


# *Using chromosome engineering on a natural isolate of the fission yeast, Schizosaccharomyces pombe, to investigate the epigenetic inheritance of the kinetochore*

Submitted in fulfilment of the conditions for the award of the degree **PhD Genetics**

**Anne Caroline Barbosa**
**Student ID 4201364**

**Supervised by William Brown**

School of Life Sciences
University of Nottingham

# CONTENTS

**Contents** ..... **i**

**List of Figures** ..... **v**

**List of Tables** ..... **ix**

**Abbreviations** ..... **x**

**Abstract** ..... **1**

**1. Introduction** ..... **2**

*1.1. Schizosaccharomyces pombe* ..... *3*

1.1.1. Phylogeny ..... 3

1.1.2. Model organism ..... 4

1.1.3. The strain CBS 2777 ..... 5

*1.2. Fission yeast cell cycle and regulation* ..... *7*

*1.3. Chromosome Engineering using recombinases* ..... *10*

1.3.1. Site-specific recombination ..... 10

1.3.2. Tyrosine Recombinases ..... 10

1.3.3. Serine Recombinases ..... 12

1.3.4. *Schizosaccharomyces pombe* plasmids ..... 14

1.3.5. Selectable markers ..... 14

1.3.6. Plasmid systems ..... 15

1.3.10. nmt1 promoter ..... 16

1.3.11. The pREP and pRIP systems ..... 16

*1.4. The Centromere* ..... *16*

1.4.1. Centromeric DNA ..... 17

1.4.5. The centromere and kinetochore proteins ..... 21

1.4.6. Kinetochore assembly in *S. pombe* ..... 24

1.4.7. Neo-centromere formation ..... 26

*1.5. Epigenetics and the centromere* ..... *27*

1.5.1. Chromatin-based epigenetics ..... 28

1.5.2. Non-coding RNAs ..... 31

1.5.3. Prions ..... 32

1.5.4. Paramutation ..... 32

1.5.5. X chromosome inactivation ..... 32
1.5.6. Silencing of the pericentromeric heterochromatin in the *Schizosaccharomyces pombe* ..... 34
1.5.7. Epigenetic inheritance of the centromere ..... 36
*1.6. Aims* ..... *38*
**2. Materials and methods ..... 39**
*2.1. Chemicals and enzymes* ..... *39*
*2.2. Fission yeast media* ..... *39*
2.2.1. Stock solutions (Forsburg, 2015) ..... 39
2.2.2. Non-selective media (Hagan, 2016, Forsburg, 2015) ..... 39
*2.3. S. pombe growth and storage* ..... *41*
2.3.1. Growth conditions ..... 41
2.3.2. Short term storage ..... 41
2.3.3. Long term storage ..... 41
*2.4. Molecular genetic methods* ..... *42*
2.4.1. High efficiency transformation of *S. pombe* ..... 42
2.4.2. Genomic DNA Extraction from *S. pombe* ..... 43
2.4.3. Gel Eletrophoresis ..... 43
2.4.4. Restriction digestion of DNA ..... 44
2.4.5. DNA extraction from agarose gels ..... 44
2.4.6. Ethanol precipitation of DNA ..... 44
2.4.7. Polymerase Chain Reaction (PCR) ..... 44
2.4.11. List of Primers ..... 46
2.4.12. PCR product sequencing ..... 47
2.4.13. Southern Blot (Southern, 1975) ..... 48
2.4.14. Random primed labeling and filter hybridization ..... 48
2.4.15. ChIP sequencing ..... 49
2.4.22. Computational analysis of ChIP sequencing data ..... 52
*2.5. Strains used or built for this project* ..... *53*
**3. Recombining mutant centromeric DNA fragments in the *S. pombe* CBS 2777 chromosome 2 using the serine recombinase from the *Streptomyces* bacteriophage ϕC31 ..... 61**
*3.1. Introduction* ..... *61*
*3.2. Integration of test sequences on chromosome 2* ..... *63*
3.2.1. Screening for site-specific recombination by PCR ..... 66

3.2.2. Sequencing the left end of the recombination site ..... 69
3.2.3. Test sequence integrity analyses ..... 70
*3.3. Targeting the left end of the centromere of chromosome 2 with a Bxb1 attB site and ura4+ gene 73*
*3.4. Integration of the mutant centromeric DNA fragments on the ade6$^{+}$ locus..... 76*
3.4.1. Confirming for the site-specific recombination PCR ..... 77
3.4.2. Test sequence integrity analyses ..... 80
*3.5. Discussion..... 81*
**4. The rate of neo-centromere formation following native centromere excision with the Bxb1 Integrase ..... 83**
*4.1. Introduction ..... 83*
*4.2. Excision of the native centromere using Bxb1 integrase..... 83*
*4.3. Average colony size and cells viability of the transformed clones ..... 85*
*4.4. Screening the ura4Δ clones..... 88*
*4.5. Southern blot analyses of the candidate swapped clones ..... 91*
*4.6. Measuring the proportion of ura4Δ clones that had centromere 2 excised ..... 97*
*4.7. Rate of centromere formation ..... 101*
*4.8. Discussion..... 104*
**5. Cnp1 binding to engineered centromeres by ChIP-seq. ..... 106**
*5.1. Introduction ..... 106*
*5.2. Analyzed clones..... 107*
*5.3. Cnp1 spreading over the test sequences..... 108*
*5.4. Centromere and neo-centromere formation after excision of the native centromere 110*
*5.5. Proportion Cnp1 binding on chromosome 2 ..... 115*
*5.6. Cnp1 binding on centromeric sequences at non-centromeric location..... 116*
*5.7. swi6 mutants..... 117*
5.7.1. Native centromere 2 excision in Δ*swi6* strains ..... 119
5.7.2. Screening the *ura4Δ* clones ..... 120
5.7.3. Measuring the proportion of *ura4*Δ clones that had centromere 2 excised 120
5.7.4. Karyotype analysis of *cen2*Δ survivors of the Not373_control$^{\Delta swi6}$ clones... 122

*5.8. Discussion ..... 123*

**6. Conclusions ..... 126**

*6.1. Further work ..... 128*

**7. References ..... 129**

**8. Appendix I ..... 148**

**9. Appendix II ..... 150**

**10. Appendix III ..... 151**

# LIST OF FIGURES

FIGURE 1.1: CONSENSUS PHYLOGENY OF FISSION YEASTS. NEIGHBOR-JOINING PHYLOGENETIC TREE BASED ON SEQUENCES OF THE D1/D2 DIVERGENT DOMAINS OF THE LSU RRNA GENE. ..... 3

FIGURE 1.2: GENOME ORGANISATION OF *S. POMBE*. ..... 6

FIGURE 1.3: CELL CYCLE OF THE FISSION YEAST *S. POMBE*. ..... 7

FIGURE 1.4: A MOLECULAR MECHANISM FOR THE REGULATION OF CDC13-ASSOCIATED KINASE ACTIVITY IN FISSION YEAST. ..... 8

FIGURE 1.5: SCHEME OF THE SITE-SPECIFIC RECOMBINATION PROMOTED BY TYROSINE RECOMBINASE FAMILY EXEMPLIFIED THROUGH THE CRE-LOXP SYSTEM SHOWING THE HOLLIDAY-JUNCTION (HJ). ..... 11

FIGURE 1.6: SCHEMATIC DESCRIPTION OF SITE-SPECIFIC RECOMBINATION BETWEEN THE BACTERIOPHAGE ATTACHMENT SITE *ATTP* IN THE PLASMID AND THE BACTERIAL ATTACHMENT SITE IN THE *ATTB* GENOME. ..... 12

FIGURE 1.7: A GENERIC SCHEME FOR THE SERINE RECOMBINASE REACTION MECHANISM. ..... 13

FIGURE 1.8: SEQUENCES OF THE ATTACHMENT SITES FOR THE *STREPTOMYCES* PHAGE ϕC31. ..... 14

FIGURE 1.9: GENERAL STRUCTURE OF THE SACCHAROMYCES CEREVISIAE CENTROMERE ..... 18

FIGURE 1.10: DIAGRAM OF THE GENERAL ORGANISATION OF THE CENTROMERIC REGIONS OF THE CHROMOSOMES 1, 2 AND 3 OF *S. POMBE*. ..... 19

FIGURE 1.11: HUMAN CENTROMERE ORGANISATION OF α-SATELLITES AND HIGHER-ORDER REPEATS. ..... 20

FIGURE 1.12: CENH3-CONTAINING CHROMATIN IN DIFFERENT ORGANISMS. ..... 21

FIGURE 1.13: ORGANIZATION OF KINETOCHORES. ..... 22

FIGURE 1.14: THE RECRUITMENT OF PROTEINS DURING KINETOCHORE ASSEMBLY IN *SCHIZOSACCHAROMYCES POMBE*. . 26

FIGURE 3.1: ANALYSIS OF CENP-A CHROMATIN IN HACS. ..... 61

FIGURE 3.2: CENTROMERE OF THE CHROMOSOME 2 OF THE *S. POMBE* STRAIN CBS 2777. ..... 63

FIGURE 3.3: DNA PLASMID NL314, A ϕC31 INTEGRASE EXPRESSION PLASMID. ..... 63

FIGURE 3.4: DNA PLASMIDS USED FOR THE SITE-SPECIFIC RECOMBINATION ..... 64

FIGURE 3.5: TEST SEQUENCES USED TO BUILD THE PLASMIDS FOR SITE-SPECIFIC RECOMBINATION. ..... 65

FIGURE 3.6: SCREENING *URA4Δ* CLONES. ..... 66

FIGURE 3.7: MAP OF THE PRIMERS USED FOR THE SITE-SPECIFIC RECOMBINATION SCREENING. ..... 67

FIGURE 3.8: SCREENING THE LEFT END OF THE RECOMBINATION SITE FOR THE NOTT373 CLONES BY PCR. ..... 68

FIGURE 3.9: SUMMARY GEL ELECTROPHORESIS CONTAINING ONE CLONE OF EACH RECOMBINED STRAIN AND THE START STRAIN AS A CONTROL. ..... 68

FIGURE 3.10: GEL ELECTROPHORESIS OF THE PCR FOR SITE-SPECIFIC RECOMBINATION SCREENING IN THE RIGHT END OF THE RECOMBINATION SITE FOR THE NOTT373 CLONES USING PRIMERS 779 AND 646. ..... 69

FIGURE 3.11: TRACE AND MAP OF THE *ATTL* PCR PRODUCT SEQUENCING. ..... 69

FIGURE 3.12: RESTRICTION ENZYME SITE MAPS OF THE RECOMBINED STRAINS. ..... 70

FIGURE 3.13: FILTER HYBRIDIZATION ANALYSIS OF THE FRAGMENTS FROM THE ENZYMATIC DIGESTION WITH *BAMH*I AND *PST*I FOR THE NOTT373 RECOMBINED STRAINS. ..... 71

FIGURE 3.14: FILTER HYBRIDIZATION ANALYSIS OF THE FRAGMENTS FROM THE ENZYMATIC DIGESTION *ECORI* FOR THE CLONES WITH THE ALPHOID FRAGMENT INSERTION AND THE NOTT373 START STRAIN AS CONTROL; HYBRIDISED WITH A PROBE FOR THE ALPHOID DNA FRAGMENT. ..... 72
FIGURE 3.15: TARGETING THE LEFT-HAND SIDE OF THE ENDOGENOUS CENTROMERE WITH *URA4+* GENE AND BXB1 *ATTB* SITE. THE TARGETED CLONES WERE SCREENED BY PCR USING THE PRIMERS 893 AND 884. ..... 73
FIGURE 3.16: GEL ELECTROPHORESIS OF THE PCR SCREENING OF THE CLONES TARGETED ON THE LEFT OF THE ENDOGENOUS CENTROMERE. ..... 74
FIGURE 3.17: LONG RANGE PCR TO SCREEN THE CLONES TARGETED ON CHROMOSOME 2. ..... 75
FIGURE 3.18: SOUTHERN BLOT ANALYSIS OF THE FRAGMENTS FROM THE ENZYMATIC DIGESTION *SALI* FOR THE NOTT373 CLONES TARGETED ON THE LEFT OF THE ENDOGENOUS CENTROMERE AND HYBRIDISED WITH A PROBE FOR THE *URA4+*. ..... 76
FIGURE 3.19: MAP OF THE PRIMERS USED FOR THE SITE-SPECIFIC RECOMBINATION SCREENING ON NOTT376. ..... 78
FIGURE 3.20: SCREENING THE RIGHT END OF THE RECOMBINATION SITE FOR THE NOTT376 CLONES BY PCR. ..... 79
FIGURE 3.21: SCREENING THE LEFT END OF THE RECOMBINATION SITE FOR THE NOTT376 CLONES BY PCR. ..... 79
FIGURE 3.22: SUMMARY GEL ELECTROPHORESIS CONTAINING ONE CLONE OF EACH RECOMBINED STRAIN AND THE START STRAIN AS A CONTROL. ..... 80
FIGURE 3.23: FILTER HYBRIDIZATION ANALYSIS OF THE FRAGMENTS FROM THE ENZYMATIC DIGESTION WITH *BAMHI* AND *PSTI* FOR THE NOTT376 RECOMBINED STRAINS. ..... 81
FIGURE 4.1: GENERAL REPRESENTATION OF THE STRATEGY TO PROMOTE THE CENTROMERE 2 EXCISION IN THE NOTT373 RECOMBINED STRAINS. ..... 84
FIGURE 4.2: PLATES OF PMG AGAR SUPPLEMENTED WITH URACIL SHOWING THE COLONIES SIZE OF THE TRANSFORMED CLONES. ..... 84
FIGURE 4.3: A GENERAL STRATEGY TO MEASURE THE AVERAGE COLONY SIZES AND THE CELLS VIABILITY. ..... 85
FIGURE 4.4: AVERAGE NUMBER OF CELLS IN A COLONY IN EACH STRAIN TRANSFORMED WITH REP81 BXB1 INTEGRASE PLASMID AS A PROPORTION OF THE CONTROL REP81 EMPTY VECTOR. ..... 87
FIGURE 4.5: AVERAGE RECOVERY OF CELLS IN EACH STRAIN TRANSFORMED WITH REP81 BXB1 INTEGRASE PLASMID AS A PROPORTION OF THE CONTROL REP81 EMPTY VECTOR. ..... 87
FIGURE 4.6: THE CLONES USED FOR THE CELLS VIABILITY TEST WERE TRANSFERRED FOR BOTH YES AND PMG SUPPLEMENTED WITH 150MG/L OF LEUCINE (FIGURE 5) TO SCREEN THE *URA4Δ* CLONES. ..... 89
FIGURE 4.7: *URA4Δ* CLONES GROWTH ON YES, PMG SUPPLEMENTED WITH LEUCINE, AND 5-FOA. ..... 90
FIGURE 4.8: MAP SHOWING THE POSITION OF THE PROBES AND RESTRICTION ENZYME SITES ON THE CENTROMERE 2 REGION OF THE STRAINS. ..... 92
FIGURE 4.9: NOTT373_CONTROL. THE GENOMIC DNA WAS DIGESTED WITH *BGLII* AND *PSTI* AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE, CENTROMERIC DNA, AND *URA4+* MARKER, RESPECTIVELY. ..... 93
FIGURE 4.10: NOTT373_CEN1. THE GENOMIC DNA WAS DIGESTED WITH *BGLII* AND *PSTI* AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE, CENTROMERIC DNA, AND *URA4+* MARKER, RESPECTIVELY. ..... 93

FIGURE 4.11: NOTT373_CEN2. THE GENOMIC DNA WAS DIGESTED WITH *BGL*II AND *PST*I AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE, CENTROMERIC DNA, AND *URA4+* MARKER, RESPECTIVELY. ..... 94
FIGURE 4.12: NOTT373_CEN3. THE GENOMIC DNA WAS DIGESTED WITH *BGL*II AND *PST*I AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE, CENTROMERIC DNA, AND *URA4+* MARKER, RESPECTIVELY. ..... 94
FIGURE 4.13: NOTT373_CEN4. THE GENOMIC DNA WAS DIGESTED WITH *BGL*II AND *PST*I AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE, CENTROMERIC DNA, AND *URA4+* MARKER, RESPECTIVELY. ..... 95
FIGURE 4.14: NOTT373_CEN5. THE GENOMIC DNA WAS DIGESTED WITH *BGL*II AND *PST*I AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE, CENTROMERIC DNA, AND *URA4+* MARKER, RESPECTIVELY. ..... 95
FIGURE 4.15: NOTT373_CEN8. THE GENOMIC DNA WAS DIGESTED WITH *BGL*II AND *PST*I AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE, CENTROMERIC DNA, AND *URA4+* MARKER, RESPECTIVELY. ..... 96
FIGURE 4.16: NOTT373_CEN3. THE GENOMIC DNA WAS DIGESTED WITH *BGL*II AND HYBRIDIZED WITH PROBES FOR THE NOURSEOTHRICIN RESISTANCE GENE AND *URA4+* MARKER, RESPECTIVELY. ..... 96
FIGURE 4.17: GENERAL REPRESENTATION SHOWING THE POSITION OF THE PRIMERS 1001 AND 1041 ON CHROMOSOME 2 OF *S. POMBE* RECOMBINED STRAIN AFTER CENTROMERE 2 EXCISION. ..... 97
FIGURE 4.18: MEASURING THE PROPORTION OF *URA4*Δ CLONES THAT HAD CENTROMERE 2 EXCISED. ..... 99
FIGURE 4.19: RATE OF CENTROMERE FORMATION. ..... 101
FIGURE 4.20: AVERAGE RATES OF EXCISION AND DEATH (K1) AND OF EXCISION AND CENTROMERE FORMATION (K2) AS A PROPORTION OF THE RATE OF CELLS GROWTH (K). ..... 104
FIGURE 5.1: LOCALISATION OF CENTROMERIC PROTEIN CNP1 ON ENGINEERED CBS 2777 STRAINS BY CHIP-SEQ OF GFP-TAGGED CNP1 ..... 109
FIGURE 5.2: PROPORTION OF CNP1 BINDING ON NOTT373 STRAINS. ..... 110
FIGURE 5.3: LOCALISATION BY CHIP-SEQ OF GFP-TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_CEN5 AFTER EXCISION OF THE NATIVE CENTROMERE. ..... 111
FIGURE 5.4: LOCALISATION BY CHIP-SEQ OF GFP-TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_CEN8 AFTER EXCISION OF THE NATIVE CENTROMERE. ..... 112
FIGURE 5.5: LOCALISATION BY CHIP-SEQ OF GFP TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_CEN1 AFTER EXCISION OF THE NATIVE CENTROMERE. ..... 112
FIGURE 5.6: LOCALISATION BY CHIP-SEQ OF GFP TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_CONTROL AFTER EXCISION OF THE NATIVE CENTROMERE. ..... 113
FIGURE 5.7: LOCALISATION BY CHIP-SEQ OF GFP TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_CEN4 VECTOR AFTER EXCISION OF THE NATIVE CENTROMERE ..... 113
FIGURE 5.8: LOCALISATION BY CHIP-SEQ OF GFP TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_CEN3 AFTER EXCISION OF THE NATIVE CENTROMERE OF CHROMOSOME 2. ..... 114

FIGURE 5.9: LOCALISATION BY CHIP-SEQ OF GFP TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_CEN2 AFTER EXCISION OF THE NATIVE CENTROMERE OF CHROMOSOME 2. ..... 114
FIGURE 5.10: LOCALISATION BY CHIP-SEQ OF GFP TAGGED CNP1 OF CENTROMERIC PROTEIN CNP1 ON NOTT373_ALPHOID AFTER EXCISION OF THE NATIVE CENTROMERE OF CHROMOSOME 2. ..... 115
FIGURE 5.11:CNP1 BINDING ON *CEN2* AND NEO-CENTROMERE AS A PROPORTION OF CNP1 BINDING ON *CEN1+CEN3*. ..... 116
FIGURE 5.12: LOCALISATION OF CENTROMERIC PROTEIN CNP1 ON CHROMOSOME 3 OF CBS 2777 BY CHIP-SEQ OF GFP TAGGED CNP1 ..... 117
FIGURE 5.13: POSITION OF THE PRIMERS 1047 AND 646 ON *S. POMBE* CHROMOSOME I. ..... 117
FIGURE 5.14: GEL ELECTROPHORESIS OF THE PCR USING PRIMERS 646 AND 1047 TO SCREEN *SWI6* TARGETED CLONES. ..... 118
FIGURE 5.15: RESTRICTION ENZYME SITE MAP OF THE CLONES BEFORE AND AFTER THE *SWI6* TARGETING. ..... 118
FIGURE 5.16: FILTER HYBRIDIZATION ANALYSIS OF THE FRAGMENTS FROM THE ENZYMATIC DIGESTION WITH *ECOR* I FOR THE *SWI6* TARGETED CLONES HYBRIDISED WITH A PROBE FOR THE REGION UPSTREAM THE GENE LOCUS. ..... 119
FIGURE 5.17: MAP SHOWING THE POSITION OF THE PRIMERS 1001 AND 1041 ON CHROMOSOME 2 IN THE *S. POMBE* RECOMBINED STRAIN BEFORE AND AFTER THE BXB1 INDUCTION. ..... 121
FIGURE 5.18: GEL ELECTROPHORESIS OF THE PCR ACROSS THE BXB1-*ATTL* SITE OF THE Δ*URA4* NOTT373_CONTROL$^{\Delta SWI6}$ CLONES. ..... 121
FIGURE 5.19: GEL ELECTROPHORESIS OF THE PCR ACROSS THE BXB1-*ATTL* SITE OF THE Δ*URA4* NOTT373_CEN5$^{\Delta SWI6}$ CLONES. ..... 122
FIGURE 5.20: PULSE FIELD GEL ELECTROPHORESIS AND FILTER HYBRIDIZATION OF THE GENOMIC DNA OF NOT373_CONTROL$^{\Delta SWI6}$ *URA4*Δ CLONES ..... 123

# LIST OF TABLES


TABLE 2.1: LIST OF PRIMERS USED IN THIS STUDY FOR GENOMIC PCR AND SEQUENCING. ..... 46

TABLE 2.2: DESCRIPTION OF THE EXPERIMENTS IN WHICH THE PRIMERS WERE USED. ..... 47

TABLE 2.3: THE STAINS USED IN THIS STUDY (DEPOSITED IN THE LAB COLLECTION) THEIR GENOTYPE AND SOURCE. BP4: BREAKPOINT 4 ON THE RE-ARRANGED CHROMOSOME 2 OF THE STRAIN CBS2777. ..... 60

TABLE 3.1: PLASMIDS USED IN THIS STUDY FOR THE ϕC31 INTEGRASE MEDIATED SITE-SPECIFIC RECOMBINATION AND THE SIZES OF THE TEST SEQUENCES. ..... 65

TABLE 3.2: PAIRS OF PRIMERS USED FOR THE SITE-SPECIFIC RECOMBINATION SCREENING BY PCR ON THE RIGHT (*attR*) AND LEFT (*attL*) ENDS OF THE RECOMBINATION SEQUENCE AND FOR SEQUENCING THE LEFT END OF THE RECOMBINATION. ..... 67

TABLE 3.3: RECOMBINED STRAINS DERIVED FROM NOTT373. ..... 72

TABLE 3.4: PAIRS OF PRIMERS USED FOR THE SITE-SPECIFIC RECOMBINATION SCREENING BY PCR ON THE RIGHT (*attR*) AND LEFT (*attL*) ENDS OF THE RECOMBINATION SEQUENCE. ..... 78

TABLE 3.5: RECOMBINED STRAINS DERIVED FROM NOTT376. ..... 80

TABLE 4.1: AVERAGE COLONY SIZE REPRESENTED AS THE NUMBER OF CELLS PER COLONY AND CELLS VIABILITY MEASURED AS THE PERCENTAGE OF ALIVE CELLS FROM AS KNOWN PLATED NUMBER OF CELLS. ..... 87

TABLE 4.2: AVERAGE PROPORTION OF *ura4Δ* CELLS IN EACH STRAIN AFTER THE INDUCTION OF THE BXB1 INTEGRASE. . 89

TABLE 4.3: SOUTHERN BLOT ANALYSIS OF THE CANDIDATE SWAPPED CLONES. ..... 97

TABLE 4.4: EXPECTED PCR PRODUCTS SIZE FOR THE EXPERIMENT USED TO CONFIRM THE PROPORTION OF *ura4*Δ CLONES THAT HAD EXCISED THE CENTROMERE 2. ..... 98

TABLE 4.5: PROPORTION OF *ura4*Δ CLONES, *ura4*Δ AND CENTROMERE EXCISED CLONES, AND REAL PROPORTION OF CENTROMERE EXCISED CLONES IN EACH STRAIN. ..... 100

TABLE 4.6: RATES OF EXCISION AND DEATH (K1/K) AND OF EXCISION AND CENTROMERE FORMATION (K2/K) AS A PROPORTION OF THE RATE OF CELLS GROWTH (K). ..... 104

TABLE 5.1: STRAINS USED FOR THE CHIP-SEQ EXPERIMENTS. ..... 108

TABLE 5.2: AVERAGE COLONY SIZE AND AVERAGE RECOVERED FOR EACH STRAIN TRANSFORMED BOTH WITH THE REP81-BXB1 AND THE REP81 EMPTY VECTOR ON Δ*swi6* STRAINS. ..... 120

TABLE 5.3: PROPORTION OF *ura4Δ* CLONES FOR EACH STRAIN AFTER THE BXB1 INDUCTION. ..... 120

TABLE 8.1: ABSOLUTE NUMBERS OF THE MEASUREMENTS OF NUMBER OF CELLS IN COLONY, VIABILITY OF CELLS, AND CENTROMERE EXCISION. ..... 149

TABLE 10.1: NUMBER OF READS IN THE CHIP-SEQ EXPERIMENT. ..... 151

# ABBREVIATIONS

| | |
|---|---|
| 5-FOA | 5-fluoroorotic acid media |
| ARS | Autonomously Replicating Sequence |
| bp | base pair |
| BSA | Bovine Serum Albumin |
| CDK | Cyclin-dependant kinases |
| Cen | Centromere |
| ChIP | Chromatin Immunoprecipitation |
| Chr | Chromosome |
| DMSO | dimethyl sulphoxide |
| dNTP | Deoxy nucleoside triphosphate |
| EDTA | Ethylene diamine tetra-acetic acid |
| GFP | Green fluorescent protein |
| kb | Kilobase pairs |
| *nat* | nourseothricin resistance gene |
| NTC | nourseothricin antibiotic |
| $OD_{600}$ | Optical density at 600nm |
| PBS | Phosphate Buffer Saline solution |
| PEG 4000 | Polyethylene glycol 4000 |
| rpm | rotations per minute |
| SDS | Sodium dodecyl sulphate |
| SSC | Saline Sodium Citrate |
| TE | Tris-EDTA |

# ABSTRACT

In many eukaryotes, kinetochores may form over a variety of unrelated centromeric sequences. This fact has led to the idea that an epigenetic process determines kinetochore location. To investigate the nature of this epigenetic process, I have established a system that allows the manipulation of the chromosomal centromeric DNA in a genetically tractable model organism, *Schizosaccharomyces pombe*. The system enables the definition of the relationship between centromeric DNA size, sequence, chromosome position and function. I have used this system to measure how binding of the conserved kinetochore protein Cnp1 (CENP-A homolog) varies as a function of the amount, sequence and position of the centromeric DNA.

In humans, cytogenetic data suggests that the number of centromere-specific CENP-A nucleosomes at each centromere is approximately uniform regardless of the sequence and length of the centromeric DNA. These observations suggest that the number of these centromere-specific nucleosomes is tightly regulated. I set out to test whether such mechanism is evolutionarily conserved in the fission yeast *S. pombe*. I did this by manipulating the amount of the centromeric DNA at one centromere and then measuring the amount of Cnp1 bound to it. The results showed that in *S. pombe*, the amount of bound Cnp1 is proportional to the amount of centromeric DNA and thus the relationship between centromeric DNA size and CENP-A binding differs from humans. During these measurements, I observed that the size and sequence of the centromeric DNA do have a role in determining Cnp1 binding to centromeric DNA but, as described by others, it is not sufficient and sometimes not necessary for functional centromere formation. The requirement of the kinetochore protein Swi6 for neo-centromere formation but not for the maintenance of a pre-established centromere was also confirmed.

# 1. INTRODUCTION

Centromeres direct the accurate segregation of chromosomes at cell division and are present at one copy on each of the chromosomes. If a chromosome lacks a centromere or the centromere functions incorrectly, the chromosome will not segregate accurately at cell division, and aneuploidy will result. Thus, centromeres play a key role in the process of inheritance. Aneuploidy can, in the case of humans, lead to congenital disabilities or, if it arises in somatic cells, to cancer. It is therefore of great importance that we understand the mechanism that ensures the presence of a single centromere on a chromosome.

Although the centromere has been extensively described by cytological and genetic approaches, defining the marks that confer its functions remain challenging. Centromeres vary cytologically between different eukaryotes. Despite this variation, centromeres hold several properties in common regarding their architecture and function. In the vast majority of the eukaryotes analysed to date, centromere specification is regarded as primarily epigenetic and depends on the presence of specialised nucleosomes containing the histone H3 variant centromere protein A (CENP-A). Specific DNA sequences are neither sufficient and nor necessary for centromere function (Earnshaw and Migeon, 1985, Steiner and Clarke, 1994, Higgins et al., 2005, duSart et al., 1997). Nonetheless, experimental work using *de novo* centromere formations assays demonstrates that in both human cells and *S. pombe* native centromeric DNA sequences are the preferred substrate for centromere formation (Catania et al., 2015, Ikeno et al., 1998, Harrington et al., 1997).

In the present work, we established a system in the model organism *Schizosaccharomyces pombe* that enables the manipulation of the centromeric DNA of one chromosome. The results show that, in *S. pombe*, the CENP-A homolog Cnp1 binds to a candidate sequence for centromere formation only when the sequence is integrated adjacent to a pre-existing centromere, and that the amount of bound Cnp1 is proportional to the centromeric sequence length. Furthermore, we have determined approximately the minimum amount of centromeric DNA necessary for centromere formation in *S. pombe* and confirmed the requirement of the heterochromatin protein Swi6 for neo-centromere formation. We conclude that, in *S. pombe*, centromere function requires the combination of Cnp1 containing nucleosomes, features of the

underlying DNA sequence, the presence of chromatin marks, and interactions with kinetochore proteins.

In this introduction, I will firstly review the phylogeny, karyotype and cell cycle of our model organism *S. pombe* strain CBS 2777, then I will discuss chromosomal engineering using site-specific recombinases, and finally, I will review the structural and functional elements related to centromere formation.

## 1.1. *Schizosaccharomyces pombe*

### 1.1.1. Phylogeny

The fungus genus *Schizosaccharomyces* is characterised by the fission division and comprises the three eight-spored species *S. japonicus*, *S. octopurus* and *S. cryophilus*, and the four-spored species *S. pombe*. This narrow group forms a monophyletic clade into the class *Schizosaccharomycetes* (figure 1.1). This class is a basal branch in *Ascomycota* phylum separated from *Saccharomyces cerevisiae* from 420 to 330 million years ago. Despite the great biological differences if compared with the budding yeasts, the fission yeasts have some features that characterise them as Ascomycetes such as the fungal-type cell wall, the closed mitosis, the sack-like sporangium and the mode of sex determination (Sipiczki, 2003b, Sipiczki, 2000).

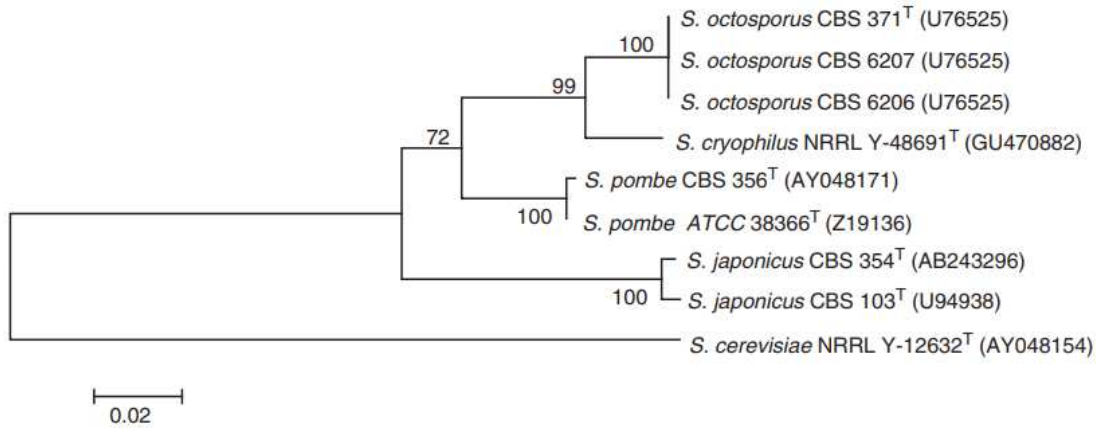



**Figure 1.1: Consensus phylogeny of fission yeasts. Neighbor-joining phylogenetic tree based on sequences of the D1/D2 divergent domains of the LSU rRNA gene.**
*Saccharomyces cerevisiae* was included as an outgroup representative. GenBank accession numbers are indicated. Numbers represent the bootstrap values from 1000 replicates (values < 50% are not shown). Scale bar represents the number of substitutions per nucleotide position. Figure adapted from (Helston et al., 2010).

### 1.1.2. Model organism

For the last 60 years, *S. pombe* has been largely exploited as a model organism for the study of the eukaryotic cell cycle and other aspects of cell biology. *S. pombe* shares with higher eukaryotes cells aspects of its cell biology processes that are not found in or are very different in *S. cerevisiae*. Examples include the cell cycle regulation machinery, cytokinesis, splicing, and protein folding processing. Aravind et al. (2000) compared more than 4000 protein sequences from fission yeast *S. pombe* with all available eukaryotic sequences and identified about 600 genes that are conserved in *S. pombe* and nonfungal eukaryotes but are missing or highly diverged in *S. cerevisiae*. The authors also reported the elimination of functionally connected groups of proteins in *S. cerevisiae* that are conserved between *S. pombe* and other eukaryotes, such as the signalosome and the spliceosome components. These results lead to the conclusion that *S. pombe* is a better model in some respects for eukaryotic cell biology than *S. cerevisiae.*

*S. pombe* was isolated for the first time from East African beer, and afterwards, it has been found in many continents in natural fermentations and fruits (Brown et al., 2011b). The most explored strain, the laboratory strain, can be traced back to a homothallic Europe isolate of *S. pombe*, which was isolated in 1921 by A. Osterwader and then developed as a genetic system by Leupold in 1950 (Brown et al., 2011b, Sipiczki, 2003a). The laboratory strain is identified as 968 and contains three chromosomes.

The laboratory strain has been widely used as an experimental system, but little is known about the variation of this species in nature. In a recent study, 81 natural isolates were analysed at the phenotypic, genotypic, and karyotypic levels and the results suggested that *S. pombe* has a variable karyotype, exists in small and incompletely isolated population and a restricted range of habitat environments (Brown et al., 2011b). In this same work, the authors characterised the strain CBS 2777 as having four chromosomes, and this is the strain that is being used for the present work.

Contributing to our understanding of *S. pombe* genetic diversity and population structure, Jeffares et al. (2015) surveyed the genomic and phenotypic variation in the fission yeast by sequencing the genomes of 161 strains isolated worldwide between the years 1912 and 2002. The analysed strains were isolated primarily from cultivated

fruit and various fermentations. Initial analysis detected 25 clusters of strains that differed by <150 SNPs. By excluding these near-identical strains, a set of 57 strains was identified that each differed by ≥1,900 SNPs. The genetic diversity within these strains was moderate ($p=3\times10^{-3}$ substitutions/site). Principal-component projection of the 'drift distance' between strains determined using 752 selected SNPs close to the linkage disequilibrium (pairwise $r^2 < 0.5$) and evenly distributed across the genome showed weak clustering of strains by geography and an evident pattern of genetic isolation by distance, with genetic and physical distances being weakly but significantly correlated. This result suggests that there is a global population structure, which has been obscured by recent dispersal and intermixing of some strains. The authors estimated that dispersal of *S. pombe* began during human antiquity (~340 BCE), indicating an evolutionarily recent worldwide dispersal, probably associated with the spreading of fermentation technologies. Furthermore, they provided a mean estimate of 1623 CE for the arrival of *S. pombe* in the Americas, coincident with European colonialism in the continent. Notably, isolates from the Americas also showed the highest genetic similarity suggesting a recent European origin for *S. pombe* in the Americas.

### 1.1.3. The strain CBS 2777

The merit of the *S. pombe* strain CBS 2777, concerning the goal of manipulating centromere sequence structure and content, is that it has two centromeres that lack flanking heterochromatin. In the laboratory strain, this heterochromatic DNA is present in a palindromic arrangement flanking the centromere central core. This palindromic organisation and the difficulty of targeting heterochromatin prevents routine manipulation of the centromeric DNA size and structure by site-specific and homologous recombination (N.Lee, PhD, University of Nottingham, 2008) in the laboratory strain.

The *S. pombe* strain CBS 2777 was isolated in Japan in 1942 from fermenting molasses. Brown et al. (2014) have characterised the sequence and organisation of CBS 2777 genome using array-based comparative genome hybridisation (CGH) to identify the components of the individual chromosomes (figure 1.2A). They showed that CBS 2777 has four chromosomes formed by a complex rearrangement of the three chromosome strain $972h^-$ (figure 1.2B). Chromosomes 1, 2 and 4 of CBS 2777 are

composed of different combinations of sequences related to those present on chromosomes I and II of the three chromosome strain. In CBS 2777 the central core of the centromere chromosome II is duplicated, so one copy is present on chromosomes 2 and the other on chromosome 4. These rearranged centromeres present in CBS 2777 are flanked by two segments of DNA, each ~100kb long, that are normally present in the centromere-distal regions of the short arms of most three chromosome strains. Neither of the segments associate with heterochromatin in either CBS 2777 or in the laboratory strain; nevertheless, the centromeric central core sequences bind kinetochore proteins in similar amounts as the laboratory strain centromeres which are flanked by heterochromatin. The absence of heterochromatin around two of the centromeres of CBS 2777 offers the possibility of their easy manipulation by site-specific and homologous recombination which would make this strain an ideal model for the study of kinetochore assembly and centromere inheritance.

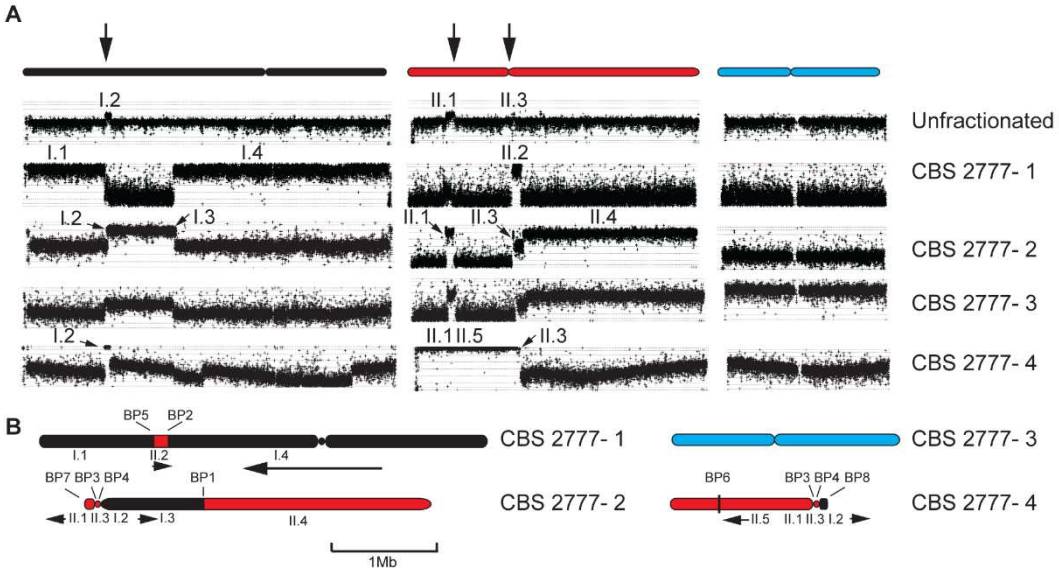



**Figure 1.2: Genome organisation of *S. pombe*.**
(A) Sequence content of size-fractionated chromosomes of CBS 2777 (for which Arabic numerals are used, with the 972h− chromosomes designated by Roman numerals). Either unfractionated CBS 2777 DNA or DNA from each of the four chromosomes that had been size-fractionated by pulsed-field gel electrophoresis was analysed by CGH. The results are aligned with respect to the three laboratory strain (972h⁻) chromosomes listed at the top of the panel and colour-coded black, red, and blue, respectively. Chromosome 3 of CBS 2777 is not detectably rearranged. Segments of the 972h⁻ chromosomes present in the CBS 2777 chromosomes are labelled according to their chromosomal origin using Roman numerals. Segment II.5 extends from II:569,645 to the imr2 left repeat of chromosome II and corresponds to an inversion to which the CGH is insensitive. The three vertical arrows at the top of the figure refer to the three regions of the laboratory strain karyotype that are duplicated around the centromeres of CBS 2777 chromosomes 2 and 4. (B) Arrangement of the sequences identified in A on the CBS 2777 chromosomes. The arrows below the chromosomes indicate whether the respective segment is in the same orientation as in the 972h⁻ genome as drawn at the top or inverted (running right

to the left). Each of the eight breakpoints (BP 1–8) defining the CBS 2777 karyotype is indicated. Figure taken from Brown et al. (2014).

## 1.2. Fission yeast cell cycle and regulation

*S. pombe* is predominantly haploid with vegetative propagation by medial fission. Conjugation and zygote formation occur in nutritional depletion if opposite mating types are available, and the resulting diploids cells usually sporulate, but diploid cells can be recovered under special conditions.

When in vegetative growth, *S. pombe* G2 phase takes 70% of the cell cycle, and G1, S and M phases each take about 10% of the division time (figure 1.3). G1 and G2 are "gaps" that ensure that everything is ready for DNA synthesis and the cell division. During S phase the DNA is replicated, and at the M phase, the segregation of the sister chromatids takes place. In eukaryotic cells, a molecular machine coordinates the cell cycles which is mainly regulated by dimers of cyclins and cyclin-dependent protein kinases. The cyclins are regulatory subunits necessary for the cyclin-dependent protein kinases activity and targeting its specific substrates. Unlike other eukaryotes, *S. pombe* has only one essential cyclin-dependent protein kinase, identified as Cdc2 (Nurse and Thuriaux, 1980) and only one essential cyclin, Cdc13, although other three cyclins (Cig1, Cig2 and Puc1) are present. The Cig1 and Cig2 also form a complex with Cdc2, although these cyclins can be substituted by Cdc13 (Fisher and Nurse, 1995), Cig1 and Cig2 are important in G1 promoting the entry into S phase. The Puc1 cyclin (Forsburg and Nurse, 1991b) associates with Cdc2 during G1 and couples the initiation of S phase with the achievement of a critical cell size.

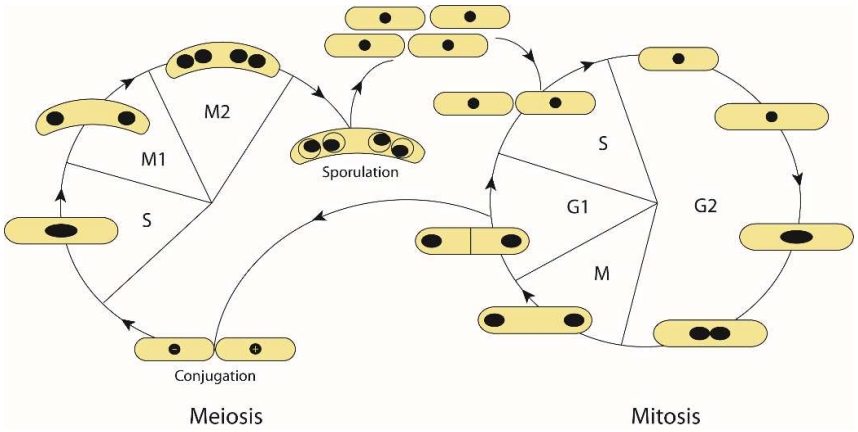



**Figure 1.3: Cell cycle of the fission yeast *S. pombe*.**
On the left is represented the meiotic cycle after the conjugation of opposite mating type cells and resulting in the formation of four spores, typical of this species. On the right is represented the

vegetative cycle. The G2 phase takes 70% of the cycle timing, and the M, G1 and S phase occur during the other 30% of the time.

During the G1, Cdc2 activity is very low, and then it rises to an intermediated level that enables the DNA replication. The peak of Cdc2 activity occurs in the M phase, and it decreases during the anaphase to the very low-level characteristic of the G1, then the cell starts a new cycle (Forsburg and Nurse, 1991a).

The dimer Cdc2/Cdc13 is regulated in three ways (figure 1.4): (1) cyclin degradation by the anaphase-promoting complex (APC); (2) reversible inhibitory phosphorylation of the Tyrosine 15 residue (Tyr15) of Cdc2 by the Wee1 and Mik1 tyrosine kinase; and (3) and binding to a reversible stoichiometric inhibitor (the Rum1 protein (Moreno and Nurse, 1994) in fission yeast). These regulators are activated as described below.

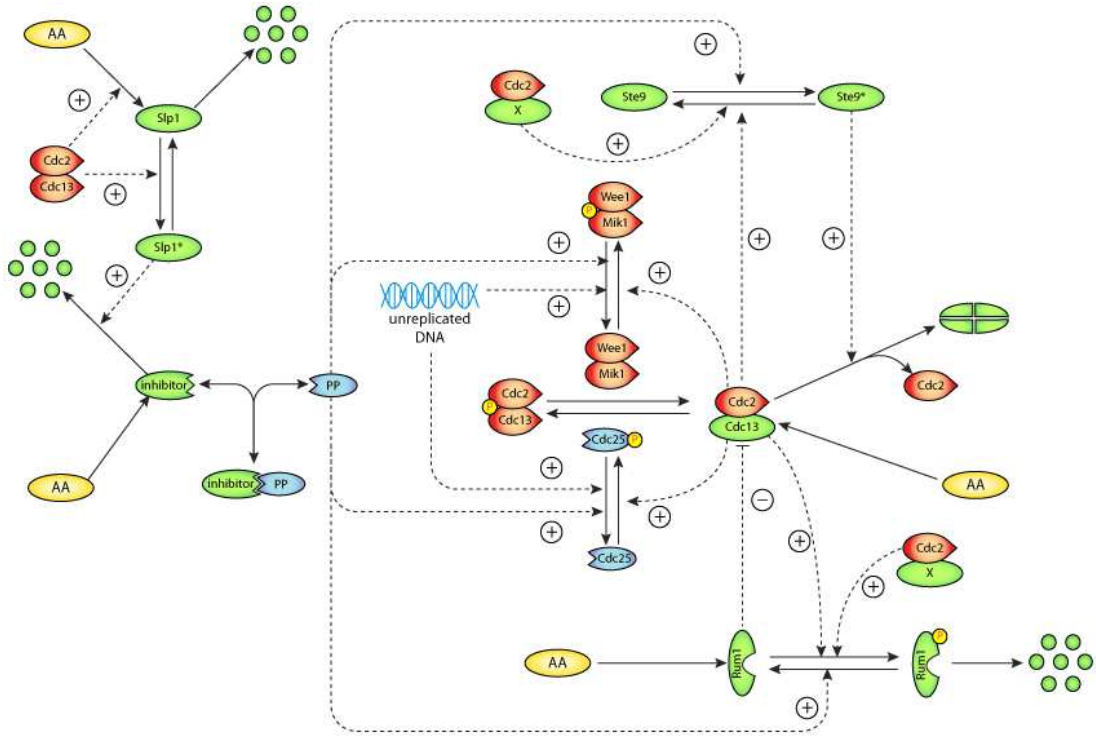



**Figure 1.4: A molecular mechanism for the regulation of Cdc13-associated kinase activity in fission yeast.**
All of the events of the fission yeast cell cycle can be orchestrated by fluctuations of the activity of a single cyclin-dependent kinase, Cdc13/Cdc2. Cdc13 is synthesised from amino acids (AA) and combines readily with catalytic subunits, Cdc2, which are assumed to be always present in excess. The activity of Cdc13/Cdc2 is modulated by Rum1 inhibition, by Tyr-15 phosphorylation (via Wee1 and Mik1, which is reversed by Cdc25 and Pyp3; the last molecule is not shown on figure), and by Ste9-dependent cyclin degradation. Other cyclins (Puc1, Cig1, and Cig2, which are represented by X on the figure) in complex with Cdc2 can assist these processes. Adapted from Sveiczer et al. (2000)

Cdc13 is continuously synthesised throughout the cell cycle, where it binds to Cdc2, and then the dimer moves into the nucleus. At early G2, the Cdc2/Cdc13 dimers

accumulate in the nucleus in an inactive form, phosphorylated by Mik1 and Wee1 kinases. Wee1 and Mik1 are inhibitory protein kinases that promote the Tyr15 phosphorylation of Cdc2 inactivating it. Mik1 does not affect the normal cell cycle progression but is critical for the S checkpoint. The absence of Wee1 allows premature activation of Cdc2/Cdc13 complex and the entry into M, generating the wee phenotype (cells as smaller than the normal) (Russell and Nurse, 1987). Wee1 is regulated by multiple protein kinases, Cdr1 is one of these proteins and has been shown to inhibit Wee1 activity in vitro (Russell and Nurse, 1987, Coleman et al., 1993, Parker et al., 1993, Wu and Russell, 1993). During the S phase, the amount of active Cdc2/Cdc13 is enough to enable the DNA synthesis and to prevent the cell from entering in mitosis. The increase of Cdc2/Cdc13 during G2 phase activates Cdc25 phosphatase that reverses Tyr15 phosphorylation on Cdc2/Cdc13 at the end of G2. The high peak of Cdc2/Cdc13 makes the transition from G2 to M phase. Therefore, to be active, the Cdc2 must be bound to a cyclin subunit, free of phosphate at Tyr15 and phosphorylated at T167. Enzyme activities capable of phosphorylating the T167 have the generic name of CDK-activating kinase or CAK (Gould, 2003). In *S. pombe* the dimer of cyclin-dependent kinase Mcs6-Mcs2 is at least partially responsible for T167 phosphorylation, and a non-essential kinase called Csk1 is predicted to takes part of the process as well (Lee et al., 1999, Hermand et al., 1998).

The APC degrades Cdc13 to stop the activity of Cdc2/Cdc13, as cells exit mitosis. Cdc13 is targeted to the APC by two auxiliary APC-binding proteins, Slp1 and Ste9 (Blanco et al., 2000). Cdc2/Cdc13 activate Slp1 by itself, after a time delay and is primarily responsible for the degradation of Cdc13 and the cells re-enter G1. Because of Cdc2/Cdc13 decreases, the Slp1 activity also falls off, then, Ste9 becomes active and takes over the degradation of Cdc13 when the cell is already in G1 phase. At the same time, Rum1 binds to the few Cdc2/Cdc13 and Cdc2/Cig2 dimers present in G1 which accumulates during late M-G1 transition. Rum 1 is not essential for the vegetative growth, but it is critical for *S. pombe* cells to arrest in G1 in response to mating factors and conjugation (Moreno and Nurse, 1994). At the beginning of a new cell cycle the Cdc2/Cig2 and Cdc2/Cdc13 become active in G1 phase, they phosphorylate some unknown protein(s) at the origin of replication, and then S-phase starts.

Although much has been learned about the regulation of the cell cycle since the discovery and characterization of Cdc2 and while this work continues to provide

a broadly applicable paradigm for cell cycle control, there are still important areas of our understanding of the cell cycle that are completely defective and, of these, the epigenetic inheritance of the kinetochore is possibly one of the most important.

## 1.3. Chromosome Engineering using recombinases

### 1.3.1. Site-specific recombination

Site-specific recombination is a type of DNA sequence recombination that occurs between specific DNA sequences that may or may not be identical. The process involves DNA breakage and reunion catalysed by enzymes called recombinases with no gain or loss of DNA and are thus described as reciprocal and conservative. Site-specific recombination systems are ubiquitous throughout the eubacteria and archaea groups but are not frequently found in eukaryotes. Site-specific recombination has been widely exploited as a tool for genome manipulation of homologous or heterologous organisms and has been used for the integration, excision, and inversion of DNA sequences. Integration results from recombination between sites on separate DNA molecules (at least one of the substrate chromosomes is circular), and it occurs with a uniquely defined orientation. For sites located on the same chromosome, the outcome is determined by their relative orientation: excision results from recombination between sites in a head-to-tail orientation; and inversion results from an exchange between inverted (head-to-head) sites.

All known recombinases fall into one of two distinct families: the tyrosine recombinases and the serine recombinases. They are classified according to the amino acid residue that forms the protein-DNA linkage during the process of recombination and has different mechanisms of action.

### 1.3.2. Tyrosine Recombinases

The feature that defines the integrase family is the tyrosine nucleophile in combination with a totally conserved set of basic amino acid residues: two arginines and a histidine, known as the ‘RHR’. Most of the studies about recombinases of this enzyme family have focused on the integrase from bacteriophage λ (Landy, 1989) which is the prototypical member of the tyrosine recombinase; *Flp* recombinase from yeast (Dixon et al., 1995); *Cre* recombinase from bacteriophage P1 (Hoess and

Abremski, 1985); and the *XerC* and *XerD* recombinases from *Escherichia coli* (Sherratt et al., 1995).

For the site-specific recombination, the tyrosine recombinases only cleave one strand of the duplex DNA at a time, each nicked strand must be joined to its partner before the second strand can be cut (figure 1.5). This produces a cross-strand intermediate known as a Holliday junction (Grindley et al., 2006).

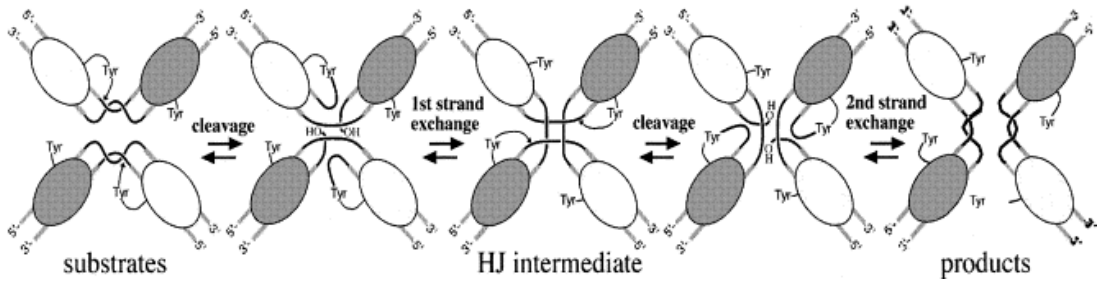



**Figure 1.5: Scheme of the site-specific recombination promoted by tyrosine recombinase family exemplified through the Cre-loxP system showing the Holliday-junction (HJ).**
Two Cre subunits bind to each loxP site and the sites associate to form a synaptic complex. Two subunits (light colored) cleave the sites with conserved Y324 to form a covalent 3'-phophotyrosine intermediate. Intermolecular attack by the partner substrates 5'-OH group completes exchange of the first pair of DNA strands and produces a HJ intermediate. The tyrosine recombinases only cleave one strand of the duplex DNA at a time, each nicked strand must be joined to its partner before the second strand can be cut. Isomerizatioin of the HJ intermediate (not shown here) results in activation of the second pair of recombinase subunits (dark colored) and exchange of the second pair of strands to form recombinant products. The products are substrate of the reaction runs in reversed since loxP is regenerated in the reaction. Diagram taken fromDiagram from Ghosh and Van Duyne (2002).

