Molecular evolution in yeast: Role of chromosomal inversions and translocations in speciation, adaptation and gene expression

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Life Sciences

2011

Samina Naseeb
Table of Contents

List of tables 7
List of figures 9
Abstract 12
Declaration 13
Copyright statement 14
Acknowledgements 15
Dedication 16
Rational for submitting the thesis in alternative format 17

Chapter 1- Introduction 18

1.1 Saccharomyces cerevisiae as a model organism 19
   1.1.1 DNA transformation and recombination in yeast 19
   1.1.2 Life cycle 20
   1.1.3 Genome of Saccharomyces cerevisiae 23
1.2 The taxonomy of yeast 25
1.3 Yeast genome evolution 28
1.4 Transposable elements (TE) and chromosomal rearrangements 30
1.5 Role of chromosomal inversions and translocations in gene order evolution, speciation and fitness 34
1.6 Co-expressed gene clusters 35
   1.6.1 Structure of the DAL cluster 36
   1.6.2 Function of DAL cluster in allantoin degradation 37
1.7 Aim of the project 40
1.8 References 41

Chapter 2- Impact of chromosomal inversions in co-expressed gene clusters 49

2.1 Foreword 50
2.2 Abstract

2.3 Background

2.4 Results

2.4.1 Transcription factor analysis of DAL cluster structure

2.4.2 Fitness assays of the wild type strains

2.4.3 Construction of the strains with different gene inversions

2.4.4 Plasmid extraction and digestion

2.4.5 Amplification of cassettes

2.4.6 Construction of the DAL2 inverted strain

2.4.6.1 Insertion of the loxP containing cassettes

2.4.6.2 Verification of cassette insertion

2.4.6.3 Cre transformation and induction

2.4.7 Construction of the strain with S. castellii like DAL cluster (S.cer-Inv1 and S.cer-Inv1-control)

2.4.7.1 Insertion of cassettes

2.4.7.2 Verification of cassette insertion

2.4.7.3 Cre transformation and induction

2.4.8 Summary of the strains constructed

2.4.9 Fitness assays of the engineered strains

2.4.10 Gene Expression Analysis of the Inverted and Non-Inverted Strains

2.4.10.1 The impact of gene inversion on expression

2.4.10.2 Effect of loxP/lox2272 insertion on gene expression

2.4.11 Expression of DAL4 antisense transcript (SUT614)

2.5 Discussion

2.6 Materials and methods

2.6.1 Strains and Media

2.6.2 Primers and Oligonucleotide probes

2.6.3 Plasmid DNA Extraction and Restriction Digestion
2.6.4 Transformation of \textit{Saccharomyces cerevisiae} by Lithium acetate protocol
2.6.5 Confirmation of the strains by colony PCR
2.6.6 Selection marker cassettes amplification by PCR
2.6.7 Construction of Inverted and Non-inverted strains
2.6.8 Fitness growth rate assay
2.6.9 Reverse Transcription
2.6.10 Real-time quantitative PCR
2.6.11 RNA extraction and Northern hybridization

2.7 References

Chapter 3- Effect of pericentric and paracentric chromosomal inversions on global gene expression

3.1 Foreword
3.2 Abstract
3.3 Background
3.4 Materials and Methods
   3.4.1 Strains and Media
   3.4.2 Fitness Growth Rate Assay
   3.4.3 Spore viability of inverted and non-inverted strains
   3.4.4 RNA extraction and whole genome transcriptome analysis
3.5 Results
   3.5.1 Strategy used to construct inverted and non-inverted strains
   3.5.2 Inversion length and Spore viability
   3.5.3 Recombination hotspots
   3.5.4 Fitness assay
   3.5.5 Whole transcriptome analysis
      3.5.5.1 Principal component analysis (PCA)
      3.5.5.2 Global expression profile of inverted and non-inverted strains
Chapter 4- Impact of chromosomal translocation on *S. paradoxus* and *S. cariocanus* speciation

4.1 Foreword
4.2 Abstract
4.3 Introduction
4.4 Materials and methods
  4.4.1 Strains and Media
  4.4.2 Oligonucleotides
  4.4.3 Plasmid DNA extraction and restriction digestion
  4.4.4 Transformation of *Saccharomyces paradoxus* by lithium acetate protocol
  4.4.5 Resistant marker cassettes amplification by PCR
  4.4.6 Construction of chromosomal translocations
  4.4.7 Confirmation of the engineered strains by colony PCR
  4.4.8 Yeast sporulation
  4.4.9 Yeast hybrid generation and tetrad analysis
4.5 Results and Discussion
  4.5.1 Construction of translocated and non-translocated strains
    4.5.1.1 Amplification of resistant marker cassette
    4.5.1.2 Insertion of cassettes
    4.5.1.3 Confirmation of cassette insertion
    4.5.1.4 *Cre* transformation and induction
  4.5.2 Spore viability of *S. cariocanus*
  4.5.3 Generation of *S. cariocanus × S. paradoxus* hybrids using the micromanipulation technique
  4.5.4 Spore viability of hybrid strains
Chapter 5- General conclusion and future work

5.1 Chapter 2. Impact of chromosomal inversions on co-expressed gene clusters 164
5.2 Chapter 3. Effect of pericentric and paracentric chromosomal inversions on global gene expression 167
5.3 Chapter 4. Impact of chromosomal translocations on *S. paradoxus* and *S. cariocanus* speciation 168
5.4 References 169

Appendix 171

References 179

Total word count: 37,136
List of Tables

1.1 Names and positions of 31 TY1 full length elements 33
2.1 Transcription factor binding sequences 55
2.2 Transcription factor binding sites for *S. cerevisiae DAL* cluster 57
2.3 Transcription factor binding sites for *S. castellii DAL* cluster 58
3.1 List of engineered strains possessing paracentric inversion 111
3.2 List of engineered strains possessing pericentric inversion 112
3.3 List of strains in which inversion was found to be lethal to the strain 113
3.4 List of strains showing the spore viability and number of recombination hotspots within the inverted region 116
3.5 Summary of growth rate analysis in C-limited and N-limited media 117
3.6 List of 10 genes in strain 25/24 which were highly changing expression 125
3.7 List of 10 genes in strain 12/13 which were highly changing expression 126
3.8 List of 10 genes in strain 5/8 which were highly changing expression 126
3.9 GO categories for Biological Function with FDR \( \leq 0.05 \) for strain 12/13 127
3.10 GO categories for Molecular Function with FDR \( \leq 0.05 \) for strain 12/13 128
3.11 GO categories for Biological Function with FDR \( \leq 0.05 \) for strain 25/24 129
3.12 GO categories for Molecular Function with FDR \( \leq 0.05 \) for strain 25/24 130
3.13 List of genes located in the vicinity of breakpoints in strain 25/24 132
4.1 Spore viability of *S. cariocanus* and *S. paradoxus* engineered strains 156
4.2 Spore viability of hybrids and number of total spores analyzed 158
A1 List of plasmids used in this study 172
A2 List of checking primers used for confirming inverted and non-inverted strains 172
A3 List of primers used for real time PCR 173
A4 List of probes used for northern blotting 173
A5 List of checking primers used for confirming translocated and non-translocated strains 174
A6 The genotypes of *Saccharomyces cerevisiae* inverted and non-inverted strains used in this study 175
A7 The genotypes of *Saccharomyces cerevisiae* strains possessing pericentric 176
and paracentric inversion used in this study

A8 The set of cassette amplifying primers for engineering inverted and non-inverted strains

A9 List of cassette amplifying primers used for engineering translocations
List of Figures

1.1 Life cycle of *Saccharomyces cerevisiae* yeast 22
1.2 Phylogenetic relationship among the *Saccharomyces* complex 27
1.3 Structure of *S. cerevisiae* Ty1 element 31
1.4 The allantoin degradation pathway 39
2.1 Gene organization of *S. cerevisiae* DAL cluster involved in Allantoin degradation and allantoin degradation pathway 53
2.2 Comparison of DAL genes between *Saccharomyces cerevisiae*, other *sensu stricto* species and *Saccharomyces castellii* 54
2.3 Size of the DAL cluster structure 56
2.4 Fitness assay for optimum allantoin concentration 59
2.5 Fitness assay of *S.castellii* WT and FY3 strains in 0.0125% allantoin medium 60
2.6 Fitness assay of WT *S. castellii* and *S. cerevisiae* lab isolate and natural isolates 61
2.7 The pZC1, pUG6 and pZC4 plasmid digest 63
2.8 The amplification of the cassettes 64
2.9 Cassettes insertion by homologous recombination 65
2.10 The colony PCR for confirmation of cassette insertion in transformant colonies selected on YPD + geneticin and YPD + hygromycin B 66
2.11 Selection of colonies after the pop-out of cassettes 67
2.12 Colony PCR of DAL2 non-inverted strains 68
2.13 Colony PCR of DAL2 inverted strains 69
2.14 The *lox2272-natNT2-lox2272* cassette insertion by homologous recombination 70
2.15 Colony PCR for confirmation of *lox2272-natNT2-lox2272* cassette insertion in transformant colonies grown on YPD + neoceothricin 71
2.16 Colony PCR of the strain with a *S. castellii* like DAL cluster (S.cer-Inv2). 72
2.17 Colony PCR of S.cer-Inv2-control strain 73
2.18 DAL cluster structure of wild type, inverted strains and non-inverted strains 74
2.19 DAL cluster structure of engineered strains possessing *S. castellii* like DAL cluster 75

9
2.20 Fitness assays of the WT and engineered strains in 0.0125% (w/v) allantoin medium.
2.21 Fitness assays of the WT and engineered strains in different media
2.22 Expression levels of \textit{DAL2}, \textit{DAL4} and \textit{DAL1} in DAL1-DAL2-DAL4-Inv and DAL1-DAL2-DAL4-NI strains
2.23 Expression levels of \textit{DAL2}, \textit{DAL4} and \textit{DAL1} in Scer-Inv1 and Scer-Inv1-control strains
2.24 Expression levels of \textit{DAL2}, \textit{DAL4} and \textit{DAL1} in DAL2-Inv and DAL2-NI strains
2.25 The map of the 5’UTR and 3’UTR in the \textit{DAL} gene cluster
2.26 Effect of \textit{loxP} and \textit{lox2272} insertion
2.27 Detection of sense and antisense \textit{DAL4} RNA transcripts in YPD medium
2.28 Detection of sense and antisense \textit{DAL4} RNA transcripts in allantoin medium
2.29 Detection of \textit{DAL4} sense and antisense transcripts in Proline + allantoin medium
3.1 The \textit{cre-loxP} recombination system to construct inverted and non-inverted strains
3.2 Recombination hotspots in the inverted region of strain 29/28
3.3 Fitness assays of strain 12/13 in different media
3.4 Fitness assays of strain 5/8 in different media
3.5 Fitness assays of strain 25/24 in different media
3.6 Fitness assays of strain 29/28 in different media
3.7 Principal component analysis (PCA) of microarray data set
3.8 Principal component analysis (PCA) of microarray data set
3.9 Comparison of global expression profile in engineered strains
3.10 Structure of the INO80 complex remodelling complex
3.11 Total number of genes changing expression on each chromosome
3.12 Total number of genes changing expression on each chromosome
4.1 The amplification of \textit{loxP-natNT2-loxP} cassette for transformation of
S. paradoxus strains

4.2 Cassettes insertion by homologous recombination 153

4.3 The colony PCR for confirmation of cassette insertion in transformant colonies grown on YPD + neoceothricin 154

4.4 The colony PCR for confirmation of translocations and non-translocations 155

4.5 The spore viability of S. cariocanus 156

4.6 Generation of hybrids and selection on SD medium with geneticin G418 157

4.7 Crosses in hybrids 159

5.1. Role of H2A.Z in DAL genes expression 166
Abstract

The University of Manchester
Samina Naseeb
Doctor of Philosophy

Molecular evolution in yeast: Role of chromosomal inversions and translocations in speciation, adaptation and gene expression

19th December 2011

Species belonging to the Saccharomyces ‘sensu stricto’ group present a good model for studying evolution due to their genetic versatility and availability of genomic data. Chromosomal rearrangements play a very important role in evolution of eukaryotes as well as prokaryotes as they can affect phenotypic characteristics and modes of speciation. In nature these rearrangements are most likely caused by highly mobile genetic elements, such as retrotransposons. Karyotypic changes can cause alterations in gene transcription causing genomic instabilities by inactivating or over-expressing particular genes. Only few studies looked at the importance of chromosomal rearrangements and their role on global and local gene expression.

The aim of this PhD project is to investigate the impact of chromosomal inversions and translocations on fitness adaptation, gene expression and speciation. I first demonstrated that a single gene inversion (of DAL2) within a co-expressed gene cluster can cause an alteration of the expression of inverted gene as well as the neighbouring genes, ultimately leading to a phenotypic change. I also showed that small and large size pericentric and paracentric inversions do not always alter the gene expression and in general have no effect on growth rate. It was also shown that effect of large inversion on gene expression is not always localized within the inversion but occurs globally. Finally, it is demonstrated that chromosomal translocations can be responsible for the reproductive isolation of Saccharomyces paradoxus and Saccharomyces cariocanus.
Declaration

The University of Manchester
PhD by published work Candidate Declaration

Candidate name: Samina Naseeb
Faculty: Life Sciences
Thesis Title: Molecular evolution in yeast: Role of chromosomal inversions and translocations in speciation, adaptation and gene expression.

Declaration to be completed by the candidate:

I declare that no portion of this work referred to in this thesis has been submitted in support of an application for another degree or qualification to this or any other university or other institute of learning.

Signed
Date
Copyright statement

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and she has given The University of Manchester the right to use such Copyright, including for administrative, promotional purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, in The University’s policy on presentation of Theses and The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations).
Acknowledgements

I would like to thank my supervisor Dr Daniela Delneri for her continued help, invaluable suggestions and provision during this work over the last three years. I would also like to express my gratitude to my advisor Dr Christopher Knight for his expertise, knowledge and valuable suggestions throughout my PhD.

I thank all the fellow researchers who have contributed in this research. In particular, I thank Dr Zorana Carter, David Minnis and Dr Leo Zeef for their input to the work presented in chapter 3.

I thank all the current and previous lab members of Delneri Lab for their extremely valuable experiences, support, and insights. In particular I would like to acknowledge Dr Zorana Carter, Sarah Hewitt and Katrazyna Hooks for spending their precious time in reading my thesis and discussing the experiments.

Finally, I would like to thank my parents, my brothers (Ashfaq & Ishtiaq), my sisters (Fouzia, Saufia, Shazia and Nazia) and all my friends for their constant motivation and support throughout the journey of PhD. They always provided a loving environment for me. Two friends I must mention are Urooj Zafar and Sabiha Khan. We’ve laughed, cried and among other things, cursed together. Thank you for your encouragement, support and most of all your humour. You both kept things light and me smiling. Very special gratitude to Valerie Price for her love, assistance and motivation not only in my PhD but in everything else as well. Many thanks are due to my dearest friend in Pakistan for all the moral support, encouragement and concern throughout this study.

I thank all the funding bodies The University of Karachi, The University of Manchester (Faculty of Life Sciences) and BBSRC (Biotechnology and Biological Sciences Research Council).
Dedication

I dedicate this work to my wonderful parents who have always been there when I needed them most and never complained about anything. I am thankful to them and especially for the ‘smart genes’ they passed on to me. They bore me, raised me, supported me, taught me, and loved me. They deserve far more credit than I can ever give them.
Rational for submitting the thesis in alternative format

The results in this thesis (chapter 2-4) are presented in the style of international peer-reviewed journal articles. Chapters 2 and 3 are written in the format of Genome Biology and chapter 4 is in format of FEMS Yeast Research. Each chapter therefore has its own list of references, Figure and Table numbers. It is acknowledged that as a consequence of this thesis format, there is some overlap and repetition of background, methodology and references. However, it is thought that the result is a clear and concise thesis summarizing the valuable work of this doctoral project.
CHAPTER 1

INTRODUCTION
1 Introduction

1.1 *Saccharomyces cerevisiae* as a model organism

*Saccharomyces cerevisiae* belongs to a group of unicellular eukaryotic fungi. It is widely known not only for its use in brewing and baking industries but also in biotechnology and research. Among all the lower eukaryotes *Saccharomyces cerevisiae* is the most ideal microorganism for studying the cellular biology of higher eukaryotes. It has been used as an experimental organism since mid-thirties of the 20th century [1, 2] and since then has received increased attention. Some properties that make *S. cerevisiae* particularly suitable for biological studies are its fast and simple growth, ease of replica plating, accessible biochemistry, mutant isolation, a well-defined genetic system with ease of genetic manipulation and the technical breakthrough of yeast transformation used in reverse genetics. Being non-pathogenic *S. cerevisiae* can also be handled with great ease and is therefore used widely in baking and brewing industries [3].

*Saccharomyces cerevisiae* has been used extensively as a model organism to investigate the molecular biology and biochemistry of eukaryotic cells, and is now also being used as a model organism in post-genomic research. It has been widely used at the genomic, proteomic and metabolic level, in gene expression studies [4]. The availability of molecular genetics techniques has helped to generate a collection of yeast mutants with single and double gene deletions. These mutants have been widely used to study gene characterization and functional profiling [5]. At the transcriptome level the use of microarrays has helped in understanding the global expression profile of yeast genes. Yeast has also been used at the proteome level to study protein-protein interactions, protein turnover and localization [6].

1.1.1 DNA transformation and recombination in yeast

One of the greatest properties of yeast is that it possesses a highly efficient DNA transformation and recombination system. Both circular and linear DNA can be transformed into yeast cells by integration into the genome. Several different
transformation protocols are available including, transformation of cells treated with lithium salts, transformation of spheroplasts and transformation by electroporation. The transforming DNA in yeast cells integrates by homologous recombination and therefore exogenous DNA can be directed to specific locations in the genome. Mutagenesis can be done by transforming the yeast with synthetic oligonucleotides. The mutant phenotypes arising after gene disruption can be used to study the function of certain proteins \textit{in vivo}, gene regulation, chromosome structure and structure-function relationships of proteins. The development of DNA transformation in yeast has been particularly useful in gene cloning and genetic engineering techniques [3].

1.1.2 Life cycle

\textit{S. cerevisiae} is also a useful model organism in genetic studies because it can exist in both haploid and diploid state. Haploid cells are of \textit{MATa} and \textit{MATα} types. The \textit{a} and \textit{α} mating types are under the control of a pair of \textit{MATa}/\textit{MATα} heterozygous alleles (Figure 1.1). During sexual reproduction the haploid cells of opposite mating types conjugate to form a heterokaryon, which is a diploid cell and capable of reproduction by budding. Conjugation occurs by contact of projections on the surface of mating cells which is followed by fusion of their plasma membrane. Meiosis is induced under starvation conditions leading to sporulation and finally to propagation of four haploid spores in an ascus (Figure 1.1). However under favourable conditions the ascus, also known as a tetrad, breaks thus segregating into a recognisable 2:2 pattern i.e. forming two haploid cells of \textit{MATa} and two haploid cells of \textit{MATα} [7].

\textit{Saccharomyces cerevisiae} can be maintained in either a heterothallic or homothallic stage. The term heterothallic is used if the gametes can only be fertilized by gametes from another individual. The term homothallic, however is used if cells derived from the same individual can fertilize each other. Wild-type yeast strains are homothallic since they have the ability to switch mating type, they can switch their mating types and fertilize each other. This condition of wild type strains is unfavourable in the laboratory as they can
switch mating type, therefore heterothallic strains which maintain their mating type and haploid state are engineered.

Apart from sexual reproduction *S. cerevisiae* and other yeasts also reproduce vegetatively by budding, a process in which a new bud grows from the side of the existing cell wall. This bud then separates from the mother cell to form a daughter cell, followed by nuclear division and cell-wall formation. Each yeast cell, on average, undergoes the budding process 12-15 times and each mother cell usually forms no more than 20-30 buds before it is no longer capable of reproducing [8].
Figure 1.1: Life cycle of *Saccharomyces cerevisiae* yeast. Yeast exists in both haploid and diploid state. The haploid gametes germinate and divide by mitosis. Two haploid cells of opposite mating type i.e. a and α fuse and form a diploid zygote that reproduces mitotically. Upon starvation, the zygote undergoes meiosis and forms an ascus containing four spores, two of each mating type. Upon favourable conditions the haploid spores can germinate and divide by mitosis.
1.1.3 Genome of *Saccharomyces cerevisiae*

*S. cerevisiae* was the first unicellular eukaryote whose genome in 1996, was fully sequenced [9]. It possesses a small genome of approximately 13Mb which is divided into 16 chromosomes [10]. All 16 chromosomes can be separated by pulsed field gel electrophoresis (PFGE) by choosing the appropriate conditions. Originally the number of protein coding genes were found to be between 5300 and 5400. The complete genome sequence now defines about 7500 open reading frames (ORFs) most of which encode specific proteins [11]. In a very recent study, Oheigeartaigh and his colleagues identified 595 additional protein coding genes in 11 yeast species using their newly designed software known as SearchDOGS, which was more precise in identifying the ORFs [12]. Approximately 1800 yeast ORFs of unknown function have homologues in the yeast or other genomes. Oliver and co-workers reported that a quarter of yeast genes that were identified by traditional methods were orphan [13]. There are about 200-300 orphan-ORFs which do not possess any known function or homology to already discovered genes and they contribute to only 5% of the total genes [14]. The majority of the protein-encoding open reading frames (~7500 ORFs) have a size of 100 to more than 4000 codons and cover nearly 70% of the total genome sequence [15]. ORFs shorter than 150 codons are designated as non-coding [16]. The protein-encoding gene in yeast is found every 2 kb compared with its complex eukaryotic relatives e.g. every 6 kb in *Caenorhabditis elegans* and 30 kb in humans [17].

Apart from the protein-coding genes, the *S. cerevisiae* genome also contains 140 ribosomal RNA genes on chromosome XII arranged in a large tandem array, 40 genes encoding small nuclear RNAs (sRNAs), 275 tRNA genes (belonging to 42 families) and Ty elements all of which are distributed across the chromosome. A very small portion of the yeast genome (approximately 4% of the total) is interrupted by introns [15]. Introns are generally located towards the start of the coding sequence at the extreme 5’ end of each gene. Most of the intron-containing genes encode for the ribosomal proteins. Up until now only two genes in the yeast genome have been recognized where two introns are present: a ribosomal protein gene *RPL6A* on chromosome VII and the *MAT* locus on chromosome III. *S. cerevisiae* is
therefore, a useful reference for comparing the sequences of animal, human and plant genes as well as multitude of unicellular organisms.

A large number of yeast mutants have been generated by disrupting yeast genes which has significantly helped to study the structure-function relationship of various proteins, protein-protein interaction, analysis of gene regulation, chromosome structure and other general questions in cell biology (Guthrie and Fink, 1991). There is a collection of yeast deletion mutant strains available, each bearing a defined mutation in one of 6000 potential protein coding genes of *S. cerevisiae* [18]. This deletion mutant collection provides the possibility to link genes with phenotypes for the quantitative analysis of phenotype. For example measurement of the intracellular concentration of metabolites in these mutant strains has revealed certain proteins active in metabolic regulation [19]. For example, Delneri and co-workers used these deletion mutant strains for competition experiments to identify the genes that show haploinsufficiency phenotypes (reduced growth rate when hemizygous) and haploproficiency phenotypes (increased growth rate when hemizygous) [20].

*S. cerevisiae* is an ideal system for functional genomics in which cell architecture and fundamental cellular mechanisms can be easily and successfully investigated. The availability of the complete genome sequence of *S. cerevisiae* stimulated much innovation in genome scale experimental approaches, including the development and implementation of DNA microarrays, systematic deletion studies and the use of hybridisation arrays. Development of all these technologies provided a way to analyze the genome at the level of mRNAs (transcriptome), proteins (proteome) and low molecular weight intermediates (metabolome). Messenger RNA molecules are not functional entities and therefore provide an indirect manner of analyzing gene function. However, proteins are final biological and functional entities and provide a direct evidence of gene function. Metabolites are also true functional entities within the cell and more than one gene may be involved in formation and degradation of a single metabolite. Various bioinformatics tools are required to elucidate the function of various genes, thus bioinformatics is a key to functional genomics [21].
1.2 The taxonomy of Yeast

*Saccharomyces cerevisiae* and other fungi were initially identified on the basis of phenotypic characters such as growth on organic compounds, fermentation of sugars and presence of morphological features. The taxonomy was difficult to deduce from this information, and therefore the taxonomist began to use molecular analysis of nuclear DNA hybridisation, electrophoretic enzyme comparison and DNA and RNA sequence comparison to elucidate taxonomy. The closely related species of *Saccharomyces cerevisiae* were then easily compared through phylogenetic analysis of sequences [22]. In recent years it has been shown that population genomics is a suitable approach for analysing phylogenetic relationships [23].

There are millions of fungal species in existence but *S. cerevisiae* is one among many yeast species that have been completely sequenced and fully described. *S. cerevisiae* belongs to the phylum Ascomycota, which is divided into three subphyla: Archiascomycotina, Pezizomycotina and Saccharomycotina (Figure 1.2). Ascomycota forms the largest phylum and this group diverged from basidiomycota 741-1195 million years ago [24]. The species in this phylum are characterized by the formation of a specialized structure, the ascus, which surrounds the spores produced during meiosis [25]. The ascomycetous yeasts can be phenotypically separated from euascomycetes by the formation of spores enclosed in a fruiting body and by the presence of budding.

Pezizomycotina includes mostly hyphal fungi such as *Neurospora crassa* whereas the Archiascomycotina, has been proposed containing yeast *Schizosaccharomyces pombe* [26]. Saccharomycotina also known as hemiascomycetes is a diverse group including more than 1000 yeast species that fall into three clusters. The first cluster known as *Saccharomyces complex* (now *Saccharomyces*), mostly consist of species from *Saccharomyces* and *Kluyveromyces* genera as shown in Figure 1.2 [27]. The second cluster consists of *Candida* species and yeasts such as *Debaryomyces hansenii* and *Lodderomyces elongisporus* [24]. These species translate CTG codons as serine rather than leucine, a change that occurred.
more than 170 million years ago [28]. The third cluster includes only one species *Yarrowia lipolytica* whose genome is sequenced [29].

The *Saccharomyces* group includes 14 clades and the yeasts belonging to the genus *Saccharomyces* are divided into two groups i.e. *Saccharomyces sensu stricto* (*Saccharomyces*) and *Saccharomyces sensu lato* (*Kazachstania*). *Saccharomyces* is a widely studied group and represents monophyletic group of interbreeding species which have an equal number of chromosomes and possess an overall phenotypic similarity [27]. The species included in this group are *S. cerevisiae*, *S. paradoxus*, *S. bayanus*, *S. pastorianus*, *S. mikatae*, *S. cariocanus* and *S. kudriavzevii* (Figure 1.2). *Kazachstania* group consists of yeast species which are non-sporulating and possess heterogeneity in their chromosomal content. These species are more distantly related to *S. cerevisiae*, *N. castellii* and *K. barnetti* [30].
Figure 1.2: Phylogenetic relationship among the *Saccharomyces* complex. The phylogenetic tree shows the yeast species resolving the “*Saccharomyces* complex” into clades, with their original designated names along with the new genera names shown on the right side of the tree. Figure adapted from [31].
1.3 Yeast genome evolution

Analysis of the complete sequence of the *Saccharomyces cerevisiae* genome has provided a pathway to study the origin of modern yeast traits. Whole-genome duplication and horizontal gene transfer (HGT) are the two molecular mechanisms that have played major roles in yeast genome evolution [32]. *Saccharomyces cerevisiae* possess approximately 10 genes which were horizontally transferred to its genome from bacteria. *URA1* is one of the genes transferred horizontally from bacteria to *Saccharomyces cerevisiae* and its other close relatives which enabled them to grow under anaerobic conditions [33]. Hall and Dietrich in 2007 reported an updated analysis of horizontally acquired genes in *Saccharomyces cerevisiae*. Apart from the 10 horizontally transferred genes that have been studied previously, they reported three more genes which were known to be transferred by HGT. Two of these genes were *BIO3* and *BIO4* involved in the biotin biosynthesis pathway. The third gene was *BDS1* encoding the alkyl-aryl sulfatase which is required for using alkyl and aryl sulphates as sulphur sources [34].

Gene and genome duplication are principle sources of complexity during evolution. According to Ohno, evolution is based on genome duplication and the duplicated copy of gene may remain in the genome resulting in gene gain or gene loss. The new gene will either acquire a new function or become non-functional or maintain function but be expressed at a different development stage [35].

