10 of the small colonies are picked and spread on a new YPAD plate. Diploid colonies are generally bigger, thus are picked first.

In both cases colonies are checked for ploidy by FACS and/or sporification.

6.7 Automated serial propagation platform

The evolving populations are maintained on top of agar surfaces in a home-made evolution chamber that controls moisture, light and temperature. Cells are kept in a 1536-well plate format on the agar surface and every fourth position is left empty. At each transfer, evolving populations are i) pinned onto the next evolution plate ii) pinned onto a scanning plate. The plate is scanned to track colony growth. Copies of evolution plates are stored as a forzen record at regular intervals. At the end of the timespan the final evolution plate is deconvoluted and populations are preserved in 96-well plates filled with 30% glycerol. The platform has since been published here [836].

6.8 Illumina sequencing

1-3 μg of extracted DAN was then supplied to the Sequencing Facility at the Wellcome Trust Sanger Institute, who sheared the DNA to 100-1,000 bp by using a Covaris E210 or LE220 (Covaris, Woburn, MA, USA) and size-selected fragments (350-450 bp) with magnetic beads (Ampure XP; Beckman Coulter). Illumina paired-end DNA library preparation were prepared by the Sanger, samples indexed and multiplexed. The DNA was sequenced on the Illumina HiSeq2500 generating 100bp paired-end reads which were aliged by the Sanger to the *S. cerevisiae* S288c assembly (R64-1-1/EF4) from Saccharomyces Genome Database (obtained from the Ensembl genome browser) using BWA[901], currently considered one of the most efficient alignment tools[860], and PCR duplicates were marked by using Picard 'MarkDuplicates'[902](see B.1.1.2).

6.9 Sequencing analysis

For parameters of all programmes used for sequencing analysis used see Appendix B.1.1.2 and for scripts written in the course of this work see `https://github.com/mareikeherzog/thesis-scripts`.

6.9.1 Quality control of DNA sequencing

Extracted DNA was tested for total volume, concentration and total amount by the sequencing facility of the Wellcome Trust Sanger Institute using gel electrophoresis and the Quant-iT^{TM} PicoGreen® dsDNA Assay Kit (ThermoFisher Scientific). The quality of the sequencing data

post-alignment was assessed using SAMTools stats (1.1+htslib-1.1), plot-bamstats, bamcheck and plot-bamcheck[903].

6.9.2 Alignment of sequencing reads to the reference genome

Fastq files were aligned to the relevant reference genome using BWA[901] and PCR duplicates marked using Picard MarkDuplicates[902] by the Wellcome Trust Sanger Institute. Where required files were realigned to a different reference genome using the same tools.

6.9.3 Variant Calling of SNPs and INDELs, Annotation and Filtering

Variant Calling was carried out using SAMTools mpileup[903], BCFtools call[903] and Scalpel[853]. Variants were annotated with Variant Effect Predictor[904] and vcf files were processed using BCFtools, VCFtools[905], BEDTools[906] and custom scripts.

6.9.4 Extracting mutational signatures

To extract mutational signtures, the SomaticSignatures[763] R package was used. A customn script was used to format all mutations from all strains into the required input format. An R wrapper script written by Kim Wong was used to run the different functionalities of the SomaticSignatures package in sequence following their methodolgy[907]. The number of signatures was set to 2-8. The normalizeMotifs function was used to normalize to whole genome trinucleotide frequencies. Signatures were also extracted using EMu[764].

6.9.5 Scripts written for this work

To analyse sequencing data software detailed in the previous sections was used (see 6.9.6 and B.1.1.2 for commands and parameters). However, some analysis steps required the use of scripts written specifically for that particular analysis. Scripts used to generate data detailed in this thesis are described below. The code for these scripts is stored in an online repository (`https://github.com/mareikeherzog/thesis-scripts`)

av_cov_bait_regions.pl A programme that takes a bam file as an input and calls a file that contains list of genomic regions. The script then uses samtools mpileup output to work out the coverage across all the bases within those regions and returns the average coverage. Written for mouse, but can be adapted to other organisms. Used in WES experiments to check the average coverage across regions covered by the baits.

bam_stats_table.pl A script turn the output from samtools stats in a table with key QC metrics to quickly check for substandard sequencing data.

bamtofastq.pl A script that generates commands like "bam2fastq -o reads#.fastq 13791_2#1.bam" for samples of interest.