The λ integrase (λ Int) catalyse unidirectional reactions between distinct substrate sites (*attB* and *attP*) (figure 1.6) and the proteins *Cre* and *Flp* that catalyse reversible reactions between identical sites. The λ *Int* has been used in the *in vitro* Gateway cloning system developed by Invitrogen however, the integrase requires the host cell accessory protein IHF and is thus unsuited to *in vivo* genome manipulation. *Cre* and *Flp* can promote site-specific recombination in the absence of additional cofactors, but both use identical substrate sites and thus, the reactions promoted by these proteins are reversible. Although these tyrosine recombinases have been widely used for the genome manipulation in a range of organism, it has mainly been used for deleting sequences integrated into the genome. *Cre* is the key reagent in the widely used floxing technique so popular amongst mouse geneticists (Nagy, 2000).

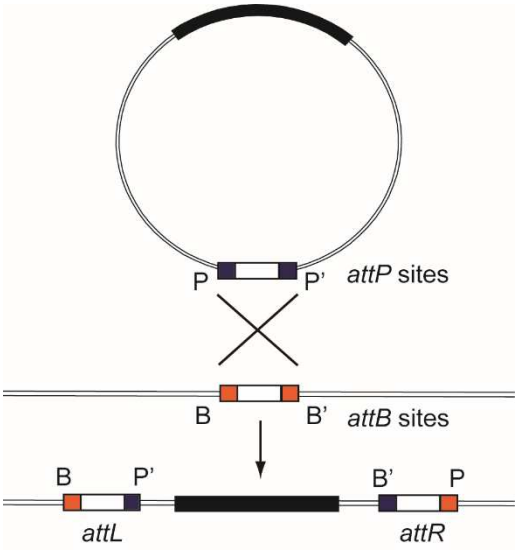



**Figure 1.6: Schematic description of site-specific recombination between the bacteriophage attachment site *attP* in the plasmid and the bacterial attachment site in the *attB* genome.**
In this way, the sequence of interest can be stably integrated into the *attB* genomic locus. The ends of the recombined sites as called *attL* on the left and *attR* on the right end.

### 1.3.3. Serine Recombinases

Site-specific recombination systems which function in heterologous hosts and which combine the simplicity of *Cre* recombinase and the directionality of λ Int were found in the serine recombinase family prompting their use as genome engineering tools (Groth et al., 2000, Thorpe and Smith, 1998, Brown et al., 2011a).

The serine recombinases are a heterogeneous family of proteins characterised by a serine nucleophile. The resolvase/invertase family are the enzymes classically used to describe the serine recombinases (Hatfull and Grindley, 1986, Stark et al., 2011). They are small proteins of only approximately 180 amino acids that are exquisitely directional and specific in their recombination outcome (Brown et al., 2011a). The first resolvase site-specific recombination systems discovered and best understood so far are those encoded by the related bacterial transposons Tn3 and γδ, required for resolution of the Tn3 transposition co-integrate and γδ transposition intermediates, and Hin and Gin invertases that control gene expression of *Salmonella flagella* antigens and tail fibres in phage Mu, respectively. These enzymes promote the DNA recombination by a double-strand break, switch, and then rebinding. All catalytic processes usually occur within a topologically specific synaptic complex with two crossover sites and four recombinase subunits and require host encoded cofactors. Once a synaptic complex is formed between the enzyme and the crossover sites, the four recombinase subunits are activated to attack the two crossover sites, forming two

double-strand breaks (figure 1.7). Then the broken ends are rearranged to bring them into a recombinant configuration (Grindley et al., 2006, Li et al., 2005).

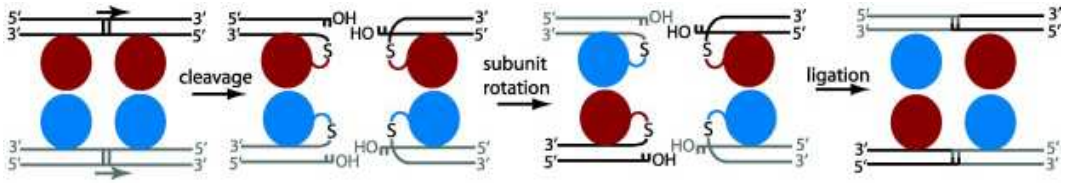



**Figure 1.7: A generic scheme for the serine recombinase reaction mechanism.**
Serine recombinases bind to a pair of recombination sites (shown here as two parallel black or grey lines with the 2 bp crossover sequences shown as two short vertical lines) as dimers (two red and two blue subunits) and then form a tetramer, bringing the DNA sites together in a synaptic complex. All four subunits are simultaneously activated to cleave the DNA, forming in each subunit a phosphoserine link to a recessed 5' end and leaving a free 2 bp 3' overhang. Subunit rotation exchanges two DNA half sites, repositioning the half sites into a recombinant format. Ligation of the DNA backbone generates the recombinant molecules. Diagram adapted from Olorunniji and Stark (2010) and (Stark et al., 2011).

Another group of serine recombinases has emerged from studies of phages, prophages and transposons from predominantly Gram-positive bacteria. These recombinases are called large serine recombinases owing to their high molecular weight compared with resolvase/invertases (Smith and Thorpe, 2002).

The large serine recombinase family are functionally, but not mechanistically, analogous to λ integrase (Smith and Thorpe, 2002). They cause recombination between the phage attachment site, *attP*, and the bacterial attachment site, *attB*, to give the hybrid sites, *attL* and *attR*. Excision is the reverse reaction, which brings together *attL* and *attR* to reform *attP* and *attB* and requires an accessory protein encoded by the phage that is termed the exisase.

ϕC31 integrase is one of the large serine recombinases, so-called because they contain the 140 amino acid catalytic domain first identified in the resolvase/invertases and a very large C-terminal domain of 300–500 amino acids compared to the small 40–50 amino acid C-terminal domain in the resolvase/invertases (Smith and Thorpe, 2002). The properties of purified ϕC31 integrase have been studied *in vitro* (Thorpe and Smith, 1998, Thorpe et al., 2000). Unlike the γδ resolvase or the Tn3 resolvase, this protein catalyses *attP/attB* recombination in the absence of host factors or other phage-encoded proteins and their reactions are not topologically restricted. Thus the *attP/attB* reaction occurs irrespective of whether the substrates are supercoiled or linear, or whether the *att* sites are on the same of different molecules and in the absence of accessory factors. These features makes the ϕC31 integrase and other similar serine recombinases important tools for genomic manipulation. The ϕC31 *att* sites have little sequence conservation between the pairs of recombination sites apart from the core

sequences. The minimal *attP* and *attB* sites are 39 and 34 bp respectively (Groth et al., 2000) and the core sequence is central (figure 1.8).




**Figure 1.8: Sequences of the attachment sites for the *Streptomyces* phage ϕC31.**
Only the upper strand of the recombination sites is shown. For each pair of sites, identical basepairs are shown in red. The bases in grey represent sequences flanking the minimal sites for *att* sites. Recombination occurs within the core sequences (shown in bold) and the two basepairs within the red boxes probably form the staggered overhangs after cleavage (Breuner et al., 2001). Inverted repeat sequences are shown by arrows. Adapted from (Smith and Thorpe, 2002).

The serine recombinase family is large and includes many enzymes with the potential to be widely used in genome engineering. Amongst these is the Bxb1 recombinase which I have also used. (Singh et al., 2013).

### 1.3.4. *Schizosaccharomyces pombe* plasmids

There is a wide variety of *S. pombe*-specific plasmids for molecular manipulation. A fission yeast plasmid usually contains a bacterial origin of replication; an antibiotic-resistance gene (that can be selected in bacteria); a yeast selectable marker (usually a nutritional marker); and an autonomous replication sequence (ARS), which is responsible for plasmid maintenance and for the high frequency of transformation.

### 1.3.5. Selectable markers

In *S. pombe* episomes are generally positive selected by prototrophic markers or by drug-resistance genes. These selectable markers are described below.

The *ura4*$^{+}$ gene is the most commonly used marker in fission yeast and encodes an orotidine monophosphate decarboxylase. The *ura4Δ* cells can be selected by the use of 5-fluoroorotic acid (5-FOA), which generates a cytotoxic compound for the *ura4*$^{+}$ cells (Luo et al., 2010). Thus, 5-FOA selects against the marker and only cells that lose the gene can survive.

The *leu1*$^{+}$ gene is another commonly used marker in *S. pombe*, it encodes a b-isopropyl malate dehydrogenase (Tinline-Purvis et al., 2009) and the *leu1-32* allele leads to leucine auxotrophy. Because the *S. cerevisiae* orthologue *LEU2* complements *leu1* auxotrophs in high copy, this marker is commonly used on episomes.

Because the *ade6*$^{+}$ gene is quite large and contains numerous restriction sites, it is seldom used as a selectable marker on plasmids. *S. pombe ade6Δ* can grow on media with low levels of adenine and it forms red or pink colonies (Thuriaux, 1985). In the absence of selection, the colonies will be white. This is a convenient plasmid marker, especially if a visual selection for plasmid maintenance is required.

Adding antibiotic resistance genes to *S. pombe* cells is useful selection method and can be used concomitantly with the nutritional markers. The kanMX6 cassette can be used to confer G418 resistance to cells (Bahler et al., 1998). Other available cassettes for fission yeast are natMX6, hphMX6 and bleMX6, which confer resistance to nourseothricin, hygromycin B and bleomycin respectively (Hentges et al., 2005).

### 1.3.6. Plasmid systems

The plasmids are designed accordingly to its function. For fission yeast molecular manipulation there are four kinds of plasmid systems available: general purpose episomes, integration vectors, expression vectors, and epitope tagging plasmids. They differ in the presence of replication origin and promoters, and in markers.

#### 1.3.7. General Purpose Episomes

Most vectors in *S. pombe* are derived from a few common plasmid backgrounds that preserve common multiple cloning sites in a lacZ+-derived polylinker, which provides blue–white selection in bacteria (Iraqui et al., 2012). The pUR18/19, pAL1, pAL19, and pBG1 plasmids are all derivatives of pUC19, have prototrophic markers and a polylinker. However, these plasmids do not provide a promoter and are generally used for cloning and expression of genomic fragments containing their own promoters.

#### 1.3.8. Integration vectors

Due to the size of *S. pombe* centromere, it cannot be placed in a typical bacterial shuttle vector, thus no single copy episomal plasmids exist in *S. pombe* and the episomes have a high frequency of loss. Therefore, experiments requiring a stable, defined copy number should be performed with integrating vectors (Iraqui et al., 2012). The best integration approach is through homologous recombination. Vectors

lacking ARS and containing a selectable marker are used for targeted homologous integration.

### 1.3.9. Expression vectors

Numerous plasmids have been designed for regulated the expression of heterologous or endogenous genes. Expression levels can vary significantly even for the same plasmid, depending upon the protein being expressed and the context of the ATG codon.

### 1.3.10. nmt1 promoter

No message in thiamine (nmt1) was the first characterised promoter in fission yeast (Olshen et al., 2004), and it remains the most commonly employed. It contains a TATA-box located 25 base pairs upstream of the transcription start site. The promoter is repressed by the presence of thiamine (vitamin B1) in the media and it is induced when thiamine is removed. This promoter was used to construct the pREP episome and the pRIP integration plasmids (Maundrell, 1993).

### 1.3.11. The pREP and pRIP systems

The pREP plasmids are pUC-based plasmids containing *S. pombe* ARS, nmt1 promoter and *LEU2* or *ura4+* as the selectable marker. The pRIP1/2/3 plasmids lack the ARS element and instead contain sup35 gene (a codon-specific tRNA nonsense suppressor), and are used for chromosomal integrations. Both pREP1/2 and pRIP1/2 have polylinkers contain ATG codons (Fridlyand et al., 2004).

## 1.4. The Centromere

The centromere is a unique region on the chromosome that is required for attachment to the mitotic spindle and chromosome segregation. Paradoxically, although this role is conserved throughout evolution, the sequences that accomplish centromere function in different organisms are not. Most budding yeasts have centromeres defined by a specific DNA sequence found on all chromosomes, whereas fission yeasts and higher eukaryotes have larger regional centromeres presenting a more complex structure. Many similarities between centromeres in different species

have been observed that highlight the unique nature and essential functions of the centromere.

### 1.4.1. Centromeric DNA

#### 1.4.2. Budding yeasts

Most fungi have the typical large regional centromeres observed in the majority of eukaryotes, where the core region of the centromere is flanked by repetitive arrays of heterochromatin (Pluta et al., 1995, Volpe et al., 2002, Chan and Wong, 2012, Hall et al., 2012). The heterochromatin is characterised by the methylation of lysine 9 of the histone H3 (H3K9me2/3). However, an early ancestor of the Saccharomycotina, a subphylum of the phylum Ascomycota, lost the enzymatic machinery (Swi6/Clr4/Epe1) for making and maintaining H3K9me2/3 heterochromatin (Malik and Henikoff, 2009, Allshire and Ekwall, 2015). The descendants of this ancestor have smaller centromeres categorised into two types: "point centromeres" and "small regional centromeres" (Roy and Sanyal, 2011, Coughlan et al., 2016). All studied species of the yeast family Saccharomycetaceae, which is part of the subphylum Saccharomycotina, have point centromeres. *Saccharomyces cerevisiae*, the first species of this family to be studied (Fitzgeraldhayes et al., 1982, Hieter et al., 1985, Hegemann and Fleig, 1993), has three conserved DNA sequences common to all chromosomes (Clarke and Carbon, 1983, Fleig et al., 1995) known as centromere DNA element (CDE), and together they form an 116–120 bp sequence (figure 1.9). These three regions are sufficient to confer the centromere function (Fitzgeraldhayes et al., 1982). CDEI is 8 bp long and is required for high-fidelity chromosome segregation; CDEII is a 78–86 bp required for chromosome segregation; and CDEIII is 26 bp long and contains seven invariant nucleotides conserved across all chromosomes, the mutation of which abolishes centromere function (McGrew et al., 1986). The CDE I–CDE III consensus is conserved among most species of the family Saccharomycetaceae (Gordon et al., 2011), except in the genus *Naumovozyma* which has point centromeres with neither absolutely conserved consensus sequences nor chromosomal locations (Kobayashi et al., 2015).

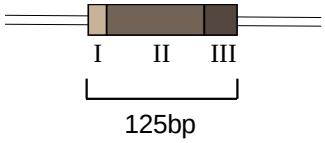



**Figure 1.9: General structure of the Saccharomyces cerevisiae centromere.**
The 125bp S. cerevisiae centromere is highly conserved and organised in three Centromere DNA Elements (CDE): CDEI (I), CDEII (II), and CDEIII (III).

The point centromeres of *Debaryomyces hansenii* and *Scheffersomyces stipitis* contain large clusters of Ty5-like retrotransposon elements within which the precise location and structure of the centromere remains unknown (Lynch et al., 2010). Still in the subphylum Saccaromycotina, *Yarrowia lipolytica* has small centromeres (Vernis et al., 2001, Yamane et al., 2008) described as point centromeres, although the consensus sequence is much more poorly conserved among chromosomes than that in Saccharomycetaceae.

Small regional centromeres show no consensus sequence or structural features. These centromeres were firstly identified in *Candida albicans* and *C. dubliniensis* (Sanyal et al., 2004, Padmanabhan et al., 2008). The sequence of a *C. albicans* centromere has shown not to be necessary nor sufficient for centromeric function (Ketel et al., 2009), posing the question of what keeps centromeres at a conserved location throughout generations. *Clavispora* (*Candida*) *lusitaniae* has similarly non-repetitive small regional centromeres (Kapoor et al., 2015). However, *C. tropicalis* have 3–6kb inverted repeat structures at its centromeres, flanking a central 2–10kb region that binds the CENP-A homolog Cse4 (Chatterjee et al., 2016).

In the methylotrophs clade, have been studied the centromeres of the species *Kuraishia capsulata*, *Ogataea (Hansenula) polymorpha*, and *Komagataella phaffii* (*Pichia pastori*). *K. capsulata* has a 200-bp conserved consensus sequence on five of its seven chromosomes (Morales et al., 2013), and *O. polymorpha* has Ty5-like retrotransposon clusters without any obvious conservation of centromere sequence or structure (Ravin et al., 2013, Hanson et al., 2014). *K. phaffii* centromeres strongly resemble the structure of the inner repeats and central core regions of *S. pombe* centromeres and other *Schizosaccharomyces* species but they lack the additional nested arrays of outer repeats that extend the inverted repeats and are similar across all *S. pombe* chromosomes (Coughlan et al., 2016).

The best-characterized large regional centromeres with H3K9me2/3 heterochromatin fungi species are *Neurospora crassa* and *S. pombe*.

### 1.4.3. Fission yeast

Fission yeast centromeres are complex structures that are widely considered to resemble those of metazoans, this makes *S. pombe* a good model for the study of metazoan centromeres which are typically larger and in a genetic background in which only limited genetic analysis is possible. The centromeric DNA of the three fission yeast chromosomes are similar in the organization but vary in size and in the number of copies of their flanking heterochromatic repeats. In the lab strain, the chromosome I centromere (Cen1), is the smallest with ~35kb of centromeric DNA; centromere 2 (Cen2) is ~65kb in length. The biggest centromere (Cen3) occupies ~110kb. These estimates of centromere length include the flanking heterochromatin which varies in copy number between different isolates. It is not clear how the centromeres of natural isolates vary and whether there is any overall relationship between the chromosomes and the total length of the centromeric DNA including the heterochromatin (Steiner et al., 1993).

The general structure of the three centromeres is formed by the central core (*cnt*), which include the inner flanking repeats (imr) and the non-palindromic DNA, and the flanking outer repeats (*otrL* and *otrR*) which are composed by sub-domains *dg* and *dh* that have specific arrangements in each centromere (figure 1.10). The inner repeats have different sequences in each centromere. Regions of the outer repeat elements are highly homologous between the different centromeres (dg is 97% identical, whilst dh is 48% identical) (Grewal, 2000). The main difference in size between the three centromeres results from variation in number of outer repeat elements.

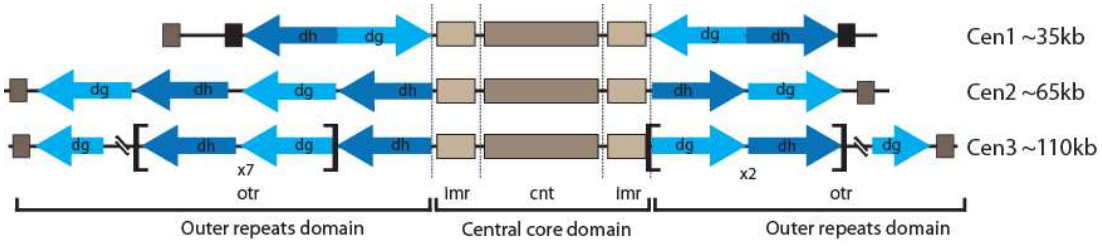



**Figure 1.10: Diagram of the general organisation of the centromeric regions of the chromosomes 1, 2 and 3 of *S. pombe*.**
The centromeres are formed by the central core (cnt) flanked by the inner repeats (imrL and imrR) and the outer repeats (otrL and otrR). The inner repeats have different sequences in each centromere and the outer repeats are composed by sub-domains dg and dh which have specific arrangements in each centromere. Figure adapted from Takahashi et al. (1992).

Studies with minichromosomes have shown that, in addition to a central core, either an outer repeat region (Steiner and Clarke, 1994) or over expression of CENP-A (Catania et al., 2015) will initiate the formation of a centromeric domain, but that neither is essential for the continued presence of the centromere and thus, once the functional state has been established, it can be propagated indefinitely.

In order to investigate the contribution of primary DNA sequence for the establishment of kinetochore in vivo, Catania et al. (2015) utilised a plasmid bearing the central domain sequence of the *S. pombe* centromere 2. The authors demonstrated that, when the centromeric chromatin protein Cnp1 is overproduced, the central domain DNA is a preferred substrate for *de novo* establishment of the Cnp1 chromatin. They have shown that, when the centromere sequence is multimerized, 2kb sub-regions with similar AT content and predicted nucleosome positioning, some of the sub-regions can form functional centromeres and others cannot. These observations indicate that central domain DNA from fission yeast centromeres contains specific information that promotes Cnp1 incorporation into chromatin, but that outer repeat enhances centromere establishment and emphasise roles for both genetic and epigenetic processes in centromere establishment.

### 1.4.4. Human

*Homo sapiens* centromeres are 0.1-4 Mb long and formed by a tandem repetitive DNA sequence known as α-satellite DNA. The α-satellite DNA is defined by a diverged 171bp motif repeated in a tandem head-to-tail fashion and has been identified at the centromeres of all normal primate chromosomes studied to date (Schueler et al., 2001, Willard and Waye, 1987). The α-satellites are hierarchically organised into higher-order repeat arrays (figure 1.11) in which a defined number of α-satellite monomers have been homogenised as a unit to yield large chromosome-specific arrays (Willard and Waye, 1987).

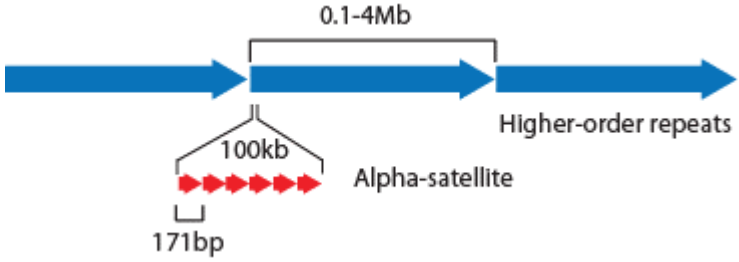



**Figure 1.11: Human centromere organisation of α-satellites and higher-order repeats.**
The α-satellite is a highly repetitive DNA hierarchically organised into higher-order repeat arrays. Figure adapted from Fukagawa and Earnshaw (2014).

### 1.4.5. The centromere and kinetochore proteins

Despite large divergence in centromere DNA sequences, many kinetochore proteins are conserved across organisms. The kinetochore is responsible for ensuring exclusive connection of centromeres to opposite spindle poles. It contains over 100 proteins, including core components and regulatory machinery to ensure their proper function.

A common feature of the functional centromeres is the existence of a CEN-specific histone H3 variant, CENP-A (named as Cse4 in *S. cerevisiae*, Cnp1 in *S. pombe*). CNP-A occupies a region ranging from several kilobases in fission yeast to about 100kb in metazoans; in contrast, the budding yeast CENP-A (CSE4) is thought to be present in only a single nucleosome at the CEN elements (figure 1.12). CENP-A chromatin is popularly regarded to function both as part of an epigenetic mark for centromere identity and as a structural foundation for the assembly of the kinetochore (Earnshaw and Rothfield, 1985, Palmer et al., 1991, Sullivan et al., 1994). Its presence has been shown to be essential for kinetochore formation at active centromeres in most eukaryotes, including dicentric chromosomes and neo-centromeres (Marshall et al., 2008, Tyler-Smith et al., 1999, Earnshaw et al., 1989, Warburton et al., 1997).

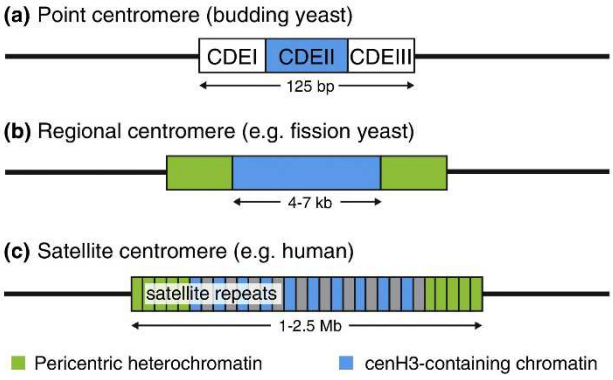



**Figure 1.12: CenH3-containing chromatin in different organisms.**
(a) The point centromere, found in budding yeast, consists of three DNA elements, CDE I–II–III. A single cenH3 nucleosome assembles on CDE II. (b)Regional centromeres, found for example in fission yeast, consist of a centromeric region that is usually flanked by pericentric heterochromatin. (c) Satellite centromeres, for example found in humans, are similar to regional centromeres in that they occupy contiguous regions of chromosomes flanked by pericentric heterochromatin, but they are composed of particular families of tandem repeats, occupied by alternating arrays of cenH3 and H3-containing nucleosomes. Adapted from Steiner and Henikoff (2015)

The deposition of CENP-A is controlled by various chaperones and assembly factors. In most vertebrates, the Mis18 complex (formed by Mis18α, Mis18β, and

M18BP1Knl2) and HJURP, which is a CENP-A specific chaperone, are involved in the CENP-A deposition (Kato et al., 2007, Foltz et al., 2009, Dunleavy et al., 2009). The HJURP-CENP-A-H4 complex recognises the Mis18 complex and enables the deposition of CENP-A (Nardi et al., 2016). In *S. pombe*, Scm3 is the HJURP homolog (Williams et al., 2009).

Proteins that specifically bind to centromere DNA may play a role in centromere specification (Nagpal and Fukagawa, 2016). In the point centromere of budding yeast, the centromere binding factor 3 (CBF3 complex) specifically binds to centromere DNA (Lechner and Carbon, 1991) and is sufficient for centromere function. In mammalian cells, regional centromeres have a conserved DNA-binding element, the CENP-B box, which associates with the centromeric protein CENP-B. In humans, the CENP-B box is a 17-bp motif found within the 171-bp α-satellite repeats (Masumoto et al., 1989). However, the CENP-B box is not completely conserved among mammalian species (Haaf et al., 1995, Kipling et al., 1995). CENP-B is thought to have to function in kinetochore assembly through association with CENP-C (Fachinetti et al., 2015).

Various proteins are recruited to establish the functional kinetochore. The kinetochore can be divided by electron microscopy into three parts: (1) the inner kinetochore, which includes proteins that recognise centromere chromatin; (2) the outer kinetochore, which includes proteins that are involved in microtubule binding; and (3) regulatory proteins (figure 1.13).

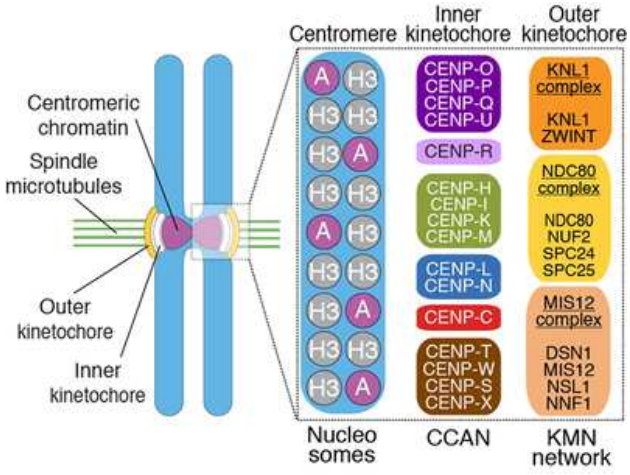



**Figure 1.13: Organization of kinetochores.**
Kinetochores are built on specialized centromeric chromatin containing the histone H3 variant CENP-A. The CCAN contains subunits (e.g. CENP-C and CENP-LN) that interact directly with CENP-A. CENP-C interacts directly with the KMN network, thus contributing to its recruitment. The Ndc80 complex is the main microtubule binder in the kinetochore. Adapted from Musacchio and Desai (2017).

The inner kinetochore is composed of the constitutive centromere-associated network (CCAN) which is formed by 16 proteins (Foltz, 2006, Okada, 2006, Izuta, 2006, Hori et al., 2008, Amano et al., 2009, Nishino et al., 2012), and is responsible for recruitment of the microtubule-associated outer kinetochore proteins during mitosis (Cheeseman et al., 2008). The CCAN is evolutionary conserved from budding and fission yeasts to humans, although most members are missing in *Caenorhabditis elegans* and *Drosophila* species (Oegema et al., 2001, De Wulf et al., 2003, Liu et al., 2005, Goshima et al., 2003, Akiyoshi and Biggins, 2010, Orr and Sunkel, 2011, Przewloka et al., 2011, Bock et al., 2012, Malvezzi et al., 2012, Drinnenberg et al., 2014). The 16 CCAN members can be divided into sub-groups based on their functions and biochemical characteristics (Okada, 2006, Hori et al., 2008): CENP-C, the CENP-T group (CENP-T-WS-X), the CENP-H group (CENP-H-I-K-L-M-N), and the CENP-O group (CENP-O-P-Q-R-U). CENP-C and CENP-T-W-S-X connect the centromere to the outer kinetochore (Przewloka et al., 2011, Nishino et al., 2012). The CENP-H group is located between the chromatin, CENP-C, and CENP-T-W-S-X. It plays roles in the maintenance and recruitment of CCAN (Cheeseman et al., 2008, Hori et al., 2013, Nagpal et al., 2015). The CENP-O group localises to the kinetochore, downstream of the CENP-H group, and plays roles in microtubule dynamics and kinetochore alignment (Hori et al., 2008, Amaro et al., 2010, Hua et al., 2011, Eskat et al., 2012, Bancroft et al., 2015).

The inner kinetochore acts as a platform for the assembly of the outer kinetochore. The outer kinetochore is the microtubule binding interface and is mainly formed by the Knl1/Mis12/Ndc80 complex (KMN), which assembles directly onto the CCAN (Cheeseman, 2004, Obuse, 2004, Cheeseman et al., 2006). The four-subunit Ndc80 complex (Spc24, Spc25, Nuf2, and Ndc80Hec1) forms a rod-shaped structure (Ciferri et al., 2008) in which the Nuf2-Ndc80Hec1 end directly binds microtubules (Wei et al., 2007, Cheeseman et al., 2006, DeLuca, 2006), and the Spc24-Spc25 complex faces the inner kinetochore (Petrovic et al., 2010). The Mis12 complex is a four-member rod-shaped complex made up of Nnf1, Mis12, Dsn1, and Nsl1 (Kline et al., 2006) that binds the Ndc80 complex through its Nsl1 subunit (Petrovic et al., 2010) and links to CENP-C through direct interaction (Przewloka et al., 2011, Screpanti et al., 2011). The inner kinetochore protein CENP-T also directly binds the Ndc80 complex (Nishino et al., 2013). The main role of Knl1 is regulation of the kinetochore-microtubule interaction, by recruiting other outer kinetochore components associated

with microtubule dynamics and the spindle assembly checkpoint (SAC) (Cheeseman and Desai, 2008).

Proper recruitment of the outer kinetochore components is necessary for correct chromosome segregation. There are feedback mechanisms for recruitment of outer kinetochore components. They ensure proper chromosome–microtubule attachment and chromosome alignment before starting chromosome segregation. The overall control of outer kinetochore recruitment is dependent on CDK1 kinase activity, which mainly regulates cell cycle progression. CDK1 is also involved in the recruitment of Aurora B kinase to the kinetochore (Tsukahara et al., 2010). Aurora B is a member of the chromosome passenger complex (CPC) and has a role in controlling kinetochore-microtubule attachments (Biggins et al. 1999; Cheeseman (Cheeseman, 2002) et al. 2002, 2006; Tanaka et al. 2002; DeLuca et al. 2006; Liu et al. 2009; Kim et al. 2010; Welburn et al. 2010).

Kinetochore function is distinct from centromeric chromatin. However, because kinetochores assemble on the surface of centromeres, it can be difficult to separate the functions of the two. Aspects of kinetochore function are beginning to be well understood, including microtubule binding, chromosome movement, and checkpoint signalling. (Cheeseman and Desai, 2008; Santaguida and Musacchio, 2009; Perpelescu and Fukagawa, 2011; Vleugel et al., 2012). However, the organisation and functions of centromeric chromatin remain less understood subjects of active study.

### 1.4.6. Kinetochore assembly in *S. pombe*

The primary function of a kinetochore is to attach the chromosome to the dynamic plus ends of spindle microtubules, the kinetochores are also associated with the maintenance of cohesion between sister chromatids till anaphase onset (Cleveland et al., 2003, Cheeseman and Desai, 2008). Additionally, a kinetochore is involved in the recruitment of the spindle assembly checkpoint machinery that monitors the kinetochore attachment to the microtubules and initiates signals to prevent cell cycle progression if an error persists.

Although the exact nature of kinetochore architecture in *S. pombe* is still not clear, studies indicate that it resembles the structure of human kinetochore, which comprises three layers: inner, middle and outer layers (Baum et al., 2006). Proteins

that form the inner layer interact directly with the centromeric DNA, while the outer layer proteins form the chromosomal attachment sites of the microtubules. Proteins in the middle layer act as linkers between the inner and outer kinetochores (Cheeseman and Desai, 2008).

Kinetochore assembly is considered to be initiated by CENP-A deposition, which involves epigenetic factors and takes place at different stages of the cell cycle. In *S. pombe* it occurs at S and G2 phases (Takayama et al., 2008, Bernad et al., 2009). The non-histone DNA-binding chaperone Scm3 (Williams et al., 2009, Pidoux et al., 2009) is required for kinetochore assembly in *S. pombe*, it is part of a protein family that has the propensity to bind to the centromeric DNA and contains a histone chaperone domain, which is required for CENP-A deposition (Williams et al., 2009). Another important protein for the inner kinetochore assembly is CENP-C, member of the evolutionarily conserved CENP-C family (Saitoh, 1992), it contains a long conserved region known as the CENP-C box, which is essential for the kinetochore localisation through determination of CENP-A localisation (Black et al., 2007a). Other proteins that affect the Cnp1 binding in *S. pombe* are Ams2 at S phase, Hip1 at G2 phase and Mis6 (Chen et al., 2003, Saitoh et al., 1997).

CENP-B is a highly conserved protein in humans and mice and is specifically localised at the centromere (Earnshaw et al., 1985). This protein binds to the 17-bp motif of the CENP-B box sequence in alphoid DNA at its aminoterminal region and forms homodimers at its carboxy-terminal region (Masumoto et al., 1989, Yoda et al., 1992). In *S. pombe*, Abp1, Cbh1 and Cbh2 are homologous to CENP-B and have redundant functions in the regulation of various nuclear processes (Murakami et al., 1996, Halverson et al., 1997, Lee et al., 1997, Nakagawa et al., 2002, Baum and Clarke, 2000, Irelan et al., 2001).

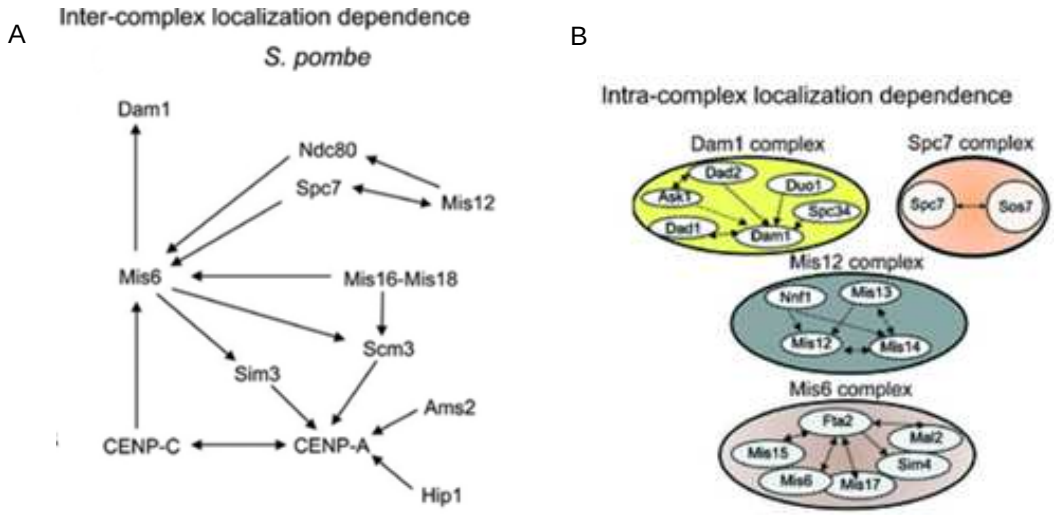



**Figure 1.14: The recruitment of proteins during kinetochore assembly in *Schizosaccharomyces pombe*.**
(A) Relative presence of various kinetochore proteins from outer to inner layers (top to bottom). Arrows indicate the localization dependence of a protein/protein complex to another during the assembly process. Arrowheads point towards the component, which is dependent on the other protein. (B) Composition of various major kinetochore complexes. Dotted arrows show intracomplex dependency. The double-headed arrows indicate mutual dependency between proteins and protein complexes. Adapted from (Roy et al., 2013)

### 1.4.7. Neo-centromere formation

Generally, the centromere has a fixed localisation on the centromere and does not change through the cells generations. However, occasionally, a centromere may appear at a new position on the chromosome; such centromeres are called neo-centromeres. The discovery and characterization of a neo-centromere on human chromosome 10 by duSart et al. (1997) showed that the DNA present at a neo-centromere is not in any way similar to that found at native centromeres and thus demonstrated that centromeres are inherited epigenetically. The mechanism of this epigenetic inheritance and its relationship to the process of neo-centromere formation is not understood (Buscaino et al., 2010).

Genomic alterations can remove or disrupt centromere function, and neo-centromeres can restore chromosome segregation faithfully. In rare cases, neo-centromeres form without a deletion but concomitant with inactivation of the endogenous centromere (Amor et al., 2004). They are particularly common in specific chromosome regions (Marshall et al., 2008). A specialised chromatin environment may favour neo-centromere formation. Neo-centromeres have been observed in diverse organisms including humans, flies, plants, and in the yeasts *C. albicans* and *S. pombe* (Ketel et al., 2009, Ishii et al., 2008, Shang et al., 2013, Heun, 2006, Topp et

al., 2009, Nasuda et al., 2005) are stable structures in both mitosis and meiosis (Tyler-Smith et al., 1999, Amor et al., 2004) and often show no detectable homology to native centromeric DNA.

Synthetic neo-centromeres can help to elucidate the cell cycle dynamics of centromeric components. Neo-centromeres have been engineered on endogenous centromere removal (Ishii et al., 2008), and in chicken cells, they typically form near the original centromere (Shang et al., 2013) possibly as a result of 'seeding' events around the original excised site. In *D. melanogaster*, ectopic centromeres have been generated following induced chromosome breakage (Olszak et al., 2011), which leads to neo-centromere formation in heterochromatic regions, indicating that the chromatin environment at ectopic sites influences neo-centromere establishment. Further studies of the nature of neo-centromere are crucial for the elucidation of centromere establishment and maintenance.

## 1.5. Epigenetics and the centromere

Epigenetics has been defined in many ways. Two of the most widely used current definitions are:

- The perpetuation of gene expression and functional differences across cell divisions without changes in DNA sequence (Heard and Martienssen, 2014).
- The structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states (Bird, 2007).

The first has the merit of precision and highlights the notion of inheritance while the second reflects common usage particularly by biochemists and cell biologists and avoids the questions posed by any instability of an epigenetic state. It suffers however from the disadvantage that it encompasses almost any protein or DNA modification that changes the expression of a gene.

The centromere is often described as being epigenetically inherited because even within a single species, centromeres can form over a variety of apparently unrelated different DNA sequences and a particular sequence may or may not function as a centromere in different individuals of the species. As regards the first definition of epigenetics, the description of the centromere as epigenetically inherited poses a

particular problem. In so far as cells may be placed in culture and propagated indefinitely, centromeres are apparently able to be inherited with undetectable levels of inaccuracy. Thus, the first definition forces us to confront the problem of the mechanism of centromere inheritance and in particular how it can be inherited in an apparently error-free way. The second definition ignores inheritance and leaves open the possibility that the centromeric state is set once in every cycle. However, independently of any definition of epigenetics, we do not know how the kinetochore, is able to stably associate with a variety of different DNA sequences in a way that persists over an apparently unlimited number of cell divisions.

The term epigenetics has been applied to many different mechanisms by which gene expression is modified (Heard and Martienssen, 2014). These discussions have influenced ideas about the epigenetic inheritance of the kinetochore. Sometimes this influence is not explicit. Consequently, it is rare to find clear statements about how the kinetochore is propagated from one cell cycle to the next with close to 100% fidelity however in order to understand something about current thinking about the problem of centromere inheritance it is necessary to consider this background.

### 1.5.1. Chromatin-based epigenetics

The primary organizational unit of chromatin is the nucleosome, which consists of ~145 bp of DNA tightly wrapped around an octamer consisting of two copies of each of the histone proteins H2A, H2B, H3, and H4 (Luger et al., 1997). While composed of simple units, chromatin is a complex structure that is regulated at the level of configuration, organization, and dynamics. Consistent with its complexity, a large variety of processes exist that help regulating chromatin homeostasis and dynamics in cells. The close association of chromatin and its modifications to the genome of the cells makes it an excellent candidate for driving epigenetic inheritance. Three of the major mechanisms are DNA methylation, incorporation of histone variants, and modification of histone proteins.

DNA methylation

DNA methylation is the covalent addition of a methyl group to the DNA backbone. Methylation of cytosine in the 5 position is the canonical epigenetic mark. The DNA of many, but not all eukaryotes, contain 5 methyl cytosine (5 Me-C) (Suzuki

and Bird, 2008). Important exceptions are the model organisms budding yeast, fission yeast and the nematode *Caenorhabditis elegans.* In both plants and animals, the 5 Me-C occurs in the palindromic sequence CpG while in most plants 5 Me-C is also found in the related near palindrome CNG where N is any residue. The palindromic nature of these sequences creates the possibility for the inheritance of the methylated state as a result of the action of a methylase which, like the methylases of restriction systems in procaryotes recognizes the hemi-methylated palindromic sequence. Such methylases have been found in the plants and animals that contain 5 Me-C and are referred to as maintenance methylases. Their action enables a methylated state to be transmitted from cell generation to the next. The efficiency of the maintenance methylase activity is approximately 0.96 per site per cell division (Laird et al., 2004). This efficiency is high enough to explain how cytosine methylation can contribute to epigenetic phenomena such as the inheritance of the inactive state by an inactive X chromosome in female mammals and the various phenotypes associated with hypomethylated sequences in epigenetic recombinant inbred lines in plants (Heard and Martienssen, 2014). The mechanism of inheritance of CpG and CNG methylation patterns has been generalized to other systems of epigenetic inheritance such as histone modifications by the application of a specific vocabulary including the terms mark, writer and reader. In the specific case of CpG methylation, the mark is the 5MeC, the writer is the maintenance methylase, and the reader is the transcriptional process inhibited by cytosine methylation. The pattern of methylation can differ between different cell types of the same organism and change during differentiation (Reik et al., 2001). Thus, sequence determinants are not sufficient to explain the existing pattern of DNA methylation.

Histone variants

Multiple variants exist for each of the canonical histone proteins that form the nucleosomes (Talbert and Henikoff, 2012). Some of the described histone variants have distinct properties, one example is the histone H3.3, which differs from its canonical H3.1 counterpart by only 5 amino acids, yet its dynamics and regulation are drastically different (Smith and Stillman, 1989, Ray-Gallet et al., 2002, Ahmad and Henikoff, 2002).

In the context of the centromere, evidence shows that the histone H3 variant CENP-A is the epigenetic mark for centromere identity. The first evidence in favour

of this term was the observation that CENP-A, unlike other histones, which are displaced by protamines, is specifically retained during spermatogenesis and occupies distinct foci within the sperm nucleus (Palmer et al., 1990). The simplest conclusion is that this CENP-A triggers centromere formation in the male pronucleus post-fertilization. Furthermore, loss of CENP-A is lethal and results in severe defects of chromosome segregation in all species analyzed (Stoler et al., 1995, Buchwitz et al., 1999, Henikoff et al., 2000, Blower and Karpen, 2001, Black et al., 2007b). Conversely, CENP-A is sufficient for the recruitment of the centromere and kinetochore proteins (Foltz, 2006, Heun, 2006, Liu et al., 2006, Okada, 2006, Mendiburo et al., 2011). Together, these observations have, for many years, spurred the hypothesis that CENP-A is primarily responsible for specifying centromeric identity. It is unclear, however, how the Cenp-A is inherited between one generation and the next with an error rate so low that the mark is maintained for thousands of generations.

Histone modifications

In addition to modifying the histone variant composition of nucleosomes, each of the histones can undergo a large number of posttranslational modifications. Common modifications on histones include acetylation, methylation, phosphorylation, ubiquitylation, citrullination, biotinylation, and ADP-ribosylation (Khare et al., 2012). Most posttranslational modifications exist in the N-terminal histone tails (Khare et al., 2012). In some cases, a single residue is known to exist in multiple different modified forms. Many modifications are shown to correlate with specific functional states, such as high or low gene-activity, splicing, DNA repair, and DNA replication (Bannister and Kouzarides, 2011).

Histone modifiers have been implicated in the transgenerational epigenetic inheritance of sterility and longevity in *C. elegans* (Katz et al., 2009, Greer et al., 2011). In flies, environmental stress induces disruption of heterochromatin over several generations (Seong et al., 2011). In mice, levels of trimethylated histone 3 lysine 27 (H3K27me3), a repressive histone mark, are lower at specific loci in the sperm of offspring from males fed a low protein diet (Carone et al., 2010). These observations suggest that histone modifications – and possibly chromatin states – may be central to mechanisms of transgenerational epigenetic inheritance in many

organisms. The exact sequence of events in which a change in histone marks could lead to transgenerational inheritance of traits is not yet known.

### 1.5.2. Non-coding RNAs

RNAs of multiple types have been implicated in epigenetic inheritance across generations. These include maternal stores of mRNAs and long noncoding lncRNAs, as well as small RNAs that interfere with transcription (siRNAs and piRNAs), mRNA stability, or translation (miRNA) via RNAi. Some of these small RNAs are strong candidates for triggering inheritance, as they guide DNA and histone modification in plants, animals, and fungi.

The RNA interference

The RNA interference (RNAi) machinery appears to be necessary for both maintenance of chromatin state and heritability of non-coding RNAs. In plants and fungi, RNAi pathways in the nucleus can repress target genes at the transcriptional level by guiding epigenetic modification of chromatin (Castel and Martienssen, 2013). Similar pathways have been found in the germline of metazoans. These findings have revealed a conserved nuclear role for RNAi in transcriptional gene silencing (TGS). Because it occurs in the germline, TGS can lead to transgenerational inheritance in the absence of the initiating RNA, but it is dependent on endogenously produced small RNA.

The substrate for nuclear RNAi has been shown to be RNA molecules, but these must be in close proximity to the locus from which they arose so that epigenetic modification can be specific. This has led to a model of co-transcriptional gene silencing (CTGS), whereby nuclear small RNAs target nascent RNA molecules from RNA polymerases, and the effector complexes themselves interact with and regulate the transcriptional machinery. The participation of nuclear RNAi in genome maintenance and DNA repair shows that there are other roles that nuclear small RNAs and their effectors have outside those involved in classical transcriptional silencing. The more we learn about nuclear RNAi, the more apparent it becomes that RNAi has a fundamental role in gene regulation and genome maintenance from one generation to the next.

### 1.5.3. Prions

Protein-based mechanisms could serve as an alternate means of transmitting adaptive phenotypes across generations (Halfmann and Lindquist, 2010). Many wild yeast strains harbour a variety of proteins that behave as prions and confer adaptive phenotypes to the responses of different yeast strains to a range of environmental stresses such as osmotic, oxidative, pH, ethanol as well as stress from antifungal drugs and DNA damaging agents (Halfmann et al., 2012). These observations raise the possibility that endogenous prions may be transmitted through meiosis and fertilization and perhaps underlie modes of transgenerational inheritance.

### 1.5.4. Paramutation

The phenomenon of paramutation consists in the interaction between alleles, whereby one allele induces a heritable epigenetic change in the other allele of a specific trait (Chandler, 2007). The change can be in the pattern of DNA methylation or histone modifications (Haring et al., 2010). The allele inducing the change is said to be paramutagenic, while the allele that has been epigenetically altered is termed paramutable (Chandler, 2007). A paramutable allele may have altered levels of gene expression, which may continue in offspring which inherit that allele, even though the paramutagenic allele may no longer be present (Chandler, 2007). Paramutation was primarily studied in maize (Chandler, 2007, Hollick, 2012), in which individual alleles at three different colour gene loci gave rise to epialleles with reduced pigmentation. These epialleles silenced other alleles in heterozygotes, more or less permanently. Silencing was allele-specific, dose-dependent, and temperature dependent and occurred shortly after fertilization. The phenomenon has been described in a number of other systems, including animal systems like *Drosophila melanogaster* and mice (Chandler, 2007, de Vanssay et al., 2012).

### 1.5.5. X chromosome inactivation

X chromosome inactivation (Lyon, 1961) is a dosage compensation mechanism present in the cells of female mammals that results in expression of X linked genes, present in two copies, at levels that are similar to male cells where they are present in just one copy. X chromosome inactivation is specifically required for the viability of the female mammalian embryo.

The inheritance of an inactive state by an inactive X chromosome ($X_i$) is a canonical example of an epigenetic process. A particularly clear example of X chromosome inactivation is provided by the "tortoise shell" cat. These animals which are almost exclusively female, have the variegated appearance as result of the random inactivation of either one of two X chromosomes that carry different alleles of a gene that determines the colour of the fur of the cat. One allele codes for orange, the other for black fur. The random choice as to which chromosome to inactivated is made during the early embryonic development (E5.5 in the mouse) and thereafter the inactive state is propagated epigenetically. As a consequence clones of cells express one or other of the fur colour genes and the animal is a mosaic with respect to these and other allelic genes on the X chromosome. The result is the variegated appearance that gives rise to the name of these animals.

X chromosome inactivation is initiated by the expression of a gene, *Xist* (Brown et al., 1991, Borsani et al., 1991), mapping to the so-called X chromosome inactivation centre (XIC) (Brown et al., 1991). At the time of its isolation, *Xist* was regarded as a novel type of gene because it encodes a long non-coding RNA of approximately 17-19kb (Brown et al., 1992) (Brockdorff et al., 1992). Expression of *Xist* is exclusively from the inactive X chromosome where it initiates the process of X-inactivation. Following expression of *Xist*, the *Xist* RNA localizes to the nucleus where it binds in *cis* and coats the inactive X chromosome (Clemson et al., 1996). In differentiated cells silencing of $X_i$ is independent of *Xist* and conditional deletion of *Xist* from the inactive X in mouse embryonic fibroblasts fails to disrupt X inactivation (Csankovszki et al., 2001). As a result of X chromosome inactivation, expression of most of the genes on one of the two chromosomes is extinguished. Exceptions are the genes in the two pseudo-autosomal regions and a few other genes elsewhere along the chromosome.

*Xist* RNA includes internally repetitive sequences termed repeats A-F and some of these have been shown to be ligands for proteins that participate in the X chromosome silencing pathway (reviewed in (Monfort and Wutz, 2017, Lu et al., 2017) and in the binding of the inactive X chromosome to the cell lamina. Proteins interacting with *Xist* have been identified by proteomic methods following cross-linking either with formaldehyde or with UV (reviewed in (Lu et al., 2017)). These approaches have led to the identification of between 10 and 200 *Xist* associated RNA binding proteins. Genetic screens (Monfort and Wutz, 2017) have also been used to

identify factors that are necessary for the ability of *Xist* to silence gene expression when placed at an autosomal locus in stem cells in culture. This combination of biochemical and genetic approaches has thus both confirmed the role of several proteins identified during previous studies and defined the repertoire of proteins playing a possible role in X chromosome inactivation. Together they indicate that X inactivation is, as yet, a poorly understood process of great complexity.

X chromosome inactivation is important to current thinking about epigenetics because it demonstrates unambiguously that *cis*-acting epigenetic inheritance is possible and that an epigenetic state is capable of being inherited for many tens of cell divisions. However, it is not clear whether the epigenetic inheritance of centromere identity has much to do with the inheritance of the inactive X chromosome for two reasons. Firstly the centromere is inherited without significant levels of instability for a limitless number of cell divisions while the inactivated state of an inactive X chromosome is often reversible (Tukiainen et al., 2017). Secondly, X inactivation is a chromosome-wide phenomenon while the centromere occupies just a small number of nucleosomes. Thus one might easily envisage that X inactivation arises as a result of the propagation of cooperative interactions between the RNA protein complexes responsible for the inactivated state while such interactions on the centromere would need to be constrained so as to restrict the centromere to a particular region.

### 1.5.6. Silencing of the pericentromeric heterochromatin in the *Schizosaccharomyces pombe*

The centromeres in the laboratory strain and most natural isolates of *S. pombe* are flanked by long tandemly repeated arrays of a repeated sequence termed dGdH. Transgenes inserted into these sequences show variegated patterns of expression (Allshire et al., 1994) which resembles the classical phenomenon of position effect variegation seen in *Drosophila melanogaster*. Mutations which disrupt silencing at the mating type locus, for example in the genes *clr1*, *clr2*, *clr3*, *clr4*, *swi6* and *rik1* also disrupt the silencing seen in transgenes placed in the pericentromeric repeats (Allshire, 1995). These mutations also reduce the accuracy of chromosome segregation demonstrating that integrity of the pericentric heterochromatin is important for this process. Swi6p is a chromodomain protein and is 46% identical to the *Drosophila* HP1 protein that is a major component of *Drosophila* heterochromatin. Swi6p localizes to the pericentromeric heterochromatin of *S. pombe* (Ekwall et al., 1995) suggesting that

the mechanism of heterochromatin formation at the pericentromeric repeats of *S. pombe* is evolutionarily conserved. Clr4p is a chromodomain-containing histone H3 K9 methyl transferase of the Su(var)3-9 family (Rea et al., 2000) and swi6p/HP1 specifically binds H3K9me3 (Bannister et al., 2001) indicating that understanding the targeting of Clr4p to the heterochromatin was key to understanding how the heterochromatin was formed. Given the fact that the pericentromeric heterochromatin of *S. pombe* has an evolutionarily conserved structural component and is important for chromosome segregation it has been the focus of intense scientific interest.

