Title
Genome-wide Analysis of Yeast Meiotic Recombination Landscape

Author
Chen, Stacy Yen-chun

Publication Date
2009

Peer reviewed|Thesis/dissertation

Genome-wide Analysis of Yeast Meiotic Recombination Landscape

by

Stacy Yen-chun Chen

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

This work would not have been possible without the support and encouragement of many people in my life.

First, I wish to thank my thesis advisor Jennifer Fung for her mentorship throughout my graduate career. This thesis would not have been brought to fruition without her tremendous help and guidance. I especially thank her for providing the resources required to complete the expensive microarray and sequencing experiments in this thesis, and to allow me the independence I needed to conduct my research in the second half of my graduate education. I would like to thank the members of my thesis committee, Joseph DeRisi, Hiten Madhani, and Wallace Marshall, for all their advice and support. Joe, in particular, provided helpful insights on the deep sequencing project and set up me to members of his lab when I had technical questions. To the members of my qualifying exam committee, David Morgan, Joseph DeRisi, Hiten Madhani, and Geeta Narlikar, and my first year rotation advisors, Carol Gross, Anita Sil, and Jeff Cox, I extend my appreciation for your guidance during my first year of graduate school.

I want to thank the members of Fung lab for their technical, intellectual and moral support throughout the years. I especially like to thank Carol Anderson for the generous emotional and moral support and for answering my countless “what do you think I should do” questions. Those conversations were vital to my well-being towards the last year of my graduate school. I thank Ashwini Oke for all her assistance with the sample preparation and bioinformatics work on the deep sequencing project. I owe much to her for being able to graduate on time. In addition to helping me with science, her gentle and

reassuring presence in the lab had always provided a calming effect on me. I want thank Jay Sandler for performing countless hours of yeast tetrad dissection and SNP genotyping PCRs that directly supported my research.

I want to thank Elizabeth Blackburn and the members of her lab for taking me under their wings and providing me with the support and resources of a second lab. Thank you for all your technical advice and intellectual input. I will miss the many engaging and heated lunchroom discussions that ranged from politics and religion to parenthood.

I thank all my classmates for helping me get through the first year of at UCSF. I especially like to thank Tet Matsuguchi, who has helped me troubleshoot so many of my failed experiments, and for being there for me during many tough times. I want to thank Claire Rowe and Erin Quan Toyama for been wonderful roommates and for introducing me to American pop culture via their favorite TV shows. I would like to thank Michelle Dimon for collaborating with me on the deep sequencing project, and for her responsiveness and patience with all my emails and questions. To Sue Adams and Danny Dam of the Tetrad Graduate Program, thank you for taking care of all the administrative details, making sure that we didn't miss the important deadlines, and for helping us get through each program requirement for graduation. You guys are simply the best.

I want to thank my teachers and friends at the Dharma Realm Buddhist Association, Berkeley Buddhist Monastery, and the Dharma Realm Buddhist Young Adults. Thank you for providing an inspiring, nourishing, positive, and supportive spiritual community for practice and study. I would like to thank Reverend Heng Sure and Marty Verhoeven as I would not be writing this acknowledgement today had they not

encouraged me to finish when I wanted to give up. Special thanks go to Marty Verhoeven, for the hundreds of Friday night meditation classes that have been integral to my spiritual growth and well-being. No words could express my gratitude towards my late spiritual teacher, Venerable Master Hsuan Hua. His compassion and wisdom inspired me to never stop trying to be the best human being I could possibly be. Your presence has been sorely missed by all of us.

I would not be the person I am today without the love of my parents and the sacrifices they have made in sending me and my brothers to America with the dream of giving us the best education and opportunities in life. I am grateful for your support, understanding, and enduring love.

Finally, I want to thank my dear friend and partner, Loc Huynh, whose love and encouragements have sustained me throughout my graduate years.

Acknowledgements of Published Materials and Collaborators

Chapter 2

Chapter 2 was originally written and submitted for publication to the Methods of Molecular Biology, DNA Recombination series, as “Mapping of Crossover Sites Using DNA Microarrays”. A revised manuscript was submitted to the series editor on November 24, 2009. At the time of completion of this thesis, the date of publication was yet unknown and a formal request for reprint was submitted to the Springer Permissions Department, Humana Press. A conditional permission for reprint was granted through an email correspondence with the managing editor of the Humana Press, Patrick Marton. The text of the email is provided on page viii.

Chapter 4 and Appendix 1

Chapter 4 was previously published as “Global Analysis of the Meiotic Crossover Landscape.” Stacy Y. Chen, Tomomi Tsubouchi, Beth Rockmill, Jay S. Sandler, Daniel R. Richards, Gerben Vader, Andreas Hochwagen, G. Shirleen Roeder and Jennifer C. Fung. Developmental Cell, September 2008, volume 15, pages 401-15. It was also published online on August 7th, 2008 in advance of the print journal. The supplemental materials of the paper are included as Appendix 1 of this thesis. The paper is reprinted here with the permission of Elsevier. The permission for reprint is provided on page ix. Tomomi Tsubouchi constructed and dissected the *zip4* mutants, Beth Rockmill performed the experiments in Figures 6C, 6D, and 6E, and Jay Sandler performed some of the tetrad dissections in Table S4. Daniel Richards wrote the computer software Allelescan used to

analyze microarray data. Gerben Vader performed the Southern blot analysis in Figure 6F and S5. Andreas Hochwagen supervised Gerben Vader, and Shirleen Roeder supervised Tomomi Tsubouchi and Beth Rockmill. Jennifer Fung directed and supervised the research.

Chapter 5

Michelle Dimon performed the analysis of the Solexa sequencing data and the detection of SNPs and indels. Ashwini Oke assisted with the sample preparation of *sgs1* and *pCLB2-MMS4* mutants and manually annotated the SNPs between S96 and YJM789 near chromosome ends. Joseph DeRisi supervised Michelle Dimon. Jennifer Fung supervised and directed the research.

Gmail by Google

Stacy Chen <stacychen@gmail.com>

Request for permission to include MiMB paper as part of PhD thesis

Marton, Patrick, Springer US <Patrick.Marton@springer.com> **Tue, Jan 5, 2010 at 8:06 AM**
To: Stacy Chen <stacychen@gmail.com>

Hi Stacy:

You'll need the formal permission from the Springer permissions department which can take a long time. In the meantime, go ahead and use this in your PhD dissertation. Just make sure to note where this was originally published.

Best,

Patrick

—

Patrick J. Marton

Managing Editor, Springer Protocols

Humana Press

—
233 Spring Street

New York| NY | 10013-1578 | USA
tel: + 1 212-620-8415

fax: +1 212-463-0742

Patrick.Marton@Springer.com

From: Stacy Chen [mailto:stacychen@gmail.com]
Sent: Monday, January 04, 2010 11:21 PM
To: Marton, Patrick, Springer US
Subject: Re: Request for permission to include MiMB paper as part of PhD thesis

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 14, 2009

This is a License Agreement between Stacy Y Chen ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Stacy Y Chen
Customer address	2295 McGee Ave Berkeley, CA 94703
License Number	2327731330307
License date	Dec 14, 2009
Licensed content publisher	Elsevier
Licensed content publication	Developmental Cell
Licensed content title	Global Analysis of the Meiotic Crossover Landscape
Licensed content author	Stacy Y. Chen, Tomomi Tsubouchi, Beth Rockmill, Jay S. Sandler, Daniel R. Richards, Gerben Vader, Andreas Hochwagen, G. Shirleen Roeder and Jennifer C. Fung
Licensed content date	16 September 2008
Volume number	
Issue number	
Pages	0
Type of Use	Thesis / Dissertation
Portion	Full article
Format	Both print and electronic
You are an author of the Elsevier article	Yes
Are you translating?	No
Order Reference Number	
Expected publication date	Jan 2010

Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
Value added tax 0.0%	0.00 USD
Total	0.00 USD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and
he permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx , As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com

All content posted to the web site must maintain the copyright information line on the bottom of each image
You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the

material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**: None

v1.6

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 30 days of the license date. Payment should be in the form of a check or money order referencing your account number and this license number 2327731330307.
If you would prefer to pay for this license by credit card, please go to http://www.copyright.com/creditcard to download our credit card payment authorization form.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2327731330307 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Genome-wide Analysis of Yeast Meiotic Recombination Landscape

By Stacy Yen-chun Chen

Abstract

At the heart of meiosis is meiotic recombination where programmed double-strand breaks are repaired into either crossovers (COs) or noncrossovers (NCOs). COs promote successful chromosome segregation during the first meiotic division by establishing chiasmata, which are physical connections between homologous chromosomes that provide the tension to properly align chromosomes on the meiosis I spindle. Homologs lacking COs may result in nondisjunction, leading to aneuploid gametes. The number and distribution of COs are tightly regulated to ensure a successful meiotic division. Despite the importance of COs, the mechanisms underlying CO control remain elusive, largely due to the difficulty in determining CO distribution on a genome-wide level.

In this thesis, we describe two methods for mapping the distribution of COs and NCOs genome-wide using two polymorphic *Saccharomyces cerevisiae* strains, S96 and YJM798. First, we used DNA microarrays to identify ~8000 polymorphic markers in the progeny of S96 and YJM789. Eight meiotic mutants were studied: *zip1*, *zip2*, *zip3*, *zip4*, *msh4*, *spo16*, *ndj1*, and *sgs1*. We demonstrated that many aspects of the CO behavior—such as CO level, CO interference, CO homeostasis, chromatid interference, and the behavior of COs near centromeres and telomeres—could be evaluated simultaneously using this method. We showed for the first time that CO homeostasis occurred in wild-type strains. We also identified Zip1 as important for CO suppression at the centromeres.

Using next-generation sequencing, we identified ~54,000 markers and studied the recombination landscape in wild-type and three meiotic mutant tetrads: *msh4*, *sgs1*, and *pCLB2-MMS4*. We demonstrated that next-generation sequencing is a powerful tool for mapping the genome-wide landscape of meiotic recombination events. When coupled with multiplexing, sequencing drastically reduces the cost to lower than that of microarrays, making it possible for large scale experiments involved in studying meiotic mutants. We showed that complex gene conversion motifs near sites of crossing over could be identified and used to unlock the molecular mechanisms and regulations that govern the distribution and formation of recombination events. This technique will prove to be an invaluable contribution to the meiosis field and will help advance our understanding of meiotic recombination in the near future.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Meiosis and Meiotic Recombination

Meiosis marks one of the greatest evolutionary breakthroughs of life on earth. Prior to meiosis, life was perpetuated by identical duplication of the genetic materials from one cell to the next. The advent of meiosis presented the first biological process where genetic materials are reliably reshuffled and recombined in a controlled and highly orchestrated manner to ensure genetic diversity in future generations.

During meiosis, cells undergo one round of DNA replication followed by two rounds of cellular divisions (meiosis I and II), producing four haploid gametes from one diploid parent. A series of events is carefully choreographed to ensure the successful and faithful segregation of genetic materials in each meiotic division (Roeder, 1997; Zickler and Kleckner, 1999). A key event in this process is meiotic recombination, which initiates with the induction of programmed double-strand breaks (DSBs) across the whole genome by the Spo11 protein (Keeney et al., 1997). DSBs are repaired by recombining with a homologus chromosome into one of the two recombinant molecules: crossovers (COs) or noncrossovers (NCOs). COs involve reciprocal exchange of genetic information between homologous chromosomes, which leads to an exchange of the flanking markers between two parents/homologs (Paques and Haber, 1999). A NCO is a type of gene conversion that involves nonreciprocal transfer of genetic information from one homolog to another. NCOs result in a non-Mendelian (non-2:2) segregation of alleles in the four meiotic progeny.

Why would the cell go through such an elaborate and expensive process to generate and then repair DSBs? There are at least two functions of meiotic recombination. First, recombination promotes greater genetic diversity by producing new combinations of parental alleles (Whitby, 2005). Second, crossing over between homologous chromosomes establishes chiasmata, which provide the physical linkages that facilitate proper alignment of homologs on the first meiotic spindle and ensure proper chromosome segregation in meiosis I (Roeder, 1997). Mutations that result in aberrant number and/or distribution of COs have been linked to chromosome nondisjunction and incidences of aneuploidy (Hassold, 2007). In humans, aneuploidy causes infertility, miscarriage, and various developmental defects.

Therefore, understanding the mechanism involved in crossing over has been an area of intense study in the field of meiosis. COs are controlled on several levels: 1) the distribution of COs along and among the chromosomes is affected by local recombination hotspots and coldspots (Blitzblau et al., 2007; Buhler et al., 2007; Gerton et al., 2000), 2) the spacing of COs is controlled by interference which also plays a role in distributing COs (Hillers, 2004), 3) the overall number of COs is regulated by adjusting the ratio of COs to NCOs via a mechanism called “homeostasis” to ensure the formation of sufficient number of COs in each meiosis (Martini et al., 2006), and 4) multiple pathways are involved in promoting CO resolution from DSBs (Martinez-Perez and Colaiacovo, 2009; Whitby, 2005).

Distribution of DSBs

As the precursor of COs, DSBs are not distributed evenly throughout the eukaryotic genome. Global mapping of DSBs revealed that DSBs occur in regions of hotspots and coldspots. By accumulating unresected DSBs in a *rad50S* background, DSBs were found to be enriched in promoter-containing intergenic regions and absent near the centromeres and telomeres (Gerton et al., 2000). Buhler et al. mapped DSBs in *dmc1* mutants, where DSBs are resected into single-stranded tails but remain unrepaired, and reported the surprising result that DSB hotspots are also found near centromeres, where genetic studies have indicated that CO formation is usually repressed (Buhler et al., 2007; Lambie and Roeder, 1986). Although crossing over is cold at telomeres, Blitzblau et al. reported DSB enrichment in regions 20-120 kb near the telomeres (Blitzblau et al., 2007). This hint at the possibility of a telomere-based mechanism where elevated COs formations in subtelomeric regions may contribute to the formation of obligate COs (Blitzblau et al., 2007). In general, although CO nonuniformity parallels DSB nonuniformity along the chromosome, exact overlap does not occur. How CO control results in the CO landscape deviating from the DSB landscape is a major avenue of pursuit in the field.

CO Interference

During the early part of the twentieth century, while constructing the genetic linkage map for *Drosophila,* Muller reported the observation that the occurrence of one crossing over event “interferes” with the occurrence of another crossing over event on the same chromosome, and coined the phenomenon with the term “interference” (Muller, 1916). Since that time, CO interference has been shown to exist in many eukaryotes (Hillers,

2004; Jones, 1984). Over the years, two important features of CO interference emerged. First, each chromosome has at least one CO, termed “obligate CO”, to ensure proper segregation in every chromosome. Second, COs are distributed non-randomly along the length of the chromosome such that two COs seldom occur in close proximity. The second feature presents an intriguing regulatory mechanism in which the absolute positioning of COs are not fixed, but the relative positions between COs in each meiosis is controlled with more regular spacing than expected randomly. Loss of interference, even in the presence of normal CO frequencies, will result in chromosome missegregation.

The precise mechanism of CO interference is unknown. However, several models have been proposed. One simple model predicts that anti-crossover signals are transmitted from initial sites of CO along the chromosomes such that no additional COs can be established in the immediate vicinity (Maguire, 1988). The inhibitory signals would be strongest near the site of initial COs and decrease proportionally as signals travel farther from the original CO site. A second model predicts that each CO is separate by a fixed number of NCOs, suggesting that the cell is somehow able to “count” the number and the types of recombination events (Foss et al., 1993). Yet, another model suggests that meiotic chromosomes undergo intrinsic mechanical stress as a result of chromatin expansion during meiotic division (Kleckner et al., 2004). The formation of a CO would relieve the stress along the chromosome bi-directionally in regions immediately adjacent to the CO, obviating the need for additional COs to occur nearby. Although many models attempt to explain the mechanism involved in implementing CO interference, the exact mechanism is still unknown.

Several challenges are involved in studying CO interference (Hillers, 2004). First, since interference is a measure of regulation between COs, recombination events need to be measured in multiple adjacent intervals along the chromosome to assess the extent of CO interference. Ideally, an assay that can measure genome-wide recombination map of a single meiosis is required to truly measure the global effect of interference. Second, since interference is a probabilistic effect, a large number of recombination data needs to be gathered to measure interference. These challenges need to be addressed before real progress in our understanding of interference can be made.

CO Homeostasis

The cell not only controls the distribution of COs throughout the genome, the number of COs is also regulated. Martini et al. reported the discovery of CO homeostasis, where the overall number of COs is modulated to ensure sufficient numbers of COs are formed (Martini et al., 2006). Normal levels of COs are maintained despite lowering the overall number of DSB-initiating events. A series of *spo11* mutant alleles were used, where the level of DSBs are ~80%, ~30%, and ~20% of the wild type. The authors observed that COs tend to be maintained at the expense of NCOs, suggesting that the cells possess an internal counter to measure the overall number of COs in the genome. Presumably, CO homeostasis reduces the chance of nondisjunction by ensuring that sufficient numbers of COs are made. The mechanism involved in establishing CO homeostasis and the relationship it has to CO interference remain unknown.

The Double-Strand Break Repair Model

How do cells repair DSBs during meiosis? A canonical model is the DSB repair model proposed by Szostak et al. (Sun et al., 1991; Szostak et al., 1983). The model suggests that recombination begins by the formation of DSBs. Degradation occurs at the 5'ends of each broken ends and processes each end into a single-stranded 3' tail. One of the single-stranded 3' tail then invades a homologous chromosome to form a displacement loop (D-loop). The single-end invasion results in a patch of heteroduplex DNA. DNA synthesis extends the invading strand using the recipient strand as a template, effectively enlarging the D-loop. The enlarged D-loop "captures" and anneals the single-stranded tail of the other broken DSB end, forming a second patch of heteroduplex DNA. DNA synthesis repairs the gap on the second broken end using the D-loop as the template. Ligation of the new-synthesized strands with the broken strands connects the two homologs into a joint molecule known as the double Holliday junctions. The double Holliday junctions can be cleaved in two different manners to produce either a CO or a NCO.

Evidence for many intermediates proposed by the DSB repair model has been observed. The double Holliday junctions predicted by the DSB repair model has been observed in budding yeast using electron microscopy (Bell and Byers, 1979). It can also be visualized using two-D gel analysis at specific loci (Allers and Lichten, 2001b; Schwacha and Kleckner, 1995). Physical evidence of the presence of heteroduplex DNA near sites of DSBs were reported by Allers and Lichten (2001), while Hunter and Kleckner reported evidence for single-end invasion molecules (Hunter and Kleckner, 2001).

Synthesis-dependent Strand Annealing

The DSB repair model predicts that the double Holliday junction is an intermediate of both COs and NCOs. However, evidence from physical assays suggested otherwise. Allers and Lichten have shown that the NCOs are detected at the same time as the double Holliday junctions, and that the formation of COs occurs after the formation of both the NCOs and the double Holliday junctions (Allers and Lichten, 2001a). In *ndt80* mutants, where cells arrest in meiosis with unresolved Holliday junctions, only CO levels are reduced while the NCO levels remain unaffected. This evidence points to the proposal that COs and NCOs are regulated differently, with double Holliday junctions as the primary intermediate for the formation of COs and not NCOs. It is now widely accepted that NCOs are formed through a mechanism called synthesis-dependent strand annealing, where after single-end invasion and synthesis of the nascent DNA, dissociation of the D-loop occurs (Whitby, 2005). The invading strand with the nascent DNA then anneals back to the other DSB end and repair the DSB gap into NCOs.

The Two Pathways of CO Resolution

What molecular mechanisms are involved in CO formation? A search for proteins involved in CO formation has led to the proposal that more than one pathway exists for the repair of DSBs into COs (Argueso et al., 2004; Martinez-Perez and Colaiacovo, 2009; Whitby, 2005). A set of meiosis-specific proteins (Zip1, Zip2, Zip3, Mer3, Msh4, and Msh5), collectively referred to as the ZMM proteins, promote single-end invasion and double Holliday junction formation (Borner et al., 2004). In the absence of ZMMs, single-stranded intermediates become transient and unstable and can no longer be

detected on the two-dimensional gels. CO formations are also reduced but not abolished in these mutants.

The formation of most non-ZMM COs depends on Mus81 and Mms4 proteins. However, the Mus81/Mms4-dependent COs do not resolve through double Holliday junction intermediates as proposed in the DSB repair model. De los Santos et al. have reported that the number of double Holliday junctions is reduced in *mms4* mutants (de los Santos et al., 2003). Using physical assays and electron microscopy, Cromie et al. have reported that single Holliday junctions, not double Holliday junctions, are the predominant intermediate in fission yeast *Schizosaccharomyces pombe*, where the majority of COs are formed through the Mus81/Mms4-dependent pathway (Cromie et al., 2006).

Another line of evidence for multiple CO pathways comes from the study of CO interference. CO interference ensures that COs are evenly distributed along the chromosomes. CO interference depends on the recruitment of ZMM proteins to sites of DSBs, and *zmm* mutants display a loss of interference in the remaining COs that are presumably formed through the Mus81/Mms4 pathway (Martinez-Perez and Colaiacovo, 2009; Whitby, 2005). On the other hand, Mus81/Mms4-dependent COs are not subject to CO interference, and CO interference persists in *mus81* and *mms4* mutants (de los Santos et al., 2003). However, residual levels of COs are still detected in *mms4* and *msh5* (a ZMM protein) double mutants, suggesting that there may be yet another undiscovered third pathway for CO resolution in budding yeast (Argueso et al., 2004; de los Santos et al., 2003).

Anti-crossover Agents: Sgs1 and Mus81

In addition to CO-promoting pathways, anti-CO agents have also been found. Sgs1, a RecQ family helicase, has long been implicated to suppress CO formation in both mitotic and meiotic cells. *sgs1* mutants display an elevated level of COs without affecting the level of gene conversions (Rockmill et al., 2003). However, elimination of Sgs1 rescues the CO defect in *zmm* mutants. The antagonistic relationship between Sgs1 and ZMM proteins led to the model that ZMMs act to stabilize early recombination intermediates and protects it from the anti-CO activities of Sgs1.

Oh and colleagues have shown that Sgs1 prevents the formation of aberrant multi-chromatid joint molecules that are formed from both ends of the DSB invading a different homologous chromosome (Oh et al., 2007). Hence, Sgs1 functions to prevent the formation of closely spaced COs that may impede successful CO segregation. Recently, two studies have shown that Mus81 also plays a role in collaborating with Sgs1 in resolving aberrant joint molecules (Jessop and Lichten, 2008; Oh et al., 2008). Using an inducible expression system in a *sgs1 mus81* double mutant, Jessop and Lichten reported that Sgs1 expression reduced the extent of joint molecule formation and that Mus81 expression restored joint molecule resolution and CO formation (Jessop and Lichten, 2008). These observations showed that Sgs1 and Mus81 collaborate to eliminate joint molecules and promote chromosome segregation.

Overview of the Work Described in This Thesis

This thesis was motivated by the need to develop an assay to assess the many levels of crossover control a genome-wide level. At the start of this thesis, Winzeler et al. (1998)

described a method of globally mapping CO events in budding yeast using the method of direct allelic scanning, taking advantage of polymorphisms between two yeast strains, S96 (a S288c derivative) and YJM789 (a clinical isolate from the lungs of an immuno-compromised patient). At the time, the genome of YJM789 was not yet completely sequenced and assembled, and a list of polymorphisms between the S288c and YJM789 was not yet available, so direct sequencing could not be used to pinpoint polymorphisms as a means to determine the parental origin of DNA isolated from various strains. Therefore, the Affymetrix S98 expression array was used to detect the differences in hybridization between the two yeast strains; probe sequences that exhibited clear differences could be used as markers to follow the inheritance of polymorphisms in the progeny. Chapter 2 of this thesis describes in detail the step-by step procedure involved in preparing the genomic DNA for microarray hybridizations. This chapter was written and submitted for publication to the Methods of Molecular Biology, DNA Recombination series.

We detected around 8,000 polymorphic markers between the S96 and YJM789 using DNA microarrays. The number of markers is significant compared to traditional genetic techniques in which less than a dozen genetic markers are scored in each experiment. New bioinformatics tools had to be developed to identify CO and gene conversion events from the 8,000 marker profiles of the four-spore tetrads. Chapter 3 describes the programming strategy employed in developing the Python-based program, CrossOver. CrossOver identifies and computes the location of eight types of CO motifs and seven types of gene conversion tracts. Functions in CrossOver that compute various analyses of the meiotic recombination events and control are also outlined and explained.

Chapter 4 is the reprint of a paper we published, reporting the global analysis of meiotic CO landscape in wild type meiosis and in eight meiotic mutants (Chen et al., 2008). The CO and NCO levels near centromeres and telomeres were examined, and CO interference was calculated for wild type and all mutants using inter-CO distances and gamma distribution function. CO homeostasis was found to be reduced in *zip2* and *zip4* mutants. We also observed that the suppression of COs and NCOs in pericentrimeric regions are abolished in *zip1* mutant. The published supplementary materials are included in Appendix 1 of this thesis.

As the YJM789 genome sequence was completed and alignment to the S288c genome was analyzed, as much as 60,000 SNPs and 6,000 indels were found between the two strains (Wei et al., 2007). With the introduction of next-generation sequencing, we set out to map the recombination products between the two strains using deep sequencing in order to obtain the resolution needed to determine whether COs leave identifiable motifs that can pinpoint the type of recombination pathway that lead to a CO's formation. Chapter 5 reports the assay we have developed for mapping polymorphic markers between S96 and YJM789 using deep sequencing and multiplexing. We report the results from sequencing two wild type tetrads and one *msh4*, *sgs1*, and *pCLB2-MMS4* mutant tetrads. Finally, chapter 6 reviews the findings described in this thesis and proposes future directions that stem from the work described here.

References

Allers, T., and Lichten, M. (2001a). Differential timing and control of noncrossover and crossover recombination during meiosis. Cell *106,* 47-57.

Allers, T., and Lichten, M. (2001b). Intermediates of yeast meiotic recombination contain heteroduplex DNA. Mol Cell *8*, 225-231.

Argueso, J.L., Wanat, J., Gemici, Z., and Alani, E. (2004). Competing crossover pathways act during meiosis in Saccharomyces cerevisiae. Genetics *168*, 1805-1816.

Bell, L., and Byers, B. (1979). Occurrence of crossed strand-exchange forms in yeast DNA during meiosis. Proc Natl Acad Sci USA *76*, 3445-3449.

Blitzblau, H.G., Bell, G.W., Rodriguez, J., Bell, S.P., and Hochwagen, A. (2007). Mapping of Meiotic Single-Stranded DNA Reveals Double-Strand-Break Hotspots near Centromeres and Telomeres. Curr Biol.

Borner, G.V., Kleckner, N., and Hunter, N. (2004). Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell *117*, 29-45.

Buhler, C., Borde, V., and Lichten, M. (2007). Mapping Meiotic Single-Strand DNA Reveals a New Landscape of DNA Double-Strand Breaks in Saccharomyces cerevisiae. PLoS Biol *5*, e324.

Chen, S.Y., Tsubouchi, T., Rockmill, B., Sandler, J.S., Richards, D.R., Vader, G., Hochwagen, A., Roeder, G.S., and Fung, J.C. (2008). Global analysis of the meiotic crossover landscape. Dev Cell *15*, 401-415.

Cromie, G.A., Hyppa, R.W., Taylor, A.F., Zakharyevich, K., Hunter, N., and Smith, G.R. (2006). Single Holliday junctions are intermediates of meiotic recombination. Cell *127*, 1167-1178.

de los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J., and Hollingsworth, N.M. (2003). The Mus81/Mms4 endonuclease acts independently of double-Holliday junction

resolution to promote a distinct subset of crossovers during meiosis in budding yeast. Genetics *164*, 81-94.

Foss, E., Lande, R., Stahl, F.W., and Steinberg, C.M. (1993). Chiasma interference as a function of genetic distance. Genetics *133*, 681-691.

Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O., and Petes, T.D. (2000). Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA *97*, 11383-11390.

Hassold, T. (2007). The origin of human aneuploidy: where we have been, where we are going. Hum Mol Genet *16*, 203-208.

Hillers, K.J. (2004). Crossover interference. Curr Biol *14*, R1036-1037.

Hunter, N., and Kleckner, N. (2001). The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. Cell *106*, 59-70.

Jessop, L., and Lichten, M. (2008). Mus81/Mms4 endonuclease and Sgs1 helicase collaborate to ensure proper recombination intermediate metabolism during meiosis. Mol Cell *31*, 313-323.

Jones, G.H. (1984). The control of chiasma distribution. In Controlling Events in Meiosis, C.W. Evans, and H.G. Dickinson, eds. (Cambridge, The Company of Biologists Ltd.), pp. 293-320.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell *88*, 375-384.

Kleckner, N., Zickler, D., Jones, G.H., Dekker, J., Padmore, R., Henle, J., and Hutchinson, J. (2004). A mechanical basis for chromosome function. Proc Natl Acad Sci U S A *101*, 12592-12597.

Lambie, E.J., and Roeder, G.S. (1986). Repression of meiotic crossing over by a centromere (CEN3) in Saccharomyces cerevisiae. Genetics *114*, 769-789.

Maguire, M.P. (1988). Crossover site determination and interference. J Theor Biol *134*, 565-570.

Martinez-Perez, E., and Colaiacovo, M.P. (2009). Distribution of meiotic recombination events: talking to your neighbors. Curr Opin Genet Dev *19*, 105-112.

Martini, E., Diaz, R.L., Hunter, N., and Keeney, S. (2006). Crossover homeostasis in yeast meiosis. Cell *126*, 285-295.

Muller, H. (1916). The mechanism of crossing over. Amer Nat *50*, 193-434.

Oh, S.D., Lao, J.P., Hwang, P.Y., Taylor, A.F., Smith, G.R., and Hunter, N. (2007). BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. Cell *130*, 259-272.

Oh, S.D., Lao, J.P., Taylor, A.F., Smith, G.R., and Hunter, N. (2008). RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. Mol Cell *31*, 324-336.

Paques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev *63*, 349-404.

Rockmill, B., Fung, J.C., Branda, S.S., and Roeder, G.S. (2003). The Sgs1 helicase regulates chromosome synapsis and meiotic crossing over. Curr Biol *13*, 1954-1962.

Roeder, G.S. (1997). Meiotic chromosomes: it takes two to tango. Genes Dev *11*, 2600-2621.

Schwacha, A., and Kleckner, N. (1995). Identification of double Holliday junctions as intermediates in meiotic recombination. Cell *83*, 783-791.

Sun, H., Treco, D., and Szostak, J.W. (1991). Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the *ARG4* recombination initiation site. Cell *64*, 1155-1161.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25-35.

Wei, W., McCusker, J.H., Hyman, R.W., Jones, T., Ning, Y., Cao, Z., Gu, Z., Bruno, D., Miranda, M., Nguyen, M., *et al.* (2007). Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. Proc Natl Acad Sci U S A *104*, 12825-12830.

Whitby, M.C. (2005). Making crossovers during meiosis. Biochem Soc Trans *33*, 1451-1455.

Winzeler, E.A., Richards, D.R., Conway, A.R., Goldstein, A.L., Kalman, S., McCullough, M.J., McCusker, J.H., Stevens, D.A., Wodicka, L., Lockhart, D.J., *et al.* (1998). Direct allelic variation scanning of the yeast genome. Science *281*, 1194-1197.

Zickler, D., and Kleckner, N. (1999). Meiotic chromosomes: integrating structure and function. Annu Rev Genet *33*, 603-754.

Chapter 2: Sample Preparation for Mapping of Crossover Sites Using DNA Microarrays

Stacy Y. Chen, Jennifer C. Fung

Department of Biochemistry and Biophysics

University of California, San Francisco, San Francisco, California, USA

Abstract

Crossovers (COs) play an essential role in promoting successful chromosome segregation during meiosis. Crossing over generates chiasmata, which are physical bridges between homologs that provide the appropriate tension to properly align chromosomes on the meiosis I spindle. Homolog pairs that fail to cross over can result in meiosis I nondisjunction, leading to aneuploid gametes. Therefore, the number and distribution of crossovers are tightly regulated to ensure that each chromosome pair receives at least one CO. Here, we describe a DNA microarray-based method to map CO distribution genome-wide, on a cell-by-cell basis, allowing for rapid and accurate analysis of multiple aspects of CO control.

1. Introduction

Meiosis is the beginning stage of sexual reproduction during which one diploid parent undergoes two rounds of cellular division to produce four haploid progeny (*1, 2*). Recombination between homologous chromosomes during the first meiotic division is essential for successful chromosome segregation. Meiotic recombination leads to the

formation of crossovers (COs) and noncrossovers (NCOs) (*3*). Crossing over creates chiasmata, which are interhomolog associations that provide the necessary tension to correctly align homologs on the meiosis I spindle. Defects in crossing over lead to meiosis I nondisjunction, resulting in the production of aneuploid gametes (*4*).

To ensure that each pair of homologous chromosomes receives at least one CO, the spatial distribution and the number of COs is highly orchestrated. Some examples of CO control include CO interference and CO homeostasis. In most eukaryotes, CO interference regulates the spatial positioning of COs along a chromosome such that a CO event in one region reduces the likelihood of another one occurring nearby (*5*-*7*). This results in a nonrandom and more evenly spaced distribution of COs across the genome where the strength of interference diminishes as a function of distance.

CO homeostasis controls the number of COs in a single meiosis, whereby the normal level of COs is maintained despite fluctuations in the overall number of recombination-initiating events (*8*). Martini et al. observed that when the overall recombination-initiating events are reduced, CO levels are maintained at the expense of NCOs. CO homeostasis may function to reduce the occurrence of nonexchange chromosomes by ensuring that a sufficient number of COs are made.

One major difficulty in understanding CO control in meiosis has been the lack of an efficient and accurate method for determining CO distribution genome-wide and on a cell-by-cell basis. Here, we describe a microarray-based approach for mapping CO distribution using the method of direct allelic variation scanning of the genome that has been adapted to analyze multiple aspects of CO control (*see* **Note 1**) (*9, 10*). This method identifies sequence polymorphisms between two strains of yeast *Saccharomyces*

cerevisiae—S96 and YJM789. Using the polymorphic markers, the parental origin of the meiotic progeny at each of the detectable sequence polymorphic loci is determined. The reciprocal CO events (and a subset of NCOs and gene conversions) can be mapped by following the inheritance pattern of allelic markers in the four haploid progeny strains. Multiple aspects of the CO landscape can thus be analyzed, including the genome-wide interference level, which can be calculated using the distribution of distances between adjacent COs and the gamma distribution function (*11, 12*), as well as CO homeostasis, which can be determined by the correlation between the number of COs and NCOs for each meiotic event (*10*).

2. Materials

2.1. Isolation of Four-Spore Viable Tetrads

1. *Saccharomyces cerevisiae* strains: S96 (*MAT***a** *ho lys5*) and YJM789 (*MATα ho::hisG lys2 cyh*). Strains were obtained from Dr. Elizabeth Winzeler of the Scripps Research Institute, La Jolla, CA, USA.
2. YPAD plates: dissolve 20 g dextrose, 20 g bactopeptone, 10 g yeast extract, and 20 g agar in water to a final volume of 940 mL. Sterilize by autoclaving. Add 50 mL of 10 m*M* sterile adenine solution and 10 mL of 20 m*M* sterile uracil solution. Pour 20 mL into each Petri plate. Allow media to cool and solidify at room temperature.
3. Amino acid mix: 2.4 g adenine, 21.0 g arginine, 6.0 g glutamic acid, 2.6 g histidine, 2.4 g inositol, 31.2 g isoleucine, 15.8 g leucine, 5.4 g lysine, 9.0 g methionine, 4.8 g phenylalanine, 6.6 g serine, 7.2 g threonine, 4.8 g tryptophan,

1.2 g tyrosine, 1.4 g uracil, and 7.2 g valine.

4. Sporulation plates: dissolve 2 g yeast extract, 1 g dextrose, 20 g potassium acetate, 1 g amino acid mix, and 20 g agar in water to a final volume of 1 L. Sterilize by autoclaving. Pour into Petri plate as described for YPAD plates.
5. Zymolyase 100T (Seikagaku Corporation, Tokyo, Japan) stock solution is prepared at 10 mg/mL in 5% (w/v) dextrose solution and is stored in single use aliquots at -20ºC.
6. Ascus digestion solution is freshly prepared each time from the Zymolyase stock solution to a working solution of 0.05 mg/mL in 1 *M* sorbitol.

2.2. Allele-specific Primer Extension Colony PCR of Four-Spore Viable Tetrads

1. Overnight yeast cultures of all spores from four-spore viable tetrads which had been grown on a YPAD plate.
2. 0.02 *M* NaOH solution.
3. PCR primers (**Table 2-1**; synthesized by Integrated DNA Technologies, Carolville, IA): standard desalted and resuspended in sterile ddH_2O to 100 μ*M* stock solution. Store at -20˚C.
4. *Taq* PCR Core Kit (Qiagen, Germantown, MD), specifically: *Taq* DNA polymerase (5 unit/μL), CoralLoad PCR buffer (10x), Q-solution (5x), and dNTP mix (10 m*M* of each dNTP). Store at -20˚C.
5. DNA HyperLadder^{TM} IV (Bioline, Taunton, MA). Store at 4˚C.
6. 1x Tris/Acetate/EDTA (TAE) buffer: 40 m*M* Tris-acetate, pH 8.5, 2 m*M* EDTA .
7. Agarose gel: 1.5% (w/v) agarose, 1x TAE buffer, 0.5 μg/mL ethidium bromide.

2.3. Isolation of Yeast Genomic DNA

1. YPAD media is prepared in a similar procedure as YPAD plates, omitting the agar.
2. TE: 10 m*M* Tris • HCl, pH 8.0, 1 m*M* EDTA, pH 8.0. Store at room temperature.
3. Buffer Y1 (yeast lysis buffer): 1 *M* sorbitol, 100 m*M* EDTA, 14 m*M* β-mercaptoethanol. Store at 2 to 8ºC.
4. Zymolyase 100T is dissolved at 10 mg/mL in 5% (w/v) dextrose solution and is freshly made each time.
5. Buffer G2 (digestion buffer): 800 m*M* guanidine HCl, 30 m*M* Tris • HCl, pH 8.0, 30 m*M* EDTA, pH 8.0, 5% Tween-20, 0.5% Triton X-100. Store at 2 to 8ºC or room temperature.
6. RNaseA is dissolved at 100 mg/mL in 0.01 *M* sodium acetate, pH 5.2, followed by heating at 100ºC for 15 minutes. Allow solution to cool slowly to room temperature before adding 0.1 volume of 1 *M* Tris • HCl, pH 7.4 to adjust the pH. Store in aliquots at -20ºC.
7. Proteinase K (Invitrogen, Carlsbad, CA) is dissolved at 20 mg/mL in sterile ddH_2O and is freshly made each time.
8. Buffer QBT (equilibration buffer): 750 *M* NaCl, 50 m*M* MOPS, pH 7.0, 15% isopropanol, 0.15% Triton X-100. Store at 2 to 8ºC or room temperature.
9. Genomic-tip 500/G (Qiagen, Valencia, CA).
10. Buffer QC (wash buffer): 1.0 *M* NaCl, 50 m*M* MOPS, pH 7.0, 15% isopropanol. Store at 2 to 8ºC or room temperature.
11. Buffer QF (elution buffer): 1.25 *M* NaCl, 50 m*M* Tris • HCl, pH 8.5, 15% isopropanol. Store at 2 to 8ºC or room temperature.

12. Isopropanol.
13. Glass microcapillary pipettes (10 μL) (VWR International, West Chester, PA): Pipettes are sealed on one end by flaming.
14. 70% (v/v) ethanol.

2.4. Fragmentation of DNA Using Deoxyribonuclease I

1. Deoxyribonuclease I (DNase I), Amplification Grade, 1 U/μL (Invitrogen).
2. 10x One-Phor-All Buffer *Plus* (discontinued item from GE Healthcare, Chalfont Saint Giles, United Kingdom): 100 m*M* Tris-acetate, pH 7.5, 100 m*M* magnesium acetate, 500 m*M* potassium acetate. Store at 4ºC.
3. 25 m*M* $CoCl_2$ solution (in package contents of the terminal transferase used for biotinylation in section 2.5).
4. 10,000x SYBR® Green I Nucleic Acid Gel Stain (Invitrogen). Store at 4ºC and shield from light.
5. Agarose gel: 2% (w/v) UltraPure agarose 1000 (Invitrogen), 1x TAE buffer, with SYBR® Green I Nucleic Acid Gel Stain. Shield from light.
6. DNA HyperLadder^{TM} IV (Bioline, Taunton, MA). Store at 4˚C.
7. Loading buffer: 7.5% (v/v) glycerol solution.

2.5. Biotinylation of DNA Fragments and Microarray Analysis

1. Bio-N^6-ddATP, 1 m*M* (Enzo Life Sciences, Inc., Farmingdale, NY).
2. Terminal transference, recombinant, 400 U/μL (Roche, Indianapolis, IN).
3. 12X MES stock solution (1.22 *M* MES, 0.89 *M* [Na+]): Dissolve 70.4 g MES free acid monohydrate and 193.3 g MES Sodium salt in 800 mL sterile ddH_2O. Mix and adjust volume to 1 L. pH should be between 6.5 to 6.7 without adjustments.

Filter sterilize using 0.2 μm filter. Store at 4ºC and shield from light. Discard if solution becomes yellow.

4. 2X hybridization buffer (200 m*M* MES, 2*M* [Na+], 40 m*M* EDTA, 0.02% Tween-20): Mix 8.3 mL 12X MES stock solution, 17.7 mL 5M NaCl, 4.0 mL 0.5 *M* EDTA, pH 8.0, 0.1 mL 10% Tween-20, and add 19.9 mL sterile ddH_2O to bring it to a final volume of 50 mL. Store at 4ºC and shield from light.
5. Herring Sperm DNA, 10 mg/mL, (Promega, Madison, WI).
6. Acetylated BSA, 20 mg/mL, (Invitrogen). Stored at -20ºC.
7. Control Oligo B2, 3 n*M*, (Affymetrix, Santa Clara, CA).
8. GeneChip® Yeast Genome S98 Array (Affymetrix, Santa Clara, CA).
9. Prepare wash buffer A, wash buffer B, Streptavidin Phycoerythrin (SAPE) stain and antibody solutions according to the *GeneChip® Expression Analysis Technical Manual* (*13*).
10. GeneChip® Hybridization Oven 645 (Affymetrix, Santa Clara, CA).
11. GeneChip® Fluidics Station 450 (Affymetrix, Santa Clara, CA).
12. GeneChip® Scanner 3000 (Affymetrix, Santa Clara, CA).
13. Affymetrix® Microarray Suite Software (Affymetrix, Santa Clara, CA).

2.6. Data Analysis using Allelescan and CrossOver Software

1. Cell File Conversion Tool. To download http://www.affymetrix.com/support/developer/tools/affytools.affx. PC only. This software can also be obtained from the Fung lab by emailing jennifer.fung@ucsf.edu.
2. Affymetrix Power Tools. To download:

http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx#1_2.

3. System requirements for Windows: Windows XP or higher, Intel Pentium 4 processor and above, 680 MB of disk space, and 512 MB of RAM.
4. System requirements for Macintosh: Mac OS X 10.5.5 and above, any Intel-based Macs, 360 MB of disk space, and 512 MB of RAM.
5. MATLAB® 6.5 with the Statistics Toolbox™ (MathWorks, Natick, MA).
6. Allelescan (Davis Lab, Stanford University). Requests for software should be directed to Dr. Daniel Richards of Ingenuity Systems, Inc., Redwood City, CA, USA.
7. Python 2.5 or higher (http://www.python.org).
8. CrossOver (Fung Lab, University of California, San Francisco). Requests for software should be directed to Dr. Jennifer Fung at jennifer.fung@ucsf.edu.

3. Methods

S96 and YJM789 haploid parental strains are mated and four meiotic progeny are isolated via tetrad dissection by selecting for those which are four-spore viable. Four-spore viable tetrads are prescreened for possible errors in the selection procedure or abnormal genome-wide missegregation using the allele-specific primer extension colony PCR. Tetrads that show a normal 2:2 segregation of parental alleles in the majority of SNP (single-nucleotide polymorphism) primer sets tested by colony PCR are selected for further microarray analysis.

Genomic DNA is isolated from the parental strains, S96 and YJM789 haploids, and their meiotic progeny, the four-spore viable tetrads. Genomic DNA is fragmented using DNase I, and end-labeled with biotin-N6-ddATP using terminal deoxynucleotidyl transferase before hybridizing to Affymetrix GeneChip® Yeast Genome S98 Array using the GeneChip® Hybridization Oven 645. The arrays are stained with R-phycoerythrin-streptavidin and amplified with biotinylated antistreptavidin antibody using and GeneChip® Fluidics Station 450, and scanned using the laser confocal scanner, GeneChip® Scanner 3000.

Microarray experiment data are analyzed using the software Allelescan, a microarray analysis platform that analyzes genomic DNA hybridization data and identifies sequence polymorphisms between samples from two distinct genetic backgrounds using their differential hybridization signal intensities (*14*). Meiotic progeny generated from the two parental backgrounds can be genotyped at each of the polymorphic markers and a segregation profile is generated for the four-spore tetrad.

CrossOver is developed as a downstream analysis tool to process the segregation profile from Allelescan in order to determine locations of meiotic recombination events. In addition, CrossOver performs various analyses that address questions of particular interest to meiotic recombination. CrossOver can compute crossover densities for each chromosome, occurrence of nonexchange chromosomes, inter-crossover distances, CO-to-centromere and CO-to-closest telomere distances, gene conversion tract lengths, correlation coefficients of the number of COs and NCOs for each meiosis, and parameters of the gamma distribution function for inter-crossover distances (*10*).

3.1. Isolation of Four-Spore Viable Tetrads

1. Streak out *Saccharomyces cerevisiae* yeast strains S96 and YJM789 from frozen stock onto YPAD plates and grow overnight at 30ºC. Select and patch a few single colonies from each parent to proceed with and mate. Yeast mating is most efficient when parent cells are from fresh cultures.
2. Use sterile toothpicks to mix a small, but equal amount of S96 and YJM789 parent cells on a YPAD plate, creating a patch of cells of about 5 mm in diameter. Allow cells to mate at 30ºC for 4 to 6 hours.
3. Transfer the mixture of S96 and YJM789 cells from the YPAD plate to a sporulation plate using a sterile toothpick. Incubate cells at 30ºC for 3 to 5 days until sufficient numbers of tetrads have formed. Cultures with fewer than 1% tetrads are difficult to dissect. Sporulation frequency for 4-spore asci is approximately 15%.
4. Prepare fresh ascus digestion solution from the zymolyase stock solution. To prepare cultures for dissection, resuspend a small dab of cells (about 1 mm in diameter) from the sporulation plate in 50 µL freshly prepared ascus digestion solution. Incubate for 10 minutes at 30ºC. Add 100 µL of sterile water and mix gently.
5. To prepare a dissection plate, hold a fresh YPAD plate at a 45º angle and gently spot 15 µL of zymolyase-treated cells along a line down the center of the plate. Allow the liquid solution to dry on plate.
6. Dissect tetrads on a yeast dissection microscope. Incubate dissected plates at 30ºC. After 3 days at 30ºC, colonies should be of sufficient size to determine viability. Only four-spore viable tetrads are selected for further analysis (*see* **Note**

2).

3.2. Allele-specific Primer Extension Colony PCR of Four-Spore Viable Tetrads

1. **Table 2-1** shows 23 sets of allele-specific PCR primers which display strong allele specificity in allele-specific primer extension PCR. Each primer set assesses SNP genotype at two different SNP positions, approximately 200 bp apart. Three primers are designed for each primer set: 1) a common forward primer that anneals to both the S96 and the YJM789 template; 2) a YJM789-specific reverse primer that anneals to the first SNP approximately 200 bp from the common forward primer; and 3) a S96-specific reverse primer that anneals to the second SNP approximately 400 bp from the common forward primer (*see* **Figure 2-1**). Allele-specific primers are designed to match only one of the two possible SNP allele sequences at the 3'-terminal nucleotide, allowing efficient amplification of the matched SNP nucleotide, but not the mismatched allele. To increase allele specificity, a single-base mismatch is sometimes introduced to both allele-specific primers 3 or 4 bases inward from the 3' end of the primer, causing further destabilization for primers that may have annealed to the wrong allele (*15*). Primer sequences within each primer set were selected to have similar melting temperatures, of approximately 54˚C. The resulting PCR reaction generates a YJM789-specific product of roughly 200 bp and a S96-specific product about 400 bp in length, which can easily be resolved from each other on a 1.5% agarose gel (*see* **Figure 2-2**).
2. To set up allele-specific primer extension colony PCR for one full four-spore tetrad and one primer set (*see* **Note 3**), take a small dab of overnight yeast culture

from each spore of the tetrad using a sterile pipette tip. Resuspend cells from each spore in a separate PCR tube containing 5 µL 0.02 *M* NaOH solution. Transfer cell mixture to a PCR machine and boil for 10 minutes at 99˚C. Cool to 4˚C.

3. Select the primer set to test with by PCR and its corresponding individual primers. Mix equal amounts of 100 µ*M* common forward primer, 100 µ*M* S96-specific reverse primer and 100 µ*M* YJM789-specific reverse primer to create a master primer mix consisting of all 3 primers (*see* **Note 4**).
4. Once the cells from step 2 have cooled to 4˚C, add the following to each of the four spore cell mixtures: 10 µL 5x Q-solution, 5 µL 10x CoralLoad PCR buffer, 1 µL 10 m*M* dNTP mix, 1 µL of master primer mix from step 3, 25.5 µL sterile ddH_2O, and 0.5 µL *Taq* DNA polymerase to a final volume of 50 µL.
5. Run the following PCR program: initial denaturing step at 94˚C for 5 minutes, followed by 35 cycles of denaturing at 94˚C for 30 seconds, annealing at 54˚C for 30 seconds, and extending at 72˚C for 1 minute, and a final extension at 72˚C for 10 minutes.
6. Prepare a 1.5% TAE agarose gel. Run 5 µL of each of the finished PCR reaction along side with 5 µL of HyperLadder^{TM} IV (or any DNA ladder that include the 200 to 400 bp range) until the YJM789-specific band (~200 bp) is resolved from the S96-specific band (~400 bp). Assess the S96 and YJM789 allele segregation of the tetrad (*see* **Note 3** and **Figure 2-2**).

3.3. Isolation of Yeast Genomic DNA

1. Yeast genomic DNA is isolated using the Qiagen Genomic-tip 500/G. The following procedures are adapted from the *Qiagen Genomic DNA Handbook* (*16*).

Using a sterile toothpick, make a circular patch of yeast culture approximately 1 inch in diameter on a YPAD plate. Grow at 30ºC overnight.

2. In a 1 L flask, inoculate the entire patch of overnight yeast culture in 150 mL YPAD liquid media. Grow culture for ~18 hours on platform shaker at 30ºC to a cell density of approximately 3 x 10^8 cells/mL (*see* **Note 5** for alternative inoculation method).
3. Harvest 100 mL of culture by centrifuging at 3000-5000 x g for 5 minutes. Discard the supernatant.
4. Resuspend cell pellet in 12 mL TE buffer and transfer cells to a 50 mL conical tube. Centrifuge again at 3000-5000 x g for 5 minutes to remove residual YPAD media. Discard the supernatant (*see* **Note 6**).
5. Resuspend cell pellet in 12 mL of Buffer Y1. Vortex to resuspend cells thoroughly. Add 1 mL of zymolyase solution. Rotate on roller drum at 30ºC for 1 to 1.5 hours.
6. Following zymolyase digestion, centrifuge cells at 5000 x g for 20 minutes at 4ºC. During centrifugation, add 30 µL of RNase A solution to 15 mL of buffer G2 and prepare fresh Proteinase K solution. Slowly decant supernatant after centrifugation to avoid disturbing the pellet. Discard supernatant.
7. Add 15 mL of buffer G2 (with RNase A) to the spheroplast pellet from step 6. Resuspend pellet thoroughly by pipeting. A homogeneous suspension is critical for efficient lysis of the spheroplasts.
8. Add 400 µL of Proteinase K solution and mix gently by inverting. Incubate at 50ºC for at least 1 hour and centrifuge at 5000 x g for 20 minutes at 4ºC. During

centrifugation, place a Qiagen Genomic-tip 500/G over a waste collector tube using a tip holder and equilibrate Genomic-tip by adding 10 mL of Buffer QBT.