The genome of *Saccharomyces cerevisiae* possesses great redundancy which became clear after its genome was fully sequenced [36]. Comparison of genome sequences of extant species revealed three different kinds of duplications: large duplicated segments with conserved gene orientation known as cluster homology regions (CHRs), genes duplicated singly in multiple copies and dispersed throughout the genome, and subtelomeric regions consisting of middle repetitive elements [37].

The origin of gene duplications became clear after the release of *Saccharomyces cerevisiae* complete genome sequence. Wolfe and Shields (1997) were first to analyze the event of
whole-genome duplication [38]. They observed that in many places in the yeast genome a series of genes on one chromosome had a series of paralogs on another chromosome usually in the same order and conserved transcriptional orientation along the chromosome. They identified 55 such duplicated regions in the *Saccharomyces cerevisiae* genome and centromere-to-telomere orientation was almost always conserved in each pair of genes. According to that study the modern genome of *Saccharomyces cerevisiae* originated from duplication of the whole genome [39]. An updated map of duplicated regions was given in 1999 by Seoighe and Wolf who showed that 52 pairs of duplicated regions remained unchanged and identified further 32 paired regions [40]. Kellis *et al.* in 2004 confirmed that *S. cerevisiae* originated from a complete duplication of eight ancestral chromosomes and it later gained functional ploidy by a massive loss of almost 90% of its duplicated genes.

Recent studies using different yeast genome sequences have revealed that entire duplicated genome existed at some point during yeast’s evolutionary past and that the whole genome duplicated at some stage followed by rearrangements principally reciprocal translocations and gene loss [41] and [42]. Duplicate genes arise by a number of different mechanisms, for example by retro transposition [43], replication errors [44], unequal crossing-over [45] and by non-disjunction [46]. However all independent groups of organisms studied so far have used gene duplications as the primary source of generating novel gene structures [47]. During the initial periods of whole genome duplication it was found that in some species, such as *Kluyveromyces lactis* and *Saccharomyces kluwerii*, the gene order remained the same even after whole genome duplication [41]. This finding was confirmed in 2004 after the publication of the complete genome sequence of *Ashbya gossypii* [48], *Kluyveromyces waltii* [49] and *Kluyveromyces lactis* [29]. Each of these species possesses a gene order similar to the existing ancestor of *Saccharomyces cerevisiae*.

Duplicated gene sequences examined from different *Saccharomyces* yeast genomes have shown that upon duplication, each duplicated gene has been altered independently through different mechanisms i.e. specialisation, deletion or differentiation of a single copy and this can have significant impact on their evolution [50]. Wolfe (2006) found that after whole genome duplication, large numbers of genes were deleted and only few survived [51].
Among the duplicated genes that survived, often one member had rapidly evolved into a novel gene with a derived function. The genes that do not possess any function gradually disappeared from the genome, for example genes with splicing and RNA interference functions have been lost from hemiascomycetes due to lack of introns in them [52]. It is observed that similar functional genes have been preserved in the duplication events. For example, transcription factors/kinases were preserved in duplicate after WGD in yeast [53], plants [53] and animals [54].

Genome duplication can alter chromosomal rearrangement either by doubling the chromosome number or by changing the gene location on the chromosome triggering evolutionary changes. It can occur either by doubling the chromosome number (autopolyploidization) or by hybridization between two species (allopolyploidization) in which the new species have sum of the number of chromosome number in its parents. A comparison of the *Saccharomyces cerevisiae* genome with *Kluyveromyces waltii* showed that the whole genome duplication involved an 8- chromosome ancestral genome doubling to form a 16- chromosome descendent[51]. It is therefore important to study chromosomal rearrangements and their role in evolution, speciation and adaptation.

1.4 Transposable elements (TE) and chromosomal rearrangements

Transposable elements (TEs) are mobile genetic sequences present in almost all the species ranging from prokaryotes to eukaryotes. The human genome is comprised of 35 % TEs [55], the maize genome consist of more than 50% TEs [56] and 3.1% of the yeast genome [57] is made up of TEs. In eukaryotes, there are two classes of TEs based upon their structure and mechanism of transposition. Class I elements encode for reverse transcriptase (RT) and their mode of transposition is mediated by RNA. This class of TEs include retrotransposons, retroposons and retrointrons. The class II elements include transposons and their mode of transposition is mediated by DNA [58].

The transposons in yeast are referred as “Ty” elements. The *S. cerevisiae* genome consists of five classes of Ty elements which are Ty1 [59], Ty2 [60], Ty3 [61], Ty4 [62] and Ty5
All the five classes possess similar structure consisting of TYA1 and TYB1 genes encoding for “gag” and “pol” genes.

Ty1 elements are the most abundant with 32 copies present (Table 1.1) contributing to at least 2.1% of \textit{S. cerevisiae} genome. They are \textasciitilde5.9 kb long consisting of a \textasciitilde5.2 kb long unique sequence known as epsilon sequence and flanked by \textasciitilde330 bp Long Terminal Repeat (LTR) sequence also known as delta sequence on each end (Figure 1.3). The homologous recombination between the two LTRs results in deletion of the unique epsilon sequence leaving behind just the delta sequence [57]. The highly repetitive Ty1 elements of the \textit{S. cerevisiae} genome are analogous to abundantly repetitive human Alu and LINE sequences. A recent study done on Ty912 element (identified as a Ty1 element) located on the non-essential terminal region of chromosome V showed that this element can increase the rate of gross chromosomal rearrangements to \textasciitilde380fold [63].

![Figure 1.3: Structure of \textit{S. cerevisiae} Ty1 element.](image)
The Ty1 elements are \textasciitilde5.9kb long. They consist of two genes \textit{gag} and \textit{pol} and these are flanked by Long Terminal Repeat (LTR) sequence on each end.

It has been shown that each type of Ty element integrates into its own specific site. Ty1 and Ty3 prefer to integrate upstream of RNA polymerase III transcription [64] whereas Ty5 preferably integrates near the telomers [65].

TEs being highly mobile elements play a very important role in evolution as they can rearrange the genome of an organism by homologous recombination. They can also integrate into new sites in the genome bringing mutations in the gene coding sequence and transcription. These rearrangements can either be inter-chromosomal or intra-
chromosomal. The inter-chromosomal rearrangements include translocations which involves the crossing over between two Ty elements located on two non-homologous chromosomes. Translocation can either be monocentric, dicentric or acentric depending upon the orientation of Ty elements. The intra-chromosomal rearrangements include deletions, duplication and inversions. Homologous recombination between the two repeats that are present in direct orientation flanking a segment of DNA will result in deletion of the DNA segment. An inversion occurs by crossover between the two repeats present in the inverted orientation. It can either be a paracentric inversion or pericentric inversions depending upon the location of the centromere. An inversion is said to be paracentric if the centromere is outside the inversion whereas inversions spanning the centromere are pericentric [66].

These rearrangement events are quite common and range from part of a gene to hundreds of genes causing genomic instabilities by inactivating the gene, changing the gene expression or forming genes encoding for altered proteins. These changes not only cause genome evolution but may often result in genetic diseases and cancers. Some examples of the genetic diseases occurring due to chromosomal loss or chromosomal addition include Cystic Fibrosis, Thalassaemia, Fragile X syndrome, Turner syndrome, Haemophilia A and Down syndrome [67].
Table 1.1: Names and positions of 31 TY1 full-length elements [57].

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome number</th>
</tr>
</thead>
<tbody>
<tr>
<td>YARCTy1-1</td>
<td>1</td>
</tr>
<tr>
<td>YBLWTy1-1</td>
<td>2</td>
</tr>
<tr>
<td>YBRWTy1-2</td>
<td>2</td>
</tr>
<tr>
<td>YDRCTy1-1</td>
<td>4</td>
</tr>
<tr>
<td>YDRCTy1-2</td>
<td>4</td>
</tr>
<tr>
<td>YDRCTy1-3</td>
<td>4</td>
</tr>
<tr>
<td>YDWRty1-4</td>
<td>4</td>
</tr>
<tr>
<td>YDWRty1-5</td>
<td>4</td>
</tr>
<tr>
<td>YERCty1-1</td>
<td>5</td>
</tr>
<tr>
<td>YERCty1-2</td>
<td>5</td>
</tr>
<tr>
<td>YGRCty1-2</td>
<td>7</td>
</tr>
<tr>
<td>YGRCty1-3</td>
<td>7</td>
</tr>
<tr>
<td>YGRWTy1-1</td>
<td>7</td>
</tr>
<tr>
<td>YHRCTy1-1</td>
<td>8</td>
</tr>
<tr>
<td>YJRWty1-1</td>
<td>10</td>
</tr>
<tr>
<td>YJRWty1-2</td>
<td>10</td>
</tr>
<tr>
<td>YLRCty1-1</td>
<td>12</td>
</tr>
<tr>
<td>YLRCty1-2</td>
<td>12</td>
</tr>
<tr>
<td>YLRWTy1-3</td>
<td>12</td>
</tr>
<tr>
<td>YMLWTy1-1</td>
<td>13</td>
</tr>
<tr>
<td>YMLWTy1-2</td>
<td>13</td>
</tr>
<tr>
<td>YMRCTy1-3</td>
<td>13</td>
</tr>
<tr>
<td>YMRCTy1-4</td>
<td>13</td>
</tr>
<tr>
<td>YNLCty1-1</td>
<td>14</td>
</tr>
<tr>
<td>YNLWTy1-2</td>
<td>14</td>
</tr>
<tr>
<td>YOLWTy1-1</td>
<td>15</td>
</tr>
<tr>
<td>YORWTy1-2</td>
<td>15</td>
</tr>
<tr>
<td>YPLWTy1-1</td>
<td>16</td>
</tr>
<tr>
<td>YPRCTy1-2</td>
<td>16</td>
</tr>
<tr>
<td>YPRCTy1-4</td>
<td>16</td>
</tr>
<tr>
<td>YPRWTy1-3</td>
<td>16</td>
</tr>
</tbody>
</table>
1.5 Role of chromosomal inversions and translocations in gene order evolution, speciation and fitness

Species are defined as groups of entities living in nature separated by genetic and phenotypic differences, capable of interbreeding and producing fertile offspring. The formation of new and distant species is referred as speciation. Speciation occurs by different mechanisms. Geographic isolation, chromosomal rearrangements and sequence divergence are the processes involved in speciation [68].

It has been proposed that reciprocal chromosomal translocations in yeasts can reinforce reproductive isolation. There are six species included in *Saccharomyces sensu stricto* group i.e. *S. cerevisiae*, *S. paradoxus*, *S. bayanus*, *S. cariocanus*, *S. mikatae* and *S. kudriavzevii*. Three out of six species are known to possess specific reciprocal translocations. *S. bayanus* possess three reciprocal and one non-reciprocal translocation, *S. cariocanus* has four reciprocal translocations, one in *S. mikatae* isolate IFO1815 and two in IFO1816 [68]. Chromosomal rearrangements between the species will prevent the transfer of a complete haploid set of genes in the gametes produced by F1 hybrid meiosis [69]. The sterility of the hybrids formed between *S. paradoxus* and *S. cariocanus* can be due to chromosomal rearrangement as the two species differ from each other by the presence of four reciprocal translocation. However, chromosomal rearrangement is not always the cause of F1 hybrid sterility in yeast as the other yeast species differ by fewer or no chromosomal translocations but produce sterile hybrids. This phenomenon was further supported by Delneri *et al.* who engineered reciprocal translocations in the genome of *S. cerevisiae* making it collinear with *S. maikatae*. Crosses between wild type *S. cerevisiae* and engineered *S. cerevisiae* showed a 25-50% reduction in spore viability compared to a cross between engineered *S. cerevisiae* with its collinear *S. mikatae* produced no fertile hybrids indicating that there is something other than chromosomal rearrangement causing the hybrid sterility.

Chromosomal inversions are also found to affect the rates of speciation, adaptation and evolution of sex chromosomes. Human and its closest relative, the chimpanzee diverged
approximately 6 million years ago. A comparison between the genome of humans and chimps revealed that the DNA sequences of the two species are 98% identical [70]. They differ from each other by nine pericentric inversions affecting chromosomes 1, 4, 5, 9, 12, 15, 16, 17 and 18 which permanently separated the two lineages [71].

Recombination between inverted and non-inverted chromosomes may result in the generation of unbalanced gametes carrying chromosomal duplications or deficiencies. These unbalanced zygotes may not be viable, thereby generating a barrier between the species that differ for the rearrangement. Another property of inversions is reduced recombination between inverted and non-inverted regions due to reduced crossing-over of inverted regions among the recombinant gametes [72]. Both types of inversions, paracentric as well as pericentric, are found to play a role in speciation; however, pericentric inversions are less common than paracentric inversions. Inversions may result in speciation by the generation of deleterious meiotic products producing inviable gametes. According to another view, inversions may lead to speciation by reducing recombination and thus protecting the genomic region from introgression [73].

1.6 Co-expressed gene clusters

Genes in prokaryotes are found to be present in clusters whereas the clustering of genes is less common among the eukaryotes [74]. Gene clustering and co-transcription of genes in operons is common in prokaryotes but in eukaryotic genomes it has been found to a lesser extent with the exception of nematodes and trypanosomes [75]. Although most eukaryotic genomes lack operons, they contain some physical gene clusters that are related in function and not in sequence [76]. One example of this was found in oat plants where at least four genes involved in species-specific antimicrobial defence pathway show complete cosegregation [77]. Similarly co-expressed genes are found to be clustered in eukaryotes. Caron and co-workers found that highly expressed genes are co-localized in humans [78]. About 20% of Drosophila genes lie in clusters of co-expressed genes [79]. Kruglyak and Tang found a very high correlation of expression patterns for adjacent genes in yeast. Co-regulated genes in yeast and other higher eukaryotes are kept in the same
chromosomal region to make them available for transcription more efficiently as a group [80]. These co-expressed neighbouring genes are therefore found not to be rearranged during evolution [81]. Co-localization of genes encoding proteins in the same metabolic pathway [82] and significant physical clustering of essential genes [83] has been reported in *S. cerevisiae*.

Hall and Dietrich reported that in *S. cerevisiae* genes of related function are not usually clustered and they identified only 14 pairs of adjacent, functionally related genes [34]. There are two large groups of gene clusters in *S. cerevisiae*, the *GAL* cluster [84] and the *DAL* cluster [85]. A lot of research has been carried out on the *GAL* gene cluster. The *GAL* genes (*GAL1, GAL10* and *GAL7*) are involved in galactose assimilation and are found to be clustered in the genome of almost all the yeast species in which they are present. This gene cluster has been lost in three yeast species (*C. glabrata, A. gossypii* and *K. waltii*). It is not yet not clear what ecological changes permitted this loss in *A. gossypii* and *K. waltii* but is understood in *C. glabrata*. *C. glabrata* is a mammalian pathogen and therefore does not encounter galactose in this environment [84].

1.6.1 Structure of the *DAL* cluster

The *DAL* cluster is the largest metabolic gene cluster in yeast which seems to have been formed recently through a set of simultaneously occurring genomic rearrangements [86]. The *DAL* cluster in *S. cerevisiae* consists of six adjacent genes encoding for proteins involved in allantoin degradation [85]. Six out of eight genes required for allantoin degradation were initially scattered throughout the yeast genome then over the time of evolution they clustered at a single subtelomeric site in an ancestor of *S. cerevisiae* and *S. castellii*. The *DAL* gene cluster is completely conserved in other *Saccharomyces sensu stricto* species and in *S. castellii*. There are no *DAL* clusters in more distantly related hemiascomycetes species. Other hemiascomycetes contain homologs of the six *DAL* genes scattered at six separate chromosomal locations. Species that have the *DAL* cluster form a monophyletic group and those which do not posses this cluster form paraphyletic group [27]. Two of the *DAL* genes, *DAL4* and *DAL7*, are duplicates genes located elsewhere in *S.
cerevisiae genome. The DAL4 gene encoding a allantoin permease is a duplicate of FUR4 gene encoding for uracil permease. Both DAL4 and FUR4 are members of purine transporter family. They possess almost identical lengths and approximately 70% sequence identity. The DAL7 gene in Saccharomyces cerevisiae, which encodes for malate synthase and is involved in allantoin degradation is a duplicate of the MLSI gene which also encodes malate synthase, but is involved in the glyoxylate cycle in utilization of non-fermentable carbon sources. The other four DAL genes DAL1, DAL2, DAL3 and DCG1, seem to have transposed to their new location in the cluster site. These four genes are single copy genes and are thought to be produced by gene duplications followed by the deletion of the original gene. These six genes are located on chromosome IX of S. cerevisiae in a 9.4kb region near the right telomere. It is thought that random rearrangements moved the DAL genes in a cluster at the sub-telomeric location which was then strongly favoured by the natural selection. The genomic reorganization of the DAL genes coincided with the reorganization of the purine degradation pathway. Allantoin degradation pathway has also shown a phylogenetic correlation between the presence of a DAL cluster and loss of the urate oxidase (UOX) and urate permease (UAP) genes [86].

1.6.2 Function of DAL cluster in allantoin degradation

The DAL cluster enables yeast to use allantoin, which is a degradation product of purines and is used as a non-preferred nitrogen source. Allantoin, the first product of purine (adenine and guanine) catabolism in many organisms, serves as a sole source of nitrogen for S. cerevisiae. Allantoin degradation involves five enzymatic steps and four transport systems by which it is subsequently converted to a mixture of four molecules of ammonia, two molecules of CO₂ and one molecule of glyoxylate (Figure 1.4). In the first step, the hydantoin ring of allantoin is opened via hydrolytic reaction catalyzed by allantoinase, the product of DAL1, resulting in the production of allantoate. Allantoate is then degraded by allantoicase, the product of DAL2, yielding one molecule of urea and ureidoglycolate. Ureidoglycollate is then hydrolyzed by ureidoglycolate hydrolase, the product of DAL3, to glyoxylate and a second molecule of urea. In most of the organisms urea is further degraded to ammonia by urease but S. cerevisiae does not contain urease activity. Thus in
*S. cerevisiae* the urea generated in first two steps undergoes a reaction with CO$_2$ in the presence of biotin and ATP resulting in the formation of allophanate, which in turn is converted to ammonia and CO$_2$ [87]. These two reactions are catalyzed by a bi-functional urea carboxylase-allophanate hydrolase, the product of a single gene, *DUR1*, 2. Urea carboxylase is a biotin requiring enzyme whose activity is separated in two steps, first the activation of CO$_2$ and then the subsequent transfer of CO$_2$ to urea. In addition to these enzymes required for allantoin degradation, the yeast cell is provided with a permease for the transport of allantoate, the product of *DAL5*, and a high-affinity permease for urea, the product of *DUR3* (Wong and Wolf, 2005).
Figure 1.4: The allantoin degradation pathway. *Saccharomyces cerevisiae* imports both allantoin and allantoate. Allantoin is converted to allantoate by allantoinase (product of *DAL1*) which is then degraded to produce ureidoglycolate and one molecule of urea. In the final stage of the pathway ureidoglycolate is converted to malate and urea is degraded to ammonia. Allantoin exists as two stereoisomers (R and S). Other yeasts import urate using urate permease, *UAP* and urate oxidase, *UOX* which is shown in grey. Figure based on [86].

Yeast species that do not possess a *DAL* cluster (*Yarrowia lipolytica*, *Candida albicans*, *Kluyveromyces waltii* and *Zygosaccharomyces rouxii*) obtain their allantoin by importing it using urate oxidase (*UOX*) whereas species that have the *DAL* cluster import allantoin from outside the cell using allantoin permease, the product of *DAL4*, which is a duplicated copy of *FUR4*. The subsequent degradation steps involve the same *DAL* pathway in all *Saccharomyces* species, but in *S. cerevisiae* and *S. castellii*, genes have been reorganized into a cluster and *MLS1* has been duplicated to produce *DAL7*. This biochemical reorganization was the result of selection to minimize oxygen consumption during the
evolution that resulted to bypass *UOX*, as the reaction catalyzed by *UOX* requires molecular oxygen as a substrate. *S. cerevisiae*, other *sensu stricto* species and *S. castellii* can grow in anaerobic conditions by fermentation whereas other species except for *S. kluyveri* are strict aerobes [32].

1.7  **Aim of the Project**

The overall aim of this project was to study the role of chromosomal rearrangements on fitness adaptation and local and global gene expression.

The first question we asked was whether there are any phenotypic and transcriptional effects of small size (single, double and triple genes) inversions within blocks of co-expressed genes in the *DAL* cluster. The species belonging to the *S. cerevisiae* “*sensu stricto*” group compared to the yeast *S. castellii* present different orientations of the *DAL* gene cluster and my aim was to understand whether these differences have an evolutionary adaptive origin.

The second aim was to study the impact of large size pericentric and paracentric inversions on phenotypic fitness, spore survival and global gene expression profile. This analysis was performed on strains of *Saccharomyces cerevisiae* in which sections of chromosomes between repetitive sequences of TY elements were inverted.

Our third aim was to understand the role that translocations have in the speciation process between two ‘*sensu stricto*’ species, *S. paradoxus* and *S. cariocanus* that differ from each other by the presence of four chromosomal translocations. Genomic data show that *S. cariocanus* and *S. paradoxus* possess no significant sequence divergence. This suggests that these yeasts belong to the same species and the reproductive isolation between them is not due to the lack of recombination in meiosis I but due to translocations. I intended to study whether the reproductive barriers are caused by the translocations and whether the use of the biological species concept is appropriate in this case.
1.8 References


46.


CHAPTER 2 (Paper 1)

IMPACT OF CHROMOSOMAL INVERSIONS ON CO-EXPRESSED GENE CLUSTERS
Impact of chromosomal inversions on co-expressed gene clusters

Samina Naseeb & Daniela Delneri

To be submitted to Genome Biology journal.

2.1 Foreword

This chapter is laid out in the format of Genome Biology journal. SN performed the experimental work, analysed the results and wrote the manuscript, DD designed and supervised the project, and reviewed the manuscript.

2.2 Abstract

Background: Chromosomal rearrangements such as deletions, translocations, duplications and inversions occur readily in nature during genome evolution. They can be adaptive in nature as well as deleterious causing several different genetic diseases by altering the gene expression or gene copy numbers. One major question, which has not been fully answered, is how do chromosomal inversions effect fitness adaptation and gene expression especially when located within a co-expressed gene cluster.

Results: Here we studied the effects of single, double, or triple gene inversions in of DAL metabolic cluster. We showed that DAL2 inversion has an impact both on the phenotype as well as expression of the neighbouring genes (DAL1 and DAL4). We also demonstrated that change in expression of DAL4 were due to the inversion of DAL2 and it is not related to the amount of anti-sense transcript of DAL4 (SUT614).

Conclusions: This finding implies that expression of genes in the DAL metabolic cluster depend on the order of some of its genes. A single gene inversion (of DAL2) may alter the expression of genes in its vicinity belonging to the DAL metabolic cluster as well as can alter the phenotype of the strain.
2.3 Background

Chromosomal rearrangements include several different events: deletions, duplications, translocations and inversions. These rearrangement events are quite common in yeast and range from a part of a gene to hundreds of genes resulting in different rates of genome evolution. In this study, we focused on chromosomal inversions. Chromosomal inversions have been found in almost all organisms ranging from prokaryotes to eukaryotes and its rate vary among different lineages. They have been identified in grasshoppers [1], rodents [2], fruitflies [3], marine snails and plants [4]. Moreover, chromosomal inversions have also been found in humans. Stefansson and co-workers in 2005 reported a 900-kb inversion at chromosomal location 17q21.31 in humans [5]. Comparison of the human and chimpanzee genomes elucidated 66 inversions of more than 25 kb in length [6]. Chromosomal inversions in yeast are predominantly small sometimes including only one gene and can generate a new gene order putting the relevant loci in proximity. One example of chromosomal inversions in yeast reinforced the separation of Candida albican species from Saccharomyces cerevisiae. Seoigh and co-workers reported that 1,100 single-gene inversions have occurred since the divergence of two species [7].

It is important to characterize inversions as they can affect the gene expression adjacent to the breakpoints [8]. To study the role of inversions on fitness and gene expression in yeast the approach taken was to study the role of gene inversions on co-expressed gene clusters. It has been shown that expression of a gene is dependent on the location of gene both in prokaryotes as well as in eukaryotes ranging from yeast to filamentous fungi, plants as well as in animals. In prokaryotes clusters of genes form operons leading to a strong co-expression and in eukaryotes physically linked genes are co-expressed [9-11]. Co-expression of genes in prokaryotes is relatively simple however in eukaryotes it is dependent on several mechanisms such as chromatin structure, bidirectionally active
promoters, genomic neighbourhood, transcription factors and other regulatory units [12, 13].

There are two groups of gene clusters in *S. cerevisiae*, the *GAL* cluster [14] and the *DAL* cluster [15]. The *DAL* cluster is the largest metabolic gene cluster in yeast which seems to have been formed through a set of simultaneously occurring genomic rearrangements during evolution [16]. This cluster has been of interest in this piece of research.

The *DAL* cluster in *S. cerevisiae* consists of six adjacent genes encoding for the proteins involved in allantoin degradation. This group of genes enables yeast to use allantoin, which is a degradation product of purines and is used as a non-preferred nitrogen source. Allantoin is converted through a series of steps into ammonia which is a simpler form of nitrogen readily used by yeast (Figure 2.1). A total of eight genes are required for allantoin degradation. *DAL5* (located on chromosome X and encodes for allantoate permease) and *DUR1, 2* (located on chromosome II and encodes for a protein which is responsible for converting urea to ammonia) are the two genes which are not located in the *DAL* cluster. Six out of eight genes required for allantoin degradation were initially scattered throughout the yeast genome then during the period of evolution they clustered at a single subtelomeric site of chromosome IX in an ancestor of *S. cerevisiae* and *S. castellii*. The *DAL* gene cluster is completely conserved in *Saccharomyces* “sensu stricto” species and in *S. castellii*. There are no *DAL* clusters in more distantly related hemiascomycetes species [17].
Figure 2.1: Gene organization of *S. cerevisiae* DAL cluster involved in Allantoin degradation and allantoin degradation pathway. Panel A shows the allantoin degradation pathway. Allantoin is converted to allantoate by allantoinase (product of DAL1p) which is then degraded to produce ureidoglycolate and then glyoxalate and urea. In the final stage of the pathway glyoxalate is converted to malate and urea is degraded to ammonia. Panel B shows the organization of DAL genes located on chromosome IX in *S. cerevisiae* as well as the transcription regulators of these genes. The product of DAL-encoded gene is shown below each gene. The green arrows indicate the direction of transcription of genes. The positive regulators of genes are shown in blue indicated with + sign and the negative regulators are shown in orange (labelled with −).

The order of DAL genes is the same in *S. cerevisiae* and other sensu stricto species, however a rearrangement in gene order occurred in *S. castellii* relative to *S. cerevisiae* and other sensu stricto species during evolution. An inversion in DAL1, DAL2 and DAL4 genes occurred, DAL2 was then re-inverted such that its orientation was kept the same as in *S. cerevisiae* and other sensu stricto species (Figure 2.2).
Figure 2.2: Comparison of DAL genes between *Saccharomyces cerevisiae*, the *sensu stricto* species and *Saccharomyces castellii*. The six DAL genes are located on chromosome IX at the same position and orientation in *S. cerevisiae* and the rest of the *sensu stricto* species. However, DAL1, DAL2 and DAL4 genes in *S. castellii* were inverted during evolution, whereas other DAL genes remained in their positions. The orange arrows indicate the genes which have not been inverted, the green arrows show genes having single inversion and blue arrow is for gene with double inversion.

In this study, our aim was to understand the phenotypic and transcriptional effects of inversions within the DAL cluster. The species belonging to the *Saccharomyces “sensu stricto”* group compared to the *S. castellii* present different orientation of the genes within this cluster and we intended to explore whether this difference has an evolutionary adaptive origin. Therefore, we decided to construct various inverted and non-inverted strains possessing single, double or triple gene inversions as well as a strain resembling the *S. castellii* DAL structure in *S. cerevisiae* background.

We analyzed the fitness of the wild type, inverted and non-inverted strains in allantoin containing medium that is responsible for triggering the expression of the DAL genes. We showed that the strain with DAL2 inversion possessed a lower growth rate as compared to the non-inverted and the wild type strains. Our data on gene expression indicates that the DAL4 gene is down regulated in the DAL2 inverted strains. Therefore, we also studied the
expression of both *DAL4* sense and anti-sense transcripts in YPD as well as in allantoin and proline containing medium. We showed that *DAL4* antisense level does not change upon *DAL2* gene inversion.