Our current understanding of how *S. pombe* pericentromeric heterochromatin silences inserted transgenes and forms heterochromatin emerged from the demonstration by Volpe and colleagues of the role played by the RNAi machinery in the process of silencing (Volpe et al., 2002) and the simultaneous discovery that sequences within the repeats are themselves transcribed and processed into short RNA molecules (Reinhart and Bartel, 2002). Volpe and colleagues (Volpe et al., 2002) demonstrated that mutations in genes involved in the RNAi pathway, argonaute (*ago1),* dicer *(dcr)* and RNA dependent RNA polymerase *(rdp1)* relaxed silencing of pericentromeric transgenes and leads to a loss of H3K9 methylation in the pericentromeric heterochromatin suggesting that the RNAi machinery targeted Clr4p to the centromeric heterochromatin. There is now a good understanding of how this occurs (Castel and Martienssen, 2013, Jain et al., 2016). Specifically; divergent transcription of the dGdH repeats generates double-stranded RNA which is processed into short dsRNA molecules by Dicer. These short dsRNA molecules can then bind the RITS complex composed of the Piwi domain containing RNAase argonaute, the chromodomain protein Chp1 and an argonaute interacting protein Tas3. Argonaute has a slicing activity which specifically cleaves one of the two strands of the ds RNA to generate a bound single-stranded unpaired RNA that can target RITs to the nascent RNA molecules on the dGdH template. The bound RITS complex targets the CLR complex to the dGdH chromatin. The CLR complex is a seven protein complex that includes Clr4p (Kuscu et al., 2014) and once targeted to the dGdH chromatin promotes H3K9 methylation and thus binding of swi6p/HP1. This pathway is important for our understanding of heterochromatin formation because it shares many features with silencing and heterochromatin formation pathways found in other organisms although details differ in the different systems (Castel and Martienssen, 2013).

### 1.5.7. Epigenetic inheritance of the centromere

The centromeres all studied organisms, with the exception of *S. cerevisiae* and some closely related species, specific DNA sequences are neither necessary nor sufficient for centromere identity. This is best exemplified by the rare occurrence, within the human population, of neo-centromeres: functional centromeres that have repositioned to atypical loci on the chromosome (Amor et al., 2004, Marshall et al., 2008, duSart et al., 1997). Furthermore, neo-centromeric loci are stably inherited throughout many divisions and even over multiple human generations. Thus, it is clear that centromeres are epigenetically defined and inherited.

Unlike most well-studied epigenetic traits, neo-centromeric loci can be transgenerationally inherited, five separate centromere repositioning events took place between zebra (*Equus burchelli*) and donkey (*Equus asinus*), which diverged from each other less than one million years ago, a very short window of evolutionary time (Carbone et al., 2006). Moreover, multiple donkey chromosomes exist where the typical centromeric satellite DNA is present at a genomic locus that is distinct from the active centromere, which is formed on complex DNA (Piras et al., 2010), arguing that centromere repositioning was the result of neocentromere formation. Similarly, non-repetitive centromeres have been found on specific chromosomes in multiple other equine species (Carbone et al., 2006, Wade et al., 2009, Piras et al., 2010), chicken (Shang et al., 2010), and orangutan (Locke et al., 2011). In light of this, it has been argued that neocentromere formation and centromere repositioning are one and the same phenomenon, observed at different timescales (Capozzi et al., 2008). Taken together, the evidence listed above likely makes centromeres the most stable epigenetic trait known to date.

The primary candidate for epigenetic specification of centromeres is the histone variant CENP-A, which replaces canonical H3 in centromeric nucleosomes (Henikoff et al., 2000, Fukagawa and Earnshaw, 2014, Mendiburo et al., 2011, Guse et al., 2011). CENP-A chromatin is sufficient for recruitment of the downstream centromere and kinetochore complexes (Guse et al., 2011, Mendiburo et al., 2011, Foltz, 2006, Okada, 2006). In addition, CENP-A is stably transmitted at centromeres through both mitotic and meiotic divisions (Jansen et al., 2007, Dunleavy et al., 2012), and its assembly is tightly cell cycle controlled (Jansen et al., 2007, Schuh et al., 2007). Importantly, targeting of this protein to an ectopic site of the genome is sufficient to

initiate an epigenetic feedback loop, recruiting more CENP-A to this site (Mendiburo et al., 2011).

The mechanism by which the centromeres epigenetically inherited is yet to be elucidated. Evidences suggest that the quantitative inheritance of CENP-A across cell division cycles ensures the epigenetic centromere inheritance (Jansen et al., 2007, Bodor et al., 2013). Soluble CENP-A specifically bind to HJURP, a CENP-A-specific chaperone (Dunleavy et al., 2009, Foltz et al., 2009), before been deposited into centromeres. The Mis18 complex, which contains Mis18a, Mis18b, and M18BP1, contributes to the licensing of centromeric chromatin for CENP-A deposition (Fujita, 2007) during the G1 phase of the cell cycle (Jansen et al., 2007). The Mis18 complex direct binding to CENP-A nucleosomes has been demonstrated in chicken cells and (Hori, 2017) in *Xenopus* cells (French et al., 2017). The binding of the Mis18 complex with pre-existing CENP-A nucleosomes is proposed to be a prototype marker for new CENP-A deposition, and suggested to be the explanation to how new CENP-A recognizes the centromere position (French et al., 2017, Hori et al., 2017).

CENP-A nucleosomes are highly enriched at the centromere (Bodor et al., 2014), but are also distributed at low levels throughout chromatin. Non-centromeric CENP-A has been observed in budding yeast (Camahort et al., 2009), chicken DT40 cells (Shang et al., 2013), human RPE cells (Bodor et al., 2014), and HeLa cells (Lacoste et al., 2014). This indicates that there is no restriction of CENP-A assembly exclusive to centromeres (Bodor et al., 2014) and that the enrichment of CENP-A nucleosomes at the centromeric region could explain why centromeres are only formed at a single position on each chromosome (Bodor, 2015). When the chromosomes separate, the CENP-A molecules at the centromere are randomly divided between the two copies. In this way, the memory of the centromere location is maintained. If the number of copies of CENP-A inherited by one of the chromosomes drops below a threshold value, a centromere will not form. CENP-A misincorporation into non-centromeric regions can promote neocentromere formation and chromosome instability (Mendiburo et al., 2011, Scott and Sullivan, 2014). To replenish centromeric chromatin following replicative dilution, existing CENP-A nucleosomes seed new CENP-A assembly during G1 (Westhorpe et al., 2015). Therefore, the number of CENP-A molecules seems to be important for controlling both the formation and the inheritance of the centromere. Bodor et al. (2014) suggest that centromeric CENP-A levels are regulated by a mass-action mechanism, where the

loading efficiency is independent of the expression levels. This mechanism ensures that with fluctuating expression levels, CENP-A remains mainly centromere restricted, and may prevent potential neocentromere seeding. The authors propose that the ample number of CENP-A nucleosomes facilitates a robust epigenetic signal that can absorb fluctuations in CENP-A inheritance and assembly in order to faithfully maintain centromere identity (Bodor et al., 2014).

Despite their considerable biological interest these mechanisms do not explain how the kinetochore, is able to stably associate with a variety of different DNA sequences in a way that persists over an apparently unlimited number of cell divisions.

## 1.6. Aims

A long-term goal is to understand the genetic and epigenetic features of the cell that are involved in centromere and kinetochore *de novo* specification and also their inheritance.

This thesis focuses on the development of a genetically tractable model that allow the study of some of the aspects that could be involved in the kinetochore inheritance. The developed system was used to test the idea that exists an evolutionarily conserved mechanism that regulates the amount of the centromere-specific histone H3 variant (CenH3) bound at centromeres independently of the length of the underlying centromeric DNA, as suggested to exist in human by previous work (Okamoto et al., 2007). I also addressed the question whether CenH3 binding requires a prior epigenetic mark. It was tested by integrating centromeric DNA at a non-centromeric locus and then, analysing the CenH3 binding to it. Finally, in order to analyse the idea that heterochromatin is required for *de novo* kinetochore formation as demonstrated some works (Ishii et al., 2008) , I deleted the gene encoding the *S. pombe* chromodomain protein Swi6.

# 2. MATERIALS AND METHODS

## 2.1. Chemicals and enzymes

Standard lab chemicals were purchased from Thermo Fisher Scientific, Roche and Sigma-Aldrich. New England Biolabs (NEB) supplied all the restriction enzymes. Taq DNA polymerase, including appropriate buffers were purchased from Thermo Fisher Scientific. Zymolase 20T and 100T were obtained from Sigma-Aldrich.

## 2.2. Fission yeast media

### 2.2.1. Stock solutions (Forsburg, 2015)

Stock solution of salt, vitamins, and minerals were prepared, filter sterilize and stored at room temperature.

50x Salt stock solution

52.5 g/l of $MgCl_2$, 0.735 g/l of $CaCl_2$, 50 g/l of KCl, and 2 g/l of $Na_2SO_4$

1000x Vitamin stock solution

1 g/l of pantothenic acid, 10 g/l of nicotinic acid, 10 g/l of inositol, 0.01 g/l of biotin

10,000x Mineral stock solution

5 g/l of boric acid, 4 g/l of $MnSO_4$, 4 g/l of $ZnSO_4$, 2 g/l of $FeCl_2$, 0.4 g/l of molybdic acid, 1 g/l of KI, 0.4 g/l of $CuSO_4$, 10 g/l of citric acid

### 2.2.2. Non-selective media (Hagan, 2016, Forsburg, 2015)

All non-selective media were autoclaved at 115°C for 10 minutes.

Yeast extract with supplements (YES)

5 g/l of yeast extract, 30 g/l of glucose, supplemented with 250 mg/l of Adenine, 250 mg/l of Histidine, 250 mg/l of Leucine, 250 mg/l of Lysine hydrochloride and 250 mg/l of Uracil. Solid media is made by adding 2% of agar.

Pombe Glutamate Medium (PMG)

3 g/l of potassium hydrogen phthalate, 2.2 g/l of $Na_2HPO_4$, 3.75 g/l of L-glutamic acid monosodium salt (Sigma-Aldrich), 20 g/l of glucose, 20 ml/l of 50x salt stock solution, 1 ml/l of 1000x vitamin stock solution, 0.1 ml/l of 10,000x mineral stock solution. Solid media is made by adding 2% of agar.

Transformation medium

1 g/l of Potassium hydrogen phthalate, 511 mg/l of $Na_2HPO_4$, 125 mg/l of L-glutamic acid, 2% w/v of glucose, 332 mg/l of $KH_2PO_4$, 162 mg/l of $MgSO_4$, 41 mg/l of NaCl, 655 mg/l of $CaCl_2$, 332 mg/l of Ammonium Sulphate, 240 mg/l of Potassium Acetate, 1 ml of 1000x vitamin stock solution, 0.1 ml of 10,000x mineral stock solution, 250 mg/l of Uracil, 250 mg/l of Leucine .

### 2.2.3. Selective media

These agents were used to select for or against the indicated marker on a plasmid or in the genome.

Yeast Extract Supplemented + nourseothricin antibiotic (YES+NTC)

YES media was supplemented with 100 μg/ml of nourseothricin (Sigma-Aldrich) antibiotic for resistance to nourseothricin resistance gene (*nat*). The YES was autoclaved at 115 °C for 10 minutes and then the antibiotic was added. Solid media is made by adding 2% of agar to the YES.

Pombe Glutamate Medium + Leucine (PMG + Leu)

PMG was supplemented by 225 μg/ml of Leucine for the growth of *leu1* mutants. The medium was autoclaved at 115 °C for 10 minutes. Solid media is made by adding 2% of agar.

Pombe Glutamate Medium + Uracil (PMG + Ura)

PMG supplemented by 225 μg/ml of Uracil for the growth of *ura4* mutants. The medium was autoclaved at 115 °C for 10 minutes. Solid media is made by adding 2% of agar.

5-fluoroorotic acid (5-FOA) media

5-FOA selects against *ura4*$^{+}$ generating a toxic metabolite in Ura+ strains.

0.1% of 5-FOA, 40 g/l of agar, 7g/l of Yeast Nitrogen Base (YNB) with $NH_4$ and without amino acids (Difco), and 2% of dextrose supplemented with 25 mg/l of Uracil, 225 mg/l of Adenine, 225 mg/l of Leucine, and 225 mg/l of Histidine, 10 g agar in 250 ml of distilled $H_2O$ was autoclave for 15 minutes, with stir bar, in 1 litter bottle. The sterilized agar was corrected for loss of $H_2O$ by adding sterile distilled $H_2O$, and store at 50°C. The appropriate amounts of YNB with $NH_4$ and without amino acids, 5-FOA, and aminoacids were added to distilled $H_2O$, to make a 250 ml solution which was filter sterilized. YNB/amino acids/5-FOA sterile solution was added to the sterile agar and mixed carefully. The plates were poured and stored at 4°C protected from light.

## 2.3. *S. pombe* growth and storage

### 2.3.1. Growth conditions

2 ml cultures were set up in test tubes with the required media using cells from a single colony. The tubes were incubated in rotatory shaker at 150 rotations per minute (rpm) at a constant temperature of 32°C overnight. This starter culture was used for setting up cultures on the day of the experiment. Cultures were grown in 250 ml Erlenmeyer flasks containing 50-100 ml of the required media incubated in rotatory shaker at 150 rpm at a constant temperature of 32°C to reach the desired cell density. Cells density was calculated based on $OD_{600}$ readings on spectrophotometer.

### 2.3.2. Short term storage

Yeast strains were streaked to single colonies on YES media plates. Plates were incubated at 32ºC for 2-3 days and could then be stored at 4ºC for up to two weeks.

### 2.3.3. Long term storage

Strains were stored by picking up a single colony and re-suspending it in 1 ml of 25% glycerol and frozen at -80ºC for long-term storage.

Cells were grown in glass test tube containing 2 ml of YES overnight to reach saturation. 500µl of the culture were added 2 cryovials tubes containing 500µl of 50% glycerol each, labelled, and frozen down at -80°C.

## 2.4. Molecular genetic methods

### 2.4.1. High efficiency transformation of *S. pombe*

*S. pombe* was transformed by high efficiency Lithium Acetate transformation using the protocol adapted from Okazaki et al. (1990). Yeast cells were transformed with a linear DNA for targeted integration into the genome by homologous recombination, or with plasmids to promote targeted integration in to the genome by site-specific recombination mediated by recombinase, and to express recombinases. Cultures were grown in 250 ml Erlenmeyer flasks containing 50-100 ml of the transformation medium incubated in rotatory shaker at 150 rpm at a constant temperature of 32°C. The cells were ready for transformation at an $OD_{600}$ of 1.0-2.0 (density of 0.5 -2 x $10^7$ cells/ml). Once the cells reached the required $OD_{600}$, they were harvested by centrifugation at 454 x*g* for 5 minutes at room temperature to pellet the cells. Supernatant was discarded and the pellet was re-suspended in 50 ml of sterile water and centrifuged at 454 x*g* for 5 minutes at room temperature. The cells were then re-suspended in 2 ml of 0.1 M lithium acetate (adjusted to pH 4.9 with acetic acid) and 1 ml transferred into an Eppendorf tube. 100-900 µg of plasmid DNA diluted in Tris-EDTA (TE) buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0) and in 10% of 1 M lithium acetate was added into the tube, and mixed by gentle vortexing. The combined weight of the mixture of cell and DNA was measured and 2.6 µl of 50% (w/v) Polyethylene glycol 4000 (PEG 4000) of the mixture was added. The tube was mixed by gentle vortexing, and incubated with mixing at 32°C for 60 minutes. After the incubation, 0.43 µl of dimethyl sulphoxide (DMSO) per milligram of the original mixture of DNA and cells was added to the mixture of cell, DNA and PEG 4000, taken to a heat shock in water bath at 40°C for 5 minutes, and then centrifuged at 1960 x*g* for 2 minutes in an benchtop centrifuge. The supernatant was removed by aspiration.

When the plasmid DNA contained an auxotrophic marker, the cells were resuspend in 1 ml of YES broth by pipetting up and down with a Gilson Pipetman P1000, the cell suspension was transferred into a 250 ml Erlenmeyer flask containing

50 ml of YES broth, and incubated in rotatory shaker at 150 rpm at a constant temperature of 32°C for 4 hours. After the incubation, the cells were centrifuged at 454 x*g* for 5 minutes at room temperature and re-suspended in 1.2 ml of YES broth. The plates were incubated at 32°C for 7 days, until colonies were clearly visible.

When the plasmid DNA contained nourseothricin antibiotic, after the heat shock and centrifugation, the supernatant was removed by aspiration. The cells were re-suspended in 2 ml of sterile water, and 0.3 ml of the cells were plated on each of four YES agar plates, and spread evenly using glass beads. The plates were incubated at 32°C for 24 hours and then replicated on YES added 100 μg/ml of nourseothricin antibiotic.

### 2.4.2. Genomic DNA Extraction from *S. pombe*

2 ml cultures were set up in test tubes with YES broth using cells from a single colony. The tubes were incubated in rotatory shaker at 150 rpm at a constant temperature of 32°C overnight. The cells were harvested by centrifugation at 341 x*g* for 5 minutes at room temperature to pellet the cells. The cells were then spheroplasted using Zymolyase 20T in 100 μl of 1 M sorbitol and 50 mM EDTA, incubated for two hours at 37°C, mixing periodically. Then the cells were concentrated by centrifugation at 341 x*g* for 5 minutes at room temperature, resuspended in 0.2 ml of DNAzol, and vortex mixed. The DNA was precipitated with an equal volume of ethanol. The crude DNA was treated with ribonuclease for 2 hours and then pronase at 200μg/ml in 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 0.1% SDS for 12 hours. The DNA was extracted three times with a 1:1 mixture of phenol and chloroform and finally precipitated with ethanol prior to use.

### 2.4.3. Gel Eletrophoresis

For conventional DNA electrophoresis gels were made with high melting temperature agarose (SeaKem® LE – catalogue number 50004) at 0.8%-1.5% w/v concentrations in 1 × TAE (1.2 g/l acetic acid, 6.3 g/l Tris-HCl, 470 mg/l EDTA, pH 8.0) or 1 × E buffer (370 mg/l $Na_2HPO_4$, 384 mg/l $NaH_2PO_4$, 292 mg/l EDTA, pH 8.0). Vertical gel boxes (17 cm × 17 cm and 4 mm thickness) were used for size fraction of high molecular weight DNA for Southern blotting and size fraction of DNA before gel purification of DNA fragments (Qiagen). Horizontal gel tanks bought

commercially were used for PCR and general analysis of restriction fragments. Gels were stained in Sterile Distilled Water (SDW) with approximately 100 µg/ml Ethidium bromide for 10 minutes and de-stained with SDW for 5 minutes. DNA was then visualised under UV light.

### 2.4.4. Restriction digestion of DNA

Restriction digestion of DNA was carried out according to the manufacturer's recommendations (NEB). All the digests were carried out in a total volume of 10 µl - 40 µl, containing 2-10 units of enzyme per 10 µl reaction. All digests were carried out for 4 – 8 hours at 37ºC, unless different temperature was optimized for the enzyme.

### 2.4.5. DNA extraction from agarose gels

If required, following electrophoresis, the DNA fragments were excised from the gel using a scalpel. QIAGEN gel extraction kit was used to extract the DNA from the gel following the manufacturer's instructions. The excised gel slice was melted at 50ºC in the binding solution provided in the kit, which allows binding of the DNA to the silica membrane of the column. The bound DNA was then washed with the wash buffer (provided by QIAGEN) supplemented with ethanol and eluted in 35 µl or 50 µl of elution buffer (provided by QIAGEN) depending on the concentration of the product required.

### 2.4.6. Ethanol precipitation of DNA

Ethanol precipitation of DNA allows the purification and concentration of the sample from a dilute solution by removing the unwanted salts. 0.3 M concentrations of sodium acetate and 3x volumes of 100% ethanol were added to the DNA sample and supernatant was discarded after centrifugation at 1960 x*g* for 10 minutes at room temperature. The resultant pellet was washed with 70% ethanol, centrifuged at 1960 x*g* for 1 minute at room temperature and the supernatant was discarded. The pellet was air dried and re-suspended in an appropriate volume of TE buffer.

### 2.4.7. Polymerase Chain Reaction (PCR)

PCRs were performed either to amplify specific DNA fragments. Standard *Taq* DNA polymerase (NEB) was mostly used, however Expand High Fidelity PCR

System (Roche) and LongAmp® *Taq* DNA Polymerase (NEB) were used to amplify fragments bigger than 4 kb.

### 2.4.8. *Taq* DNA Polymerase

A standard PCR reaction was carried out in a 10-50 µl reaction volume consisting of 10- 200 ng of template DNA, dNTPs (dATP, dCTP, dTTP, dGTP) at a final concentration of 0.25 mM each, 1 µM each of two oligonucleotide primers (forward and reverse primers), 1 unit of Taq polymerase in 1X reaction buffer (supplemented with $MgCl_2$ supplied by NEB) made to final volume with sterile distilled water. A typical PCR program consists of an initial denaturing step at 94°C for 2 minutes followed by 33 cycles of DNA denaturation at 94 °C for 10-15 seconds, primer annealing at 45°C-50°C for 10-15 seconds and primer elongation at 68°C for 1 minute per amplified 1 kb of DNA. All PCR reactions were run on a Bio Rad DNA Engine Thermal Cycler.

### 2.4.9. LongAmp® *Taq* DNA Polymerase (NEB)

The Long Amp was used to amplify fragments bigger than 4 kb. The genomic DNA was amplified with of 1X LongAmp *Taq* Reaction Buffer containing 2 mM $MgCl_2$, 300 µM dNTPs, 0.4 µM Forward Primer, 0.4 µM Reverse Primer, 10 ng – 500 ng genomic DNA, and 1 unit LongAmp *Taq* DNA Polymerase. PCR was carried out for 33 cycles using a initial denaturing step at 94°C for 2 minutes, 94ºC denaturation step for 15-30 seconds, an annealing step at 45°C-60°C for 15-60 seconds, elongation step at 65°C for 4-10 minutes, and a final elongation at 65°C for 10 minutes.

### 2.4.10. Long-range PCR

The Roche Expand long-range PCR was also used to amplify fragments bigger than 4kb. The genomic DNA was amplified with of 1X Expand *Taq* Reaction Buffer containing 17.5 mM $MgCl_2$, 350 µM dNTPs, 0.3 µM Forward Primer, 0.3 µM Reverse Primer, 10 ng – 500 ng genomic DNA, and 1 unit Expand *Taq* DNA Polymerase. PCR was carried out using an initial denaturation at 92° for 2 minutes followed by10 cycles using 92ºC denaturation step for 10 seconds, annealing step at 50°C for 10 seconds, and elongation step at 65°C for 12 minutes. Then, 25 cycles using, 94ºC denaturation step for 10 seconds, annealing step at 50°C for 10 seconds, and elongation step at 65°C for 12 minutes and increasing 20 seconds in each cycle. The final elongation was used at 65°C for 10 minutes.

### 2.4.11. List of Primers

| Code | Sequence |
|---|---|
| **WB35** | TAGTCGGTACCGAGTATTTCAAAGCTATTCAGC |
| **WB67** | TAGTCGCGGCCGCTGAGAAAGTCTTTGCTGATA |
| **106** | GCCTAAATCAAAATGAAGCTGA |
| **325** | ACGTGGATCCAGTATATTGGTCAATTAGAACCAA |
| **579** | CTACGTTTTTTGTTTCTTGCAGTTCGAAAAATCGTGCACATTTG |
| **581** | ATTTAGTCCAAGCTTTCGCGTGTTTGCTTACTTTTATTTTC |
| **585** | GTTTACAAATTTGTTTACTGAGCATTCAGCAAATTGCAAAAG |
| **646** | AAGACTGTCAAGGAGGGTATTCTG |
| **728** | CAAGTAGAATGAATTCGAATTG |
| **776** | TGGGTACCACTCTTGACGAC |
| **777** | GCCGAGTACGAGATGACCAC |
| **779** | CTATTACGAAGTAATTTCCGCTTC |
| **804** | CTTTTTAGAATACATTATACAATCTAG |
| **805** | GCTCACTACAATAAACAACTTTG |
| **884** | GATAAAACTAATTTTGTTTCGAATC |
| **893** | AAAGTTTCGTCAATATCACAG |
| **909** | AATGCACACATCACAAAGAACTATGTC |
| **1001** | GATTCGATACTAACGCCGCCATCCAG |
| **1030** | ACGCCCCCAACTGAGAGAACTCAAGG |
| **1041** | TTTAAAAATAATTGCTGTGAACTTTAGTAG |
| **1047** | CGAATTTCTCAAGAGGCAACATAG |
| **1119** | GTTGTAAGGTTCCCGCAAAA |
| **1120** | CATAATGGCAAATTATGGCAA |

**Table 2.1: List of primers used in this study for genomic PCR and sequencing.**

The primers were used as described in the table below:

| Primers | Usage |
|---|---|
| **WB35 and WB67** | make a probe for the *ura4*$^{+}$ gene |
| **106** | long range PCR to screen the Break Point 3 (BP3) (fig. 1.2) targeting on chr 4 |
| **325** | check the site-specific recombination at the left end (*attL*) of Nott373_cen7, Nott373_cen8, Nott376_cen7, and Nott376_cen8 |
| **579** | check the site-specific recombination at the left end (*attL*) of Nott373_cen1, Nott373_cen4, Nott376_cen1, and Nott376_cen4 |
| **581** | check the site-specific recombination at the left end (*attL*) of Nott373_cen3 and Nott376_cen3 |
| **585** | check the site-specific recombination at the left end (*attL*) of Nott373_cen2, Nott376_cen2, Nott373_cen5, and Nott376_cen5 |
| **646** | check the site-specific recombination at the right (*attR*) end of Nott373 and Nott376 |
| **728** | check the site-specific recombination at the left end (*attL*) of Nott373 clones |
| **776 and 777** | make the probe for the nourseothricin resistance gene |
| **779** | check the site-specific recombination at the right (*attR*) end of Nott373 clones |
| **804** | check the site-specific recombination at the right (*attR*) end of Nott376 clones |
| **805** | check the site-specific recombination at the left end (*attL*) of Nott376 clones |
| **884 and 893** | screen the Nott373 recombined clones targeted on the BP3 |
| **909** | checking the site-specific recombination at the left end (*attL*) of Nott373_alphoid and Nott376_alphoid |
| **1001** | checking the site-specific recombination at the left end (*attL*) of Nott373control |
| **1030** | long range PCR to screen the BP3 targeting on chromosome 2 |
| **1041** | check the excision of the native centromere 2 |
| **1047** | monitor *swi6* targeting |
| **1119 and 1120** | make the probe to check the *swi6* gene targeting |

**Table 2.2: Description of the experiments in which the primers were used.**

### 2.4.12. PCR product sequencing

Sequencing of PCR products was carried out in 10 µl containing 0.5 µl of BigDye Terminator v. 3.1 mix (Applied Biosystems), 2 µl of dilution buffer (Applied Biosystems), 5-10 pmol of primer, and 0.2 µg of template DNA in a final reaction volume of 10 µl. Cycle conditions were an initial denaturation at 96°C for 2 minutes, then 25 cycles of 96°C for 30 seconds, 50°C for 15 seconds, and 60°C for 4 minutes. Excess dye terminators were removed using CleanSeq magnetic bead sequencing reaction clean up kit from Beckman Coulter (Agencourt

Bioscience). Sequencing was performed by DBS Genomics, University of Dunham on an Illumina HiSeq.

### 2.4.13. Southern Blot (Southern, 1975)

Southern blot combines transfer of electrophoresis-separated DNA fragments to a filter membrane and subsequent fragment detection by hybridization with radioactive probe.

The gel containing size fractionated DNA was treated with denaturation solution (20 g/l of NaOH, 29.2 g/l of NaCl) for 40 minutes. Then it was treated with neutralising solution (158.6 g/l Tris-HCl, 87.7 g/l NaCl, pH 7.4) for 2 x 30 minutes. The blotting apparatus consisting of the blotting tank containing 10 × Saline Sodium Citrate buffer (SSC) (44.1 g/l of trisodium citrate, 87.6 g/l of NaCl, pH 7.0) and pre-soaked wick. Filter paper cut 2 cm larger than the size of the gel was washed in 2 × SSC buffer (8.82 g/l of trisodium citrate, 17.52 g/l of NaCl, pH 7.0) and laid on top of the wick. The pre-treated gel was laid on top of the filter paper. Parafilm was placed around the gel so that the buffer moved up through the gel only by capillary action and not from the surrounding region. GeneScreen® (PerkinElmer Life Sciences) nylon membrane was washed in 2 × SSC buffer, excess moisture was removed by placing on filter paper, and the membrane was then placed on top of the gel. Another filter paper soaked in 2 × SSC buffer was placed on top of the nylon membrane. Finally, stacks of paper towels were placed on the filter paper, to induce the capillary action overnight.

The nylon membrane bound to DNA transferred from the gel was removed washed in 2 × SSC buffer and allowed to air dry. The nylon membrane was then baked to 65°C in a vacuum oven followed by exposure to short wavelength UV (254 nm) with doses between 0.6-0.8 kJ/m$^2$ for 9 seconds to cross-link the DNA fragments and the nylon membrane.

### 2.4.14. Random primed labeling and filter hybridization

Random primed labeling, based on Feinberg and Vogelstein (1983) is a method of incorporating radioactive nucleotides along the length of a fragment of DNA. In this method a DNA fragment is denatured by heating in a boiling water bath. Then, random sequence oligonucleotides are annealed to both strands. Klenow fragment

polymerase is then used to extend the oligonucleotides, using three cold nucleotides and one radioactively labeled nucleotide provided in the reaction mixture, to produce a uniformly labeled double-stranded probe.

100 ng of DNA probe and random hexamers ($dN_6$) in sterile distilled water was denatured at 95°C for 7 minutes then incubated on ice for 2 minutes. Oligo labelling buffer, 200 μM of premixed solution containing the sodium salts of dATP, dGTP, and dTTP, 3 μl of $^{32}$P dCTP (30 μCu) and 1 unit of Klenow fragment DNA polymerase were added, the reaction mixture vortexed and incubated at 37°C for 1 hour. 90 μl of NTE (29.22 g/l of NaCl, 1.21 g/l of Tris-HCl, 14.6 g/l of EDTA, pH 8.0) was added to the labelled probe reaction mixture. The reaction mixture was then run through a NTE equilibrated Sephadex® G50 (Sigma-Aldrich) spun column to separate the labelled DNA from unincorporated α32-P dCTP. The labelled DNA was then denatured at 95°C for 7 minutes and added to the nylon membrane pre-hybridising in 20 ml of Church mix (71 g/l of Sodium phosphate, 7% SDS, pH 7.4) at 60ºC.

Hybridisation was carried out overnight. The filter was washed in the appropriate wash buffer (based on the stringency of the required wash) 3 x 20 minutes.

When probing for repeated sequences, low stringency wash buffer (0.5% SDS, 0.2x SSC) was used but when probing for single copy gene high stringency wash buffer (0.5% SDS, 0.4x SSC) was used. Filters were dabbed partially dry by paper towel, laid on waxed Benchcote sheet DNA side up and wrapped with cling film.

Exposure to intensifying screen was carried out for 6 hours and scanned on a phosphoimager and viewed by ImageQuant TL 8.1 (GE Healthcare Life Sciences) computer software.

In order to hybridise with different probes, the filter was washed with boiling in a solution of 0.5% of SDS and 0.1x TE buffer for 30 minutes.

### 2.4.15. ChIP sequencing

The strains were tagged on the N-terminal of the centromeric specific protein Cnp1 (CENP-A homolog) using Green Fluorescent Protein (GFP) as described by Coffman et al. (2011). The tagged protein was submitted to chromatin immunoprecipitation (ChIP) following the protocol described below. Control

experiments in which no antibody was used established that the ChIP was antibody- and tag-specific.

### 2.4.16. Library preparation

Fixation of the cells:

150 ml of log-phase growing cells ($5x10^6$ – $2x10^7$ cells/ml) were taken, harvested by centrifugation at 454 x*g* for 5 minutes at room temperature, supernatant removed, pellet washed 2 times with 25 ml of cold Phosphate Buffer Saline solution (PBS – 137mM NaCl, 2.7mM KCl, 10mM $Na_2HPO_4$, $KH_2PO_4$ 1.8mM), and re-suspended in 50 ml of PBS. 1.35 ml of 37% paraformaldehyde was added, swirled slowly for 30 minutes at room temperature, quench fixation was carried out by the addition of 2.5 ml of 2.5 M glycine and swirled for 5 minutes at room temperature. The cells were transferred to a 50 ml centrifuge tube, pelleted by centrifugation for 5 minutes at 1900 x*g* at 4°C, and washed 2 times with 25 ml of cold PBS.

### 2.4.17. Preparation of the magnetic beads for immunoprecipitation

For each sample was used 45 µl of 30 mg/ml Dynabeads™ protein A (Thermo Fisher). The Dynabeads™ were put in an Eppendorf tube and centrifuged at 1960 x*g* in a microcentrifuge for 1 minute at 4°C, the supernatant was removed, and re-suspend in 0.5 ml of cold PBS containing 5 mg/ml of Bovine Serum Albumin (BSA). The washing step was repeated and then the Dyna beads were re-suspended in 15 µl of the BSA in PBS solution, added anti-GFP antibody (ab290 - Abcam), and incubated overnight with rotation at 4°C. Before use it was washed 2 times with the BSA in PBS solution.

### 2.4.18. Preparation of the chromatin extract

40 µl of a 25X stock solution of protease inhibitor (PI) cocktail (Roche 1873580) was added to 1 ml of lysis buffer (150mM NaCl, 50mM HEPES-KOH, 1% Triton X-100, 0.1% Sodium deoxycholate, 1mM EDTA, pH 7.5), the fixed cells pellet were resuspended in 0.8 ml of cold lysis buffer with PI, and transferred to pre-cooled tube, filled with acid-washed glass beads, and the cells were lysed by shaking in an MP Bio Fast-Prep 24 at maximum speed 12 times for 1 minute each . The bottom of the tube was punctured with a needle and placed in a 2 ml Eppendorf. The tubes were centrifuged at 706 x*g* for 10 minutes at 4°C to separate the chromatin from the glass

beads. The chromatin was transferred to a fresh 1.5 ml Eppendorf tube and the lysate sonicated using a probe sonicator on ice to shear the chromatin 10 times for 25 seconds. The cells debris were separated by centrifugation for 5 minutes at 17640 x*g* at 4°C, and the supernatant transferred (chromatin extract) to a fresh tube. 15µl magnetic beads + antibodies were added and incubated overnight at 4°C while inverting.

### 2.4.19. Magnetic beads wash after the immunoprecipitation

The beads were washed using a magnetic stand as follows:

- 2 times 1 ml of cold lysis buffer
- 2 times 1 ml of cold lysis buffer plus 360mM NaCl
- 2 times 1 ml of cold wash buffer (10mM Tris-HCl pH 8, 250mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 1mM EDTA)
- 2 times 1 ml of cold TE, pH 8.

The beads were harvested by centrifugation at 706 x*g* in a microfuge for 3 minutes, the supernatant removed, 40 µl of elution buffer (TES – 50mM Tris-HCl pH 8, 1.5% SDS, 1mM EDTA) was added, and incubated at 65°C for 15 minutes. Concentrated by centrifugation for 1 minute at 17,640 x*g*, the supernatant transferred to a fresh Eppendorf tube, and 4 volumes of TE/1% SDS was added. The chromatin was incubated overnight at 65°C to reverse the crosslink.

### 2.4.20. DNA Clean-up

100 µl of the chromatin extract was taken and added 97 µl of TE, 1 µl of 100 mg/ml dextran, and 4 µl of 100 mg/ml RNAase; incubated at 37°C for 2 hours, then 4 µl of 50 mg/ml proteinase K was added and incubated at 50°C for 4 hours. NaCl was added to a final concentration of 200mM, mixed and extracted 2 times with the same volume of phenol/chloroform/isoamyl alcohol; precipitated with EtOH and re-suspend in 25 µl of TE.

### 2.4.21. Sequencing

The 20 ChIP samples, including controls and inputs, were multiplexed using Illumina DNA barcoding so they could be sequenced as a pool in Illumina HiSEq 2000 using 100 bp paired-end reads by the project manager at the High-Throughput Genomics Wellcome Trust Centre for Human Genetics, Oxford, UK.

### 2.4.22. Computational analysis of ChIP sequencing data

The reads were firstly processed using the Sickle (Josh, 2011) with standard parameters for quality trimming and the Scythe (Buffalo, 2011) with standard parameters for PCR adapter trimming. The trimmed reads were 75 basepairs on average. The BWA-mem (Li and Durbin, 2009) was used for all the alignments with standard parameters, to the genome *Escherichia coli* strain K12, sub-strain MG1655 (GenBank accession number U00096) to remove possible contaminants.

The genome of the engineered strains was built on Vector NTI® (Thermo Fisher Scientific) using the previously sequenced genome of the start strain CBS 2777 (Brown et al., 2014) and the sequence of the plasmids used for the engineering. The reads unmapped to the *E. coli* genome were aligned, using BWA-mem with standard parameters, to the correspondent *S. pombe* genome. This resulted in a primary mapped reads. Then the reads mapped in the correct orientation and within insert size were selected using Samtool flags. Unmapped reads were removed from the bam files and the total number of reads mapping to each chromosome were then determined using the following chromosome coordinates:

- chromosome I: 2926051-2939353
- chromosome II: 125168-133320
- chromosome III: 1092658-1106291
- chromosome IV: 1619693-1627846

In some cases, it was observed the formation of a neo-centromere on chromosome II, plotting the whole chromosome 2 coverage for the reads it was observed that the neo-centromere was formed over the same region for all cases, so it was possible to determine the coordinates for this region, which was:

- chromosome II neo-centromere: 3855830-3867330

Coverage vectors where created for using the R package Bioconductor (Huber et al., 2015), then coverage plots were created in R. The plots were smoothed using the default running medians scatter plot smoothing in R. For coverage plots across entire chromosomes data points were plotted every 1000 base pairs, and to plot the centromeric regions data points were plotted every 10 base pairs.

## 2.5. Strains used or built for this project

| Strain | Genotype | Source |
|---|---|---|
| Nott 76 | Wild type | CBS2777 |
| Nott 273 | *CBS 2777 leu1-D32 ura4-D18* | Laboratory collection |
| Nott 363 | *CBS 2777 leu1-D32 ura4-D18* GFP-*cnp1* | Laboratory collection |
| Nott 373 | *CBS 2777 leu1-D32 ura4-D18* GFP-*cnp1* cen2R:: ΦC31 *attB ura4*$^{+}$ ΦC31 *attB* | Laboratory collection |
| Nott 376 | *CBS 2777 leu1-D32 ura4-D18* GFP-*cnp1 ade6D*:: ΦC31 *attB ura4*$^{+}$ ΦC31 *attB* | Laboratory collection |
| Nott 453 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 1* | This study |
| Nott 454 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 3* | This study |
| Nott 455 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 6* | This study |
| Nott 459 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB nat - clone 2* | This study |
| Nott 460 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB nat - clone 3* | This study |
| Nott 461 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB nat - clone 4* | This study |
| Nott 465 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen1 nat - clone 1* | This study |
| Nott 466 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen1 nat - clone 5* | This study |
| Nott 467 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen1 nat - clone 6* | This study |
| Nott 468 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen2 nat - clone 2* | This study |
| Nott 469 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen2 nat - clone 8* | This study |
| Nott 470 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen3 nat - clone 3* | This study |
| Nott 471 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen3 nat - clone 5* | This study |
| Nott 472 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen4 nat - clone 3* | This study |
| Nott 473 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen4 nat - clone 4* | This study |
| Nott 474 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen4 nat - clone 5* | This study |
| Nott 475 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen5 nat - clone 1* | This study |

| | | |
|---|---|---|
| Nott 476 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen5 nat - clone 5* | This study |
| Nott 477 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen5 nat - clone 6* | This study |
| Nott 478 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen7 nat - clone 3* | This study |
| Nott 479 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen7 nat - clone 4* | This study |
| Nott 480 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen8 nat - clone 6* | This study |
| Nott 481 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB alphoid nat - clone 3* | This study |
| Nott 482 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB alphoid nat - clone 4* | This study |
| Nott 483 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB alphoid nat - clone 5* | This study |
| Nott 484 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen1 nat - clone 1* | This study |
| Nott 485 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen1 nat - clone 2* | This study |
| Nott 486 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen1 nat - clone 3* | This study |
| Nott 487 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen2 nat - clone 1* | This study |
| Nott 488 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen2 nat - clone 2* | This study |
| Nott 489 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen2 nat - clone 3* | This study |
| Nott 490 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen4 nat - clone 1* | This study |
| Nott 491 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen4 nat - clone 2* | This study |
| Nott 492 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen4 nat - clone 3* | This study |
| Nott 493 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat - clone 1* | This study |
| Nott 494 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat - clone 3* | This study |
| Nott 495 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat - clone 4* | This study |
| Nott 496 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB alphoid nat - clone 1* | This study |
| Nott 497 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB alphoid nat - clone 2* | This study |
| Nott 498 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB alphoid nat - clone 3* | This study |
| Nott 499 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen1 nat - clone 2* | This study |

| | | |
|---|---|---|
| Nott 500 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen1 nat - clone 5* | This study |
| Nott 501 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen1 nat - clone 7* | This study |
| Nott 502 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen2 nat - clone 1* | This study |
| Nott 503 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen2 nat - clone 2* | This study |
| Nott 504 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen2 nat - clone 3* | This study |
| Nott 505 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen4 nat - clone 1* | This study |
| Nott 506 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen4 nat - clone 2* | This study |
| Nott 507 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen4 nat - clone 4* | This study |
| Nott 508 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat - clone 3* | This study |
| Nott 509 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat - clone 5* | This study |
| Nott 510 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat - clone 6* | This study |
| Nott 511 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB alphoid nat - clone 1* | This study |
| Nott 512 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB alphoid nat - clone 2* | This study |
| Nott 513 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB alphoid nat - clone 4* | This study |
| Nott 514 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen8 nat - clone 1* | This study |
| Nott 515 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen3 nat - clone 2* | This study |
| Nott 516 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen7 nat - clone 1* | This study |
| Nott 517 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen3 nat - clone 1* | This study |
| Nott 518 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen3 nat - clone 10* | This study |
| Nott 519 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen8 nat - clone 1* | This study |
| Nott 520 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen8 nat - clone 2* | This study |
| Nott 521 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen3 nat - clone 1* | This study |
| Nott 522 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen3 nat - clone 2* | This study |
| Nott 523 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen3 nat - clone 3* | This study |

| | | |
|---|---|---|
| Nott 524 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen3 nat - clone 4* | This study |
| Nott 525 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen7 nat - clone 1* | This study |
| Nott 526 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen7 nat - clone 2* | This study |
| Nott 527 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen7 nat - clone 3* | This study |
| Nott 528 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen8 nat - clone 1* | This study |
| Nott 529 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 ade6D:: Bxb1 attB Cen8 nat - clone 2* | This study |
| Nott 530 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 1* | This study |
| Nott 531 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 2* | This study |
| Nott 532 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 3* | This study |
| Nott 533 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 4* | This study |
| Nott 534 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat - clone 5* | This study |
| Nott 584 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen1 nat - clone 1* | This study |
| Nott 585 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen2 nat - clone 14* | This study |
| Nott 586 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen4 nat - clone 32* | This study |
| Nott 587 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen4 nat - clone 36* | This study |
| Nott 588 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen5 nat - clone 1* | This study |
| Nott 589 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB alphoid nat - clone 1* | This study |
| Nott 590 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB alphoid nat - clone 8* | This study |
| Nott 591 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen3 nat - clone 25* | This study |
| Nott 592 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen3 nat - clone 29* | This study |
| Nott 593 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen8 nat - clone 3* | This study |
| Nott 594 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen8 nat - clone 4* | This study |
| Nott 595 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen8 nat - clone 5* | This study |
| Nott 596 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB nat - clone 2* | This study |

| | | |
|---|---|---|
| Nott 597 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB nat - clone 4* | This study |
| Nott 598 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB nat - clone 6* | This study |
| Nott 628 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 1* | This study |
| Nott 629 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 2* | This study |
| Nott 630 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB nat - clone 3* | This study |
| Nott 631 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 4* | This study |
| Nott 632 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 5* | This study |
| Nott 633 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 6* | This study |
| Nott 634 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 7* | This study |
| Nott 635 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 8* | This study |
| Nott 636 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 9* | This study |
| Nott 637 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ nat - clone 10* | This study |
| Nott 638 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 1* | This study |
| Nott 639 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 2* | This study |
| Nott 640 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 3* | This study |
| Nott 641 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 4* | This study |
| Nott 642 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 5* | This study |
| Nott 643 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 6* | This study |
| Nott 644 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 7* | This study |
| Nott 645 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 8* | This study |
| Nott 646 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 9* | This study |
| Nott 647 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen1 nat - clone 10* | This study |
| Nott 648 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 1* | This study |
| Nott 649 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 2* | This study |

| | | |
|---|---|---|
| Nott 650 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 4* | This study |
| Nott 651 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 5* | This study |
| Nott 652 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 6* | This study |
| Nott 653 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 7* | This study |
| Nott 654 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 8* | This study |
| Nott 655 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 9* | This study |
| Nott 656 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 10* | This study |
| Nott 657 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen2 nat - clone 11* | This study |
| Nott 658 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 1* | This study |
| Nott 659 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 2* | This study |
| Nott 660 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 3* | This study |
| Nott 661 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 4* | This study |
| Nott 662 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 5* | This study |
| Nott 663 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 6* | This study |
| Nott 664 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 7* | This study |
| Nott 665 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 8* | This study |
| Nott 666 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 9* | This study |
| Nott 667 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 10* | This study |
| Nott 668 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 1* | This study |
| Nott 669 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 2* | This study |
| Nott 670 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 3* | This study |
| Nott 671 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 4* | This study |
| Nott 672 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 5* | This study |
| Nott 673 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 6* | This study |

| | | |
|---|---|---|
| Nott 674 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 7* | This study |
| Nott 675 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 8* | This study |
| Nott 676 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 9* | This study |
| Nott 677 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen4 nat - clone 10* | This study |
| Nott 678 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 1* | This study |
| Nott 679 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 2* | This study |
| Nott 680 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 3* | This study |
| Nott 681 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 4* | This study |
| Nott 682 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 5* | This study |
| Nott 683 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 6* | This study |
| Nott 684 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 7* | This study |
| Nott 685 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 8* | This study |
| Nott 686 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 9* | This study |
| Nott 687 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen5 nat - clone 10* | This study |
| Nott 688 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 1* | This study |
| Nott 689 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 2* | This study |
| Nott 690 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 3* | This study |
| Nott 691 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 4* | This study |
| Nott 692 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 5* | This study |
| Nott 693 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 6* | This study |
| Nott 694 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 7* | This study |
| Nott 695 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 8* | This study |
| Nott 696 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 9* | This study |
| Nott 697 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ Cen3 nat - clone 10* | This study |

| | | |
|---|---|---|
| Nott 698 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 1* | This study |
| Nott 699 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 2* | This study |
| Nott 700 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 3* | This study |
| Nott 701 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 4* | This study |
| Nott 702 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 5* | This study |
| Nott 703 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 6* | This study |
| Nott 704 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 7* | This study |
| Nott 705 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 8* | This study |
| Nott 706 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 9* | This study |
| Nott 707 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2Δ alphoid nat - clone 10* | This study |
| Nott 723 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat* ΔSwi6 - clone 3 | This study |
| Nott 724 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat* ΔSwi6 - clone 4 | This study |
| Nott 725 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB Cen5 nat* ΔSwi6 - clone 5 | This study |
| Nott 726 | *CBS 2777 leu1-D32 ura4-D18 GFP-cnp1 cen2R:: Bxb1 attB nat* ΔSwi6 - clone 14 | This study |

**Table 2.3: The stains used in this study (deposited in the lab collection) their genotype and source.** BP4: breakpoint 4 on the re-arranged chromosome 2 of the strain CBS2777.

# 3. RECOMBINING MUTANT CENTROMERIC DNA FRAGMENTS IN THE *S. POMBE* CBS 2777 CHROMOSOME 2 USING THE SERINE RECOMBINASE FROM THE *STREPTOMYCES* BACTERIOPHAGE ϕC31

## 3.1. Introduction

In humans, centromeres are formed by the tandemly repeated sequence called alphoid DNA. Alphoid DNA shows length polymorphism (Lee et al., 1997) consistent with its tandemly repeated nature. Okamoto et al. (2007) demonstrated that different lengths of human alphoid DNA binds approximately the same amount of CENP-A (figure 3.1), suggesting the existence of an epigenetic mechanism that regulates the assembly and maintenance, and ensures a fixed amount of CENP-A independently of the amount of centromeric DNA (Brown and Xu, 2009).

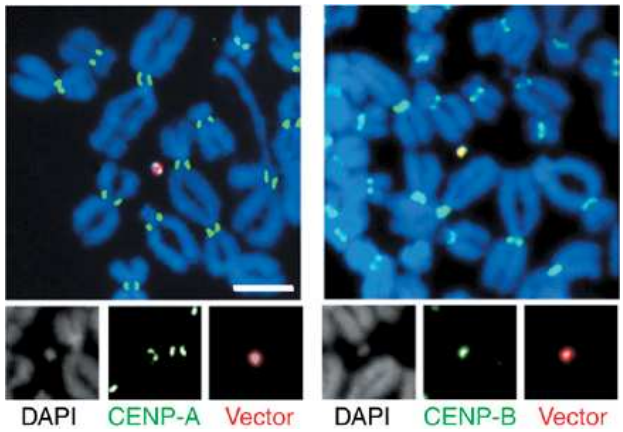



**Figure 3.1: Analysis of CENP-A chromatin in HACs.**
Indirect immunostaining and FISH analysis of xxx. Green fluorescent secondary antibodies were used to detect CENP-A and CENP-B. Adapted from (Okamoto et al., 2007).

If such a mechanism exists then it would likely to be evolutionarily conserved. In my project, I wanted to determine whether such a mechanism existed in the genetically tractable model organism; the fission yeast *S. pombe*. One way to address the question is to vary the amount of centromeric DNA. If an evolutionarily conserved system that regulates the number of centromeric nucleosomes exists, then the number of centromeric nucleosomes will be fixed, independently of the amount of centromeric DNA. The re-arranged centromeres of the *S. pombe* strain CBS 2777 allowed me to test this prediction. Chromosomes 2 and 4 of the CBS 2777 contain centromeric

central core sequences but lack the outer repeats domain that is present in the centromeres of the commonly used laboratory strain. In CBS 2777, the centromere central core sequence is flanked by two segments of DNA, each ~100kb long, that are normally present in centromere-distal regions of the short arms of the laboratory strain chromosomes (Brown et al., 2014). The absence of the outer repeats domains enables one to efficiently modify the centromeric flanking regions by a combination of sequence targeting and site-specific recombination techniques.

In this chapter, the large serine recombinase ϕC31 was used as a tool for promoting the integration DNA fragments into the centromere of chromosome 2 of the *S. pombe* strain CBS 2777.

Six different fragment sizes of *S. pombe* centromeric DNA ranging from 1kb to the size of the native centromere; 9.4kb, and a 5.7kb fragment of human Y chromosome alphoid DNA have been placed adjacent to the endogenous centromere of CBS 2777 chromosome 2 together with an *attB* site for the Bxb1 integrase. After the introduction of the mutant centromeric DNA and an *attB* site for the Bxb1 integrase into the right flank of the centromere sequence, targeting was used to place a Bxb1 *attP* site into the left flank of the centromere, completing the construction of a system that allowed subsequent excision of the native centromere, but left behind different sizes of candidate centromeric DNA. In this way, I have been able to assemble centromeres of different sizes and to establish whether centromere protein binding is independent of the amount of underlying centromeric DNA. The *S. pombe* centromeric sequences that I introduced were derived from the laboratory strain and in addition were mutated with 4 base pair deletions every 1kb in order to facilitate discrimination from the resident centromeric DNA (WRA Brown, unpublished). We assumed that these mutations does not affect the Cnp1 binding because there is a natural polymorphism between the laboratory strain and the strain CBS 2777. The introduction of human alphoid DNA let me analyse sequence specificity of the DNA sequence for the kinetochore assembly. I also integrated an empty vector thus allowing excision of the native centromere without leaving any candidate centromeric sequence behind.

## 3.2. Integration of test sequences on chromosome 2

The system was built using the strain Nott373 from the laboratory collection (table 2.2), which was derived from the *S. pombe* isolate CBS 2777. Nott373 has the Cnp1 protein tagged at the amino terminus with GFP and is *leu1*Δ and *ura4*Δ. The strain was targeted on the right-hand side of the centromere of chromosome 2 with a cassette containing a pair of *attB* sites for the ϕC31 integrase flanking an *ura4* gene as a selectable marker (figure 3.2).

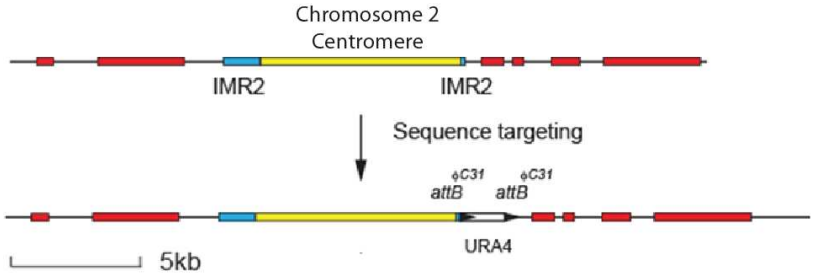



**Figure 3.2: Centromere of the chromosome 2 of the *S. pombe* strain CBS 2777.**
The centromere 2 was targeted with *attB* sites for the ϕC31 integrase and *ura4+* gene.

The ϕC31 integrase was used for integrating candidate centromeric DNA fragments in the targeted region shown in figure 3.2. Firstly, the strain was transformed with a plasmid named NL314, this is a REP1 plasmid carrying the autonomously replicating sequence 1 (ARS1), the *Saccharomyces cerevisiae LEU2* gene as a selectable marker, the nmt1 promoter, and a codon optimised ϕC31 integrase gene (figure 3.3). PMG medium supplemented with 150 μg/ml of uracil and 5 μg/ml of thiamine selected for the clones carrying the marker and the thiamine repressed the nmt1 promoter. We obtained clones from this transformation which were used for the site-specific recombination mediated by the ϕC31 integrase.

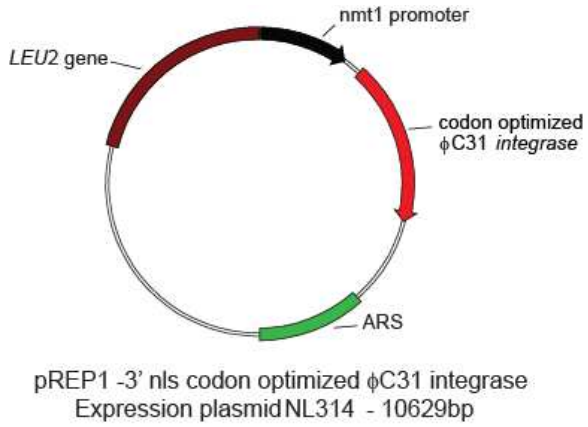



**Figure 3.3: DNA plasmid NL314, a ϕC31 integrase expression plasmid.**
The plasmid is derived from pRep1 and expresses a ϕC31 gene that is codon optimised for expression.

