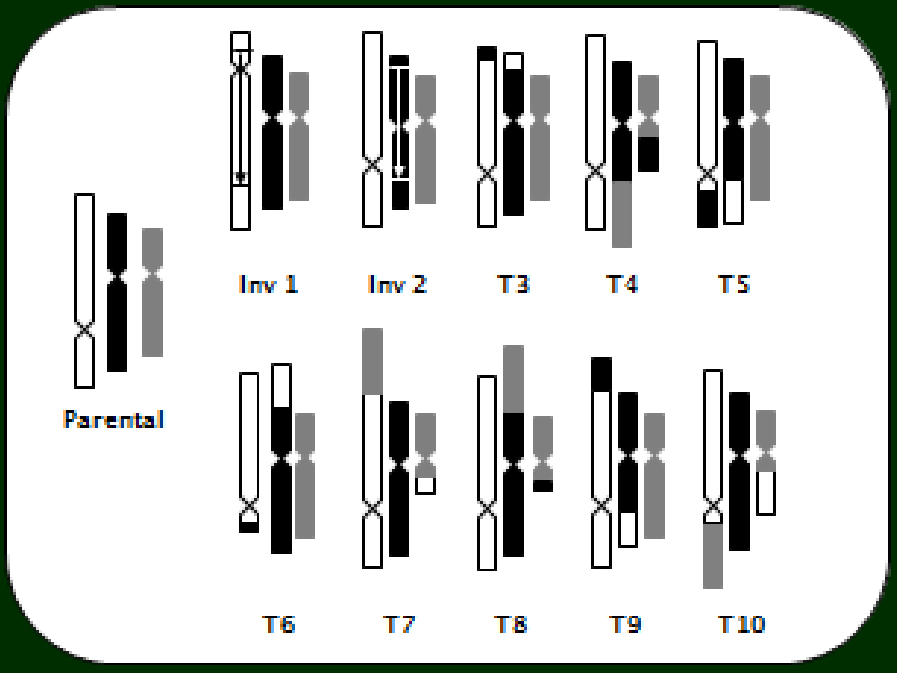
Chromosomal structure: a selectable trait for evolution

Ana Teresa dos Santos Avelar



Dissertation presented to obtain the Ph.D degree in Evolutionary Biology

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras, November, 2012

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA /UNL

Knowledge Creation

Chromosomal structure:

a selectable trait for evolution

Ana Teresa dos Santos Avelar

Dissertation presented to obtain the Ph.D degree in Evolutionary Biology

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Research work coordinated by:

FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Oeiras, Novembro 2012

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

This dissertation was sponsored by Fundação para a Ciência e Tecnologia.

Apoio financeiro da FCT e do FSE no âmbito do Quadro Comunitário de apoio,
BD nº SFRH/BD/33214/2007.

Declaração

Esta dissertação é o resultado do meu próprio trabalho que foi desenvolvido entre Janeiro de 2007 e Março de 2012 no laboratório do Doutour Miguel Godinho Ferreira, Instituto Gulbenkian de Ciência em Oeiras, Portugal, no âmbito do Programa Doutoral do Instituto Gulbenkian de Ciência PGD2007. Este trabalho foi co-supervisionado pela Doutora Isabel Gordo, Instituto Gulbenkian de Ciência, Oeiras, Portugal. Todas as colaborações estão indicadas em cada capítulo, nas secções de 'Material and Methods' e 'Acknowledgments'.

Os Capítulos 2 e 3 descrevem resultados que foram submetidos no mesmo artigo científico: **Avelar A.T**., Gordo I. and Ferreira M.G.. Gene organization is a selectable trait maintained by antagonistic pleiotropy. Foram separados em dois Capítulos unicamente para permitir uma discussão mais aprofundada bem como a inclusão de resultados não publicados.

Os resultados do Capítulo 4 motivaram futuro trabalho que levará à publicação do resultado descrito.

Esta dissertação teve o apoio financeiro da Fundação para a Ciência e a Tecnologia (FCT) através da bolsa BD nº SFRH/BD/33214/2007 e do projecto PTDC/BIA-EVF/114460/2009.

Acknowledgments

This thesis has been the result of a very fruitful collaborative effort between my two supervisors and me. I would like to thank Miguel for all his enthusiasm throughout the years and for teaching me how to be a scientist. Thank you Miguel for all the opportunities you created for me throughout the years and for taking me as Ph.D student in the beginning of your P.I. career with a risky project. Another big thanks to Isabel Gordo for all the time she gave me during the several unscheduled interruptions and for her great excitement about the work. Thank you both!

All my laboratory "baby-steps" I owe them to my dear "*fissionary*" contemporaneous Hugo Almeida, Clara Reis and Tiago Carneiro who taught me all when I arrived in the lab and were happy to discuss with me my project. Albeit the tight space, it was a great fun to be part of the Telomere and Genome Stability lab! Thank you also to the more recent lab people, Joana Nabais, Catarina Henriques, Madalena Carneiro for all the discussion during the lab meetings. To Maria Gallo, who was my company in long hours in the lab and become a friend. To Lilia Perfeito, for her genuine interest and curiosity in biology and her helpful comments and corrections of the first version of my thesis.

Thank you to all that were available to discuss this project, in particular to the Evolutionary Biology Lab and my thesis committee Lars Jensen and José Pereira Leal.

I have a number of other people that where a great help in my daily-laboratory life. Among them are the Zeng Ho technicians Paulo Duarte and Sónia Rosa who were more than just colleagues. A big thanks to my dear Sónia for all the good moments and support. To all the people of the Zeng Ho wing that were there from 2008 to 2012 and made it a great place to work! I would also like to thank Gil Neto and in particular Renato Alves who spent a

great number of hours helping with Linux and all the programming part. I also thank Rui Gardner and Telma Lopes for all the support with FACS.

To IGC, for being such a lively and collaborative place to work. To Prof. António Coutinho and Henrique Teotónio as the former director of the IGC PhD program, for choosing me as an IGC Ph.D student and allowing me to freely express my scientific curiosities.

To my sweet friends Bárbara Jezowska, Francisca Vasconcelos, Mariana Silva and Clara Pereira who are fantastic people and made my last year a more complete one. An enormous thanks to Inês Ferreira who was my first Ph.D friend and my guidance throughout all the burocratic things I had to do.

To all the precious people “behind the scene” in particular to my companion Ben, my mum, my father and my brother who have been my constant inspiration and strength. You all know how important you are for me! Ben thanks for being present in all moments including the hardest and showing me how to move on. You are all amazing people and I am privileged to be around you! A final thanks to all my other friends, in particular Claúdia and David for showing me new perspectives.

“Evolutionary progress depends on a successful equilibrium between two antagonistic factors of the population: genetical stability and genetical variability.”

A Brito da Cunha, 1960

Table of Contents

Summary

Evolution is driven by biological diversity, which is displayed by different phenotypes. These phenotypes arise as a coordinated response to the genetic composition of each organism. Chromosomal rearrangements (CRs), such as inversions and translocations, are a type of mutation contributing both to between and within species phenotypic variation. Additionally, they are a prominent feature of several types of cancer, in particular lymphomas. However, unlike other types of mutations, the effects of inversions and translocations have not yet been directly quantified. The objective of this thesis is to quantify the mitotic and meiotic effects of CRs and to understand if chromosomal diversity is an important macromutation for the generation of biological diversity.

Initially, we asked whether chromosomal rearrangements are a polymorphic mutation in the fission yeast *Schizosaccharomyces pombe*. We found, like others, that karyotype differences are very common in *S. pombe* isolates in spite of nucleotide diversity of the order observed within species diversity. This fact led us to test the genetic isolation between the natural isolates by scoring hybrid viabilities in pairwise crosses. We found that in some cases hybrid viability was severely impaired. These results prompted us to measure the meiotic and mitotic effects of single CRs in an otherwise isogenic background.

Using *S. pombe* as a model, we generated a collection of strains containing either a pericentric inversion or a reciprocal translocation. These rearrangements were generated so that gene sequence was not altered, and, therefore, allowed us to address uniquely the weight of the rearrangements. We scored meiotic viability for different pairwise crosses of parental and rearranged strains and characterized the amount of homologous recombination. Heterozygotic crosses involving inversions and reciprocal translocations caused drops in hybrid viability up to around 50%. Homologous recombination was impaired in inversions, as expected, but not in translocations.

To quantify the mitotic fitness effects of the different rearrangements, we performed competition assays in different environments. We found a strong genotype by environment interaction on the effects of chromosome structure with changes in gene expression throughout the genome. Also, we found several cases of antagonistic pleiotropy, whereby the effects of CRs change with the mode of reproduction: CRs can have a deleterious effect at meiosis but exhibit a growth advantage during mitosis.

Finally, we asked whether the rate of adaptation in a rearranged background is different from that of its parental isolate. In order to do so, we attempted to measure the beneficial mutation rate in an inverted and a non-rearranged genotype, and found evidence that the distribution of fitness effects is dependent on the genomic background.

Our work constitutes the first direct determination of the fitness effects caused by chromosome structure and suggests that it is possible to maintain polymorphic CRs in nature despite their lethality in meiosis. This selection is environmentally dependent and occurs at gene level.

Sumário

A evolução é impulsionada pela diversidade biológica que se manifesta em diferentes fenótipos. Estes fenótipos surgem como uma resposta coordenada à composição genética de cada organismo. Rearranjos Cromossómicos (CRs), tais como inversões e translocações, são um tipo de mutações que contribuem para variação fenotípica quer entre espécies quer dentro da própria espécie. Além disso, são uma característica proeminente de vários tipos de cancro em particulares linfomas. Contudo, ao contrário de outros tipos de mutações, os efeitos de inversões e translocações não foram ainda directamente quantificados. O objectivo desta tese é quantificar os efeitos meióticos e mitóticos dos CRs e compreender se a diversidade cromossómica é uma importante macro-mutação para a geração de diversidade biológica.

Inicialmente, testámos se CRs são polimórficos na levedura de fissão *Schizosaccharomyces pombe*. Observámos, tal como outros, que diferenças cariotípicas são altamente predominantes em isolados naturais de *S. pombe*, apesar da diversidade nucleotidíca estimada estar na ordem da diversidade observada dentro de espécie. Para testar se as diferenças genómicas dos isolados naturais contribuem para isolamento genético, quantificámos a viabilidade dos híbridos em cruzamentos emparelhados. Encontrámos que, nalguns casos, a viabilidade é severamente reduzida. Estes resultados motivaram-nos a medir os efeitos meióticos e mitóticos de CRs cuja diferença no fundo genético são apenas os rearranjos.

Utilizando *S. pombe* como modelo, criámos uma colecção de estirpes contendo ou uma inversão pericentrica ou uma translocação recíproca. Estes rearranjos foram gerados de forma a que a sequência genética não fosse alterada permitindo-nos, desta forma, abordar exclusivamente os efeitos dos rearranjos. Quantificámos a viabilidade meiótica para diferentes combinações entre estirpes parentais e rearranjadas e caracterizámos a quantidade de

recombinação homóloga. Cruzamentos heterozigóticos envolvendo inversões pericentricas e translocações reciprocas causaram uma abaixamento na viabilidade dos híbridos até cerca de 50%. No caso de inversões, a recombinação homóloga é, tal como esperado, reduzida mas não no caso de translocações.

Para quantificar os efeitos de fitness mitóticos dos diferentes rearranjos, realizámos ensaios de competição em diferentes ambientes. Encontrámos uma forte interacção entre genótipo e fenótipo provenientes da estrutura cromossómica com alterações na expressão génica ao longo do genoma. Além disso, encontrámos vários casos de pleiotropia antagonística, segundo o qual os efeitos dos CRs mudam de acordo com o modo de reprodução: CRs podem ter um efeito prejudicial na meiose mas exibem uma vantagem de crescimento durante a mitose.

Por último, testámos se a taxa de adaptação varia para um fundo genético rearranjado versus um não-rearranjado. Para isso, tentámos medir a taxa de mutação benéfica num estirpe contendo uma inversão pericentrica e um genótipo não-rearranjado, e encontrámos provas de que a distribuição dos efeitos de fitness é dependente do fundo genómico. Este trabalho constitui a primeira determinação directa dos efeitos de fitness causadas pela estrutura dos cromossomas e sugerem que é possível manter CRs polimórficos na natureza apesar de sua letalidade em meiose. Esta selecção é ecologicamente dependente e ocorre no nível do gene.

List of abbreviations

BIR - Break Induced Repair

C- Control

C-GFP – Control expressing Green Fluorescent Protein

Chr - Chromosome

CI – Confidence Interval

CMT1A - Charcot-Marie-Tooth disease type 1A

CO - Crossing-Over

CR – Chromosomal Rearrangement

DSB – Double Strand Break

EMM - Edinburgh Minimal Media

ESR - Environmental Stress Response

EVT – Extreme Value Theory

FoSTES – Fork stalling and Template Switching

HNPP - Hereditary neuropathy with liability to pressure palsies

HR – Homologous Recombination

HU - Hydroxyurea

I – Inverted population

Inv – Inversion

Kb - Kilobases

LCR – Low copy repeat

Mb - Megabases

MBC - Carbendazim

MMBIR - Microhomology- Mediated Break Induced Repair

Myr – Million years

NAHR – Non-Allelic Homologous Recombination

NHEJ – Non-Homologous End Joining

NI – Natural Isolate

P- Parental

PCR – Polymerase Chain Reaction

P-GFP – Parental expressing Green Fluorescent Protein

PFGE - Pulse-Field Gel Electrophoresis

qRT-PCR – Quantitative Reverse Transcription PCR

RBM - Random Breakage Model

FBM - Fragile Breakage Model

SE – Standard Error

SPB - Spindle Pole Body

T- Translocation

TE – Transposable Element

WC- PFGE- Whole-Chromosome Pulse Field Gel Electrophoresis

YES – Yeast Extract with Supplements

Chapter 1 – General Introduction

1. Chromosomal Rearrangements.

Biological diversity is manifested in many different formats and solutions. All this diversity is encoded in the genetic background of each organism. Scientists have for long dedicated most of the research to study how the linear sequence of DNA is translated into phenotypes. Nowadays, this perspective is moving towards the analysis of other types of regulation. For instance, one of those is epigenetics, the study of the modifications that occur on top of the DNA linear sequence and regulate several processes such as gene expression and checkpoint activation. However, one very important, and often neglected aspect of genetic diversity, is the structure of the linear sequence itself. Variations in structure will change the position of genes and consequently their genetic and molecular environment. Therefore, alteration of the chromosome structure should inevitably lead to biological diversity.

Chromosomal Rearrangements (CRs) are a class of mutation contributing to variation both within and between species. Additionally, many human diseases are caused by CRs rather than point mutations within genes. For example, duplications and deletions are responsible for several human pathologies such as α –thalassemia (Lupski, 1998) and translocations are a predominant feature of several cancers in particularly of many human lymphomas (Nambiar and Raghavan, 2011). CRs are in fact a major characteristic of cancer cells. In this particular environment, genomic variation might be a major driver for the different solutions to escape cellular control found by this type of cells. In spite of this, chromosomal structure has not yet been properly addressed as a form of phenotypic diversity. Two very important questions are whether changes uniquely in structure lead to different phenotypes and if so, is this through the alteration in the regulation of genes?

Another very relevant aspect of CRs is their presence as polymorphism within and between species. However, CRs are known to be deleterious at the meiotic level in organisms with sexual reproduction. This makes it surprising that CRs exist as polymorphisms and are segregating in natural populations

such as in *Drosophila*. A major unsolved question is how chromosomal variation can be maintained in nature in spite of strong deleterious effects at meiosis.

Given all the above, the study of chromosome structure remains as a gap in the understanding of biological diversity. In order to filling parts of this gap, I have used the fission yeast *Schizosaccharomyces pombe* as an experimental model to understand how CRs can contribute to biological diversity.

1.1 Types of chromosome rearrangements.

CRs are changes in the number and/or organization of the linear DNA sequence that alter the structure of individual chromosomes. CRs exist in the form of deletions, duplications, aneuploidies, transpositions, inversions and translocations (Figure 1.1). A deletion is a loss of chromosomal segments which can vary in size from thousands to hundreds of thousands of base pairs (Figure 1.1a) whereas duplication is a mutation in which part of the chromosome has been doubled (Figure 1.1b). The doubling or deletion of the entire chromosome(s) is termed aneuploidy (Figure 1.1c).

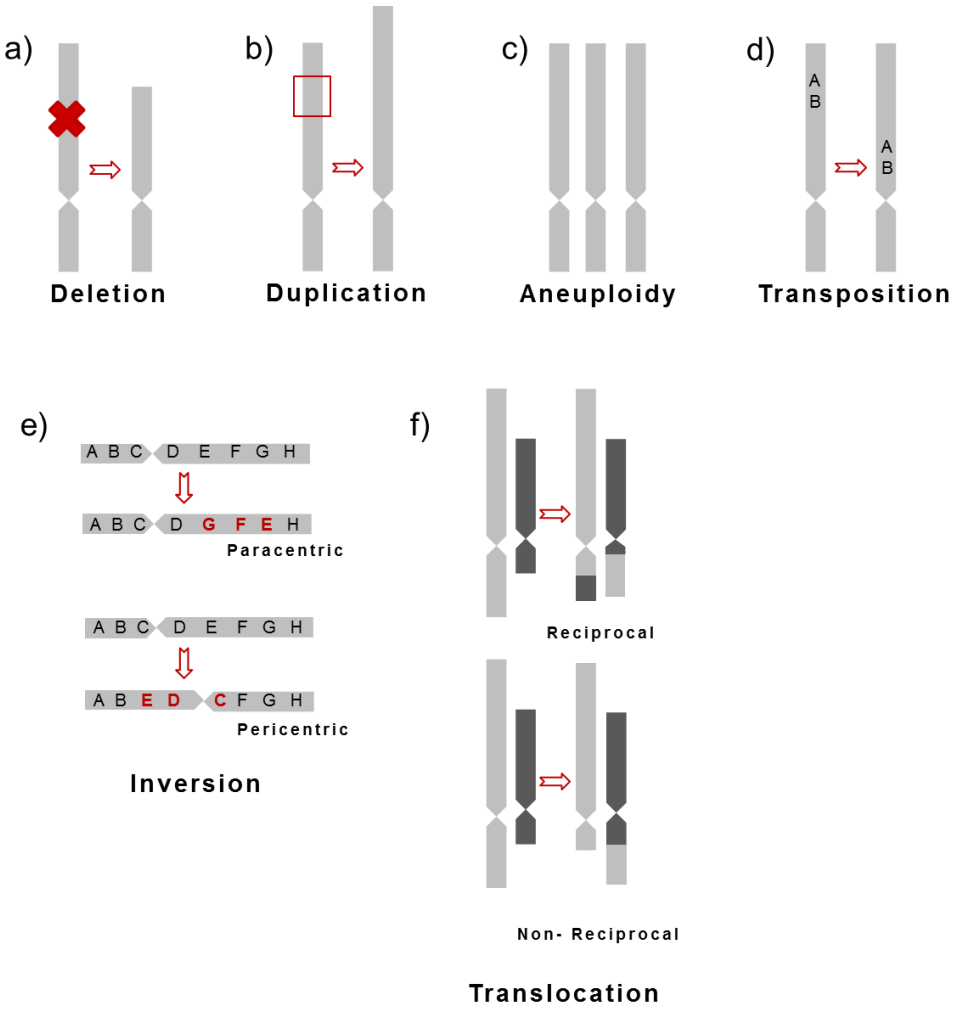


Figure 1.1. **Chromosomal rearrangements**. **a**, duplication. **b**, deletion. **c**, aneuploidy. **d**, transposition. **e**, inversion. **f**, translocation.

In transpositions, inversions and translocations, DNA is moved from its original place. In the case of transpositions, short DNA segments are moved from one position in the genome into another (Figure 1.1d). These short segments are termed transposable elements (TEs) and can perform the transposition itself by use of an encoded transposase enzyme (Alberts *et al*., 2002).

Inversions are a 180 degrees movement of a segment of DNA and consequently change the order of genes along the chromosome (Figure 1.1e). Inversions can be pericentric if they include the centromere or paracentric if they do not include the centromere. Lastly, translocations correspond to the movement of genetic material between non-homologous chromosomes or within the same chromosome (Figure 1.1f). Translocations are designated as reciprocal when there is movement of DNA from both chromosomes involved in the translocation event. Robertsonian translocations are a particular case of reciprocal translocation where there is reciprocal exchange between two acrocentric chromosomes (chromosomes in which the centromere is located quite near one of the ends of the chromosome) to generate a large metacentric chromosome (a chromosome where the centromere is located at the middle of the two arms which have roughly the same length) and small chromosome. Non-reciprocal translocations occur if only a segment from one of the chromosomes is dislocated the translocation (Figure 1.1f). For clarification consider the following non-homologous chromosomes AB●CDEFG and MN●OPQRS where ● represents the centromere and the letters represent chromosome loci. If the DNA fragment EFG moves from the first to the second chromosome without any transfer of DNA from the second chromosome, a non-reciprocal translocation takes places producing the AB●CD and MN●OPEFGQRS chromosomes. A two-way exchange of fragments between the two chromosomes results in a reciprocal translocation that can, for instance, give rise to chromosomes AB●CDQRS and MN●OPEFG.

The different types of CRs have different properties. Duplications, deletions and aneuploidies will inevitably lead to changes in DNA content, while in transpositions, inversions and translocations this occurs in the breakpoint region. However, gene order is always altered in transpositions, inversions and translocations.

1.2 Natural diversity and rates of chromosomal evolution.

The first chromosomal rearrangement segregating in natural populations was described by Dobzhansky and Sturtevant in *Drosophila pseudoobscura* by cytological identification of meiotic loops in polytene chromosomes (Dobzhansky and Sturtevant, 1938). Since then *Drosophila* is the most well characterized genus. An estimated three quarters of all the species in the genus are polymorphic for paracentric inversions (Krimbas and Powell, 1992). *D. pseudoobscura* populations are a remarkable example which displays high levels of polymorphism for inversions associated with environmental adaptation (Anderson *et al*., 1991; Schaeffer *et al*., 2003). Increasing evidence for selection on chromosomal inversions exists for a number of other organisms such as the mice (Lyon, 2003), mosquitos (Coluzzi *et al.*, 2002) and humans (Stefansson *et al.*, 2005). In yeast, CRs such as duplications and translocations, are often part of an adaptive response of yeast subjected to different environmental (Dunham *et al*., 2002; Infante *et al*., 2003) and stress conditions (Adams et al., 1992; Dhar et al., 2011). This type of chromosomal variation is widely present in industrial strains (Codón and Benitez, 1995; Querol *et al.*, 2003; Akao *et al.*, 2011).

The most recent examples of inversion polymorphisms associated with ecotype adaptation occur in the yellow monkeyflower *Mimulus guttatus* (Lowry and Willis, 2010) and in the threespine sticklebacks (Jones *et al.*, 2012). In the yellow monkeyflower *Mimulus guttatus*, two different arrangements of a chromosomal locus occur in either the annual or the perennial populations distributed in western North America. This inversion polymorphism is associated with traits involved in the life-history transition and consequently local adaptation as well as with three ecological reproductive isolating barriers (Lowry and Willis, 2010). In the case of threespine stickleback, populations have emerged in different locations repeatedly. This recurrent evolution makes it possible to characterize the divergent loci and gene expression changes between the marine and freshwater populations (Jones *et al.*, 2012). Jones *et al.* show that

a polymorphic inversion differentiating the two populations changes the exons of *KCNH4*, a potassium transporter gene which encodes for a voltage-gated potassium channel protein. This causes a duplication of a 3' exon that likely leads to different transcripts and gene products in the different populations, therefore attributing *KCNH4* with ecotype-specific isoforms (Jones *et al*., 2012).

CRs are also present between different species. Nowadays, extensive literature exists describing these differences among several groups of organisms as a way to study the mechanisms behind chromosomal evolution. For instance, whole genome inference studies propose that CRs were involved in the steps that lead to the modern *Saccharomyces cerevisiae* from the *Saccharomyces* ancestral (Gordon *et al*., 2009). Comparative genomic studies use synteny blocks in order to compare genomes from different species. However, synteny has been used in the literature with two different meanings. In the time of comparative genetic maps, synteny originally described the co-localization of several markers along the same chromosome. Nowadays with the advance of genome sequencing, synteny refers to the preservation of gene order between homologous genes along chromosome segments in different species. The presence of synteny blocks between different species makes it possible the use of algorithms that search for the most likely ancestral karyotype based on which CRs might have occurred between the species in study. Therefore, one can make predictions on the molecular mechanisms and types of CRs that separate such species.

Through the use of synteny blocks, Ranz and colleagues compared the largest chromosomal element from *Drosophila repleta* with its homologous in *Drosophila melanogaster*, the Muller's element E (Ranz *et al*., 2001). A Muller element (A to F) corresponds to one of the six chromosomal arms in *Drosophila* which can be found in all the species of the genus. Muller elements share high levels of synteny blocks but have been internally rearranged mostly by the fixation of paracentric inversions (Bhutkar *et al*., 2008). Ranz *et al.* report-

ed 114 ±14 fixed paracentric inversions between *D. repleta* and *D. melanogaster* which have diverged 40-62 million years (Myr) and represents the more diverged lineages inside the *Drosophila* genus (Ranz *et al*., 2001). Until then a rate of 0.06567 disruptions/ Mb/ Myr was the highest rate of CRs inferred. However, a whole genome comparative study with the genomes from twelve different *Drosophila* species (Bhutkar *et al.*, 2008) estimated a much higher rate than previously calculated (Ranz *et al*., 2001; González *et al*., 2002; Bartolomé and Charlesworth, 2006) but still within the same order of magnitude depending on the pair of species considered and on the particular chromosomal region. For example, considering the same chromosome size and divergence time as Ranz *et al.* (Ranz *et al*., 2001), Bhutkar obtain a 0.165 disruptions/Mb/Myr for Muller element E whereas Ranz *et al.* had estimated rate 0.0848 disruptions/Mb/Myr for the equivalent.

Yeasts along with *Drosophila* have been extensively characterized by their enormous chromosomal reshuffling with rates that are quite variable between the different phylogenetic lineages (Liti and Louis, 2005; Fischer *et al*., 2006; Dujon, 2010). Prior to the whole genome sequencing, Fischer *et al.* characterized the CRs that exist between the *sensu stricto* group (Fischer *et al.,* 2000). Taking the *S. cerevisiae* genome as reference, Fischer described one non-reciprocal translocation with *Saccharomyces bayanus* and nine reciprocal translocations with *S. bayanus* and *Saccharomyces cariocanus*, involving 13 out of the 16 *S. cerevisiae* chromosomes using Pulse-Field Gel Electrophoresis and Southern blot analysis (Fischer *et al*., 2000). Interestingly, the two strains of *Saccharomyces mykatae* used in this study contain two different reciprocal translocations despite involving one of the same chromosomes. A whole genome sequence comparison study appeared later using eleven yeast species representing the entire Hemiascomycete phylum (Fischer *et al*., 2006). Phylogenetic analysis of six representatives from each of the different lineages revealed specific genome rearrangement rates for each lineage. The *Debaryomyces hansenii* branch, in which *Candida albicans* is included,

showed the highest rate for gene inversion with an average of 35% of inversion for all genes analysed. In fact, the opportunist pathogens *C. albicans* and *Candida glabrata* show the more unstable genomes (Fischer *et al*., 2006) which supports the observations from laboratory and clinical isolates (Selmecki *et al*., 2006, 2009; Poláková *et al*., 2009).

Other experimental organisms have been used for comparative genomic studies such as the soil nematodes *Caenorhabditis briggsae* and *Caenorhabditis elegans*. These two species have an estimated divergence time of around 100 Myr yet they remain morphologically highly similar and occupy the same ecological niche (Cutter *et al.*, 2009). Also, they contain the same chromosome number and despite extensive colinearity given by large blocks of synteny, these two nematodes contain marked CRs between them. A first study by Coghlan and Wolfe using only the 13% of known sequence of *C. briggsae* at the time, estimated a rearrangement rate of 0.4–1.0 chromosomal breakages/Mb/Myr and inferred a ratio of translocations to inversions to transpositions to be 1:1:2 (Coghlan and Wolfe, 2002). A latter study by Hillier *et al.* has not confirmed the presence of translocations and reported a much lower number of transposition events (Hillier *et al*., 2007). However, Hillier's initial analysis found 244 translocation events which were excluded after a more stringent analysis which eliminates repetitive sequences. It is possible that Coghlan and Wolfe by extrapolating the CR rate to the whole genome based solely on 13% sequencing at the time lead to an overestimation of the rate. However, despite excluding repeated sequences, Hillier's analysis covers all the genome of both species. Nonetheless, both studies agree that in *Caenorhabditis* the chromosomal arms are undoubtedly more recombinogenic and sustain CRs as opposed to the centers, which are more static.

Several estimates for the rate of CRs have also been done in the mammalian lineage. As observed for other taxa, the rate of CRs is different across the different branches of the mammalian lineage. Bourquet *et al.* performed a syntenic block analysis and estimated a rate of 3.2 and 3.5 chromosomal rear-

rangements/Myrs on the mouse and rat branch, respectively, from the murid rodent ancestor (Bourque *et al*., 2004). A much smaller rate of 1.6 chromosomal rearrangements/Myr was obtained for the human branch using the murid rodent ancestor (Bourque *et al.*, 2004). This analysis implies that rearrangements occur much more frequently in murid rodents than in humans. Interestingly, they do seem to have a high tendency for chromosomal rearrangements as witnesses by the immense chromosomal radiation of the house mice in Madeira Island (Britton-Davidian *et al.*, 2007).

An extension of the analysis of Bourquet *et al.* was made by Murphy *et al.* who compared the rate of chromosomal evolution between eight mammalian species (Murphy *et al.*, 2005). The increased number of species analysed (cat, cattle, dog, pig, horse and the already rat, mouse and human) allowed Murphy and colleagues to infer the genome architecture of the ancestors of three mammalian lineages. The authors deduced many lineage specific rearrangements, in particular inversions for the case of primates. Recent work has reinforced the importance of chromosomal inversions as a source of variation in primate genome evolution. Using a genome wide comparison of chimpanzee and human, Feuk and colleagues validated a minimum of 23 inversions bigger than 25 kb differencing this two (Feuk *et al*., 2005). Moreover, the 23 inversions were then tested against human samples and 3 out of the 23 are polymorphic with one allele matching the human configuration and other matching the chimpanzee configuration. As it is the case for other organisms, inversions are the dominant CR with very few translocations predicted (Zhao and Bourque, 2009). Nowadays, the current global estimate for the rate of CRs in the vertebrate lineage is 2 rearrangements/Myr (Drillon and Fischer, 2011) which is close to the above described estimates for the mammalian genome evolution (3.2, 3.5 and 1.6 chromosomal rearrangements per million years on the mouse, rat and human branch, respectively, from the murid rodent ancestor (Bourque *et al*., 2004, 2005). An emergent theme from all works across the different taxa is that the rate of rearrangements is different across

the different lineages and organisms (Fischer *et al*., 2006; Zhao and Bourque, 2009; Drillon and Fischer, 2011).

1.3 Mechanisms behind the formation of Chromosomal Rearrangements.

As pointed out before, CRs are at the basis of several genomic disorders in human pathologies. Consequently, the molecular mechanisms that lead to the generation of CRs have been widely studied in human cells (Gu *et al*., 2008) albeit similar mechanisms being responsible for CRs in other organisms (Mieczkowski *et al.*, 2006). To date, three major types of mechanisms have been proposed: Non-Allelic Homologous Recombination (NAHR), Non-Homologous End Joining (NHEJ) and Fork Stalling and Template Switching (FoSTeS) also known as Microhomology-Mediated Break Induced Repair (MMBIR).

Non-Allelic Homologous Recombination is a type of homologous recombination that occurs between repetitive homologous sequences of DNA which are not alleles (Figure 1.2). This ectopic recombination occurs mainly through ribosomal and transfer RNA genes, TEs, segmental duplications, multigene families and repeats associated with telomeres and centromeres using the same type of cellular machinery as meiotic recombination. In human genomes NAHR seems to be the predominant mechanism for the generation of CRs in particular for recurrent rearrangements (Bondeson *et al.*, 1995; Kidd *et al*., 2008). Recurrent rearrangements are rearrangements in which the same fragment and breakpoint responsible for the pathology are present in unrelated individuals. In NAHR, the alignment between the repeats is dependent on the size of the repeats and the distance between them but also on the degree of homology and orientation with respect to each other (Gu *et al*., 2008). Depending on the orientation of the repeats, NAHR can lead either to duplications and deletions or to inversions. If the sequences are located on the same chromosome in the same orientation, recombination between the

direct repeats will result in a deletion or duplication of the fragment between the loci (Figure 1.2a). If on the other hand they are in inverted orientations the consequence is an inversion of the segment between the two loci (Figure 1.2b). NAHR can also lead to translocations if the repeats are on different chromosomes (Ou *et al.*, 2011) (Figure 1.2c).

Inversions produced by NAHR are responsible, for instance, for the disruption of the factor VIII gene which leads to severe haemophilia A (Lakich *et al*., 1993). Another well described example of the different outcomes of NAHR using the same sequences are two human neuropathies, the Charcot-Marie-Tooth disease type 1A (CMT1A) and Hereditary neuropathy with liability to pressure palsies (HNPP) (Lupski *et al*., 1991; Pentao *et al.*, 1992; Chance *et al*., 1993, 1994). CMT1A results from the heterozygous duplication of a 1.4 Megabases (Mb) fragment whereas HNPP results from a heterozygous deletion of the same genomic segment (Lupski and Stankiewicz, 2005).

Ectopic recombination has also been pointed out as the possible mechanism giving rise to most of the paracentric inversions that separate *D. pseudoobscura* and *D. melanogaster* (Richards *et al.*, 2005). In yeast from the *sensu stricto* group, duplicated genes, TEs and tRNAs have been found at the breakpoints of rearrangements between the different species which indicates that they may have originated from ectopic recombination (Fischer *et al.*, 2000; Kellis *et al*., 2003; Mieczkowski *et al*., 2006). In addition, recombination between the transposable Ty elements in yeast is a common mechanism that leads to the formation of CRs in laboratory evolution experiments (Dunham *et al*., 2002) and in wine strains of yeast (Rachidi *et al*., 1999).

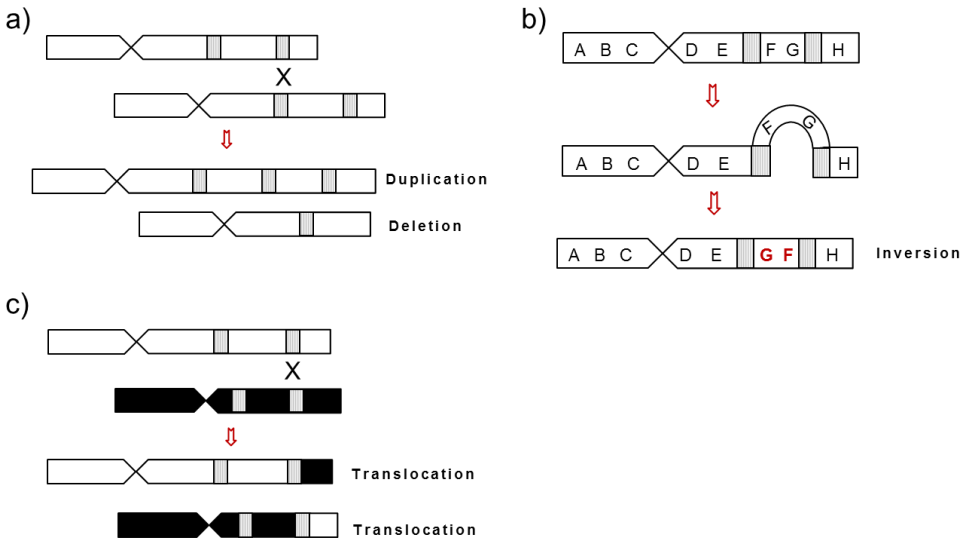


Figure 1.2. **Non-Allelic Homologous Recombination is involved in the formation of CRs**. NAHR occurs between repetitive sequences, such as low-copy repeats (LCR) and depending on the orientation and length of the LCR it leads to different CRs. **a**, deletions and duplications. **b**, translocations. **c**, inversions. Squares with grey vertical bars represent the LCRs.

Other types of molecular mechanisms than ectopic recombination are responsible for non-recurrent rearrangements in which breakpoints are not shared among unrelated individuals in pathological conditions. Some types of non-recurrent rearrangements can occur through NHEJ which requires the creation of a double strand break (DSB) (Figure 1.3). NHEJ is one of the two major repair mechanisms for DSBs in several different organisms, including bacteria, yeast and humans (Lieber, 2010). DNA ends generated by DSBs caused by exposure to DNA damaging agents do not have ligatable termini. Random DNA double strand breaks are known to occur due to oxidative free radicals, ionizing radiation or topoisomerase failure. These ends are minimally processed by specific proteins that are part of the NHEJ-mechanism and physical juxtaposed for the ligation. NHEJ ligates DNA ends with little or no homology. Consequently, it can lead to major errors if the cell experiences

more than one DSB. In this situation, previously unlinked DNA molecules may be joined resulting in CRs.

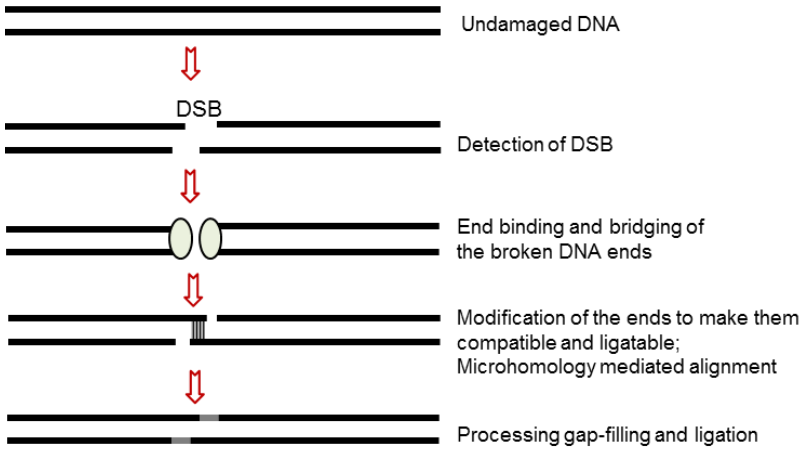


Figure 1.3. **Non-Homologous End-Joining**. When a double-strand break (DSB) occurs it can be repaired by NHEJ. During this mechanism of repair, some addition or deletion of bases may be required. The two thick black lines represent the two DNA strands. Grey circles represent NHEJ-specific proteins. Adapted from (Gu *et al*., 2008).

DSBs also occur as part of a physiological response in somatic cells of the vertebrate immune system (Nambiar and Raghavan, 2011). In these cases, DSBs are formed during the V(D)J recombination mechanism that occurs at lymphocytes as a mechanism for generating the diversity of receptors for binding to antigens. The incorrect occurrence of the V(D)J recombination mechanisms leads to the formation of DSBs in ectopic places and erroneous NHEJ leading to lymphomas (Aplan, 2006).

Despite their predominant role in the formation of CRs, NAHR and NHEJ are not sufficient to explain all rearrangements observed in nature and even at the breakpoints of complex rearrangements in pathological processes. The presence of duplicated segments at breakpoints suggests that another mechanism might be involved (Murphy *et al.*, 2005; Ranz *et al.*, 2007).

Inverted duplicated sequences at breakpoint regions have been described for the polymorphic inversion In(3R)P in *D. melanogaster* (Matzkin *et al*., 2005), the polymorphic inversion In(2L)a in *Anopheles gambiae* (Sharakhov *et al.*, 2006) and for the pericentric inversion fixed between *Pan troglodytes* chromosome 10 and the homologous *Homo sapiens* chromosome 12 (Kehrer-Sawatzki *et al*., 2005). In these cases, duplications seem to arise as a consequence of the inversions. Ranz *et al.* characterized the sequences at the breakpoints of inversion In(3R)84F1;93F6-7 in *D. melanogaster* and *D. yakuba* and came up with a new probable molecular mechanism that lead to its formation (Ranz *et al*., 2007). The authors proposed a model whereby staggered breaks in structurally unstable genomic regions are the initiating factor (Figure 1.4). According to this model, after the production of two staggered breaks at different locations of the same chromosome, the repair mechanism does not re-join 5' ends to their own 3' but to the 3' end of the other breakpoint. Later, the remaining gaps are filled up and inverted duplications originate from both ends of the inversion.

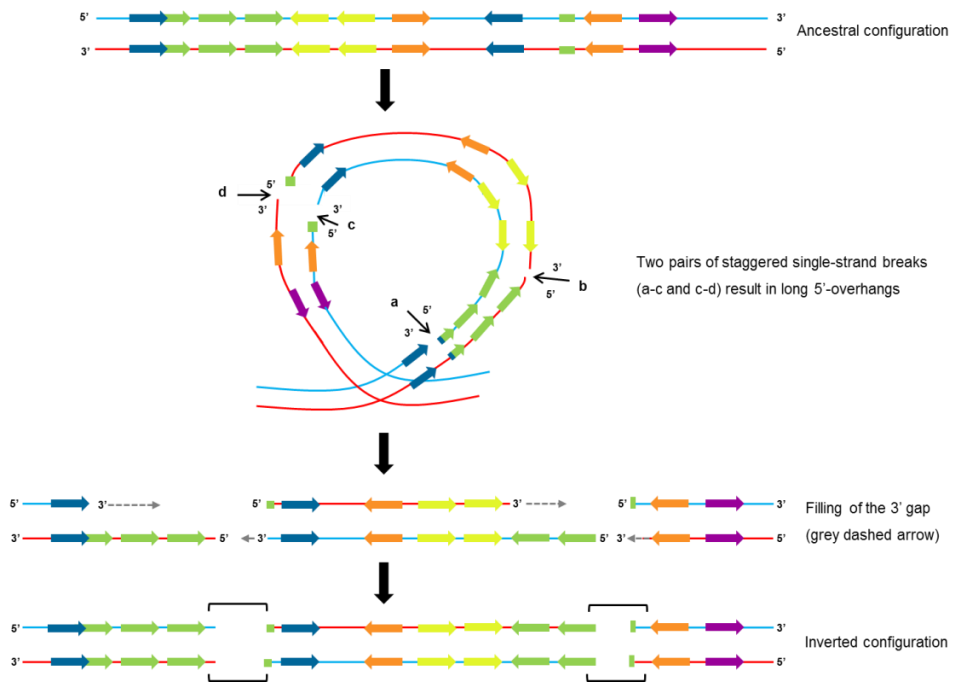


Figure 1.4. Isochromatid model proposed by Ranz and colleagues with staggered single-strand breaks giving rise to an inversion. Two pairs of staggered single-strand breaks (a-b and c-d) result in long 5'-overhangs, which can then be filled in (grey dashed arrow); when followed by non-homologous end joining, this may result in an inversion flanked by inverted duplications of the sequences between the paired single-strand breaks. Adapted from (Ranz *et al*., 2007).

Other more complex replication-based mechanism has been proposed based on observations from non-recurrent rearrangements associated with genomic diseases such as the Pelizarus- Merzbacher disease (Lee *et al.*, 2007; Hastings *et al*., 2009). Breakpoint junctions in these cases present very complex sequences that seemed to result from a skipped backward and then skipped forward DNA replication after a replication fork collapse. The proposed model is termed FoSTeS/MMBIR and is an extension of the Break-Induced Repair (BIR) mechanism (Figure 1.5). BIR is a replication-based repair mechanism specific for restarting replication at collapsed replication

forks. The initial stalling of the replication fork is caused by unstable genomic regions such as low copy repeats (Lee *et al.*, 2007), and palindromic DNA or stem-loop structures (Bhutkar *et al.*, 2008; Nambiar and Raghavan, 2011; Liu *et al*., 2012). BIR is initiated when the replicative helicase encounters a nick on the template strand and it is an accurate type of repair that involves long stretches of homology between the DNA sequences. BIR at DSBs has been shown to explain some chromosomal translocation events in genetic mutations of yeast (Haber, 2006). Even though BIR is very similar to FoSTeS/MMBIR, FoSTeS is an inexact mechanism based solely on microhomology and consequently liable to CRs. The type of CR generated by these mechanisms depends on the strand that will be invaded. For instance, translocations would be formed is the switch goes to a different chromosome and deletions when there is a switch to a position ahead of the fork collapse (Hastings *et al*., 2009).