### 2.4 RESULTS

#### 2.4.1 Transcription factor analysis of *DAL* cluster structure

Prior to the construction of inverted and non-inverted strains we looked for the transcription factor (TF) binding sites lying within the *DAL* cluster of *S.cerevisiae* and *S.castellii* (Table 2.1). The binding sequences of TFs were obtained from Harbinson *et al.* (2004) [18].

<table>
<thead>
<tr>
<th>Factor</th>
<th>Transcriptional Effect</th>
<th>Binding Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>GAT1</em></td>
<td>Activator</td>
<td>AGATAAAG/GATAAA</td>
</tr>
<tr>
<td><em>GLN3</em></td>
<td>Activator</td>
<td>GATAAGA</td>
</tr>
<tr>
<td><em>DAL82</em></td>
<td>Activator</td>
<td>GATAAG</td>
</tr>
<tr>
<td><em>HAP2</em></td>
<td>Activator</td>
<td>CCAAT</td>
</tr>
<tr>
<td><em>DAL80</em></td>
<td>Repressor</td>
<td>GATAA</td>
</tr>
<tr>
<td><em>SUM1</em></td>
<td>Repressor</td>
<td>AGY<em>GW</em>CACAAAAA</td>
</tr>
</tbody>
</table>

From an evolutionary perspective, the level to which promoter elements are conserved is important. To investigate how much conservation there is between the *DAL* cluster in *S. cerevisiae* and *S. castellii* we compared the sizes of individual genes, intergenic region and the entire cluster. We found that the distribution of intergenic regions between the two
species was highly varied but that the overall size of the cluster differed by only a single base pair (Figure 2.3).

**S. cerevisiae**

![Diagram of S. cerevisiae DAL cluster]

**S. castellii**

![Diagram of S. castellii DAL cluster]

**Figure 2.3: Size of the gene ORFs in DAL cluster.** A schematic representation of the sizes of genes, intergenic regions and whole cluster. The upper panel shows *S. cerevisiae* DAL cluster and the lower panel shows *S. castellii* DAL cluster. The genes on waston strand are coloured red and those on crick strand are coloured blue.

The transcription factor binding sites present within the intergenic regions of the *DAL* cluster in *S.cerevisiae* and *S.castellii* are shown in Table 2.2 & 2.3. It was observed that *HAP2* transcription factor is present upstream of *DAL1, DAL2, DAL3* and *DAL4* only in *S. castellii* and not in *S. cerevisiae*. *S. castellii* possess only two transcription factors binding upstream of *DAL4* and *DAL1* however, in *S. cerevisiae* several transcription factors are located upstream of *DAL4* and *DAL1*.
Table 2.2: Transcription factor binding sites for *S. cerevisiae* DAL cluster

<table>
<thead>
<tr>
<th>TF</th>
<th>Binding sites upstream of DAL4</th>
<th>TF</th>
<th>Binding sites upstream of DAL1</th>
<th>TF</th>
<th>Binding sites upstream of DAL2</th>
<th>TF</th>
<th>Binding sites upstream of DCG1</th>
<th>TF</th>
<th>Binding sites upstream of DAL7</th>
<th>TF</th>
<th>Binding sites upstream of DAL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUM1</td>
<td>-265 to 255</td>
<td>DAL82</td>
<td>-363 to 359</td>
<td>DAL82</td>
<td>-202 to 196</td>
<td>DAL82</td>
<td>-148 to 154</td>
<td>DAL82</td>
<td>-266 to 226</td>
<td>DAL82</td>
<td>-38 to -32</td>
</tr>
<tr>
<td>DAL82</td>
<td>-251 to 245</td>
<td>GAT1</td>
<td>-364 to 360</td>
<td>GAT1</td>
<td>-202 to 197</td>
<td>GAT1</td>
<td>-148 to 155</td>
<td>DAL82</td>
<td>-254 to 199</td>
<td>GAT1</td>
<td>-38 to -33</td>
</tr>
<tr>
<td>GAT1</td>
<td>-251 to 246</td>
<td>GLN3</td>
<td>-364 to 359</td>
<td>GLN3</td>
<td>-204 to 196</td>
<td>GLN3</td>
<td>-146 to 154</td>
<td>DAL82</td>
<td>-167 to 161</td>
<td>GLN3</td>
<td>-40 to -32</td>
</tr>
<tr>
<td>GLN3</td>
<td>-251 to 245</td>
<td>DAL82</td>
<td>-381 to 376</td>
<td>DAL82</td>
<td>-179 to 173</td>
<td>DAL82</td>
<td>-158 to 164</td>
<td>GLN3</td>
<td>-167 to 159</td>
<td>DAL82</td>
<td>-122 to 116</td>
</tr>
<tr>
<td>DAL82</td>
<td>-365 to 359</td>
<td>GAT1</td>
<td>-381 to 377</td>
<td>GAT1</td>
<td>-180 to 174</td>
<td>HAP2</td>
<td>-177 to 181</td>
<td>DAL82</td>
<td>-210 to 204</td>
<td>DAL82</td>
<td>-162 to 156</td>
</tr>
<tr>
<td>GAT1</td>
<td>-365 to 360</td>
<td>GLN3</td>
<td>-381 to 376</td>
<td>GLN3</td>
<td>-179 to 171</td>
<td>GAT1</td>
<td>-210 to 205</td>
<td>GAT1</td>
<td>-162 to 155</td>
<td>GLN3</td>
<td>-162 to 154</td>
</tr>
<tr>
<td>GLN3</td>
<td>-365 to 359</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table representing the transcription factors (TF) and their binding sites at the intergenic region of *DAL4, DAL1, DAL2, DCG1, DAL7* and *DAL3* in *S. cerevisiae*. 
Table 2.3: Transcription factor binding sites for *S. castellii* DAL cluster

<table>
<thead>
<tr>
<th>TF</th>
<th>Binding sites upstream of DAL2</th>
<th>TF</th>
<th>Binding sites upstream of DAL4</th>
<th>TF</th>
<th>Binding sites upstream of DAL1</th>
<th>TF</th>
<th>Binding sites upstream of DCg1</th>
<th>TF</th>
<th>Binding sites upstream of DAL7</th>
<th>TF</th>
<th>Binding sites upstream of DAL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAL82</td>
<td>-120 to -115</td>
<td>GAT1</td>
<td>-46 to -42</td>
<td>HAP2</td>
<td>-20 to -16</td>
<td>DAL82</td>
<td>-126 to -120</td>
<td>GAT1</td>
<td>-291 to -286</td>
<td>GAT1</td>
<td>-188 to -184</td>
</tr>
<tr>
<td>GAT1</td>
<td>-120 to -116</td>
<td>HAP2</td>
<td>-307 to -303</td>
<td>DAL82</td>
<td>-238 to -233</td>
<td>GAT1</td>
<td>-125 to -120</td>
<td>GAT1</td>
<td>-291 to -287</td>
<td>HAP2</td>
<td>-206 to -202</td>
</tr>
<tr>
<td>GLN3</td>
<td>-120 to -115</td>
<td>GAT1</td>
<td>-333 to -329</td>
<td>GAT1</td>
<td>-239 to -233</td>
<td>GLN3</td>
<td>-126 to -120</td>
<td>GLN3</td>
<td>-291 to -286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAL82</td>
<td>-140 to -134</td>
<td>GLN3</td>
<td>-238 to -233</td>
<td>GAT1</td>
<td>-159 to -155</td>
<td>DAL82</td>
<td>-537 to -531</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAT1</td>
<td>-140 to -135</td>
<td>DAL82</td>
<td>-268 to -263</td>
<td>GAT1</td>
<td>-536 to -531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLN3</td>
<td>-140 to -134</td>
<td>GAT1</td>
<td>-267 to -263</td>
<td>GLN3</td>
<td>-537 to -531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAP2</td>
<td>-222 to -218</td>
<td>GLN3</td>
<td>-268 to -263</td>
<td>HAP2</td>
<td>-420 to -416</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table representing the transcription factors (TF) and their binding sites at the intergenic region of *DAL2, DAL4, DAL1, DCg1, DAL7* and *DAL3* in *S. castellii*.  

58
2.4.2 Fitness assays of the wild type strains

The DAL cluster in *Saccharomyces castellii* differs from *Saccharomyces cerevisiae* by inversion of three DAL genes (Fig 2.2). Firstly, the growth rate of the wild type *S. castellii* and wild type *S. cerevisiae* (Fy3) in rich medium (YPD), F1 medium containing nicotinic acid, and allantoin + nicotinic acid containing medium was assigned. Nicotinic acid was added as a supplement to enable the growth of *S. castellii* in the F1 medium. We tested *S. castellii* growth in different “drop-out” medium and nicotinic acid was found to be promoting its growth.

Before using allantoin in our medium, we tested different concentrations of the allantoin on the growth rate of our wild type strain. It was found that high concentration (0.4% (w/v)) is deleterious to the strain and therefore a low concentration of (0.0125% (w/v)) was used in the medium for further experiments (Figure 2.4).

![Figure 2.4: Fitness assay for optimum allantoin concentration.](image)

**Figure 2.4: Fitness assay for optimum allantoin concentration.** Growth rate of a wild type *S. cerevisiae* strain, FY3, was measured in different concentrations of allantoin containing media using the Optima plate reader over the course of 40 hours. Higher concentration of allantoin 0.4% (w/v), shown in blue, was found to be toxic to the cells, therefore, an optimum concentration of 0.0125% (w/v) of allantoin shown in red was used for our fitness assays. Each point represents the mean average of 5 technical repeats. Error bars are plotted at +/-1 standard deviation.
The wild type *S. castellii* was growing much slower than *S. cerevisiae* in both allantoin and F1 medium (Figure 2.5a & 2.5b).

**Figure 2.5: Fitness assay of *S. castellii* WT and FY3 strains in 0.0125% allantoin medium.** Panel A the growth profiles of both strains in F1 medium. Panel B growth profiles of both strains in allantoin medium. Each point represents the mean average of three technical repeats for five independent biological replicas. Error bars are at 95% confidence intervals.

It was also observed that *S. castellii* grows slower than both the laboratory as well as natural isolates of *S. cerevisiae* in F1 and allantoin containing medium, however no difference in growth was observed in YPD medium (Figure 2.6)
Figure 2.6: Fitness assay of WT *S. castellii* and *S. cerevisiae* lab isolate (FY3) and natural isolates (T73, 9 and 96.2) in different media. Panel A the growth profiles of all strains in YPD medium were same. Panel B & C growth profiles in 0.0125% allantoin and F1 medium. *S. castellii* is showing lower growth rate in both media compared to the laboratory and natural isolates of *S. cerevisiae*. Each point represents the mean average of five technical repeats for each strain. Error bars are at 95% confidence intervals.
These results indicated that *S. cerevisiae* can utilize allantoin in much more growth efficient way than *S. castellii* wild type strain. We therefore used *S. cerevisiae* as a background strain to construct our inverted and non-inverted strains.

### 2.4.3 Construction of the strains with different gene inversions

To study the impact of gene inversion on gene expression, we constructed seven inverted and non-inverted strains using the cre-*lox*P system which are shown in Figure 2.18 and 2.19 [19]. Wild type *S. cerevisiae* (FY3) was the background strain used to construct all inverted and non-inverted strains. The steps included were the extraction and digestion of the plasmid containing cassette, amplification and transformation of the desired cassette, transformation and induction of the cre encoding gene followed by the selection of the inverted and non-inverted strains.

### 2.4.4 Plasmid extraction and digestion

The plasmids used in this study are listed in the appendix Table A1. The plasmid DNA was prepared and extracted using the QIAgen DNA extraction kit and validated by digesting them using the restriction enzyme *Eco*RV, *Eco*RI or *Bgl*II depending upon the restriction site giving single cut. The digested and un-digested DNA was then visualised under UV light immediately after performing gel electrophoresis (Figure 2.7).
Figure 2.7: The pZC1, pUG6 and pZC4 plasmid digest. Panel A 0.8 % (w/v) agarose gel showing the plasmid pZC1 (4235bp) and plasmid pUG6 (4009bp) containing the *loxP-hphNT1-loxP* and *loxP-kanMX-loxP* cassettes, respectively. The plasmids were linearised by digesting with restriction endonuclease *EcoRV*. Panel B agarose gel (0.8% (w/v)) showing the plasmid pZC4 (3827bp) containing cassette *lox2272-natNT2-lox2272* which was digested with *BglII*.

### 2.4.5 Amplification of specific cassettes

The cassettes used in this study *loxP-kanMX-loxP*, *loxP-hphNT1-loxP* and *lox2272-natNT2-lox2272* were amplified by PCR using the cassette specific amplifying primers listed in appendix Table A8. These primers contained approximately 45bp flanking sequence of both ends homologous to the flanking region of interest within the *DAL* cluster genes. The correct size of the PCR amplified product was confirmed on 0.8 % (w/v) agarose gel by gel electrophoresis (Figure 2.8).
Figure 2.8: The amplification of the cassettes. The exact size of all amplified cassettes was confirmed by gel electrophoresis. Panel A the PCR product of loxP-hphNT1-loxP (1826bp) Panel B the PCR product of loxP-kanMX-loxP (1613bp) and Panel C the PCR product of lox2272-natNT2-lox2272 cassette (1539bp).

2.4.6 Construction of the DAL2 inverted strain

2.4.6.1 Insertion of the loxP containing cassettes

The amplified product of the gene specific cassette (approximately 0.1-1 µg) was transformed in *Saccharomyces cerevisiae* (FY3) strain by lithium acetate transformation protocol. The *loxP-kanMX-loxP* cassette was inserted immediately upstream of the promoter region of *DAL2* and *loxP-hphNT1-loxP* cassette was inserted just downstream of the terminator of *DAL2* to invert the *DAL2* gene. Cassettes were inserted at their specific loci by homologous recombination as shown in Figure 2.9.

The control strains were also generated which had a single cassette either upstream of the promoter or downstream of the terminator region of wild type FY3 strain. The
Transformants were selected on YPD plates containing either geneticin G418 or hygromycin B antibiotics depending upon the resistance marker on the transformed cassette.

**Figure 2.9: Cassettes insertion by homologous recombination.** The amplified loxP-kanMX-loxP cassette was inserted by homologous recombination in FY3 immediately upstream of DAL2 promoter and downstream of DAL4 terminator. Green block arrows indicate the direction of gene and small triangular arrows are for loxP sequences. All the other cassettes were inserted following a similar strategy at their respective positions.

### 2.4.6.2 Verification of cassette insertion

Detection of the correct insertion of the cassettes in *S. cerevisiae* genome was done by diagnostic colony PCR using the checking primers listed in the appendix Table A2. Four primers were designed for each colony PCR assay. The forward and reverse primers were specific to the 5’ and 3’ regions of the cassettes whereas the forward and reverse primers were designed to be gene specific and anneal to the upstream and downstream coding region of the gene. The PCR products of amplified cassette were verified by gel electrophoresis in order to confirm the predicted size of the product if integration had taken place (Figure 2.10).
Figure 2.10: The colony PCR for confirmation of cassette insertion in transformant colonies selected on YPD + geneticin and YPD + hygromycin B. All the gels shown in this figure were made as 0.8% (w/v) agarose, lane L in each gel is for Hyperladder I (Bioline) and lane NC is for the negative control of respective primer pairs. Panel A shows the location of $\text{loxP-kanMX-loxP}$ and $\text{loxP-hphNT1-loxP}$ insertion upstream and downstream of $\text{DAL2}$ along with the cassette and gene specific primer binding regions. The block green coloured arrows indicate the direction of the genes, the small arrows indicate the primer binding sites and triangles show the $\text{loxP}$ sequences. Panel B lanes a-e are the transformant colonies confirmed by primer pair $\text{DAL4F} + \text{P3K4-up}$ giving a product of band size 830 bp, lanes f-j are the transformant colonies confirmed by primer pair $\text{DAL2R} + \text{P4K4-down}$ giving a product of band size 1040 bp. Panel C lanes a-d are the transformant colonies confirmed by primer pair $\text{DAL2FC} + \text{hph-down}$ giving a product of band size 800 bp, lanes e-g are the transformant colonies confirmed by primer pair $\text{DCG1 RC} + \text{hph-up}$ giving a product of band size 998 bp.
2.4.6.3  Cre transformation and induction

The cassette removal and gene inversion was performed by transformation and induction of cre gene containing plasmid. Colonies that lost both marker cassettes (Figure 2.11) were selected and confirmed by colony PCR. The colonies that have lost resistance to geneticin and hygromycin B were tested for gene non-inversion (Figure 2.12) and inversion (Figure 2.13) by colony PCR and the size of PCR products estimated by 0.8% (w/v) agarose gel electrophoresis.

Figure 2.11: Selection of colonies after the pop-out of cassettes. The strains harbouring the marker cassettes were transformed and induced with cre-recombinase containing plasmid. After cre induction the colonies that have lost both cassettes i.e. loxP-<i>kanMX</i>-loxP and loxP-<i>hphNT1</i>-loxP were selected. Panel A shows the growth of all colonies on YPD medium. Panel B shows the selection for the loss of loxP-<i>kanMX</i>-loxP cassette after cre induction. Panel C shows the selection for the loss of loxP-<i>hphNT1</i>-loxP cassette after cre induction. The blue boxes indicate few colonies that are growing on YPD but are unable to grow on YPD + geneticin and YPD + hygromycin B because they have lost both marker cassettes.
Figure 2.12: Colony PCR of DAL2 non-inverted strains. The non-inverted strains were confirmed by diagnostic colony PCR using the gene specific checking primers and product size matched to the predicted size. Panel A shows the direction of non-inverted DAL genes (green arrows), the loxP scars (blue triangles) and the gene specific primer binding regions (black arrows). Panel B shows 0.8% (w/v) agarose gel that represents as follows: Lane L is for Hyperladder I (Bioline), lanes 1-2 are the transformant colonies confirmed by primer pair DAL4F + DAL2R giving a product of band size 1.1kb and lanes 3-4 are the transformant colonies confirmed by primer pair DAL2FC + DCG1RC giving a product of band size 1.1kb
Figure 2.13: Colony PCR of DAL2 inverted strains. The DAL2 inverted strains were confirmed by diagnostic colony PCR using the gene specific checking primers and product size matched to the predicted size. Panel A shows the direction of non-inverted DAL genes (green arrows), the inverted DAL2 (red arrow), the loxP scars (blue triangles) and the gene specific primer binding regions (black arrows). Panel B shows 0.8% (w/v) agarose gel that presents as follows: Lane L is for Hyperladder I (Bioline), lanes 1-5 are the transformant colonies confirmed by primer pair DAL24FC + DCG1_RC giving a product of band size 995 bp and lanes 6-10 are the transformant colonies confirmed by primer pair DAL4F + DAL2-R giving a product of band size 1.1k b. NC in each case represents the negative control of their respective primers.

2.4.7 Construction of the strain with S. castellii like DAL cluster (S.cer-Inv1 and S.cer-Inv1-control)

This strain was made by using the DAL2 inverted strain having loxP scars upstream of the promoter region and downstream the terminator region of DAL2.
2.4.7.1 Insertion of cassettes

The \textit{lox2272-natNT2-\textit{lox}2272} cassette was inserted upstream of the promoter region of \textit{DAL2} in \textit{DAL2} inverted strain to replace the \textit{loxP} scar (Figure 2.14).

\begin{figure}[ht]
\centering
\includegraphics[width=\textwidth]{figure2.14.png}
\caption{The \textit{lox2272-natNT2-\textit{lox}2272} cassette insertion by homologous recombination.} The amplified \textit{lox2272-natNT2-\textit{lox}2272} cassette was inserted by homologous recombination in \textit{DAL2} inverted strain immediately upstream of \textit{DAL2} promoter to replace the \textit{loxP} sequence. Green block arrows indicate the original direction of genes, red block arrows show the inverted genes, small blue triangular arrows are for \textit{loxP} sequences and small pink triangles are for \textit{lox}2272 sequence.
\end{figure}

The second \textit{lox2272-natNT2-\textit{lox}2272} cassette was inserted downstream of the terminator of \textit{DAL1} but in the opposite direction to the first cassette to invert the \textit{DAL1}, \textit{DAL2} and \textit{DAL4} genes. The first cassette was popped-out before inserting the second \textit{lox2272-natNT2-\textit{lox}2272} cassette to enable easier selection of the transformants. Cassettes were inserted at their specific loci by homologous recombination.

The control strains were also generated which had single cassette i.e. \textit{lox2272-natNT2-\textit{lox}2272} at upstream of the \textit{DAL2} promoter and in other case \textit{lox2272-natNT2-\textit{lox}2272} at downstream of the \textit{DAL1} terminator region. The transformants were selected on neocothricin containing plates.
2.4.7.2 Verification of cassette insertion

Detection of the correct insertion of the cassettes in *S. cerevisiae* genome was done by diagnostic colony PCR using the checking primers listed in appendix Table A2. The amplified PCR products were checked by 1% (w/v) agarose gel electrophoresis to confirm the predicted size of the product i.e. 575bp and 1000bp (Figure 2.15).

Figure 2.15: Colony PCR for confirmation of *lox2272-natNT2-lox2272* cassette insertion in transformant colonies grown on YPD + neocotricin. Panel A shows the location of *lox2272-natNT2-lox2272* insertion at upstream of *DAL2* and downstream of *DAL1* along with the cassette and gene specific primer binding regions (black arrows). The block green coloured arrows indicate the direction of the non-inverted genes and red block arrow is for single inverted gene, and blue and pink triangles show the *loxP* and *lox2272* sites respectively. Panel B 1% (w/v) agarose gel that represents as follows: Lane L is for Hyperladder I (Bioline), lanes 1-7 are the transformant colonies confirmed by primer pair DAL1_FC + nat-down giving a product of band size 575 bp, lanes 8-15 are the transformant colonies confirmed by primer pair DCG1_RC + nat-up giving a product of band size 1000 bp and lane NC in each case is negative control of the respective primers.

2.4.7.3 Cre transformation and induction

The cassette removal and gene inversion was performed by transformation and induction of *cre* gene containing plasmid. The colonies that have lost resistance to drug selection, i.e. the selection marker, were tested for gene inversion by colony PCR and predicted size
confirmed by agarose gel electrophoresis (Figure 2.16). Similarly, the respective control strains were also confirmed by colony PCR as shown in Figure 2.17. In this case, the control strain was the one having DAL2 gene inverted only once while the experimental inverted strain was that one having DAL2 gene inverted twice.

A)

B)

Figure 2.16: Colony PCR of the strain with a S. castellii like DAL cluster (S.cer-Inv2). The strain S.cer-Inv2 was confirmed by diagnostic colony PCR using the gene specific checking primers and product size matched to the predicted size. Panel A shows the direction of single DAL genes (red arrows), the double inverted DAL2 (yellow arrow), the loxP scars (blue triangles), the lox2272 scars (pink triangles) and the gene specific primer binding regions (black arrows). Panel B shows 0.8% (w/v) agarose gel that presents as follows: Lane L is for Hyperladder I (Bioline), lanes 1-7 are the transformant colonies confirmed by primer pair DAL21_FC + DCG1_RC giving a product of band size 990 bp and NC represents the negative control of the respective primers.
Figure 2.17: Colony PCR of S.cer-Inv2-control strain. The strain S.cer-Inv2-control was confirmed by diagnostic colony PCR using the gene specific checking primers and product size matched to the predicted size. Panel A shows the direction of non-inverted DAL genes (green arrows), the single inverted DAL2 (red arrow), the loxP scars (blue triangles), the lox2272 scars (pink triangles) and the gene specific primer binding regions (black arrows). Panel B 0.8% (w/v) agarose gel that presents as follows: Lane L is for Hyperladder I (Bioline), lanes 1-7 are the transformant colonies confirmed by primer pair DAL21_RC + DAL1_FC giving a product of band size 771 bp and NC represents the negative control of the respective primers.

2.4.8 Summary of the strains constructed

In total, we constructed two strains with a single gene inversion (DAL2-Inv and DAL3-inv), a double gene inversion (DAL3-DAL7-Inv), two triple gene inversions (DAL1-DAL2-DAL4-Inv and DAL3-DAL7-DCG1-Inv) along with their respective control non-inverted strains i.e. DAL2-NI, DAL3-NI, DAL3-DAL7-NI, DAL1-DAL2-DAL4-NI and DAL3-DAL7-DCG1-NI which were collinear to the S. cerevisiae strain. These strains contain two loxP scars at the inversion breakpoints (Figure 2.18a, 2.18b). All the strains were constructed using the same strategy as described above.
Figure 2.18: DAL cluster structure of wild type, inverted strains and non-inverted strains. loxP-kanMX-loxP, and loxP-hphNT1-loxP cassettes were inserted in FY3 strain at the inversion breakpoints to construct the single, double and triple inverted strains using the cre-loxP recombination mechanism. Panel A structure of DAL cluster in the wild type and the inverted strains. Panel B structure of DAL cluster in non-inverted strains. The red blocks indicate the inverted genes, green blocks indicate the non-inverted genes and the blue triangles represent the loxP scars.

The DAL1-DAL2-DAL4-Inv and DAL2-Inv strains were used as parent strains to further construct two strains both having DAL structure similar to S. castellii DAL cluster (S.cer-Inv1 and S.cer-Inv2) along with the control strains which were identical to their parent strains i.e. DAL1-DAL2-DAL4-Inv and DAL2-Inv except for the two extra lox2272 scars (Scer-Inv1-control and Scer-Inv2-control) (Fig 2.19).
Figure 2.19: DAL cluster structure of engineered strains possessing S. castellii like DAL cluster. \textit{lox2272-natNT2-lox2272} cassettes were inserted in DAL1-DAL2-DAL4-Inv strain and DAL2-Inv strain at the inversion breakpoints to construct a final strain having DAL cluster similar to \textit{S. castellii} DAL cluster. Scer-Inv1 strain constructed using DAL1-DAL2-DAL4-Inv strain as a parent strain and Scer-Inv1-control which is similar to the parent strain except for the extra \textit{lox2272} scars. Scer-Inv2 strain constructed using DAL2-Inv strain as a parent strain and Scer-Inv2-control which is similar to the parent strain, however, carrying extra \textit{lox2272} scars. The red blocks indicate the single inverted genes, green blocks indicate the non-inverted genes, the double inverted genes are in yellow, the blue triangles represent the \textit{loxP} scars, the violet triangles are for the \textit{lox2272} scars and the orange boxes show that these two strains have similar DAL structure with difference in the position of \textit{loxP}s and \textit{lox2272}s.

2.4.9 Fitness assays of the engineered strains

The fitness of inverted and non-inverted strains was determined in rich medium (YPD), minimal medium containing allantoin as nitrogen source and F1 medium.

When comparing the FY3, DAL1-DAL2-DAL4-Inv, DAL1-DAL2-DAL4-NI strains in YPD, F1 and allantoin containing medium, no major difference was found in the growth rates (Figure 2.20). However, we found a significant drop in fitness of Scer-Inv1 in allantoin containing medium compared to other strains. It could be hypothesised that this decrease in growth rate is probably due to the inversions (Figure 2.20).
The next question to address was whether single inversion of \textit{DAL2} affects the fitness of our strain. It was observed that the \textit{DAL2} gene inversion affects the growth rate which drops significantly in comparison to the WT and DAL2-NI strains in allantoin containing medium. The same effect was observed in medium containing 0.1% proline + 0.0125% allantoin (Figure 2.21). The Scer-Inv2 strain showed to overcome reduced fitness affect and possessed significantly higher growth rate in comparison to the Scer-Inv2-control strain (Figure 2.21d). The growth rates of the inverted and non-inverted strains in N-limited medium were no different of inverted, non-inverted and WT strains (data not shown).
Figure 2.20: Fitness assays of the WT and engineered strains in 0.0125% (w/v) allantoin medium. Panel A represents the structure of DAL genes in strains tested for fitness. The red block arrow is for single inversion, and yellow for double inversion. Panel B the growth rate of DAL1-DAL2-DAL4-inv strain was found to be similar to the WT and non-inverted strain in YPD medium. Panel B growth rate of DAL1-DAL2-DAL4-inv strain was also observed to be similar to the WT and non-inverted strain in allantoin medium. Panel C growth rate of Scer-Inv1 (yellow line) was found to be lower than Scer-Inv1-control (blue line) as well as to the WT and parent strain (DAL1-DAL2-DAL4-Inv). Each point represents the average of three technical repeats in five independent biological repeats. Error bars are at 95% confidence intervals.
Figure 2.21: Fitness assays of the WT and engineered strains in different media. **Panel A** represents the structure of DAL genes in strains tested for fitness. The red block arrow is for single inversion, green for no inversion and yellow for double inversion. The growth rate of DAL2-inv strain was found to be lower than WT and non-inverted strain in 0.0125% (w/v) allantoin medium (**Panel B**), in 0.1% (w/v) proline + 0.0125% (w/v) allantoin medium (**Panel C**). **Panel D** the fitness of Scer-Inv2 and Scer-Inv2-control was observed to be higher in comparison to the WT and parent strain (DAL2-Inv) in 0.0125% allantoin medium. Each point represents the mean average of their technical repeats in five independent biological repeats. Error bars are at 95% confidence intervals.
2.4.10 Gene Expression Analysis of the Inverted and Non-Inverted Strains

To compare the expression level of the *DAL* genes between inverted and non-inverted strains, we performed real-time quantitative PCR. To investigate if the expression change is due to the gene inversion and *loxP* insertion, we also checked the expression of *DAL* genes in wild type (Fy3) and non-inverted strains. For each strain, cells were collected at exponential phase (OD$_{600nm}$=0.5), RNA was extracted and expression was analysed by quantitative real time PCR. For each sample, the housekeeping gene *ACT1* RNA was amplified and all data were normalized to *ACT1* (using the MJ Opticon Monitor3 analysis software provided by Bio-Rad). The expression pattern of the *DAL* genes can be affected either by gene inversion or by insertion of extra sequence, we therefore decided to study both parameters.