9. Gently pour supernatant into a fresh 50 mL conical tube and discard the pellet. Vortex for exactly 8 seconds at top speed. Add 10 mL of Buffer QBT to the supernatant, and vortex again for 2 more seconds to mix. Pour mixture into the equilibrated Genomic-tip and allow it to slowly drip through the column by gravity. See **Note 7** if the column becomes clogged.
10. Wash the Genomic-tip by adding 30 mL of Buffer QC. Repeat wash. While waiting for the wash buffer to drip through, prewarm Buffer QF in 50ºC water bath.
11. To collect eluate, place Genomic-tip over a clean 50 mL conical tube using a tip holder provided by the manufacturer. Elute DNA with 15 mL of prewarmed Buffer QF.
12. Precipitate DNA by adding 10.5 mL of room temperature isopropanol to the eluate. Gently invert the tube 10 to 15 times until white web-like precipitated DNA appears.
13. Using a sealed glass microcapillary pipette, gently spool the precipitated DNA and transfer to a 1.5 mL microcentrifuge tube containing 200 µL of 70% ethanol. Nutate DNA for five minutes before spinning for a few seconds in a microcentrifuge at top speed.
14. Gently remove the supernatant with a pipet. Briefly air-dry the pellet for less than 5 minutes. Take caution to not overdry the pellet. Overdried pellets become difficult to redissolve later.

15. Depending on pellet size, add 100 to 150 µL of TE, pH 8.0, to the pellet. The optimal target DNA concentration is around 1 µg/µL. Dissolve the DNA overnight on a nutator at room temperature.
16. Next day, place DNA in 50ºC water bath for 1 hour to aid in dissolving the DNA. Flick the tube occasionally to help resuspension. If the pellet remains undissolved or if the DNA solution appears murky, *see* **Note 8**.
17. Measure DNA concentration with a spectrophotometer such as NanoDrop™. Take two readings to ensure consistency in DNA concentration. Widely different readings indicate the presence of undissolved DNA. Adjust DNA concentration to around 1 µg/µL with TE. Refer to **Note 8** for resuspending undissolved DNA. Store DNA sample at -20ºC, or proceed to DNase I digestion.

3.4. Fragmentation of DNA Using Deoxyribonuclease I

1. Prepare a boiling water bath for deactivation of DNase I. Alternatively, set a heat block or program a PCR machine to 100ºC.
2. Prepare appropriate dilutions of DNase I (*see* **Note 9**).
3. For each DNase I reaction, dilute 15 µg of genomic DNA in sterile ddH_2O to a volume of 36.8 µL, then add 4.5 µL of 10x One-Phor-All Buffer *Plus* and 2.7 µL of 25 m*M* $CoCl_2$ solution to a total volume of 44 µL.
4. Add 1 µL of diluted DNase I to the reaction tube. Thoroughly mix the tube with gentle flicks. (If you will be using a boiling water bath to deactivate DNase I, place a lid clamp on the tube at this step.) Immediately transfer the tube to a 37ºC water bath and incubate for 5 minutes.
5. Place the sample in the boiling water bath or in a 100 ºC heat block for 10 minutes

to deactivate DNase I and to convert dsDNA to ssDNA. Snap cool DNA sample on ice for 10 minutes immediately after boiling to retain DNA in single-stranded state. Quick spin the sample to collect any condensation that may have gathered on the sides of the tube.

6. Prepare a 2% TAE agarose gel with SYBR® Green I nucleic acid stain. Dilute SYBR® Green I stock solution 1:10,000 in the melted agarose solution just prior to pouring the gel. Shield from light.
7. Combine 1 μL of each digestion sample with 2 μL of 7.5% glycerol loading buffer. Load all 3 μL onto agarose gel along side with 3 μL of HyperLadder^{TM} IV. Run agarose gel until the range 30 to 100 bp is resolved.
8. If multiple DNase I digests were performed for a genomic DNA sample, select the best DNase I digest with which to proceed. Genomic DNA fragments should be under 100 bp. See **Figure 2-3** for an ideal fragmentation pattern (also *see* **Note 10**). Repeat DNase I fragmentation with genomic sample that does not show desirable fragmentation in any of the digests. Adjust DNase I concentration accordingly. Alternatively, one can consider increasing or reducing fragmentation time to achieve the desirable degree of fragmentation.

3.5. Biotinylation of DNA Fragments and Microarray Analysis

1. Add 1.5 μL biotin-N^6-ddATP and 1.5 μL terminal transferase to the DNase I-digested sample. Incubate at 37ºC for 1.5 hours to biotinylate DNA fragments.
2. In a boiling water bath, boil sample for 12 minutes and snap cool on ice for 10 minutes.
3. To prepare sample for hybridization, add 150 μL 2X hybridization buffer, 3 μL

Herring sperm DNA (10 mg/mL), 7.5 µL acetylated BSA (20 mg/mL), 5 µL control oligonucleotide B2, and 87.5 µL of water to a final volume of 300 µL. Transfer sample to a 2 mL screw top tube.

4. Find a local genomics core facility that provides service for processing the Affymetrix GeneChip® microarrays. Alternatively, contact Affymetrix for array processing services available in your area (*17*).
5. Follow standard protocol procedures for hybridization, washing and staining, and scanning of the GeneChip® Yeast Genome S98 Array as described in *GeneChip® Expression Analysis Technical Manual* (*13*), with the following exceptions:
    a. In preparing samples for hybridization, incubate samples at 99ºC for 10 minutes, then transfer to ice for 5 minutes before centrifuging at top speed for 3 minutes.
    b. Load 210 µL of sample onto GeneChip® Yeast Genome S98 Array. Avoid taking any particulates that may have settled at the bottom of the tube.
    c. Incubate array in hybridization oven at 42ºC, 60 rpm for 18 hours.
    d. Use fluidics protocol "EukGE-WS2v4_450".
6. After the array has been scanned, experiment data are generated by the Affymetrix® Microarray Suite Software in files with the following extensions: .exp, .dat, .cel and .chp. Only .cel files, which contain probe location and signal intensity data, are needed to proceed with the downstream analysis.

3.6. Data Analysis using Allelescan and CrossOver Software

1. If your .cel file is version 4, convert it from version 4 to version 3. If your .cel file was generated using Affymetrix® GeneChip® Operating Software, download Cel

File Conversion Tool for the conversion. If your .cel file was generated using Affymetrix® GeneChip® Command Console® Software, use the apt-cel-convert tool from the Affymetrix Power Tools for the conversion.

2. Install MATLAB® and the Statistics Toolbox™ onto your computer.
3. Copy the Allelescan software folder onto your computer.
4. Run allelescan.m file using MATLAB®.
5. Following the instructions in the *Allelescan Users Manual*, create a new project, identify locations of sequence polymorphic markers among samples, genotype one four-spore tetrad, and determine the segregation inheritance pattern of the tetrad (*14*). Also *see* **Note 11.**
6. See Figure 3 in Chen et al., 2008, for a sample segregation profile (*10*). Save the dump_segregation.txt file for further analysis using the CrossOver analysis software.
7. Install Python onto your computer.
8. Copy the CrossOver software folder onto your computer.
9. Following the instructions in the readme.txt file located inside the "doc" folder of CrossOver, copy the dump_segregation.txt file from Allelescan into the "segfiles" folder in CrossOver and change the filename according to the readme.txt file. Run CrossOver following the documentation given in the readme.txt file.

4. Notes

1. This paper presents a method which utilizes the Affymetrix GeneChip® Yeast Genome S98 Array. For an alternative approach using the Affymetrix custom tiling array, see reference (*18*).
2. Approximately 45% of all dissected tetrads for the S96/YJM789 diploid strain are four-spore viable tetrads (*10*). Dissect enough tetrads to obtain the target number of four-spore viable tetrads.
3. Here we describe the allele-specific PCR procedure to test one primer set in all 4 spores of a tetrad. To test multiple primer sets for the same tetrad, increase the initial resuspension volume for each spore accordingly. Our lab generally tests 8 different primer sets for each tetrad prior to microarray analysis. Only tetrads that display a 2:2 segregation of S96 and YJM789 alleles in at least 7 primer sets tested will be chosen for downstream microarray analysis.
4. Store the unused master primer mix at -20˚C for future use. However, repeated freezing and thawing will slowly degrade the master primer mix. Aliquot the master primer mix into smaller quantity to reduce the number of freeze and thaw cycles.
5. Alternatively, inoculate a 5 mL starter culture in YPAD liquid media from a single colony of fresh yeast culture. Grow at 30ºC for 6 to 8 hours. Measure the OD of the 5 mL starter culture. Calculate the volume of cells needed to set up a 500 mL culture at an OD of 0.005. Inoculate culture in YPAD liquid media in a 2.8 L flask. Grow for ~14 hours on platform shaker at 30ºC. In the next step, harvest all 500 mL of yeast culture by centrifugation.

6. Overloading the Genomic-tip with an excess of yeast culture leads to clogging of the tip, and underloading results in low DNA yield. Visually inspect the size of the cell pellet after the TE wash. We find that a cell pellet of 5 mL reliably yields a generous amount of DNA.
7. Highly concentrated genomic DNA lysates may clog the column, leading to decreased flow rate. Gentle and slow positive pressure may be applied to facilitate flow. When applying positive pressure, do not exceed the recommended flow rate of 20 to 40 drops/min.
8. If parts of the DNA pellet remain stubbornly undissolved, or if the DNA solution appears murky due to precipitated salt, spin DNA in a tabletop microcentrifuge at top speed for 5 to 10 minutes. Transfer the supernatant to a clean tube and measure the DNA concentration using a spectrophotometer. To recover additional DNA from the pellet, judiciously add small amounts of TE to see if it would progressively aid in dissolving DNA. Take caution to not dilute DNA to a final concentration of less than 500 ng/μL.
9. Because DNase I activity may vary between different batches, we recommend testing a range of DNase I dilutions to find the dilution that gives the most desirable digestion pattern. To start off, try 1:4, 1:3 and 1:2 dilutions of DNase I in sterile ddH_2O. Since the quality of genomic DNA can also affect the efficiency of fragmentation by DNase I, we recommend fragmenting every genomic DNA sample with at least two dilutions of DNase I, those which showed the best fragmentation patterns in the test sample. Please note that long term storage of diluted DNase I in water is not recommended. DNase I should be freshly diluted

before use to ensure consistent enzymatic activity. As an alternative to using varying dilutions of DNase I, one can also vary the incubation time for DNase I digestion and determine a time that yields the most desirable digestion pattern.

10. Oligo length of the probe set for the Affymetrix GeneChip® Yeast Genome S98 Array is 25 bp. Over-digestion of genomic DNA leads to nonspecific hybridization on the microarray and high background, while under-digestion results in low signal intensity. Therefore, it is crucial to fragment samples within 30 to 100 bp to ensure ideal hybridization.
11. Allelescan requires at least three independent hybridizations from each parent (S96 and YJM789) in order to locate sequence polymorphisms. Additional samples provide better marker determination, resulting in increased number of identified markers. Seven to eight thousand polymorphic markers between the S96 and YJM789 strains are within the maximum possible detection range for the Affymetrix S98 array.

Acknowledgements

We thank Carol Anderson for optimizing an alternative yeast inoculation method for genomic DNA extraction. We thank Ashwini Oke for her assistance in performing the DNase I digestion for this publication. We also thank Mike Pollard for critical reading of the manuscript.

References

1. Roeder, G. S. (1997) Meiotic chromosomes: it takes two to tango, *Genes Dev. 11*, 2600-2621.

2. Zickler, D., and Kleckner, N. (1999) Meiotic chromosomes: integrating structure and function, *Annu. Rev. Genet. 33*, 603-754.

3. Whitby, M. C. (2005) Making crossovers during meiosis, *Biochem. Soc. Trans. 33*, 1451-1455.

4. Hassold, T. (2007) The origin of human aneuploidy: where we have been, where we are going, *Hum. Mol. Genet. 16*, 203-208.

5. Hillers, K. J. (2004) Crossover interference, *Curr. Biol. 14*, R1036-1037.

6. Jones, G. H. (1984) The control of chiasma distribution, *Symp. Soc. Exp. Biol. 38*, 293-320.

7. Muller, H. (1916) The mechanism of crossing over, *Amer. Nat. 50*, 193-434.

8. Martini, E., Diaz, R. L., Hunter, N., and Keeney, S. (2006) Crossover homeostasis in yeast meiosis, *Cell 126*, 285-295.

9. Winzeler, E. A., Richards, D. R., Conway, A. R., Goldstein, A. L., Kalman, S., McCullough, M. J., McCusker, J. H., Stevens, D. A., Wodicka, L., Lockhart, D. J., and Davis, R. W. (1998) Direct allelic variation scanning of the yeast genome, *Science 281*, 1194-1197.

10. Chen, S. Y., Tsubouchi, T., Rockmill, B., Sandler, J. S., Richards, D. R., Vader, G., Hochwagen, A., Roeder, G. S., and Fung, J. C. (2008) Global analysis of the meiotic crossover landscape, *Dev. Cell 15*, 401-415.

11. McPeek, M. S., and Speed, T. P. (1995) Modeling interference in genetic recombination, *Genetics 139*, 1031-1044.

12. Zhao, H., Speed, T. P., and McPeek, M. S. (1995) Statistical analysis of crossover interference using the chi-square model, *Genetics 139*, 1045-1056.

13. Affymetrix. (2005) GeneChip Expression Analysis Technical Manual with Specific Protocols for Using the GeneChip Hybridization, Wash, and Stain Kit.

14. Richards, D. (2004) Allelescan Users Manual.

15. Okimoto, R., and Dodgson, J. B. (1996) Improved PCR amplification of multiple specific alleles (PAMSA) using internally mismatched primers, *Biotechniques 21*, 20-22, 24, 26.

16. Qiagen. (2001) Qiagen Genomic DNA Handbook.

17. Affymetrix. http://www.affymetrix.com.

18. Mancera, E., Bourgon, R., Brozzi, A., Huber, W., and Steinmetz, L. M. (2008) High-resolution mapping of meiotic crossovers and non-crossovers in yeast, *Nature 454*, 479-485.

Figure Legends

Figure 2-1. A schematic of the design for allele-specific primer extension PCR.
The common primer anneals to both the S96 and the YJM789 genome. SNP sites are engineered at the 3'-terminal nucleotide of each allele-specific primer, leading to amplification of only one of the two SNP alleles. Allele-specific primers are designed at two separate SNP positions (indicated by ▼ and ▽), approximately 200 bp apart. The resulting PCR yields two allele-specific bands with a 200 bp difference in size (shown in dotted line), which can then be visualized on the agarose gel. A single nucleotide internal mismatch is engineered in the allele-specific primers to enhance specificity by further

destabilizing allele primers that may have annealed to the wrong allele (*15*). Positions of mismatch are denoted by an asterisk (*). Primers containing mismatch at the 3'-terminal nucleotide do not successfully amplify and are illustrated in gray.

Figure 2-2. Allele-specific primer extension colony PCR for S96 and YJM alleles.
(A) Allele-specific primer extension PCR results for S96 (S) and YJM789 (Y) parental haploid strains. PCR primers are designed so that the S96 allele-specific band is approximately 200 bp longer than the YJM789 allele-specific band. PCR of four primer sets (PS) are shown. (B) Allele-specific primer extension PCR is performed for a four-spore tetrad using the same four primer sets as shown for the parents. Four spores are indicated by a, b, c, and d. PCR products from all four primer sets show 2:2 segregation of the SNP alleles.

Figure 2-3. Fragmentation pattern of genomic DNA.
Multiple dilutions of DNase I were used in fragmenting genomic DNA samples. 15 μg of genomic DNA was incubated with various dilutions of DNase I for 5 minutes. Shown are 1 μl of each digestion sample resolved on a 2% TAE agarose gel stained with SYBR Green I nucleic acid stain. In this sample, digestion using the 1:4 dilution of DNase I, indicated with an asterisk, reveals the most ideal fragmentation pattern.

Table 2-1. Allele-specific primer sequences.
Shown are primer sequences for 23 primer sets. Genomic coordinates are approximated in kilo-base pair. Each primer set assesses SNP genotype at two SNP positions. Three

primers are designed for each primer set: a common primer, a YJM789-specific primer (which anneals to the first SNP) and a S96-specific primer (which anneals to the second SNP). Primer sequences are given in the 5’ to 3’ direction. Mismatches internal to the 3’ end of the primer, when present, are underlined. The 3’-terminal nucleotide of each allele-specific primer is the position of the SNP and matches only one of the two possible SNP sequences.

Figure 2-1

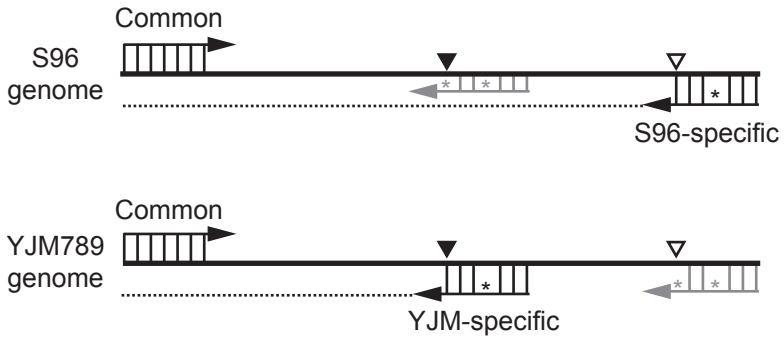


Figure 2-2

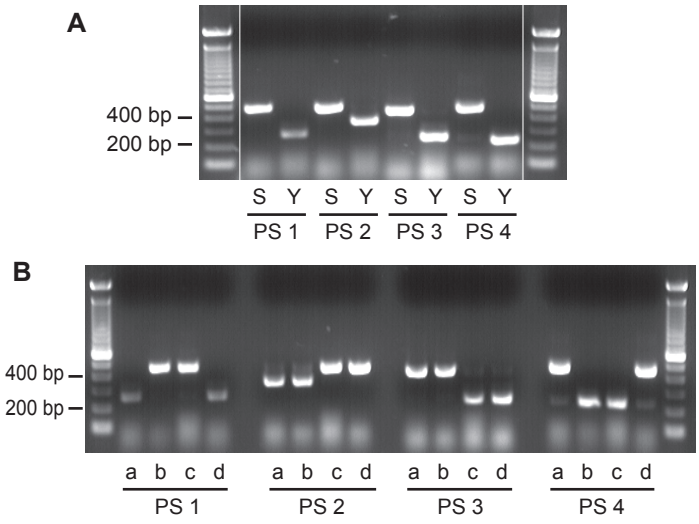


Figure 2-3

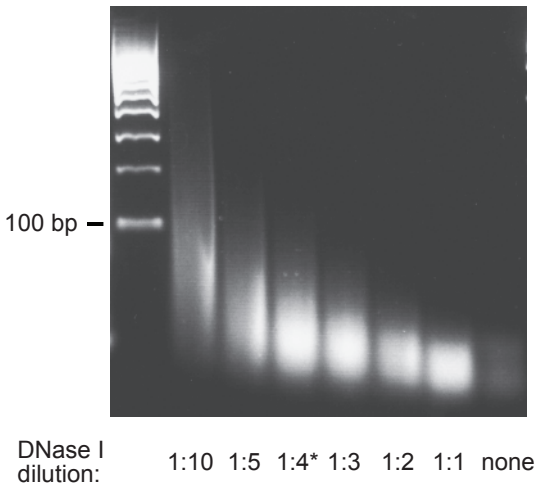


Table 2-1. Allele-specific primer sequences.

Chr.	Pos. (kbp)	Primer Type	Primer Sequence (5' to 3')
II	673	Common	GGCTATTGATGCGATAAATAAAGGC
		S96-specific	TTGGTTCTACGATACTGGGTGAC
		YJM789-specific	TTCCACATATCTTTTGAAAAGAGTCGTA
III	82	Common	GCCGAGAGTATCACTGATTCAAGG
		S96-specific	CGCTGTTAGGTGGCTTTTTTACAGTA
		YJM789-specific	CACTTTCAGTCCCTTTTTCCTCCT
IV	344	Common	GTAATTCTACCTAGCCCACCAC
		S96-specific	GCATATCGTATGATTCGACTACAGACG
		YJM789-specific	CTTATCTAAGCTGATACCAGGGTATA
IV	1261	Common	CGGATACCAAGATTGTCATACTCACTAAAG
		S96-specific	CTTAATGGGTATGAATATATTCTTGTTTATTCTCC
		YJM789-specific	GGTGAATGTAAAATTAATACGGCGGTAAC
V	458	Common	GCGATAATTGACCTTTTCCAAGGAC
		S96-specific	GGTCCCTTATAAACGTATGAAGTGTAG
		YJM789-specific	GTTTCTTAGGCAATCTAGTAATGTTG
VI	239	Common	CATATGTATACACATATACATATCTGTACATACTC
		S96-specific	GATAGCTGCCCATCGAAATACGTTT
		YJM789-specific	GATTATAGATACCCACGACTGGTTGAAA
VII	773	Common	GGGTGATAATACATACTCCCCATC
		S96-specific	GTTGGGATTCCATTGTTAATAACACTAG
		YJM789-specific	CATGGAAAACCGGATTTCTAGGAAGGAAG
VIII	359	Common	GGTGAATAATGAAGATTGGGTGAATAATTTG
		S96-specific	GTGATAATACACTACTAATGTGACTACTAGTAGAC
		YJM789-specific	GCTGTGATAATTATTCATAGAAATATTACAGAGCATA
VIII	413	Common	CGCAAGACTTTCTTCACCAATACTTTG
		S96-specific	CATTTACTTCACTTCGTAGCAATGTTAAG
		YJM789-specific	GGCATGCATACTGGGACGT
IX	98	Common	GGCCAATGAGCAAAAATTTAGGC
		S96-specific	CAAATTGGAAGCAAAGAGAAAGGTTTC
		YJM789-specific	CCTCCCCGTTACAGTTTAGACTG
IX	191	Common	CTCGAAAGTGCTACCCACTGC
		S96-specific	GGGACGAAAAGAGCAGCTGTATTAACG
		YJM789-specific	GGGTTTATTACTTCAGGGAACTTTCTGGTT
X	137	Common	GAAATAGTAATCCCAACGCACTCATCCGC
		S96-specific	CTTCTGAAAATAATCTTGAAATGGCATGATATGAATCTA
		YJM789-specific	GGTGAACAGGTGCATTTTGAGAAGA
X	148	Common	GTAATGACTATACGTATAAGGAAAATTAAGAAAAGGC
		S96-specific	CACCACAACAAGCTATGCTATAC
		YJM789-specific	GGTGCTATCAGTAAAAGGAAGGAGAACAT
X	516	Common	GTAAATCAGTATAGTAATGTCCTTCGGATGG
		S96-specific	GGATGTACCTAAAATACAGCAAACAAAGCGTT
		YJM789-specific	CACGCAAGCCATCACCCGATA
X	622	Common	CCATCCAGAGTATACATCGAAGG
		S96-specific	CACTTCAATCCTTTCAAGAAGACATAT
		YJM789-specific	GAAGAATCTTTGAAGACTGGTAATCCT
X	627	Common	CTGTGAACCTTAGAAATCCTCTATGC
		S96-specific	CGTCCAACCTGCCCATCACCCT
		YJM789-specific	AATGATGAGATAATTAACCCAACAGCCGG
XI	394	Common	CGTGTGGCTGCCTCTAAGAATTAAACTTC
		S96-specific	CCATTGATCATTTGCACAAATCATTGAAC

		YJM789-specific	GCTTCGCTCAATAAAAAAAGATCTTCATCGG
XI	624	Common	GAGGAGTTCAACAATGAACTGC
		S96-specific	ATGAATCCTTTTGGGCAGGATT
		YJM789-specific	AGTTTTTCACCGGAAAGTAACGGAATA
XI	320	Common	GTATAAGTGCATACTAACATACTGTGTACGTAC
		S96-specific	GACATGAACGACGTTTTGGGAAAAATAAC
		YJM789-specific	CTAAGAGAAGATTCGGGTTTTAATTTAAGGTT
XII	574	Common	GTTGAAGCACTGCCTCCAG
		S96-specific	GATCGAAGGAAACTAAAAGAGGTTTGATGTCAG
		YJM789-specific	GCGCCAAACAAGGGATGG
XII	780	Common	CATGGAGGCTAGACATGACTAATG
		S96-specific	CAGTCGATCTCTTGCCCTAG
		YJM789-specific	CCTTTTGTTCAATGGCAGAATTTCTATGCA
XIII	216	Common	GACCGCTATGCGTCTGATGT
		S96-specific	CAGCTGATAAAGAACACTGATCATGACA
		YJM789-specific	CCTTTTGGATCTTCTGTCTTTGAGCT
XIII	802	Common	CCAGCAGGGAAGCCATTAAATAG
		S96-specific	CTAGGTGAGTAGACTAACCGATCC
		YJM789-specific	GTATTTGAGAAGGGGGTTTAACACTAACA

Chapter 3: CrossOver Analysis Tool

Stacy Y. Chen, Jennifer C. Fung

Department of Biochemistry and Biophysics

University of California, San Francisco

San Francisco, CA 94158

Abstract

As microarray-based techniques replace the more traditional genetic assays in mapping meiotic recombination products, new bioinformatics tools are needed to process the large quantity of data produced by microarrays. The CrossOver program was developed to fulfill that need. It is an analysis tool that determines the location of meiotic recombination events based on the genotypic profile at each polymorphic marker from a full four-spore viable tetrad in budding yeast. CrossOver identifies and computes the location of eight types of crossover (CO) motifs and seven types of gene conversion (GC) tracts. In addition, it calculates: CO density of the whole genome, CO densities of each chromosome, occurrence of nonexchange chromosomes, the distribution of inter-recombination events, centromere-to-recombination events, telomere-to-recombination events, the distribution of GC tract lengths, and chromatid interference. Also computed are the correlation coefficient between the number of COs and NCOs of each meiosis over an ensemble of tetrads as an indicator of CO homeostasis, and the parameters of the gamma distribution function for inter-CO distances, which are indicators of CO interference. This chapter describes the various types of COs and GCs identified by

CrossOver and the programming strategy employed in searching for those recombination motifs. The various functions and output data types of CrossOver are also explained.

Introduction

As microarray technology replaces genetic tetrad dissection in determining meiotic recombination products between two polymorphic yeast strains, new bioinformatics tools are needed to analyze the large amount of microarray data. Traditionally, less than a dozen genetic markers are analyzed using tetrad dissection in a given analysis, restricted by the number of different auxotrophic or drug resistant markers engineered in a given genome (Malkova et al., 2004). Dissection data are often scored by hand and calculated, at best, by using a simple spreadsheet program for the recombination and interference values. Our use of microarray application yields approximately 8000 sequence polymorphic markers between yeast *Saccharomyces cerevisiae* strains S96 and YJM789—a number that makes manual tallying of genetic markers an impossible task. The CrossOver program is developed as the new bioinformatics tool to fulfill this need.

Microarray data is first analyzed with an existing software Allelescan, developed by Daniel Richards in Dr. Ron Davis' Lab at Stanford (Winzeler et al., 1998). Allelescan is a microarray analysis platform that analyzes genomic DNA hybridization data, and identifies sequence polymorphisms using differential hybridization signal intensities between samples from two distinct genetic backgrounds. Meiotic progeny from a cross between two different parental backgrounds can be genotyped at each of the polymorphic markers. In the end, Allelescan generates a marker segregation profile that records the marker allele of each polymorphic marker for all 4 spores of a given tetrad. This

segregation file can then be used for downstream analysis of meiotic recombination events. (See step 1 under “Programming Strategy” for an example of a segregation file.)

The CrossOver program is developed in our lab as a downstream analysis tool to process the marker segregation profile and determine the positions of meiotic recombination events (Chen et al., 2008). It scans through the marker segregation pattern of a full four-spore tetrad to identify 8 types of CO and 7 types of GC motifs. For each recombination event, the program documents the positions of the event, which chromatid(s) are involved in recombination, the tract length of a GC event, and the number of polymorphic markers involved in identifying the event. In addition, CrossOver performs analyses to characterize several aspects of meiotic recombination. CrossOver can compute CO and GC density of each of the 16 chromosome as well as the entire genome, occurrence of nonexchange chromosomes, inter-recombination event distances, recombination event-to-centromere and closest telomere distances, and the ratio of two-, three- or four-strand double COs for evaluation of chromatid interference. CO homeostasis is assessed by the correlation coefficient between the number of CO and noncrossover (NCO) of an individual meiosis across the ensemble of tetrads; and CO interference is evaluated by the parameters of the gamma distribution function for inter-CO distances.

The purpose of this chapter is to explain what CrossOver program does and the logic behind programming CrossOver. This chapter describes the various types of COs and GCs identified by CrossOver and gives schematic drawings for each recombination type. It goes through an in-depth discussion of the programming strategies employed in search for each recombination motif, accompanied by a detail flowchart of the

programming logic to help users understand how each recombination motif is selected. Lastly, CrossOver functions that compute various analyses of the meiotic recombination events are outlined and explained.

Results

Eight Types of COs and Seven Types of GCs as Delineated by CrossOver Program

During early meiosis, programmed double-strand breaks (DSB) are generated by Spo11 proteins across the whole genome (Keeney et al., 1997). DSBs are repaired through a series of recombination steps that results two main recombination products: COs and/or GCs (Baudat and de Massy, 2007; Whitby, 2005). COs are reciprocal exchange of genetic information between two homologous chromosomes. They form chiasmata which provide the physical link that holds homologs together during the meiosis I division (Roeder, 1997). GCs, on the other hand, are nonreciprocal transfers of genetic information from one homolog to another and do not produce chiasmata. They generate non-Mendelian type (non-2:2) segregation of alleles in the four meiotic progeny within a tetrad.

COs are intimately associated with GCs. Genetic experiments have found that sites of COs are often physically linked to GCs (Borts and Haber, 1989; Symington and Petes, 1988). Short GCs can be found on the two homologous chromosomes that are involved in crossing over. These GCs presumably share the same DSB that gives rise to a CO. GCs associated with COs arise when mismatch repair proteins correct heteroduplex DNA near the site of crossing over (Allers and Lichten, 2001; Borts et al., 2000). COs that do not have an associated GC may appear due to the lack of detectable allelic

markers at the region of crossing over, or due to the repair of heteroduplex DNA that results in restoration of alleles instead of conversion (Paques and Haber, 1999). GCs not associated with a CO are called NCOs. They arise from DSB events that are independent of COs. In this section, different types of COs and GCs identified by CrossOver will be outlined and explained. Figure 3-1a illustrates the various types of CO motifs and figure 3-1b shows different types of GC tracts.

Type 0 CO: Single CO without associated GC

This is the simplest type of all COs identified by the CrossOver program. DNA is reciprocally exchanged between the same pair of adjacent allelic markers on both chromatids that were involved in crossing over. All markers involved in identifying the CO segregated 2:2. Thus, no GC is detected at the CO junction. As mentioned earlier, failure of detection for a GC may be due to the lack of detectable allelic markers near the exact site of crossing over.

Type 1 CO: Single CO with associated GC

The type 1 CO motif includes a GC at the CO junction between the two chromatids involved in crossing over. A majority of COs in yeast are believed to have an associated GC. A misalignment in the exchange of parental alleles between the two chromatids results in a short stretch of non-Mendelian (non-2:2) segregation of alleles at the site of crossing over. This short stretch of non-2:2 markers occurring within the juncture of a crossover marks the location of the GC associated with a CO. These GCs arise from

mismatch repair of the heteroduplex DNA at the site of Holliday junctions, and they share the same DSB that gave rise to the CO (Allers and Lichten, 2001).

Type 2 CO: Single CO accompanied by a NCO

In type 2 COs, the two chromatids involved in crossing over give a clean CO junction without a detectable GC as in the case of type 0 COs. However, a small stretch of markers with non-2:2 segregation of alleles is detected on one of the chromatids not involved in crossing over, resulting in the detection of a NCO near a CO site. The NCO tract length either overlaps the pair of adjacent markers where the CO exchange occurred, or contains at least one of the two adjacent markers. The CrossOver program categorizes this type of NCO that accompanies a CO as type 6 GCs. See below for more description of type 6 GCs.

Type 3 CO: Single CO with associated GC, accompanied by a NCO

A type 3 CO is essentially the same as a type 2 CO, except the CO has an associated GC as in case of type 1 COs. The CO is associated with a GC at the CO junction, and a NCO tract is detected to overlap parts (or all) of the GC on a chromatid not involved in crossing over. Type 3 CO presents a special scenario where a short stretch of markers that segregated 4:0 or 0:4 occurs at the site of crossing over, overlapping either both ends of the GC tract that is associated with CO or only one end of the GC tract. Because of the appearance of the short stretch of 4:0 or 0:4 markers, it is listed as a separate CO type by the CrossOver program.

Type 4 CO: Single CO associated with complex pattern of GC(s) and/or NCO(s) not described by types 0 to 3 COs

Single COs that do not fit in any of the above types of COs are sorted as type 4. It includes CO with discontinuous stretches of GCs at or near the CO site, and/or continuous or discontinuous intervals of NCO that appear on the two chromatids not involved in crossing over. A rough estimation of the CO location is given by the CrossOver program for users to inspect the marker segregation profile at these sites manually.

Type 5 CO: Double COs without associated GC

Double COs are cases where two COs are detected at the same location in the genome and involve all four chromatids in the CO exchange. The switch of allele from one parent to the other occurred between the same two adjacent allelic markers in all four chromatids. Thus, both COs have the same genome location. No GC is detected to associate with either CO. This type of double COs arises from two independent DSB events on two different chromatids.

Type 6 CO: Double COs with associated GC(s)

Type 6 double COs are similar to type 5 double COs, except in at least one of the two COs, the switch of parent alleles occurred between different pairs of adjacent allelic markers in the two chromatids involved in the exchange, leading to the detection of associated GC. Type 6 double COs may include one CO with associated GC or both COs may be associated with a GC.

Type 7 CO: Double COs with complex GC pattern

Type 7 double COs contains associated GC(s) of discontinuous conversion tract(s). Any double COs that do not fall in the category of either type 5 or type 6 double COs are sorted as type 7. Just as in the case of single CO with complex GC pattern (type 4 CO), a rough estimation of the CO location for both COs is given by the CrossOver program for users to inspect the marker segregation profile of these sites manually.

Type 8 CO: Single CO associated with GC, accompanied by a NCO (selected from two COs within close range)

Depending on which parental allele the NCO near a CO with an associated GC is converted into, type 3 COs may also appear as two distinct CO events within close proximity of each other. When the conversion of NCO results in a 4:0 or 0:4 segregation of allelic markers, it is categorized as type 3 COs. However, if the conversion of NCO results in 3:1 or 1:3 segregation of alleles, then the combination of a CO with associated GC and accompanied by a NCO may appear to be two separate COs exchanging between 3 chromatids (a 3-strand double COs). It is difficult to distinguish between the apparent double COs in this case and actual double COs. The CrossOver program allows the user to make this decision. Users may choose to categorize two COs found within a user-defined range that exchange between 3 chromatids as type 8 COs, a CO with associated GC accompanied by a NCO, rather than actual two distinct COs. Both type 8 and type 3 COs contain a CO associated with a GC and accompanied by a NCO, the difference lies

in type 8 COs that do not contain any markers with a 4:0 or 0:4 segregation pattern. Otherwise, both types of COs are essentially the same.

Type 0 GC: NCO

NCOs are GC events that give rise to a non-reciprocal and non-Mendelian type allele segregation among the four chromatids, specifically in the 3:1 or 1:3 segregation pattern. They are not associated with COs. They are thought to be the recombination product of the synthesis dependent strand annealing (SDSA) pathway, and each NCO event arises from an independent DSB event (Merker et al., 2003).

Type 1 GC: GC associated with CO

As mentioned in the discussion for type 1 COs, GC events that accompany a CO are byproducts of mismatch repair and arise from the same DSB that gave rise to the CO associated with it. They are a result of the correction of heteroduplex DNA by the mismatch repair proteins near the site of crossing over. Restoration of heteroduplex DNA from the strand invasion step may give rise to non-2:2 segregation and nonuniform patterns of GC near the CO junction.

Type 2 GC: Premeiotic GC

Mitotic GCs generated while the cell was at the diploid stage prior to the premeiotic replication will result in markers with a 4:0 or 0:4 segregation of alleles in the four spore tetrad stage. This type of conversion event is not due to the programmed DSBs induced at the onset of meiosis, but result from mitotic recombination . Break-induced replication

(BIR; see type 3 and 4 GCs) that contains 4:0 or 0:4 markers are filtered out and not counted as a premeiotic GC. Markers segregated 4:0 or 0:4 from the type 3 CO (CO with associated GC and accompanied by a NCO) are also filtered out.

Type 3 GC: Single BIR of 3:1 or 1:3 allele segregation

BIR is a GC event that extends all the way to the end of the chromosome arm (Paques and Haber, 1999). These conversion tracts can be very large, extending from the site of DSB all the way to the end of the chromosome, possibly a couple hundred kilobases in length. Type 3 GCs specifically include single BIRs that display a 3:1 or 1:3 segregation of alleles.

Type 4 GC: Double BIRs of 4:0 or 0:4 allele segregation

Type 4 GCs are cases where there are two BIRs with both BIR tracts extending to the end of the same chromosome arm, resulting in a GC tract with 4:0 or 0:4 segregation of alleles at the end of the chromosome arm. The two BIR events originate from two independent DSBs that occurred on two separate chromatids, and convert the particular chromosome end of all four chromatids to the same allele genotype. This is a rare event but is occasionally observed in meiotic mutants.

Type 5 GC: NCO that appears as two adjacent COs with close proximity

Depending on the allele restored from the repair of heteroduplex DNA in GC events, a NCO product may appear to be two distinct but nearby COs (as an apparent double COs) that exchanged between the same two chromatids (Paques and Haber, 1999). Type 5 GC

events are determined by selecting for two COs located within a user-defined distance of one and another and exchanged between the same two chromatids in both COs. In this case, both exchange events arise from a single DSB and is actually a NCO in disguise due to the different possible outcomes of allele restoration of the mismatch repair.

Type 6 GC: NCO near a CO on a non-crossing over chromatid

Type 6 GCs are NCOs that occur within a user-specified distance of a CO site on a chromatid not involved in crossing over. This scenario may arise from both ends of a DSB invading different chromatids, resulting in a multi-chromatid joint molecule that is frequently observed in *sgs1* mutants (Oh et al., 2007). One invading strands eventually resolved into a CO, while the second invading strand resolved into a NCO on a separate chromatid. This CO motif may also arise from a heteroduplex rejection of an invading strand that eventually resolves via crossing over through invasion of a different chromatid strand (Borts and Haber, 1987). A third possibility may be that the CO and the nearby NCO arise from two completely independent DSB events in close proximity with one and another on two separate chromatids—a plausible scenario at DSB hotspots. This type of NCOs are categorized as a separate type so the users may chose include or not include them as real NCOs that come from an independent DSB event.

Type 7 GC: NCO near a CO on a crossing over chromatid

Type 7 GCs are NCOs that occur within a user-specified distance of a CO site on a chromatid involved in crossing over. This type of GC may arise from mismatch repair of

short patches of heteroduplex DNA near double Holliday junctions (Allers and Lichten, 2001). In this case, the NCO arises from the same DSB that resolved into the CO.

Programming Strategy

This section will explain the strategy employed by CrossOver in selecting for the various types of COs and GCs delineated in Figures 3-1a and 3-1b. The descriptions here are intended to help users understand how each type of COs and GCs are derived and the strategies involved in doing so. It is not intended to be a comprehensive discussion on how each variable in the CrossOver program is calculated. The strategy involved in selecting for COs will be explained first. Single COs are explained in steps 6 to 11 and the double COs in steps 12 to 14. Special handling of single COs that are within a close range of each other are explained in steps 15 to 17. GCs that appear at or near CO junctions, such as in types 1, 5, and 6 GCs, will be explained in the steps that elucidate the selection of the particular CO the GCs accompany. Lastly, the remaining types of GCs, such as NCOs, NCOs near COs, BIRs and pre-meiotic GCs, will be explained in steps 18 to 24. A schematic flowchart of the strategy is shown in Figure 3-2.

Step 0. System requirements

CrossOver version 3.0 is written in Python 2.5 (http://www.python.org) and uses the numerical package NumPy 1.3.0 (http://numpy.scipy.org) for a variety of statistical analyses. It is developed and tested on Eclipse (http://www.eclipse.org), a development platform maintained by an open source community, but can also be run directly from command line in the Unix system.

Step 1. Read input file

The CrossOver program is designed to read and process the marker segregation profile of a four-spore tetrad. Below is an example of the first few lines of the segregation profile of a wild-type tetrad in the input format specified for CrossOver:

(chr#)	(posn)	[genotypes (0=yjm,1=s96)]:		WTx5a	WTx5b WTx5c WTx5d
1	414	1	0	1	0
1	482	1	0	1	0
1	483	1	0	1	0
1	3335	1	0	1	0
1	7401	1	0	1	0

Each row indicates one allelic marker, and gives the chromosome number, the genome coordinate in base pairs and the genotype for each spore of a tetrad at that particular allelic marker. A genotype of “0” indicates an YJM789 allele at that marker position, and “1” indicates a S96 allele.

Step 2. Sort and index all markers

Allelic markers are first sorted by its chromosome number and then by genome position in increasing order. An ordered numerical index is assigned to the sorted marker list, where 0 indicates the first marker and each subsequent marker gets an index of one plus the previous index.

Step 3. Sort markers into three types based on their segregation ratio

For each allelic marker, the segregation ratio between the S96 and YJM alleles among the 4-spore tetrad is determined. Five segregation ratios of S96:YJM789 are possible: 0:4, 1:3, 2:2, 3:1, 4:0. All markers are then sorted into three types based on their segregation ratio: a) markers that segregated 2:2; b) markers that segregated either 3:1 or 1:3; and lastly c) markers that segregated 4:0 or 0:4.

Step 4. Identify switch in genotype between S96 and YJM alleles along each chromatid using the 2:2 segregation marker list

To identify CO sites, we start by analyzing the list of markers that segregated 2:2 and ignore, for now, the markers that segregated otherwise. The strategy is to first identify the pair of 2:2 markers that give the approximate CO position, and then add back the non-2:2 markers to recalculate the exact CO location and determine the GC pattern(s), if any, near the site of crossing over.

CrossOver sequentially scans through the list of markers that segregated 2:2 in each spore. It then identifies pairs of markers where a switch in genotype occurred, either from 0 to 1 or 1 to 0. From here on, a change in genotype between two markers will be referred to as a “genotype switch”, and two markers that define the switch will be called “2:2 switch markers.” Notice that the indices of the 2:2 marker list will not to be contiguous due to the filtering of non-2:2 markers. Thus the indices of the pair of markers that marks the switch in genotype may or may not be sequential in number. Marker pairs not sequential in indices suggest additional non-2:2 markers at the CO junction and therefore imply the presence of a GC associated with the CO. From here on, a switch in

genotype from the S96 allele to the YJM789 allele will be referred to as the “1-0 switch”, and a switch from YJM789 to S96 allele will be referred to as the “0-1 switch”.

Step 5. Identify genotype switches that share the same marker pair indices

Because only markers that segregated 2:2 are scanned through at this stage, any switch in genotype in one chromatid will be accompanied by a reciprocal genotype switch in another chromatid between the same two markers. For example, a 1-0 switch will have a 0-1 switch between the exact same two markers on a different chromatid. The next step is to identify and pair opposing genotype switches from different chromatids that occurred between the same pair of 2:2 markers.

The interval specified by the pair of 2:2 markers indicates the relative location of a CO. If four genotype switches are found between the same two markers, with one in every chromatid (two chromatids with 1-0 switch and two with 0-1 switch), then the interval specified by the two markers delineates the relative position of two COs, with each CO arising from an independent DSB event and exchanges DNA between a different set of chromatids. Single COs with two opposing genotype switches are recorded in a list separate from the double COs that have four genotype switches. The two lists will be analyzed separately.

Finding single COs: steps 6 ~ 11

Step 6. Identify type 0 CO: single CO without associated GC

For COs with non-detectable associated GC, the pair of 2:2 switch markers that specify the relative CO location would have sequential marker indices, indicating that there is no

additional non-2:2 markers present at the CO junction. Failure of detection for a GC may be due to the lack of detectable allelic markers near the exact site of crossing over. These COs are categorized as type 0 COs, which are “single COs without associated GC.” The location of CO is calculated as the midpoint of the two 2:2 switch markers (Figure 3-3a).

Step 7. For switch pairs that do not have sequential marker indices, add back the non-2:2 markers. Identify and tally the number of genotype switches within the interval marked by the 2:2 switch markers along each chromatid.

For COs with associated GC and/or NCO at the site of crossing over, marker indices for the 2:2 switch markers will not be sequential in number. To further analyze these types of COs, the non-2:2 markers that were filtered out in step 4 are added back in this step. Each chromatid is scanned for changes in genotype within the interval marked by the 2:2 switch markers. This is followed by recording the direction of genotype change and the indices of the non-2:2 switch markers for each chromatid. The number of switches found in each of the chromatids is recorded in a 1 by 4 matrix notation [a, b, c, d], where a, b, c, and d are the number of switches in each of the four spores at the particular CO junction.

Step 8. Identify type 1 CO: single CO with associated GC

Type 1 COs are single COs with associated GC. They have GC tracts at the CO junction on the two chromatids involved in the exchange. Therefore, the two chromatids involved in crossing over will each have one genotype switch (with opposite change in genotype), and the two chromatids not involved in crossing over will have no genotype switch. The

number of switches in type 1 CO is indicated by the following matrix notation: [0, 0, 1, 1].

To calculate the location of type 1 CO, CrossOver first calculates the midpoint of the two switch markers of each genotype switch, and then takes the midpoint of the two midpoints as the CO location (Figure 3-3b). The position for the GC associated with this CO will share the same genome position. GC tract length is calculated by taking the difference between the two midpoints of the two pairs of switch markers. In addition, a maximum possible tract length is calculated by taking the difference between the two outermost markers of four switch markers from the two genotype switches. Conversely, the minimum possible tract length is calculated by taking the difference between the two inner markers of the four switch markers (Figure 3-3b). Please note that a GC indicated by only one non-2:2 allelic marker will have the same innermost marker and therefore gives a minimum possible tract length of 0 using this method of calculation.

Step 9. Identify type 2 CO: single CO accompanied by a NCO

The hallmark of type 2 CO is the presence of a NCO on a chromatid not involved in crossing over. COs in type 2 are not associated with GCs. The markers that specify the NCO tract overlaps the two switch markers that identify the CO, resulting in the presence of non-2:2 markers between the 2:2 switch markers. The number of switches in this type of CO is indicated by the following matrix notation: [0, 1, 1, 2]. The two chromatids involved in crossing over will each have one genotype switch of opposing genotype change with both genotype switches aligned and occurring between the same two markers, indicating no GC present between the CO junction. The one chromatid with 2

genotype switches will have two switches of opposing genotypic change, indicating the presence of a NCO.

The location of CO is calculated the same way as type 0 CO, using the midpoint of the pair of markers that defined the genotype switch on both chromatids. The position of NCO is determined by taking the midpoint of the two midpoints of the switch markers that defined the NCO interval (Figure 3-3c). NCO tract lengths are calculated similar to GC tract lengths in step 8.

Step 10. Identify type 3 CO: single CO with associated GC, accompanied by a NCO

Similar to type 2 COs, type 3 COs also have a NCO present at the CO junction on a chromatid that is not involved in the CO exchange. The main difference here is that the CO has an associated GC, which means the indices of the two switch marker pairs will not align between the two chromatids involved in CO exchange. Type 3 COs will have the same matrix notation as type 2: [0, 1, 1, 2]. CO position and associated GC position and tract lengths are determined as described in step 8 for the type 1 COs. NCO positions and tract lengths are determined as described in step 10 for the type 2 COs. It is interesting to note here that type 3 COs will always contain a stretch of 4:0 or 0:4 markers at the CO junction. These 4:0 or 0:4 markers are not categorized as pre-meiotic GCs, but are part of the marker pattern that result from a type 3 CO with associated GC and NCO combination.

Step 11. Identify type 4 CO: single CO with complex pattern of GC(s) and/or NCO(s)

COs that do not fit the patterns describe for type 0 to 3 COs are categorized as type 4. These COs do not have sequential 2:2 switch marker indices and they do not have segregation matrix of either [0, 0, 1, 1] or [0, 1, 1, 2]. They are categorized as type 4 COs and are intended for users inspect each case manually. The estimated position given for type 4 CO is calculated using the midpoint of the two midpoints of the 2:2 genotype switch markers.

Finding double COs: steps 12 ~ 14 (continued from step 5)

Step 12. Identify type 5 CO: double COs without associated GC

Steps 12 through 14 identify two COs that occurred between the same pair of 2:2 switch marker indices and exchanges all four chromatids. A genotype switch is found in all four chromatids between the same pair of adjacent 2:2 markers. Type 5 double COs does not have detectable GC tracts between either COs. Therefore, the two 2:2 markers that marked the CO junction will have sequential indices, and the genotype switches will align in all four chromatids. In this case, both COs will have the same CO position, which is calculated as the midpoint of the two switch markers. However, we cannot unambiguously identify which chromatid exchanged with which, information that is used to calculate chromatid interference at a later step. Therefore, each CO is randomly assigned with a unique pair of chromatids of opposing genotype switch for the purpose of chromatid interference calculation.

Step 13. Identify type 6 COs: double COs with 1 or 2 associated GC(s)

For double COs that do not have sequential 2:2 switch markers, add back the non-2:2 markers and find the two markers where the genotype switch occurred in each of the four chromatids. Select for double COs with the switch marker matrix of [1, 1, 1, 1], where each chromatid has only one genotype switch. When assigning which two chromatids are involved in crossing over, sort all four switches by genome location, and then pair up the first 1-0 switch with the first 0-1 switch and the second 1-0 switch with the second 0-1 switch. If any of the same type genotype switches have the same position, randomly assign them to a partner of an opposing genotype switch. After assigning CO exchange partners, CO position, associated GC position and tract lengths are calculated as described for type 0 and/or 1 CO.

Step 14. Identify type 7 CO: double COs with complex GC pattern

Any double COs that do not have sequential 2:2 switch marker indices and do not display the switch marker matrix of [1, 1, 1, 1] are listed in this category. These COs contain complex GC associated with either one or both of the double COs. An estimated CO location is given by calculating the midpoint of the two midpoints of the 2:2 switch marker pairs. Users are expected to inspect these COs manually. Chromatids are randomly paired for CO exchange partners in the same way as for type 5 COs for the purpose of chromatid interference calculation.

Finding two COs within a user-defined range: steps 15~17

Step 15. Identify two single COs within a user-defined range

Single COs that appear within close proximity of each other may not be true COs. Two adjacent COs that exchange between the same two pair of chromatids (known as 2-strand double COs) may be a NCO (Paques and Haber, 1999), whereas closely positioned 3-strand double COs may be a CO accompanied by a NCO. Assuming interference would reduce the likelihood of COs within close proximity of each other (Hillers, 2004), the program allows users the option to filter out two single COs located within a user-defined range and categorize them accordingly. CrossOver selects only single COs that are within a user-defined range of each other and ignores double COs in this selection. However, since double COs are not included in this analysis, any single COs that are within a close range of a double CO will not be detected. Users working with meiotic mutants where interference is abolished or unpredictable CO pattern is displayed are advised to examine the CO pattern manually and use the output of this program as a reference only.

The default range for selecting COs within a close range is set at 6 kb. This was determined by investigating the distances between adjacent COs in two wild-type tetrads whose recombination events were mapped through next-generation sequencing (see chapter 5 of this dissertation). While most inter-CO distances are larger than 19kb, a small subset of inter-CO distances cluster between 0.9 kb to 5.9 kb. Therefore, we hypothesize that the small subset of double COs with a inter-CO distance under 6 kb are actually not two COs, but one of the above possible alternative events depending on the number of chromatids involved in the exchange.

Step 16. Two-strand double COs within close proximity

For two single COs that fall within a user-defined range of each other, if the two COs exchange between the same two strand of chromatids, it is categorized as a NCO that appears as two COs. Both COs are deleted from the CO list, and a NCO is added to type 5 GC. The midpoint between the positions of the two COs is taken to be the position of the new NCO, and the difference between the two CO positions is the tract length for the NCO.