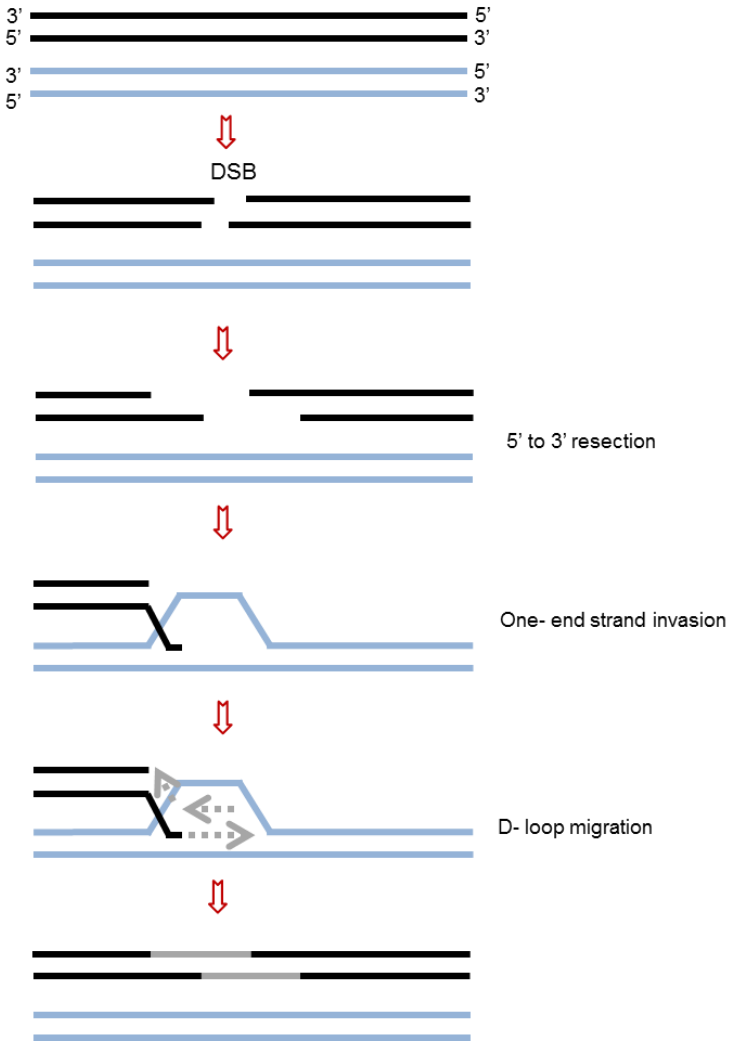


Figure 1.5. Break-induced Repair. A nick on the template strand sensed by the replicative helicase is the trigger to initiate BIR. The 5' end of the broken arm is resected by an exonuclease to leave a 3' overhang that invades a homologous sequence. This 3' end primes DNA synthesis and establishes a new replication fork consisting of both leading and lagging strand synthesis using the invaded sequence as template. Due to its low processivity, the newly extended DNA sequence separates and its 3' end reinvades again the homologous sequence and the process is repeated. After cycles of invasion, extension and separation the replication fork becomes more processive and replication proceeds until the end of the chromosome arm or the replicon. Adapted from (Hastings *et al.*, 2009).

1.4 Models of chromosomal evolution.

Models for chromosomal evolution are predominantly based on studies from comparative genomics. Several of these studies observe the same region of a chromosome to be used in the formation of different CRs. This breakpoint reuse might be a consequence of the use of the mechanisms described in the previous section. Even though the causes of breakpoint reuse remain unclear they are frequently associated with segmental duplications or TEs. Apart from unstable or repetitive sequences in the formation of CRs, physical proximity between the breakpoints can also be a major factor (Wijchers and de Laat, 2011). For instance, formation of translocations requires interaction with partner chromosomes. Several studies have shown a positive correlation between spatial proximity of regions that frequently undergo translocations and translocation frequency (Bickmore and Teague, 2002; Cornforth *et al.*, 2002; Parada and Misteli, 2002).

A model for chromosomal evolution, whether random or based on hotspot areas, is still under large debate. For a long time the Random Breakage Model (RBM) prevailed as the accepted form of genome evolution (Nadeau and Taylor, 1984). This model, as the name indicates, postulates that rearrangements occur randomly in the genome without any particular hotspot sites. However, initial studies of Pevzner and Tesler tested the RBM comparing at the time the draft genomes of human and mouse (Pevzner and Tesler, 2003). The high amount of very small synteny blocks allowed the authors to refute the model since it does not fit in the predicted exponential distribution of the RBM. Unlike the RBM, Pevzner and Tesler proposed a new model, the Fragile Breakage Model (FBM) suggesting that rearrangements occur in particular genomic areas (hotspots) which are fragile and more prone to breakage. A non-random breakage model for the generation of CRs is supported by work in amniotes by two independent groups (Bourque *et al*., 2004; Larkin *et al*., 2009). Another important observation extracted from their work is the amount of breakpoint reuse between humans and mice (Pevzner

and Tesler, 2003). Breakpoints in phylogenomic analysis correspond to the sequence between two synteny blocks. Consequently, breakpoint reuse signifies that the same genome regions rather than few specific nucleotides are involved in the rearrangement. Breakpoint reuse has been observed by other comparative sequence analysis in mammals (Murphy *et al*., 2005) and *Drosophila* (Ranz *et al*., 2007). The work developed by Murphy and colleagues has described an initial very high rate of 20% breakpoint reuse in the mammalian lineages (Murphy *et al*., 2005) which suggests that the same genomic region in different mammalian lineages is giving rise to several independent rearrangements. However this number is now set around 8% (Ma *et al.*, 2006; Larkin *et al*., 2009) which is lower thanks to the higher resolution of map comparisons.

As more genomes are being sequenced with higher resolution, tests to both models can be achieved and increased models are even appearing (Alekseyev and Pevzner, 2010). In addition it will probably be important to merge the ideas from genomic sequences prone to rearrangements with transcriptional activity at those areas. For instance, an emerging idea is that chromosomal size (Bickmore and Teague, 2002), the relative positioning of chromosomes (Meaburn *et al*., 2007) and chromatin composition (Fudenberg *et al*., 2011) influence the formation of CRs.

1.5 Implications of chromosomal rearrangements in meiosis and speciation.

In nature, the propagation of an individual's genetic background occurs by sexual or asexual reproduction. In organisms with sexual reproduction, the process of formation of gametes is called meiosis. Meiosis is a cellular process whereby whole genome replication is followed by two nuclear divisions.

Meiosis is important for the generation of genetic diversity and it involves, for most known organisms, obligatory homologous recombination

(HR) which shuffles alleles between chromosomes. In evolutionary terms, recombination is also important for the acceleration of adaptation as it generates combinations of alleles in the same genetic background that would otherwise have to appear independently in asexual divisions. HR is fundamental for guaranteeing the correct segregation of chromosomes during the two rounds of nuclear divisions: reductional segregation of homologous chromosomes in the first division and equational segregation of sister chromatids in the second division. In the absence of meiotic recombination homologs missegregate and the resulting aneuploid gametes give rise to defective or unviable progeny.

Meiosis is divided in two major steps denominated Meiosis I and Meiosis II (Figure 1.6). In the first step, homologous chromosomes are duplicated and condensed followed by alignment at the metaphase plate. At this stage chiasma occur between homologous allelic regions which will undergo recombination. Chiasmata (singular, chiasma) are structures that allow for the physical alignment and conjoining of chromosomes to resist the bipolar pull of the spindle apparatus. The tension created by the spindle serves as a signal for the correct orientation of the chromosomes and for proper segregation. Without recombination, homologs are not joined, tension is not generated and chromosomes eventually move to the poles at random. Moreover, as homologous chromosomes interact along their entire length, recombination between non-allelic loci with similar sequences is reduced thus avoiding ectopic recombination (Niwa *et al*., 2000; Davis and Smith, 2003).

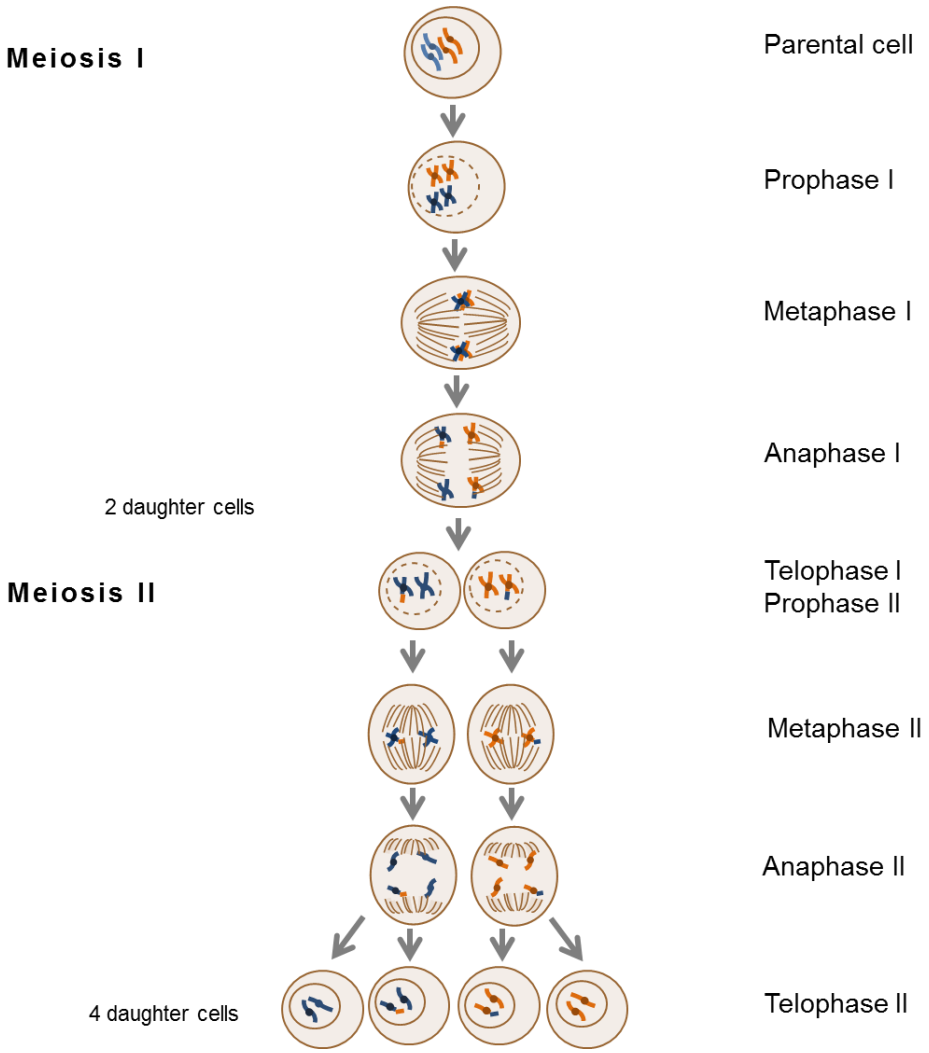


Figure 1.6. **Meiosis**. For simplicity, the parental cell contains two homologous chromosomes; the blue lines represent the parental and the orange the maternal chromosomes. At the beginning of meiosis, the chromosomes replicate and each consists of two closely associated sister chromatids. This is followed by condensation of the chromosomes (Prophase I). The homologs then pair and form the chiasmata allowing for crossing-over. The homologous chromosome pairs line up in the equator of the spindle (Metaphase I) and separate (Anaphase I). This is followed by cytokinesis resulting into two daughter cells (Telophase I). In the second phase of meiosis, the sister chromatids that form each chromosome are separated leading to four daughter cells with haploid content. Adapted from (Alberts *et al*., 2002).

At meiosis the pairing of the homologous chromosomes and the formation of the chiasmata creates a bivalent structure. In the case of heterozygotic carriers of CRs, the chromosomes are not homologous throughout the length of the entire chromosome which creates problems at Prophase I. For example, reciprocal translocations form a tetravalent structure with their respective normal homologous chromosome, allowing for complete homosynapsis of the chromosomes involved (Figure 1.7). In a normal situation in bivalents, homologous centromeres separate and move towards opposite poles in Anaphase I. However, in tetravalents, Anaphase I can occur in one of the following forms:

- Alternate segregation of centromere, which leads to the production of dyads with balanced karyotypes and four viable gametes after the second meiotic division (Figure 1.7a) and thus to the best case scenario.
- Adjacent type I, where segregation of adjacent centromeres causes the formation of unbalanced karyotypes and consequently of four dead gametes at the end of Meiosis II. This is the most common segregation type in humans for reciprocal translocations (Anton *et al*., 2008) (Figure 1.7b).
- Adjacent type II, is similar to adjacent type I of segregation (Figure 1.7c).
- 3:1 disjunction: leads to the formation of a cell with three chromosomes and another with only one chromosome at the end of Anaphase I (Figure 1.7d). This type of disjunction is rare.

If, for instance, there is a single or unequal number of chiasmata in the region between a centromere and the translocation breakpoint followed by alternate segregation, the final outcome will be four gametes where two will contain unbalanced information and the other two will have either the wild-type or the rearranged configuration (Loidl *et al.*, 1998)(Figure 1.7e).

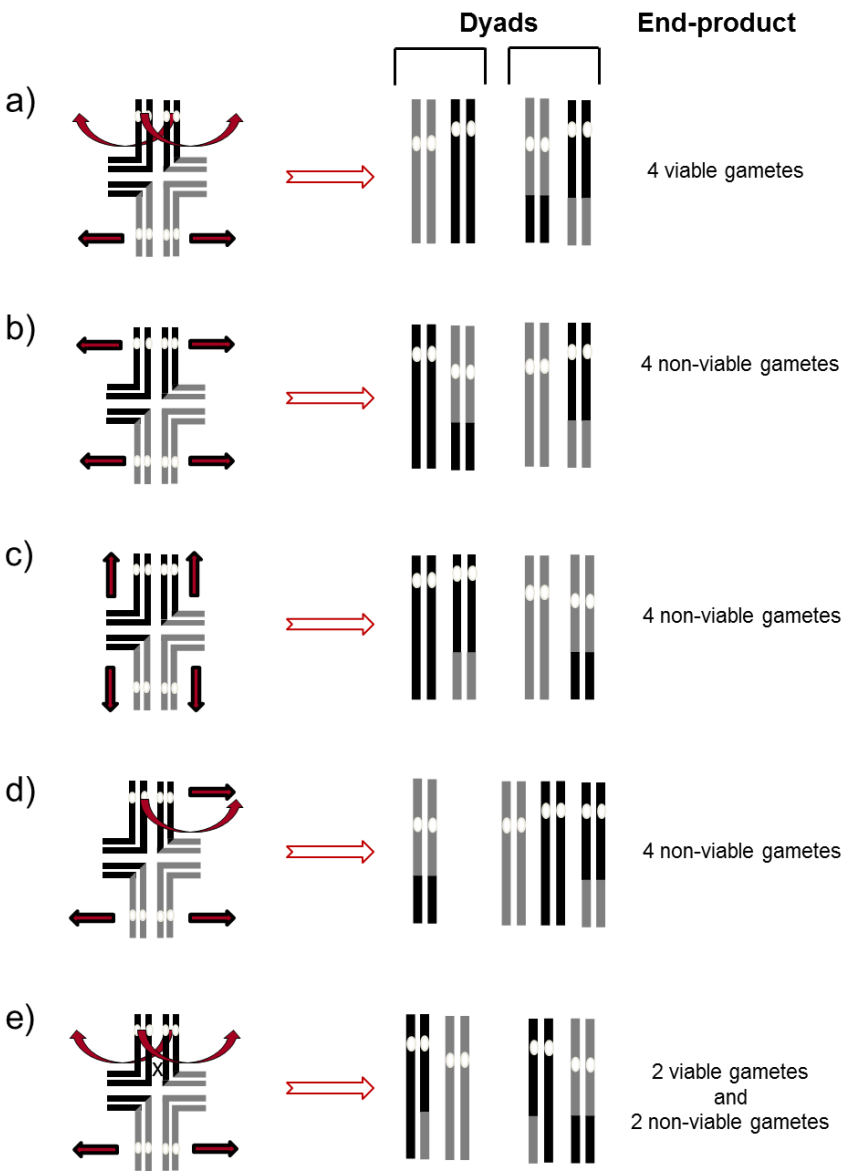


Figure 1.7. Segregation of homologous chromosomes in tetravalent structures. Tetravalents are formed at Prophase I of Meiosis I in heterozygotic crosses with reciprocal translocations. Depending on the resolution of the centromeres, meiotic products will contain different types of DNA content. **a**, alternate-segregation. **b**, adjacent-type I. **c**, adjacent-type II. **d**, 3:1 disjunction **e**, alternate segregation with chiasmata. Rectangles represent chromosomes, white circles represent the centromeres. Adapted from (Loidl *et al*., 1998).

Other types of CRs such as inversions cause different problems at meiosis when in heterozygosity (Figure 1.8). Upon the pairing of homologous sequences, the inverted regions will form a loop with the homologous sequences in the collinear chromosome. Crossing-overs outside the inversion or an even-number of crossing-overs will cause no problems. However, a single crossing-over inside the loop or a uneven crossing-over inside the inversion will result in the production of duplications and deletions and consequently, unviable gametes. The type of karyotypes will be different depending on whether the inversion is paracentric (Figure 1.8a) or pericentric (Figure 1.8b).

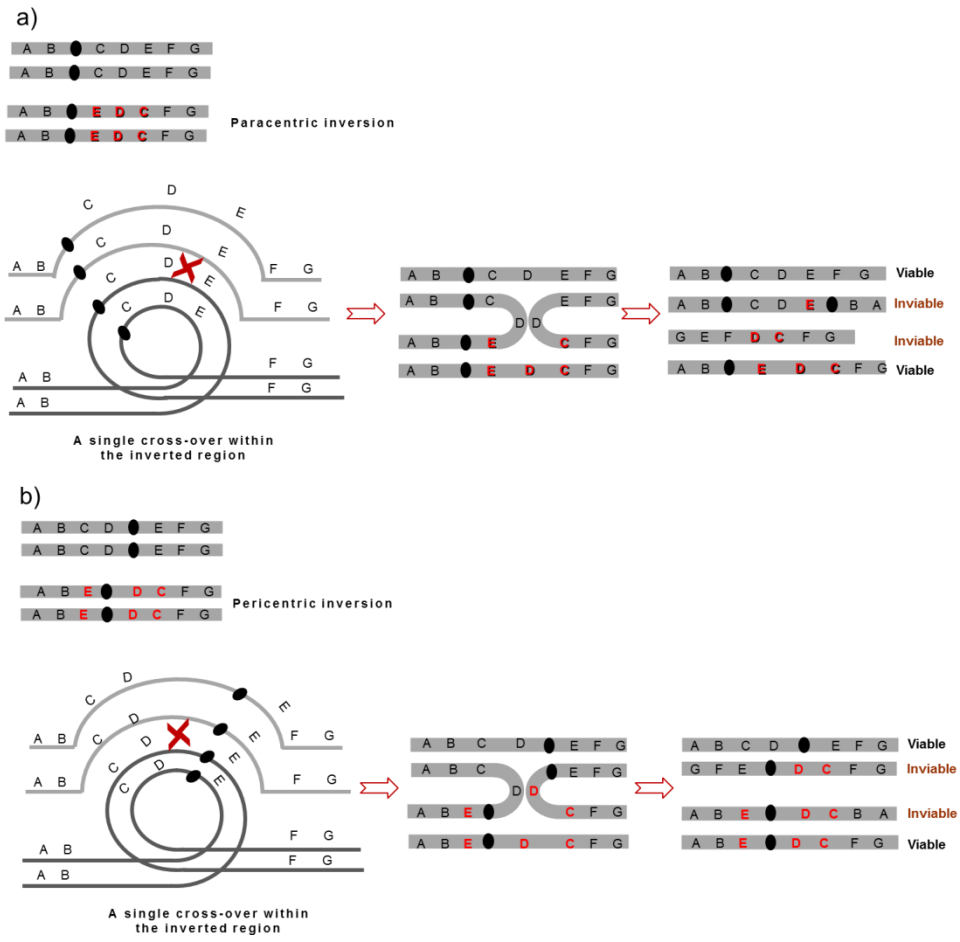


Figure 1.8. **Meiotic pairing of inversions**. Chromosomal inversions will form a loop with its corresponding homolog at Prophase I. Scheme represent the consequences of a single crossing-over during homologous recombination occurring inside the inversions. **a**, paracentric inversions. One of the chromatids gets two centromeres while the other has none. This will lead to a dicentric chromosome and a acentric one which will be lost. Only the parental configurations will survive. **b**, pericentric inversions. One of the chromatids gets a large duplication whereas the other has a deletion. Only the parental configurations will survive. Based on images from (Griffiths et al., 2000).

The capacity of CRs to cause problems in meiosis and consequently to act as a genetic barrier has long captured the attention of evolutionary biologists (White, 1969; King, 1993). This can be particular important in parapatry in which species are a continuous with no specific extrinsic barrier to gene flow and in sympatry where species overlap geographic areas. In these cases, a mechanism must counteract gene flow for the initial separation of populations. CRs have been proposed as a possible mechanism involved in speciation due to the reduced recombination occurring between chromosomes carrying different arrangements thus acting as a barrier to gene flow (Schaeffer *et al*., 2003).

Chromosomal speciation models based on the suppressed recombination properties of CRs were initially proposed by Rieseberg *et al.* from observation in *Helianthus* species (Rieseberg, 2001) and by Noor *et al.* in *Drosophila* (Noor *et al.*, 2001). Rieseberg measured a significantly lower introgression at rearranged versus collinear regions in two naturally hybridizing species of sunflower, *H. annuus* and *H. petiolaris* (Rieseberg *et al*., 1999). Alongside, Noor *et al.* mapped the loci involved in reproductive isolation between *D. pseudoobscura and D. persimilis* to genomic locations that have fixed CRs between these species (Noor *et al*., 2001).

These proposals where after theorized by Navarro and Barton's model focusing on the accumulation of incompatibilities in rearranged versus collinear genomes (Navarro and Barton, 2003). Later, Kirkpatrick and Barton extended the model by incorporating a divergent selection regime which explains how a CR can increase in frequency and be fixed in different populations in the presence of gene flow (Kirkpatrick and Barton, 2006). According to the model, an inversion can increase in frequency if it hitchhikes with beneficial mutations that increase adaptation of a subpopulation. Additionally, a CR can prevent migration of disadvantageous alleles from another subpopulation into this one given that they occur in corresponding proximal areas to the inversion. This way, subpopulations will remain adapted

to their local environment with different alleles since they will not migrate between the subpopulations as a result of the reduced recombination (caused by the CR) between the different genome configurations.

A prediction from all recombination suppression models is that gene flow across rearranged regions is reduced and consequently higher divergence is expected near the breakpoints. Evidence supporting the models came from work developed in *Drosophila* in which, as initially observed by Noor, loci involved in both pre-and post-zygotic isolation between the sympatric pair *D. pseudoobscura and D. persimilis* map to inversion regions (Brown *et al*., 2004; McGaugh and Noor, 2012). Nonetheless, recent work has provided confounding evidence. Despite the divergence in overall rearranged areas between sunflower species, the signal for adaptation was actually stronger in the collinear regions (Strasburg *et al.*, 2009).

Consequently, it is important to disentangle and understand the role of CRs in adaptation separately from that of speciation. Regardless of the several chromosomal speciation models (Jackson, 2011) they all consider that CRs can be a driving force for speciation due to their role as a major molecular mechanism for the reduction of gene flow between different genetic backgrounds (Faria and Navarro, 2010). The recombination models now are in need of incorporating other factors (Stevison *et al.*, 2011). As pointed out by Jackson (Jackson, 2011), it is important to validate each individual element of the different models as empirical data nowadays is not sufficient to establish which factors inside the chromosomal speciation theory are indeed part of the speciation process or just a by-product of evolutionary time. For instance, the fact that sympatric species have marked differences at chromosomal breakpoints that are not seen in allopatric species is only a suggestion more than an indication of the validity of the models (Noor *et al*., 2001; Brown *et al*., 2004). Moreover, it is important to notice that not all CRs need to be related to positive selection; they might simply have been neutral or not been involved in the speciation process. It is thus, necessary to understand the fitness

consequences of CRs, their intrinsic properties and to establish and validate whether they can indeed initiate speciation processes by testing directly the Kirkpatrick and Barton's model.

2. *Schizosaccharomyces pombe* as a model organism.

The fission yeast *Schizosaccharomyces pombe* was first isolated from African millet beer in 1890 (Lindner, 1893). Since then it became a popular model system for both cellular and molecular biology studies in particular after the isolation of cell cycle mutants in 1970 by the Nobel laureate Paul Nurse and colleagues (Nurse, 2000).

S. pombe is unicellular haploid yeast that divides by fission and contains a small genome of approximately 14 Mb divided in three chromosomes (Chr.I, 5.7Mb; Chr.II, 4.6Mb; Chr.III, 3.5Mb)(Fan *et al.*, 1989). *S. pombe* cells are cylindrical and grow by increasing their length from approximately 8 to 15 μm while maintaining a constant diameter of 3-4 μm (Brunner and Nurse, 2000). Consequently, the newly born cells are the shortest and the cells just about to divide are the longest. Division is accomplished by medial fission producing two essentially identical daughter cells (Figure 1.9). Mitotic cell cycle in fission yeast has distinctive G1 (10%), S (10%), G2 (70%) and M (10%) phases. *S. pombe* chromosomes, like many other eukaryotes, adopt a Rabl configuration where centromeres cluster adjacent to the spindle pole body (SPB) and the telomeres are attached to the nuclear membrane at the opposite side of the nucleus, in relation to the SPB (Funabiki *et al.*, 1993). In *S. pombe* constitutive heterochromatin is present at the nuclear periphery, like in human cells (Olsson and Bjerling, 2011).

Fission yeast spends most of its time in mitosis but when it is subjected to nitrogen starvation it will express mating pheromone which trigger conjugation (mating) and formation of a diploid (Figure 1.9). Diploids in fission yeast are unstable so soon after karyogamy (nuclear fusion) of the two haploid

cells, the diploid immediately undergoes meiosis leading to the production of four haploid spores contained inside a ascus called a tetrad. Analysis of the haploid progeny from one ascus reveals all of the products of a single meiotic recombination event at each locus analysed.

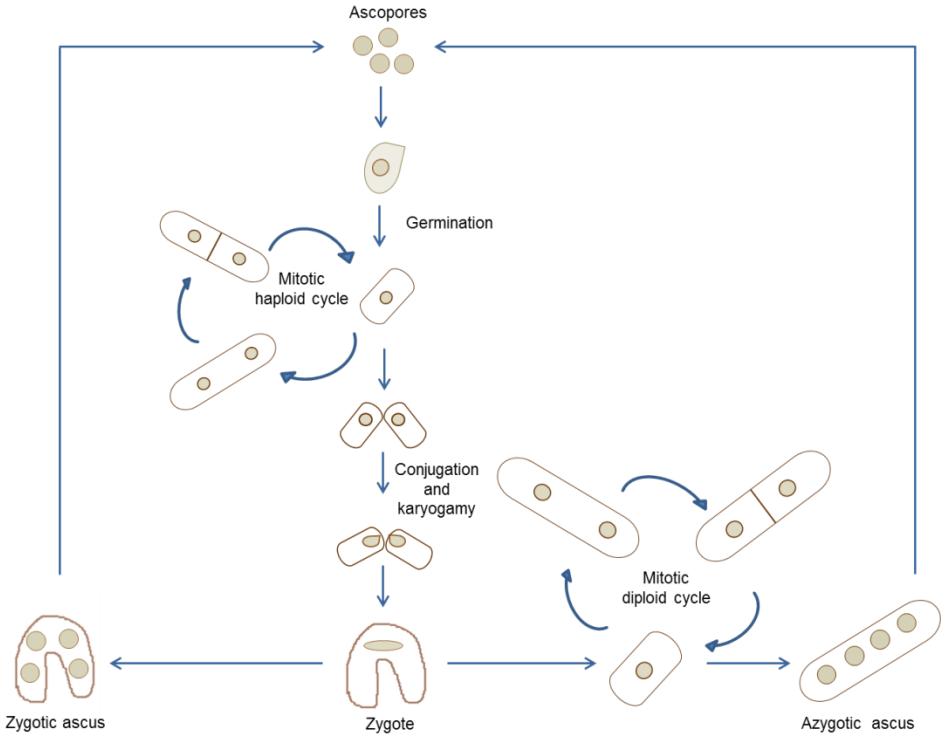


Figure 1.9. Fission yeast life cycle. During the mitotic haploid cycle, cells grow at the tips maintaining a constant diameter. Upon nitrogen starvation, cells can undergo mating and form a diploid cell (zygote) that immediately undergoes meiosis leading to the formation of a zygotic ascus (tetrad). Sporulation of this tetrad will release four haploid spores that will germinate and enter the mitotic cycle. Given that fission yeast is preferentially haploid, the diploid mitotic cycle can be forced in the laboratory by the combination of genetic markers. Adapted from Nielsen and Egel, 2007.

During meiosis, meiotic pairing of the homologous chromosomes leads to the switch from a Rabl configuration of the chromosomes at the mitotic

nucleus to the meiotic bouquet (Wells *et al*., 2006). Proper alignment and formation of chiasma between homologous chromosomes occurs during nuclear oscillations. Telomere clustering and nuclear oscillation are important mechanisms which occur in meiotic prophase and lead to the formation of the bouquet structure (Yamamoto and Hiraoka, 2001). This structure is important to assure HR and prevent ectopic recombination between non-homologous regions (Niwa *et al*., 2000; Wells *et al*., 2006).

S. pombe does not have crossover interference (Munz, 1994) a phenomenon whereby a crossing-over (CO) in one place decreases the likelihood of a crossing-over in the nearby interval. Interference results in widely spaced crossovers along chromosomes and is important for small chromosomes as it assures that even the smallest one will receive at least one crossing-over. Even though fission yeast does not have CO interference, crossovers are nearly uniformly distributed along the three chromosomes, thus reducing the probability of zero crossovers and therefore extremely low faulty segregations (Fox and Smith, 1998). There are between 11 and 17 crossovers per chromosome in a wild-type *S. pombe* meiosis (Munz, 1994).

S. pombe is an attractive model for chromosome biology studies as it shares several characteristics with human genome structure like heterochromatic factors such as Swi6/HP1, chromatin modifiers like the histone methyltransferase Clr4/Su_Var2-9, centromere proteins such as CENP-B and components of RNAi apparatus such as Argonaute and Dicer (Forsburg, 2003). Pericentromeric heterochromatin is lacking in the budding yeast but is present in both humans and fission yeast chromosomes. Chromosomes of fission yeast have average inter-origin distance of 14 kb (Heichinger *et al*., 2006) and large heterochromatic centromeres (~40 to 100 kb)(Pidoux and Allshire, 2004). Several genes in fission yeast have at least one intron and sometimes multiple introns despite their small length (most are <100 bp in length)(Wood *et al*., 2002). The telomere structure and its protein components are also more

conserved from humans to fission yeast than to *S. cerevisiae* making it a suitable model for telomere biology studies (Moser and Nakamura, 2009).

S. pombe is a well suited model for studies of genetic basis of phenotypic variation since its biology and genetics are well known. As a microorganism fission yeast is an excellent model for evolutionary biology studies due to its short generation time and high population sizes. Also, as it is a haploid, phenotypes will be visible immediately.

3. Aim of this thesis.

The main purpose of this thesis is to quantify the fitness effects of CRs and to understand how they can contribute to biological diversity using the fission yeast *Schizosaccharomyces pombe* as an experimental model.

In Chapter 2 I described the presence of CRs as a macro-mutation present in natural isolates of *S. pombe* and provide evidence for their implication in reproductive isolation of these isolates.

In Chapter 3 I quantify the two major fitness components, meiotic and mitotic, of engineered *S. pombe* strains containing either a pericentric inversion or a reciprocal translocation.

Finally, in Chapter 4 I attempt to measure the mutation rate and distribution of effects of beneficial mutations of a collinear and a inverted genetic background to address whether genomic background affects adaptation rate.

Chapter 2 – Chromosomal Rearrangements in natural isolates of *S. pombe*.

2.1 Abstract

Chromosomal rearrangements (CRs) are genome modifications that encompass large pieces of DNA. In the 1930s, Dobzhansky first described a chromosome inversion polymorphism segregating in a natural *Drosophila* population. Nowadays, it is estimated that three quarters of all species of *Drosophila* are polymorphic for inversions. Nonetheless, the question still remains open: can CRs contribute to the processes of adaptation and speciation? Here I show that CR are pervasive in natural isolates of *Schizosaccharomyces pombe*, a signature of natural variation that has been previously missed in this species. My analysis of natural isolates of the fission yeast *S. pombe* revealed that, remarkably, several strains display CRs in the form of translocations and inversions. Moreover, nucleotide sequence analysis reveals almost no difference between them. I tested whether these strains exhibit some degree of genetic isolation and surprisingly viability of hybrids between the different isolates is highly reduced, suggesting the beginning of a speciation process.

2.2 Introduction.

As White elegantly wrote in 1977 “eventually the story of chromosomal mechanisms and its evolution will have to be entirely rewritten in molecular terms” (White, 1977). For most of the history of molecular biology, single nucleotide polymorphisms (SNPs) have been the major genetic measure used to classify biological diversity. Increasing studies with improved resolution based on genomic sequence data are providing valuable data on the enormous diversity of CRs in different taxa (Emanuel and Saitta, 2007). In fact, genomes are nowadays considered more fluid than ever before since the time of Dobzansky (Kirkpatrick, 2010).

The genus *Drosophila* has long been a model for cytogenetic studies of genome evolution along with the yeast group *Saccharomyces sensu stricto*.

The genome of *Saccharomyces* and other hemiascomycetes has been restructured by many CRs (Dujon *et al*., 2004; Gordon *et al.*, 2009). Inference studies have proposed that CRs are involved in a series of steps that led to the modern *S. cerevisiae* from the ancestral *Saccharomyces* (Gordon *et al*., 2009). The same is thought to have happened in *Drosophila* (Bhutkar *et al*., 2007, 2008) and *Caenorhabditis* (Coghlan and Wolfe, 2002). Similarly, extensive genome reshuffling in the presence of synteny blocks is a feature of the differences between humans and chimpanzees genome which differ by around 1500 inversions (Feuk *et al*., 2005).

Chromosomal variation is also observed within species as a polymorphic trait. For example, wine strains of *S. cerevisiae* have a remarkably high chromosomal diversity in terms of length and ploidy (Querol *et al*., 2003). This enormous genomic diversity is also found in other natural and even clinical isolates of *S. cerevisiae* (Carreto *et al.*, 2008). Polymorphic CRs such as inversions are often associated with adaptation to specific environmental conditions in *D. pseudoobscura* (Anderson *et al*., 1991; Schaeffer *et al*., 2003), *Anopheles gambiae* (Coluzzi *et al*., 2002; Pombi *et al.*, 2008), the yellow monkeyflower *Mimulus guttatus* (Lowry and Willis, 2010) and the recent case of the threespine sticklebakcs (Jones *et al*., 2012). Frequently, different populations share the same breakpoints (Ranz *et al.*, 2007) as it is the case of some inversions, preferentially found in certain chromosomes (Pevzner and Tesler, 2003; Flores *et al*., 2007). This enormous genomic diversity is suggestive that CRs are a mechanism for the generation of biological diversity even within species that can be a starting point to a speciation process, as suggested by chromosomal speciation models (Faria and Navarro, 2010).

Speciation, for sexual organisms, corresponds to the evolution of reproductive isolation. There are several models of speciation and for one in particular CRs are the fundamental unit. Chromosomal speciation models postulate that rearrangements are the driving force for reproductive isolation by reducing gene flow between different genetic backgrounds. Mutations can

exchange freely only in regions not involved in the rearrangement. This selective gene flow will allow for the accumulation of different mutations in different genomic backgrounds. Data collected from natural populations showing suppressed recombination in rearranged areas of the genome (Schaeffer *et al*., 2003) indicates that CRs can act as a genetic filter between populations (Ayala and Coluzzi, 2005). CRs, in particular inversions, have been implicated as a mechanism for the preservation of co-adapted gene complexes (Noor *et al*., 2001). Formalization of chromosomal speciation was initially done by Navarro and Barton (Navarro and Barton, 2003) and further extended by Kirkpatrick and Barton (Kirkpatrick and Barton, 2006). The best example for the importance of inversions in suppression of recombination is the mammalian sex chromosomes where a series of inversions led to a largely reduced recombination rate of the Y chromosome with the X chromosome (Lahn and Page, 1999; Kirkpatrick, 2010).

Several works have been already developed in the *Saccharomyces sensu stricto* group and *S. cerevisiae* still remains the most studied yeast. However, *S. pombe* is a potentially very interesting model organism for evolutionary biology studies due to its small genome of around 14 Mb distributed in only three chromosomes of 5.7, 4.6 and 3.5 Mb (Wood *et al*., 2002). Fission yeast *S. pombe* was originally isolated from African millet beer but its ecology it is rather unknown. A recent whole genome study has compared the nucleotide diversity between two of the most used *S. pombe* strains, L972 and NCYC132, and found a nucleotide diversity of 1% (Rhind *et al*., 2011). A later study analysed 81 natural isolates of *S. pombe* (Brown *et al.*, 2011) and found nucleotide diversity for fission yeast consistent with the variation described by Rhind and colleagues.

Given the importance and the wide distribution of CRs in several species I proposed to determine its presence or absence in *S. pombe*. I postulated that if they can be an important source of variability that leads to different adaptive behaviours they should occur in natural isolates. Moreover, if they can

contribute to an incipient speciation process I expect, in case of detection of CRs, to see a decline in meiotic viability.

2.3. Material and methods.

2.3.1 Strains and media. The strains used in this study are listed in Tables 2.1 and 2.2. Standard media and growth conditions were used throughout this work (Moreno *et al*., 1991) with minor modifications (see protocols below). Cultures were grown in rich media Yeast Extract with Supplements (YES) or Edinburgh Minimal Media (EMM) at 32ºC with appropriate shacking, unless otherwise stated. Tetrad dissection for meiotic viability analysis was made in YES solid media. Strains were constructed using standard genetic techniques for fission yeast (Bähler *et al*., 1998). Strains generated in this study were made, when appropriate, by transformation of specific DNA fragments or by mating with available strains.

Media composition:

- Yeast extract (YE): 0.5% (w/v) Oxoid yeast extract, 3.0% (w/v) glucose.
- Yeast extract with supplements (YES): YE+ 225 mg/l adenine, arginine, histidine, leucine, uracil and lysine hydrochloride.
- Malt extract (ME): 3% (w/v) Bacto-malt extract (Difco). Supplements added as for YES. Adjusted to pH 5.5 with NaOH.
- Edinburgh Minimal Medium (EMM): 3 g/l potassium hydrogen phthallate, 2.2 g/l Na2HPO4, 20 ml/l salts (stock x 50), 1 ml/l vitamins (stock x 1000), 0.1 ml/l minerals (stock x 10,000).
- Supplements: Adenine, Arginine, Leucine, Lysine, Histidine and NH4 stocks prepared at 7.5 g/l (14 ml added to EMM) and Uracil at 2.5g/l (42 ml added to EMM).
- Salts x 50: 52.5 g/l MgCl2.6H20, 0.735 mg/l CaCl2.2H20, 50 g/l KCl, 2 g/l Na2SO4.

- Vitamins x 1000: 1 g/l pantothenic acid, 10 g/l nicotinic acid, 10 g/l inositol, 10 mg/l biotin.
- Minerals x 10.000: 5 g/l boric acid, 4 g/l MnSO4, 4g/l $ZnSO4.7H_2O$, 2 g/l $FeCl_2.6H_2O$, 0.4 g/l molybdic acid, 1 g/l KI, 0.4 g/l $CuSO_4.5H_2O$, 10 g/l citric acid. After autoclaving, a few drops of preservative were added (1:1:2, chlorobenzene: dichloroethane: chlorobutane).

For EMM, supplements were added after autoclaving. 2% (w/v) Glucose was added from a filtered stock of 40% (w/v) Glucose. 2% (w/v) Maltose and 2% (w/v) Raffinose added from filtered stocks of 20% (w/v) of each.

Solid media was made by adding 2% Difco Bacto Agar. All media was prepared in bulk. Sterilization was made by autoclaving at 10 psi for 20 min. The media was stored in 500 ml bottles and agar was remelted in a microwave oven before using.

2.3.2 Yeast cells transformation. Transformation of yeast cells was performed using the lithium acetate protocol as described previously (Moreno *et al*., 1991). Accordingly, cells were incubated overnight in rich media. The following day they were diluted and grown until reach log phase. Cells were pelleted and washed once in 1ml of LiAc-TE (0.1 M Lithium acetate, 10 mM Tris pH 7.5, 1 mM EDTA). Cells were resuspended in 1 ml of LiAc-Te and to a 100ul/per transformation, 10μl of carrier salmon sperm DNA (10 mg/ml, Stratagene), previously heated 5 minutes at 100ºC, and 10μl (containing 1 μg of DNA) and the mix was incubated for 5 minutes at RT. 260 μl of LiAc-TE-PEG (LiAc-TE plus 40% PEG4000) was added and mixed gently. Cells were incubated for 45 minutes at 32ºC. A volume of 43 μl of DMSO (Sigma) was added and mixed gently followed by a 7 minutes heat-shock at 42ºC. Cells were pelleted for 1 minute at 3000 rpm, washed once with YES or EMM and let to recover O/N at 32ºC (for the case of Insertions or knockouts) or plated

directly in selective media (In the case of plasmid DNA). Correct integration was confirmed by PCR analysis.

2.3.3 Construction of a non-exchangeable mating type cassette. In order to disrupt the mating type switch that occurs in h90 strains (which is the case for the natural isolates), a stable non-switchable *mat1* locus was constructed both for *mat1-P* and *mat1-M* as described previously (Arcangioli and Klar, 1991)(Figure 2.1a). Sequences of *mat1-M*, *mat1-P*, *H2* and *H1* were taken from (Kelly *et al*., 1988). Primer sequences are listed in Table 2.4. The plasmid pFa6a-3HA-natMX6 was used as the backbone for the insertion of fragment PCR1 and PCR2 (Figure 2.1b). This was the simplest available plasmid with the natMX6 resistance cassette. Sequence PCR1 was inserted in the *HindIII* and *PacI* (replacing the 3HA fragment) and sequence PCR2 was inserted in the *EcoRI* site. PCR1 and PCR2 products were amplified from pON104 (*mat1-P*) and pON107 (*mat1-M*) plasmids, kindly provided by Olaf Neilsen (U. Copenhagen). For PCR1 the following pairs of primers were used: pcr1 mat1-P F/ pcr1 mat1-P R and pcr1 mat1-M F/ pcr1 mat1-M R. The forward and reverse primers contain, respectively, a *HindIII* and a *BglII* recognition site. For PCR2 the following pairs of primers were used: pcr2 mat1-P F/pcr2 mat1-M R and pcr2 mat1-M F/pcr2 mat1-M R. The forward and reverse primers contain, respectively, the *PmeI* and *EcoRV* recognition site. DNA amplifications were made using the *PfuI* Proofreading DNA Polymerase (Promega). PCR1-P/M and PCR2-P/M were initially cloned into a pGEMT (Promega) and a pTOPO (Invitrogen) cloning vector, respectively. Confirmation of these clonings was made by enzymatic restriction of p-GEMT-PCR1 with *HindIII* and *BglII* (both from Fermentas) and PCR2 with *EcoRI* (Fermentas). I performed a *EcoRI/ScaI* digestion, independently, in the PFa6a-3HA-natMX6-PCR-2M and PFa6a-3HA-natMX6-PCR-2P to test for correct direction of insertion. Both plasmids were sequenced with Ttef forw and pFa6 Rev primers. Once the positive clones were confirmed, I did a

HindIII and *BglII* restriction in pGEMT-pcr1M, pGEMT-pcr1P, pFa6a-pcr2P and pFa6a-pcr2M and a ligation of the fragments according to its combination. Both pGEMT-pcr1P and pGEMT-pcr1M plasmids were sequenced in the part of interest to confirm the integrity of the fragments: T7, 1matP 539F, 1matM 602F, 1matM 1175 F, 1matM1651F, SP6; T7, 1matP 539F, 1matP 1124F, 1matP 1625 F, SP6.

The final constructs, pFa6a-pcr1P-natMX6-pcr2P and pFa6a-pcr1M-natMX6-pcr2M, were used to transform homothallic *S. pombe*. Enzymatic DNA digestion with *HindIII* (NEB) and *EcoRV* (Fermentas) digestion release the stable *mat-1* cassette. To confirm loss of the switching capacity, strains were platted in ME for three days and checked for the absence of mating and meiotic products. Genomic integration was tested by a PCR amplification using the MM, MP and MT primers and also the T tef forw/mat locus end rev pair.

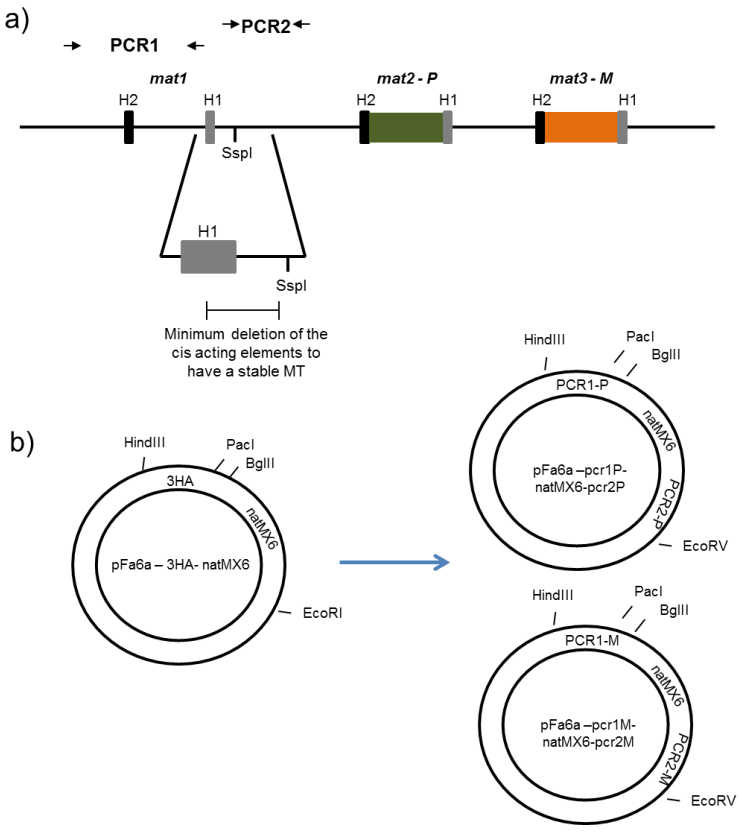


Figure 2.1. Mating-type switching disruption. a, schematic representation of mating-type locus in *S. pombe*. PCR1 and PCR2 correspond to the sequences clones in order to disrupt the minimal 50bp of the homology region H1. **b**, schematic drawing of clonings. Adapted from Benoit Arcangioli (Arcangioli and Klar, 1991).