The plasmids used for the site-specific recombination were previously engineered by William Brown. The plasmids are pFA6 vectors (Bahler et al., 1998), including a natMX6 marker, and carrying the candidate centromeric DNA fragments (figure 3.4), two *attP* sites for the ϕC31 integrase, and a single *attB* site for the Bxb1 integrase that will be used for the excision of the native centromere sequence; and the nourseothricin resistance gene as selectable marker in reverse position with respect to the polarity of the ϕC31 *attP* sites so that transcription was towards the integrated DNA at the centromere.

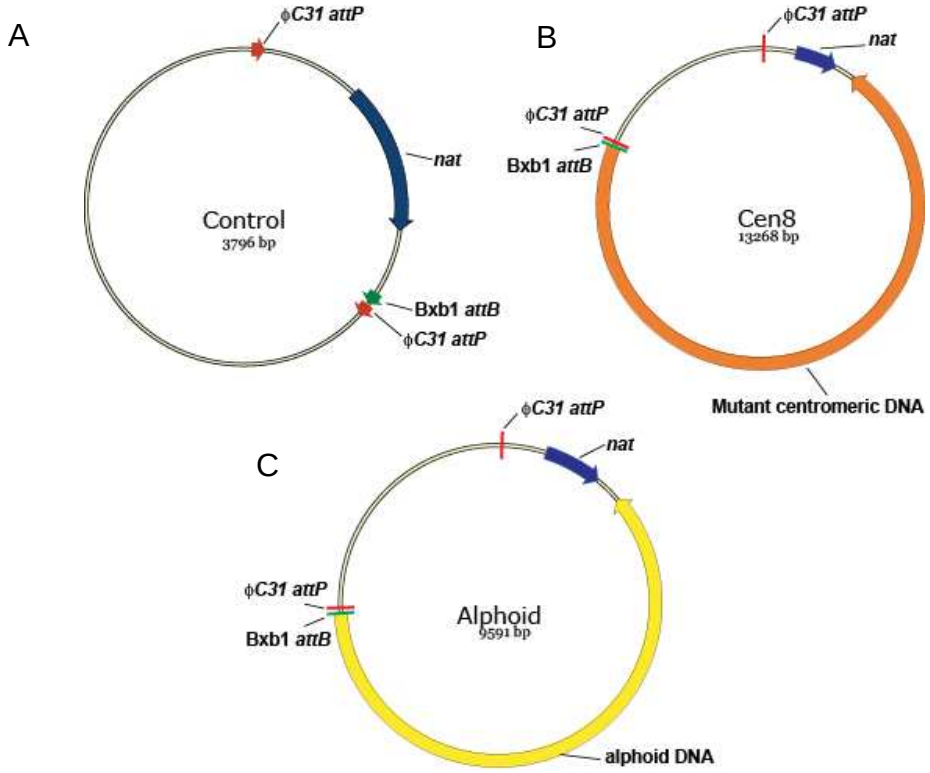



**Figure 3.4: DNA plasmids used for the site-specific recombination.**
The engineered plasmids contain a pair of *attP* sites for the ϕC31 integrase; an *attB* site for the Bxb1 integrase; and nourseothricin resistance gene (nat). (A) Control: plasmid containing no candidate centromeric sequence. (B) Cen8: plasmid containing a 9.4kb mutant *S. pombe* centromeric DNA sequence, this is one of the seven plasmids carrying *S. pombe* centromeric DNA, the other candidate sequences range from 1kb to an inverted version of the 9.4kb test sequence. (C) Alphoid: plasmid containing the 5.7kb alphoid DNA repeat from the human Y chromosome.

In preliminary experiments, plasmids containing the resistance gene oriented so that transcription was away from the integration site position were used and the results showed that the marker gene was inverted by re-arrangement after integration suggesting that placing a promoter adjacent to the kinetochore is selected against.

Nine different plasmids were engineered (table 3.1), seven of which contain *S. pombe* mutant centromeric DNA ranging from 1kb to the size of the central core of the native centromere that is 9.4kb (figure 3.4B). The centromeric DNA fragments are

mutant due to the 4 bp deletions present every 100 kb across the sequence that enables it to be distinguished from the endogenous centromeres by sequence polymorphism. The position of the deletions were randomly selected. Another donor plasmid contained a 5.7kb fragment of human Y chromosome alphoid DNA (figure 3.4C). Finally, an empty vector was used as a control (figure 3.4A), thus the nourseothricin resistance gene will be next to the centromere.

| Plasmid | Description | Size of the centromeric DNA |
|---|---|---|
| Control | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-ϕC31*attP* | 0kb (control) |
| Cen1 | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**Cen1**-ϕC31*attP* | 4kb |
| Cen2 | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**Cen2**-ϕC31*attP* | 3.6kb |
| Cen3 | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**Cen3**-ϕC31*attP* | 2kb |
| Cen4 | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**Cen4**-ϕC31*attP* | 1kb |
| Cen5 | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**Cen5**-ϕC31*attP* | 6.8kb |
| Cen7 | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**Cen7**-ϕC31*attP* | 9.4kb (inverted) |
| Cen8 | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**Cen8**-ϕC31*attP* | 9.4kb |
| Alphoid | pFA6a-natMX6-ϕC31*attP*-Bxb1*attB*-**alphoid**-ϕC31*attP* | 5.7kb |

**Table 3.1: Plasmids used in this study for the ϕC31 integrase mediated site-specific recombination and the sizes of the test sequences.**
kb = kilobase pairs.

Figure 3.5 summarises the sequence organisation of the inserts in these plasmids. Cen7 and Cen8 comprise the centromere central core of the laboratory strain, including the inner repeats. The difference between these two plasmids is the orientation of the sequence. Cen1, Cen2, Cen3, Cen4, and Cen5 are smaller fragments of the centromere central core of the laboratory strain, without the inner repeats.

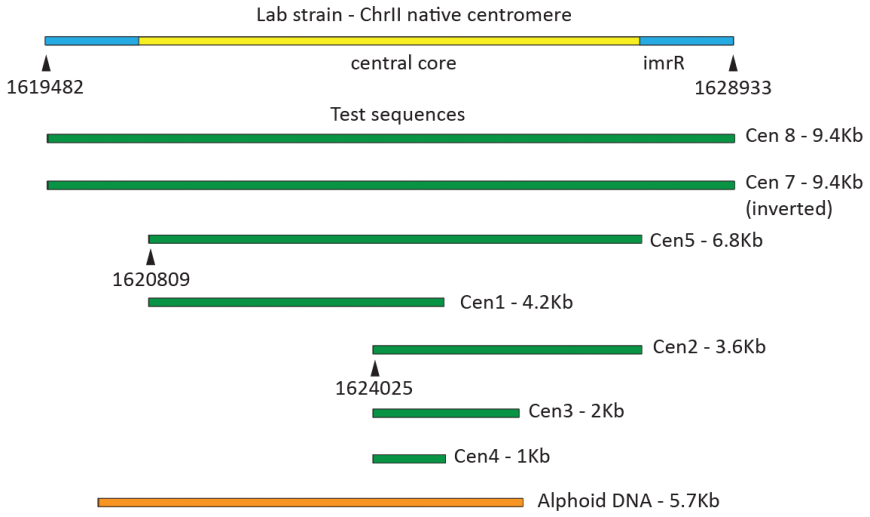



**Figure 3.5: Test sequences used to build the plasmids for site-specific recombination.**

The mutant centromeric sequences are fragments from the chromosome II of the *S. pombe* laboratory strain and contain 4 bp deletions across the sequence. The alphoid DNA fragment is derived from the human Y chromosome (Tylersmith and Brown, 1987).

The transformations were performed as described in the material and methods section with the Nott373 strain containing the ϕC31 integrase. The concentration of the cells was 1-2x$10^7$ cell/ml and concentration of the plasmids was 100 μg. The cells were plated on YES added 100 μg/ml of nourseothricin. The recombined strains are *ura4*Δ and were screened on PMG agar lacking Uracil. The clones that have lost the *ura4+* gene are the ones that do not grow on PMG+LEU medium (Figure 3.7).

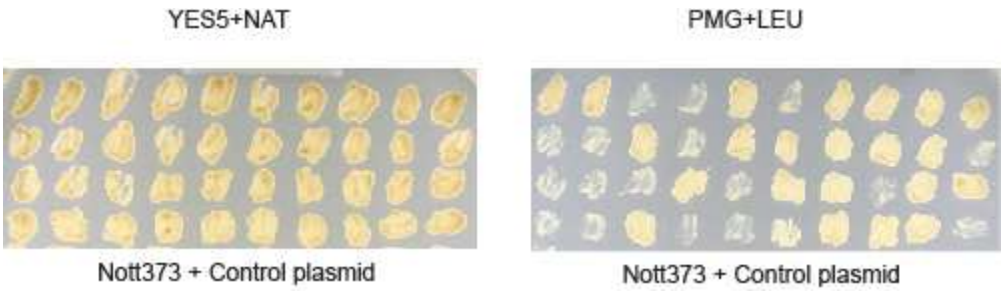



**Figure 3.6: Screening *ura4Δ* clones.**
On the left are the clones selected on the medium YES added 10μg/mL of nourseothricin (NAT). On the right are the same clones tested on the medium PMG supplemented by 225μg/mL of Leucine (LEU). The clones that have lost the *ura4+* gene do not grow on PMG+LEU medium.

The *ura4*Δ clones were named using the number of the start strain (Nott373) followed by the name of the plasmid. The transformation with each of the nine plasmids gave *ura4*Δ clones which were screened for site-specific recombination by PCR, as described in the next section.

### 3.2.1. Screening for site-specific recombination by PCR

The *ura4Δ* clones were submitted to PCR reactions for checking the site-specific recombination in both *attL* and *attR* sites of the recombinant sequence. The nucleotides sequence of the primers are listed in table 2.1 at material and methods section, the expected fragment sizes are detailed in table 3.2, and the localisation of the primers in the genome are shown in figure 3.7. Figures 3.8, 3.9, and 3.10 show the gel electrophoresis of the PCR products.

| ***attR*** | **Primers** | **Size** |
|---|---|---|
| All the strain | 779 + 646 | 208 bp |
| ***attL*** | **Primers** | **Size** |
| Nott373_control | 728 + 1001 | 278 bp |
| Nott373_cen1 | 728 + 579 | 300 bp |
| Nott373_cen2 | 728 + 585 | 700 bp |
| Nott373_cen3 | 728 + 581 | 600 bp |
| Nott373_cen4 | 728 + 579 | 300 bp |
| Nott373_cen5 | 728 + 585 | 700 bp |
| Nott373_cen7 | 728 + 325 | 1500 bp |
| Nott373_cen8 | 728 + 325 | 1500 bp |
| Nott373_alphoid | 728 + 909 | 250 bp |

**Table 3.2: Pairs of primers used for the site-specific recombination screening by PCR on the right (*attR*) and left (*attL*) ends of the recombination sequence and for sequencing the left end of the recombination.**
For the *attR* screening, the same pair of primers was used for all the strains. For the *attL* screening, the reverse primer was specific for each plasmid used for the transformation.

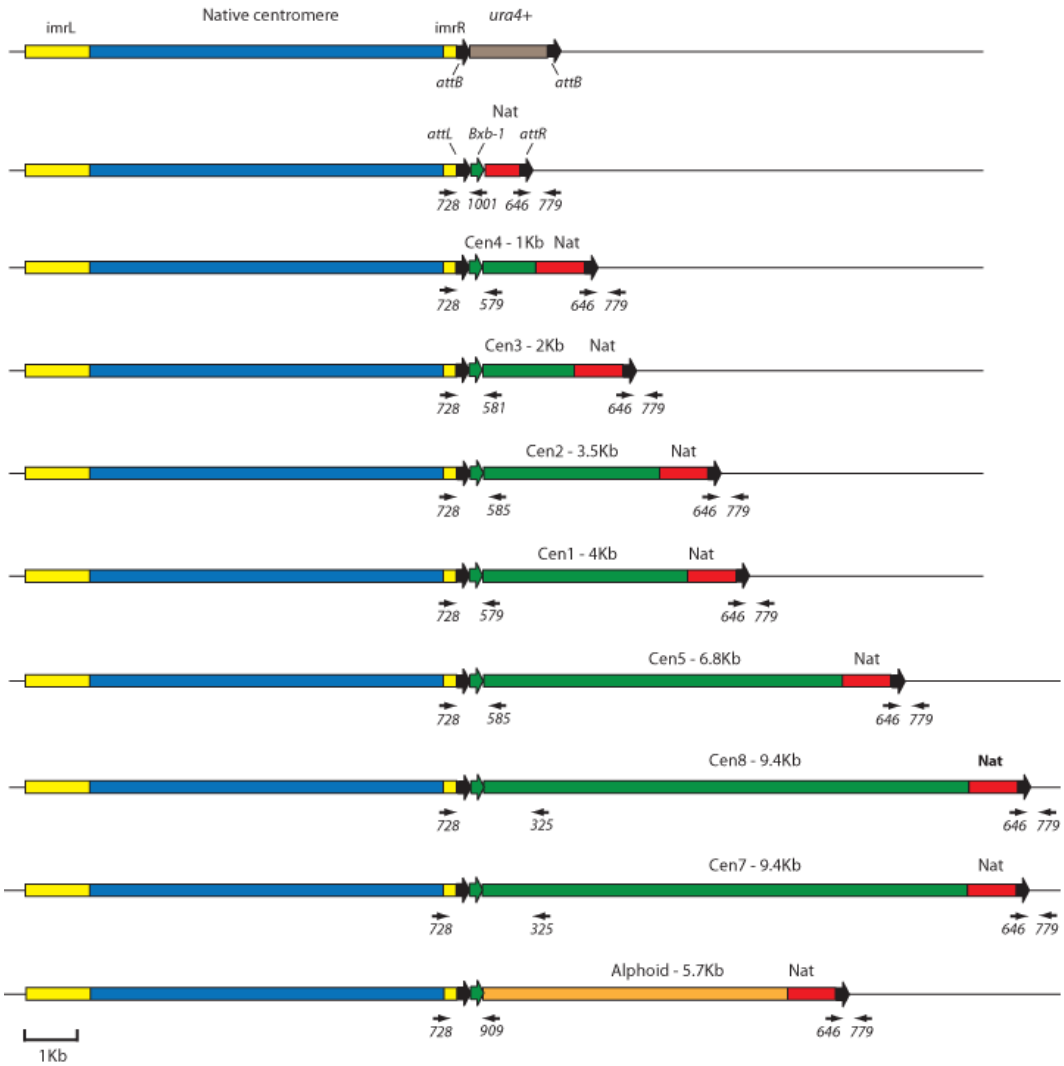



**Figure 3.7: Map of the primers used for the site-specific recombination screening.**
The clones were screened for site-specific recombination by PCR using pairs of primers to amplify the *attL* and *attR* regions.

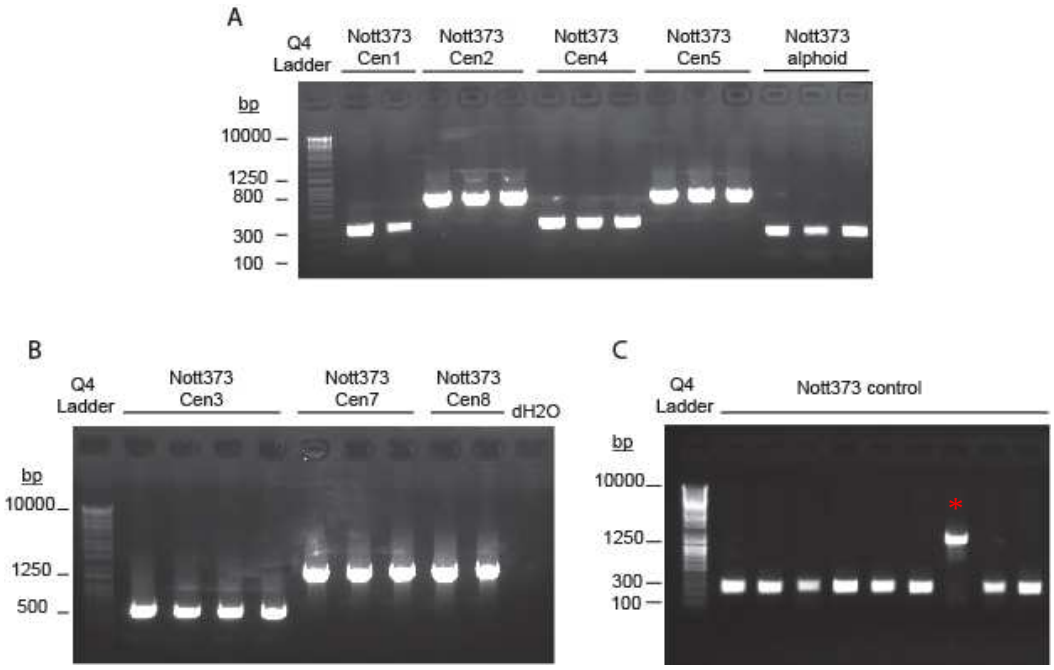



**Figure 3.8: Screening the left end of the recombination site for the Nott373 clones by PCR.**
A, B and C - Gels electrophoresis of the PCRs for site-specific recombination screening in the left end of the recombination site for the Nott373 clones using the primer 728 and the specific primer for each strain. * *ura*Δ clone that was not correctly recombined, that clone was discarded.

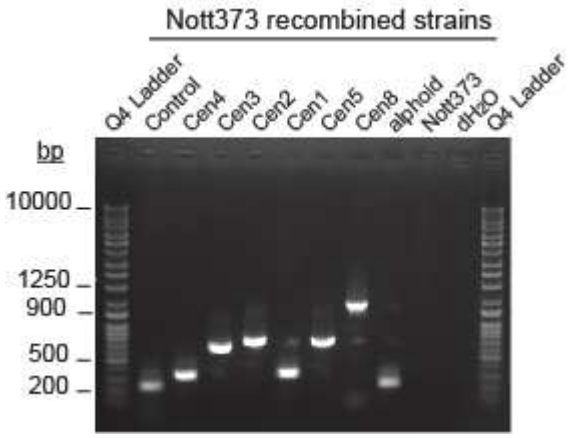



**Figure 3.9: Summary gel electrophoresis containing one clone of each recombined strain and the start strain as a control.**

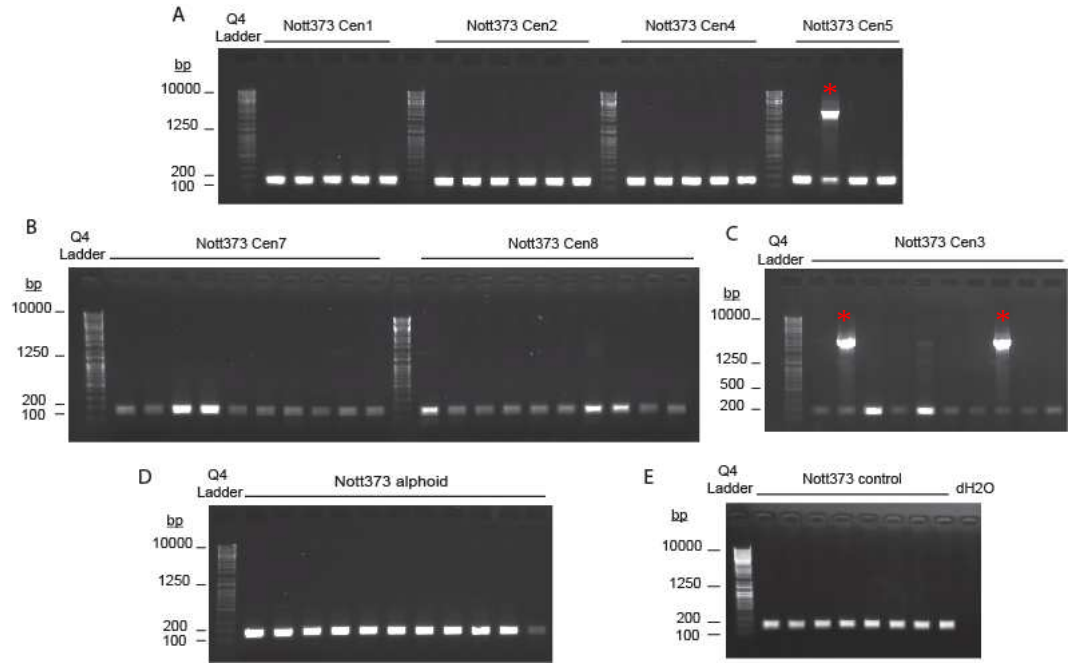



**Figure 3.10: Gel electrophoresis of the PCR for site-specific recombination screening in the right end of the recombination site for the Nott373 clones using primers 779 and 646.**
**uraΔ* clone that was not correctly recombined, those clones were discarded.

We obtained the expected product size for all the nine strains, meaning that the site-specific recombination caused no damage in the *attL* and *attR* sites.

### 3.2.2. Sequencing the left end of the recombination site

The PCR products of the *attL* region of the strains were purified on agarose gel and sequenced to confirm the integrity of this region. The sequencing was performed using the forward primer 728 (table 2.1). The sequences showed intact structures for the *attL* site followed by the *attB* site for the Bxb1 integrase in all the strains (figure 3.11).

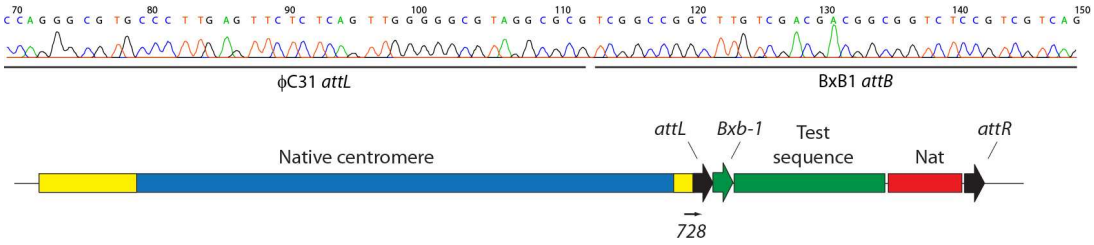



**Figure 3.11: Trace and map of the *attL* PCR product sequencing.**
The sequencing was performed using the forward primer 728. The trace shows intact structures of the ϕC31 *attL* followed by the Bxb1 *attB*.

In summary, the nine different candidate sequences were integrated into chromosome 2 along with the *nat* marker and an *attB* site for the Bxb1 recombinase. The clones were screened by selective medium and by PCR, and the sequencing of the

PCR product demonstrated that there was no damage of left and right ends of the recombination.

### 3.2.3. Test sequence integrity analyses

Then, the recombined strains of Nott373 were analysed by restriction enzyme digestion and filter hybridization in order to verify the integrity of the test sequences. Two sites were particularly useful for this purpose. There is a *Bam*HI site 100 bp centromere proximal of the *attL* site derived from the original *attB ura4*$^{+}$ *attB* cassette, and a *Pst*I site 900 bp from the start of the nourseothricin resistance gene. Neither of these sites is present in the centromeric sequences and thus, restriction enzyme digestion *Bam*HI and *Pst*I enabled me to confirm the integrity of the mutant centromeric DNA inserts and the presence of the nourseothricin resistance gene in the strains. The genomic DNA of the start strain Nott373, the strain recombined with the control plasmid (Nott373_control), and the strains recombined with the *S. pombe* mutant centromeric DNA fragments were therefore digested with the restriction enzymes *Bam*HI and *Pst*I, size fractionated by a vertical agarose gel electrophoresis gel, and transferred to nylon filters. The restriction enzyme maps predicted for the correctly integrated test sequences are shown in figure 3.12.

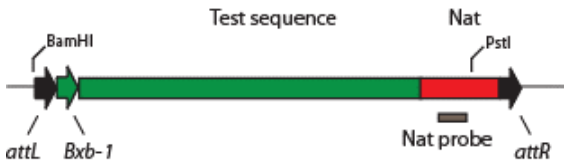
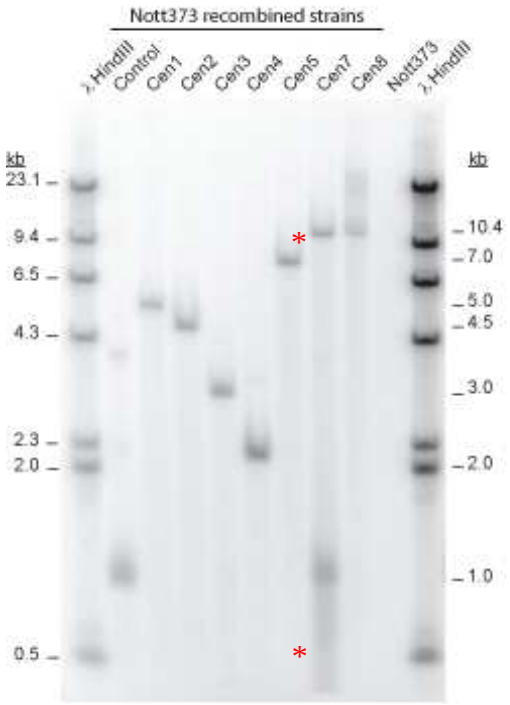



**Figure 3.12: Restriction enzyme site maps of the recombined strains.**
The *Bam*HI site is 100 bp before the test sequence and the *Pst*I is the nourseothricin resistance gene 900 bp from the beginning of the gene.

The filters were hybridised with a probe for the nourseothricin resistance gene. Results show that, except for the Nott373cen7, all the strains hybridised with the nourseothricin resistance gene in a single copy and showed the expected size if DNA fragment (figures 3.13). The strain Nott373cen7 presented two copies of the nourseothricin resistance gene, so this strain was not used in the following experiments.




| Position | Strain | Expected size |
|---|---|---|
| 1 | λ*Hind*III | |
| 2 | Nott373_control | 1 kb |
| 3 | Nott373_cen1 | 5 kb |
| 4 | Nott373_cen2 | 4.5 kb |
| 5 | Nott373_cen3 | 3 kb |
| 6 | Nott373_cen4 | 2 kb |
| 7 | Nott373_cen5 | 7 kb |
| 8 | Nott373_cen7 | 10.4 kb |
| 9 | Nott373_cen8 | 10.4 kb |
| 10 | Nott373 | |
| 11 | λ*Hind*III | |

**Figure 3.13: Filter hybridization analysis of the fragments from the enzymatic digestion with *BamH*I and *Pst*I for the Nott373 recombined strains.**
The clones were hybridised with a probe for the nourseothricin resistance gene. * The Nott373_cen7 presented 2 copies of the nourseothricin resistance gene, this clone was not used in the following experiments.

The alphoid DNA contains a site for *Pst*I and thus the integrity of this insert was confirmed by digestion with *Eco*RI which cuts at sites immediately flanking the alphoid DNA but not within the insert (figure 3.14A). The filter was hybridised with a probe for the alphoid test sequence. The expected fragment size (5.7 kb) was obtained for two clones of the Nott373_alphoid (figure 3.14B). The figure 3.14B also shows the filter of the strain Nott376_alphoid. This strain will be described in the next section.

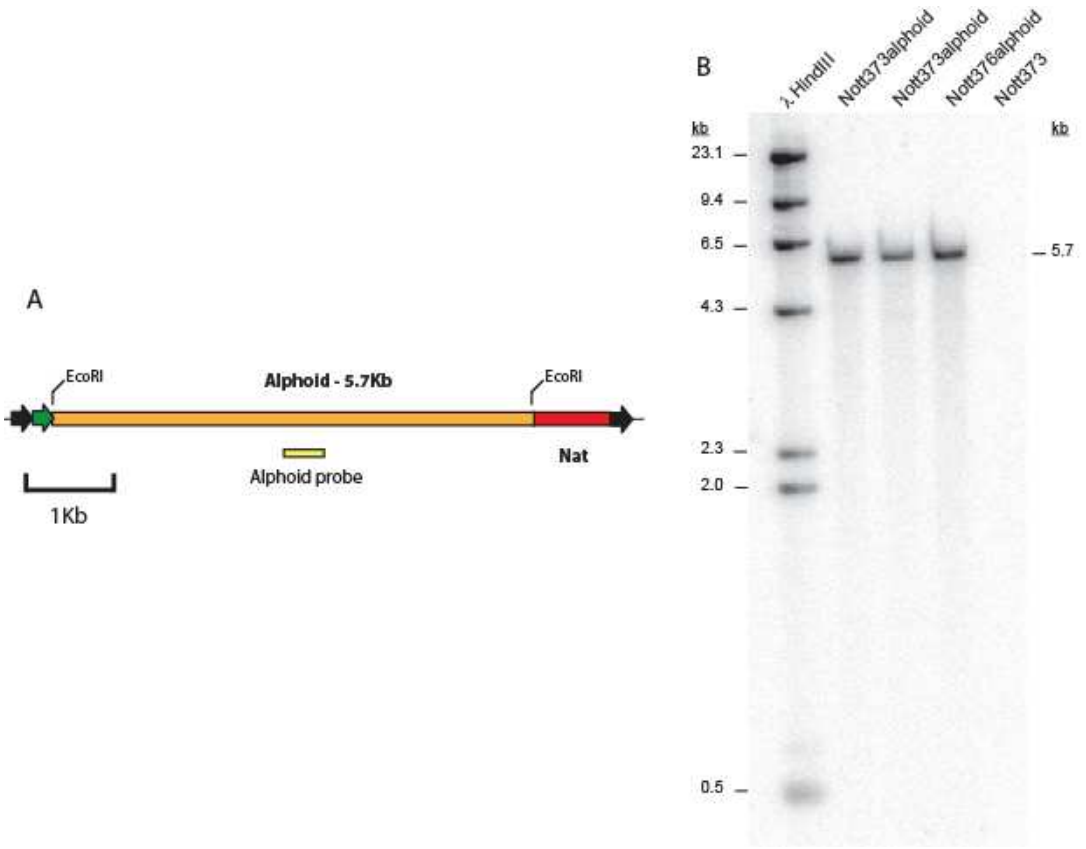



**Figure 3.14: Filter hybridization analysis of the fragments from the enzymatic digestion *EcoR*I for the clones with the alphoid fragment insertion and the Nott373 start strain as control; hybridised with a probe for the alphoid DNA fragment.**
(A) Restriction enzyme site maps of the recombined strain; (B) Filter and hybridised with the alphoid probe.

In this way, I engineered a set of strains in which different sequences had been placed adjacent to the native centromere. The recombined strains were named using the number of the start strain (Nott373) followed by the name of the plasmid and deposited at the laboratory collection (table 3.3).

| Laboratory collection | Strain name | Description |
|---|---|---|
| Nott 531 | Nott373_control | Nott373-chr2R:Bxb1*attB*-*nat* |
| Nott 484 | Nott373_cen1 | Nott373-chr2R:Bxb1*attB*-**cen1**-*nat* |
| Nott 487 | Nott373_cen2 | Nott373-chr2R:Bxb1*attB*-**cen2**-*nat* |
| Nott 515 | Nott373_cen3 | Nott373-chr2R:Bxb1*attB*-**cen3**-*nat* |
| Nott 490 | Nott373_cen4 | Nott373-chr2R:Bxb1*attB*-**cen4**-*nat* |
| Nott 493 | Nott373_cen5 | Nott373-chr2R:Bxb1*attB*-**cen5**-*nat* |
| Nott 516 | Nott373_cen7 | Nott373-chr2R:Bxb1*attB*-**cen7**-*nat* |
| Nott 519 | Nott373_cen8 | Nott373-chr2R:Bxb1*attB*-**cen8**-*nat* |
| Nott 496 | Nott373_alphoid | Nott373-chr2R:Bxb1*attB*-**alphoid**-*nat* |

**Table 3.3: Recombined strains derived from Nott373.**
Laboratory collection number, names used in this work, and description of the strain.

## 3.3. Targeting the left end of the centromere of chromosome 2 with a Bxb1 *attB* site and *ura4+* gene

In the section 3.2 we built strains in which a set of candidate sequences together with a Bxb1 *attB* site had been placed on the right-hand side of the centromere of the chromosome 2 of CBS 2777. We wanted to assay the functionality of the sequences. The way we chose to do this was to delete the native centromere and determine whether the candidate sequence could support centromere formation. In to do this, we engineered a chromosome in which the left hand side (see BP3 on figure 1.2B) was targeted with a Bxb1 *attP* site and the counter selectable marker *ura4* (figure 3.15)

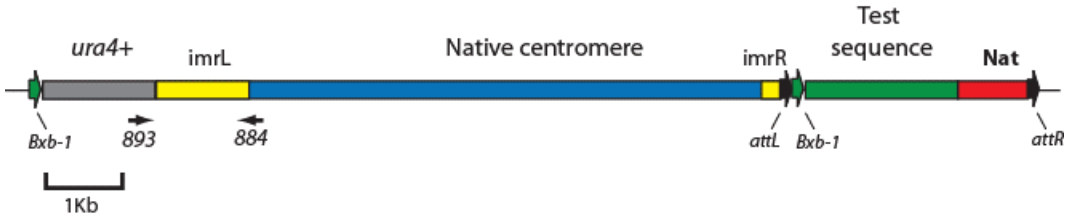



**Figure 3.15: Targeting the left-hand side of the endogenous centromere with *ura4+* gene and Bxb1 *attB* site.** The targeted clones were screened by PCR using the primers 893 and 884.

The targeted clones were screened by PCR using the 893 forward primer for the *ura4*$^{+}$ and the 884 reverse primer for the endogenous centromere (table 2.1, figure 3.16), the clones correctly targeted on the left of the native centromere were expected to show a fragment of approximately 900 bp. We obtained right targeted clones for all the nine Nott373 recombined strains.

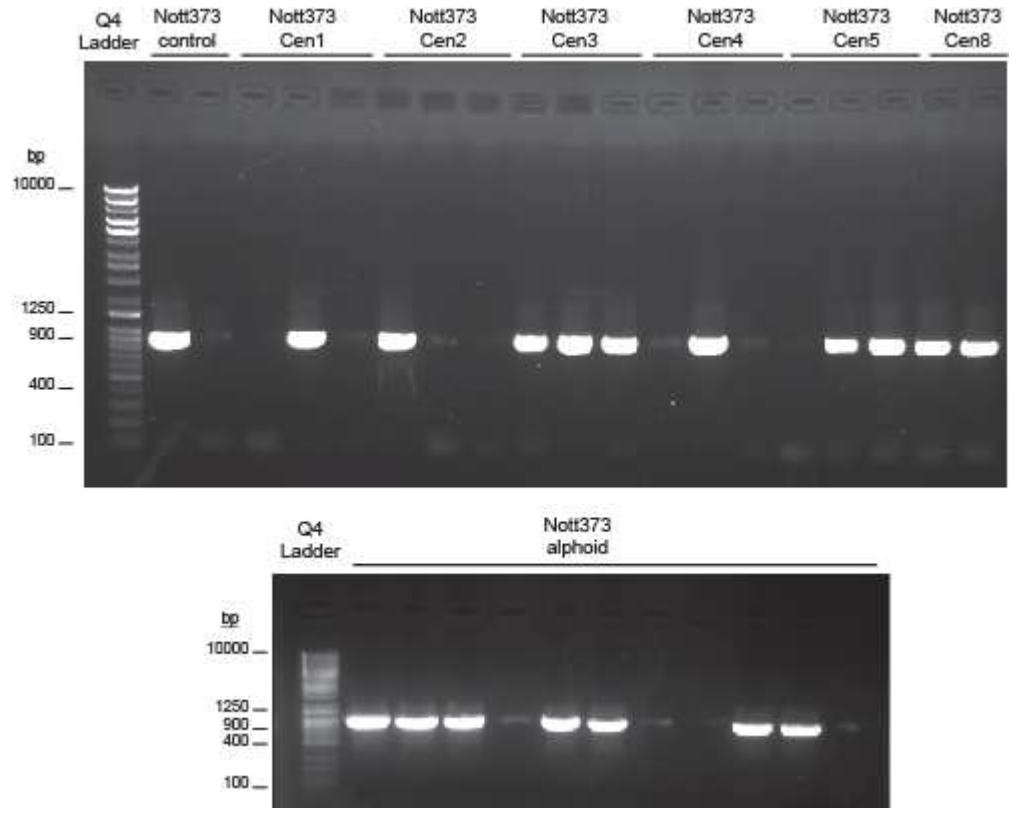



**Figure 3.16: Gel electrophoresis of the PCR screening of the clones targeted on the left of the endogenous centromere.**
The targeted clones amplified a fragment of ~900 bp.

As mentioned previously, the central core of the centromere of chromosome II of the *S. pombe* lab strain 968h[90] is duplicated in the genome of CBS 2777 (Brown et al., 2014). One copy is present on both chromosomes 2 and 4 of this strain, thus the targeting could occur in any of these two chromosomes. Long range PCR tests were performed to screen the Nott373_cen2 clones targeted on chromosome 2 using primers 893 and 1030, and primers 893 and 1030 for chromosome 4 (figure 3.17A). Eight clones of the Nott373_cen2 were tested for both targeting on chromosome 2 and chromosome 4 (figure 3.17B). The clones 1 and 6 showed to be target only on chromosome 2, the clone 2 was targeted on both chromosome 2 and chromosome 4, and the clones 3,4,5,7, and 8 were targeted only on chromosome 4.

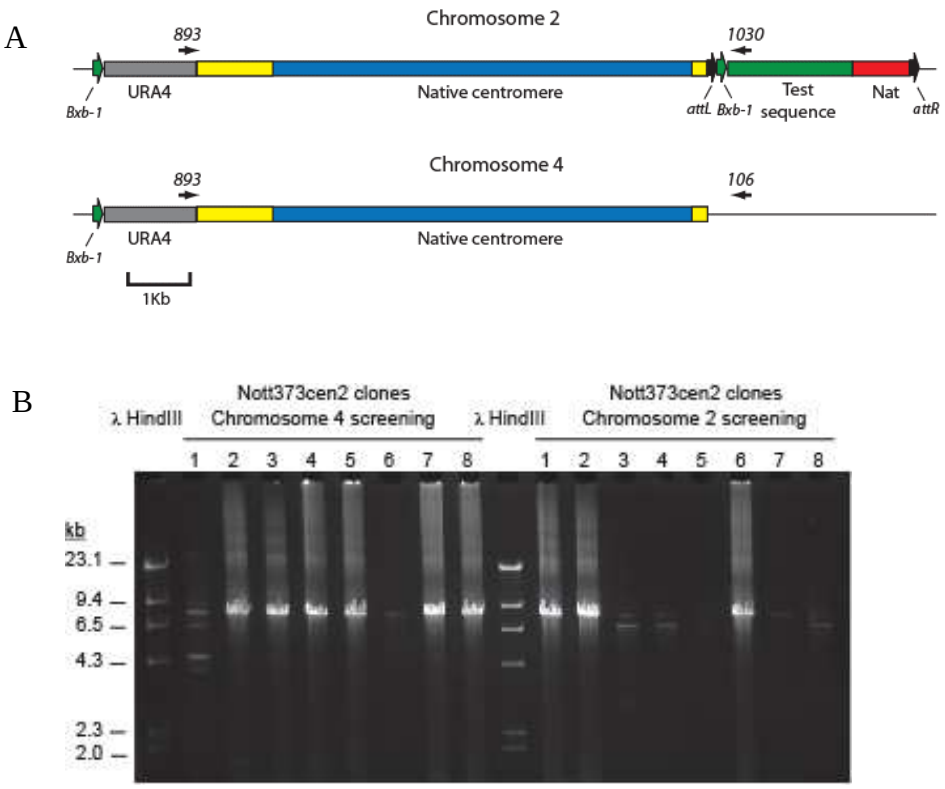



**Figure 3.17: Long range PCR to screen the clones targeted on chromosome 2.**
A- Map showing the position of the primers 893 and 1030 on chromosome 2, and primers 893 and 106 on chromosome 4. B- Gel electrophoresis of the Long Range PCR products from the screening of the Nott373_cen2 clones. The clones 1 and 6 were targeted only on chromosome 2.

The clones targeted on chromosome 2 of the other Nott373 recombined strains were screened using restriction enzyme digestions followed by gel electrophoresis and filter hybridization. One clone of the Nott373_cen2 that showed to be targeted on chromosome 4 by long range PCR was used as a control. In order to carry out the gel electrophoresis and hybridization analysis, the genomic DNA of the clones was digested with *Sal*I, separated by a vertical electrophoresis gel, and transferred to nylon filters. The filters were hybridised with a probe for the *ura4*$^{+}$ (figure 3.18A). The clones targeted on chromosome 2 show a fragment of approximately 10.1kb and the clones targeted on chromosome 4 show a fragment of approximately 13.3kb (figure 3.18B). Two clones of the Nott373_cen4 were targeted on chromosome 4, all the others tested clones were targeted on chromosome 2.

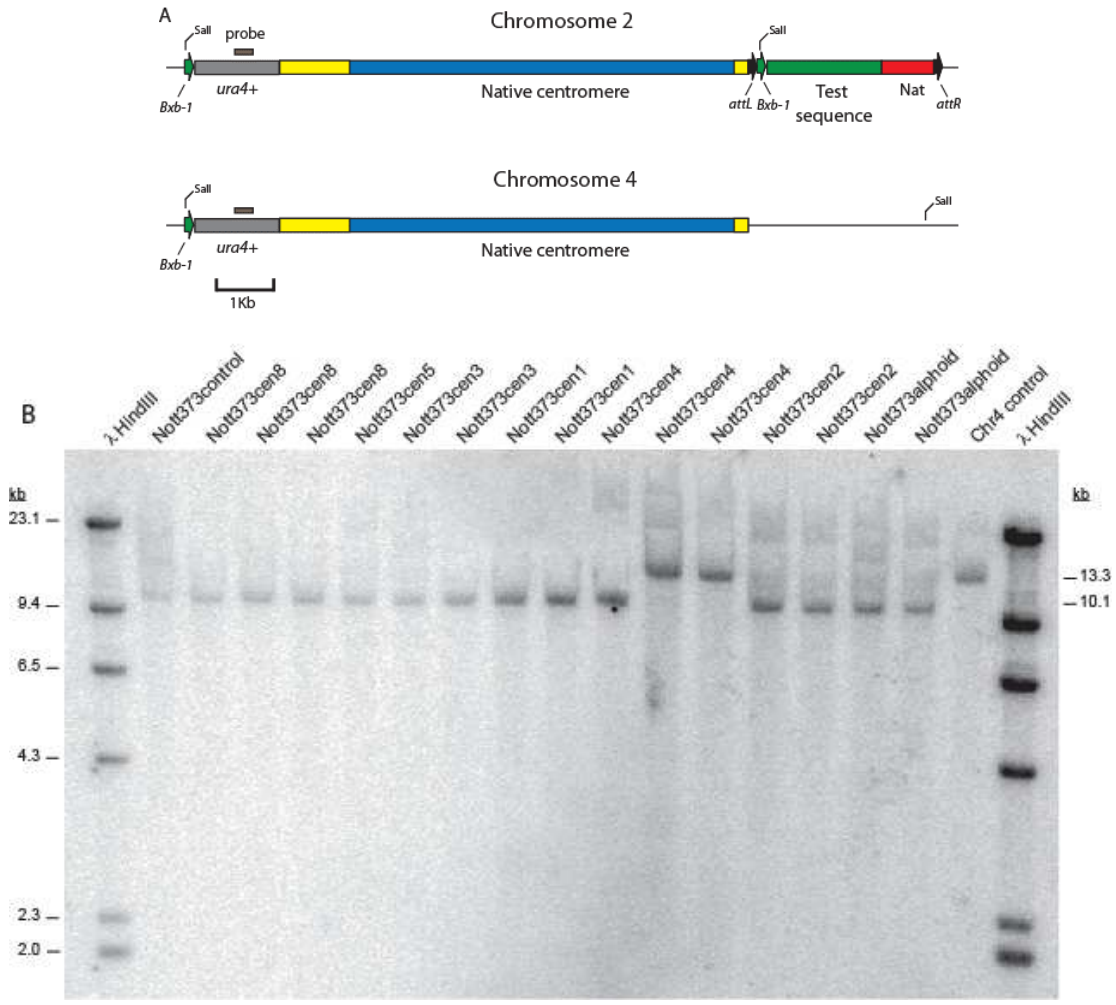



**Figure 3.18: Southern blot analysis of the fragments from the enzymatic digestion *Sal*I for the Nott373 clones targeted on the left of the endogenous centromere and hybridised with a probe for the *ura4+*.**
(A) Restriction enzyme site maps of the strains targeted on chromosome 2 and on chromosome 4; (B) Filter hybridised with the *ura4+* probe.

With the integration of the Bxb1 *attB* site on the left-hand side of the native centromere we completed a system that allows us to excise the native centromere DNA upon subsequent the Bxb1 integrase expression.

## 3.4. Integration of the mutant centromeric DNA fragments on the *ade6*$^{+}$ locus

In order to observe whether or not the kinetochore can bind on centromeric DNA in regions distant from the natural centromere, test sequences were inserted at the *S. pombe ade6+* gene locus.

The gene *ade6*$^{+}$ is located on chromosome III (1316291-1318035) of the *Schizosaccharomyces pombe* strain CBS 2777. The laboratory collection strain Nott376 (table 2.2) expresses a Cnp1 protein tagged at the amino terminus with GFP, and the *ade6* was targeted with *ura4*$^{+}$ marker and *attB* sites for the ϕC31 integrase. The test sequences were integrated into the targeted region of Nott376.

As before, ϕC31 integrase was expressed using the plasmid NL314 (figure 3.2). Site-specific recombination was performed using 50 ml of the cells of the NL314 transformed cells grown a density of $1.5x10^{7}$ cells/ml and 100 μg of each vector. The *ura4Δ* clones were screened on PMG supplemented with 225 μg/ml of both leucine and adenine. The *ura4Δ* clones obtained from the transformation were submitted to PCR reactions for checking the site-specific recombination in both *attL* and *attR* sites of the recombinant sequence, as described in the next section.

### 3.4.1. Confirming for the site-specific recombination PCR

The sequences of the primers used in the PCR reaction are listed in table 2.1 in the material and methods section, the localisation of the primers in the genome are shown in figure 3.19, and the pairs of primers used for each sample and the expected fragment sizes are detailed in table 3.4.

We obtained the right product size for all the nine strains. Figures 3.20, 3.21, and 3.22 show the gel electrophoresis of the PCR products.

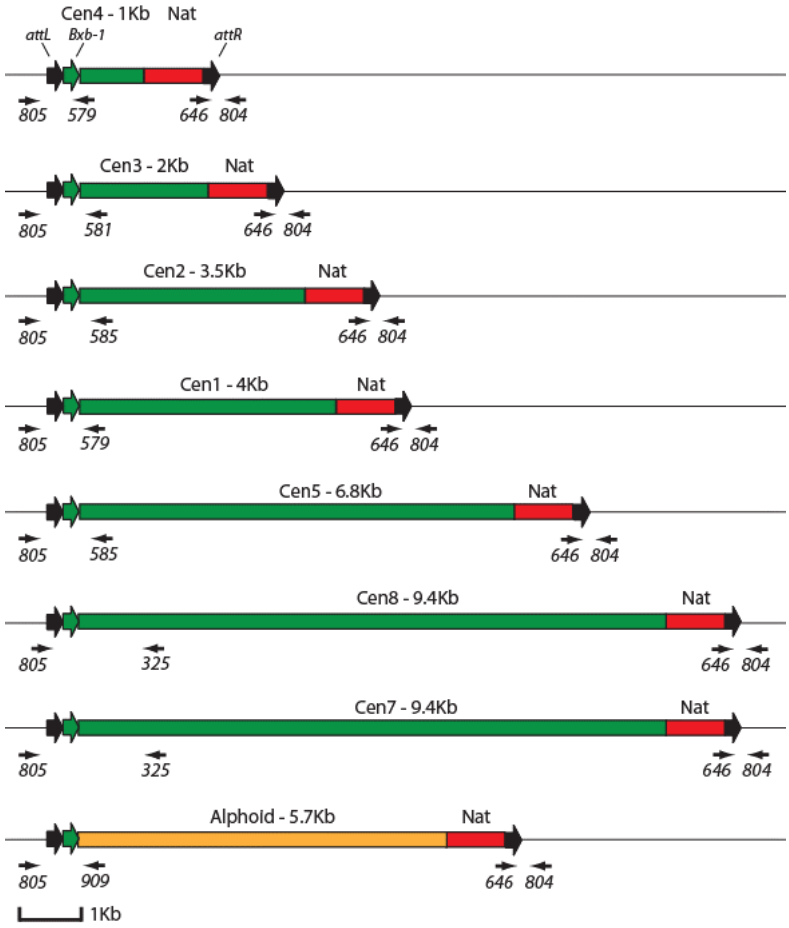



**Figure 3.19: Map of the primers used for the site-specific recombination screening on Nott376.**
The clones were screened for site-specific recombination by PCR using pairs of primers to amplify the *attL* and *attR* regions.

| ***attR*** | **Primers** | **Size** |
|---|---|---|
| All the strains | 804 + 646 | 200 bp |
| ***attL*** | **Primers** | **Size** |
| Nott376_control | 805 + 1001 | 278 bp |
| Nott376_cen1 | 805 + 579 | 300 bp |
| Nott376_cen2 | 805 + 585 | 700 bp |
| Nott376_cen3 | 805 + 581 | 600 bp |
| Nott376_cen4 | 805 + 579 | 300 bp |
| Nott376_cen5 | 805 + 585 | 700 bp |
| Nott376_cen7 | 805 + 325 | 1500 bp |
| Nott376_cen8 | 805 + 325 | 1500 bp |
| Nott376_alphoid | 805 + 909 | 250 bp |

**Table 3.4: Pairs of primers used for the site-specific recombination screening by PCR on the right (*attR*) and left (*attL*) ends of the recombination sequence.**
For the *attR* screening, the same pair of primers was used for all the strains. For the *attL* screening, the pair of primers varied according to the plasmid used for the transformation.

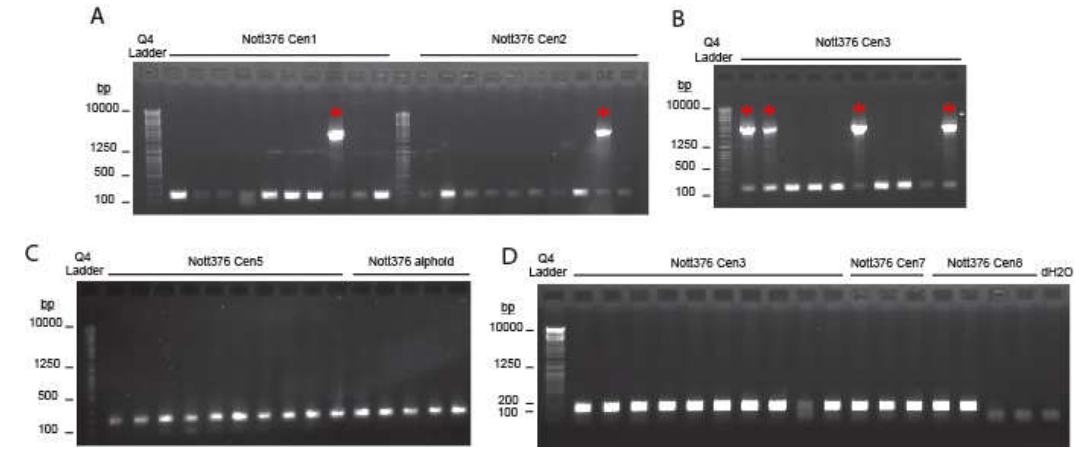



**Figure 3.20: Screening the right end of the recombination site for the Nott376 clones by PCR.**
A, B, C and D - Gel electrophoresis of the PCR for site-specific recombination checking in the right end of the recombination site for the Nott376 clones. **ura4*Δ clone that was not correctly recombined, those clones were not used in the following experiments.

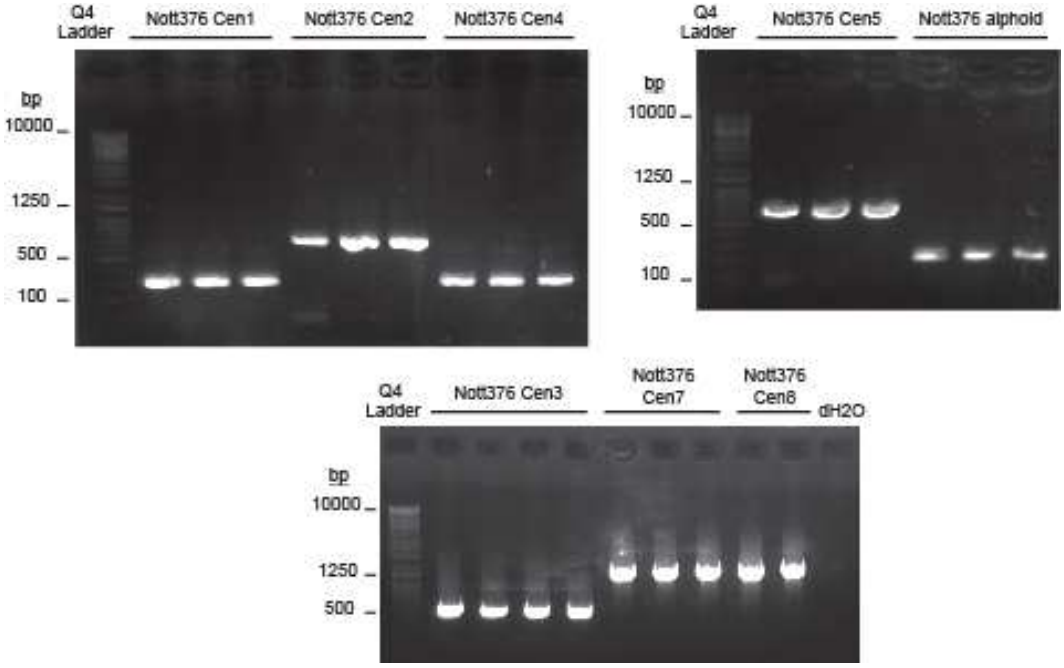



**Figure 3.21: Screening the left end of the recombination site for the Nott376 clones by PCR.**
A, B C - Gel electrophoresis of the PCR for site-specific recombination checking in the left end of the recombination site for the Nott376 clones using the primer 805 and the specific primer for each strain.

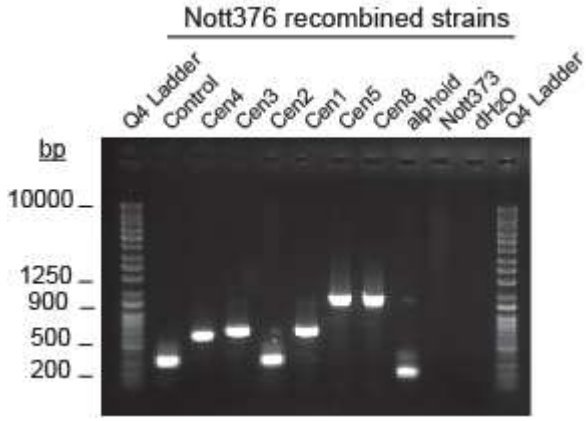



**Figure 3.22: Summary gel electrophoresis containing one clone of each recombined strain and the start strain as a control.**

The recombined strains derived from Nott376 were named with the number of the start strain followed by the name of the plasmid and deposited at the laboratory collection (table 3.5).