Step 17. Three-strand double COs within close proximity

The scenario where two single COs that fall within a user-defined distance of each other and exchange between 3 chromatids is categorized as type 8 CO, which is CO with associated GC and accompanied by a NCO on a third chromatid. Type 8 is essentially the same as type 3 CO. The only difference lies in the genotype of the NCO with respect to the genotype of the GC associated with CO. If the NCO and GC associated with CO have the same genotype, it displays a 4:0 or 0:4 marker segregation pattern which is the hallmark of type 3 COs. But if the NCO and the GC associated with CO have opposite genotypes, then it will appear as a two COs within close proximity and thus be categorized as type 8 CO. The two single COs are deleted from the CO list. A switch marker matrix is calculate for the interval between the two COs and should display the configuration of [0, 1, 1, 2]. The two chromatids with one genotype switch are involved in crossing over. The chromatid with two genotype switches contains the NCO. CO position, GC position and tract length, as well as NCO position and tract length, are calculated as described in type 3 CO.

Finding additional GCs: steps 18 ~ 22 (continued from step3)

Step 18. Sort consecutive non-2:2 markers into GC tracts

Two different lists of non-2:2 markers were identified at step 3: markers that segregated 3:1 or 1:3, and markers that segregated 4:0 or 0:4. In both lists, GC tracts were identified by grouping markers with consecutive indices together. A single marker without adjacent non-2:2 markers will be a tract of its own.

Step 19. Identify type 3 and 4 GCs: single and double BIRs

In both lists from step 18, CrossOver identifies GC tracts that contain the first or the last marker of each chromosome. Tracts in the 3:1 or 1:3 marker list that fall at either the end of the chromosome will be categorized as type 3 GCs, which are BIR tracts that result from a single BIR event. Tracts from the 4:0 or 0:4 marker list will be listed as type 4 GCs, which are BIR tracts with two BIR events. The position of both types of BIRs is estimated by the midpoint between the end of the chromosome on one end of the tract and the midpoint of the genotype switch markers that defines the other end of the BIR tract (Figure 3-3d). The tract length will be the difference between the two.

Step 20. Filter any non-2:2 marker tract that falls within a CO junction

Any non-2:2 marker tracts that fall within a CO junction, as defined by the 2:2 switch markers of a CO (see step 5), are already analyzed along with the COs. Therefore, it would be redundant to analyze it further and are filtered out at this step.

Step 21. Identify conversion tracts near sites of CO

GC tracts with 3:1 or 1:3 marker segregation that are found within a close range of a CO site may have arisen from heteroduplex DNA near the double Holliday junctions and share the same DSB as the CO. Therefore, these conversion tracts are sorted into separate categories from the NCOs.

The default range for selecting conversion tracts within a close range of a CO is set at 5 kb. This was determined by investigating the CO-to-NCO distance for COs whose immediate recombination neighbor is a NCO in two wild-type tetrads whose recombination events were mapped through next-generation sequencing (see chapter 5 of this dissertation). While most CO-to-NCO distances are larger than 10 kb, a small population of distances clustered between 2 to 5 kb. Thus, we hypothesize that the NCOs within 5 kb range of a crossing over site may have arisen from the same double-strand break that resolved into the CO.

Step 22. Identify type 0 GCs: NCO

GC tracts with 3:1 or 1:3 marker segregation that are not within a close range of a CO site are categorized as type 0 GCs—the NCOs. NCO position is calculated by taking the midpoint of the two midpoints of the genotype switch markers that defined the NCO. Tract length is the difference between the two midpoints.

Step 23. Identify type 6 and type 7 GCs

GC tracts with 3:1 or 1:3 marker segregation that are found within a close range of a CO site are further sorted into two categories depending on which chromatid the GC is on with respect to the two chromatids involved in crossing over. GCs that are on a chromatid

involved in crossing over are categorized as type 7 GCs, and GCs that are on a chromatid not involved in crossing over are sorted as type 6 GCs. The GC tract length is determined as described for a NCO.

Step 24. Identify type 2 GCs: premeiotic GC

GC tracts with 4:0 or 0:4 marker segregation that are not BIRs and do not fall within a CO junction are categorized as type 2 GCs—the premeiotic GCs. The position and tract lengths of a premeiotic GC is calculate similar to a NCO (see step 21).

Functions of the CrossOver Program

The following section describes the different outputs of the CrossOver program. It will also introduce additional functions that calculate different metrics of the various meiotic recombination products.

Raw outputs for COs and GCs

CrossOver outputs a raw data file that contains the list of all the COs found. It also identifies the chromosomal position, the CO type, which two chromatids did the CO exchange occurred, and if any, the tract length of associated GC and the number of markers that defined the associated GC.

CrossOver also outputs a raw data file that contains the list of all GCs. For each conversion tract, it identifies the chromosomal location, the conversion type, the tract length, the minimum and maximum possible tract lengths, the number of markers

converted in each tract, and the number of YJM alleles found in the first marker of the tract (for calculation of allele parity).

Statistical analysis of COs and GCs

CrossOver also creates summary reports detailing the key statistics of COs and GCs. The CO report contains the follow calculations:

- CO statistics: total count, average count per tetrad, and the standard deviation.
- CO with associated GC statistics: total count, average count per tetrad, standard deviation, and the percentage of CO with associated GC.
- CO per chromosome statistics: for each of the 16 chromosomes, it calculates the average number of CO per chromosome, the standard deviation, and the CO density measured in cM per kb.
- Nonexchange chromosome (E0) calculation: a tally of the number of tetrads that do not have at least one CO for each of the 16 chromosomes.
- CO types analysis: total number of COs per type, average number of COs per type, the standard deviation, and the percentage of each type of COs out of the total.

For each of the seven GC types, CrossOver outputs the following calculations for the GC report:

- GC type statistics: total count of the GC type, average number of the GC type per tetrad, and the standard deviation.
- GC per chromosome statistics: average number of the GC type per chromosome and standard deviation.

- Chromosomes without a certain type of GC: a tally of the number of tetrads that do not have the GC type for each of the 16 chromosomes (a matrix similar to the E0 calculation for COs).
- Tract lengths statistics: the average, medium and standard deviation of three types of tract lengths are calculated (Figures 3-3b and 3-3c): a) tract length calculated by taking the difference between the two midpoints of the genotype switch markers on either end of the tract; b) the minimum possible tract length calculated by taking the difference between the two inner markers of the four switch markers; c) the maximum possible tract length calculated by taking the difference between the two outer markers of the four switch markers.
- Number of markers per tract statistics: the average, median and the standard deviation of the number of markers per GC tract, and the number and percentage of tracts with greater than 1 marker.
- Parity calculation: calculates the number of YJM789 allele (out of all four chromatids) of the first marker of the GC tract. Reports the number of GC tracts whose first marker has 0, 1, 2, 3 or 4 alleles of YJM789. First marker is used based on the assumption that the segregation ratio remains the same throughout the tract.

Calculation of inter-event distances

The distance between adjacent meiotic recombination products has been used as a measurement for the strength of interference between recombination products (Chen et al., 2008; Mancera et al., 2008). Strong interference would allow few products to have

close neighbors while low or no interference would display a distribution of recombination products similar to that of a random distribution (Hillers, 2004). The CrossOver program allows users to indicate which type of recombination products to use as the input to calculate inter-event distances. The flexibility allows the calculation of CO interference, NCO interference, or interference between COs and NCOs, which is estimation for DSB interference. The program outputs a list of distances in which users can then use to graph histogram plots on a conventional graphing software.

Calculation of centromere-to-event and telomere-to-event distances

CrossOver also calculates the distance between user-specified recombination products to the centromere or to the nearest telomere. Centromere-to-event distances can be used to evaluate whether a particular type of event is repressed or elevated near the centromere as compared to a known control. On the other hand, nearest telomere-to-event distances can be used to evaluate whether a particular type of event is repressed or elevated near the telomere as compared to a known control. The program outputs a list of distances in both calculations. In our study of the *zip1* mutant, we observed a relief of centromeric repression in both COs and NCOs (Chen et al., 2008). One *zip1* tetrad was sufficient to demonstrate the drastic relief of centromeric repression in *zip1* mutant.

Interference

Interference is measured by fitting inter-event distances with a gamma distribution function characterized by a shape (γ) and a scale (β) parameter (Chen et al., 2008). The gamma distribution describes the distribution of intervals between successive random

events, and has been a choice in the meiosis field to describe the distribution of intervals between successive meiotic recombination products. The shape parameter (γ) is used as a measure of the strength of interference, where a value of $\gamma = 1$ corresponds to no interference and $\gamma > 1$ indicates positive interference (McPeek and Speed, 1995; Zhao et al., 1995b). CrossOver calculates the shape (γ) and scale (β) parameter based on the inter-event distances.

Chromatid Interference

Chromatid interference is a measurement of whether a CO event between any two nonsister chromatids affects the probability of those chromatids being involved in neighboring COs (Zhao et al., 1995a). To calculate chromatid interference, the ratio of two-, three-, and four-strand double COs between adjacent COs is tallied. The expected ratio of two-, three-, and four-strand double COs for the case of no chromatid interference is 1:2:1. Genetic studies have revealed no chromatid interference for wild type tetrads (Perkins, 1962). The CrossOver program outputs the observed two-, three-, and four-strand double COs ratios in raw counts, and also gives the expected ratio in counts for the case of no chromatid interference based on the number of two CO neighbors. The users can then use these numbers for a Chi-square test to determine whether there is chromatid interference.

CO Homeostasis

CO homeostasis is a phenomenon by which the number of COs are maintained within a narrow range of fluctuation despite fluctuations in the number of DSBs from cell to cell

(Martini et al., 2006). A good test for CO homeostasis is the correlation between the number of COs and NCOs of each meiosis across an ensemble of tetrads, and can be assessed by the correlation coefficient which measures the intensity of association between two variables (Chen et al., 2008). An ideal CO homeostasis would give a correlation coefficient of zero, and no homeostasis would result in a correlation coefficient of one. CrossOver outputs a list of CO and NCO counts for each tetrad, and calculates the correlation coefficient between the two across a family of user-specified tetrads. At least 20 tetrads are needed to obtain statistically accurate correlation coefficients.

Conclusion

Advances in biotechnology such as microarrays and next-generation sequencing have opened opportunities in the field of genetics in unprecedented ways. Scientists studying meiosis move from using a handful of genetic markers to thousands, if not tens of thousands, of polymorphic markers detected using microarray and next-generation sequencing. The enormous flood of data necessitates the need for new bioinformatics tools to analyze meiotic recombination products. The CrossOver program was created to address this need.

A large number of polymorphic markers are now available for mapping meiotic recombination products in high resolution, at a genome-wide basis, in ways never seen before by traditional yeast geneticists. Different types of COs and GCs can now be elucidated in high resolution. This chapter explains the various motifs of meiotic recombination products identified by the CrossOver program. Several analyses of

recombination products are also explained and illustrated. This program is a powerful tool not only in analyzing meiotic recombination products of wild type strains, but is equally powerful in analyzing mutants with defects in meiotic recombination. Comparisons between meiotic mutants and wild-type strains will shed insight into the function lost in each mutation. The following two chapters of this dissertation describe the results from experiments using the microarray (chapter 4) and next-generation sequencing (chapter 5) technologies in analyzing the recombination products of wild type and various meiotic mutant progenies.

Acknowledgements

We thank Tetsuya Matsuguchi for programming advice in the early months of developing CrossOver, and Alan Barber for useful and practical lessons in the python programming language. Special thanks go to Ashwini Oke for helpful discussions on the different types of COs.

References:

Allers, T., and Lichten, M. (2001). Intermediates of yeast meiotic recombination contain heteroduplex DNA. Mol Cell *8*, 225-231.

Baudat, F., and de Massy, B. (2007). Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res *15*, 565-577.

Borts, R.H., Chambers, S.R., and Abdullah, M.F. (2000). The many faces of mismatch repair in meiosis. Mutat Res *451*, 129-150.

Borts, R.H., and Haber, J.E. (1987). Meiotic recombination in yeast: alteration by multiple heterozygosities. Science *237*, 1459-1465.

Borts, R.H., and Haber, J.E. (1989). Length and distribution of meiotic gene conversion tracts and crossovers in Saccharomyces cerevisiae. Genetics *123*, 69-80.

Chen, S.Y., Tsubouchi, T., Rockmill, B., Sandler, J.S., Richards, D.R., Vader, G., Hochwagen, A., Roeder, G.S., and Fung, J.C. (2008). Global analysis of the meiotic crossover landscape. Dev Cell *15*, 401-415.

Hillers, K.J. (2004). Crossover interference. Curr Biol *14*, R1036-1037.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell *88*, 375-384.

Malkova, A., Swanson, J., German, M., McCusker, J.H., Housworth, E.A., Stahl, F.W., and Haber, J.E. (2004). Gene conversion and crossing over along the 405-kb left arm of Saccharomyces cerevisiae chromosome VII. Genetics *168*, 49-63.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., and Steinmetz, L.M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature *454*, 479-485.

Martini, E., Diaz, R.L., Hunter, N., and Keeney, S. (2006). Crossover homeostasis in yeast meiosis. Cell *126*, 285-295.

McPeek, M.S., and Speed, T.P. (1995). Modeling interference in genetic recombination. Genetics *139*, 1031-1044.

Merker, J.D., Dominska, M., and Petes, T.D. (2003). Patterns of heteroduplex formation associated with the initiation of meiotic recombination in the yeast Saccharomyces cerevisiae. Genetics *165*, 47-63.

Oh, S.D., Lao, J.P., Hwang, P.Y., Taylor, A.F., Smith, G.R., and Hunter, N. (2007). BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. Cell *130*, 259-272.

Paques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev *63*, 349-404.

Perkins, D.D. (1962). Crossing-over and interference in a multiply marked chromosome arm of Neurospora. Genetics *47*, 1253-1274.

Roeder, G.S. (1997). Meiotic chromosomes: it takes two to tango. Genes Dev *11*, 2600-2621.

Symington, L.S., and Petes, T.D. (1988). Expansions and contractions of the genetic map relative to the physical map of yeast chromosome III. Mol Cell Biol *8*, 595-604.

Whitby, M.C. (2005). Making crossovers during meiosis. Biochem Soc Trans *33*, 1451-1455.

Winzeler, E.A., Richards, D.R., Conway, A.R., Goldstein, A.L., Kalman, S., McCullough, M.J., McCusker, J.H., Stevens, D.A., Wodicka, L., Lockhart, D.J., *et al.* (1998). Direct allelic variation scanning of the yeast genome. Science *281*, 1194-1197.

Zhao, H., McPeek, M.S., and Speed, T.P. (1995a). Statistical analysis of chromatid interference. Genetics *139*, 1057-1065.

Zhao, H., Speed, T.P., and McPeek, M.S. (1995b). Statistical analysis of crossover interference using the chi-square model. Genetics *139*, 1045-1056.

Figure Legends

Figure 3-1. Different types of COs and GCs identified by CrossOver.

(A) A schematic illustration of the nine different types of COs the CrossOver program can identify. Red and blue bars represent homologous chromosomes from different parental strains. Possible variations of each CO type are shown. Type 0 through type 4 are single COs, and type 5 through type 7 are double COs. Type 8 describes a special situation where two COs within close proximity of one and other may be the product of one CO accompanied by a NCO.

(B) A schematic illustration of the seven types of GCs that can be identified by CrossOver. For each GC type, conversion tract(s) are illustrated in blue. Possible variations of each GC type are included.

Figure 3-2. A flowchart diagram delineating the programming strategy of CrossOver. The step-by-step strategy employed in identifying different types of COs (A) and GCs (B) is delineated. CO(-) denotes CO without an associated GC, whereas CO(+) denotes CO with associated GC. Circled numbers point to key steps in the analysis and correspond to the step numbers used in the programming strategy section of the text. Representative drawing for each type of CO and GC is shown.

Figure 3-3. Examples of how CO and GC positions and tract lengths are computed.

Examples of how CO position, GC position, GC tract length, maximum possible GC tract length, minimum possible GC tract length, and the number of markers that defined a GC tract are computed. The following types of recombination products are chosen as examples: (A) type 0 CO: single CO without associated GC, (B) type 1 CO: single CO with associated GC, (C) type 2 CO: single CO without associated GC accompanied by a NCO, and (D) type 3 GC: single BIR defined by markers segregated 3:1 or 1:3. Gray boxes surround the genotype switches that define a CO junction. Yellow and blue boxes surround genotype switches that define a NCO and a BIR respectively. The midpoint of two markers are marked with a “>”.

Figure 3-1

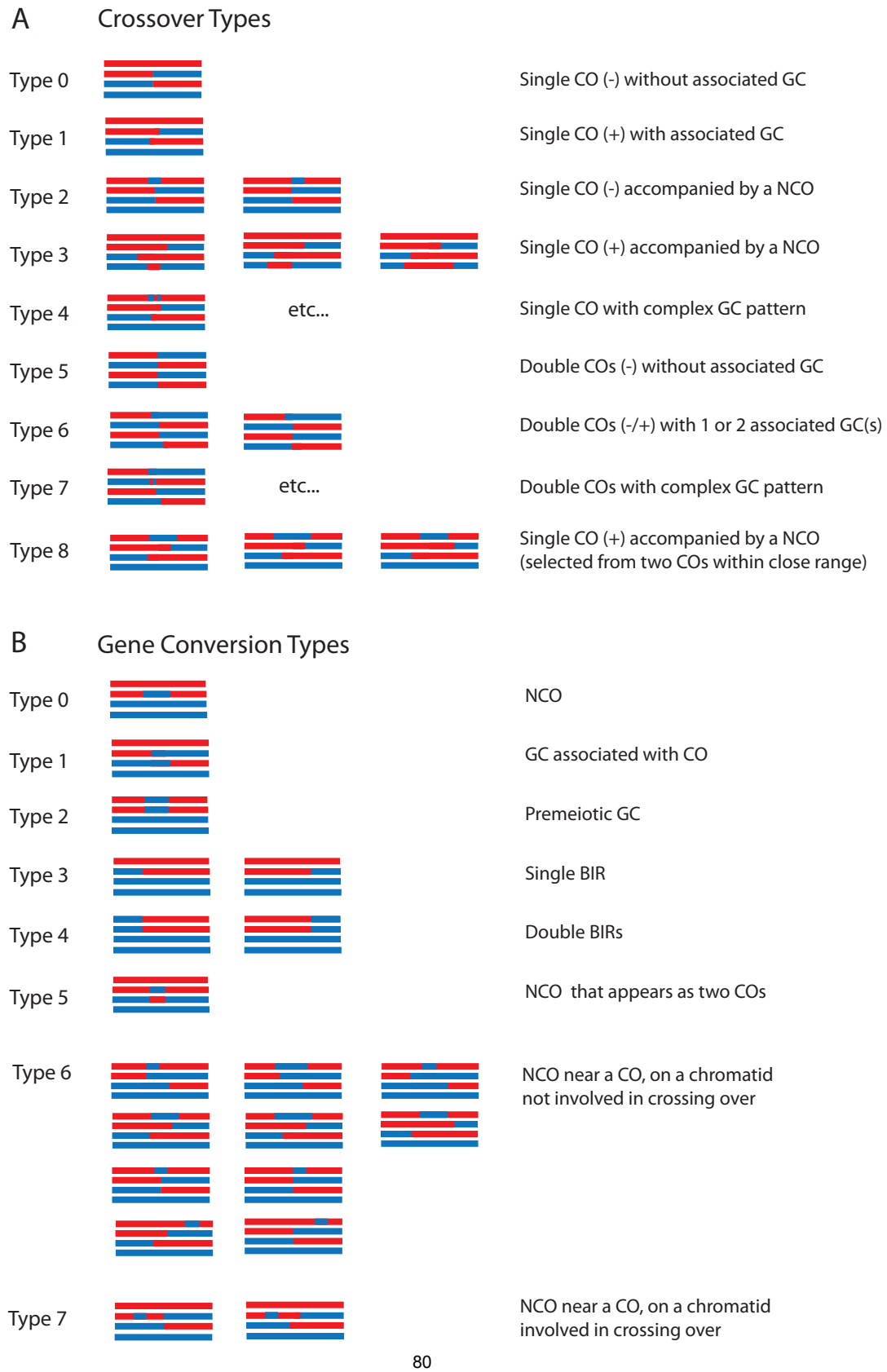

A
Crossover Types
Type 0
Single CO (-) without associated GC
Type 1
Single CO (+) with associated GC
Type 2
Single CO (-) accompanied by a NCO
Type 3
Single CO (+) accompanied by a NCO
Type 4
etc...
Single CO with complex GC pattern
Type 5
Double COs (-) without associated GC
Type 6
Double COs (-/+) with 1 or 2 associated GC(s)
Type 7
etc...
Double COs with complex GC pattern
Type 8
Single CO (+) accompanied by a NCO
(selected from two COs within close range)
B
Gene Conversion Types
Type 0
NCO
Type 1
GC associated with CO
Type 2
Premeiotic GC
Type 3
Single BIR
Type 4
Double BIRs
Type 5
NCO that appears as two COs
Type 6
NCO near a CO, on a chromatid
not involved in crossing over
Type 7
NCO near a CO, on a chromatid
involved in crossing over

Figure 3-2a

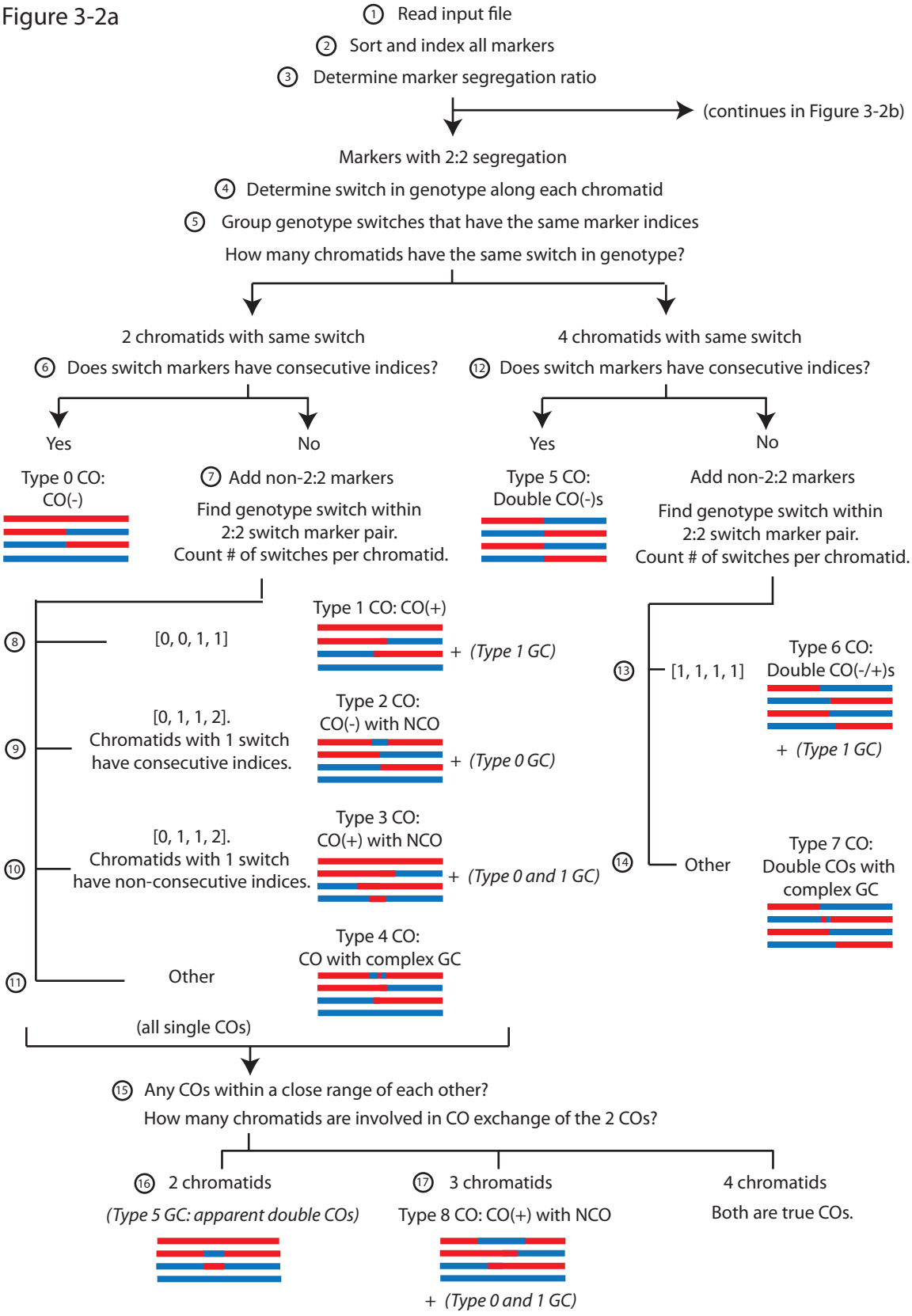
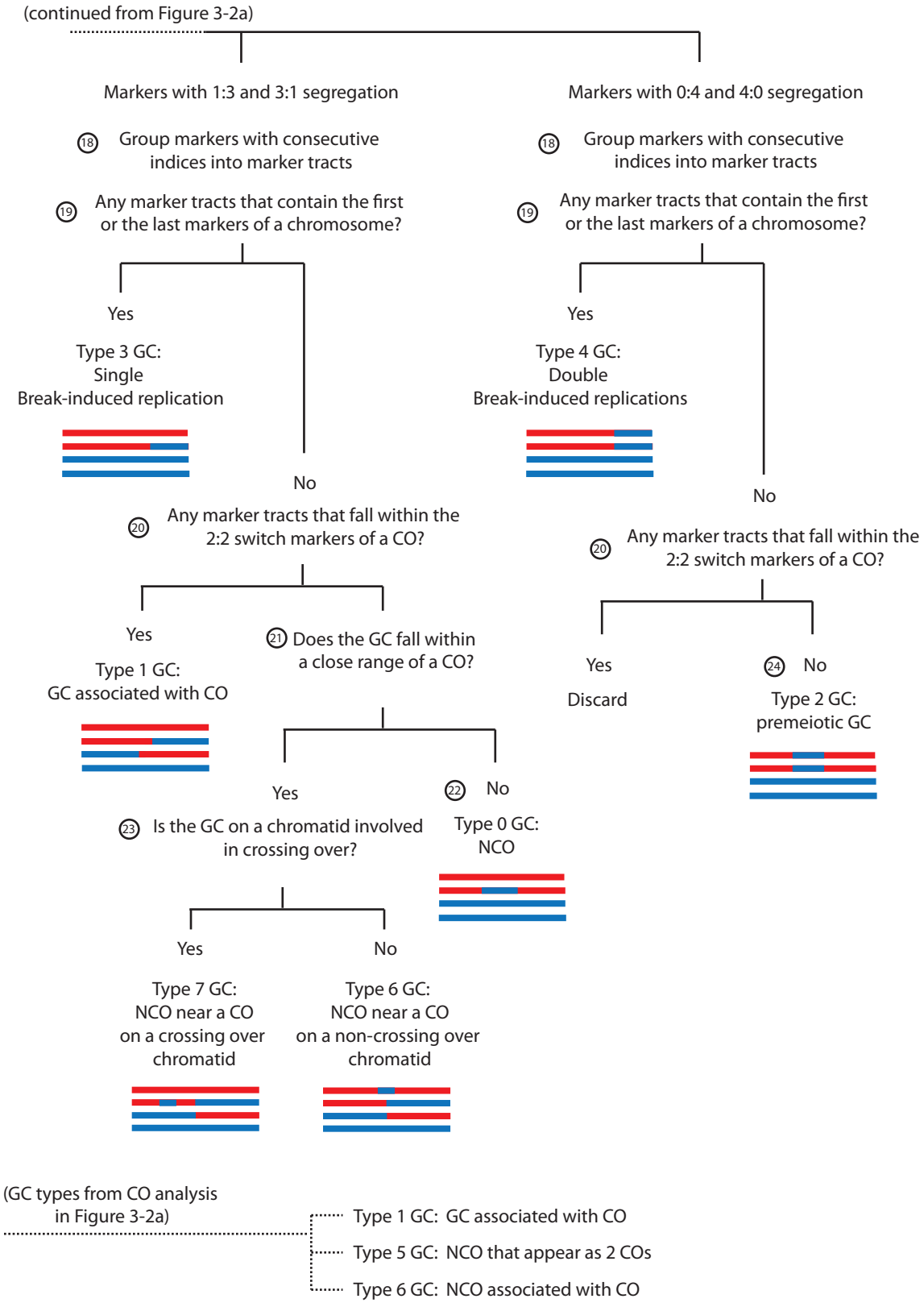

1 Read input file
2 Sort and index all markers
3 Determine marker segregation ratio
(continues in Figure 3-2b)
Markers with 2:2 segregation
4 Determine switch in genotype along each chromatid
5 Group genotype switches that have the same marker indices
How many chromatids have the same switch in genotype?
2 chromatids with same switch
4 chromatids with same switch
6 Does switch markers have consecutive indices?
12 Does switch markers have consecutive indices?
Yes
No
Yes
No
Type 0 CO: CO(-)
7 Add non-2:2 markers
Find genotype switch within 2:2 switch marker pair. Count # of switches per chromatid.
Type 5 CO: Double CO(-)s
Add non-2:2 markers
Find genotype switch within 2:2 switch marker pair. Count # of switches per chromatid.
8 [0, 0, 1, 1]
Type 1 CO: CO(+)
+ (Type 1 GC)
9 [0, 1, 1, 2]. Chromatids with 1 switch have consecutive indices.
Type 2 CO: CO(-) with NCO
+ (Type 0 GC)
10 [0, 1, 1, 2]. Chromatids with 1 switch have non-consecutive indices.
Type 3 CO: CO(+) with NCO
+ (Type 0 and 1 GC)
11 Other
Type 4 CO: CO with complex GC
(all single COs)
13 [1, 1, 1, 1]
Type 6 CO: Double CO(-/+)s
+ (Type 1 GC)
14 Other
Type 7 CO: Double COs with complex GC
15 Any COs within a close range of each other?
How many chromatids are involved in CO exchange of the 2 COs?
16 2 chromatids
(Type 5 GC: apparent double COs)
17 3 chromatids
Type 8 CO: CO(+) with NCO
+ (Type 0 and 1 GC)
4 chromatids
Both are true COs.

Figure 3-2b



(GC types from CO analysis in Figure 3-2a)

- Type 1 GC: GC associated with CO
- Type 5 GC: NCO that appear as 2 COs
- Type 6 GC: NCO associated with CO

Figure 3-3

A Type 0 CO: Single CO without associated GC

Chr#	Position		Spore 1	Spore 2	Spore 3	Spore 4
1	23813		1	0	1	0
1	23814		1	0	1	0
1	23815	> 26896	1	0	1	0
1	29977		1	1	0	0
1	29980		1	1	0	0
1	32650		1	1	0	0

CO position = 26896

B Type 1 CO: Single CO with associated GC

Chr#	Position		Spore 1	Spore 2	Spore 3	Spore 4
8	315879		1	0	0	1
8	315881	> 322754	1	0	0	1
8	329627		1	0	1	1
8	329633	>333711.5	1	0	1	1
8	337790		1	0	1	0
8	337796		1	0	1	0

CO position = (322754 + 333711.5) / 2
GC tract length = 333711.5 - 322754
GC tract length$_{max}$ = 337790 - 315881
GC tract length$_{min}$ = 329633 - 329627
GC marker # = 2

C Type 2 CO: Single CO(-) accompanied by a NCO

Chr#	Position		Spore 1	Spore 2	Spore 3	Spore 4
8	46056	> 46062	1	1	0	0
8	46068		0	1	0	0
8	46074	>51156.5	0	1	0	0
8	56239		0	0	1	0
8	56242	>56240.5	1	0	1	0
8	56245		1	0	1	0

CO position = 51156.5
NCO tract length = 56240.5 - 46062
NCO tract length$_{max}$ = 56242 - 46056
NCO tract length$_{min}$ = 56239 - 46068
NCO marker # = 3

D Type 3 GC: Single BIR

Chr#	Position		Spore 1	Spore 2	Spore 3	Spore 4
10	690770		1	0	1	0
10	690776	>697359.5	1	0	1	0
10	703943		0	0	1	0
10	704893		0	0	1	0
10	704905		0	0	1	0
10	704908		0	0	1	0
10	704920		0	0	1	0
10	708148		0	0	1	0

End of Chr 10: 745745
BIR position = (745745 + 697359.5) / 2
BIR tract length = 745745 - 697359.5
BIR tract length$_{max}$ = 745745 - 690776
BIR tract length$_{min}$ = 745745 - 703943
BIR marker # = 6

Chapter 4: Global Analysis of the Meiotic Crossover Landscape using DNA Microarrays

Stacy Y. Chen[1], Tomomi Tsubouchi[2,3], Beth Rockmill[2,5], Jay S. Sandler[1], Daniel R. Richards[4], Gerben Vader[7], Andreas Hochwagen[7], G. Shirleen Roeder[2,5,6] and Jennifer C. Fung[1*]

[1]Department of Biochemistry and Biophysics
University of California, San Francisco, San Francisco, California, USA
[2]Howard Hughes Medical Institute
Yale University, New Haven, Connecticut, USA
[3]Present address:
Marie Curie Research Institute, DNA Recombination Group
Oxted, Surrey RH8 0TL, UK
[4]Ingenuity Systems, Inc.
Redwood City, CA 94063, USA
[5]Department of Molecular, Cellular, and Developmental Biology
Yale University, New Haven, CT 06520, USA
[6]Department of Genetics
Yale University, New Haven, Connecticut, USA
[7]Whitehead Institute for Biomedical Research
Cambridge, Massachusetts 02142

Abstract

Tight control of the number and distribution of crossovers is of great importance for meiosis. Crossovers establish chiasmata, which are physical connections between homologous chromosomes that provide the tension necessary to align chromosomes on the meiotic spindle. Understanding the mechanisms underlying crossover control has been hampered by the difficulty in determining crossover distributions. Here, we present a microarray-based method to analyze multiple aspects of crossover control simultaneously and rapidly, at high resolution, genome-wide, and on a cell-by-cell basis. Using this approach, we show that loss of interference in *zip2* and *zip4*/*spo22* mutants is accompanied by a reduction in crossover homeostasis, thus connecting these two levels of crossover control. We also provide evidence to suggest that repression of crossovers at telomeres and centromeres arise from different mechanisms. Lastly, we uncover a surprising new role for the synaptonemal complex component, Zip1, in repressing crossing over at the centromere.

Introduction

As part of sexual reproduction, diploid parents undergo meiosis to produce gametes with a haploid complement of chromosomes (Roeder, 1997; Zickler and Kleckner, 1999). Central to this process is the segregation of homologous chromosomes at the first meiotic division. During prophase I, a high level of recombination is induced through the formation of double-strand breaks (DSBs) via the Spo11 protein (Keeney et al., 1997). A significant fraction (~half in budding yeast) of DSB repair events is accompanied by crossing over. Crossovers (COs) establish chiasmata, which are physical connections

between homologs that promote proper chromosome segregation by correctly aligning chromosomes on the meiosis I spindle. Failure to sustain a CO on each pair of chromosomes can result in the production of aneuploid gametes; in humans, this leads to infertility, miscarriage and developmental disabilities (Hassold, 2007).

To ensure that each chromosome pair receives at least one CO, crossing over is highly regulated. In most organisms, the spatial distribution of COs is tightly controlled through a process known as CO interference (Hillers, 2004; Jones, 1984; Muller, 1916). Interference ensures that COs are distributed nonrandomly along chromosome pairs to attain a more regular spacing between COs than would be expected for a random distribution. As a result, COs seldom occur close together.

Another manifestation of CO control is CO homeostasis, first described by Martini et al. (2006) as the means whereby normal levels of COs are maintained despite lowering the overall number of DSB-initiating events. CO homeostasis presumably reduces the chances of nondisjunction by ensuring that sufficient numbers of COs are made. Still unknown is how CO homeostasis is achieved or what its relationship is to interference since no mutants have been described that affect this process.

In spite of the importance of CO control, its molecular mechanisms remain elusive due, in large part, to the lack of an efficient and accurate way of measuring CO distribution. A typical method for measuring interference in budding yeast requires the manual dissection of tetrads containing the four progeny of a single meiosis. Only those tetrads that produce four viable spores are then scored for a limited number of genetic markers. Each tetrad is classified as having the parental ditype, tetratype or nonparental ditype (NPD) arrangement of markers for each interval. To calculate interference, a NPD

ratio is determined, which is the number of NPDs observed (~ equivalent to double COs) divided by the number of NPDs expected based on the frequency of tetratypes (~ equivalent to single COs) if COs were distributed randomly (Papazian, 1952). Accurate measurement of the NPD ratio requires dissection of large numbers of 4-spore viable tetrads (typically hundreds to thousands), making the assessment of interference relatively difficult (Ott, 1991). Furthermore, meiotic mutants with defects in crossing over typically show poor spore viability, drastically reducing the number of 4-spore viable tetrads that can be obtained. As a result, mutants that might affect interference (e.g., mutants in recombination, chromosome structure and synaptonemal complex assembly) are not routinely analyzed for interference defects.

An alternative method for measuring COs, that can be applied to analyze interference, is direct allelic variation scanning of the genome (Winzeler et al., 1998). This method uses the nucleotide sequence variation between two yeast strains to evaluate the parental origins of progeny DNA resulting from a cross between them. By hybridizing total genomic DNA from the two different strains of yeast to high-density oligonucleotide arrays, Winzeler and coworkers identified a total of 3714 markers capable of distinguishing between the two strains. The inheritance pattern of these markers in the progeny strains was used to locate COs. The distribution of distances between adjacent COs can be used to measure interference. The advantage of this method is that very few 4-spore viable tetrads would be needed to analyze interference, since interference would be assessed from all COs genome-wide, rather than from a few marked intervals.

Of the few mutants that have been examined genetically for loss of interference, at least two affect proteins that are components of the synapsis initiation complex (SIC), namely Msh4 and Zip4/Spo22 (hereafter referred to as Zip4) (Novak et al., 2001; Tsubouchi et al., 2006). SICs promote chromosome synapsis by facilitating polymerization of Zip1, a major building block of the synaptonemal complex (Sym et al., 1993). Zip2, Zip3, Zip4, Msh4 and Msh5 (a.k.a., ZMM proteins (Lynn et al., 2007)) are all components of the SIC. Mutations affecting all known SIC components reduce crossing over, and SICs display interference (Fung et al., 2004), suggesting that SICs are the same as, or associated with, a large CO-promoting complex assembly known as the late recombination nodule (Carpenter, 1988).

Until recently, a reasonable assumption would have been that all SIC mutants show the same level of interference since all show a similar reduction in crossing over. However, a recent study of the *zip4* mutant (Tsubouchi et al., 2006) reported evidence for negative interference, which differs from the absence of interference found for *msh4* (Novak et al., 2001; Sym and Roeder, 1994). Negative interference implies a different kind of nonrandom distribution, where COs are clustered together, instead of being spaced far apart. However, apparent negative interference can arise from variations in CO frequencies within a population of cells showing no interference (Sall and Bengtsson, 1989), an aspect that is difficult to assess genetically. Adding to the confusion, a more recent study of interference in several ZMM mutants reports normal interference for *zip4* (Shinohara et al., 2008). A key benefit of the microarray analysis is its ability to address whether variations in recombination exist within a population since the analysis is

performed on a cell-by-cell basis, unlike genetic measurements that are inherently population-based; thus, apparent vs. true negative interference can be distinguished.

Besides interference and homeostasis, many organisms have additional mechanisms to modify the CO landscape that can potentially influence CO control. Both recombination hotspots (Petes, 2001) and the suppression of COs near telomeres (Su et al., 2000) and centromeres (Lambie and Roeder, 1986) are known to contribute to the nonuniformity of CO distribution. Crossing over near centromeres and/or too far from them can be detrimental to chromosome segregation and increases the risk of producing aneuploid progeny (Koehler et al., 1996a; Lacefield and Murray, 2007; Lamb et al., 1996; Rockmill et al., 2006). Analysis of COs in the vicinity of telomeres and centromeres could be greatly aided by a genome-wide approach in which crossing over near these chromosomal landmarks can be easily assessed.

In this chapter, we show that mapping COs by DNA microarrays is a powerful approach for assessing CO control. We show that all metrics of crossing over previously determined genetically can be recapitulated with this genomic approach. Gene conversions (GCs) can also be assessed, but in a more limited fashion than COs. For the first time, we identify mutants, *zip2* and *zip4*, that show a reduction in CO homeostasis. Our analyses of COs and NCOs (GCs not associated with COs) at telomeres and centromeres suggest that different mechanisms are responsible for CO repression at these sites. At telomeric ends, COs are repressed by changing the relative proportions of COs vs. NCOs, while COs near centromeres are reduced most likely by favoring repair between sister chromatids versus inter-homolog repair. Finally, we show that this centromeric repression is dependent on Zip1.

Results

Genome-wide Analysis of Recombination Using DNA Microarrays

The tetrads genotyped in this study resulted from a cross between a standard laboratory strain, S96 (an S288c derivative), and a clinical isolate, YJM789 (Wei et al., 2007). The sequence difference between these strains (0.6%) is high enough to achieve the resolution required to detect COs, but not so high as to act as a barrier to recombination (Figure A1-1 and Appendix 1). Spore viability and sporulation frequency are provided in Table A1-1 for strains derived from these parents. Sequence differences between the two parental strains were used to determine the parental origin of progeny DNA in each tetrad.

In this study, about 8000 markers (probe sequences), whose hybridizations show differential signals between the two parental strains, were scored. The mean distance between markers is 1.5 kb (~0.5 cM); overall, markers are uniformly distributed across the genome with only a few noticeable gaps (Figure 4-1A). The distribution of inter-marker distances is shown in Figure 4-1B. Because 4-spore viable tetrads are examined, markers showing reciprocal exchange can be unambiguously identified as COs; markers showing 3:1 and 1:3 configurations are identified as GCs, whereas 4:0 and 0:4 configurations often indicate premeiotic recombination events.

Microarray data from 26 wild-type tetrads show that, on mean, 98.0% of the markers segregate 2:2; 2.0% of the markers segregate 1:3 or 3:1, and less than 0.1% of the markers segregate 4:0 or 0:4 (Table A1-2), in good agreement with genetic data that reports 95% of markers segregating 2:2 and 4.8% showing non-2:2 segregations (Fogel et al., 1978).

Good Agreement Found for CO Frequency and Density

Examination of CO frequency reveals a mean of 95 (± 10 SD) COs per meiosis (Figure 4-1C, Table 4-1), on par with the 86 COs per meiosis computed from map distances compiled from several genetic studies (Cherry et al., 1997) (Yeast Genome Database). The slightly greater value for COs seen here may be due to a better overall marker resolution compared to the marker resolution of the genetic map. Alternatively, the slight increase in map distance might reflect increased numbers of events due to repeated cycles of heteroduplex rejection characteristic of polymorphic strains (Borts and Haber, 1987). Figure 4-1C shows good agreement of CO frequency on a per chromosome basis. A plot of CO density against chromosome size reveals that smaller chromosomes have a higher density of COs than larger chromosomes (Figure 4-1D), a trend consistent with previous genetic observations (Kaback et al., 1992).

No Chromatid Interference

Unlike standard genetic analysis using phenotypic markers, the microarray approach allows a straightforward analysis of chromatid interference (where a CO between any two nonsister chromatids affects the probability of those chromatids being involved in neighboring COs) since the chromatids involved in each CO are known. Previous studies report no chromatid interference in wild-type strains as assayed by the ratio of two-, three-, and four-strand double COs between adjacent COs (Perkins, 1962). In wild type (Table 4-1), we see no difference from the 1:2:1 ratio expected for no chromatid interference ($\chi^2 = 1.46, P = 0.5$), consistent with previous genetic studies.

Repression of COs near Telomeres and Centromeres

Telomere- and centromere-proximal regions have reduced CO frequency relative to the rest of the chromosome (Lambie and Roeder, 1986, 1988; Su et al., 2000). To determine whether our microarray data detects a reduction in COs in these regions, we examined the distribution of telomere-CO and centromere-CO distances. The distance between every CO and the nearest chromosome end (determined from SGD) was obtained and the resulting histogram is shown in Figure 4-2A. We observe a 7-fold repression within 20 kb of the chromosome end, as compared to regions further away from the telomeres. Elevated CO levels as compared to what was expected for a simulated distribution were seen 20-140 kb away from the chromosome end (Figure 4-2A), in agreement with a recent study of crossing over at chromosome ends (Barton et al., 2008). To determine whether this elevation of CO frequency is due to the inclusion of small chromosomes that have a higher CO density than other chromosomes, we reanalyzed the telomere-CO distances excluding the four smallest chromosomes (Figure 4-2B). Removal of the smallest chromosomes eliminated most of the observed elevation in CO frequency; however, some elevation of CO frequency remained, though at defined intervals 40-60 kb and 140-160 kb away from the ends.

Recent analyses of genome-wide DSB hotspot distributions (Blitzblau et al., 2007b; Buhler et al., 2007) reported a ~2-fold repression of DSBs within 20 kb of the chromosome end. Such a repression of DSBs could contribute to the observed lower level of crossing over. However, when telomere-NCO distances were examined, no concomitant repression of NCOs is seen in the 20 kb region nearest the chromosome end,

instead the NCO level is within the range predicted by the simulation and in accordance with the level found in neighboring intervals (Figures 4-2C and 4-2D). The fact that DSBs level are repressed, but NCO levels remain unchanged, suggests that the repression of COs reflects a change in the CO:NCO ratio (in favor of NCOs) rather than an alteration in overall levels of DSBs or a switch from inter-homolog to inter-sister repair.

Centromeric repression of meiotic recombination has been well documented in budding yeast (Lambie and Roeder, 1986) and other higher eukaryotes (Hassold et al., 1996; Koehler et al., 1996b). To test whether CO repression at the centromere can be seen in the wild-type distribution of COs, we measured the centromere-CO distance for every CO. Figure 4-2E and 4-2F show that crossing over within 10 kb from the centromere is decreased 6-fold, compared to neighboring intervals greater than 10 kb away. Unlike at the telomere, measurements of centromere-NCO distances do show a repression of NCO frequency (6-fold) at the most proximal interval to the centromere (Figures 4-2G and 4-2H). Therefore, CO repression is less likely to occur via modification of the CO:NCO ratio as at telomeres, but is more likely to result from mechanisms that either alter the number of DSBs or change the bias from inter-homolog to inter-sister repair. Thus, the mechanisms by which CO repression is attained at the centromere and the telomere appear to be different.

Determination of CO Interference with Only a Few Tetrads

The foregoing results show that the microarray-based analysis can recapitulate previous measurements of CO frequency. But can microarray-based measurements recapitulate numerical estimates of interference in wild type? Inspection of the microarray results

shows that wild-type COs are relatively evenly spaced and no chromosome is without at least one CO (Figure 4-3A), indicating that CO distribution is regulated in a manner qualitatively consistent with the existence of interference (compare with Figure 4-3B showing a loss of interference). Quantitative comparison is more difficult because the NPD ratio, which is a well-known metric for interference, is an inherently population-based measure, requiring large numbers of tetrads for reliable statistics. Because our measurements are based on analyzing a small number of tetrads, we could not directly calculate the NPD ratio for any given marker pair with statistical accuracy. Instead, to determine whether the level of interference obtained by microarrays is quantitatively similar to that obtained genetically via NPD ratios, we employed a method in which interference measured by inter-CO distances is converted into a NPD ratio using Monte Carlo simulation.

Briefly, inter-CO distances were measured and fitted with a gamma distribution function characterized by a shape (γ) and scale (β) parameter. The gamma distribution arises in statistical studies of the distributions of intervals between successive random events; hence, it is a natural choice for a distribution to describe the intervals between successive COs. The gamma distribution is a useful tool for estimating interference levels since γ itself can be used as a measure of the strength of interference. A value of $\gamma = 1$ corresponds to no interference whereas $\gamma > 1$ indicates positive interference with larger values of gamma indicating stronger interference (McPeek and Speed, 1995; Zhao et al., 1995). Experimentally obtained inter-CO distances are well fit by the gamma distribution for wild type (Figure 4-4A; $\chi^2 = 4.2$, $P > 0.99$, Figure A1-2A (smaller bin size)) and for *zip4* (Figure 4B; $\chi^2 = 0.63$, Figure A1-2B).

The parameters of the gamma function do not directly tell us the value expected for the NPD ratio; hence, we used a simulation-based approach to estimate the NPD ratio from the gamma distribution. From the best-fit parameters of the gamma distribution, a conditional probability function (hazard function) was determined that gives the probability of a CO arising at a particular distance from a pre-existing CO (Figure 4-4C, details on the gamma distribution is given in the supplementary material). This function was then used as the basis for simulating CO positions for a large population of tetrads to back-calculate a simulated value for the NPD ratio (see Supplemental Procedures for details on the simulation of NPD ratios). Applying this analysis to wild-type inter-CO distances, a best-fit gamma value of 1.94 was found; this in turn, gave a simulated NPD ratio value of 0.38, which is in good correspondence with the mean NPD ratio of 0.32 obtained from published values of wild-type interference for intervals with a mean size of 30 cM. The gamma value of 1.94 concurs with a previously reported gamma value ($\gamma \sim 2$) for *Saccharomyces cerevisiae* (Foss and Stahl, 1995), confirming that interference in budding yeast is not as strong as in other organisms, such as Drosophila ($\gamma \sim 4$) (calculated in Foss and Stahl, 1995), *Arabidopsis thaliana* ($\gamma \sim 3$) (Copenhaver et al., 2002) or *Mus musculus* ($\gamma \sim 10$) (Broman et al., 2002; de Boer et al., 2006). Although this analysis encompassed data from all 26 wild-type tetrads, we find that even 3 tetrads provide a sufficient number of inter-CO distances (~ 250) to assess interference levels (data not shown). Figure A1-3 shows interference calculated from our microarray data by an adaptation of the method devised by Malkova et al. (2004) to measure the extent of interference on adjacent intervals. The maximum effective distance over which interference extends is ~150 kb in agreement with the 154.2 kb reported in the Malkova

study (Figure A1-3A). The effective distance over which interference acts can also be obtained directly from the hazard function (Figure 4-4C).

One final aspect of interference that could be tested is whether GCs unassociated with a CO, hereafter referred to as NCOs, show a lack of interference. Studies in fungi report that NCOs, unlike COs, do not exhibit interference (Malkova et al., 2004; Mortimer and Fogel, 1974). To see if a similar effect is seen with NCOs observed in DNA microarrays, we first eliminated any GCs associated with the formation of a CO ($GC_{CO}s$) before calculating distances between the remaining NCOs. Because there were on mean only 50 detectable GCs per tetrad (31 $GC_{CO}s$, 19 NCOs, Table 4-1), many more wild-type tetrads were needed (~26) to accumulate enough inter-NCO distances to measure interference. The NCOs observed do not exhibit interference (γ =1.1, corresponding to a predicted NPD $ratio_{sim}$ = 0.9). By this analysis, NCOs observed by microarrays behave as expected based on tetrad analysis.

CO Homeostasis Measured from Microarray Data

CO homeostasis assures that CO numbers are maintained within a narrow range of fluctuation despite fluctuations in the number of DSBs from cell to cell. Analysis of the correlation between COs and NCOs provides a test for CO homeostasis by reporting the level of correlation between NCOs and COs in individual tetrads, over the ensemble of tetrads. The correlation coefficient is not a measure of quantitative change of one variable with respect to another, but it is a measure of intensity of association between two variables (see Experimental Procedures for more details). For ideal homeostasis, the number of COs would be independent of the number of NCOs, giving a correlation

coefficient of zero. No homeostasis would result in a correlation coefficient of one. The wild-type correlation coefficient is -0.07, indicating nearly ideal homeostasis, in agreement with an earlier observation for CO homeostasis (Martini et al., 2006).

Marker Resolution Influences GC Detection

Markers showing 3:1 or 1:3 configurations indicate a GC event. In general, contiguous markers having the same pattern of 3:1 or 1:3 chromatid arrangements are considered to be part of a single GC event. The mean number of events and mean tract length for both GC_{CO}s and NCOs are provided in Table 4-1; however, caution is warranted before making comparisons with the GC data if detection issues are not taken into account.

Although there is excellent detection of COs, GC detection is limited by our current marker density. If the mean GC tract length is 1.5 kb (Borts and Haber, 1987), but the mean distance between markers is only 1.5 kb (Figure 4-1B), our study will underestimate the actual frequency of GCs because some strand exchange events will fail to include a scorable marker. The GC comparisons presented here in this study either takes into account the detection issue or are not unduly affected by the detection limitation.