2.3.4 Meiotic viability. For meiotic crosses cells were grown overnight in EMM minimal media with all supplements and diluted the day after. When cells reached exponential phase, 1 ml of culture was centrifuged and washed twice in 1 ml of EMM without supplements. The final pellet was resuspended in 100 μl EMM without supplements. This was made for both h- and h+ cultures. 50 μl of each mating type was mixed in a eppendorfs and pipette up and down. The desired volume was pipetted into a ME plate, allowed it to dry and incubated at 25ºC for 2-3 days. When tetrads were formed, an aliquot of

the cells from the ME plate was putted into 40 μl of sterile H_2O. A drop of this mix was placed in a corner of a YES plate, letting it slide into the plate, making a line from one side of the plate to the other. The plate was allowed to dry and the tetrads were dissected in a micromanipulator equipped with a glass needle (Singer Instruments). The plate was incubated for 6 hours to allow the release of the spores from the tetrads. After, the spores were dissected and the plate was incubated for 5 days at 32ºC, time after which colony formation was scored. Meiotic viability was given as the number of colonies formed divided by the number of colonies expected.

2.3.5 Pulse Field Gel Electrophoresis. PFGE Protocol was performed as described in (Ferreira and Cooper, 2001) with some modifications. All solutions were prepared with miliQ-water and filtered. Filter tips were used throughout to avoid action of DNases. Exponentially growing cells ($0.5\text{-}1x10^7$ cells/ml) were centrifuged in RNAse/DNase free 50 ml falcon tubes for 3 minutes at 3000 rpm. The cells were collected into eppendorfs (pellets could be frozen at -20ºC at this step) and cell pellets washed twice in SP1 buffer (1.2 M D-sorbitol, 50 mM sodium citrate, 50 mM $Na_2HPO_4 \cdot 7H_2O$, 40 mM EDTA pH 5.6. pH was adjusted with citric acid/phosphoric acid respectively). When cells were resuspended for the second time, cell density was estimated by reading a 50x dilution = 20 μl cells + 980 μl water. 1-2 μl of zymolyase-20T (300 ng/100 μl SP1 buffer) was gently added and the mixture was incubated at 37ºC for 10-40 minutes. Meanwhile, a 1% low melting point agarose in TSE (0.9 M Sorbitol, 10 mM Tris-HCl pH 7.5, 45 mM EDTA) was melted and cooled to 42ºC. After 10 minutes, the cells were checked for zymolyase treatment at the light microscope (looking for ghost cells by adding 1% SDS to 10 μl cells on a microscope slide and compared with cells without SDS treatment). If lysis had not begun, a new aliquot of Zymolyase was added. When cells had undergone zymolyase treatment (50% lysis for exponential cells and about 25% for

nitrogen-starved cells), spheroplasts were spinned down at 3000 rpm for 1minute. Supernatant was removed and spheroplasts were gently resuspended in the previously prepared LMP agarose in TSE (to give a final concentration of 1 x 10^8 cells in 100 μl which makes 2-3 plugs (to the volume the OD was multiplied by 500 for exponentially growing cells; e.g. OD=0.3 would be 150 μl LMP agarose). The cell suspension was dispensed into 90 μl plug molds and allowed to solidify at 4ºC for 10 min. The plugs were transfer from the molds into 15 ml tubes, covered with 3 ml solution of 0.25 M EDTA, 50 mM Tris-HCl pH 7.5 and 1% SDS and incubated at 55ºC for 90 minutes. After this time, the solution was removed. The plugs were incubated with a solution of 0.5 M EDTA pH 9.5, 1% Lauryl Sarcosine, 1mg/ml proteinase K at 55ºC for 24 hours (this solution can be left for several days, but one o/n is enough). The next day, plugs were incubated in 7ml T10xE (10 mM Tris-HCl pH7.5, 10 mM EDTA) at RT for 5min (quick wash) and then for half an hour. This wash was repeated at least one time. After, the plugs were incubated in T10xE containing 0.04mg/ml PMSF (in DMSO, to inactivate Proteinase K) at 55ºC for 1 hour. The solution was removed and plugs incubated in T10xE at 25ºC for half an hour. This wash was repeated at least one time. The plugs were store in 0.5M EDTA, 10 mM Tris-HCl pH 9.5 at 4ºC if they were used for running whole chromosomes.

For running whole chromosomes, plugs were pre-equilibrated in 1x TAE for 1 hour. Plugs were loaded into a 0.8% Pulse-Field Certified Agarose (Biorad) gel in 1xTAE and sealed with excess molten agarose. Electrophorese was made at 14ºC in 1x TAE.

For running *NotI* or *SfiI* digested chromosomes, plugs were pre-equilibrated in 500 μl *NotI* buffer (10 mM NaCl, 5 mM Tris-HCl [pH 7.9], 1 mM MgCl2, 0.1 mM DTT, 100 μg/ml BSA) or *SfiI* buffer (10 mM Tris, 10 mM MgCl2, 50 mM NaCl2 and 100 μg/ml BSA) for 1 hour at 37ºC. The buffer was then replaced with 500 μl fresh buffer plus 5μl/per tube of the respective

enzyme and incubated overnight at 37ºC. The plugs were kept at 4 ºC until running the gel (they can be kept at 4ºC for several days). For long term storage, plugs can be stored at this stage in 0.5M EDTA, 10 mM Tris-HCl pH 9.5 at 4ºC. Plugs were loaded into a 1% Pulse-Field Certified Agarose (Biorad) gel in 0.5x TBE buffer and sealed with excess molten agarose. Electrophorese was made at 14ºC in 0.5x TBE.

All PFGE were performed on a Biorad CHEF-DR III System. For whole chromosome PFGE the "Angle Ramp: *S. pombe*" program was used:

	Block 1	Block 2	Block 3
Switch Time:	1,200 sec.	1,500 sec.	1,800 sec.
Run Time:	24 hours	24 hours	24 hours
Angle:	96°	100°	106°
Voltage Gradient:	2 V/cm	2 V/cm	2 V/cm

For *NotI* and *SfiI* PFGE gels the following program was used: Switch Time: 60-120 sec; Run Time: 24 hours; Voltage Gradient: 6 V/cm. After electrophoresis, DNA was visualized by ethidium bromide staining, and gels were processed for Southern blotting.

2.3.6 Southern Blotting. Southern blot analysis was performed as described (Ferreira and Cooper, 2001). The protocol was performed as follow: PFGE agarose gel were incubated in 0.25N HCl (10 ml HCl in 500 ml H_2O) for 15 minutes followed by 30 minutes incubation in Blot 1 (20g NaOH, 87.6g NaCl in 1 litter of H_2O) and 1 hour in Blot 2 (77g NH4Ac, 0.8 NaOH in 1 litter H_2O). Afterwards the gel was washed for 5 minutes in 6x SSC (stock 20x SSC: 175.3g NaCl, 88.2 g Sodium Citrate in 1 litter H_2O; pH adjusted to 7 with HCl). Membrane was activated for transfer Duralon UV-membrane (nylon, uncharged membrane from Stratagene) with 10 minutes incubation in Blot 2. A stack of paper towels of about 3 inches was placed on the bench and on top of these a 3 MM Whatman paper with the same size as the gel. Next, the

activated membrane was placed and on top of it the gel was placed with the wells up. The whole stack was covered with saran wrap and a glass plate placed on top of the stack with two full 500 ml bottle on top of the glass plate to act as a weight. The gel transferred overnight or for a minimum of 3 hours. After transfer, the membrane was allowed to dry at room-temperature. The DNA was cross-linked to membrane by using the Autocross link from Stratalinker® UV Crosslinker. For radioactivity labelling of the membrane, the membrane was placed in a hybridization bottle and pre-incubated for 30 minutes at 65ºC in Church- Gilbert solution (1% BSA, 1 mM EDTA, 7% SDS, 0.5 M Na_2HPO_4, 4 ml H_3PO_4 85%, up 1 litter with H_2O). Different probes labelled with ^{32}P using the Prime-it II random primer labelling kit (Stratagene) were used for Southern blotting. Probes were prepared as following: 25 ng of DNA template, 10 μl of random oligo primers and 23 μl of H_2O and heat-boiled for 5 minutes. Spinned down and added 10 μl of 5x dCTP buffer, 2.5 μl dCTP labeled nucleotides and 1 μl of (-) Exo Klenow (Stratagene). The reaction was incubated at 37ºC for 7 minutes. Meanwhile, columns were washed twice in 250 μl of TE 1x centrifuging at 3000 rpm for 1 minute. 50 μl of TE 1x were added to the reaction and transferred to the column. The column was centrifuged 3 minutes at 3000 rpm. The probe was boiled for 5 minutes, spinned down, added to membrane and incubated overnight at 65ºC. The next day, the probe mix was discarded in a radioactive waste container, rinsed once with MiliQ-water and once with Wash Solution (2x SSC, 0.1% SDS). The membrane was kept in a larger volume of new Wash Solution for 30 minutes at 65ºC in a rotating oven. After, the membrane was placed in a saran wrap and putted down on a phosphoimager screen overnight to develop signal. The signal was read in a PhosphorImager apparatus (Molecular Dynamics). Southern blot probes were design to hybridize specific sites of the chromosome. Probes were generated by PCR amplification (Table 2.6).

2.3.7 DNA Sequencing. Colony PCR was performed to obtain genomic DNA for PCR amplification. Briefly, a portion of a colony grown in solid medium was mixed with Z buffer (2.5 mg/ml of Zymolase, 1.2 M Sorbitol, 0.1 M Sodium phosphate pH 7.4 and a 1:10 of 10x Lyzing enzymes stock 100 mg/ml). This mix was incubated in a thermocycler for 30 minutes at 30ºC followed by 5 minutes at 100ºC. Primers are described in Table 2.5. Genetic sequences were amplified with PCR master mix (Roche) and then purified with Wizard SV Gel and PCR clean up system (Promega). DNA sequencing was performed according to the manufacturer's instructions (Applied Biosystems). Briefly, 70 to 90 ng of the purified PCR DNA were added to 2 μl of buffer, 2 μl of Terminator Ready Reaction Mix (BigDye® Terminator v1.1 Cycle Sequencing Kit), 1 μl of primer and up to 10 μl of deionized water. For Cycle Sequencing, the tubes were placed in the thermal cycler being thermal cycling as follows: 96ºC 1 min, 25 cycles of 96ºC 10 sec, 50ºC for 5 sec and 60ºC for 4 min. The sequencing reaction was placed into 1,5ml eppendorfs and precipitated for a minimum of 4 hours with 10 μl of H20, 2 μl of 3 M Sodium Acetate pH4.6 and 50 μl of 95% ethanol. The mix was centrifuged for 30 minutes at 14000 rpm 4ºC. The supernatant was carefully aspirated and the pellet was washed with 250 μl of 70% ethanol and centrifuged again for 15 minutes. Finally, the supernatant was carefully aspirated and the pellet allowed drying.

DNA sequencing quality and analysis was performed using 4Peaks and DNAStrider software, respectively. DNA sequence alignment was performed in MUSCLE from EBI and π values calculated using DnaSP (Rozas and Rozas, 1999).

2.4. Results.

2.4.1 Natural isolates contain chromosomal rearrangements.

CRs exist as polymorphic traits in natural populations and can even sometimes be associated with adaptation. A recent example was described in *Mimulus guttatus* were different arrangements are implicated in the different ecological response to the environment (Lowry and Willis, 2010). If CRs are indeed an important macro-mutation leading to phenotypic variation I would expect to find them in natural isolates of different species. In order to test it, I gathered several natural isolates of *S. pombe* from different geographic regions (Table 2.1) and analysed their karyotypic profile by Pulse-Field Gel Electrophoresis (PFGE)(Figure 2.2). *S. pombe* contains three chromosomes with sizes of approximately 5.7, 4.6 and 3.5 Mb (Fan *et al*., 1991; Garkavtsev and Mizukami, 1997). Since L972 is the most widely used strain and the first to have its genome sequenced, I considered it as the reference genome.

Whole-chromosome PFGE (WC-PFGE) analysis suggests that all strains, except NCYC132, have the expected karyotypes (Figure 2.2). The second band, corresponding to reference Chr.II is absent in strain NCYC132. This strain has a band markedly lower than expected, which could be caused by a major deletion of a chromosomal portion. Another hypothesis would be a translocation of a large DNA amount into another chromosome which would alter the size of both chromosomes. As for the third band, it is expected to have slight variations in size as it corresponds to Chr.III which contains the polymorphic rDNA loci.

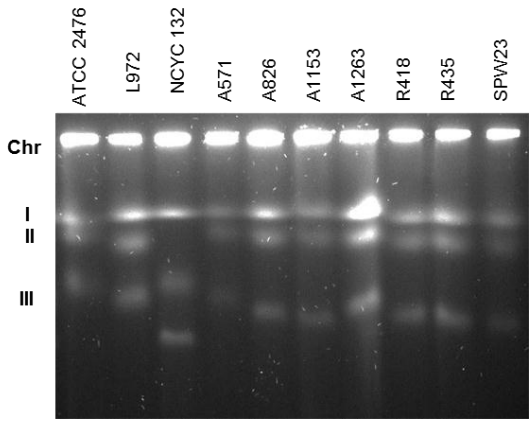


Figure 2.2. WC-PFGE analysis of *S. pombe* chromosomes. L972 as a reference strain contains three chromosomes: Chr.I 5.7Mb, Chr.II 4.6 Mb and Chr.III 3.5 Mb. Chr.III contains the rDNA cluster which is polymorphic in size for *S. pombe*.

In order to test whether there was a CR in strain NCYC132, I designed probes for the arms of the three chromosomes as well as for each centromere (Figure 2.3a and Table 2.6). Each centromere in *S. pombe* contains a unique sequence region, *imr* (Takahashi *et al*., 1992) making it possible to distinguish between them.

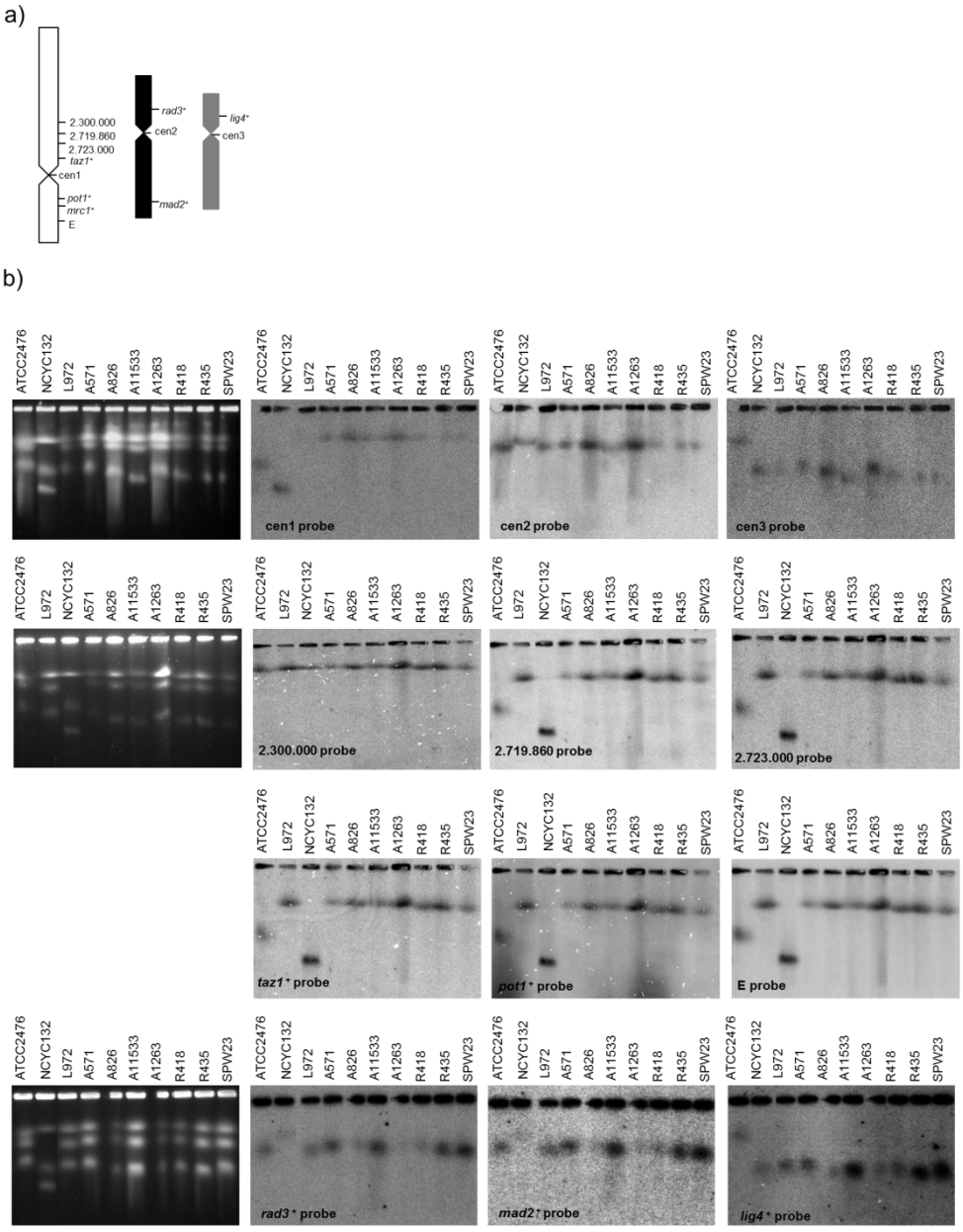


Figure 2.3. Southern blot analysis reveals the presence of translocations in two natural isolates of *S. pombe.* **a**, localization of probes used for Southern blot. **b**, Southern Blot analysis reveals a translocation for strains ATCC2476 and NCYC132.

The centromere I (cen1) probe should hybridize to the highest molecular weight band of WC-PFGE because it corresponds to the centromere of the largest chromosome. However, Southern blotting analysis shows that the cen1 probe hybridizes to the third band of the WC-PFGE both for strain ATCC2746 and strain NCYC132 (Figure 2.3b). Consistent with this, probes E, *pot1*$^{+}$, *taz1*$^{+}$, 2.723.000 and 2.719.860 (all of which are located in the Chr.I of the reference L972, Figure 2.3b) also hybridize to the third band in WC-PFGE in strains ATCC2476 and NCYC132. The only probe belonging to Chr.I that does not co-localize with all the above is probe 2.300.000 (Figure 2.3b). Probe 2.300.000 hybridizes to the first band in both ATCC2476 and NCYC132 strains. In ATCC2476, the first band co-localizes with cen3, *lig4*+ and *ura4*+ probes (Chr.III probes for reference strain) while the first band of NCYC132 co-localizes to cen2, *rad3*+ and *mad2*+ (Chr.II probes for reference strain). Thus, from probe 2.300.000 on, the arm of Chr.I has translocated to Chr.III in ATCC2476 and to Chr.II in NCYC132. These results suggest that the translocation breakpoint must be localized between region 2.719.860 and 2.300.000. They also point out to a probably common fragile site in Chr.I even though the translocated arm moved to different sites.

The translocation described for NCYC132 has also been found in a study that came out in the meantime looking at the same strains (Brown *et al*., 2011). Brown and colleagues found that this translocations is actually reciprocal and is present in more five other isolates spread world-wide.

No other translocation was detected using probes for the arms of Chr.II and Chr.III. For the remaining strains, all the probes hybridize to the expected site as in L972 in the WC-PFGE. Thus, I have not been able to detect any translocation for the other strains.

This first analysis using WC-PFGE would not allow visualising other types of CRs such as inversions. Inversions do not cause major alterations in DNA content and would not therefore alter the size of individual

chromosomes. In order to map other possible rearrangements, I have performed a more detailed karyotypic analysis by doing a *NotI* restriction-PFGE. *NotI* is an infrequent cutter that releases thirteen well described fragments in the L972 strain, designated by the letters A to Q (Figure 2.4a)(Fan *et al*., 1991).

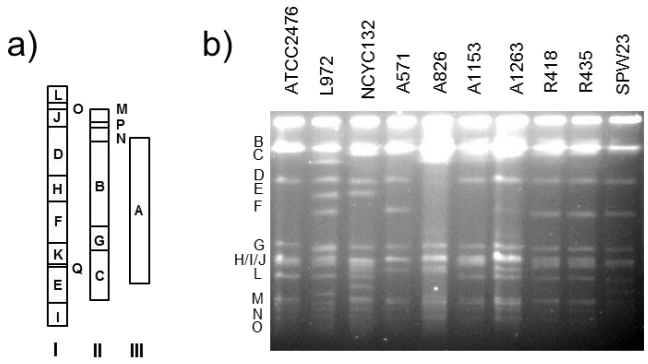


Figure 2.4. *NotI*-restriction PFGE analysis of *S. pombe* chromosomes. **a**, localization of *NotI* restriction sites in reference strain L972. **b**, *NotI*-restriction PFGE of *S. pombe* natural isolates. *NotI* map figure adapted from (Garkavtsev and Mizukami, 1997).

NotI profiles of natural isolate strains revealed a remarkable variable profile (Figure 2.4b) suggesting the probable presence of other CRs than translocations. However, this variation could be attributed to polymorphic variation in the described *NotI* sites. In order to test if the differences were caused by polymorphic variation, I performed a PCR amplification followed by a *NotI* restriction of almost all conserved *NotI* sites described (Figure 2.5 and Table 2.7). PCR amplification followed by *NotI* restriction revealed that the sites are conserved between the strains, rejecting the hypothesis of polymorphic variation.

The variation in the *NotI* restriction profile could therefore be caused by reorganization of the linear DNA sequence as it is the case of inversion. An inversion would alter the localization of the *NotI* restriction sites and consequently the sizes of the released fragments.

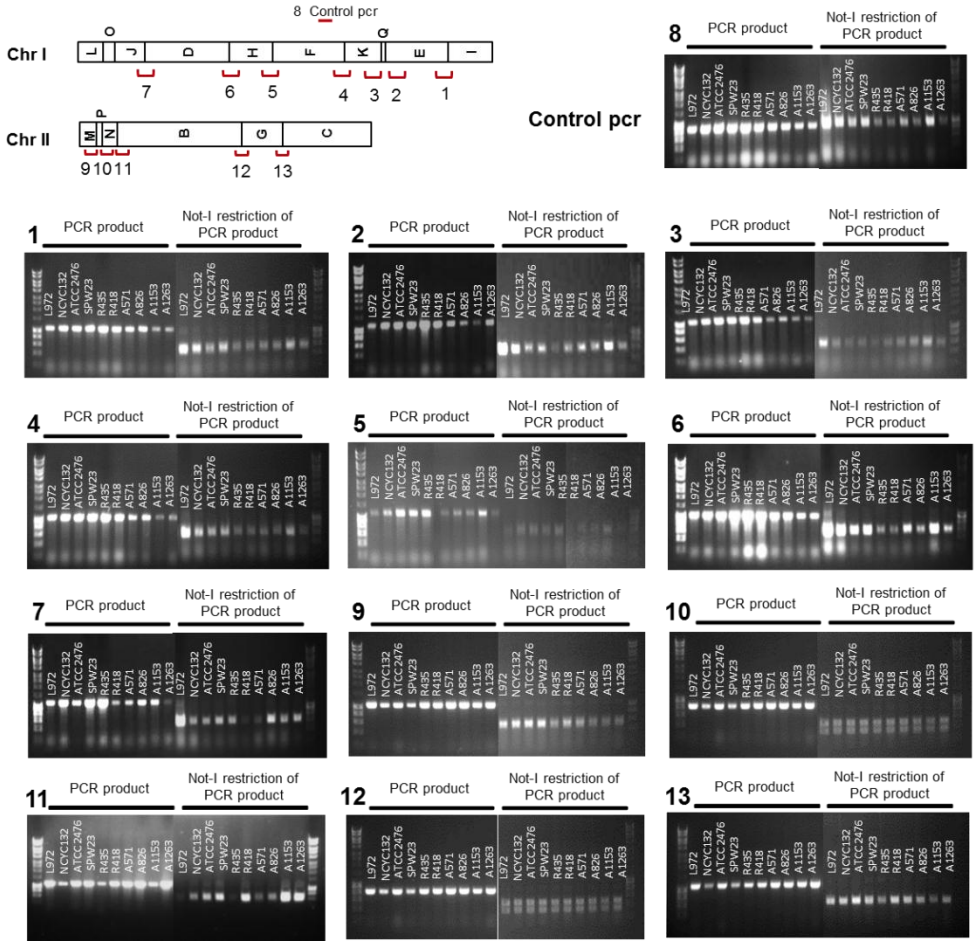


Figure 2.5. PCR amplification and *NotI* restriction of the known *NotI* recognition sites. *NotI* restriction sites are conserved in the natural isolate strains. *NotI* map figure adapted from (Garkavtsev and Mizukami, 1997).

In order to test for the presence of an inversion, I performed Southern blot analysis in the *NotI* restriction-PFGE using the same probes as before.

In the reference strain L972, these probes hybridize in chromosomal regions corresponding to the tips of some of the largest *NotI* fragments (fragments E, F and H). Either one of these two *NotI* fragments (E and/or F) was absent in most strains (Figure 2.4b). The two probes of each fragment will co-localize if there was no disruption in the fragment. This happens in the reference strains were probe E co-localizes with probe *pot1*$^{+}$, probe *taz1*$^{+}$ with probe 2.723.000 and probe 2.719.000 with probe 2.300.000 (Figure 2.6a and 2.6b). If the two probes from each tip of a fragment do not co-localize in the same size band, this means that there was a disruption within the fragment. Single duplications or deletions would still make the probes to co-localize as they would only alter the final size of the fragment.

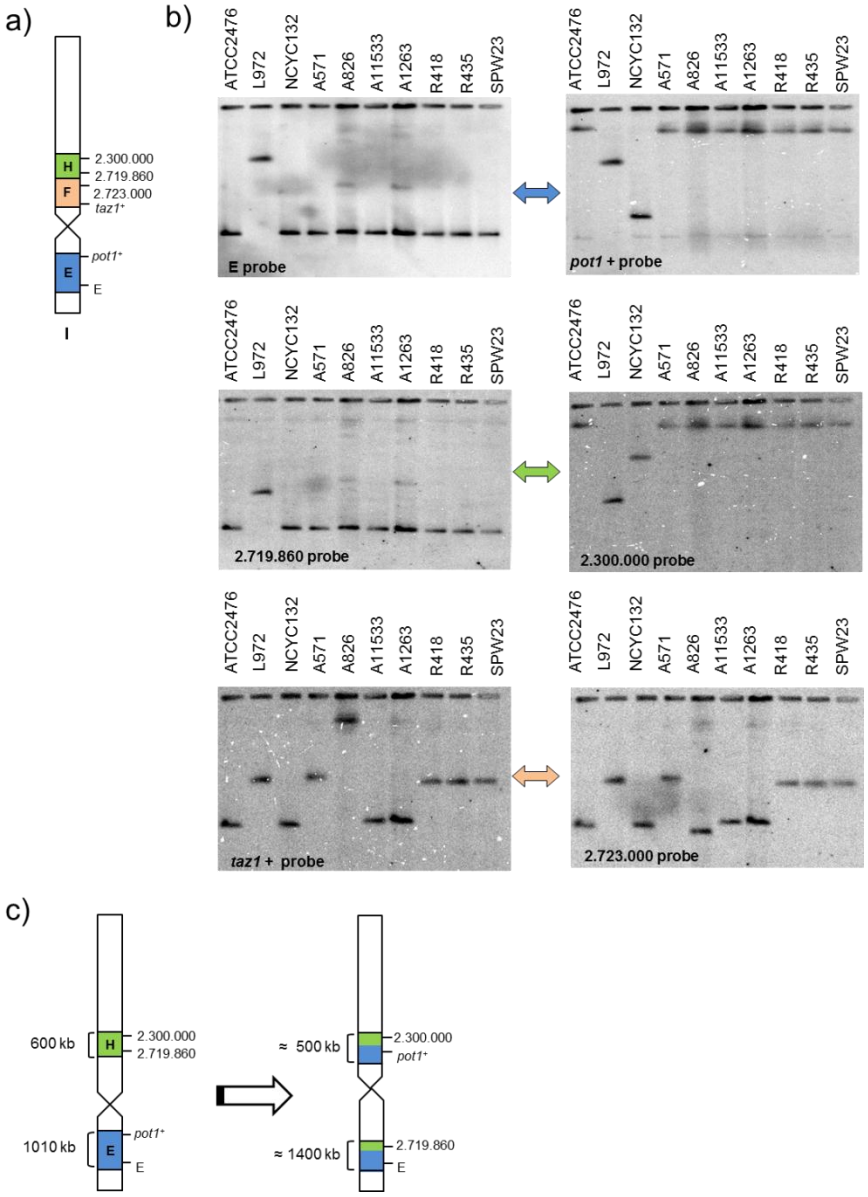


Figure 2.6. Southern blot analysis of *NotI*-restriction PFGE of *S. pombe* chromosomes. a, localization of probes used for Southern blot. **b**, Southern Blot analysis reveals co-localization of signal corresponding to a pericentric inversion between E and H fragments. Coloured arrows indicate the sense of the probes that co-localize in the reference strain L972. **c**, representation of the pericentric inversion between *NotI* fragments E and H.

In all strains except L972, the probes for either fragment E and H do not co-localize. Instead, probe E co-localizes with probe 2.719.860 and they both hybridize to a fragment smaller than the E or H fragment. Likewise, probe *pot1*$^{+}$ co-localizes with probe 2.300.000 and these hybridize to a bigger fragment (Figure 2.6b). Given that the *NotI* restriction sites for these fragments are conserved (Figure 2.5), the only explanation for this results is a pericentric inversion. This inversion has breakpoints inside the reference fragment H (between 2.300.000 bp and 2.719.860 bp) and the reference fragment E (4.129.529 bp and 4.949.448 bp). With this analysis I was able to show that there is, at least, one common inversion in all strain except L972, between the H and E *NotI* fragments (Figure 2.6c).

The pericentric inversion here described has also been mapped by Brown and colleagues that have determined the precise location of the breakpoints: 2.683.519-2.638.704 and 4.911.470-4.911.584.

There could be however other CRs since other bands are also altered in the *NotI*-restriction PFGE. This is the case for the band corresponding to fragment F which is absent in strains ATCC2476, NCYC132, A826 and A1153 (Figure 2.4b).

In strains ATCC2476, NCYC132, A1153 and A1263, probes *taz1*+ and 2.723.000 co-localize but hybridize to a smaller fragment than the reference F fragment (Figure 2.6b). This suggests either the occurrence of a deletion or the formation of a new *NotI* site inside the fragment. In order to discard between the two hypotheses it would be necessary to sequence the region between the two probes. I have not pursuit this further.

Regarding strains A826, the probes *taz1*+ and 2.723.000 do not co-localize which indicates the presence of a more complex rearrangement. This has not been further investigated.

Overall, consistent to what was recently observed (Brown *et al.*, 2011), I found that CRs such as translocations and a pericentric inversion are

pervasive in natural populations of *S. pombe*. Based solely on my data, I was able to determine the following different karyotypes (Figure 2.7).

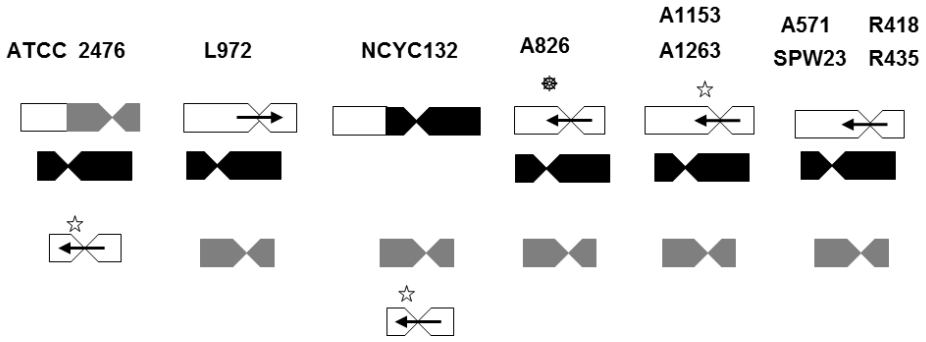


Figure 2.7. Chromosomal rearrangements in natural isolates of *S. pombe*. Schematic representation of the karyotype from the different natural isolate strains. Symbols in strains ATCC2476, NCYC132, A1153, A1263 and A826 represent CRs not fully characterized.

2.4.2 Natural isolates exhibit low genetic diversity.

Taking into account the highly variable karyotypes, I could expect a high degree of genetic diversity. In order to measure the genetic divergence between the natural isolate strains, I sequenced one nuclear (*his3*$^+$) and one mitochondrial (*cox1*$^+$) gene, as well as a small centromeric region of chromosome II (cen2) which has been described as highly polymorphic (Steiner *et al*., 1993).

Nucleotide diversity (average of nucleotide differences per site between two randomly chosen DNA sequences), denoted by π, for the different sequences are listed in Table 2.3. Comparisons were made between all the isolates (π) or between all the isolates and a single outgroup species (π *S. octosporus*, π *S. cryophilus* and π *S. japonicus*).

The gene *his3*$^+$ contains three introns of 55 bp, 185 bp and 59 bp. The introns occupy 20% of the total length of the entire unspliced sequence. Yet, the nucleotide variability for this gene is only 0.5×10^{-3}. Brown and colleagues

have also obtained a low value of nucleotide diversity of 1.6 x 10^{-3} for the largest intron in *S. pombe* (761 bp, 45% of total unspliced sequence).

The value of nucleotide diversity that I have obtained for the intergenic region in cen2 is 6.8 x 10^{-3}, roughly similar to 7.0 x 10^{-3} calculated in the study of Brown despite my smaller sample size. The nucleotide diversity described for this polymorphic region is consistent to what has been estimated in a recent whole genome comparison study between *S. pombe* L972 and NCYC132 strains (Rhind *et al*., 2011). Rhind and colleagues estimated a 8.9 x10^{-3} nucleotide diversity from pairwise comparison of 4-fold degenerate sites between L972 and NCYC132 (Rhind *et al*., 2011). Additionally, it is close to the 5.0 x 10-3 for some natural isolates of *S. cerevisiae* (Fay and Benavides, 2005) and the 5.65 x 10^{-3} described for the world-wide populations of *S. cerevisiae* (Liti *et al*., 2009). So, even though these strains have CRs between them, their nucleotide diversity is still in the same order of magnitude of the species diversity.

2.4.3 Meiotic viability is impaired in heterozygotic crosses between natural isolates.

We have found, like others (Brown *et al*., 2011), that CRs are pervasive in natural populations of *S. pombe* in spite of nucleotide diversity of the order observed within species diversity. This prompted to question whether these karyotypic differences were causing genetic isolation between the natural isolates. In order to measure genetic isolation, I scored hybrid viabilities in pairwise crosses.

Most of *S. pombe* natural isolates are *h90* at the mating type locus, which means they can interchange between h+ and h- mating-types. *h90* strains will preferentially cross with themselves which makes it virtually impossible to perform intercrosses. In order to perform pairwise crosses is necessary to have stable mating-type strains.

The mating-type information is expressed from the *mat1* locus. This locus can exist as P-information (h+) or M-information (h-). These two specific P- and M- DNA cassettes are stored as silenced cassettes in the *mat2-P* and *mat3-M* locus, respectively (Nielsen and Egel, 2007). In the switching process, the P- or M- specific DNA cassette at *mat1* is replaced by the alternative version, which is copied from one of the unexpressed storage loci. All the cassettes are flanked by identical pair of homology boxes named H1 and H2 which are essential for the switching process. In order to stop the switching and obtain stable mating-types, it is necessary to remove the H1 sequence that flanks the *mat1* locus (Figure 2.1a)(Arcangioli and Klar, 1991).

We have disrupted the switchable mating type locus of *h90* strains by removing a minimal sequence from the homology region H1 in all the natural isolates (Figure 2.1a). Once I produced separate h- and h+ mating types for each strain, I performed both intra and inter pairwise crosses. Meiotic viabilities were measured after 5 days of incubation at 32ºC after tetrad dissection. The number of viable meiotic products was divided by the expected resulting from the number of tetrads dissected.

We have chosen strains representative of each chromosomal configuration. I was unable to observe tetrads with strain ATCC2476. As predicted, all strains show high meiotic viability for homozygotic crosses (around 84 ± 6%) except A826 which shows an unexpectedly low viability (57 ± 9 %)(Figure 2.8).

These strains seem to be starting an isolation process as meiotic viability decreases sharply in crosses between different karyotypes. Surprisingly, almost all of the heterozygotic crosses with strain L972 show a severe lethality with viabilities close to 1%. Heterozygotic crosses with strain NCYC132 also have very low meiotic viability. In this case, the less severe heterozygotic cross is with strain SPW23 producing only 43 ± 13% viable offspring.

It seems from this analysis that a difference in an inversion in chromosome I between L972 and all the other strains causes a more severe barrier than a translocation in NCYC132 and the other Brazilian strains (A571, A826, A1153, A1263, R418, R435 and SPW23). All heterozygotic crosses with L972 produce offspring with a maximum of 10 ± 6% viability whether that with NCYC132, are on average 23 ± 9% viable. Additionally, there seems to be a trend towards geographic origin. The Brazilian strains, which come from the same geographic region, are more similar in karyotype and have heterozygotic offspring viabilities closer to homozygotic viabilities.

	Europe	Africa	America						
	L972	**NCYC132**	**A826**	**A1263**	**A1153**	**R435**	**SPW23**	**A571**	**R418**
L972	96 ±4%								
NCYC132	1 ±2%	87 ±9%							
A826	4 ±5%	14 ±8%	57 ±9%						
A1263	5 ±6%	25 ±8%	50 ±13%	72 ±11%					
A1153	3 ±4%	20 ±8%	59 ±8%	63 ±7%	82 ±6%				
R435	0 ±0%	13 ±9%	40 ±11%	51 ±10%	76 ±8%	85 ±8%			
SPW23	10 ±6%	43 ±13%	73 ±10%	72 ±11%	62 ±9%	74 ±10%	89 ±6%		
A571	4 ±4%	n.d.	51 ±11%	n.d.	52 ±11%	n.d.	n.d.	95 ±5%	
R418	1 ±2%	n.d.	n.d.	n.d.	69 ±11%	94 ±6%	79 ±11%	62 ±12%	92 ±7%

Figure 2.8. Hybrid lethality in natural isolates of *S. pombe*. Meiotic viability from crosses between natural isolates. Each cross represents the analysis of 14 to 43 tetrads (containing the four meiotic products). Errors correspond to 2 x Standard Error (S.E.) of a binomial distribution.

2.5 Discussion and Conclusions.

CRs have been described both in natural populations and as a consequence of adaptation to specific laboratory conditions (Dunham *et al*., 2002; Coyle and Kroll, 2008). A very interesting natural example occurs in populations of sticklebacks where an inversion is possibly responsible for the adaptation to either marine or freshwater habitats (Jones *et al*., 2012). Therefore, CRs can appear as a positively selectable macro-mutation. If this is

a general property of different species, I would expect to find them in natural isolates of *S. pombe*.

I have analysed the chromosomal diversity and the meiotic viability of a sample of ten natural isolates of *S. pombe* collected from different geographic regions. Regardless of the small sample size, I have observed high karyotypic variability in face of low nucleotide diversity. This probably accounts for the low levels of meiotic viability obtained in some heterozygotic crosses. Surprisingly, all strains share a common inversion on chromosome I (with the exception of the reference strain) despite their different geographic location. This inversion has also been found in almost all strains studied by Brown and colleagues when analysing a collection of 40 natural isolates of *S. pombe* (Brown *et al*., 2011). Given that the inversion is the most frequent chromosomal configuration, the most likely scenario is that the non-inverted genomic configuration of strain L972 is the derided configuration.

In addition to the pericentric inversion, I detected two translocations involving a common chromosomal arm. The one which is present in strain NCYC1232 is known to be present in at least five other strains spread world-wide (Brown *et al*., 2011). According to these authors, the widespread distribution of the strains with this haplotype suggests that the karyotype variant is long standing relative to the other rearrangements and may therefore define an incipient species. Interestingly, I show that heterozygotic crosses with NCYC132 are highly reduced, likely as a consequence of the rearrangement in question, which supports the hypothesis of incipient speciation by Brown and colleagues.

Even though I have not determined the specific breakpoints for the translocations, the same arm has recurrently been used in both cases. This data is suggestive of what is found also in *D. melanogaster* (Ranz *et al*., 2007) and *Anopheles gambiae* (Pombi *et al*., 2008) where the same chromosomal arm/region is predominantly re-used in some species suggesting a selection for these rearrangements. The same phenomenon

could occur in the *S. pombe* natural isolates in which this specific breakage in chromosome I is related to an adaptive event.

Along with the karyotypic diversity, I have calculated the nucleotide diversity for two chromosomal loci and one mitochondrial gene. I calculated an average nucleotide diversity of 6.8 x 10^{-3} at a intergenic locus of centromere 2 and 0.5 x10^{-3} for the nuclear gene *his3*$^{+}$. These two sequences are located in close proximity in chromosome II. The difference in diversity is likely due to purifying selection occurring on the *his3*$^{+}$ gene. I have measured far higher diversity in the mitochondrial gene *cox1*$^{+}$ (14,5 x10^{-3}) which might reflect selection acting on this important metabolic gene. Still, the nucleotide diversity I have obtained for the two nuclear sequences is in the same order as that of *Saccharomyces* (Fay and Benavides, 2005; Liti *et al*., 2009). Also, despite the low coverage from my sequence analysis, I obtained values similar to those obtained with a whole genome analysis comparison between L972 and NCYC132 (Rhind *et al*., 2011) and in a very recent study comparing eighty-one natural isolates of *S. pombe* (Brown *et al*., 2011).

The above results led me to test the genetic isolation between the natural isolates by scoring hybrid viabilities in pairwise crosses. As predicted, all strains show high meiotic viability for homozygotic crosses, except A826 which shows an unexpectedly low viability. The low level of homozygotic viability has been observed before in crosses of *Saccharomyces paradoxus* (Liti *et al*., 2006). Proper meiosis requires HR for the formation of a physical connection between the homologous chromosomes (Ding *et al*., 2010). Mutations that lower meiotic recombination decrease viability by causing non-disjunction and consequently leading to the production of aneuploidy gametes. I can imagine that perhaps this strain is not so proficient at HR due to the incapacity to execute the processes underlying HR and therefore meiosis is not well processed. Interestingly, this effect seems to be compensated when this strain is crossed with another isolate such as SPW23 which does not have the same chromosomal configuration. It is therefore surprising that

heterozygotic crosses of some of the Brazilian strains with A826 have higher values than its homozygotic cross. These results raise the possibility that a gene involved in meiosis is affected specifically in A826. This could be tested by over-expression of homologous recombination genes in order to try to complement the phenotype.

According to Liti *et al.*, the gradient of gamete viability is correlated with sequence divergence (Liti *et al*., 2006). However, in my study, CRs are likely the strongest factor influencing the gradient in hybrid viability. Still, sequence data is coherent with meiotic viability data: L972 is the most isolated strain both in meiosis and in sequence (for cen2 intergenic region and *cox1+* mitochondrial gene, Appendix B). The interspecific crosses between L972 and any other natural isolate are close to zero. An extremely low meiotic viability of ≤ 1% exists within the *Saccharomyces sensu stricto* (Naumov, 1987). Despite the low viability for heterozygotic involving L972 or NCYC132, they only differ 1% in total genome nucleotide sequence (Rhind *et al.*, 2011) and are part of the same species according to classic classification. A striking similar example but among two close species it that of *S. cariocanus* and *S. paradoxus*. *S. cariocanus* is very similar in sequence to *S. paradoxus* but they differ by four translocations and the meiotic viability between them is extremely low, around 5% (Liti *et al*., 2006). In *Saccharomyces* the low viability of hybrids within the *senso stricto* group has been attributed to the mismatch repair system (Greig *et al*., 2003). The mismatch repair system is responsible for detection and editing of improperly paired bases formed during replication and for correcting mutagenic lesions resulting from modified bases. However, the differences obtained when removing the mismatch repair system account for only a small percentage of the viability (Greig *et al*., 2003) and a large difference is still left to be explained. In the case of L972 and NCYC132 *S. pombe* strains, I propose that the chromosomal differences between them are a major factor for the low viability of hybrids. In this situation, these strains are likely beginning a speciation process which makes them a powerful tool to

study the mechanisms underlying this process. Therefore, supporting the hypothesis that CRs promote speciation, my results show that CRs could act as one of the initial genetic barriers.

