Meiotic recombination and its implications for plant breeding

T.G. (Erik) Wijnker

Thesis committee

Promotors
Prof. dr. J.H.S.G.M. de Jong
Personal chair at the Laboratory of Genetics, Wageningen University

Prof. dr. ir. M. Koornneef
Personal chair at the Laboratory of Genetics, Wageningen University and
Director at the Max Planck Institute for Plant Breeding Research (MPIZ) Cologne, Germany

Co-promotor
Dr. R.H.G. Dirks
Research manager at the Rijk Zwaan Breeding Company, Fijnaart.

Other members
Prof. dr. ir. E. Jacobsen, Wageningen University, Wageningen
Prof. dr. H. Puchta, University of Karlsruhe, Karlsruhe, Germany
Dr. A. Schnittger, University of Strasbourg, Strasbourg, France
Prof. dr. M.E. Schranz, Wageningen University, Wageningen

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences (EPS)

Meiotic recombination and its implications for plant breeding

T.G. (Erik) Wijnker

Thesis
Submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropf,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 6 February 2013
at 4 p.m. in the Aula.

T.G. (Erik) Wijnker
Meiotic Recombination and its implications for plant breeding

Thesis, Wageningen University, Wageningen, NL (2013)
With references, with summaries in Dutch and English

ISBN 978-94-6173-440-2

Contents

Genesis

Het was de zesde dag. Adam stond klaar.
Hij zag de eiken met hun volle greep
in het niets. Macht is een kwestie van vertakkingen.
hij had de bergen gezien, opbergruimtes van
alleen maar zichzelf, hoge leegstaande kelders.
En herten. Met poten zo dun als stethoscopen
stonden ze te luisteren aan de borst van de aarde,
en zodra ze iets hoorden, liepen ze weg,
de uitvinding van het pizzicato met zich meenemend,
verten in. Herten.
En hij had de zee gezien, het laden en het lossen van drukte,
waar je rustig van werd. En de lege, hetzerige gebaren
van de wind, van kom mee, kom mee, en niemand volgde.
En diepte, afgronden waar je moeilijk van werd. En zwijgen,
want dat deed het allemaal, en te groot zijn.
En toen zei God: en nu jij. Nee, zei Adam.

H. de Coninck
(In: 'De hectaren van het geheugen', 1985)

CHAPTER 1

General introduction

Meiosis generates variation

Life evolves by adapting continuously to its ever-changing environment. Sexual reproduction evolved in complex organisms (Eukaryotes) to facilitate this adaptation by actively changing allele combinations on chromosomes. A greater part of this variability is generated in the highly specialised process of meiosis. The unique combination of homologous pairing, crossover formation and balanced segregation of recombinant chromosomes creates novel alleles and novel allele combinations on chromosomes. As a major driving force of all this genetic variation, meiosis can be placed at the very heart of evolution, domestication, and also of plant breeding. The key to the creation of new crop varieties lies in the systematic exploration of genetic variation and the selection of new phenotypes. While meiotic recombination provides plant breeders with countless numbers of allele combinations, the unpredictability of such variation leads to time consuming breeding practices (Dirks *et al.* 2009; Wijnker and de Jong 2008). Improving breeding thus requires improving both the processes of variation generation and discovery as well as methods of selection. Only if we are able to understand the mechanisms governing meiotic recombination, we will be able to control meiotic recombination, and manage it for breeding purposes. I will therefore raise questions of how plants generate variation during meiosis, determine what patterns underlie this variation and to what extent this can help improving the management of variation for plant breeding.

As explained above, meiosis plays a pivotal role in creating novel genetic variants/genotypes by recombining variation that before existed in different genotypes. It does so in two consecutive divisions and along few entirely different processes. It generates new allele and chromosome combinations, while reducing the somatic chromosome number by half during gamete production (Gerton and Hawley 2005). This is achieved in a controlled sequence of unique events starting with an S-phase in which chromosomes are duplicated, generating two sister-chromatids that remain joined together at their centromeres until the metaphase of the second division. Following duplication, plants produce about 100-500 double strand breaks (DSBs) along the whole length of their chromosomal DNA (Pawlowski *et al.* 2003; Sanchez-Moran *et al.* 2007), a process that initiates homology search and recombination. At the same time, chromosomes progressively condense while assembling proteinaceous lateral axes that subsequently serve as a base for homologous chromosomes to join together along their entire lengths. This joint tripartite proteineous structure is known as the synaptonemal complex (SC) (Moses 1968; Westergaard and von Wettstein 1972).

While chromosomes begin to synapse, DSBs are processed and repaired through homologous recombination (HR). This is a specific DNA repair pathway in which DNA repair proteins use homologous DNA sequences as template for DSB repair (Puchta 2005). While mostly leading to genuine repair, HR in meiosis also generates true reciprocal exchanges between non-sister chromatids. These so-called meiotic crossovers originate when a chromatid is repaired by joining one end of it to the other end of a homologous

(non-)sister chromatid, thus generating a recombinant chromatid. Recombinant chromosomes (i.e., the chromatids) thus consist of new allele combinations of both homologues. While the paired homologues disjoin, their SC disassembles and homologues remain together only at sites of their crossovers, which are now microscopically visible as chiasmata in cell complements at late meiotic prophase I (diplotene to metaphase I). Here we see evidently that homologues are joined by at least one chiasma, which is essential for proper orientation of the bivalents on the metaphase I divisional plane. Shortly later, at anaphase I, the half-bivalents (homologues) segregate equally to their poles, and then undergo a second division in which sister centromeres disjoin, similar to mitosis. The tetrad stage marks the end of meiosis, when the meiocyte contains four spores that each carry half the chromosome number of the parent. These spores can then further develop into gametes (like pollen and egg cells) that can then fuse with one another, thus re-establishing the somatic chromosome number.

Crossover recombination ensures that the dual roles of meiosis are properly executed:

- crossovers form chiasmata that are important for joining homologous chromosome together, enabling them to segregate equally to opposite poles at the first meiotic division and reduce the chromosome number of the cell to half (the reductional division).
- crossovers generate new variation directly by creating new allele combinations on chromosomes, while facilitating random chromosome assortment.

Meiosis generates novel genetic combinations in a highly regulated way, but interestingly, the precise outcome of the genetic content of their daughter cells is always different by the reshuffling of alleles during meiosis. The total genetic blending by recombination and independent chromosome assortment, depends on the number and position of the crossover sites (intrachromosomal recombination) and the number of chromosomes (interchromosomal recombination). While the chromosome number is normally fixed, crossover incidence varies between bivalents, chromatids, cells, sexes, individuals and species (Baudat *et al.* 2010; Lenormand and Dutheil 2005). As a result, no meiotic products are identical.

Meiosis and breeding

While the generation of random variation is pivotal to breeders, it also poses them with tough challenges because meiotic recombination, which is essential for generating favourite combination of valuable traits, also produces allele combinations that are undesired and hence useless in the breeding program. Plant breeders found a way of dealing with this by preserving specific allele combinations in (near) homozygotes such as inbred families and doubled haploids (Forster *et al.* 2007). Although the homologues of inbreds still form crossovers at meiosis, the identical homologues ensure that no new allele combinations will be generated. Plant breeders produce and hold vast collections of homozygous (often -inbred lines,) that are used in controlled crosses to generate het-

erozygous F1 hybrids. The varieties of most crops today, are F1's, which are tested on their performance in different environments. Breeders need to switch between steps that require recombination (for the generation of new allele and chromosome combinations) and steps in which the new allele and chromosome combinations are fixed for later use, i.e., the generation of homozygous breeding material.

One of the classical breeding strategies for broadening the genetic base of crops is introgressive hybridisation. Desirable traits from a wild relative or donor species (specific accessions) are transferred to the recipient crop by intra- or interspecific hybridisation followed by backcrossing and consecutive selections. The introgression of a specific locus e.g. conferring disease resistance, drought resistance etc. is often far from straightforward. Chromosomes contain thousands of linearly arranged loci, and introgression of precisely one locus requires the occurrence of crossovers on either side of that locus, after which repeated backcrosses are required for eliminating all other unwanted alleles at other loci. When the homoeologous donor and recipient chromosome regions differ in collinearity (e.g. by inversions or translocations or repeat content), the chance of a crossover very close to the region of interest is low, if not zero. This problem, known as linkage drag, is one of the major problems in breeding programs aiming at replacement of desired wild genes to the crop species.

The number and positions of crossovers on the chromosome pairs are subjected to processes, which regulate or confine their formation. First of all, the number of crossovers is highly restricted. In all cases, bivalents, even for the smallest chromosomes have always at least one crossover, which is needed for balanced disjoin of the half bivalents at anaphase I. Without that mechanism, achiasmatic chromosomes (univalents) will segregate randomly, giving rise to spores that contain unbalanced, variable chromosome numbers, many of which are not viable or form aneuploid offspring. Meiosis with large numbers of crossover events are uncommon; in many cases not more than 1-2 crossovers per chromosome arm are observed. In *Arabidopsis*, for example, female meiosis experiences hardly more than one crossover per homologue pair, whereas male meiosis has about two (Giraut *et al.* 2011). In addition, a bivalent with more chiasmata per chromosome arm, often tends to keep these crossover events well separated by a mechanism known as genetic (crossover- or chiasma-) interference. Most crossovers are so called "interference dependent", i.e belonging to the subset of class I crossovers that comprises the far majority in *Arabidopsis* (Mercier *et al.* 2005) (Higgins *et al.* 2004), tomato (Lhuissier *et al.* 2007) maize (Falque *et al.* 2009) as well as in other model organisms like mouse (de Boer *et al.* 2006) and budding yeast (de los Santos *et al.* 2003). The second class of crossovers are are formed through a different pathway and are not subjected to crossover interference. Only a small proportion (i.e. 10-15%) of crossovers in the species mentioned above originate from the second pathway, however, but some organisms like fission yeast that lack the class I crossover pathway all repair their DSBs through the interference insensitive pathway (Osman *et al.* 2003).

Another mechanism that shapes the recombination landscape is the way in which single copy and repetitive sequences are organised along the chromosome. In most eukaryotic species satellite and tandem repeats as well as "gypsy" and other types of long terminal repeat (LTR-) retrotransposons are most prevalent in heterochromatin areas around the centromeres (pericentromeres) and at distal chromosome ends, but are rare in euchromatin (Gaut *et al.* 2007; Stack 1984). This differentiation of repeats in heterochromatin and single copy sequences in euchromatin has a clear effect on where crossovers reside. To some extent this can be intuitively understood, since sequences for which on the homologue no matching sequence exist, will not be able to repair the break from a non-sister template (Goldfarb and Lichten 2010).
There is much more to this story that we currently do not understand well. In many species, crossovers have a strong preference for euchromatic regions, whereas heterochromatin (like pericentromere domains in larger genomes like tomato) have a much lower likeliness of containing crossovers (Sherman and Stack 1995). In the model plant *Arabidopsis*, crossover frequencies are in fact somewhat enhanced close to the centromere (Giraut *et al.* 2011). Heterochromatic regions being more highly methylated, led researchers to question what the effect of DNA de-methylation is on crossover frequencies. Recombination was reported to increase in euchromatin but remains equal in heterochromatin or even decreases (Melamed-Bessudo and Levy 2012; Mirouze *et al.* 2012; Yelina *et al.* 2012), whereas very centromere proximal crossovers might increase (Yelina *et al.* 2012). A more straightforward link between recombination and chromatin modifications is H3K4 methylation of histones, which mouse and human associate with crossover and recombination hotspot sites (Baudat *et al.* 2010; Borde *et al.* 2009; Grey *et al.* 2011). The same was reported for yeast (Borde *et al.* 2009), but might not hold true (Tischfield *et al.* 2012).