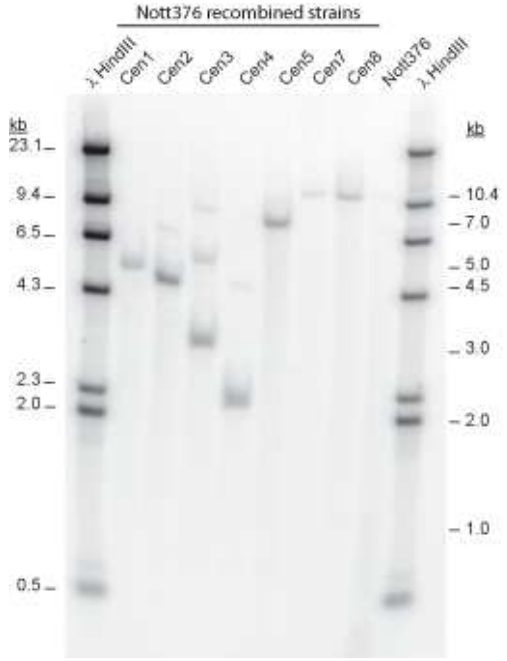
| Laboratory collection | Strain name | Description |
|---|---|---|
| Nott 499 | Nott376_cen1 | Nott373- *ade6Δ*::Bxb1*attB*-**cen1**-*nat* |
| Nott 502 | Nott376_cen2 | Nott373- *ade6Δ*::Bxb1*attB*-**cen2**-*nat* |
| Nott 521 | Nott376_cen3 | Nott373- *ade6Δ*::Bxb1*attB*-**cen3**-*nat* |
| Nott 505 | Nott376_cen4 | Nott373- *ade6Δ*::Bxb1*attB*-**cen4**-*nat* |
| Nott 508 | Nott376_cen5 | Nott373- *ade6Δ*::Bxb1*attB*-**cen5**-*nat* |
| Nott 525 | Nott376_cen7 | Nott373- *ade6Δ*::Bxb1*attB*-**cen7**-*nat* |
| Nott 528 | Nott376_cen8 | Nott373- *ade6Δ*::Bxb1*attB*-**cen8**-*nat* |
| Nott 510 | Nott376_alphoid | Nott373- *ade6Δ*::Bxb1*attB*-**alphoid**-*nat* |

**Table 3.5: Recombined strains derived from Nott376.**
Laboratory collection number, names used in this work, and description of the strain.

### 3.4.2. Test sequence integrity analyses

The recombined strains of Nott376 were analysed by restriction enzyme digestion and filter hybridization in order to verify the integrity of the test sequences. As done for the Nott373 recombined strain, the genomic DNA of the start strain Nott376, and the strains recombined with the *S. pombe* mutant centromeric DNA fragments were digested with the restriction enzymes *Bam*HI and *Pst*I, size fractionated by a vertical agarose gel electrophoresis gel, and transferred to nylon filters. The restriction enzyme maps predicted for the correctly integrated test sequences are shown in figure 3.12.

The filters were hybridised with a probe for the nourseothricin resistance gene. All the strains hybridised with the nourseothricin resistance gene in a single copy and showed the expected size if DNA fragment (figures 3.23).




| Position | Strain | Expected size |
|---|---|---|
| 1 | λ*Hind*III | |
| 2 | Nott373_cen1 | 5 kb |
| 3 | Nott373_cen2 | 4.5 kb |
| 4 | Nott373_cen3 | 3 kb |
| 5 | Nott373_cen4 | 2 kb |
| 6 | Nott373_cen5 | 7 kb |
| 7 | Nott373_cen7 | 10.4 kb |
| 8 | Nott373_cen8 | 10.4 kb |
| 9 | Nott373 | |
| 10 | λ*Hind*III | |

**Figure 3.23: Filter hybridization analysis of the fragments from the enzymatic digestion with *BamH*I and *Pst*I for the Nott376 recombined strains.**
The clones were hybridised with a probe for the nourseothricin resistance gene.

## 3.5. Discussion

The objective of this chapter of my work was to develop a molecular system in the *S. pombe* strain CBS 2777 to test the predictions of kinetochore spreading and Cnp1 binding. CBS 2777 is particularly useful for these experiments because the lack of flanking repeats in two of its four chromosomes, making it easier to manipulate. The system was built by site-specific recombination using vectors engineered with serine recombinase attachment sites. The test sequences were placed on the right-hand side of the centromere on chromosome 2. The experimental strains were then targeted on the left-hand side of the endogenous centromere of chromosome 2 with *ura4*$^+$ and *attB* site for Bxb1 integrase. As a control for this system, the mutant centromeric DNA fragments and the alphoid fragment were inserted at the CBS 2777 *ade6*$^+$ locus, on chromosome 3.

The results of this chapter demonstrate that a combination of gene targeting by homologous recombination followed by site-specific recombination promoted by

ϕC31 integrase enables efficient integration of DNA fragments into the *S. pombe* genome and allow manipulation of the centromere sequence content in this organism. This observation is consistent with previous studies showing that the integrase work in vertebrate cells (Dafhnis-Calas et al., 2005, Malla et al., 2005, Belteki et al., 2003).

The significance of these experiments is that they provide a system which should enable the study of the requirements for centromere inheritance and kinetochore assembly using the genetic approaches that are possible in *S. pombe*. The different fragment sizes should allow me to test the minimum test sequence size required for an efficient centromere formation after the native centromere excision, and the human alphoid DNA enable me to analyse specificity of the DNA sequence for the kinetochore assembly. Chapter four describes such experiments.

# 4. THE RATE OF NEO-CENTROMERE FORMATION FOLLOWING NATIVE CENTROMERE EXCISION WITH THE BXB1 INTEGRASE

## 4.1. Introduction

In the previous chapter, a system was described that allowed the native centromeric region of the chromosome 2 in the CBS 2777 derived strains to be sequentially modified in order to examine the centromere activity of potentially functional sequences. In the first step, an *attB* site for the Bxb1 recombinase, together with a candidate sequence, were placed on one side of the native centromeric DNA and, in the second, an *ura4*$^{+}$ gene, together with a Bxb1 *attP* site, were placed on the other side. In this chapter, the Bxb1 recombinase is expressed in a Rep81 plasmid to mediate the excision of the native centromeric DNA in order to permit the activity of the candidate sequences to be tested.

## 4.2. Excision of the native centromere using Bxb1 integrase

In order to promote the excision of the native centromere 2 of the Nott373 recombined strains, the Bxb1 integrase was expressed from a plasmid in which an REP81 plasmid was used to drive expression of a codon-optimised gene that encoded a version of the Bxb1 integrase tagged with a nuclear localisation signal. The plasmid included an autonomously replicating sequence 1 (ARS1) and to allow selection, a *Saccharomyces cerevisiae LEU2* gene (figure 4.1). The transformations were performed as described in the material and methods section, using 100 ml of cultures at a density of $2x10^{7}$ cell/ml and 900 μg of the plasmid. After recovery, the cells were plated on PMG medium supplemented with 250mg/l of uracil, this medium selected for the clones containing the *LEU2* gene. An REP81 empty vector plasmid was used as a control for the integrase expression plasmid.

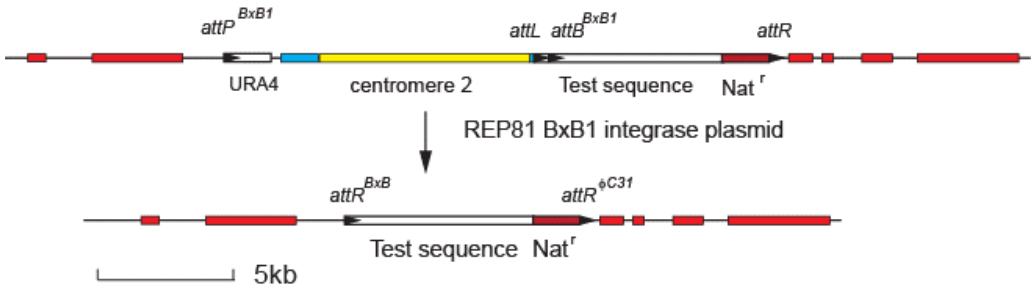



**Figure 4.1: General representation of the strategy to promote the centromere 2 excision in the Nott373 recombined strains.**
The Bxb1 integrase was induced in an REP81 plasmid.

As result, the size of the colonies recovered following transformation with the REP81 Bxb1 integrase plasmid varied according to the size and sequence of the test DNA that had been integrated adjacent to the native centromere. Strains with no test sequence, or carrying test sequences of 4 kb or less of *S. pombe* centromeric DNA, or a fragment of the human Y alphoid DNA gave smaller colonies than the REP81 empty vector controls (figure 4.2A), while strains carrying test sequences with more than 4 kb of *S. pombe* centromeric DNA gave colonies that were similar in size to both the transformation with the empty REP81 vector or the REP81 Bxb1 integrase plasmid (figure 4.2B). This result suggested that most cells of the strains that gave small colonies were dying after excision of the centromere 2.

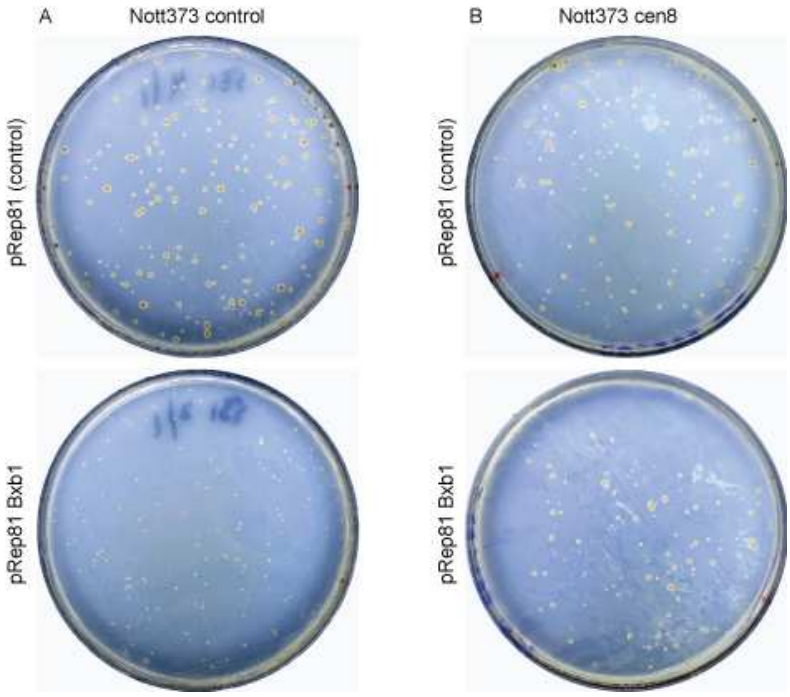



**Figure 4.2: Plates of PMG agar supplemented with uracil showing the colonies size of the transformed clones.**

(A) Colonies of Nott373_control transformed with REP81 empty vector (top) and REP81 Bxb integrase plasmid (bottom). (B) Colonies of Nott373_cen8 transformed with REP81 empty vector (top) and REP81 Bxb integrase plasmid (bottom).

## 4.3. Average colony size and cells viability of the transformed clones

In order to confirm the idea that most cells of the strains that gave small colonies were dying after excision of the centromere 2, I measured the average number of cells in a colony, the cells viability, and the proportion of cells that had the centromere 2 excised in each strain following transformation with the REP81 Bxb1 integrase plasmid and the REP81 empty vector.

Figure 4.3 shows the experimental procedures involved in the measurement of the average colony size and the viability of the cells resulting from the transformation with the control and integrase plasmids. For each strain, a group of 50 colonies from the same transformation plate was randomly selected from both control and experiment transformations, the cells were re-suspended in PMG broth, and the number of cells was counted three times in a modified Neubauer chamber. A known number of cells was then plated on YES and the number of recovered colonies counted in order to measure the cells viability. The experiment was performed in for 3 independent transformation with REP81 Bxb1 and REP81 empty vector.

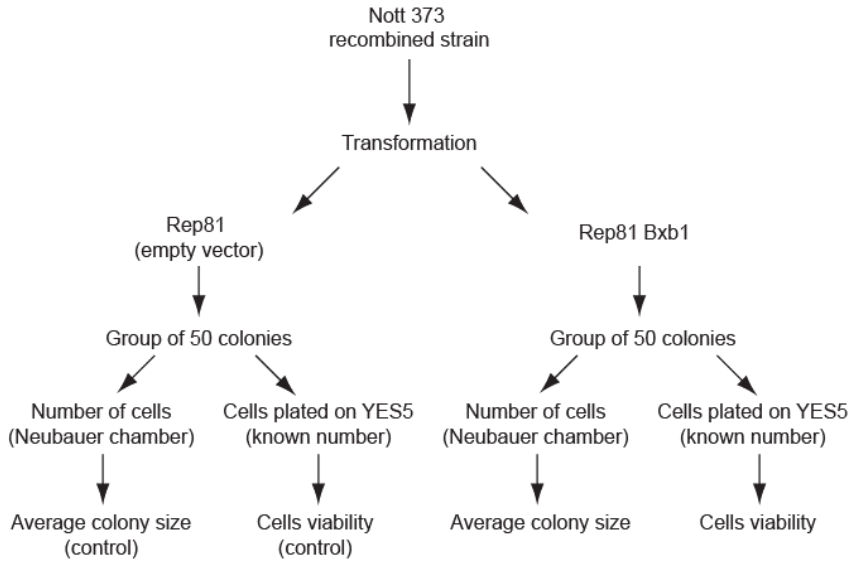



**Figure 4.3: A general strategy to measure the average colony sizes and the cells viability.**
The average colony size was measured counting the number of cells in Neubauer chamber and the cells viability was measured by counting the number of survived cells on YES agar in which a known number of cells was plated.

Table 4.1 and figures 4.4 and 4.5 show the average number of number of cells in a colony and viability of cells for the 3 experiments of each strain. The absolute number of each experiment are shown in Appendix I. Strains with no test sequence or carrying *S. pombe* mutant centromeric DNA until the length 2 kb as test sequence presented a number of cells in a colony that ranged from 3% to 12% of the control REP81 empty vector. The cells viability of these strains not higher than 47% of the control. The strain carrying 3.6 kb and 4.2 kb of *S. pombe* mutant centromeric DNA gave colonies around 40% of the number of cells observed in the control of the same strain. The cells viability was 74% of the control for the 4.2 kb strain and bigger than the control for the 3.6 kb strain, this unexpected result is explained in the next section. Strains carrying test sequences of more than 4kb presented, on average, number of cells in a colony and cells viability similar to the control. The colonies of the strain harbouring the alphoid DNA adjacent to the native centromere had, on average, 20% of the number of cells in the control colonies and presented 30% of the control in cells viability. The start strain did not show differences in colony size and cell viability upon the transformation with the integrase or the empty vector, indicating that the Bxb1 integrase is not toxic for the *S. pombe* cells.

| Strain (test sequence) | Plasmid | Average number of cells in the colony | Viability (proportion) |
|---|---|---|---|
| **Nott373 (start strain)** | REP81 | $2.7x10^6$ | 0.56 |
| | REP81Bxb1 | $2.9x10^6$ | 0.47 |
| **Nott373_alphoid (5.8kb)** | REP81 | $6.3x10^6$ | 0.4 |
| | REP81Bxb1 | $1.3x10^6$ | 0.12 |
| **Nott373_control (0kb)** | REP81 | $5.6x10^6$ | 0.4 |
| | REP81Bxb1 | $0.17x10^6$ | 0.13 |
| **Nott373_cen4 (1kb)** | REP81 | $3x10^6$ | 0.31 |
| | REP81Bxb1 | $0.3x10^6$ | 0.11 |
| **Nott373_cen3 (2kb)** | REP81 | $3.5x10^6$ | 0.38 |
| | REP81Bxb1 | $0.44x10^6$ | 0.06 |
| **Nott373_cen2 (3.6kb)** | REP81 | $1x10^6$ | 0.38 |
| | REP81Bxb1 | $0.36x10^6$ | 0.18 |
| **Nott373_cen1 (4.2kb)** | REP81 | $1x10^6$ | 0.35 |
| | REP81Bxb1 | $0.38x10^6$ | 0.26 |
| **Nott373_cen5 (6.8kb)** | REP81 | $4.1x10^6$ | 0.32 |
| | REP81Bxb1 | $4.3x10^6$ | 0.34 |
| **Nott373_cen8 (9.4kb)** | REP81 | $2.6x10^6$ | 0.3 |
| | REP81Bxb1 | $3.6x10^6$ | 0.3 |

**Table 4.1: Average colony size represented as the number of cells per colony and cells viability measured as the percentage of alive cells from as known plated number of cells**.
The results as shown for each strain transformed with the REP81 empty vector and the REP81 Bxb1 integrase plasmid.

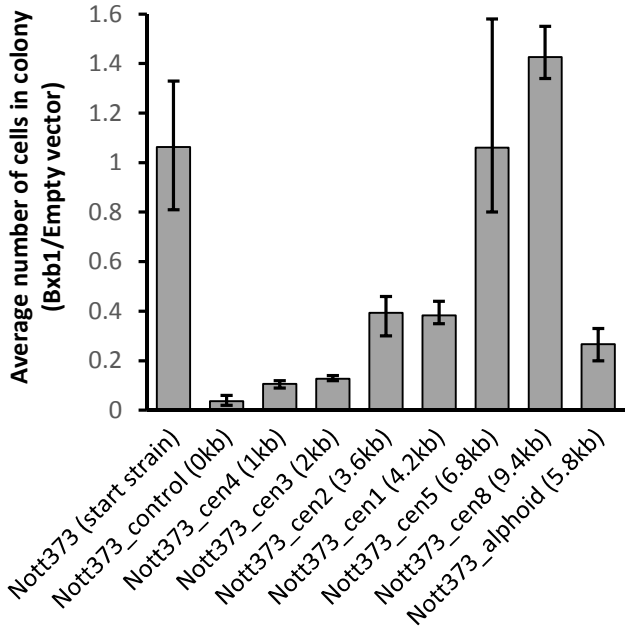



**Figure 4.4: Average number of cells in a colony in each strain transformed with REP81 Bxb1 integrase plasmid as a proportion of the control REP81 empty vector.**
The bar shows the range of values observed. The proportion is very low when the strain carries 4kb or less of *S. pombe* centromeric DNA as test sequence, or when it carries a fragment of the human alphoid DNA as test sequence. The start strain (Nott373) and the strains carrying *S. pombe* centromeric DNA of 6.8kb and 9.5kb gave colonies similar in size for transformation with the control and experiment plasmids.

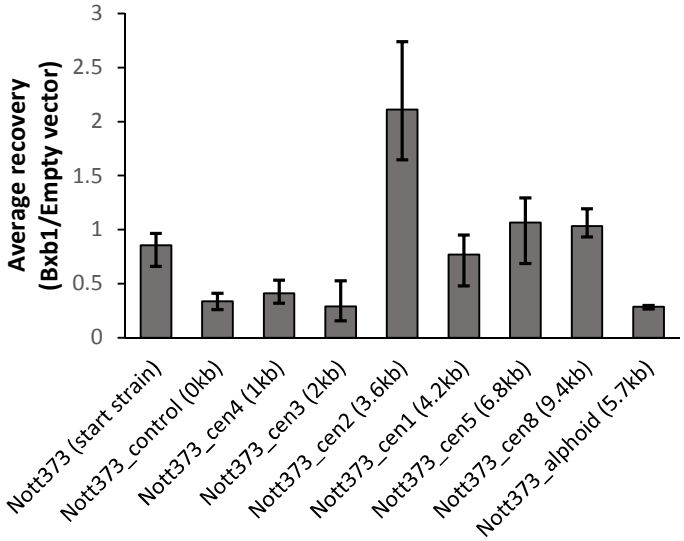



**Figure 4.5: Average recovery of cells in each strain transformed with REP81 Bxb1 integrase plasmid as a proportion of the control REP81 empty vector.**
The bar shows the range of values observed. The proportion is very low when the strain until 2kb of *S. pombe* centromeric DNA as test sequence, or when it carries a fragment of the human alphoid DNA as

test sequence. The start strain (Nott373) and the strains carrying *S. pombe* centromeric DNA of 6.8kb and 9.5kb gave colonies similar in size for transformation with the control and experiment plasmids.

Considering that there is no difference in growth rate among the strains, this result confirm our idea that most cells of the strains that gave small colonies were dying after excision of the centromere 2. The results suggested that the length and nature of the candidate sequences have roles on the efficiency of centromere maintenance after native centromere excision.

## 4.4. Screening the *ura4Δ* clones

The clones recovered on the YES agar for the viability test were transferred for both YES and PMG supplemented with 150mg/l of leucine (figure 4.6). The DNA fragment between the Bxb1 recognition sites includes the native centromere 2 and the *ura4*$^{+}$ marker, placed on the left-hand side of the native centromere, hence both regions are expected to be excised once the Bxb1 integrase is induced (see figure 4.1). The *ura4Δ* clones do not grow on the PMG supplemented with leucine medium enabling identification of clones in which the native centromere 2 had been deleted. 100 clones of each strain were tested for the loss of *ura4*$^{+}$ gene in this way. The test was performed in triplicates.

Nott373_cen3 and Nott373_alphoid presented less than 1% of *ura4Δ* cells from the tested ones, Nott373_control and Nott373_cen4 had 3% or less of *ura4Δ* cells, for Nott373_cen1 the percentage of *ura4Δ* cells was 69%, and Nott373_cen5 and Nott373_cen8 presented more than 98% of *ura4Δ* cells (table 4.2).

The strain Nott373_cen2 was unusual in that the colonies recovered on YES following excision of the native centromere were of two sizes: 38% were big colonies and 62% were small colonies. When tested for the loss of the *ura4*$^{+}$ gene on PMG supplemented with leucine medium, it was found that 95% of the small colonies were *ura4Δ*, and 99% of the big colonies had the *ura4*$^{+}$ gene. Thus, in total, the Nott373_cen2 presented 59% of *ura4Δ* cells (table 4.2). We think that, when the native centromere is excised, the 3.6kb test sequence carried by the Nott373_cen2 forms a weak centromere arresting the cells in mitosis therefore, forming small colonies. This idea also explain the higher viability of cells shown on figure 4.5. Further tests are necessary to confirm this hypothesis.

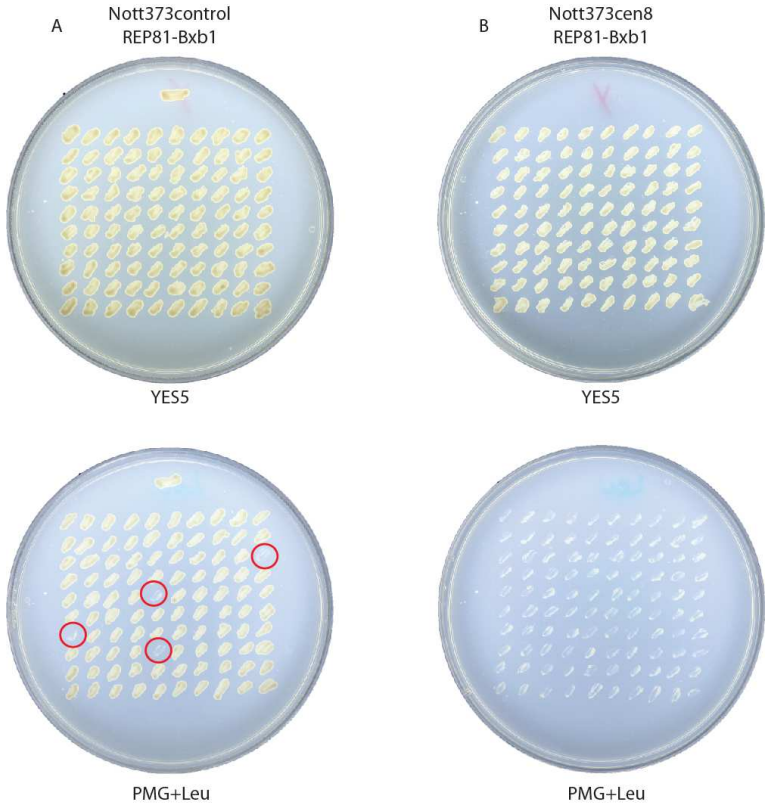



**Figure 4.6: The clones used for the cells viability test were transferred for both YES and PMG supplemented with 150mg/l of Leucine (figure 5) to screen the *ura4Δ* clones.**
(A) Nott373_control presented 3% of *ura4Δ* clones. (B) Nott373_cen8 had above 98% of *ura4Δ* clones.

| Strain | *ura*Δ (proportion) |
|---|---|
| Nott373_control (0kb) | 0.03 |
| Nott373_cen4 (1kb) | 0.03 |
| Nott373_cen3 (2kb) | 0 |
| Nott373_cen2 (3.6kb) | 0.59 |
| Nott373_cen1 (4.2kb) | 0.69 |
| Nott373_cen5 (6.8kb) | 0.986 |
| Nott373_cen8 (9.5kb) | 0.986 |
| Nott373_alphoid (5.8kb) | 0 |

**Table 4.2: Average proportion of *ura4Δ* cells in each strain after the induction of the Bxb1 integrase.**
The proportion of *ura4Δ* cells is not higher than 0.03 when the test sequence has 2 kb or less. This proportion increases to 0.59 when the test sequence is 3.6 kb, 0.69 for the 4.2 kb test sequence, and to 0.986 for the 6.8 kb and 9.5 kb teat sequences. The proportion of *ura4Δ* cells in the Nott373_alphoid was 0 in this experiment.

The loss of the *ura4*$^{+}$ gene was confirmed by plating the clones on 5-Fluoroorotic acid agar (5-FOA) which generates a toxic metabolite in Ura+ cells. Figure 4.7 illustrates the growth of the strains Nott373_control and Nott373_cen8 *ura4*Δ clones on YES, PMG supplemented with Leucine, and 5-FOA mediums.

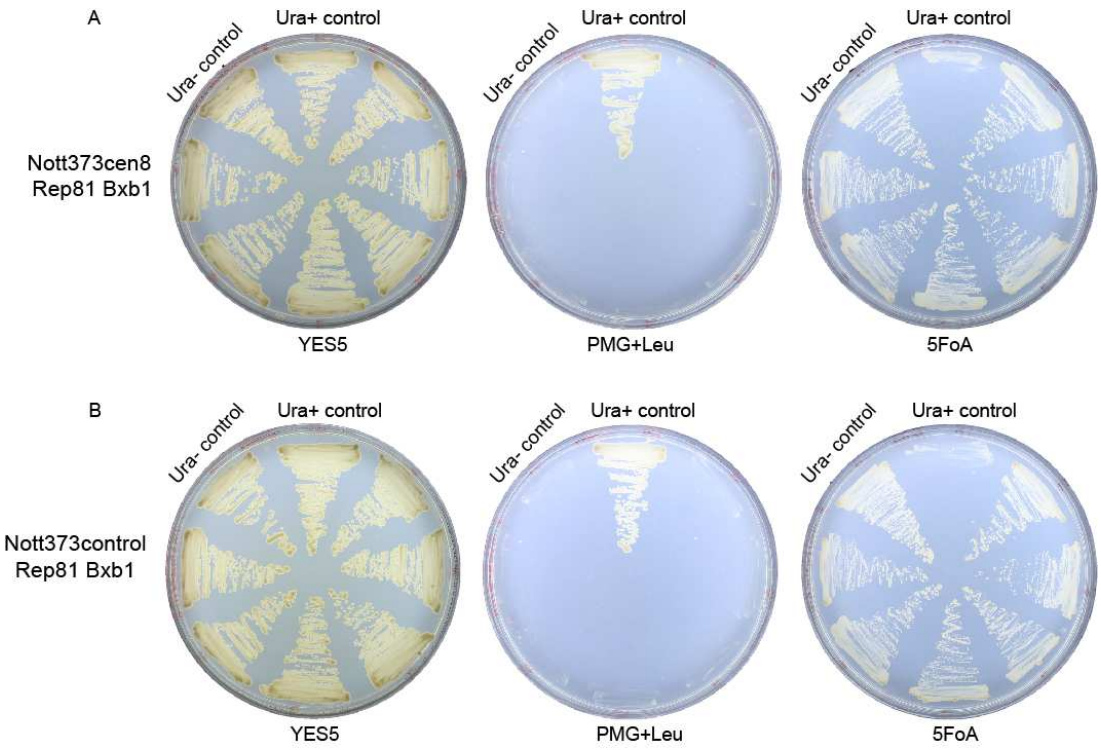



**Figure 4.7: *ura4Δ* clones growth on YES, PMG supplemented with Leucine, and 5-FOA.**
(A) *ura4Δ* Nott373_cen8 clones. (B) *ura4Δ* Nott373_control clones.

In summary, the small size of the colonies, the low viability of the cells, and low proportion of *ura4*Δ cells in the strains carrying either no test sequence, *S. pombe* centromeric DNA with less than 3.6kb in size, or the alphoid DNA confirmed the idea that the majority of the cells of these strains were dying after the centromere 2 excision. The strain harbouring the 3.6kb test sequence presented small colonies after Bxb1 integrase induction, the proportion of viable cells and of *ura4*Δ colonies was higher than the in the colonies from the previously described strains, but the centromere formation is likely to be unstable, suggested by the variable colonies sizes on YES. The strain carrying 4.2kb of *S. pombe* centromeric DNA showed colony sizes similar to the 3.6kb strain after transformation, but higher proportion of viable cells and of *ura4*Δ colonies. Finally, the strains carrying 6.8kb and 9.4kb of *S. pombe* centromeric DNA gave colonies that were similar in size and cells viability upon transformation with either the empty REP81 vector or the REP81 Bxb1 integrase plasmid and contained a high proportion of *ura4*Δ cells.

## 4.5. Southern blot analyses of the candidate swapped clones

In order to confirm deletion of the native centromere 2, ten *ura4Δ* clones of each strain were submitted to restriction enzyme digestions followed by gel electrophoresis and filter hybridization. The genomic DNA of the clones carrying mutant *S. pombe* centromeric DNA (Nott373_cen1, Nott373_cen2, Nott373_cen3, Nott373_cen4, Nott373_cen5, Nott373_cen8) and the Nott373_control was digested with the enzymes *Bgl*II, which cuts the DNA flanking the left-hand side of the native centromere 2, and *Pst*I, which cuts the nourseothricin resistance gene approximately 900 bp from the start of the gene (figure 4.8 A and B). The genomic DNA of the clones derived from Nott373_alphoid strain were digested with *Bgl*II, which cuts the DNA flanking the test sequence (figure 4.8C). The digested genomic DNA was size fractionated by a vertical agarose gel electrophoresis and transferred to nylon filters. The filters were hybridized with the probes for the nourseothricin resistance gene, the *S. pombe* centromeric DNA of chromosome 2, and the *ura4+* marker, respectively (figure 4.7). The probes were washed out from the filters with boiling 0.5% SDS and 0.1x TE buffer before hybridization with the next probe. Not all the probe can be removed with this process, so the filters showed some background on the following hybridization. The controls were the strains not transformed with the REP81 Bxb1 integrase plasmid.

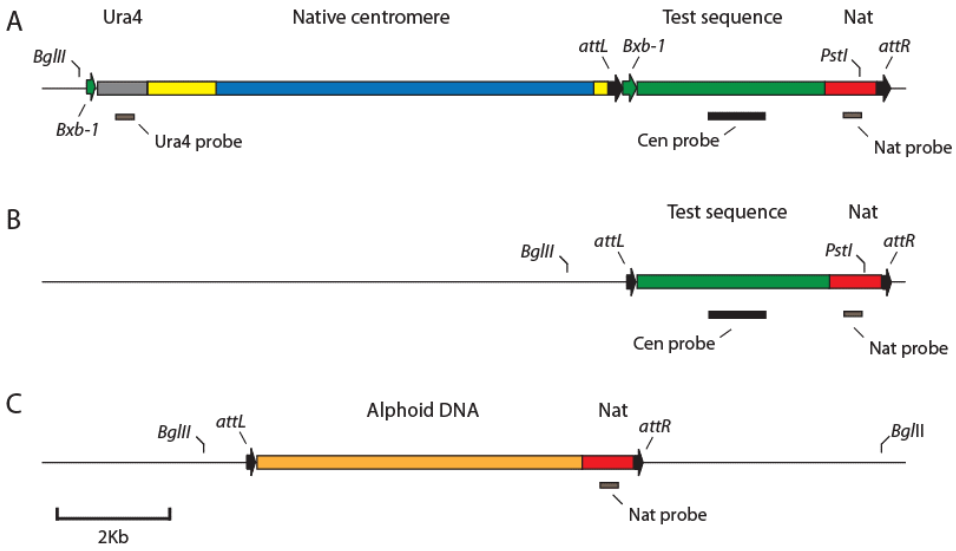



**Figure 4.8: Map showing the position of the probes and restriction enzyme sites on the centromere 2 region of the strains.**
The Ura4 probe hybridises with the *ura4*$^{+}$ marker, the Cen probe hybridises with the *S. pombe* mutant centromeric DNA (test sequences), and the nat probe hybridises with the nourseothricin resistance gene. A- General representation of the strains before the Bxb1 integrase induction. The genomic DNA of these strains was digested with *Bgl*II and *Pst*I and used as controls. B- General representation of the clones carrying *S. pombe* mutant centromeric DNA as test sequences after the Bxb1 integrase induction. The genomic DNA of these clones was digested with *Bgl*II and *Pst*I. C- Nott373alphoid strain after the Bxb1 induction. The genomic DNA of the clones derived from this strain was digested with *Bgl*II.

Figures 4.9 to 4.16 show the hybridised filters and the expected DNA fragments sizes for each probe. Nott373_control showed one clone harbouring the *ura4*$^{+}$ gene (figure 4.9). Nott373_cen2, Nott373_cen3, Nott373_cen4, and Nott373_alphoid presented *ura4*Δ clones in which the centromere 2 have not been excised (figure 4.11, 4.12, 4.13, and 4.16); the Nott373_cen2 and the Nott373_cen3 presented clones in which the DNA fragment sizes were not the predicted nor the size of the control, suggesting that these clones suffered rearrangements in the DNA of the chromosome 2 (figures 4.11 and 4.12). All the tested clones of Nott373_cen1, Nott373_cen5 and Nott373_cen8 had the centromere 2 excised (figures 4.0, 4.14, and 4.15).

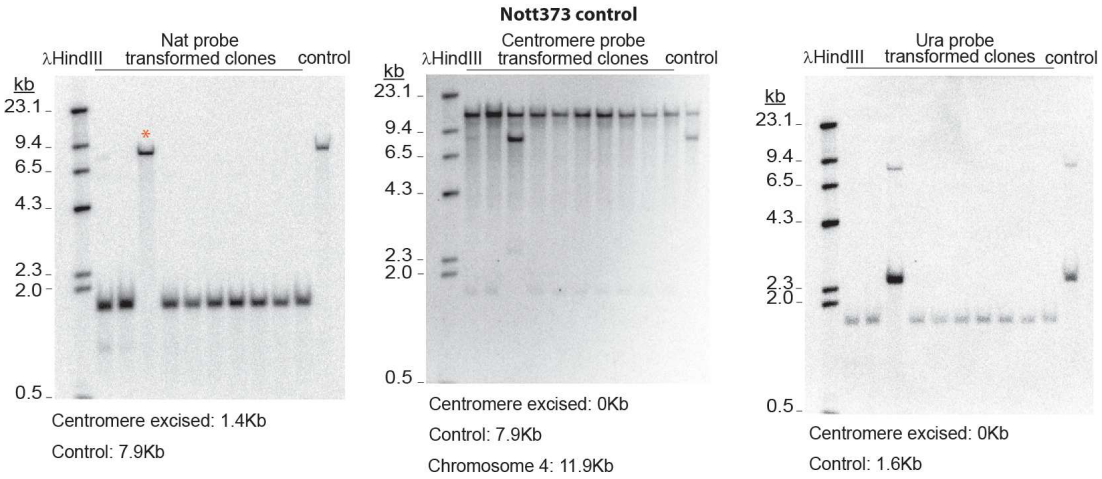



**Figure 4.9: Nott373_control. The genomic DNA was digested with *Bgl*II and *Pst*I and hybridized with probes for the nourseothricin resistance gene, centromeric DNA, and *ura4+* marker, respectively.**
The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot. *Clone in which the centromere 2 was not excised.

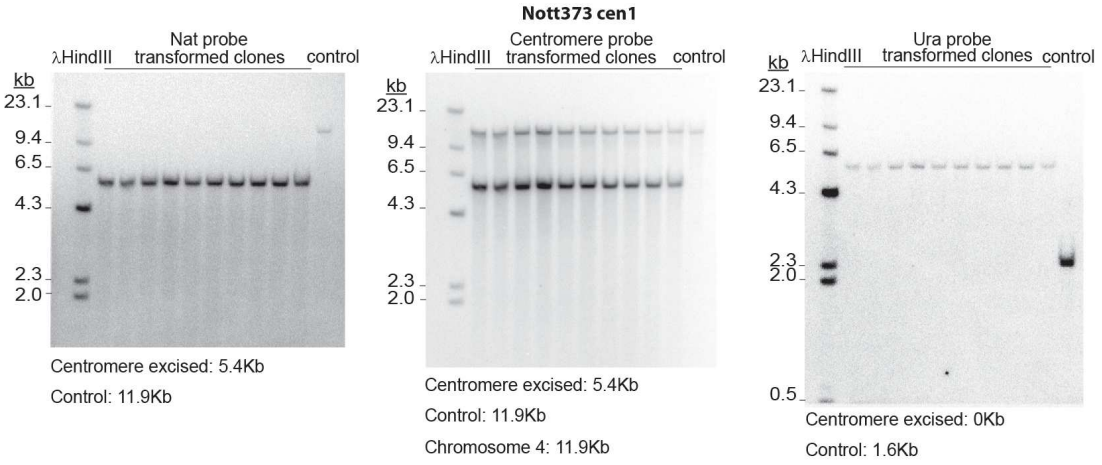



**Figure 4.10: Nott373_cen1. The genomic DNA was digested with *Bgl*II and *Pst*I and hybridized with probes for the nourseothricin resistance gene, centromeric DNA, and *ura4+* marker, respectively.**
The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot.

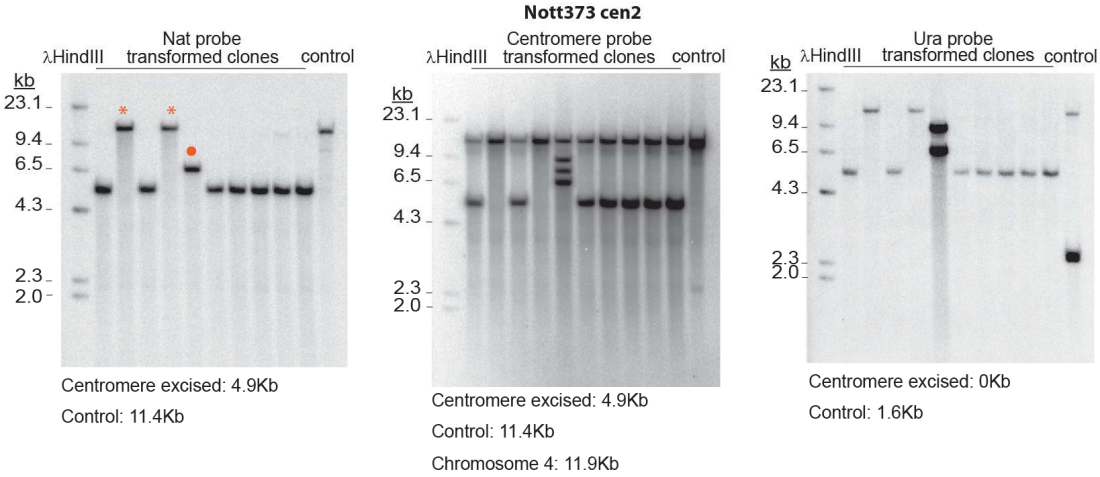



**Figure 4.11: Nott373_cen2. The genomic DNA was digested with *Bgl*II and *Pst*I and hybridized with probes for the nourseothricin resistance gene, centromeric DNA, and *ura4*+ marker, respectively.**
The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot. * Clones in which the centromere 2 was not excised. ○ Clone that probably had the chromosome 2 re-arranged.

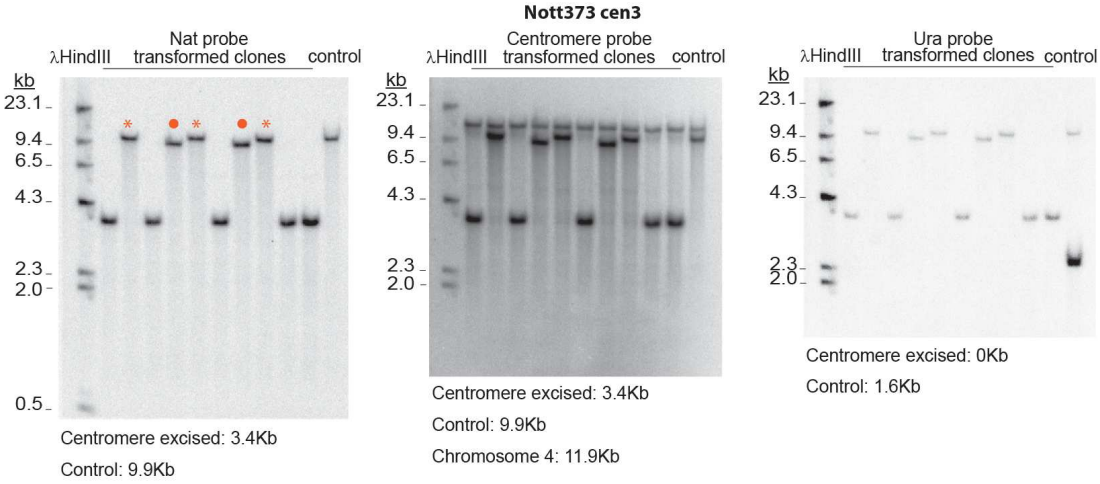



**Figure 4.12: Nott373_cen3. The genomic DNA was digested with *Bgl*II and *Pst*I and hybridized with probes for the nourseothricin resistance gene, centromeric DNA, and *ura4*+ marker, respectively.**
The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot. * Clones in which the centromere 2 was not excised. ○ Clones that probably had the chromosome 2 re-arranged.

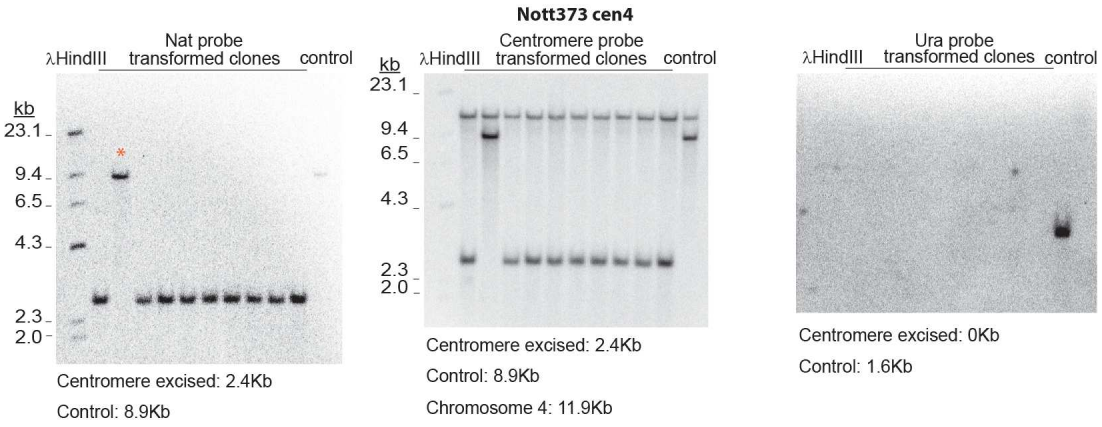



**Figure 4.13: Nott373_cen4. The genomic DNA was digested with *Bgl*II and *Pst*I and hybridized with probes for the nourseothricin resistance gene, centromeric DNA, and *ura4*+ marker, respectively.**
The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot. * Clone in which the centromere 2 was not excised.

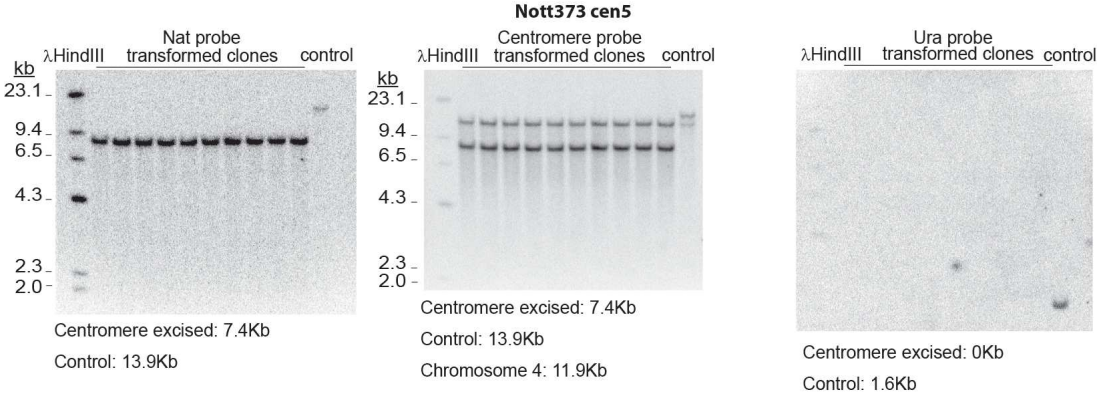



**Figure 4.14: Nott373_cen5. The genomic DNA was digested with *Bgl*II and *Pst*I and hybridized with probes for the nourseothricin resistance gene, centromeric DNA, and *ura4*+ marker, respectively.**
The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot.

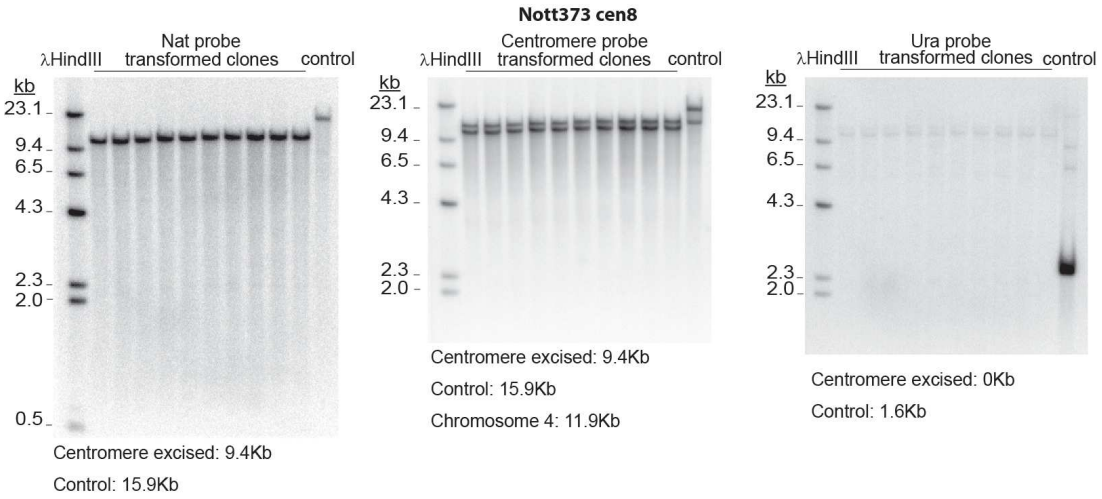



**Figure 4.15: Nott373_cen8. The genomic DNA was digested with *Bgl*II and *Pst*I and hybridized with probes for the nourseothricin resistance gene, centromeric DNA, and *ura4+* marker, respectively.**
The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot.

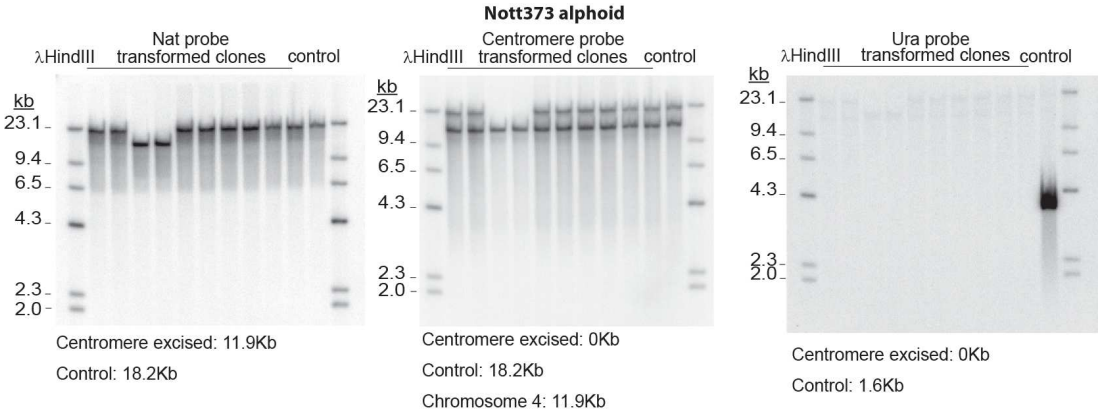



**Figure 4.16: Nott373_cen3. The genomic DNA was digested with *Bgl*II and hybridized with probes for the nourseothricin resistance gene and *ura4+* marker, respectively.**
Two of the ten tested clones had the native centromere 2 excised. The control is the untransformed strain. The sizes of the DNA fragments predicted to arise following centromere excision are written under the image of each blot. The third and fourth tested clones had the centromere 2 excised, the other 8 clones not.

These results demonstrated directly that the Bxb1 integrase has excised the native centromere without additional reagents. The fact that I analysed only 10 clones makes it difficult to estimate the efficiency of the process from this data. Table 4.3 summarises the Southern blot data.

| Strain | Swapped* |
|---|---|
| Nott373_control (0kb) | 9/10 |
| Nott373_cen4 (1kb) | 9/10 |
| Nott373_cen3 (2kb) | 5/10 |
| Nott373_cen2 (3.5kb) | 7/10 |
| Nott373_cen1 (4.2kb) | 10/10 |
| Nott373_cen5 (6.8kb) | 10/10 |
| Nott373_cen8 (9.5kb) | 10/10 |
| Nott373_alphoid (5.8kb) | 2/10 |

**Table 4.3: Southern blot analysis of the candidate swapped clones.**
*Swapped: Number of *ura4Δ* clones that had the native centromere accurately swapped from chromosome 2 expressed as Number of swapped clones/Number of analysed clones.

## 4.6. Measuring the proportion of *ura4*Δ clones that had centromere 2 excised

The fact that I nalysed only 10 clones by Southern blot analysis meant that an accurate estimate of the proportion of *ura4Δ* clones that had centromere 2 excised was not possible. In order to determine this estimate, 40 clones of each strain were also tested by PCR. The forward primer used aligned outside the Bxb1 *attR* site and the reverse primer aligned to the nourseothricin resistance gene (figure 4.17). The sequences of the primers are listed in table 2.1 in the material and methods section and the expected DNA fragment sizes for each strain are detailed in table 4.4.

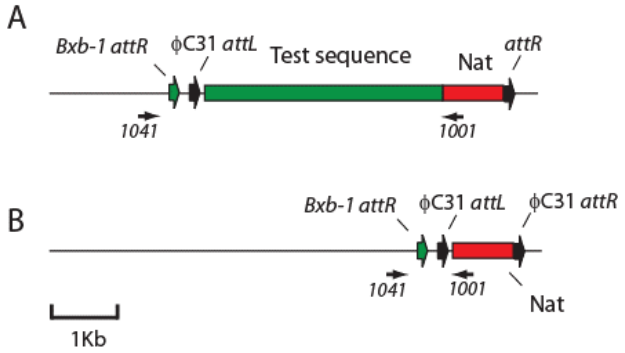
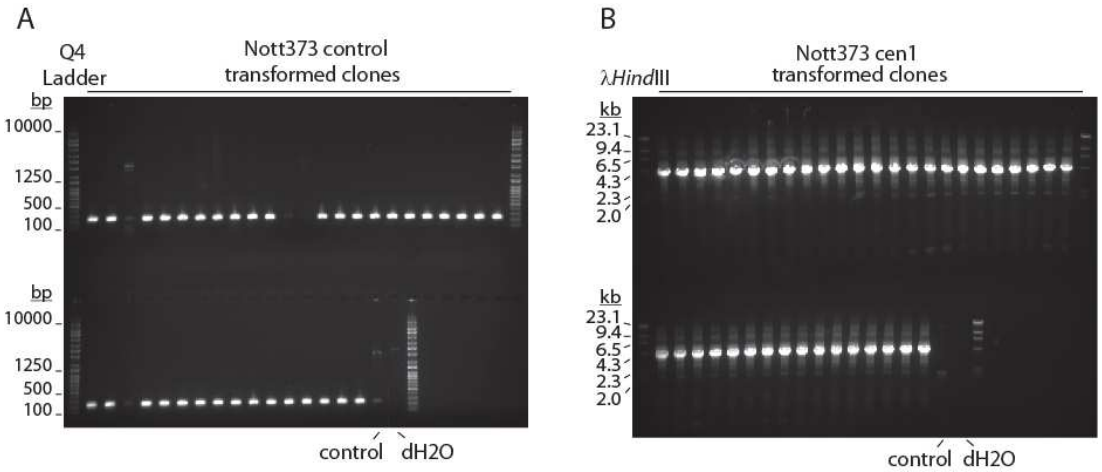



**Figure 4.17: General representation showing the position of the primers 1001 and 1041 on chromosome 2 of *S. pombe* recombined strain after centromere 2 excision.**
(A) Strains carrying test sequence adjacent to the centromere 2 locus (Nott373_cen1, Nott373_cen2, Nott373_cen3, Nott373_cen4, Nott373_cen5, Nott373_cen8, and Nott373_alphoid). (B) Strain that does not carry test sequence (Nott373_control).

| Strain | Size |
|---|---|
| Nott373_control | 250bp |
| Nott373_cen1 | 4250bp |
| Nott373_cen2 | 3750bp |
| Nott373_cen3 | 2250bp |
| Nott373_cen4 | 1250bp |
| Nott373_cen5 | 6250bp |
| Nott373_cen8 | 9650bp |
| Nott373_alphoid | 5950bp |

**Table 4.4: Expected PCR products size for the experiment used to confirm the proportion of *ura4Δ* clones that had excised the centromere 2.**

90% or more than of the *ura4Δ* clones of the strains Nott373_control, Nott373_cen1, Nott373_cen4, Nott373_cen5, and Nott373_cen8 had the centromere 2 excised (figure 4.18 A, B, D, E, and F). Nott373_cen3 presented 72.5% of centromere excision (figure 4.17C), Nott373_cen2 gave 60% of centromere excision (WRA Brown; personal communication), and the Nott373_alphoid had 25% of the *ura4Δ* clones with the centromere excised from chromosome 2 (figure 4.18G). As the PCR products of the Nott373_alphoid clones were very weak, the electrophoresis gel was transferred to a nylon filter and hybridised with a probe for the alphoid DNA to confirm that (figure 4.18H).