2.4.10.1 The impact of gene inversion on expression

To study the direct affect of gene inversion on expression we compared the inverted strains with the non-inverted control strains. In all the expression experiments, we applied Tukey’s HSD (Honestly Significant Difference) test to analyze the significant difference between the inverted and non-inverted candidates.

Firstly, we inverted *DAL1*, *DAL2* and *DAL4* genes and then checked the expression of each of these genes in the inverted strain (DAL1-DAL2-DAL4-Inv) compared to the non-inverted strain (DAL1-DAL2-DAL4-NI). We observed that *DAL4* and *DAL1* expression was relatively increased in the inverted strain, however, there was no significant change in the *DAL2* expression level (Figure 2.22).

We measured the expression of these genes in S.cer-Inv1 strain (in which *DAL2* was re-inverted to mimic *S.castellii* DAL cluster structure) and compared them to the S.cer-Inv1-control strain. Surprisingly, we didn’t observe any change in *DAL2* expression level nor of *DAL1* expression (Figure 2.23). However, *DAL4* expression was reduced by 50% in S.cer-Inv1 strain in comparison to S.cer-Inv1-control strain (Figure 2.23). This indicated that
DAL4 regulation is affected by disturbing DAL2 orientation.

Figure 2.22: Expression levels of DAL2, DAL4 and DAL1 in DAL1-DAL2-DAL4-Inv and DAL1-DAL2-DAL4-NI strains. Green bar represents the non-inverted strain and red bar represents the inverted strain. Error bars, SD, are from three technical repetitions for three independent biological samples. Relative normalized fold expression was calculated by using ΔΔCt method and ACT1 as a reference gene.
Figure 2.23: Expression levels of *DAL2*, *DAL4* and *DAL1* in Scer-Inv1 and Scer-Inv1-control strains. Blue bar represents the Scer-Inv1-control strain and yellow bar represents the Scer-Inv1 strain. Error bars, SD, are from three technical repetitions for three independent biological samples. Relative normalized fold expression was calculated by using ΔΔCt method and *ACT1* as a reference gene.
Our previous results showed that double inversion of \textit{DAL2} alters the expression of \textit{DAL4}. To confirm the effect of \textit{DAL2} inversion on \textit{DAL1} and \textit{DAL4}, we decided to measure the expression of a single inversion of \textit{DAL2} against that of \textit{DAL1}, \textit{DAL4} and \textit{DAL2}. We observed 50\% decline of \textit{DAL1} and \textit{DAL2} gene expression in the DAL2-inv strain (Figure 2.24) and \textit{DAL4} expression level was reduced to 90\% in the inverted strain (Figure 2.24c).

Figure 2.24: Expression levels of \textit{DAL2}, \textit{DAL4} and \textit{DAL1} in DAL2-Inv and DAL2-NI strains. Green bar represents the non-inverted strain and red bar represents the inverted strain. Error bars, SD, are from three technical repetitions for three independent biological samples. Relative normalized fold expression was calculated by using \textit{ΔΔCt} method and \textit{ACT1} as a reference gene.
2.4.10.2 Effect of \textit{loxP/lox2272} insertion on gene expression

Short sequence of \textit{loxP} and \textit{lox2272} consisting of 34bp were inserted at the breakpoints of the inversions without disrupting the 5'UTR and 3'UTR of genes. The sites of \textit{loxP} insertion, the binding sites of transcription regulatory factors and the regions of 5'UTR and 3'UTR are shown in the Figure 2.25. We checked the expression of \textit{DAL1}, \textit{DAL2} and \textit{DAL4} in our parent strain and the control strain having \textit{loxPs} and \textit{lox2272s}. In few strains, we observed significant change in expression of \textit{DAL} genes (Figure 2.26) however, in most of the cases, we did not observe any significant change in expression of \textit{DAL} genes upon insertion of \textit{loxP} or \textit{lox2272} (data not shown). Any change in expression occurring due to insertion of \textit{loxPs} or \textit{lox2272s} was taken into consideration by normalizing the expression of inverted strains with the control non-inverted strains.
Figure 2.25: The map of the 5’UTR and 3’UTR in the DAL gene cluster. Representation of the DAL genes sense transcripts (green arrows), antisense transcripts (blue arrows), sites of loxP insertions (red arrows) and transcription factor binding sites (black arrows). The number in square boxes indicates 5’UTR and 3’UTR regions.
Figure 2.26: Effect of *loxP* and *lox2272* insertion. Panel A Two *loxP*s were inserted at 3’ end of *DAL1* and *DAL2* in FY3 strain to invert *DAL1*, *DAL2* and *DAL4* genes. We observed a significant change in expression of *DAL4* by the insertion of *loxP*s. Panel B Two *lox2272*s were inserted at 5’ and 3’ end of *DAL2* in DAL1-DAL2-DAL4-Inv strain to invert *DAL2*. A significant change in *DAL1* expression was observed upon *lox2272* insertion. Panel C Two *loxP*s were inserted at 5’ and 3’ end of *DAL2* in FY3 strain to invert *DAL2* showing a significant change in *DAL2* expression upon *loxP* insertion.
2.4.11 Expression of *DAL4* antisense transcript (*SUT614*)

*Saccharomyces cerevisiae* possess several transcripts which either exist as stable unannotated transcripts (SUTs) or are rapidly degraded and are cryptic unstable transcripts (CUTs) [20]. Steinmetz’s bidirectional promoter database showed that three out of six *DAL* genes in the cluster possess stable unannotated transcripts which are *SUT195*, *SUT614* and *SUT196*.

To study the impact of gene inversion on antisense RNA, we performed northern analysis of *DAL4* and *SUT614* in DAL2-Inv and DAL2-NI strains. We observed that *DAL4* anti-sense RNA (*SUT614*) is highly expressed in allantoin containing medium in both inverted and non-inverted strains as compared to the mRNA. We also checked the expression of both *DAL4* and *SUT614* in YPD medium for both strains. We observed that *DAL4* mRNA is not expressed in YPD (Figure 2.27) however, in allantoin containing medium the expression is slightly enhanced (Figure 2.28).

It has previously been reported that proline supplemented with hydantoin acetate (HAA) serves as a good nitrogen source in the medium for *DAL4* expression [15]. We therefore decided to study the expression of *DAL4* sense and antisense transcripts in medium containing 0.1% (w/v) proline + 0.0125% (w/v) allantoin of our DAL2-Inv, DAL2-NI and FY3 strains. We observed that *DAL4* sense and antisense RNA is highly expressed in FY3 and DAL2-NI strains in comparison to actin, however, *DAL4* mRNA showed much lower expression in the inverted strain (Figure 2.29).
Figure 2.27: Detection of sense and antisense *DAL4* RNA transcripts in YPD medium. 20µg of total RNA samples from DAL2-Inv and DAL2-NI strains was analysed. The strains were grown in YPD medium to exponential phase (OD$_{600nm}$=0.5). RNA extraction and northern blot analysis were performed as described in materials and method. Probes used were *DAL4* sense, *DAL4* antisense and *ACT1*. *DAL4* sense and antisense signal of inverted and non-inverted strain compared to actin in YPD medium. Results of a representative experiment are shown. Independent experiments showed similar expression profiles of *DAL4*.

![Diagram](image1)

Figure 2.28: Detection of sense and antisense *DAL4* RNA transcripts in allantoin medium. 20µg of total RNA samples from DAL2-Inv and DAL2-NI strains was analysed. The strains were grown in allantoin medium to exponential phase (OD=0.5). RNA extraction and northern blot analysis were performed as described in materials and method. Probes used were *DAL4* sense, *DAL4* antisense and *ACT1*. *DAL4* sense and antisense signal of inverted and non-inverted strain compared to actin in allantoin medium. Results of a representative experiment are shown. Independent experiments showed similar expression profiles of *DAL4*.

![Diagram](image2)
Figure 2.29: Detection of \textit{DAL4} sense and antisense transcripts in \textit{Proline + allantoin medium}. \textit{DAL4} sense and antisense signal of inverted and non-inverted strain compared to actin is shown. 20µg of total RNA samples from FY3, DAL2-Inv and DAL2-NI strains was analysed. The strains were grown in medium to exponential phase (OD\textsubscript{600nm}=0.5). RNA extraction and northern blot analysis were performed as described in materials and method. Probes used were \textit{DAL4} sense, \textit{DAL4} antisense and \textit{ACT1}. Results of a representative experiment are shown. Independent experiments showed similar expression profiles of \textit{DAL4}.

2.5 Discussion

\textit{Saccharomyces castellii} \textit{DAL} cluster differs from \textit{Saccharomyces cerevisiae} and other \textit{sensu stricto} species by gene inversion. This cluster was formed during the period of evolution in \textit{sensu stricto} species and \textit{S. castellii} and is highly conserved among all the species. The overall size of the \textit{DAL} cluster is highly conserved between \textit{S. cerevisiae} and \textit{S. castellii} differing by only 1bp. However, there is no size conservation between independent genes and the non-coding regions. We observed that sizes of \textit{DAL1} and \textit{DAL3} differ significantly between \textit{S. cerevisiae} and \textit{S. castellii}. \textit{DAL1} is 106bp (7%) bigger in \textit{S. castellii} than in \textit{S. cerevisiae} and \textit{DAL3} is 172bp (29%) smaller in \textit{S. castellii} than in \textit{S. cerevisiae}.

Analysis of the \textit{DAL} cluster in \textit{S. cerevisiae} showed the presence of an \textit{ARS} sequence overlapping the 3’ ends of \textit{DAL2} and \textit{DCG1}. We did not find any \textit{ARS} sequence in \textit{S. castellii}. The region between \textit{DAL2} and \textit{DCG1} in \textit{S. castellii} has been disrupted due to \textit{DAL2} inversion. This inversion could have possibly disrupted \textit{ARS} as it overlaps \textit{DAL2} and \textit{DCG1}.

The \textit{DAL} cluster is expressed under nitrogen limited conditions to degrade allantoin.
which is a non-preferable source of nitrogen to a simpler form of nitrogen i.e. ammonia. Allantoin was thought to be the most suitable nitrogen source in the medium as it is the first substrate in the metabolic pathway [21]. The previous work on allantoin metabolism in \textit{S. cerevisiae} used the medium containing 0.2% allantoin concentration [22]. The data in this study suggested that a higher concentration of allantoin (0.4% (w/v)) was deleterious to growth and the cells were happily growing in low and medium concentrations of allantoin (0.0125% (w/v)-0.2% (w/v)). We therefore used 0.0125% (w/v) allantoin in our experiments.

Our results showed that the laboratory strains and natural isolates of \textit{S. cerevisiae} are much better growers in allantoin containing medium than wild type \textit{S. castellii}. We then selected a wild type \textit{S. cerevisiae} lab strain FY3 to construct the inverted and non-inverted strains.

The three genes (\textit{DAL1}, \textit{DAL2} and \textit{DAL4}) in \textit{DAL} metabolic cluster of \textit{S. castellii} are shown to be inverted and \textit{DAL2} has been inverted twice [16]. We aimed to study the role of these inversions on fitness and gene expression, therefore, constructed various inverted and non-inverted strains including a strain possessing an \textit{S. castellii} like \textit{DAL} cluster. It is not known which gene in the cluster was inverted first, we therefore adopted two approaches to construct \textit{S. castellii} like \textit{DAL} cluster in \textit{S. cerevisiae} background (Figure 2.19).

This study demonstrated that the inversion of \textit{DAL1}, \textit{DAL2} and \textit{DAL4} does not have any effect on fitness in YPD, F1 or in allantoin medium, however, a negative fitness effect was observed when only \textit{DAL2} gene was inverted. The two strains S.cer-Inv1 and S.cer-Inv2 mimicking \textit{S. castellii} \textit{DAL} structure showed a significant change in fitness. The S.cer-Inv1 inversion made from re-inverting \textit{DAL2} in \textit{DAL1}-\textit{DAL2}-\textit{DAL4}-Inv strain showed a negative fitness effect which was similar to the effect observed by inverting just \textit{DAL2}. However the S.cer-Inv2 inversion made from inverting \textit{DAL1}, \textit{DAL2} and \textit{DAL4} in \textit{DAL2}-inv strain showed a positive fitness effect. It is not known which \textit{DAL} gene was first inverted in nature but since \textit{DAL2} inversion had a selective disadvantage on fitness we hypothesized that the order of these inversions can be that \textit{DAL2} was inverted first and because it possessed a negative
fitness effect subsequently *DAL1*, *DAL2* and *DAL4* were inverted leading to the formation of the *S. castellii* like *DAL* cluster showing positive fitness.

The six *DAL* genes involved in allantoin metabolism got clustered on the right arm of chromosome IX during the period of evolution which was then environmentally favoured by the yeast species [16]. Previous literature have shown that *DAL4* mutants grow normally in medium containing either ammonia, allantoate, arginine or asparagines as nitrogen sources however they are incapable of growth in allantoin containing medium [23].

Allantoin is a purine derivative formed by oxidation of xanthine to urate and then to allantoin in purine degradation pathway. These reaction are catalyzed by xanthine dehydrogenase (*XDH*) and urate oxidase (*UOX*). *XDH* is not present in yeast species and therefore they import urate, allantoin or allantoate into the cell to use purine derivatives as nitrogen source. The yeast species which do not posses *DAL* cluster are capable of using urate as a nitrogen source using urate permease (*UAP*) to import it and *UOX* to oxidize it and the enzymes of *DAL* metabolic pathway to degrade it to urea [24]. *S. cerevisiae* and *S. castellii* do not have *UAP* and *UOX* and therefore are unable to use urate and thus import allantoin as nitrogen source using allantoin permease gene *DAL4*. Once imported allantoin is then degraded to simpler form of nitrogen using the same *DAL* pathway genes present in all yeasts; however, in *S. cerevisiae* and *S. castellii* the *DAL* genes have been organized into a cluster. The ability to import allantoin instead of urate ruled out the oxygen requiring step performed by *UOX* leading to the biochemical reorganization of purine degradation pathway. The species which possess the *DAL* cluster have the ability to grow under oxygen limiting conditions [16].

All the genes in *DAL* metabolic pathway are sensitive to nitrogen catabolite repression (NCR) and are only expressed when readily used nitrogen sources such as ammonia, glutamine and asparagine are not available for utilization [25]. The NCR-sensitive transcription of allantoin pathway genes is mediated by three types of cis-acting and trans-acting factors which are: 1) *UAS_{NTR}* binding either Gln3p or Gat1p [26]. They are present upstream of all allantoin pathway genes (*DAL1, DAL2, DAL3, DAL4* and *DAL7*). 2) *URS_{GATA}* which is associated with DAL80p [27]. These sequences down-
regulate the \textit{DAL} gene expression. 3) UIS\textsubscript{ALL} binding DAL81p and DAL82p [28]. \textit{DAL5} and \textit{DAL3} are inducer independent genes.

In this study, we showed that the strains \textit{S.cer} Inv1 and DAL2-Inv harbour reduced expression of \textit{DAL4} in allantoin containing medium. Analyzing the growth rate of these two strains revealed a reduced fitness effect, which could be explained by the down regulation of \textit{DAL4} as this gene is responsible for allantoin uptake by the cell.

It has been shown that expression of \textit{DAL1}, \textit{DAL2} and \textit{DAL4} is dependent upon functional \textit{GLN3}, \textit{DAL82} and \textit{DAL81}, however \textit{DAL80} null mutants showed enhanced expression of \textit{DAL4} [25]. To construct our inverted strains we inserted \textit{loxP} and \textit{lox2272} sequences at various positions of the genes (Figure 2.25). Our results show that expression of these genes is dependent upon one another. We inserted a \textit{loxP} sequence at 3’end of \textit{DAL1} such that none of the transcription sites at \textit{DAL1} were disrupted and therefore we didn’t observe any change in expression of \textit{DAL1} gene upon insertion of the \textit{loxP} sequences. The \textit{DAL2} possess an \textit{ARS} sequence overlapping 3’ end of \textit{DAL2} and \textit{DCG1}. The \textit{loxP} sequence was inserted immediately after the stop codon of \textit{DAL2} by repeating the \textit{ARS} sequence in our primer so that it is not disrupted. We observed a significant change in expression of \textit{DAL2} as well as \textit{DAL4} after insertion of the \textit{loxP} sequences (Figure 2.26). No significant change in expression was observed for both \textit{DAL2} and \textit{DAL4} in the strain having \textit{loxP} sequence inserted at the 5’ end of \textit{DAL2}. This \textit{loxP} sequence was inserted such that none of the transcription sites for \textit{DAL2} and \textit{DAL4} were disrupted.

In yeasts and other higher eukaryotes gene clusters can either be formed by genetic linkage or by physical clustering of the genes which are co-regulated [29]. Co-expression of the genes can either be because of bidirectional promotors [20] or due to regulation at chromatin level [30]. The \textit{DAL} cluster seems to have been formed by both processes i.e. co-regulation and genetic linkage. The \textit{DAL} cluster may have been formed by genetic linkage keeping the \textit{DAL} gene alleles in close proximity on the same chromosome. Therefore the \textit{DAL} genes are co-regulated and change in expression of one gene is dependent on the expression of other gene [31]. It has also been shown that gene order in eukaryotes is related to gene function and is not random. Our results
demonstrate that by inverting \textit{DAL1}, \textit{DAL2} and \textit{DAL4} the expression of \textit{DAL1} and \textit{DAL4} is enhanced (Figure 2.22) however when \textit{DAL2} is inverted twice it reduces the expression of \textit{DAL4} (Figure 2.23b). It was also observed that \textit{DAL2} inversion not only effects its own gene expression but also the expression of neighbouring genes (Figure 2.24). These genes are therefore located next to each other and each of the allele alone is detrimental in absence of other [32].

As we see from the Fig 2.1 \textit{DAL1}, \textit{DAL2} and \textit{DAL4} require functional \textit{GLN3}, \textit{DAL82} and \textit{DAL81} for transcription. Although our molecular modification done to engineer the inverted and non-inverted strains did not disrupt the transcription factors, here can be a possibility that inversion of these genes disrupt the function of these transcription factors which in turn disrupt the function of gene.

Co-expression of closely located genes is not only dependent upon transcription factors or chromatin structure but is also dependent upon bidirectionally active promoters [33]. It has previously been shown that \textit{DAL4} possess an unannotated transcript \textit{SUT614}. To test the role of bidirectional promoter in \textit{DAL4} we looked at the expression of \textit{SUT614} in our \textit{DAL2}-Inv and non-inverted strain. Cooper et al. [15] and Sumrada and Cooper [34] demonstrated that \textit{DAL4} expression is induced by hydantoine acetate an analog of allantoin, or oxaluric acid (OXLU) an analogue of allophanate both of which are non-metabolizable inducers of allantoin degradation pathway. We therefore checked the expression of \textit{DAL4} and \textit{SUT614} in allantoin medium and in medium containing mixture of proline and allantoin. As expected, we observed different levels of expression for \textit{DAL4} and \textit{SUT614} in different media. It has been reported that the induced expression of \textit{DAL4} is strain specific, different strains show different response to presence or absence of the inducer and nitrogen source. The expression analysis of \textit{DAL4} and \textit{SUT614} in proline+allantoin medium exhibited that \textit{DAL4} expression is reduced in the inverted strain when compared to its counterpart non-inverted and wild type FY3 strains. However, \textit{SUT614} was equally highly expressed in all the three strains (Figure 2.29). This suggests that inverting \textit{DAL2} reduces the expression of its neighboring gene \textit{DAL4} without altering its antisense transcript \textit{SUT614}.
2.6 Materials and methods

2.6.1 Strains and Media

All strains were engineered with the FY3 background and they were maintained on YPD medium containing 2% (w/v) yeast extract, 1% (w/v) peptone and 2% (w/v) glucose. The transformants were grown on YPD-agar containing the desired antibiotics i.e 300µg/ml geneticin (GibcoBRL), 100 µg/ml cloNAT (Werner BioAgents, Jena, Germany), 10 µg/ml phleomycin (InvivoGen) and 300 µg/ml hygromycin B (Duchefa Biochemie) for selection of the kanMX, natNT2, pCre-ble and hphNT1 markers. A full list of engineered strains is provided in the appendix (Table A6). Mineral salt medium was prepared as previously described [35]. 0.0125% (w/v) allantoin was used as nitrogen source.

2.6.2 Primers and Oligonucleotide probes

Gene sequences were obtained from SGD and PCR primers were designed using the Primer3 programme. Strand specific oligonucleotide probes were manually designed keeping the GC content to 40% and length to 34-40 bp. The BLAST tool of SGD was used to check the specificity of each probe and primer. Primer and probe sequences are provided in the appendix (Table A2-A4 & A8).

2.6.3 Plasmid DNA Extraction and Restriction Digestion

QIAprep spin miniprep kit was used to extract the plasmid DNA following the manufacturer's guidelines (QIAGen catalogue no. 27104). The list of plasmids used is provided in the appendix (Table A1). The extracted DNA was digested by specific restriction enzymes. The reaction mixture consisted of 10 units of the restriction enzyme, 100-200 ng of plasmid DNA and an enzyme specific buffer. The final volume was adjusted to 20 µl by addition of sterile dH2O. The reaction mixture was incubated at 37°C for 1 hour and DNA was visualised under UV light immediately after agarose gel electrophoresis.
2.6.4 Transformation of Saccharomyces cerevisiae by Lithium Acetate protocol

The Saccharomyces cerevisiae strain was transformed using the high efficiency yeast transformation protocol [36]. The transformation mix consisted of 240 µl of 50% (w/v) PEG<sub>3350</sub>, 36 µl of 0.1M lithium acetate, 50 µl of single-stranded carrier DNA (2mg/ml) and 34 µl plasmid DNA (0.1-1 µg). The mixture was dissolved by gentle vortexing until the pellet was resuspended and incubated for heat shock at 42°C for 40 minutes. The transformants were allowed to grow in YPD for 4-12 hours before platting out on appropriate selectable medium containing geneticin, hygromycin or neocotricin antibiotic. The plates were incubated at 30°C for 2-3 days. The transformant colonies were tested by colony PCR.

2.6.5 Confirmation of the strains by colony PCR

Colony PCR experiments were done to confirm the insertion of cassettes at correct position in the genome and to verify for the inverted and non-inverted genes. Different sets of primers were used for each specific reaction (appendix Table A2). The template was prepared by dissolving a small amount of colony in 38.5 µl of sterile milliQ H<sub>2</sub>O. The cell suspension was heated at 95°C for 10 minutes and the ruptured cells were allowed to cool on ice for 5 minutes. The PCR mixture was prepared by adding 0.4 mM of dNTPs, PCR buffer provided by the supplier, 1.5mM of MgCl<sub>2</sub>, 50 pmol of each primer and 5 units Roche Taq polymerase. 11 µl of the reaction mixture was added in 38.5 µl of cell suspension to bring up the final volume to 50 µl. The thermocycling conditions used were 3 minutes at 95°C, followed by 35 cycles of 45 seconds at 95°C to denature DNA, 45 seconds at 55–58°C for primer annealing and 1–3 minutes at 72°C for DNA elongation. Finally a cycle of 5 minutes at 72°C was added to ensure full extension of DNA. 15 µl of the PCR product was loaded on 0.8% (w/v) or 1% (w/v) agarose gel electrophoresis in 1xTAE and visualized by ethidium bromide staining on a transilluminator.

2.6.6 Selection marker cassettes amplification by PCR

The cassettes used in this study were loxP-kanMX-loxP, loxP-hphNT1-loxP and
These cassettes were amplified using the gene specific primers (appendix Table A8). The loxP-kanMX-loxP cassette was amplified according to Delneri et al. 2003 [37]. The PCR reaction mixture consisted of 1µl template DNA, 0.4mM of dNTPs, 5x PCR buffer provided by the supplier, 1.5mM of MgCl₂, 3µM of each primer and 5 units Roche Taq Polymerase. The final volume was adjusted to 50µl by sterile MilliQ water. The loxP-hphNT1-loxP cassette amplification had the same reaction mixture as loxP-KanMx-loxP apart from buffer which was 10x buffer I (500mM Tris/HCl pH 9.2, 22.5mM MgCl₂, 160mM (NH₄)₂SO₄ and the cycling conditions used were; an initial start of 3min at 95°C, followed by 10 cycles of 30 sec at 95°C for DNA denaturation, 30 sec at 55°C for primer annealing and 2 min 40 seconds at 68°C for DNA extension. This was followed by 20 additional cycles for same DNA denaturation, primer annealing and extension with an addition of 20 sec/cycle. A last step of 5 minutes at 72°C was included to ensure full extension of DNA [38]. The reaction conditions for amplifying lox2272-natNT2-lox2272 were same as loxP-hphNT1-loxP but the reaction mixture was made using GC-RICH PCR kit (Roche catalogue no. 12 140 306 001 ) which consisted of; Mix1: 0.2 mM dNTPs, 0.2-0.5 µg of template, 5µl of resolution solution containing DMSO (provided in the kit) and 0.2 µM of each primer. The final volume of mix1 was made up to 35µl by sterile distilled water. Mix 2 consisted of 5 x PCR buffer, 1 µl Taq Polymerase and final volume made up to 15 µl by sterile milliQ water.

2.6.7 Construction of Inverted and Non-inverted strains

The resistance gene marker cassettes containing the loxP loci were inserted in the genome by PCR-mediated gene replacement method [39]. Transformation was done by lithium acetate method [36]. The strains bearing cassettes were then transformed with Cre-recombinase containing plasmid. The Ce-recombinase enzyme was induced by first growing the cells overnight in YP-raffinose medium and then in YP-galactose for 2-3 hrs. The colonies were verified for inversion and non-inversion by colony PCR.

2.6.8 Fitness Growth Rate Assay

Growth rate of all the inverted and non-inverted strains was determined using
FLUOstar optima microplate reader. Cells were grown to stationary phase in YPD, minimal and allantoin containing medium. Optical density of the cultures was measured at 595nm and then diluted to OD_{595nm} = 0.1 in pre-warmed respective medium. 240 µl of the diluted cultures was transferred to each well of 96 well plate including the medium controls. The OD measurement was taken by the microplate reader at 30°C for 40 hours. The optical density of the cultures was measured every 5 minutes at 595 nm with 1 minute linear shaking just before every OD measurement. Growth curves were plotted using the Optima data analysis programme and standard deviations were calculated in Excel.

2.6.9 Reverse Transcription

Total RNA for reverse transcription was extracted using the Qiagen RNA extraction kit following the manufacturer's instructions. RNA concentration was determined using the nanodrop spectrophotometer (ND-1000), quality and integrity of RNA was checked by electrophoresis on 1.5% agarose gel. Whenever required RNA was treated with DNaseI (Fermentas) prior to cDNA synthesis as described by the manufacturer. 1µg of total RNA was reverse transcribed to cDNA in a 20µl reaction mixture by Qiagen reverse transcription kit using the random primers.