Understanding and controlling meiotic variation

Systematic studies of inheritance have a long history, tracing back to the pea experiments by Gregor Mendel in 1865 (Mendel 1866 (for 1865)) and the discovery of genetic linkage in 1915 (Sturtevant 1915). Because recombination is a stochastic process, the precise assessment of crossover frequencies was never easy, since it requires the study of relatively large populations of individuals or meiotic cells. The comparison of mapping populations (like populations derived through male and female meiosis) can point to sex differences in recombination frequencies (Vizir and Korol 1990). Alternatively, the cytological structures of crossover sites (e.g., late recombination nodules in the electron microscope, or immune detection of MLH1) can be counted directly on chromosomes in developing meiocytes (Chelysheva *et al.* 2010). In the last decade a variety of new tools have been developed and can help the study of meiotic recombination and facilitate crossover (and CO-interference-) assessment at much faster rates. These are replacing classical morphological and DNA markers and multilocus platforms such as AFLPs with generic genotyping systems (like BeadArray or KASPar assays) that can be used to ef-

ficiently genotype large offspring numbers. This allows for the subsequent construction and study of recombination landscapes (Giraut *et al.* 2011; Mirouze *et al.* 2012). The discovery of the *Arabidopsis* QUARTET mutant was a leap forward, and allowed mapping in relation to centromere positions, and the analysis of the products of single meiosis events (Copenhaver *et al.* 1999) This approach was later improved by expressing fluorescent proteins under the control of pollen specific promoters. Pollen tetrads now allow the quantifications of crossovers and interference (Berchowitz and Copenhaver 2008). Similar fluorescent transgene-based marker systems were developed for seeds (Melamed-Bessudo *et al.* 2005) to measure recombination frequencies in large populations at the seed level. The study of meiotic recombination has certainly been more intense in recent years, but is still far from easy and quick. The use of pollen tetrads requires that all studies have to be performed in a quartet background. Fluorescent markers must be introduced to desired backgrounds, and even the use of high throughput genotyping platforms can become rather costly when detailed recombination studies are scaled up to the level of large sized populations.

The interest of breeders in meiotic recombination lies in the panoply of applications. One of the most compelling ones, is to increase the recombination frequency which would speed up breeding schemes that now require very large populations to obtain sufficient recombination events. The discovery of mutants that show increased recombination has taken a long time, which is easily understood when one considers the difficulty of performing screens for such processes. Methods for inducing significant increases in recombination frequencies are not established (Wijnker and de Jong 2008), but the recent discovery of that the mutation of the crossover suppressor FANCM leads to threefold increases in recombination frequencies in *Arabidopsis* is very promising (Crismani *et al.* 2012). FANCM is an apparent suppressor of class II crossover formation, and its mutation elevates recombination through this interference-independent pathway. The temporal suppression of such genes will speed up some breeding programs significantly. Another recent discovery that merits attention is the action of methyltransferase PRDM9, a protein with zinc-finger domain in mouse and human meiosis, which directs meiotic recombinases to specific sequences (Baudat *et al.* 2010). This methyltransferase can bind to specific DNA sequences during meiosis, and directly (through binding of recombinases to PRDM9) or indirectly (through H3K4 methylation) defines about 40% of crossover hotspots (Baudat *et al.* 2010). Although we cannot currently assess whether such a protein is easily transferable to plants, it at least suggests that we might develop similar methods to direct recombination to sites of interest in the future (Bogdanove *et al.* 2010).

Modifying the recombination frequencies or site-directed crossover formation provide the tools for the steering of genetics. An eminent question is then what processes govern pairing between sequences of the homologues or homoeologues, even if they differ to some extent in their DNA sequences? Breeders frequently try to enrich their breeding material with alien introgressions from related species. Introgressed segments are

less likely to engage in crossover formation, for which the reasons are not all too clear. Recent studies showed that SNP polymorphisms do not limit crossover formation per se (Salome *et al.* 2012). Extensive research in allohexaploid wheat, shows that the homoeologues from the A, B and D genome in wild type wheat do not form crossovers, unless the *Ph*1 locus is deleted or mutated. *Ph1* (Preferential pairing of Homologues) is a presumed epi-allele of a cyclin dependent kinase (CDK), a cell cycle regulator that mediates chromatin remodelling during the meiotic prophase. Under wild type conditions, CDK activity is constitutively low, leading to differential condensation of chromatin between homoeologues. It was shown that this is accompanied by much more similar condensation patterns in the subtelomere heterochromatin (Colas *et al.* 2008). When *Ph1* is functionally absent, CDK activity increases (because expression of the gene is regulated from one of the homoeologous chromosomes), and crossovers are formed between homoeologous chromosomes (Greer *et al.* 2012). A recent study showed that *ph1* mutant like effects can be induced in wheat by the artificial upregulation of kinase activity through application of the phosphatase inhibitor okadaic acid (Knight *et al.* 2010).

The most dramatic changes affecting HR in meiosis are found in inversion heterozygotes. Such segments completely abolish meiotic recombination between the non-homologous regions (Wijnker and de Jong 2008). The study of chromosome collinearity (and genome evolution) that shows such inversions, has a long history in cytogenetic research. The recent use of bacterial artificial chromosomes (BACs) as fluorescent probes in FISH (Fluorescent in-situ hybridisation) studies, has recently shown its use in plants (Lou *et al.* 2010; Peters *et al.* 2012; Szinay *et al.* 2012). BACs of known genomic position in one species can be hybridized to nuclei of related species to study large-scale chromosome rearrangements, which may be a major cause of linkage drag and thereby can many be a major obstacle for the introgression of single loci of interest from a donor- into a recipient background. Although there are currently no known methods for making inversions undone, simply learning about the presence of inversions is of great help to help breeders understand why some crosses will not lead to expected outcomes.

The rapidly advancing methods of next generation sequencing and micro-arrays have recently propelled our understanding of meiosis even further. The studies allow the precise dissection of whole meiotic tetrads for the presence of crossovers and gene-conversions in fission yeast (Lu *et al.* 2012; Mancera *et al.* 2008; Qi *et al.* 2009). Gene-conversions are recognized as footprints of so-called non-crossovers (NCOs): non-reciprocal nucleotide exchanges between homologues as a result of HR events that did not lead to crossovers. Such studies give unparalleled insights into the precise outcomes of meiotic recombination. It was shown that NCOs contribute to about 1% of new variation in yeast (Mancera *et al.* 2008) whereas in plants NCOs contribute only a fraction of that. High density genotyping can now even help making personalised recombination maps as was shown for a human male (Wang *et al.* 2012).

While crossovers and random assortment of chromosomes at meiotic prophase I are beneficial for the generation of new variation, they will also generate such large variation that the genome of a heterozygote becomes completely scrambled in its gametes. Prevention of this scrambling is not possible, and breeders have but one way to preserve valuable allele combinations, which is the preservation of these in homozygous lines. A well-known method for achieving homozygosity is using a series of recurrent backcrosses in which a locus of segment of choice is introgressed into a known recipient genotype. Repeated cycles of selfing will also lead to homozygosity but requires many generations. The quickest method surely is the use of doubled haploids in which the genotype of a gamete is immediately fixed by growing the gamete directly into a homozygous diploid plant. A strategy holding much promise for the development of efficient haploid induction in plants was the discovery that targeted alterations to the CENTROMERE HISTONE 3 can lead to the production of haploid plants. *Arabidopsis* with engineered CENH3 "tail swaps" (i.e., carrying the protein tail of regular histones) can be used as inducer lines to generate haploids after crossing with a donor plant (Ravi and Chan 2010). Alternative ways have previously been proposed for the direct fixation of genotypes, among which apomixis has received most attention in literature (van Dijk 2008). Apomictic plants reproduce clonally through seeds, requiring the formation of a gamete that is genetically identical to the mother plant. Recently it was shown that meiosis can be modified to generates such gametes, by mutating single or multiple genes simultaneously to turn meiosis into a mitosis (d'Erfurth *et al.* 2009; Olmedo-Monfil *et al.* 2010; Ravi *et al.* 2008). As expected, the combination of such mutants with the haploid inducing potential of CENH3 tailswap inducers could be used to achieve a synthetic form of apomixis, in which these diploid gametes were grown into clonal offspring (Marimuthu *et al.* 2011). The fixation of heterozygous genomes is a great tool for turning segregating heterozygotes directly into clonally propagating lines, however does imply the intrinsic problem that thereafter any cross with such a plant would again imply the segregation of traits which makes improvement of the variety impossible.

Alternative strategies have been proposed for the fixation of complex genotypes, the so-called "reverse breeding" strategies. Reverse breeding strategies aim at doing the exact opposite of traditional breeding: instead of crossing two lines to generate a heterozygote, these strategies attempt to create homozygous parents from a starting heterozygote. The best known example by now is surely reverse breeding in which crossover recombination is suppressed in a heterozygote of choice. This leads to the (binomial) distribution of non-recombinant chromosomes into gametes. Since crossovers are absent, most of the gametes of such a plant are unviable due to aneuploidy, but balanced gametes are formed by chance at low frequency, and these can be grown into (doubled) haploids or used in pollination experiments. In such a DH population, all plants carry precisely half of the (unscrambled) chromosomes of the chosen heterozygote. By simply selecting among these plants, a combination of DHs can be identified that together per-

fectly reconstitute the genome of the starting heterozygote (Dirks *et al.* 2009; Wijnker *et al.* 2012) . An alternative approach has been suggested, in which diploid gametes are used that arise from a meiosis that omitted the second meiotic division. Such gametes are diploid, but – different from mitotic divisions - are homozygous for their centromeres and genes linked to the centromeres. When grown directly into plants, such plants would fixate directly the centromere parts of the starting heterozygote, and in subsequent generations (either through selfing of the production of DHs of these plants) homozygous parents can be obtained.

Layout of this thesis

During my research I have addressed both fundamental as well as more applied aspects of meiotic recombination. In chapter two I discuss some of the most basic aspects of meiotic recombination by examining (male) *Arabidopsis* meiosis at the highest detail. During early meiotic prophase *Arabidopsis* produces a surprisingly large number of 120-235 double strand breaks (DSBs) along the chromosomes axes (Chelysheva *et al.* 2010; Sanchez-Moran *et al.* 2007). And though essential for the formation of crossovers (Grelon *et al.* 2001), the far majority of DSBs do not lead to the formation of meiotic crossovers, since only about 10 crossover events are observed in a typical meiosis. Using whole genome sequencing, we took a detailed look at meiosis products: both by sequencing doubled haploid offspring as well as whole meiotic tetrads, in which the genomes of all four offspring from a single meiosis event were sequenced. The excess DSBs are repaired presumably through very small gene-conversion tracts. In addition, we show that CO-sites in *Arabidopsis* preferentially localize at sites of constitutively open chromatin, and we describe the surprising presence of consensus motifs at recombination sites in *Arabidopsis*, suggesting that crossovers may be directed to, or be promoted by specific sequences.