To estimate the proportion of GCs detected out of all GCs, we divided the mean number of NCOs (18.6) by an estimate of the total expected number of NCOs (66.1) based on a higher resolution tiling array analysis of the same wild-type strain (Mancera et al., 2008). This calculation results in a detection level of 28% of the actual number of GC events compared to the 70% detection of NCOs by Mancera et al. (2008). Since detection is not equal for GCs with small vs. long GC tract lengths, the subpopulation we do detect

will be biased towards GCs with longer tract lengths (Figure A1-4). One implication of this unequal detection of GC tracts is that any comparison made where there is a potential difference in GC tract lengths between the two populations must factor in how the change in detection might affect the comparison.

Conversion tract lengths differ between COs and NCOs (Baudat and de Massy, 2007). The medians of GC_{CO} and NCO tract lengths of wild type were compared (Tables 4-1 and A1-3). GC_{CO} tract lengths (4.4 kb) were found to be significantly larger than NCO tract lengths (3.9 kb), in agreement with observations in mice and humans (Guillon et al., 2005; Jeffreys and May, 2004).

CO Levels in Mutants Agree with Genetic Data, Except for *zip1*

To test the usefulness of the microarray analysis in measuring CO control in mutants, we looked at eight mutants with known or potential interference defects. The *zip1, zip4, msh4, ndj1* and *sgs1* mutants have been previously shown to be defective in interference, albeit to different extents (Chua and Roeder, 1997; Novak et al., 2001; Rockmill et al., 2003; Sym and Roeder, 1994; Tsubouchi et al., 2006) (Figure 4-5A). We also included *zip2* and *zip3*, whose gene products are part of the SIC (Agarwal and Roeder, 2000; Chua and Roeder, 1998), but whose levels of interference were unknown at the initiation of this study. In addition, we analyzed a mutation in the *SPO16* gene, which has recently been shown to encode a SIC component; the *spo16* mutant has been reported to show normal levels of CO interference (Shinohara et al., 2008).

In general, the change in CO levels for the mutants as determined by microarray agrees with values reported in prior genetic studies (Table 4-1) and with the genetic data

obtained in this study (Table A1-4). The only notable exception is *zip1*. Instead of the two-fold decrease in COs found for *zip1* in genetic and physical studies (Storlazzi et al., 1996; Sym et al., 1993), *zip1* shows an increase in COs (110 COs/tetrad) compared to wild type (95 COs/tetrad; Table 4-1). Our hypothesis is that in the case of *zip1*, only a selected population of cells (a subset with high levels of crossing over) produces tetrads in the S96/YJM789 diploid. Indeed, the frequency of asci containing four spores is orders of magnitude lower in *zip1* than in the mutants affecting SIC proteins (Table A1-1), suggesting the *zip1* mutant has additional difficulties not experienced by the SIC mutants. Consistent with the notion that we are looking at a selected subset of meioses in *zip1*, we find a 2-fold increase in NCOs in *zip1* as compared to other ZMM mutants (i.e. *zip2* and *zip4*).

Changes in GC Tract Lengths

All mutants, except *msh4* and *ndj1,* show increased NCO frequencies (Tables 4-1 and A1-5). Because an increase in NCO tract length could give rise to an apparent increase in NCO frequency, NCO tracts lengths were examined using a nonparametric multi-comparison median test (Levy, 1979) to determine if NCO tract lengths are significantly different between the different strains (Table A1-6). Only for wild type, *zip1, zip2, zip4* and *sgs1* were sample sizes large enough to perform this test. The results show that the NCO tract lengths of *zip1, zip2* and *zip4* are significantly greater than that of wild type (Table A1-6). Whether the ~2-fold difference seen in NCO frequencies in *zip2* and *zip4* vs. wild type can be entirely attributed to the increase in tract length remains to be seen. However, it is doubtful that an increase in tract length is responsible for the additional 2-

fold increase (above *zip2* and *zip4*) in NCO frequency seen for *zip1*, since no significant differences were seen among tract lengths for *zip1, zip2* and *zip4*. The same conclusions can be drawn for GC_{CO}s tract lengths (Table A1-6). Because the median NCO tract lengths in *sgs1* does not differ from that of wild type, the increase in NCO frequency in *sgs1* is likely a true increase in the number of NCOs and not an artifact of detection.

Analysis of CO Interference, Chromatid Interference and E_0s in Mutants

A representative example of CO distributions for a mutant (*zip4*) with reduced interference is shown in Figure 4-3B. In *zip4*, where loss of interference is expected (Tsubouchi et al., 2006), COs are less evenly spaced, despite the overall reduced number of COs. Figure 4-5A plots the array-derived interference values against the mean genetic values for all mutants. For comparison, published measurements of interference assayed genetically were used, except for *zip2*, *zip3* and *spo16* for which tetrads were dissected (Table A1-4). In most cases, microarray-based interference levels for the mutants agree well with the genetic data (Figure 4-5A). The two exceptions are *zip4* and *ndj1*. The *zip4* mutant shows a loss of interference, not normal interference or negative interference, both of which have been reported in different studies (Shinohara et al., 2008; Tsubouchi et al., 2006). In *ndj1*, wild-type interference is found, instead of a moderate decrease in interference (Chua and Roeder, 1997). The *spo16* mutant shows a decrease in interference similar to that shown by the other SIC mutants (Figure 4-5A). Examination of chromatid interference in the mutants showed no significant difference in the 1:2:1 ratio expected for no chromatid interference (Table 4-1). Lastly, all mutants show increased numbers of E_0s, defined as chromosome pairs that lack any COs (Supplemental

Results, Table A1-7). Because only 4-spore viable tetrads were examined in our microarray analysis, the number of E_0s represents a minimal estimate of the E_0 frequency. E_0s are seen more frequently for smaller chromosomes, although E_0s for larger chromosomes are observed as well. In the majority of tetrads, zero or one E_0 was the norm, although four E_0s are observed in one *msh4* tetrad (data not shown).

Negative Interference in *zip4* Mutant May Arise from Variations in CO Frequency

The *zip4* mutant has been reported to display negative interference, a phenomenon that can be explained either by the tendency of COs to cluster or by variation in CO frequency within a cell population having no interference (see Introduction). The latter effect can arise because measurements of NPD ratios require the assumption of a known and constant CO frequency. It is impossible to assess the cell-to-cell variations in CO frequency using population-based genetic techniques. However, the microarray approach enables analysis of individual meioses and thus is uniquely powerful in addressing such questions.

To assess whether *zip4* has true or apparent negative interference, CO number was examined on a tetrad-by-tetrad basis to look for outliers as evidence for the existence of a separate population of *zip4* tetrads with a higher CO frequency. Figure 4-5B shows the distribution of CO numbers per meiosis for wild type, *zip4* and *zip2* for which larger numbers of tetrads were available. An outlier is observed only for *zip4* and not for wild type or *zip2*. Table A1-8 shows how the inclusion of the outlier results in less interference than in the case where the outlier has been excluded. Although apparent negative interference arises when there is a variation in the recombination frequency within a

population having no interference, the effect is the greatest when only a small fraction of the population (<10%) has much larger recombination levels relative to the rest of the population (Figure 1 in Sall and Bengtsson, 1989). This is exactly what is seen in *zip4*, where one out of 34 tetrads exhibits a higher level of crossing over than the remainder of the population (Figure 4-5B). Taken together with the facts that regional clustering is not apparent in the CO spatial distribution (data not shown) and a loss of interference is observed by our approach, these considerations suggest that the negative interference observed genetically may result from the existence of more than one population of tetrads, rather than actual clustering of COs.

CO Homeostasis is Perturbed in *zip2* and *zip4*

CO homeostasis analysis was confined to mutants with sufficient number of tetrads, namely *zip2* and *zip4*. Any change in CO homeostasis would be reflected as an increase or decrease in the correlation coefficient. For both *zip2* and *zip4*, a decrease in CO homeostasis is indicated by significant increases in correlation coefficients, 0.44 and 0.34 respectively, as compared to -0.07 in the wild-type control (Figure 4-5C).

Centromeric Repression of Recombination Is Relieved in a *zip1* Mutant

Do any of the mutants relieve the telomeric or centromeric repression of COs? Of the eight mutants examined, only the *zip1* mutant has any effect at the centromere. Crossing over in *zip1* is no longer repressed in the 10 kb region closest to the centromere and is comparable to the levels of crossing over more distal to the centromere (Figure 4-6A). No relief of telomeric repression is seen in any of the mutants tested (data not shown).

Does Zip1 affect crossing over per se or does it prevent DSBs from occurring near centromeres? To answer this question, we compared the frequency of NCOs proximal to the centromere in wild type and *zip1*. In contrast to wild type (Figures 4-2G and 4-2H), the frequency of NCOs for *zip1* within 10 kb nearest the centromere is equal to the frequencies found in noncentromeric regions (Figure 4-6B), thus paralleling the increase in COs seen in *zip1*. The CO:NCO ratio in this proximal interval is not significantly different between wild type (1.31 +/-1.0 SE) and *zip1* (0.75 +/- 0.18) and thus is unaffected by the *zip1* mutation.

Genetic Measurements Confirm that NCO Levels Change at Centromeres in *zip1*

Given that our *zip1* strain shows higher levels of crossing over than expected based on genetic and physical data, it is possible that the high level of COs at centromeres is true only for the subpopulation of *zip1* cells that exhibit the overall high levels of crossing over. To address this concern, we performed a genetic analysis of recombination near the centromere of chromosome III in a BR1919 strain. The haploid parents are identical throughout the genome, except for a small number of well-defined genetic markers. In this strain background, the sporulation efficiency and spore viability of *zip1* is comparable to that of the SIC mutants.

To assay the level of recombination at the centromere, we used a strain carrying *URA3* heteroalleles adjacent to the centromere of chromosome III so that gene convertants (i.e., Ura^+ prototrophs) could be selected (Figure 4-6C). We found an 8-fold increase in Ura^+ recombinants in *zip1* relative to wild type (Figure 4-6D), strongly supporting the idea that interhomolog recombination is increased at centromeres in *zip1*

mutants. In comparison, no such increase was found for *zip2*. Thus, this genetic analysis concurs with our microarray analysis; moreover, it shows that the result is not inherent to the multiply heterozygous diploid nor is it a consequence of the aberrantly high levels of recombination observed in the *zip1* tetrads used for the microarray study.

Flanking markers were used to determine whether the selected GC events are associated with crossing over (Figure 4-6C). In wild type, Ura^+ gene convertants are associated with crossing over on mean 35% of the time (i.e., flanking marker exchange occurs in 35% of the Ura^+ spores) (Figure 4-6E). In *zip1*, only 18% on mean have associated COs, consistent with the two-fold reduction in crossing over reported in *zip1*. The frequency of crossing over also decreases in two centromere-distal intervals on chromosome III (Figure 4-6E), as expected for *zip1*. These results indicate that the fraction of DSB repair events resolved as COs is not increased in the centromere-adjacent interval in *zip1* and thus cannot be responsible for the increase in centromere-proximal COs in *zip1*. This concurs with our observation from the microarray analysis that the CO:NCO ratio is unchanged.

Alternatively, more DSBs occurring in the most centromere-proximal interval could explain the increased number of COs observed in the *zip1* mutant. Contradictory to that notion, no increase in DSB hotspots is seen at the most centromere-proximal region by a genome-wide study of DSB hotspots in a *dmc1 zip1* mutant (Blitzblau et al., 2007b). Three chromosomes examined by Southern analysis in a *zip1 dmc1* mutant also do not exhibit any increase in DSB activity in centromere-proximal regions as compared to the *dmc1* control (Figure 4-6F). Since neither a change in the CO:NCO ratio nor a change in

the number of DSBs is observed, these results implicate a shift from inter-sister to inter-homolog repair as the reason for the increase in COs at the centromere in a *zip1* mutant.

Discussion

Evaluation of the Microarray Approach

The microarray-based genome-wide detection system for COs is a powerful approach for gaining information about CO control for several reasons. First, many aspects of CO behavior can be evaluated simultaneously: information about CO and GC levels, CO interference, CO homeostasis, chromatid interference and crossing over in relationship to telomeres and centromeres, can all be obtained at the same time. Second, because COs are monitored genome-wide, many fewer tetrads are needed to generate statistically significant data compared to the hundreds to thousands of tetrads needed to get similar data genetically using conventional phenotypic markers. Third, analysis of CO control is relatively rapid; data can be acquired within two weeks of making a mutant diploid strain. Finally, cell-to-cell variations can be assessed, permitting the detection of important fluctuations that would otherwise be missed in assays looking at means in large populations.

On the other hand, there are some limitations to the microarray technique. In the microarray method, only a global determination of CO control can be assessed, since the data are derived from a relatively small number of tetrads. Only with a large number of tetrads can local variations in different intervals along a chromosome or among different chromosomes be measured. In fact, local vs. global observations of interference might account for differences found between genetic and microarray measurements. This could

explain how *spo16* and *zip4* could be observed to have normal interference in one study (Shinohara et al., 2008), but show a reduction in interference in our study. Interestingly, we do see a large local variation in interference for *spo16* in our BR1919-8B lab strain (Table A1-4). Completely opposing values of interference are observed in the two intervals we examined; the *HIS4-LEU2* interval shows a loss of interference (NPD = 1.3), whereas the *LEU2-MAT* interval shows normal interference (NPD = 0.35). There was no discordance in the one common interval between our study and that of Shinohara; both studies report normal interference in the *LEU2-MAT* interval. Another example where local vs. global evaluations might differ is in *ndj1*, which was shown previously to have somewhat impaired interference (Chua and Roeder, 1997), but exhibits normal interference in our genomic analysis. One possibility for the difference in interference seen in *ndj1* is the potential variation in interference on small vs. large chromosomes since the genetic study was carried out on only one small chromosome (III). Local variations might also account for the negative interference of *zip4* rather than the existence of a subpopulation, since technically, there is no statistically significant difference between 1 outlier in 34 tetrads (*zip4*) vs. 0 outliers in 26 tetrads (wild type or *zip2*).

zip2 and zip4 Affect CO Homeostasis

Analysis of a series of *SPO11* alleles with decreasing frequencies of DSBs revealed the existence of CO homeostasis in an otherwise wild-type strain (Martini et al., 2006). Our observation that wild type shows no correlation between COs and NCOs confirms that CO homeostasis is part of normal CO control. It has been proposed that the molecular

mechanism that gives rise to CO interference may also be responsible for CO homeostasis (Martini et al., 2006). This hypothesis predicts that any observed loss of interference would be accompanied by a concomitant loss of homeostasis. Supporting this notion, we see a reduction of CO homeostasis in two mutants (*zip2* and *zip4*) that show reduced interference. However, although interference was almost completely abolished in these mutants, the reduction of CO homeostasis was more modest, suggesting that the connection between CO homeostasis and interference is more complex.

CO Prevention at the Centromere

Centromere-proximal crossing over contributes to aneuploidy in budding yeast due to precocious separation of sister chromatids (PSSC) at meiosis I (Rockmill et al., 2006). In *Drosophila* and humans, COs near the centromere also predispose a chromosome to segregate aberrantly (Hassold and Hunt, 2001; Koehler et al., 1996a), suggesting that prevention of COs near centromeres may be critical for the proper alignment of homologs. Our finding that centromeric repression of crossing over depends on Zip1 is consistent with the timing and localization of Zip1 on meiotic chromosomes. Tsubouchi and Roeder (2005) showed that Zip1 holds chromosomes together in pairs at their centromeres, early in meiotic prophase when the homology search is underway. Early in prophase I, many nonhomologous centromeres couplings are found, but these decrease as chromosomes find their correct partners. Important to the homology search is DSB formation by the Spo11 protein, resulting strand invasions reactions that likely stabilize and define a homologous pair. Because centromere coupling initially takes place between

nonhomologous centromeres, there may be a need to suppress homology assessment at centromeres. The Zip1-dependent bias towards inter-sister vs. inter-homolog recombination near centromeres may act to limit homology searches nearby and promote searches in more distal regions.

Microarray Mapping of COs and NCOs

Recently, a similar method using tiling arrays with a median distance of 78 bp between consecutive markers was used to map meiotic COs and NCOs in wild type and *msh4* for the same S96/YJM789 hybrid used in our study (Mancera et al, 2008). In agreement with our analyses, their study reports that wild-type strains show interference and *msh4* strains have lost interference. Particularly noteworthy is that the higher resolution of their study permitted a better analysis of the relationship between COs and NCOs and a more accurate assessment of NCO tract lengths and frequencies. The high resolution CO and NCO maps revealed the existence of genomic locations with distinct preferences for COs or NCOs. Although limited in resolution for NCOs, our observation that GC_{CO}s have larger tract lengths than NCOs is confirmed by their study. Our in-depth analyses of CO control in wild type and several mutants and our extensive analysis of telomeres and centromeres, together with the high resolution analysis of NCOs of Mancera et al. (2008), clearly demonstrates the power of this microarray-based approach for future studies of meiotic recombination.

Materials and Methods

Strains

Haploid yeast strains S96 and YJM789 were used in this study (Winzeler et al., 1998). Deletion strains were constructed by PCR-mediated gene replacement using the pFA6a-kanMX6 plasmid as the template (Longtine et al., 1998). Genotypes of strains are listed in Table A1-9. In all but *zip1*, haploid strains were mated and zygotes were picked after 4 hrs and allowed to grow on YPAD plates for < 3 days to minimize mismatch repair before transferring to 2% potassium acetate sporulation plates at 30ºC. Tetrads were dissected after 3-5 days. For *zip1*, because the sporulation of 4-spore tetrads was so low, zygotes were taken en masse and patched to a sporulation plate after 6-8 hours of mating.

Southern Analysis

To induce synchronous meiosis, strains were pre-inoculated at OD_{600} = 0.3 in BYTA medium (50mM potassium phthalate, 1% yeast extract, 2% bactotryptone, 1% potassium acetate), grown for 16 hours at 30°C, washed twice, and resuspended at OD_{600} = 1.9 in SPO medium (0.3% potassium acetate). Southern analysis was performed as described by Blitzblau et al. (2007a).

Sample Preparation

Genomic DNA was purified from 100 ml of overnight YPAD culture using a Qiagen genomic-tip 500/G following the Qiagen genomic DNA handbook with the slight modification of extending zymolyase and protease K digestion to 1 hour. 15 μg of genomic DNA was digested to 50- to 100-bp fragments and end-labeled as previously described (Winzeler et al., 2003). Labeled DNA fragments were then hybridized to Affymetrix Yeast Genome S98 arrays (Gladstone Institute, San Francisco CA).

Data Analysis

Marker designations and CO locations were determined using the Allelescan software. In our CrossOver software, programs were written to generate the distributions for our analysis using the output segregation file from Allelescan. Analyses of chromatid interference and GC_{CO}s and NCOs are within the CrossOver software. A description of the interference analysis and the simulation algorithm is provided in Supplemental Procedures.

Genetics

Genetic analyses of *zip1* near the centromere were carried out as described by Rockmill et al. (2006).

Correlation Coefficient Analysis

Using the inherent DSB fluctuation expected on a cell-to-cell basis, we assayed the intensity of the association between COs and NCOs to assess what might be homeostatically controlled. CO homeostasis was measured by a lack of statistical association between fluctuations in CO number and NCO number. We quantified the extent of statistical association between the two numbers using the Pearson's correlation coefficient, a measure of statistical association between two random variables that generates values in the range of -1.0 to 1.0. Statistical significance between the mutant and wild type was determined using a analysis comparing a control correlation coefficient to each other mutant correlation coefficient (Zar, 1984). If the control set of data is B and

each other group of data is A, we can compute $q = (z_B - z_A)/SE$ where $z = 0.5 * \ln((1+r)/(1-r))$, r is the correlation coefficient and $SE = \sqrt{1/(n_A-3) + 1/(n_B-3)}$ in the case where sample sizes (n_A and n_B) are not the same. The critical value for the q statistic is given in Figure 4-5C. Below, examples are provided for the various potential relationships between COs and NCOs in the face of fluctuating DSBs.

Positive correlation coefficient (fixed CO/NCO ratio). In the case of a fixed CO/NCO ratio, cells with lower numbers of DSBs would be expected to show correspondingly low numbers of NCOs and COs, and cells with higher numbers of DSBs would be expected to show correspondingly high numbers of NCOs and COs, thus giving a positive correlation coefficient.

Negative correlation coefficient (fixed CO + NCO). A negative correlation coefficient would be indicative of maintaining the overall total of NCOs and COs such that an increase in one comes at the expense of the other. This would be expected if the number of DSBs did not vary between cells, but instead the frequency of resolving a DSB as either a CO or an NCO was variable.

Zero correlation coefficient (CO level maintained or NCO level maintained). A correlation coefficient of zero can have either of two meanings. It could mean that the two variables that are being tested for correlation have absolutely nothing to do with each other. Alternatively, if there is a known relationship expected between two variables that is established by other data, it could mean that one variable is being controlled (homeostasis) and the other variable is not. In the case of COs and NCOs, given the fact that both are derived from DSBs, rules out the possibility that COs and NCOs have nothing to do with each other. A zero correlation coefficient could therefore mean that

either the CO level is homeostatically controlled or the NCO level is homeostatically controlled. Since the coefficient of variation (CV = SD/mean) for NCOs is larger than for COs (CV_{NCO} = 0.45 vs CV_{CO} = 0.10), it suggests that homeostatic control is exerted on the COs.

Acknowledgments

We give thanks to Amy MacQueen and Wallace Marshall for critical reading of the manuscript. We also thank Wallace Marshall and Tetsuya Matsuguchi for technical and programming advice. S.C is supported by a Genentech Fellowship. J.F is supported by the American Cancer Society Research Scholar Award (RSG CCG 110688) and the UCSF Sandler Fellows Program.

References

Agarwal, S., and Roeder, G.S. (2000). Zip3 provides a link between recombination enzymes and synaptonemal complex proteins. Cell *102*, 245-255.

Barton, A.B., Pekosz, M., Kurvathi, R.S., and Kaback, D.B. (2008). Meiotic Recombination at Chromosome Ends in Saccharomyces cerevisiae. Genetics. (Epub ahead of print).

Baudat, F., and de Massy, B. (2007). Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res *15*, 565-577.

Bishop, D.K., Park, D., Xu, L., and Kleckner, N. (1992). *DMC1*: a meiosis-specific yeast homolog of E. coli *recA* required for recombination, synaptonemal complex formation, and cell cycle progression. Cell *69*, 439-456.

Blitzblau, H.G., Bell, G.W., Rodriguez, J., Bell, S.P., and Hochwagen, A. (2007a). Mapping of Meiotic Single-Stranded DNA Reveals Double-Strand-Break Hotspots near Centromeres and Telomeres. Curr Biol.

Blitzblau, H.G., Bell, G.W., Rodriguez, J., Bell, S.P., and Hochwagen, A. (2007b). Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. Curr Biol *17*, 2003-2012.

Borts, R.H., and Haber, J.E. (1987). Meiotic recombination in yeast: alteration by multiple heterozygosities. Science *237*, 1459-1465.

Broman, K.W., Rowe, L.B., Churchill, G.A., and Paigen, K. (2002). Crossover interference in the mouse. Genetics *160*, 1123-1131.

Buhler, C., Borde, V., and Lichten, M. (2007). Mapping Meiotic Single-Strand DNA Reveals a New Landscape of DNA Double-Strand Breaks in Saccharomyces cerevisiae. PLoS Biol *5*, e324.

Carpenter, A.T.C. (1988). Thoughts on recombination nodules, meiotic recombination, and chiasmata. In Genetic Recombination, R. Kucherlapati, and G.R. Smith, eds. (Washington, D.C., American Society for Microbiology), pp. 529-548.

Cherry, J.M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., and Riles, L. (1997). Genetic and physical maps of Saccharomyces cerevisiae. Nature *387 (6632 Suppl)*, 67-73.

Chua, P.R., and Roeder, G.S. (1997). Tam1, a telomere-associated meiotic protein, functions in chromosome synapsis and crossover interference. Genes Dev *11*, 1786-1800.

Chua, P.R., and Roeder, G.S. (1998). Zip2, a meiosis-specific protein required for the initiation of chromosome synapsis. Cell *93*, 349-359.

Copenhaver, G.P., Housworth, E.A., and Stahl, F.W. (2002). Crossover interference in Arabidopsis. Genetics *160*, 1631-1639.

de Boer, E., Stam, P., Dietrich, A.J., Pastink, A., and Heyting, C. (2006). Two levels of interference in mouse meiotic recombination. Proc Natl Acad Sci U S A *103*, 9607-9612.

Fogel, S., Mortimer, R., Lusnak, K., and Tavares, F. (1978). Meiotic gene conversion: a signal of the basic recombination event in yeast. Cold Spring Harb Symp Quant Biol *43*, 1325-1341.

Foss, E.J., and Stahl, F.W. (1995). A test of a counting model for chiasma interference. Genetics *139*, 1201-1209.

Fung, J.C., Rockmill, B., Odell, M., and Roeder, G.S. (2004). Imposition of crossover interference through the nonrandom distribution of synapsis initiation complexes. Cell *116*, 795-802.

Guillon, H., Baudat, F., Grey, C., Liskay, R.M., and de Massy, B. (2005). Crossover and noncrossover pathways in mouse meiosis. Mol Cell *20*, 563-573.

Hassold, T. (2007). The origin of human aneuploidy: where we have been, where we are going. Hum Mol Genet *16*, 203-208.

Hassold, T., and Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet *2*, 280-291.

Hassold, T.H., Abruzzo, M., Adkins, K., Griffin, D., Merrill, M., Millie, E., Saker, D., Shen, J., and Zaragoza, M. (1996). Human aneuploidy: Incidence, origin, and etiology. Environ Mol Mutagen *28*, 167-175.

Hillers, K.J. (2004). Crossover interference. Curr Biol *14*, R1036-1037.

Huitema, B.E. (1974). Three multiple comparison procedures for contrasts among correlation coefficients. Proc Soc Statist Sect, Amer Statist Assoc, 1974, 336-339.

Jeffreys, A.J., and May, C.A. (2004). Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat Genet *36*, 151-156.

Jones, G.H. (1984). The control of chiasma distribution. Symp Soc Exp Biol *38*, 293-320.

Kaback, D.B., Guacci, V., Barber, D., and Mahon, J.W. (1992). Chromosome size-dependent control of meiotic recombination. Science *256*, 228-232.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell *88*, 375-384.

Koehler, K.E., Boulton, C.L., Collins, H.E., French, R.L., Herman, K.C., Lacefield, S.M., Madden, L.D., Schuetz, C.D., and Hawley, R.S. (1996a). Spontaneous X chromosome MI and MII nondisjunction events in Drosophila melanogaster oocytes have different recombinational histories. Nat Genet *14*, 406-414.

Koehler, K.E., Hawley, R.S., Sherman, S., and Hassold, T. (1996b). Recombination and nondisjunction in humans and flies. Hum Mol Genet *5*, 1495-1504.

Lacefield, S., and Murray, A.W. (2007). The spindle checkpoint rescues the meiotic segregation of chromosomes whose crossovers are far from the centromere. Nat Genet *39*, 1273-1277.

Lamb, N.E., Freeman, S.B., Savage-Austin, A., Pettay, D., Taft, L., Hersey, J., Gu, Y., Shen, J., Saker, D., May, K.M.*, et al.* (1996). Susceptible chiasmate configurations of chromosome 21 predispose to non-disjunction in both maternal meiosis I and meiosis II. Nat Genet *14*, 400-405.

Lambie, E.J., and Roeder, G.S. (1986). Repression of meiotic crossing over by a centromere (CEN3) in Saccharomyces cerevisiae. Genetics *114*, 769-789.

Lambie, E.J., and Roeder, G.S. (1988). A yeast centromere acts in cis to inhibit meiotic gene conversion of adjacent sequences. Cell *52*, 863-873.

Levy, K.J. (1979). Pairwise comparisons associated with the K independent sample median test. Amer Statist *33*, 138-139.

Longtine, M.S., III, A.M., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Phillippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. Yeast *14*, 953-961.

Lynn, A., Soucek, R., and Borner, G.V. (2007). ZMM proteins during meiosis: crossover artists at work. Chromosome Res *15*, 591-605.

Malkova, A., Swanson, J., German, M., McCusker, J.H., Housworth, E.A., Stahl, F.W., and Haber, J.E. (2004). Gene conversion and crossing over along the 405-kb left arm of Saccharomyces cerevisiae chromosome VII. Genetics *168*, 49-63.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., and Steinmetz, L.M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature. (Epub ahead of print).

Martini, E., Diaz, R.L., Hunter, N., and Keeney, S. (2006). Crossover homeostasis in yeast meiosis. Cell *126*, 285-295.

McPeek, M.S., and Speed, T.P. (1995). Modeling interference in genetic recombination. Genetics *139*, 1031-1044.

Mortimer, R., and Fogel, S. (1974). Genetical interference and gene conversion. In Mechanisms in Recombination, R. Grell, ed. (New York, Plenum Press), pp. 263-275.

Muller, H. (1916). The mechanism of crossing over. Amer Nat *50*, 193-434.

Novak, J.E., Ross-Macdonald, P., and Roeder, G.S. (2001). The budding yeast Msh4 protein functions in chromosome synapsis and the regulation of crossover distribution. Genetics *158*, 1013-1025.

Ott, J. (1991). Analysis of Human Genetic Linkage (Baltimore, Johns Hopkins University Press).

Papazian, H.P. (1952). The analysis of tetrad data. Genetics *37*, 175-188.

Perkins, D.D. (1962). Crossing-over and interference in a multiply marked chromosome arm of Neurospora. Genetics *47*, 1253-1274.

Petes, T.D. (2001). Meiotic recombination hot spots and cold spots. Nat Rev Genet *2*, 360-369.

Rockmill, B., Fung, J.C., Branda, S.S., and Roeder, G.S. (2003). The Sgs1 helicase regulates chromosome synapsis and meiotic crossing over. Curr Biol *13*, 1954-1962.

Rockmill, B., Voelkel-Meiman, K., and Roeder, G.S. (2006). Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in Saccharomyces cerevisiae. Genetics *174*, 1745-1754.

Roeder, G.S. (1997). Meiotic chromosomes: it takes two to tango. Genes Dev *11*, 2600-2621.

Sall, T., and Bengtsson, B.O. (1989). Apparent negative interference due to variation in recombination frequencies. Genetics *122*, 935-942.

Shinohara, M., Oh, S.D., Hunter, N., and Shinohara, A. (2008). Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. Nat Genet *40*, 299-309.

Storlazzi, A., Xu, L., Schwacha, A., and Kleckner, N. (1996). Synaptonemal complex (SC) component Zip1 plays a role in meiotic recombination independent of SC polymerization along the chromosomes. Proc Natl Acad Sci USA *93*, 9043-9048.

Su, Y., Barton, A.B., and Kaback, D.B. (2000). Decreased meiotic reciprocal recombination in subtelomeric regions in Saccharomyces cerevisiae. Chromosoma *109*, 467-475.

Sym, M., Engebrecht, J., and Roeder, G.S. (1993). ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. Cell *72*, 365-378.

Sym, M., and Roeder, G.S. (1994). Crossover interference is abolished in the absence of a synaptonemal complex protein. Cell *79*, 283-292.

Tsubouchi, T., and Roeder, G.S. (2005). A synaptonemal complex protein promotes homology-independent centromere coupling. Science *308*, 870-873.

Tsubouchi, T., Zhao, H., and Roeder, G.S. (2006). The meiosis-specific Zip4 protein regulates crossover distribution by promoting synaptonemal complex formation together with Zip2. Dev Cell *10*, 809-819.

Wei, W., McCusker, J.H., Hyman, R.W., Jones, T., Ning, Y., Cao, Z., Gu, Z., Bruno, D., Miranda, M., Nguyen, M.*, et al.* (2007). Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. Proc Natl Acad Sci U S A *104*, 12825-12830.

Winzeler, E.A., Castillo-Davis, C.I., Oshiro, G., Liang, D., Richards, D.R., Zhou, Y., and Hartl, D.L. (2003). Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics *163*, 79-89.

Winzeler, E.A., Richards, D.R., Conway, A.R., Goldstein, A.L., Kalman, S., McCullough, M.J., McCusker, J.H., Stevens, D.A., Wodicka, L., Lockhart, D.J.*, et al.* (1998). Direct allelic variation scanning of the yeast genome. Science *281*, 1194-1197.

Wu, H.Y., and Burgess, S.M. (2006). Ndj1, a telomere-associated protein, promotes meiotic recombination in budding yeast. Mol Cell Biol *26*, 3683-3694.

Zar, J. (1984). Biostatistical Analysis, 2nd edn (Upper Saddle River, Prentice Hall).

Zhao, H., Speed, T.P., and McPeek, M.S. (1995). Statistical analysis of crossover interference using the chi-square model. Genetics *139*, 1045-1056.

Zickler, D., and Kleckner, N. (1999). Meiotic chromosomes: integrating structure and function. Annu Rev Genet *33*, 603-754.

Figure Legends

Figure 4-1. Characterization of Crossover Distribution in Wild Type

(A) Marker distribution for the S96/YJM789 strain shown for all 16 chromosomes. Vertical bars indicate the location of markers. (B) Plot of frequency of inter-marker distances. Over 78% of the markers are spaced less than 2 kb apart. Mean distance is 1.5 kb. (C) Mean number of COs per chromosome and total COs per meiosis were compared

between microarray data and genetic map data obtained from the Saccharomyces Genome Database (SGD). Error bars denote 95% confidence interval (C.I.) of the microarray data. (D) Comparison of CO density between microarray and genetic data. 95% C.I.s are shown for microarray data.

Figure 4-2. CO and NCO Distributions near Telomeres and Centromeres in Wild Type

Distribution of COs and NCOs relative to the nearest telomere (A-D) or centromere (E-H). Microarray data from wild type is plotted against a simulated distribution that incorporates interference but assumes a uniform CO landscape along the chromosome. (B), (D), (F) and (H) show distributions without the 4 smallest chromosomes (1, 3, 6 and 9). Error bar = SD.

Figure 4-3. CO Distribution Pattern for Wild Type and *zip4*

Shown are CO distributions from representative tetrads from wild type (WT) (A) and *zip4* (B). Black vertical bars indicate the location of COs, and blue bars indicate centromeres. S96 parental origin is displayed in green; YJM789 parental origin is shown in red. Yellow (S96) and magenta (YJM789) indicate less confidence (<99% probability) in the designation of marker origin. Yellow and magenta sections at the ends of chromosomes are extrapolations from the last known marker nearest the end.

Figure 4-4. Determination of Interference

Comparison of the experimental and best-fit gamma distribution for inter-CO distances for wild type with normal interference (A) and *zip4* with reduced interference (B). $\gamma = 1$

indicates no interference, while $\gamma > 1$ indicates positive interference. (C) Hazard functions are calculated from the best-fit gamma distribution parameters for wild type (WT) (*solid line*) and *zip4* (*dotted line*).

Figure 4-5. *zip4* and *zip2* Show Reduced CO Homeostasis

(A) Comparison of interference determined by microarray (simulated NPD ratio) and genetic approaches (NPD ratio). Genetic NPD ratios were obtained by averaging published NPD ratios; simulated NPD ratios were determined in this study (Table A1-4). Error bars = SD. Best fit gamma values are shown. $P > 0.05$ shows that the best fit inter-CO distribution fits well with the experimental distribution, as determined by chi-square analysis.(B) Dispersion of CO number per meiosis for WT (n = 26; in gray), *zip4* (n = 34; in black) and *zip2* (n = 26; in white). Black vertical arrow indicates the outlier *zip4* tetrad with 126 COs. (C) Comparison of a control correlation coefficient (wild type) against mutants using an analog to the Dunnett's test (Huitema, 1974). Correlation coefficients were calculated based on the numbers of COs and NCOs. q' denotes critical value of $q_{0.05,\infty,3}$. $q > q'$, rejects the hypothesis that correlations are the same.

Figure 4-6. Centromere-Proximal CO Repression Is Relieved in a *zip1* Mutant

Comparison of centromere-proximal COs (A) and NCOs (B) in wild type and *zip1*. (C) Chromosome III markers in a strain used to genetically measure GCs and associated crossing over at the centromere (BR4633, Rockmill et al., 2006). (D) Frequency of Ura^+ gene convertants from random spores for wild type, *zip1* and *zip2*. SDs are shown. (E) Frequency of COs associated with Ura^+ gene convertants for random spores for wild

type, *zip1* and *zip2*. Fold change relative to wild type is indicated above the bars. (F) *dmc1Δ* (NKY1455, (Bishop et al., 1992) and *dmc1Δ zip1Δ* (YAH2650, (Blitzblau et al., 2007b) cells were induced to undergo meiosis, and samples were collected at the indicated time points. Genomic DNA was digested and analyzed by Southern blot. The following restriction enzymes and probes (SGD coordinates) were used: *CEN2*, SacI, II:231,552-232,350; *CEN4*, SpeI, IV:448,180-449,164; *CEN15*, SphI/NheI, XV:331,713-332,402 (Blitzblau et al., 2007b). Black arrowheads indicate major DSB sites. *CEN4* is located adjacent to *YDL001W*, off the bottom of the gel. Quantification of DSB frequencies is provided in Figure S5.

Table 4-1. Summary of Crossover and Gene Conversion Data

For chromatid interference, ratios were normalized to the 2-strand double (s.d.) CO. Chi-square analysis for wild type and all mutants showed no difference from the expected 1:2:1 ratio if there were no chromatid interference. The average number of markers involved in detecting each gene conversion is reported for NCO and $GC_{CO.}$ Percentage of NCO and GC_{CO} detected by more than 1 marker is shown.

[a]Sym and Roeder (1994), [b]Chua and Roeder (1998), [c]Agarwal and Roeder (2000), [d]Tsubouchi et al. (2006), [e]Novak et al. (2001), [f]Chua and Roeder (1997), [g]Wu and Burgess (2006), and [h]Rockmill et al. (2003)

Figure 4-1

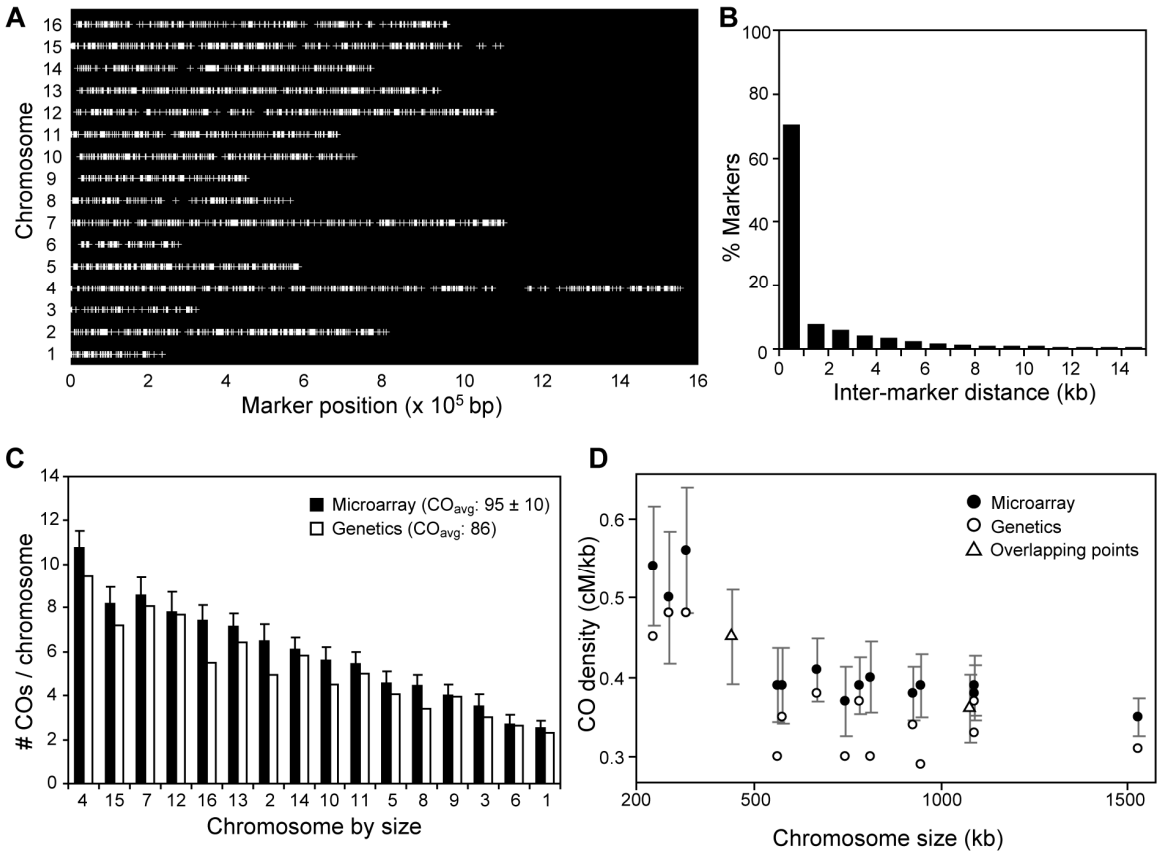

A
Chromosome
Marker position (x 10^5 bp)
B
% Markers
Inter-marker distance (kb)
C
COs / chromosome
Microarray (CO_avg: 95 ± 10)
Genetics (CO_avg: 86)
Chromosome by size
D
CO density (cM/kb)
Microarray
Genetics
Overlapping points
Chromosome size (kb)

Figure 4-2

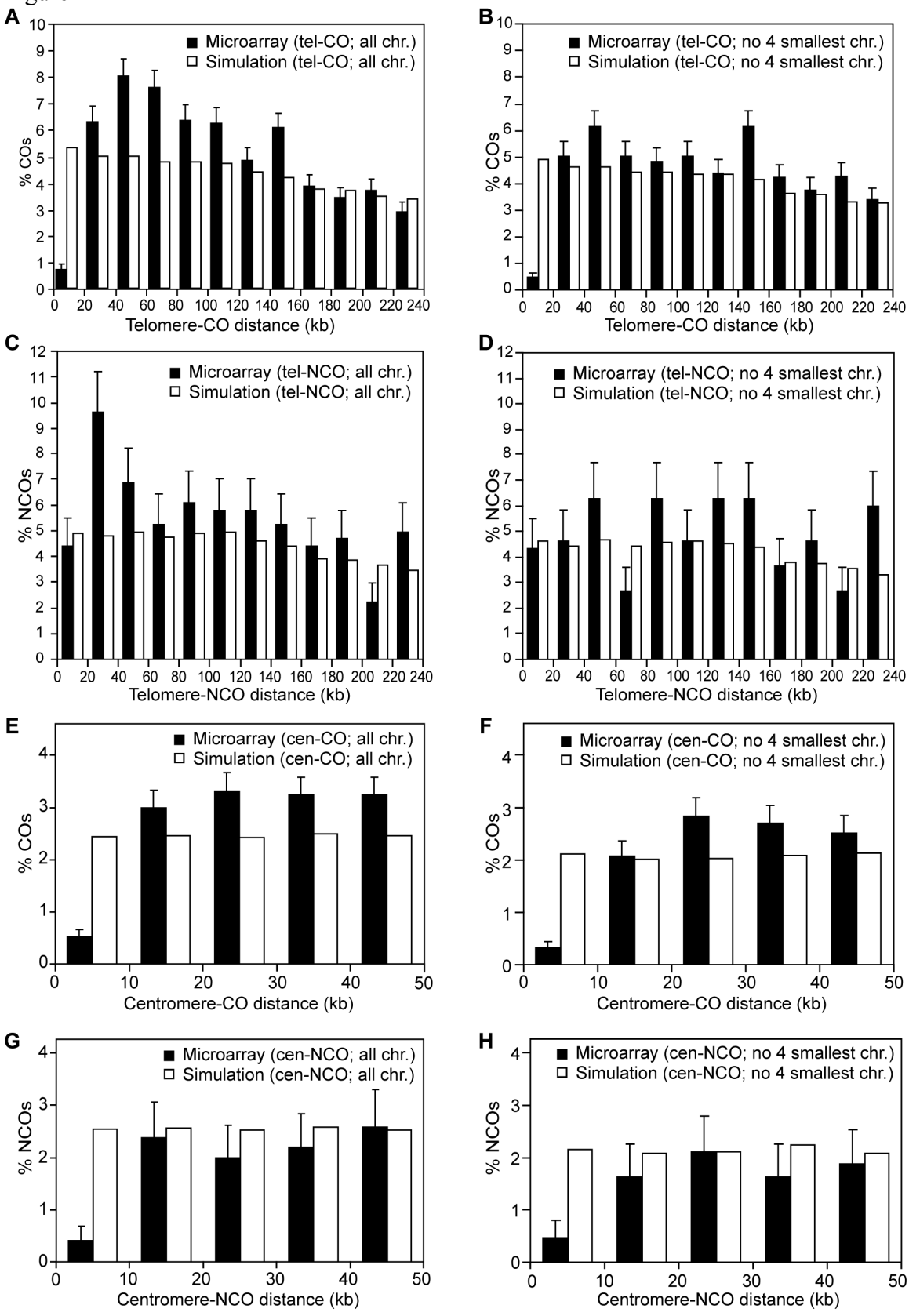

A
Microarray (tel-CO; all chr.)
Simulation (tel-CO; all chr.)
% COs
Telomere-CO distance (kb)
B
Microarray (tel-CO; no 4 smallest chr.)
Simulation (tel-CO; no 4 smallest chr.)
% COs
Telomere-CO distance (kb)
C
Microarray (tel-NCO; all chr.)
Simulation (tel-NCO; all chr.)
% NCOs
Telomere-NCO distance (kb)
D
Microarray (tel-NCO; no 4 smallest chr.)
Simulation (tel-NCO; no 4 smallest chr.)
% NCOs
Telomere-NCO distance (kb)
E
Microarray (cen-CO; all chr.)
Simulation (cen-CO; all chr.)
% COs
Centromere-CO distance (kb)
F
Microarray (cen-CO; no 4 smallest chr.)
Simulation (cen-CO; no 4 smallest chr.)
% COs
Centromere-CO distance (kb)
G
Microarray (cen-NCO; all chr.)
Simulation (cen-NCO; all chr.)
% NCOs
Centromere-NCO distance (kb)
H
Microarray (cen-NCO; no 4 smallest chr.)
Simulation (cen-NCO; no 4 smallest chr.)
% NCOs
Centromere-NCO distance (kb)

Figure 4-3

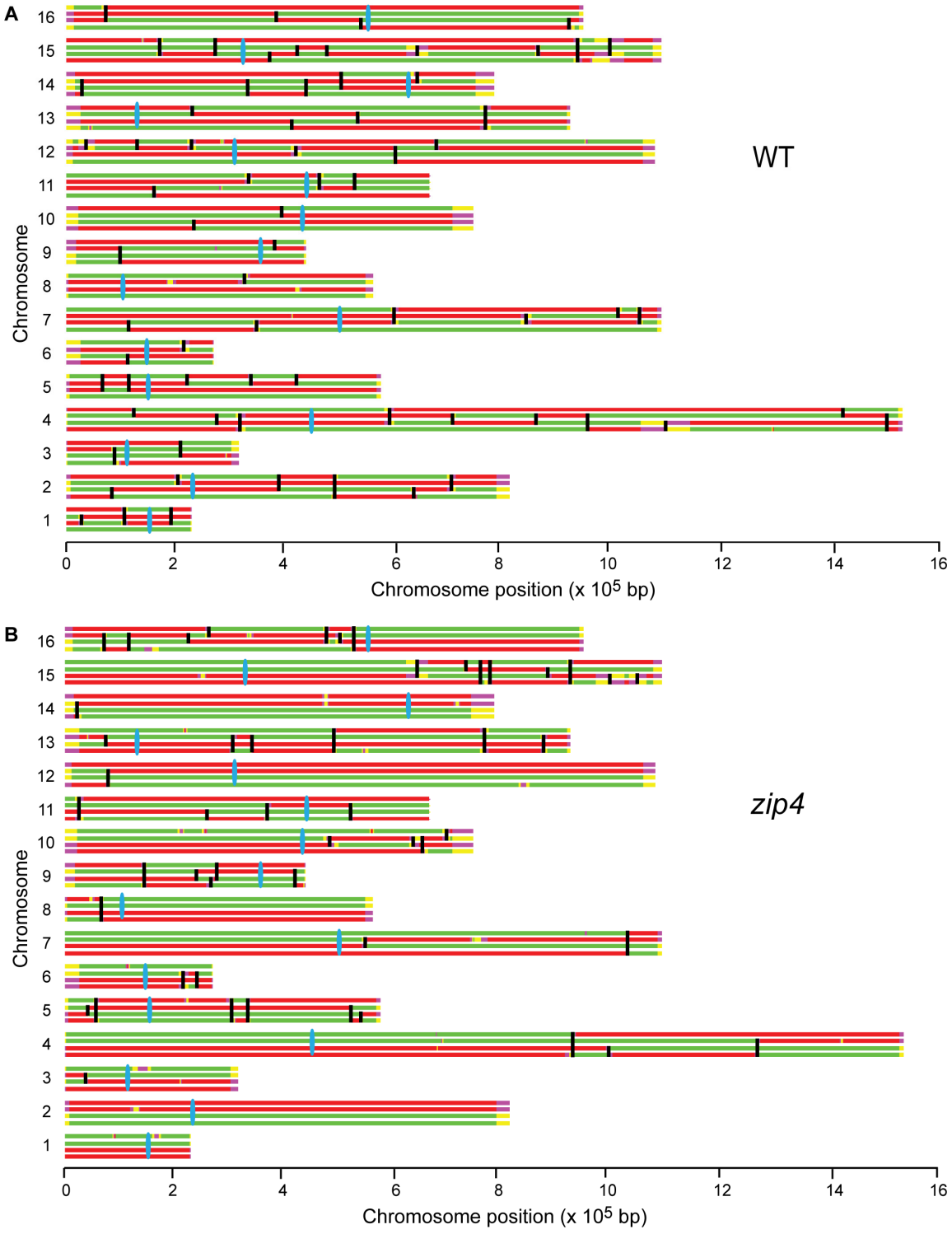

A
WT
Chromosome
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
2
4
6
8
10
12
14
16
Chromosome position (x 10^5 bp)
B
zip4
Chromosome
Chromosome position (x 10^5 bp)