2.6.10 Real-time quantitative PCR

The quantitative real time PCR (qPCR) was performed on the cDNA of wild type strain, inverted and non-inverted engineered strains using the Quantitect real time PCR kit from Qiagen. All real-time PCRs were performed on Chromo4 gradient thermocycler in 96 well plate from Biorad. Real time PCR primers were synthesized by MWG-Eurofins (HPSF purified). Each primer was designed to amplify a 250-300 bp fragment. Optimized reactions were carried out in 50 µl final volume containing 10 ng/µl of cDNA, 5 pmole of each primer and 25µl of 2x quantitect syber green. The qPCR conditions were used with an initial denaturation of 3min at 95°C followed by 35 cycles consisting of 95°C for 45 sec, 58°C for 45 sec and 72°C for 3 min with a final extension of 5 min at 72°C. Melting curves were analyzed from 55°C to 95°C at a rate of 0.2°C/2sec. Actin (ACT1) was used as a housekeeping reference gene. Serial dilutions
(10^{-1}-10^{-5}) of actin DNA was used for generating standard curve. The expression of each gene was estimated using the Ct Values. All real time PCRs were tested in triplicate and each experiment was done on three independent biological replicas. The overall standard deviation is shown in results. A blank (with no RT) was also included in each experiment.

2.6.11 RNA extraction and Northern hybridization

Total RNA was extracted from yeast strains using Trizol (Invitrogen, catalogue # 155-96-018) as described by the manufacturer. 20 µg of the total RNA was loaded in each slot and resolved on 1% (w/v) formaldehyde agarose gel [40]. Samples were prepared by adding RNA loading dye from Fermentas which contained ethidium bromide. The RNA was transferred on the nylon membrane with panther semidry electroblotter HEP-1 in accordance to the manufacturer's manual using 1XTBE as transfer buffer. The RNA was fixed on the membrane by UV irradiation for 1min using the UV crosslinker (XL-1500) 1200 µJ/cm^2 for 60 sec. 5 pmole oligo was labelled with [32p] ATP using T4 polynucleotide kinase (Fermentas, catalogue #EK003) and membrane was hybridized at 37°C in oligo hyb solution (0.17Moles Na_2HPO_4, 0.079Moles NaH_2PO_4, 35g SDS, 1ml 0.5M EDTA, dH_2O to 500 ml and warm to dissolve, pH ~7.2) for overnight. The membranes were washed at 42°C in 6 X SSC for 10min and 2 X SSC, 0.1% SDS for 10 min. The final washing was done at room temperature in 6 X SSC. Membranes were exposed to phosphoimager screen for 1-3 days and band intensities were quantified using Quantitect programme from Biorad.

2.7 References


25. Daugherty JR, Rai R, el Berry HM, Cooper TG: Regulatory circuit for responses of nitrogen catabolic gene expression to the GLN3 and DAL80


CHAPTER 3 (Paper 2)

EFFECT OF PERICENTRIC AND PARACENTRIC CHROMOSOMAL INVERSIONS ON GLOBAL GENE EXPRESSION
3. Effect of pericentric and paracentric chromosomal inversions on global gene expression

S Naseeb, Z Carter, D Minnis, L Zeef and D Delneri

To be submitted to Genome Biology

3.1 Foreword

This chapter contains results which will be submitted to Genome Biology journal. SN performed the gene expression and fitness assays (SD medium) and wrote the manuscript, LZ and SN analyzed the array data (LZ= performed data normalization and PCA analysis; SN = performed all the other analysis mentioned in manuscript), ZC constructed all the strains, ZC and DM performed the fitness assays (N-limited and C-limited media) and spore viability experiments and DD designed and supervised the project and reviewed the manuscript.

3.2 Abstract

**Background:** Chromosomal inversions play an important role in genome evolution. The inversion breakpoints can disrupt transcription factor binding motifs or open reading frames altering the gene expression. They can also reorganize the genome causing structural problems during meiosis. It is therefore important to study the impact of inversions on gene expression at global level.

**Results:** Here, we analyzed the effect of pericentric and paracentric chromosomal inversions on fitness, crossing over and gene expression. We demonstrated that pericentric and paracentric inversions did not have any effect on the phenotypic fitness. We showed that strains possessing small size inversions had no effect on crossing over and they had spore viability of 92-95% whereas some of the medium and large size inversions reduced the spore viability to 50%. We also showed that effect of inversions on gene expression is random. Sometimes inversions have no effect on gene expression however in some cases it not only alters the expression profile of the genes located
within the breakpoints but also of the genes present in the neighbourhood of the breakpoints as well as in the other parts of the genome.

**Conclusion:** This finding suggests that, the impact of inversions on spore viability mostly depends upon the size of the inversion. Medium and large inversions are lethal to the strain as compared to the small inversions. The effect of inversion on gene expression is independent of the size and type of the inversion. Some large size inversions changed the expression globally whereas others had no impact on gene expression.

**Keywords**

Pericentric inversion, paracentric inversion, transposon, *Saccharomyces cerevisiae*

### 3.3 Background

Genome rearrangements amongst the yeast species are common. The highly similar genomic contents of yeast genomes can be arranged differently in distinctive or same species. These differences in the genome structure have been shown to contribute towards genome-wide genetic incompatibilities, which have been shown to reinforce reproductive isolation [1].

Dunham *et al.* [2] characterised genomic rearrangements in *Saccharomyces cerevisiae* strain which was evolved to 100-150 generations in glucose limited chemostate cultures. They showed that many of these rearrangements were flanked by transposon sequences which acted as the recombination sequences indicating the breakpoints were nearby Ty elements.

There are different types of chromosomal rearrangements: inversions, translocations, duplications and deletions. Most commonly studied and easiest to observe are translocations and duplications since these differences are easily spotted on Pulse Field Gel Electrophoresis (PFGE). However, inversions cannot be seen on PFGE and whole genome sequence of a strain of interest needs to be available for identifying inversions. Not many cases of inversions have been annotated in yeast and one possibility for that
can be due to tightly packed genome of yeast, which does not accommodate such rearrangements.

The first evidence of chromosomal inversions was published in 1921 by Sturtevant who studied rearrangement of genes in Drosophila [3]. The current literature shows that inversions are now reported in almost all organisms ranging from prokaryotes to eukaryotes. They have been reported in plants [4], rodents [5], snails [6] and fruit flies [7]. Moreover it has been identified that human and chimpanzees genome differ by nine pericentric inversions affecting chromosomes 1, 4, 5, 9, 12, 15, 16, 17 and 18 which permanently separated two lineages [8]. In 2008, Lee and his co-authors showed that these inversions were generated by the recombination of retrotransposons known as interspersed element-1 (LINE-1) and Alu elements present in human genome. They also identified 252 inversions by comparing the genome sequences of human and chimpanzees and reported that 44% of the 252 inversions were caused by the retrotransposons [9].

Mobile genetic elements (MGE) are segments of DNA that can move within the genome and they are of several different types i.e. transposons, retrotransposons, DNA transposons, insertion sequences, plasmids, bacteriophages and group II introns. The presence of these highly repetitive sequences could theoretically cause a series of rearrangements such as duplications, inversions and translocations.

The aim of this project is to study the impact of pericentric and paracentric inversions on the global gene expression profile and phenotype. We chose to create inversions in the proximity of Ty elements because they act as recombination sequences. Among the five classes of Ty elements (Ty1-Ty5) existing in yeast, we considered the Ty1 because it is the most abundant retrotransposon with 31 copies per haploid cell [10]. Strains of *Saccharomyces cerevisiae* in which sections of chromosomes were inverted were available in Dr. Delneri’s lab (unpublished data). The inversions were created between repetitive sequences of Ty1 elements where they are more likely to occur in nature. In this piece of research, we present the phenotypic and transcriptional analysis of the engineered inverted and non-inverted strains. We show that both small and large inversions have no significant change on the phenotypic fitness of strains. We found
that small inversions result in no impact on crossing over however, some of the strains carrying medium and large inversions had reduced spore viability, to 50% of normal. To better understand the role of pericentric and paracentric inversions at the transcriptome level, we performed microarray analysis of inverted and non-inverted individuals. We showed that chromosomal inversions do not always alter the global expression profile. It was observed that in some cases inversions did not show any effect on gene expression however in one strain inversion not only changed the expression of genes located within the inversion but also of the genes which were present in the immediate vicinity of inversion breakpoints and of the genes located far away from the breakpoints.

3.4 Materials and Methods

3.4.1 Strains and Media

All inverted and non-inverted yeast strains used in this study were previously constructed in the Delneri Lab (data unpublished). A complete list of the strain names and their genotype is given in Table A7 (appendix). The strains were maintained on YPD medium containing 2% (w/v) Bacto-yeast extract, 1% (w/v) Bacto-peptone and 2% (w/v) glucose. The mineral salt medium (carbon limited and nitrogen limited) were prepared as described previously [11]. The inverted and non-inverted strains were sporulated on minimal sporulation medium containing 1% (w/v) potassium acetate and 2% (w/v) Bacto-agar along with the auxotrophic supplements.

3.4.2 Fitness Growth Rate Assay

Growth rate of all the inverted and non-inverted strains was determined using FLUOstar optima microplate reader. Cells were grown from a starting OD$_{595}$ of 0.1 to stationary phase in Carbon and Nitrogen limited F1 growth media. The OD measurements were taken by the microplate reader at 30°C for 72 hours. The optical density of the cultures was measured every 5 minutes at 595 nm with 1 minute linear shaking just before every OD measurement. In each case, three biological replicates of each strain (each with three technical replicates) were used for a total of nine sets of
growth values per strain. These were averaged and blank-corrected before being plotted. Growth curves were plotted using the Optima data analysis programme and standard deviations were calculated in Excel.

3.4.3 Spore viability of inverted and non-inverted strains

All of the inverted and non-inverted strains (in background BY4741) were crossed with BY4742 using a microneedle. At least three biological replicas of an inverted and three of non-inverted strains were crossed in each case. The hybrids were grown in pre-sporulation medium (0.8% (w/v) Bacto-yeast extract, 0.3% (w/v) Bacto-peptone and 2% (w/v) glucose) at 30°C for 12 hours. The cells were washed twice with sterile milliQ H₂O and plated on minimal sporulation medium (5g potassium acetate, 0.5g Bacto-yeast extract, 0.025% (w/v) glucose, 10g agar in 500ml of distilled H₂O). The sporulation plates were incubated at 20°C for 5-7 days for the formation of tetrads. The tetrads were dissected using Singer MSM-300 micromanipulator. Spore viability was calculated based on the percentage of colonies that had grown for each variant of the strain out of a possible 240 dissected spores.

3.4.4 RNA extraction and whole genome transcriptome analysis

The inverted and non-inverted strains were grown in SD medium (0.67% (w/v) of bacto yeast nitrogen base without amino acids, 2% (w/v) glucose and 2% (w/v) bacto-agar) to mid-log phase (OD₅₉₅nm = 0.5). Total RNA was extracted in triplicate by using Trizol reagent following the manufacturer’s instructions (Invitrogen, catalogue # 155-96-018) and quantified using the nanodrop ultra-low-volume spectrophotometer (Nanodrop Technologies). The quality of total RNA was determined by gel electrophoresis on 1.5% (w/v) agarose gel in TAE buffer and using Agilent Bioanalyzer 2100 (Agilent Technologies Ltd, UK). The mRNA expression levels in the inverted and non-inverted strains were analysed using a GeneChip Yeast Genome 2.0 Array (Affymetrix) for three replicates each for the inverted and the non-inverted strains [12]. The arrays were run according to the manufacturer’s instructions. Array data were normalized and summarized using dChip software [13]. The background correction, quantile normalization, and gene expression analysis were performed using RMA in
Bioconductor [14]. Principal component analysis (PCA) was performed with the Partek Genomics Solution (version 6.5, Copyright 2010, Partek Inc., St. Charles, MO, USA). A false-discovery correction was applied to \( p \)-values to produce a \( q \)-value [15]. Genes with significant up-regulation or down-regulation were identified at \( q \leq 0.2 \). A complete microarray data set was submitted to MIAME (Minimum Information About a Microarray Experiment).

### 3.5 Results

Ty elements are an important source of genome evolution. In nature these retrotransposons can lead to intrachromosomal crossovers or interchromosomal crossovers resulting in gene deletions, inversions, duplications or translocations of genomic material which in turn may alter the gene expression. Changes in the expression level can bring about many structural and morphological variations contributing to phenotypic and genetic diversity in a population leading to differentiation and evolution of species. It is therefore of utmost importance to study the mechanisms behind the changes in gene expression. To study the phenotypic and transcriptional changes occurring due to pericentric and paracentric inversions between Ty elements a total of 16 strains carrying inversions were previously constructed in BY4741 background (Delneri lab, data unpublished). Seven of the constructed strains possessed paracentric inversions of different sizes on different chromosomes and five possessed pericentric inversions of different sizes on different chromosomes (Table 3.1 & 3.2).

#### 3.5.1 Strategy used to construct inverted and non-inverted strains

Strains carrying pericentric and paracentric inversion were constructed using the cre-loxP system [16]. Wild type \( S. cervevisiae \) (BY4741) was the background strain used to construct all the inverted and non-inverted strains. The steps included were amplification and transformation of the desired resistance marker cassette, transformation and induction of \( cre \) and selection of the inverted and non-inverted strains (Figure 3.1).
Figure 3.1: The cre-loxP recombination system to construct inverted and non-inverted strains. Panel A represents the amplification of loxP-kanMX and loxP-hphNT1 cassettes using the primers containing 40bp sequence homologous to the Ty1 element (yellow and red blocks) and 20bp sequence specific to the cassette (red arrow). Panel B upon transformation the amplified cassette was inserted in the genome by homologous recombination. Panel C the confirmation of cassettes at their specific loci was done by diagnostic colony PCR. Primer pair a + b was used to confirm the position of loxP-kanMX and primers c + d were used to confirm the position of loxP-hphNT1 in the chromosome. Panel D strain possessing resistant marker cassette was transformed by cre containing plasmid. The cre-recombinase enzyme was induced to recombine the loxP sites and provoke inversion. Panel E the strains possessing inversion and no inversion was confirmed by diagnostic colony PCR using the primers c + f and e + b for inversion, primers a + b and c + d for no inversion.

3.5.2 Inversion length and Spore viability

In order to assess whether the size and position of the inversion on the chromosome had any effect on the spore viability, each of the constructed inverted and non-inverted strains (in BY4741 background) were crossed with BY4742. At least three biological
replicas of inverted and three of non-inverted strains were crossed in each case. The strains were sporulated on minimal sporulation medium and the tetrads dissected.

It was observed that small size pericentric and paracentric inversions didn’t have much effect on the spore viability however some of the medium and large size inversions (strains 12/13, 25/24 and 29/28) decreased the spore viability to 50% in comparison to the non-inverted strains (Table 3.1 & 3.2). In strains 31/29, 6/8, 4/8 and 30/28 we were unable to find inversions and therefore it is postulated that these inversions are lethal (Table 3.3).

3.5.3 Recombination hotspots

Since we observed 50% loss of spore viability in the inverted individuals of strain 12/13, 25/24 and 29/28, our next approach was to look for the presence of hotspots of recombination within the inverted region of these strains compared with strains which did not show any significant change in the spore viability. In 2008, Mancera et al. generated a full map of recombination hotspots for crossovers (CO) in *S. cerevisiae* [17]. We used genomic coordinates of these recombination hotspots and compared them with the inversion boundaries of our engineered strains. A map of inverted region for strain 29/28 showing the location of CO hotspots is shown in Figure 3.2. All strains were mapped by the same method.
Table 3.1: List of engineered strains possessing paracentric inversion. The size, location of the inversion and percentage of spore viability is reported.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Chromosome number</th>
<th>Location of inversion and cassette insertion</th>
<th>Type of inversion</th>
<th>Size of inversion (kb)</th>
<th>Spore viability of inverted strains (%)</th>
<th>Spore viability of non-inverted strains (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/13</td>
<td>chVII</td>
<td>YGRCTy1-2::loxP-kanMX + YGRWTy1-1::loxP-hphNT1</td>
<td>Paracentric</td>
<td>26</td>
<td>86</td>
<td>94.93</td>
</tr>
<tr>
<td>6/7</td>
<td>chIV</td>
<td>YDRCTy1-3::loxP-kanMX + YDRWTy1-4::loxP-hphNT1</td>
<td>Paracentric</td>
<td>108</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>5/7</td>
<td>chIV</td>
<td>YDRCTy1-2::loxP-kanMX + YDRWTy1-4::loxP-hphNT1</td>
<td>Paracentric</td>
<td>217</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td>12/13</td>
<td>chVII</td>
<td>YGRCTy1-3::loxP-kanMX + YGRWTy1-1::loxP-hphNT1</td>
<td>Paracentric</td>
<td>228</td>
<td>52</td>
<td>98.75</td>
</tr>
<tr>
<td>5/8</td>
<td>chIV</td>
<td>YDRCTy1-2::loxP-kanMX + YDRWTy1-5::loxP-hphNT1</td>
<td>Paracentric</td>
<td>328</td>
<td>95</td>
<td>92.9</td>
</tr>
<tr>
<td>25/24</td>
<td>chXIV</td>
<td>YNLWTy1-2::loxP-kanMX + YNLCTy1-1::loxP-hphNT1</td>
<td>Paracentric</td>
<td>428</td>
<td>48</td>
<td>94</td>
</tr>
<tr>
<td>4/7</td>
<td>chIV</td>
<td>YDRCTy1-1::loxP-kanMX + YDRWTy1-4::loxP-hphNT1</td>
<td>Paracentric</td>
<td>450</td>
<td>89</td>
<td>94</td>
</tr>
</tbody>
</table>

The strains column represents the names given to the engineered strains in the laboratory. The chromosome number column indicates the chromosome on which inversion was created. The 3rd column shows the Ty elements where the $\text{loxP-kanMX}$ and $\text{loxP-hphNT1}$ cassettes were inserted to generate inversions. The strains highlighted in blue were chosen for microarray analysis.
Table 3.2: List of engineered strains possessing pericentric inversion. The size, location of the inversion and percentage of spore viability is shown.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Chromosome number</th>
<th>Location of inversion and cassette insertion</th>
<th>Type of inversion</th>
<th>Size of inversion (kb)</th>
<th>Spore viability of inverted strains (%)</th>
<th>Spore viability of non-inverted strains (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21/22</td>
<td>chXIII</td>
<td>YMLWTy1-2::loxP-kanMX + YMRCTy1-3::loxP-hphNTI</td>
<td>Pericentric</td>
<td>167</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>20/22</td>
<td>chXIII</td>
<td>YMLWTy1-1::loxP-kanMX + YMRCTy1-3::loxP-hphNTI</td>
<td>Pericentric</td>
<td>173</td>
<td>85.6</td>
<td>95</td>
</tr>
<tr>
<td>21/23</td>
<td>chXIII</td>
<td>YMLWTy1-2::loxP-kanMX + YMRCTy1-4::loxP-hphNTI</td>
<td>Pericentric</td>
<td>182</td>
<td>90.42</td>
<td>98.33</td>
</tr>
<tr>
<td>20/23</td>
<td>chXIII</td>
<td>YMLWTy1-1::loxP-kanMX + YMRCTy1-4::loxP-hphNTI</td>
<td>Pericentric</td>
<td>189</td>
<td>90.42</td>
<td>96.67</td>
</tr>
<tr>
<td>29/28</td>
<td>chXVI</td>
<td>YPRCTy1-2::loxP-kanMX + YPLWTy1-1::loxP-hphNTI</td>
<td>Pericentric</td>
<td>754</td>
<td>48</td>
<td>94</td>
</tr>
</tbody>
</table>

The strains column represents the names given to the engineered strains in the laboratory. The chromosome number column indicates the chromosome on which inversion was created. The 3rd column shows the Ty elements where the *loxP-kanMX* and *loxP-hphNTI* cassettes were inserted to generate inversions. The strains highlighted in blue were chosen for microarray analysis.
Table 3.3: List of strains in which inversion was found to be lethal to the strain. The size of inversion is shown.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Chromosome number</th>
<th>Location of inversion and cassette insertion</th>
<th>Type of inversion</th>
<th>Size of inversion (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>31/29</td>
<td>chXVI</td>
<td>YPRWTy1-3::loxP-kanMX + YPRCTy1-2::loxP-hphNT1</td>
<td>Inversion was lethal</td>
<td>46</td>
</tr>
<tr>
<td>6/8</td>
<td>chIV</td>
<td>YDRCTy1-3::loxP-kanMX + YDRWTy1-5::loxP-hphNT1</td>
<td>Inversion was lethal</td>
<td>225</td>
</tr>
<tr>
<td>4/8</td>
<td>chIV</td>
<td>YDRCTy1-1::loxP-kanMX + YDRWTy1-5::loxP-hphNT1</td>
<td>Inversion was lethal</td>
<td>567</td>
</tr>
<tr>
<td>30/28</td>
<td>chXVI</td>
<td>YPRCTy1-4::loxP-kanMX + YPLWTy1-1::loxP-hphNT1</td>
<td>Inversion was lethal</td>
<td>800</td>
</tr>
</tbody>
</table>
Figure 3.2: Recombination hotspots in the inverted region of strain 29/28. The crossover (CO) hotspots were mapped to the inverted region of strain 29/28 (having inversion at YPLWTy1T1 and TPRCTy1T2) and are shown in the bottom track of the map. The black bar at the top of map indicates the region and size of inversion at chromosome XVI. Beginning and end of the black bar indicated inversion breakpoints. All the transcripts overlapping inversion are shown. The image was imported from University of California Santa Cruz (UCSC) genome browser [18] with hotspots data from Mancera et al. (2008) added as a custom track.
Number of recombination hotspots in all of constructed strains along with their percentage of spore viability and inversion length are shown in Table 3.4. Highlighted in bold red in Table 3.4 are the strains which had ≥ 2 recombination hotspots in their inverted region. In these strains, the inversions were either lethal or showed a 50% drop in spore viability. Four hotspots were found in strain 30/28 and inversion was observed to be lethal in it. Three hotspots were found in strain 25/24 and 29/28 having large size inversion of 428kb and 754kb with 48% drop in spore viability. Strain 12/13 also showed 50% drop in spore viability and had two recombination hotspots within the inverted region (Highlighted in bold red in Table 3.4). The strains with no significant change in spore viability had one or no hotspots in their inverted region of the chromosome.
Table 3.4: List of strains showing the size and type of the inversion, percentage of spore viability and number of recombination hotspots within the inverted region.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Chromosome number</th>
<th>Type of inversion</th>
<th>Size of inversion (kb)</th>
<th>Spore viability of inverted strains (%)</th>
<th>Spore viability of non-inverted strains (%)</th>
<th>Recombination Hotspots within the inverted region</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/13</td>
<td>chVII</td>
<td>Paracentric</td>
<td>26</td>
<td>86</td>
<td>94.93</td>
<td>0</td>
</tr>
<tr>
<td>6/7</td>
<td>chIV</td>
<td>Paracentric</td>
<td>108</td>
<td>95</td>
<td>97</td>
<td>0</td>
</tr>
<tr>
<td>5/7</td>
<td>chIV</td>
<td>Paracentric</td>
<td>217</td>
<td>94</td>
<td>97</td>
<td>1</td>
</tr>
<tr>
<td>12/13</td>
<td>chVII</td>
<td>Paracentric</td>
<td>228</td>
<td>52</td>
<td>98.75</td>
<td>2</td>
</tr>
<tr>
<td>5/8</td>
<td>chIV</td>
<td>Paracentric</td>
<td>328</td>
<td>95</td>
<td>92.9</td>
<td>1</td>
</tr>
<tr>
<td>25/24</td>
<td>chXIV</td>
<td>Paracentric</td>
<td>428</td>
<td>48</td>
<td>94</td>
<td>3</td>
</tr>
<tr>
<td>4/7</td>
<td>chIV</td>
<td>Paracentric</td>
<td>450</td>
<td>89</td>
<td>94</td>
<td>2</td>
</tr>
<tr>
<td>21/22</td>
<td>chXIII</td>
<td>Pericentric</td>
<td>167</td>
<td>80</td>
<td>92</td>
<td>0</td>
</tr>
<tr>
<td>20/22</td>
<td>chXIII</td>
<td>Pericentric</td>
<td>173</td>
<td>85.6</td>
<td>95</td>
<td>0</td>
</tr>
<tr>
<td>21/23</td>
<td>chXIII</td>
<td>Pericentric</td>
<td>182</td>
<td>90.42</td>
<td>98.33</td>
<td>0</td>
</tr>
<tr>
<td>20/23</td>
<td>chXIII</td>
<td>Pericentric</td>
<td>189</td>
<td>90.42</td>
<td>96.67</td>
<td>0</td>
</tr>
<tr>
<td>29/28</td>
<td>chXVI</td>
<td>Pericentric</td>
<td>754</td>
<td>48</td>
<td>94</td>
<td>3</td>
</tr>
<tr>
<td>31/29</td>
<td>chXVI</td>
<td>Inversion was lethal</td>
<td>46</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>6/8</td>
<td>chIV</td>
<td>Inversion was lethal</td>
<td>225</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>4/8</td>
<td>chIV</td>
<td>Inversion was lethal</td>
<td>567</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>30/28</td>
<td>chXVI</td>
<td>Inversion was lethal</td>
<td>800</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
</tbody>
</table>

The columns highlighted in red represent the strains which showed 50% drop in spore viability and possessed ≥ 2 recombination hotspots.
3.5.4 Fitness assay

To study the effect of inversions on strain fitness, we checked the growth rate of engineered strains in Nitrogen limited (N-limited) and Carbon limited (C-limited) medium. No difference in growth rate of inverted and non-inverted strains was observed in C-limited medium. However in N-limited medium one inverted strain (5/8) showed a decrease in growth rate in comparison to the non-inverted strains (Table 3.5).

Table 3.5: Summary of growth rate analysis in C-limited and N-limited media.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Growth rate in C-limited media</th>
<th>Growth rate in N-limited media</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inverted</td>
<td>Non-inverted</td>
</tr>
<tr>
<td>6/7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25/24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4/7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21/22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20/22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21/23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20/23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29/28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The red arrow indicates decrease in the growth rate and the blue arrow shows increase in the growth rate. The strains highlighted in blue were chosen for microarray analysis.

The fitness assays of the four strains (12/13, 5/8, 25/24 and 29/28) which were processed for microarray analysis are shown in Figure 3.3-3.6. The growth of these four strains was also checked in SD medium and no difference in growth was observed.
Figure 3.3: Fitness assays of strain 12/13 in different media. The growth of inverted and non-inverted strains was found to be same in N-limited medium (Panel A), C-limited medium (Panel B) and SD medium (Panel C). Each point represents the mean average of three technical repeats for three independent biological replicas. Error bars are at 95% confidence intervals.
Figure 3.4: Fitness assays of strain 5/8 in different media. The growth of inverted strain was found to be slightly lower than the non-inverted strain in N-limited medium (Panel A), however no difference in growth rate of inverted and non-inverted strains was observed in C-limited medium (Panel B) and SD medium (Panel C). Each point represents the mean average of three technical repeats for three independent biological replicas. Error bars are at 95% confidence intervals.
Figure 3.5: Fitness assays of strain 25/24 in different media. The growth of inverted and non-inverted strains was found to be same in N-limited medium (Panel A), C-limited medium (Panel B) and SD medium (Panel C). Each point represents the mean average of three technical repeats for three independent biological replicas. Error bars are at 95% confidence intervals.
Figure 3.6: Fitness assays of strain 29/28 in different media. The growth of inverted and non-inverted strains was found to be same in N-limited medium (Panel A), C-limited medium (Panel B) and SD medium (Panel C). Each point represents the mean average of three technical repeats for three independent biological replicas. Error bars are at 95% confidence intervals.
Based on our spore viability and fitness data of 16 engineered strains, we selected 4 inverted and their respective non-inverted strains to determine the effect of inversions on global gene expression. Three of the selected strains (12/13, 25/24 and 29/28) showed a 50% decline in spore viability and one (5/8) had 95% spore viability equivalent to its respective non-inverted partner (highlighted in blue Table 3.1 & 3.2).

3.5.5 Whole transcriptome analysis

The whole transcriptome analysis was done on engineered inverted and non-inverted individuals. Cells were collected at exponential phase (OD$_{595\text{nm}}$=0.5), RNA was extracted and expression was analysed by performing yeast Affimatrix array.

3.5.5.1 Principal component analysis (PCA)

Before comparing the expression of genes between inverted and non-inverted strains, the PCA was performed on all four strains harbouring genomic inversions to understand the relationship between the samples by identifying the similarities and differences in the data set. Two components were applied on PCA. The first principal component (PC#1) deals with the actual biology of the experiment by distinguishing between the treatments and separating the inverted individuals from the control non-inverted ones. The second principal component (PC#2) represents the overall variance between the three biological replicates for each of the inverted and non-inverted samples.