In conclusion, the data here presented shows that *S. pombe* is a species with high degree of plasticity in its genome. High genome diversity has also been described at a detailed level using aCGH for wild isolates of *S. cerevisiae* (Carreto *et al*., 2008) also suggesting that genomes are more plastic than once thought. Although CRs are not the sole driver for biological processes such as speciation, it does increase biological diversity upon which natural selection can act. It is important to point that some of these strains might already be undergoing a speciation event as meiotic viability is very low. If this is happening this represents an example where chromosomal variability is preceding nucleotide diversity in the evolutionary process. Therefore, even though pre-zygotic and ecological barriers to gene flow are generally thought to arise before chromosomal barriers this does not need to be the case.

2.6 Acknowledgements. I thank the Portuguese Yeast Culture Collection and CA Rosa (UF Minas Gerais) for the natural isolate strains, B Arcangioli (I Pasteur) for advices on disruption of switchable mating-type, Olaf Neilsen (U. Copenhagen) for the h+ and h- expressing plasmids and H Teotónio for suggestions on the combination of meiotic crosses. I would also thank to P Duarte (IGC) with media preparation and C.C. Reis (IGC) for technical assistance on PFGE. This work was supported by the Portuguese Fundação para a Ciência e a Tecnologia (FCT) PTDC/BIA-EVF/114460/2009. A.T.A. was funded by an FCT PhD scholarship SFRH/BD/33214/2007.

2.7 Tables.

Table 2.1. Natural Isolate strains information.

Strain name	Origin	Geographical location
L972	Grape Juice	Europe
ATCC 2456	Rhum	Europe
NCYC132	African Millet Beerd	Africa
UFMG-A571	Must of Brazilian cachaça	Brazil
UFMG-826	Must of Brazilian cachaça	Brazil
UFMG-1153	Must of Brazilian cachaça	Brazil
UFMG-1263	Must of Brazilian cachaça	Brazil
UFMG-R418	Frozen pulp of Eugenic uniflore (“pi-tanga”, Myrtaceae)	Brazil
UFMG-R435	Frozen pulp of Eugenic uniflore (“pi-tanga”, Myrtaceae)	Brazil
UFMG-SPW23	Frozen pulp of Eugenic uniflore (“pi-tanga”, Myrtaceae)	Brazil

Table 2.2. Natural Isolate strains.

Strain name	Genotype	Common name	Creator
MGF 832	*NCYC132 h90::mat1-M-natMX6*	NCYC-M	This study
MGF 833	*NCYC132 h90::mat1-P-natMX6*	NCYC-P	This study
MGF 836	*R435 h90::mat1-M-natMX6*	R435-M	This study
MGF 837	*R435 h90::mat1-P-natMX6*	R435-P	This study

MGF 838	*SPW23 h90::mat1-M-natMX6*	SPW23-M	This study
MGF 839	*SPW23 h90::mat1-P-natMX6*	SPW23-M	This study
MGF 842	*ATCC2476 h90::mat1-M-natMX6*	ATCC2476-M	This study
MGF 843	*ATCC2476 h90::mat1-P-natMX6*	ATCC2476-P	This study
MGF 846	*L972 h90::mat1-M-natMX6*	L972-M	This study
MGF 847	*L972 h90::mat1-P-natMX6*	L972-P	This study
MGF 848	*A826 h90::mat1-M-natMX6*	A826-M	This study
MGF 849	*A826 h90::mat1-P-natMX6*	A826-P	This study
MGF 850	*A1153 h90::mat1-M-natMX6*	A1153-M	This study
MGF 851	*A1153 h90::mat1-P-natMX6*	A1153-P	This study
MGF 852	*A1263 h90::mat1-M-natMX6*	A1263-M	This study
MGF 853	*A1263 h90::mat1-P-natMX6*	A1263-P	This study

Table 2.3. Nucleotide diversity for natural isolates. Numbers in parentheses represent Standard Deviation.

	***his3*$^+$ (nuclear gene) x 10^{-3}**	***cen2*$^+$ (intergenic region) x 10^{-3}**	***cox1*$^+$ (mitochondrial gene) x 10^{-3}**
π	0.5 (0.1)	6.8 (3.7)	14.5 (10.0)
π *S. octosporus*	53.1 (41.6)	n.s.	57.2 (35.1)
π *S. cryophilus*	54.8 (42.9)	n.s.	58.5 (36.1)
π *S. japonicus*	62.2 (48.7)	n.s.	66.0 (41.2)

Table 2.4. Primers for mat cassete construction.

Primer name	Sequence 5' → 3'
1mat-m 602 f	cccaatttactgaaccctgc
1mat-m 1175 f	cggaaagacactacgtctac
1mat-m 1651 forw	gccatactgttttagaggg
1mat-p 539 frow	gggtaggaacaaagaaagag
1mat-p 1124 forw	gcgtattatggcttggtgta
1mat-p 1625 forw	ggcttttgttgcactatgac
pcr1 mat1-m forw	cccaagctttagacgggaattctttaggg
pcr1 mat1-m rev	gaagatctggagaaagactatacattta
pcr1 mat-p forw	cccaagcttgtttattctatgactacagg
pcr1 mat-p rev	gaagatctagaaaacaaaggagaaagac
pcr2 mat1-m forw	gggtttaaacagatatttgcttcgctacgc
pcr2 mat1-m rev	gcgatatcagacagttaaaattggcggg
pcr2 mat1-p forw	gggtttaaacccctctaacgagatatttgc
mat locus end rev	tgtgaatggagttggagagg
matl-m 98 r	aataatgtcagcagaagacc
matl-m 1065 f	ttactgccctgattctatcg
MM	tacgttcagtagacgtagtg
MP	acggtagtcatcggtcttcc
MT1	agaagagagagtagttgaag
ptopo +200 r	ctcactatagggcgaattgg
ptopo -200 f	gagttagctcactcattagg

Table 2.5. Primers for DNA sequencing.

Primer name	Sequence 5' → 3'
cen2 6714 forw	taggtatttccaagggctcc
cen2 6846 forw	cacatcgtagtaattggtcc
cen2 7350 forw	cgcattctctcaagatgtcg
cen2 7555 forw	gattcgaaggtctttactcg
cen2 7960 rev	gtcttccttttaaccaggc
cox1 -250 forw	tgatctataggttcgagccc
cox1 187 forw	gttgcaatctcagcacatgg
cox1 620 forw	gagtgtgtttaatgaagagg
cox1 1169 forw	gatttattgaagcagaaggg
cox1 1637 forw	acttctcatagtggtccagc
cox 2000 rev	tagcccatgcaaataatggc
cox1 2127 forw	gaacctcagagactttacgc
cox1 2549 forw	aataatcattggctagctgg
cox1 3048 forw	ttaattatgccagctttcgg
cox1 3495 forw	tggagtccaaaaatgtttgg
cox1 4026 rev	gtgttctctcacatataagg
his3 -20bp f	gatatcacctgcatcttagg
his3 -20bp r	ggtgtgtttccgtatagttc
his3 f	cattgcgtatcactgttggc
his3 r	cgcacagcgataaggctgaa
his3 354 f	tggttcggttatttccgtgg
his3 795 f	gaatggaatcgttgtcgtcg
his3 1256 f	aggaggtaagcctagtaacg

Table 2.6. Primers for Southern blot probes.

Primer name	Sequence 5' → 3'
imr1l rev	aagtattaagttttaatgcc
imr1l forw	gaaagtatacttagattttcg
imr2l rev	agacacaagcattgttctcg
imr2l forw	tcacatggtcgataatttcc
imr3l rev	ctttgataactgaacagcgg
imr3l forw	tgaaatcgtatatgactgcg
lig4 screen f	gcgagtactaacatattagg
lig4 screen rev	aaatggtgctgtaagttgcc
mcr1 2474 forw	ggtagcatcaaagttgtccc
mcr1 +1514 rev	gccttgtccttcacttttcc
rad3 6219 forw	ggatggcaacttatacccat
3' utr rad3 r ko	cggatgatgactcagaaaca
rad32 screen f	aataagcgacgagaattacc
rad32 screen rev	agtggagtacgttacatcgc
5' mad2 f251	cccaaactgacattattacg
mad2 r 556	acttcgctgtctttatcagc
taz1 seq1620 top	cgttcgggtacacgcagg
taz1 3' r	gatggcatatgtataaagac
e probe start forw	ttcatcatcgccgtattccg
e probe start rev	tggtgctataagggtttgcg
2300000 chrI forw	acgatcacgatcaacttagc
2300000 chrI rev	tctcgccaattcaagagacc
2719860 forw chrI	gaggaattagatgtggaggg
2719860 rev chrI	ctaggtcgcaactatactcg
2723000 forw chrI	ttgcgtatgccattcaatcg
2723000 rev chrI	agatatggcttggactcacc

Table 2.7. Primers for *NotI*-sites amplification.

PCR product	Primer name	Sequence 5' → 3'
#1	5073760 forw	ccgggacatttatttacagc
	5073760 rev	taggtgattacatgaaagcg
#2	4076940 forw	aaagcaacgtgcaatctccg
	4076940 rev	gggttattttaagggcgagc
#3	4072320 forw	tcaaacggcatagtgaatcc
	4072320 rev	gatacaagataatctgcagc
#4	3597760 forw	taacaaatgagacaaggagg
	3597760 rev	taacgaggattagtgcatcg
#5	2722260 forw	cggtaatatttcaggtatgc
	2722260 rev	ttcagattggcagttactgc
#6	2130280 forw	tgtgtgaattttgtgtgtcg
	2130280 rev	gcatcacttagtagatgtgc
#7	1024620 forw	agctttgtcacaggataagg
	1024620 rev	atctctgcttcctttgttgg
#8	f probe forw	gcttcttcaatggcctttcg
	f probe rev	tcggaaagtttacggacacg
#9	202302 chrII forw	acataccgtttctacttcgc
	202302 chrII rev	cactggatatcgtgttttcc
#10	289360 chrII forw	ctgtgtactttacgacatcg
	289360 chrII rev	ttgcaaggagaagtttcacc
#11	465160 chrII forw	acaattgtgccggaagtagc
	465160 chrII rev	acgcttgctgttcttcttcc
#12	2406620 chrII	tgtgtcgctacaaattctcg
	2406620 chrII rev	ctgtttggaattctctgtcc
#13	3023560 chrII	gtgattgtaccaaagaagcg
	3023560 chrII rev	tatgccaacagtatcttggc

Chapter 3 – Meiotic and mitotic fitness of genetically engineered CRs.

3.1 Abstract.

Several lines of evidence point to the importance of CRs in adaptation, both in wild populations and in cancer cells. However, in these studies it is difficult to disentangle the contribution of gene alterations from chromosomal differences as both are usually simultaneously present. This is even harder when one wants to study the impact of CRs for meiosis and hybrid sterility.

Saccharomyces cerevisiae and *Schizosaccharomyces pombe* are different yeast species that have diverged millions of years ago. Since the genome of *S. pombe* resembles more human genomes in structure, this model organism is well suited for understanding the cellular mechanisms that are altered with chromosomal structural alterations.

I sought to determine if CRs can be a selectable trait in the absence of adaptation. In order to do so I constructed a range of ten different karyotypes, including two pericentric inversions and eight reciprocal translocations, in *S. pombe*. Analysis of these strains demonstrates that even in the absence of nucleotide sequence variation, gene expression is altered. I also show that CRs can act as an isolating barrier for incipient speciation processes in cases where rearrangements are only slightly deleterious, or even beneficial on mitotic fitness. I found several cases of antagonistic pleiotropy where CRs with deleterious effects during meiosis have beneficial effects during mitosis. My results constitute the first direct determination of the fitness effects of these mutations and importantly suggest a mechanism of selection that can explain their polymorphism in nature.

3.2 Introduction.

Aneuploidies are one of the most observed macro-mutations in experimental evolution studies (Dunham *et al*., 2002; Gresham *et al*., 2008; Dhar *et al*., 2011) or as mechanism for drug resistance in yeast (Selmecki *et al.*, 2006, 2009; Poláková *et al*., 2009). Nonetheless, CRs, particularly translocations, are a prominent feature of several cancers and specifically the major cause of most leukemias and lymphomas (Rowley, 2001). In these cases, chromosomal translocations lead either to altered gene expression, for instance by the juxtaposition of oncogenes near promoters/enhancer elements, or to fusion of genes. Consequently this results in the deregulation of cellular proteins especially those coded by proto-oncogenes and tumour suppressor genes, which are critical functional regulators of the cell (Mitelman *et al*., 2004; Nambiar and Raghavan, 2011). The most common example is the Philadelphia chromosome first described in 1961 by Nowel and Hyngerford in chronic myelogenous leukemia patients (Nowell and Hungerford, 1960). This is a reciprocal translocation, creating an extended chromosome 9 (der 9), and a shortened chromosome 22 (the Philadelphia chromosome). The t(9;22) translocation event of chromosomes 9 and 22 at the *BCR* and *ABL* genes, respectively, gives rise to an oncogenic protein: a constitutively active *ABL* tyrosine signalling kinase resulting from the fusion of the two genes (Nambiar and Raghavan, 2011). However, translocations can also occur in healthy individuals and t(14;18) is the most commonly reported one with a prevalence of around 30-60% based on data from European and American individuals (Limpens *et al*., 1995; Fuscoe *et al*., 1996). Additionally, some healthy individuals contain a very small percentage of cells in which the t(14;18) breakpoints are equal to those detected in patients of follicular lymphoma (Limpens *et al*., 1995; Fuscoe *et al*., 1996).

Cancer stands as an example where CRs may play a role in adaptation at the cellular level. Cancer cells possess highly disorganized genomes as compared to normal cells (Albertson, 2006). Even though this phenomenon is considered an abnormality these cells manage to be highly successful in terms of proliferation and adaptation. Indeed, they become independent entities as they manage to subvert the cellular processes that characterize normal tissues. CRs are frequently encountered in tumours which suggest that they have adaptive properties. Moreover, the same specific rearrangement occur in different individuals across similar tumours which is indicative of a some mechanism in the choice of rearrangement partner (Wijchers and de Laat, 2011).

There is an increase in research activity aimed at understanding the mechanism underlying the formation of chromosomal translocations and inversions. Particularly, it is suggested that genomes are non-randomly arranged within the nuclear space (Olsson and Bjerling, 2011) and chromosome organization in the nucleus and chromatin status may influence the impact of chromosome structure in the cell (Meaburn *et al*., 2007; Wijchers and de Laat, 2011). This non-random organization also has profound effects on gene expression. The nucleus has distinct structural and functional compartments and individual chromosomes occupy specific positions in the nucleus known as chromosomes territories (Lanctôt *et al*., 2007). Translocations and inversions change the localization of genes; in these cases the chromosomes, parts of chromosomes or individual genes might not move according to a normal regulation of the cell and therefore transcripts and/or replication does not occur properly or at the regular speed (Hiratani *et al.*, 2008). In the case of *Schizosaccharomyces pombe* the constitutive heterochromatin is found at the nuclear periphery, similar to mammalian cells (Olsson and Bjerling, 2011). I expect that CRs in *S. pombe* will modify the normal nuclear localisation of genes and possibly of

their heterochromatic state and expression level in a manner similar to what happens in cancer human cells (Harewood *et al*., 2010).

Chromosomal rearrangements in the form of inversions are widely distributed both between species and within species, segregating as polymorphisms (Hoffmann and Rieseberg, 2008). In the 1930s, Dobzhansky first described a chromosome inversion polymorphism segregating in a natural *Drosophila* population (Dobzhansky and Sturtevant, 1938). Since then *Drosophila* has been the best characterized group and a full analysis of their karyotype is available since the release of the sequences from 12 different *Drosophila* species (Bhutkar *et al*., 2008). Indeed, inversions are highly present in nature differentiating between close species such as the tomato and potato genome (Bonierbale *et al*., 1988), the human and the chimpanzee (Feuk *et al*., 2005; Tuzun *et al*., 2005; Flores *et al*., 2007) and in several species of fungi (Perkins, 1997). Inversions are also present as a polymorphic variation within species in several organisms for example, the fruitlies *Rhagoletis* (Feder *et al.*, 2003b)*, Anopheles gambiae* (Brooke *et al*., 2002), *Drosophila montana* and *Drosophila virilise* (Morales-Hojas *et al.*, 2007) and *D. pseudoobscura* (Anderson *et al*., 1991). Importantly, *Drosophila* data show that frequencies of large inversions usually vary clinally and/or seasonally and therefore are influenced by selection in at least some environments (Krimbas and Powell, 1992). Evidence for selection on inversion polymorphism is also increasing in other organisms (Feder *et al*., 2003a; Stefansson *et al.*, 2005; Selmecki *et al.*, 2009). Furthermore, even in experimental laboratory evolution studies, the appearance of similar recurrent genomic rearrangements suggests that they may be adaptive and responsible for the increased fitness of these strains (Dunham *et al*., 2002). For instance, adaptation in yeast to glucose-limited (Dunham *et al*., 2002) and phosphate limited media (Adams *et al*., 1992) is achieved by the appearance of CRs. Recently, the

same phenomenon was also observed in resistance to salt stress (Dhar *et al*., 2011). Interestingly, Dhar *et al.* observed that single nucleotide mutations are not always the most common mutations leading to an adaptive phenotype and that ploidy variation is a form of resistance to saline stress.

In this work I measured the meiotic and mitotic fitness of two pericentric inversions and eight different reciprocal translocations in the fission yeast *Schizosaccharomyces pombe*. *S. pombe* is a great model for chromosomal studies as its chromosomes are structurally similar to those of human. For instance, the nucleosome, the basic organisation unit of chromatin, is similar between human and yeast cells (Olsson and Bjerling, 2011). In addition it is a haploid organism with a generation time between 2 to 4 hours, depending on the environment, and with a small genomic content of three chromosomes with a total size of approximately 14Mb (Wood *et al.*, 2002). In *S. pombe*, the centromeres are attached to the SPB and the mating type region is closely associated with the centromeres. The telomeres are attached to the nuclear membrane at the opposite side of the cell nucleus, as compared to the SPB, in proximity to the nucleolus (Chikashige *et al.*, 1997). Fission yeast is also a great model organism for meiosis studies since analysis of the fours spores coming from each tetrad reveals all of the products of a single meiotic recombination event.

3.3 Materials and Methods.

3.3.1 Strains and media. The strains used in this study are listed in Table 3.1. Standard media and growth conditions were used throughout this work (Moreno *et al*., 1991) with minor modifications (see detailed protocols in Chapter 2 Materials and Methods). Cultures were grown in rich media (YES) or minimal media (EMM) at 32ºC with appropriate shacking, unless otherwise stated. Tetrad dissection for meiotic viability analysis was per-

formed in YES solid media. For competition assays, media containing drugs was prepared fresh daily. Hydroxyurea (Sigma- Aldrich) was directly added in powder to YES at the desired amount. An MBC (Carbendazim 97% Sigma- Aldrich) stock was prepared fresh in DMSO (Sigma- Aldrich) at a concentration of 5mg/ml and then a 1:10 dilution of this stock was used to prepare the final media. Caffeine was stock at -20ºC in a concentration of 80mg/ml in sterile miliQ-water. To make 3% EtOH, this percentage was added to fresh media from a 100% EtOH solution. To prepare YES pH=3, a 5 M HCl was added to YES and the pH measured with a pH testing tape. Strains were constructed using standard genetic techniques for fission yeast (Bähler *et al*., 1998). Strains generated in this study were made, when appropriate, by transformation of specific fragments or by mating with available strains.

3.3.2 Yeast cells transformation. Transformation of yeast cells was performed using the lithium acetate protocol as described previously (Chapter 2 of this thesis; Moreno *et al*., 1991).

3.3.3 Construction of Genetically Engineered Strains. For the parental strains, different combinations of N. Kleckner's (U. Harvard) strains were crossed in order to obtained different haploid strains with selected combination of the loxP cassette at the auxotrophic marker (Figure 3.1a and Table 3.1). MGF1070 and MGF2098 were generated by protoplast fusion of strains MGF539 and MGF542 has described in Gingold, 1986. For sporulation of the diploids, cells were transformed with the pON104 (*mat1-P*) plasmid. The haploid strains containing the loxP and KanMX cassettes at different auxotrophic markers (N. Kleckner strains) were all h-. In order to obtain the parental strains containing the cassettes at two auxotrophic markers (breakpoints) it was necessary first to create a diploid. Protoplast

fusion was the initial protocol followed in order to generate the diploids which after sporulation would allow for the selection of the parental haploid containing both the markers. However, this protocol was highly variable and the parental P5 was the only to be successfully constructed with this protocol. I have therefore changed the strategy to insert the h+ information using the cassettes constructed in Chapter 2 of this thesis. Strains were selected based on the simultaneous presence of padh1-loxP-kanMX6 and loxP-ura4-kanMX6 fragment after PCR amplification. These strains are referred to as parental strains. For generation of rearranged strains, parental strains were transformed with pREP81-Cre and pREP81 (as a control) and plated directly in selective media EMM –leu with 4 mM thiamine. Once colonies appeared, single colonies were inoculated overnight in liquid media EMM -leu and plated in EMM -ura 18/20 hours after for selection of the recombination event. Single colonies were re-stricked into EMM -ura and then to YES and genotyped. Cells were frozen at this step while the *ura4*+ phenotype was also being confirmed by PCR using the following pairs of primers: 3'utr adh1 F and Kanmx 880rev (2000bp in parental, 3000bp in rearranged); 3'utr adh1 F and ura671 rev (product present only in a rearranged strain). Independently, PCR amplification using different pairs of primers according to the strain's specific auxotrophic markers (arg7 -200 forw, arg1 -200 forw, his1-100 forw, lys3 -200 forw and lys4 -200 forw) in combination with KanMX6 880 rev will give a 2000bp and 2500bp PCR products for parentals breakpoint configuration and a 3000bp and 1500bp PCR products for the rearranged configuration. Primer sequences are listed in Table 3.3. All the strains were immediately frozen as soon as the genotype was confirmed. As parental strains are *ura4*$^{-}$ and CR strains are *ura4*$^{+}$, I had to obtain *ura4*+ Parental strains. I call these Control strains (Figure 3.1b and Table 3.1). They were created by crossing Parental strains to the prototrophic strain L972.

Selection was then performed both on plate and PCR using the pair of primers described above. In order to integrate the GFP cassette, a new resistance marker had to be used. The available GFP plasmids contained the already in use kanMX6 cassette. Therefore, the kanMX6 was replaced for the hphMX6 marker in the pks168 (pFa6a-kanMX6-P3nmt1-GFP) and pks394 plasmids (pFa6a kanMX6-P3nmt1-mCherry)(Snaith *et al*., 2005). Briefly, pFa6a-3HA- hphMX6 and pFa6a-KanMX6-P3nmt1-GFP were digested with the restriction enzymes *PmeI* (NEB) and *BglII* (Fermentas). Confirmation of the clonings for the replacement with hphMX6 was made by EcoRI digestion which cuts the hphMX6 fragment but not the KanMX6.

GFP containing strains were created by insertion of 3nmt1-GFP-hphMX6 fragment near one of the breakpoint loci (Figure 3.1b and Table 3.1). Primers are listed in Table 3.4. For generation of this fragment, a PCR amplification of the final plasmid (pFa6a-3nmt1-GFP-hphMX6) was made using the following primers: *arg7* - arg7 locus integr 80mer forw and arg7 locus integr 80mer rev; *his1* - his1 locus integr 80mer forw and his1 locus integr 80mer rev; *lys3* - lys3 locus integr 80mer forw and lys3 locus integr 80mer rev; *lys4* - s and lys4 locus integr 80mer rev. PCR program using 100mers: 1) 95ºC 5 min; 2) 95ºC 1 min; 3) 45ºC 2 min; 4) 72ºC x extension time (usually 30 sec for each 500bp); 5) Go to 2, 5 times; 6) 95ºC 30 sec; 7) 50ºC 30 sec; 8) 72ºC x extension time; 9) Go to 6, 35 times; 10) 72ºC 10 min; 11) 4ºC 10 min; 12) end.

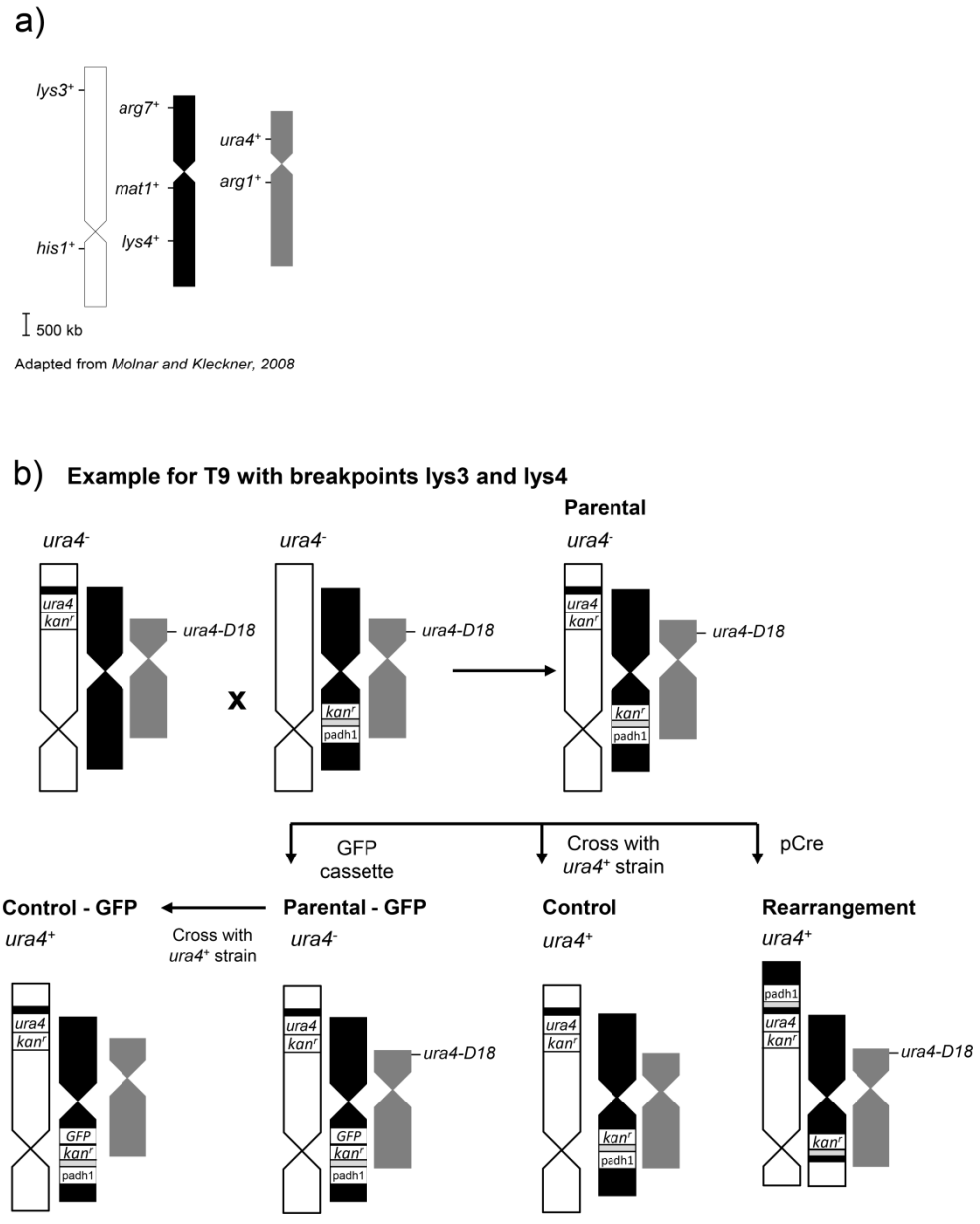


Figure 3.1. Construction of genetically engineered chromosomal rearrangements. a, localization of auxotrophic markers used for breakpoints. **b**, phylogeny for the construction of the different strains. Example for control and translocation 9 with breakpoints at locus *lys3*$^{+}$ and *lys4*$^{+}$.

3.3.4 Meiotic viability. For meiotic crosses cells were grown overnight in EMM minimal media with all supplements and diluted the day after. When cells reached exponential phase, 1 ml of culture was centrifuged and washed twice in 1 ml of EMM without supplements. The final pellet was resuspended in 100 μl EMM without supplements. This was made for both h- and h+ cultures. 50 μl of each mating type was mixed in a eppendorfs and pipette up and down. The desired volume was pipetted into a ME plate, allowed it to dry and incubated at 25ºC for 2-3 days. When tetrads were formed, an aliquot of the cells from the ME plate was putted into 40 μl of sterile H_2O. A drop of this mix was placed in a corner of a YES plate, letting it slide into the plate, making a line from one side of the plate to the other. The plate was allowed to dry and the tetrads were dissected in a micromanipulator equipped with a glass needle (Singer Instruments). The plate was incubated for 6 hours to allow the release of the spores from the tetrads. After, the spores were dissected and the plate was incubated for 5 days at 32ºC, time after which colony formation was scored. Meiotic viability was given as the number of colonies formed divided by the number of colonies expected.

3.3.5 Pulse Field Gel Electrophoresis. PFGE Protocol was performed as described in Ferreira and Cooper, 2001 with some modifications (Chapter 2 of this thesis).

3.3.6 Competitive fitness assays. To measure fitness cost of the CR, a FACS-based competitive assay was done. Unlabelled rearranged and control strain were independently competed against its specific reference GFP-labelled strain (Control-GFP). To do so, both CR and Control strain were grown in YES liquid medium for 24 hours at 32ºC with aeration and shacking. The day after cultures were diluted and when they have reached

exponential phase they were mixed at 1:1 ratios with the fluorescently labelled control strain. Fresh media was added to make initial cell density in the order of 1 x 10^5 cells/ml. The effective ratios of the mixture were measured using CyAn™ ADP Analyzer (Beckman Coulter, Inc.). Competitions were performed in a final volume of 600 μl of media in a 96-deep well plate (VWR) incubated in a mini-shaker (VWR) with 850 rpm. Unless otherwise stated, cells were grown at 32ºC. Every 24 hours cells were diluted 1:30 to new fresh media. This was repeated three times. After 4 days, corresponding to approximately 20 generations, the final ratios of the unlabelled to labelled strain were measured. Selection coefficients were calculated using the formula: $\Delta S = \frac{1}{t} \times ln\left(\frac{ratio\ t0}{ratio\ tf}\right)$ where $ratio = \frac{frequency\ of\ unlabeled\ strain}{frequency\ of\ labeled\ strain}$ for t_0 (initial, time zero generations) and t_f (final, time 20 generations). The selection coefficient was estimated as an average of at least twenty independent competition assays. The number of GFP-positive (reference) and non-fluorescent (experimental) cells was determined using a CyAn™ ADP Analyzer (Beckman Coulter, Inc.) counting 10.000 total cells for each sample.

3.3.7 qRT-PCR. Total RNA was isolated from exponentially growing cultures of each strain with RiboPure-Yeast kit (Ambion). Synthesis of cDNA was performed using MMLV Reverse Transcriptase (Promega). A mix of 1 μg of RNA, 0.5 μl of random primers (Promega) and up to 15 μl of H_20 DEPC treated water was incubated in a thermocycler for 5 minutes at 70ºC followed by 5 minutes on ice. For each sample, it was after added 5 μL Reverse Transcriptase Buffer, 1.25 μl dNTP 10 mM, 1 μl Reverse Transcriptase (MMLV RT 200un), 2.75 μl H_20 DEPC treated for experiment (RT). In the case of the control, no reverse transcriptase was added; in this case the volume of H_20 DEPC treated was 3.75 μl. PCR program: 25ºC for

10 min; 37ºC for 60 min; 85ºC for 10 min; 4ºC forever. cDNA was kept at -20ºC for short term and at -80ºC for long term storage. The cDNA control reaction for qPCR amplications was RNA that went through the cDNA cycle but without addiction of the reverse transcriptase enzyme. Quantitative PCR was performed using SYBERGREEN PCR Mater Mix (Applied Biosystems) following the manufacturer's specifications. Primers for qRT-PCR were designed manually for an average size product of 130bp and 50% GC content. All qRT-PCR reactions were performed in a 10 μl final volume and three replicates per sample (technical replicas) and set up in a 384-well plate. Primers are listed in Table 3.6. Quantitative DNA amplification was carried out in the ABI 7900 HT machine (Applied Biosystems) in a two-step process as follows: 2 min at 50ºC, a denaturation step of 10 min at 95°C and 40 cycles of 95°C for 30 sec each, with annealing and extension at 72°C for 1 min and 30 sec respectively. A final dissociation cycle of 15 sec at 95ºC, 15 sec at 55ºC and 15 sec at 95ºc was added. The reporter signals were analysed using the SDS 2.4 software (Applied Biosystems). The C_T value of the gene coding for actin (*act1+*) was used for normalization of variable cDNA levels and induction factors were determined for each gene and condition.

3.3.8 Propidium Iodide FACS analysis. FACS was performed on ethanol-fixed cells using standard procedures. To prepare cells for FACS analysis, 500 μl of fixed cells were centrifuged for 3 minutes at 3000 rpm. Supernatant was removed and cells washed twice in 1 ml of a 50 mM NaCitrate solution (filtered stock at 0.5 M). After the final wash, the pellet was resuspended in 500 μl of 50 mM NaCitrate solution containing 3 μl of RNAse (Sigma) and incubated at 37ºC for 5 minutes. For FACS detection, propidium iodide was added to an aliquot of cell suspension to make a final concentration of 4 μg/ml (Propidium Iodine stock of 4 mg/ml in H_2O, filtered and

stored in dark at -20°C). Cells were processed immediately (they can also be conveniently stored overnight at 4°C in the dark before processing the next day).

3.3.9 Growth assays. Growth curves were performed in a 96-deep well plate (VWR) format. Briefly, cells were grown overnight in rich in a volume of 600 μl per well in 96-deep well plates. The day after, cells were diluted 1:30 to the appropriate media and time points were taken, approximately, every 2 hours. Cell O.D. was measured in a final volume of 100 μl in 96-well flat bottom at 600 nm in a Victor III apparatus (Perkin Elmer).

3.3.10 Viability Assays. Drug sensitivity analysis was performed as described in Miller *et al*., 2006. Briefly, exponentially growing cells in YES rich media were collected and counted. Five-fold serial dilutions from a starting concentration of 1×10^7 cells/ml are prepared and 5μl spots of the serial dilutions are spotted into the specific media containing agar plate.

3.4 Results.

3.4.1 Genetically Engineered Chromosomal Rearrangements reduce meiotic fitness.

CRs have been for long time pointed out has a highly significant macro-mutation. Yet, a systematic characterization of their effects has never, to my knowledge, been done. I have proposed to quantify the fitness effects of CRs, both for meiosis and mitosis. In order to do so, I constructed a collection of strains, each containing a single CR (Figure 3.2a). The strains were generated using the Cre-LoxP system inserted at different auxotrophic markers (Figure 3.1a)(Molnar and Kleckner, 2008). All strains are isogenic with respect to their parental except for the *ura4*$^+$. Parental strains are *ura4*$^-$ and in order to be fully isogenic, I crossed them with a

ura4$^+$ strain and selected for the presence of this marker. I have named this strain, Control strain. CR and Control are thus isogenic except for the rearrangement (Table 3.1). I should point out that the *ura4*$^+$ is being expressed from a constitutive promoter in the CRs and from the endogenous promoter in the Controls. Thus, all the quantifications made in this work show the relative selection coefficient value to the control (collinear configuration) for the CR in question, either an inversion or a translocation. For simplicity, from now on I will refer to the control strains as parental.

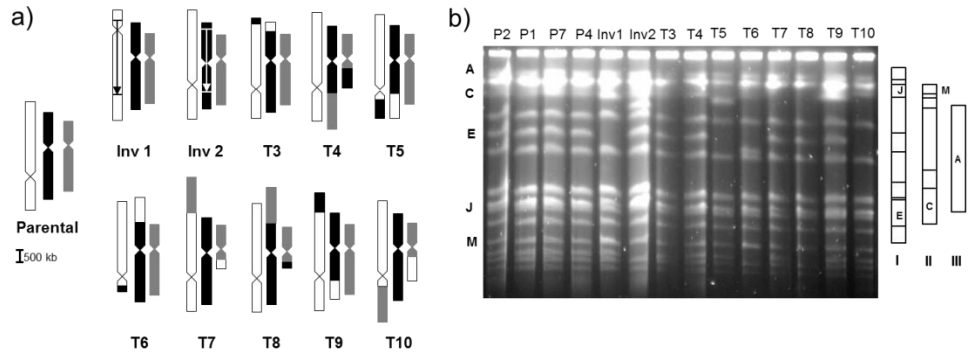


Figure 3.2. Genetically engineered chromosomal rearrangements. a, collection of strains containing rearrangements. **b**, *NotI*-restriction profile of the chromosomal engineered strains. Some rearrangements are not detectable in the *NotI*-restriction PFGE profile since the sizes of the altered bands are too large (e.g. T4, T7) or too small (e.g. T8) and this in lower resolution areas, or the news bands formed have a similar size to parental strain fragments (e.g. T3, T5, and T9). In Inv1, T5, T6 and T10 disrupt the E fragment which is very visible. Consider the following examples: in T4, fragment A changes from 3500bp to 2708bp and fragment C from 1525bp to 2317bp; in T7 fragment J from 500bp to approx. 2390bp and fragment A from 3500bp to 2140bp; in T8 the M fragment 240 to 1900bp and A from 3500bp and 1900bp; in T9 fragment C from 1525bp to 960bp (same size as E fragment) and J from 500bp to 1595bp (same as C fragment).

In order to exclude possible genomic instability problems or aneuploidies with these strains, I analysed their DNA profiles and have not detected gross deviations to normality (Figure 3.3).

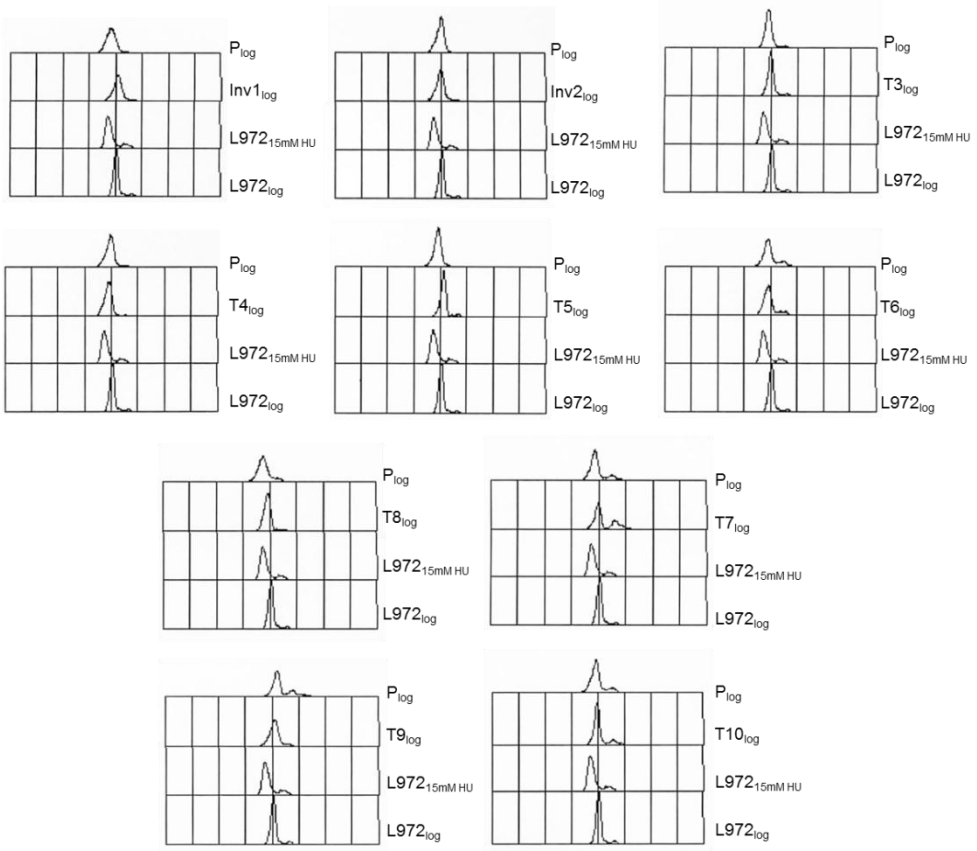


Figure 3.3. FACS profiles of DNA content in chromosomal rearranged strains. All strains were analysed for their DNA content in exponentially growing cultures. *S. pombe* spends most of its cell cycle in G2 phase (2C DNA content). Only upon nutritional starvation, e.g. NH_4 starvation, it will arrest the cell cycle at the G1 phase (1C DNA content). An early S phase arrest can also be obtained using drugs that arrest DNA replication such as hydroxyurea (HU). L972 strain was used as a control both for exponentially growing and early S phase -arrested cells. All strains contain normal 2C DNA content; no aneuploidies were observed.

To quantify the exact impact of a CR for hybrid viability, I performed the following combinations of meiotic crosses: parental with parental (P x P), parental with rearranged (P x R) and rearranged with rearranged (R x R)(Figure 3.4a). I have not seen a bias of lethality or segregation distortion either for the parental or for the rearrangement breakpoints as both are equally represented in the progeny (Figure 3.4b).

Parentals of Inv1 and Inv2, have high offspring viability of 87 ± 7% and 92 ± 4%, respectively. Both inversions 1 and 2 have also high homozygotic R x R offspring viability of 86 ± 3% and 90 ± 5%, respectively. However, there is a drop of 37 ± 13% and 40 ± 13% viability in heterozygotic crosses P x R with inversions relative to that of P x P crosses.

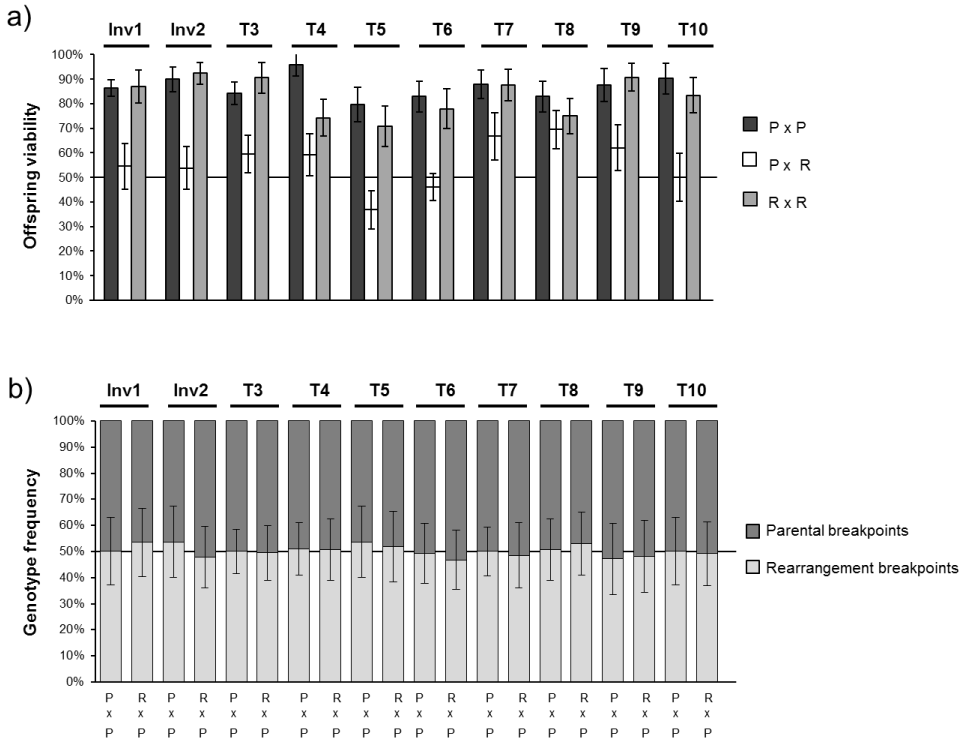


Figure 3.4. Meiotic viability of chromosomally rearranged strains. a, meiotic viability of homozygotic and heterozygotic crosses. Abbreviations: P – parental strain; R – rearranged strain. Error bars represent 2 x S. E. for a binomial distribution of n > 20 tetrads (> 100 individuals) analysed. **b**, Frequency of breakpoints from Parental or Rearranged in the progeny. Error bars represent 2 x S. E. for a binomial distribution.

Taking into account Mendel's law for independent segregation of loci this would be the expected survival of heterozygotic meiosis with inversion when there is no linkage of loci. The lethality comes in fact from the recombinant products that generate imbalanced gametes. Moreover, this value is in agreement with the high percentage of 2-spore tetrad distribution in heterozygotic crosses observed for both of these inversions (Figure 3.5).