Figure 4-4

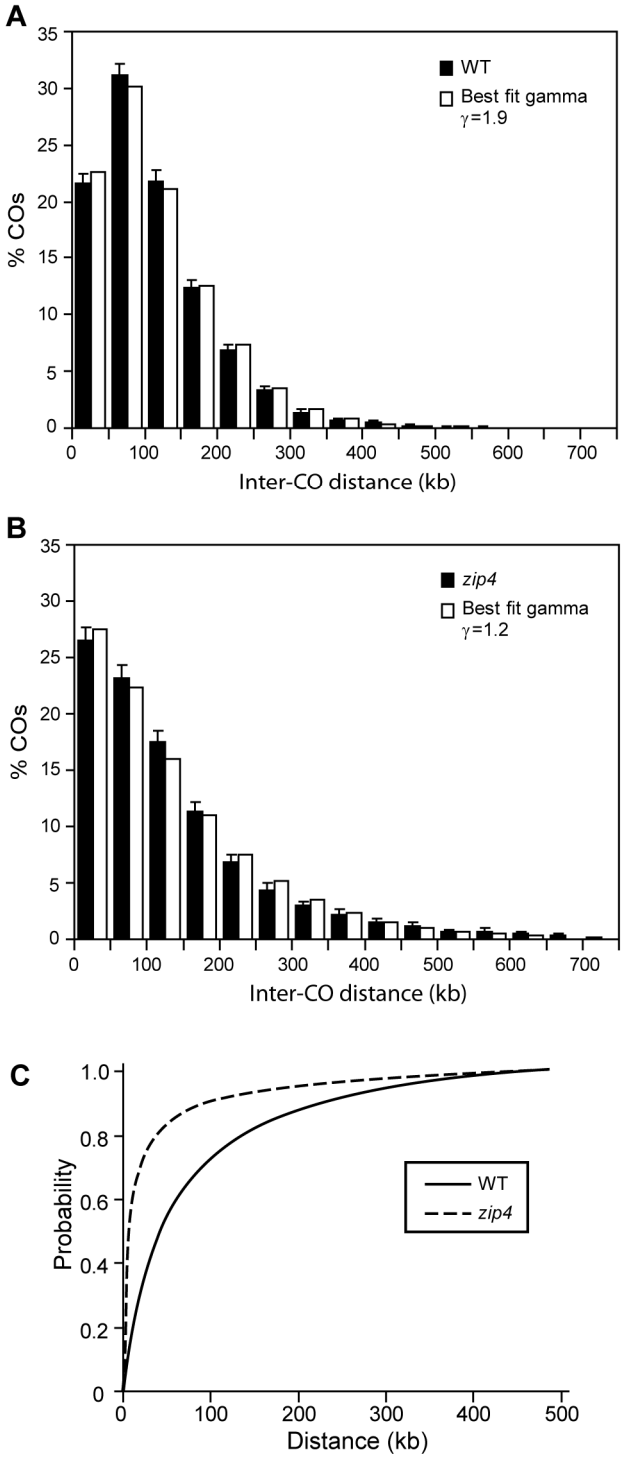
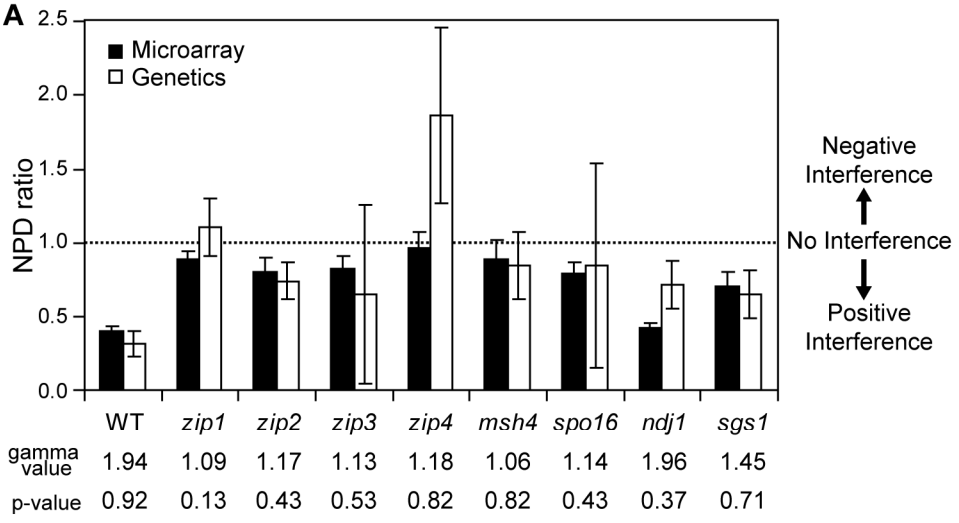
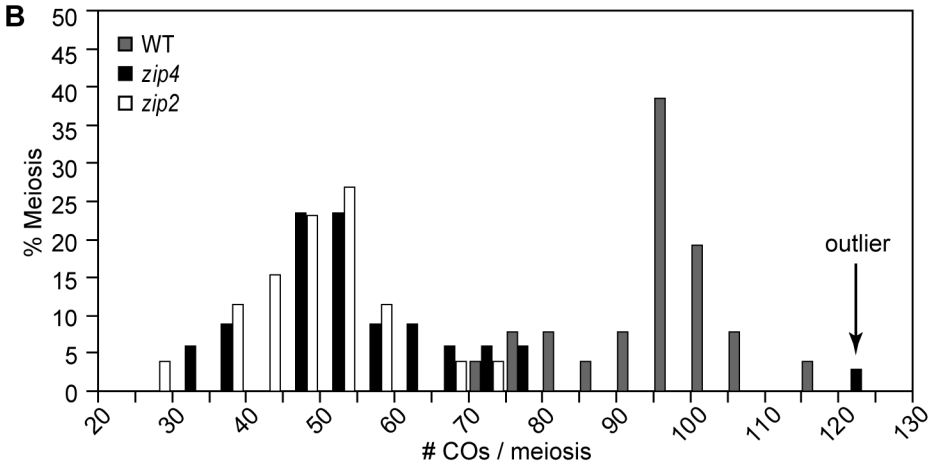

A
WT
Best fit gamma
γ=1.9
% COs
Inter-CO distance (kb)
B
zip4
Best fit gamma
γ=1.2
% COs
Inter-CO distance (kb)
C
WT
zip4
Probability
Distance (kb)

Figure 4-5





C

	Correlation Coefficient				
	WT	mutant	*q*	*q'*	Conclusion
WT vs. *zip4* (all data)	-0.07	0.34	2.09	1.64	Reject H_0
WT vs. *zip4* (no outlier)	-0.07	0.67	3.89	1.64	Reject H_0
WT vs. *zip2*	-0.07	0.44	2.44	1.64	Reject H_0

Figure 4-6

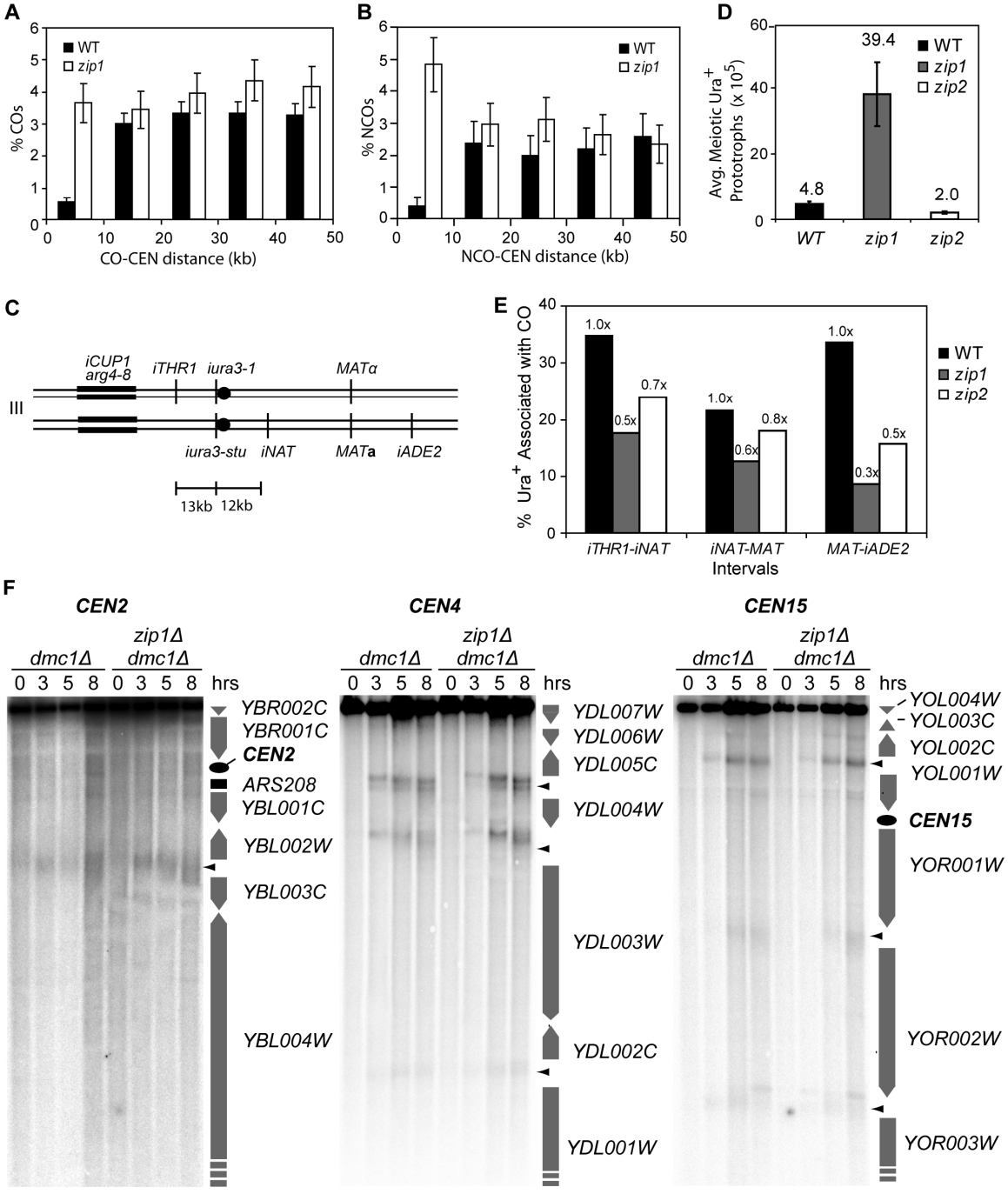

A
% COs
WT
zip1
CO-CEN distance (kb)
B
% NCOs
NCO-CEN distance (kb)
D
Avg. Meiotic Ura+ Prototrophs (x 10^5)
39.4
4.8
2.0
zip2
C
iCUP1
arg4-8
iTHR1
iura3-1
MATα
III
iura3-stu
iNAT
MATa
iADE2
13kb
12kb
E
% Ura+ Associated with CO
1.0x
0.5x
0.7x
0.6x
0.8x
0.3x
0.5x
iTHR1-iNAT
iNAT-MAT
MAT-iADE2
Intervals
F
CEN2
CEN4
CEN15
dmc1Δ
zip1Δ dmc1Δ
0 3 5 8
hrs
YBR002C
YBR001C
CEN2
ARS208
YBL001C
YBL002W
YBL003C
YBL004W
YDL007W
YDL006W
YDL005C
YDL004W
YDL003W
YDL002C
YDL001W
YOL004W
YOL003C
YOL002C
YOL001W
CEN15
YOR001W
YOR002W
YOR003W

Table 4-1. Summary of Crossover and Gene Conversion Data

	Recombination		Crossover Count				Chromatid Interference				
Strains	Array	Genetics	# Tetrads	Total	Mean	SD	2-s.d.	3-s.d.	4-s.d.	Ratio	*p*-value
WT	1.0	1.0	26	2474	95.2	10.2	538	1010	510	1.0 : 1.9 : 1.0	0.48
zip1	1.2	0.4 – 0.8[a]	9	987	109.7	18.1	245	416	187	1.0 : 1.7 : 0.8	0.05
zip2	0.5	0.5 – 0.8[b]	26	1272	49.0	9.3	232	442	218	1.0 : 1.9 : 0.9	0.77
zip3	0.6	0.3 – 0.8[c]	8	416	52.0	8.3	78	142	76	1.0 : 1.8 : 1.0	0.77
zip4	0.6	0.3 – 0.9[d]	34	1927	56.7	17.0	341	667	362	1.0 : 2.0 : 1.0	0.45
msh4	0.4	0.4 – 0.6[e]	11	405	36.8	8.7	66	113	60	1.0 : 1.7 : 0.9	0.60
spo16	0.6	0.5	8	422	52.8	12.4	96	131	72	1.0 : 1.4 : 0.8	0.07
ndj1	1.0	0.9 – 1.4[f,g]	7	695	99.3	19.0	174	267	149	1.0 : 1.5 : 0.9	0.03
sgs1	1.1	1.3[h]	11	1163	105.7	14.0	266	476	248	1.0 : 1.8 : 0.9	0.34

	Gene Conversion not associated with CO (NCO)							Gene Conversion associated with CO (GC_{CO})							Mean % GC_{CO}/CO
	Count		Tract Length (kb)			Markers per NCO		Count		Tract Length (kb)			Markers per GC_{CO}		
	Mean	SD	Mean	SD	Med	Mean	% NCO >1 Marker	Mean	SD	Mean	SD	Median	Mean	% GC_{CO} >1 Marker	
WT	18.6	8.7	5.1	3.8	3.9	3.3	79	30.7	5.1	5.9	5.2	4.4	3.1	65	32.5
zip1	71.3	17.4	6.3	5.2	5.1	3.9	79	61.4	9.5	7.9	7.9	6.3	4.1	73	57.3
zip2	30.3	10.1	5.8	4.4	4.7	4.1	78	26.0	6.8	7.6	7.7	5.8	4.2	72	53.0
zip3	37.0	8.9	5.9	5.3	4.6	4.1	83	29.0	4.1	7.4	6.4	5.8	3.9	70	57.1
zip4	35.1	14.8	6.3	5.4	5.0	3.8	78	28.3	11.0	7.3	6.5	5.7	4.0	71	51.0
msh4	17.2	6.5	5.5	5.1	4.2	2.9	68	17.0	6.5	6.5	3.9	5.8	3.6	72	45.9
spo16	33.9	8.5	6.6	5.1	5.3	3.7	82	28.1	7.9	7.9	5.9	6.4	3.5	70	54.3
ndj1	21.6	8.4	6.2	5.6	5.3	5.0	85	36.0	10.9	6.2	5.0	4.9	3.6	63	37.5
sgs1	36.4	25.4	6.4	8.8	4.1	4.2	87	39.5	10.0	6.0	5.2	4.8	3.7	70	36.3

Chapter 5: Mapping Meiotic Recombination Events using Deep Sequencing Reveals New Insights into Recombination Mechanisms

Stacy Y. Chen[1], Michelle Dimon[2,3], Ashwini Oke [1], Joseph L. DeRisi [2,4], and Jennifer C. Fung [1]

[1]Department of Obstetrics, Gynecology and Reproductive Sciences

University of California, San Francisco, San Francisco, California, USA

[2]Department of Biochemistry and Biophysics

University of California, San Francisco, San Francisco, California, USA

[3]Biological and Medical Informatics Program

University of California, San Francisco, San Francisco, California, USA

[4]Howard Hughes Medical Institute, San Francisco, California, USA

Abstract

At the heart of meiosis is the process of meiotic recombination where programmed double-strand breaks (DSBs) are repaired into either crossovers (COs) or noncrossovers. Previously, we have mapped meiotic recombination events using ~8000 polymorphic markers identified between two yeast strains, the S96 and YJM789, by microarrays. While microarrays can be used to successfully identify the majority of COs, a significant portion of gene conversions remain undetected due to conversion tract lengths below the detection limit. Next-generation sequencing offers the opportunity to map all polymorphisms between the two yeast strains and not just the markers whose sequences were printed on the microarray probes, thereby providing a higher level of detection.

Using high throughput sequencing coupled with multiplexing, we identified ~54,000 markers in the non-multiplexed tetrads and ~30,000 markers in tetrads sequenced with 4 multiplexed samples. To better understand the details of how DSBs are repaired, we mapped recombination events in two wild-type tetrads, one *msh4* and one *sgs1* deletion mutant tetrad, and a tetrad carrying a meiosis-specific null allele of *pCLB2-MMS4*. In this study, we identified complex gene conversion tracts near sites of crossing over. Specifically, we found two types of conversion tract motifs near COs, and showed that the ratios of the two motifs are varied in meiotic mutants. Multiple recombination pathways are known to exist and involve different molecular mechanisms in CO resolution. These conversion motifs provide the footprints of how each CO was processed and resolved, and can be used to identify COs derived from different recombination pathways. Studying the recombination landscape and CO motifs in meiotic mutants will ultimately provide insights into the mechanisms of how COs are processed and the role each mutant plays in different pathways.

Introduction

During meiosis, a diploid parent undergoes two rounds of meiotic division to produce four progeny (Roeder, 1997; Zickler and Kleckner, 1999). At the heart of this process is meiotic recombination, where reciprocal exchange of genetic material between homologous chromosomes results in the reassortment of alleles from both parents, contributing to greater genetic diversity in the next generation. Crossing over also results in the formation of physical connections that hold the homologs together during the first meiotic division, absence of which leads to meiosis I nondisjunction. In addition to

crossovers (COs), meiotic recombination also results in gene conversion (GCs), which is a nonreciprocal exchange between homologous chromosomes that creates a short patch of homogenized alleles between the homologs (Paques and Haber, 1999). Developing an efficient and cost-effective assay for mapping the two products of meiotic recombination, is the first step towards understanding the mechanism and control that underlie meiotic recombination.

Until very recently, yeast geneticists relied mainly on traditional assays such as manual tetrad dissection and two-dimensional gel electrophoresis to identify recombination events (Malkova et al., 2004; Oh et al., 2009; Shinohara et al., 2008). These techniques limit the detection of events to a small number of genetic markers and could not reveal the global recombination landscape of the genome. Furthermore, analysis such as measuring GC tract lengths and assessing CO interference and homeostasis could not be performed on a genome-wide scale using such traditional techniques.

The advent of microarray technology made it possible to map recombination products genome-wide on a cell-by-cell basis. Winzeler and colleagues were the first to report a genome-wide mapping of COs in yeast meiosis (Winzeler et al., 1998). Our lab, as well as the Steinmetz Lab, has extended this study and utilized the sequence polymorphisms between two yeast strains, S96 (a S288c derivative) and YJM789 (a clinical isolate), to probe for more detailed meiotic recombination events in the 4-spore progeny of S96 and YJM789 (Chen et al., 2008; Mancera et al., 2008). The YJM789 genome contains approximately 60,000 SNPs and 6,000 indels with respect to the S288c genome (Wei et al., 2007). Using Affymetrix yeast S98 expression array, our lab was

able to detect around 8,000 markers genome-wide. We mapped CO sites and a small number of GC tracts in 26 wild-type tetrads and applied the same technique to eight meiotic mutants: *zip1, zip2, zip3, zip4, msh4, spo16, ndj1,* and *sgs1* (Chen et al., 2008). CO interference and homeostasis were also examined. Mancera and colleagues from the Steinmetz laboratory used high density custom-designed tiling arrays and genotyped around 52,000 markers in 51 wild-type tetrads and 5 tetrads carrying the *msh4* null mutation (Mancera et al., 2008). A handful of single spores and dyads carrying the *mms4* null mutation were also analyzed. The authors reported that the high marker resolution allowed the detection of most CO sites and ~70% of the GC tracts. Based on this work, the authors published a high resolution genome-wide CO and NCO map in yeast meiosis. They also reported evidence for interference between COs and NCOs, a phenomenon previously known only to exist between COs.

While both studies have demonstrated the tremendous power microarray technology has in mapping meiotic recombination products, several limitations remain. The number of identifiable polymorphic markers is confined by the number and the initial design of the probe sequences, and that polymorphisms not present among the probe sequences would never be identified as potential markers. Furthermore, microarray data are inferred from the differential hybridization between probes to target sequences, and are not a direct interrogation of the actual DNA sequence. Most importantly, while custom tiling arrays can achieve high marker resolution, they are extremely costly, making genetic screening for meiotic mutants with altered recombination landscape an expensive undertaking.

Recent developments in next-generation DNA sequencer have transformed the field of biological sciences through their ability to produce gigabases of sequence information in a single run, offering the most cost-effective sequencing technology to date (Ansorge, 2009; Metzker, 2010). Next-generation sequencing would also circumvent the limitations posed by microarrays in that sequence polymorphisms are directly sequenced instead of inferred from differential hybridization between probes and target sequences. Sequencing also allows the opportunity to indentify all polymorphisms between the two yeast strains and not just the subset printed on the array probes. Furthermore, the development of a multiplex barcoding system that allows simultaneous sequencing of multiple samples in one sequencing lane has the potential to reduce the cost of sequencing to lower than that of the array technology but with increased marker resolution (Lefrancois et al., 2009).

At the onset of the current study, no laboratories have reported the use of next-generation sequencing in identifying meiotic recombination products. Very recently, Qi and colleagues have reported using the Roche 454-FLX sequencer to identify recombination products in one four-spore tetrad from a cross between the *Saccharomyces cerevisiae* strains S288c and RM11-1a (Qi et al., 2009). They identified over 46,000 single-nucleotide polymorphisms (SNPs), and detected 91 COs and 21 NCOs. However, only one tetrad was sequenced in this study, and no meiotic mutants were investigated either. The authors also did not explore the possibility of multiplexing multiple samples in one sequencing reaction.

At the time of writing this manuscript, by using the Solexa Genome Analyzer (GA) II from Illumina, we have successfully sequenced the S96 and YJM789 parental

strains, two wild-type tetrads, as well as one tetrad from each of the mutant carrying the *msh4, sgs1 and pCLB2-MMS4* mutation. Around 54,000 polymorphic markers, including both SNPs and indels, were identified on average in non-multiplexed tetrads. We demonstrated that the barcoding system can be successful applied to up to four yeast samples per sequencing lane, and identified ~30,000 markers on average in tetrads sequenced with 4-multiplexed samples. Multiplexing help reduce the cost of analyzing one full tetrad to approximately $500, significantly lower than that of using custom tiling arrays which can cost around $860.

Overall, our study provides the most extensive analysis of meiotic recombination products in terms of the number of markers detected, the number of meiotic mutants sequenced, and the type of analysis performed. More wild-type and meiotic mutant tetrads are continually being sequenced and analyzed in our laboratory, with the goal of elucidating the different pathways and mechanisms governing meiotic recombination control and decision.

Results

Mapping Sequence Polymorphisms using Next-generation Sequencing and Multiplexing

We identified ~ 60,000 SNPs and indels between two polymorphic *Saccharomyces cerevisiae* strains, S96 and YJM789, by high-throughput sequencing using Illumina's GA II sequencing platform. For each parent strain, 7 to 8 million sequence reads of 43 bases in length are generated, totaling around 330 to 350 megabases of sequence information (Table 5-1). Approximately 90% of the sequence reads can be mapped uniquely to either

the S96 or the YJM789 reference genome, leading to a genome-wide coverage of 25x on average. To map meiotic recombination events utilizing the ~60,000 polymorphic markers identified between the S96 and YJM789 parental strains, progenies of 5 meioses from crosses between the two parents are also sequenced: two wild-type tetrads, one *msh4* null allele tetrad, one *sgs1* null allele tetrad, and a tetrad with a meiosis-specific null allele of *MMS4* (*pCLB2-MMS4),* which contains the *CLB2* promoter in place of the endogenous *MMS4* promoter (see Materials and Methods).

To reduce the cost of sequencing, we tested multiplexing two, four, and eight sample libraries in one sequencing lane. Figure 5-1 shows the experimental workflow for multiplexing four yeast spores in one lane. Short fragments of yeast genomic DNA are ligated with adapter oligos that contain one of the eight three-nucleotide barcodes: TGT, CAT, ACT, GTT, AGT, GAT, CTT, and TGT (See Materials and Methods for barcodes design and procedures in library preparation). PCR libraries from differentially barcoded DNA samples are pooled together in equimolar ratios for sequencing. Table 5-2 lists the names of the sample and the corresponding barcodes that are pooled into one sequencing reaction. After sequencing, sequence reads are divided according to the barcode at the first three bases of each read, and aligned to the two reference genomes (S288c and YJM789) for SNP and indel detection. (See Materials and Methods for detailed explanations on read mapping and marker detection).

The sequencing results from the 0, 2, 4, and 8 multiplexed samples are compared (Table 5-1). Eight spores from 2 tetrads—WTx30 and *msh4*x8—were sequenced without multiplexing. Around 60,000 markers were identified in these spores—the same level of markers detected from sequencing the two parental strains. Four spores of the WTx46

tetrad were sequenced with two multiplexed samples in each lane. These spores also display a similar number of markers as the non-multiplexed spores. In contrast, spores from tetrads (WTx46, *sgs1*x1, and *pCLB2-MMS4*x1) that were multiplexed with 4 samples have a reduced average marker number of ~50,000, while spores from tetrads (*sgs1*x1, and *pCLB2-MMS4*x1) multiplexed with 8 samples have a even further reduced average marker number of ~25,000 markers (Figure 5-2A and B; Table 5-1).

To map recombination events among 4 spores of a single tetrad, only markers that are identified and genotyped in all four spores may be used, as it is impossible to unambiguously determine a conversion event at a specific SNP location if only few of the 4 spores are genotyped at that SNP site. A list of SNPs and indels genotyped in all four spores of a tetrad are compiled for each tetrad and used to deduce meiotic recombination events. This list is referred to as the "filtered marker list." Tetrads sequenced without multiplexing and those sequenced with two multiplexed samples have comparable number of filtered markers of around 54,000 to 53,000 markers, respectively (Figure 5-2A and B; Table 5-3). Tetrads sequenced by multiplexing 4 samples have, on average, around 30,000 filtered markers. Subsequently, tetrads sequenced by multiplexing 8 samples only yielded an average of 3,000 filtered markers—a number considered too low for high-resolution identification of meiotic recombination events on a genome-wide level. Thus, tetrads sequenced with 8-multiplexed samples are not included for downstream analysis of recombination events.

The coverage for individual markers in the filtered marker list is analyzed for each tetrad. For each marker, the number of reads aligned to that particular SNP or indel is tallied. Table 5-3 shows the mean, median, maximum and minimum marker coverage

for each spore. Marker coverage is between 25.5x to 35.2x for spores sequenced without multiplexing, with the minimum coverage of 5x per marker. Predictably, increasing the number of multiplexed samples would lead to a corresponding decrease in marker coverage. Figure 5-3 shows a histogram of the read coverage frequency for each of the representative tetrad selected from 0, 2, 4, and 8 multiplexed samples. Multiplexing 2 samples reduces marker coverage to 16.7x to 18.3x—still a relatively high coverage level. Spores that were multiplexed with 4 samples have a wider range of marker coverage, depending on the efficiency of each barcodes and the concentration of sequencing library loaded into the flowcell. We find that PCR amplification of libraries barcoded with TCT and CTT are less efficient than the rest. Sequencing samples barcoded with TCT and CTT yielded only 5.2x to 5.6x average marker coverage, while samples labeled with the rest of the barcodes ranged between 7.1x to 10.5x in average marker coverage. The recent study conducted by the Ma laboratory reports sequencing of one meiotic tetrad using the Roche GS20/FLX platform that yielded a mere genome-wide coverage of 3.6x to 4.9x (a marker coverage for the filtered marker list, if calculated, would be even less) (Qi et al., 2009). The marker coverage levels of our study, including the spores that are multiplexed with four samples, are much higher than the single tetrad sequenced by the Ma Laboratory.

Sequencing Reveals a High Resolution Mapping of Polymorphic Markers Genome-wide

Next, we analyzed the genome-wide distribution of more than 54,000 polymorphic markers found between S96 and YJM789 in this study. The filtered marker list from

tetrad WTx30 is chosen as an example for the following marker distribution analysis. Figure 5-4 shows the genome-wide distribution of markers from filtered marker list of WTx30. The marker distribution has a median distance of 78 bp between two consecutive markers, with a mean distance of 207 bp (SD = 579; Table 5-4A). Ninety-seven percent of inter-marker distances are within 1 kb, with 74% of all distances fall within 200 bp (Figure 5-5). We also investigated marker gaps greater than 2 kb. There are 563 inter-marker distances greater than 2 kb in size, covering approximately 2.3 Mb (19%) of the yeast genome (Table 5-4B), with the largest gap of 40.5 kb in length. Table 5-4 also shows the inter-marker distance analysis performed for all the tetrads analyzed in this study.

To assess whether recombination near chromosome ends can be detected using our current marker list, we investigated the distribution of markers near chromosome ends. Figure 5-6 shows the distribution of markers within 15 kb of the left and right ends of all 16 chromosomes. No markers are detected within 15 kb of the left chromosome end of chromosomes 6, 9, 10, 12, and 14 and the right chromosome end of chromosomes 7, 8, 10, 12, and 15. Positions of telomeres and subtelomeric elements for each chromosome end in the S288c strain are plotted as reference (Ed Louis laboratory website). In a few incidences, markers are detected in the subtelomeric elements, an example can be seen on the right end of chromosome 5. Table 5-5 shows the distance from the last detectable marker on each chromosome end to the end of that chromosome arm. Our analysis shows that although recombination events may be detected near many chromosome ends, we would miss recombination on the few chromosomes lacking detectable markers near the ends.

High-resolution mapping of Recombination Events in Two Wild-type Tetrads

Using more than 53,000 sequence polymorphisms genotyped by high-throughput sequencing, we mapped meiotic recombination events in two wild type tetrads, WTx30 and WTx46 (Figures 5-7 and 5-8). We identified a total of 190 CO events from two tetrads, averaging 95.0 COs per meiosis (SD=2.8; Table 5-6A). We observed 63.5% of all COs have detectable GC tract associated with crossing over. The average conversion tract length for GC associated with CO is 2,160 bp, with the median length of 1,892 bp. The smallest conversion tract detected is 228 bp (detected by 1 marker), while the lengthiest tract is 6,630 bp. Overall, 77.5% of all GC tracts associated with COs are detected by more than 1 polymorphic marker, with an average of 8.8 markers detecting each conversion tract. For statistics on the two individual tetrads analyzed, see Table 5-6A.

We detected a total of 98 noncrossovers (NCOs) from two tetrads, averaging 49 NCOs per tetrad (SD=11.3; Table 5-6B). The average NCO tract length is 1,979 bp (SD=1,729), while the median NCO tract length is 1,563 bp. Although the median NCO tract lengths are slightly shorter than that of GC tracts associated with COs, the difference is not significant in the two wild-type tetrads we sequenced (Wilcoxon rank-sum test, P= 0.1341). The shortest NCO detected was 32 bp (detected by 2 adjacent markers), and the longest NCO was 8,294 bp. On average, 68.4% of NCOs were detected by more than 1 polymorphic marker, averaging 7 markers per one NCO tract. Table 5-6B also shows the NCO statistics for the two individual wild type tetrads.

Although only two WT tetrads were sequenced and analyzed thus far, the observed recombination events in the current study are on par with the study conducted

by Mancera et al. Mancera and colleagues studied the recombination in 51 meioses and reported an average of 90.5 COs and 46.2 NCOs per meiosis, with 30.1% of COs without detectable conversion tracts (2008). They also observed a difference in tract lengths between NCO and GC associated with COs. Their reported median conversion tract length of GCs associated with CO was 2.0 kb and the median NCO tract length was 1.8 kb. The difference in their medians is statistically significant (Wilcoxon rank-sum test, $P < 0.0001$) (Mancera et al., 2008).

Complex Gene Conversion Patterns Detected near Crossovers

Owing to the unprecedentedly high number of markers detected in this study, a more thorough analysis of GC patterns near COs may now be conducted. COs accompanied by complex pattern of conversion tracts are observed. Distinct conversion tracts that appear to be independent NCOs and are not associated with a CO are observed near CO sites, both near COs with associated GC and those without detectable associated GCs. These conversion tracts can be found either on the two strands of chromatid that were involved in crossing over, or on the strands that were not involved in crossing over (Figure 5-9).

We measured the CO-to-NCO distance for COs whose immediate recombination neighbor is a NCO. Two populations of distance emerged. While most distances are larger than 10 kb, a small population of distances clustered between 2 to 5 kb. We hypothesize that the NCOs within 5 kb range of a crossing over site may have arisen from the same double-strand break that resolved into the CO. NCOs that appeared on the two crossing over chromatids may have originated from mismatch repair of short patches of heteroduplex DNA found near a double Holliday junction (Allers and Lichten, 2001).

Meanwhile, NCOs on chromatids not involved in crossing over may have arisen from single strand invasion of both ends of a double-strand break in which one strand invasion would later be resolved into a NCO (Oh et al., 2007). Therefore, we choose not to include NCOs within 5 kb of a CO site towards the final NCO count in this study. NCOs within 5 kb of a CO site are henceforth referred to as "complex GCs near COs".

In the two wild-type tetrads examined, 5.8% (11 cases) of all COs are accompanied by a complex GC within 5 kb on a non-crossing over chromatid, and 12.6% (24 cases) of all COs are accompanied by a complex GC within 5 kb on one of the crossing over chromatids. The average tract length for complex GCs near COs on a non-crossing over chromatid is 4,041 bp (SD=4,087), with a median length of 1,992 bp. On average, 72.7% of these complex GCs are detected with more than 1 marker. In comparison, complex GCs near COs on a crossing over chromatid have a much shorter tract length, an average of 1,167 bp (SD=888) with a median length of 908 bp. Around 70.8% of this type of complex GCs are detected with more than 1 marker. Figure 5-9 shows a comparison of the two types of complex GC patterns near COs for the two individual wild-type tetrads, and tables 5-7A and B gives the tract length measurements of the complex GCs near COs for each tetrad.

Detection of Conversion Tracts is slightly Reduced in Tetrad Multiplexed with Four Libraries, while the Detection of Crossovers Remains the Same

Although multiplexing may reduce the cost of sequencing, it is unclear how many samples one can multiplex before detection level starts to attenuate. One wild-type tetrad, WTx46, was sequenced first by multiplexing two samples in one lane (Figure 5-8) and

repeated with multiplexing four samples (Figure 5-10). Since the same tetrad was sequenced twice varying the number of samples multiplexed, a direct comparison can be made between the sequencing results. The tetrad with 2 multiplexed samples yielded 53,201 filtered markers, while the 4 multiplexed samples yielded 32,764 markers, a 38.4% reduction in the overall number of markers (Table 5-3). Analysis of the inter-marker distance shows that the average distances between adjacent markers increased from 212 bp to 348 bp (Table 5-4A). However, this reduction in the number of markers did not affect the number of COs detected between the two repeats. Ninety-three COs were detected in both cases (Table 5-6A). The positions of the same CO detected from the two tetrads were compared. The average difference between the CO positions in two repeats is 225 bp (SD = 501 bp). Twenty-eight COs out of 93 have the exact same CO position between the two repeats.

We then compared the number of GC tracts detected in two repeats. The tetrad with 2 multiplexed samples has 59 GCs associated with CO, while the tetrad with 4 multiplexed samples has 57—a slight lost in detection of 2 GC tracts (3.4%). A more apparent reduction is observed in the detection of NCOs. Fifty-seven NCOs were detected in tetrad with 2-multiplexed samples, while 49 NCOs were found in the 4-multiplexed samples, a modest reduction of 14% (Table 5-6B).

Overall, conversion tract lengths have lengthened in the tetrad sequenced with 4 multiplexed samples, suggesting that the GCs that went undetected in the 4-multiplex samples are tracts with shorter lengths. The average tract length for GCs associated with COs lengthened from 2,159 bp to 2,365 bp, a 10% increase in length. The average NCO tract length also increased from 1,975 bp to 2,509 bp, a 27% increase in length. Since the

overall number of filtered markers is less in the tetrad with 4-multiplexed samples, the number of markers involved in detecting each conversion tract is also reduced. For GCs associated with COs, the average number of markers involved in detecting each tract went from 9.7 markers to 6.2 markers. A similar effect is observed in the number of markers detecting NCOs, which went from 8.1 to 6.3 markers.

Complex GC patterns near COs are compared between the two differentially multiplexed tetrads. Nearby GCs on the non-crossing over strands are generally longer in length than the GCs on crossing over strand (Table 5-7). In both wild type tetrads, WTx30 (not multiplexed) and WTx46 (multiplexed with 2 samples), the average tract length for complex GCs near COs on a crossing over strand is more than 50% shorter than the GCs near CO on a non-crossing over strand (WTx30: 744 bp vs. 2,433 bp; WTx46: 1,379 bp vs 4,175 bp; Table 5-7). Consequently, no change is found in the number of GCs near COs on a non-crossing over strand detected between the two differentially multiplexed tetrads. However, as the average tract length for GCs on a crossing over strand is much shorter in comparison, the number of GCs near COs on a crossing over strand is reduced from 16 to 12. A corresponding decrease in the average number of markers involved in tract detection is also observed.

In summary, while the number and position of CO events can be accurately detected by tetrads sequenced by 4-multiplex samples, a small number of shorter conversion tracts that are only revealed through non-2:2 allele segregation of a single or a few markers may go undetected in the 4-multiplexed samples. Detection of large conversion tracts should remain unchanged. Multiplexing 4 samples in one lane verses multiplexing 2 samples reduces the cost of sequencing by half from around $1,000 to

$500 per tetrad. For experiments involving sequencing large number of tetrads, double the number of tetrads can be sequenced with the same cost, sacrificing a slight reduction in the detection of conversion tracts. For this reason, with the exception of *msh4* mutant tetrad, which was sequenced prior to the development of multiplexing in our lab, we have chosen to multiplex all 4 spores of the same tetrad in one sequencing lane.

msh4 mutant lacks Complex Gene Conversions near COs on the Crossing Over Chromatids

To assess the effect of meiotic mutations on recombination products, we mapped the recombination events in a tetrad carrying the *msh4* null allele (Figure 5-11). Genetic studies have shown that Msh4 is required for wild-type level crossing over in meiosis, but not required for GC events or meiotic mismatch repair (Ross-Macdonald and Roeder, 1994). Cells carrying the *msh4* null mutation also have decreased CO interference (Novak et al., 2001). Here, one tetrad with the *msh4* null mutation was sequenced without multiplexing. As reported in the genetic studies, the number of COs reduced from the wild-type average of 95 COs to 35 COs, while the number of GCs associated with COs remains high, at 77%, even slightly higher than the wild-type average of 63.5% (Table 5-6A). The slight increase in the detection level of GCs associated with COs may be due to the fact that the average conversion tract length for GCs associated with COs is longer in the *msh4* mutant—2,899 bp in *msh4* and 2,161 bp in wild types (Wilcoxon rank-sum test, $P = 0.00701$)—allowing the detection of more tracts in *msh4* mutant as compared with wild types. Similar to the results from genetic studies, the number of NCOs in *msh4* mutant remains high. Fifty-seven NCOs are detected with an average conversion tract

length of 2,068, closely similar to the wild type average of 63.5 NCOs and tract length of 1,979 bp (Table 5-6B).

One surprising finding of the *msh4* mutant was revealed from the analysis of complex GC patterns near COs. Similar to wild type, a small number of complex GCs were found near COs on the chromatids not involved in crossing over (Table 5-7A). However, no complex GCs were found near COs on the chromatids that were involved in crossing over, the only tetrad examined in this study that exhibit such behavior. This surprising result may help elucidate the molecular mechanism of CO formation in the absence of Msh4 (see discussion).

Unusually High Number of Conversion Events Detected in *pCLB2-MMS4* Tetrad

Genetic studies have indicated that Mms4 functions in concert with Mus81 to create a small class of non-interfering COs, which do not have double Holliday junctions as a recombination intermediate (Cromie et al., 2006; de los Santos et al., 2003; Whitby, 2005). Since Mms4 also have a role in DNA metabolism of vegetative cells, a meiosis-specific null allele of *mms4*, *pCLB2-MMS4* (see Materials and Methods), was constructed to investigate the role of Mms4 in meiotic recombination. One tetrad carrying the *pCLB2-MMS4* mutation was sequenced by multiplexing four samples in one lane (Figure 5-12). However, two of the *pCLB2-MMS4* spores were labeled with barcodes TCT and CTT, the two barcodes that did not PCR amplify as efficiently as the rest (data not shown). As a result, fewer reads from these two libraries were sequenced in the multiplexed sample, leading to a reduced number of filtered markers for the *pCLB2-MMS4* tetrad as compared to other tetrads sequenced by multiplexing 4 libraries (Table 5-3).

Sequencing data show that the tetrad carrying the *pCLB2-MMS4* mutation displays a slightly elevated number of COs, at around 114 (Table 5-6A), a deviation from previous genetic studies of the *mms4* mutant in SKI background, which reported a 2-fold reduction in COs (de los Santos et al., 2001). Comparing to WTx46, which was also sequenced with 4-multiplexed samples, the percentage of GCs associated with COs of the *mms4* mutant remains very similar to wild type. However, the average tract length for GCs associated with CO of the *mms4* mutant is around 3,930 bp, nearly twice the size of the wild-type GCs associated with COs (Wilcoxon rank-sum test, $P = 0.000522$). Surprisingly, more than a 4-fold increase in the number of NCOs is detected in the *pCLB2-MMS4* tetrad, with the average NCO tract nearly 900 bp shorter than that of WTx46 (1,412 bp vs. 2,509 bp; Wilcoxon rank-sum test, $P < 0.0001$; Table 5-6B). The 4-fold increase in the number of NCOs is paralleled with a 2-fold increase in the number of complex GCs near COs, both on the chromatids involved in crossing over as well as on chromatids not involved in crossing over, resulting in ~45% of COs with complex GC patterns nearby (Table 5-7; Figure 5-9). The unusually high frequency of conversion events was also observed by Mancera and colleagues, who genotyped 6 dyads and 8 single *mms4* spores using custom tiling arrays and have discovered these spores to exhibit frequent conversion events (Figure 5 of Mancera et al., 2008). The mechanism contributed to the high level of conversion events observed in *mms4* mutants remains unclear.

Increased Crossing Over Detected in Helicase Mutant *sgs1*

Sgs1 is a RecQ family DNA helicase that has been shown to suppress crossing over and prevent the formation of joint molecules comprised of 3 or 4 interconnected chromatids (Jessop and Lichten, 2008; Jessop et al., 2006; Oh et al., 2007; Oh et al., 2008). Joint molecules arise from the two ends of a double-strand break invading different chromatids. To investigate recombination products in *sgs1* mutants, we sequenced a tetrad carrying the *sgs1* null allele by multiplexing four spores (Figure 5-12). Comparing to WTx46 (multiplexed with 4 samples), *sgs1*x1 shows an elevated number of COs, at around 103. The level of COs with associated GCs remains similar to WTx46, at 61.7% (Table 5-6). In contrast, the number of NCOs is elevated in *sgs1*x1, with 87 NCOs in *sgs1*x1 and 49 NCOs in WTx46. NCO tracts also appear to be shorter on average at around 1,664 bp compared to WTx46 which is around 2,509 bp (Wilcoxon rank-sum test, $P = 0.05489$). A slight elevation in the number of complex GCs near COs is also detected on the chromatids not involved in crossing over, while the number of GCs near COs on the crossing over chromatids remains comparable to wild type.

Closely Spaced Double Crossovers Increased in *sgs1* mutant

Closely spaced double COs involving exchange between 2 chromatids may be a NCO that appears as an apparent double CO (Paques and Haber, 1999), while closely spaced double COs involving exchange between 3 chromatids may arise from a single CO with a nearby NCO on a non-crossing over strand (see gene conversion types 5 and 6 in Figure 3-1). CO interference in wild-type cells reduces the likelihood of two COs spaced closely together (Hillers, 2004). We investigated the distances between adjacent COs in wild-type tetrads. While most inter-CO distances are larger than 19kb, a small subset of inter-

CO distances cluster between 0.9 kb to 5.9 kb. A similar trend is observed in *msh4* mutant. Assuming closely spaced double COs to be rare under wild-type level of CO interference, we hypothesize that the small subset of double COs with a inter-CO distance under 6 kb are actually not two COs, but one of the above possible alternative events depending on the number of chromatids involved in the exchange, and are categorized as such in our analysis.

The *sgs1* mutant tetrad also displays a subset of inter-CO distances that are smaller than 6 kb, while the rest of the inter-CO distances are greater than 13 kb. However, *sgs1* mutant differs from wild type and *msh4* in that an elevated number of closely spaced double COs under 6 kb was observed. WTx30 and WTx46 (2 multiplexed) have 7% and 5% of inter-CO distances less than 6 kb, respectively, while *sgs1*x1 has 20% of inter-CO distances less than 6 kb (data not shown). This elevation in the number of closely spaced double COs was also observed by Oh and colleagues in a genetic study using a *sgs1-ΔC796* truncation mutant (Oh et al., 2007).

Abnormal Break-induced Replication Products Discovered in *sgs1* Mutant

Surprising aberrations were observed on chromosome 11 of the *sgs1* mutant tetrad. Initial analysis of the chromosome using markers in the filtered marker list revealed a chromosome with strikingly few markers as compared to other tetrads (Figure 5-13; Figure 5-14A). We then examined the chromosome by incorporating all markers from each spore of the tetrad that were genotyped, not simply the filtered markers. Peculiarly, two chromatids stand out as having numerous short patches of conversion tracts clustered towards three ends of the four chromosome arms (Figure 5-14B).

We proceeded to investigate the number of reads aligning to each reference genome at these short patches of conversion tracts. A closer examination of the reads aligned to these markers suggests that both parental alleles exist in the sequence reads at these markers. Large numbers of marker were not genotyped due to the ambiguity in genotyping markers with similar number of reads mapping to both the S288c and YJM789 reference genome. An example of the marker read counts for a small segment of chromosome 11 is shown (Table 5-8). Since each yeast spore was isolated from dissecting a 4-spore tetrad into single spores and was streaked to single colony prior to DNA extraction, the observation that both parental alleles co-exist over long stretches of chromosome 11 is extremely puzzling.

Since conversion tracts cluster towards the ends of the chromosome in three of the four arms, it is likely that break-induced replication was a contributing cause to this abnormality (Paques and Haber, 1999). However, why do both parental alleles exist? One possible explanation may be that duplicate copies of chromosome 11 are present in one spore, with each copy of the chromosome originated from a different parent. Further investigation is required to solve this puzzling finding.

Discussion

High-throughout Sequencing Coupled with Multiplexing is a Powerful and Cost-effective Tool for Mapping Meiotic Recombination Events

We have demonstrated that next-generation sequencing is a powerful tool for mapping the genome-wide landscape of meiotic recombination products. Our study presents highest marker resolution (~54,000) for mapping meiotic recombination events in yeast

Saccharomyces cerevisiae reported to date, surpassing the study conducted by Mancera et al using custom high density tiling arrays by more than 2,000 markers (Mancera et al., 2008). Our study also exceeds the marker resolution reported recently by the Ma group using next-generation sequencing by 8,000 markers (Qi et al., 2009). The high marker density identified between S96 and YJM789 strains allows detection of not only CO events but a great number of conversion tracts as well. Complex GCs near crossing over sites that previously went undetected may now be studied. Furthermore, the capability to multiplex 4 spores in one sequencing lane reduces the cost by three-fourth, to the extent that the cost of sequencing becomes lower than that of using microarrays. The tremendous reduction in cost allows for large scale experiments for the study of meiotic mutants. This technique will prove to be an invaluable contribution to the meiosis field and will help advance our understanding of meiotic recombination in the foreseeable future.

Mapping Recombination Events in Meiotic Mutants Provide Invaluable Insights into Recombination Pathways

One exciting aspect of the sequencing technique is the ability to map complex GC patterns near sites of crossing over. Short conversion tracts found within 5 kb of COs have been identified on chromatids involved in crossing over as well as on those uninvolved in the exchange. These GC motifs provide the footprints of how a CO was processed and resolved, and can be used to identify COs derived from different recombination pathways.

The current recombination model postulates that COs are generated by two independent pathways: the Msh4/Msh5-dependent pathway and the Mus81/Mms4-dependent pathway (Martinez-Perez and Colaiacovo, 2009; Whitby, 2005). Specifically, the model predicts that COs generated from the Msh4/Msh5-depedent pathways are subject to CO interference, while the COs generated from the Mus81/Mms4-depedent pathway are not. In addition, genetic and physical assays have shown that COs from the Msh4/Msh5-dependent pathway are resolved through a double Holliday junctions intermediate, while the COs from the Mus81/Mms4-dependent pathway are resolved through a single Holliday junction intermediate (Cromie et al., 2006). The Msh4/Msh5-depedent pathway is predicted to be the main CO pathway in *Saccharomyces cerevisiae,* while the Mus81/Mms4-dependent pathway is the predominate pathway in *Schizosaccharomyces pombe*.

The canonical double-strand break repair model of recombination predicts that COs resolved through double Holliday junctions may leave behind short tracts of heteroduplex DNA near the site of CO on the two strands of chromatid involved in crossing over (Sun et al., 1991; Szostak et al., 1983). These heteroduplex tracts may then be repaired into short conversion tracts near site of COs. On the other hand, several models exist for the resolution of single Holliday junctions in the Mus81/Mms4-dependent pathway (Cromie et al., 2006; Hollingsworth and Brill, 2004; Whitby, 2005). Models presented by Whitby, as well as Hollingsworth and Brill, predict that COs from the Mus81/Mms4-dependent pathway would have no distinct patches of heteroduplex DNA near the site of crossing over, while the model proposed by Cromie et al. predicts

the presence of distinct heteroduplex DNA tracts near COs resolved through the Mus81/Mms4-dependent pathway.

Sequencing data from the *msh4* and *pCLB2-MMS4* mutant tetrads offer us the opportunity to distinguish between these models. Our data shows that no complex conversion tracts were found near COs in *msh4* mutant on the chromatids involved in crossing over, while this particular type of complex GC pattern was enhanced in the *pCLB2-MMS4* mutant tetrad. In the *msh4* null mutant, the COs generated by Msh4/Msh5-dependent pathway would be eliminated, leaving behind only the COs generated via the Mus81/Mms4-dependent pathway. Our data from sequencing one *msh4* mutant tetrad supports the model presented by Whitby, Hollingsworth and Brill in that no complex GCs were found near COs on the crossing over chromatids in the *msh4* mutant tetrad. However, it is also possible that the conversion tracts near COs generated via the Mus81/Mms4-dependent pathway is much shorter than those generated by the Msh4/Msh5-dependent pathway, and are thus below the detection limit of our marker resolution.

Our data also supports the double-strand break repair model in that in the absence of the Mus81/Mms4-dependent pathway, all COs made will have resulted from the Msh4/Msh5-depedent pathway where a double Holliday junctions is an intermediate. Short conversion tracts are predicted to be present near a subset of COs. Thus, a higher percentage of total COs in the *pCLB2-MMS4* mutant would have detectable conversion tracts near CO sites on the two crossing over chromatids compared to the wild-type tetrads. Our data from the *pCLB2-MMS4* mutant concurs with this prediction.

Sgs1 and Mus81/Mms4 Collaborate to Resolve Multichromatid Joint Molecules

Our observation of the recombination patterns in the *sgs1* mutant tetrad also supports the current understanding of Sgs1 emerged from genetic and biochemical studies. Oh and colleagues have detected high levels of multichromatid joint molecules and an increased level of closely spaced double COs in a truncated mutation of Sgs1 (Oh et al., 2007). The authors proposed that multichromatid joint molecules appear when both ends of a double-strand break invade and prime DNA synthesis from different templates. The authors also suggested that these interconnected chromatids typically occur in wild-type cells where Sgs1 plays an important role in the resolution of these joint molecules. Joint molecules involving three or four chromatids may eventually be resolved into closely spaced double COs, or a CO with nearby conversion tracts on chromatids not involved in crossing over. In *sgs1* mutant, where there is a delay in the resolution of joint molecules, an increase in the number of closely spaced double COs and complex GCs near COs on non-crossing over strands may appear. Our analysis of recombination patterns in one *sgs1* mutant tetrad correlates with these predictions precisely.

Recently, two studies have shown that Mus81/Mms4 collaborates with Sgs1 in resolving aberrant joint molecules (Jessop and Lichten, 2008; Oh et al., 2008). These studies indicate that Sgs1 limits the formation of aberrant joint molecules, while Mus81/Mms4 are required for the resolution of joint molecules. In our study of *sgs1* and *pCLB2-MMS4* mutants, we observed an increase in the number of COs with complex GC tracts on the non-crossing over chromatids, which supports the role Mms4 has in collaborating with Sgs1 to resolve joint molecules. One possibility is that in the *sgs1* mutant, aberrant joint molecules are formed in greater numbers than in wild type,

overwhelming the ability of Mus81/Mms4 to resolve joint molecules. Therefore, joint molecules persist longer as recombination intermediates, allowing some joint molecules to be resolved as two closely spaced COs. The delay in joint molecule resolution also increases the window of opportunity for the formation of conversion tracts at locations where joint molecules previously connect the chromatids. In the *pCLB2-MMS4* mutant, Sgs1 prevents the formation of joint molecules, but some joint molecules still form at the wild type level. In the absence of *mms4*, joint molecule resolution is delayed, again leading to an increase in the number of complex GC tracts observed on the non-crossing over chromatids.

The use of next-generation sequencing in mapping meiotic recombination events opens an exciting doorway for analysis of different recombination motifs. Coupled with multiplexing, this technique allows mapping of recombination events in large number of tetrads and opens the possibility for studying many meiotic recombination mutants. Studying the recombination landscape and CO motifs in meiotic mutants will ultimately provide insights into the mechanisms of how COs are processed.

Materials and Methods

Yeast Strains, Cell growth and Genomic DNA Extraction

Table 5-9 lists the yeast strains used in this study. Haploid yeast strains carrying the *MSH4* and *SGS1* deletion in both parental backgrounds are constructed as described (Chen et al., 2008). Haploid strains with *pCLB2-MMS4* in the S96 and YJM789 background were constructed in a one step PCR-based gene replacement method using

the pFA6a-pCLB2-Nat plasmid as the template. The pFA6a-pCLB2-Nat plasmid was constructed by replacing the KanMX6 cassette in pFA6a-pCLB2-KanMX6 plasmid (Lee and Amon, 2003) with the bacteria gene *nat*, which confers nourseothricin resistance. Fifty bases upstream of the *MMS4* gene are removed and replaced with *pCLB2*. Haploid cells are mated and zygotes are dissected to select for four-spore tetrads. Yeast strains are grown and harvested for genomic DNA extraction as described in Chapter 2 of this dissertation.