In chapter three focus shifts from crossover formation to the regulation of CO formation and the placement of crossovers along the chromosome axis. Crossover formation is studied in an allelic series of hypomorphic mutants of CDKA;1, the main cell-cycle regulator of *Arabidopsis*. These hypomorphis mutants have varying levels of CDKA;1 activity, and help to dissect the requirement of this cell-cycle for meiosis. CDKA;1 activity is required for a variety of processes in early meiosis (in chromosome condensation, sister-chromatid cohesion and crossover formation) as well as late meiosis (i.e. in the progression through meiosis II). One of the studied hypomorphic alleles provides insights into the regulation of crossover positioning. In a nutshell, our data suggest that low CDKA;1 activity leads to low levels of crossover interference, and the distal positioning of crossovers. Inversely, high CDKA;1 activity leads to high interference, coupled to a placement of crossovers in the middle of chromosomes. This together suggests that CDKA;1 activity is an active player shaping the recombination landscape, and could very well cause differences in recombination landscapes between sexes (heterochiasmy).

In chapter four attention shifts to a yet higher level of integration, by closely examining the chromosome structure between related species in the genus Solanum. Within this genus there is a variety of important crops (tomato, potato, eggplant) for which there are extensive breeding programs. A proper understanding of the way in which the genomes of these species are organised (i.e. to what extent the chromosomes are co-linear), is very important. From a breeding point of view, the collinearity of chromosomes of different species ultimately defines whether traits (genes of alleles) can be introgressed from one species into the other, or whether inversions would lead to extensive linkage drag resulting from inversions. From a more evolutionary point of view, the co-linearity of chromosomes can be used to trace chromosome evolution.

Having examined some of the most fundamental aspects of meiosis that influence meiotic recombination, I will then focus more on breeding itself. The identification of more and more meiotic genes in *Arabidopsis* and other plants in the last decade (Mercier and Grelon 2008), it becomes more and more clear that there is a huge potential in modifying meiosis for breeding. In chapter five the various possibilities of modifying meiosis are reviewed from a breeding perspective. The possibilities of increasing meiotic recombination, inducing recombination in (homoeologous) regions that would normally not recombine, the (im-)possibilities of breeding with chromosomes that show high sequence divergence or inversions and the possibilities of suppressing meiotic variation for breeding.

Addressing all possible ways of modifying meiosis through experiments lies well beyond the scope of this (and probably any) thesis. Instead, we focused on developing one strategy in high detail. Chapter six introduces reverse breeding: an anticipated breeding method based on the suppression of crossover recombination. It is explained how the suppressing crossover recombination in combination with doubled haploid technology can bring about breeding strategies that allow breeders to fix an uncharacterized heterozygous genotype by constructing homozygous parents for it. It would allow breeders to select interesting heterozygotes that can be brought into breeding programs by (re-) constructing parental lines. Simultaneously, reverse breeding provides a relatively simple method for the production of chromosome substitution lines that have great uses in the genetic dissection of complex traits.

Penultimate chapter seven describes the practise of reverse breeding, by presenting a proof of concept in the model plant *Arabidopsis*. *Arabidopsis* is the species of choice for the plant geneticist: as it can easily be transformed, haploids can be readily generated and its genome sequence is completely known.. We demonstrate, that reverse breeding in *Arabidopsis* is perfectly feasible, and as expected, generates parental lines from a starting heterozygote.

Last chapter eight then discusses the various topics addressed in this thesis, and identifies new challenges and possibilities using knowledge and insights on meiosis for future

breeding applications. The possibilities that reverse breeding generates for the dissection of complex traits and heterosis will therefore be extensively discussed.

Literature

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B (2010) PRDM9 Is a major determinant of meiotic recombination hotspots in humans and mice. Science 327:836

Berchowitz LE, Copenhaver GP (2008) Fluorescent *Arabidopsis* tetrads: a visual assay for quickly developing large crossover and crossover interference data sets. Nature Protocols 3:41

Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: finding plant genes for disease and defense. Current Opinion in Plant Biology 13:394

Borde V, Robine N, Lin W, Bonfils S, Geli V, Nicolas A (2009) Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. The EMBO Journal 28:99

Chelysheva L, Grandont L, Vrielynck N, le Guin S, Mercier R, Grelon M (2010) An easy protocol for studying chromatin and recombination protein dynamics during *Arabidopsis thaliana* meiosis: immunodetection of cohesins, histones and MLH1. Cytogenetic and Genome Research 129:143

Colas I, Shaw P, Prieto P, Wanous M, Spielmeyer W, Mago R, Moore G (2008) Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. Proceedings of the National Academy of Sciences 105:6075

Copenhaver GP, Nickel K, Kuromori T, Benito M-I, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD, McCombie WR, Martienssen RA, Marra M, Preuss D (1999) Genetic definition and sequence analysis of Arabidopsis centromeres. Science 286:2468

Crismani W, Girard C, Froger N, Pradillo M, Santos JL, Chelysheva L, Copenhaver GP, Horlow C, Mercier R (2012) FANCM limits meiotic crossovers. Science 336:1588

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R (2009) Turning meiosis into mitosis. PLoS Biology 7:e1000124

de Boer E, Stam P, Dietrich AJJ, Pastink A, Heyting C (2006) Two levels of interference in mouse meiotic recombination. Proceedings of the National Academy of Sciences 103:9607

de los Santos T, Hunter N, Lee C, Larkin B, Loidl J, Hollingsworth NM (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

Dirks R, Van Dun K, De Snoo CB, Van Den Berg M, Lelivelt CLC, Voermans W, Woudenberg L, De Wit JPC, Reinink K, Schut JW, Van Der Zeeuw E, Vogelaar A, Freymark G, Gutteling EW, Keppel MN, Van Drongelen P, Kieny M, Ellul P, Touraev A, Ma H, De Jong H, Wijnker E (2009) Reverse breeding: a novel breeding approach based on engineered meiosis. Plant Biotechnology Journal 7:837

Falque M, Anderson LK, Stack SM, Gauthier F, Martin OC (2009) Two types of meiotic crossovers coexist in maize. The Plant Cell 21:3915

Forster BP, Heberle-Bors E, Kasha KJ, Touraev A (2007) The resurgence of haploids in higher plants. Trends in Plant Science 12:368

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK (2007) Recombination: an underappreciated factor in the evolution of plant genomes. Nature Reviews Genetics 8:77

Gerton JL, Hawley RS (2005) Homologous chromosome interactions in meiosis: diversity amidst conservation. Nature Reviews Genetics 6:477

Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mézard C (2011) Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. PLoS Genetics 7:e1002354

Goldfarb T, Lichten M (2010) Frequent and efficient use of the sister chromatid for DNA double-strand break repair during budding yeast meiosis. PLOS Biology 8:e1000520

Greer E, Martín AC, Pendle A, Colas I, Jones AME, Moore G, Shaw P (2012) The *Ph1* locus suppresses Cdk2-type activity during premeiosis and meiosis in wheat. The Plant Cell 24:152

Grelon M, Vezon D, Gendrot G, Pelletier G (2001) AtSPO11-1 is necessary for efficient meiotic recombination in plants. the EMBO Journal 20:589

Grey C, Barthès P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B (2011) Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLOS Biology 9:e1001176

Higgins JD, Armstrong SJ, Franklin FCH, Jones GH (2004) The *Arabidopsis* MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. Genes and Development 18:2557

Knight E, Greer E, Draeger T, Thole V, Reader S, Shaw P, Moore G (2010) Inducing chromosome pairing through premature condensation: analysis of wheat interspecific hybrids. Functional and Intergrative Genomics 10:603

Lenormand T, Dutheil J (2005) Recombination difference between sexes: a role for haploid selection. PLoS Biology 3:e63

Lhuissier FGP, Offenberg HH, Wittich PE, Vischer NOE, Heyting C (2007) The mismatch repair protein MLH1 marks a subset of strongly interfering crossovers in tomato. The Plant Cell 19:862

Lou Q, Iovene M, Spooner D, Buell C, Jiang J (2010) Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence in situ hybridization mapping. Chromosoma 119:435

Lu P, Han X, Qi J, Yang J, Wijeratne AJ, Li T, Ma H (2012) Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg *erecta* and all four products of a single meiosis. Genome Research 22:508

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

Marimuthu MPA, Jolivet S, Ravi M, Pereira L, Davda JN, Cromer L, Wang L, Nogué F, Chan SWL, Siddiqi I, Mercier R (2011) Synthetic clonal reproduction through seeds. Science 331:876

Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. Proceedings of the National Academy of Sciences 109:981

Melamed-Bessudo C, Yehuda E, Stuitje AR, Levy AA (2005) A new seed-based assay for meiotic recombination in *Arabidopsis thaliana*. The Plant Journal 43:458

Mendel JG (1866 (for 1865)) Versuche über Pflanzenhybriden. Verhandlungen des Naturforschenden Vereins in Brünn, Abhandlungen 4:3

Mercier R, Jolivet S, Vezon D, Huppe E, Chelysheva L, Giovanni M, Nogué F, Doutriaux M-P, Horlow C, Grelon M, Mézard C (2005) Two meiotic crossover classes cohabit in Arabidopsis: one is dependent on MER3, whereas the other one is not. Current Biology 15:692

Mercier R, Grelon M (2008) Meiosis in plants: ten years of gene discovery. Cytogenetics and Genome Research 120:281

Mirouze M, Lieberman-Lazarovich M, Aversano R, Bucher E, Nicolet J, Reinders J, Paszkowski J (2012) Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. Proceedings of the National Academy of Sciences 109:5880

Moses MJ (1968) Synaptinemal complex. Annual Review of Genetics 2:363

Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada J-P (2010) Control of female gamete formation by a small RNA pathway in *Arabidopsis*. Nature 464:628

Osman F, Dixon J, Doe CL, Whitby MC (2003) Generating crossovers by resolution of nicked holliday junctions: a role for Mus81-Eme1 in meiosis. Molecular Cell 12:761

Pawlowski WP, Golubovskaya IN, Cande WZ (2003) Altered nuclear distribution of recombination protein RAD51 in maize mutants suggests the involvement of RAD51 in meiotic homology recognition. The Plant Cell 15:1807

Peters SA, Bargsten JW, Szinay D, van de Belt J, Visser RGF, Bai Y, de Jong H (2012) Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. The Plant Journal 1:602

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. Journal of Experimental Botany 56:1

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

Ravi M, Chan SWL (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464:615

Ravi M, Marimuthu MPA, Siddiqi I (2008) Gamete formation without meiosis in *Arabidopsis*. Nature 451:1121

Salome PA, Bomblies K, Fitz J, Laitinen RAE, Warthmann N, Yant L, Weigel D (2012) The recombination landscape in *Arabidopsis thaliana* F2 populations. Heredity 108:447