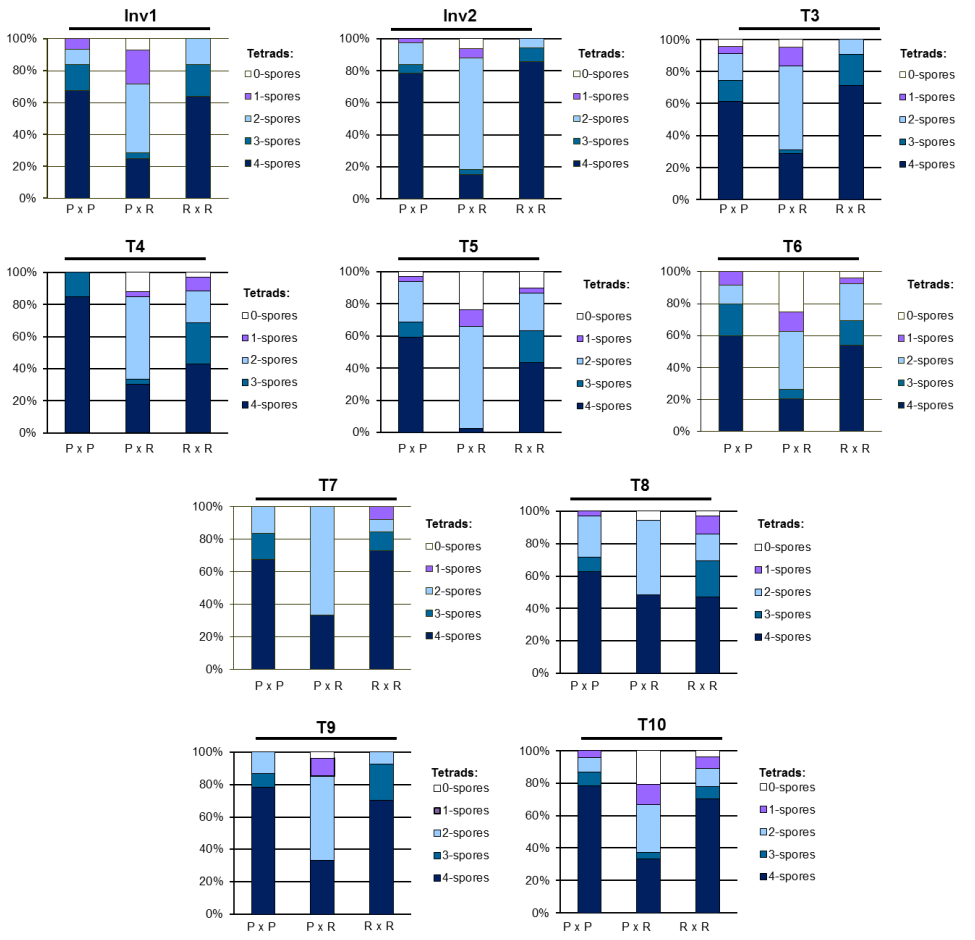


Figure 3.5. Tetrad numbers distribution. Distribution of the number of tetrads according to the number of spores that formed a colony. More than 20 tetrads (> 100 individuals) analysed.

Heterozygotic crosses involving translocations T3, T5, T6, T9 and T10 lead to the average losses in hybrid viability relative to the P x P cross of 31 ± 12%, 54 ± 14%, 44 ± 10%, 29 ± 16% and 45 ± 15%, respectively. Translocations T4, T7 and T8 are the ones affecting the meiosis less. Translocation 7 causes an average drop in viability in heterozygotic crosses relative to the P x P of 24 ± 16% and T4 and T8 have almost no observable

effect. Translocation T4 *per se* causes an immediate impairment in meiosis as R x R for this rearrangement has an average lethality of 23 ± 10%. This value does not drop further in the heterozygotic P x R cross. As for translocation T8, this seems to be an interesting exception as there is no reduction in meiotic viability in P x R crosses. In addition to what is described above, with the exception of T4, all rearrangements to not seem to have meiotic problems on their own. Also, the less severe ones are still able to produce 4-spore tetrads in P x R crosses (Figure 3.5). This suggests that their chromosomal configuration is probably not disrupting recombination hotspots which would otherwise lead to a higher production of lethal genotype combinations. The opposite seems to be true as the more deleterious translocations produce mainly 2 and 0-spore tetrads, thus putting forward the hypothesis that perhaps, in this case, the breakpoints are at places of higher recombination.

3.4.2. Inversions and translocation lock the loci involved in the rearrangement but only inversions reduce homologous recombination.

In order to quantify the impairment of CRs in homologous recombination, I measured the number of recombinant progeny for two loci: the breakpoint locus and an internal locus, the mating type locus (*mat1*$^+$). The breakpoint loci have different size cassettes in the P or R strains and this was used to distinguish the breakpoint case in the progeny. However, apart from the breakpoints, the only other genetic difference in a meiotic cross between P and R was the mating-type locus which could be either h+ or h-. Therefore, only rearrangements involving chromosome II were considered as *mat1*+ is located in chromosome II at approximately position 2.1Mb. Since both *lys4*$^+$ and *arg7*$^+$ are more than 1Mb apart from *mat1*$^+$, I predicted a 50% recombination of these alleles in collinear chromosomes.

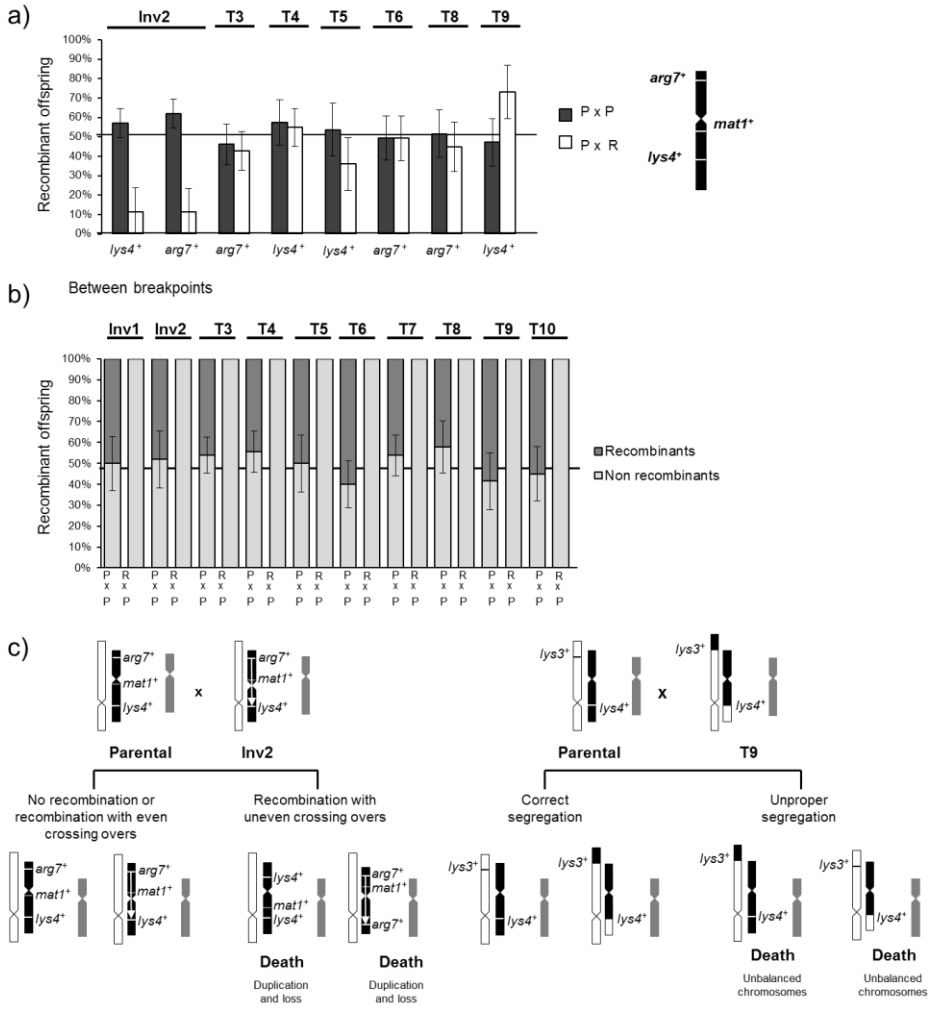


Figure 3.6. Recombination between breakpoint loci. a, frequency of recombination between *lys4*$^+$ or *arg7*$^+$ breakpoint locus and the *mat1*$^+$ locus. Only rearrangements involving chromosome II were considered. Error bars represent 2 x S.E. for a binomial distribution (n> 52 individuals) analysed. **b**, frequency of recombination between the two loci involved in rearrangements. Hetererozygotic crosses with strains containing any of the CRs leads to the production of hybrids that contain in the breakpoints either one of the two parental combinations (parental or rearranged). Therefore, inversions and translocations cause total linkage between genes close to the breakpoint. Error bars represent 2 x S.E. for a binomial distribution (n> 52 individuals) analysed. **c**, outcomes of meiotic recombination between breakpoints loci.

I observed, as expected, free recombination between the markers *lys4*$^{+}$/*mat1*$^{+}$ and *arg7*$^{+}$/*mat1*$^{+}$ in homozygotic P x P crosses as denoted by recombinant frequency around 50% (Figure 3.6a). All heterozygotic P x R crosses for translocations maintain recombination levels around the frequency observed for its respective control crosses. Thus, translocations do not reduce homologous recombination. As far as inversions are concerned, recombination between any of the two breakpoint loci and *mat1*$^{+}$ is significantly reduced in heterozygotic crosses. In this case, heterozygotic crosses with the pericentric inversion Inv2 produced only 11 ± 7% recombinants (Figure 3.6a). It is important to stress out that this reduction seen in homologous recombination is likely a result of the death of certain recombinant spores that have unbalanced DNA content. In order to be viable, the recombinants have to originate from meiotic crossing-overs with an even number of crossing-overs inside the inversion or one inside and the other outside. Uneven numbers of crossing over inside the inversion will lead to unbalanced DNA content. Hence, this reduction likely does not arise from an active mechanism of reduction of recombination but reflects of the structural impediment of pairing and resolution between inversions and collinear chromosomes. My results reinforce what has been observed in literature for reduction of gene flow inside pericentric inversions in *D. melanogaster* (Coyne *et al*., 1991). Nevertheless, the reduction in recombination I measured is not as high as in natural populations, which is coherent, since Inv1 and Inv2 have not suffered any evolution and to accumulate possible epistatic interactions that would further decrease gene flow.

I next investigated how breakpoint loci segregate in heterozygotic crosses to test if there is linkage between them. In order to address this point, I quantified the frequency of the segregation of the loci involved in the rearrangement both for P x P and for P x R crosses by scoring the number

of progeny containing either the two parental or rearranged breakpoint sequences or a combination of both. In P x P, as in R x R crosses, there should be an independent segregation of these loci as the homologous chromosomes can pair to form a normal bivalents in metaphase I. This would not the case for CRs where structural constraints that occur during Prophase I and Metaphase I can cause random segregation of chromosomes and therefore of breakpoints which leads to production of unbalanced gametes. In heterozygotic crosses involving any of the ten different rearrangements, the only surviving progeny was that where the breakpoint structure was both either from the parental or from the rearrangement (Figure 3.6b and 3.6c). This means that no viable recombinants between the breakpoints were recovered. Recombination between breakpoints leads to the production of DNA unbalance in the meiotic products and consequently lethality. Due to this, in heterozygosity, CRs prevent recombination between the breakpoints loci and the ones in its proximity. Consequently, CRs can lock loci that are far apart.

Altogether, these results suggest that CRs affect meiosis differently at different stages depending on the type of rearrangement.

3.4.3 Chromosome structure is a selectable trait.

CRs can arise as a selected macro-mutation in yeast adapting to specific environments in such as glucose-limited chemostats (Dunham *et al*., 2002). Previous studies have pointed out the relevance of several types of CRs for adaptation in resistance to starvation (Coyle and Kroll, 2008), antifungal drugs (Selmecki *et al.*, 2009) and to sulfites used in wine production (Pérez-Ortín *et al*., 2002).

I have quantified mitotic fitness of CRs in genetically engineered yeast strains containing either an inversion or a translocation. In order to do so, I designed a competitive fitness assay based on previous assays used

for *S. cerevisiae* (Desai *et al*., 2007). In this assay, a 1:1 population is prepared containing the rearranged and the GFP-labeled isogenic parental and allowed to grow for 24 hours in rich media (YES), at which time the population is diluted and initiates a new growth cycle. This dilution process is repeated three times allowing the cells to undergo approximately 20 generations. The same experiment is done with the unlabelled isogenic parental strain and the above GFP-labelled counterpart in order to calculate the weight of the GFP marker. As it is shown in Figure 3.7, inversions and translocation have variation for fitness *per se* without any previous adaptation. Thus, CRs have intrinsic phenotypic diversity with neutral, deleterious and beneficial effects.

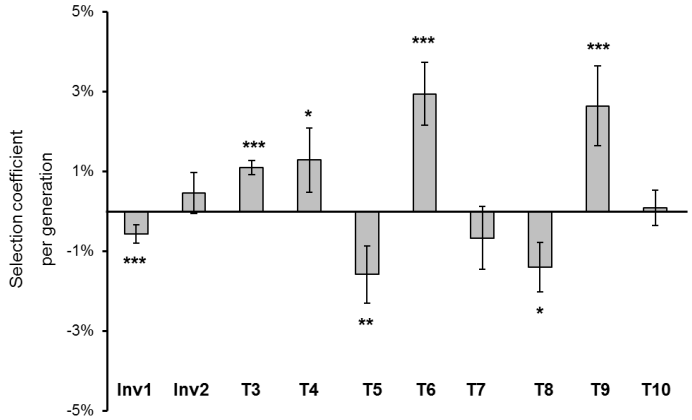


Figure 3.7. Chromosome structure is a selectable trait. Selective coefficients in YES glucose for strains with CRs. Selection coefficients are relative to parental strain. Averages were calculated from more than 10 independent experiments each with two independent clones. Error bars represent 2 x S.E.. (*** for p< 0.001, ** for p< 0.01 and * for p< 0.05 according to Mann-Whitney U-test).

These phenotypic differences could arise as a result of changes in gene expression caused by the modifications in gene order. Specifically, the genetic engineering of these strains caused no disruption in any gene,

and so they are isogenic with their parental except for the rearrangement. If there is a change in gene expression, I would expect to detect it near the breakpoints. The sites of recombination were introduced using an antibiotic resistance cassette (*Kan*^{r}) that confers resistance to *geneticin* and can be used to assay for gene expression (Figure 3.8a and 3.8b).

To test this hypothesis, I performed a functional spot assay where serial dilutions of exponentially growing cells are spotted into increasing amounts of *geneticin* (Figure 3.8b). I observed that rearrangements respond differently to increasing amounts of *geneticin*, suggesting that there is differential expression near the breakpoints. Increased sensitivity to *geneticin* points to a down-regulation of the *kanMX6*$^+$ resistance gene. Even when there is lethality with 400mg/l of *geneticin,* the sensitivity of the different rearrangements does not seem to be the same. This alteration appears to be specific to the combination of chromosomal arms that are displaced. In the case of T5 and T6, T7 and T8, one of the two arms that participate in the translocation is the same but moving to different locations, giving rise to different phenotype. Conversely, Inv2 and T5, T6 and T8, T7 and T9 share the same phenotype and in this case different arms are moved into a common location. In Inv2, T3, T5, T7 and T9, the inverted or translocated arms, are moved into either the *lys3*$^+$ or *lys4*$^+$ breakpoint region. The parental and translocation 9 involve both of these loci and consistently they have the more severe phenotype. This seems to suggest that the region between *lys3*$^+$ or *lys4*$^+$ and the centromere will more likely negatively affect gene expression.

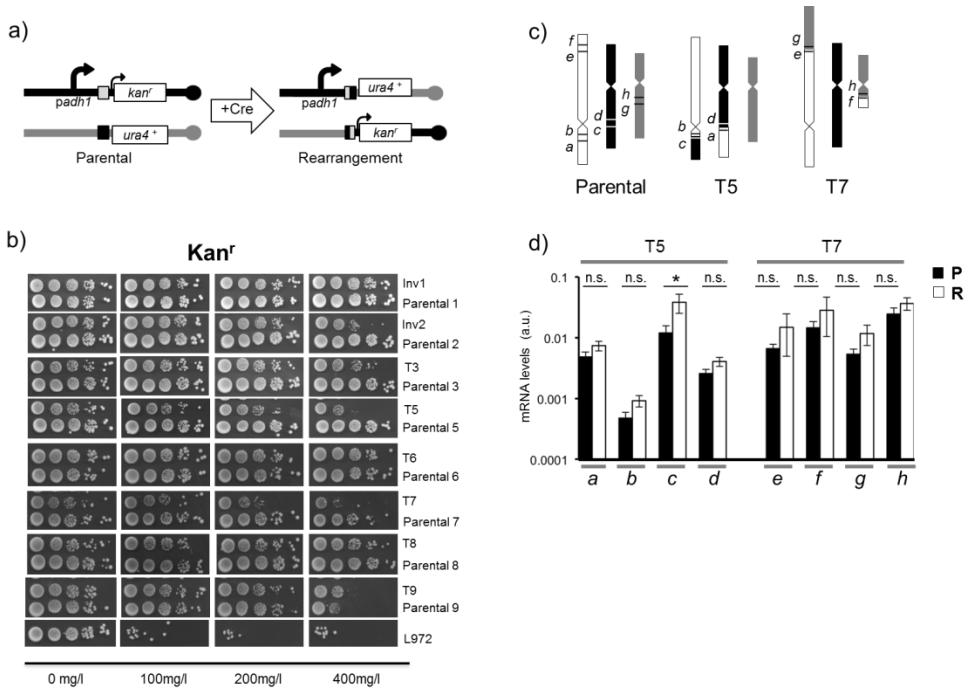


Figure 3.8. CRs change local gene expression levels. a, Cre/loxP recombination assay. Recombination between the promoter construct (padh1-loxP-kanMX6) and the gene construct (loxP-ura4-kanMX6) results in the formation of the rearrangement and *ura4*$^{+}$ expression. Black and grey boxes represent loxP sequences, the ovals indicate centromeres. **b**, Gene expression in breakpoints assayed through *geneticin* resistance. **c**, Schematic representation of genes near the breakpoints in the parental strain and their localization in T5 and T7. **d**, mRNA levels relative to controls near the breakpoints. mRNA was extracted from exponentially growing cultures in YES glucose. Letters correspond to the following genes: a- *rec10*$^{+}$, b-*cut7*$^{+}$, c-*rrp12*$^{+}$, d-*mrpl16*$^{+}$, e-*SPAC2F7.02c*$^{+}$, f-*SPAC227.17C*$^{+}$, g-*bit6*$^{+}$, h-*SPCC777.12c*$^{+}$. Error bars represent S.E. for n between 3 to 6 independent experiments (* for p< 0.05 according to Mann-Whitney U-test).

To directly evaluate differential gene expression near the breakpoints, I quantified mRNA levels of genes surrounding the breakpoint in translocations T5 and T7 (Figure 3.8c and 3.8d). These two rearrangements were chosen randomly. I found a statistically significant gene expression for *rrp12*$^{+}$ near the breakpoint in translocation T5. This

gene moves from the *lys4*$^{+}$ to *his1*$^{+}$ loci, which if the hypothesis above is correct, I would expect it to be up-regulated in T5 which indeed is (Figure 3.8d). Although I cannot make a clear correlation with the arms that are moving and the phenotype, my results show that chromosome structure is a selectable trait *per se* with changes at the gene expression level.

3.4.4 Chromosome structure is an environmentally-dependent selectable trait with changes in gene expression.

Fitness is an environmentally dependent parameter and therefore I tested whether the selective coefficient of the engineered CRs will change with environmental variations. In order to test it, I performed the same type of competitive assay as described above using ten different environmental conditions. These conditions were established in a wild-type prototrophic strain so that growth would be suboptimal (Figure 3.9). Suboptimal conditions were used as a way to mimic environmental variations that can occurs in nature and at the same time to potentiate possible phenotypic variation among the different genotypes. The drugs were used to target specific cellular processes. A wild-type prototrophic strain was used only for simplicity and to establish uniform conditions independent of the auxotrophic markers.

I observed that all the conditions tested alter some phase of the growth in relation to the optimum (Figure 3.9). The only condition that has not caused growth-rate differences in the wild-type prototrophic strain was the temperature of 36ºC. However, this temperature was still tested as the maximum temperature where fission yeast can live.

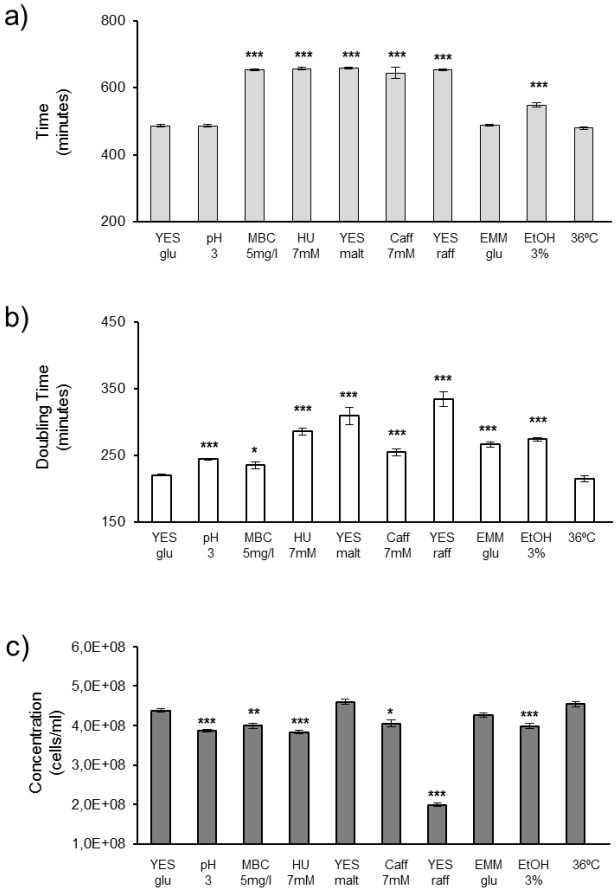


Figure 3.9. Mitotic growth of wild-type L972 *S. pombe* in different environments. **a**,**b**,**c**, measurements are for lag (**a**) and exponential (**b**) phases and carrying capacity (**c**). Statistical comparisons were all made against growth in YES gluc. Error bars represent 2 x S.E. for n=8 independent experiments (* for p< 0.05, ** for p< 0.01 and *** for p< 0.001 according to Mann-Whitney U-test).

CRs have environmentally-dependent fitness (Figure 3.10, Appendix F). Moreover, the same rearrangement can be either advantageous or disadvantageous with changing conditions; this is the case for inversion 1, translocations T4 and T5 that change signal depending on the media were they are growing. The scale defined for the heat map considers values from

-2.5% to +2.5% indistinguishable which are therefore considered neutral. Even though with this scale there is already a considerable phenotypic variability, for more detailed analysis of fitness values is represented in graphic form in Appendix F.

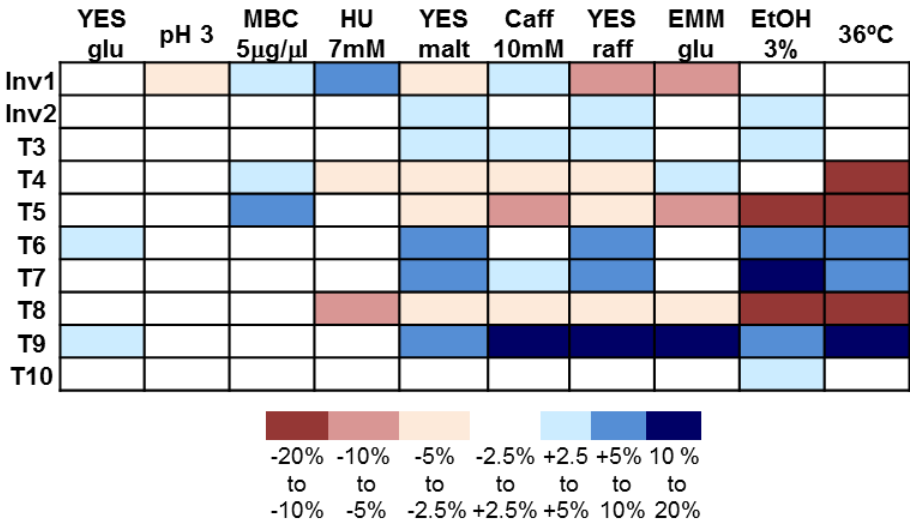


Figure. 3.10. Mitotic fitness of CRs in different environments. Selective coefficients for strains with CRs in different growth conditions. Selective coefficients are relative to parental strain. Coloured boxes represent the average of n>20 independent experiments using two independent clones. Environments were ordered by increased variance in fitness.

As it was shown before, CRs cause changes in gene expression near the breakpoints in rich media. I hypothesized that gene expression changes could be occurring genome-wide due to the variance of fitness of each rearrangement in the different media tested. I wanted to test if CRs can change gene expression in locations far away from their breakpoint. To assess this I chose two CR-containing strains based on their opposite response to 3% EtOH, an environment that occurs naturally for fission yeast (Gomes *et al*., 2002). I measured mRNA levels for genes involved in stress and ethanol responses (Alexandre *et al*., 2001; Chen *et al*.,

2003)(Figure 3.11b and Table 3.5). Genes statistically different from the parental are represented in Figure 3.11b as well as a sample representative of different sites of the chromosomes. I have analysed independent exponentially growing populations with no exposure (T_0) and with 15 or 30 minutes (T_{15} and T_{30}) exposure to 3% EtOH.

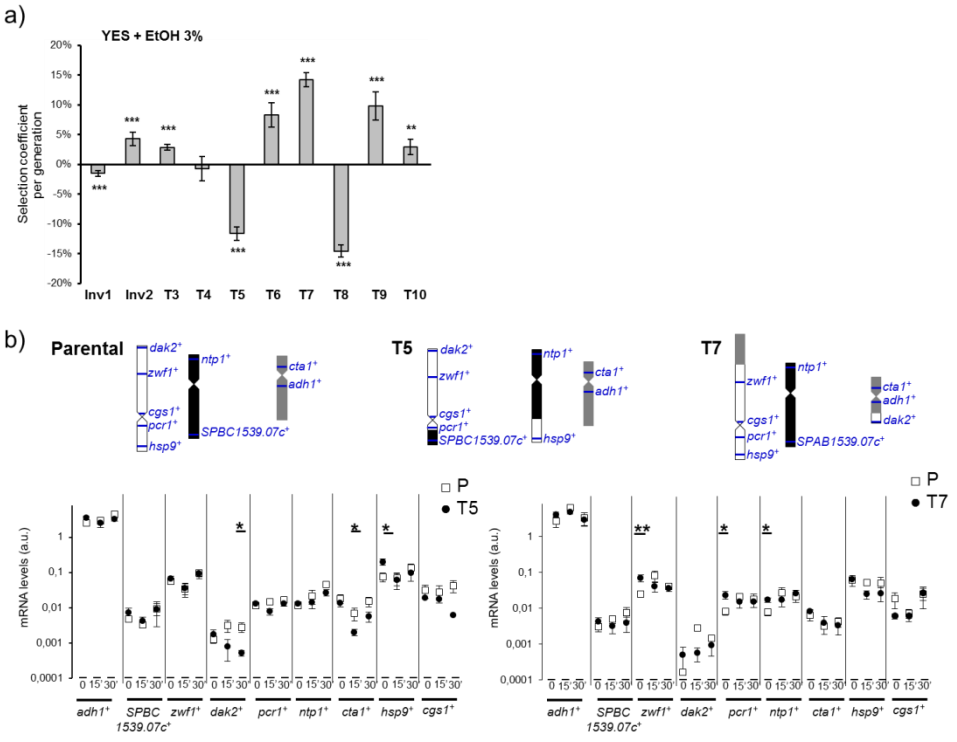


Figure. 3.11. Rearrangements cause changes in gene expression.. a, Mitotic fitness in rich media containing 3% ethanol. Error bars represent 2x(S.E.. (*** for p< 0.001 according to Mann-Whitney U-test). **b**, mRNA levels for genes involved in stress and ethanol consumption. mRNA was extracted from exponentially growing cultures without (T_0) and with 15 and 30 minutes (T_{15} and T_{30}) exposure to 3%EtOH. Error bars represent S.E. for n between 3 to 6 independent experiments. (* for p< 0.05 according to Mann-Whitney U-test).

Analysis of 21 genes, involved either in stress or metabolic responses, revealed that some of these genes already have differential expression in translocations (Figure 3.11b, Appendix G). Interestingly, *hsp9*$^{+}$ is up-regulated in translocation T5 in rich media. The *hsp9*$^{+}$ gene is located in the translocated arm of chromosome I in T5 but very far apart from the breakpoint. The same type of phenomenon is observed in translocation T7 where genes *zwf1*$^{+}$, *pcr1*$^{+}$ and *ntp1*$^{+}$ are also differentially expressed in the translocation in absence of the EtOH challenge. *ntp1*$^{+}$ is on chromosome II which is a chromosome not involved in the formation of this translocation. Thus, CRs alter the expression of genes in regions not involved in the rearrangement even in the absence of any environmental challenge.

In the presence of 3% EtOH, *adh4*$^{+}$ and *SPAC13A11.06*$^{+}$ respond significantly different in the translocation 7 (Appendix G). These genes are located near the extremity of chromosome I, in the arm involved in the translocation. *dak2*$^{+}$ is also an interesting example. It responds differently in T5 and it is located in the opposite chromosomal site of the breakpoint. This data shows that chromosome structure can alter gene expression of genes located far away from the breakpoint.

As far as metabolic and stress responses in the presence of 3% EtOH, there is a slight alteration of the metabolic and signalling response. However, I did not detect a strong environmental stress response (ESR), defined by a set of genes that is induced or repressed at least twofold upon environmental stress (Chen *et al*., 2003). There has been no major increase in ESR-responsive genes which points out to the fact that 3% EtOH is sufficient to disturb growth but not to generate a stress response. Another gene whose levels have not been coherent with a stress response is *hsp9*$^{+}$. Heat shock proteins increase upon several stresses studied so far (Chen *et al*., 2003). Hsp9 levels are maintained throughout time, except in T5, which decrease as this rearrangements causes and up-regulation of

hsp9^{+}*.* The lower levels of two important proteins, catalase and dihydroxyacetone kinase (encoded by *cta1*$^{+}$ and *dak2*$^{+}$, respectively), could perhaps be important to explain the deleterious phenotype of T5 in media containing 3% EtOH.

3.5 Discussion and Conclusions.

3.5.1 Chromosomal Rearrangements as vehicles for incipient genetic separation.

CRs have for long interested evolutionary biologists for their potential in creating a genetic barrier to the fixation of genetic differences between populations. In order to quantify the immediate drop in meiotic viability I have constructed ten different strains of *S. pombe* each containing a different balanced CR. Except of translocation T4, all rearrangements do not cause meiotic problems in homozygosity as seen by the high viability of intraspecific crosses for the different rearrangements.

In heterozygosity rearrangements cause drops in meiotic viability up to 50% with exception of translocation T8 where there is virtually no effect. Possibly, despite the chromosomal rearrangement, the tetravalent structure facilitates the resolution of the structure in such a way that alternate-segregation is the predominant segregation mode. In addition, I have not seen any segregation distortion where heterokaryotypes would preferentially produce gametes carrying either the wild-type or the inversion. Nonetheless, the different rearrangements have slightly different costs in meiosis. Different losses in viability have also been observed in previous work using *S. cerevisiae* from 50% to 80% in heterozygotic crosses of reciprocal translocations caused by non-allelic recombination (Loidl *et al*., 1998). Also, Delneri and colleagues have measured the contribution of two specific reciprocal translocations for the speciation between *S. cerevisiae* and *S. mykatae* (Delneri *et al*., 2003). By

engineering the genome of wild-type *S. cerevisiae*, the authors observed an improvement in the viability of the cross between *S. mikatae* and the engineered *S. cerevisiae*, named Sct1/2. As a control, heterozygotic crosses of wild-type *S. cerevisiae* and a Sct1 strain, containing a single translocation between arm VI and VII or Sct2 with a single translocation between chromosome VI and XVI, led to viabilities in the order of 60 and to 50%, respectively. Therefore, even in this other model organism, the different location of the rearrangement, albeit of the same type, can lead to differences in viability.

The different levels of viability among the different rearrangements can reflect problems at either one or both major steps of meiosis: homologous recombination and segregation of homologous chromosomes which are both fundamental for the production of viable offspring at meiosis.

In the case of translocations, the proper resolution of the quadrivalents formed at metaphase I is important for the formation of dyads with balanced chromosomes and after four viable meiotic products (see General Introduction of this thesis for resolution of quadrivalents). The metaphase I orientation of the centromeres of the quadrivalent formed in heterozygotes for a reciprocal translocation depends on the size of the interchange, the nature of the chromosomes involved and the nature and location of the chiasma (Loidl *et al*., 1998; Davis and Smith, 2003). The differences in the viabilities of the meiosis for the different translocation might come from the way quadrivalents are resolved at anaphase of meiosis I. In order to properly evaluate the type of segregation that is predominant in each translocation, a representation of the distribution of the different tetrads formed prior to germination would be necessary. Nonetheless, as I have mainly tested meiotic viability of four-spore' tetrads and analysed the distribution of their viable offspring this reflects not only

the disjunction at anaphase I but the survival of spores after the entire meiotic process. However, as tetrads were chosen randomly, some contained spores with unusual size and/or morphology (data not shown) which is consistent with the occurrence of adjacent disjunction of the centromeres of the quadrivalents at anaphase I leading to unbalanced dyads and consequently to four dead spores.

I have not detected the simultaneous presence of Parental and CRs breakpoints at chromosomes contrary to what has been observed in meiotic segregation of crosses between *S. cerevisiae* with translocation and wild-type (Loidl *et al*., 1998). Neither PCR amplification of the different cassettes neither selection on plate identified the presence of disomic colonies (containing extra chromosomes) in cases where I retrieved only two viable spores. Also, I have not identified any non-disjunction cells or cells where anaphase I had led to the production of imbalanced genetic information. In *S. cerevisiae*, asci with two viable spores were frequently produced by translocation heterozygotes; both 3:1 segregation and the occurrence of an uneven number of cross-over in the interstitial region between the centromeres and the translocation breakpoint of the quadrivalent lead to 50% viability of the products of a meiotic division (Loidl *et al*., 1998). I have only observed the formation of offspring containing either the parental or the rearranged breakpoints, consistent with the formation of balanced gametes. This might reflect the difference between *S. cerevisiae* and *S. pombe* as the first one is diploid and the second is haploid; also *S. pombe* is lethal for any disomy except for chromosome III (Niwa and Yanagida, 1985). Nonetheless, homologous recombination is intact in these rearrangements for the contiguous regions, thus chiasma is occurring in these meiosis. However, recombinants for the breakpoints are lethal as they generate meiotic products with imbalanced DNA content which in *S. pombe* is lethal. Therefore, CRs lock the loci that in close

proximity with the rearrangement breakpoint. This is particularly interesting in the case of translocations, where the loci locked are in different chromosomes.

The drop in homologous recombination is, as expected, high for inversions. Davis and Smith found that, in the absence of homologous recombination, fission yeast is still able to segregate faithfully homologous chromosomes during meiosis, at MI stage, at a rate higher than it would be expected by chance (Davis and Smith, 2003). Thus, even if there is reduced recombination in inversions, *S. pombe* segregation is still more faithful than expected by chance suggesting there is some mechanism that assures maximum genome stability. Therefore, in the strains of *S. pombe* analysed in this work, the main type of segregation might be alternate centromere segregation. Adjacent type I segregation might also occur giving rise to the tetrads with no viable spores.

As it can be observed, T9 seems to cause a slight increase in homologous recombination between the *lys4*$^{+}$ and the *mat1*$^{+}$ loci. Interestingly, as *S. pombe* does not have crossover interference (Munz, 1994) and therefore the number and location of crossing-overs is random, perhaps this chromosomal configuration favours the increase of DSB hotspots that start the recombination process. Cross overs may be adjusted to compensate lethality and allow favourable crossing-overs inside the inversion.

The results here presented suggest that chromosomal rearrangements affect meiosis at different stages depending on the type of rearrangement. In the long-term, I would expect mutations to accumulate differentially in an inverted and collinear genome backgrounds in the regions close to the rearrangement where there is suppression of recombination. Perhaps this initial decrease in recombination could be extended to the entire length of the chromosome, as it has been measured

in *D. melanogaster* populations (Coyne *et al.*, 1991). Would this be true, CRs could be strongly implicated as an incipient genetic mechanism for the speciation process, as proposed initially by White (White, 1977) and then extended by others (Noor *et al*., 2001; Rieseberg, 2001; Navarro and Barton, 2003; Kirkpatrick and Barton, 2006). Translocations link loci in different chromosomes whereas inversions link loci in the same chromosome near the breakpoints.

These strains can provide valuable information in studies of meiosis namely to better map and understand meiotic pairing at MI stage and the importance of chromosome configuration for segregation. Furthermore, it can be a source for studies of hotspots for crossing-over. The strains containing inversions can also be useful for studies of inter and intrachromosomal effects and their contribution to nucleotide variation (Stevison *et al*., 2011). Interestingly, I could imagine a scenario like the one proposed by Ding and colleagues where "particular molecular patterns along each chromosome provide a chromosomal barcode for the recognition of homologous chromosomes" (Ding *et al*., 2010).

3.5.2 Chromosomal Rearrangements are a selectable trait with environment dependent fitness.

I have studied fitness effects of eight different translocations and two inversions in *S. pombe* and their environmental variation. My results demonstrate that chromosome structure is a selectable trait even on optimal media. I have observed both advantageous and disadvantageous effects for translocations and inversions in different environments.

Several CRs occur as a consequence of adaptation in different environmental conditions (Adams *et al.*, 1992; Gresham *et al.*, 2008) but most of these include duplication of genes important for glucose uptake (Brown *et al*., 1998) or alteration of specific promoter leading to

overexpression of metabolic genes (Dunham *et al*., 2002). To my knowledge, the most directed study to date to measure fitness of CRs has been made by Colson and colleagues (Colson *et al*., 2004). In this study the authors measured the adaptation of a translocation that is occurring in *S. mikatae* which differs from *S. cerevisiae* by the presence of two reciprocal translocations. Despite the importance of this work, this rearrangement has been selected by natural selection throughout the evolutionary period that separates the two species. Thus, a major difference in my work is that I have generated random balanced CRs which do not cause amplification or deletion of any gene neither change promoter sequences.

In the case of aneuploidies, gene dosage is often correlated with improved expression and resistance. For instance, Pavelka and colleagues measured increased resistance to 4-NQ to be correlated with an aneuploidy of the chromosome which contains the resistance gene; deletion of this gene in the aneuploid strain restored the sensitivity to wild-type levels (Pavelka *et al*., 2010). In the case of inversions or translocation there is no gene dosage alteration so I cannot attribute the increase in gene expression to unbalanced gene content. Therefore, the phenotype observed in my strains comes mainly from repositioning of chromosome domains and reordering of sequences that can impact gene expression and/or DNA replication.

DNA replication, gene expression or DNA repair and recombination are genomic processes that require the coordinated temporal and spatial action of multiple players. If mitotic effects of CRs are only structural, I would expect a constant sign for the selection coefficient across the different environments. This effect would be exacerbated in drugs targeted for cell cycle phases such as HU (which works in S-phase) and MBC (which works in M-phase). On the other hand, I would expect a variation in

fitness across the different environments when the main alteration caused by the CRs is gene expression. This was indeed what I observed; I found changes in gene expression located several kilobases away from the breakpoints, on opposite arms of the chromosome or on different chromosomes altogether. I choose two CRs causing opposite selection coefficients across the different environments in order to quantify the extent of alteration of gene expression caused by the rearrangements. Even in the absence of any environmental challenge, *hsp9*$^{+}$ is up-regulated in translocation T5 and *zwf1*$^{+}$, *pcr1*$^{+}$ and *ntp1*$^{+}$ in translocation T7. *ntp1*$^{+}$ is on chromosome II which is a chromosome not involved in the formation of this translocation. These results highlight the importance of gene positioning and the fact that genomes are not randomly organized within the cell nucleus (Misteli, 2007). Moreover, *dak2*$^{+}$ and *hsp9*$^{+}$ that respond differently in T5 and *adh4*$^{+}$ and *SPAC13A11.06*$^{+}$ in T7, are located in chromosomes involved in the translocations but both very far from the breakpoints. Another extreme example if the response in *cta1*$^{+}$ expression in T5; *cta1* is located in chromosome III that does not participate in the formation of T5.

I have not observed a global stress response suggesting that CRs do not induce known gene stress pathways. It has been previously observed that a common gene expression stress response are not an obligatory property of aneuploid strains (Pavelka *et al*., 2010). I have not observed an ESR to 3% ethanol suggesting that the amount of ethanol used is sufficient to disturb growth and cause changes in gene expression but does not lead to a robust ESR. Some genes such as *ntp1*$^{+}$, *tps1*$^{+}$ and *pcr1*$^{+}$ increase upon EtOH exposure as expected but others do not change significantly such as *cgs1*$^{+}$ and *pka1*$^{+}$, which increase dramatically in stress responses as part of the cAMP signalling. For instance, *pyr1*$^{+}$ changes little in ESR but increases at least two-fold in the presence of ethanol in *S. cerevisiae* (Alexandre *et*

al., 2001). Instead, I do not see any change for *pyr1*$^{+}$ in the strains here tested.

Previous studies have tried to obtain a correlation between breakpoints and gene expression levels by looking at genes near breakpoints that distinguish humans and chimpanzees (Marquès-Bonet *et al*., 2004; Muñoz and Sankoff, 2012). These studies are useful to understand gene expression differences between two karyotipically similar species. However, it is not trivial to extract absolute information on the alteration of CRs for gene expression because there was no isogenicity on the background. A probably fairer comparison has been made on gene expression from cell lines from human individuals with the reciprocal translocation t(11;22)(q23;q11)(Harewood *et al*., 2010). This translocation is the most common constitutional reciprocal translocation in man and balanced carriers are phenotypically normal except for decreased fertility, an increased spontaneous abortion rate and a possible predisposition to breast cancer in some families. The authors have measured differential gene expression in chromosome 11 and distributed throughout the genome and associated these facts with the movement of the chromosome to a more central position of the nucleus. Like the previous study, I was also able to measure differences in gene expression in chromosomes that were also not involved in the translocation, suggesting that CRs affect *trans* regulation and are therefore not restricted to the breakpoint area.

There are several examples of gene regulation by chromosome interactions. The importance of structure in nuclear organization and gene expression is highlighted in works such as the one by Spilianakis and colleagues. In naïve T cells, the TH2 regulatory locus on chromosome 11 physically interacts with the *Ifng* locus on chromosome 10 (Spilianakis *et al.*, 2005). Upon stimulation of naïve T cells, these two locus physically separate and *Ifng* transcription initiates. Another example comes from

studies in sensory neurons where the choice for expression of a particular odorant receptor results from the interaction of a single *trans*-acting enhancer element in chromosome 14 with the selected receptor in another chromosome (Lomvardas *et al*., 2006). This study shows that an interaction between a promoter and a specific enhancer located on different chromosomes can lead to the choice of the receptor to be expressed.

In *S. pombe* chromosomes often tend to have specific interactions (Molnar and Kleckner, 2008) which could be disrupted by CRs and cause the observed changes in phenotype. Although I have not characterized the nuclear organization of the rearranged chromosomes, I can imagine that they are disrupting the normal nuclear positioning of several genes or modules of genes. This could be the mechanism by which CRs lead to different gene expression patterns and consequently phenotypes and establish a link for their adaptive behaviour. It will be interesting to describe whether CRs disrupt interactions such as the ones described and to asses the phenotypic impact of these alterations in other organisms. This way, it could be answered how CRs affect *trans* regulations.

I cannot exclude that translocations and inversions are affecting other processes than gene expression. For instance, in budding yeast, chromosome length is known to influence mitotic segregation and cell division (Murray *et al*., 1986). Chromosomal size can also influence topological stress induced by DNA replication (Kegel *et al*., 2011). Thus, chromosome stability and, consequently, fitness is influenced by chromosome length. In order to test this hypothesis, I consider the strains containing the largest (6.96 Mb) and the smallest (1.93 Mb) chromosomes generated by this work (T7 and T8 respectively, Figure 3.2a and Table 3.2). If both topological stress and segregation stability were translated into slower growth, I would expect that strains carrying longest chromosomes to have deleterious phenotypes even in rich media. However, T7 and T8 have

the opposite results for fitness in most tested conditions (Figure 3.10 and Appendix F), arguing that chromosome size is not the major factor behind my observations. In addition, translocations where chromosome size is not significantly altered, such as T5, exhibit some of the most deleterious fitness effects (Figure 3.10 and Appendix F).

In conclusion, using *S. pombe* as a model system, I show that variation in genome structure has fitness consequences as important as typical nucleotide changes. These fitness consequences are environmentally dependent with changes in gene expression throughout the genome. This work suggests that certain CRs may be maintained as polymorphisms in nature despite their low meiotic viability in heterozygotic crosses. Disadvantages in sexual cycle can be overcome by CR-dependent advantageous effects during asexual growth. Likewise, disadvantages in one environment can be balanced by advantages in another. CRs can be fundamental for the generation of biological diversity at the karyotype level and may constitute a building block for further differentiation. Chromosome structure is highly dynamic and has a complex organization that clearly does not dependent only on the linear sequence.

3.6 Acknowledgements. I thank M Molnar and N Kleckner (U. Harvard) for generously providing the Cre-loxP strains, J Mata (U. Cambridge) for technical assistance with RNA purification and gene expression analysis, P Duarte and S Rosa (IGC) for technical assistance, M Carneiro (IGC) for assistance with qRT-PCR, R Gardner and T Lopes (IGC) for technical support with flow cytometry. This work was supported by the Portuguese Fundação para a Ciência e a Tecnologia (FCT) PTDC/BIA-EVF/114460/2009. A.T.A. was funded by an FCT PhD scholarship SFRH/BD/33214/2007.

3.7 Tables.

Table 3.1. Strains used in this study.