Adapter Design for Illumina Sample Library

Illumina uses two proprietary adapters for single read sequencing on the GAII platform. Each adapter is ligated to one end of the DNA fragment before PCR amplification, allowing the PCR library to bind to the flowcell surface. Adapters used in this study are synthesized by joining the blunt ends of the two Solexa adapters with a single base of deoxyuridine (personal communications with Sam Coopers of the Yamamoto Lab, UCSF). Deoxyuridine is inserted on both strands of the adapter oligos. Both oligos are also modified with a phosphate at the 5' end (see Table 5-10 for adapter sequences). End-joined adapters are then ligated to blunt-ended and A-tailed genomic DNA fragments prepared as described in the standard Solexa genomic DNA library preparation protocol (http://www.illumina.com). Joining the ends of the two adapters yields a circular product after the ligation step, which increases the ligation efficiency, as well as ensuring that each DNA fragment will contain a different solexa sequencing adapter at either end. The ligated product is then incubated with *E. coli* uracil-DNA glycosylase (New England Biolabs, Ipswich, MA, USA), which catalyses the release of uracil from the deoxyuridine

base inserted between the end-joined adapters, rendering the DNA more susceptible to scission. Hydrolysis of the uracil also prevents DNA polymerase from amplifying across the joined adapter ends in the follow-up PCR amplification step, thus preventing PCR cross contamination.

Barcode Design for Multiplexed Samples

Three letter barcodes are added to end of the Illumina's proprietary adapter where sequencing initiates, resulting in sequences that contain the three letter barcodes at beginning of every read—a method similar to the ChIP-Seq multiplexing reported by Lefrancois and colleagues (Lefrancois et al., 2009). Two sets of 4 barcodes are designed for this study: 1) ACT, CAT, TGT, GTT; 2) GAT, AGT, TCT, CTT. Barcodes are designed with two principles in mind. First, within each set of 4 barcodes, no two barcodes have the same nucleotide at the first two base positions, allowing for an even distribution of A, T, G, and C at the first two positions in the pooled sequencing sample. The Solexa image analysis and base-calling software uses the first two cycles to calculate the matrix and instrument offsets. If the first two bases are homogenous, the offsets could not be calculated and will lead to poor image analysis and base-calling for the entire read (personal communications, Illumina Technical Support). Second, to prevent one barcode from been miscalled into another, the barcodes are designed for the detection of one-base sequencing error. The barcodes have a Hamming distance (Hamming, 1950) of two, which measures the minimum number of errors required to transform one barcode into another. Third, all barcodes end with a T overhang to allow the ligation of the adapters to

the A-tailed DNA fragments as described in Illumina's genomic DNA library preparation protocol.

Only one of the two sets of 4 barcodes are used when multiplexing 2 and 4 samples. For the first reason described above, when multiplexing only two yeast samples in one sequencing lane, each yeast sample is labeled with two of the four barcodes to ensure an even distribution of all 4 bases in the pooled library in the first two cycles. When multiplexing four yeast samples, each yeast sample is barcoded with one of the 4 barcodes. Both sets of barcodes are used when multiplexing 8 samples, in which case, Hamming distance becomes one and sequencing error in the barcode would not be detectable. It should be noted here that samples labeled with TCT and CTT barcodes (*pCLB2-MMS4*x1c and *pCLB2-MMS4*x1d) yielded less than half the number of total reads as samples labeled with other barcodes. Since multiplexing 8 samples in one lane yielded too few markers for meiotic recombination analysis, the second set of barcodes will no longer be used in future experiments.

Preparation of Illumina DNA library

Genomic DNA libraries are prepared as described in the Illumina's "Preparing Samples for Sequencing Genomic DNA" protocol with the following exceptions. Five µg of genomic DNA is resuspended in 150 µL of TE and fragmented using the Bioruptor UCD-200 (Diagenode, Belgium) water bath sonicator for 30 min using the following cycle: 1 min sonication, 1 min rest. End-repair and A-tailing of fragmented DNA are as described in the Illumina protocol. DNA fragments are then ligated in a 1:10 molar ratio to the adapters synthesized as described in the "Adapter Design for Illumina Sample Library" of

this chapter's Materials and Methods section. Ligation reaction is incubated with 4 μL of uracil-DNA glycosylase (New England Biolabs, Ipswich, MA, USA) in a final volume of 60 μL for 1 hr at 37ºC. Unligated adapters are then removed using QIAquick gel extraction kit (Qiagen), and samples are eluted in 30 μL EB buffer. Three μL of the eluted post-gel extraction library is used as templates for PCR according to the conditions described in the Illumina protocol. For amplifying multiplexed samples, the corresponding 3 letter barcode is added to the 3' end of Illumina's proprietary PCR primer with the 5' to 3' sequence of "AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT". PCR amplified DNA libraries are cleaned up using the QIAquick PCR Purification kit, and eluted in 30 μL of EB buffer. DNA concentrations are measured using a Nanodrop spectrophotometer and diluted to a working concentration of 10 nM. Samples for multiplexing are pooled at this stage. All 4- and 8-multiplexed samples were submitted for sequencing at the Vincent J. Coates Genomics Sequencing Laboratory of the California Institute for Quantitative Biosciences at the University of California, Berkeley. The remaining samples were processed at the Center for Advanced Technology at the University of California, San Francisco.

Mapping Sequence Reads

Multiplexed samples are first sorted into separate pools according to the barcode at the beginning of each read and trimmed of the barcode bases before alignment. Using Bowtie 0.11.3 (http://bowtie-bio.sourceforge.net) short read aligner, raw sequence reads generated from the Solexa Pipeline software suite are aligned to a merged reference

genome, containing both the S288c (genbank format 20090321) and the YJM789 (version AAFW02000000) single genomes. A merged reference genome is used in the Bowtie alignment to allow reads containing sequence polymorphisms to align to the best matched genome. Reads containing no sequence polymorphisms will align equally well to both single genomes in the merged reference genome, resulting in two alignment hits from Bowtie. The default of two mismatches are allowed. No indels are tolerated by Bowtie. The Bowtie results are then filtered and divided into S96 and YJM matches. Reads with more than 2 alignments to the merged genome are discarded. Of the reads with 2 alignments to the merged genome, reads that aligned twice to either single genome (S288c or YJM789) are also discarded. These criteria ensure that only reads aligned uniquely to either single genome or reads that aligned once to each single genome are kept for further analysis. Bowtie alignment results for the merged reference genome are separated into two single genomes, S288c and YJM789, for downstream SNP and indel analysis.

Compiling a Reference List of Known SNPs and Indels

A list of previously identified and published polymorphisms between the S288c and YJM789 yeast strains was obtained from the Steinmetz Lab (Wei et al., 2007). This list was cross checked with our re-sequencing of the S96 and YJM789 parental strains. Markers with inconsistent genotyping between the published list and our re-sequencing results are corrected with the genotype determine from our re-sequencing data. Over a thousand markers were corrected from the published list.

The published genome sequences for YJM789 is assembled into large contigs rather than full length chromosomes, and regions near chromosome ends are often covered by small contigs whose alignment to the S288c genome were not fully analyzed (personal communication with Wei Wu, Steinmetz lab). The list of polymorphisms obtained from the Steinmetz lab contains only polymorphisms detected in YJM789 contigs larger than 10 kb. As a result, much of the polymorphisms near chromosome ends and subtelomeric regions are not identified between the two strains. From the Bowtie alignments of our parental strains, we annotated SNPs that appeared in the region between the last polymorphic marker of each chromosome arm (as annotated by the Steinmetz Lab) and the end of that chromosome arm. More than two hundred new SNPs were discovered and identified.

Genotyping SNPs and Indels

The Illumina Pipeline software computes a base quality score for each base of every Solexa sequence read using an error model generated from a control sequencing lane with control sequence. Using the base quality scores generated from Illumina's pipeline, we calculate a marker quality score for each known SNP or indel in each single reference genome (S96 and YJM789). The marker is genotyped in each reference genome based on the marker quality score calculated for that reference genome. The final genotype of the marker is then determined from the two marker genotypes of the single reference genomes.

To genotype a SNP

For each reference genome, the bases that aligned to each SNP position are sorted into 5 possible nucleotide calls: A, T, G, C, or N. For each possible nucleotide call, a "nucleotide score" is the difference between the summation of the base quality scores of the bases with that particular nucleotide call, and the summation of the base quality scores of all other bases that do not have that particular nucleotide call.

For each sequencing sample, a marker quality score threshold is determined from the overall genome coverage and read length of that particular sequencing sample as described in the "Determination of Quality Score Threshold" section of the Materials and Methods. For each SNP marker, if the highest nucleotide score is above the marker threshold set for the sequencing sample, then the SNP is genotyped as that nucleotide in that single reference genome. A marker genotype is determined if the nucleotide call matches the known SNP genotype of either S96 or YJM789 strains. For example, if a SNP base is genotyped as an "A" in a sequencing sample and the SNP sequence for the S96 and YJM789 reference genomes are "T" and "A", respectively, then the SNP is genotyped as an YJM789 SNP. If none of the "nucleotide scores" of a SNP exceeds the threshold, the marker is not genotyped at the SNP position in that reference genome.

The marker genotype calls from the two reference genomes are then compared, and a final SNP genotype call is made in the following fashion. SNPs with agreeing marker genotype calls from both reference genomes are given that genotype. SNPs with conflicting marker genotype calls from the two reference genomes are not considered further. For SNPs in which only one reference genome was genotyped, a final marker genotype is made only if the genotype call matches the reference genome from which it originates. For example, if a marker is genotyped as an S96 SNP according to the Bowtie

alignment of the S288c reference genome and not genotyped in the YJM789 reference genome, then the SNP is given a final marker call of the S96 genotype. This criterion follows the fact that a read bearing an S96 SNP would preferentially match to the S288c reference genome, rather than the YJM789 reference genome. The reverse is true for an YJM789 SNP.

To genotype an indel

Marker quality scores for indels are calculated in a similar fashion as the SNP markers with the following modifications. Since an insertion may span several bases, one reference base is selected for the purpose of calculating the marker quality score for that insertion. The reference base for an insertion is the center base of an odd-length insertion and the right-most base of the two center bases of an even-length insertion. For a deletion, the marker quality score is calculated using the average of the base quality scores of the two bases directly bordering the deleted region. To ensure that the two bases directly bordering the deleted region are adjacent bases on a sequence read and that no insertion is present in between them, the first and last three bases of the sequence read are not included in analysis. Because Bowtie tolerates up to two mismatches, it is possible for a read with no sequencing errors to match at an indel site if three or less bases lie on one side of the indel.

Two criteria are applied for indel detection. First, since a deletion in one reference genome is an insertion in the opposite reference genome, an indel marker of a yeast sample can only be genotyped in either reference genome, but not both. Therefore, an indel is called if the marker quality score passes the threshold in only one of the reference genomes. Second, an ideal indel is one that displays a strong preferential alignment

between the two genomes, resulting in a high quality score in one genome and a low quality score in the opposite genome. Therefore, we impose a second criterion: the quality score of the indel from the genome that passes the threshold needs to be at least twice the quality score of the same indel from the genome that does not pass the threshold.

Determination of Quality Score Threshold

The critical threshold for marker quality scores is determined using reads simulated from the error rate model of an existing Solexa run. A section of ~700 kb of chromosome 2 is used for the simulation. A chimera chromosome consisting of sequences from both reference genomes is generated by introducing a crossover every 140 kb. The opensource Maq software is used for the simulation (http://maq.sourceforge.net). The simulator randomly introduces substitutions to the chimera genome based on the error model of an existing Solexa run, and generates a pool of simulated reads with specified read length and genome coverage level. Due to the variety of genome coverage levels and read lengths in our real experiment, simulation is performed for a number of conditions. Figure 5-15A lists the various coverage levels and the corresponding read lengths simulated. The name of the actual sample whose error model was used for each simulation is also included. Simulation at each coverage level is repeated 6 times.

At each coverage level, simulated reads are aligned to the merged reference genome using Bowtie and identified for known SNPs and indels as described in the above sections using the following marker threshold levels: 20, 50, 80, 100, 150, 200, 250, 300, 400, and 500. The final marker genotype calls at every threshold are cross checked with

the actual marker genotypes from the chimera chromosome used in the simulation. The number of incorrectly genotyped markers in each simulation at each coverage level and threshold is tallied. Table 5-11 shows the percentage of correctly and incorrectly genotyped markers, as well as the number of incorrectly genotyped markers, in the simulations of the 700 kb chimera chromosome. The threshold level chosen for each coverage level in the current study is indicated and highlighted in blue. Figure 5-15B plots the average number of incorrectly genotyped markers in the 700 kb chimera chromosome, along with the standard error of the mean from the 6 repeated simulations. Figure 5-15A summarizes the threshold chosen for each coverage level in the current study. In general, the threshold that gave the highest percentage of correctly identified markers is chosen with the following constraints: a) the average number of incorrectly indentified markers in the chimera chromosome is less than or equal to 3, and b) threshold levels 20 and 50 are not considered due to the low number of reads (1 or 2 reads) required to reach the thresholds.

Identification of Meiotic Recombination Products

COs and conversion tracts are identified and analyzed using the CrossOver software, version 3.1. A detailed explanation for how CrossOver selects for recombination events are discussed in Chapter 3 of this dissertation.

Acknowledgements

We thank Angelika Amon for the pFA6a-pCLB2-KanMX6 plasmid, and Carol Anderson for the construction of *pCLB2-MMS4* strains and critical reading of this chapter. We

thank Samantha Coopers for sharing the protocol on genomic PCR library preparation for solexa sequencing. We also thank Katherine Sorber for her helpful advice on the solexa library preparation in the early stage of this project. This work would not be possible without the help from Clement Chu on sequencing using the Illumina Genome Analyzer II. Special thanks go to James Haber for stimulating discussion and helpful suggestions.

References

Allers, T., and Lichten, M. (2001). Intermediates of yeast meiotic recombination contain heteroduplex DNA. Mol Cell *8*, 225-231.

Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. N Biotechnol *25*, 195-203.

Chen, S.Y., Tsubouchi, T., Rockmill, B., Sandler, J.S., Richards, D.R., Vader, G., Hochwagen, A., Roeder, G.S., and Fung, J.C. (2008). Global analysis of the meiotic crossover landscape. Dev Cell *15*, 401-415.

Cromie, G.A., Hyppa, R.W., Taylor, A.F., Zakharyevich, K., Hunter, N., and Smith, G.R. (2006). Single Holliday junctions are intermediates of meiotic recombination. Cell *127*, 1167-1178.

de los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J., and Hollingsworth, N.M. (2003). The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. Genetics *164*, 81-94.

de los Santos, T., Loidl, J., Larkin, B., and Hollingsworth, N.M. (2001). A role for MMS4 in the processing of recombination intermediates during meiosis in Saccharomyces cerevisiae. Genetics *159*, 1511-1525.

Hamming, R. (1950). Error Detecting and Error Correcting Codes. Bell System Technical Jounral *26*, 147-160.

Hillers, K.J. (2004). Crossover interference. Curr Biol *14*, R1036-1037.

Hollingsworth, N.M., and Brill, S.J. (2004). The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions. Genes Dev *18*, 117-125.

Jessop, L., and Lichten, M. (2008). Mus81/Mms4 endonuclease and Sgs1 helicase collaborate to ensure proper recombination intermediate metabolism during meiosis. Mol Cell *31*, 313-323.

Jessop, L., Rockmill, B., Roeder, G.S., and Lichten, M. (2006). Meiotic chromosome synapsis-promoting proteins antagonize the anti-crossover activity of sgs1. PLoS Genet *2*, e155.

Lee, B.H., and Amon, A. (2003). Role of polo-like kinase *CDC5* in programming meiosis I chromosome segregation. Science *300*, 482-486.

Lefrancois, P., Euskirchen, G.M., Auerbach, R.K., Rozowsky, J., Gibson, T., Yellman, C.M., Gerstein, M., and Snyder, M. (2009). Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. BMC Genomics *10*, 37.

Malkova, A., Swanson, J., German, M., McCusker, J.H., Housworth, E.A., Stahl, F.W., and Haber, J.E. (2004). Gene conversion and crossing over along the 405-kb left arm of Saccharomyces cerevisiae chromosome VII. Genetics *168*, 49-63.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., and Steinmetz, L.M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature *454*, 479-485.

Martinez-Perez, E., and Colaiacovo, M.P. (2009). Distribution of meiotic recombination events: talking to your neighbors. Curr Opin Genet Dev *19*, 105-112.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat Rev Genet *11*, 31-46.

Novak, J.E., Ross-Macdonald, P., and Roeder, G.S. (2001). The budding yeast Msh4 protein functions in chromosome synapsis and the regulation of crossover distribution. Genetics *158*, 1013-1025.

Oh, S.D., Jessop, L., Lao, J.P., Allers, T., Lichten, M., and Hunter, N. (2009). Stabilization and electrophoretic analysis of meiotic recombination intermediates in Saccharomyces cerevisiae. Methods Mol Biol *557*, 209-234.

Oh, S.D., Lao, J.P., Hwang, P.Y., Taylor, A.F., Smith, G.R., and Hunter, N. (2007). BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. Cell *130*, 259-272.

Oh, S.D., Lao, J.P., Taylor, A.F., Smith, G.R., and Hunter, N. (2008). RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. Mol Cell *31*, 324-336.

Paques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev *63*, 349-404.

Qi, J., Wijeratne, A.J., Tomsho, L.P., Hu, Y., Schuster, S.C., and Ma, H. (2009). Characterization of meiotic crossovers and gene conversion by whole-genome sequencing in Saccharomyces cerevisiae. BMC Genomics *10*, 475.

Roeder, G.S. (1997). Meiotic chromosomes: it takes two to tango. Genes Dev *11*, 2600-2621.

Ross-Macdonald, P., and Roeder, G.S. (1994). Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. Cell *79*, 1069-1080.

Shinohara, M., Oh, S.D., Hunter, N., and Shinohara, A. (2008). Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. Nat Genet *40*, 299-309.

Sun, H., Treco, D., and Szostak, J.W. (1991). Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the *ARG4* recombination initiation site. Cell *64*, 1155-1161.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25-35.

Wei, W., McCusker, J.H., Hyman, R.W., Jones, T., Ning, Y., Cao, Z., Gu, Z., Bruno, D., Miranda, M., Nguyen, M.*, et al.* (2007). Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. Proc Natl Acad Sci U S A *104*, 12825-12830.

Whitby, M.C. (2005). Making crossovers during meiosis. Biochem Soc Trans *33*, 1451-1455.

Winzeler, E.A., Richards, D.R., Conway, A.R., Goldstein, A.L., Kalman, S., McCullough, M.J., McCusker, J.H., Stevens, D.A., Wodicka, L., Lockhart, D.J.*, et al.* (1998). Direct allelic variation scanning of the yeast genome. Science *281*, 1194-1197.

Zickler, D., and Kleckner, N. (1999). Meiotic chromosomes: integrating structure and function. Annu Rev Genet *33*, 603-754.

Figure Legends

Figure 5-1. A schematic diagram of the sequencing workflow for multiplexing four yeast samples.

Genomic DNA from each spore is purified and fragmented using sonication. Samples from each spore are end-repaired and A-tailed before ligating to a uniquely barcoded adapter. Ligated products are incubated with uracil-DNA glycosylase and gel purified to remove unligated adapters before PCR amplification. After PCR, the four barcoded PCR libraries are pooled into one sequencing sample for sequencing on Illumina's GAII sequencer. Sequence reads are sorted into four spores using the barcodes present at the first three bases of each read. Reads from each barcode pool are then aligned to the two reference genomes (S288c and YJM789). SNPs and indels are genotyped for each spore prior to identification of recombination events in the 4-spore tetrad.

Figure 5-2. The number of markers identified in non-multiplexed and multiplexed sequencing samples.

(A) The average number of markers identified in tetrads sequenced with non-multiplexed samples and those multiplexed with 2, 4 and 8 samples are plotted. Non-filtered markers include all markers that were genotyped in each spore sample. Filtered markers represent only the markers that were genotyped in all 4 spores of a tetrad. Red squares show the average number of markers for the non-filtered marker list, while blue diamonds indicate

the average number of markers for the filtered marker list. Error bars are standard deviations (SD).

(B) The average number of non-filtered and filtered markers for all tetrads are shown. The names of the tetrads sequenced in each multiplex category are shown. SD is given for each marker average. Tetrad WTx46 is the only tetrad sequenced with 2-multiplexed samples, hence no SD was given for the average number of filtered markers.

Figure 5-3. Comparison of marker coverage levels between non-multiplexed and various multiplexed tetrads.

A histogram of marker coverage level is shown for a representative tetrad selected from each of the following categories: (A) non-multiplexed, (B) 2-multiplexed, (C) 4-multiplexed, and (D) 8-multiplexed tetrads. The histogram shows the number of reads aligned to each marker of the filtered marker list for each spore of the representative tetrad. An increase in the number of multiplexed samples leads to a corresponding decrease in marker coverage level.

Figure 5-4. Genome-wide distribution of polymorphic markers.

Polymorphic markers comprised of SNPs and indels from the filtered marker list of the WTx30 tetrad are shown. Blue vertical bars indicate the location of markers. Black vertical bars indicate the two ends of each chromosome. Yellow circles mark the location of centromeres.

Figure 5-5. Frequency of inter-marker distances.

(A) The filtered marker list from the WTx30 tetrad is used for the calculation of inter-marker distances. 97% of inter-marker distances are within 1 kb in size, and 2% of inter-marker distances are between 1 to 2 kb. The inset graph shows that 74% of inter-marker distances are within 200 bp. (B) A histogram of the frequency of inter-marker distances is shown for inter-marker distances between 0 and 200 bp in 20 bp bin size.

Figure 5-6. Distribution of polymorphic markers near chromosome ends.
Markers within 15 kb of the left (A) and right (B) chromosome ends are shown for each of the 16 chromosomes for markers from the filtered marker list of the WTx30 tetrad. Telomeres are shown in red. X' and Y' subtelomeric elements (coordinates obtained from Ed Louis lab website) are shown in green and blue, respectively. Black vertical bars indicate the marker positions.

Figure 5-7. Marker segregation profile of the WTx30 tetrad.
The marker segregation profile for all 16 chromosomes of all 4 spores of the WTx30 tetrad is shown. The WTx30 tetrad was sequenced without multiplexing and has 54,292 filtered markers. Markers genotyped as S96 are shown in blue, while markers genotyped as YJM789 markers are shown in red. Black vertical bars indicate the ends of the chromosomes.

Figure 5-8. Marker segregation profile of the WTx46 tetrad (2-multiplexed).
The marker segregation profile for all 16 chromosomes of all 4 spores of the WTx46 tetrad is shown. The WTx46 tetrad was sequenced with 2-multiplexed samples, and

53,201 filtered markers were identified. Markers genotyped as S96 are shown in blue, while markers genotyped as YJM789 are shown in red. Black vertical bars indicate the ends of the chromosomes.

Figure 5-9. Crossover and complex gene conversion motifs.

Four types of CO patterns are analyzed and the frequency of each motif is plotted for all tetrads. The four motifs are: COs without associated GCs, COs with associated GCs, COs with complex GCs on the non-crossing over stands, and COs with complex GCs on the crossing over stands. The 4 motifs are not mutually exclusive. Each COs is identified as either with associated GC or without associated before taking into account for nearby conversion tracts. The COs with nearby complex GCs may contain both COs with or without GCs at the CO junction.

Figure 5-10. Marker segregation profile of the WTx46 tetrad (4-multiplexed).

The marker segregation profile for all 16 chromosomes of all 4 spores of the WTx46 tetrad is shown. The WTx46 tetrad was sequenced with 4-multiplexed samples, and 32,764 filtered markers were identified. Markers genotyped as S96 are shown in blue, while markers genotyped as YJM789 are shown in red. Black vertical bars indicate the ends of the chromosomes.

Figure 5-11. Marker segregation profile of the *msh4*x8 tetrad.

The marker segregation profile for all 16 chromosomes of all 4 spores of the *msh4*x8 tetrad is shown. The *msh4*x8 tetrad was not multiplexed, and 54,450 filtered markers

were identified. Markers genotyped as S96 are shown in blue, while markers genotyped as YJM789 are shown in red. Black vertical bars indicate the ends of the chromosomes.

Figure 5-12. Marker segregation profile of the *pCLB2-MMS4*x1 tetrad (4-multiplexed).
The marker segregation profile for all 16 chromosomes of all 4 spores of the *pCLB2-MMS4*x1 tetrad is shown. The *pCLB2-MMS4*x1 tetrad was sequenced with 4-multiplexed samples, and 23,426 filtered markers were identified. Markers genotyped as S96 are shown in blue, while markers genotyped as YJM789 are shown in red. Black vertical bars indicate the ends of the chromosomes.

Figure 5-13. Marker segregation profile of the *sgs1*x1 tetrad (4-multiplexed).
The marker segregation profile for all 16 chromosomes of all 4 spores of the *sgs1*x1 tetrad is shown. The *sgs1*x1 tetrad was sequenced with 4-multiplexed samples, and 35,063 filtered markers were identified. Markers genotyped as S96 are shown in blue, while markers genotyped as YJM789 are shown in red. Black vertical bars indicate the ends of the chromosomes. Chromosome 11 appears to have strikingly few markers as compared to other chromosomes in the *sgs1* tetrad. Further analysis of the chromosome 11 markers are shown in figure 5-14 and table 5-8.

Figure 5-14. Distribution of filtered and non-filtered markers on chromosome 11 of the *sgs1*x1 tetrad.
The filtered markers on chromosome 11 of the *sgs1*x1 tetrad appears significantly less dense compare to other chromosomes of the same tetrad (Figure 5-13). The non-filtered

markers on chromosome 11 of *sgs1*x1 are plotted for comparison. The distribution of non-filtered marker reveals numerous short patches of conversion tract on spores A and D of *sgs1*x1.

Figure 5-15. Marker quality score thresholds used in this study.

The marker quality score thresholds chosen for each genome coverage level in the current study are summarized in (A). The number of incorrectly identified markers is averaged over six simulation cycles for the 700 kb chimera chromosome. Standard error of the mean is given. The read length and the error model used in the simulation of each coverage level are listed. For each genome coverage level, the average number of incorrectly identified markers and the standard error of the mean are plotted in (B).

Table 5-1. Summary of the number of reads, the genome coverage level, and the number of markers identified in each sequencing sample.

The following information and statistics is shown for each yeast sample sequenced: the number of multiplexed samples, the barcodes of the library, the read length, number of reads generated, the number and the percentage of total reads that mapped to either the S288c or the YJM789 genome, and the number of unfiltered markers genotyped for each strain. Also shown are the number of megabases sequenced, which is calculated by multiplying the number of mapped reads to the read length, and the mean genome coverage, which is the number of bases sequenced divided by the size of the yeast genome.

Table 5-2. DNA libraries pooled for multiplexing.

DNA libraries that were pooled into one sequencing lane for multiplexing are shown, along with the barcodes used labeling in each sample. Since pool 1 and 2 are multiplexed with libraries from only two spores, to ensure an even distribution of all 4 bases in the pooled library in the first two sequencing cycles, the DNA from each of the two spores is labeled with two barcodes from the first barcode set. In pools 1 to 5, only one of the two sets of barcodes is used. Both sets of barcodes were used in pool 6 where 8 samples were multiplexed in one lane.

Table 5-3. Coverage level analysis for filtered markers.

The number of filtered markers identified for each tetrad is shown. Marker coverage is calculated based on the number of reads aligned to each filtered marker of the tetrad. The mean marker coverage level and the median marker coverage level are shown for each spore. The maximum coverage level indicates the coverage level of the marker with the highest number of reads, and the minimum coverage level indicates the coverage level of the marker with the lowest number of reads.

Table 5-4. Analysis of inter-marker distances.

(A) The total, mean, and median of the number of inter-marker distance are shown for all tetrads sequenced in this study. The maximum and minimum distances between two markers are calculated.

(B) Inter-marker distances larger than 2 kb are analyzed and the statistics are given. The genome covered by distances greater than 2 kb are shown for each tetrad.

Table 5-5. Distance of the last marker to the chromosome end.

The leftmost and rightmost markers on every chromosome are identified in the WTx30 tetrad, and the distances between the leftmost and rightmost markers to their respective chromosome ends are shown. The two ends of each chromosome are defined according to the S288c genome sequences.

Table 5-6. Summary of COs, GCs associated with COs, and NCOs data.

(A) The number of COs and GCs associated with COs are shown for each tetrad. The tract lengths of GCs associated with COs are analyzed, and the average, median, maximum and minimum tract lengths are shown. The average and the median number of markers per GC tract associated with CO are given, as well as the number and percentage of tracts identified with greater than 1 marker.

(B) The number of NCOs identified for each tetrad is shown. The statistics for the NCO tract length and the number of markers per NCO tract are also given. The parity of NCOs indicates the ratio of YJM789 conversion tracts to S96 conversion tracts. The parity of NCOs for each tetrad is provided in counts.

Table 5-7. Summary of complex GCs near COs data.

The tract length data and the number of markers involved in the detection of each tract in the complex GCs found on the non-crossing over chromatids (A) and the crossing over chromatids (B) are shown for each tetrad. The count, average, median, maximum, and

minimum are shown for the tract length analysis. The average, median, and number of tracts detected by greater than 1 marker are shown for the marker analysis.

Table 5-8. Marker read count for a segment of chromosome 11 of *sgs1*x1.

The number of reads that aligned to the S288c and the YJM789 reference genomes are shown for a small section of chromosome 11 of the *sgs1*x1 tetrad. In each marker shown, the number of reads that aligned to the two reference genomes are relatively comparable, a strong indication that both parental alleles are present in the sequencing sample.

Table 5-9. Yeast strains.

The name and the genotype of yeast strains used in the current study are shown.

Table 5-10. Barcoded adapter oligo sequences.

Oligo sequences of 8 pairs of barcoded adapters designed for the current study are shown. All oligos are modified with an internal deoxyuridine indicated by a "/U/", as well as a 5' phosphate that facilitates the ligation of adapter oligos to the DNA library. Oligos are designed based on Illumina's the proprietary adapter sequences.

Table 5-11. Thresholds simulation results.

Simulation is performed using a 700 kb chimera chromosome for 10 marker quality score thresholds at 7 different genome coverage levels. The percentage of correctly and incorrectly genotyped markers averaged from 6 simulation repeats is shown for each simulation condition. The average the number of incorrectly genotyped markers from the

simulation is also included. Thresholds chosen for this study are indicated and highlighted in red.

Figure 5-1

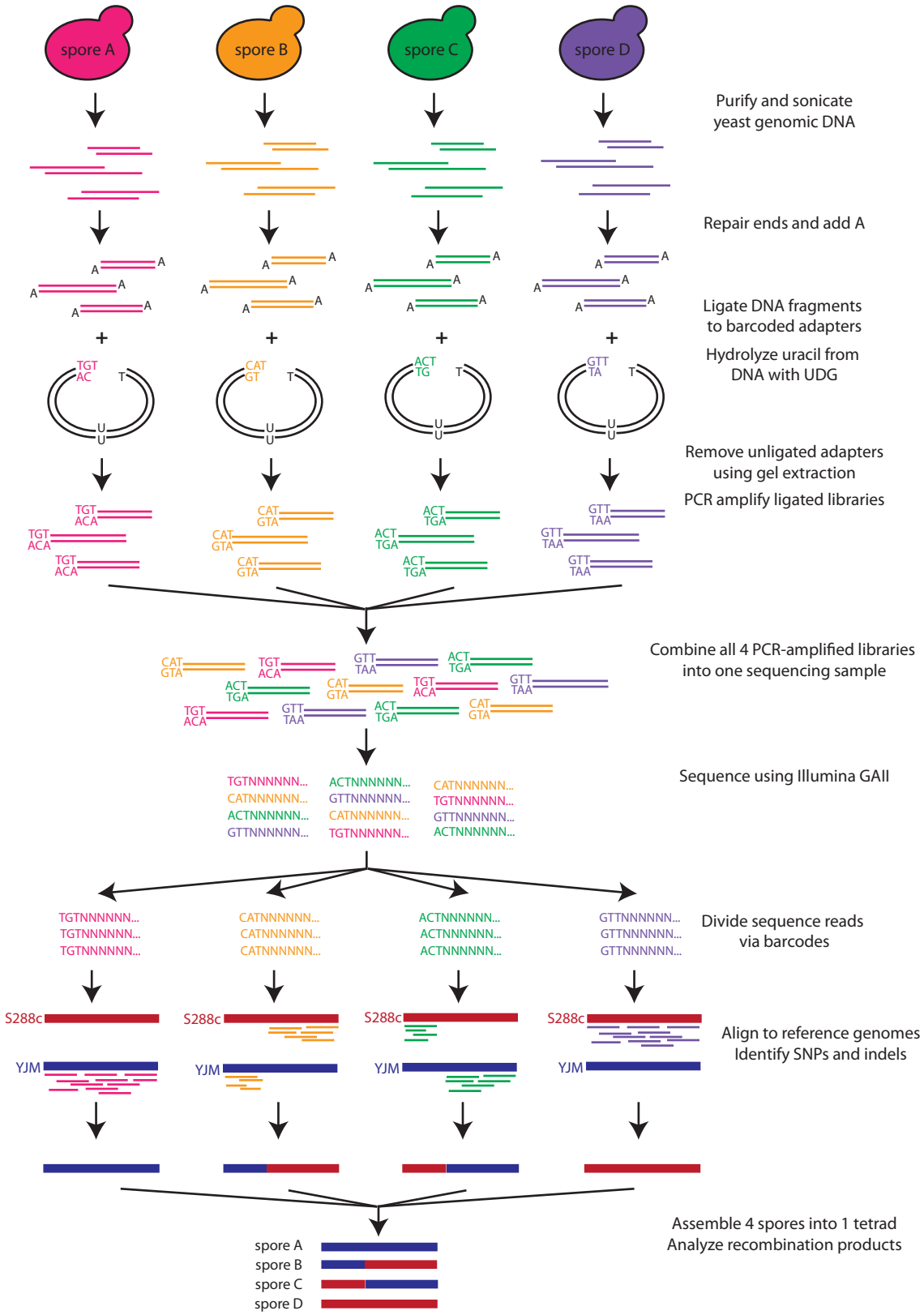
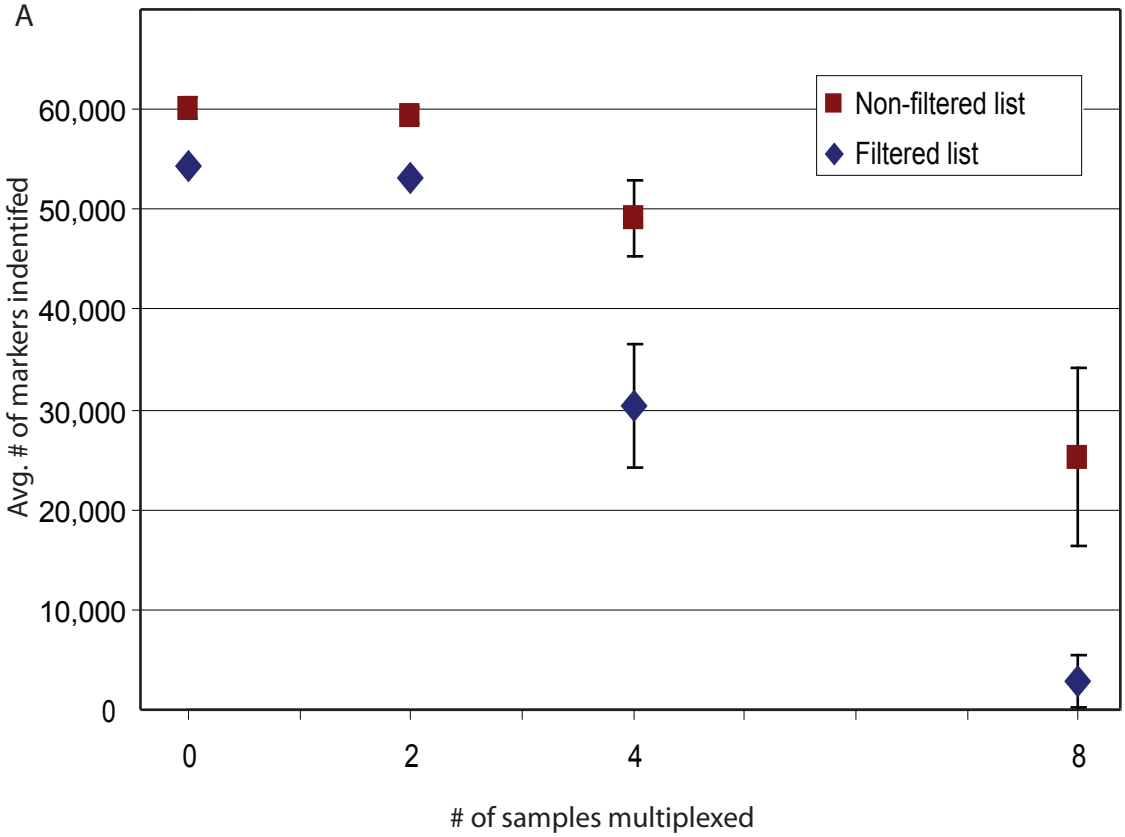

spore A
spore B
spore C
spore D
Purify and sonicate yeast genomic DNA
Repair ends and add A
Ligate DNA fragments to barcoded adapters
Hydrolyze uracil from DNA with UDG
Remove unligated adapters using gel extraction
PCR amplify ligated libraries
Combine all 4 PCR-amplified libraries into one sequencing sample
Sequence using Illumina GAII
Divide sequence reads via barcodes
S288c
YJM
Align to reference genomes
Identify SNPs and indels
Assemble 4 spores into 1 tetrad
Analyze recombination products
spore A
spore B
spore C
spore D

Figure 5-2

A



B

Tetrad Name(s)	# Samples multiplexed	Avg. Marker # (non-filtered)	SD	Avg. Marker # (filtered list)	SD
WTx30, *msh4*x8	Not multiplexed	60,147	1,038	54,371	112
WTx46	2	59,417	836	53,201	N/A
WTx46, *sgs1*x1, *pCLB2-MMS4*x1	4	49,096	3,738	30,418	6,163
*sgs1*x1, *pCLB2-MMS4x1*	8	25,243	8,969	2,827	2,584

Figure 5-3

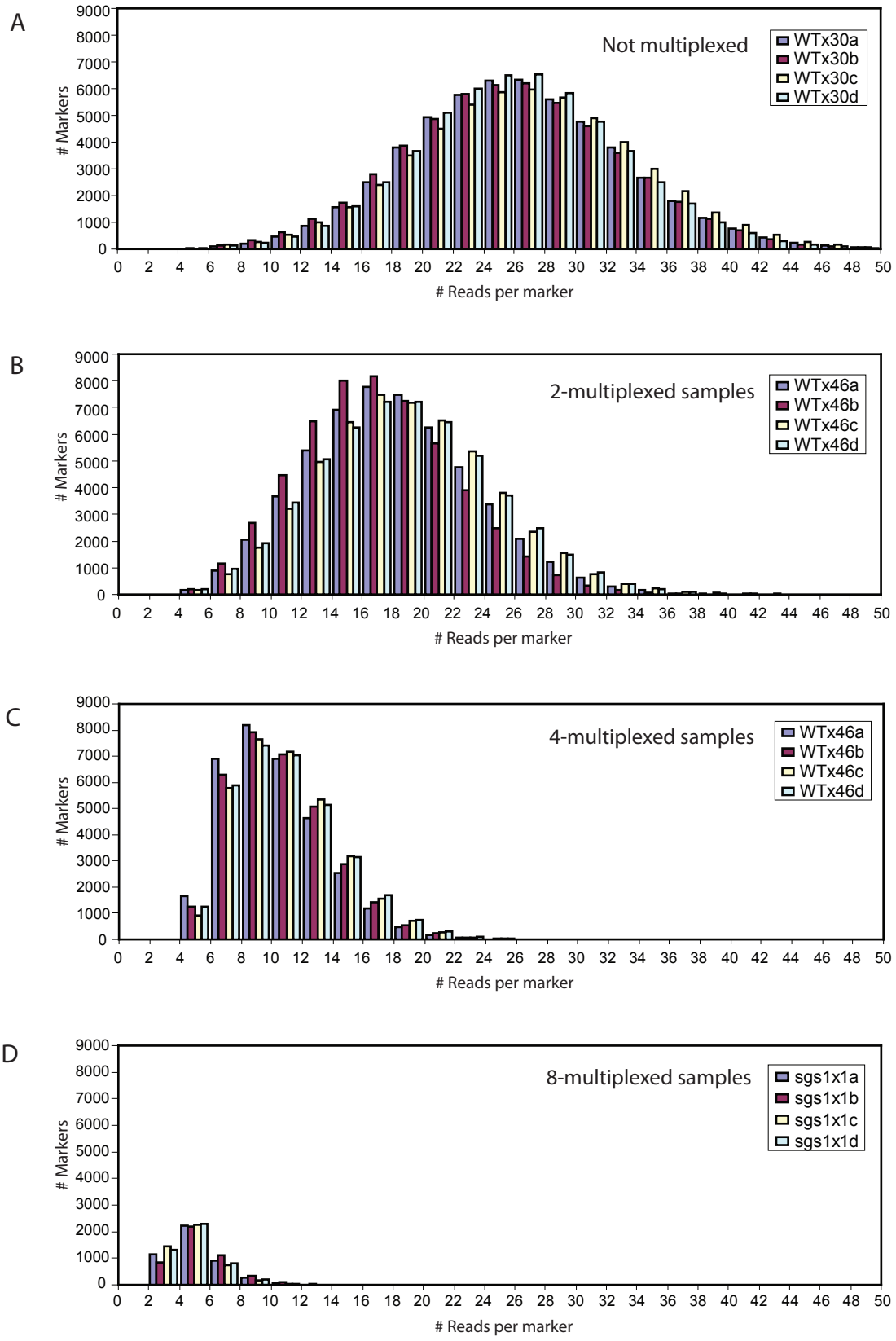

A
Not multiplexed
WTx30a
WTx30b
WTx30c
WTx30d
Markers
Reads per marker
B
2-multiplexed samples
WTx46a
WTx46b
WTx46c
WTx46d
Markers
Reads per marker
C
4-multiplexed samples
WTx46a
WTx46b
WTx46c
WTx46d
Markers
Reads per marker
D
8-multiplexed samples
sgs1x1a
sgs1x1b
sgs1x1c
sgs1x1d
Markers
Reads per marker

Figure 5-4

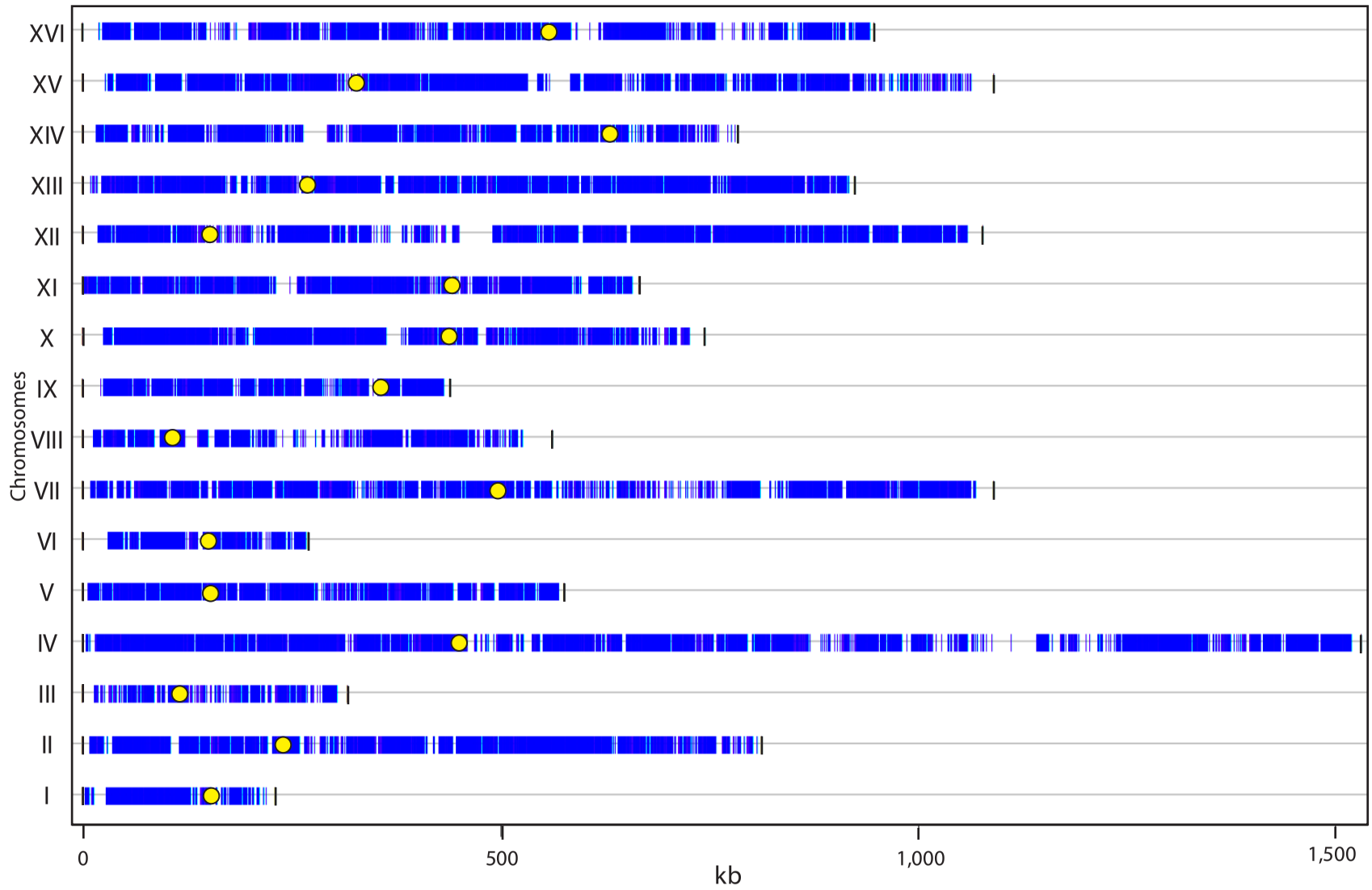
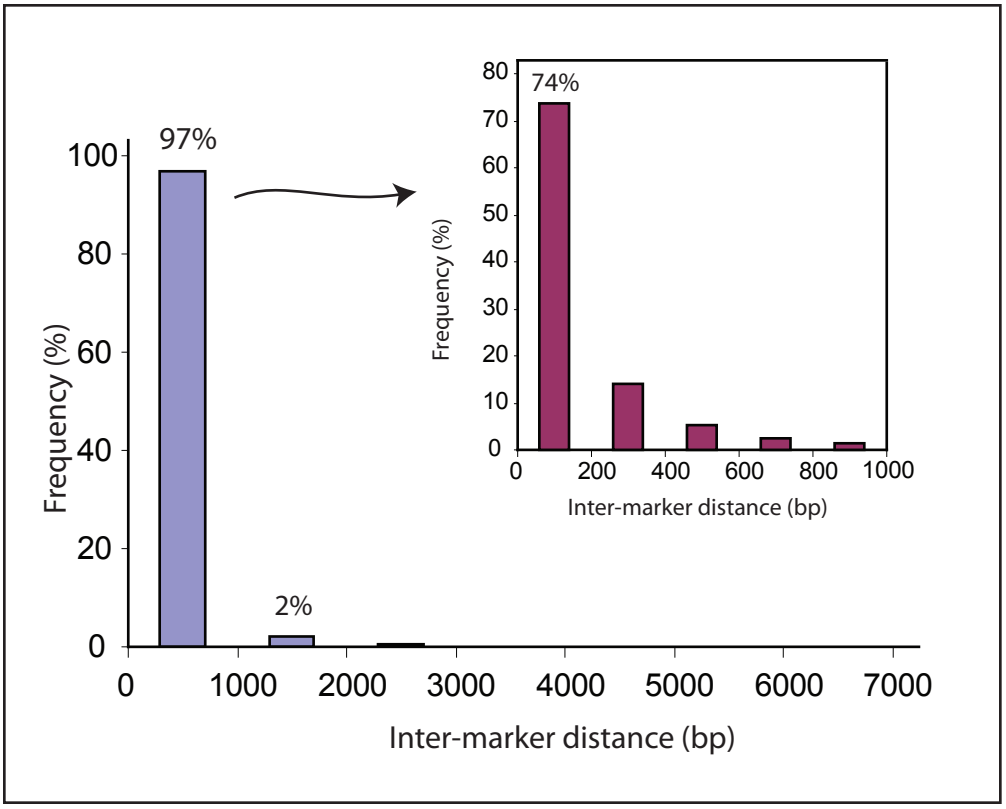

XVI
XV
XIV
XIII
XII
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Chromosomes
0
500
1,000
1,500
kb

Figure 5-5

A


97%
2%
74%
Frequency (%)
Inter-marker distance (bp)
0
20
40
60
80
100
0
1000
2000
3000
4000
5000
6000
7000
0
10
20
30
40
50
60
70
80
0
200
400
600
800
1000

B

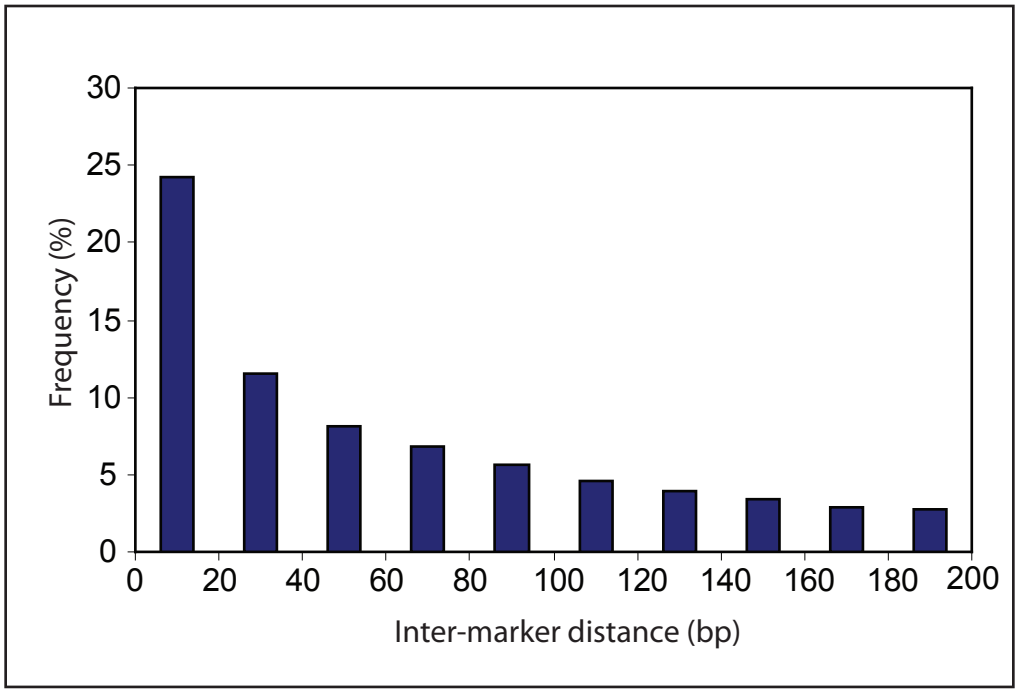
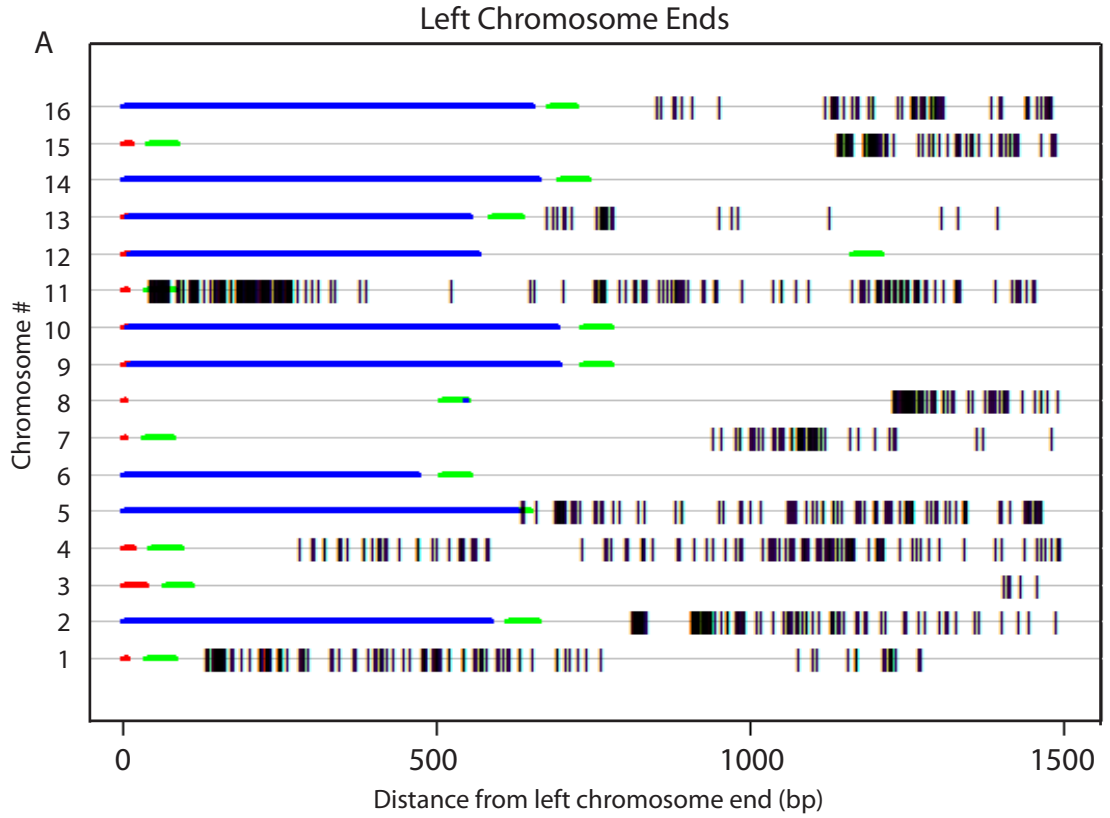