Sanchez-Moran E, Santos J-L, Jones GH, Franklin FCH (2007) ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in *Arabidopsis*. Genes and Development 21: 2220

Sherman JD, Stack SM (1995) Two-dimensional spreads of synaptonemal complexes from solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics 141:683

Stack SM (1984) Heterochromatin, the synaptonemal complex and crossing over. Journal of Cell Science 71:159

Sturtevant AH (1915) The behavior of the chromosomes as studied through linkage. Molecular and General Genetics MGG 13:234

Szinay D, Wijnker E, van den Berg R, Visser RGF, de Jong H, Bai Y (2012) Chromosome evolution in Solanum traced by cross-species BAC-FISH. New Phytologist 195:688

Tischfield SE, Keeney S, (2012) Scale matters: The spatial correlation of yeast meiotic DNA breaks with histone H3 trimethylation is driven largely by independent colocalization at promoters. Cell Cycle 11:1496

van Dijk PJ (2008) Biotechnology: A hold on plant meiosis. Nature 451:1063

Vizir I, Korol AB (1990) Sex difference in recombination frequency in *Arabidopsis*. Heredity 65:379

Wang J, Fan HC, Behr B, Quake Stephen R (2012) Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell 150:402

Westergaard M, von Wettstein D (1972) The synaptinemal complex. Annual Reviews of Genetics 6

Wijnker E, de Jong H (2008) Managing meiotic recombination in plant breeding. Trends in Plant Science 13:640

Wijnker E, van Dun K, de Snoo CB, Lelivelt CLC, Keurentjes JJB, Naharudin NS, Ravi M, Chan SWL, de Jong H, Dirks R (2012) Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. Nature Genetics 44:467

Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, Miller N, Drouaud J, Grelon M, Copenhaver GP, Mezard C, Kelly KA, Henderson IR (2012) Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. PLoS Genetics 8:e1002844

On simply enjoying science…

[…]

See the world in green and blue
See China right in front of you
See the canyons broken by cloud
See the tuna fleets clearing the sea out
See the Bedouin fires at night
See the oil fields at first light
And see the bird with a leaf in her mouth
After the flood all the colours came out

It was a beautiful day
Don't let it get away
Beautiful day

[…]

U2

(In: 'Beautiful day'. All that you can't leave behind, 2000)

Chapter 2

Whole-genome sequencing of (non–) crossover sites reveals that meiotic recombination targets structurally open chromatin in *Arabidopsis thaliana*

Authors:
T.G. Wijnker[1,5], Geo J. Velikkakam[2,5], Jia Ding[2], Vimal Rawat[2], Beth Rowan[3], Frank Becker[1], Bruno Huettel[2], C. Bastiaan de Snoo[4], Daniel F.W. de Jong[4], Hans de Jong[1], Detlef Weigel[3], Maarten Koornneef[1,2], Joost J.B. Keurentjes[1] and K. Schneeberger[2].

1) Laboratory of Genetics, Wageningen University, Wageningen, The Netherlands.
2) Max Planck Institute for Plant Breeding Research, Köln, Germany
3) Max Planck Institute for Developmental Biology, Tuebingen, Germany
4) Rijk Zwaan R&D Fijnaart, Fijnaart, The Netherlands
5) These authors contributed equally to this research

Correspondence to Joost.Keurentjes@wur.nl or Schneeberger@mpipz.mpg.de

Abstract

Meiotic recombination has long been studies in the model plant *Arabidopsis* but major questions still remain about the repair of double strand breaks (DSBs), the number and tract lengths of gene conversions and the positioning of recombination events. We set out to answer these questions using whole-genome sequencing in the most detailed study of *Arabidopsis* meiosis so far. Deep sequencing of five complete meiotic tetrads and ten homozygous doubled haploids reveals the precise genetic makeup of crossover and non-crossover positions. A typical tetrad displays about ten COs, six CO associated gene conversions (GCs) and five to six non-crossovers. CO-associated GCs have significantly longer conversion tracts than NCOs (~500 vs ~153-~282 bp). The latter estimate is unreliable because of the relatively low SNP density in Arabidopsis. CO recombination sites occur in the genome irrespective of promoter- or otherwise annotated regions, but preferentially occur near poly-A sequences which are known to be free from nucleosomes. Such a localization pattern is highly reminiscent of meiosis in yeast (Saccharomyces cerevisiae). Surprisingly, we find a palindromic CTTCTTCTTCT microsatellite repeat overrepresented at CO sites, which has high similarity to a (TL1) known binding site of a heat shock transcription factor.

Introduction

Homologous recombination during the meiotic prophase is initiated by the formation of double strand breaks (DSBs). When the repair of these breaks is directed through a non-sister chromatid this can lead to the formation of crossovers (COs), non-reciprocal exchange of chromosome segments between non-sisters. Alternatively DSBs can also be resolved as non-crossovers (NCOs). In that case strand invasion leads to heteroduplex formation (Allers *et al.* 2001), which are subsequently corrected and, depending of the direction of repair, give rise to a 2:2 or 3:1 segregation of loci in tetrads. Loci that segregate in a 3:1 ratio are also known as gene conversions (Zickler 1934). Gene conversions not only result from NCOs, but can also form at crossover sites, known as CO-associated gene conversions (CO-associated GCs; see Figure 1). In the yeast, *Saccharomyces cerevisiae*, COs are thought to form through a recombination intermediate known as the double Holiday Junction (Schwacha and Kleckner 1995; Szostak *et al.* 1983), whereas NCOs are currently thought to be formed mainly through an alternative pathway, known as Synthesis Dependent Strand Annealing (SDSA) (McMahill *et al.* 2007). Excellent reviews provide further details on these recombination pathways: (Filippo *et al.* 2008; Mazón *et al.* 2010; Osman *et al.* 2011).

The central question on how DSBs are being repaired during meiosis is still open. Arabidopsis forms between 120-235 DSBs per meiotic event (Chelysheva *et al.* 2007; Sanchez-Moran *et al.* 2007; Vignard *et al.* 2007) which is in far excess of the what actually becomes manifested as the nine to ten COs that are commonly observed (Armstrong and Jones 2003; Lu *et al.* 2012). Most breaks must thus employ other modes of repair. Mouse and yeast, that produce ~250 and ~150 DSBs respectively, also produce relatively

high numbers of DSBs (Buhler *et al.* 2007; Moens *et al.* 1997), whereas *Drosophila melanogaster* and the nematode *Caenorhabditis elegans* produce substantially less DSBs, with ~20 DSBs (Jang *et al.* 2003) and ~12 DSBs (Mets and Meyer 2009) per genome.

Yeast meiosis is has been studied thoroughly using tetrad studies and estimates revealed that the 150 DSBs are resolved in about 90 COs and approximately 60 NCOs (Mancera *et al.* 2008; Qi *et al.* 2009). The fraction of nucleotides subject to gene conversion in yeast during a single meiosis adds up to 1% of the total genome size, highlighting the significant contribution of GC to meiotic recombination in that species. In most eukaryotes the analyses of gene conversion in meiotic tetrads are difficult or impossible because spores separate after meiosis (Sun *et al.*, 2012), but the discovery of the quartet mutant in *Arabidopsis thaliana* (Preuss *et al.* 1994) makes studies into gene conversion rates feasible sun (Lu *et al.* 2012; Sun *et al.* 2012). Sun et al (2012) recently estimated that 3.5×10−4 conversions take place per locus per meiosis.

Most data provided by Sun *et al.* (2012) could however not distinguish NCOs from CO-associated GCs. Lu *et al.* (2012), in which eight offspring from two complete Arabidopsis meiotic tetrads were sequenced, showed that both NCOs and CO-associated GCs occur. Gene conversions events were found to accompany about half of CO events and the tract length was estimated at ~558 bp. Four NCOs were recovered in their study with tract lengths ranging from one (three events) to 1799 bp. This low number of NCOs is remarkable in view of the ~120 - ~235 DSBs formed. One theory (Lu *et al.*, 2012) suggests that DSBs are predominantly repaired via the (identical) sister chromatid instead of the homologous chromosome (hence leaving no detectable NCOs). Alternatively, heteroduplexes arising during NCO formation might be preferentially repaired towards the broken strand, thereby restoring the original genotype, and preventing the detection of gene conversion. Finally, NCOs might have very short tract lengths in plants and their detection might require higher levels of polymorphisms for their detection than present in *Arabidopsis*, which amounts to ~1 per 200 bp on average (Lu *et al.* 2012). However, the latter hypothesis seems difficult to reconcile with their claimed long NCO of 1799 bp. A reliable estimate of CO conversion tracts (COCTs) and NCOs would be most helpful in derivation of new hypotheses. Ratios for lengths of COCTs and NCOs have previously been estimated for human and yeast, in which COCTs were found to be consistently longer: ~460 and 55-290 bp respectively in human (Jeffreys and May 2004) and 2 Kb and 1.8 Kb in yeast (Mancera *et al.* 2008).

Other intriguing questions with respect to CO formation concern their precise placement. From yeast it is known that COs predominantly occur in promoter regions (Pan *et al.* 2011; Wu and Lichten 1994), presumably associated with the presence of open, accessible chromatin through the positioning of nucleosomes at the onset of meiosis (Berchowitz *et al.* 2009; Pan *et al.* 2011). For plants this is as of yet unclear, although the pattern of transposon insertion, which localize to open chromatin, shows a remarkably similar distribution to meiotic recombination landscapes in maize, suggesting that open chroma-

tin might strongly correlate with DSB-formation (Liu *et al.* 2009). A completely different relation between recombination hotspots in *Arabidopsis* and transposable elements was suggested through a study of hotspots derived from haplotypes of natural accessions (Horton *et al.* 2012). These authors showed that historical hotspots preferentially localize in regions annotated as transposable elements. A relation to open chromatin was however not suggested. Meiotic CO hotspots have been shown to also occur in *Arabidopsis* experimental populations (Giraut *et al.* 2011; Yelina *et al.* 2012), but what determines the placement of these hotspots in plants is not known.

Recombination hotspots in human and mouse were recently shown to be closely associated with specific histone marks such as H3K4me3 methylation, which is actively imposed onto chromosomes in early meiotic prophase. This presumably is the result of PRDM9 (Baudat *et al.* 2010; Buard *et al.* 2009; Smagulova *et al.* 2011), a methyltransferase with a zinc finger DNA binding domain. As to whether H3K4me3 methylation also defines recombination hotspots in yeast is still under investigation with conflicting findings reported thus far (Buard *et al.* 2009; Tischfield *et al.* 2012). The involvement of the DNA binding methyltransferase PRDM9 in CO formation in some (but not all) mammals (Muñoz-Fuentes *et al.* 2011; Oliver *et al.* 2009) also points to the fact that specific sequence motifs may influence the propensity of a sequence to form DSBs and thereby become a hotspot for DSB and CO formation. Common sequence motifs have so far been reported for yeasts (Wahls and Davidson 2010; White *et al.* 1993), *Arabidopsis* (Horton *et al.* 2012) and several *Drosophila* species (Comeron *et al.* 2012; Miller *et al.* 2012; Stevison and Noor 2010). Although the precise ways in which such sequences promote DSB formation (and hotspot localization), studies in yeast suggest that at least part of such sequences provide binding sites for specific transcription factors that promote meiotic hotspot formation (Wahls and Davidson 2010; White *et al.* 1993) like PRDM9 in human and mouse. Inversely, the presence of a recombination hotspot itself has been suggested to change the local DNA itself, leading to an enrichment of GC content at the hotspot site. This has was suggested to result from preferential repair of heteroduplex mismatches from A/T to G/C basepairs, a phenomenon known as biased gene conversion (BGC) (Galtier *et al.* 2001). However, to date no evidence for BGC or enriched GC content has been reported for *Arabidopsis* (Giraut *et al.* 2011; Horton *et al.* 2012).