	Genotype	Common name	Creator
L972	*Wild-type Schizosaccharomyces pombe*		
MGF 527	*h- leu1-32 ura4-D18 ade6-M210*	NK 118	N. Kleckner
MGF 2097	*h- arg1::padh1-loxP-kanMX6 leu1-32 ura4-D18 ade6-M 216*	NK 5AU/2	N. Kleckner
MGF 529	*h- lys3::padh1-loxP-kanMX6 his1::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 1A2U (111)	N. Kleckner
MGF 530	*h- arg7::padh1-loxP-kanMX6 lys4::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M216*	NK 3A4U (113)	N. Kleckner
MGF 533	*h- lys3::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 1U/2 (134)	N. Kleckner
MGF 534	*h- lys3::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M216*	NK 1U/2 (146)	N. Kleckner
MHG 535	*h– his1::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 2U/1 (140)	N. Kleckner
MGF 537	*h- arg7::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 3U/1 (141)	N. Kleckner

MGF 538	*h- arg7::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M216*	NK 3U/2 (125)	N. Kleckner
MGF 539	*h- his1::padh1-loxP-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 2A/1 (122)	N. Kleckner
MGF 540	*h- his1::padh1-loxP-kanMX6 leu1-32 ura4-D18 ade6-M216*	NK 2 A/2 (158)	N. Kleckner
MGF 541	*h- lys4::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 4U/1 (147)	N. Kleckner
MGF 542	*h- lys4::loxP-ura4-kanMX6 leu1-32 ura4-D18 ade6-M216*	NK 4U/2 (148)	N. Kleckner
MGF 544	*h- arg7::padh1-loxP-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 3A/1 (162)	N. Kleckner
MGF 546	*h- lys4::padh1-loxP-kanMX6 leu1-32 ura4-D18 ade6-M210*	NK 4A/1 (161)	N. Kleckner
MGF 2079	*h- arg1::padh1-loxP-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M 216*	NK 5AU/2-M	This study
MGF 2080	*h+ arg1::padh1-loxP-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M 216*	NK 5AU/2-P	This study
MGF 2081	*h- arg7::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 141-M	This study

MGF 2082	*h+ arg7::loxP-ura4-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 141-P	This study
MGF 2083	*h- lys3::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 134-M	This study
MGF 2084	*h+ lys3::loxP-ura4-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 134-P	This study
MGF 2085	*h- lys3::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M216*	NK 146-M	This study
MGF 2086	*h+ lys3::loxP-ura4-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M216*	NK 146-P	This study
MGF 2087	*h- his1::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 140-M	This study
MGF 2088	*h- his1::loxP-ura4-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 140-P	This study
MGF 2089	*h- lys4::padh1-loxP-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 161-M	This study
MGF 2090	*h+ lys4::padh1-loxP-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 161-P	This study
MGF 2091	*h-arg7::padh1-loxP-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 162-M	This study

MGF 2092	*h- arg7::padh1-loxP-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 162-P	This study
MGF 2093	*h- his1::padh1-loxP-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M216*	NK 158-M	This study
MGF 2094	*h- his1::padh1-loxP-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M216*	NK 158-P	This study
MGF 2095	*h- lys4::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 147-M	This study
MGF 2096	*h- lys4::loxP-ura4-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210*	NK 147-P	This study
MGF 1497	*h- ade6-M210 mat1-M::mat1-M-natMX6 leu1-32*	NK 118-M	This study
MGF 1498	*h+ ade6-M210 mat1-P::mat1-P-natMX6 leu1-32*	NK 118-P	This study
MGF 1495	*h- his1::padh1-loxP-kanMX6 lys3::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210*	Parental 1	This study
MGF 1496	*h+ his1::padh1-loxP-kanMX6 lys3::loxP-ura4-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210*	Parental 1	This study
MGF 1503	*h- lys3::padh1-loxP- kanMX6R his1::loxP-ura4-kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210*	Control 1	This study

MGF 1521	*h+ his1::padh1-loxP-kanMX6R lys3::loxP-ura4-kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M210 his1-gfp-hphMX6R*	Parental 1- GFP	This study
MGF 1720	*h- lys3::padh1-loxP-kanMX6 his1::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ade6-M210*	Control 1-GFP	This study
MGF 1122	*h- his1::loxP- kanMX6R lys3::padh1-loxP-ura4+ - kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18*	Inv 1	This study
MGF 1123	*h+ his1::loxP- kanMX6R lys3::padh1-loxP-ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18*	Inv 1	This study
MGF 826	*h- arg7::padh1-loxP-kanMX6 lys4:: loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M216*	Parental 2	This study
MGF 827	*h+ arg7::padh1-loxP-kanMX6 lys4:: loxP-ura4-kanMX6 mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M216*	Parental 2	This study
MGF 1563	*h- arg7::padh1-loxP- kanMX6R lys4::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M216 ura4+*	Control 2	This study
MGF 1318	*h- arg7::padh1-loxP- kanMX6R lys4:: loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M216 ura4-D18 lys4-GFP-hphMX6R*	Parental 2- GFP	This study
MGF 1513	*h- arg7::padh1-loxP- kanMX6R lys4::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 lys4-gfp-hphMX6R*	Control 2- GFP	This study
MGF 824/1492	*h- arg7::loxP- kanMX6R lys4:: padh1-loxP-ura4+ - kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M216 ura4-D18*	Inv 2	This study

MGF 825	*h+ arg7::loxP- kanMX6R lys4:: padh1-loxP- ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M216 ura4-D18*	Inv 2	This study
MGF 1744	*h- lys3::loxP-ura4- kanMX6R arg7::padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18*	Parental 3	This study
MGF 1745	*h+ lys3::loxP-ura4- kanMX6R arg7::padh1-loxP- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18*	Parental 3	This study
MGF 1515	*h- lys3::loxP-ura4- kanMX6R arg7:: padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210*	Control 3	This study
MGF 1485	*h- lys3::loxP-ura4- kanMX6R arg7:: padh1-loxP- kanMX6R leu1-32 ade6-M216 ura4-D18 matM- natMX6R lys3-gfp- hphMX6R*	Parental 3- GFP	This study
MGF 1505	*h- lys3::loxP-ura4- kanMX6R arg7:: padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ lys3-gfp- hphMX6R*	Control 3-GFP	This study
MGF 1138	*h+ arg7:: loxP- kanMX6R lys3::padh1-loxP- ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M216 ura4-D18*	T3	This study
MGF 1793	*h- arg7:: loxP- kanMX6R lys3::padh1-loxP- ura4+ - kanMX6R ade6-M210 leu1-32 mat1-M::mat1-M-natMX6 ura4-D18*	T3	This study
MGF 1726	*h- arg1::padh1-loxP- kanMX6R lys4::loxP- ura4- kanMX6R leu1-32 ade6-M210 ura4-D18*	Parental 4	This study
MGF 1727	*h+ arg1::padh1-loxP- kanMX6R lys4::loxP- ura4- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18*	Parental 4	This study

MGF 1728	*h- arg1::padh1-loxP- kanMX6R lys4::loxP-ura4-k kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210*	Control 4	This study
MGF 1581	*h- arg1::padh1-loxP-k kanMX6R lys4::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R lys4-gfp- hphMX6R*	Parental 4- GFP	This study
MGF 1633	*h- arg1::padh1-loxP- kanMX6R lys4::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM- natMX6R lys4-gfp- hphMX6R*	Control 4-GFP	This study
MGF 1099	*h- arg1::loxP- kanMX6R lys4::padh1-loxP-ura4+ - kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	T4	This study
MGF 1783	*h+ arg1::loxP- kanMX6R lys4::padh1-loxP-ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	T4	This study
MGF 1070	*h- his1::padh1-loxP- kanMX6R lys4::loxP-ura4-k kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	Parental 5	This study
MGF 2098	*h+ his1::padh1-loxP- kanMX6R lys4::loxP-ura4-k kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	Parental 5	This study
MF 1279	*h- his1::loxP- kanMX6R lys4::padh1-loxP-ura4-kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM*	Control 5	This study
MGF 1489	*h- his1::padh1-loxP- kanMX6R lys4::loxP-ura4-kanMX6R leu1-32 ade6-M210 ura4-D18 matM- natMX6R lys4-gfp-hphMX6R*	Parental 5- GFP	This study
MGF 1582	*h-his1::padh1-loxP- kanMX6R lys4::loxP-ura4-kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM- natMX6R lys4-gfp-hphMX6R #5*	Control 5-GFP	This study

MGF 1089	*h- his1::loxP- kanMX6R lys4::padh1-loxP-ura4+ - kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM*	T5	This study
MGF 1106	*h+ his1::loxP- kanMX6R lys4::padh1-loxP-ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18*	T5	This study
MGF 1729	*h- arg7::loxP-ura4- kanMX6R his1::padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ura4-D18 ade6-M210 matM- natMX6R*	Parental 6	This study
MGF 1132	*h+ arg7::loxP-ura4- kanMX6R his1::padh1-loxP- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ura4-D18 ade6-M216 matM- natMX6R*	Parental 6	This study
MGF 1561	*h- arg7::loxP-ura4- kanMX6R his1::padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM- natMX6R*	Control 6	This study
MGF 1551	*h+ arg7::loxP-ura4- kanMX6R his1::padh1-loxP- kanMX6R leu1-32 ura4-D18 ade6-M216 matM- natMX6R his1-gfp- hphMX6R*	Parental 6- GFP	This study
MGF 1574	*h- arg7::loxP-ura4- kanMX6R his1::padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM- natMX6R his1-gfp- hphMX6R #7*	Control 6-GFP	This study
MGF 1164/1794	*h- his1::loxP- kanMX6R arg7::padh1-loxP-ura4+ - kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	T6	This study
MGF 1165	*h+ his1::loxP- kanMX6R arg7::padh1-loxP-ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M216 ura4-D18 matP- natMX6R*	T6	This study
MGF 1749	*h- arg1::padh1-loxP- kanMX6R lys3::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	Parental 7	This study

MGF 1072	*h+ arg1::padh1-loxP- kanMX6R lys3::loxP-ura4- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	Parental 7	This study
MGF 1567	*h- arg1::padh1-loxP- kanMX6R lys3::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM-natMX6R*	Control 7	This study
MGF 1522	*h- arg1::padh1-loxP- kanMX6R lys3::loxP-ura4- kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM-natMX6R lys3-gfp- hphMX6R*	Parental 7- GFP	This study
MGF 1736	*h- arg1::padh1-loxP- kanMX6R lys3::loxP-ura4- kanMX6R 6 mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM- natMX6R lys3-gfp- hphMX6R*	Control 7-GFP	This study
MGF 1795	*h- arg1::padh1-loxP- kanMX6R lys3::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM-natMX6R*	T7	This study
MGF 1108	*h+ arg1::padh1-loxP- kanMX6R lys3::loxP-ura4- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4+ matM-natMX6R*	T7	This study
MGF 1069	*h- arg1::padh1-loxP- kanMX6R arg7::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM-natMX6R*	Parental 8	This study
MGF 1068	*h+ arg1::padh1-loxP- kanMX6R arg7::loxP-ura4- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18 matM-natMX6R*	Parental 8	This study
MGF 1565	*h- arg1::padh1-loxP- kanMX6R arg7::loxP-ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM- natMX6R*	Control 8	This study
MGF 1549	*h- arg1::padh1-loxP-kanMX6 arg7::loxP-ura4-kanMX6 mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R arg7-gfp- hphMX6R*	Parental 8- GFP	This study

MGF 1576	*h- arg1::padh1-loxP- kanMX6R arg7::loxP-ura4- kanMX6R leu1-32 ade6-M210 arg7-gfp-hphMX6R*	Control 8-GFP	This study
MGF 1058	*h- arg1:: loxP- kanMX6R arg7:: padh1-loxP-ura4+ - kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18*	T8	This study
MGF 1060	*h+ arg1:: loxP- kanMX6R arg7:: padh1-loxP-ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18*	T8	This study
MGF 1746	*h- lys3::loxP-ura4- kanMX6R lys4:: padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M216 ura4-D18 matM- natMX6R*	Parental 9	This study
MGF 1747	*h+ lys3::loxP-ura4- kanMX6R lys4:: padh1-loxP- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M216 ura4-D18 matP- natMX6R*	Parental 9	This study
MGF 1511	*h+ lys3::loxP-ura4-k kanMX6R lys4:: padh1-loxP- kanMX6R leu1-32 ade6-M216*	Control 9	This study
MGF 1487	*h- lys3::loxP-ura4- kanMX6R lys4:: padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M216 ura4-D18 matP- natMX6R lys4- gfp -hMX6R*	Parental 9- GFP	This study
MGF 1507	*h+ lys3::loxP-ura4- kanMX6R 6 lys4:: padh1-loxP- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M216 lys4-gfp- hphMX6*	Control 9-GFP	This study
MGF 1792	*h- lys4:: loxP- kanMX6R lys3::padh1-loxP-ura4+ - kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M216 ura4-D18*	T9	This study
MGF 1796	*h+ lys4:: loxP- kanMX6R lys3::padh1-loxP-ura4+ - kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M216 ura4-D18*	T9	This study

MGF 1737	*h- arg1::padh1-loxP- kanMX6R his1::loxP- ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	Parental 10	This study
MGF 1739	*h+ arg1::padh1-loxP- kanMX6R his1::loxP- ura4- kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18 matM- natMX6R*	Parental 10	This study
MGF 1751	*h- arg1::padh1-loxP- kanMX6R his1::loxP- ura4- kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM- natMX6R*	Control 10	This study
MGF 1381	*h- arg1::padh1-loxP- kanMX6R his1::loxP- ura4-kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18 his1-gfp- hphMX6R*	Parental 10- GFP	This study
MGF 1740	*h- arg1::padh1-loxP- kanMX6R his1::loxP- ura4-kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4+ matM-natMX6R #3*	Control 10-GFP	This study
MGF 1797	*h- arg1::loxP- KanMX6R his1::padh1-loxP- ura4-kanMX6R mat1-M::mat1-M-natMX6 leu1-32 ade6-M210 ura4-D18*	T10	This study
MGF 1112	*h+ arg1::loxP- KanMX6R his1::padh1-loxP- ura4-kanMX6R mat1-P::mat1-P-natMX6 leu1-32 ade6-M210 ura4-D18*	T10	This study

Table 3.2. Chromosomal sizes of engineered CRs.

Strain	Chr. I (bp)	Chr. II (bp)	Chr. III (bp)
Reference	5600	4500	3500
Inv1	5600	4500	3500
Inv2	5600	4500	3500
T3	5070	4725	3500
T4	5600	5395	2615
T5	5300	4810	3500
T6	4616	5495	3500
T7	6955	4510	2145
T8	5600	6080	1930
T9	6070	4040	3500
T10	6158	4510	2915

Table 3.3. Primers for breakpoint analysis.

Primer name	Sequence 5' → 3'
arg7 -200 forw	ttgagtacttgctatccacg
arg1 -200 forw	tgcaactgaacatacgatgg
his1 -100 forw	gactgctttttcgacattgg
lys3 -200 forw	agtttttggtctccttcgcc
lys4 -200 forw	atacatacgtgcatccttgc
3'utr adh1 f	cttcaatttctctactccgc
kanmx 880r	cgcatcaaccaaaccgttat
ura 671 rev	aatgatgatatcgctaccgc
pfa6a rev	gactcactatagggagaccg

Table 3.4. Primers for integration of GFP and mCherry.

Primer name	Sequence 5' → 3'
arg7 locus integr 80mer forw	gtggtgctaaaagcaaatataaaaaatggtgtatgatattttcct aattaagttactaaaaatctaattgcataaaatcagaattcgagctcgtttaaac
arg7 locus integr 80mer rev	atttttaattatatttcatttccaaacgtgtctctctgcattgttttgcaa aatcatttattcacaacatcgaaaaaatgatattaccctgttatcccta
his1 locus integr 80mer forw	agctgacctgctttaatatttatcgtcagttaaagtgtcgaacgactg caacgaaaactgaattagtaaggaaaaaaagagaattcgagctcgtttaaac
his1 locus integr 80mer rev	tttgcattcgatcttcgaacgttgatgtaaagagaccggtttatcctct aattttaattatatttaaatatataaaagtgatattaccctgttatcccta
lys3 locus integr 80mer forw	atatatttgactaatttaaatatgataatatacacatttctctgcaaacctcc tttatacttttcgccggtcattgattagaattcgagctcgtttaaac
lys3 locus integr 80mer rev	atattggggaaagtttcaagatttatatttaccatataggtaagtggcaa atgcaggatgaaatcctgtttcttaattatatattaccctgttatcccta
lys4 locus integr 80mer forw	tttttattctattaaaataaaatattttaatttgaaaatttatattag cacgctcaaggaattacttccttgctacttcgaattcgagctcgtttaaac
lys4 locus integr 80mer rev.	ggggtttaaattatttttttacaaaaggtaataataggtttgaaa aattatataaaaatacgcattaactttacataaaatatattaccctgttatccc cta
pfa6 rev	gactcactatagggagaccg
his1 11130 forw	agaaaaagccgctttgatcc
his1 12038 rev	attgagagtgccatctgagc
lys4 11366 forw	cctgcttcaagccttattgc

lys4 122929 rev	atcaagaaatcgctcgtgcc
lys3 2204 forw	ttagacggcattttcgatgc
lys3 2863 rev	aagcctttccattatcttgg
arg7 2482 forw	gtaaaatgggttcttcatgc
arg7 3366 rev	ggtatgcatacaatcaacat
t tef forw	tgtcgattcgatactaacgc
t tef forw new version	gcgttagtatcgaatcgaca
hph rev	tgatacacatggggatcagc

Table 3.5. Genes for qRT-PCR.

Gene	Function	Chromosomal location
atf1	transcription factor, Atf-CREB family	II: 1547661..1549361
pka1	cAMP-dependent protein kinase catalytic subunit	II: 396992..398530
hsp9	Heat shock protein 9	I: 5 320 399..5 320 605
tps1	alpha,alpha-trehalose-phosphate synthase [UDP-forming]	I : 3479314..3480855
ntp1	Neutral trehalase	II : 207078..209285
zwf1	Glucose 6-phosphate 1-dehydrogenase involved in the PPP	I: 1 454 911..1 456 529

adh1	alcohol dehydrogenase	III: 1591359..1592411
adh4	Alcohol dehydrogenase 4	I: 156548..157816
pcr1 (= mts2)	transcription factor	I : 4252403..4252918
ctt1 (=cta1)	Catalase	III: 56927..58465
hxk2	hexokinase 2	I: 2665816..2667183
pyr 1	pyruvate carboxylase	II: 2190874..2194431
spac13a11.06	pyruvate decarboxylase	I: 585 597..587 312
spac922.07c	aldehyde dehydrogenase	I: 5487116..5488606
spac9e9.09c	aldehyde dehydrogenase	I: 4453560..4455071
spac22a12.16	ATP-citrate synthase subunit 2	I: 1186314..1187851
cgs1	cAMP-dependent protein kinase regulatory subunit	I: 3643239..3644742
dak1	dihydroxyacetone kinase	I: 1177065..1178807
dak2	dihydroxyacetone kinase	I: 64904..66679
ght5	hexose transporter	III: 212313..213953
spbc1539.07c	glutathione-dependent formaldehyde dehydrogenase	II: 4373675..4375075
bit61	TORC2 subunit	III: 1611296..1612564
spcc777.12c	thioredoxin family protein	III: 1617210..1618046
spac2f7.02c	NLI interacting factor family phosphatase	I: 532432..533409
spac227.17c	conserved protein	I: 529801..530166

rrp12	rRNA processing protein	II: 3502416..3505505
mrpl16	mitochondrial ribosomal protein subunit L16	II: 3508286..3508933
rec10	meiotic recombination protein Rec10	I: 4301402..4303777
cut7	nesin-like protein	I: 4306719..4309976
act1	actin	II: 1476149..1477276

Table 3.6. Primers for qRT-PCR.

Primer name	Sequence 5' → 3'
act1 forw	cggtattgtcaacaactggg
act1rev	agtcatcttctcacggttgg
adh1 forw	actcttcttgcggtaactgc
adh1 rev	tggcaatgcagtagtgttgg
adh4 forw	tcaattggtggtggttctgc
adh4 rev	caagggaagttgaggtttgg
atf1 forw	gggtcgcaacaatttaacgg
atf1rev	ccattctcggcattttgtcc
bit61 forw	taacctggttcccatatccg
bit61 rev	aatagtgcaaagcctcaccc
ctt1 forw	acaaatgaagaagccgctgc
ctt1 rev	ctcgaaagcacttcttcagg
cut7 forw	ttggtcacgaagacgtatgc
cut7 rev	ttaccggttccagtttgtcc
dak1 forw	ctggcaccttggtaatttgc
dak1 rev	tgatacgtcatctgcaacgg
dak2 forw	agcttcattggatcactgcg
dak2 rev	atttgaaggttccaggctcg

ght5 forw	ttctttggtcgtcgtatgcc
ght5 rev	acacaatcataacggcaccg
hsp9 forw	atgtctgatcccgcaagaaagtc
hsp9 rev	cgtaggcaccagtaatggattcc
hxk2 forw	accttcctatgggtttcacc
hxk2 rev	gcaaagtccttgttgatggc
mrpl16 forw	tggggagaatatggtatgcg
mrpl16 rev	tacgcatactggcacattgc
ntp1 forw	gcaaaggattatggacgtcg
ntp1 rev	ttcttcgagtaagggcatcc
pcr1 forw	tcaacatatggcttaccggc
pcr1 rev	tgttgatttggagggagacc
pka1 forw	gcaagtgcccattctgttgg
pka1 rev	gcgaatgtctttgatggctg
pyr1 forw	caaggcttaatgggtcaacc
pyr1 rev	gcagcaggaaggaatttacc
rec10 forw	caatgcggaaactgtcaagc
rec10 rev	cgttctgtaaaatgccctcg
rrp12 forw	cgacgagtttcatcttttgcc
rrp12 rev	caactcaaaagccatagcacgc
spac2f7.02c forw	tggagtcagcaattagtccc
spac2f7.02c rev	ggttctgttggttgttcacc
spac9e9.09c forw	tgttcctgaatgtggcttcc
spac9e9.09c rev	aggccaccttatcaatgtcc
spac13a11.06 forw	gtattccttctgatgctggc
spac13a11.06 rev	acttcttgctcaactgctgg
spac22a12.16 forw	ggtgcctcagttgtttatgc
spac22a12.16 rev	taagtttggccgtcagtagg

spac227.17c forw	cacttttggtcagtatgcgg
spac227.17c rev	ttcgacttggttgacagacacc
spac922.07c forw	gactatgcggtgaaatctgc
spac922.07c rev	cagctaaagtatccgcatgc
spbc1539.07c forw	ggtgtcacaactggatttgg
spbc1539.07c rev	atgattctcgaagcaccagc
spcc777.12c forw	tttctcctcaggttcttccc
spcc777.12c rev	caagtgcctttccaaagacc
tdh1 forw	ttccgttgttgacttgaccg
tdh1 rev	cggtcaagtcaacaacggaa
tps1 forw	tgacctttcaatggttgggc
tps1 rev	ccaaaaagacgggaatagcg
zwf1 forw	caaagagcctattggtacgg
zwf1 rev	ggtttccatggtcaaaaggg

Chapter 4 – Beneficial mutation rate in *S. pombe*.

4.1. Abstract.

Natural selection is one of the driving forces of evolution. Nowadays, the available experimental data is not enough to understand how the distribution of fitness effects changes both with genetic background and environment. In order to address the contribution of genetic background to the distribution of effects of beneficial mutations, I performed an evolutionary experiment and applied a theoretical model designed by Illingworth and Mustonen. I estimated a beneficial mutation rate in two different genetic backgrounds of the fission yeast, *Schizosaccharomyces pombe* of 3.3×10^{-9} beneficial mutations per genome per generation for one background and 4.3×10^{-9} for a rearranged genome background. Strikingly, the data suggests that the distribution of beneficial mutations in these two genetic backgrounds is different. However, the two genomes are not directly comparable since they do not have the same auxotrophic markers. These results are highly encouraging and further work should be done to validate that two genomes coding for the same genetic information but with a different gene order, have different beneficial mutation rate.

4.2. Introduction.

Evolution is a constant dynamic process that involves changes in the frequencies of alleles within populations. These frequency changes are dependent on four major forces: natural selection, mutation, genetic drift and gene flow. Genetic drift is a stochastic process of loss of variability in populations and gene flow is the transfer of genes or alleles between populations and genomes. Different alleles are generated by mutation. Mutations are, therefore, the primary source of variation upon which natural selection and genetic drift will act. A mutation is said to be deleterious or beneficial when it causes a change in a phenotype that affects survival and/or reproduction. In these conditions, the fitness effect is negative or positive, respectively. If on the other hand, a mutation does not alter the phenotype, it is

said to be neutral. The fitness effects of mutations permits to estimate the fate and diversity of a population.

Adaptation, the ability to cope and survive in the surrounding environment, occurs by the fixation of beneficial mutations. Adaptation was long thought to occur by selective sweeps, where the appearance in population of a strong beneficial mutation would replace the previous dominant genotype (Atwood *et al*., 1951). In this model of periodic selection, the clonal population is kept until a new stronger beneficial mutation arises that replaces the previous population. Thus, in classical population genetics, adaptation is considered to be a slow process, punctuated by short periods of fitness increases accompanied with fixation events. Periodic selection was the dominant model until the arrival of the clonal interference by Gerrish and Lenski (Gerrish and Lenski, 1998). In the clonal interference model, mutations occur at a high rate with different fitness effects. Since there is a high input of beneficial mutations, fixation takes longer to occur as all the mutations have to compete with one another until one gets fixed. In this scenario, the stronger the mutation, the more changes it has to survive drift and to fix more rapidly. In the clonal interference model, adaptation is largely dominated by the strong effect of mutations. Clonal interference has been experimental demonstrated by some authors (de Visser and Rozen, 2006; Perfeito *et al*., 2007; Kao and Sherlock, 2008).

A caveat of this model is that it does not take into account the simultaneous appearance of multiple mutations in the same clone. Desai and Murray have approached this problem by designing a model that considers multiple beneficial mutations occurring in the same clone (Desai *et al*., 2007). The authors point out that in large populations where clonal interference is important double mutants will usually appear. Consider the case in which, before a highly beneficial mutation B fixates, a less fit mutation A occurring in a fraction of the population can get another mutation C and that the combined effect of A and C exceeds that of B. In this case mutation A (along with C) can

in fact get fixed. The limitation of this multiple mutations model is that it assumes that all beneficial mutations have the same effects, which is not a biological reality. Despite the good fit of the model to experimental results, Desai and Murray raise the point to the problem that adaptation likely occurs by both clonal interfere among mutations of different effect and competition between multiple mutation clones. Thus, considering each model separately is an incomplete approach to evolution in asexual populations.

A major objective for evolutionary biologists is trying to predict the fitness effects of beneficial mutations and to determine the distribution of beneficial mutations to be able to know, for instance, if mutations of small effect are more common than the ones with large effect (Orr, 2003). Considering any type of distribution for fitness effects of mutations, Gillespie highlighted that, in most cases the wild-type allele will be highly fit and so its fitness will be close the extreme right tale of fitness distributions (Figure 4.1)(Gillespie, 1984).

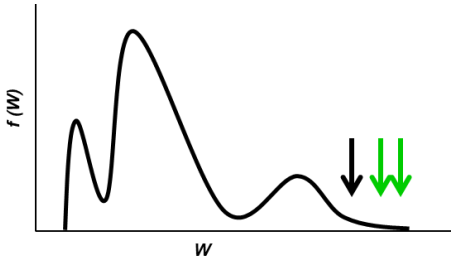


Figure 4.1. **Arbitrary distribution of fitness effects at a random locus.** The real shape of this distribution is, in reality, almost always unknown for any locus. W, fitness value, f(W) frequency of certain fitness value. Black arrow represents the wild-type position and the two green arrows represent beneficial mutations (extreme draws from the distribution). Adapted from (Orr, 2005).

Additionally, a genome that provides fitness greater than the wild-type is a rare event and therefore shall lie on the extreme right tail of distributions of fitness effects of mutations. Gillespie applied extreme value theory (E.V.T., part of statistics that deals with the extreme deviations from the median of probability distributions) to try to postulate a general form for the distribution of beneficial mutations. Gillespie concluded that the tails of all-Gumbel type distributions (a very flexible type of distribution that includes many familiar distributions, including the normal) are exponential (Gillespie, 1984). This theory was later followed by theoretical work developed by Orr (Orr, 2002). Nonetheless, the available data is still controversial on what concerns the distribution of fitness effects. Experimental studies are roughly consistent with the exponential distribution (Imhof and Schlötterer, 2001; Kassen and Bataillon, 2006). However, distributions that do not fit the exponential have also been described (Sanjuán *et al.*, 2004; Rokyta *et al*., 2008; MacLean and Buckling, 2009; Bataillon *et al*., 2011). As a result, as noted by Orr, there is still not enough data to make general empirical conclusions about the distribution of beneficial mutational effects. Interestingly, in the model system *Schizosaccharomyces pombe* the distribution of fitness effects, either neutral, deleterious or beneficial has never been addressed.

Recent application of modelling to experimental evolution experiments has allowed the estimation of parameters such as selection coefficients and the rate of beneficial mutations. One such model was designed by Hegreness and the result being that, independently of the underlying mutational distribution, the distribution of fixed beneficial mutations is approximate to a unimodal shape (Hegreness *et al*., 2006). Hegreness concluded that population evolutionary trajectories can be simplified to two unique parameters: effective selection coefficient and an effective rate of beneficial mutations. However, even if the model allows for the simplification of experimental data, it is important to be aware of the reduction in information. Also, in this model only the first mutational event is considered and so

frequencies at later times are discarded. Contrary to the model of Hegreness, a recent model by Illingworth and Mustonen, allows for the inclusion of different selective coefficient of beneficial mutations at different times and to fit all the experimental data (Illingworth and Mustonen, 2012). This model also estimates the time and the selective coefficient of beneficial mutations from marker frequency in evolving haploid populations independently of unknown distributions.

I have applied the model developed by Illingworth and Mustonen to my own experimental data of evolved population of the fission yeast *S. pombe* to estimate a distribution and mutation rate of beneficial mutations. Microorganisms simplify the study of mutations as they have short generation times and large populations sizes. Also, since microbial populations can undergo hundreds of generations in a relative small amount of time, beneficial mutations are sure to arise. I have used two different genetic backgrounds of *S. pombe* in order to address the question of how the distribution of fitness effects depends on the genetic background specifically changed by a chromosomal rearrangement.

4.3. Material and Methods.

4.3.1. Strains and media. Strains used in this study are listed in the Table 4.2. Strains containing the reference genome (like wild-type fission yeast L972) are designated Control (C) and those containing a pericentric inversion in chromosome I, Inv1 (Figure 4.2). In order to discard the influence of mating type during the experiment I included both h+ and h- mating types in the experimental evolution. Therefore, in the initial experimental setup there is 25% of each of the following strains: C1-GFP-h+, C1-GFP-h-, C1-mCherry-h+ and C1-mCherry-h-. The same rational was used for the inverted strains: Inv1-GFP-h+, Inv1-GFP-h-, Inv1-mCherry-h+ and Inv1-mCherry-h-. The *S. pombe* strains were constructed by the genetically engineering of specific locus as

described in Supplementary Information using standard genetic techniques for fission yeast (Bähler *et al*., 1998). Transformation of yeast cells was performed using the lithium acetate protocol as described elsewhere (Moreno *et al*., 1991). These strains had no previous history in the evolutionary environment used in this study. For details of strain construction see Chapter 3 of this thesis. EMM minimal medium is prepared as described in Chapter 2 of this thesis.

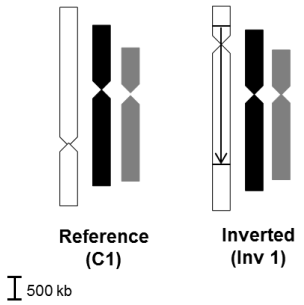


Figure 4.2. Representation of karyotypes used in this work. Inv1 is a pericentric inversion on chromosome I. Both strains are isogenic with the exception of the rearrangement.

4.3.2. Experimental Evolution. Independent cultures of C1-GFP-h+, C1-GFP-h-, C1-mCherry-h+, and C1-mCherry-h- were grown in YES liquid medium for 24 hours at 32ºC with aeration and shacking. The following day cells where mixed 1:1:1:1 ratio and fresh media was added to make this initial cell density around 4 x 10^6 cells/ml. This mixture was then split into 15 populations that were each grown in 600 µl of media in a 96-deep-well plate at 32ºC with constant agitation of 850rpm in a mini-shaker (VWR). All populations were allowed to evolve by asexual propagation for approximately 330 generations. The effective ratio of the mixture was measured by flow-cytometry using a Cyan ADP cell analyser (Beckman Coulter, Inc.). A heterogeneous minimal media containing two different carbon sources was

used in this experiment: EMM minimal media prepared with all supplements, 1% maltose, 1% raffinose and 3% YE. This media contain two carbon sources that are not optimum for *S. pombe* and therefore constitute a challenge where adaptation is sure to arise. Every 24 hours cells were diluted 1:31 to new fresh media using a multichannel micropipette. The daily 31-fold dilutions during transfer of the populations to fresh medium allowed ~5 cell generations/day. Under these conditions, I calculated an effective population size, N_e, of 1.5×10^7 (Lenski *et al*., 1991). GFP/mCherry ratios were monitored approximately every 15 generations by measuring the fluorescence from each well on the plate using a Cyan ADP cell analyser (Beckman Coulter, Inc.) and counting 10.000 total cells for each sample. Measurements were taken immediately after the daily dilution. The same experimental design was done for the populations of strains with inversion in chromosome I.

4.3.3 Analysis of marker trajectories and inferred parameters. The frequency change of the two marked subpopulations over time reflects the invasion of a subpopulation due to hitchhiking (selection of alleles based solely on their linkage to other alleles) with beneficial mutations. I have used the model of Illingworth and Muston to estimate the time of appearance, effect and number of beneficial mutations in each population using the available free online software Optimist (Illingworth and Mustonen, 2012). The inferred trajectories for each subpopulation were estimated by the method and a model was obtained for each mutation parameter. For each simulation, I have chosen 10 rounds of optimisation. The best model to fit the observed frequencies was chosen by the lowest Akaike's information criterion (A.I.C.= $2k - 2log$LL, where k is the number of parameters, k = 3 x #mutations, and LL the likelihood estimated in the method)(Tables 4.3 and 4.4). I run both programs (first_multi and sys_multi) in order to obtain the best parameters for the model. The final output files containing the selection coefficients were

used to design the distribution of beneficial mutations and calculate the beneficial mutation rate.

The model has been designed for a maximum population size of 200 individuals (for plate counting). However, since I have used FACS analysis, the number of individuals obtained for each time point is much higher (~10.000). An error of % was calculated as the error associated with frequency measurements in FACS. Population sizes for 1% error were estimated according to the formula $n = \frac{p\,(1-p)}{\left(\frac{error}{2}\right)^2}$, where p is the frequency of the marker. Parameters obtained from the method are listed in Tables 4.5 and 4.6. I have also performed all the analysis considering the model's maximum population size of 200 colonies. These results were the same and I chose to represent only those for a population size for which the error is 1%.

We estimated a mutation rate for each genotype using the formula $Ub = \frac{k}{Ne.2E(sa).T}$ where *k* is the total number of observed mutations, $E(s_a)$ is the mean selection coefficient, *T* is the number of generations multiplied by the number of populations and N_e the population effective size ($N_e = N_0$ x *g*, where g is the number of generations per day).

4.4. Results and Discussion.

4.4.1. Trajectories of marker frequencies estimated by the model.

I have used a common laboratory strategy to characterize beneficial mutations by allowing the propagation of isogenic strains labelled either with a constitutive neutral marker GFP or mCherry. I have followed 15 independent lines of a collinear genome (Control, C) and of a rearranged (Inverted, Inv1) strain of *S. pombe* to adapt to a given laboratory environment for approximately 330 generations and followed the frequency of fluorescent markers at periodic intervals (Figures 4.3a and 4.3b). The changes in marker frequencies depend upon the rate and strength at which beneficial mutation

occur in the subpopulation. Beneficial mutations occurring anywhere in the genome will drag the fluorescent neutral marker (Kao and Sherlock, 2008). This can only be used for asexual populations where there is no recombination. The observed dynamics is represented in Figure 4.3. As expected, there is signature of appearance of beneficial mutations because of the frequency changes. Beneficial mutations deviate from these ratios at around generation 100 for control populations and at around generation 150 for rearranged populations.

I have used the model of Illingworth and Muston to fit the trajectories of the marker frequencies of each population using the online free available software Optimist (http://www.sanger.ac.uk/resources/software/optimist/). For each population, I have obtained theoretical frequencies based on inferred selection coefficients and time of appearance for beneficial mutations. As seen in the examples of Figure 4.3c, the model describes well the observation.

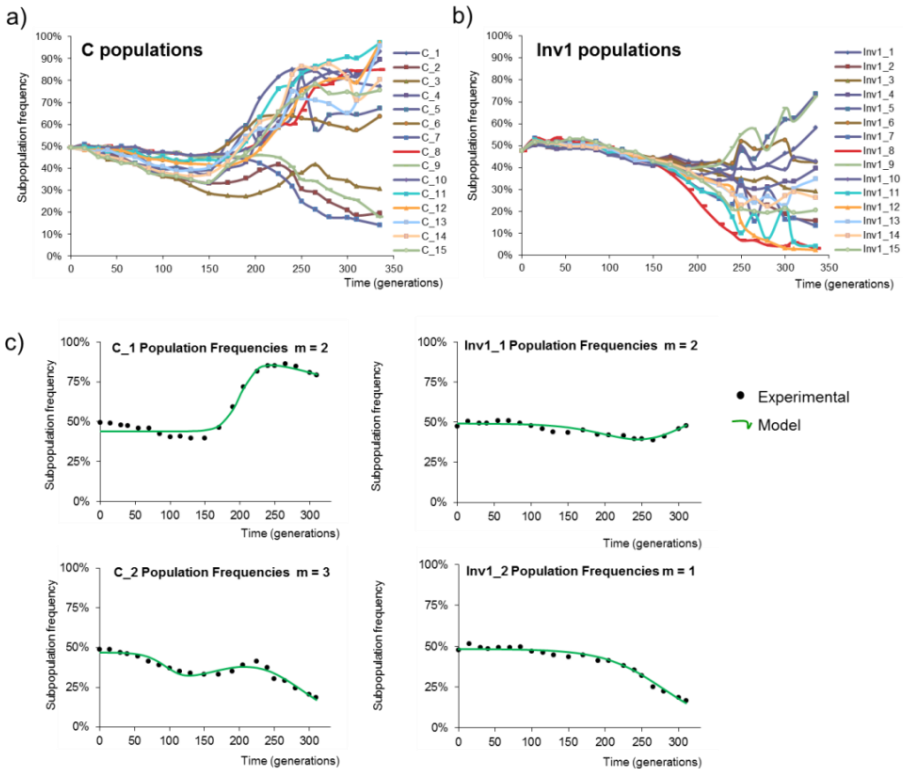


Figure 4.3. Evolution of Control (C) and inverted (Inv1) populations. a, **b**, experimental marker frequencies for one of the two evolving subpopulations of *S. pombe* in control (**a**) and in rearranged (**b**) populations. **c**, four examples of experimental (black dots) and inferred (red and green solid line) marker frequencies.

4.4.2. Mutation rate but not distribution of fitness effects is the same for the two genetic backgrounds.

By applying the model described by Illingworth and Mustonen it is possible to infer population parameters such as the time of appearance and the selection coefficients of beneficial mutations in experimental populations of microorganisms (Tables 4.5 and 4.6). These parameters in turn make it possible to draw a distribution of fitness effects of beneficial mutations (see below).

Table 4.1. Values estimated from the model.

Parameters	Control populations	Inverted populations
K	36	28
$E(s_a)$	0.0775	0.0466
N_e	$1.5 \times 10^{+7}$	$1.5 \times 10^{+7}$
T	310 x 15	310 x 15
Beneficial mutation rate U_b	3.3×10^{-9}	4.3×10^{-9}

The model inferred a total of 36 events in the Control population and 28 in the Inv1 population. With these parameters, I have calculated a mutation rate of $U_b = 3.3 \times 10^{-9}$ beneficial mutations per genome per generation for Control background and a slightly different $U_b = 4.3 \times 10^{-9}$ for the Inv1. These values are from the same order of magnitude as described in a *Eschericia coli* study (Imhof and Schlötterer, 2001) although the mean effect of mutations in that study was only 0.02 compared to the ~0.08 and ~0.05 found here. The U_b rates here found are 100 times smaller than those estimated in the majority of studies using large effective population sizes of 10^{+7} bacteria (Gordo *et al.*, 2011).

Taking into account the large effective population size used ($N_e \approx 1.5 \times 10^{+7}$) the effect of clonal interference is significant and not taken into account in my calculations. I am therefore aware of the probable underestimation of U_b since populations were large enough for the simultaneous presence of different genotypes at very low frequency in the population. Smaller population sizes allow the detection of small effect mutations that might be lost in competition. Using small population sizes (10^{+4}), the estimated beneficial mutation of *E. coli* is of 2×10^{-5} (Perfeito *et al*., 2007). Even so, the large N_e used in my

work is in the order of most of the published data and therefore I can compare this data to previous ones.

Single mutations with selection coefficients lower than 0.02 could not be detected which could also lead to an underestimation of the beneficial mutation rate of arising mutations. Technically this is because many newly arising beneficial mutations of small effect are lost in competition with mutations of large effect and do not have time to increase in frequency in order to be detected (Orr, 2003).

Despite the similar beneficial mutation rate between the two different genetic backgrounds, the distribution of the beneficial mutations is significantly different ($p=7.46 \times 10^{-7}$, Kolmogorov-Smirnov)(Figure 4.4). The mean for distribution of fitness effects both populations is also significantly different ($p = 3.7 \times 10^{-6}$, Kruskal-Wallis test). Despite this, I cannot state that the different distributions for the effects are caused uniquely by the rearrangement. The control strain used in this experiment does not have the same auxotrophic markers as the rearranged background and so the two populations cannot be directly compared. The cassettes used for the generation of the inversion are inserted at the *his1*$^{+}$ and the *lys3*$^{+}$ loci whereas the control used has the parental cassettes inserted in the *lys4*$^{+}$ and the *arg7*$^{+}$ loci. This result is nonetheless highly encouraging and suggests that just by changing the order of genes along a chromosome there is not only an immediate fitness variation as measured in Chapter 3, but also the propensity to accumulate different mutations than its collinear counterpart.

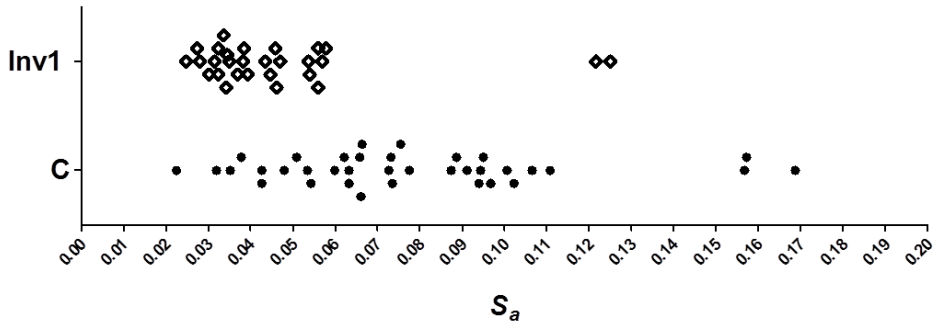


Figure 4.4 Distribution of fitness effects beneficial of mutations. C, control populatios; Inv1, rearranged populations.

The method described by Illingworth and Mustonen allows for the inference of both the selection coefficients of beneficial mutations but also for their time at which each beneficial mutation occurred and in which population taking into q_0 (set to 0.1% of the population). I have used the temporal information to try to extract a pattern of temporal distribution of beneficial mutations. New arising mutations occur throughout time in both backgrounds (Figure 4.5). However, this data clearly shows the measurement of beneficial mutations at time 0 generations (Table 4.5 and 4.6). Almost all of the 15 independent lines have an effect from the fluorescent marker which is therefore not neutral. The algorithm designed by Illingworth and Mustonen does not allow for the use of non-neutral markers. In order to exclude the non-neutrality of the marker it would be necessary to re-write the algorithm. This result implies not only the non-neutrality of the markers but also that some beneficial mutations have been overestimated the number of mutations and their selective factor. Still, comparison can be made as this effect is occurring in both control and rearranged populations.

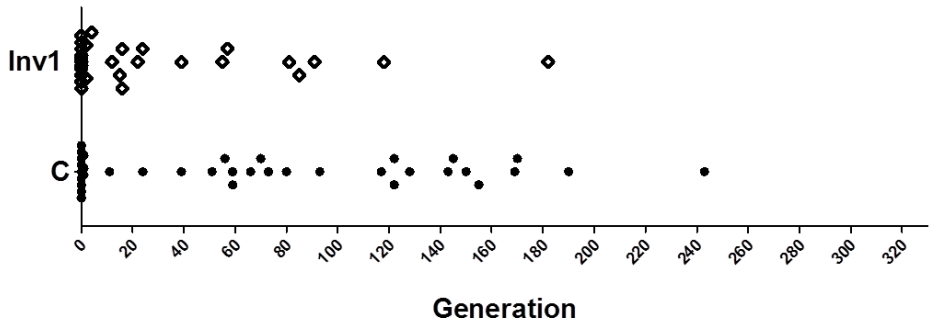


Figure 4.5. Distribution of time of appearance of beneficial mutations. C, control populatios; Inv1, rearranged populations.