The PCA results of the microarray data for strain 12/13 and 25/24 showed that the three replicates of inverted and non-inverted strains cluster separately on component 1 validating our ability to separate the inverted individuals from the non-inverted ones and revealing distinct expression profiles for each. In strain 12/13 the three biological replicates showed similar pattern of clustering on component 2 indicating that there isn’t any significant difference among the biological replicates. Similar pattern of clustering was also observed among the biological replicates of strain 25/24 (Figure 3.7).
The PCA of strain 29/28 showed that there was no major variance among the three biological replicates of inverted and non-inverted individuals as they clustered separately on component 2. However the PCA of strain 5/8 did not show strong correlation among the three biological replicates of inverted and non-inverted strain (Figure 3.8).

**Figure 3.7:** Principal component analysis (PCA) of microarray data set. Overall variation for the three inverted replicas (blue circles) and three non-inverted control replicas, where each spot represents an individual array. **Panel A** is PCA analysis for strain 12/13 and **Panel B** is PCA analysis for strain 25/24.
Figure 3.8: Principal component analysis (PCA) of microarray data set. Overall variation for the three inverted replicas (blue circles) and three non-inverted control replicas, where each spot represents an individual array. Panel A is PCA analysis for strain 5/8 and Panel B is PCA analysis for strain 29/28.

3.5.5.2 Global expression profile of inverted and non-inverted strains

For comparisons of the global expression profile between engineered inverted and non-inverted candidates, the differentially expressed genes were identified using a
combination of statistical score and fold change. The change in gene expression was considered to be significant if the q-value was ≤ 0.2.

Our expression data showed that strain 25/24 possessed highest number (753 genes) of differentially expressed genes (either up or down regulated) by 1.5-4.0 folds. The results from strain 12/13 identified 425 genes that were differentially expressed (either up or down regulated) by 1.5-6.0 fold. The strain 5/8 showed only 26 genes that were differentially expressed (either up or down regulated) by very small fold change i.e. 1.5-1.8 fold. We did not observe any significant differential expression between inverted and non-inverted individuals for strain 29/28. The top ten genes in all 3 strains with significant change in expression between inverted and non-inverted strains are shown in Table 3.6-3.8.

Table 3.6. List of 10 genes in strain 25/24 which were highly changing expression

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>Gene Symbol</th>
<th>Chromosomal Location</th>
<th>Fold-Change (Inv/Cntl)</th>
<th>q.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAR028W</td>
<td>---</td>
<td>chr1</td>
<td>-1.63</td>
<td>0.02</td>
</tr>
<tr>
<td>YAL037W</td>
<td>---</td>
<td>chr1</td>
<td>-1.53</td>
<td>0.02</td>
</tr>
<tr>
<td>YBL060W</td>
<td>YEL1</td>
<td>chr2</td>
<td>1.85</td>
<td>0.02</td>
</tr>
<tr>
<td>YBL035C</td>
<td>POL12</td>
<td>chr2</td>
<td>1.38</td>
<td>0.02</td>
</tr>
<tr>
<td>YDR261W</td>
<td>---</td>
<td>chr10</td>
<td>1.6</td>
<td>0.02</td>
</tr>
<tr>
<td>YBR088C</td>
<td>POL30</td>
<td>chr2</td>
<td>2.09</td>
<td>0.02</td>
</tr>
<tr>
<td>YBR223C</td>
<td>TDP1</td>
<td>chr2</td>
<td>-1.64</td>
<td>0.02</td>
</tr>
<tr>
<td>YDR527W</td>
<td>RBA50</td>
<td>chr4</td>
<td>1.63</td>
<td>0.02</td>
</tr>
<tr>
<td>YDL101C</td>
<td>DUN1</td>
<td>chr4</td>
<td>1.87</td>
<td>0.02</td>
</tr>
<tr>
<td>YDL003W</td>
<td>MCD1</td>
<td>chr4</td>
<td>1.91</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Microarray data showing genes with significant change in expression (q-value) with calculated fold change between the inverted and non-inverted strains.

125
Table 3.7. List of 10 genes in strain 12/13 which were highly changing expression

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>Gene Symbol</th>
<th>Chromosomal Location</th>
<th>Fold Change (Inv/Cntl)</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>YER054C</td>
<td><em>GIP2</em></td>
<td>chr5</td>
<td>-5.11</td>
<td>0.12</td>
</tr>
<tr>
<td>YDR085C</td>
<td><em>AFR1</em></td>
<td>chr4</td>
<td>-2.26</td>
<td>0.12</td>
</tr>
<tr>
<td>YBL086C</td>
<td>---</td>
<td>chr2</td>
<td>-2.00</td>
<td>0.12</td>
</tr>
<tr>
<td>YMR081C</td>
<td><em>ISF1</em></td>
<td>chr13</td>
<td>-3.63</td>
<td>0.13</td>
</tr>
<tr>
<td>YML128C</td>
<td><em>MSC1</em></td>
<td>chr13</td>
<td>-3.15</td>
<td>0.13</td>
</tr>
<tr>
<td>YGL229C</td>
<td><em>SAP4</em></td>
<td>chr7</td>
<td>-2.35</td>
<td>0.13</td>
</tr>
<tr>
<td>YLR258W</td>
<td><em>GSY2</em></td>
<td>chr12</td>
<td>-2.18</td>
<td>0.13</td>
</tr>
<tr>
<td>YGL183C</td>
<td><em>MND1</em></td>
<td>chr7</td>
<td>-1.53</td>
<td>0.13</td>
</tr>
<tr>
<td>YHR153C</td>
<td><em>SPO16</em></td>
<td>chr8</td>
<td>1.93</td>
<td>0.17</td>
</tr>
<tr>
<td>YDL154W</td>
<td><em>MSH5</em></td>
<td>chr4</td>
<td>-1.47</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Microarray data showing genes with significant change in expression (q-value) with calculated fold change between the inverted and non-inverted strains.

Table 3.8. List of 10 genes in strain 5/8 which were highly changing expression

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>Gene Symbol</th>
<th>Chromosomal Location</th>
<th>Fold Change (Inv/Cntl)</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>YNL144C</td>
<td>---</td>
<td>chr14</td>
<td>-1.6837</td>
<td>0.06</td>
</tr>
<tr>
<td>YOR273C</td>
<td><em>TPO4</em></td>
<td>chr15</td>
<td>-1.5593</td>
<td>0.06</td>
</tr>
<tr>
<td>YLR327C</td>
<td><em>TMA10</em></td>
<td>chr12</td>
<td>-1.7194</td>
<td>0.08</td>
</tr>
<tr>
<td>YOL155W</td>
<td>---</td>
<td>chr15</td>
<td>-1.5942</td>
<td>0.08</td>
</tr>
<tr>
<td>YFR015C</td>
<td><em>GSY1</em></td>
<td>chr6</td>
<td>-1.6883</td>
<td>0.10</td>
</tr>
<tr>
<td>YHL021C</td>
<td><em>AIM17</em></td>
<td>chr8</td>
<td>-1.4803</td>
<td>0.10</td>
</tr>
<tr>
<td>YPL014W</td>
<td>---</td>
<td>chr16</td>
<td>-1.706</td>
<td>0.10</td>
</tr>
<tr>
<td>YDR277C</td>
<td><em>MTH1</em></td>
<td>chr4</td>
<td>-1.4739</td>
<td>0.11</td>
</tr>
<tr>
<td>YDR254W</td>
<td><em>CHL4</em></td>
<td>chr4</td>
<td>-1.4597</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Microarray data showing genes with significant change in expression (q-value) with calculated fold change between the inverted and non-inverted strains.
In order to see if there was any particular class of genes affected by inversion, we performed Gene Ontology (GO) analysis [19]. GO enrichment analysis was performed using the web interface for High-Throughput GoMiner programme (http://discover.nci.nih.gov/gominer/htgm.jsp) on all the genes that were significantly differentially expressed in strains 12/13, 25/24 and 5/8. We observed that in all three strains, inversions affect the expression of genes belonging to different categories. In strains 12/13 and 25/24 we found certain GO terms enriched in two categories biological processes and molecular function. However, we did not find any significant enrichment of any particular class for strain 5/8 both for biological processes and molecular function. The GO term was classified as significant if the false discovery rate (FDR) was ≤ 0.05 [20]. GO terms enriched in biological processes and molecular function for strains 12/13 and 25/24 are shown in Table 3.9-3.12.

Table 3.9: GO categories for Biological Function with FDR ≤ 0.05 for strain 12/13

<table>
<thead>
<tr>
<th>GO Category</th>
<th>Description</th>
<th>False Discovery Rate (FDR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0007039</td>
<td>Vacuolar protein catabolic process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0009628</td>
<td>Response to abiotic stimulus</td>
<td>0.0001</td>
</tr>
<tr>
<td>GO:0009266</td>
<td>Response to temperature stimulus</td>
<td>0.00018</td>
</tr>
<tr>
<td>GO:0006112</td>
<td>Energy reserve metabolic process</td>
<td>0.0003</td>
</tr>
<tr>
<td>GO:0034605</td>
<td>Cellular response to heat</td>
<td>0.00037</td>
</tr>
<tr>
<td>GO:0009408</td>
<td>Response to heat</td>
<td>0.001</td>
</tr>
<tr>
<td>GO:0005978</td>
<td>Glycogen biosynthetic process</td>
<td>0.0013</td>
</tr>
<tr>
<td>GO:0034637</td>
<td>Cellular carbohydrate biosynthetic process</td>
<td>0.002</td>
</tr>
<tr>
<td>GO:0016051</td>
<td>Carbohydrate biosynthetic process</td>
<td>0.0027</td>
</tr>
<tr>
<td>GO:0005991</td>
<td>Trehalose metabolic process</td>
<td>0.0031</td>
</tr>
<tr>
<td>GO:0030163</td>
<td>Protein catabolic process</td>
<td>0.01</td>
</tr>
<tr>
<td>GO:0006914</td>
<td>Autophagy</td>
<td>0.015</td>
</tr>
<tr>
<td>GO:0044264</td>
<td>Cellular polysaccharide metabolic process</td>
<td>0.017</td>
</tr>
<tr>
<td>GO:0009250</td>
<td>Glucan biosynthetic process</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Table 3.10: GO categories for Molecular Function with FDR ≤ 0.05 for strain 12/13

<table>
<thead>
<tr>
<th>GO Category</th>
<th>Description</th>
<th>False Discovery Rate (FDR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0019203</td>
<td>Carbohydrate phosphatase activity</td>
<td>0</td>
</tr>
<tr>
<td>GO:0008194</td>
<td>UDP-glycosyltransferase activity</td>
<td>0.027</td>
</tr>
<tr>
<td>GO:0003825</td>
<td>Alpha-trehalose-phosphate synthase activity</td>
<td>0.028889</td>
</tr>
<tr>
<td>GO:0046527</td>
<td>Glucosyltransferase activity</td>
<td>0.03</td>
</tr>
<tr>
<td>GO:0015294</td>
<td>Solute:cation symporter activity</td>
<td>0.03125</td>
</tr>
<tr>
<td>GO:0004805</td>
<td>Trehalose-phosphatase activity</td>
<td>0.031429</td>
</tr>
<tr>
<td>GO:0030234</td>
<td>Enzyme regulator activity</td>
<td>0.031667</td>
</tr>
<tr>
<td>GO:0016758</td>
<td>Transferase activity transferring hexosyl groups</td>
<td>0.0325</td>
</tr>
<tr>
<td>GO:0019208</td>
<td>Phosphatase regulator activity</td>
<td>0.036364</td>
</tr>
<tr>
<td>GO:0019888</td>
<td>Protein phosphatase regulator activity</td>
<td>0.036667</td>
</tr>
<tr>
<td>GO:0003873</td>
<td>6-phosphofructo-2-kinase activity</td>
<td>0.0375</td>
</tr>
<tr>
<td>GO:0015293</td>
<td>Symporter activity</td>
<td>0.04</td>
</tr>
<tr>
<td>GO Category</td>
<td>Description</td>
<td>False Discovery Rate (FDR)</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>GO:0031323</td>
<td>Regulation of cellular metabolic process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0065007</td>
<td>Biological regulation</td>
<td>0</td>
</tr>
<tr>
<td>GO:0080090</td>
<td>Regulation of primary metabolic process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0060255</td>
<td>Regulation of macromolecule metabolic process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0050789</td>
<td>Regulation of biological process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0050794</td>
<td>Regulation of cellular process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0019222</td>
<td>Regulation of metabolic process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0019219</td>
<td>Regulation of nucleoside nucleotide and nucleic acid metabolic process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0051171</td>
<td>Regulation of nitrogen compound metabolic process</td>
<td>0</td>
</tr>
<tr>
<td>GO:0006350</td>
<td>Transcription</td>
<td>0</td>
</tr>
<tr>
<td>GO:0010468</td>
<td>Regulation of gene expression</td>
<td>0.000526</td>
</tr>
<tr>
<td>GO:0006139</td>
<td>Nucleobase nucleoside nucleotide and nucleic acid metabolic process</td>
<td>0.0005</td>
</tr>
<tr>
<td>GO:0009889</td>
<td>Regulation of biosynthetic process</td>
<td>0.0009</td>
</tr>
<tr>
<td>GO:0031326</td>
<td>Regulation of cellular biosynthetic process</td>
<td>0.0013</td>
</tr>
</tbody>
</table>
Table 3.12: GO categories for Molecular Function with FDR ≤ 0.05 for strain 25/24

<table>
<thead>
<tr>
<th>GO Category</th>
<th>Description</th>
<th>False Discovery Rate (FDR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0030515</td>
<td>SnoRna binding</td>
<td>0</td>
</tr>
<tr>
<td>GO:0003724</td>
<td>Rna helicase activity</td>
<td>0.0001</td>
</tr>
<tr>
<td>GO:0030528</td>
<td>Transcription regulator activity</td>
<td>0.00013</td>
</tr>
<tr>
<td>GO:0004004</td>
<td>Atp-dependent rna helicase activity</td>
<td>0.00016</td>
</tr>
<tr>
<td>GO:0008186</td>
<td>Rna-dependent atpase activity</td>
<td>0.0002</td>
</tr>
<tr>
<td>GO:0043566</td>
<td>Structure-specific dna binding</td>
<td>0.00032</td>
</tr>
<tr>
<td>GO:0042577</td>
<td>Lipid phosphatase activity</td>
<td>0.00033</td>
</tr>
<tr>
<td>GO:0016563</td>
<td>Transcription activator activity</td>
<td>0.00035</td>
</tr>
<tr>
<td>GO:0022890</td>
<td>Inorganic cation transmembrane transporter activity</td>
<td>0.00037</td>
</tr>
<tr>
<td>GO:0008175</td>
<td>Trna methyltransferase activity</td>
<td>0.002</td>
</tr>
<tr>
<td>GO:0043495</td>
<td>Protein anchor</td>
<td>0.0025</td>
</tr>
<tr>
<td>GO:0003709</td>
<td>RNA polymerase III transcription factor activity</td>
<td>0.0027</td>
</tr>
<tr>
<td>GO:0016462</td>
<td>Pyrophosphatase activity</td>
<td>0.01</td>
</tr>
<tr>
<td>GO:0016817</td>
<td>Hydrolase activity acting on acid anhydrides</td>
<td>0.012</td>
</tr>
<tr>
<td>GO:0016818</td>
<td>Hydrolase activity acting on acid anhydrides</td>
<td>0.012</td>
</tr>
</tbody>
</table>

We also analyzed the expression of genes within the inversion breakpoints, in the immediate surrounding areas of the inversion (10 genes on either side of the breakpoints) and all other genes in the genome that do not lie within these two categories. This was conducted by comparing the significant fold change expression between the mean inverted versus the non-inverted strains.

It was observed that the number of differentially expressed genes within the inversion, in the immediate surrounding or in the other genes within the genome was not significant for strains 5/8, 12/13 and 29/28. However, strain 25/24 showed a significant number of genes with changed expression not only within the inversion but also in the surrounding of inversion breakpoints as well as in other genes of the genome (Figure 3.9).
Figure 3.9: Comparison of global expression profile in engineered strains. Bar graphs for strain 29/28 (panel A), 5/8 (panel B) and 12/13 (panel C) and 25/24 (panel D) representing changes in gene expression for the genes within the inversion breakpoints, 10 genes on the immediate surroundings of the breakpoints and all the other genes in the genome. Percentage of genes significantly changing expression is represented by blue bars and non-significant genes by red bar.
List of genes significantly changing expression and located in the vicinity of the inverted region for strain 25/24 are shown in Table 3.13. It was observed that most of the genes located on the right side of the breakpoint were down-regulated with exception of one gene \( \text{POP3} \). Similarly, the genes located on left side of the breakpoint were up-regulated with exception of \( \text{POR1} \) and \( \text{MTQ1} \). It is possible that inversion has moved some enhancer elements on the left side of the breakpoint which therefore resulted in positive regulation of the genes located in that region.

Table 3.13: List of genes located in the vicinity of breakpoints in strain 25/24

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Gene Symbol</th>
<th>Fold-Change (Inv vs. NI)</th>
<th>q.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>YNL055C</td>
<td>( \text{POR1} /// \text{POR2} )</td>
<td>-1.06518</td>
<td>0.156629</td>
</tr>
<tr>
<td>YNL056W</td>
<td>( \text{OCA2} )</td>
<td>1.10763</td>
<td>0.178809</td>
</tr>
<tr>
<td>YNL058C</td>
<td>-</td>
<td>-1.20544</td>
<td>0.06801</td>
</tr>
<tr>
<td>YNL059C</td>
<td>( \text{ARP5} )</td>
<td>1.4864</td>
<td>0.065693</td>
</tr>
<tr>
<td>YNL061W</td>
<td>( \text{NOP2} )</td>
<td>1.68019</td>
<td>0.041329</td>
</tr>
<tr>
<td>YNL062C</td>
<td>( \text{GCD10} )</td>
<td>1.62596</td>
<td>0.05729</td>
</tr>
<tr>
<td>YNL063W</td>
<td>( \text{MTQ1} )</td>
<td>-1.39527</td>
<td>0.026852</td>
</tr>
<tr>
<td>YNL064C</td>
<td>( \text{YDJ1} )</td>
<td>1.30816</td>
<td>0.147473</td>
</tr>
<tr>
<td>YNL275W</td>
<td>( \text{BOR1} )</td>
<td>-1.12439</td>
<td>0.171483</td>
</tr>
<tr>
<td>YNL277W</td>
<td>( \text{MET2} )</td>
<td>-1.49674</td>
<td>0.042813</td>
</tr>
<tr>
<td>YNL278W</td>
<td>( \text{CAF120} )</td>
<td>-1.22818</td>
<td>0.09793</td>
</tr>
<tr>
<td>YNL280C</td>
<td>( \text{ERG24} )</td>
<td>-1.10546</td>
<td>0.124835</td>
</tr>
<tr>
<td>YNL282W</td>
<td>( \text{POP3} )</td>
<td>1.27647</td>
<td>0.183329</td>
</tr>
<tr>
<td>YNL283C</td>
<td>( \text{WSC2} )</td>
<td>-1.27817</td>
<td>0.053328</td>
</tr>
</tbody>
</table>

Microarray data showing genes with significant change in expression (q-value) with calculated fold change between the inverted and non-inverted strains.

The \( \text{ARP5} \) gene highlighted in yellow in Table 3.13 was found to be up-regulated by 1.5 fold in inverted strains relative to the non-inverted one. Arp5 is an important actin related protein and is a component of INO80 chromatin remodelling complex (Figure 3.10). This complex regulates the transcription of approximately 20% of the yeast genes. It remodels the nucleosome on the promoter to make DNA accessible for transcription. Since the expression of \( \text{ARP5} \) was being changed in the inverted strain, it
might be one of the reasons for observing the massive change in gene expression in strain 25/24.

![Diagram of the INO80 complex](image)

**Figure 3.10: Structure of the INO80 complex remodelling complex.** INO80 complex consists of inositol requiring protein Ino80, actin Act1, actin related protein Arp4 and Arp5, TATA binding protein Taf14 and other proteins Ies1, Ies2, Ies3, Ies4, Ies5, Ies6, Rvb1 and Rvb2 (Figure taken from Bao and Shen (2010)).

In all four strains, different size inversion was created on a different chromosome. We observed the number of genes with altered expression on all the chromosomes. It was found that, in all four strains, the differentially expressed genes were distributed throughout the genome rather than on the chromosome having inversion or its neighbouring chromosome. This is graphically represented in Figure 3.11 & 3.12.
Figure 3.11: Total number of genes with changed expression on each chromosome. Number of genes changing expression on each chromosome relative to the total number of genes located on the chromosome was analysed for strains 29/28 (red) and 5/8 (yellow).
Figure 3.12: Total number of genes with changed expression on each chromosome. Number of genes changing expression on each chromosome relative to the total number of genes located on the chromosome was analysed for strains 12/13 (green) and 25/24 (blue).
3.6 Discussion

In this piece of research, we studied the role of small and large size pericentric and paracentric inversions on crossing over, strain fitness and global gene expression. Yeast strains carrying different sizes of pericentric and paracentric inversions were previously constructed in the Delneri lab (Data unpublished) and were available for use. These inversions were made between yeast transposable elements “Ty1” in different chromosomes. Because of their mobility and repetitive nature they play a very important role in recombination leading to several different types of chromosomal rearrangements. Since they are more likely to cause a chromosomal rearrangement in nature, they were chosen for the construction of chromosomal inversions. All inversions were made in laboratory yeast strain BY4741. A total of seven strains possessing paracentric inversions and five strains possessing pericentric inversions were studied (Table 3.1 & 3.2).

Chromosomal inversion and the length of chromosome being affected can effect the crossing over in a chromosome. Research on chromosomal inversions in Drosophila species report that unsuccessful inversions are usually of small size whereas successful ones are mostly the medium sized inversions. The detrimental effects of an inversion increases with the increasing length of chromosomal inversion due to increased chances of double crossovers. Therefore, natural selection will favour those inversions which are of intermediate size. However, selection does not only act on physical length of inversion but also on recombination length [19, 20]. Caceres and his co-authors in 1998 studied the relationship between inversion size and recombination map length in Diptera species and they observed that physical length of successful inversions is different among different species [21]. We also observed different effects of different size inversions on our strains. We found that in some strains (strain 4/7) large size inversions have no significant effect on crossing over giving a spore viability of 89-94%. Whereas in others the large size inversion was observed to be either detrimental to the strain viability (strain 4/8 and 30/28) or was reducing the spore viability to 50% (strain 25/24 and 29/28).
Our results showed that small size paracentric and pericentric inversions did not have a significant change in the spore viability of inverted strains in comparison to the non-inverted strains. Some of the strains (strain 12/13) carrying medium size inversion showed a 50% drop in the spore viability relative to non-inverted control strains which had a spore viability of 92-98% (Table 3.1). However, other strains (5/7 and 5/8) with medium size inversions possessed 95-97% spore viability both in inverted as well as non-inverted individuals (Table 3.1 & 3.2).

In four of our strains (strain 31/29, 6/8, 4/8 and 30/28), we were unable to find any inverted individuals. In these strain all the colonies tested by diagnostic colony PCR were non-inverted. To check if the inversion was lethal or it was simply not occurring we extracted the genomic DNA of these strains immediately after cre recombinase induction before plating them out on YPD medium. It was found that inversion was happening in the genome however it was lethal to the strain and therefore no inversion was detected after plating the induced cultures on YPD medium. It was observed that inversions in these strains was disrupting some essential genes required for the mismatch repair in mitosis and meiosis, RNA splicing and processing and genes required for sister chromatid cohesion. It is also possible that some important transcription regulatory factors might have been disrupted by inversion, thereby making the inversion lethal to the strain.

It has previously been reported that in individuals having successful inversions, the recombination map length is inversely correlated to physical size of inversion. As the recombination map length increases the length of successful inversions decreases. However no such relationship was observed between the physical length of unsuccessful inversions and recombination map length [23].

Recombination not only has positive effect on the progeny by increasing the genetic diversity but it can also have detrimental effects by separating a favourable allele combination and causing chromosomal rearrangements that are lethal. It has been shown that crossing over in yeast and other eukaryotes is more likely to occur at regions of chromosomes having recombination hotspots. These are the regions on a chromosome where recombination event mostly occurs [22]. We observed that all the
strains (11/13, 6/7, 5/7, 5/8, 21/22, 20/22, 21/23, 20/23, 31/29 and 6/8) which had one or none recombination hotspot within the inverted region possessed 85-95% spore viability in both inverted and non-inverted strains. Whereas the strains which possessed ≥ 2 recombination hotspots within the inverted region showed either 50% drop in spore viability or the inversion was lethal to the strain with exception of strain 4/7. Strain 4/7 was the only strain that had 2 recombination hotspots but showed no significant change in spore viability (Table 3.1). Our data showed that in most cases the higher the number of recombination hotspots within the inverted region the more harmful it is for the strain. It can be one of the reasons for seeing a massive drop in spore viability in some of our strains.

Gene expression is affected by various factors and expression of a gene is not only controlled by its promoters and transcription factors but also by other factors such as chromatin remodelling. Since many of these regulatory factors are located far away on the chromosomes altering the expression of more than one gene, it is therefore of interest to study the gene expression of not only the gene undergoing modification but also of surrounding genes [25].

The study of gene expression in strains possessing paracentric and pericentric inversions indicated high number of differentially expressed genes i.e. 753 in strain 25/24 and 425 in strain 12/13. Functional interpretation of these genes performed using High-Throughput GoMiner programme provided the list of statistically most represented GO terms. The analysis showed that in both strains there was enrichment of different GO terms for biological processes and molecular function. It was found that strain 25/24 which showed significant change in global expression profile had enrichment of GO molecular function terms like, “transcription activator activity”, “transcription regulator activity”, “RNA polymerase III transcription factor activity” and biological function terms like “transcription regulation”, “regulation of biological functions” and “regulation of gene expression”. These terms all indicate that inversion in this strain caused more disruption of transcription factors as compared to the strains 12/13, 5/8 and 29/26 thereby changing the global expression profile in this particular strain.
It has previously been reported that genomic neighbourhoods and genome position also regulates the gene expression and plays a crucial role in genome evolution [26]. This was first studied in Drosophila [27] and later in Mouse [28, 29] and human genome [30]. Flagfeldt et al. (2009) studied the impact of gene position and genomic neighbourhood on gene regulation in *Saccharomyces cerevisiae*. They studied the lacZ expression level by inserting the lacZ constructs at different sites in the genome. They observed that at some of the sites which were in close proximity to autonomously replicating sequence (ARS) the lacZ expression was very high whereas at certain other sites which were close to the telomeres, expression level was reduced [31].

Our expression data for strains 29/28, 5/8, 12/13 and 25/24 showed that chromosomal inversion not always change the expression of the genes. The comparison of percentage of genes changing expression in all strains to the percentage of genes not changing expression showed that 100% genes in strain 29/28 were not having any significant change in expression. Only 2% genes in strain 5/8 and 20% in strain 12/13 were found to be changing expression and these genes were either located within the inversion or far away from the breakpoint (Figure 3.9). In all the three strains, we didn’t observe any significant change in the expression of the genes present in the neighbourhood of inversion breakpoints which is in accordance to the Meadow’s study done on testis gene in *Drosophila* species. Meadow et al. studied the impact of neighbourhood continuity on testis gene expression in engineered inverted and the non-inverted Drosophila individuals disrupting the neighbourhood genes. They found no change in the expression of neighbourhood genes in the inverted individuals proving that neighbourhood organisation is not always a major contributor to gene expression [32].

However, one of our strains (25/24) showed strikingly high number of genes changing expression. These genes were not only located within the inverted region but also in the surroundings of the breakpoint and as well as in all other parts of the genome. It was observed that 75% of the genes located in immediate vicinity of the breakpoint were changing expression and this number was significantly higher than the number of genes changing expression within in the inverted region (Figure 3.9). This supports the idea that inversions not only change the expression of the genes solely present in the
inverted region but also significantly alters the expression of the neighbouring genes and genes located elsewhere in the genome.