The possibility of making doubled haploids (DH) in *Arabidopsis*, that provides the opportunity to directly turn (F1 derived-) gametes into homozygous plants, allows the easy detection of recombination footprints like NCOs (Ravi and Chan 2010). Also the possibility of studying F1-derived meiotic tetrads could be of great help (Preuss *et al.* 1994) to study CO-associated GCs and NCOs as was done by Lu *et al* (2012). Here we use both approaches in combination with a whole-genome sequencing approach at high depth to accurately estimate gene conversion tract lengths and frequencies for both COs and NCOs. Using this approach, we can obtain a detailed view of the localization of COs and NCOs

in *Arabidopsis* meiosis, and relate these to genomic features like histone methylation, consensus sequences and regions known to represent open chromatin.

Materials and Methods

Plant materials

Doubled haploid *Arabidopsis* lines were selected from a previously established Columbia (Col) – Landsberg *erecta* (L*er*) DH population (Wijnker *et al.* 2012). For the generation of meiotic tetrads, a Col - L*er* hybrid in a quartet1 background was made by crossing *qrt1* -/- Col (N660403) to L*er* (N8050) lines which were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC) (http://Arabidopsis.info/). Single meiotic quartets from this F1 were picked up with a single hair under a dissecting microscop and transferred onto individual style of a virgin flower (Copenhaver *et al.* 2000) of a male sterile Cape Verde Islands (CVI), that was selected from a EMS mutation screen (by M. Koornneef). These plants were grown under long day (14 hours light) conditions in a growth chamber. Over 700 unique pollinations were made accordingly. The resulting siliques were harvested individually and when four seeds were recovered from one silique, they were grown under short day conditions (8 hours light) to maximize rosette size before harvesting. Plants were genotyped using a previously described SNP marker set (Wijnker *et al.* 2012) to verify that all markers segregate in the expected tetrad 2:2 ratio.

Sample- and library preparation and Sequencing

DNA samples of parental lines, DHs and tetrad offspring were extracted from fully grown rosettes using a CTAB method, with a nuclei extraction step to remove mitochondrial and chloroplast DNA. Rosette leaves (0.5-1 gram) were ground to a fine powder in liquid nitrogen using mortar and pestle and transferred to a 15-mL polyethylene centrifuge tube containing 10 mL of ice-cold nuclei extraction buffer, consisting of 10 mM TRIS-HC1 pH 9.5, 10 mM EDTA pH 8.0, 100 mM KC1, 500 mM sucrose, 4 mM spermidine, 1 mM spermine and 0.1% beta-mercaptoethanol. The suspended tissue was mixed thoroughly with a wide-bore pipette and filtered through (by a brief spin at less than 100g for 5 seconds) two layers of Miracloth (CalBiochem, http://www.merckmillipore.com/) to an ice-cold 50-mL polyethylene centrifuge tube. Lysis Buffer (2 mL), consisting of 100mM Tris ph7.5, 0.7M NaCl, 10 mM EDTA, 1% BME (2-mercaptoethanol) and 1% CTAB in MQ water, was added to the filtered suspension and mixed gently for 2 min on ice. The nuclei were pelleted by centrifugation at 2000g for 10 min at 4 °C. 500 µL CTAB extraction buffer was added to nuclei pellet, mixed well by inverting tube and incubated for 30 min at 60 °C. Samples were cooled at RT for 5 min. after which 350 uL chloroform/iso amyl alcohol (24:1) was added and inverted and mixed gently for about 5 min and spun for 6k rpm for

10 min. The upper layer (450 µL) was transferred to a new 2 mL tube containing 450 µL isopropanol, and mixed by inverting several times and pellet DNA by spinning 13000 rpm for 3 min. After washing the DNA pellet in 75% EtOH, the DNA was resuspended in sterile DNase free water (containing RNaseA 10 µg/mL). The sample was incubated at 65 °C for 20 min to destroy any DNases, and stored at 4 °C until use. DNA concentration and quality was determined with a Nanodrop 1000 (Peqlab; http://www.peqlab.de), a Qubit® 2.0 Fluorometer (Life Technologies™; http://www.lifetechnologies.com/) and on a 1% agarose gel. DNA samples were concentrated to more than 50 ng × μL^{-1} with a speed-vac when necessary.

Library preparation and sequencing

At least 500 ng of high-quality (260/280 ratio > 1,8) genomic DNA was fragmented on a COVARIS S2 (http://covarisinc.com/) to achieve a mean length of 300 bp followed by DNA purification with PCR purification columns (QIAquick PCR purification kit, QIAGEN; www.qiagen.com/). Libraries were generated using the Illumina Genomic DNA TruSeq sample kit (www.illumina.com/) according to the manufacturer's instruction (TruSeq DNA sample preparation v2 guide, Illumina) with Sciclone G3 robotics using the TruSeq DNA protocol by Caliper (Caliper Life Sciences, Hopkinton, MA; www.caliperls.com/). Libraries were quality assessed with a Bioanalyzer (Agilent 2100, Agilent, www.genomics.agilent.com/) and then quantified by fluorometry, immobilized and processed onto a flow cell with a cBot (Illumina) followed by sequencing-by-synthesis in 100 bp paired-end runs on a Illumina Genome Analyzer GAIIx (Col parent of DH lines, DH lines 1 to 4) or an Illumina HiSeq2000 (Ler parent of DH lines, Col (qrt), L*er* (qrt) and CVI, DH 5 to 10 and five deep-sequenced tetrads). After the run read data were extracted with the software package CASAVA (Version 1.8, Illumina; www.illumina.com/) to prepare the .fas/.qual read pairs. Sequencing yield per sample is listed in Supplementary Table S1.

Resequencing analysis

We applied the resequencing pipeline SHORE to each of the samples (Ossowski *et al.* 2008). First, reads were trimmed based on quality values. High quality reads were then aligned against the reference sequence using GenomeMapper as alignment tool (Arabidopsis_Genome_Initiative 2000; Schneeberger *et al.* 2009). After using read pair information to remove repetitive reads, we performed consensus calling to assess reference position-specific counts for all alleles, in addition to a SHORE quality score (Supplementary Figure 1).

Marker definition for DH genome analysis

Based on the resequencing results of the parental accessions, Col-0 and L*er*, we defined a set of high quality markers that we used for initial genotyping of all DH samples. Positions that account for valid markers had support for the reference allele within the resequencing of one Col-0 sample with a SHORE quality value of 25 or greater and support for a homozygous SNP in the resequencing analysis of L*er* with a resequencing quality value of 40. Even though the reference sequence was based on the genome of Col-0, valid markers also included the rare cases where the resequencing of Col-0 supports a non-reference allele (with a SHORE quality value of 25 or greater) and Ler features the reference allele (again quality value of 25 or greater). The coverage at all valid marker positions was required to be between 50 and 150 read alignments for the resequencing of L*er*, in order to remove regions with different copy number in the genome of L*er*, in comparison to the reference sequence. This identified 438,919 positions as markers, which were used to genotype the double haploid lines.

Genotyping followed by CO and NCO assignment in DH genomes

The recombinant genomes of DH lines were established at all previously identified marker positions by assigning one of five different states to each marker. These five states included both parental alleles (if there was a resequencing quality score higher than 15 and more than three read alignments), ambiguous markers (if evidence for both alleles was present), the presence of a third allele and finally all markers with less than three alignments were assigned a non-informative state. This initial genotyping was used to identify COs by counting consecutive markers of the same genotype. Blocks with at least 25 consecutive markers were used as "seeds" and COs were called in between seeds with different genotypes. COs positions were improved by extending the seeds until the nearest block of opposing markers, featuring more markers than the preceding block, is identified. COs assessment partitioned the genomes in regions of different parental backgrounds, which descended from the respective parent.

NCOs positions were identified by identification of markers differ from the parental background. Only markers with unique support for one of the parental alleles were considered (**Supplementary Table S2**). These initial NCOs were then filtered for occurrence of the same NCO within more than one DH line, in order to exclude NCO annotations that are due to systematic errors. This last step would exclude NCOs that happen independently and convert the same marker (**Supplementary Table S2**). The total amount of these occurrences was 42. Manually inspections of the short read alignments at these positions showed that the coverage at each single marker was drastically reduced, or the resequencing indicated large-scale disruptions.