4.5. Conclusions.

This work shows for the first time an evolution experiment using the fission yeast *Schizosaccharomyces pombe*. I have extracted population genetics parameters by fitting my experimental data with a theoretical model proposed by Illingworth and Mustonen (Illingworth and Mustonen, 2012). I have quantified a beneficial mutation rate and mean selection coefficient that are at least ten times lower than most of the values calculated for several species of bacteria (Gordo *et al*., 2011) as well as for *S. cerevisiae* where the estimated beneficial mutation rate is in the order of 10^{-5} (Desai *et al*., 2007).

This work suggests that just by changing the order of genes along a chromosome the distribution of fitness effects of beneficial mutations is different than in its collinear counterpart. This work does not however allow for a clear statement that the difference distributions calculated are caused uniquely by the rearrangement since the collinear genome (control) used in this experiment does not have the same auxotrophic markers as the inverted background. In order to address this question properly, it will be necessary to repeat this experiments using the proper control containing the breakpoints at the same auxotrophic markers as the inversion. Additionally, in order to detect the maximum beneficial mutations one should include more neutral markers

and smaller population sizes. This would allow detecting small effect mutations that might be lost in competition and perhaps approximate the values here estimated from those of *S. cerevisiae*.

If verified, this result can have important consequences in the speed of adaptation of one genotype versus the other if they are competing for the same resources. The different distribution of mutational effects in the two backgrounds can have major implications if one considers speciation processes, for instance. When these two genotypes meet, some mutations will not be exchanged freely as a result of problems in meiosis (see Chapter 3 of this thesis). Additionally, if the fitness effect of beneficial mutations arising in each background are indeed different, this could imply that the process of speciation between these two backgrounds would perhaps be faster than our expectations.

If verified, this work opens a new question: how can genome with exactly the same gene content but only an alteration in the order of these genes, alter the effects of beneficial mutations? Additionally, my observations that the genomic background can alter the beneficial mutation rate could be extended to all the different genomic configurations generated in Chapter 3 of this thesis. This would increase the knowledge of how much the beneficial mutation rate depends on the backgrounds.

4.6. Acknowledgements. I would like to thank I Gordo (IGC) for help on the model analysis and in writing the statistical analysis. I also thank J Sousa (IGC) for help on Mathematica 8.0 and R Alves (IGC) and G Neto (IGC) for help on computation.

4.7. Tables.

Table 4.2. Strains used in this study.

	Genotype	Common name	Creator
MGF 1337	*h- arg7::padh1-loxP-kanMX6 lys4::loxP-ura4-kanMX6 leu1-32 mat1-M::mat1-M-natMX6 ade6^{+} lys4-gfp-hphMX6R*	**Control - gfp**	This work
MGF 1318	*h+ arg7::padh1-loxP-kanMX6 lys4::loxP-ura4-kanMX6 leu1^{+} mat1-P::mat1-P-natMX6 ade6-M216 lys4-gfp-hphMX6R*	**Control - gfp**	This work
MGF 1336	*h- arg7::padh1-loxP-kanMX6 lys4::loxP-ura4-kanMX6 leu1-32 mat1-M::mat1-M-natMX6 ade6^{+} lys4-gfp-hphMX6R*	**Control-mCherry**	This work
MGF 1388	*h+ arg7::padh1-loxP-kanMX6 lys4::loxP-ura4-kanMX6 leu1-$^{+}$ mat1-P::mat1-P-natMX6 ade6-M216 lys4-gfp-hphMX6R*	**Control-mCherry**	This work
MGF 1339	*h+ his1::loxP- kanMX6^{R} lys3::padh1-loxP-ura4^{+} - kanMX6^{R} leu1-32 ade6-M216 ura4-D18 mat1-P::mat1-P-natMX6 his1-gfp-hphMX6^{R}*	**Inv1 - gfp**	This work
MGF 1341	*h- his1::loxP- kanMX6^{R} lys3::padh1-loxP-ura4^{+} - kanMX6^{R} leu1-32 ade6-M216 ura4-D18 mat1-M::mat1-M-natMX6 his1-gfp-hphMX6^{R}*	**Inv1 - gfp**	This work
MGF 1343	*h+ his1::loxP- kanMX6^{R} lys3::padh1-loxP-ura4^{+} - kanMX6^{R} leu1-32 ade6-M216 ura4-D18 mat1-P::mat1-P-natMX6 his1-gfp-hphMX6^{R}*	**Inv1-mcherry**	This work
MGF 1342	*h- his1::loxP- kanMX6^{R} lys3::padh1-loxP-ura4^{+}- kanMX6^{R} leu1-32 ade6-M216 ura4-D18 mat1-M::mat1-M-natMX6 his1-mCherry-hphMX6^{R}*	**Inv1-mcherry**	This work

Table 4.3. Data for fitting of model for Control populations.

Population	Model	LL	mut	K	AIC
C_1	m = 1	-2804,66	1	3	5615,32
C_1	m = 2	-446	2	6	904
C_1	m = 3	-178	3	9	374
C_2	m = 3	-258,47	3	9	534,94
C_3	m = 3	-313	3	9	644,224
C_4	m = 4	-237	4	12	498,718
C_5	m = 1	-81,4463	1	3	168,8926
C_5	m = 2	-503,297	2	6	1018,594
C_5	m = 3	-286	3	9	590
C_6	m = 1	-1469,67	1	3	2945,34
C_6	m = 2	-238	2	6	487,554
C_7	m = 2	-165	2	6	342,464
C_8	m = 1	-330,37	1	3	666,74
C_8	m = 2	-198	2	6	408,25
C_9	m = 1	-206	1	3	418,048
C_10	m = 2	-164	2	6	340,886
C_11	m = 1	-1193,01	1	3	2392,02
C_11	m = 2	-231	2	6	473,542
C_12	m = 2	-429	2	6	869,056
C_12	m =3	-334,12	3	9	686,24
C_13	m = 2	-612	2	6	1236
C_13	m = 3	-600	3	9	1218
C_14	m = 3	-65,5144	3	9	149,0288
C_14	m = 2	-737	2	6	1486
C_14	m = 3	-366	3	9	750
C_15	m = 3	-295	3	9	608,106

Table 4.4. Data for fitting of model for inverted (Inv1) populations.

Population	Model	LL	mut	K	AIC
Inv1_1	m = 2	-197	2	6	406,012
Inv1_2	m = 1	-205	1	3	415,086
Inv1_3	m = 1	-520	1	3	1046
Inv1_3	m = 2	-138,42	2	6	288,84
Inv1_4	m = 1	-868	1	3	1741,796
Inv1_4	m = 2	-188	2	6	387,932
Inv1_5	m = 2	-217	2	6	446,792
Inv1_6	m = 1	-770	1	3	1546
Inv1_6	m = 2	-364	2	6	739,238
Inv1_7	m = 2	-214	2	6	440,472
Inv1_8	m = 2	-183	2	6	377,438
Inv1_9	m = 2	-274	2	6	560,176
Inv1_10	m = 2	-209	2	6	429,396
Inv1_11	m = 1	-1193,01	1	3	2392,02
Inv1_11	m = 2	-356	2	6	724,178
Inv1_12	m = 1	-475,252	1	3	956,504
Inv1_12	m = 2	-400	2	6	812
Inv1_12	m =3	-235	3	9	488,532
Inv1_13	m = 1	-1271,75	1	3	2549,5
Inv1_13	m = 2	-245	2	6	501,642
Inv1_14	m = 1	-1262,11	1	3	2530,22
Inv1_14	m = 2	-181	2	6	373,502
Inv1_15	m = 1	-1193,01	1	3	2392,02
Inv1_15	m = 2	-199	2	6	410,056

Table 4.5. Parameters for Control populations with total size n estimated for error = 1%.

Population	Selection Coefficient	Time of appearance	Population where it appeared
C_1	0,0874141	122	0
C_1	0,0939606	150	1
C_2	0,0967471	80	1
C_2	0,0660901	0	1
C_2	0,0735264	24	0
C_3	0,0542673	0	1
C_3	0,102261	169	1
C_3	0,062142	39	0
C_4	0,15673	73	0
C_4	0,11076	0	1
C_4	0,168719	143	1
C_4	0,106503	1	0
C_5	0,0943478	11	0
C_5	0,0886651	0	1
C_6	0,0911377	117	0
C_6	0,0950017	128	1
C_7	0,0598633	190	0
C_7	0,0378121	70	1
C_8	0,0479787	66	0
C_8	0,0319646	0	1
C_9	0,0224832	0	1
C_10	0,0775566	170	1
C_10	0,0754538	145	0
C_11	0,042651	1	1
C_11	0,0632535	59	0
C_12	0,0657819	93	0

C_12	0,0351928	0	1
C_12	0,0663504	122	1
C_13	0,042651	1	1
C_13	0,0632535	59	0
C_14	0,0508617	0	1
C_14	0,157188	243	1
C_14	0,0731491	56	0
C_15	0,0534621	0	1
C_15	0,100624	155	1
C_15	0,0726892	51	0

Table 4.6. Parameters for inverted (Inv1) populations with total size n estimated for error = 1%.

Population	Selection Coefficient	Time of appearance	Population where it appeared
Inv1_1	0,038235	85	0
Inv1_1	0,0280241	0	1
Inv1_2	0,0248213	0	1
Inv1_3	0,121607	2	0
Inv1_3	0,125031	4	1
Inv1_4	0,0537324	2	0
Inv1_4	0,0559528	0	1
Inv1_5	0,0461654	81	0
Inv1_5	0,0323736	0	1
Inv1_6	0,056957	39	0
Inv1_6	0,0539597	24	1
Inv1_7	0,0344435	12	0
Inv1_7	0,0384802	0	1
Inv1_8	0,0577655	182	0
Inv1_8	0,0350286	0	1

Inv1_9	0,0559435	57	0
Inv1_9	0,0459037	16	1
Inv1_10	0,0470074	22	0
Inv1_10	0,0446981	0	1
Inv1_11	0,027434	15	0
Inv1_11	0,0342711	0	1
Inv1_12	0,0314932	16	1
Inv1_13	0,0369403	55	0
Inv1_13	0,0336507	0	1
Inv1_14	0,0434916	118	0
Inv1_14	0,0302029	0	1
Inv1_15	0,0393065	91	0
Inv1_15	0,0323941	0	1

Chapter 5 – General Discussion

Biological diversity is the main driver of evolution. This diversity is manifested by different phenotypes which arise as a complex response to the genetic composition of each organism. Mutations at single nucleotides and those encompassing several megabases such as inversions, for instance, make up the genetic diversity of organisms. This diversity at the genomic level contributes both for standing genetic variation within populations and at a greater extent to divergence between species.

A great part of research in biology is dedicated to understanding how genetic diversity is translated into phenotypic variability. In this thesis, I have addressed how genomic variation in the form of CRs affects phenotypic variation.

5.1 Chromosomal Rearrangements in *S. pombe* natural isolate strains.

In this work, I describe a high prevalence of CRs in natural isolates of the fission yeast *S. pombe* thus reinforcing its importance as a polymorphic trait. In particular, I mapped a pericentric inversion in Chr.I which is present in all the strains analysed in relation to the reference strain L972. This inversion has also been recently described for other isolates of *S. pombe* (Brown *et al.*, 2011). Since these strains were collected from different geographic locations, I hypothesize that the inversion event appeared independently in different isolates. However, I also consider that the most likely scenario is that in which the genomic configuration for region in the strain L972 is actually the derived one.

I have also observed independent translocations in two out the ten strains analysed one of which has also been described in five other strains spread world-wide (Brown *et al.*, 2011). Interestingly, one of the breakpoints in the two translocations here described is localized in a close, not-determined, region of Chr.I. This can be indicative of selection acting on this particular chromosomal disruption in a way similar to what happens in *Drosophila* and *Anopheles*, where both fixed and polymorphic inversions are found in specific

chromosome arms (Krimbas and Powell, 1992; Pombi *et al*., 2008). However, contrary to what has been seen in *Drosophila* (Krimbas and Powell, 1992; Bhutkar *et al*., 2008) and mammalian species (Stefansson *et al*., 2005; Zhao and Bourque, 2009), translocations are not a rare event in natural isolates of *S. pombe* (Chapter 2 of this thesis; Brown *et al*., 2011). Even if smaller inversions are mapped in the future, translocations are still a frequent type of CRs in *S. pombe*. One possible reason for their high prevalence might be related to the life cycle of *S. pombe*. Fission yeast is a microorganism which divides mitotically without obligatory engaging in the sexual cycle. Therefore, if the rearrangement in question has a mitotic advantage and it is rarely "seen" by meiosis it is likely to survive. On the other hand, the rearrangements occur in strains that are geographically isolated and perhaps would never have contact in natural situations which would increase its fixation probability. This point has not been addressed here and in order to tackle it is necessary to understand how CRs can increase in frequency and not be lost once they appear in the population.

In what concerns the rearrangements in *S. pombe* isolates, the data available is not sufficient to postulate a probable mechanism for the formation of the rearrangements. Work by Brown *et al.* has not found repetitive sequences at some studied breakpoints which suggest that no transposable-like element or duplicated sequences were involved in the formation of the rearrangements (Brown *et al*., 2011). However, the presence of microhomology sequences in two out of five sequences involved in the exchanges (Brown *et al*., 2011) could mean that a replication based mechanism such as the MMBIR is involved.

The natural isolates of *S. pombe* are a powerful source to study the mechanism and rate of appearance of CRs in fission yeast. Additionally, they remain as tools to test the amount of gene flow that is occurring in natural populations and to correlate it with the proximity to CRs. According to theory (Faria and Navarro, 2010; Jackson, 2011)and to my own data (Chapter 3), I

would expect higher nucleotide divergence near the breakpoints than in collinear regions, since these are prevented from gene flow.

Finally, for this part of the work, two important questions are still open. The first one is if these CRs adaptive and the second is which cellular mechanisms are altered between the strains as a result of these rearrangements.

5.2 Can CR appear as an early mechanism for genetic separation?

An extremely important component of fitness for organisms with sexual reproduction is their ability to produce viable offspring. This will determine whether its genes are maintained or extinct from the population. In order to be maintained, the genome has to be faithfully propagated which causes a theoretical problem for carriers of CRs. CRs are associated with low hybrid viabilities and gene flow in heterozygotic crosses (Noor *et al*., 2001; Schaeffer *et al*., 2003; Brown *et al*., 2004). This property of CRs, along with their ubiquity between different species led to the proposal that CRs could serve as an initial genetic barrier for the speciation process (Noor *et al*., 2001; Rieseberg, 2001; Kirkpatrick and Barton, 2006, 2006; Hoffmann and Rieseberg, 2008).

I wondered whether CRs in the natural isolates had phenotypic consequences at the meiotic level. In order to address this, I have quantified the meiotic viabilities of several heterozygotic crosses between the different natural isolate strains of *S. pombe*. These strains seem to be undergoing reproductive isolation given that some are almost incapable of producing viable offspring in heterozygotic crosses. In some cases the viability of hybrids can be as low as 1%. I propose that this could be attributed to the chromosomal differences between them since I show that high levels of karyotypic diversity occur in face of low nucleotide diversity. Probably, the accumulation of these CRs affects both meiotic HR and segregation, thus leading to the production of gametes with unbalanced DNA content. However, I cannot state that CRs are the single factor contributing to the low viability of

the hybrids in spite of nucleotide diversity of the order observed within species diversity. Several mechanisms can contribute to reproductive isolation (Maheshwari and Barbash, 2011). It is possible that some nuclear or mitochondrial genes could be causing genic incompatibilities in parallel (Lee *et al*., 2008; Presgraves, 2010). For instance, intron variation in *cox1*$^{+}$ has been involved in a nuclear-mitochondrial incompatibility causing reproductive isolation between *S. cerevisiae* and two other species, *S. bayanus* and *S. paradoxus* (Chou *et al*., 2010). The nucleotide variation measured here for the mitochondrial gene *cox1*+ could thus also be strongly associated to the hybrid lethality measured among the natural isolates.

For the above, it would be impossible to clearly ascertain the impact on reproduction of these particular CRs because they are the product of natural selection associated with a given genome diversity. Also, up to now there was no clear measurement of meiotic viability and HR with “clean” CRs. For these reasons, I constructed ten genetically engineered unbiased CRs in *S. pombe* departing from a clean common background. These ten rearrangements included two pericentric inversions and eight reciprocal translocations. I have quantified drops in hybrid viability up to 54 ± 14% in heterozygotic crosses involving CRs. Chromosomal rearrangements seem to affect meiosis at different stages depending on the type of rearrangement. Inversions lower HR as predicted (Navarro *et al*., 1997) but not translocations. The both types of CRs prevent recombination at loci near the breakpoint. These results support that CRs are a mechanisms to create a genetic barrier for gene flow and revive CRs as a testable mechanism for initiation of speciation processes as previously proposed (Noor *et al*., 2001; Rieseberg, 2001; Navarro and Barton, 2003; Kirkpatrick and Barton, 2006; Hoffmann and Rieseberg, 2008; Faria and Navarro, 2010).

5.3 If CRs are impairing meiosis how can they be maintained in nature?

So far there was no clear measurement of the fitness distribution of single CRs. Previous work by Colson *et al.* has suggested that CRs are adaptive as they contribute to higher fitness of yeast growing in nutrient limited-environments (Colson *et al*., 2004). Colson and colleagues measured the fitness effects of two reciprocal translocations that distinguish *S. mikatae* from *S. cerevisiae*. However, the two translocations studied exist as a result of natural selection throughout the evolutionary period that separates the two species.

In this thesis I have extended this knowledge by designing CRs that have not been exposed to selective pressures neither disrupt any gene sequence. I found that chromosome structure is a selectable trait *per se* and that the phenotypic diversity of CRs is as rich as for SNPs: CRs can be neutral, deleterious and even beneficial. Moreover, this selection is environmentally-dependent and acts on structure *per se* without the need of hitchhiking other genes involved in the adaptation.

However, the implications of the meiotic data would go against the maintenance of CRs in nature. Yet, I do see them in natural populations. One of the main criticisms to the chromosomal speciation theory has to do with the heterozygous state of the CRs. On one hand, if CRs are weakly underdominant, then they would not contribute significantly to hybrid sterility. On the other hand, if they are strongly underdominant, they would be immediately eliminated. In my thesis I show that these scenarios are only theoretical. I demonstrate that CRs can lead to enough meiotic divergence without underdominance of the rearrangement. In practice, I observed that CRs cause genetic isolation by lowering hybrid viability and HR but at the same time can be highly fit in mitosis. This happens, for instance, in the case of translocation T6 which is quite lethal in meiosis but in mitosis, for instance, in YES maltose media can out-compete the collinear karyotype.

The results present here suggest that CRs can exist by a phenomenon of antagonistic pleiotropy. Antagonistic pleiotropy was first defined by George Williams in 1957 as way to explain the trade-off between genes involved in reproduction that are beneficial early in life but disadvantageous in a later phase contributing to senesce and ageing (Williams, 1957). Antagonistic pleiotropy occurs when one gene that controls more than one trait has a beneficial fitness effect in at least one trait but a deleterious one in another trait. As seen in this thesis, CRs affect both the asexual and the sexual life cycles. I saw several examples where effects in the asexual life cycle would compensate the deleterious effects in the sexual cycle. Likewise, disadvantages in one environment can be balanced by advantages in another. Yet, the selective value of the CR depends on how it affects the total reproductive probability. Therefore, the definitive experimental proof combining the two life stages in a single experiment is still lacking. Additionally, it is important to ascertain whether this trade-off is frequency-dependent, which would be fundamental to establish patterns of appearance and maintenance of CRs in nature.

5.4 How do CRs cause the phenotypic diversity?

We now know that chromosome structure is highly dynamic and that its complex organization does not rely only on the linear sequence. In the literature, CRs have been described to cause alterations in gene expression. This occurs in human pathological conditions where the CR leads to disruption of a gene or promoter sequence (Rowley, 2001; Nambiar and Raghavan, 2011). Additionally, it has also been shown to some extent that these gene expression differences are due to alterations in position of genes and relative position to regulatory regions (Lanctôt *et al*., 2007).

In this work, I show that chromosomal diversity without disruption of gene or promoter sequences leads to a surprisingly enormous phenotypical diversity. I show that without disruption of neither of these sequences, gene

expression is altered just as a consequence of structural reorganization. This point out to the relevance of chromosomal context and that *trans* and/or *cis* regulation might be disturbed as a result of alterations in chromosome structure.

Gene positioning has implications for gene expression and can consequently be the cause of phenotypic variation observed in this work. This opens one question: how are genes relocated relative to one another and to regulatory regions inside the nucleus? I envision that application of technology such as 3C or 4C techniques would provide clues on the relocation of genes and its relation with differences in expression (Takizawa *et al*., 2008). But what if genes move around in the chromosome but do not change their relative nuclear context? Then I would consider alterations in the local or global chromatin status. This could, for instance, alter promoter occupancy in the rearrangements versus that of parental strains.

We could also explore other mechanisms behind the phenotypic variation. One of those may involve alteration of replication times. Replication can be slightly altered when chromosomes are rearranged. For instance, the rearrangements could cause alterations in replication fork stability. In this case, the genetic background can be a source of differentiation on its own if it is changing the mutational context. Faria and Navarro pointed out that CRs could promote changes in the patterns of mutations due to changes of relative position of genes inside CRs to replication origins (Faria and Navarro, 2010). Interestingly, my work suggests that a pericentric inversion on Chr.I has the same order of beneficial mutation rate that a collinear strain but the distribution of the fitness effects between the two backgrounds will be different. This is intriguing on its own and further work should be done in order to address whether CRs do indeed affect the distribution of beneficial effects of mutations and to map where these occur in the genome, particularly in relation to the rearrangement. Additionally, the same quantifications for deleterious mutations could be done. In these future experiments, whole genome

sequencing would be highly complementary and would allow to establish the genetic basis of adaptation (in the case of beneficial mutations) and to distinguish between one mutation with high effect and several mutations of smaller effect.

Concluding remarks.

This work establishes the importance of chromosomal structure when interpreting natural variability. Using *S. pombe* as model I show that chromosomes are dynamic structures upon which natural selection acts on. The results here presented shed light into the relevance of chromosomal structure for biological diversity and at the same time open several questions and avenues of research. Additionally, the strains here generated are tools to test important questions from different fields of biology.

Understanding genetic variation has, for me, enormous challenges. The biggest one is to understand how all the biological layers inside a cell and afterwards in an organism coordinate to transduce the genotype into a phenotype. Ideally, in the future, the merger of several different research areas will fill in these layers and help us to go back and forward on the phenotype-to-genotype interactions.

References

Adams, J., Puskas-Rozsa, S., Simlar, J., and Wilke, C.M. (1992). Adaptation and major chromosomal changes in populations of Saccharomyces cerevisiae. Curr. Genet. *22*, 13–19.

Akao, T., Yashiro, I., Hosoyama, A., Kitagaki, H., Horikawa, H., Watanabe, D., Akada, R., Ando, Y., Harashima, S., Inoue, T., *et al.* (2011). Whole-Genome Sequencing of Sake Yeast Saccharomyces cerevisiae Kyokai no. 7. DNA Res.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell (Garland Science).

Albertson, D.G. (2006). Gene amplification in cancer. Trends Genet. *22*, 447–455.

Alekseyev, M.A., and Pevzner, P.A. (2010). Comparative genomics reveals birth and death of fragile regions in mammalian evolution. Genome Biol. *11*, R117.

Alexandre, H., Ansanay-Galeote, V., Dequin, S., and Blondin, B. (2001). Global gene expression during short-term ethanol stress in Saccharomyces cerevisiae. FEBS Lett. *498*, 98–103.

Anderson, W.W., Arnold, J., Baldwin, D.G., Beckenbach, A.T., Brown, C.J., Bryant, S.H., Coyne, J.A., Harshman, L.G., Heed, W.B., and Jeffery, D.E. (1991). Four decades of inversion polymorphism in Drosophila pseudoobscura. Proc. Natl. Acad. Sci. U.S.A. *88*, 10367–10371.

Anton, E., Vidal, F., and Blanco, J. (2008). Reciprocal translocations: tracing their meiotic behavior. Genet. Med. *10*, 730–738.

Aplan, P.D. (2006). Causes of oncogenic chromosomal translocation. Trends Genet *22*, 46–55.

Arcangioli, B., and Klar, A.J. (1991). A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient mat1 switching in Schizosaccharomyces pombe. EMBO J *10*, 3025–3032.

Atwood, K.C., Schneider, L.K., and Ryan, F.J. (1951). Periodic selection in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. *37*, 146–155.

Ayala, F.J., and Coluzzi, M. (2005). Chromosome speciation: humans, Drosophila, and mosquitoes. Proc. Natl. Acad. Sci. U.S.A. *102 Suppl 1*, 6535–6542.

Bähler, J., Wu, J., Longtine, M.S., Shah, N.G., Mckenzie III, A., Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast *14*, 943–951.

Bartolomé, C., and Charlesworth, B. (2006). Rates and Patterns of Chromosomal Evolution in Drosophila pseudoobscura and D. miranda. Genetics *173*, 779–791.

Bataillon, T., Zhang, T., and Kassen, R. (2011). Cost of adaptation and fitness effects of beneficial mutations in Pseudomonas fluorescens. Genetics *189*, 939–949.

Bhutkar, A., Gelbart, W.M., and Smith, T.F. (2007). Inferring genome-scale rearrangement phylogeny and ancestral gene order: a Drosophila case study. Genome Biol. *8*, R236.

Bhutkar, A., Schaeffer, S.W., Russo, S.M., Xu, M., Smith, T.F., and Gelbart, W.M. (2008). Chromosomal Rearrangement Inferred From Comparisons of 12 Drosophila Genomes. Genetics *179*, 1657–1680.

Bickmore, W.A., and Teague, P. (2002). Influences of chromosome size, gene density and nuclear position on the frequency of constitutional translocations in the human population. Chromosome Res. *10*, 707–715.

Bondeson, M.L., Dahl, N., Malmgren, H., Kleijer, W.J., Tönnesen, T., Carlberg, B.M., and Pettersson, U. (1995). Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. Hum. Mol. Genet. *4*, 615–621.

Bonierbale, M.W., Plaisted, R.L., and Tanksley, S.D. (1988). RFLP Maps Based on a Common Set of Clones Reveal Modes of Chromosomal Evolution in Potato and Tomato. Genetics *120*, 1095–1103.

Bourque, G., Pevzner, P.A., and Tesler, G. (2004). Reconstructing the Genomic Architecture of Ancestral Mammals: Lessons From Human, Mouse, and Rat Genomes. Genome Res. *14*, 507–516.

Bourque, G., Zdobnov, E.M., Bork, P., Pevzner, P.A., and Tesler, G. (2005). Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. Genome Res *15*, 98–110.

Britton-Davidian, J., Catalan, J., Lopez, J., Ganem, G., Nunes, A.C., Ramalhinho, M.G., Auffray, J.C., Searle, J.B., and Mathias, M.L. (2007). Patterns of genic diversity and structure in a species undergoing rapid chromosomal radiation: an allozyme analysis of house mice from the Madeira archipelago. Heredity (Edinb) *99*, 432–442.

Brown, C.J., Todd, K.M., and Rosenzweig, R.F. (1998). Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol. Biol. Evol. *15*, 931–942.

Brown, K.M., Burk, L.M., Henagan, L.M., and Noor, M.A.F. (2004). A test of the chromosomal rearrangement model of speciation in Drosophila pseudoobscura. Evolution *58*, 1856–1860.

Brown, W.R.A., Liti, G., Rosa, C., James, S., Roberts, I., Robert, V., Jolly, N., Tang, W., Baumann, P., Green, C., *et al.* (2011). A Geographically Diverse Collection of Schizosaccharomyces Pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. G3 *1*, 615–626.

Brunner, D., and Nurse, P. (2000). New concepts in fission yeast morphogenesis. Philos Trans R Soc Lond B Biol Sci *355*, 873–877.

Carreto, L., Eiriz, M.F., Gomes, A.C., Pereira, P.M., Schuller, D., and Santos, M.A. (2008). Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics *9*, 524.

Chance, P.F., Abbas, N., Lensch, M.W., Pentao, L., Roa, B.B., Patel, P.I., and Lupski, J.R. (1994). Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. Hum. Mol. Genet. *3*, 223–228.

Chance, P.F., Alderson, M.K., Leppig, K.A., Lensch, M.W., Matsunami, N., Smith, B., Swanson, P.D., Odelberg, S.J., Disteche, C.M., and Bird, T.D. (1993). DNA deletion associated with hereditary neuropathy with liability to pressure palsies. Cell *72*, 143–151.

Chapter 2 of this thesis.

Chen, D., Toone, W.M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N., and Bähler, J. (2003). Global transcriptional responses of fission yeast to environmental stress. Mol. Biol. Cell *14*, 214–229.

Chikashige, Y., Ding, D.Q., Imai, Y., Yamamoto, M., Haraguchi, T., and Hiraoka, Y. (1997). Meiotic nuclear reorganization: switching the position of centromeres and telomeres in the fission yeast Schizosaccharomyces pombe. EMBO J *16*, 193–202.

Chou, J.-Y., Hung, Y.-S., Lin, K.-H., Lee, H.-Y., and Leu, J.-Y. (2010). Multiple molecular mechanisms cause reproductive isolation between three yeast species. PLoS Biol. *8*, e1000432.

Codón, A.C., and Benitez, T. (1995). Variability of the Physiological Features and of the Nuclear and Mitochondrial Genomes of Bakers' Yeasts. Systematic and Applied Microbiology *18*, 343–352.

Coghlan, A., and Wolfe, K.H. (2002). Fourfold Faster Rate of Genome Rearrangement in Nematodes Than in Drosophila. Genome Res *12*, 857–867.

Colson, I., Delneri, D., and Oliver, S.G. (2004). Effects of reciprocal chromosomal translocations on the fitness of Saccharomyces cerevisiae. EMBO Rep. *5*, 392–398.

Coluzzi, M., Sabatini, A., Della Torre, A., Deco, D., Angela, M., and Petrarca, V. (2002). A Polytene Chromosome Analysis of the Anopheles Gambiae Species Complex. Science *298*, 1415–1418.

Cornforth, M.N., Greulich-Bode, K.M., Loucas, B.D., Arsuaga, J., Vázquez, M., Sachs, R.K., Brückner, M., Molls, M., Hahnfeldt, P., Hlatky, L., *et al.* (2002). Chromosomes are predominantly located randomly with respect to each other in interphase human cells. J Cell Biol *159*, 237–244.

Coyle, S., and Kroll, E. (2008). Starvation induces genomic rearrangements and starvation-resilient phenotypes in yeast. Mol. Biol. Evol. *25*, 310–318.

Coyne, J.A., Aulard, S., and Berry, A. (1991). Lack of Underdominance in a Naturally Occurring Pericentric Inversion in Drosophila Melanogaster and Its Implications for Chromosome Evolution. Genetics *129*, 791–802.

Cutter, A.D., Dey, A., and Murray, R.L. (2009). Evolution of the Caenorhabditis elegans genome. Mol. Biol. Evol. *26*, 1199–1234.

Davis, L., and Smith, G.R. (2003). Nonrandom homolog segregation at meiosis I in Schizosaccharomyces pombe mutants lacking recombination. Genetics *163*, 857–874.

Delneri, D., Colson, I., Grammenoudi, S., Roberts, I.N., Louis, E.J., and Oliver, S.G. (2003). Engineering evolution to study speciation in yeasts. Nature *422*, 68–72.

Desai, M.M., Fisher, D.S., and Murray, A.W. (2007). The speed of evolution and maintenance of variation in asexual populations. Curr. Biol. *17*, 385–394.

Dhar, R., Sägesser, R., Weikert, C., Yuan, J., and Wagner, A. (2011). Adaptation of Saccharomyces cerevisiae to saline stress through laboratory evolution. J. Evol. Biol. *24*, 1135–1153.

Ding, D.-Q., Haraguchi, T., and Hiraoka, Y. (2010). From meiosis to postmeiotic events: alignment and recognition of homologous chromosomes in meiosis. FEBS J. *277*, 565–570.

Dobzhansky, T., and Sturtevant, A.H. (1938). Inversions in the Chromosomes of Drosophila Pseudoobscura. Genetics *23*, 28–64.

Drillon, G., and Fischer, G. (2011). Comparative study on synteny between yeasts and vertebrates. C. R. Biol. *334*, 629–638.

Dujon, B. (2010). Yeast evolutionary genomics. Nat. Rev. Genet. *11*, 512–524.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., *et al.* (2004). Genome evolution in yeasts. Nature *430*, 35–44.

Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., and Botstein, D. (2002). Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. *99*, 16144–16149.

Emanuel, B.S., and Saitta, S.C. (2007). From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. Nat. Rev. Genet. *8*, 869–883.

Fan, J.B., Chikashige, Y., Smith, C.L., Niwa, O., Yanagida, M., and Cantor, C.R. (1989). Construction of a Not I restriction map of the fission yeast Schizosaccharomyces pombe genome. Nucleic Acids Res *17*, 2801–2818.

Fan, J.B., Grothues, D., and Smith, C.L. (1991). Alignment of Sfi I sites with the Not I restriction map of Schizosaccharomyces pombe genome. Nucleic Acids Res. *19*, 6289–6294.

Faria, R., and Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends in Ecology & Evolution *25*, 660–669.

Fay, J.C., and Benavides, J.A. (2005). Evidence for domesticated and wild populations of Saccharomyces cerevisiae. PLoS Genet. *1*, 66–71.

Feder, J.L., Berlocher, S.H., Roethele, J.B., Dambroski, H., Smith, J.J., Perry, W.L., Gavrilovic, V., Filchak, K.E., Rull, J., and Aluja, M. (2003a). Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. Proc Natl Acad Sci U S A *100*, 10314–10319.

Feder, J.L., Roethele, J.B., Filchak, K., Niedbalski, J., and Romero-Severson, J. (2003b). Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, Rhagoletis pomonella. Genetics *163*, 939–953.

Ferreira, M.G., and Cooper, J.P. (2001). The Fission Yeast Taz1 Protein Protects Chromosomes from Ku-Dependent End-to-End Fusions. Molecular Cell *7*, 55–63.

Feuk, L., MacDonald, J.R., Tang, T., Carson, A.R., Li, M., Rao, G., Khaja, R., and Scherer, S.W. (2005). Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet. *1*, e56.

Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., and Louis, E.J. (2000). Chromosomal evolution in Saccharomyces. Nature *405*, 451–454.

Fischer, G., Rocha, E.P.C., Brunet, F., Vergassola, M., and Dujon, B. (2006). Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages. PLoS Genet. *2*, e32.

Flores, M., Morales, L., Gonzaga-Jauregui, C., Domínguez-Vidaña, R., Zepeda, C., Yañez, O., Gutiérrez, M., Lemus, T., Valle, D., Avila, M.C., *et al.* (2007). Recurrent DNA inversion rearrangements in the human genome. Proc. Natl. Acad. Sci. U.S.A. *104*, 6099–6106.

Forsburg, S.L. (2003). Overview of Schizosaccharomyces pombe. Curr Protoc Mol Biol *Chapter 13*, Unit 13.14.

Fox, M.E., and Smith, G.R. (1998). Control of meiotic recombination in Schizosaccharomyces pombe. Prog. Nucleic Acid Res. Mol. Biol. *61*, 345–378.

Fudenberg, G., Getz, G., Meyerson, M., and Mirny, L.A. (2011). High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. Nat. Biotechnol. *29*, 1109–1113.

Funabiki, H., Hagan, I., Uzawa, S., and Yanagida, M. (1993). Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. J. Cell Biol. *121*, 961–976.

Fuscoe, J.C., Setzer, R.W., Collard, D.D., and Moore, M.M. (1996). Quantification of t(14;18) in the lymphocytes of healthy adult humans as a possible biomarker for environmental exposures to carcinogens. Carcinogenesis *17*, 1013–1020.

Garkavtsev, I., and Mizukami, T. (1997). Integrated map of the Schizosaccharomyces pombe genome. Chromosoma *106*, 254–265.

Gerrish, P.J., and Lenski, R.E. (1998). The fate of competing beneficial mutations in an asexual population. Genetica *102-103*, 127–144.

Gillespie, J.H. (1984). Molecular Evolution Over the Mutational Landscape. Evolution *38*, 1116–1129.

Gomes, F.C.O., Pataro, C., Guerra, J.B., Neves, M.J., Corrêa, S.R., Moreira, E.S.A., and Rosa, C.A. (2002). Physiological diversity and trehalose accumulation in Schizosaccharomyces pombe strains isolated from spontaneous fermentations during the production of the artisanal Brazilian cachaça. Can. J. Microbiol. *48*, 399–406.

González, J., Ranz, J.M., and Ruiz, A. (2002). Chromosomal elements evolve at different rates in the Drosophila genome. Genetics *161*, 1137–1154.

Gordo, I., Perfeito, L., and Sousa, A. (2011). Fitness effects of mutations in bacteria. J. Mol. Microbiol. Biotechnol. *21*, 20–35.

Gordon, J.L., Byrne, K.P., and Wolfe, K.H. (2009). Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern Saccharomyces cerevisiae genome. PLoS Genet. *5*, e1000485.

Greig, D., Travisano, M., Louis, E.J., and Borts, R.H. (2003). A role for the mismatch repair system during incipient speciation in Saccharomyces. J. Evol. Biol. *16*, 429–437.

Gresham, D., Desai, M.M., Tucker, C.M., Jenq, H.T., Pai, D.A., Ward, A., DeSevo, C.G., Botstein, D., and Dunham, M.J. (2008). The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast. PLoS Genet *4*, e1000303.

Griffiths, A.J., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. (2000). An Introduction to Genetic Analysis (New York: W.H.Freeman).

Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics *1*, 4.

Haber, J.E. (2006). Transpositions and translocations induced by site-specific double-strand breaks in budding yeast. DNA Repair (Amst.) *5*, 998–1009.

Harewood, L., Schütz, F., Boyle, S., Perry, P., Delorenzi, M., Bickmore, W.A., and Reymond, A. (2010). The Effect of Translocation-Induced Nuclear Reorganization on Gene Expression. Genome Res. *20*, 554–564.

Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation. PLoS Genet *5*, e1000327.

Hegreness, M., Shoresh, N., Hartl, D., and Kishony, R. (2006). An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. Science *311*, 1615–1617.

Heichinger, C., Penkett, C.J., Bähler, J., and Nurse, P. (2006). Genome-wide characterization of fission yeast DNA replication origins. EMBO J. *25*, 5171–5179.

Hillier, L.W., Miller, R.D., Baird, S.E., Chinwalla, A., Fulton, L.A., Koboldt, D.C., and Waterston, R.H. (2007). Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. *5*, e167.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. PLoS Biol *6*, e245.

Hoffmann, A.A., and Rieseberg, L.H. (2008). Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? Annual Review of Ecology, Evolution, and Systematics *39*, 21–42.

Illingworth, C.J.R., and Mustonen, V. (2012). A Method to Infer Positive Selection from Marker Dynamics in an Asexual Population. Bioinformatics.

Imhof, M., and Schlötterer, C. (2001). Fitness Effects of Advantageous Mutations in Evolving Escherichia Coli Populations. PNAS *98*, 1113–1117.

Infante, J.J., Dombek, K.M., Rebordinos, L., Cantoral, J.M., and Young, E.T. (2003). Genome-Wide Amplifications Caused by Chromosomal Rearrangements Play a Major Role in the Adaptive Evolution of Natural Yeast. Genetics *165*, 1745–1759.

Jackson, B.C. (2011). Recombination-suppression: how many mechanisms for chromosomal speciation? Genetica *139*, 393–402.

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., *et al*. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature *484*, 55–61.

Kao, K.C., and Sherlock, G. (2008). Molecular Characterization of Clonal Interference during Adaptive Evolution in Asexual Populations of Saccharomyces cerevisiae. Nat Genet *40*, 1499–1504.

Kassen, R., and Bataillon, T. (2006). Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat. Genet. *38*, 484–488.

Kegel, A., Betts-Lindroos, H., Kanno, T., Jeppsson, K., Ström, L., Katou, Y., Itoh, T., Shirahige, K., and Sjögren, C. (2011). Chromosome length influences replication-induced topological stress. Nature *471*, 392–396.

Kehrer-Sawatzki, H., Sandig, C.A., Goidts, V., and Hameister, H. (2005). Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. Cytogenet. Genome Res. *108*, 91–97.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241–254.

Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988). Four mating-type genes control sexual differentiation in the fission yeast. EMBO J. *7*, 1537–1547.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., *et al*. (2008). Mapping and sequencing of structural variation from eight human genomes. Nature *453*, 56–64.

King, M. (1993). Species Evolution: The Role of Chromosome change (Cambridge University Press).

Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. PLoS Biol *8*, e1000501.

Kirkpatrick, M., and Barton, N. (2006). Chromosome Inversions, Local Adaptation and Speciation. Genetics *173*, 419–434.

Krimbas, C.B., and Powell, J.R. (1992). Drosophila Inversion Polymorphism (CRC Press).

Lahn, B.T., and Page, D.C. (1999). Four Evolutionary Strata on the Human X Chromosome. Science *286*, 964–967.

Lakich, D., Kazazian, H.H., Antonarakis, S.E., and Gitschier, J. (1993). Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nature Genetics *5*, 236–241.

Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation

of gene expression in three dimensions. Nature Reviews Genetics *8*, 104–115.

Larkin, D.M., Pape, G., Donthu, R., Auvil, L., Welge, M., and Lewin, H.A. (2009). Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res.

Lee, H.-Y., Chou, J.-Y., Cheong, L., Chang, N.-H., Yang, S.-Y., and Leu, J.-Y. (2008). Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. Cell *135*, 1065–1073.

Lee, J.A., Carvalho, C.M.B., and Lupski, J.R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell *131*, 1235–1247.

Lenski, R.E., Rose, M.R., Simpson, S.C., and Tadler, S.C. (1991). Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. American Naturalist *138*, 1315–1341.

Lieber, M.R. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. Annual Review of Biochemistry *79*, 181–211.

Limpens, J., Stad, R., Vos, C., de Vlaam, C., de Jong, D., van Ommen, G.J., Schuuring, E., and Kluin, P.M. (1995). Lymphoma-associated translocation t(14;18) in blood B cells of normal individuals. Blood *85*, 2528–2536.

Lindner, P. (1893). Schizosaccharomyces pombe n. sp. neuer Gärungserreger. Wochenschr f Brauerei *10*, 1298 – 1300.

Liti, G., Barton, D.B.H., and Louis, E.J. (2006). Sequence diversity, reproductive isolation and species concepts in Saccharomyces. Genetics *174*, 839–850.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., *et al.* (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337–341.

Liti, G., and Louis, E.J. (2005). Yeast evolution and comparative genomics. Annu. Rev. Microbiol. *59*, 135–153.

Liu, P., Carvalho, C.M., Hastings, P., and Lupski, J.R. (2012). Mechanisms for recurrent and complex human genomic rearrangements. Curr. Opin. Genet. Dev. *22*, 211–220.

Loidl, J., Jin, Q.W., and Jantsch, M. (1998). Meiotic pairing and segregation of translocation quadrivalents in yeast. Chromosoma *107*, 247–254.

Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell *126*, 403–413.

Lowry, D.B., and Willis, J.H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. *8*,.

Lupski, J.R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet. *14*, 417–422.

Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A., *et al.* (1991). DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell *66*, 219–232.

Lupski, J.R., and Stankiewicz, P. (2005). Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes. PLoS Genet *1*, e49.

Lyon, M.F. (2003). Transmission ratio distortion in mice. Annu. Rev. Genet. *37*, 393–408.

Ma, J., Zhang, L., Suh, B.B., Raney, B.J., Burhans, R.C., Kent, W.J., Blanchette, M., Haussler, D., and Miller, W. (2006). Reconstructing contiguous regions of an ancestral genome. Genome Res. *16*, 1557–1565.

MacLean, R.C., and Buckling, A. (2009). The Distribution of Fitness Effects of Beneficial Mutations in Pseudomonas aeruginosa. PLoS Genet *5*, e1000406.

Maheshwari, S., and Barbash, D.A. (2011). The Genetics of Hybrid Incompatibilities. Annual Review of Genetics *45*, 331–355.

Marquès-Bonet, T., Cáceres, M., Bertranpetit, J., Preuss, T.M., Thomas, J.W., and Navarro, A. (2004). Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. Trends Genet. *20*, 524–529.

Matzkin, L.M., Merritt, T.J.S., Zhu, C.-T., and Eanes, W.F. (2005). The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion In(3R)Payne in Drosophila melanogaster. Genetics *170*, 1143–1152.

McGaugh, S.E., and Noor, M.A.F. (2012). Genomic impacts of chromosomal inversions in parapatric Drosophila species. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *367*, 422–429.

Meaburn, K.J., Misteli, T., and Soutoglou, E. (2007). Spatial genome organization in the formation of chromosomal translocations. Semin Cancer Biol *17*, 80–90.

Mieczkowski, P.A., Lemoine, F.J., and Petes, T.D. (2006). Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae. DNA Repair (Amst.) *5*, 1010–1020.

Miller, K.M., Rog, O., and Cooper, J.P. (2006). Semi-conservative DNA replication through telomeres requires Taz1. Nature *440*, 824–828.

Mitelman, F., Johansson, B., and Mertens, F. (2004). Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nat. Genet. *36*, 331–334.

Molnar, M., and Kleckner, N. (2008). Examination of interchromosomal interactions in vegetatively growing diploid Schizosaccharomyces pombe cells by Cre/loxP site-specific recombination. Genetics *178*, 99–112.