budding_yeast_gene_name_conversion.pl A script that takes a list of S. cerevisiae systematic gene names and returns their standard name and a description. This basic operation has been re-purposed for other scripts that handle S. cerevisiae vcf files. The equivalent for S. pombe has also been written (fission_yeast_gene_name_conversion.pl).

consequence_display.pl This script will go through a vcf file and count the consequences of the mutations that were called. If a mutation is associated with more than one consequence the one deemed more sever will be displayed. (Severity is indicated by the order of consequences in the array e.g. a gained stop codon is judged more severed than an inframe deletion). Other variations of this script have been written to deal with multi-sample vcfs, distinguish between SNVs and INDELs or categorise mutations as 'coding', 'intronic', 'regulatory' and 'non-coding'.

coverage_of_gene_mouse.pl This script can be used to get the coverage across all exons of a specific gene for mouse WES bam files. A variation to do the same for human sequencing data has also been written.

filter_bait_regions.pl This is used in the analysis of mouse WES or targeted exon sequencing experiments. The script takes a vcf file (variant calling from the experiment in question) and a bed file that contains the genomic locations of the regions of interest (in a standard WES experiment that would be a file containing the location of all mouse exons). The script then removes all variants from the vcf file that do not fall within a region of interest.

gt-filter.pl These custom filters for vcf-annotate allow filtering of vcf files on three metrics. Genotypes set to . for samples with DP < 10, Genotypes set to . for samples with GQ < 95 and a minimum value of MQ>30 is required. Written with the help of Dr. Thomas Keane and Shane McCarthy.

intersect_vcf_mutlists.pl This script was used to check whether all mutations introduced into simulated genomes were actually found by the simulated sequencing and subsequent analysis and are present in vcf files or whether mutations found in the analysis were present in the mutation lists.

mask-hets.pl This custom filters for vcf-annotate will set genotypes to . for all mutations that are heterozygous e.g. 0/1, 1/2, etc. mask-homs.pl to remove homozygous mutations has been written, too.

merge_bams_samtools.pl A script that takes a list of bam file locations and, if the sample names of two successive bam files are the same, merges them into one bam file.

rDNA_cnv_estimate.pl A programme that will estimate the copy number for rDNA repeats.

raindrop_plot_distances_morechr.pl A script that takes a vcf file of mutations and outputs the distances between mutations in a way that they can be plotted with gnuplot to make a raindrop plot.

remove_shared_variants.pl A script that takes a multi-sample vcf file and removes mutations that occur in more than one sample. A variant to only remove mutations present in all samples has been written.

samtools.stats.cov.pl A script that takes samtools stats output and computes how many nucleotides have a coverage less than 5 or a coverage less than 10.

subset_loop_no_conversion.pl A script that takes a multi-sample vcf file as input and utilizes the vcf-subset command to separate the vcf file into its samples.

ty-realign.sh A script that takes a list of bam files, locates the corresponding fastq files and realigns them to the Ty custom reference genome.

vcf_stats_table_all.pl A script that will take the output of vcf-stats and output a table with the information such as INDEL_Count, SNV_Count, Transitions, Transversions, C>T, A>G, A>T, C>G, G>T, A>C as well as different lengths of small INDELs.

vcf_to_gene_list.pl A script that turns a vcf file into a table of mutations that affect genes. The information printed is: the type of mutation (SNV or INDEL), the chromosome, the position, the gene (its systematic and common name and a description), the consequence of the mutation (e.g. frameshift mutation), the number of homozygous and heterozygous mutations found across all samples in the vcf file, the names of samples carrying the mutation.

6.9.6 Step-by-step workflow of variant analysis

After quality control and alignent to a reference genome, analysis to extract variants present in samples that are not present in controls was carried out with the following steps and commands (see also B.1.1.2 for command parameters and 6.9.5 used):

Step1: Variant calling was performed against a reference genome

- S. cerevisiae: samtools mpileup -f Saccharomyces_cerevisiae.EF4.69.dna_sm.toplevel.fa -g -t DP,DV -C50 -pm3 -F0.2 -d10000 sample.bam | bcftools call -vm -f GQ > sample.vcf
- Mouse: samtools mpileup -f GRCm38_68.fa -g -t DP,DV -C50 -pm3 -F0.2 -d10000 sample.bam | bcftools call -vm -f GQ > sample.vcf
- optional (INDELs only): scalpel –somatic –normal control.bam –tumor sample.bam –bed WES_regions.bed –ref genome.fa