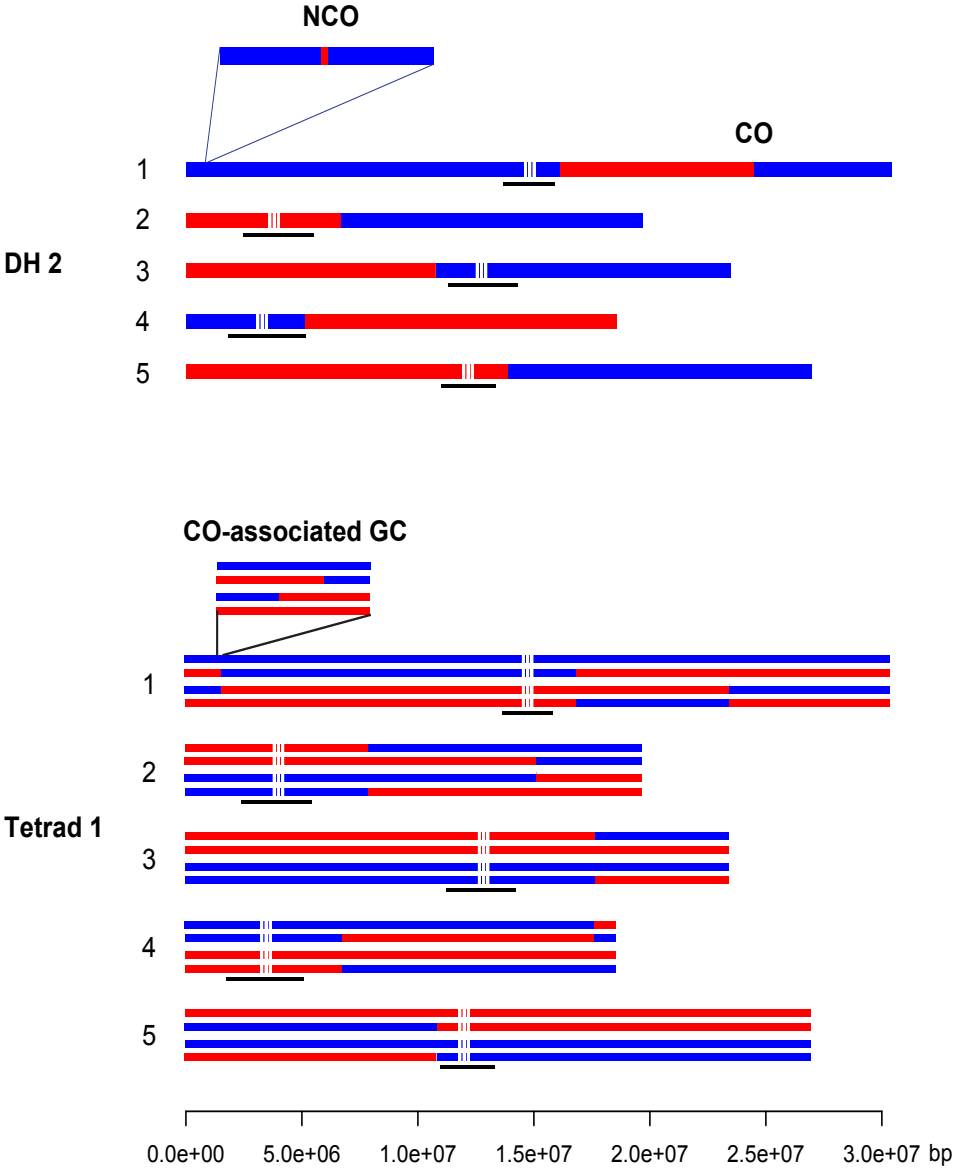


Figure 1: Genome composition of a doubled haploid (above) and tetrad (below). Chromsome numbers are indicated on the left. Landsberg *erecta* alleles given in blue, Columbia in red. Enlargement in DH shows the presence of a NCO. Inset in tetrad shows a CO-associated GC. Note here the 3:1 segregation of Col-alleles. Black bars represent the preicentromere areas.

Marker definition for tetrad genome analysis

In contrast to the DH analysis, three different parents contributed to the tetrad offspring. This thus required a different set of markers for genotyping. Valid markers included all positions with resequencing scores higher than 24 in the analysis of Col-0 and CVI and a non-reference allele with a quality score of 40 within the analysis of Ler. Naturally, the alleles of Col-0 and L*er* needed to be different at valid markers.

Genotyping and CO and NCO assignment in tetrad genomes

Initial genotyping and CO location identification was performed as outlined for the Col-L*er* DH lines. In contrast to the DH analysis, tetrad samples involve a third genotype, CVI (*i.e.* a Col-L*er* F1 was crossed onto a CVI female receptor and so is heterozygous). In order to avoid unreliable resequencing calls we excluded all marker positions that overlapped with annotated transposable elements and those markers, which reside in regions that are likely to have a different copy number in any of the parents, as identified within the SHORE resequencing analysis of the parental alleles. For NCO detection we need to distinguish two different scenarios. In the first case a putatively converted allele differs from the recipient parent (CVI) and at the same time differs from the parental genotype. NCOs can be confidently identified by observing an additional, unexpected genotype (type 1 NOCs). In the second case the recipient parental allele is different from the expected allele of the second allele and hence the absence of an expected allele is required for identification of NCOs (type 2 NCOs). However, confidently identifying absence is more challenging than identifying the presence of alleles.

In detail, valid type 1 NCOs required a minimal quality score of 35 and in addition a minimum coverage of 25 alignments per marker, of which at least 15 had to support the unexpected converted allele. In contrast, type 2 NCOs were only scored if there were more than 50 short read alignments present, of which no more than two represented the putatively absent allele. For these NCO predictions we applied hard cutoffs, even though the coverage between the different lines varies. This accounts for the fact that genomes sequenced at low coverage do not provide the power for the recalling the same number of events. In fact, for the estimation of the total amount of NCOs we considered only those markers that featured enough sequencing reads and quality scores high enough as to reveal NCOs. This drastically reduces number of markers (**Supplementary Table 3**), however it allows our analysis to settle on a set of markers for which we can confidently say that they have or have not undergone a conversion. Like for the NCOs identified in the DH samples, we filtered for those NCOs that were identified in more than one background and removed 178 NCOs from our analysis.

Refinement of tract lengths of COs and NCOs

Marker definition based on next-generation resequencing usually suffers from incompleteness. Conservative thresholds that aim at reducing false positives naturally exclude a substantial fraction of the real polymorphisms. In particular, more complex changes are unlikely included within the set of polymorphisms that are confidently identified by short read alignments. Consequently, our marker sets are unlikely to include all types of polymorphisms existing between the parental accessions. In order to get a complete picture of the COCTs and conversion tracts of NCOs that were identified with the sparse marker as defined above, we manually refined all CO and NCO tracts by visually parsing the short read alignments. This allowed us to explore all polymorphisms within each of the conversions tracks and to reveal the real extent of all tracts identified.

Statistics

We used the 2-sample Kolmogorov-Smirnov test (2-sided) to test whether the observed size distributions of CO-associated GCs and NCO tract lengths are different. Since sample sizes differed (32 and 28 observations respectively), the critical D-value (α=0.01) was calculated as Dcrit.=1.63*$\sqrt{}$ ((32+29)/(32*29)) = 0.443.

Functional annotation of polymorphism within conversion tracts

The functional annotation of CO sites in DH lines and tetrads follows TAIR10 (http://www.Arabidopsis.org/). In tetrad offspring, the annotation of a CO position can in specific cases differ between two offspring, when the flanking SNPs of a CO conversion tract are differently annotated (**Figure 1**). To compare the annotation of COs in tetrads and DH lines, we used the annotation of DH CO sites and the annotation of one offspring per CO position from the tetrad data.

Motif search

We searched for consensus motifs at CO-sites identified in the tetrad offspring employing highly restrictive to inclusive search methods. Performing motif searches only on sequences revealed by CO-associated GCs ensures that the sequences were subjected to DSB repair. As GC are not necessarily centered on the location of the respective DSB, additional inclusion of flanking sequences of different lengths can increase the probability of including sequences in which the actual DSB was formed. These sets are described in **Table 1** and range from most restrictive (Set 1) to most inclusive (Set 8). Candidate motifs were identified with MEME (Bailey *et al.* 2006), which was ran with the "zoops" model, while correcting for the genomic background. Searches were done for all eight sets above. For matching a found consensus sequence against known transcription factor binding sites, we used JASPAR (Portales-Casamar *et al.* 2009) (http://jaspar.cgb.ki.se) and TRANSFAC (Matys *et al.* 2003) (www.gene-regulation.com/pub/databases.html#transfac).

Set	Sequences	
1	32	converted SNPs, with 50 bp on either side
2	32	converted SNPs, with 250 bp on either side
3	32	converted SNPs, with 500 bp on either side
4	33	converted COCT SNPs plus 50 bp on either side, plus maximal COCTs smaller than 200 bp
5	40	converted COCT SNPs plus 250 bp on either side, plus maximal COCTs smaller than 500 bp
6	51	converted COCT SNPs plus 1000 bp on either side, plus maximal COCTs smaller than 2000 bp
7	52	All CO tracts
8	52	All CO tracts with 500 bp on either side

Table 1: Description of different sets used for MEME motif searches in tetrad offspring.

Nucleosome occupancy

For determining nucleosome occupancy at CO sites, sequences of their positions and their flanking regions were pasted to NXsensor (www.sfu.ca/~ibajic/NXSensor/) (Luykx *et al.* 2006). For determining the minimal distance from a CO-site to the nearest nucleosome free area (**Figure 8**), annotated the distance from the CO-midpoint to nucleosome-free areas as defined by An, n ≥ 10 (poly-A); Tn, n ≥ 10 (poly-T) and (C/G)3-N2-(C/G)3-N2-(C/G)3.

RESULTS

Experimental design and data analysis

NCOs leave a genetic footprint in the form of converted single nucleotide polymorphisms (SNPs) or insertion/deletions, and such small changes are most easily detected through whole-genome sequencing of homozygous samples. We therefore selected ten random doubled haploids from a DH population made from the accessions Columbia (Col) and Landsberg *erecta* (L*er*) (Wijnker *et al.* 2012). These DH offspring carry the genome complement of one meiotic spore. DH offspring do not allow the study of CO-associated GCs, since information on sister chromatids was found to be essential for defining such gene conversions (see below).

We therefore generated and sequenced offspring of five spore tetrads (the four spores from one meiosis) by crossing the four pollen that result from one meiosis onto a receptor line of the accession Cape Verde Islands (CVI). As a male we used a Col-L*er* F1, homozygous for a mutation in *quartet1*, a mutation that prevents the separation of pollen

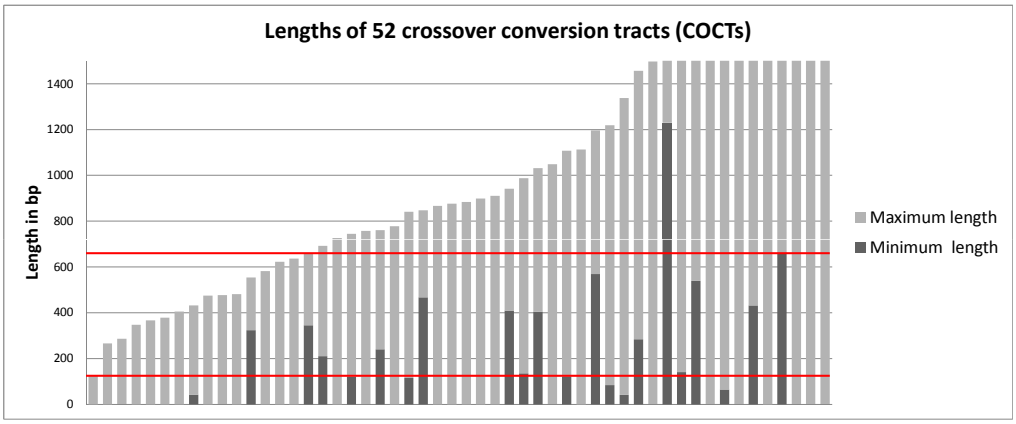


Figure 2: The minimum and maximum lengths of 52 crossover conversion tract (COCT) lengths, sorted by maximum length with the smallest shown left. The smallest maximal length (127 bp) marks the lower limit of our COCT length estimate (lower red line). The dark base of different bars represents the minimal COCT lengths. The upper limit of our COCT tract length estimate (upper red line) is marked by a minimal tract length of 663 (fourth from left). The largest minimal tract length may represent an exceptional case (see text for details). Little over half COCTs have very short or no minimum lengths and are not visible in the graph (e.g. the 7 COCTs shown left). Please note that maximum tract lengths are only shown for lengths < 1500 bp.

after meiosis (Preuss *et al.* 1994). These tetrads allow the precise examination of CO sites and associated GCs, as well as the identification of NCO events. However, heterozygosity of the offspring complicates their analysis for NCOs. In short, we need to distinguish two different scenarios. First, the identification of NCOs, where a putatively converted allele differs from the recipient parent (CVI) and at the same time differs from the parental genotype, can be identified by observing an additional, unexpected genotype (type 1). In the second scenario, where the recipient parental allele is different from the expected allele of the second allele, the absence of an expected allele is required for identification of NCOs (type 2). Lu *et al.* (2012) raised the same issue, when they backcrossed a Col x L*er* F1 onto a Col recipient parent, but solved it by removing the type 2 NCOs from their analysis completely, which comes at the costs of excluding around 50% of the genome (Lu *et al.* 2012). In our case we have called both types of NCOs using strict coverage requirements when calling type 2 NCOs (see materials and methods).