Morales-Hojas, R., Päällysaho, S., Vieira, C., Hoikkala, A., and Vieira, J. (2007). Comparative polytene chromosome maps of D. montana and D. virilis. Chromosoma *116*, 21–27.

Moreno, S., Klar, A., and Nurse, P. (1991). Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. Meth. Enzymol. *194*, 795–823.

Moser, B.A., and Nakamura, T.M. (2009). Protection and replication of telomeres in fission yeast. Biochem. Cell Biol. *87*, 747–758.

Muñoz, A., and Sankoff, D. (2012). Detection of gene expression changes at chromosomal rearrangement breakpoints in evolution. BMC Bioinformatics *13*, S6.

Munz, P. (1994). An Analysis of Interference in the Fission Yeast Schizosaccharomyces Pombe. Genetics *137*, 701–707.

Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L., *et al.* (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science *309*, 613–617.

Murray, A.W., Schultes, N.P., and Szostak, J.W. (1986). Chromosome length controls mitotic chromosome segregation in yeast. Cell *45*, 529–536.

Nadeau, J.H., and Taylor, B.A. (1984). Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci U S A *81*, 814–818.

Nambiar, M., and Raghavan, S.C. (2011). How does DNA break during chromosomal translocations? Nucleic Acids Res. *39*, 5813–5825.

Naumov, G.I. (1987). Genetic basis for classification and identification of the ascomycetous yeasts. Studies in Mycology *30*, 469–475.

Navarro, A., and Barton, N.H. (2003). Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. Evolution *57*, 447–459.

Navarro, A., Betran, E., Barbadilla, A., and Ruiz, A. (1997). Recombination and Gene Flux Caused by Gene Conversion and Crossing over in Inversion Heterokaryotypes. Genetics *146*, 695–709.

Nielsen, O., and Egel, R. (2007). Sex in Fungi: Molecular Determination and Evolutionary Implications. Chapter 8. The mat genes of Schizosaccharomyces pombe: expression, homothallic switch, and silencing. (ASM Press).

Niwa, O., Shimanuki, M., and Miki, F. (2000). Telomere-led bouquet formation facilitates homologous chromosome pairing and restricts ectopic interaction in fission yeast meiosis. EMBO J. *19*, 3831–3840.

Niwa, O., and Yanagida, M. (1985). Triploid meiosis and aneuploidy in Schizosaccharomyces pombe: an unstable aneuploid disomic for chromosome III. Current Genetics *9*, 463–470.

Noor, M.A.F., Grams, K.L., Bertucci, L.A., and Reiland, J. (2001). Chromosomal Inversions and the Reproductive Isolation of Species. PNAS *98*, 12084–12088.

Nowell, P.C., and Hungerford, D.A. (1960). A minute chromosome in human chronic granulocytic leukemia. Science *132*, 1497.

Nurse, P. (2000). A long twentieth century of the cell cycle and beyond. Cell *100*, 71–78.

Olsson, I., and Bjerling, P. (2011). Advancing our understanding of functional genome organisation through studies in the fission yeast. Curr Genet *57*, 1–12.

Orr, H.A. (2002). The population genetics of adaptation: the adaptation of DNA sequences. Evolution *56*, 1317–1330.

Orr, H.A. (2003). The Distribution of Fitness Effects Among Beneficial Mutations. Genetics *163*, 1519–1526.

Orr, H.A. (2005). The genetic theory of adaptation: a brief history. Nat. Rev. Genet. *6*, 119–127.

Ou, Z., Stankiewicz, P., Xia, Z., Breman, A.M., Dawson, B., Wiszniewska, J., Szafranski, P., Cooper, M.L., Rao, M., Shao, L., *et al*. (2011). Observation and prediction of recurrent human translocations mediated by NAHR between nonhomologous chromosomes. Genome Res *21*, 33–46.

Parada, L., and Misteli, T. (2002). Chromosome positioning in the interphase nucleus. Trends Cell Biol. *12*, 425–432.

Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I., and Lupski, J.R. (1992). Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. Nat. Genet. *2*, 292–300.

Pérez-Ortín, J.E., Querol, A., Puig, S., and Barrio, E. (2002). Molecular Characterization of a Chromosomal Rearrangement Involved in the Adaptive Evolution of Yeast Strains. Genome Res *12*, 1533–1539.

Perfeito, L., Fernandes, L., Mota, C., and Gordo, I. (2007). Adaptive Mutations in Bacteria: High Rate and Small Effects. Science *317*, 813–815.

Perkins, D.D. (1997). Chromosome rearrangements in Neurospora and other filamentous fungi. Adv. Genet. *36*, 239–398.

Pevzner, P., and Tesler, G. (2003). Human and Mouse Genomic Sequences Reveal Extensive Breakpoint Reuse in Mammalian Evolution. PNAS *100*, 7672–7677.

Pidoux, A.L., and Allshire, R.C. (2004). Kinetochore and heterochromatin domains of the fission yeast centromere. Chromosome Res. *12*, 521–534.

Poláková, S., Blume, C., Zárate, J.Á., Mentel, M., Jørck-Ramberg, D., Stenderup, J., and Piškur, J. (2009). Formation of new chromosomes as a virulence mechanism in yeast Candida glabrata. Proc Natl Acad Sci U S A *106*, 2688–2693.

Pombi, M., Caputo, B., Simard, F., Di Deco, M.A., Coluzzi, M., Torre, A. della, Costantini, C., Besansky, N.J., and Petrarca, V. (2008). Chromosomal plasticity and evolutionary potential in the malaria vector Anopheles gambiae sensu stricto: insights from three decades of rare paracentric inversions. BMC Evol Biol *8*, 309.

Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. Nature Reviews Genetics *11*, 175–180.

Querol, A., Fernández-Espinar, M.T., del Olmo, M. lí, and Barrio, E. (2003). Adaptive evolution of wine yeast. Int. J. Food Microbiol. *86*, 3–10.

Rachidi, N., Barre, P., and Blondin, B. (1999). Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of Saccharomyces cerevisiae. Mol. Gen. Genet. *261*, 841–850.

Ranz, J.M., Casals, F., and Ruiz, A. (2001). How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus Drosophila. Genome Res. *11*, 230–239.

Ranz, J.M., Maurin, D., Chan, Y.S., von Grotthuss, M., Hillier, L.W., Roote, J., Ashburner, M., and Bergman, C.M. (2007). Principles of Genome Evolution in the Drosophila melanogaster Species Group. PLoS Biol *5*, e152.

Rhind, N., Chen, Z., Yassour, M., Thompson, D.A., Haas, B.J., Habib, N., Wapinski, I., Roy, S., Lin, M.F., Heiman, D.I., *et al.* (2011). Comparative functional genomics of the fission yeasts. Science *332*, 930–936.

Richards, S., Liu, Y., Bettencourt, B.R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M.J., Chen, R., Meisel, R.P., *et al.* (2005). Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res. *15*, 1–18.

Rieseberg, L.H. (2001). Chromosomal rearrangements and speciation. Trends Ecol. Evol. (Amst.) *16*, 351–358.

Rieseberg, L.H., Whitton, J., and Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics *152*, 713–727.

Rokyta, D.R., Beisel, C.J., Joyce, P., Ferris, M.T., Burch, C.L., and Wichman, H.A. (2008). Beneficial Fitness Effects Are Not Exponential for Two Viruses. J Mol Evol *67*, 368–376.

Rowley, J.D. (2001). Chromosome translocations: dangerous liaisons revisited. Nat. Rev. Cancer *1*, 245–250.

Rozas, J., and Rozas, R. (1999). DnaSP Version 3: An Integrated Program for Molecular Population Genetics and Molecular Evolution Analysis. Bioinformatics *15*, 174–175.

Sanjuán, R., Moya, A., and Elena, S.F. (2004). The Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in an RNA Virus. PNAS *101*, 8396–8401.

Schaeffer, S.W., Goetting-Minesky, M.P., Kovacevic, M., Peoples, J.R., Graybill, J.L., Miller, J.M., Kim, K., Nelson, J.G., and Anderson, W.W. (2003). Evolutionary genomics of inversions in Drosophila pseudoobscura: evidence for epistasis. Proc. Natl. Acad. Sci. U.S.A. *100*, 8319–8324.

Selmecki, A., Forche, A., and Berman, J. (2006). Aneuploidy and isochromosome formation in drug-resistant Candida albicans. Science *313*, 367–370.

Selmecki, A.M., Dulmage, K., Cowen, L.E., Anderson, J.B., and Berman, J. (2009). Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. PLoS Genet. *5*, e1000705.

Sharakhov, I.V., White, B.J., Sharakhova, M.V., Kayondo, J., Lobo, N.F., Santolamazza, F., Della Torre, A., Simard, F., Collins, F.H., and Besansky, N.J. (2006). Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the Anopheles gambiae complex. Proc. Natl. Acad. Sci. U.S.A. *103*, 6258–6262.

Snaith, H.A., Samejima, I., and Sawin, K.E. (2005). Multistep and multimode cortical anchoring of tea1p at cell tips in fission yeast. EMBO J. *24*, 3690–3699.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature *435*, 637–645.

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G., *et al.* (2005). A common inversion under selection in Europeans. Nat. Genet. *37*, 129–137.

Steiner, N.C., Hahnenberger, K.M., and Clarke, L. (1993). Centromeres of the fission yeast Schizosaccharomyces pombe are highly variable genetic loci. Mol Cell Biol *13*, 4578–4587.

Stevison, L.S., Hoehn, K.B., and Noor, M.A.F. (2011). Effects of inversions on within- and between-species recombination and divergence. Genome Biol Evol *3*, 830–841.

Strasburg, J.L., Scotti-Saintagne, C., Scotti, I., Lai, Z., and Rieseberg, L.H. (2009). Genomic Patterns of Adaptive Divergence between Chromosomally Differentiated Sunflower Species. Mol Biol Evol *26*, 1341–1355.

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. (1992). A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. Mol Biol Cell *3*, 819–835.

Takizawa, T., Meaburn, K.J., and Misteli, T. (2008). The meaning of gene positioning. Cell *135*, 9–13.

Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., *et al.* (2005). Fine-scale structural variation of the human genome. Nature Genetics *37*, 727–732.

de Visser, J.A.G.M., and Rozen, D.E. (2006). Clonal Interference and the Periodic Selection of New Beneficial Mutations in Escherichia coli. Genetics *172*, 2093–2100.

Wells, J.L., Pryce, D.W., and McFarlane, R.J. (2006). Homologous chromosome pairing in Schizosaccharomyces pombe. Yeast *23*, 977–989.

White, M.J.D. (1969). Chromosomal Rearrangements and Speciation in Animals. Annual Review of Genetics *3*, 75–98.

White, M.J.D. (1977). Animal Cytology and Evolution (CUP Archive).

Wijchers, P.J., and de Laat, W. (2011). Genome organization influences partner selection for chromosomal rearrangements. Trends Genet. *27*, 63–71.

Williams, G.C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. Evolution *11*, 398.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., *et al.* (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871–880.

Yamamoto, A., and Hiraoka, Y. (2001). How do meiotic chromosomes meet their homologous partners?: lessons from fission yeast. Bioessays *23*, 526–533.

Zhao, H., and Bourque, G. (2009). Recovering genome rearrangements in the mammalian phylogeny. Genome Res *19*, 934–942.

Appendices

Appendix A – *SfiI*-PFGE of chromosomes from Natural Isolates of *S. pombe*.

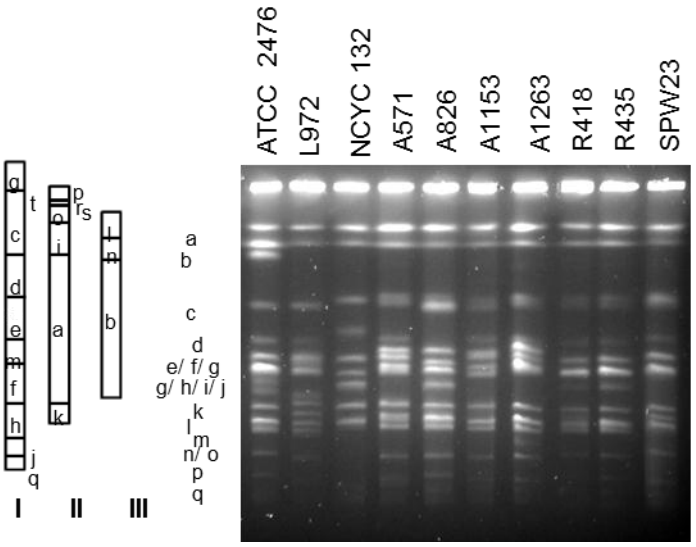


Figure. A1. *SfiI* restriction PFGE analysis of Natural Isolates from *S. pombe*.

Appendix B- Tree representation of nucleotide diversity.

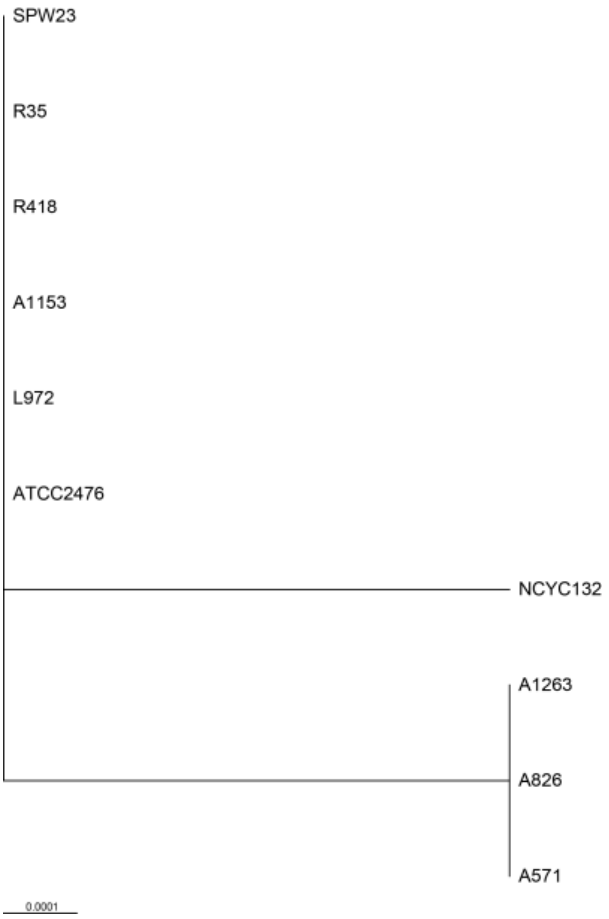


Figure B1. Tree representation for *his3*$^{+}$ alignment.

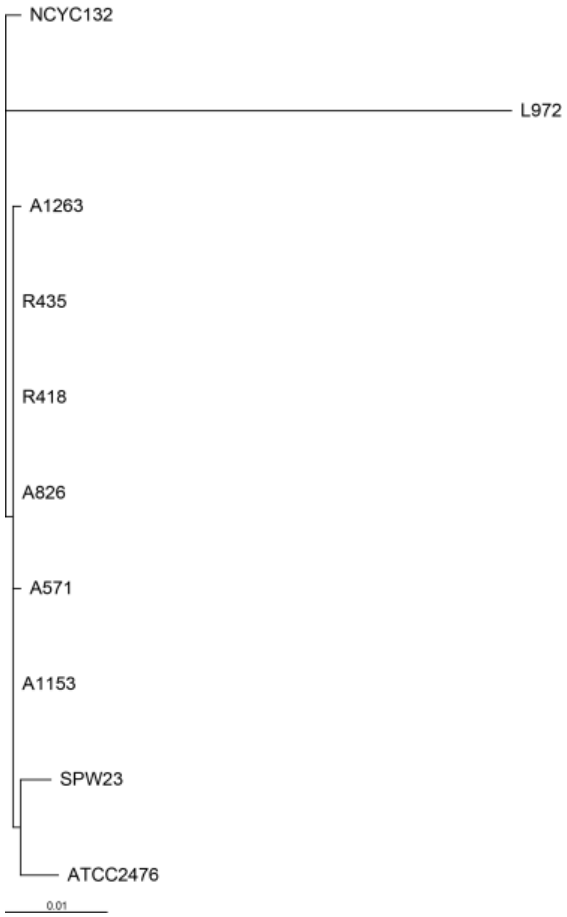


Figure B2. Tree representation for cen2 alignment.

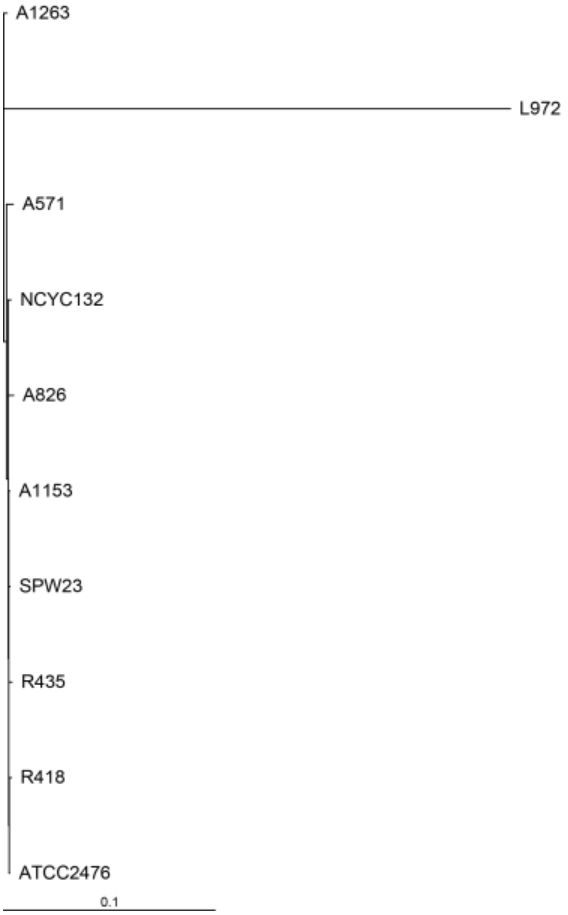


Figure B3. Tree representation for *cox1*$^{+}$ alignment.

Appendix C – Tetrad analysis of Natural Isolates of *S. pombe*.

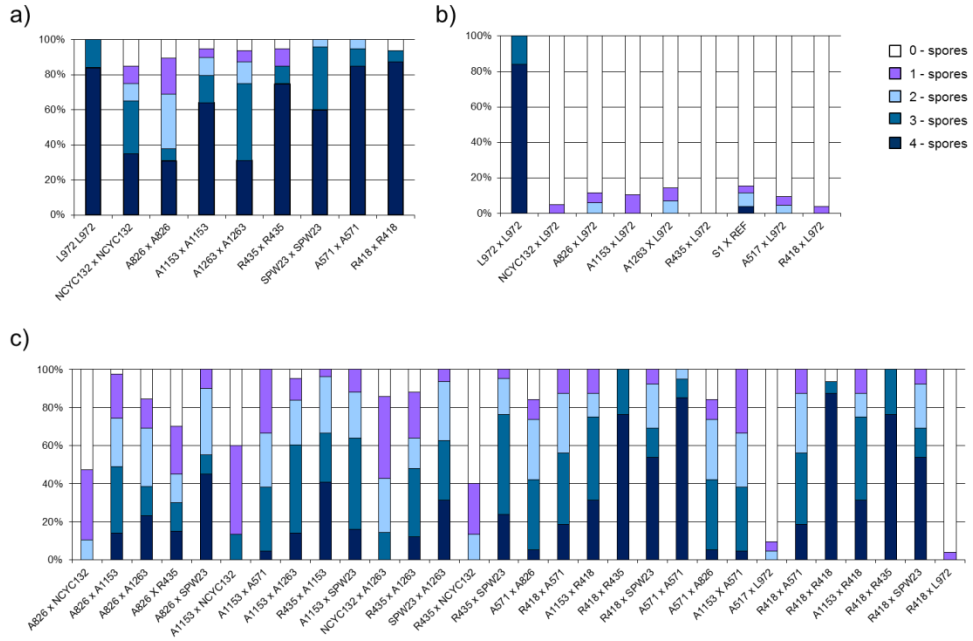


Figure C1. Tetrad numbers distribution. Distribution of the number of tetrads according to the number of spores that formed a colony. **a**, homozygotic crosses. **b**, heterozygotic crosses with L972 (reference strain). **c**, heterozygotic crosses between all the different strains.

Appendix D – Growth rate of genetically engineered CRs.

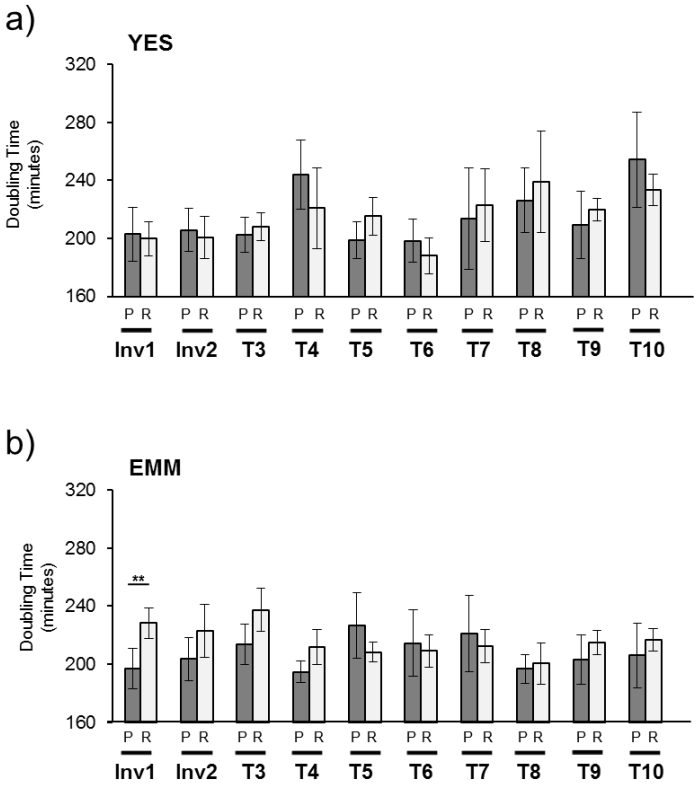


Figure. D1. Growth-rate of Parental (P) and Rearranged (R) strains in rich-media (YES) and minimal media (EMM). Error bars represent 2xS.E. for n>6 independent experiments (* for p< 0.05, ** for p< 0.01 according to Mann-Whitney U-test).

Appendix E – Calibration of growth conditions used for competition assays.

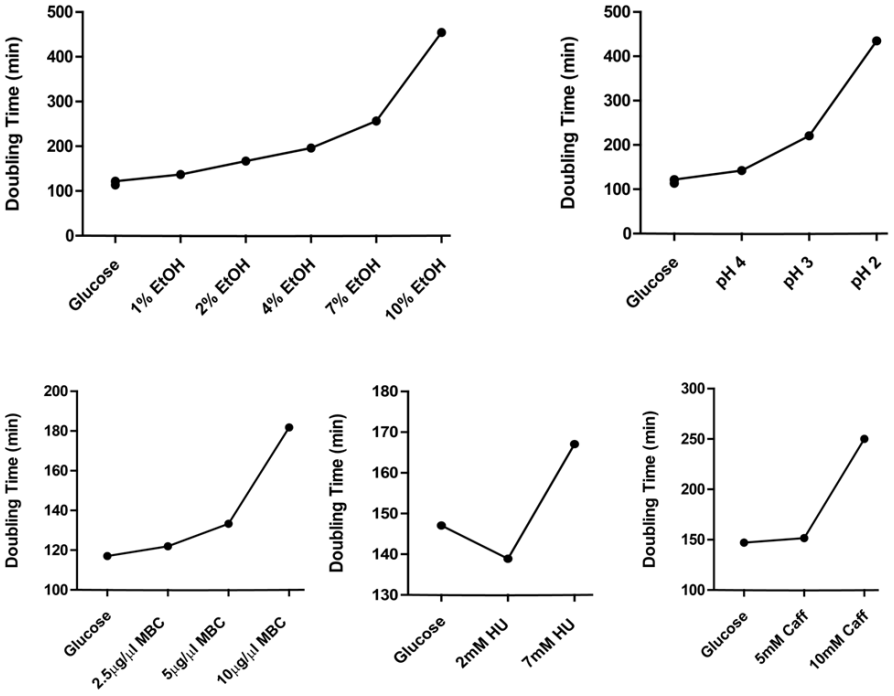
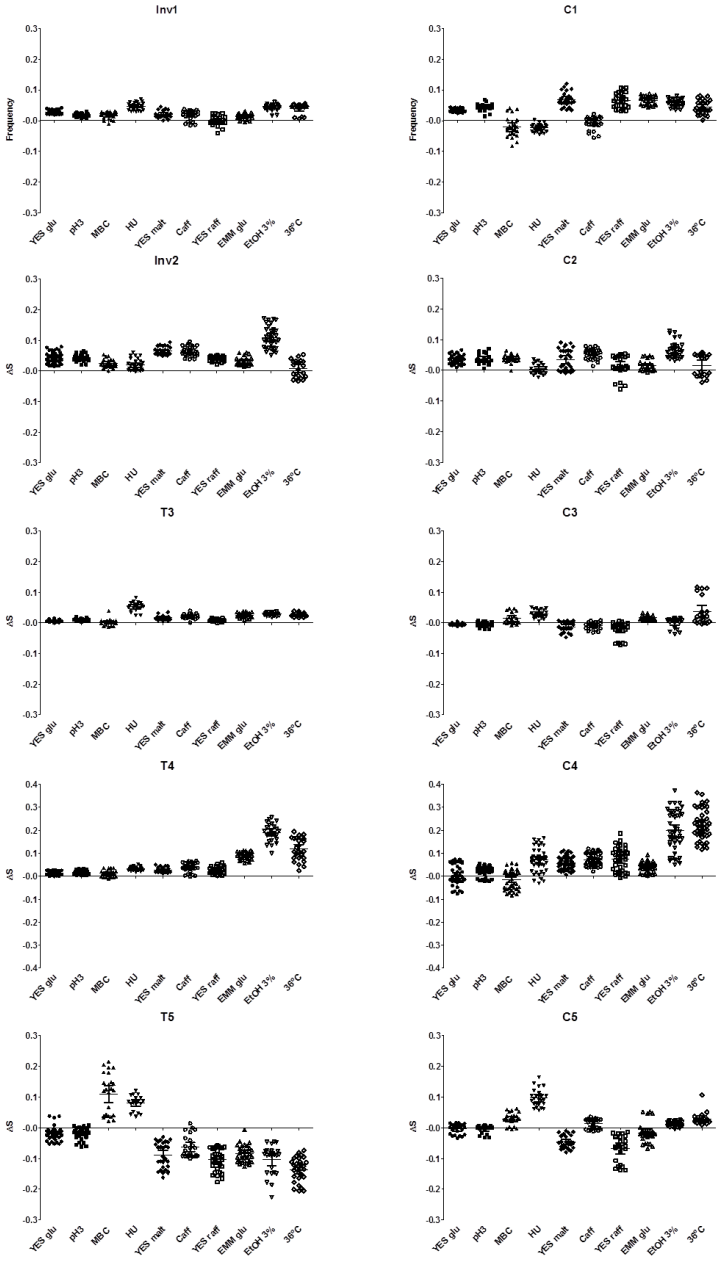


Figure E1. Average doubling times for growth of auxotrophic strain in the respective test media.

Appendix F – Mitotic fitness analysis of genetically engineered CRs.



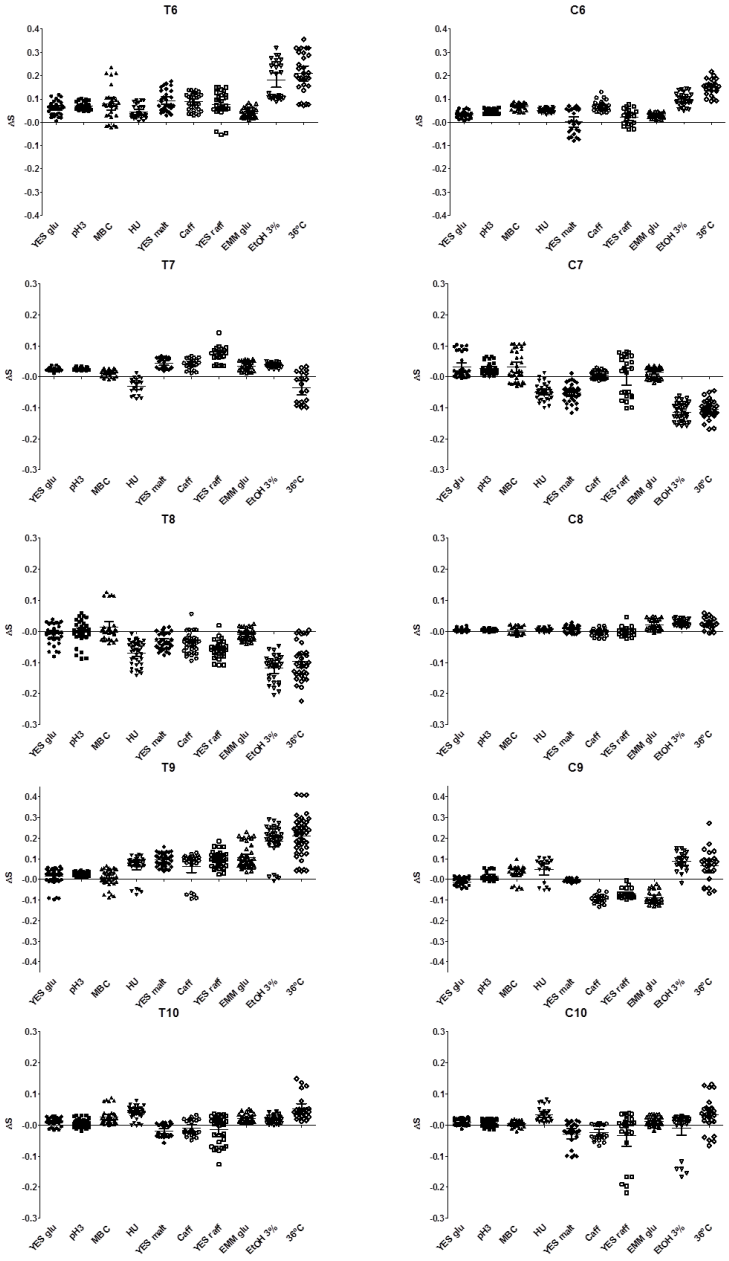


Figure F1. Delta-S (*ΔS*) values for independent competitions of rearranged and control versus the GFP-labelled control. Error bars represent mean with 95% C.I.

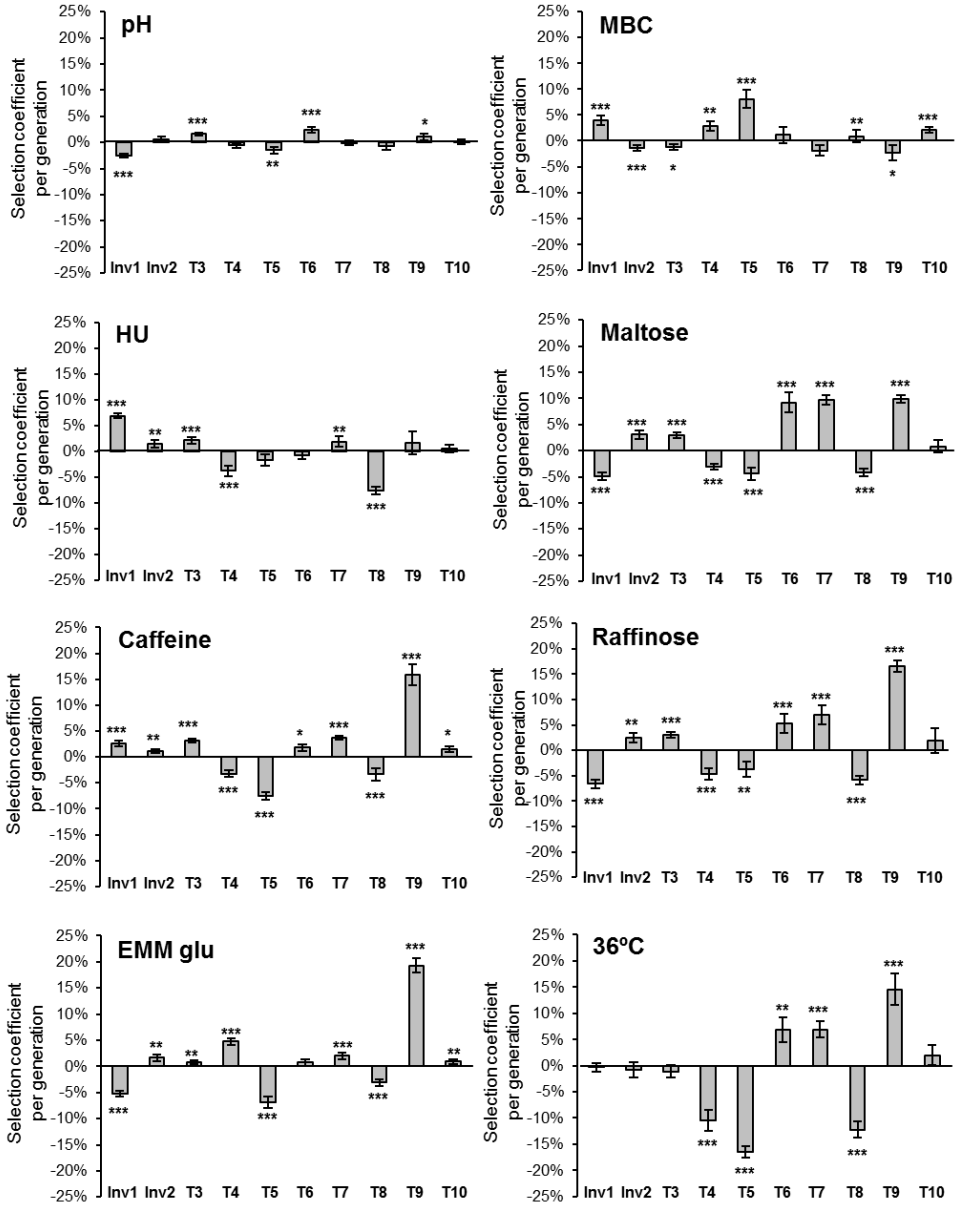


Figure F2. Average selection coefficients (*ΔS*) values for all environment competitions. Averages were calculated from more than 10 independent experiments each with two independent clones. Error bars represent 2 x S.E.. (* for p< 0.05, ** for p< 0.01 *** for p< 0.001 according to Mann-Whitney U-test).

Appendix G – Gene expression analysis.

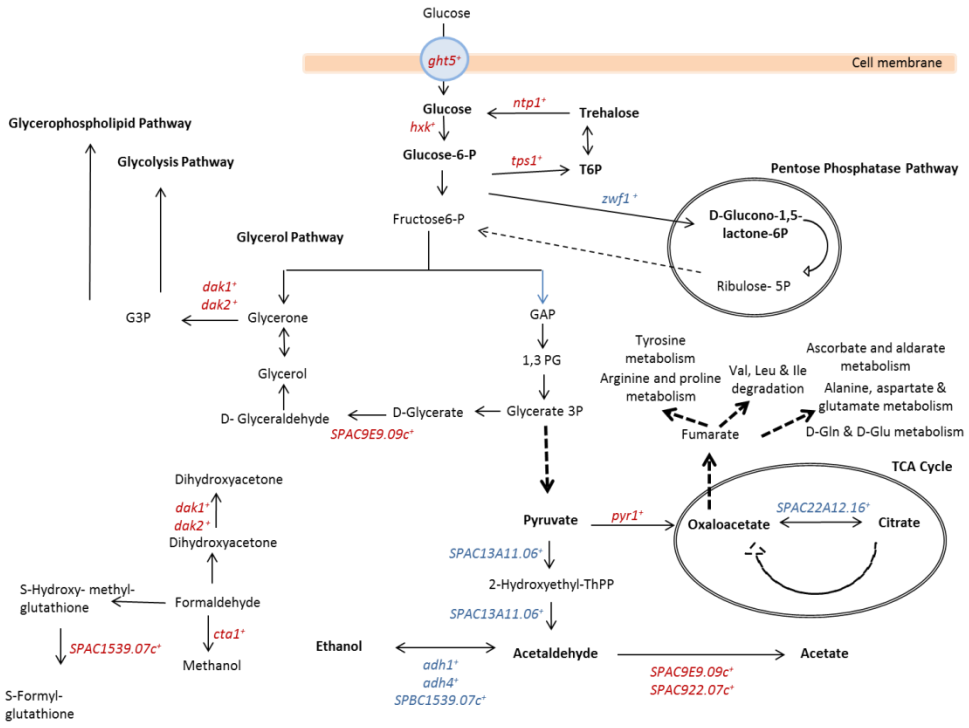


Figure G1. Metabolic pathways for the genes analysed. Gene names are highlighted in blue (repressed) and in red (over-expressed) according to their expression state upon ESR (Alexandre *et al*., 2001; Chen *et al*., 2003)

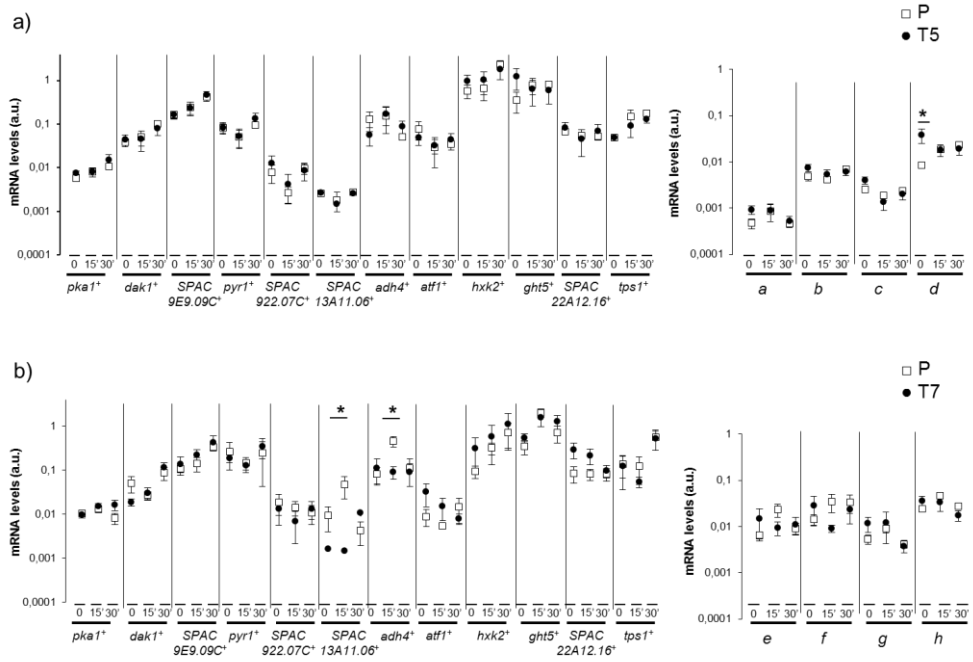


Figure G2. mRNA levels for genes involved in stress and ethanol consumption. **a,** Parental and translocation T5**. b,** Parental and translocation T7. mRNA was extracted from exponentially growing cultures without (T0) and with 15 and 30 minutes (T_{15} and T_{30}) exposure to 3%EtOH. Error bars represent S.E. for n between 3 to 6 independent experiments. (* for $p < 0.05$ according to Mann-Whitney U-test).

Appendix H – Beneficial mutation rate in *S. pombe.*

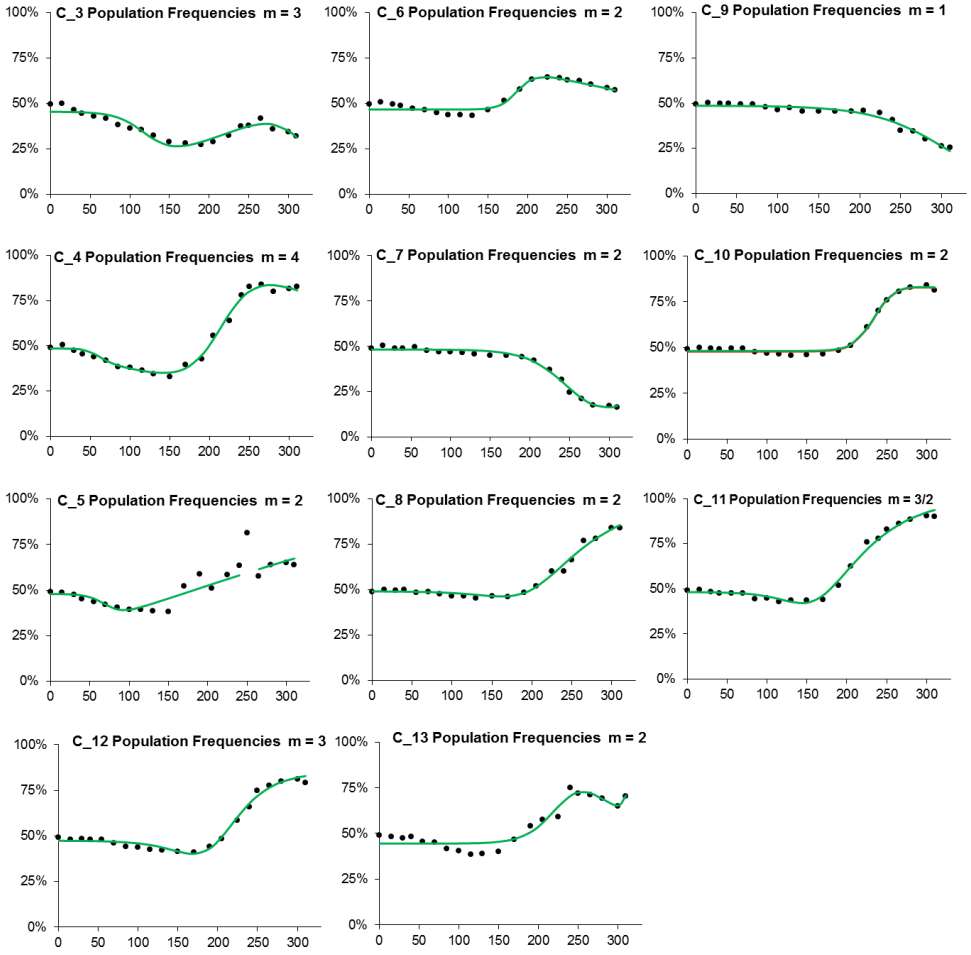


Figure H1. Experimental and model frequencies for Control (C) populations.

Black dots, experimental data; green line, model.

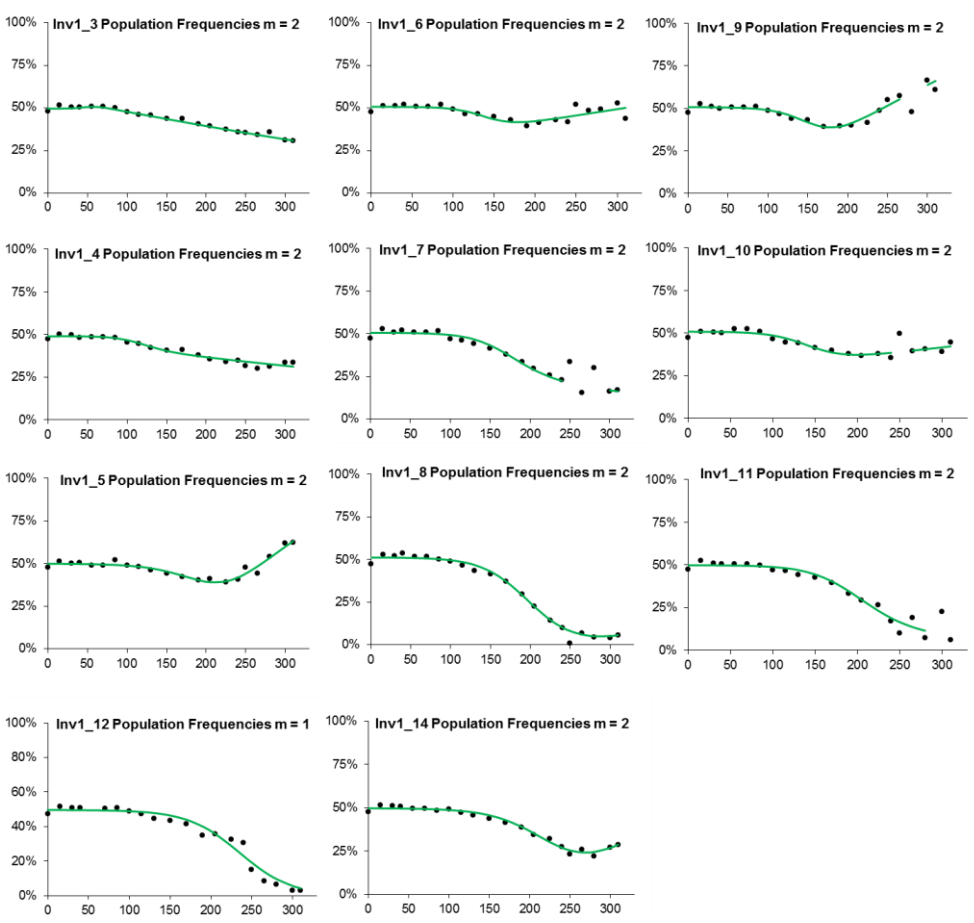


Figure H2. Experimental and model frequencies for Inverted (Inv1) populations.
Black dots, experimental data; green line, model.