We also observed that one of the genes involved in chromatin remodelling complex “ARP5” was being up regulated in the inverted strain 25/24. ARP5 is a part of INO80 complex which play an important role in DNA replication, transcriptional regulation and DNA repair [33]. We found that ARP5 was not the only gene in this complex changing the expression but other INO80 complex genes i.e. ARP4, IES2, IES3, IES4, RVB1, RVB2 and TAF14 all were showing a significant change in expression. In the inverted strain 25/24, all these genes were 1.2-1.5 fold up-regulated. INO80 complex is known to regulate the transcription of PHO5 (phosphate responsive gene) and INO1 (inositol responsive gene) [34, 35]. Our microarray data showed that both these genes were being significantly down-regulated by 1.1 fold in the inverted individuals relative to the non-inverted ones. Since inversion in this particular strain (25/24) was changing the expression of most of the INO80 complex genes therefore the expression of other genes which are being regulated by INO80 complex was also disrupted.

3.7 References


CHAPTER 4 (Paper 3)

IMPACT OF CHROMOSOMAL TRANSLOCATIONS ON S. PARADOXUS AND S. CARIOCANUS SPECIATION
4. Impact of chromosomal translocations on *S. paradoxus* and *S. cariocanus* speciation

Samina Naseeb and Daniela Delneri

To be submitted to FEMS Yeast Research

4.1 Foreword

This chapter is laid out in the format of FEMS Yeast Research. SN performed the experimental work, analyzed the results and wrote the manuscript, DD designed and supervised the project, and reviewed the manuscript.

4.2 Abstract

The *Saccharomyces* ‘sensu stricto’ species are reproductively isolated and possess post-zygotic barrier. Sequence divergence is one of the mechanisms for post-zygotic isolation among the species. *S. cariocanus* and *S. paradoxus* are reproductively isolated and show high sequence identity. The two species were recently thought to be the same species. However, these two species differ in their genome by the presence of four reciprocal translocations, therefore the sterility of hybrids between *S. paradoxus* and *S. cariocanus* can be due to chromosomal rearrangements. The aim of this study is to understand whether the translocations, rather than the sequence divergence, are responsible for the reproductive isolation of these two species. It is known that one reciprocal translocation reduces the spore viability to 50% therefore the presence of four translocations will reduce the spore viability of *S. cariocanus* × *S. paradoxus* hybrids to 6.25% and by restoring one translocation the spore viability should increase to 12.5%. We engineered *S. paradoxus* strains N-17 and N-44 possessing one translocation on chrIX/chrXV and crossed the translocated as well as the non-translocated strains with *S. cariocanus*. The translocated hybrids gave a spore viability of 2.7% which was slightly higher than the non-translocated hybrids giving a spore viability of 1.7%. This shows that isolation between two species can be due to
translocations and can be further confirmed by restoring the other three translocations and testing the spore viability.

Keywords

_Saccharomyces paradoxus, Saccharomyces cariocanus_, speciation, Cre/loxP, reciprocal translocation

4.3 Introduction

Speciation has always been a hot topic of discussion for the evolutionary geneticists and it has been defined in different ways. According to the biological species concept the species are believed to be a group of interbreeding individuals reproductively isolated from other such groups. The phylogenetic species concept is based on the evolutionary history and includes the group whose members are descended from a common ancestor possessing a combination of certain defining, or derived, traits. Speciation in yeast occurs by one primary mechanism which is accumulation of sequence differences (sequence divergence) which impair recombination. Chromosomal rearrangements can lead to the formation of inviable spores due to chromosomal missegregation (Fischer et al., 2000). In 1998 Ryu and his co-authors detected two reciprocal translocations at duplicated _RPL2_ loci between _S. cerevisiae_ and _S. bayanus_ (Ryu et al., 1998). In 2000 Fischer et al. detected one non-reciprocal translocation and three reciprocal translocation in _S. bayanus_, one reciprocal translocation in _S. miakatae_ and four reciprocal translocations in _S. cariocanus_. There are six species and one natural hybrid included in _Saccharomyces ‘sensu stricto’_ group, _S. cerevisiae, S. paradoxus, S. cariocanus, S. bayanus, S. kudriavzevii, S. mikatae_ and _S. pastorianus_ (a hybrid species formed from a cross between _S. cerevisiae_ and _S. bayanus_). Species within the _sensu stricto_ group possess postzygotic barriers and therefore result in ≤ 1% spore viability. It has been shown that these postzygotic barriers are due to the sequence divergences. A previous study done on sequence divergence of _sensu stricto_ group species showed relationship between sequence diversity and reproductive isolation (Liti et al., 2006). It was also shown that species _S. cariocanus_ and _S. paradoxus_, which are reproductively isolated, showed no significant
sequence divergence (Liti et al., 2009). However, these two species differ in their genome for the presence of four translocations (Fischer et al., 2000).

The aim of this project is to see whether the translocations, rather than the sequence divergence, are responsible for the reproductive isolation of these two species. To study this we decided to construct a chromosomal translocation in three strains of S. paradoxus (N-17, N-44 and YPS138) mimicking S. cariocanus translocation at chromosome IX and chromosome XV using the cre-loxP mechanism described previously (Delneri et al., 2003). The engineered translocated (N-17-T1 & N-44-T1) and control non-translocated (N-17-NT1 & N-44-NT1) strains were then crossed with the wild type S. cariocanus. It was hypothesized that if translocations are causing reproductive isolation then undoing one translocation should raise the spore viability from 6.25% to 12.5%. We showed that by undoing one out of four translocations of S. cariocanus in S. paradoxus there is a small but significant increase on the spore viability of translocated strains. Therefore the isolation between the two species can be due to the presence of translocation.

4.4 Materials and methods

4.4.1 Strains and Media

The yeast strains used in this study were S. cariocanus and S. paradoxus The S. paradoxus strains used were N-44 (Far Eastern), N-17(European) and YPS138(North American) and were obtained from Liti lab (Cubillos et al., 2009). The strains S.par-N44-T1 × S. car, S.par-N44-NT1 × S. car, S.par-N17-T1 × S. car and S.par-N17-NT1 × S. car were constructed in this study. All strains were maintained on YPD medium containing 2% (w/v) Bacto-yeast extract, 1% (w/v) pepton and 2% (w/v) glucose. The translocants were selected on glucose minimal medium (SD) containing 0.67% (w/v) yeast nitrogen base (Difco), 0.5% (w/v) ammonium sulphate and 2% (w/v) glucose (amino acids were added at a concentration of 20 mg/l) containing the desired antibiotics i.e. 300 µg/ml geneticin (GibcoBRL).
4.4.2 Oligonucleotides

Gene sequences for *S. paradoxus* were obtained from SGRP and PCR primers were designed using the Primer3 programme. The BLAST tool of SGRP was used to check the specificity of each primer. The oligonucleotide sequences are provided in appendix (Table A5 & A9).

4.4.3 Plasmid DNA Extraction and Restriction Digestion

QIAprep spin miniprep kit was used to extract the plasmid DNA following the manufacturer's guidelines (QIAgen catalogue no. 27104). The extracted DNA was digested by specific restriction enzymes. The reaction mixture consisted of 10 units of the restriction enzyme, 100-200 ng of plasmid DNA and an enzyme specific buffer. The final volume was adjusted to 20 µl by addition of sterile dH2O. The reaction mixture was incubated at 37°C for 1 hour and DNA was visualised under UV light immediately after agarose gel electrophoresis.

4.4.4 Transformation of *Saccharomyces paradoxus* by Lithium Acetate protocol

The *Saccharomyces paradoxus* strain was transformed using the high efficiency yeast transformation protocol (Gietz and Schiestl, 2007) with some modifications. The transformation mix consisted of 240 µl of 50% (w/v) PEG3350, 36 µl of 0.1M lithium acetate, 25 µl of single-stranded carrier DNA (2mg/ml) and 34 µl plasmid DNA (0.1-1 µg). The mixture was dissolved by gentle vortexing until the pellet was resuspended and incubated at 30°C for 30minutes before incubating at 42°C for 20 minutes for heat shock. The transformants were plated out on appropriate selectable medium containing geneticin (Formedium™) or neoceothricin (Werner BioAgents) antibiotic. The plates were incubated at 30°C for 2-3 days. The transformants were tested by colony PCR.

4.4.5 Resistant marker cassettes amplification by PCR

The resistant marker cassette used in this study was *loxP-natNT2-loxP* (Carter and Delneri, 2010). The cassette was amplified using the gene specific primers (appendix
Table A9) according to (Delneri et al., 2003) with some modifications. The PCR reaction mixture consisted of 1µl template DNA, 0.4 mM of dNTPs, 5x PCR buffer provided by the supplier, 1.5 mM of MgCl₂, 3 µM of each primer and 5 units Roche Taq Polymerase. The final volume was adjusted to 50µl by sterile MilliQ water. The cycling conditions used were; an initial start of 3 min at 95°C, followed by 10 cycles of 30 sec at 95°C for DNA denaturation, 30 sec at 55°C for primer annealing and 2 min 40 seconds at 68°C for DNA extension. This was followed by 20 additional cycles for same DNA denaturation, primer annealing and extension with an addition of 20 sec/cycle. A last step of 5 minutes at 72°C was included to ensure full extension of DNA (Janke et al., 2004).

### 4.4.6 Construction of chromosomal translocations:

Among the four translocations present in *S. cariocanus*, one was generated in *S. paradoxus* using the cre-λoxP recombination system. The resistance gene marker cassettes containing the λoxP loci were inserted in the genome by PCR-mediated gene replacement method (Wach et al., 1994). The strains bearing cassettes were transformed with cre-recombinase containing plasmid. The cre-recombinase enzyme was induced by firstly growing the cells overnight in YP-raffinose medium and then in YP-galactose for 2-3 hrs. The colonies were verified for translocation and no translocation by colony PCR.

### 4.4.7 Confirmation of the engineered strains by colony PCR

The engineered translocated and control non-translocated strains were verified by colony PCR. Different sets of primers were used for each specific reaction (supplementary table). The template was prepared by dissolving a small amount of colony in 38.5 µl of sterile MilliQ H₂O. The cell suspension was heated at 95°C for 10 minutes and the ruptured cells were allowed to cool on ice for 5 minutes. The PCR mixture was prepared by adding 0.4 mM of dNTPs, PCR buffer provided by the supplier, 1.5 mM of MgCl₂, 50 pmol of each primer and 5 units Roche Taq polymerase. 11 µl of the reaction mixture was added in 38.5 µl of cell suspension to bring up the final volume to 50 µl. The thermocycling conditions used were 3 minutes at 95°C,
followed by 35 cycles of 45 seconds at 95°C to denature DNA, 45 seconds at 55–58°C for primer annealing and 1–3 minutes at 72°C for DNA elongation. Finally a cycle of 5 minutes at 72°C was added to ensure full extension of DNA. 15 µl of the PCR product was loaded on 1% or 0.1% (w/v) agarose gel electrophoresis in 1xTAE and visualized by ethidium bromide staining on a transilluminator.

### 4.4.8 Yeast sporulation

The *Saccharomyces* hybrid strains (S.par-N44-T1 × S. car, S.par-N44-NT1 × S. car, S.par-N17-T1 × S. car and S.par-N17-NT1 × S. car) and wild type *S. cariocanus* were grown in rich pre-sporulation medium (0.8% (w/v) Bacto-yeast extract, 0.3% (w/v) Bacto-peptone and 50ml of 40% (w/v) glucose) overnight at 30°C and washed twice with sterile MilliQ H2O before platting them on minimal sporulation medium (5 g potassium acetate, 0.5 g bacto yeast extract, 0.313 ml of 40% (w/v) glucose, 10g agar in 500ml of distilled H2O). The plates were incubated at 20°C for 5-7 days to force them into meiosis and formation of tetrads containing four haploid spores. The haploid spores were visible under light microscope after one week of incubation.

### 4.4.9 Yeast hybrid generation and tetrad analysis

The haploid cells of engineered *S. paradoxus* (MAT a) strains S. par-N44-T1, S. par-N17-T1, S. par-N44-NT1 and S. par-N17-NT1 were put in contact with spores of *S. cariocanus* using a microneedle to create the hybrids. Prior to creating the hybrids, the tetrads of *S. cariocanus* were dissolved in 1.2 M sorbitol and 5 mg/ml zymolase and incubated at 37°C for 8 minutes to digest the ascus wall. The *S. cariocanus* tetrads and *S. paradoxus* haploid cells were streaked vertically opposite to each other on a thin YPD agar plate. The tetrads were then dissected and four spores separated on the grid using the microneedle. The *S. paradoxus* cells were brought in close contact with the spores to allow mating. The mating plates were incubated at 30°C for two days and replica plated on SD minimal medium containing resistant antibiotic geneticin G418 for selection of the hybrids (minimal medium for *S. cariocanus* selection and genetacin for *S. paradoxus* selection). The hybrids were sporulated on sporulation medium for 7-10
days. Spore viability was checked by dissecting the tetrads using Singer MSM-300 micromanipulator.

4.5 Results and Discussion

Reciprocal translocations can reinforce the reproductive isolation barrier in yeast. Delneri et al. in 2003 proposed that reciprocal translocations do not infer the process of speciation; however, once species have been formed by some other means they may lead to reproductive isolation between the species. S. cariocanus and S. paradoxus have same genome sequence but they differ by the presence of four reciprocal translocations located on eight different chromosomes. We aim to study whether the reproductive barriers between the two species are caused by the translocations and that whether the use of the biological species concept is appropriate in this case. To answer this question we constructed chromosomal translocation at chromosome IX in S. paradoxus strains N-17, N-44 and YPS138 to create one out of four reciprocal translocations found in S. cariocanus using the cre-loxP recombination mechanism, so that only three translocations were different. Successful translocants were obtained for N-17 and N-44; however, we were unable to find any translocant for YPS138. Therefore N-17 and N-44 were processed for this study.

4.5.1 Construction of translocated and non-translocated strains

To study the impact of chromosomal translocations on reproductive isolation, we constructed translocated and non-translocated strains using the cre-loxP recombination system (Delneri et al., 2000). S. paradoxus (N-17 and N-44) was the background strain used to construct all translocated and non-translocated strains. The steps included for the construction of strains are shown below.

4.5.1.1 Amplification of resistant marker cassette

The resistant marker cassette loxP-natNT2-loxP was amplified by PCR using the cassette specific amplifying primers listed in appendix Table A9. These primers contained approximately 45 bp flanking sequence of both ends homologous to the
flanking region within the chromosome IX and chromosome XV. The correct size of the PCR amplified product was confirmed on 0.8 % (w/v) agarose gel by gel electrophoresis (Figure 4.1).

Figure 4.1: The amplification of \textit{loxP-natNT2-loxP} cassette for transformation of \textit{S. paradoxus} strains. Panel A shows the amplification of \textit{loxP-natNT2-loxP} (1563bp) using the primers consisting of 20bp specific to the cassette and 40bp specific to upstream region of \textit{MNT3} at chromosome IX. Panel B shows 0.8% (w/v) agarose gel and the lanes consist of: lane HL1 is for Hyperladder I (Bioline) and lanes 1-3 are for \textit{loxP-kanMX-loxP} gene of 1563bp size with overhangs of \textit{MNT3} upstream region.

4.5.1.2 Insertion of cassettes

The amplified product of the gene specific cassette (approximately 0.1-1 µg) was transformed in \textit{Saccharomyces paradoxus} (N-17 and N-44) strain by lithium acetate transformation protocol. The translocation breakpoints in \textit{S. cariocanus} were previously mapped by (Fischer et al., 2000). The \textit{loxP-natNT2-loxP} cassette was inserted at the translocation breakpoints which was immediately upstream of YIL014w
at chrIX and YOL055c at chrXV. Cassettes were inserted at their specific loci by homologous recombination as shown in Figure 4.2. The transformants were selected on YPD plates containing neoceothricin antibiotic.

Figure 4.2: Cassettes insertion by homologous recombination. Panel A shows the insertion of amplified loxP-natNT2-loxP cassette by homologous recombination in N-44 at chrIX and XV. Panel B shows the chromosomes IX and XV consisting of loxP-natNT2-loxP cassette at their specific locations. Chromosome IX is indicated in blue and XV in brown, loxP sequences are arrows. The cassette was inserted in similar strategy at its respective positions in strain N-17.

4.5.1.3 Confirmation of cassette insertion

Verification of the correct insertion of the cassettes in all strains was done by diagnostic colony PCR using the checking primers listed in appendix Table A5. Four primers were designed for each colony PCR experiment. The nat-up and nat-down primers were specific to the 5' and 3' regions of the loxP-natNT2-loxP cassette whereas the forward and reverse primers were designed to be gene specific (Figure 4.3A1 and 4.3B1). If the cassette was inserted successfully at chromosome IX then a band size of 1000bp is expected with primer pairs chk9Tf and natTup (Figure 4.3A2) and chk9Tr with natTdown will give a band size of 598bp (Figure 4.3). In a similarly strategy the successful insertion of the cassette at chromosome XV will be tested which is shown in Figure 4.3. The amplified PCR products were run on 1% (w/v) agarose gel electrophoresis to confirm the predicted size of the product (Figure 4.3).
Figure 4.3: The colony PCR for confirmation of cassette insertion in transformant colonies grown on YPD + neocothricin. All the gels shown in this figure were made as 1% (w/v) agarose, lane HL1 in each gel is for Hyperladder I (Bioline) and lane NC is for the negative control of respective primer pairs. Panel A1 and B1 show the location of loxP-natNT2-loxP insertions in chromosome IX and XV along with the cassette and gene specific primer binding regions. Panel A2 the lanes 1-13 are the transformant colonies containing loxP-natNT2-loxP at chromosome IX confirmed by primer pair chk9-f+nat-up giving a product of band size 1000 bp. Panel A3 the lanes 1-9 are the transformant colonies containing loxP-natNT2-loxP at chromosome IX confirmed by primer pair chk9-r+nat-down giving a product of band size 598 bp. Panel B2 lanes 1-6 are the transformant colonies having loxP-natNT2-loxP at chromosome XV confirmed by primer pair chk15-f+nat-up producing a product of band size of 1000bp. Panel B3 the lanes 1-4 are the transformant colonies containing loxP-natNT2-loxP at chromosome XV confirmed by primer pair chk15-r+nat-down producing a product of band size 800 bp.

4.5.1.4 Cre transformation and induction

The cassette removal and chromosomal translocation was performed by transformation and induction of cre gene containing plasmid. Colonies that lost the marker cassettes
were selected and confirmed for translocation as well as for lack of translocation by diagnostic colony PCR and predicted size confirmed by 1% (w/v) agarose gel electrophoresis (Figure 4.4).

**Figure 4.4: The colony PCR for confirmation of translocations and non-translocations.** All the gels shown in this figure were made as 1% (w/v) agarose, lane HL1 in each gel is for Hyperladder I (Biolaine) and lane NC is for the negative control of respective primer pairs written on the top of each gel. **Panel A1** shows the chromosomes IX and XV after translocation along with their specific primer binding regions to confirm for translocation. **Panel A2** shows the lanes 1-4 for the translocant colonies confirmed by primer pair chk9-f+chk15-r producing a product of band size 515 bp. **Panel A3** shows the lanes 1-8 for the translocant colonies confirmed by primer pair chk15-f+chk9-r giving a product of band size 615 bp. **Panel B1** shows the chromosomes IX and XV without translocation along with their specific primer binding regions to confirm no translocation. **Panel B2** shows the lanes 1-4 for the non-translocant colonies confirmed by primer pair chk9-f+chk9-r giving a product of band size of 764 bp. **Panel B3** the lanes 1-8 are the non-translocant colonies confirmed by primer pair chk15-f+chk15-r producing a product of band size 398 bp.

The strains possessing translocation were designated as S. par-N17-T1 and S. par-N44-T1 and the non-translocated strains were designated as S. par-N17-NT1 and S. par-N44-NT1.
4.5.2 Spore viability of *S. cariocanus*

We first tested the spore viability of *S. cariocanus*. It showed a spore viability of 81.25% as shown in table 4.1. An example of plate showing spore viability is shown in Figure 4.5

**Table 4.1: Spore viability of *S. cariocanus* and *S. paradoxus* engineered strains**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Total spores dissected</th>
<th>Number of viable spores</th>
<th>% of viable spores</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. cariocanus</em></td>
<td>384</td>
<td>312</td>
<td>81.25</td>
</tr>
</tbody>
</table>

![S. cariocanus](image)

**Figure 4.5: The spore viability of *S. cariocanus*.** Figure showing the spore viability of *S. cariocanus*. *S. cariocanus* was sporulated and tetrads were dissected using micromanipulator on YPD agar plate. All plates were incubated for 2-3 days at 30°C.

4.5.3 Generation of *S. cariocanus* × *S. paradoxus* hybrids using the micromanipulation technique

Each of the engineered translocated and non-translocated cells was crossed with spores of *S. cariocanus* to see the impact of restored translocation. *S. cariocanus* is wild-type yeast that is able to switch mating type. To cross it with our *S. paradoxus* translocated and non-translocated haploid (MAT a) strains, *S. cariocanus* underwent meiosis upon sporulation and formed tetrads with four haploid spores. The spores were separated and brought in contact with *S. paradoxus* (translocated and non-translocated) haploid cells.
with the help of micromanipulator needle to construct hybrids. Twenty spore-cell matings were done for all crosses on enriched medium (YPD) plates to observe growth for all colonies. The hybrids formed were selected by replica plating on minimal medium (for selection of *S. cariocanus*) containing genaticin antibiotic for analysing the presence of kanamycin marker in *S. paradoxus* (Figure 4.6). The hybrids were designated as *S*.par-N44-T1 × *S*. car, *S*.par-N44-NT1 × *S*. car, *S*.par-N17-T1 × *S*. car and *S*.par-N17-NT1 × *S*. car. We also tested the spore viability of parental strains i.e. *S*.par-N44-T1 × *S*. par-N44-NT1, *S*. par-N44-NT1 × *S*. par-N44-NT1, *S*.par-N17-T1 × *S*.par-N17-NT1, *S*.par-N17-NT1 × *S*.par-N17-NT1. All other hybrids were constructed in similar method.

![Figure 4.6: Generation of hybrids and selection on SD medium with geneticin G418. Panel A shows the growth of all colonies for strain S.par-N44-NT1 × S. car on rich medium (YPD) after crossing with microneedle. Panel B shows the growth of hybrid colonies of S.par-N44-NT1 × S. car after being replica plated on minimal medium containing urea and geneticin (SD + urea + G418).](image)

### 4.5.4 Spore viability of hybrid strains

Each of the hybrid strain was sporulated for 7 days and the spores of these hybrids were then dissected to check the effect of reciprocal translocation on spore viability. The control non-translocated strains of N-44 and N-17 showed spore viability of 82.29 and 84.38% respectively, however, the translocated strains possessed 50% drop in spore viability. The hybrids of *S*. par-N44-T1 × *S*. car showed an approximate cumulative spore viability of 2.46 % and the non-translocated ones (*S*.par-N44-NT1 × *S*. car) had
slightly lower spore viability of 1.7%. The hybrids of S. par-N17-T1 × S. car showed spore viability of 2.77% whereas the non-translocated ones (S.par-N17-NT1 × S. car) possessed spore viability of 1.54%. A total of 192 and 648 spores were dissected for each hybrid (Table 4.2). An example of a dissection plate for *S. paradoxus* × *S. cariocanus* hybrid is shown in Figure 4.7.

**Table 4.2: Spore viability of hybrids and number of total spores analyzed.**

<table>
<thead>
<tr>
<th>Hybrid strains</th>
<th>Total spores dissected</th>
<th>Number of viable spores</th>
<th>% of viable spores</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. par-N44-NT1 × S. par-N44-NT1</td>
<td>192</td>
<td>158</td>
<td>82.29</td>
</tr>
<tr>
<td>S.par-N17-NT1 × S.par-N17-NT1</td>
<td>192</td>
<td>162</td>
<td>84.38</td>
</tr>
<tr>
<td>S.par-N44-T1 × S. par-N44-NT1</td>
<td>192</td>
<td>102</td>
<td>53.13</td>
</tr>
<tr>
<td>S.par-N17-T1 × S.par-N17-NT1</td>
<td>192</td>
<td>99</td>
<td>51.56</td>
</tr>
<tr>
<td>S.par-N44-T1 × S. car</td>
<td>648</td>
<td>16</td>
<td>2.47</td>
</tr>
<tr>
<td>S.par-N44-NT1 × S. car</td>
<td>648</td>
<td>11</td>
<td>1.70</td>
</tr>
<tr>
<td>S.par-N17-T1 × S. car</td>
<td>648</td>
<td>18</td>
<td>2.78</td>
</tr>
<tr>
<td>S.par-N17-NT1 × S. car</td>
<td>648</td>
<td>10</td>
<td>1.54</td>
</tr>
</tbody>
</table>
Figure 4.7: Crosses in hybrids. Dissection of spores from strains obtained from crosses between S.par-N44-T1 × S. car (Panel A), S.par-N44-NT1 × S. car (Panel B), S.par-N17-T1 × S. car (Panel C) and S.par-N17-NT1 × S. car (Panel D). 648 tetrads in total were dissected for each hybrid using micromanipulator and incubated for 3-4 days to check the growth of viable spores.

All the Saccharomyces ‘sensu stricto’ species are able to mate in the laboratory (Naumov et al., 2000). However, in postmating reproductive isolation the F1 hybrids are made normally but are sterile when undergoing meiosis. Most of the F1 hybrids are sterile and the gametes made from them are in-viable or give a spore viability of less than 1%. It has been shown that the hybrids formed between S. paradoxus and S. cariocanus give a spore viability of 0.66% (Fischer et al., 2000) however, Liti et al., 2006 observed a spore viability of approximately 5% for the hybrids of S. paradoxus and S. cariocanus. Mating between the same species produces spores of 100% or ~80% viability (Greig et al., 2002). The sterility of F1 hybrids can be due to chromosomal rearrangements preventing the gametes from receiving a full haploid set of genes. Chromosomal translocations may block the meiotic pairing of chromosomes resulting in chromosomal missegregation. This model has previously been reported showing that restored colinearity in the genome of S. cerevisiae relative to S. mikatae significantly increased the hybrid fertility (Delneri et al., 2003).
Sequence divergence between homologous chromosomes can lead to failure in recombination at meiosis resulting in formation of inviable aneuploid spores (Scannell et al., 2007). In 2006 Liti et al. demonstrated that the species of *S. cariocanus* (UFRJ50791 and UFRJ50816) appeared within the North American lineage of *S. paradoxus* on the basis of phylogenetic analysis among strains of *S. paradoxus* and *S. cariocanus*. They also reported that *S. cariocanus* as a member of *S. paradoxus* species particularly the American sub populations as the hybrids between the two species produce viable spores (4.81% for a cross between Sc UFRJ50791 and Sp YPS138 whereas 5.09% for a cross between Sc UFRJ50816 and Sp YPS125). It was also shown that in the hybrids of these two species no aneuploidy existed in the viable spores.

According to our hypothesis since *S. cariocanus* possesses four translocations therefore the hybrids of *S. paradoxus* and *S. cariocanus* should possess a spore viability of 6.25%. Reducing the genomic rearrangements from four to three between the two species should lead to an increase in the spore viability (from 6.25% expected when four translocations are present to 12.5% expected when three translocations are present). Our results show that *S. cariocanus* gave spore viability of 81% and similarly the engineered strains also possessed spore viability of 75-82% (Table 4.1) therefore a cross between the two species should reduce the spore viability to 4.6-5.01%. By restoring one out of four translocations, the spore viability should increase to 9.2-10%.

We observed that by reversing the *S. cariocanus* translocation chr9/chr15 in *S. paradoxus* and crossing the translocated/non-translocated strains with *S. cariocanus* the increase in the spore viability was significant (Two sample t-test: p<0.001). From the data so far obtained we conclude that presence of translocations in *S. cariocanus* relative to *S. paradoxus* is responsible for the reproductive isolation of the two species. In all 648 spores dissected for our crosses we observed a similar pattern of increase from 1.5% for the hybrids with no translocation to 2.7% for the hybrids with translocation. This suggests the possibility that by reversing all of the four translocations we may have a higher increase in the spore viability.
4.6 References


CHAPTER 5

GENERAL CONCLUSION AND FUTURE WORK
5. **General conclusion and future work**

A survey of the literature shows that the genomes of yeast species have changed massively during the period of evolution due to chromosomal rearrangements including deletion, duplication, translocation and inversion of genes. In this study, I focused on chromosomal inversions and translocations and investigated their effect on fitness adaptation, gene expression and speciation. In second and third chapter, we showed the effect of small and large size inversions on phenotypic fitness, recombination and gene expression. In fourth chapter, we studied the role of chromosomal translocations in speciation process.