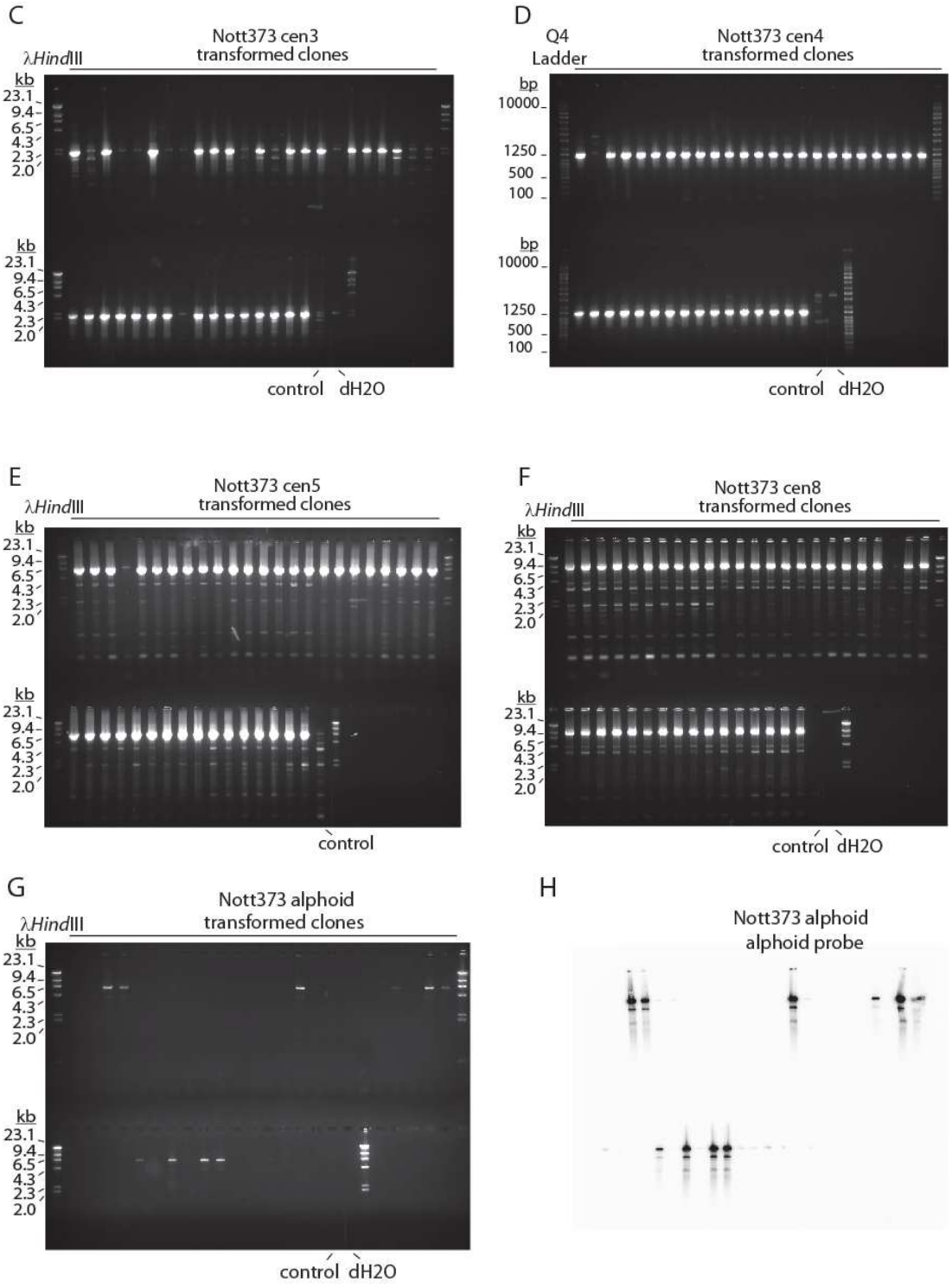



**Figure 4.18: Measuring the proportion of *ura4Δ* clones that had centromere 2 excised.**
A – G: Gels electrophoresis of the PCRs across the test sequences. H – Filter hybridization of the gels electrophoresis containing the Nott373alphoid clones (G).

The proportion of cells that had the centromere 2 excised in each strain was corrected by multiplying the proportion of *ura4Δ* clones by the proportion of centromere 2 excised obtained in the PCR experiment (table 4.5). The proportion of cells in which centromere 2 was correctly excised was 3% or lower when there is no

test sequence adjacent to the centromere 2, when the test sequence is the *S. pombe* mutant centromeric DNA with less than 3.6kb, or when it is the human alphoid DNA. This proportion increases to 35% when the test sequence is 3.6kb, then to approximately 70% when the test sequence is 4.2kb, and to 96% when the test sequence is 6.8kb or 9.4kb. These results corroborate for the idea that the candidate sequence length influences the centromere maintenance after the native centromere excision

| Strain | *ura*Δ cells | swapping checked by Southern blot | swapping checked by PCR | cen excised cells |
|---|---|---|---|---|
| **Nott373_control** | 0/100 | 9/10 | 36/40 | <0.029 |
| | 0/100 | | | <0.029 |
| | 10/100 | | | 0.1 |
| **Nott373_cen4** | 5/100 | 9/10 | 39/40 | 0.049 |
| | 2/100 | | | 0.02 |
| | 3/100 | | | 0.029 |
| **Nott373_cen3** | 0/100 | 5/10 | 29/40 | <0.029 |
| | 0/100 | | | <0.029 |
| | 0/100 | | | <0.029 |
| **Nott373_cen2** | 69/100 | 7/10 | 6/10 | 0.414 |
| | 49/100 | | | 0.294 |
| | 63/100 | | | 0.378 |
| **Nott373_cen1** | 52/100 | 10/10 | 40/40 | 0.52 |
| | 82/100 | | | 0.82 |
| | 73/100 | | | 0.73 |
| **Nott373_cen5** | 98/100 | 10/10 | 39/40 | 0.956 |
| | 99/100 | | | 0.965 |
| | 97/100 | | | 0.946 |
| | 99/100 | | | 0.965 |
| | 100/100 | | | 0.975 |
| **Nott373_cen8** | 100/100 | 10/10 | 39/40 | 0.975 |
| | 94/100 | | | 0.917 |
| | 100/100 | | | 0.975 |
| **Nott373_alphoid** | 0/100 | 2/10 | 10/40 | <0.029 |
| | 0/100 | | | <0.029 |
| | 1/100 | | | 0.003 |

**Table 4.5: Proportion of *ura4*Δ clones, *ura4*Δ and centromere excised clones, and real proportion of centromere excised clones in each strain.**

## 4.7. Rate of centromere formation

The differences in colonies size, cells viability and proportion of cells that had the centromere 2 excised in each strain suggest that the centromere excision was not occurring in every cell cycle and that the size and nature of the centromeric sequence influence the efficiency of centromere formation. With the data collected above, it was possible to calculate the frequency that each strain formed the centromere after the native centromere disruption.

After the Bxb1 integrase induction, in a cell cycle there were 3 possible events that could be measured (figure 4.19):

1- No excision of the centromere (*n*), which was measured by the number of Ura+ cells.
2- Excision and centromere formation (*s*), measured by the number of *ura4Δ* cells that had the centromere excised
3- Excision and death (*x*), measured by comparing the colony size of the strains transformed with the Bxb1 integrase plasmid and colony size of the strains transformed with the empty vector.

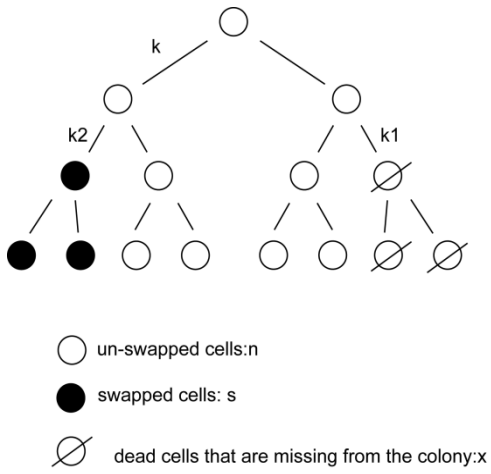



**Figure 4.19: Rate of centromere formation.**
Events that could be measured after the Bxb1integrase induction: 1- no excision of the centromere 2 (white circle); 2- excision of the centromere 2 (black circle); 3- excision of the centromere 2 and consequent death of the cell (crossed circle). *k*, $k_1$ and $k_2$ are the rates of each event.

Knowing n, s and x the rate of excision and death ($k_1$), and the rate of excision and centromere formation ($k_2$) could be calculated as a proportion of the doubling rate of the cells (*k*) as following:

The variation in number of cells that did not have the centromere 2 excised in a colony in a cell cycle can be calculates as:

$\frac{dn}{dt} = (k - k_1 - k_2)\mathrm{n}$ (1)

The variation of the number of cells have the centromere 2 excised and survived in a colony and the number of dead cells in a colony are, respectively:

$\frac{ds}{dt} = k.s + k_2.n$ (2)

$\frac{dx}{dt} = k.x + k_1.n$ (3)

From (2) we can get the equation that determine the number of cells that excise and form a centromere ($s$):

$\frac{ds}{dt} - k.s = k_2.n$

$e^{-\int kdt}\frac{\partial s}{\partial t} - k^{-\int kdt}.s = k_2.e^{-\int kdt}.n$

$e^{-kt}\frac{ds}{dt} - k.e^{-kt}.s = k_2.e^{-kt}.n$

$\frac{d}{dt}(e^{-kt}.s) = k_2.e^{-kt}.n(t)$

$s.e^{-kt} = k_2\int e^{-kt}.n(t)dt$

$s.e^{-kt} = k_2\int e^{-kt}.e^{(k-k_1-k_2)t}dt$

$s.e^{-kt} = k_2\int e^{-(k_1+k_2)t}\,dt$

$s.e^{-kt} = \frac{-k_2e^{-(k_1+k_2)t}}{(k_1+k_2)} + c$

When $s = 0$,

$c = \frac{k2}{(k1+k2)}$

$se^{-kt} = \frac{k2}{(k1+k2)}(1 - e^{-(k1+k2)t})$

So,

$s = e^{kt}\frac{k2}{(k1+k2)}(1 - e^{-(k1+k2)t})$

From (3), the number of cells that excise the centromere and die is

$x = e^{kt}\frac{k1}{(k1+k2)}(1 - e^{-(k1+k2)t})$

And the number of cells that do not excise the centromere is

$n = e^{kt}(e^{-(k1+k2)t})$

Calculation examples are on Appendix II.

The calculation of $k_1/k$ (rate of excision and death) and $k_2/k$ (rate of excision and centromere formation) were done in triplicate for each strain (table 4.6). The results show that when the centromeric DNA length is 6.8kb or 9.5kb the rate of excision and centromere formation is 1 in every 5 cell cycles and the death is not detectable. When the centromeric DNA length is 2kb or less the rate of excision and centromere formation is $10^{-4}$ or lower and the death is 1 in every 5 cycles. When the length is 3.6kb or is 4.2kb the centromere formation rate range from 1 every 5 cycles and 1 every 100 cycles and the death is lower than what was observed for smaller fragments. Figure 4.20 shows the average of the measurements.

These observation demonstrate that the length of the centromeric DNA has a role on centromere formation.

| Strain | Length of centromeric DNA | k1/k | k2/k |
|---|---|---|---|
| **Nott373_control** | 0kb | 0.32 | 0 (<0.00006) |
| | | 0.32 | 0 (<0.00018) |
| | | 0.25 | 0.00059 |
| **Nott373_cen4** | 1kb | 0.22 | 0.00042 |
| | | 0.2 | 0.0002 |
| | | 0.21 | 0.00029 |
| **Nott373_cen3** | 2kb | 0.17 | 0 (< 0.0031) |
| | | 0.27 | 0 (< 0.00009) |
| | | 0.25 | 0 (< 0.00012) |
| **Nott373_cen2** | 3.6kb | 0.04 | 0.038 |
| | | 0.01 | 0.022 |
| | | 0.02 | 0.037 |
| **Nott373_cen1** | 4.2kb | 0.09 | 0.032 |
| | | 0.15 | 0.056 |
| | | 0.22 | 0.034 |
| **Nott373_cen5** | 6.8kb | nd | 0.176 |
| | | 0.041 | 0.192 |
| | | nd | 0.189 |
| | | 0.004 | 0.17 |
| | | 0.008 | 0.239 |
| **Nott373_cen8** | 9.5kb | nd | 0.25 |
| | | nd | 0.22 |
| | | nd | 0.25 |
| **Nott373_alphoid** | 5.7kb | 0.16 | 0 (<0.00008) |
| | | 0.16 | 0 (<0.00009) |
| | | 0.33 | 0.00004 |

**Table 4.6: Rates of excision and death (k1/k) and of excision and centromere formation (k2/k) as a proportion of the rate of cells growth (k).**

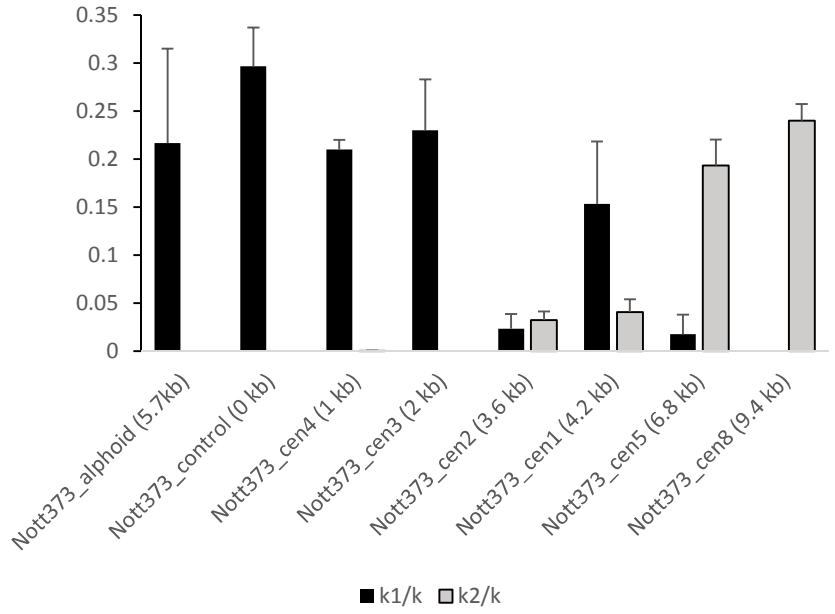



**Figure 4.20: Average rates of excision and death (k1) and of excision and centromere formation (k2) as a proportion of the rate of cells growth (k).**
When the centromeric DNA length is 6.8kb or 9.5kb k2/k is around 0.2, what means 1 in every 5 cell cycles and the death (k1/k) is not detectable for the 9.5 kb test sequence and lower than 0.05 in the 6.8 kb test sequence. When the centromeric DNA length is 2kb or less the rate of excision and centromere k2/k is is almost undetectable k1/k is 1 in every 5 cycles. When the length is 3.6kb or is 4.2kb the centromere formation rate range from 1 every 5 cycles and 1 every 100 cycles and the death is lower than what was observed for smaller fragments. The error bars show the standard deviation.

## 4.8. Discussion

In this chapter, I have described a system that allows the native centromere to be excised from a chromosome that has been engineered so that it contains a test sequence adjacent to this native centromere. In principal, this should allow one to assay the ability of different sequences to function as centromeres. Working on this assumption, we have established a framework for using the approach to assay the centromeric activity of different sequences. This has established that the Bxb1 integrase deletes the centromere about once every four or five cell cycles and that centromeres are formed over the candidate sequences with nearly absolute efficiency if the candidate sequence is equal to or greater than 6.8kb in size. However, the centromere excision and centromere formation is only detectable at frequencies of $5x10^{-3}$ or less if the candidate is 3.6kb or smaller in length or is composed of human Y

alphoid DNA. The threshold is about 4kb because at this length the candidate supports centromere formation (as judged by the recovery of viable cells) at an intermediate rate. The ability of a sequence to support centromere is sequence and species specific as human Y alphoid DNA shows no activity in our assay.

# 5. CNP1 BINDING TO ENGINEERED CENTROMERES BY CHIP-SEQ.

## 5.1. Introduction

In most eukaryotes, the defining feature of centromeres is the presence of nucleosomes containing the centromere protein A (CENP-A) (Foltz et al., 2006). Although CENP-A is required for centromere function, the mechanism by which it promotes and stabilises the centromere in eukaryotic cells is not yet known.

Initially, I set out to test the idea that exists an evolutionarily conserved mechanism that regulates the amount of Cnp1 bound at centromeres independently of the length of the underlying centromeric DNA, as seems to hold in human (Irvine et al., 2004). In order to do this, I used the natural isolate of *S. pombe*, CBS 2777 and serine recombinases to engineer a system in which I could manipulate the amount of centromeric DNA on CBS 2777 chromosome 2, as described in previous chapters. In this chapter, I describe measurements of the amount and position of bound Cnp1 by sequencing the DNA co-immunoprecipitated (ChIP-seq) with the GFP amino terminus tagged protein Cnp1 from strains containing different amounts of centromeric DNA in the centromere of chromosome 2.

The kinetochores of fission yeast and many other eukaryotes are known to be embedded in the pericentromeric heterochromatin (Karpen and Allshire, 1997, Grewal and Jia, 2007) which has been suggested to be an epigenetic mark that contributes to centromere identity and function (Cleveland et al., 2003, Henikoff et al., 2001). This pericentric heterochromatin is marked by the presence of methylated histone H3K9 (histone H3 methylated on lysine 9) and several chromodomain proteins that bind to this modified histone. The RNA-induced transcriptional silencing (RITS) complex, a kind of RNA interference (RNAi) by which short RNA molecules trigger the downregulation of transcription of a particular gene, associates with H3K9 leading to further H3K9 methylation and the recruitment of Swi6. The Swi6 is a chromodomain protein, homolog of mammalian and *Drosophila* heterochromatin protein 1 (HP1) (Ekwall et al., 1995). It localises to the centromeric outer repeat regions, and also to the telomeres and silent mating type locus (Ekwall et al., 1995, Partridge et al., 2000, Nakayama et al., 2000, Noma et al., 2001).

In order to analyse the heterochromatin requirements for neo-centromere formation, I deleted the gene encoding the chromodomain protein Swi6 in strains Nott373_control and Nott373_cen5. The native centromere 2 of the *swi6* mutants strains were then excised the by Bxb1 recombinase induction. The measurements of the amount and position of bound Cnp1 by ChIP-seq in these strains are also described in this chapter.

## 5.2. Analyzed clones

The clones used for the ChIP-seq experiments are listed in table 5.1.

| Strain | Genotype | Source |
|---|---|---|
| Nott588 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen5 nat - clone 1* | This study |
| Nott589 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB alphoid nat - clone 1* | This study |
| Nott590 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB alphoid nat - clone 8* | This study |
| Nott591 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen3 nat - clone 25* | This study |
| Nott592 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen3 nat - clone 29* | This study |
| Nott593 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen8 nat - clone 3* | This study |
| Nott594 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen8 nat - clone 4* | This study |
| Nott595 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB Cen8 nat - clone 5* | This study |
| Nott596 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB nat - clone 2* | This study |
| Nott597 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2L:: Bxb1 attP ura4⁺ cen2R:: Bxb1 attB nat - clone 4* | This study |
| Nott631 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D nat - clone 4* | This study |
| Nott632 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D nat - clone 5* | This study |
| Nott638 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen1 nat - clone 1* | This study |
| Nott639 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen1 nat - clone 2* | This study |
| Nott648 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen2 nat - clone 1* | This study |
| Nott650 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen2 nat - clone 4* | This study |
| Nott653 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen2 nat - clone 7* | This study |
| Nott654 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen2 nat - clone 8* | This study |
| Nott658 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen3 nat - clone 1* | This study |
| Nott660 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen3 nat - clone 3* | This study |

| | | |
|---|---|---|
| Nott671 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen4 nat - clone 4* | This study |
| Nott674 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen4 nat - clone 7* | This study |
| Nott678 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen5 nat - clone 1* | This study |
| Nott679 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen5 nat - clone 2* | This study |
| Nott688 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen3 nat - clone 1* | This study |
| Nott689 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D Cen3 nat - clone 2* | This study |
| Nott700 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D alphoid nat - clone 3* | This study |
| Nott701 | *CBS2777 leu1-D32 ura4-D18 GFP-cnp1 cen2D alphoid nat - clone 4* | This study |

**Table 5.1: Strains used for the ChIP-seq experiments.**

## 5.3. Cnp1 spreading over the test sequences

ChIP-seq was carried out for Nott373_cen5, Nott373_cen8, and Nott373_alphoid (table 3.2) before centromere 2 excision to examine the consequences of increasing centromeric DNA. As a control, the strain containing the integrated empty vector (Nott373_control) was also analysed. All analyses were carried out in duplicate. The analysis of Nott 373_control confirmed the results of the analysis of Brown et al. (2014) and showed that, in chromosome 2, Cnp1 binds primarily to the centromere central core and flanking sequences chromosome (II: 125168-133320) (figure 5.1 A and B), and to minor sites at the telomere (green arrows on figure 5.1) and at the glyceraldehyde 3-phosphate gene *tdh1* (green arrows on figure 5.1).

Analysis of the Cnp1 data show that increasing the amount of *S. pombe* centromeric DNA, the Cnp1 spreads over the test sequences of Nott373_cen5 (6.8kb test sequence) (Figure 5.1 C and D) and Nott373_cen8 (9.5kb test sequence) (Figure 5.1 E and F) binding proportionally more Cnp1 (Figure 5.2). In *S. pombe*, human alphoid DNA does not bind Cnp1 (Figure 5.1 G and H), suggesting that there are specific DNA sequence requirements for Cnp1 binding in the fission yeast.

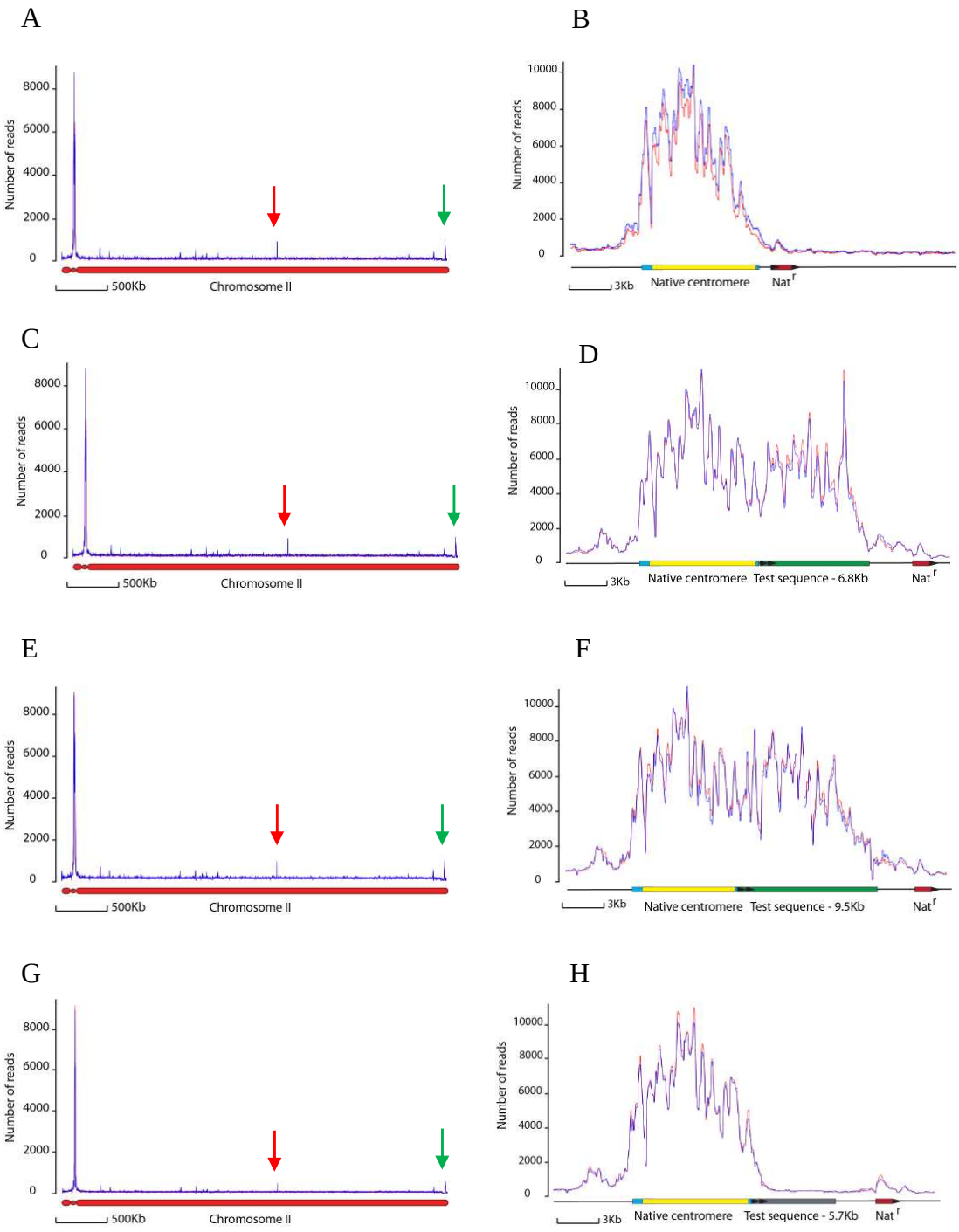



**Figure 5.1: Localisation of centromeric protein Cnp1 on engineered CBS 2777 strains by ChIP-seq of GFP-tagged Cnp1.**

The ChIP-seq experiment was carried out in duplicate and the samples are color-coded in red and blue. Major binding sites are localised on the centromeric region, and a minor binding site is observed on *tdh1* gene (red arrows) and telomeric region (green arrows). (A-B) Cnp1 binding on chromosome 2 and on centromere 2 of Nott373_control, respectively. Strains Nott596 and Nott597 (C-D) Cnp1 biding on chromosome 2 and on centromere 2 of Nott373_cen5, respectively. Strains Nott588. Cnp1 spreads over the 6.8kb *S. pombe* mutant centromeric DNA. (E-F) Cnp1 binding on chromosome 2 and on centromere 2 of Nott373_cen8, respectively. Strains Nott591 and Nott592. Cnp1 spreads over the 9.5kb *S. pombe* mutant centromeric DNA. (G-H) Cnp1 binding on chromosome 2 and on centromere 2 of Nott373_alphoid, respectively. Cnp1 does not spread over the 5.7kb human alphoid DNA. Only those reads in which the mates map within 180 bp and 250 bp of one another were used in this mapping.

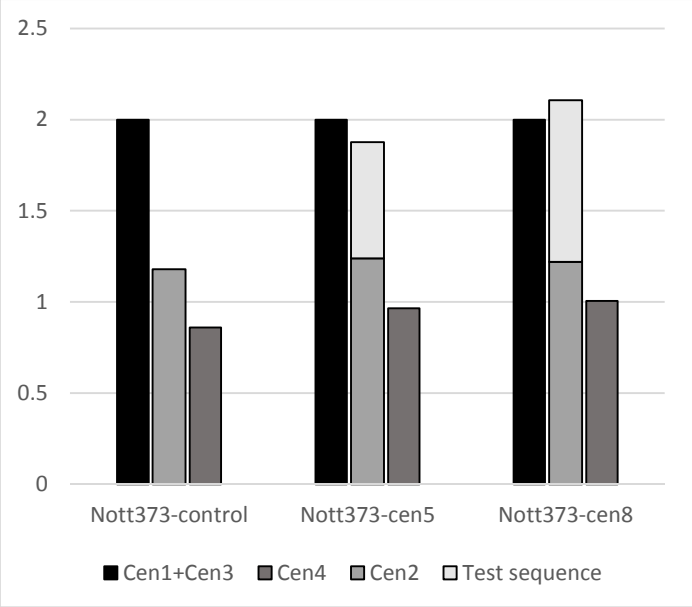



**Figure 5.2: Proportion of Cnp1 binding on Nott373 strains.**
The total number of sequencing reads was normalised to the sum of reads of chromosomes 1 and 3 of each sample. Increasing the length of *S. pombe* centromeric DNA the amount of Cnp1 bound increases proportionally.

## 5.4. Centromere and neo-centromere formation after excision of the native centromere

ChIP-seq on tagged Cnp1 was carried out for the engineered Nott373 strains after excision of the native centromere. The analysis was carried out in duplicate with the exception of the products of excision from the Nott373_cen2 that had 4 samples sequenced because of the apparent heterogeneity of the clones explained on section 4.4. Strains carrying test sequences of 4kb or more were able to bind the Cnp1 and form a functional centromere over the test sequence. Functional centromeres formed over the test sequence were observed for both samples of Nott373_cen5 (6.8kb test sequence) (Figure 5.3), Nott373_cen8 (9.5kb test sequence) (Figure 5.4), and for one of the Nott373_cen1 (4.2kb test sequence) (Figure 5.5A).

The second Nott373_cen1 sample (Figure 5.5B), three samples of Nott373_cen2 (3.6kb test sequence) (Figure 5.9A), and both analysed sample of Nott373_cen3 (2kb test sequence) (Figure 5.8), Nott373_cen4 (1kb test sequence) (Figure 5.7), Nott373_control (0kb test sequence) (Figure 5.6), and Nott373_alphoid (5.8kb human alphoid DNA) (Figure 5.10) formed a neo-centromere at a similar location, close to the distal telomere on chromosome 2. Thus, 12 strains in total formed a neo-centromere over the same 13kb subtelomeric region. These results are similar to

those of Ishii et al. (2008) who observed neo-centromere formation in the subtelomeric regions of chromosome 1 following excision of the native centromere.

One of the Nott373_cen2 samples analysed presented no Cnp1 major binding site on chromosome 2 (Figure 5.9B), suggesting that the chromosome 2 of this sample has been lost by fusion with another chromosome. We analysed 10 Ura- derivatives of each excision reaction by pulsed field gel electrophoresis and detected a re-arrangement and loss of chromosome 2 in one of the Ura- products of the excision of the native centromere from Nott373_cen2 sample. The pattern of ethidium bromide staining of the gel suggested that the chromosome 2 had fused with chromosome 3. Ishii et al. (2008) observed fused chromosomes following excision of the native centromere of chromosome I. They referred to these as Type II survivors but did not quantitate their frequency relative to those with a three chromosome karyotype; the so-called type I survivors.

Samples that had formed neo-centromere presented a minor Cnp1 binding site in the DNA flanking the pre-existent centromere 2 (Figures 5.5B, 5.6, 5.7, 5.8, 5.9A, 5.10). The region is comprised of two fragments of approximately 1kb that flank the left and the right ends of centromeres 2 and 4. This region is identical to the centromere flanking DNA on chromosome 4 and we assume that the binding that we are detecting on the derived chromosome 2 reflects an ectopic assignment of ChIP-seq reads derived from chromosome 4 (Figure 5.6B).

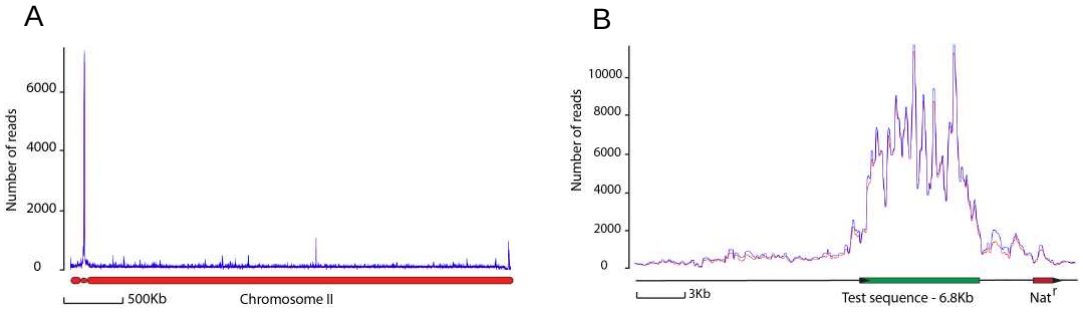



**Figure 5.3: Localisation by ChIP-seq of GFP-tagged Cnp1 of centromeric protein Cnp1 on Nott373_cen5 after excision of the native centromere.**
Strains Nott678 and Nott679. The ChIP-seq experiment was carried out in duplicate and the samples were color-coded in red and blue. (A) Cnp1 binding on Nott373_cen5 chromosome 2 after the native centromere excision. Major binding sites are localised on the centromeric region. (B) Cnp1 binding on Nott373_cen5 centromere 2. The 6.8kb test sequence binds Cnp1 and forms a functional centromere.

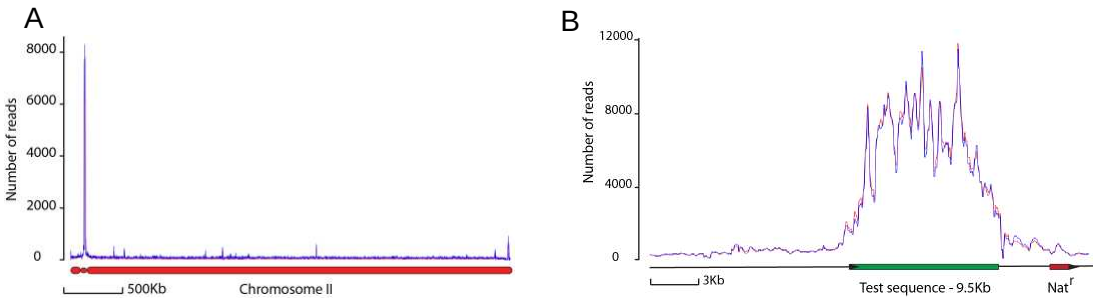



**Figure 5.4: Localisation by ChIP-seq of GFP-tagged Cnp1 of centromeric protein Cnp1 on Nott373_cen8 after excision of the native centromere.**
Strains Nott688 and Nott689. The ChIP-seq experiment was carried out in duplicate and the samples were color-coded in red and blue. (A) Cnp1 binding on Nott373_cen8 chromosome 2 after the native centromere excision. Major binding sites are localised on the centromeric region. (B) Cnp1 binding on Nott373_cen8 centromere 2. The 9.5kb test sequence binds Cnp1 and forms a functional centromere.

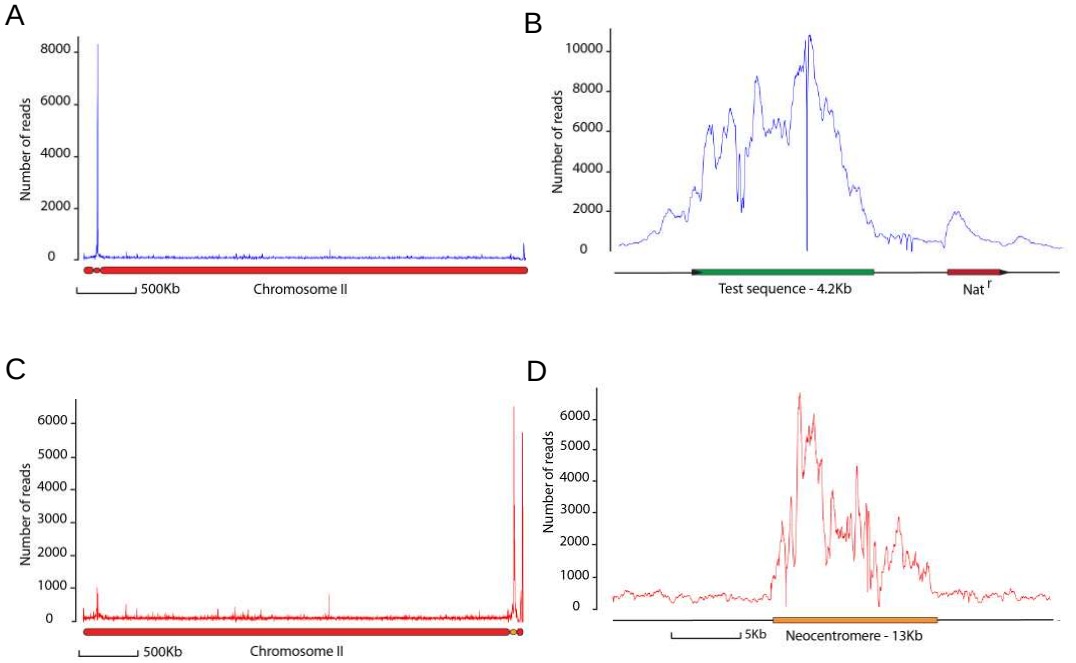



**Figure 5.5: Localisation by ChIP-seq of GFP tagged Cnp1 of centromeric protein Cnp1 on Nott373_cen1 after excision of the native centromere.**
(A) Strain Nott638. Cnp1 binding on Nott373_cen1 chromosome 2 after the native centromere excision. Major binding sites are localised on centromeric region. (B) Strain Nott638. Cnp1 binding on Nott373_cen8 centromere 2. The 4.2kb test sequence binds Cnp1 and forms a functional centromere. (C) Strain Nott639. Cnp1 binding on Nott373_cen1 chromosome 2 after the native centromere excision. Major binding sites are localised at the subtelomeric region. (D) Strain Nott639. Cnp1 binding on Nott373_cen1 subtelomeric region, where it formed a neo-centromere.

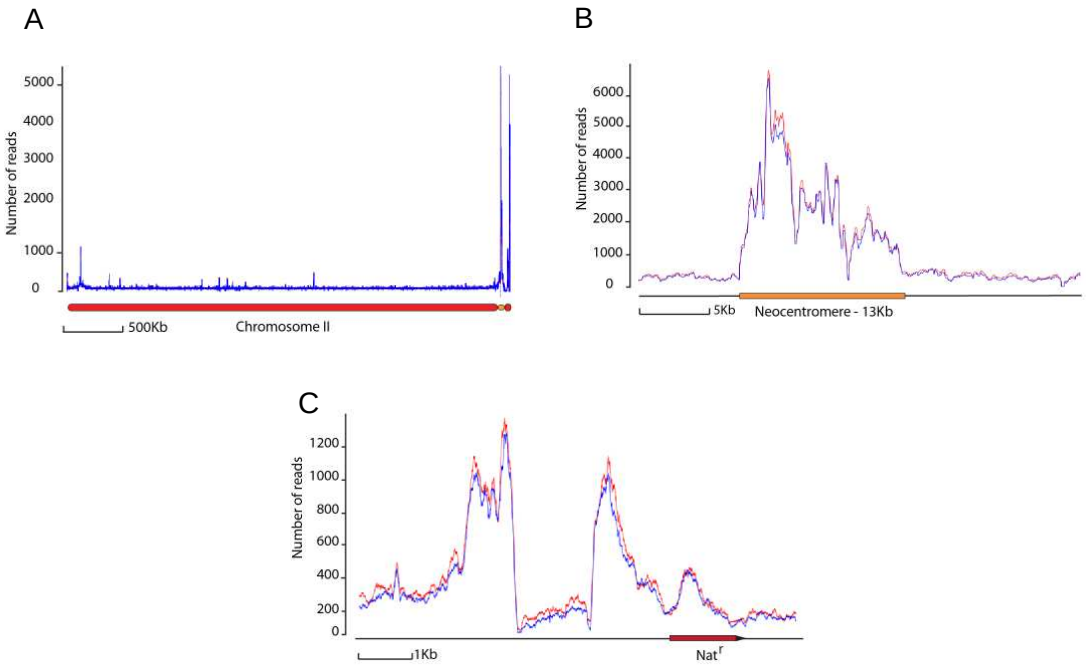



**Figure 5.6: Localisation by ChIP-seq of GFP tagged Cnp1 of centromeric protein Cnp1 on Nott373_control after excision of the native centromere.**
Strains Nott631 and Nott632. The ChIP-seq experiment was carried out in duplicate and the samples were color-coded in red and blue. (A) Cnp1 binding on Nott373_control chromosome 2 after the native centromere excision. Major binding sites are localised at the subtelomeric region. (B) Cnp1 binding on Nott373_control subtelomeric region, where it formed a neo-centromere. (C) Cnp1 binding on Nott373_control centrome 2 flanking reagion. Minor Cnp1 binding site was observed at the region of the pre-existent centromere 2. The site is formed by two fragments of approximately 1kb that flank both ends of centromere 2 and centromere 4. We judge that these reads are miss assigned to this region are consequence of sequence identity with the centromere flanking DNA of chromosome 4.

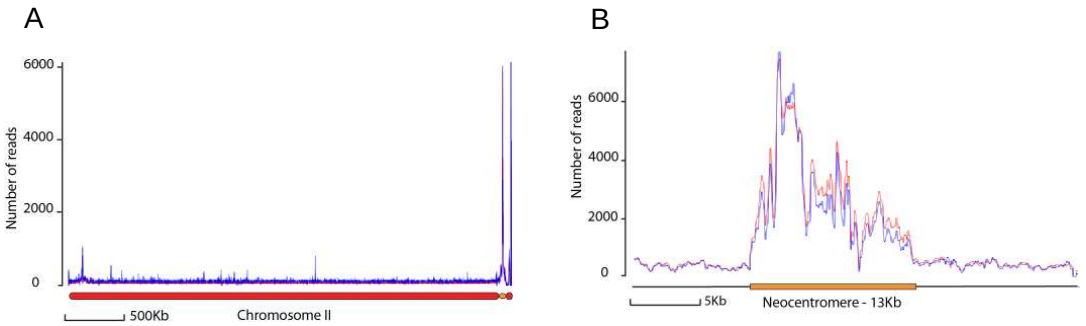



**Figure 5.7: Localisation by ChIP-seq of GFP tagged Cnp1 of centromeric protein Cnp1 on Nott373_cen4 vector after excision of the native centromere.**
Strains Nott671 and Nott672. The ChIP-seq experiment was carried out in duplicate and the samples were color-coded in red and blue. (A) Cnp1 binding on Nott373_cen4 chromosome 2 after the native centromere excision. (B) Cnp1 binding on Nott373_cen4 subtelomeric region, where it formed a neo-centromere.

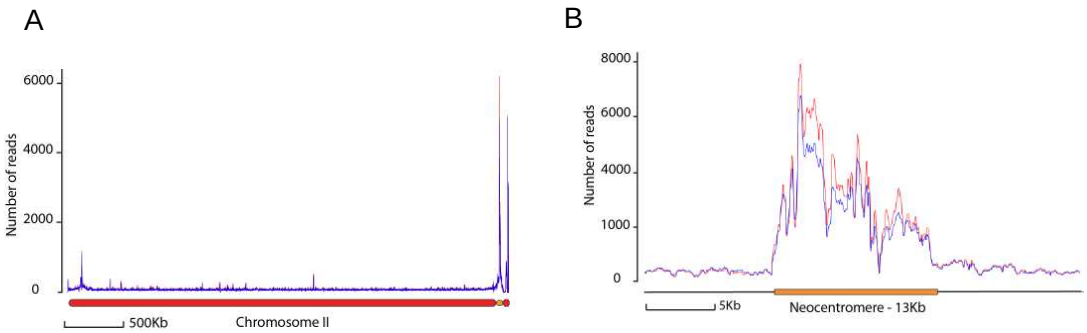



**Figure 5.8: Localisation by ChIP-seq of GFP tagged Cnp1 of centromeric protein Cnp1 on Nott373_cen3 after excision of the native centromere of chromosome 2.**
Strains Nott658 and Nott 660. The ChIP-seq experiment was carried out in duplicate and the samples were color-coded in red and blue. (A) Cnp1 binding on Nott373_cen3 chromosome 2 after the native centromere excision. (B) Cnp1 binding on Nott373_cen3 subtelomeric region, where it formed a neo-centromere.

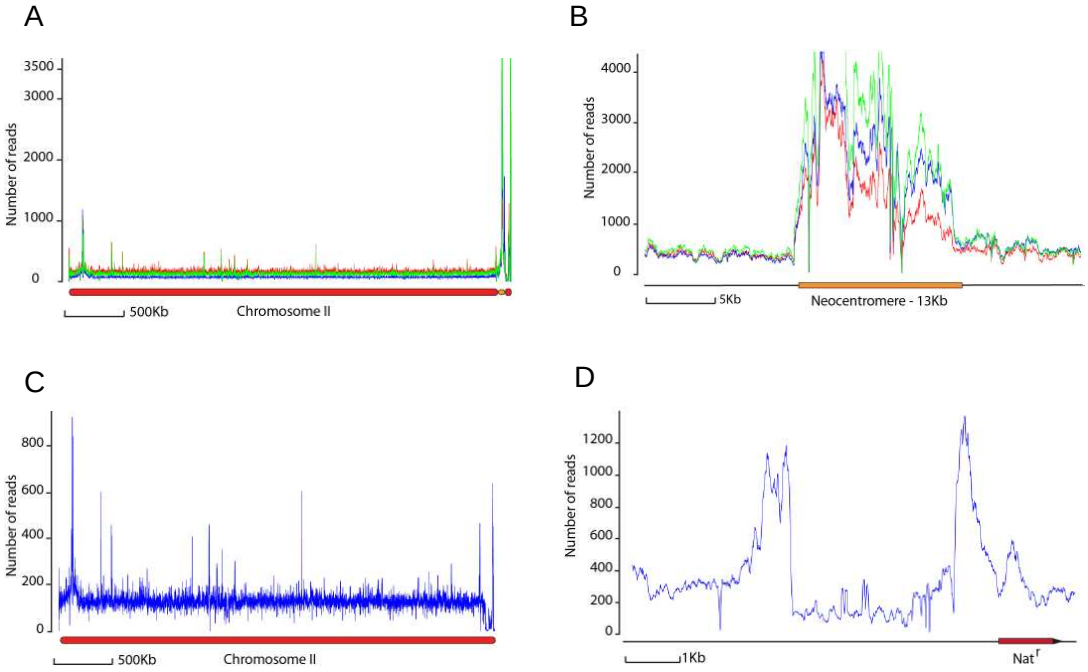



**Figure 5.9: Localisation by ChIP-seq of GFP tagged Cnp1 of centromeric protein Cnp1 on Nott373_cen2 after excision of the native centromere of chromosome 2.**
(A) Strains Nott648, Nott653, and Nott654 illustrated with the individual samples color-coded in blue, red and green. Cnp1 binding on Nott373_cen2 chromosome 2 after the native centromere excision. (B) Cnp1 binding on Nott373_cen2 subtelomeric region of strains Nott648, Nott653, and Nott654, where it formed a neo-centromere. (C) Strain Nott650. Cnp1 binding on Nott373_cen2 chromosome 2 after the native centromere excision. No major binding was observed in this strain. Pulsed field gel analysis showed that this strain had fused chromosome 3. This clone presented no major Cnp1 binding site. (D) Cnp1 binding on Nott373_cen2 centrome 2 flanking region. Minor Cnp1 binding site was observed at the region of the pre-existent centromere 2. The site is formed by two fragments of approximately 1kb that flank both ends of centromere 2 and centromere 4. We judge that these reads are miss assigned to this region are consequence of sequence identity with the centromere flanking DNA of chromosome 4.

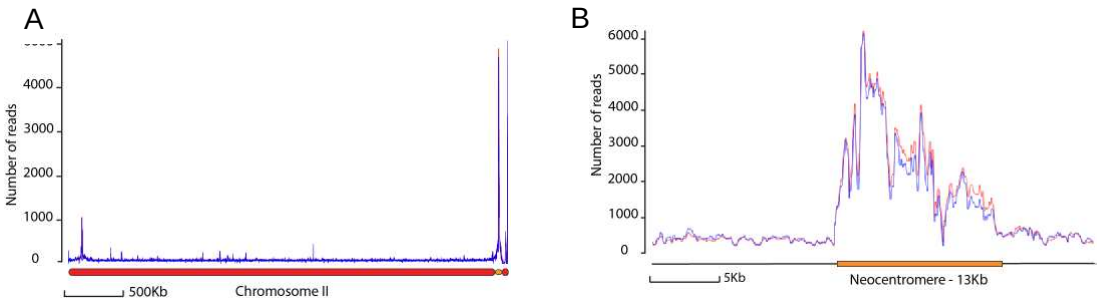



**Figure 5.10: Localisation by ChIP-seq of GFP tagged Cnp1 of centromeric protein Cnp1 on Nott373_alphoid after excision of the native centromere of chromosome 2.**
Strains Nott700 and Nott701. The ChIP-seq experiment was carried out in duplicate and the samples were color-coded in red and blue. (A) Cnp1 binding on Nott373_alphoid chromosome 2 after the native centromere excision. (B) Cnp1 binding on Nott373_alphoid subtelomeric region, where it formed a neo-centromere.

## 5.5. Proportion Cnp1 binding on chromosome 2

I measured the total amount of Cnp1 binding to the centromeric DNA and normalised this binding by comparison to the total amount of Cnp1 binding to centromere 1 (*cen1*) and centromere 3 (*cen3*) while assuming that the amount of normalised binding to centromere 4 (*cen4*) was constant. The *cen1+cen3* mean number of reads in the Nott373control was 1819594.5. We used this number to get the normalisation factor for each sample with the equation:

$$(cen1 + cen3)/1819594.5$$

The results of this analysis (figure 5.11) show that in *S. pombe*, the amount of Cnp1 binding to the centromere is proportional to the length of the centromeric DNA, rather than being independent of the total amount of centromeric DNA as is apparently the case in vertebrates. In figure 5.11, the centromeric sequences of 4.2kb, 6.8kb, and 9.5kb are formed by test sequences only (strains Nott373_cen1, Nott373_cen5, and Nott373_cen8, respectively, after centromere 2 excision), the 8.2kb sequence is the natural centromere 2 (Nott373_control), and the 14.9kb and the 17kb centromeric sequences are the sums of native centromere and test sequences (Nott373_cen5 and Nott373_cen8 strains before *cen2* excision). Regression analysis shows a positive correlation between the centromeric DNA length and the proportion of Cnp1 binding ($R^2$=0.99). This result suggests that there is no regulatory mechanism for the Cnp1 binding on the native centromere of *S. pombe.*

The minimum centromeric DNA length observed to be able to bind Cnp1 and form a functional centromere is 4.2kb. Centromeric sequences of less than 4.2kb form

neo-centromeres in a rate of $10^{-3}$ or lower. The neo-centromere binds less Cnp1 per kilobase of DNA, maybe because this region is a weaker binder for the Cnp1, or the neo-centromere has a different pathway for the kinetochore formation.

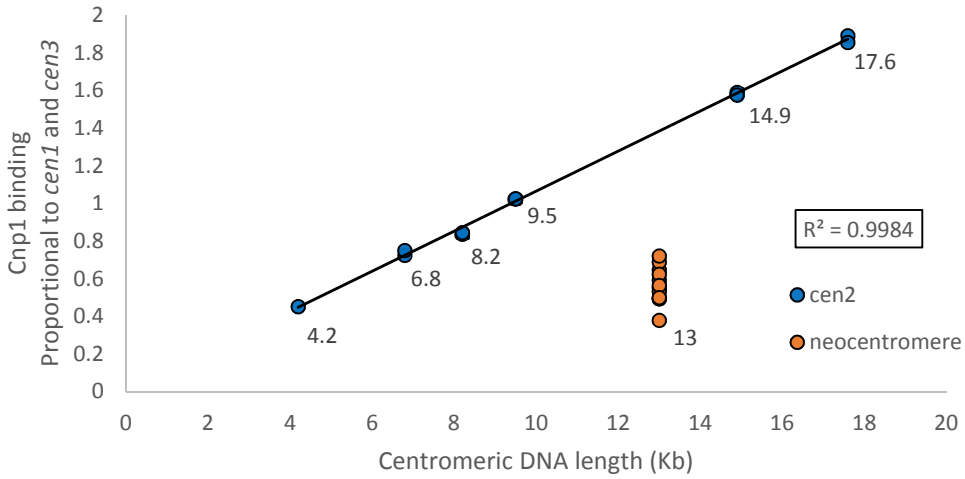



**Figure 5.11:Cnp1 binding on *cen2* and neo-centromere as a proportion of Cnp1 binding on *cen1+cen3*.**
The minimum centromeric DNA length able to bind Cnp1 is 4.2kb. There is a positive correlation between the centromeric DNA length and the proportion of Cnp1 ($R^2$=0.9984). Neo-centromeres are formed at a 13kb subtelomeric region and binds proportionally less Cnp1.

## 5.6. Cnp1 binding on centromeric sequences at non-centromeric location

Test sequences were also placed at the *ade6* locus on chromosome 3 of *S. pombe* CBS 2777 as described in Chapter 1. It allowed me to examine whether the presence of the centromeric DNA alone is sufficient for Cnp1 binding and centromere formation. ChIP-seq for the GFP-tagged Cnp1 was carried out for Nott376_cen8 (carrying 9.4kb mutant centromeric sequence at *ade6* locus) and Nott376_alphoid (5.7kb human alphoid DNA at *ade6* locus). The ChIP-seq analysis was carried out in duplicate. Centromere central core and flanking sequences (III: 1092658-1106291) were the major binding sites of this protein on chromosome 3 of both strains (figure 5.12) and there is no Cnp1 binding at the *ade6* locus in these strains, indicating that the presence of the centromeric sequence is not enough to recruit the Cnp1 binding. Ectopic Cnp1 binding was also observed at the telomeric regions corresponding to the rDNA. The reason why this might be so is as yet unclear.

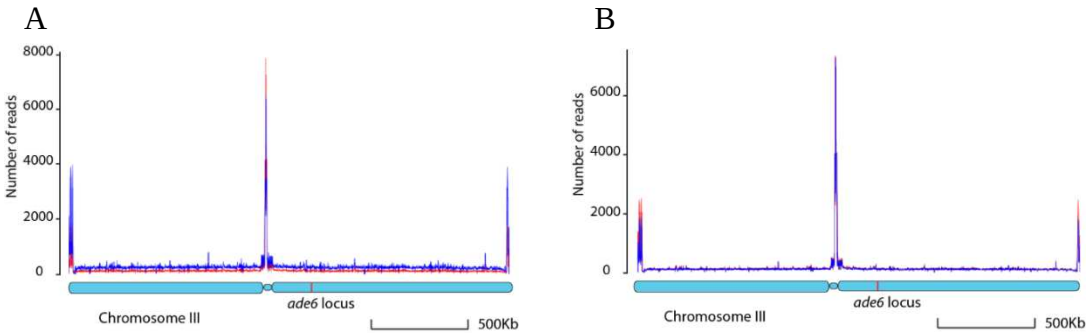



**Figure 5.12: Localisation of centromeric protein Cnp1 on chromosome 3 of CBS 2777 by ChIP-seq of GFP tagged Cnp1.**
The ChIP-seq experiment was duplicated, and the samples are color-coded. Major Cnp1 binding site is observed at the centromeric region. The sequenced strains show no Cnp1 binding on the centromeric DNA at the ade6 locus. (A) Cnp1 binding on chromosome 3 of Nott376_cen8, which carries 9.4kb test sequence on ade6 locus. (B) Cnp1 binding on chromosome 3 of Nott376_alphoid, which carries 5.7kb test sequence on ade6 locus.