Frequency (%)
Inter-marker distance (bp)
0
5
10
15
20
25
30
0
20
40
60
80
100
120
140
160
180
200

Figure 5-6


A
Left Chromosome Ends
Chromosome #
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
500
1000
1500
Distance from left chromosome end (bp)

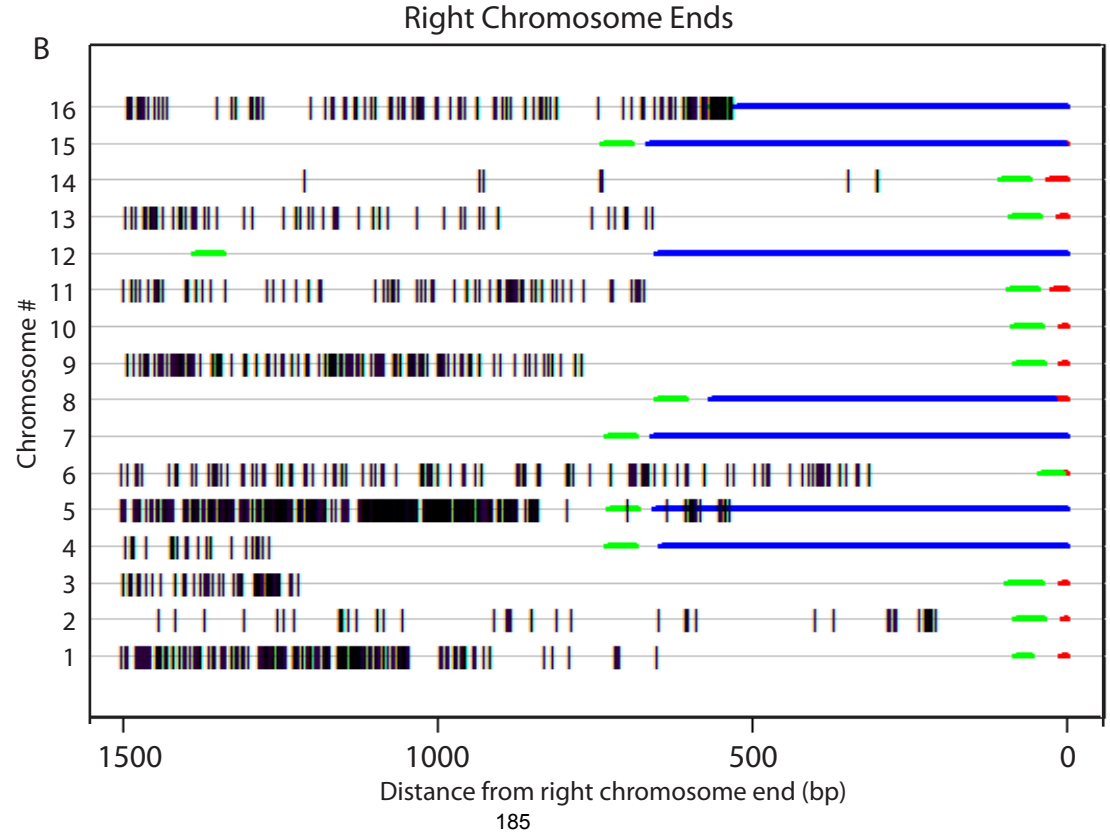

B
Right Chromosome Ends
Chromosome #
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
1500
1000
500
0
Distance from right chromosome end (bp)

Figure 5-7

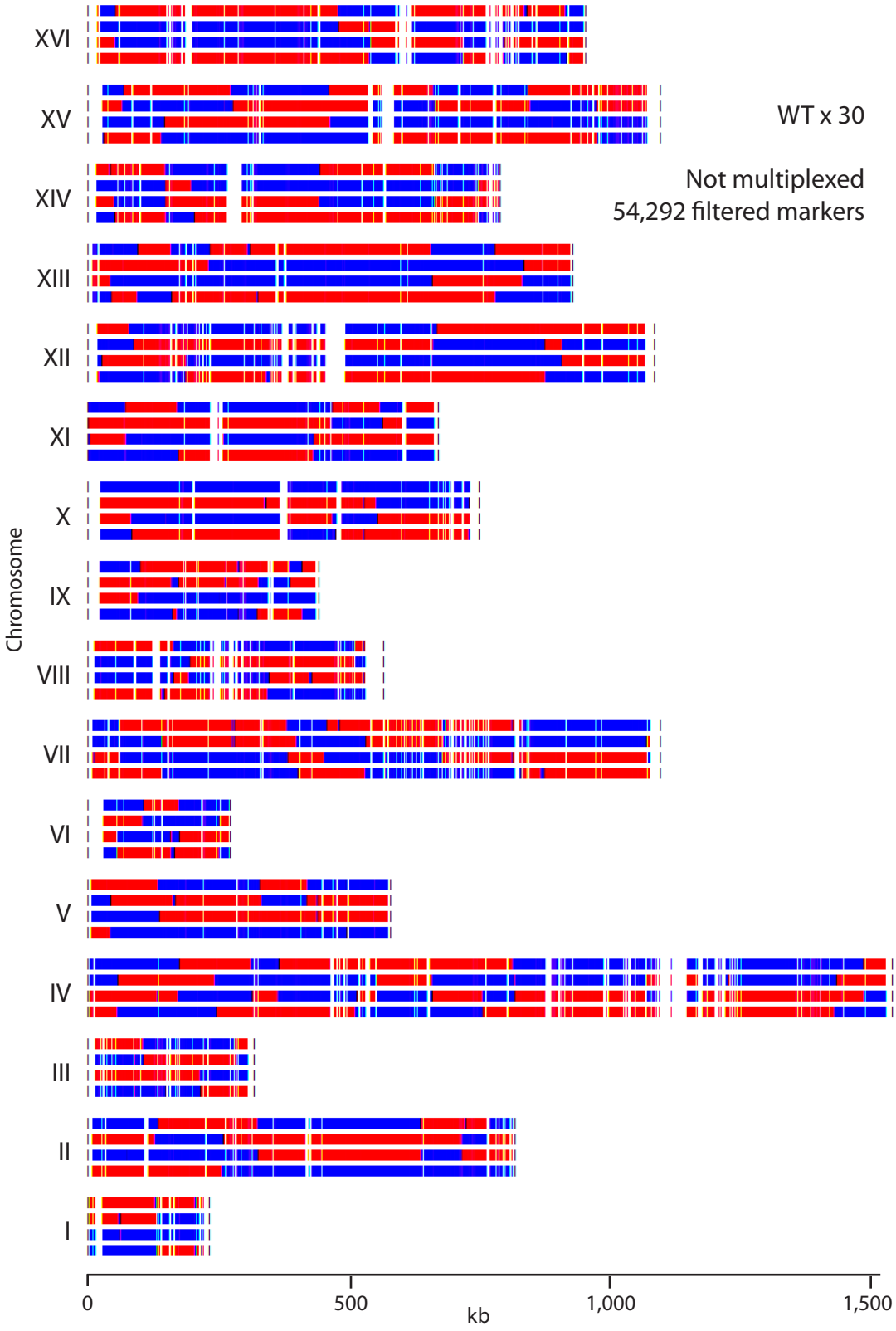

XVI
XV
XIV
XIII
XII
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Chromosome
WT x 30
Not multiplexed
54,292 filtered markers
0
500
1,000
1,500
kb

Figure 5-8

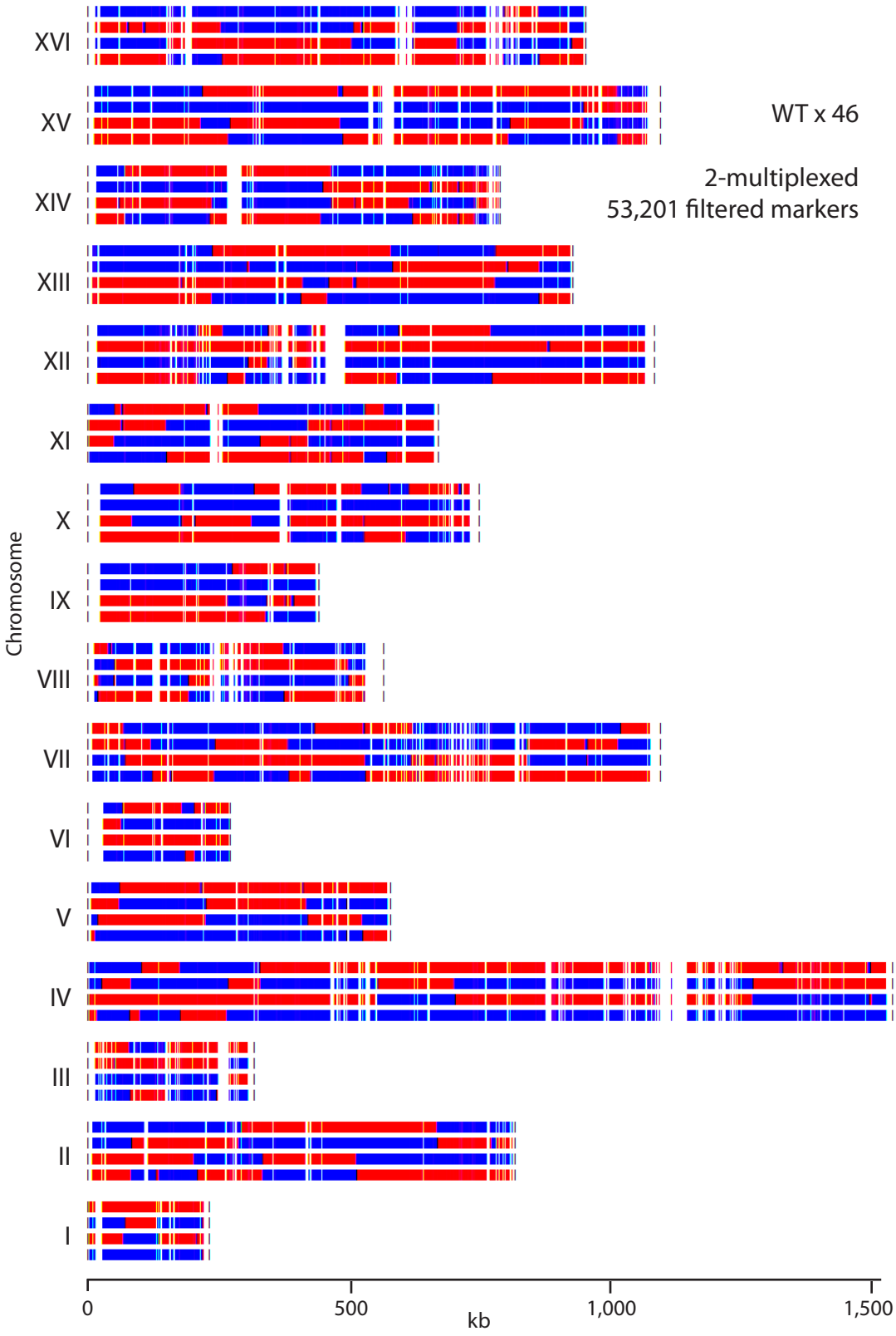

XVI
XV
XIV
XIII
XII
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Chromosome
WT x 46
2-multiplexed
53,201 filtered markers
0
500
1,000
1,500
kb

Figure 5-9

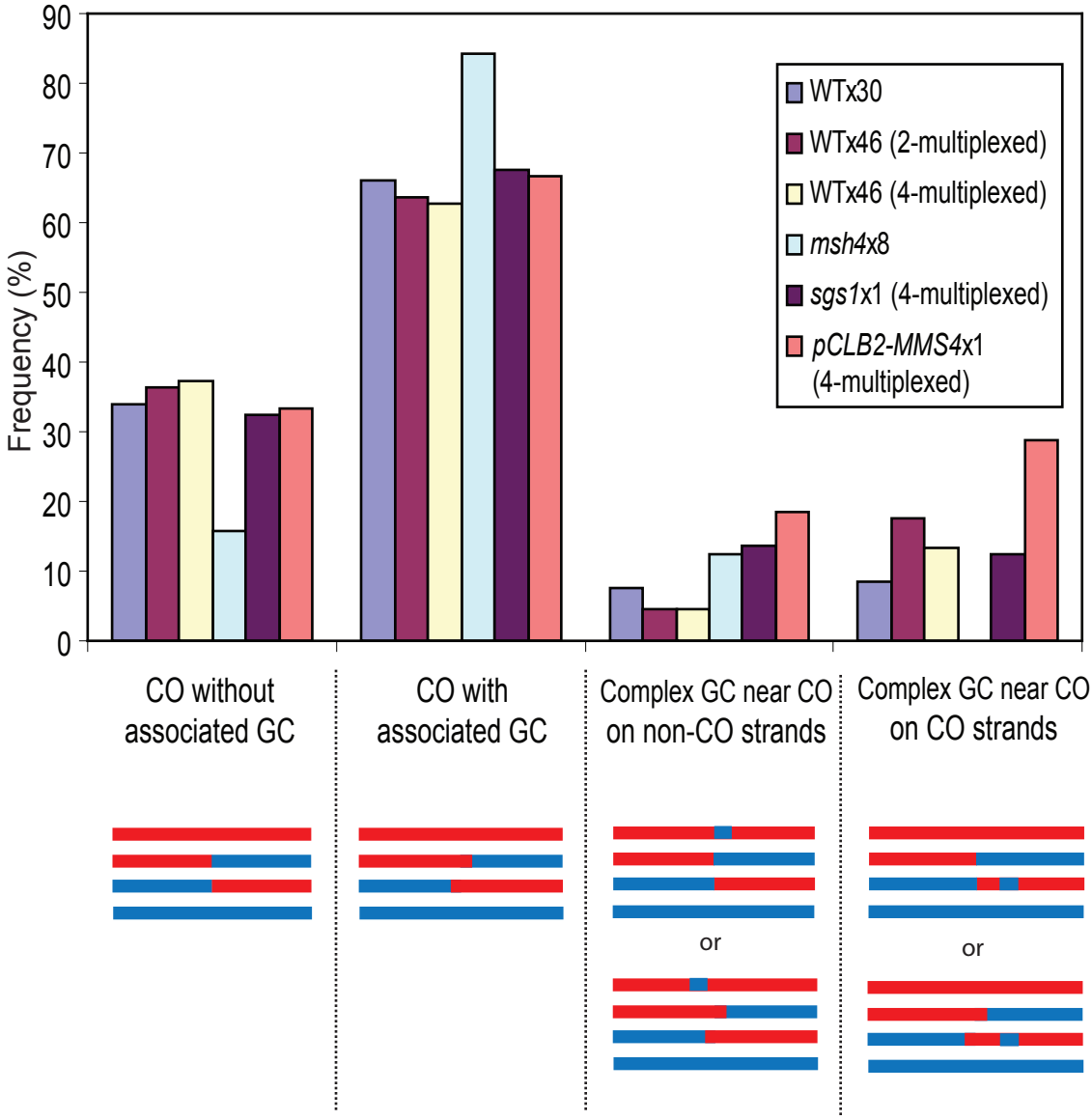

90
80
70
60
50
40
30
20
10
0
Frequency (%)
WTx30
WTx46 (2-multiplexed)
WTx46 (4-multiplexed)
msh4x8
sgs1x1 (4-multiplexed)
pCLB2-MMS4x1 (4-multiplexed)
CO without associated GC
CO with associated GC
Complex GC near CO on non-CO strands
Complex GC near CO on CO strands
or
or

Figure 5-10

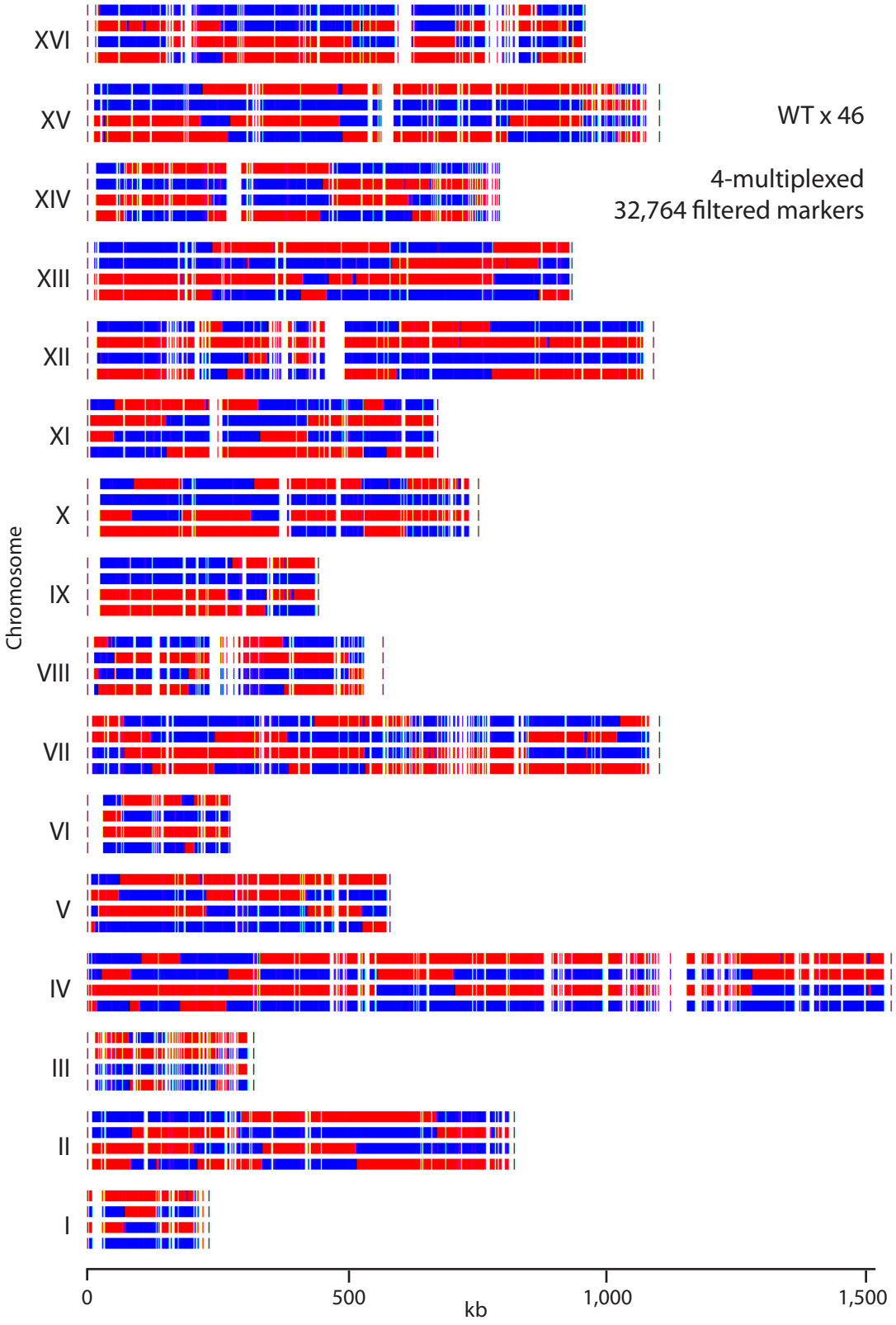

XVI
XV
XIV
XIII
XII
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Chromosome
WT x 46
4-multiplexed
32,764 filtered markers
0
500
1,000
1,500
kb

Figure 5-11

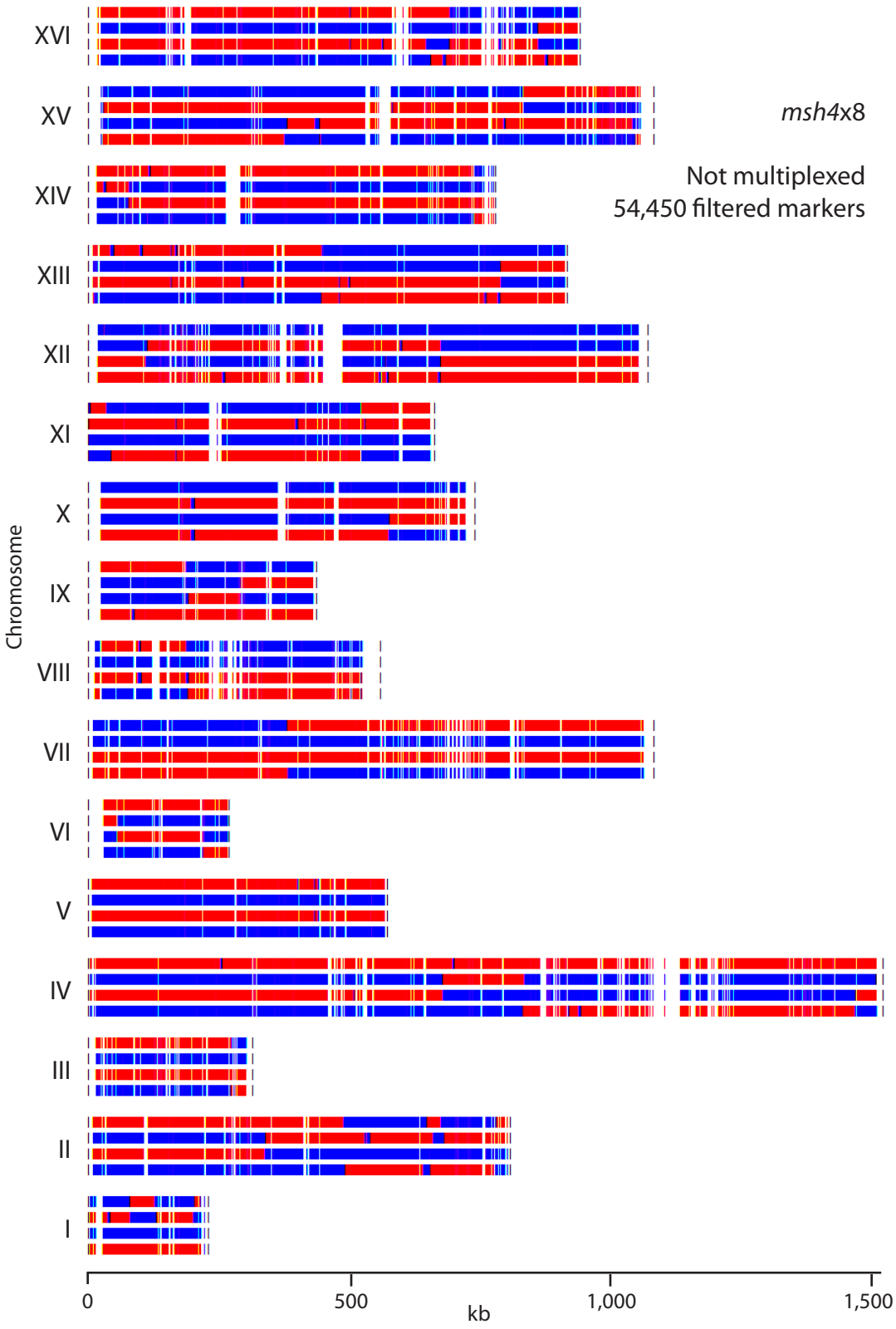

XVI
XV
XIV
XIII
XII
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Chromosome
msh4x8
Not multiplexed
54,450 filtered markers
0
500
1,000
1,500
kb

Figure 5-12

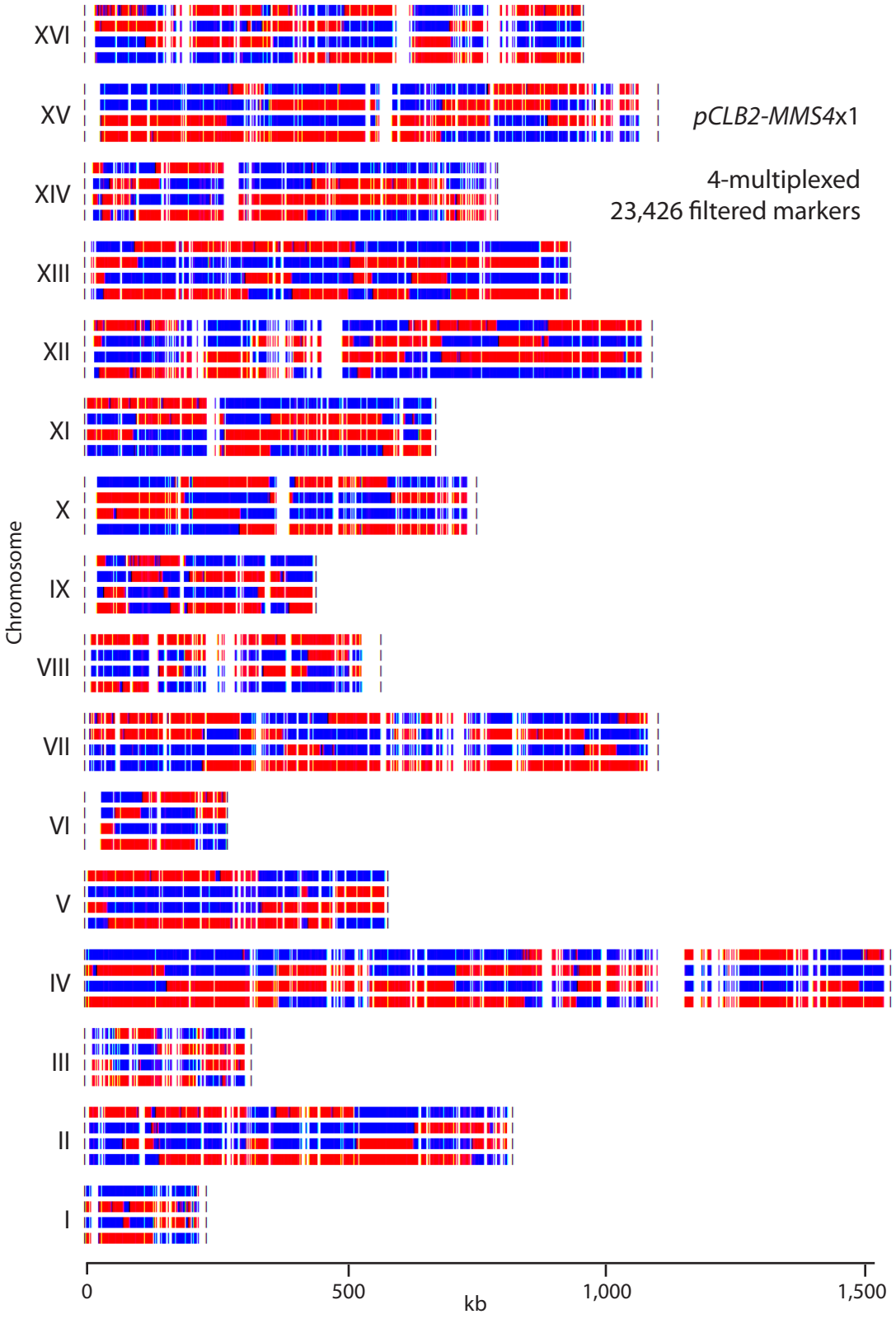

XVI
XV
XIV
XIII
XII
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Chromosome
pCLB2-MMS4x1
4-multiplexed
23,426 filtered markers
0
500
1,000
1,500
kb

Figure 5-13

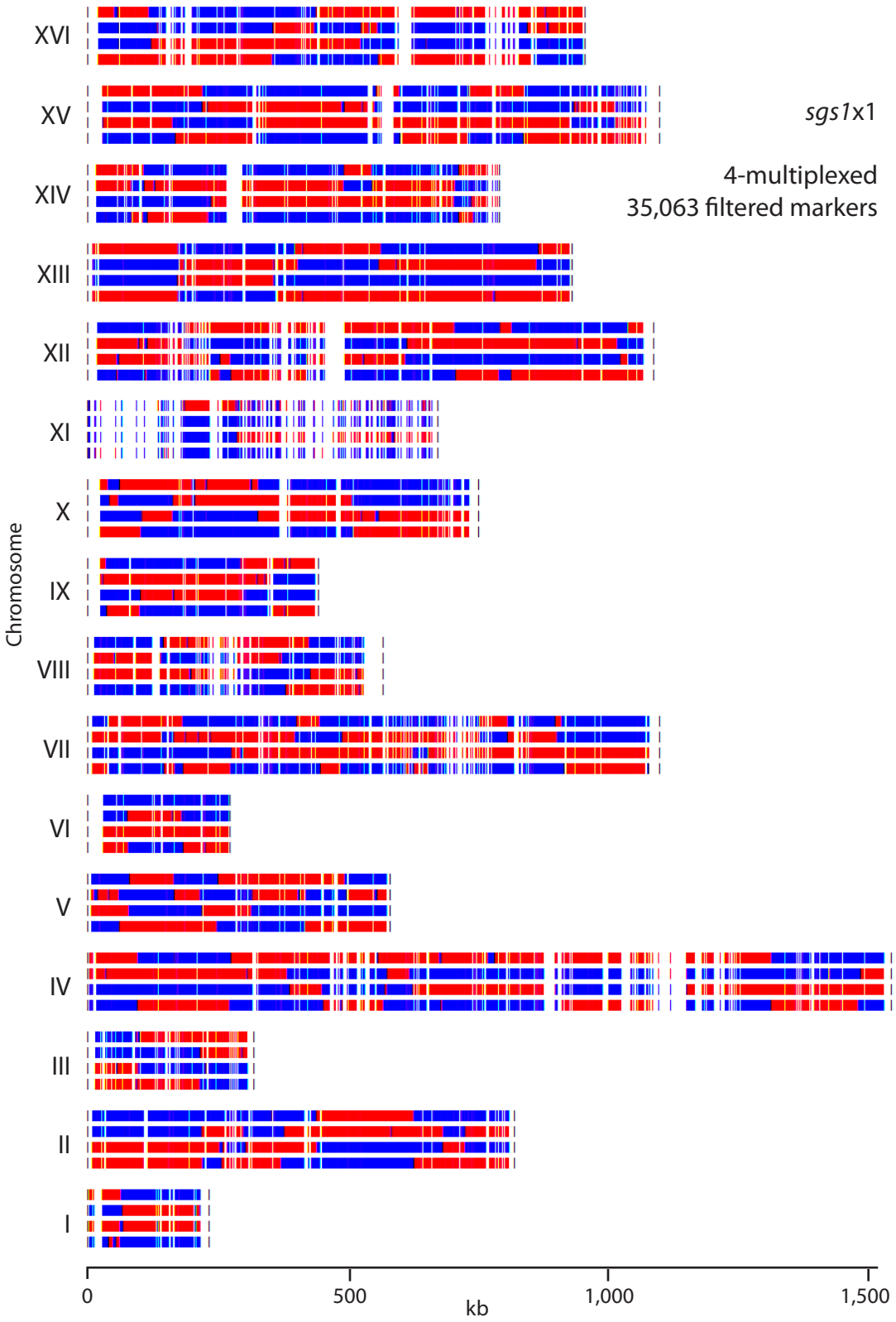

XVI
XV
XIV
XIII
XII
XI
X
IX
VIII
VII
VI
V
IV
III
II
I
Chromosome
sgs1x1
4-multiplexed
35,063 filtered markers
0
500
1,000
1,500
kb

Figure 5-14

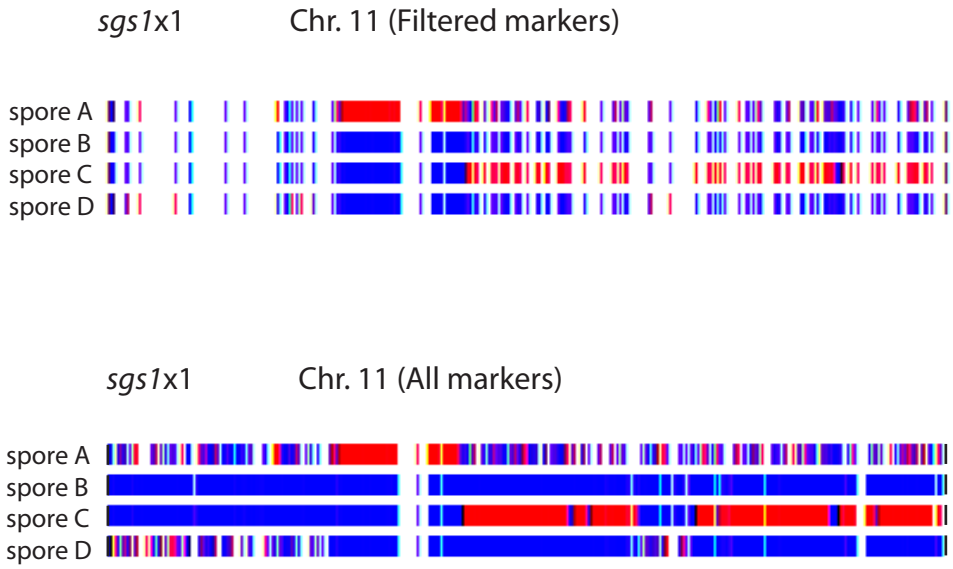

*sgs1*x1
Chr. 11 (Filtered markers)
spore A
spore B
spore C
spore D
*sgs1*x1
Chr. 11 (All markers)
spore A
spore B
spore C
spore D

Figure 5-15

A

Coverage Level	Threshold	Avg. # of incorrect markers	Standard error of the mean	Read Length (bp)	Error Model
3	80	0.83	0.31	33	WTx46a
5	80	2.67	0.80	33	WTx46a
7	80	3.00	0.58	33	WTx46a
9	150	1.67	0.67	33	WTx46a
18	150	2.83	0.87	43	*msh4*x8a
26	150	1.60	0.56	43	S96
35	150	2.33	0.61	43	*msh4*x8a

B

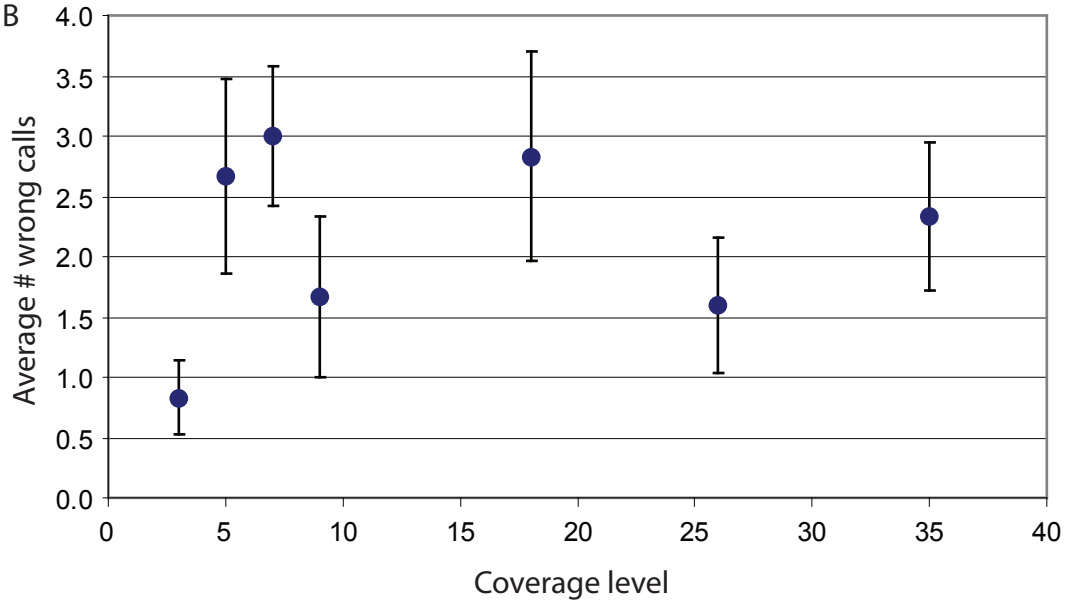


Table 5-1. Summary of the number of reads, the genome coverage level, and the number of markers identified in each sequencing sample.

Strain name	# of samples per lane	Bar-codes	Read length (bp)	# Reads	# Bps (MB)	# Mapped reads	% Mapped reads	Mean cove-rage	# Markers detected
S96	1	none	43	7676516	330	6837623	89.07	24.36	59204
YJM789	1	none	43	8185240	352	7399253	90.4	26.36	61466
WTx30a	1	none	43	8149873	350	7195292	88.29	25.63	61167
WTx30b	1	none	43	8297940	357	7466135	89.98	26.6	59091
WTx30c	1	none	43	8146882	350	7378977	90.57	26.29	59071
WTx30d	1	none	43	8285236	356	7588260	91.59	27.03	61194
*msh4*x8b	1	none	43	9642431	415	8770921	90.96	31.24	59381
*msh4*x8a	1	none	43	8438421	363	7498887	88.87	26.71	60950
*msh4*x8c	1	none	43	10950039	471	9942476	90.8	35.42	59188
*msh4*x8d	1	none	43	11873473	511	10796528	90.93	38.46	61134
WTx46a	2	ACT, TGT	40	6255681	250	5778052	92.36	19.15	60213
WTx46b	2	CAT, GTT	40	5928702	237	5411834	91.28	17.93	58838
WTx46c	2	ACT, TGT	40	6239270	250	5664098	90.78	18.77	60051
WTx46d	2	CAT, GTT	40	6076922	243	5496123	90.44	18.21	58564
WTx46a	4	ACT	33	3676946	121	3388267	92.15	9.26	49202
WTx46b	4	GTT	33	3868480	128	3522513	91.06	9.63	49880
WTx46c	4	TGT	33	3825152	126	3462712	90.52	9.47	50733
WTx46d	4	CAT	33	3763177	124	3395189	90.22	9.28	49999
*sgs1*x1a	4	TGT	33	2949510	97	2449339	83.04	6.7	52232
*sgs1*x1b	4	GTT	33	3815423	126	3364122	88.17	9.2	50273
*sgs1*x1c	4	CAT	33	3151607	104	2671275	84.76	7.3	51816
*sgs1*x1d	4	ACT	33	3009306	99	2582171	85.81	7.06	51839
*pCLB2-MMS4*x1a	4	GAT	33	4189720	138	3761300	89.77	10.28	48618
*pCLB2-MMS4*x1b	4	AGT	33	4450032	147	3932534	88.37	10.75	51556
*pCLB2-MMS4*x1c	4	TCT	33	1960357	65	1742196	88.87	4.76	42454
*pCLB2-MMS4*x1d	4	CTT	33	1952593	64	1737125	88.97	4.75	40549
*sgs1*x1a	8	TGT	33	1382911	46	1120502	81.02	3.06	27525
*sgs1*x1b	8	GTT	33	1531504	51	1324401	86.48	3.62	32340
*sgs1*x1c	8	CAT	33	1332155	44	1102751	82.78	3.01	26258
*sgs1*x1d	8	ACT	33	1344629	44	1128291	83.91	3.08	29053
*pCLB2-MMS4*x1a	8	GAT	33	1333906	44	1187608	89.03	3.25	29086
*pCLB2-MMS4*x1b	8	AGT	33	1462984	48	1276598	87.26	3.49	34906
*pCLB2-MMS4*x1c	8	TCT	33	721666	24	630378	87.35	1.72	11430
*pCLB2-MMS4*x1d	8	CTT	33	731376	24	642691	87.87	1.6	11349

Table 5-2. DNA libraries pooled for multiplexing.

Pool	# samples multiplexed	First set of barcodes				Second set of barcodes			
		ACT	TGT	CAT	GTT	GAT	AGT	TCT	CTT
1	2	WTx46a	WTx46a	WTx46b	WTx46b				
2	2	WTx46c	WTx46c	WTx46d	WTx46d				
3	4	WTx46a	WTx46b	WTx46c	WTx46d				
4	4	*sgs1*x1d	*sgs1*xa	*sgs1*x1c	*sgs1*x1b				
5	4					*pCLB2-MMS4*x1a	*pCLB2-MMS4*x1a	*pCLB2-MMS4*x1a	*pCLB2-MMS4*x1a
6	8	*sgs1*x1d	*sgs1*xa	*sgs1*x1c	*sgs1*x1b	*pCLB2-MMS4*x1a	*pCLB2-MMS4*x1a	*pCLB2-MMS4*x1a	*pCLB2-MMS4*x1a

Table 5-3. Coverage level analysis for filtered markers.

Strain Name	# of samples per lane	# filtered markers	Mean Coverage	Median Coverage	SD	Max Coverage	Min Coverage
WTx30a	1	54292	25.9	26	6.94	98	5
WTx30b	1	54292	25.5	25	7.08	221	5
WTx30c	1	54292	26.2	26	7.22	85	5
WTx30d	1	54292	25.7	26	6.77	226	5
*msh4*x8b	1	54450	30.2	30	9.68	78	5
*msh4*x8a	1	54450	25.5	25	6.80	95	5
*msh4*x8c	1	54450	34.1	34	10.37	229	5
*msh4*x8d	1	54450	35.2	35	10.11	257	5
WTx46a	2	53201	17.7	17	5.49	137	5
WTx46b	2	53201	16.7	16	5.15	51	5
WTx46c	2	53201	18.3	18	5.71	53	5
WTx46d	2	53201	18.2	18	5.74	125	5
WTX46a	4	32764	9.9	9	3.25	53	5
WTX46b	4	32764	10.2	10	3.30	35	5
WTX46c	4	32764	10.5	10	3.38	33	5
WTX46d	4	32764	10.5	10	3.48	62	5
*sgs1*x1a	4	35063	7.1	7	3.01	43	3
*sgs1*x1b	4	35063	10.2	10	3.87	40	5
*sgs1*x1c	4	35063	7.5	7	3.04	36	3
*sgs1*x1d	4	35063	7.5	7	3.10	32	3
*pCLB2-MMS4*x1a	4	23426	10.8	10	3.37	69	5
*pCLB2-MMS4*x1b	4	23426	11.8	11	4.01	71	5
*pCLB2-MMS4*x1c	4	23426	5.6	5	2.06	17	3
*pCLB2-MMS4*x1d	4	23426	5.2	5	1.91	16	3
*sgs1*x1a	8	4654	4.8	4	1.83	36	3
*sgs1*x1b	8	4654	5.2	5	1.95	17	3
*sgs1*x1c	8	4654	4.5	4	1.55	24	3
*sgs1*x1d	8	4654	4.6	4	1.54	20	3
*pCLB2-MMS4*x1a	8	1000	4.8	4	1.66	23	3
*pCLB2-MMS4*x1b	8	1000	4.9	5	1.86	24	3
*pCLB2-MMS4*x1c	8	1000	3.8	3	1.00	9	3
*pCLB2-MMS4*x1d	8	1000	3.7	3	0.93	8	3

Table 5-4. Analysis of inter-marker distances.

A.

Tetrad Name	# samples per lane	# of inter-marker distances	Mean (bp)	Median (bp)	SD (bp)	Max. (bp)	Min. (bp)
WTx30	1	56223	207	78	579	40490	1
msh4x8	1	56319	207	78	579	40490	1
WTx46	2	55075	212	80	595	40490	1
WTx46	4	33471	348	130	852	40490	1
sgs1x1	4	35828	324	113	899	40748	1
pCLB2-MMS4x1	4	23919	485	179	1146	53159	1
sgs1x1	8	4759	2431	811	5521	149437	1
pCLB2-MMS4x1	8	1005	11164	5454	16268	122979	1

B.

Tetrad Name	# samples per lane	# of inter-marker distances > 2kb	Mean (bp)	Median (bp)	SD (bp)	Genome Covered (Mb)
WTx30	1	563	4122	3047	3433	2.32
msh4x8	1	561	4131	3047	3442	2.32
WTx46	2	556	4208	3179.5	3560	2.34
WTx46	4	796	4153	3146.5	3332	3.31
sgs1x1	4	791	4445	3256	3840	3.52
pCLB2-MMS4x1	4	1005	4188	3082	3649	4.21
sgs1x1	8	1480	6678	4339.5	8434	9.88
pCLB2-MMS4x1	8	702	15747	9545	17580	11.05

Table 5-5. Distances of the last marker to the chromosome end.

Chr #	Distance of rightmost marker to the right chromosome end (bp)	Distance of leftmost marker to the left chromosome end (bp)
1	1334	6504
2	8161	2032
3	14069	12195
4	2839	12666
5	6382	5327
6	30187	3093
7	9430	21945
8	12337	35872
9	19891	7670
10	24374	15847
11	439	6700
12	17583	16246
13	6785	6568
14	15947	2959
15	11422	19361
16	8547	5278

Table 5-6. Summary of COs, GCs associated with COs, and NCOs data.

A.

Tetrad	# samples per lane	CO#	GCco#	%GCco	GCco Tract length (bp)					# Markers / tract				
					Average	Median	SD	Min	Max	Average	Median	SD	# tracts with >1 marker	% tracts with >1 marker
Combined WT	1 & 2	95	60.5	63.5	2160	1892	1480	228	6630	8.8	6	9.5	93	77.5
WTx30	1	97	62	63.9	2161	1908	1375	274	6630	8.0	5	8.4	48	77.4
WTx46	2	93	59	63.4	2159	1862	1584	228	6006	9.7	6	10.6	45	77.6
WTx46	4	93	57	61.3	2365	2175	1598	281	7383	6.2	4	7.2	41	71.9
*msh4*x8	1	35	27	77.1	2899	2726	1301	811	6071	14.4	9	13.7	24	88.9
*sgs1*x1	4	103	65	61.7	2519	1933	2098	92	10060	7.9	5	7.9	54	83.1
*mms4*x1	4	114	72	61.4	3930	3341	2874	157	14284	8.0	4	8.6	57	79.2

B.

Tetrad	# samples per lane	Independent NCOs: Tract length (bp)						# Markers / tract					Parity (YJM:S96)	
		#	Avg	Median	SD	Min	Max	Avg	Median	SD	# tracts with >1 marker	% tracts with >1 marker	1:3	3:1
Combined WT	1 & 2	49	1979	1563	1729	32	8294	7.5	3	9.8	67	68.4	41	55
WTx30	1	41	1985	1658	1582	53	7067	6.6	4	6.7	31	75.6	16	25
WTx46	2	57	1975	1527	1824	32	8294	8.1	3	11.5	36	63.2	25	30
WTx46	4	49	2509	1703	1997	83	8294	6.3	3	8.2	33	67.3	23	24
*msh4*x8	1	57	2068	1746	1804	136	12542	8.9	5	12.4	42	73.7	36	19
*sgs1*x1	4	87	1664	637	2169	66	10850	5.7	1	11.9	41	47.1	40	43
*mms4*x1	4	225	1412	518	2408	26	25273	3.1	1	6.7	71	31.6	100	121

Table 5-7. Summary of complex GCs near COs data.

A.

Tetrad	# samples per lane	Complex GC near CO (on non-CO strands): Tract length (bp)						# Markers / tract				
		Count	Avg	Median	SD	Min	Max	Avg	Median	SD	# tracts with >1 marker	% tracts with >1 marker
Combined WT	1 & 2	11	4042	1992	4087	151	14901	10.3	10	9.8	8	72.7
WTx30	1	7	2433	1868	2018	151	6142	7.7	4	7.1	4	57.1
WTx46	2	4	6857	5878	5133	770	14901	14.8	10	11.9	4	100.0
WTx46	4	4	4904	5864	2501	770	7119	7.5	6	5.9	4	100.0
*msh4*x8	1	4	2161	2255	687	1125	3008	9.5	11	6.3	3	75.0
*sgs1*x1	4	13	5730	1945	9255	74	32718	6.8	3	9.2	9	69.2
*mms4*x1	4	20	3553	1908	4299	59	16144	5.4	1	9.6	9	45.0

B.

Tetrad	# samples per lane	Complex GC near CO (on CO strands): Tract length (bp)						# Markers / tract				
		Count	Avg	Median	SD	Min	Max	Avg	Median	SD	# tracts with >1 marker	% tracts with >1 marker
Combined WT	1 & 2	24	1167	909	888	120	3742	5.9	2	8.0	17	70.1
WTx30	1	8	744	363	719	298	2534	4	2	4.5	5	62.5
WTx46	2	16	1379	1389	889	120	3742	6.9	3	9.1	12	75.0
WTx46	4	12	1787	1844	1161	33	4047	3.8	1	5.3	5	41.7
*msh4*x8	1	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
*sgs1*x1	4	12	749	419	761	99	2930	3.5	1	6.8	3	25.0
*mms4*x1	4	31	1468	1185	1337	162	6007	3.8	2	3.5	19	61.3

Table 5-8. Marker read count for a segment of chromosome 11 of *sgs1*x1.

Chr #	Marker Position	# reads aligned to S288c genome	# reads aligned to YJM789 genome
Chr 11	653750	8	12
Chr 11	653866	8	8
Chr 11	654043	12	13
Chr 11	654214	12	10
Chr 11	654427	8	5
Chr 11	654442	6	5
Chr 11	654580	6	4
Chr 11	654623	7	5
Chr 11	655479	4	12
Chr 11	655619	8	10
Chr 11	655678	1	8
Chr 11	655748	8	4
Chr 11	655763	3	5
Chr 11	655772	3	6
Chr 11	655864	12	5
Chr 11	656146	9	7
Chr 11	656230	12	10
Chr 11	656293	8	4
Chr 11	656374	13	12
Chr 11	656402	8	8
Chr 11	656748	6	2
Chr 11	656940	8	6
Chr 11	656941	8	6
Chr 11	656949	6	6
Chr 11	657079	8	8
Chr 11	657142	4	4
Chr 11	657307	10	8
Chr 11	657439	7	7
Chr 11	657450	7	5
Chr 11	657476	4	4
Chr 11	657579	3	4
Chr 11	657610	0	3
Chr 11	657633	3	2
Chr 11	657659	3	2
Chr 11	657664	3	2
Chr 11	657717	2	3
Chr 11	657813	1	3
Chr 11	657820	3	3

Table 5-9. Yeast strains.

Strain	Genotype
S96	*MAT***a** *ho lys5*
YJM789	*MATα ho::hisG lys2 cyh*
SYC1110	S96 but *msh4::kanMX6*
SYC1111	YJM789 but *msh4::kanMX6*
SYC1120	S96 but *sgs1::kanMX6*
SYC1121	YJM789 but *sgs1::kanMX6*
yCA007	S96 but *pCLB2-MMS4*
yCA008	YJM789 but *pCLB2-MMS4*
JCF1566	Spore a from WTx30, a cross between S96 and YJM789
JCF1567	Spore b from WTx30, a cross between S96 and YJM789
JCF1568	Spore c from WTx30, a cross between S96 and YJM789
JCF1568	Spore d from WTx30, a cross between S96 and YJM789
JCF1701	Spore a from WTx46, a cross between S96 and YJM789
JCF1702	Spore b from WTx46, a cross between S96 and YJM789
JCF1703	Spore c from WTx46, a cross between S96 and YJM789
JCF1704	Spore d from WTx46, a cross between S96 and YJM789
SYC1222	Spore a from *msh4*x8, a cross between SYC1110 and SYC1111
SYC1223	Spore b from *msh4*x8, a cross between SYC1110 and SYC1111
SYC1224	Spore c from *msh4*x8, a cross between SYC1110 and SYC1111
SYC1225	Spore d from *msh4*x8, a cross between SYC1110 and SYC1111
SYC1122	Spore a from *sgs1*x1, a cross between SYC1120 and SYC1121
SYC1123	Spore b from *sgs1*x1, a cross between SYC1120 and SYC1121
SYC1124	Spore c from *sgs1*x1, a cross between SYC1120 and SYC1121
SYC1125	Spore d from *sgs1*x1, a cross between SYC1120 and SYC1121
JCF3354	Spore a from *pCLB2-MMS4*x1, a cross between yCA007 and yCA008
JCF3355	Spore b from *pCLB2-MMS4*x1, a cross between yCA007 and yCA008
JCF3356	Spore c from *pCLB2-MMS4*x1, a cross between yCA007 and yCA008
JCF3357	Spore d from *pCLB2-MMS4*x1, a cross between yCA007 and yCA008

Table 5-10. Barcoded adapter oligo sequences.