5.1 **Chapter 2. Impact of chromosomal inversions on co-expressed gene clusters**

Chromosomal inversions and their potential role in fitness adaptation and expression processes in yeasts are not well studied. Therefore in our research, we studied the effects of inversion on a co-expressed gene cluster (the *DAL* cluster) using yeast as a model organism. It has been shown that *Saccharomyces castellii* *DAL* cluster possesses inversions of three *DAL* genes relative to *Saccharomyces cerevisiae* and other *sensu stricto* species. To study the role of these inversions on fitness and gene expression, we constructed several inverted and non-inverted strains possessing single, double, or triple gene inversions as well as a strain having a *DAL* cluster similar the *S. castellii* *DAL* cluster in *S. cerevisiae* background (Fig 2.18 & 2.19). The *DAL* cluster consists of six *DAL* genes located on chromosome IX at a subtelomeric region in *S. castellii*, *S. cerevisiae* and other *sensu stricto* species. Three of the *DAL* genes (*DAL1*, *DAL2* and *DAL4*) have been inverted in *S. castellii* and *DAL2* has been inverted twice such that its position in the cluster is changed but it has maintained its orientation (in the same direction) (Figure 2.2). It is not known which of the *DAL* genes in *S. castellii* got inverted first. We therefore adopted two different approaches for engineering *S. castellii* like *DAL* cluster in *S. cerevisiae* background (Figure 2.19).

We showed that the phenotypic fitness of *S. cerevisiae* reduces by inverting *DAL2* however there is a relative gain in fitness when *DAL2* is re-inverted along with
inversion of \textit{DAL1} and \textit{DAL4} to mimic \textit{S. castellii} like \textit{DAL} cluster in \textit{S. cerevisiae} background (Figure 2.21). On the other hand, we did not observe any change in fitness in the strain having \textit{DAL1-DAL2-DAL4} inverted, however, by re-inverting \textit{DAL2} showed a disadvantage on strains fitness. We can speculate that in nature \textit{DAL2} might first have undergone inversion in \textit{S. castellii} and due to fitness disadvantage it would have re-inverted to restore the fitness effect and this inversion would then have been favored by \textit{S. castellii} and it retained in its genome. It has been shown in literature that inversion play a very important role in local adaptation, and inversions that get adapted to the environment get selective advantage and are therefore retained in the genome [1].

Genes in \textit{DAL} metabolic pathway are tightly clustered and it has been reported that one reason for clustering of genes is due to epistatic selection for linkage. The genes were linked to avoid the accumulation of toxic substance glyoxalate which is an intermediate product made in the \textit{DAL} pathway and is toxic to yeast. \textit{DAL3} product is responsible for the production of glyoxalate which is in turn removed by \textit{DAL7}. That might be the reason for the two genes to function together. Previous studies have shown that \textit{DAL3} activity is reduced in absence of \textit{DAL7} [2, 3].

Our study showed that insertion of a small sequence of 34bp in the intergenic region of \textit{DAL} cluster leads to alteration of the gene expression because these genes are tightly clustered (Figure 2.26). Expression studies on inverted and non-inverted strains showed that by inverting a single gene in \textit{DAL} metabolic cluster of yeast, it not only changes the expression of the gene being inverted, but the expression of the neighboring genes is also altered (Figure 2.24). We also showed that inversion of \textit{DAL2} reduces the expression of \textit{DAL4} in the inverted strain and this alteration is not due to the change in the expression of its anti-sense transcript \textit{SUT614} (Fig 2.29).

The sub-telomeric \textit{DAL} cluster is placed within Htz1-activated domain (HZAD) which consists of histone H2A variant H2A.Z (Htz1). H2A.Z plays an important role in the transcription of genes that are normally in repressed form and are highly induced under specific growth conditions. H2A.Z binds to the nucleosomes in the promoter region of inactive genes, in order to keep the chromatin structure available for transcription [4]. It also helps in transcription by preventing the binding of silencing protein to
euchromatin. In the absence of H2A.Z, the silencing proteins spread to the neighboring genes thereby interfering in their normal expression [5].

H2A.Z is present in region of $DAL1$, $DAL2$, $DCG1$ and $DAL3$ genes of $DAL$ metabolic cluster (Figure 5.1). It is thought to act as an anti-silencing factor by preventing the spread of Sir protein and therefore activating the expression of these genes. It is possible that inversion of $DAL1$ and $DAL2$ in our engineered strains might have affected the H2A.Z factor which in turn altered the expression of these genes. The next step in this project would be to study the effect of H2A.Z on expression of $DAL$ genes in inverted and non-inverted strains. This could be done by deleting the H2A.Z factor from inverted and non-inverted individuals and comparing the expression of H2A.Z deleted inverted mutants with the H2A.Z non-deleted inverted mutant using the Real-Time quantitative PCR.

**Figure 5.1: Role of H2A.Z in $DAL$ genes expression.** The genes of $DAL$ metabolic cluster are located within Htz1- activated domain (HZAD) which consists of H2A histone variant H2A.Z. H2A.Z binds at promoter regions of $DAL1$, $DAL2$, $DAL3$ and $DCG1$. It activates the expression of these genes by preventing the spread of Sir protein (silencing protein) into the euchromatin regions. The pink block arrows indicate the direction of genes and H2A.Z is represented by blue boxes. Figure taken from [6].
5.2 Chapter 3. Effect of pericentric and paracentric chromosomal inversions on global gene expression

The results from Chapter 2 established that small size inversions consisting of just one or few genes can change the phenotypic profile as well as gene expression. This led us to study the effects of large pericentric and paracentric inversions on strain’s fitness, spore viability and global transcription level. To study this, we used previously constructed strains in Delneri’s lab possessing pericentric and paracentric chromosomal inversions as small as 26kb and as big as 754kb between Ty1 and Ty2 elements on different chromosomes. We observed that both small and large size inversions did not change the phenotypic fitness of strains in nitrogen limited, carbon limited and SD medium. Global expression analysis showed that most inversions did not alter the overall expression profile however, one large inversion not only altered the expression of genes located within the inverted region but also changes the expression of genes present in the vicinity of breakpoints and in the other parts of the genome (Figure 3.9).

Inversions can cause hybrid sterility in the heterozygotes. They can lead to the formation of unbalanced gametes during meiosis due to crossing over in the inverted regions causing hybrid sterility. It has been shown that single or odd number of crossovers between the breakpoints of a pericentric inversion will produce unbalanced gametes carrying deletions, insertions, zero or two centromeres. Such pericentric heterozygous individuals will have lower fitness in nature [7].

We crossed the inverted strains possessing small and large pericentric and paracentric inversions with the non-inverted candidates and found that size of the inversions affects the spore viability. It was observed that medium and large size paracentric as well as large size pericentric inversions reduce the spore viability to 50% in comparison to the non-inverted ones, whereas small size inversions had no effect on the spore viability.

By aligning the hot spots of recombination with the inverted region of our strains we found that the strains showing loss in spore viability possessed at least two recombination hotspots in the inverted region with the exception of one strain 4/7. The
strains with no significant change in spore viability either had none or one recombination hot spot in the inverted region (Table 3.4).

Our expression data analysis showed that there was no particular class of genes being affected by inversions and different genes in the genome were randomly changing the expression. In strain 12/13 the expression of three genes (MND1, SPO16 and MSH5) that are involved in meiosis was significantly changed and this strain also showed a 50% drop in spore viability. MND1 is involved in chromosome pairing during recombination and it was 1.5 fold down-regulated in the inverted strain. Similarly, MSH5 was also down-regulated by 1.5 fold and it is responsible for crossover formation during meiosis. Change in the expression of these genes could be a reason for the altered spore viability of this strain. It was found that the global expression profile of strain 25/24, possessing a large size inversion was massively altered in the inverted individuals relative to the non-inverted ones. ARP5 is an important gene involved in chromatin remodeling and it was up-regulated by 1.5 fold in the inverted strain. ARP5 is a component of INO80 complex. This complex is known to play a very important role in transcriptional regulation, DNA replication and repair. It may be that ARP5 expression was altered due to inversions which lead to the change in global expression of other genes in the genome. This study will be finished by confirming the change in expression of MND1, MSH5, SPO16 and ARP5 by quantitative real time PCR.

5.3 Chapter 4. Impact of chromosomal translocations on S. paradoxus and S. cariocanus speciation

Chromosomal translocations have previously been extensively studied in yeast and they are found in many yeast species. They play a very important role in reproductive isolation by rearranging the genome on one species relative to the other species. S. paradoxus and S. cariocanus differ by the presence of four reciprocal translocations. It is known that one reciprocal translocation reduces the spore viability to 50%. Due to the presence of four reciprocal translocations in S. cariocanus, the hybrids between S. paradoxus and S. cariocanus should give a reduced spore viability of 6%. By restoring one reciprocal translocation the spore viability should increase to 12.5%. We tested this
hypothesis by creating one reciprocal translocation involving chromosomes IX and XV in *S. paradoxus* strains N-44 and N-17. The translocated and non-translocated strains were crossed with *S. cariocanus* and the hybrids were sporulated. We showed that there was a small but significant increase in spore viability of translocated hybrids of N-17 and N-44 relative to the non-translocated ones. The spore viability was increased from 1.5% (of non-translocated hybrids) to 2.7% (of translocated hybrids) shown in Table 4.2. This indicates that reproductive isolation between *S. paradoxus* and *S. cariocanus* can be due to the chromosomal translocations and not sequence divergence.

It is known from a previous study that N-17 is a European strain, N-44 is Far Eastern and YPS138 is American strain [8]. YPS138 is the closest relative to *S. cariocanus* and it would be worth restoring the translocation in YPS138 and test the spore viability of its translocated hybrid with *S. cariocanus*.

### 5.4 References


APPENDIX (A)
### Table A1. List of plasmids used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Gene</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUG6</td>
<td>Kan</td>
<td>Winzeler et al., 1999</td>
</tr>
<tr>
<td>pFA6a-loxP-hphNT1-loxP(pZC1)</td>
<td>hph</td>
<td>Carter and Delneri, 2010</td>
</tr>
<tr>
<td>pFA6a-loxP-natNT2-loxP(pZC2)</td>
<td>nat</td>
<td>Carter and Delneri, 2010</td>
</tr>
<tr>
<td>pFA6a-lox2272-natNT2-lox2272(pZC4)</td>
<td>nat</td>
<td>Carter and Delneri, 2010</td>
</tr>
<tr>
<td>pSH-ble</td>
<td>cre</td>
<td>Guldener et al., 1996</td>
</tr>
</tbody>
</table>

### Table A2. List of checking primers used for confirming inverted and non-inverted strains.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence 5'-3'</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAL1_FC</td>
<td>AACCCATCCGCTCTGAGT</td>
<td>58.2</td>
</tr>
<tr>
<td>DAL1_RC</td>
<td>AATTGCAGCCGTGGAGAC</td>
<td>56</td>
</tr>
<tr>
<td>DCG1_RC</td>
<td>GGATGGCACGCTGTTTCT</td>
<td>56</td>
</tr>
<tr>
<td>Dal2_FC</td>
<td>GCTCTATGACGAGGCCGAAG</td>
<td>59.4</td>
</tr>
<tr>
<td>hph-up</td>
<td>AGTTCGGTTTCAGGCAAGT</td>
<td>56.7</td>
</tr>
<tr>
<td>hph-down</td>
<td>AGCATCAGCTCATCAGAG</td>
<td>56.7</td>
</tr>
<tr>
<td>DAL4.F</td>
<td>CTGTCCCTGGTACTCCCGTA</td>
<td>61.4</td>
</tr>
<tr>
<td>DAL2.R</td>
<td>CTATGACGAGGCCGAAG</td>
<td>59.4</td>
</tr>
<tr>
<td>P3K4up</td>
<td>AACGTGAGTCTTTTCCTTACC</td>
<td>60.1</td>
</tr>
<tr>
<td>P4K4down</td>
<td>TCTCCTTCATTACAGAAGC</td>
<td>57.3</td>
</tr>
<tr>
<td>nat_down</td>
<td>GCTGACCGTCAGAGACAT</td>
<td>58.2</td>
</tr>
<tr>
<td>nat_up</td>
<td>ATGTCCTCGACGGTCAGC</td>
<td>58.2</td>
</tr>
<tr>
<td>D1</td>
<td>CCCCAAGAAGCACTGTTGT</td>
<td>57</td>
</tr>
<tr>
<td>D2</td>
<td>AGCAATTGTCACTGTTTCC</td>
<td>55</td>
</tr>
<tr>
<td>D3</td>
<td>TGCAGCCACATGAAGTACTTGA</td>
<td>57</td>
</tr>
<tr>
<td>D4</td>
<td>TTGAGGAGTCGTCGATGTA</td>
<td>57</td>
</tr>
<tr>
<td>D5</td>
<td>TGGTTAAGGGGATCTGGTGT</td>
<td>57</td>
</tr>
<tr>
<td>D6</td>
<td>GGGAAGAGCACGCCCTCCAC</td>
<td>61</td>
</tr>
<tr>
<td>D7</td>
<td>GGATGGCCAGCGCTTCTTCT</td>
<td>56</td>
</tr>
<tr>
<td>D8</td>
<td>AACTGGGAGGGAGTCTTCTT</td>
<td>59</td>
</tr>
</tbody>
</table>
Table A3. List of primers used for real time PCR.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence 5'-3'</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT-DAL1_R</td>
<td>GTCATCCCGTTGTTAGGCAGT</td>
<td>59.4</td>
</tr>
<tr>
<td>RT-DAL1_F</td>
<td>AGGCTGCTCCTTTTGTCGATA</td>
<td>57.3</td>
</tr>
<tr>
<td>RT-DAL2_R</td>
<td>CCCGTCTTTGACTTACCAA</td>
<td>57.3</td>
</tr>
<tr>
<td>RT-DAL2_F</td>
<td>CATTACCCTGGAAGGTGTGCT</td>
<td>57.3</td>
</tr>
<tr>
<td>RT-DAL4_F</td>
<td>GGGCTGGTGTTCTTTGATT</td>
<td>57.3</td>
</tr>
<tr>
<td>RT-DAL4_R</td>
<td>CTTGGAAACAAAGCCGTCAT</td>
<td>55.3</td>
</tr>
<tr>
<td>RT-DCG1_F</td>
<td>TCGAAGTCAATGACGGTGTC</td>
<td>57.3</td>
</tr>
<tr>
<td>RT-DCG1_R</td>
<td>CTTGTTTCCTGCCCATCAAT</td>
<td>55.3</td>
</tr>
</tbody>
</table>

Table A4. List of probes used for northern blotting.

<table>
<thead>
<tr>
<th>Probes</th>
<th>Sequence 5'-3'</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAL4-sense</td>
<td>ATAGTATGGTGATGCGTGCAAGCCTGGTAGTGCAACGC</td>
<td>74.6</td>
</tr>
<tr>
<td>DAL4-antisense</td>
<td>GCGTTGCACTAACCAGGGCTTGACGCAATACCATAT</td>
<td>74.6</td>
</tr>
<tr>
<td>New DAL4-sense1</td>
<td>TGTCCAACCCTCCACATCTATTGTGTCATATACA</td>
<td>67.1</td>
</tr>
<tr>
<td>New DAL4-antis1</td>
<td>TGTATATGACACAATAGATGTGGAGGTTGGACA</td>
<td>67.1</td>
</tr>
<tr>
<td>New DAL4-sense2</td>
<td>GTCCCTGTTCAATTACTTTATATTTCCTTGCTGGCA</td>
<td>67.1</td>
</tr>
<tr>
<td>New DAL4-antis2</td>
<td>TGCCAGGCAAGAAATATAAGTAATTGAACCAGGAC</td>
<td>67.1</td>
</tr>
<tr>
<td>New DAL4-sense5</td>
<td>AAGATTTGGAAAGAATGAGTTACGTAGGATGACCT</td>
<td>67.1</td>
</tr>
<tr>
<td>New DAL4-antis5</td>
<td>AGGTACATCCCTACGTAACTCTATTTCTCCAATCTT</td>
<td>67.1</td>
</tr>
<tr>
<td>DAL4-sense3</td>
<td>TGGAGACCACCTACACCAGAGGTACTAGGCTGGTGTT</td>
<td>74.6</td>
</tr>
<tr>
<td>DAL4-antisense3</td>
<td>AACACACCAGCCTCAGTACCTTTGCTGGTGTAAGTGGTGCTCCA</td>
<td>74.6</td>
</tr>
<tr>
<td>New Actin-sense1</td>
<td>ATCTATCGTGGTAGACAAGACACCAAGGTA</td>
<td>68.2</td>
</tr>
<tr>
<td>New Actin-antis1</td>
<td>TACCTTGGTGCTCTTGCTACCGACGATAGAT</td>
<td>68.2</td>
</tr>
<tr>
<td>New Actin-sense2</td>
<td>AACTTTCAACGTCCAGCCTTTCTACGTTTCCAT</td>
<td>67</td>
</tr>
<tr>
<td>New Actin-antis2</td>
<td>ATGGGAAACGTAAGAGGCTGGAAGCCTGAAGGT</td>
<td>67</td>
</tr>
<tr>
<td>New Actin-antis3</td>
<td>AGGAATGATCTTTGCCTCAGGATGACG</td>
<td>68.2</td>
</tr>
<tr>
<td>Actin-sense</td>
<td>GCTTCATCCTCACGATTTGGCTCTCCATCAAGGTC</td>
<td>73.6</td>
</tr>
</tbody>
</table>
Table A5. List of checking primers used for confirming translocated and non-translocated strains

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence 5'-3'</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>chk 9f</td>
<td>CAGGTTTGCTGAGTCATTGC</td>
<td>57.3</td>
</tr>
<tr>
<td>chk 9r</td>
<td>CCTTTTGGCAACCCAACCCCTTA</td>
<td>55.3</td>
</tr>
<tr>
<td>nat_down</td>
<td>GCTGACCGTGCAGGACAT</td>
<td>58.2</td>
</tr>
<tr>
<td>nat_up</td>
<td>ATGTCCTCGAGTGTCAGC</td>
<td>58.2</td>
</tr>
<tr>
<td>N44-chk9-r</td>
<td>GCACTTCAACTCGGCTGC</td>
<td>&gt;75</td>
</tr>
<tr>
<td>YPS138-chk9-r</td>
<td>CTTTTTGCAGCCAACCCTT</td>
<td>54.5</td>
</tr>
<tr>
<td>YPS138-chkxv-r</td>
<td>GCGATTAACGTCGCACGC</td>
<td>61.4</td>
</tr>
<tr>
<td>chk15-f</td>
<td>GCTGCAAGATGGATTGATCAT</td>
<td>55.3</td>
</tr>
<tr>
<td>15R-1</td>
<td>CTAAAGGATGGAAGCGCG</td>
<td>56</td>
</tr>
</tbody>
</table>
Table A6. The genotypes of *Saccharomyces cerevisiae* inverted and non-inverted strains used in this study.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY3 (<em>S. cerevisiae</em>)</td>
<td>MATα ura3-52</td>
<td>Brachmann <em>et al.</em>, 1998</td>
</tr>
<tr>
<td>T73 (<em>S. cerevisiae</em>)</td>
<td>Natural isolate</td>
<td>Eladio Barrio laboratory</td>
</tr>
<tr>
<td>9 (<em>S. cerevisiae</em>)</td>
<td>Natural isolate</td>
<td>Eladio Barrio laboratory</td>
</tr>
<tr>
<td>96.2 (<em>S. cerevisiae</em>)</td>
<td>Natural isolate</td>
<td>Eladio Barrio laboratory</td>
</tr>
<tr>
<td>DAL2-Inv</td>
<td>FY3 pDAL2::loxP + tDAL2::loxP (inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL2-NI</td>
<td>FY3 pDAL2::loxP + tDAL2::loxP (non-inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL3-Inv</td>
<td>FY3 pDAL3::loxP + tDAL3::loxP (inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL3-NI</td>
<td>FY3 pDAL3::loxP + tDAL3::loxP (non-inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL3-DAL7-Inv</td>
<td>FY3 tDAL3::loxP + pDAL7::loxP (inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL3-DAL7-NI</td>
<td>FY3 tDAL3::loxP + pDAL7::loxP (non-inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL1-DAL2-DAL4-Inv</td>
<td>FY3 tDAL1::loxP + tDAL2::loxP (inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL1-DAL2-DAL4-NI</td>
<td>FY3 tDAL1::loxP + tDAL2::loxP (non-inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL3-DAL7-DCG1-Inv</td>
<td>FY3 tDAL3::loxP + pDCG1::loxP (inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>DAL3-DAL7-DCG1-NI</td>
<td>FY3 tDAL3::loxP + pDCG1::loxP (non-inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>S.cer-Inv1</td>
<td>FY3 tDAL1::loxP + pDAL2::lox2272 + tDAL2::lox2272 (inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>S.cer-Inv1 control</td>
<td>FY3 tDAL1::loxP + pDAL2::lox2272 + tDAL2::lox2272 (non-inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>S.cer-Inv2</td>
<td>FY3 tDAL1::loxP + pDAL2::lox2272 + tDAL2::loxP (inverted)</td>
<td>This study</td>
</tr>
<tr>
<td>S.cer-Inv2 control</td>
<td>FY3 tDAL1::loxP + pDAL2::lox2272 + tDAL2::loxP (non-inverted)</td>
<td>This study</td>
</tr>
</tbody>
</table>
Table A7. The genotypes of *Saccharomyces cerevisiae* strains possessing pericentric and paracentric inversion used in this study.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BY4741</td>
<td>MATa his3Δ1; leu2Δ0; met15Δ0; ura3Δ0</td>
<td>Brachmann <em>et al.</em>, 1998</td>
</tr>
<tr>
<td>21/22</td>
<td>BY4741 YMLWTy1-2::loxP-kanMX + YMRCTy1-3::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>21/23</td>
<td>BY4741 YMLWTy1-2::loxP-kanMX + YMRCTy1-4::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>20/22</td>
<td>BY4741 YMLWTy1-1::loxP-kanMX + YMRCTy1-3::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>20/23</td>
<td>BY4741 YMLWTy1-1::loxP-kanMX + YMRCTy1-4::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>29/28</td>
<td>BY4741 YPRCTy1-2::loxP-kanMX + YPLWTy1-1::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>30/28</td>
<td>BY4741 YPRCTy1-4::loxP-kanMX + YPLWTy1-1::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>11/13</td>
<td>BY4741 YGRCTy1-2::loxP-kanMX + YGRWTy1-1::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>12/13</td>
<td>BY4741 YGRCTy1-3::loxP-kanMX + YGRWTy1-1::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>6/7</td>
<td>BY4741 YDRCTy1-3::loxP-kanMX + YDRWTy1-4::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>6/8</td>
<td>BY4741 YDRCTy1-3::loxP-kanMX + YDRWTy1-5::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>5/7</td>
<td>BY4741 YDRCTy1-2::loxP-kanMX + YDRWTy1-4::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>5/8</td>
<td>BY4741 YDRCTy1-2::loxP-kanMX + YDRWTy1-5::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>4/7</td>
<td>BY4741 YDRCTy1-1::loxP-kanMX + YDRWTy1-4::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>4/8</td>
<td>BY4741 YDRCTy1-1::loxP-kanMX + YDRWTy1-5::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>25/24</td>
<td>BY4741 YNLWTy1-2::loxP-kanMX + YNLCTy1-1::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
<tr>
<td>31/29</td>
<td>BY4741 YPRWTy1-3::loxP-kanMX + YPRCTy1-2::loxP-hphNT1</td>
<td>Delneri Laboratory</td>
</tr>
</tbody>
</table>
Table A8. The set of cassette amplifying primers for engineering inverted and non-inverted strains

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence 5'-3'</th>
<th>Tm(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dal1_F</td>
<td>AACACCATTGGGTCAAAACTTTGCTTTGATTCTTAGACGTTAACactatagggagaccggcag</td>
<td>&gt;75</td>
</tr>
<tr>
<td>Dal1_R</td>
<td>TTTCTATAGAAATTTCTTTAATAAAATTTTGCAAACCTTTAGTcgtacgctgaggtgac</td>
<td>71.5</td>
</tr>
<tr>
<td>DAL2_SN_(F)</td>
<td>CCAGATGGAGGAGTGAAAAAGATAAGAGTTTGGGGGGTACTGAcactatagggagaccggcaga</td>
<td>&gt;75</td>
</tr>
<tr>
<td>DAL2_SN_(R)</td>
<td>AGAGTGCATTGGTTTAAAATACAGTAGTTAAGTATTTATCAGTACCCCCCCACTCTcgtacgctgaggtgac</td>
<td>&gt;75</td>
</tr>
<tr>
<td>ChIXDAL4/2ins.F</td>
<td>GTACTGATGATTAGCTAAAACCGTCAATCAACCTACAAACTTGCGC cgtacgctgaggtgac</td>
<td>72.8</td>
</tr>
<tr>
<td>ChIXDAL4/2ins.R</td>
<td>GTCAAAGATAAGATGTCGGAATTATCCGGAGTTCTGATAGGCTC cactatagggagaccggcag</td>
<td>72.8</td>
</tr>
<tr>
<td>lox2272_F</td>
<td>GTCAAAGATAAGATGTCGGAATTATCCGGAGTTCTGATAGGCTC cactatagggagaccggcag</td>
<td>&gt;75</td>
</tr>
<tr>
<td>lox2272_R</td>
<td>AGAGTGCATTGGTTTAAAATACAGTAGTTAAGTATTTATCAGTACCCCCCCACTCTcgtacgctgaggtgac</td>
<td>&gt;75</td>
</tr>
<tr>
<td>DAL1+ lox2272</td>
<td>TTTCTATAGAAATTTCTTTAATAAAATTTTGCAACTTTAGTcgtacgctgaggtgacg</td>
<td>72.8</td>
</tr>
</tbody>
</table>
### Table A9. List of cassette amplifying primers used for engineering translocations

<table>
<thead>
<tr>
<th>Probes</th>
<th>Sequence 5’-3’</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amp 9f</td>
<td>CAAGCATCTAACCTGGTTTAAATCAGATAATTCAGGAACAGCTGAAGCTTACG</td>
<td>&gt;75</td>
</tr>
<tr>
<td>Amp 9r</td>
<td>TCGGCTGCTTATATCATTCTATATTAATAGCTTGAATAGGCAAACAGCTTAGGTGATCTG</td>
<td>&gt;75</td>
</tr>
<tr>
<td>N44-ampMNT3-f</td>
<td>CAAGCATCTAACCTGGTTTAAATTACGATAATTCAGGAACAGCTTACG</td>
<td>74.4</td>
</tr>
<tr>
<td>N44-ampMNT3-r</td>
<td>TCGGCTGCTTATATCATTCTATATTAATAGCTTGAATAGGCAAACAGCTTACG</td>
<td>&gt;75</td>
</tr>
<tr>
<td>YPS138-ampMNT3-f</td>
<td>CAATCATCTAACCTGATGTGAATTCAGGAACAGCTTACG</td>
<td>72.9</td>
</tr>
<tr>
<td>YPS138-ampMNT3-r</td>
<td>TCGGCTGCTTATATCATTCTATATTAATAGCTTGAATAGGCAAACAGCTTACG</td>
<td>&gt;75</td>
</tr>
<tr>
<td>YPS138_SN_amp.f</td>
<td>CTTTACTCAGATTTTGTTCAGCTTATAATCTATATAAGATCTGGCGATAGGAGACCCGGCA</td>
<td>75</td>
</tr>
<tr>
<td>YPS138_SN_amp.r</td>
<td>ATTAGTTTATAGGCCAGATTGCCGGCTTTCCAGCTATAGCTAGGGAGACCCGGCA</td>
<td>&gt;75</td>
</tr>
<tr>
<td>N-17_SN_amp.f</td>
<td>ATTGTGTTGTATATCTCAGATTTGATAGGACCCGGCA</td>
<td>&gt;75</td>
</tr>
<tr>
<td>N-17_SN_amp.r</td>
<td>GTTTCTACTACAGTGTTCAGCTATATCTAGGTTGAAATCTGCTAGGTGACGCTGAGG</td>
<td>&gt;75</td>
</tr>
<tr>
<td>N-44_SN_amp.f</td>
<td>GGGCCCAACGGATGGCAACGTGGTTTACCACAAACTAGCCAAGATGGGAGGCCA</td>
<td>&gt;75</td>
</tr>
<tr>
<td>N-44_SN_amp.r</td>
<td>CTGAGGTACCTCATTTGGAGATCACAGCAACAAATAGCTAGGTGACGCTAGGTGACG</td>
<td>&gt;75</td>
</tr>
</tbody>
</table>
References