## 5.7. *swi6* mutants

The gene that encodes the Swi6 protein was targeted with hygromycin-resistant gene (*hph*) on strains Nott373control, which carries no test sequence, and Nott373_cen5, harbouring the 6.8kb test sequence. The targeting construction was built by William Brown. The plasmid was linearized using the restriction enzyme *Hind*III, then lithium acetate transformation was carried out as described in the materials and methods section, using 50 ml of culture in a density of $2.5x10^7$ cell/ml and 250 μg of the linearized plasmid. The transformed cells were plated on YES agar supplemented with 100 μg/ml of hygromycin. The hygromycin-resistant clones were screened for targeting by PCR using primers 1047 and 646 (table 2.2 and figure 5.13), the predicted size of the PCR product was 738bp (figure 5.14).

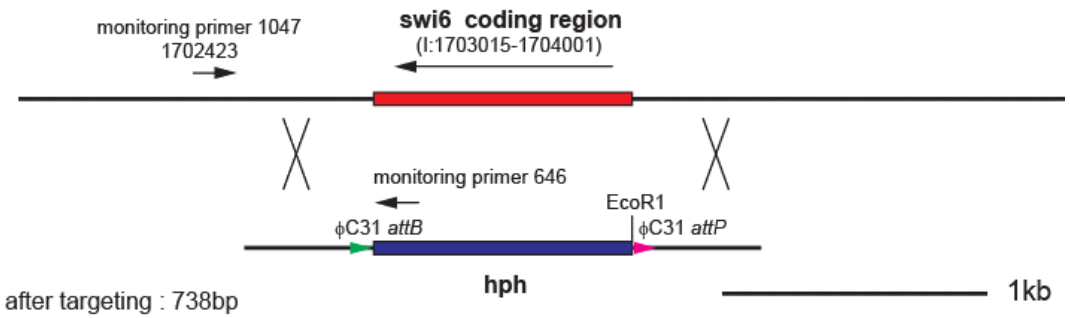



**Figure 5.13: Position of the primers 1047 and 646 on *S. pombe* chromosome I.**

The pair of primers was used for screening the clones which the *swi6* gene was targeted. After the gene targeting, the predicted fragment size amplified by the pair of primers is 738bp.

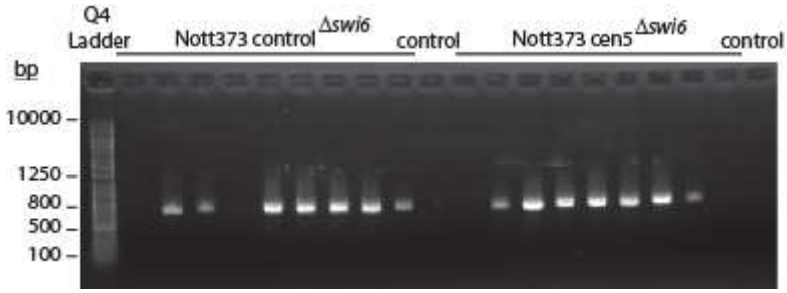



**Figure 5.14: Gel electrophoresis of the PCR using primers 646 and 1047 to screen *swi6* targeted clones.**
Not transformed strains Nott373_control and Nott373_cen5 are the controls.

The *swi6* targeted clones were analysed by restriction enzyme digestion and filter hybridization. The genomic DNA of the clones was digested with the restriction enzyme *EcoR*I which cuts at sites flanking the *swi6* locus. Then the DNA was size fractionated by a vertical agarose gel electrophoresis and transferred to nylon filters. The restriction enzyme map for the clones before and after the gene targeting are shown in figure 5.15. The filters were hybridised with a probe for a region upstream the *swi6* gene, the predicted size for the targeted clones is 5.8kb. Results of the hybridization are showed on figures 5.16A and 5.16 B for the strains Nott373_control and Nott373_cen5, respectively.

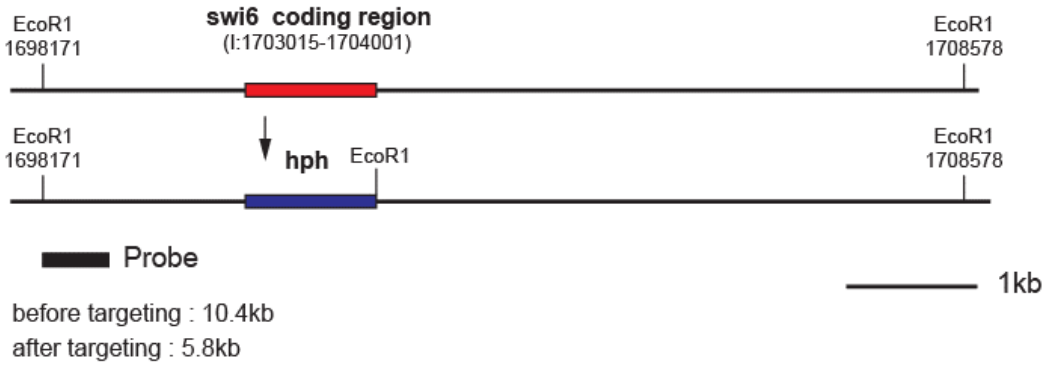



**Figure 5.15: Restriction enzyme site map of the clones before and after the *swi6* targeting.**
The *EcoR*I sites are 4.8kb upstream the targeted gene locus and 4.6kb downstream the *swi6* locus. The probe hybridises to a region upstream the *swi6* locus.

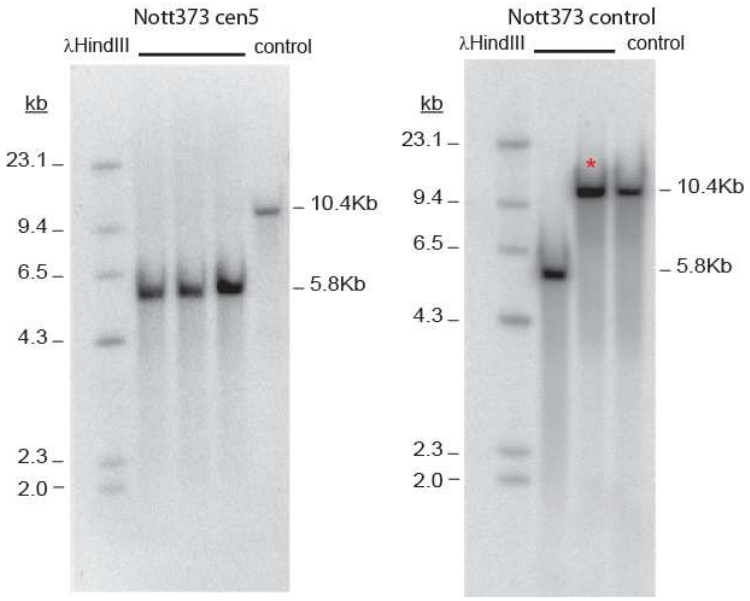



**Figure 5.16: Filter hybridization analysis of the fragments from the enzymatic digestion with *EcoR* I for the *swi6* targeted clones hybridised with a probe for the region upstream the gene locus.**
(A) *swi6* targeted clones derived from the Nott373_control, the not targeted strain is the control; (B) *swi6* targeted clones derived from the Nott373_cen5, the not targeted strain is the control. * clone not correctly targeted and was not used in the following experiments.

### 5.7.1. Native centromere 2 excision in Δ*swi6* strains

In order to promote the excision of the *cen2* in Nott373$^{\Delta swi6}$ strains, the Bxb1 integrase was expressed from an REP81 plasmid as described in chapter 4. REP81 empty vector was used as a control for the integrase expression plasmid. The transformations were performed as described in the material and methods section, using 100 ml of cultures at a density of $2x10^7$ cell/ml and 900 μg of the plasmid. After recovery, the cells were plated on PMG medium supplemented with 250mg/l of uracil, this medium selected for the clones containing the *LEU2* gene.

As observed for the strains with intact heterochromatin, following transformation with the REP81-BxB1, Nott373_control$^{\Delta swi6}$ gave smaller colonies than the REP81 empty vector controls while Nott373_cen5$^{\Delta swi6}$ gave colonies that were similar in size after transformation with either the empty REP81 vector or the REP81-Bxb1 integrase plasmid.

The colony size, the cells viability of the transformed strains were measured as described in chapter 4 (figure 4.3). The experiment was performed five times. Table 5.2 shows the average colony size and viability of Nott373_control$^{\Delta swi6}$ and Nott373_cen5$^{\Delta swi6}$ transformed both with the pREP81-Bxb1 and the pREP81 empty

vector and compares it with the same strains with intact heterochromatin. Neither growth retardation nor reduced cell viability was apparently observed in either Nott373_control$^{\Delta swi6}$ or Nott373_cen5$^{\Delta swi6}$ clones.

| Strain (test sequence) | Plasmid | Average colony size | Viability |
|---|---|---|---|
| **Nott373_control** | Rep81 | $5.6x10^6$ | 0.4 |
| **(0kb)** | Rep81Bxb1 | $1.7x10^5$ | 0.13 |
| **Nott373_control$^{\Delta swi6}$** | Rep81 | $4.3x10^6$ | 0.39 |
| **(0kb)** | Rep81Bxb1 | $9.3x10^5$ | 0.19 |
| **Nott373_cen5** | Rep81 | $4.1x10^6$ | 0.32 |
| **(6.8kb)** | Rep81Bxb1 | $4.3x10^6$ | 0.34 |
| **Nott373_cen5$^{\Delta swi6}$** | Rep81 | $4.6x10^6$ | 0.51 |
| **(6.8kb)** | Rep81Bxb1 | $3.7x10^6$ | 0.38 |

**Table 5.2: Average colony size and average recovered for each strain transformed both with the REP81-Bxb1 and the REP81 empty vector on Δ*swi6* strains.**

### 5.7.2. Screening the *ura4Δ* clones

The clones recovered on the YES plates used for the viability test were transferred for both YES and PMG supplemented with 150mg/l of Leucine, the *ura4*Δ clones do not grow on this medium. 100 clones of each strain were tested and the number of *ura4*Δ clones counted, the test was performed five times. No significant differences in the proportion of *ura4*Δ cells were observed between the strains containing the intact and disrupted *swi6* gene (table 5.3).

| Strain | *ura4*Δ |
|---|---|
| Nott373_control (0kb) | 0.03 |
| Nott373_control$^{\Delta swi6}$ (0kb) | 0.018 |
| Nott373_cen5 (6.8kb) | 0.986 |
| Nott373_cen5$^{\Delta swi6}$ (6.8kb) | 0.97 |

**Table 5.3: Proportion of *ura4Δ* clones for each strain after the Bxb1 induction.**

### 5.7.3. Measuring the proportion of *ura4*Δ clones that had centromere 2 excised

In order to screen *ura4*Δ clones that had excised the *cen2*, 40 independent clones of each Δ*swi6* strain were tested by PCR using LongAmp®. If *cen2* is excised, the primers 1001 and 1041 amplifies the Bxb1-*attL* in Not373_control$^{\Delta swi6}$ and through

the test sequence in Not373_cen5$^{\Delta swi6}$ (figure 5.17). The sequences of the primers are listed in table 2.1 in the material and methods section. The Δ*swi6* strains before Bxb1 induction and $dH_2O$ were used as negative controls.

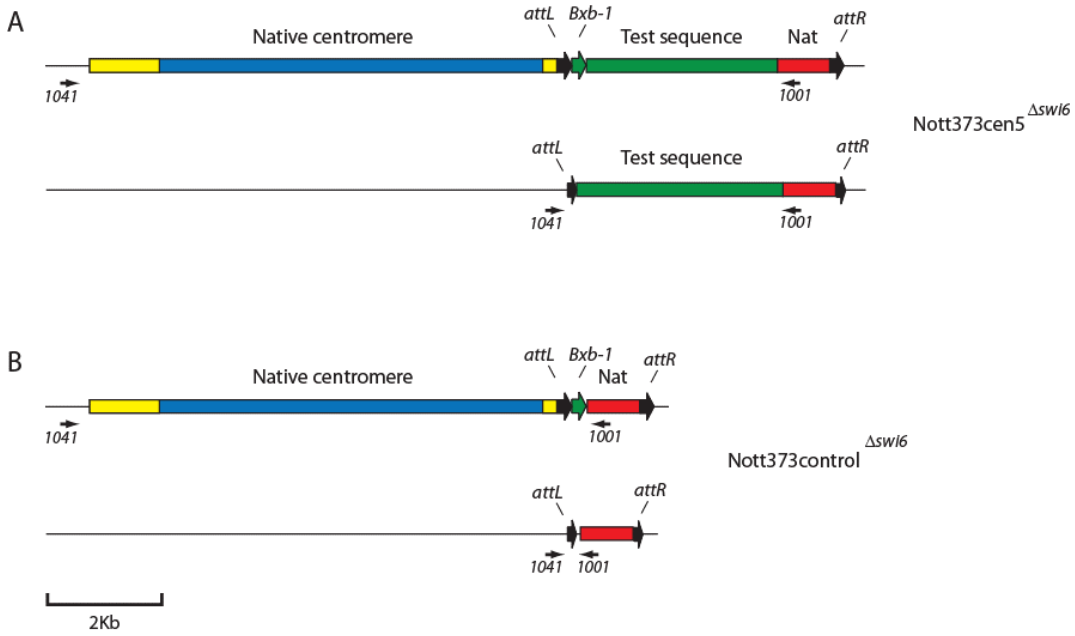



**Figure 5.17: Map showing the position of the primers 1001 and 1041 on chromosome 2 in the *S. pombe* recombined strain before and after the Bxb1 induction.**
(A) Nott373_cen5$^{\Delta swi6}$ (B) Nott373_control$^{\Delta swi6}$.

The predicted DNA fragment size for the *ura4*Δ Not37_3control$^{\Delta swi6}$ clones was 250bp when the *cen2* is excised. 28 of the 40 tested clones amplified a fragment of approximately 9kb, and the remaining 12 clones have not amplified any fragment. This result showed that *cen2* have not been excised in any of the tested clones (figure 5.18).

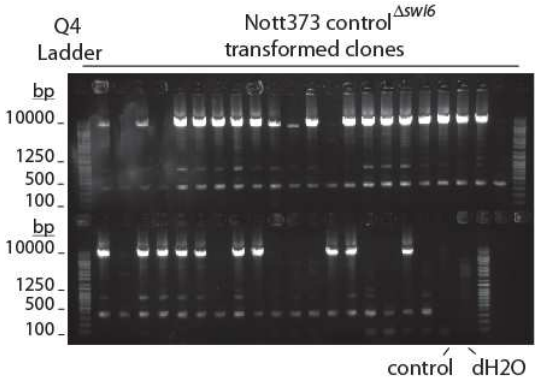



**Figure 5.18: Gel electrophoresis of the PCR across the Bxb1-*attL* site of the Δ*ura4* Nott373_control$^{\Delta swi6}$ clones.**
The expected size for the Δcen2 was 250bp.

The expected DNA fragment size for the *ura4*Δ Not373_cen5$^{\Delta swi6}$ clones was 6250bp when the *cen2* is excised. 39 of the 40 analysed clones amplified the expected fragment (figure 5.19), showing that *cen2* was excised and the centromere was formed over the 6.8kb test sequence as observed in chapter 4.

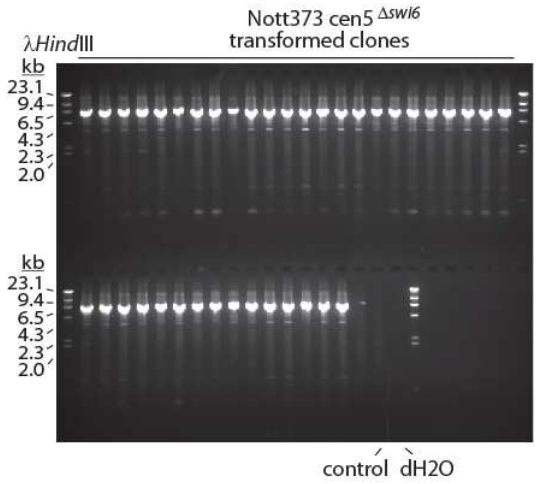



**Figure 5.19: Gel electrophoresis of the PCR across the Bxb1-*attL* site of the Δ*ura4* Nott373_cen5$^{\Delta swi6}$ clones.**
The expected size for the Δcen2 was 6250bp.

### 5.7.4. Karyotype analysis of *cen2*Δ survivors of the Not373_control$^{\Delta swi6}$ clones

In order to understand better what happened on *cen2* in the *ura4*Δ Not373_control$^{\Delta swi6}$ clones, William Brown analysed by pulsed field gel electrophoresis the chromosomes of 44 non-independent clones. These clones were isolated by plating on 5-FOA a pull of cells transformed with the Rep81-Bxb1 plasmid. The undigested chromosomes were subjected to pulsed-field gel electrophoresis (PFGE), followed by ethidium bromide (EtBr) staining and Southern blotting using a probe that hybridises with *cen2* and *cen4* because they have identical DNA sequences in *S. pombe* CBS 2777 (figure 5.20). The result shows that, in 37 clones, only three chromosomes rather than four were detected and the chromosome 2 was fused to chromosome 4. The remaining 7 clones have not excised the *cen2*, presenting 4 chromosomes on pulse field gel electrophoresis and the probe hybridised with both chromosomes 2 and 4. If Swi6 had no role in the neo-centromere formation we would expect to observe about 90% of the *ura4*Δ clones forming neo-centromere

as observed for the Not373_control with intact *swi6* (see figure 4.17A). The results suggest that Swi6 is required for the neo-centromere formation.

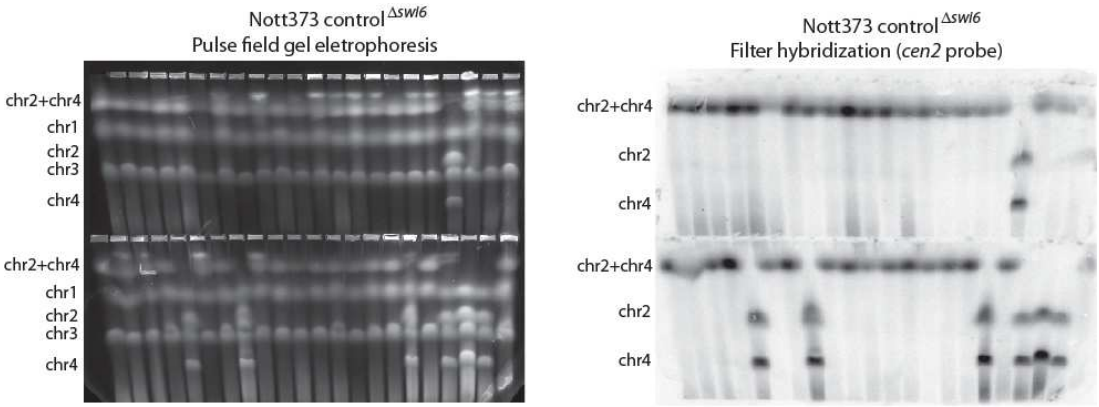



**Figure 5.20: Pulse Field Gel electrophoresis and filter hybridization of the genomic DNA of Not373_control$^{\Delta swi6}$ *ura4*Δ clones.**

## 5.8. Discussion

The results of this chapter disproved our initial hypothesis that Cnp1 binding was regulated by an epigenetic mechanism and, assuming that the different regions of the centromeric DNA have the same capacity of Cnp1 binding, we established that the amount of Cnp1 varies in proportion to the length of the underlying centromeric DNA sequence specific manner for lengths of centromeric DNA in excess of 4kb,. Sequences containing less than about 4kb of centromeric DNA and human Y alphoid DNA do not bind Cnp1 consistent with the failure of these sequence to seed centromere formation following deletion of the pre-existing centromere as described in chapter 4. I also addressed the question whether Cnp1 binding required a prior epigenetic mark and confirmed that it does so; thus when a 9.5kb tract of centromeric DNA was integrated at a non-centromeric locus; the *ade6* locus on chromosome 3 of CBS 2777 then no binding of Cnp1 was detectable by ChIP-seq.

The ChIP-seq analysis shows that strains carrying less than 4kb *S. pombe* centromeric DNA or the human alphoid DNA formed neo-centromeres. The subtelomeric region was the unique observed locus where neo-centromere formed after centromere 2 disruption on CBS 2777. Based on this result, it could be suggested that the subtelomeric heterochromatin region adjacent to the neo-centromeric loci is required for *de novo* kinetochore formation as indicated in previous works (Cleveland

et al., 2003, Henikoff et al., 2001, Folco et al., 2008, Durand-Dubief and Ekwall, 2008). To assess this possibility, I analysed whether a heterochromatin-defective strain is able to form neo-centromere and to maintain the centromere over the test sequence after the native centromere 2 disruption. Results show that neo-centromere formation is affected by the deletion of *swi6* but the centromere formation over the test sequence is maintained.

The experimental analysis described in this chapter led to several important conclusions about the relationship between Cnp1 binding on the one hand and centromere size, sequence and location on the other.

First of all, we observe that in *S. pombe* Cnp1 binding to a candidate centromeric sequence only occurs when the sequence is integrated adjacent to a pre-existing centromere. This is consistent with the results of work in several eukaryotes (Earnshaw and Migeon, 1985, Steiner and Clarke, 1994, Higgins et al., 2005, duSart et al., 1997) analysed to date which has shown that the centromere is specified epigenetically and that specific DNA sequences are not sufficient for centromere function. An example of the type of observation that revealed the necessity of an epigenetic mark for centromere formation came from work on human samples containing dicentric chromosomes in which one centromere was functionally inactivated without changes to its underlying DNA sequence (Earnshaw and Migeon, 1985).

Secondly, we observe that Cnp1 binding is proportional to the amount of centromeric DNA. Considering that there is a natural polimosphism between the centromeric DNA of the laboratory strain and the CBS 2777, we assume that different regions of the centromeric DNA bind Cnp1 with the same efficiency. Thus, the length of the centromeric DNA is what controls the amount of the Cnp1 binding. Moreover, Cnp1 spreads over the *S. pombe* centromeric sequences when placed adjacent to the native centromere. Together, these results suggest the existence of a seeding mechanism, where the presence of Cnp1 nucleosomes and the availability of centromeric DNA promote the binding of more Cnp1 units on the centromere.

Thirdly, we observe specificity in the binding of Cnp1 to a sequence when it is placed adjacent to a native centromere. Human Y alphoid DNA, a centromere sequence, do not bind Cnp1 when placed adjacent to the native centromere. It could be suggested to express the human Cnp1 homolog CENP-A to see binding over the alphoid DNA. But what we aimed to test in this experiment was the sequence

specificity, so if we have expressed CENP-A and have seen binding we would not be able to draw conclusions about sequence specificity, or, if we have not observed CENP-A binding we could not make any conclusion because the human protein requires chaperones and other signals that are not found in S. pombe.

We also have determined the approximate size of the minimum amount of centromeric DNA necessary to function as a centromere after excision of the native centromere. Assuming that different regions of the centromeric DNA does not favour or damage the capacity of Cnp1 binding, we observed that test sequences of 4kb or more in length are able to bind Cnp1 and form functional centromere with high efficiency, as demonstrated on Chapter 4. In contrast, strains containing test sequences of less than 4kb do not bind Cnp1 after centromere 2 excision. In the experiments in which sequences of this size were present as candidate sequences the number of survived cells decreased from 74% of the control in the Nott373_cen1 (4.2kb test sequence) to 47% of the control in the Nott373_cen2 (3.6kb test sequence) or less in the strains carrying smaller fragments of *S. pombe* centromeric DNA, and the surviving cells were found to have formed neo-centromes on the centromere distal subtelomeric DNA. These observations suggest that the minimum size of centromeric DNA able to form a functional centromere is about 4kb in *S. pombe*. The neo-centromere formation that we have detected confirms the results of others that specific centromeric DNA sequences are not necessary for kinetochore assembly in *S. pombe*.

Finally, *swi6* has been deleted from Nott373_control (0kb test sequence) and Nott373_cen5 (6.8kb test sequence) strains, and it was analysed the consequences of native centromere 2 disruption in these strains. In the Nott373_control$^{\Delta swi6}$, neo-centromere formation was undetectable and, all the survivors presented fused chromosomes or the centromere 2 was not excised. This result was consistent with the work of Ishii et al. (2008) who demonstrated that neighboring heterochromatin is essential for *de novo* kinetochore formation at the non-centromeric locus, but not for the maintenance of the centromere, in agreement with previous works (Folco et al., 2008, Durand-Dubief and Ekwall, 2008).

It remains unclear, however, whether the neo-centromeric DNA shares features important for centromere function with the native centromeric DNA or whether biding at the two sequences occurs as a result of different mechanisms. Our system will allow this question to be addressed.

# 6. CONCLUSIONS

The work described in this thesis was aimed at developing a system that enables the relationship between the sequence and length of a segment of DNA and its ability to support kinetochore formation to be defined in the model organism *Schizosaccharomyces pombe*. The methodology that I have developed uses one unidirectional serine recombinases to integrate a candidate sequence and a second to delete the pre-existing centromere. Thus the approach that I have developed can be described as "centromere swapping".

In summary, the main conclusions of our work are:

- ϕC31 integrase catalyse efficient and accurate site-specific integration of DNA of up to 13kb into *S. pombe* genome. This observation is consistent with previous studies showing that the integrase work in vertebrate cells (Dafhnis-Calas et al., 2005, Malla et al., 2005, Belteki et al., 2003).
- It was observed that, when nourseothricin resistance gene is oriented so that transcription is away from the integration site position, it induces the inversion of the marker gene by re-arrangement, suggesting that placing a promoter adjacent to the kinetochore is selected against.
- The Bxb1 integrase deletes a 9kb stretch of native centromeric DNA at a rate of about once every four or five cell cycles in *S. pombe*.
- Centromere formation depends on the nature and size of the candidate sequence: candidate sequences equal or greater than 6.8kb in size forms centromere over the candidate sequences with nearly absolute efficiency; the size threshold that supports centromere formation is about 4kb, bellow that, the centromere excision and centromere formation is only detectable at frequencies of $5x10^{-3}$ or less.
- The ability of a sequence to support centromere in *S. pombe* appears to be sequence and species specific as human Y alphoid DNA shows no activity in our assay. It would be of interest however to assay other members of the alphoid DNA family and the centromeric DNA from other organisms such as mouse and *Arabidopsis thaliana.*

- ChIP-seq experiments have shown that the amount of the CENP-A binding to the centromere is proportional to the amount of the underlying centromeric DNA.
- The centromere spreads over the *S. pombe* centromeric sequences when placed adjacent to the native centromere suggesting a seeding mechanism, where the presence of Cnp1 nucleosomes and the availability of centromeric DNA promote the binding of more Cnp1 units at the centromere.
- In *S. pombe*, there is no evidence of the existence of a mechanism which controls the amount of Cnp1 binding to the centromere as seems to hold in human (Irvine et al., 2004).
- The heterochromatic protein Swi6 is essential for *de novo* kinetochore formation at the non-centromeric locus, but not for the maintenance of the centromere, in agreement with previous works (Ishii et al., 2008, Folco et al., 2008, Durand-Dubief and Ekwall, 2008).
- The system has proven to be efficient, this should allow one to assay the ability of different sequences to function as centromeres.

The major contribution of my work to the centromere research field is methodologic. We have established a new system that allow the engineering of centromeres. This system will enable the relationship between size, sequence and chromosome position to be defined for the first time in any eukaryote.

The present work alters the view of the centromere specification once it demonstrates that, in *S. pombe*, the amount Cnp1 bound is proportional to the length of the centromeric DNA. Quantitative fluorescence microscopy study suggests that, in human, the amount of nucleosomes containing centromere-specific histone H3 variant do not correspond to the length of the centromeric DNA (Irvine et al., 2004, Okamoto et al., 2007). The methodology used in my thesis, based on ChIP-seq data, makes a more direct measurement of the Cnp1 bound to the DNA.

As might be expected from the functional conservation among eukaryotes, centromere and kinetochore proteins generally share substantial homology even between evolutionarily distant organisms (Cheeseman and Desai, 2008). In contrast, however, centromeric DNA sequences are highly variable in size and complexity (Marshall et al., 2008, Eichler, 1999). Despite the lack of DNA sequence conservation

among species, the presence of repetitive elements may be a characteristic feature of the called ‘regional centromeres’ in nature (Marshall et al., 2008, Henikoff et al., 2001), so we can expect that the findings of this work are extensible to eukaryotes carrying these regional centromeres.

Our current view of how centromeres are specified is based on studies in many diverse eukaryotic species. It can be difficult to form a model using every piece of evidence, yet all the results discussed here are particularly helpful in examining aspects of centromere specification. Key features of future findings will need to be viewed in the context of our current knowledge of centromere specification in various eukaryotic species in order to further elucidate the complex mechanisms of this process.

## 6.1. Further work

Further studies are necessary to understand the centromeric DNA features that are important for native centromere and neo-centromere formation. These features can be common for native centromere and neo-centromere or the two kind of centromeres could use different mechanisms. The system developed in our work will allow these question to be addressed.

Therefore, the next step will consist of exploiting the system generated in this work to assay the centromeric activity of different sequences. In order to achieve that, it will be necessary to assemble a whole set of plasmids with different types and lengths of candidate DNA sequences. Then, these sequences will be integrated either adjacent to the native centromere or at the subtelomeric region by site specific recombination mediated by the ϕC31 integrase. The Cnp1 binding will be analysed by ChIP-seq once the old centromere has been excised by the Bxb1 integrase. It would also be of interest to apply this approach to other unicellular eukaryotes in order to understand the sequence determinants of kinetochore formation in these systems and to compare then with *S. pombe.*

# 7. REFERENCES


AHMAD, K. & HENIKOFF, S. 2002. Histone H3 variants specify modes of chromatin assembly. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 16477-16484.

AKIYOSHI, B. & BIGGINS, S. 2010. Cdc14-Dependent Dephosphorylation of a Kinetochore Protein Prior to Anaphase in Saccharomyces cerevisiae. *Genetics,* 186**,** 1487-U631.

ALLSHIRE, R. C. 1995. ELEMENTS OF CHROMOSOME STRUCTURE AND FUNCTION IN FISSION YEAST. *Seminars in Cell Biology,* 6**,** 55-64.

ALLSHIRE, R. C. & EKWALL, K. 2015. Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. *Cold Spring Harbor Perspectives in Biology,* 7.

ALLSHIRE, R. C., JAVERZAT, J. P., REDHEAD, N. J. & CRANSTON, G. 1994. POSITION EFFECT VARIEGATION AT FISSION YEAST CENTROMERES. *Cell,* 76**,** 157-169.

AMANO, M., SUZUKI, A., HORI, T., BACKER, C., OKAWA, K., CHEESEMAN, I. M. & FUKAGAWA, T. 2009. The CENP-S complex is essential for the stable assembly of outer kinetochore structure. *Journal of Cell Biology,* 186**,** 173-182.

AMARO, A. C., SAMORA, C. P., HOLTACKERS, R., WANG, E., KINGSTON, I. J., ALONSO, M., LAMPSON, M., MCAINSH, A. D. & MERALDI, P. 2010. Molecular control of kinetochore-microtubule dynamics and chromosome oscillations. *Nature Cell Biology,* 12**,** 319-U42.

AMOR, D. J., BENTLEY, K., RYAN, J., PERRY, J., WONG, L., SLATER, H. & CHOO, K. H. A. 2004. Human centromere repositioning "in progress". *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 6542-6547.

ARAVIND, L., WATANABE, H., LIPMAN, D. J. & KOONIN, E. V. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 11319-11324.

BAHLER, J., WU, J. Q., LONGTINE, M. S., SHAH, N. G., MCKENZIE, A., STEEVER, A. B., WACH, A., PHILIPPSEN, P. & PRINGLE, J. R. 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast,* 14**,** 943-951.

BANCROFT, J., AUCKLAND, P., SAMORA, C. P. & MCAINSH, A. D. 2015. Chromosome congression is promoted by CENP-Q- and CENP-E-dependent pathways. *Journal of Cell Science,* 128**,** 171-184.

BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Research,* 21**,** 381-395.

BANNISTER, A. J., ZEGERMAN, P., PARTRIDGE, J. F., MISKA, E. A., THOMAS, J. O., ALLSHIRE, R. C. & KOUZARIDES, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature,* 410**,** 120-124.

BAUM, M. & CLARKE, L. 2000. Fission yeast homologs of human CENP-B have redundant functions affecting cell growth and chromosome segregation. *Molecular and Cellular Biology,* 20**,** 2852-2864.

BAUM, M., SANYAL, K., MISHRA, P. K., THALER, N. & CARBON, J. 2006. Formation of functional centromeric chromatin is specified epigenetically in Candida albicans. *Proc Natl Acad Sci U S A,* 103**,** 14877-82.

BELTEKI, G., GERTSENSTEIN, M., OW, D. W. & NAGY, A. 2003. Site-specific cassette exchange and germline transmission with mouse ES cells expressing phi C31 integrase. *Nature Biotechnology,* 21**,** 321-324.

BERNAD, R., SANCHEZ, P. & LOSADA, A. 2009. Epigenetic specification of centromeres by CENP-A. *Exp Cell Res,* 315**,** 3233-41.

BIRD, A. 2007. Perceptions of epigenetics. *Nature,* 447**,** 396-8.

BLACK, B. E., JANSEN, L. E., MADDOX, P. S., FOLTZ, D. R., DESAI, A. B., SHAH, J. V. & CLEVELAND, D. W. 2007a. Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol Cell,* 25**,** 309-22.

BLACK, B. E., JANSEN, L. E. T., MADDOX, P. S., FOLTZ, D. R., DESAI, A. B., SHAH, J. V. & CLEVELAND, D. W. 2007b. Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Molecular Cell,* 25**,** 309-322.

BLANCO, M. A., SANCHEZ-DIAZ, A., DE PRADA, J. M. & MORENO, S. 2000. APC(ste9/srw1) promotes degradation of mitotic cyclins in G(1) and is inhibited by cdc2 phosphorylation. *Embo Journal,* 19**,** 3945-3955.

BLOWER, M. D. & KARPEN, G. H. 2001. The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nature Cell Biol.,* 3**,** 730-739.

BOCK, L. J., PAGLIUCA, C., KOBAYASHI, N., GROVE, R. A., OKU, Y., SHRESTHA, K., ALFIERI, C., GOLFIERI, C., OLDANI, A., DAL MASCHIO, M., BERMEJO, R., HAZBUN, T. R., TANAKA, T. U. & DE WULF, P. 2012. Cnn1 inhibits the interactions between the KMN complexes of the yeast kinetochore. *Nature Cell Biology,* 14**,** 614-+.

BODOR, D. 2015. *Quantitative Centromere Epigenetics.* PhD, Universidade Nova de Lisboa.

BODOR, D. L., MATA, J. F., SERGEEV, M., DAVID, A. F., SALIMIAN, K. J., PANCHENKO, T., CLEVELAND, D. W., BLACK, B., SHAH, J. V. & JANSEN, L. E. T. 2014. The quantitative architecture of centromeric chromatin. *Elife,* 3.

BODOR, D. L., VALENTE, L. P., MATA, J. F., BLACK, B. E. & JANSEN, L. E. T. 2013. Assembly in G1 phase and long-term stability are unique intrinsic features of CENP-A nucleosomes. *Molecular Biology of the Cell,* 24**,** 923-932.

BORSANI, G., TONLORENZI, R., SIMMLER, M. C., DANDOLO, L., ARNAUD, D., CAPRA, V., GROMPE, M., PIZZUTI, A., MUZNY, D., LAWRENCE, C., WILLARD, H. F., AVNER, P. & BALLABIO, A. 1991. CHARACTERIZATION OF A MURINE GENE EXPRESSED FROM THE INACTIVE X-CHROMOSOME. *Nature,* 351**,** 325-329.

BROCKDORFF, N., ASHWORTH, A., KAY, G. F., MCCABE, V. M., NORRIS, D. P., COOPER, P. J., SWIFT, S. & RASTAN, S. 1992. THE PRODUCT OF THE MOUSE XIST GENE IS A 15 KB INACTIVE X-SPECIFIC TRANSCRIPT CONTAINING NO CONSERVED ORF AND LOCATED IN THE NUCLEUS. *Cell,* 71**,** 515-526.

BROWN, C. J., BALLABIO, A., RUPERT, J. L., LAFRENIERE, R. G., GROMPE, M., TONLORENZI, R. & WILLARD, H. F. 1991. A GENE FROM THE REGION OF THE HUMAN X-INACTIVATION CENTER IS EXPRESSED

EXCLUSIVELY FROM THE INACTIVE X-CHROMOSOME. *Nature,* 349**,** 38-44.

BROWN, C. J., HENDRICH, B. D., RUPERT, J. L., LAFRENIERE, R. G., XING, Y., LAWRENCE, J. & WILLARD, H. F. 1992. THE HUMAN XIST GENE - ANALYSIS OF A 17 KB INACTIVE X-SPECIFIC RNA THAT CONTAINS CONSERVED REPEATS AND IS HIGHLY LOCALIZED WITHIN THE NUCLEUS. *Cell,* 71**,** 527-542.

BROWN, W. R. A., LEE, N. C. O., XU, Z. Y. & SMITH, M. C. M. 2011a. Serine recombinases as tools for genome engineering. *Methods,* 53**,** 372-379.

BROWN, W. R. A., LITI, G., ROSA, C., JAMES, S., ROBERTS, I., ROBERT, V., JOLLY, N., TANG, W., BAUMANN, P., GREEN, C., SCHLEGEL, K., YOUNG, J., HIRCHAUD, F., LEEK, S., THOMAS, G., BLOMBERG, A. & WARRINGER, J. 2011b. A Geographically Diverse Collection of Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. *G3-Genes Genomes Genetics,* 1**,** 615-626.

BROWN, W. R. A., THOMAS, G., LEE, N. C. O., BLYTHE, M., LITI, G., WARRINGER, J. & LOOSE, M. W. 2014. Kinetochore assembly and heterochromatin formation occur autonomously in Schizosaccharomyces pombe. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 1903-1908.

BROWN, W. R. A. & XU, Z.-Y. 2009. The 'kinetochore maintenance loop'—The mark of regulation? *BioEssays,* 31**,** 228-236.

BUCHWITZ, B. J., AHMAD, K., MOORE, L. L., ROTH, M. B. & HENIKOFF, S. 1999. A histone-H3-like protein in C. elegans. *Nature,* 401**,** 547-548.

BUFFALO, V. 2011. Scythe.

BUSCAINO, A., ALLSHIRE, R. & PIDOUX, A. 2010. Building centromeres: home sweet home or a nomadic existence? *Current Opinion in Genetics & Development,* 20**,** 118-126.

CAMAHORT, R., SHIVARAJU, M., MATTINGLY, M., LI, B., NAKANISHI, S., ZHU, D. X., SHILATIFARD, A., WORKMAN, J. L. & GERTON, J. L. 2009. Cse4 Is Part of an Octameric Nucleosome in Budding Yeast. *Molecular Cell,* 35**,** 794-805.

CAPOZZI, O., PURGATO, S., DI CANTOGNO, L. V., GROSSO, E., CICCONE, R., ZUFFARDI, O., DELLA VALLE, G. & ROCCHI, M. 2008. Evolutionary and clinical neocentromeres: two faces of the same coin? *Chromosoma,* 117**,** 339-344.

CARBONE, L., NERGADZE, S. G., MAGNANI, E., MISCEO, D., CARDONE, M. F., ROBERTO, R., BERTONI, L., ATTOLINI, C., PIRAS, M. F., DE JONG, P., RAUDSEPP, T., CHOWDHARY, B. P., GUERIN, G., ARCHIDIACONO, N., ROCCHI, M. & GIULOTTO, E. 2006. Evolutionary movement of centromeres in horse, donkey, and zebra. *Genomics,* 87**,** 777-782.

CARONE, B. R., FAUQUIER, L., HABIB, N., SHEA, J. M., HART, C. E., LI, R. W., BOCK, C., LI, C. J., GU, H. C., ZAMORE, P. D., MEISSNER, A., WENG, Z. P., HOFMANN, H. A., FRIEDMAN, N. & RANDO, O. J. 2010. Paternally Induced Transgenerational Environmental Reprogramming of Metabolic Gene Expression in Mammals. *Cell,* 143**,** 1084-1096.

CASTEL, S. E. & MARTIENSSEN, R. A. 2013. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet,* 14**,** 100-12.

CATANIA, S., PIDOUX, A. L. & ALLSHIRE, R. C. 2015. Sequence Features and Transcriptional Stalling within Centromere DNA Promote Establishment of CENP-A Chromatin. *Plos Genetics,* 11.

CHAN, F. L. & WONG, L. H. 2012. Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Research,* 40**,** 11178-11188.

CHANDLER, V. L. 2007. Paramutation: From maize to mice. *Cell,* 128**,** 641-645.

CHATTERJEE, G., SANKARANARAYANAN, S. R., GUIN, K., THATTIKOTA, Y., PADMANABHAN, S., SIDDHARTHAN, R. & SANYAL, K. 2016. Repeat-Associated Fission Yeast-Like Regional Centromeres in the Ascomycetous Budding Yeast Candida tropicalis. *Plos Genetics,* 12.

CHEESEMAN, I. M. 2002. Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell,* 111**,** 163-172.

CHEESEMAN, I. M. 2004. A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes Dev.,* 18**,** 2255-2268.

CHEESEMAN, I. M., CHAPPIE, J. S., WILSON-KUBALEK, E. M. & DESAI, A. 2006. The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. *Cell,* 127**,** 983-997.

CHEESEMAN, I. M. & DESAI, A. 2008. Molecular architecture of the kinetochore-microtubule interface. *Nature Reviews Molecular Cell Biology,* 9**,** 33-46.

CHEESEMAN, I. M., HORI, T., FUKAGAWA, T. & DESAI, A. 2008. KNL1 and the CENP-H/I/K complex coordinately direct kinetochore assembly in vertebrates. *Molecular Biology of the Cell,* 19**,** 587-594.

CHEN, E. S., SAITOH, S., YANAGIDA, M. & TAKAHASHI, K. 2003. A cell cycle-regulated GATA factor promotes centromeric localization of CENP-A in fission yeast. *Molecular Cell,* 11**,** 175-187.

CIFERRI, C., PASQUALATO, S., SCREPANTI, E., VARETTI, G., SANTAGUIDA, S., DOS REIS, G., MAIOLICA, A., POLKA, J., DE LUCA, J. G., DE WULF, P., SALEK, M., RAPPSILBER, J., MOORES, C. A., SALMON, E. D. & MUSACCHIO, A. 2008. Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell,* 133**,** 427-439.

CLARKE, L. & CARBON, J. 1983. GENOMIC SUBSTITUTIONS OF CENTROMERES IN SACCHAROMYCES-CEREVISIAE. *Nature,* 305**,** 23-28.

CLEMSON, C. M., MCNEIL, J. A., WILLARD, H. F. & LAWRENCE, J. B. 1996. XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear chromosome structure. *Journal of Cell Biology,* 132**,** 259-275.

CLEVELAND, D. W., MAO, Y. H. & SULLIVAN, K. F. 2003. Centromeres and kinetochores: From epigenetics to mitotic checkpoint signaling. *Cell,* 112**,** 407-421.

COFFMAN, V. C., WU, P. C., PARTHUN, M. R. & WU, J. Q. 2011. CENP-A exceeds microtubule attachment sites in centromere clusters of both budding and fission yeast. *Journal of Cell Biology,* 195**,** 563-572.

COLEMAN, T. R., TANG, Z. H. & DUNPHY, W. G. 1993. NEGATIVE REGULATION OF THE WEE1 PROTEIN-KINASE BY DIRECT ACTION OF THE NIM1/CDR1 MITOTIC INDUCER. *Cell,* 72**,** 919-929.

COUGHLAN, A. Y., HANSON, S. J., BYRNE, K. P. & WOLFE, K. H. 2016. Centromeres of the Yeast Komagataella phaffii (Pichia pastoris) Have a Simple Inverted-Repeat Structure. *Genome Biology and Evolution,* 8**,** 2482-2492.

CSANKOVSZKI, G., NAGY, A. & JAENISCH, R. 2001. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *Journal of Cell Biology,* 153**,** 773-783.

DAFHNIS-CALAS, F., XU, Z. Y., HAINES, S., MALLA, S. K., SMITH, M. C. M. & BROWN, W. R. A. 2005. Iterative in vivo assembly of large and complex transgenes by combining the activities of phi C31 integrase and Cre recombinase. *Nucleic Acids Research,* 33.

DE VANSSAY, A., BOUGE, A. L., BOIVIN, A., HERMANT, C., TEYSSET, L., DELMARRE, V., ANTONIEWSKI, C. & RONSSERAY, S. 2012. Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *Nature,* 490**,** 112-U132.

DE WULF, P., MCAINSH, A. D. & SORGER, P. K. 2003. Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes. *Genes Dev.,* 17**,** 2902-2921.

DELUCA, J. G. 2006. Kinetochore microtubule dynamics and attachment stability are regulated by Hec1. *Cell,* 127**,** 969-982.

DIXON, J. E., SHAIKH, A. C. & SADOWSKI, P. D. 1995. THE FLP RECOMBINASE CLEAVES HOLLIDAY JUNCTIONS IN TRANS. *Molecular Microbiology,* 18**,** 449-458.

DRINNENBERG, I. A., DEYOUNG, D., HENIKOFF, S. & MALIK, H. S. 2014. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife,* 3.

DUNLEAVY, E. M., BEIER, N. L., GORGESCU, W., TANG, J., COSTES, S. V. & KARPEN, G. H. 2012. The Cell Cycle Timing of Centromeric Chromatin Assembly in Drosophila Meiosis Is Distinct from Mitosis Yet Requires CAL1 and CENP-C. *Plos Biology,* 10.

DUNLEAVY, E. M., ROCHE, D., TAGAMI, H., LACOSTE, N., RAY-GALLET, D., NAKAMURA, Y., DAIGO, Y., NAKATANI, Y. & ALMOUZNI-PETTINOTTI, G. 2009. HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell,* 137**,** 485-497.

DURAND-DUBIEF, M. & EKWALL, K. 2008. Heterochromatin tells CENP-A where to go. *Bioessays,* 30**,** 526-9.

DUSART, D., CANCILLA, M. R., EARLE, E., MAO, J. I., SAFFERY, R., TAINTON, K. M., KALITSIS, P., MARTYN, J., BARRY, A. E. & CHOO, K. H. A. 1997. A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nature Genetics,* 16**,** 144-153.

EARNSHAW, W. C. & MIGEON, B. R. 1985. 3 RELATED CENTROMERE PROTEINS ARE ABSENT FROM THE INACTIVE CENTROMERE OF A STABLE ISODICENTRIC CHROMOSOME. *Chromosoma,* 92**,** 290-296.

EARNSHAW, W. C., RATRIE, H. & STETTEN, G. 1989. VISUALIZATION OF CENTROMERE PROTEINS CENP-B AND CENP-C ON A STABLE DICENTRIC CHROMOSOME IN CYTOLOGICAL SPREADS. *Chromosoma,* 98**,** 1-12.

EARNSHAW, W. C. & ROTHFIELD, N. 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma,* 91**,** 313-321.

EARNSHAW, W. S., SULLIVAN, K. F., SAUNDERS, W. S., CLEVELAND, D. W. & ROTHFIELD, N. F. 1985. IDENTIFICATION AND MOLECULAR-CLONING OF CENP-B, A HUMAN CENTROMERE PROTEIN

RECOGNIZED BY ALL HUMAN ANTICENTROMERE ANTIBODIES (ACA). *Clinical Research,* 33**,** A591-A591.
EICHLER, E. E. 1999. Repetitive conundrums of centromere structure and function. *Human Molecular Genetics,* 8**,** 151-155.
EKWALL, K., JAVERZAT, J. P., LORENTZ, A., SCHMIDT, H., CRANSTON, G. & ALLSHIRE, R. 1995. THE CHROMODOMAIN PROTEIN SWI6 - A KEY COMPONENT AT FISSION YEAST CENTROMERES. *Science,* 269**,** 1429-1431.
ESKAT, A., DENG, W., HOFMEISTER, A., RUDOLPHI, S., EMMERTH, S., HELLWIG, D., ULBRICHT, T., DORING, V., BANCROFT, J. M., MCAINSH, A. D., CARDOSO, M. C., MERALDI, P., HOISCHEN, C., LEONHARDT, H. & DIEKMANN, S. 2012. Step-Wise Assembly, Maturation and Dynamic Behavior of the Human CENP-P/O/R/Q/U Kinetochore SubComplex. *Plos One,* 7.
FACHINETTI, D., HAN, J. S., MCMAHON, M. A., LY, P., ABDULLAH, A., WONG, A. J. & CLEVELAND, D. W. 2015. DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. *Developmental Cell,* 33**,** 314-327.
FEINBERG, A. P. & VOGELSTEIN, B. 1983. A TECHNIQUE FOR RADIOLABELING DNA RESTRICTION ENDONUCLEASE FRAGMENTS TO HIGH SPECIFIC ACTIVITY. *Analytical Biochemistry,* 132**,** 6-13.
FISHER, D. & NURSE, P. 1995. CYCLINS OF THE FISSION YEAST SCHIZOSACCHAROMYCES-POMBE. *Seminars in Cell Biology,* 6**,** 73-78.
FITZGERALDHAYES, M., CLARKE, L. & CARBON, J. 1982. NUCLEOTIDE-SEQUENCE COMPARISONS AND FUNCTIONAL-ANALYSIS OF YEAST CENTROMERE DNAS. *Cell,* 29**,** 235-244.
FLEIG, U., BEINHAUER, J. D. & HEGEMANN, J. H. 1995. FUNCTIONAL SELECTION FOR THE CENTROMERE DNA FROM YEAST CHROMOSOME-VIII. *Nucleic Acids Research,* 23**,** 922-924.
FOLCO, H. D., PIDOUX, A. L., URANO, T. & ALLSHIRE, R. C. 2008. Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. *Science,* 319**,** 94-7.
FOLTZ, D. R. 2006. The human CENP-A centromeric nucleosome-associated complex. *Nature Cell Biol.,* 8**,** 458-469.
FOLTZ, D. R., JANSEN, L. E. T., BAILEY, A. O., YATES, J. R., BASSETT, E. A., WOOD, S., BLACK, B. E. & CLEVELAND, D. W. 2009. Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell,* 137**,** 472-484.
FOLTZ, D. R., JANSEN, L. E. T., BLACK, B. E., BAILEY, A. O., YATES, J. R. & CLEVELAND, D. W. 2006. The human CENP-A centromeric nucleosome-associated complex. *Nature Cell Biology,* 8**,** 458-U77.
FORSBURG, S. L. 2015. *PombeNet* [Online]. Web Page. Available: http://www-bcf.usc.edu/~forsburg/ [Accessed 16th of May 2017 2017].
FORSBURG, S. L. & NURSE, P. 1991a. CELL-CYCLE REGULATION IN THE YEASTS SACCHAROMYCES-CEREVISIAE AND SCHIZOSACCHAROMYCES-POMBE. *Annual Review of Cell Biology,* 7**,** 227-256.

FORSBURG, S. L. & NURSE, P. 1991b. IDENTIFICATION OF A G1-TYPE CYCLIN PUC1+ IN THE FISSION YEAST SCHIZOSACCHAROMYCES-POMBE. *Nature,* 351**,** 245-248.

FRENCH, B. T., WESTHORPE, F. G., LIMOUSE, C. & STRAIGHT, A. F. 2017. Xenopus laevis M18BP1 Directly Binds Existing CENP-A Nucleosomes to Promote Centromeric Chromatin Assembly. *Developmental Cell,* 42**,** 190-+.

FRIDLYAND, J., SNIJDERS, A. M., PINKEL, D., ALBERTSON, D. G. & JAIN, A. N. 2004. Hidden Markov models approach to the analysis of array CGH data. *Journal of Multivariate Analysis,* 90**,** 132-153.

FUJITA, Y. 2007. Priming of centromere for CENP-A recruitment by human hMis18[alpha], hMis18[beta], and M18BP1. *Dev. Cell,* 12**,** 17-30.

FUKAGAWA, T. & EARNSHAW, W. C. 2014. The Centromere: Chromatin Foundation for the Kinetochore Machinery. *Developmental Cell,* 30**,** 497-509.

GHOSH, K. & VAN DUYNE, G. D. 2002. Cre-loxP biochemistry. *Methods,* 28**,** 374-383.

GORDON, J. L., BYRNE, K. P. & WOLFE, K. H. 2011. Mechanisms of Chromosome Number Evolution in Yeast. *Plos Genetics,* 7.

GOSHIMA, G., KIYOMITSU, T., YODA, K. & YANAGIDA, M. 2003. Human centromere chromatin protein hMis12, essential for equal segregation, is independent of CENP-A loading pathway. *J. Cell Biol.,* 160**,** 25-39.

GOULD, K. L. 2003. Protein Kinases Driving the Cell Cycle. *In:* EGEL, R. (ed.) *The Molecular Biology of Schizosaccharomyces pombe.* Springer.

GREER, E. L., MAURES, T. J., UCAR, D., HAUSWIRTH, A. G., MANCINI, E., LIM, J. P., BENAYOUN, B. A., SHI, Y. & BRUNET, A. 2011. Transgenerational epigenetic inheritance of longevity in Caenorhabditis elegans. *Nature,* 479**,** 365-U204.

GREWAL, S. I. S. 2000. Transcriptional silencing in fission yeast. *Journal of Cellular Physiology,* 184**,** 311-318.

GREWAL, S. I. S. & JIA, S. T. 2007. Heterochromatin revisited. *Nature Reviews Genetics,* 8**,** 35-46.

GRINDLEY, N. D. F., WHITESON, K. L. & RICE, P. A. 2006. Mechanisms of site-specific recombination. *Annual Review of Biochemistry.*

GROTH, A. C., OLIVARES, E. C., THYAGARAJAN, B. & CALOS, M. P. 2000. A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 5995-6000.

GUSE, A., CARROLL, C. W., MOREE, B., FULLER, C. J. & STRAIGHT, A. F. 2011. In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature,* 477**,** 354-U136.

HAAF, T., MATER, A. G., WIENBERG, J. & WARD, D. C. 1995. PRESENCE AND ABUNDANCE OF CENP-B BOX SEQUENCES IN GREAT APE SUBSETS OF PRIMATE-SPECIFIC ALPHA-SATELLITE DNA. *Journal of Molecular Evolution,* 41**,** 487-491.

HAGAN, I., CARR, A., GRALLERT, A., NURSE, P. 2016. *Fission Yeast: A Laboratory Manual,* Cold Spring Harbor Laboratory Press.

HALFMANN, R., JAROSZ, D. F., JONES, S. K., CHANG, A., LANCASTER, A. K. & LINDQUIST, S. 2012. Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature,* 482**,** 363-U1507.

HALFMANN, R. & LINDQUIST, S. 2010. Epigenetics in the Extreme: Prions and the Inheritance of Environmentally Acquired Traits. *Science,* 330**,** 629-632.