Oligo Name	Bar-code	Modifications	Sequence (5' to 3')
SYC 413	TGT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGT
SYC 414	TGT	5' phosphate, internal dU	CAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
SYC 417	CAT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAT
SYC 418	CAT	5' phosphate, internal dU	TGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
SYC 419	ACT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTACT
SYC 420	ACT	5' phosphate, internal dU	GTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
SYC 421	GTT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTT
SYC 422	GTT	5' phosphate, internal dU	ACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
SYC 433	AGT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGT
SYC 434	AGT	5' phosphate, internal dU	CTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
SYC 435	GAT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAT
SYC 436	GAT	5' phosphate, internal dU	TCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
SYC 437	CTT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTT
SYC 438	CTT	5' phosphate, internal dU	AGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT
SYC 439	TCT	5' phosphate, internal dU	GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/U/ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCT
SYC 440	TCT	5' phosphate, internal dU	GAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/U/CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT

Table 5-11. Threshold simulation results.

Coverage level	Threshold level	% markers correctly identified	% markers incorrectly identified (x 10^{-3})	Avg. # of markers incorrectly identified in the chimera chr	Thresholds chosen for current study
3x	500	0.0	0.00	0.0	
3x	400	0.0	0.00	0.0	
3x	300	0.1	0.04	0.2	
3x	250	0.5	0.04	0.2	
3x	200	3.0	0.08	0.3	
3x	150	13.0	0.11	0.5	
3x	100	32.7	0.11	0.5	
3x	**80**	**52.4**	**0.19**	**0.8**	current threshold
3x	50	74.8	0.61	2.7	
3x	20	91.2	1.91	8.3	
5x	500	0.0	0.00	0.0	
5x	400	0.1	0.00	0.0	
5x	300	2.2	0.00	0.0	
5x	250	7.4	0.00	0.0	
5x	200	20.5	0.04	0.2	
5x	150	46.0	0.15	0.7	
5x	100	69.7	0.31	1.3	
5x	**80**	**82.8**	**0.61**	**2.7**	current threshold
5x	50	91.9	1.57	6.8	
5x	20	96.4	2.64	11.5	
7x	500	0.1	0.00	0.0	
7x	400	1.3	0.00	0.0	
7x	300	12.3	0.00	0.0	
7x	250	26.4	0.00	0.0	
7x	200	48.6	0.08	0.3	
7x	150	73.5	0.19	0.8	
7x	100	87.5	0.46	2.0	
7x	**80**	**92.6**	**0.69**	**3.0**	current threshold
7x	50	95.8	1.34	5.8	
7x	20	97.1	2.56	11.2	
9x	500	0.8	0.00	0.0	
9x	400	7.1	0.04	0.2	
9x	300	32.6	0.08	0.3	
9x	250	52.3	0.15	0.7	
9x	200	72.8	0.34	1.5	
9x	**150**	**87.5**	**0.38**	**1.7**	current threshold
9x	100	94.1	1.38	6.0	
9x	80	95.7	1.49	6.5	
9x	50	96.7	2.14	9.3	
9x	20	97.2	2.98	13.0	
18x	500	55.6	0.23	1.0	
18x	400	81.5	0.27	1.2	
18x	300	93.8	0.38	1.7	
18x	250	96.0	0.42	1.8	
18x	200	96.8	0.50	2.2	
18x	**150**	**97.2**	**0.65**	**2.8**	current threshold
18x	100	97.4	0.73	3.2	
18x	80	97.5	0.77	3.3	
18x	50	97.6	0.92	4.0	
18x	20	97.7	1.15	5.0	

Continues from Table 5-11:

Coverage level	Threshold level	% markers correctly identified	% markers incorrectly identified (x 10^{-3})	Avg. # of markers incorrectly identified in the chimera chr.	Thresholds chosen for current study
26x	500	89.8	0.11	0.5	
26x	400	95.4	0.11	0.5	
26x	300	96.8	0.19	0.8	
26x	250	97.0	0.23	1.0	
26x	200	97.2	0.34	1.5	
26x	**150**	**97.3**	**0.34**	**1.5**	current threshold
26x	100	97.5	0.38	1.7	
26x	80	97.5	0.50	2.2	
26x	50	97.6	0.84	3.7	
26x	20	97.7	1.34	5.8	
35x	500	96.7	0.31	1.3	
35x	400	97.1	0.38	1.7	
35x	300	97.3	0.42	1.8	
35x	250	97.4	0.50	2.2	
35x	200	97.5	0.54	2.3	
35x	**150**	**97.5**	**0.54**	**2.3**	current threshold
35x	100	97.6	0.61	2.7	
35x	80	97.6	0.73	3.2	
35x	50	97.6	0.77	3.3	
35x	20	97.7	1.03	4.5	

Chapter 6: Conclusions and Future Directions

Whole-genome mapping of Meiotic Recombination Events

In this thesis, we described two methods for mapping meiotic recombination events genome-wide using budding yeast strains S96 and YJM789. Chapter 2 and 4 described the method of using the Affymetrix S98 expression microarrays. We demonstrated that many aspects of the crossover (CO) behavior can be evaluated simultaneously, including CO and gene conversion levels, CO interference, CO homeostais, chromatid interference, and the relationship between COs near centromeres and telomeres. We also showed that because COs are mapped genome-wide, fewer tetrads are needed to obtain statistically significant data compared with the hundreds and sometimes thousands of tetrads required using genetic dissection of phenotypic markers.

In chapter 5, we described the second method of mapping recombination events uses Illumina's Genome Analyzer II sequencer. We demonstrated that next-generation sequencing is a powerful tool for mapping the genome-wide landscape of meiotic recombination products, and improves the marker detection level from ~8,000 markers detected using microarrays to that of ~54,000 markers. When coupled with multiplexing, sequencing drastically reduces the cost of this method to lower than that of microarrays. The tremendous reduction in cost makes it possible for large scale experiments for the study of recombination products in meiotic mutants. We also showed that studying the recombination maps of COs and NCOs in meiotic mutants can help unlock the molecular mechanisms and regulations that govern the distribution and formation of recombination events. This technique will prove to be an invaluable contribution to the meiosis field and

will help advance our understanding of meiotic recombination in the near future. Below, we summarize a few key findings and results revealed in our study.

Relationship between Chromosomal Landscapes and Recombination Events

Since the positions of COs and noncrossovers (NCOs) can be mapped, we can calculate the distance between these recombination events and chromosomal landmarks such as centromeres and telomeres. We have shown that both COs and NCOs are typically repressed within 20 kb of centromeres, and that this repression was lost in *zip1* mutants. Tsubouchi and Roeder have shown that Zip1 localizes to centromeres early in meiosis and is thought to hold the chromosomes together to aid in the homology search (Tsubouchi and Roeder, 2005). In a separate publication, the authors also showed that synapsis initiates from the centromeres and short stretches of linear Zip1 are found to be associated with centromeres early in meiosis (Tsubouchi et al., 2008). Our data suggest that in addition to holding the chromosomes together at the centromeres, Zip1 may also play a role in preventing pericentromeric recombination, which could lead to chromosome missegregation and aneuploid gametes (Rockmill et al., 2006).

CO Interference can be Calculated using a Small Number of Tetrads

We also showed that inter-CO distances can be used to calculate CO interference, providing an ultra-fast and powerful assay to measure the global CO interference in wild type as well as meiotic mutants. Traditional genetic assay of measuring CO interference requires manual dissection of hundreds, and sometimes thousands, of four-spore tetrads. Our method requires only a few tetrads to calculate CO interference level. We described

a method to calculate interference levels using inter-CO distances and gamma distribution function, and have calculated CO interference levels in meiotic mutants with as few as seven tetrads. We showed that in mutants whose interference values were measured using tetrad dissection, the interference values calculated using the microarray technique was similar to the values reported from previous genetic studies.

CO Homeostasis

By examining the correlation between the number of COs and NCOs in 26 wild type tetrads, we have shown that the CO number is not correlated with the number of NCOs, supporting the role of CO homeostasis as part of the CO control. The correlation between the number of COs and NCOs was also evaluated in 26 *zip2* tetrads and 34 *zip4* tetrads. An increase in correlation between COs and NCOs was observed in the *zip2* and *zip4* mutants, suggesting that CO homeostasis is reduced in these mutants. At the time of the study, the cost of microarrays prevented the study of CO homeostasis in more meiotic mutants. Using a multiplex barcoding system with deep sequencing, CO homeostasis can thus be evaluated in additional meiotic mutants at a reasonable cost.

Recombination motifs

Deep sequencing has revealed the presence of complex gene conversion (GC) tracts near COs. These conversion tracts are likely the result of mismatch repairs of heteroduplex DNA that arise from CO resolution. These complex GC patterns near COs provide the footprint of how each CO was resolved and hint at the possible recombination intermediates that were involved in each CO formation. We showed that the ratio

between different conversion motifs near COs is perturbed in *msh4*, *sgs1*, and *pCLB2-MMS4* mutants as compared to the wild type ratio. Our data suggest that the ratio of COs resolved from different pathways was altered in the mutants. Specifically, in *msh4* mutants, we observed no COs with complex GCs on the chromatids involved in crossing over—a CO motif that is predicted for COs resolved through a double Holliday junctions intermediate. This suggests that the remaining COs in the *msh4* mutants, which are presumably resolved through the Mus81/Mms4-dependent pathways, did not resolve through a double Holliday junction. Cromie et al. has reported that single Holliday junction was the intermediate for COs formed from Mus81/Mms4-dependent pathways (Cromie et al., 2006). Our data suggest that the mechanisms involved in resolving single Holliday junctions is likely to leave no, or very short, heteroduplex DNA tracts on the two chromatids involved in crossing over.

Both Sgs1 and Mus81/Mms4 have been implicated to have a role in resolving multichromatid joint molecules that are formed from both ends of a DSB invading different homologous chromosomes (Jessop and Lichten, 2008; Oh et al., 2008). In our study, we observed an elevated level of GCs near COs on chromatids not involved in crossing over in both the *sgs1* and the *pCLB2-MMS4* single mutants. Our data support the joint role Sgs1 and Mus81/Mms4 have in resolving joint molecules, and suggest that in the absence of either Sgs1 or Mms4, joint molecules persist longer as recombination intermediates but the other protein(s) that are still present will eventually resolve the joint molecule. However, the delay increases the window of opportunity for the formation of closely-spaced COs and the conversion tracts at locations where joint molecules previously connect the chromatids. Our study shows that studying the recombination

landscape and CO motifs in meiotic mutants provides invaluable insights into the mechanisms of how COs are processed.

References

Cromie, G.A., Hyppa, R.W., Taylor, A.F., Zakharyevich, K., Hunter, N., and Smith, G.R. (2006). Single Holliday junctions are intermediates of meiotic recombination. Cell *127*, 1167-1178.

Jessop, L., and Lichten, M. (2008). Mus81/Mms4 endonuclease and Sgs1 helicase collaborate to ensure proper recombination intermediate metabolism during meiosis. Mol Cell *31*, 313-323.

Oh, S.D., Lao, J.P., Taylor, A.F., Smith, G.R., and Hunter, N. (2008). RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. Mol Cell *31*, 324-336.

Rockmill, B., Voelkel-Meiman, K., and Roeder, G.S. (2006). Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in Saccharomyces cerevisiae. Genetics *174*, 1745-1754.

Tsubouchi, T., Macqueen, A.J., and Roeder, G.S. (2008). Initiation of meiotic chromosome synapsis at centromeres in budding yeast. Genes Dev *22*, 3217-3226.

Tsubouchi, T., and Roeder, G.S. (2005). A synaptonemal complex protein promotes homology-independent centromere coupling. Science *308*, 870-873.

Appendix 1: Supplemental Materials for Chapter 4

Stacy Y. Chen[1], Tomomi Tsubouchi[2,3], Beth Rockmill[2,5], Jay S. Sandler[1], Daniel R. Richards[4], Gerben Vader[7], Andreas Hochwagen[7], G. Shirleen Roeder[2,5,6] and Jennifer C. Fung[1*]

[1]Department of Biochemistry and Biophysics

University of California, San Francisco, San Francisco, California, USA

[2]Howard Hughes Medical Institute

Yale University, New Haven, Connecticut, USA

[3]Present address:

Marie Curie Research Institute, DNA Recombination Group

Oxted, Surrey RH8 0TL, UK

[4]Ingenuity Systems, Inc.

Redwood City, CA 94063, USA

[5]Department of Molecular, Cellular, and Developmental Biology

Yale University, New Haven, CT 06520, USA

[6]Department of Genetics

Yale University, New Haven, Connecticut, USA

[7]Whitehead Institute for Biomedical Research

Cambridge, Massachusetts 02142

Results

Minor Alterations of Crossing Over in the Hybrid Background

Studies have shown that in homeologous strains (10-30% sequence divergence), spore viability is reduced to less than 1.0% and recombination to less than 10% of normal levels (Hunter et al., 1996), calling into question whether CO control is "normal" in strains showing any amount of divergence. However, this is not a concern for the S96/YJM789 hybrid used in this study, since spore viability is relatively high (81%, n=541). We also measured recombination and interference genetically in a hybrid between one parent (YJM789) and a lab strain (BR1919-8B (also an S288 derivative)) and compared it to the homozygote (BR1919 2n). Note that the BR1919-8B strain, rather than S96, was chosen for comparison purposes since the BR1919-8B strain was already well-characterized for both recombination and interference. Although there are differences in the CO levels in the two intervals examined (typical for strains heterozygous at multiple loci), no significant difference in interference levels were found (Figure A1-1A, Figure A1-1B).

Materials and Methods

Simulations

We assume that the inter-crossover distance distribution can be described by a gamma distribution, characterized by a shape parameter γ and a scale parameter β, according to the following probability density function.

$$f(x) = \frac{\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{x}{\beta}}}{\beta\Gamma(\gamma)}$$

The corresponding cumulate distribution function is then

$$F(x) = \int_0^x \frac{\left(\frac{t}{\beta}\right)^{\gamma-1} e^{-\frac{t}{\beta}}}{\beta\Gamma(\gamma)} dt$$

letting u = t/β, and dt = β du, we obtain

$$F(x) = \frac{1}{\beta\Gamma(\gamma)} \int_0^{x/\beta} u^{\gamma-1} e^{-u} \beta du$$

$$= \frac{1}{\Gamma(\gamma)} \int_0^{x/\beta} u^{\gamma-1} e^{-u} du$$

$$= \frac{\Gamma_{(x/\beta)}(\gamma)}{\Gamma(\gamma)}$$

where the numerator is a partial gamma function evaluated for x/β.

The hazard function is defined as the function giving the probability that, given an event (in this case, a CO) at position 0, a second event will occur at position x. The hazard function is equal to the probability density function divided by 1 minus the cumulative distribution function. Therefore,

$$h(x) = \frac{f(x)}{1 - F(x)}$$

$$= \frac{\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{x}{\beta}}}{\beta\Gamma(\gamma)} \bullet \frac{1}{1 - \frac{\Gamma_{\left(x/\beta\right)}(\gamma)}{\Gamma(\gamma)}}$$

$$= \frac{\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{x}{\beta}}}{\beta\left(\Gamma(\gamma) - \Gamma_{\left(x/\beta\right)}(\gamma)\right)}$$

Evaluation of hazard function h(x) therefore requires computation of both a complete gamma function as well as a partial gamma function. To use this in the simulation, the current array describing the probability of placing the next CO at any given position is to be multiplied by the hazard function centered at the position of the most recently placed CO. Initially, the array is set so the probability of getting a CO is uniform across the entire genome. The simulation does not take into account CO hotspots. Hotspots were not incorporated due to lack of high resolution data in which the relative strengths of hotspots are known on a genome-wide basis. As COs are added sequentially, the probabilities in the array are modified by the gamma-based interference function as described below. For a gamma distribution with gamma greater than or equal to one, the hazard function is small (near zero) and then increases smoothly until it approaches one. Multiplication by this function will essentially remove a portion of the probability density function around the recently placed CO, thus decreasing the likelihood that subsequent COs will occur near that position.

The parameters of the gamma distribution can be obtained directly from the raw data (list of inter-crossover distances) using the moment estimators (Evans M et al.,

2000). An improved moment estimator with correction for small sample size (Hwang TY and Huang PH, 2002) was tested and found to yield no difference for the data used here. We therefore used the standard moment estimators:

$$\hat{\gamma} = \left(\frac{\bar{x}}{s}\right)^2$$

$$\hat{\beta} = \frac{s^2}{\bar{x}}$$

Both the average number of COs obtained from the microarray data and its variance are used to determine the total number of COs simulated.

To calculate simulated NPD ratios, the genome was divided into equal intervals and for the frequencies of PDs, TTs and NPDs were tallied from the simulated crossover distributions. 10,000 meioses were simulated for each NPD analysis. Since the value of the NPD ratio varies as a function of the interval size, the interval size was chosen as the average of the cM size used in the published genetic determinations of NPD ratios in order to best compare genetic vs. microarray NPD ratios. To obtain average simulated NPD ratios, the NPD ratio for each interval was averaged together for all intervals.

Supplemental References

Evans M, Hastings N, and B., P. (2000). Statistical Distributions, 3rd edn (New York, Wiley).

Hunter, N., Chambers, S.R., Louis, E.J., and Borts, R.H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J *15*, 1726-1733.

Hwang TY, and Huang PH (2002). On new moment estimation of parameters of the gamma distribution using its characterization. Ann Inst Statist Math *54*, 840-847.

Malkova, A., Swanson, J., German, M., McCusker, J.H., Housworth, E.A., Stahl, F.W., and Haber, J.E. (2004). Gene conversion and crossing over along the 405-kb left arm of Saccharomyces cerevisiae chromosome VII. Genetics *168*, 49-63.

Mood, A.M. (1950). Introduction to the Theory of Statistics (New York, McGraw-Hill).

Figure Legends

Figure A1-1. Characterization of Hybrid Strain

(A) Comparison of crossing over for two intervals on chromosome III measured in map distances for the BR1919-8B diploid strain and for a diploid resulting from mating BR1919-8Ba to YJM789. Map distances are significantly different (difference in map distances were greater than twice the SE) for the *LEU2-MAT* (LM) interval ($2*SE < |LM_1-LM_2|$: $0.0518 < 0.1192$), but not for *HIS4-LEU2* (HL) ($2*SE < |HL_1-HL_2|$: $0.0336 < 0.0265$). (B) Comparison of interference (1-NPD ratio) between the same strains for the same intervals as in (A) measured by tetrad dissection (n = 1000, BR1919-8B diploid; n = 515, BR1919-8Ba x YJM789). Chi-square tests show no difference in interference between the two strains (*HIS4-LEU2*: $\chi^2=0.0014$, $P > 0.95$; *LEU2-MAT*: $\chi^2=0.0305$, $P > 0.5$).

Figure A1-2. Distribution of Inter-crossover Distances for Wild type and *zip4*

Comparison of microarray and best-fit gamma distribution for inter-CO distances for wild type with normal interference (A) and *zip4* with reduced interference (B) as

presented in Figure 4A and 4B, but with a smaller bin size of 5 kb. Microarray data is presented in gray bars; simulation data is plotted as a black line.

Figure A1-3. Analysis of Interference using the Malkova Method

CO interference is analyzed using the method presented in Malkova et al. (2004). CO distributions are simulated for 10000 meioses using the hazard functions with gamma values obtained from wild type and *zip4* microarray data. Each chromosome is divided into 50 kb intervals (~15 cM) and the number of COs per interval is determined and assigned to be a PD, TT or NPD. Each interval in turn becomes a reference interval and PD:TT:NPD ratios are calculated for each adjacent interval depending on whether a CO event (TT or NPD) or no COs (PD) had occurred. A chi-square test is then performed, in each of the adjacent intervals, to determine whether the difference between the PD:TT:NPD ratio of the two sample pools (with or without a CO in the reference interval) is significant. A high *p*-value indicates no difference between the two pools and suggests no effects of interference between the reference interval and the particular adjacent interval. A *p*-value lower than 0.05 indicates significant difference between the PD:TT:NDP ratio of the two sample pools, and suggests that interference extends from the reference interval into the particular adjacent interval. Shown here is a plot of *p*-values for a representative interval (7) on Chromosome 2, indicated by a vertical black arrow, and its adjacent intervals along the chromosome for WT (A) and *zip4* (B).

Figure A1-4. Distribution of GC Tract Lengths in Wild Type

Wild-type GC tract lengths from the microarray data plotted as a histogram with 1-kb bin size. (A) Tract length distribution of GCs unassociated with a CO (NCOs) (B) Tract length distribution of GCs associated with crossovers (GC_{CO}).

Figure A1-5. Quantification of Southern Blot Analysis of DSB Hotspots

Southern analysis of DSB hotspots near the centromeres in a *zip1 dmc1* background (as shown in Figure 6F) was quantified for wild type and the *zip1* mutant. For each hotspot, the average intensity obtained from the 5-hr and 8-hr time points was used. Hotspots YBL002W/3C, YDL004W/5C and YOL001W/2C are quantified for *CEN2*, *CEN4*, and *CEN5*, respectively. Error bars depict standard deviation.

Table A1-1. Spore Viability and Sporulation Frequency for Wild Type and Mutants

Numbers of 0, 1, 2, 3, or 4 spore viable tetrads (s. v.) from tetrad dissection were counted, and the frequency of each type is reported. Sporulation frequencies were measured in the S96/YJM789 diploid strain background at least 72 hours after transferring onto plates containing sporulation medium. % total sporulation reports the frequency of cells that have completed MI or MI and MII, as determined by DAPI (4'-6'-Diamindino-2-phenylindole) staining of nuclei. All mutants show lower overall spore viability than wild type.

Table A1-2. Parity of Marker Segregation and NCO Frequencies

The parental origin (S96 or YJM789) of each marker was determined from each of the four spores making up one tetrad. The frequency of markers exhibiting one of the five

marker segregation patterns (S96:YJM789: 0:4, 1:3, 2:2, 3:1, 4:0) was determined. As expected, most of the markers show a 2:2 segregation pattern. The frequency of NCOs, exhibiting a 3:1 or 1:3 marker segregation pattern and no associated CO, is also shown. In wild type, a greater proportion of NCOs show a 1:3 pattern of marker segregation, indicating that S96 sequences are converted to YJM789 sequences more often than YJM789 sequences are converted to S96. This proportion is maintained in *ndj1* and *sgs1*. However, less of a bias towards conversion to YJM789 is seen for the rest of the mutants.

Table A1-3. Comparison of NCO vs. GC_{co} Median Tract Lengths in Wild Type and Mutants

The median test (Mood, 1950) was used to test the hypothesis H_0 that the median tract lengths of NCOs vs. GC_{co} within a strain are the same. A grand median is determined from all the samples from both populations and tract lengths above and below it is tallied for both NCOs and GC_{co}s and compared using a 2x2 contingency table by chi-square analysis. *ndj1* was not tested because of insufficient sample size. Because the distribution of tract lengths is skewed, the median was chosen over the mean as a better measure of the central tendency. The significance level is set to 0.05. For all strains examined, the hypothesis that the GC tract lengths are the same for NCOs and GC_{co}s was rejected. Conversion tract lengths are larger in GCs associated with a CO than in NCOs.

Table A1-4. Genetic Measurements of CO Frequency and Interference

P = parental ditype, T = tetratype, NPD = nonparental ditype and cM = centiMorgan. NPD_{exp} = number of NPDs expected in the absence of interference. The NPD ratio is the

number of NPDs observed relative to the number of NPDs expected. Interference is given as 1-NPD ratio.

Table A1-5. Comparison of NCO Frequencies among Strains

A Tukey multiple comparison test with unequal sample sizes was used to test the hypothesis that the average number of NCOs are the same for all the strains with a significance level of 0.05. This test performs pair-wise testing of strain A vs. strain B using means ranked in order of magnitude. A q-statistic is calculated and compared to the critical value, q'. The hypothesis that the mean NCO frequencies are the same between strain A vs. strain B is rejected if $q > q'$.

Table A1-6. Comparison of GC_{CO} and NCO Tract Length Medians between Strains

A Tukey-type multiple comparison test of the median with equal samples was used to test the hypothesis that median GC_{CO} tract lengths are the same between the mutants to a significance level of 0.05. The same test was applied to the median NCO tract lengths. Since the test requires equal sample sizes, strains with larger sample sizes than the strain with the lowest sample size had their sample size equalized by selecting a random subpopulation of tract lengths from the larger pool. Comparisons were only made with strains with large enough sample sizes. For the median NCO tract lengths, an additional comparison was made specifically for those strains (wild type (WT), *zip2* and *zip4*) with larger samples sizes so that medians could be more accurately determined and compared. If $q > q'$, the hypothesis that the medians are same is rejected.

Table A1-7. Nonexchange Chromosomes

Number of E_0s (chromosomes (chr) without a CO) found for each of the 16 chromosomes in wild type and all mutants. Chromosomes are arranged in order of increasing size. Total number of E_0s over all strains were tallied. A trend is seen that smaller chromosomes show a higher incidence of E_0s. % E_0 is the incidence of E_0s divided by the total number of chromosomes for all the tetrads in one strain background.

Table A1-8. Analysis of the Effects of the *zip4* Outlier Tetrad

One *zip4* tetrad was found to have a much higher number of crossovers (126 COs) than remaining 33 *zip4* tetrads. To determine the effects of this outlier, we analyzed the *zip4* data set with and without the outlier. Presented here is the comparison of CO count, interference, NCO count, and chromatid interference for *zip4* data with and without the outlier tetrad.

Table A1-9. Yeast Strains

Yeast strains used in this study are listed.

Figure A1-1

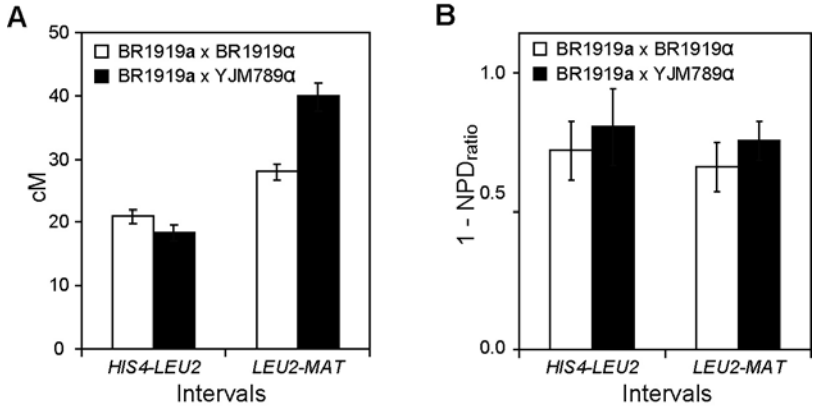

A
BR1919a x BR1919α
BR1919a x YJM789α
cM
50
40
30
20
10
0
HIS4-LEU2
LEU2-MAT
Intervals
B
BR1919a x BR1919α
BR1919a x YJM789α
1 - NPDratio
1.0
0.5
0.0
HIS4-LEU2
LEU2-MAT
Intervals

Figure A1-2

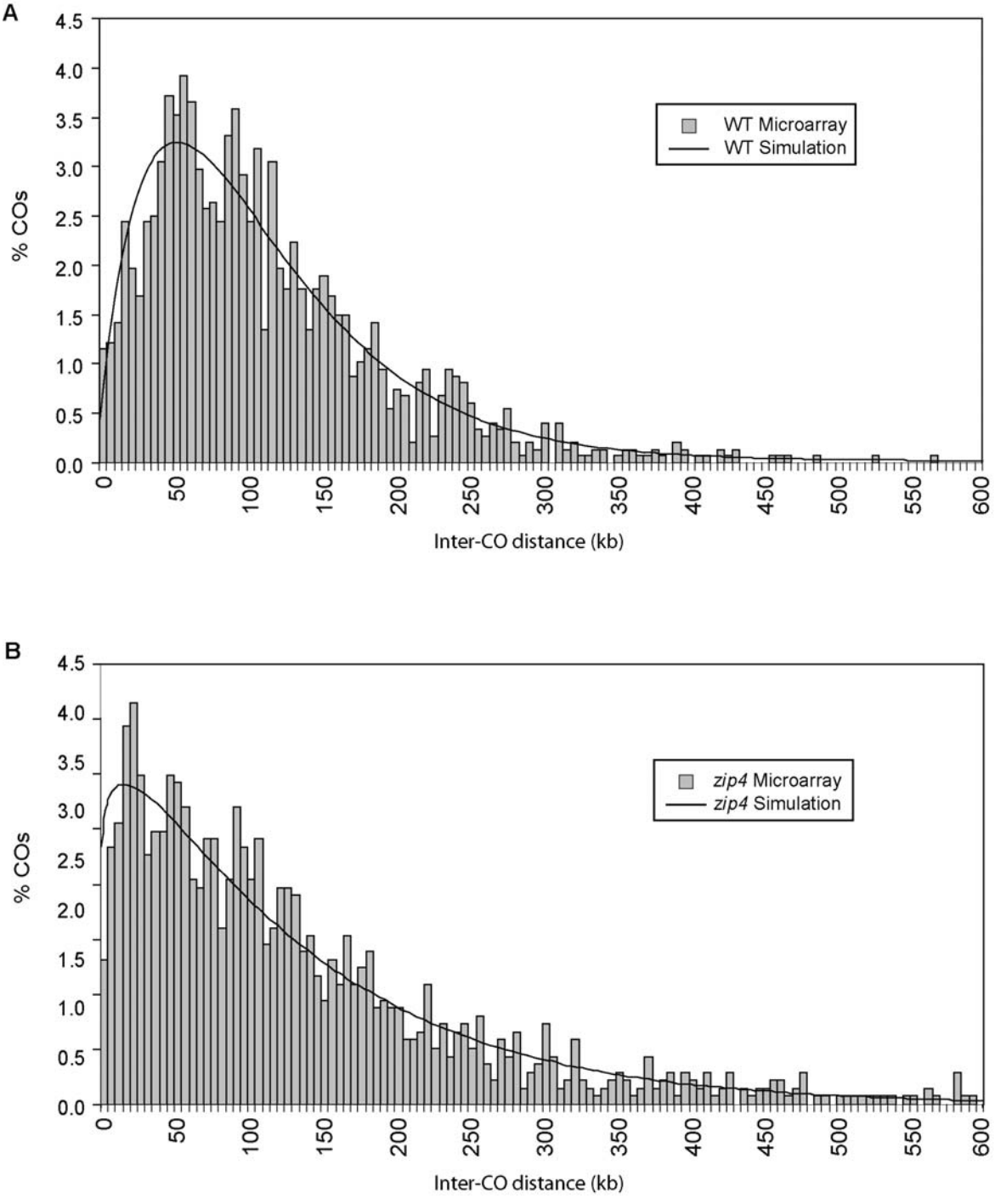

A
% COs
Inter-CO distance (kb)
WT Microarray
WT Simulation
B
% COs
Inter-CO distance (kb)
zip4 Microarray
zip4 Simulation

Figure A1-3

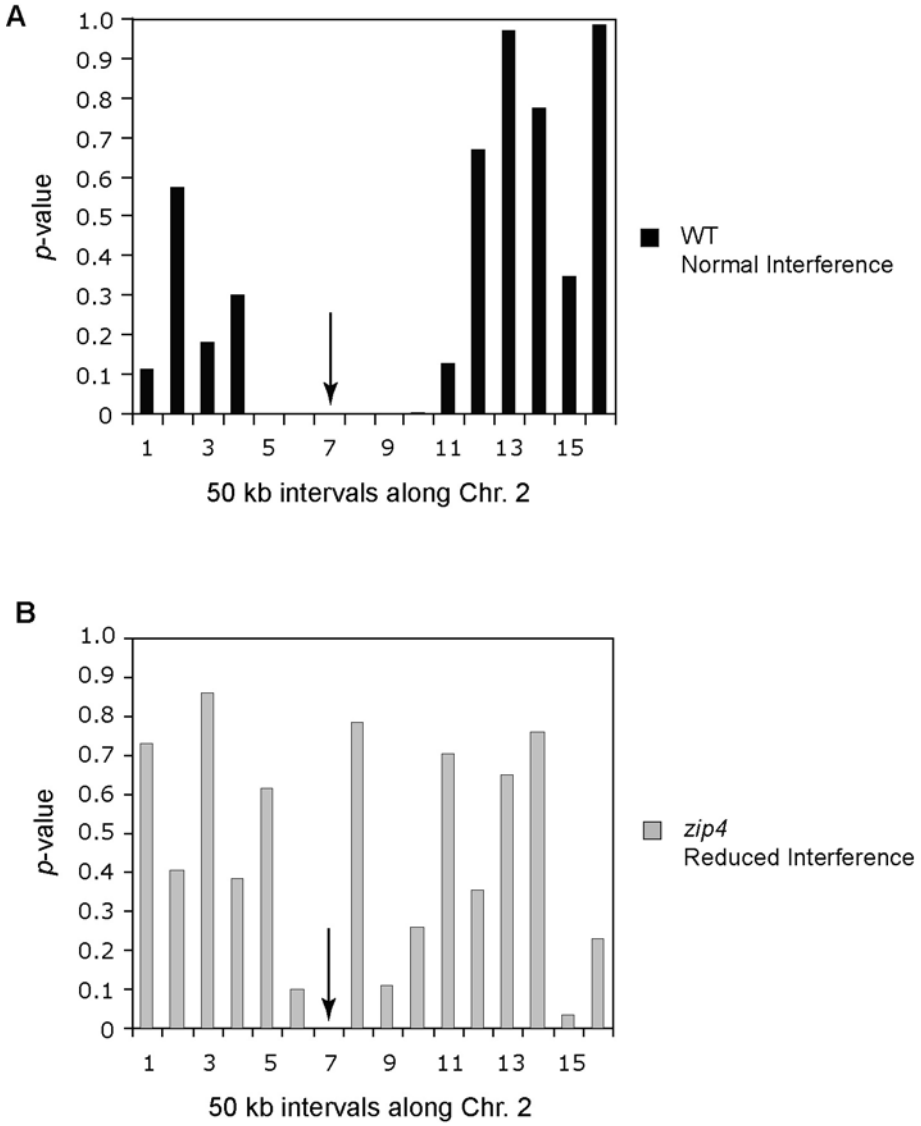

A
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p-value
1
3
5
7
9
11
13
15
50 kb intervals along Chr. 2
WT
Normal Interference
B
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p-value
1
3
5
7
9
11
13
15
50 kb intervals along Chr. 2
zip4
Reduced Interference

Figure A1-4

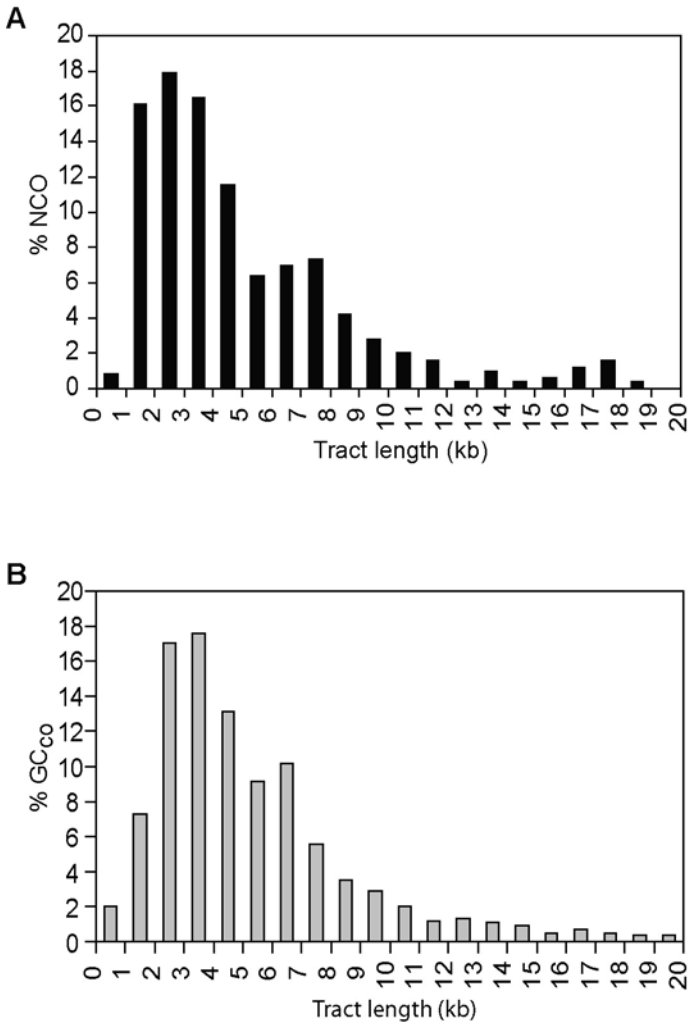


Figure A1-5

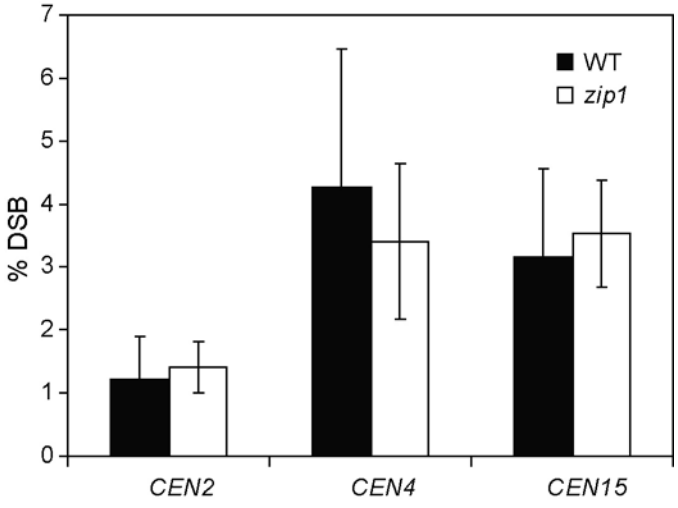


Table A1-1. Spore Viability and Sporulation Frequency for Wild Type and Mutants

	Spore Viability (%)						Sporulation Frequency (%)	
	0 s.v.	1 s.v.	2 s.v.	3 s.v.	4 s.v.	total viability	4-spore asci	total sporulation
WT	0.5	3.5	13.8	37.3	44.9	80.6	14.6	36.9
zip1	25.0	32.0	25.6	13.0	4.3	34.9	0.4	5.4
zip2	20.6	18.8	28.5	16.7	15.5	46.9	16.1	38.7
zip3	30.5	23.7	29.8	11.5	4.6	34.0	8.8	25.8
zip4	27.5	2.5	40.0	17.5	12.5	46.3	7.5	17.9
msh4	30.1	14.8	23.8	16.2	15.2	42.9	5.6	12.6
spo16	14.6	10.0	24.0	33.3	18.1	57.6	17.9	33.7
sgs1	25.4	24.9	28.5	17.1	4.2	37.4	5.2	8.4
ndj1	32.8	22.7	22.7	16.4	5.5	34.8	4.1	16.2

Table A1-2. Parity of Marker Segregation and NCO Frequencies

	S96:YJM789 Marker Segregation Frequency (%)					NCO Frequency (%)	
	0:4	1:3	2:2	3:1	4:0	S96	YJM789
WT	0.0	1.1	98.0	0.9	0.0	37.2	62.8
zip1	0.1	3.5	92.7	3.6	0.1	47.7	52.3
zip2	0.5	2.3	95.7	1.4	0.0	45.6	54.4
zip3	0.1	1.7	96.5	1.7	0.1	51.0	49.0
zip4	0.1	1.8	96.6	1.6	0.0	43.0	57.0
msh4	0.4	0.8	98.0	0.7	0.0	46.6	53.4
spo16	0.7	1.8	96.0	1.4	0.0	42.1	57.9
ndj1	0.0	1.5	96.6	1.6	0.3	38.4	61.6
sgs1	0.7	2.2	94.8	1.8	0.5	35.8	64.3

Table A1-3. Comparison of NCO vs. GC_{co} Median Tract Lengths in Wild Type and Mutants

	WT	*zip1*	*zip2*	*zip3*	*zip4*	*msh4*	*spo16*	*sgs1*
Median NCO (kb)	3.9	5.1	4.7	4.6	5.0	4.2	5.3	4.1
Median GCco (kb)	4.4	6.3	5.8	5.8	5.7	5.8	6.4	4.8
χ^2	7.0	19.8	18.0	9.4	13.8	13.9	10.3	4.4
p-value	0.008	0.000	0.000	0.002	0.000	0.000	0.001	0.036
Conclusion	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

Table A1-4. Genetic Measurements of CO Frequency and Interference

Strains	Back-ground	Interval	P	T	NPD	NPD_{exp}	NPD ratio	Interfer ence	cM
WT	BR1919 2n	*HIS4-LEU2*	615	378	7	25	0.28	0.72	21
WT	BR1919 2n	*LEU2-MAT*	519	475	15	44	0.34	0.66	28
zip2	BR1919 2n	*HIS4-LEU2*	2962	552	10	12	0.82	0.18	9
zip2	BR1919 2n	*LEU2-MAT*	2770	820	18	28	0.64	0.36	13
zip3	BR1919 2n	*HIS4-LEU2*	864	264	2	9	0.22*	0.78	12
zip3	BR1919 2n	*LEU2-MAT*	705	442	31	29	1.08	-0.08	27
spo16	BR1919 2n	*HIS4-LEU2*	1516	294	9	7	1.33	-0.33	10
spo16	BR1919 2n	*LEU2-MAT*	1431	423	5	14	0.35	0.65	12

Table A1-5. Comparison of NCO Frequencies among Strains

Strain A vs. Strain B	NCO Mean of Strain A	NCO Mean of Strain B	q	$q'_{0.05,\infty,8}$	Conclusion (H_0: Mean A = Mean B)
zip1 vs. *msh4*	71	17	25.306	4.286	reject H_0
zip1 vs. WT	71	19	28.549	4.286	reject H_0
zip1 vs. *ndj1*	71	22	20.169	4.286	reject H_0
zip1 vs. *zip2*	71	30	22.620	4.286	reject H_0
zip1 vs. *spo16*	71	34	16.617	4.286	reject H_0
zip1 vs. *zip4*	71	35	20.877	4.286	reject H_0
zip1 vs. *zip3*	71	37	5.661	4.286	reject H_0
zip1. vs. *sgs1*	71	36	15.923	4.286	reject H_0
sgs1 vs. *msh4*	36	17	9.891	4.286	reject H_0
sgs1 vs. WT	36	19	10.712	4.286	reject H_0
sgs1 vs. *ndj1*	36	22	6.220	4.286	reject H_0
sgs1 vs. *zip2*	36	30	4.424	4.286	reject H_0
sgs1 vs. *spo16*	36	34	1.975	4.286	accept H_0
sgs1 vs. *zip4*	36	35	1.929	4.286	accept H_0
sgs1 vs. *zip3*	36	37	0.333	4.286	accept H_0
zip3 vs. *msh4*	37	17	8.591	4.286	reject H_0
zip3 vs. WT	37	19	8.958	4.286	reject H_0
zip3 vs. *ndj1*	37	22	5.375	4.286	reject H_0
zip3 vs. *zip2*	37	30	3.378	4.286	accept H_0
zip3 vs. *spo16*	37	34	1.384	4.286	accept H_0
zip3 vs. *zip4*	37	35	1.129	4.286	accept H_0
zip4 vs. *msh4*	35	17	10.229	4.286	reject H_0
zip4 vs. WT	35	19	12.268	4.286	reject H_0
zip4 vs. *ndj1*	35	22	5.634	4.286	reject H_0
zip4 vs. *zip2*	35	30	3.540	4.286	accept H_0
zip4 vs. *spo16*	35	34	0.632	4.286	accept H_0
spo16 vs. *msh4*	34	17	7.101	4.286	reject H_0
spo16 vs. WT	34	19	7.239	4.286	reject H_0
spo16 vs. *ndj1*	34	22	4.038	4.286	accept H_0
spo16 vs. *zip2*	34	30	1.714	4.286	accept H_0
zip2 vs. *msh4*	30	17	7.301	4.286	reject H_0
zip2 vs. WT	30	19	8.153	4.286	reject H_0
zip2 vs. *ndj1*	30	22	3.326	4.286	accept H_0
ndj1 vs. *msh4*	22	17	2.502	4.286	accept H_0
ndj1 vs. WT	22	19	1.943	4.286	accept H_0
WT vs. *msh4*	19	17	1.078	4.286	accept H_0

Table A1-6. Comparison of GC_{CO} and NCO Tract Length Medians between Strains

Strain A vs. Strain B	n	Median GCco Tract Length of Strain A (kb)	Median GCco Tract Length of Strain B (kb)	q	$q'_{0.05,5,\infty}$	Conclusion (H_0: Median A = Median B)
zip1 vs. *zip2*	434	6.4	5.7	1.63	3.86	accept H_0
zip1 vs. *zip4*	434	6.4	5.6	2.21	3.86	accept H_0
zip1 vs. *sgs1*	434	6.4	4.8	6.43	3.86	reject H_0
zip1 vs. WT	434	6.4	4.4	6.91	3.86	reject H_0
zip2 vs. *zip4*	434	5.7	5.6	0.58	3.86	accept H_0
zip2 vs. *sgs1*	434	5.7	4.8	4.80	3.86	reject H_0
zip2 vs. WT	434	5.7	4.4	5.28	3.86	reject H_0
zip4 vs. *sgs1*	434	5.6	4.8	4.22	3.86	reject H_0
zip4 vs. WT	434	5.6	4.4	4.70	3.86	reject H_0
sgs1 vs. WT	434	4.8	4.4	0.48	3.86	accept H_0
Strain A vs. Strain B	n	Median NCO Tract Length of Strain A (kb)	Median NCO Tract Length of Strain B (kb)	q	$q'_{0.05,5,\infty}$	Conclusion (H_0: Median A = Median B)
zip1 vs. *zip2*	400	5.1	4.7	2.30	3.86	accept H_0
zip1 vs. *zip4*	400	5.1	4.4	3.17	3.86	accept H_0
zip1 vs. *sgs1*	400	5.1	4.0	4.99	3.86	reject H_0
zip1 vs. WT	400	5.1	4.0	5.37	3.86	reject H_0
zip2 vs. *zip4*	400	4.7	4.4	0.96	3.86	accept H_0
zip2 vs. *sgs1*	400	4.7	4.0	2.78	3.86	accept H_0
zip2 vs. WT	400	4.7	4.0	3.17	3.86	accept H_0
zip4 vs. *sgs1*	400	4.4	4.0	1.82	3.86	accept H_0
zip4 vs. WT	400	4.4	4.0	2.21	3.86	accept H_0
sgs1 vs. WT	400	4.0	4.0	0.38	3.86	accept H_0
zip2 vs. *zip4*	503	4.7	4.7	0.45	3.31	accept H_0
zip2 vs. WT	503	4.7	3.9	4.64	3.31	reject H_0
zip4 vs. WT	503	4.7	3.9	5.97	3.31	reject H_0

Table A1-7. Nonexchange Chromosomes

Chr #	Chr size (kb)	WT	*zip1*	*zip2*	*zip3*	*zip4*	*msh4*	*spo16*	*ndj1*	*sgs1*	sum
1	230	0	3	12	1	8	2	1	1	1	31
6	270	0	0	5	3	4	1	0	1	0	16
3	317	0	0	5	0	2	1	1	1	0	10
9	440	0	0	3	1	1	2	0	0	0	8
8	563	0	0	2	1	2	0	0	0	0	5
5	577	0	0	2	0	2	1	0	0	0	5
11	666	0	0	0	1	2	1	0	0	0	5
10	746	0	0	0	0	1	0	1	0	0	2
14	784	0	0	1	1	3	1	0	1	0	7
2	813	0	0	0	0	3	0	0	0	0	3
13	924	0	0	0	0	0	0	1	0	0	1
16	948	0	0	1	0	1	0	0	0	0	2
12	1078	0	0	5	0	0	0	0	0	0	5
7	1091	0	0	0	0	1	0	0	0	0	1
15	1091	0	0	0	0	0	1	0	0	0	1
4	1532	0	0	0	0	0	0	0	0	0	0
# Te-trads		26	9	26	8	34	11	8	7	7	
% E_0		0.0	2.1	8.7	6.3	5.5	5.7	3.9	6.3	2.7	

Table A1-8. Analysis of the Effects of the *zip4* Outlier Tetrad

		CO Count		Inter-ference	NCO Count		Chromatid Interference	
	# Tetrad	Mean	SD	NPD ratio	Mean	SD	2-s.d : 3-s.d. : 4-s.d. Ratio	*p*-value
zip4 (all data)	34	56.7	17	0.96	36.7	16.3	1.0 : 2.0 : 1.1	0.45
zip4 (no outlier)	33	54.5	12	0.90	34.7	11.8	1:0 : 1.9 : 1.1	0.43

Table A1-9. Yeast Strains

Strain	Genotype
S96	*MAT***a** *ho lys5*
YJM789	*MATα ho::hisG lys2 cyh*
JCF1850	S96 but *zip1::kanMX6*
JCF1852	YJM789 but *zip1::kanMX6*
JCF1104	S96 but *zip2::kanMX6*
JCF1106	YJM789 but *zip2::kanMX6*
SYC1104	S96 but *zip3::kanMX6*
SYC1105	YJM789 but *zip3::kanMX6*
TY461	S96 but *zip4::kanMX6*
TY462	YJM789 but *zip4::kanMX6*
SYC1110	S96 but *msh4::kanMX6*
SYC1111	YJM789 but *msh4::kanMX6*
JCF2035	S96 but *spo16::kanMX6*
JCF1210	YJM789 but *spo16::kanMX6*
SYC1120	S96 but *sgs1::kanMX6*
SYC1121	YJM789 but *sgs1::kanMX6*
SYC1112	S96 but *ndj1::kanMX6*
SYC1113	YJM789 but *ndj1::kanMX6*
BR4633	*leu2::CUP1,arg4-8 iTHR1 iura3-1* *MATα iADE2 leu2::CUP1,arg4-8* *iura3-stu iNAT iLEU2 MAT***a** *trp1-289 ade2-1 ura3-1* *trp1-289 ade2-1 ura3-1*
BR4790	BR4633 but *zip1::kanMX6*
BR4829	BR4633 but *zip2::kanMX6*
S2937	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1*
JCF406	BR1919-8B *MAT***a** *ura3-1 trp1-289 thr1-4 ade2-1*
S1561	BR1919-8B *MATα leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1*
JS03	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1 zip2::URA3* *MATα ura3-1 trp1-289 thr1-4 ade2-1 zip2::URA3*
BR3643	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1 zip3::URA3* *MATα ura3-1 trp1-289 thr1-4 ade2-1 zip3::URA3*
JS36	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1 spo16::kanMX6*

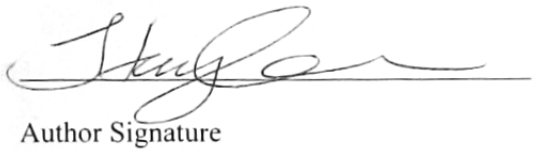
	MATα ura3-1 trp1-289 thr1-4 ade2-1 spo16::kanMX6
NKY1455	SK1 *MAT***a** *leu2::hisG his4X-LEU2-URA3 ura3 ho::LYS2 lys2 arg4-nsp dmc1::ARG4*
	MATα leu2::hisG his4B-LEU2 ura3 ho::LYS2 lys2 arg4-bgl dmc1::ARG4
YAH2650	Same as NKY1455, but *zip1::LYS2/zip1::LYS2*

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

Jan 5th, 2010

Date