Crossover conversion tracts in Arabidopsis

Sequence analysis allowed us to determine the genome composition of DHs (**Figure 1, Supplementary Figure S2**). We identified 61 COs in the ten DHs (for positions and CO-ID numbers see **Supplementary Table S4**). DH offspring typically reveal half of the expected CO numbers in a meiotic tetrad and our DHs were chosen from a previously described DH population that show CO frequencies identical to previously published populations. Most of these show a clear cut transition from one parental genotype to the other with-

out irregular patterns of genotype changes. In one CO event we detected the footprint of a complex conversion tract by the recovery of an NCO flanking the CO breakpoint. Spore tetrads provide much higher detail at the CO make-up, as they allow the segregation ratios of markers near CO breakpoints to be checked (**Figure 1**). The genome compositions of all tetrads are shown in **Supplementary Figure 3**. A 3:1 segregation ratio of markers between CO breakpoints defines CO-associated GCs which were found to accompany 32 out of 52 COs (61%) (**Supplementary Tables S5-S7**). Most COCTs show one sided (full) conversions in which all converted SNPs are solely derived from one of the strands involved (Stahl and Foss 2010). Mismatches in the conversion tract are thus favorably resolved to one side and not randomly in either way. We observed 14 and 18 full conversions to L*er* and Col respectively, which is not significantly different (χ^2 test, $\alpha < 0.05$).

Two conversion tracts show more complex patterns, in which converted SNPs are derived from both strands (CO 21 and CO 42, **Supplementary Table S6**). A detailed alignment of CO 21 is shown in **Supplementary Figure S3**. Three out of 30 investigated gametes (*i.e.*, ten DHs and 20 tetrad offspring) that together show 165 recombinant sequences show such complex patterns in our study. This corresponds to a (3/165=) 1,8 % probability of finding its footprint in any given gamete. Four COCTs (7,7%) were found to span complex, diverged regions of the *Arabidopsis* genome (CO 1, 12, 15 and 26, **Supplementary Table S6**) with single (4 and 12 bp) or multiple deletions. Two examples of COCTs spanning such diverging regions are shown in **Supplementary Figure S4**.

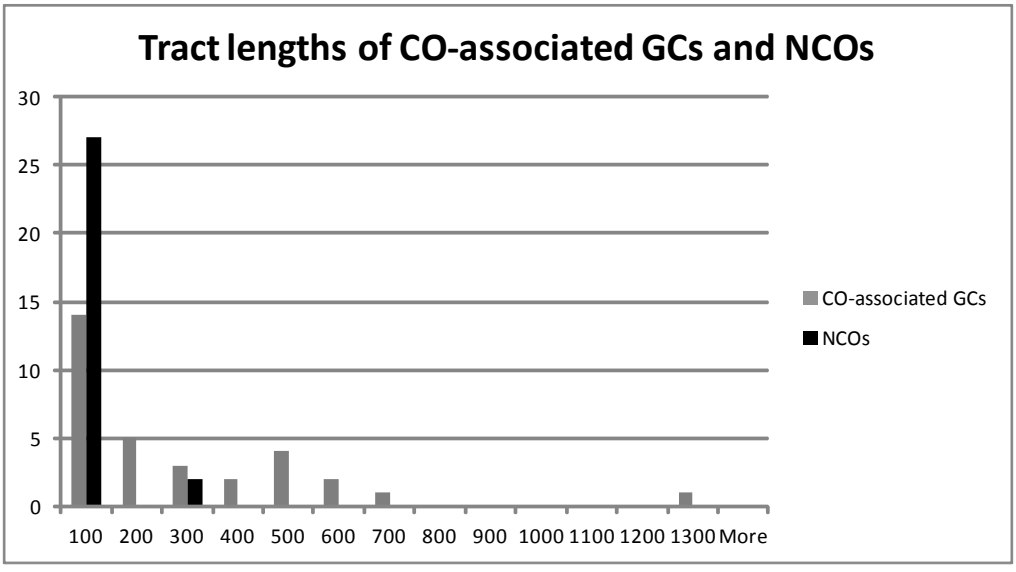


Figure 3: size distribution of CO-associated gene conversions and NCO tract lengths (in bp). Note the one long CO-associated GC conversion tract (1229 bp). The NCO tract length distribution is more skewed to the left.

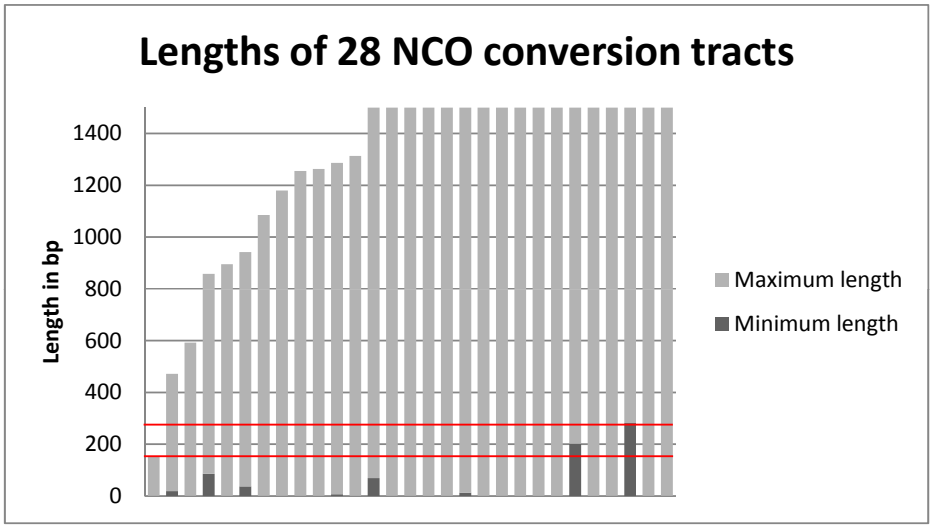


Figure 4: The minimum and maximum lengths of 28 NCO tract lengths. The tract lengths are sorted by maximum length, with the smallest shown left. The smallest maximal length (153 bp) marks the lower limit of our NCO tract length estimate (lower red line). The dark base of different bars represents the minimal NCO tract length. The upper limit of our NCO tract length estimate (upper red line) is marked by the largest minimal tract length (third from left, 282 bp). 21 NCOs have a minimum length of 1 bp and are not displayed in the graph. Please note that maximum tract lengths are only shown for lengths < 1500 bp.

Minimal COCT lengths can be derived from the distance between converted markers in a COCT, whereas the distance between the flanking markers defines the maximal tract length. Such values can be used to estimate COCT length ranges. The shortest maximal length is 127 bp, which represents a tentative lower limit to COCT lengths (see **Figure 2**). The COCTs showing converted SNPs range from 1 to 1229 bp (**Figure 2**) and most (20 out of 32 = 69%) span more than 1 converted marker. A size distribution shows that the largest minimal COCT (CO 12) is nearly ten times larger than the smallest maximal tract length (**Figure 3**). If we consider this one an exception, the second largest minimal COCT length provides a conservative upper limit to the COCT length of 663 bp. We found 16 COCTs of which the maximal length is smaller than 663 bp. COCTs in *Arabidopsis* range between 127 and 663 bp, and might average at 400 bp.

Non-crossovers in *Arabidopsis*

Among the 20 tetrad offspring we detected 28 NCOs, all of which are supported by a 3:1 segregation of markers in the tetrad as a whole. Of these, 22 show a single converted SNP while six conversion tracts span more markers, with the longest measuring 282 bp. In the ten DHs we identified a total of 11 NCOs. This equals to 1.1 per genome and is close to the 1.4 NCOs per genome we observed in the tetrads. Eight of these show only one converted SNP, whereas two others show 2 and 3 converted SNPs and have minimal tract lengths

of 86 and 69 bp (**Supplementary Table S8**).

The minimum and maximum lengths of all NCO tracts are shown in **Figure 4**. The largest minimal length (282 bp) is longer than the shortest maximum length (153 bp). The average minimal NCO tract length is 25 bp, which is smaller than the 271 bp average for CO-associated GCs. The length distribution of minimal conversion tract lengths of NCOs is shown in comparison to COCTs in **Figure 3**. NCOs are smaller than COCTs and were found to be significantly different in length (Kolmogorov-Smirnov test, $\alpha < 0.01$). We wondered whether the smaller size of NCOs relative to CO-associated GCs could be attributed to a lower SNP density at NCO sites. The average lengths of the flanking intervals for COs and NCOs is 725 and 1352 bp respectively. We found these distances to be are statistically different (t-test; two sample, unequal variance; $p < 0.01$) (**Supplemental Figure S6**). Detected NCOs thus occur in areas of low SNP density. This might have biological relevance, but could also point to a NCO detection bias towards less complex regions. The best size estimate for NCO tract lengths is 153-282 bp, but we note that size estimates become less reliable at the low SNP densities as found between (these) accessions in *Arabidopsis*.

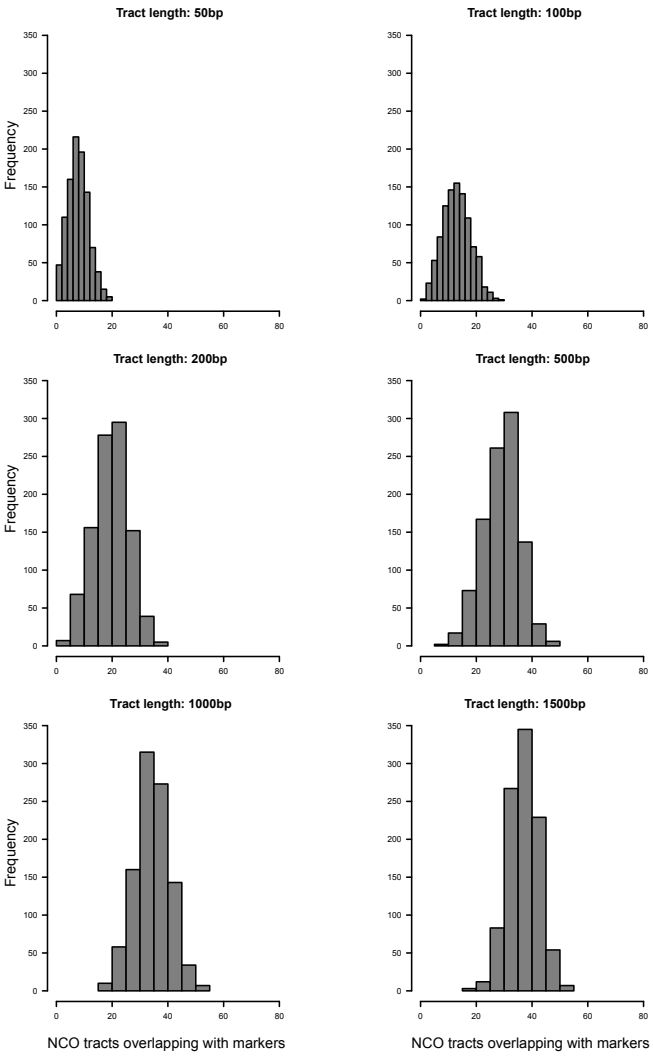


Figure 5: Expected numbers of detected NCOs per tetrad at different tract lengths. All distributions based on 1000 random samplings of 150 DSBs over the genome.