HALL, L. E., MITCHELL, S. E. & O'NEILL, R. J. 2012. Pericentric and centromeric transcription: a perfect balance required. *Chromosome Research,* 20**,** 535-546.
HALVERSON, D., BAUM, M., STRYKER, J., CARBON, J. & CLARKE, L. 1997. Centromere DNA-binding protein from fission yeast affects chromosome segregation and has homology to human CENP-B. *Journal of Cell Biology,* 136**,** 487-500.
HANSON, S. J., BYRNE, K. P. & WOLFE, K. H. 2014. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus Saccharomyces cerevisiae system. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** E4851-E4858.
HARING, M., BADER, R., LOUWERS, M., SCHWABE, A., VAN DRIEL, R. & STAM, M. 2010. The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. *Plant Journal,* 63**,** 366-378.
HARRINGTON, J. J., VANBOKKELEN, G., MAYS, R. W., GUSTASHAW, K. & WILLARD, H. F. 1997. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nature Genetics,* 15**,** 345-355.
HATFULL, G. F. & GRINDLEY, N. D. F. 1986. ANALYSIS OF GAMMA-DELTA-RESOLVASE MUTANTS INVITRO - EVIDENCE FOR AN INTERACTION BETWEEN SERINE-10 OF RESOLVASE AND SITE-I OF RES. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 5429-5433.
HEARD, E. & MARTIENSSEN, R. A. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell,* 157**,** 95-109.
HEGEMANN, J. H. & FLEIG, U. N. 1993. THE CENTROMERE OF BUDDING YEAST. *Bioessays,* 15**,** 451-460.
HELSTON, R. M., BOX, J. A., TANG, W. & BAUMANN, P. 2010. Schizosaccharomyces cryophilus sp nov., a new species of fission yeast. *Fems Yeast Research,* 10**,** 779-786.
HENIKOFF, S., AHMAD, K. & MALIK, H. S. 2001. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science,* 293**,** 1098-1102.
HENIKOFF, S., AHMAD, K., PLATERO, J. S. & VAN STEENSEL, B. 2000. Heterochromatic deposition of centromeric histone H3-like proteins. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 716-721.
HENTGES, P., VAN DRIESSCHE, B., TAFFOREAU, L., VANDENHAUTE, J. & CARR, A. M. 2005. Three novel antibiotic marker cassettes for gene disruption and marker switching in Schizosaccharomyces pombe. *Yeast,* 22**,** 1013-1019.
HERMAND, D., PIHLAK, A., WESTERLING, T., DAMAGNEZ, V., VANDENHAUTE, J., COTTAREL, G. & MAKELA, T. P. 1998. Fission yeast Csk1 is a CAK-activating kinase (CAKAK). *Embo Journal,* 17**,** 7230-7238.
HEUN, P. 2006. Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev. Cell,* 10**,** 303-315.
HIETER, P., PRIDMORE, D., HEGEMANN, J. H., THOMAS, M., DAVIS, R. W. & PHILIPPSEN, P. 1985. FUNCTIONAL SELECTION AND ANALYSIS OF YEAST CENTROMERIC DNA. *Cell,* 42**,** 913-921.

HIGGINS, A. W., GUSTASHAW, K. M. & WILLARD, H. F. 2005. Engineered human dicentric chromosomes show centromere plasticity. *Chromosome Research,* 13**,** 745-762.

HOESS, R. H. & ABREMSKI, K. 1985. MECHANISM OF STRAND CLEAVAGE AND EXCHANGE IN THE CRE-LOX SITE-SPECIFIC RECOMBINATION SYSTEM. *Journal of Molecular Biology,* 181**,** 351-362.

HOLLICK, J. B. 2012. Paramutation: a trans-homolog interaction affecting heritable gene regulation. *Current Opinion in Plant Biology,* 15**,** 536-543.

HORI, T., AMANO, M., SUZUKI, A., BACKER, C. B., WELBURN, J. P., DONG, Y. M., MCEWEN, B. F., SHANG, W. H., SUZUKI, E., OKAWA, K., CHEESEMAN, I. M. & FUKAGAWA, T. 2008. CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. *Cell,* 135**,** 1039-1052.

HORI, T., SHANG, W. H., HARA, M., ARIYOSHI, M., ARIMURA, Y., FUJITA, R., KURUMIZAKA, H. & FUKAGAWA, T. 2017. Association of M18BP1/KNL2 with CENP-A Nucleosome Is Essential for Centromere Formation in Non-mammalian Vertebrates. *Developmental Cell,* 42**,** 181-+.

HORI, T., SHANG, W. H., TAKEUCHI, K. & FUKAGAWA, T. 2013. The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *Journal of Cell Biology,* 200**,** 45-60.

HUA, S. S., WANG, Z. K., JIANG, K., HUANG, Y. J., WARD, T., ZHAO, L. L., DOU, Z. & YAO, X. B. A. 2011. CENP-U Cooperates with Hec1 to Orchestrate Kinetochore-Microtubule Attachment. *Journal of Biological Chemistry,* 286**,** 1627-1638.

HUBER, W., CAREY, V. J., GENTLEMAN, R., ANDERS, S., CARLSON, M., CARVALHO, B. S., BRAVO, H. C., DAVIS, S., GATTO, L., GIRKE, T., GOTTARDO, R., HAHNE, F., HANSEN, K. D., IRIZARRY, R. A., LAWRENCE, M., LOVE, M. I., MACDONALD, J., OBENCHAIN, V., OLES, A. K., PAGES, H., REYES, A., SHANNON, P., SMYTH, G. K., TENENBAUM, D., WALDRON, L. & MORGAN, M. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods,* 12**,** 115-121.

IKENO, M., GRIMES, B., OKAZAKI, T., NAKANO, M., SAITOH, K., HOSHINO, H., MCGILL, N. I., COOKE, H. & MASUMOTO, H. 1998. Construction of YAC-based mammalian artificial chromosomes. *Nature Biotechnology,* 16**,** 431-439.

IRAQUI, I., CHEKKAL, Y., JMARI, N., PIETROBON, V., FREON, K., COSTES, A. & LAMBERT, S. A. E. 2012. Recovery of Arrested Replication Forks by Homologous Recombination Is Error-Prone. *Plos Genetics,* 8.

IRELAN, J. T., GUTKIN, G. I. & CLARKE, L. 2001. Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. *Genetics,* 157**,** 1191-1203.

IRVINE, D. V., AMOR, D. J., PERRY, J., SIRVENT, N., PEDEUTOUR, F., CHOO, K. H. A. & SAFFERY, R. 2004. Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres. *Chromosome Research,* 12**,** 805-815.

ISHII, K., OGIYAMA, Y., CHIKASHIGE, Y., SOEJIMA, S., MASUDA, F., KAKUMA, T., HIRAOKA, Y. & TAKAHASHI, K. 2008. Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. *Science,* 321**,** 1088-1091.

IZUTA, H. 2006. Comprehensive analysis of the ICEN (interphase centromere complex) components enriched in the CENP-A chromatin of human cells. *Genes Cells,* 11**,** 673-684.

JAIN, A. K., XI, Y. X., MCCARTHY, R., ALLTON, K., AKDEMIR, K. C., PATEL, L. R., ARONOW, B., LIN, C. R., LI, W., YANG, L. Q. & BARTON, M. C. 2016. LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. *Molecular Cell,* 64**,** 967-981.

JANSEN, L. E. T., BLACK, B. E., FOLTZ, D. R. & CLEVELAND, D. W. 2007. Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.,* 176**,** 795-805.

JEFFARES, D. C., RALLIS, C., RIEUX, A., SPEED, D., PREVOROVSKY, M., MOURIER, T., MARSELLACH, F. X., IQBAL, Z., LAU, W., CHENG, T. M. K., PRACANA, R., MULLEDER, M., LAWSON, J. L. D., CHESSEL, A., BALA, S., HELLENTHAL, G., O'FALLON, B., KEANE, T., SIMPSON, J. T., BISCHOF, L., TOMICZEK, B., BITTON, D. A., SIDERI, T., CODLIN, S., HELLBERG, J., VAN TRIGT, L., JEFFERY, L., LI, J. J., ATKINSON, S., THODBERG, M., FEBRER, M., MCLAY, K., DROU, N., BROWN, W., HAYLES, J., SALAS, R. E. C., RALSER, M., MANIATIS, N., BALDING, D. J., BALLOUX, F., DURBIN, R. & BAHLER, J. 2015. The genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nature Genetics,* 47**,** 235-+.

JOSH, N. A. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33).

KAPOOR, S., ZHU, L. S., FROYD, C., LIU, T. & RUSCHE, L. N. 2015. Regional centromeres in the yeast Candida lusitaniae lack pericentromeric heterochromatin. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** 12139-12144.

KARPEN, G. H. & ALLSHIRE, R. C. 1997. The case for epigenetic effects on centromere identity and function. *Trends Genet.,* 13**,** 489-496.

KATO, T., SATO, N., HAYAMA, S., YAMABUKI, T., ITO, T., MIYAMOTO, M., KONDO, S., NAKAMURA, Y. & DAIGO, Y. 2007. Activation of holliday junction-recognizing protein involved in the chromosomal stability and immortality of cancer cells. *Cancer Research,* 67**,** 8544-8553.

KATZ, D. J., EDWARDS, T. M., REINKE, V. & KELLY, W. G. 2009. A C. elegans LSD1 Demethylase Contributes to Germline Immortality by Reprogramming Epigenetic Memory. *Cell,* 137**,** 308-320.

KETEL, C., WANG, H. S. W., MCCLELLAN, M., BOUCHONVILLE, K., SELMECKI, A., LAHAV, T., GERAMI-NEJAD, M. & BERMAN, J. 2009. Neocentromeres Form Efficiently at Multiple Possible Loci in Candida albicans. *Plos Genetics,* 5.

KHARE, S. P., HABIB, F., SHARMA, R., GADEWAL, N., GUPTA, S. & GALANDE, S. 2012. HIstome-a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Research,* 40**,** D337-D342.

KIPLING, D., MITCHELL, A. R., MASUMOTO, H., WILSON, H. E., NICOL, L. & COOKE, H. J. 1995. CENP-B BINDS A NOVEL CENTROMERIC SEQUENCE IN THE ASIAN MOUSE MUS CAROLI. *Molecular and Cellular Biology,* 15**,** 4009-4020.

KLINE, S. L., CHEESEMAN, I. M., HORI, T., FUKAGAWA, T. & DESAI, A. 2006. The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation. *J. Cell Biol.,* 173**,** 9-17.

KOBAYASHI, N., SUZUKI, Y., SCHOENFELD, L. W., MULLER, C. A., NIEDUSZYNSKI, C., WOLFE, K. H. & TANAKAL, T. U. 2015. Discovery of an Unconventional Centromere in Budding Yeast Redefines Evolution of Point Centromeres. *Current Biology,* 25**,** 2026-2033.

KUSCU, C., ZARATIEGUI, M., KIM, H. S., WAH, D. A., MARTIENSSEN, R. A., SCHALCH, T. & JOSHUA-TOR, L. 2014. CRL4-like Clr4 complex in Schizosaccharomyces pombe depends on an exposed surface of Dos1 for heterochromatin silencing (vol 111, pg 1795, 2014). *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 14307-14307.

LACOSTE, N., WOOLFE, A., TACHIWANA, H., GAREA, A. V., BARTH, T., CANTALOUBE, S., KURUMIZAKA, H., IMHOF, A. & ALMOUZNI, G. 2014. Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX. *Molecular Cell,* 53**,** 631-644.

LAIRD, C. D., PLEASANT, N. D., CLARK, A. D., SNEEDEN, J. L., HASSAN, K. M., MANLEY, N. C., VARY, J. C., JR., MORGAN, T., HANSEN, R. S. & STOGER, R. 2004. Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc Natl Acad Sci U S A,* 101**,** 204-9.

LANDY, A. 1989. DYNAMIC, STRUCTURAL, AND REGULATORY ASPECTS OF LAMBDA-SITE-SPECIFIC RECOMBINATION. *Annual Review of Biochemistry,* 58**,** 913-949.

LECHNER, J. & CARBON, J. 1991. A 240-KD MULTISUBUNIT PROTEIN COMPLEX, CBF3, IS A MAJOR COMPONENT OF THE BUDDING YEAST CENTROMERE. *Cell,* 64**,** 717-725.

LEE, C., WEVRICK, R., FISHER, R. B., FERGUSONSMITH, M. A. & LIN, C. C. 1997. Human centromeric DNAs. *Human Genetics,* 100**,** 291-304.

LEE, K. M., SAIZ, J. E., BARTON, W. A. & FISHER, R. P. 1999. Cdc2 activation in fission yeast depends on Mcs6 and Csk1, two partially redundant Cdk-activating kinases (CAKs). *Current Biology,* 9**,** 441-444.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, W. K., KAMTEKAR, S., XIONG, Y., SARKIS, G. J., GRINDLEY, N. D. F. & STEITZ, T. A. 2005. Structure of a synaptic gamma delta resolvase tetramer covalently linked to two cleaved DNAs. *Science,* 309**,** 1210-1215.

LIU, S. T., RATTNER, J. B., JABLONSKI, S. A. & YEN, T. J. 2006. Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. *J. Cell Biol.,* 175**,** 41-53.

LIU, X. K., MCLEOD, I., ANDERSON, S., YATES, J. R. & HE, X. W. 2005. Molecular analysis of kinetochore architecture in fission yeast. *Embo Journal,* 24**,** 2919-2930.

LOCKE, D. P., HILLIER, L. W., WARREN, W. C., WORLEY, K. C., NAZARETH, L. V., MUZNY, D. M., YANG, S. P., WANG, Z. Y., CHINWALLA, A. T., MINX, P., MITREVA, M., COOK, L., DELEHAUNTY, K. D., FRONICK, C., SCHMIDT, H., FULTON, L. A., FULTON, R. S., NELSON, J. O., MAGRINI, V., POHL, C., GRAVES, T. A., MARKOVIC, C., CREE, A., DINH, H. H., HUME, J., KOVAR, C. L., FOWLER, G. R., LUNTER, G.,

MEADER, S., HEGER, A., PONTING, C. P., MARQUES-BONET, T., ALKAN, C., CHEN, L., CHENG, Z., KIDD, J. M., EICHLER, E. E., WHITE, S., SEARLE, S., VILELLA, A. J., CHEN, Y. A., FLICEK, P., MA, J. A., RANEY, B., SUH, B., BURHANS, R., HERRERO, J., HAUSSLER, D., FARIA, R., FERNANDO, O., DARRE, F., FARRE, D., GAZAVE, E., OLIVA, M., NAVARRO, A., ROBERTO, R., CAPOZZI, O., ARCHIDIACONO, N., DELLA VALLE, G., PURGATO, S., ROCCHI, M., KONKEL, M. K., WALKER, J. A., ULLMER, B., BATZER, M. A., SMIT, A. F. A., HUBLEY, R., CASOLA, C., SCHRIDER, D. R., HAHN, M. W., QUESADA, V., PUENTE, X. S., ORDONEZ, G. R., LOPEZ-OTIN, C., VINAR, T., BREJOVA, B., RATAN, A., HARRIS, R. S., MILLER, W., KOSIOL, C., LAWSON, H. A., TALIWAL, V., MARTINS, A. L., SIEPEL, A., ROYCHOUDHURY, A., MA, X., DEGENHARDT, J., BUSTAMANTE, C. D., GUTENKUNST, R. N., MAILUND, T., DUTHEIL, J. Y., HOBOLTH, A., SCHIERUP, M. H., RYDER, O. A., YOSHINAGA, Y., DE JONG, P. J., WEINSTOCK, G. M., ROGERS, J., MARDIS, E. R., GIBBS, R. A., et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Nature,* 469**,** 529-533.

LU, Z. P., CARTER, A. C. & CHANG, H. Y. 2017. Mechanistic insights in X-chromosome inactivation. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 372.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature,* 389**,** 251-260.

LUO, H. E., LI, J. T., ESHAGHI, M., LIU, J. H. & KARUTURI, R. K. M. 2010. Genome-wide estimation of firing efficiencies of origins of DNA replication from time-course copy number variation data. *Bmc Bioinformatics,* 11.

LYNCH, D. B., LOGUE, M. E., BUTLER, G. & WOLFE, K. H. 2010. Chromosomal G+C Content Evolution in Yeasts: Systematic Interspecies Differences, and GC-Poor Troughs at Centromeres. *Genome Biology and Evolution,* 2**,** 572-583.

LYON, M. F. 1961. GENE ACTION IN X-CHROMOSOME OF MOUSE (MUS MUSCULUS L). *Nature,* 190**,** 372-&.

MALIK, H. S. & HENIKOFF, S. 2009. Major Evolutionary Transitions in Centromere Complexity. *Cell,* 138**,** 1067-1082.

MALLA, S., DAFHNIS-CALAS, F., BROOKFIELD, J. F. Y., SMITH, M. C. M. & BROWN, W. R. A. 2005. Rearranging the centromere of the human Y chromosome with phi C31 integrase. *Nucleic Acids Research,* 33**,** 6101-6113.

MALVEZZI, F., LITOS, G., SCHLEIFFER, A., HEUCK, A., CLAUSEN, T. & WESTERMANN, S. 2012. A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. *Molecular Biology of the Cell,* 23.

MARSHALL, O. J., CHUEH, A. C., WONG, L. H. & CHOO, K. H. A. 2008. Neocentromeres: New insights into centromere structure, disease development, and karyotype evolution. *American Journal of Human Genetics,* 82**,** 261-282.

MASUMOTO, H., MASUKATA, H., MURO, Y., NOZAKI, N. & OKAZAKI, T. 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.,* 109**,** 1963-1973.

MAUNDRELL, K. 1993. THIAMINE-REPRESSIBLE EXPRESSION VECTORS PREP AND PRIP FOR FISSION YEAST. *Gene,* 123**,** 127-130.

MCGREW, J., DIEHL, B. & FITZGERALDHAYES, M. 1986. SINGLE BASE-PAIR MUTATIONS IN CENTROMERE ELEMENT-III CAUSE ABERRANT CHROMOSOME SEGREGATION IN SACCHAROMYCES-CEREVISIAE. *Molecular and Cellular Biology,* 6**,** 530-538.

MENDIBURO, M. J., PADEKEN, J., FULOP, S., SCHEPERS, A. & HEUN, P. 2011. Drosophila CENH3 Is Sufficient for Centromere Formation. *Science,* 334**,** 686-690.

MONFORT, A. & WUTZ, A. 2017. Progress in understanding the molecular mechanism of Xist RNA function through genetics. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 372.

MORALES, L., NOEL, B., PORCEL, B., MARCET-HOUBEN, M., HULLO, M. F., SACERDOT, C., TEKAIA, F., LEH-LOUIS, V., DESPONS, L., KHANNA, V., AURY, J. M., BARBE, V., COULOUX, A., LABADIE, K., PELLETIER, E., SOUCIET, J. L., BOEKHOUT, T., GABALDON, T., WINCKER, P. & DUJON, B. 2013. Complete DNA Sequence of Kuraishia capsulata Illustrates Novel Genomic Features among Budding Yeasts (Saccharomycotina). *Genome Biology and Evolution,* 5**,** 2524-2539.

MORENO, S. & NURSE, P. 1994. REGULATION OF PROGRESSION THROUGH THE G1 PHASE OF THE CELL-CYCLE BY THE RUM1(+) GENE. *Nature,* 367**,** 236-242.

MURAKAMI, Y., HUBERMAN, J. A. & HURWITZ, J. 1996. Identification, purification, and molecular cloning of autonomously replicating sequence-binding protein 1 from fission yeast Schizosaccharomyces pombe. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 502-507.

MUSACCHIO, A. & DESAI, A. 2017. A Molecular View of Kinetochore Assembly and Function. *Biology (Basel),* 6.

NAGPAL, H. & FUKAGAWA, T. 2016. Kinetochore assembly and function through the cell cycle. *Chromosoma,* 125**,** 645-659.

NAGPAL, H., HORI, T., FURUKAWA, A., SUGASE, K., OSAKABE, A., KURUMIZAKA, H. & FUKAGAWA, T. 2015. Dynamic changes in CCAN organization through CENP-C during cell-cycle progression. *Molecular Biology of the Cell,* 26**,** 3768-3776.

NAGY, A. 2000. Cre recombinase: The universal reagent for genome tailoring. *Genesis,* 26**,** 99-109.

NAKAGAWA, H., LEE, J. K., HURWITZ, J., ALLSHIRE, R. C., NAKAYAMA, J., GREWAL, S. I. S., TANAKA, K. & MURAKAMI, Y. 2002. Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. *Genes & Development,* 16**,** 1766-1778.

NAKAYAMA, J., KLAR, A. J. S. & GREWAL, S. I. S. 2000. A chromodomain protein, Swi6, performs imprinting functions in fission yeast during mitosis and meiosis. *Cell,* 101**,** 307-317.

NARDI, I. K., ZASADZINSKA, E., STELLFOX, M. E., KNIPPLER, C. M. & FOLTZ, D. R. 2016. Licensing of Centromeric Chromatin Assembly through the Mis18 alpha-Mis18 beta Heterotetramer. *Molecular Cell,* 61**,** 774-787.

NASUDA, S., HUDAKOVA, S., SCHUBERT, I., HOUBEN, A. & ENDO, T. R. 2005. Stable barley chromosomes without centromeric repeats. *Proceedings of*

*the National Academy of Sciences of the United States of America,* 102**,** 9842-9847.

NISHINO, T., RAGO, F., HORI, T., TOMII, K., CHEESEMAN, I. M. & FUKAGAWA, T. 2013. CENP-T provides a structural platform for outer kinetochore assembly. *Embo Journal,* 32**,** 424-436.

NISHINO, T., TAKEUCHI, K., GASCOIGNE, K. E., SUZUKI, A., HORI, T., OYAMA, T., MORIKAWA, K., CHEESEMAN, I. M. & FUKAGAWA, T. 2012. CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. *Cell,* 148**,** 487-501.

NOMA, K., ALLIS, C. D. & GREWAL, S. I. S. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science,* 293**,** 1150-1155.

NURSE, P. & THURIAUX, P. 1980. REGULATORY GENES-CONTROLLING MITOSIS IN THE FISSION YEAST SCHIZOSACCHAROMYCES-POMBE. *Genetics,* 96**,** 627-637.

OBUSE, C. 2004. A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. *Nature Cell Biol.,* 6**,** 1135-1141.

OEGEMA, K., DESAI, A., RYBINA, S., KIRKHAM, M. & HYMAN, A. A. 2001. Functional analysis of kinetochore assembly in Caenorhabditis elegans. *J. Cell Biol.,* 153**,** 1209-1225.

OKADA, M. 2006. The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nature Cell Biol.,* 8**,** 446-457.

OKAMOTO, Y., NAKANO, M., OHZEKI, J. I., LARIONOV, V. & MASUMOTO, H. 2007. A minimal CENP-A core is required for nucleation and maintenance of a functional human centromere. *Embo Journal,* 26**,** 1279-1291.

OKAZAKI, K., OKAZAKI, N., KUME, K., JINNO, S., TANAKA, K. & OKAYAMA, H. 1990. HIGH-FREQUENCY TRANSFORMATION METHOD AND LIBRARY TRANSDUCING VECTORS FOR CLONING MAMMALIAN CDNAS BY TRANSCOMPLEMENTATION OF SCHIZOSACCHAROMYCES-POMBE. *Nucleic Acids Research,* 18**,** 6485-6489.

OLORUNNIJI, F. J. & STARK, W. M. 2010. Catalysis of site-specific recombination by Tn3 resolvase. *Biochemical Society Transactions,* 38**,** 417-421.

OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. & WIGLER, M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics,* 5**,** 557-72.

OLSZAK, A. M., VAN ESSEN, D., PEREIRA, A. J., DIEHL, S., MANKE, T., MAIATO, H., SACCANI, S. & HEUN, P. 2011. Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nature Cell Biology,* 13**,** 799-U166.

ORR, B. & SUNKEL, C. E. 2011. Drosophila CENP-C is essential for centromere identity. *Chromosoma,* 120**,** 83-96.

PADMANABHAN, S., THAKUR, J., SIDDHARTHAN, R. & SANYAL, K. 2008. Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, Candida albicans and Candida dubliniensis. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 19797-19802.

PALMER, D. K., O'DAY, K. & MARGOLIS, R. L. 1990. The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. *Chromosoma,* 100**,** 32-6.

PALMER, D. K., O'DAY, K., TRONG, H. L., CHARBONNEAU, H. & MARGOLIS, R. L. 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl Acad. Sci. USA,* 88**,** 3734-3738.

PARKER, L. L., WALTER, S. A., YOUNG, P. G. & PIWNICAWORMS, H. 1993. PHOSPHORYLATION AND INACTIVATION OF THE MITOTIC INHIBITOR WEE1 BY THE NIM1/CDR1 KINASE. *Nature,* 363**,** 736-738.

PARTRIDGE, J. F., BORGSTROM, B. & ALLSHIRE, R. C. 2000. Distinct protein interaction domains and protein spreading in a complex centromere. *Genes & Development,* 14**,** 783-791.

PETROVIC, A., PASQUALATO, S., DUBE, P., KRENN, V., SANTAGUIDA, S., CITTARO, D., MONZANI, S., MASSIMILIANO, L., KELLER, J., TARRICONE, A., MAIOLICA, A., STARK, H. & MUSACCHIO, A. 2010. The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *Journal of Cell Biology,* 190**,** 835-852.

PIDOUX, A. L., CHOI, E. S., ABBOTT, J. K. R., LIU, X. K., KAGANSKY, A., CASTILLO, A. G., HAMILTON, G. L., RICHARDSON, W., RAPPSILBER, J., HE, X. & ALLSHIRE, R. C. 2009. Fission Yeast Scm3: A CENP-A Receptor Required for Integrity of Subkinetochore Chromatin. *Molecular Cell,* 33**,** 299-311.

PIRAS, F. M., NERGADZE, S. G., MAGNANI, E., BERTONI, L., ATTOLINI, C., KHORIAULI, L., RAIMONDI, E. & GIULOTTO, E. 2010. Uncoupling of Satellite DNA and Centromeric Function in the Genus Equus. *Plos Genetics,* 6.

PLUTA, A. F., MACKAY, A. M., AINSZTEIN, A. M., GOLDBERG, I. G. & EARNSHAW, W. C. 1995. CENTROMERE - HUB OF CHROMOSOMAL ACTIVITIES. *Science,* 270**,** 1591-1594.

PRZEWLOKA, M. R., VENKEI, Z., BOLANOS-GARCIA, V. M., DEBSKI, J., DADLEZ, M. & GLOVER, D. M. 2011. CENP-C Is a Structural Platform for Kinetochore Assembly. *Current Biology,* 21**,** 399-405.

RAVIN, N. V., ELDAROV, M. A., KADNIKOV, V. V., BELETSKY, A. V., SCHNEIDER, J., MARDANOVA, E. S., SMEKALOVA, E. M., ZVEREVA, M. I., DONTSOVA, O. A., MARDANOV, A. V. & SKRYABIN, K. G. 2013. Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1. *Bmc Genomics,* 14.

RAY-GALLET, D., QUIVY, J. P., SCAMPS, C., MARTINI, E. M. D., LIPINSKI, M. & ALMOUZNI, G. 2002. HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Molecular Cell,* 9**,** 1091-1100.

REA, S., EISENHABER, F., O'CARROLL, N., STRAHL, B. D., SUN, Z. W., SCHMID, M., OPRAVIL, S., MECHTLER, K., PONTING, C. P., ALLIS, C. D. & JENUWEIN, T. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature,* 406**,** 593-599.

REIK, W., DEAN, W. & WALTER, J. 2001. Epigenetic reprogramming in mammalian development. *Science,* 293**,** 1089-1093.

REINHART, B. J. & BARTEL, D. P. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science,* 297**,** 1831-1831.

ROY, B. & SANYAL, K. 2011. Diversity in Requirement of Genetic and Epigenetic Factors for Centromere Function in Fungi. *Eukaryotic Cell,* 10**,** 1384-1395.
ROY, B., VARSHNEY, N., YADAV, V. & SANYAL, K. 2013. The process of kinetochore assembly in yeasts. *Fems Microbiology Letters,* 338**,** 107-117.
RUSSELL, P. & NURSE, P. 1987. NEGATIVE REGULATION OF MITOSIS BY WEE1+, A GENE ENCODING A PROTEIN-KINASE HOMOLOG. *Cell,* 49**,** 559-567.
SAITOH, H. 1992. CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell,* 70**,** 115-125.
SAITOH, S., TAKAHASHI, K. & YANAGIDA, M. 1997. Mis6, a fission yeast inner centromere protein, acts during G1/S and forms specialized chromatin required for equal segregation. *Cell,* 90**,** 131-143.
SANYAL, K., BAUM, M. & CARBON, J. 2004. Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 11374-11379.
SCHUELER, M. G., HIGGINS, A. W., RUDD, M. K., GUSTASHAW, K. & WILLARD, H. F. 2001. Genomic and genetic definition of a functional human centromere. *Science,* 294**,** 109-115.
SCHUH, M., LEHNER, C. F. & HEIDMANN, S. 2007. Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. *Curr. Biol.,* 17**,** 237-243.
SCOTT, K. C. & SULLIVAN, B. A. 2014. Neocentromeres: a place for everything and everything in its place. *Trends in Genetics,* 30**,** 66-74.
SCREPANTI, E., DE ANTONI, A., ALUSHIN, G. M., PETROVIC, A., MELIS, T., NOGALES, E. & MUSACCHIO, A. 2011. Direct Binding of Cenp-C to the Mis12 Complex Joins the Inner and Outer Kinetochore. *Current Biology,* 21**,** 391-398.
SEONG, K. H., LI, D., SHIMIZU, H., NAKAMURA, R. & ISHII, S. 2011. Inheritance of Stress-Induced, ATF-2-Dependent Epigenetic Change. *Cell,* 145**,** 1049-1061.
SHANG, W. H., HORI, T., MARTINS, N. M. C., TOYODA, A., MISU, S., MONMA, N., HIRATANI, I., MAESHIMA, K., IKEO, K., FUJIYAMA, A., KIMURA, H., EARNSHAW, W. C. & FUKAGAWA, T. 2013. Chromosome Engineering Allows the Efficient Isolation of Vertebrate Neocentromeres. *Developmental Cell,* 24**,** 635-648.
SHANG, W. H., HORI, T., TOYODA, A., KATO, J., POPENDORF, K., SAKAKIBARA, Y., FUJIYAMA, A. & FUKAGAWA, T. 2010. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Research,* 20**,** 1219-1228.
SHERRATT, D. J., ARCISZEWSKA, L. K., BLAKELY, G., COLLOMS, S., GRANT, K., LESLIE, N. & MCCULLOCH, R. 1995. SITE-SPECIFIC RECOMBINATION AND CIRCULAR CHROMOSOME SEGREGATION. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences,* 347**,** 37-42.
SINGH, S., GHOSH, P. & HATFULL, G. F. 2013. Attachment Site Selection and Identity in Bxb1 Serine Integrase-Mediated Site-Specific Recombination. *Plos Genetics,* 9.
SIPICZKI, M. 2000. Where does fission yeast sit on the tree of life? *Genome Biology,* 1**,** 4.

SIPICZKI, M. 2003a. Fission yeast phylogenesis and evolution. *Molecular Biology of Schizosaccharomyces Pombe***,** 431-443.
SIPICZKI, M. 2003b. Fission Yeast Phylogenesis and Evolution. *In:* EGEL, R. (ed.) *The Molecular Biology of Schizosaccharomyces pombe.* Springer.
SMITH, M. C. M. & THORPE, H. M. 2002. Diversity in the serine recombinases. *Molecular Microbiology,* 44**,** 299-307.
SMITH, S. & STILLMAN, B. 1989. PURIFICATION AND CHARACTERIZATION OF CAF-I, A HUMAN CELL FACTOR REQUIRED FOR CHROMATIN ASSEMBLY DURING DNA-REPLICATION INVITRO. *Cell,* 58**,** 15-25.
SOUTHERN, E. M. 1975. DETECTION OF SPECIFIC SEQUENCES AMONG DNA FRAGMENTS SEPARATED BY GEL-ELECTROPHORESIS. *Journal of Molecular Biology,* 98**,** 503-&.
STARK, W. M., BOOCOCK, M. R., OLORUNNIJI, F. J. & ROWLAND, S. J. 2011. Intermediates in serine recombinase-mediated site-specific recombination. *Biochemical Society Transactions,* 39**,** 617-622.
STEINER, F. A. & HENIKOFF, S. 2015. Diversity in the organization of centromeric chromatin. *Current Opinion in Genetics & Development,* 31**,** 28-35.
STEINER, N. C. & CLARKE, L. 1994. A NOVEL EPIGENETIC EFFECT CAN ALTER CENTROMERE FUNCTION IN FISSION YEAST. *Cell,* 79**,** 865-874.
STEINER, N. C., HAHNENBERGER, K. M. & CLARKE, L. 1993. CENTROMERES OF THE FISSION YEAST SCHIZOSACCHAROMYCES-POMBE ARE HIGHLY VARIABLE GENETIC-LOCI. *Molecular and Cellular Biology,* 13**,** 4578-4587.
STOLER, S., KEITH, K. C., CURNICK, K. E. & FITZGERALDHAYES, M. 1995. A MUTATION IN CSE4, AN ESSENTIAL GENE ENCODING A NOVEL CHROMATIN-ASSOCIATED PROTEIN IN YEAST, CAUSES CHROMOSOME NONDISJUNCTION AND CELL-CYCLE ARREST AT MITOSIS. *Genes & Development,* 9**,** 573-586.
SULLIVAN, K. F., HECHENBERGER, M. & MASRI, K. 1994. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.,* 127**,** 581-592.
SUZUKI, M. M. & BIRD, A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet,* 9**,** 465-76.
SVEICZER, A., CSIKASZ-NAGY, A., GYORFFY, B., TYSON, J. J. & NOVAK, B. 2000. Modeling the fission yeast cell cycle: Quantized cycle times in wee1(-) cdc25 Delta mutant cells. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 7865-7870.
TAKAHASHI, K., MURAKAMI, S., CHIKASHIGE, Y., FUNABIKI, H., NIWA, O. & YANAGIDA, M. 1992. A LOW COPY NUMBER CENTRAL SEQUENCE WITH STRICT SYMMETRY AND UNUSUAL CHROMATIN STRUCTURE IN FISSION YEAST CENTROMERE. *Molecular Biology of the Cell,* 3**,** 819-835.
TAKAYAMA, Y., SATO, H., SAITOH, S., OGIYAMA, Y., MASUDA, F. & TAKAHASHI, K. 2008. Biphasic incorporation of centromeric histone CENP-A in fission yeast. *Molecular Biology of the Cell,* 19**,** 682-690.
TALBERT, P. B. & HENIKOFF, S. 2012. Chromatin: Packaging without Nucleosomes. *Current Biology,* 22**,** R1040-R1043.
THORPE, H. M. & SMITH, M. C. M. 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase

family. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 5505-5510.
THORPE, H. M., WILSON, S. E. & SMITH, M. C. M. 2000. Control of directionality in the site-specific recombination system of the Streptomyces phage phi C31. *Molecular Microbiology,* 38**,** 232-241.
THURIAUX, P. 1985. DIRECT SELECTION OF MUTANTS INFLUENCING GENE CONVERSION IN THE YEAST SCHIZOSACCHAROMYCES-POMBE. *Molecular & General Genetics,* 199**,** 365-371.
TINLINE-PURVIS, H., SAVORY, A. P., CULLEN, J. K., DAVE, A., MOSS, J., BRIDGE, W. L., MARGUERAT, S., BAHLER, J., RAGOUSSIS, J., MOTT, R., WALKER, C. A. & HUMPHREY, T. C. 2009. Failed gene conversion leads to extensive end processing and chromosomal rearrangements in fission yeast. *Embo Journal,* 28**,** 3400-3412.
TOPP, C. N., OKAGAKI, R. J., MELO, J. R., KYNAST, R. G., PHILLIPS, R. L. & DAWE, R. K. 2009. Identification of a Maize Neocentromere in an Oat-Maize Addition Line. *Cytogenetic and Genome Research,* 124**,** 228-238.
TSUKAHARA, T., TANNO, Y. & WATANABE, Y. 2010. Phosphorylation of the CPC by Cdk1 promotes chromosome bi-orientation. *Nature,* 467**,** 719-U111.
TUKIAINEN, T., VILLANI, A. C., YEN, A., RIVAS, M. A., MARSHALL, J. L., SATIJA, R., AGUIRRE, M., GAUTHIER, L., FLEHARTY, M., KIRBY, A., UMMINGS, B. B. C., ASTEL, S. E. C., KARCZEWSKI, K. J., AGUET, F., BYRNES, A., LAPPALAINEN, T., REGEV, A., ARDLIE, K. G., HACOHEN, N., MACARTHUR, D. G. & CONSORTIUM, G. T. 2017. Landscape of X chromosome inactivation across human tissues. *Nature,* 550**,** 244-+.
TYLER-SMITH, C., GIMELLI, G., GIGLIO, S., FLORIDIA, C., PANDYA, A., TERZOLI, G., WARBURTON, P. E., EARNSHAW, W. C. & ZUFFARDI, O. 1999. Transmission of a fully functional human neocentromere through three generations. *American Journal of Human Genetics,* 64**,** 1440-1444.
TYLERSMITH, C. & BROWN, W. R. A. 1987. STRUCTURE OF THE MAJOR BLOCK OF ALPHOID SATELLITE DNA ON THE HUMAN Y-CHROMOSOME. *Journal of Molecular Biology,* 195**,** 457-470.
VERNIS, L., POLJAK, L., CHASLES, M., UCHIDA, K., CASAREGOLA, S., KAS, E., MATSUOKA, M., GAILLARDIN, C. & FOURNIER, P. 2001. Only centromeres can supply the partition system required for ARS function in the yeast Yarrowia lipolytica. *Journal of Molecular Biology,* 305**,** 203-217.
VOLPE, T. A., KIDNER, C., HALL, I. M., TENG, G., GREWAL, S. I. S. & MARTIENSSEN, R. A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science,* 297**,** 1833-1837.
WADE, C. M., GIULOTTO, E., SIGURDSSON, S., ZOLI, M., GNERRE, S., IMSLAND, F., LEAR, T. L., ADELSON, D. L., BAILEY, E., BELLONE, R. R., BLOCKER, H., DISTL, O., EDGAR, R. C., GARBER, M., LEEB, T., MAUCELI, E., MACLEOD, J. N., PENEDO, M. C. T., RAISON, J. M., SHARPE, T., VOGEL, J., ANDERSSON, L., ANTCZAK, D. F., BIAGI, T., BINNS, M. M., CHOWDHARY, B. P., COLEMAN, S. J., DELLA VALLE, G., FRYC, S., GUERIN, G., HASEGAWA, T., HILL, E. W., JURKA, J., KIIALAINEN, A., LINDGREN, G., LIU, J., MAGNANI, E., MICKELSON, J. R., MURRAY, J., NERGADZE, S. G., ONOFRIO, R., PEDRONI, S., PIRAS, M. F., RAUDSEPP, T., ROCCHI, M., ROED, K. H., RYDER, O. A., SEARLE, S., SKOW, L., SWINBURNE, J. E., SYVANEN, A. C., TOZAKI,

T., VALBERG, S. J., VAUDIN, M., WHITE, J. R., ZODY, M. C., LANDER, E. S., LINDBLAD-TOH, K., BROAD INST GENOME SEQUENCING, P. & BROAD INST WHOLE GENOME ASSEMBLY, T. 2009. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science,* 326**,** 865-867.

WARBURTON, P. E., COOKE, C. A., BOURASSA, S., VAFA, O., SULLIVAN, B. A., STETTEN, G., GIMELLI, G., WARBURTON, D., TYLERSMITH, C., SULLIVAN, K. F., POIRIER, G. G. & EARNSHAW, W. C. 1997. Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Current Biology,* 7**,** 901-904.

WEI, R. R., AL-BASSAM, J. & HARRISON, S. C. 2007. The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nature Struct. Mol. Biol.,* 14**,** 54-59.

WESTHORPE, F. G., FULLER, C. J. & STRAIGHT, A. F. 2015. A cell-free CENP-A assembly system defines the chromatin requirements for centromere maintenance. *Journal of Cell Biology,* 209**,** 789-801.

WILLARD, H. F. & WAYE, J. S. 1987. HIERARCHICAL ORDER IN CHROMOSOME-SPECIFIC HUMAN ALPHA-SATELLITE DNA. *Trends in Genetics,* 3**,** 192-198.

WILLIAMS, J. S., HAYASHI, T., YANAGIDA, M. & RUSSELL, P. 2009. Fission Yeast Scm3 Mediates Stable Assembly of Cnp1/CENP-A into Centromeric Chromatin. *Molecular Cell,* 33**,** 287-298.

WU, L. & RUSSELL, P. 1993. NIM1 KINASE PROMOTES MITOSIS BY INACTIVATING WEE1 TYROSINE KINASE. *Nature,* 363**,** 738-741.

YAMANE, T., OGAWA, T. & MATSUOKA, M. 2008. Derivation of consensus sequence for protein binding site in Yarrowia lipolytica centromere. *Journal of Bioscience and Bioengineering,* 105**,** 671-674.

YODA, K., KITAGAWA, K., MASUMOTO, H., MURO, Y. & OKAZAKI, T. 1992. A HUMAN CENTROMERE PROTEIN, CENP-B, HAS A DNA-BINDING DOMAIN CONTAINING 4 POTENTIAL ALPHA-HELICES AT THE NH2 TERMINUS, WHICH IS SEPARABLE FROM DIMERIZING ACTIVITY. *Journal of Cell Biology,* 119**,** 1413-1427.

# 8. APPENDIX I

| Strain | Exp | Plasmid | Average number of cells in a colony | Average number of cells recovered | Corrected number of cells in a colony | Predicted number of cells in a colony | Missing cells | Ura minus cells | Cen excision checked by Southern blot | Cen excision checked by PCR | k1/k | k2/k |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nott373_control | 1 | Rep81 | 5229167 | 0.553 | 2891729 | | | | 9/10 | 36/40 | | |
| | 1 | Rep81Bxb | 134000 | 0.143 | 19162 | 34651 | 5194516 | 0/100 (<0.029) | | | 0.32 | 0 (<0.00006) |
| | 2 | Rep81 | 7380000 | 0.28 | 2066400 | | | | | | | |
| | 2 | Rep81Bxb | 136000 | 0.095 | 12920 | 46143 | 7333857 | 0/100 (<0.029) | | | 0.32 | 0 (<0.00018) |
| | 3 | Rep81 | 4160000 | 0.359 | 1493440 | | | | | | | |
| | 3 | Rep81Bxb | 234615 | 0.148 | 34723 | 96722 | 4125277 | 10/100 | | | 0.25 | 0.00059 |
| Nott373_cen4 | 1 | Rep81 | 2800000 | 0.259 | 725200 | | | | 9/10 | 39/40 | | |
| | 1 | Rep81Bxb | 324000 | 0.083 | 26892 | 103830 | 2696170 | 5/100 | | | 0.22 | 0.00042 |
| | 2 | Rep81 | 3240000 | 0.36 | 1166400 | | | | | | | |
| | 2 | Rep81Bxb | 287719 | 0.192 | 55242 | 153450 | 3086550 | 2/100 | | | 0.2 | 0.0002 |
| | 3 | Rep81 | 2839286 | 0.288 | 817714 | | | | | | | |
| | 3 | Rep81Bxb | 325000 | 0.11 | 35750 | 124132 | 2715154 | 3/100 | | | 0.21 | 0.00029 |
| Nott373_cen3 | 1 | Rep81 | 3679245 | 0.345 | 1269340 | | | | 5/10 | 29/40 | | |
| | 1 | Rep81Bxb | 509804 | 0.182 | 92784 | 268940 | 3410305 | 0/100 (<0.029) | | | 0.17 | 0 (< 0.0031) |
| | 2 | Rep81 | 2627451 | 0.461 | 1211255 | | | | | | | |
| | 2 | Rep81Bxb | 326000 | 0.072 | 23472 | 50915 | 2576536 | 0/100 (<0.029) | | | 0.27 | 0 (< 0.00009) |
| | 3 | Rep81 | 4040000 | 0.335 | 1353400 | | | | | | | |
| | 3 | Rep81Bxb | 476000 | 0.062 | 29512 | 88096 | 3951904 | 0/100 (<0.029) | | | 0.25 | 0 (< 0.00012) |
| Nott373_cen2 | 1 | Rep81 | 1440000 | 0.44 | 633600 | | | | 7/10 | 6/10 | | |
| | 1 | Rep81Bxb | 600000 | 0.724 | 434400 | 987273 | 452727 | 69/100 | | | 0.04 | 0.038 |
| | 2 | Rep81 | 1470000 | 0.4 | 588000 | | | | | | | |
| | 2 | Rep81Bxb | 680000 | 0.78 | 530400 | 1326000 | 144000 | 49/100 | | | 0.01 | 0.022 |
| | 3 | Rep81 | 1500000 | 0.254 | 381000 | | | | | | | |
| | 3 | Rep81Bxb | 450000 | 0.696 | 313200 | 1233071 | 266929 | 63/100 | | | 0.02 | 0.037 |

| Strain | Exp | Plasmid | Average number of cells in a colony | Average number of cells recovered | Corrected number of cells in a colony | Predicted number of cells in a colony | Missing cells | Ura minus cells | Cen excision checked by Southern blot | Cen excision checked by PCR | k1/k | k2/k |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nott373_cen1 | 1 | Rep81 | 923077 | 0.32 | 295385 | | | | 10/10 | 40/40 | | |
| | 1 | Rep81Bxb | 403846 | 0.304 | 122769 | 383654 | 539423 | 52/100 | | | 0.09 | 0.032 |
| | 2 | Rep81 | 1018182 | 0.334 | 340073 | | | | | | | |
| | 2 | Rep81Bxb | 360000 | 0.292 | 105120 | 314731 | 703451 | 82/100 | | | 0.15 | 0.056 |
| | 3 | Rep81 | 1000000 | 0.392 | 392000 | | | | | | | |
| | 3 | Rep81Bxb | 363636 | 0.188 | 68364 | 174397 | 825603 | 73/100 | | | 0.22 | 0.034 |
| Nott373_cen5 | 1 | Rep81 | 3300000 | 0.329 | 1085700 | | | | 10/10 | 39/40 | | |
| | 1 | Rep81Bxb | 5200000 | 0.332 | 1726400 | 5247416 | -1947416 | 98/100 | | | nd | 0.176 |
| | 2 | Rep81 | 3940000 | 0.312 | 1229280 | | | | | | | |
| | 2 | Rep81Bxb | 4060000 | 0.251 | 1019060 | 3266218 | 673782 | 99/100 | | | 0.041 | 0.192 |
| | 3 | Rep81 | 5160000 | 0.283 | 1460280 | | | | | | | |
| | 3 | Rep81Bxb | 5340000 | 0.319 | 1703460 | 6019293 | -859293 | 97/100 | | | nd | 0.189 |
| | 4 | Rep81 | 4720000 | 0.366 | 1727520 | | | | | | | |
| | 4 | Rep81Bxb | 4060000 | 0.414 | 1680840 | 4592459 | 127541 | 99/100 | | | 0.004 | 0.17 |
| | 5 | Rep81 | 3480000 | 0.332 | 1155360 | | | | | | | |
| | 5 | Rep81Bxb | 2780000 | 0.403 | 1120340 | 3374518 | 105482 | 100/100 | | | 0.008 | 0.239 |
| Nott373_cen8 | 1 | Rep81 | 2185185.185 | 0.255 | 557222 | | | | 10/10 | 39/40 | | |
| | 1 | Rep81Bxb | 3040000 | 0.304 | 924160 | 3624157 | -1438971.815 | 100/100 | | | nd | 0.25 |
| | 2 | Rep81 | 3156862.745 | 0.336 | 1060706 | | | | | | | |
| | 2 | Rep81Bxb | 4245614.035 | 0.328 | 1392561 | 5461024 | -3275838.815 | 94/100 | | | nd | 0.22 |
| | 3 | Rep81 | 2344262.295 | 0.3 | 703279 | | | | | | | |
| | 3 | Rep81Bxb | 3640000 | 0.28 | 1019200 | 3397333 | -1053070.705 | 100/100 | | | nd | 0.25 |
| Nott373_alphoid | 1 | Rep81 | 5860000 | 0.426 | 2496360 | | | | 2/10 | 10/40 | | |
| | 1 | Rep81Bxb | 1557692 | 0.127 | 197827 | 464382 | 5395618 | 0/100 (<0.029) | | | 0.16 | 0 (<0.00008) |
| | 2 | Rep81 | 4540000 | 0.386 | 1752440 | | | | | | | |
| | 2 | Rep81Bxb | 1500000 | 0.102 | 153000 | 396373 | 4143627 | 0/100 (<0.029) | | | 0.16 | 0 (<0.00009) |
| | 3 | Rep81 | 6290909 | 0.401 | 2522655 | | | | | | | |
| | 3 | Rep81Bxb | 1269231 | 0.119 | 151038 | 376655 | 5914254 | 1/100 | | | 0.33 | 0.00004 |

**Table 8.1: Absolute numbers of the measurements of number of cells in colony, viability of cells, and centromere excision.**

# 9. APPENDIX II

**Example of calculation of the rate of centromere formation after native centromere excision:**

**Nott373_control (sample 1)**

Control (REP81 empty vector): average 5229167 cells/colony ($e^{kt}$) (1)

Un-swapped cells: 34651 cells (*n*) (2)

Swapped cells= 0 cells (*s*) (3)

Missing cells: 5194516 cells (*x*) (4)

From (1) we have: $kt = \ln 5229167$

So, $kt = 15.4698$

$$\frac{dn}{dt} = (k - k_1 - k_2)\mathrm{n}$$

$$\int \frac{dn}{dt} = \int (k - k_1 - k_2)\mathrm{n}$$

$$\int \frac{dn}{n} = \int (k - k_1 - k_2)\mathrm{t}$$

$$\ln n \ = \ (k - k_1 - k_2)\mathrm{t}$$

From (3) $k_2 = 0$;

So, $\ln n \ = \ (k - k_1)\mathrm{t}$

$$\ln n = kt - k_1 t$$

$$k_1 t = kt - \ln n$$

$$\frac{k_1}{k} = \frac{kt - \ln n}{kt}$$

$$\frac{k_1}{k} = \frac{15.4698 - 10.4531}{15.4698}$$

$\frac{k_1}{k} = 0.32$ and $\frac{k_2}{k} =$ not detectable

# 10. APPENDIX III

| Code | Strain | Cen1CC | Cen2CC | Cen2 TestSeq | Neo-centromere | Cen3CC | Chr3Ade | Cen4 |
|---|---|---|---|---|---|---|---|---|
| WTCHG_261086_293 | Nott376_cen8 | 607077 | 542130 | | 10241 | 571390 | 13579 | 552955 |
| WTCHG_261086_294 | Nott376_cen8 | 283667 | 254478 | | 11805 | 268897 | 9447 | 259616 |
| WTCHG_261086_291 | Nott376_alphoid | 552651 | 486330 | | 9872 | 516699 | 7931 | 496905 |
| WTCHG_261086_292 | Nott376_alphoid | 529940 | 466355 | | 9826 | 491797 | 6936 | 474002 |
| WTCHG_261086_289 | Nott373_control | 717037 | 580654 | | 10075 | 678743 | 2769 | 593511 |
| WTCHG_261086_290 | Nott373_control | 812563 | 654221 | | 10311 | 744153 | 2510 | 666193 |
| WTCHG_261086_287 | Nott373_cen8 | 592997 | 567939 | 511895 | 10120 | 559003 | 2930 | 576843 |
| WTCHG_261086_288 | Nott373_cen8 | 514523 | 481082 | 432983 | 10459 | 480551 | 3085 | 489245 |
| WTCHG_261086_285 | Nott373_alphoid | 984975 | 819324 | 31926 | 10701 | 894533 | 2678 | 842276 |
| WTCHG_261086_286 | Nott373_alphoid | 772485 | 634259 | 25803 | 11052 | 695812 | 2451 | 654137 |
| WTCHG_261086_284 | Nott373_cen5 | 826748 | 748756 | 521488 | 11145 | 780616 | 2950 | 764543 |
| WTCHG_261086_302 | Nott373_cen5 | 810110 | 747086 | 486034 | 14517 | 767500 | 4324 | 757448 |
| WTCHG_261086_281 | Nott373_alphoid | 735479 | | 6743 | 400402 | 669254 | 2196 | 659220 |
| WTCHG_261086_282 | Nott373_alphoid | 731266 | | 7026 | 357004 | 675979 | 2500 | 674840 |
| WTCHG_261086_279 | Nott373_cen8 | 994039 | | 977326 | 11326 | 923166 | 2840 | 895735 |
| WTCHG_261086_280 | Nott373_cen8 | 888379 | | 870146 | 13860 | 820355 | 3845 | 786962 |
| WTCHG_261086_277 | Nott373_cen5 | 762478 | | 529789 | 10325 | 707364 | 2953 | 655399 |
| WTCHG_261086_278 | Nott373_cen8 | 779407 | | 564516 | 12879 | 739873 | 4097 | 667022 |
| WTCHG_261086_275 | Nott373_cen1 | 941828 | | 409543 | 14297 | 876191 | 2509 | 736533 |
| WTCHG_261086_276 | Nott373_cen1 | 636338 | | 4956 | 404692 | 603552 | 3661 | 566594 |
| WTCHG_261086_271 | Nott373_cen2 | 361291 | | 21946 | 138061 | 333451 | 2227 | 313566 |
| WTCHG_261086_272 | Nott373_cen2 | 1015769 | | 6283 | 489158 | 933961 | 2809 | 913938 |
| WTCHG_261086_273 | Nott373_cen2 | 613107 | | 5898 | 413748 | 577968 | 3293 | 552412 |
| WTCHG_261086_274 | Nott373_cen2 | 563075 | | 5074 | 1236 | 523148 | 3022 | 466282 |
| WTCHG_261086_269 | Nott373_cen3 | 787678 | | 3156 | 546752 | 721715 | 2913 | 701543 |
| WTCHG_261086_270 | Nott373_cen3 | 706546 | | 2598 | 408049 | 661880 | 2229 | 630339 |
| WTCHG_261086_267 | Nott373_cen4 | 924794 | | 1040 | 558769 | 856537 | 2718 | 882516 |
| WTCHG_261086_268 | Nott373_cen4 | 710127 | | 1265 | 388977 | 673989 | 3983 | 645712 |
| WTCHG_261086_265 | Nott373_control | 815761 | | | 424189 | 756321 | 2927 | 718321 |
| WTCHG_261086_266 | Nott373_control | 1013926 | | | 487792 | 940474 | 3040 | 882526 |

**Table 10.1: Number of reads in the ChIP-seq experiment.**
Cen=centromere;Chr=chromosome;CC=central core; TestSeq=Test sequence; Ade=*ade6* gene locus.