Step2: Ensembl variant effect predictor (VEP) was run on the vcf files

- variant_effect_predictor.pl –species saccharomyces_cerevisiae|mus_musculus -i sample.vcf –format vcf -o sample.vep.txt –force_overwrite –database
- vcf2consequences_vep -v sample.vcf -i sample.vep.txt > sample.csqs.vcf

Step3: The vcf files were checked for expected mutations (e.g. check for deletions, polymerase mutations or other expected mutations that should be present)

Step4: Filtering

- for mouse WES: perl filter_bait_regions.pl -i sample.csqs.vcf > sample.ex.vcf
- for scalpel generated vcf files: cat sample.somatic.vcf | vcf-annotate -f gt-filter.pl > sample.filt.vcf

- bcftools norm -f Saccharomyces_cerevisiae.EF4.69.dna_sm.toplevel.fa|GRCm38_68.fa sample.csqs.vcf > sample.norm.vcf
- cat sample.norm.vcf | vcf-annotate -H -f +/q=30/Q=50/SnpGap=7 > sample.annotate.vcf
- cat sample.annotate.vcf | vcf-annotate -f gt-filter.pl > sample.gq.vcf
- optional for haploid samples: cat sample.gq.vcf | vcf-annotate -f mask-hets.pl > sample.hets.vcf

Step5: Files were subjected to vcf-subset to remove variants that did not pass filters

- subset_loop_no_conversion.pl –> carries out the following command in a loop: vcf-subset -c sample_name sample.vcf -e > sample.sub.vcf

Step6: Sample files were intersected to remove any variants not aquired in the course of the experiment

- cat sample.sub.vcf | vcf-sort > sample.sort.vcf; bgzip -f sample.sort.vcf; tabix -f -p vcf sample.sort.vcf.gz
- vcf-isec -f -a -c sample.vcf.gz control1.vcf.gz control2.vcf.gz (...) > sample.isec1.vcf (commands for mouse samples also include files with varaints obtained from sequencing mice of the same background)
- if scalpel was also used: bedtools intersect -header -a sample.vcf -b sample.somatic.filt.vcf > sample.merged.vcf; cat sample.vcf | grep "#" -v | grep "INDEL" -v >> sample.merged.vcf (these commands retain all post-filtering and intersection SNVs identified by samtools mpileup and those INDELs identified by both variant callers).
- if there are replicates for a sample (e.g. post-propagation polymerase strains had two colonies from the same line sequenced): vcf-isec -f -a sample1.isec1.vcf.gz sample2.isec1.vcf.gz > sample.merge.vcf

Step7: All sample files were merged from one experiment into one vcf file

- for x in sort.*.vcf.gz, do list=$list‘echo "$x"‘; list=$list’ ’; done
- vcf-merge $list 2>/dev/null > experiment_merge.vcf

Step8(optional): Variants present in all samples or variants present in more than one sample were removed

- perl remove_shared_variants.pl -i Experiment_merge.vcf > merge_unique.vcf

Step9: Specific outputs were produced depending on the experiment

- Output number of SNVs and INDELs: for x in *.merge.vcf; do n=$(echo $x | sed 's/.merge.vcf//g'); m=$(cat $x | grep "#" -v | wc -l); o=$(cat $x | grep "#" -v | grep "INDEL" -v | wc -l); p=$(cat $x | grep "#" -v | grep "INDEL" | wc -l); printf "$n\t$m\t$o\t$p\n" >> Mutations.results.txt; done
- Calculate mean number of mutations and the standard deviation of numbers in the Mutations.results.txt file ($2 for total number of mutations, $3 for SNVs and $4 for INDELs): for x in $sample_names; do t=‘cat Mutations.results.txt | grep $x | awk '{sum+=$2} END { print (sum/NR)}’‘; s=‘cat Mutations.results.txt | grep $x | awk '{sum+=$2; array[NR]=$2} END {for(x=1;x<=NR;x++){sumsq+=((array[x]-(sum/NR))**2);}print sqrt(sumsq/NR)}’‘; printf "$x\t$t\t\t$s\n"; printf "$x\t$t\t\t$s\n" >> Mutations.results.txt; done
- Output mutation patterns (e.g. transitions versus transversions): perl vcf_stats_table_all.pl experiment_merge.vcf > stat_table.txt
- Output a list of mutation affecting genes: perl vcf_to_gene_list.pl -i experiment_merge.vcf > genelist.txt