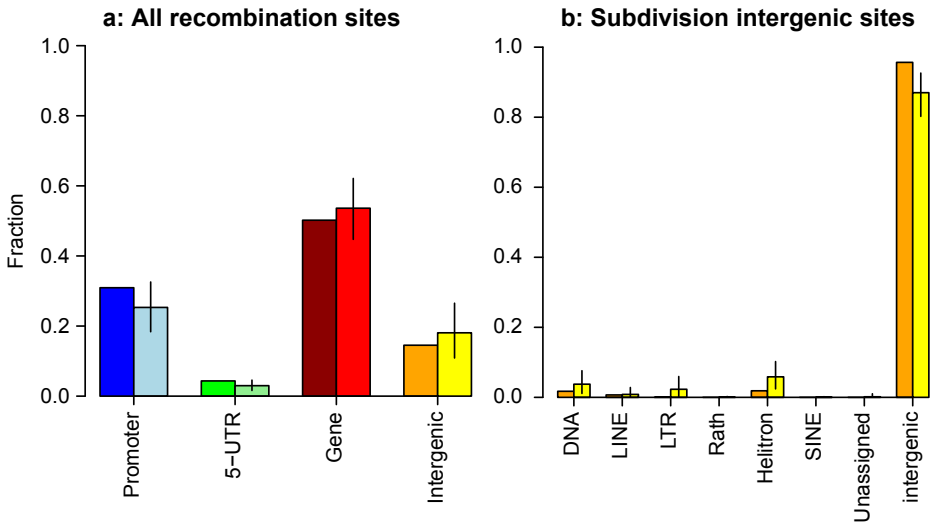


Figure 6: Annotation of CO-sites. a) Annotation of all CO sites (dark colour) compared to randomly sampled sites over the genome (light colours, with error bars). b) Precise annotation of the CO sites annotated as "intergenic" in Figure 5a. Observations given in orange, simulations in yellow. Errorbars in both a and b show 90% confidence intervals, defined by 0.5 and 0.95 quantiles based on 10,000 replications.

We observed 1.1 and 1.4 NCOs per DH and tetrad respectively and found these to be relatively small (~220 bp). Given the ~120 to ~235 DSBs that were reported in male *Arabidopsis* meiosis (Chelysheva *et al.* 2007; Sanchez-Moran *et al.* 2007; Vignard *et al.* 2007), we asked whether this number would be expected. We therefore performed random sampling throughout the *Arabidopsis* genome (excluding the pericentromere heterochromatin as defined by Giraut *et al.* determine how many NCOs would have been detected at different tract lengths (Giraut *et al.* 2011) (**Figure 5**). We assumed a conservative number of 150 DSBs, considered inter-homologue repair only and assumed that all DSBs lead to detectable NCOs. It then becomes clear becomes clear that only very small NCO tracts can explain the observed NCO numbers. At a tract length of 200 bp, we would expect ~20 NCOs per tetrad, which equals to ~5 NCOs per offspring genome. At a tract length of 50 bp, the median number of expected NCOs is eight per tetrad (*i.e.*, two per offspring genome). Under our rather conservative estimation of ~150 DSBs, we would expect to detect more NCOs that we actually do. These date could be reconciled if in the case of NCOs mismatch repair is random to either chromatid, which would be different than as we observe in COCTs. Please see our discussion for possible explanations.

CO and NCO annotation

COs do not localize randomly over the genome, but are associated with genomic features like the accessibility of DNA for recombinases, specific chromatin marks, sequence divergence and motifs that may promote CO formation. Horton *et al.* (2012) suggested

that transposable element (TE) annotated sequences are overrepresented among CO-hotspots. We therefore looked whether the annotation of sequences at CO sites in DH and tetrad offspring suggests a preference for specific regions. The annotation of CO sequences (whether they are placed in genic-, intergenic-, promoter- and 5' UTR regions) is shown in (**Figure 6**). The annotation of the 113 CO sites detected in DH and tetrad offspring are compared with randomly sampled sites over the genome, for which 95% confidence intervals were determined by bootstrapping. CO-sites in *Arabidopsis* thus do not show a preference for specifically annotated regions. Since we identified only 29 NCO sites, we did not compare the sequences of these sites.

Histone H3K4 methylation

Hotspot sites in mouse were shown to be closely associated with trimethylation of H3K4 (Buard *et al.* 2009; Grey *et al.* 2011). In *Arabidopsis* the position of a CO hotspot was also shown to co-localize with a H3K4me3 site known from somatic tissues (Yelina *et al.* 2012). We analysed whether known sites of H3K4 trimethylation coincide with the established CO positions in tetrads. Supplementary Figure S8 shows four random CO-positions in relation to known H3K4 trimethylation sites in *Arabidopsis* rosettes (Dijk *et al.* 2010). In 13 out of 52 CO sites we found the CO-position at positions rich in H3K4me3.

CO motifs and nucleosome occupancy

Previous studies showed that specific sequences are associated or regulate CO- or CO-hotspot sites a variety of species (Baudat *et al.* 2010; Comeron *et al.* 2012; Horton *et al.* 2012; Oliver *et al.* 2009). We therefore asked whether we could detect specific sequences at CO-sites in *Arabidopsis*. We used the sequences of CO positions from our tetrad data to search for consensus sequences at CO-sites. We excluded the CO positions of DH lines because tetrad CO sites are defined by two rather than only one flanking marker. Since we do not know where, if any, motifs occur, we designed various sets ranging from very strict (the minimal COCT length) to most relaxed sets as defined by the maximum conversion tract lengths (see Materals and Methods; **Table 1**).

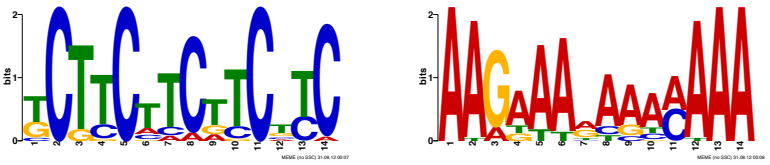


Figure 7: Two consensus motifs that are overrepresented at CO sites. The size of letters correlates to variation at that position. The CTT like cis-element (left) shows similarity to a known binding site for a heat-shock factor-like transcription factor. The poly-A like motif has no known function (right).

Our searches revealed the presence of two motifs that are significantly overrepresented at CO sites (using sets 3 and 6; Table 1). One motif is a 14 bp poly-A-like motif AAGAAA[AG]AAA[AC]AAA (e-value of 2.9e-8) that was found in 31 out of 32 sequences used in set 3. The second motif we recovered is a palindromic CTT like microsatellite [TG]CT[TC]CTTC[TG][TC]C motif (e-value of 3.5e-11) present in 23 out of 32 sequences (see **Figure 7**). The alignments of CO sequences with the recovered motifs are shown in **Supplementary Figure S8**.

We investigated whether any of the other two recovered sequences are of known function, and found the CTT-like palindrome to be near identical to the GAAGAAGAA motif of the *TRANSLOCON 1* (*TL1*) cis-element, a known binding site for a heat-shock factor-like transcription factor (Pajerowska-Mukhtar *et al.* 2012; Wang *et al.* 2005). The identified motif of 14 bp is longer than the TL1 element. Of the 23 sequences that show the 14 bp motif, the TL1 binding site is preserved with no (n=6), one (n=8) or two (n=6) mismatches.

Poly-A motifs and nucleosome occupancy

The recovery of a poly-A-like motif is reminiscent of previous reports on the presence of especially adenine-rich repeats (such as poly-A motifs) at CO-hotspot sites in *Arabidopsis* (Horton *et al.* 2012). Homogeneous poly-A motifs (>10 bp) are common in eukaryote genomes and known as sequences prohibiting nucleosome occupancy (Suter *et al.* 2000). Specific CG-rich sequences ([C/G]3-N2-[C/G]3-N2-[C/G]3) also serve the same function (Wang and Griffith 1996). Using the webtool NXSensor, we noted the regular occurrence of nucleosome-free sites in our COCTs (Luykx *et al.* 2006), and asked whether nucleosome free areas are overrepresented at CO sites. The shortest distance from all CO-midpoints to the nearest nucleosome free region was compared with a distance distribution based on randomly sampled sites (**Figure 8**). Our CO sites did occur more often near nucleosome free areas (t-test, $\alpha < 0.05$). We then asked whether the nucleosome-free poly–A sequences overlap with our identified poly-A-like motif. Interestingly, there is overlap in only 3 out of 31 cases, suggesting that the identified poly-A-like motif is a motif different from homogeneous poly-A motifs.

GC content

It has previously been suggested that mismatch repair proteins have a GC-bias by repairing heteroduplexes preferentially towards GC rather than AT bases. This can lead to GC enrichment at recombination sites (Galtier *et al.* 2001; Pessia *et al.* 2012). We calculated the GC content for 32 sequences involved in crossovers (sets 1 and 3 as described in Table 1) as well as their flanking sequences (5 Kb on either side of the CO position). The GC content for these sets is 33, 33 and 34% respectively, suggesting that GC content at CO sites

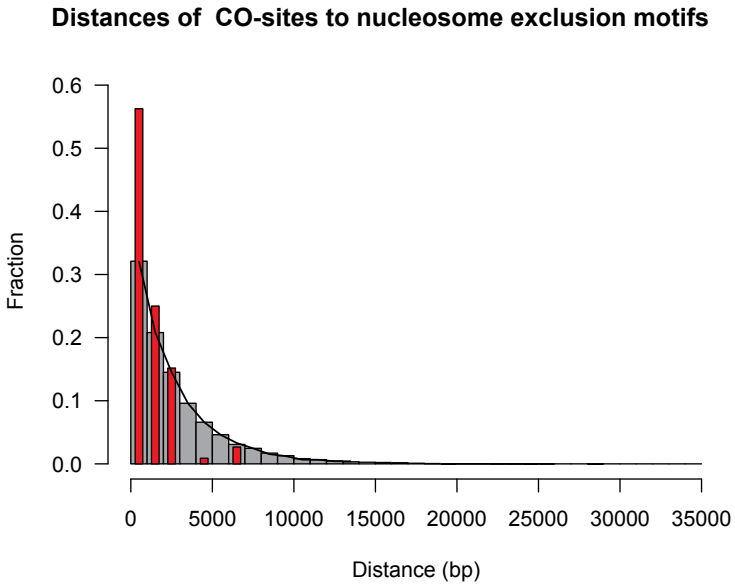


Figure 8: Overlay of normalized distributions of distance from randomly chosen sites (grey) and CO-midpoints (red, narrow bars) to the nearest nucleosome exclusion motif. The expected distribution is connected by a line. Note that CO-sites are located closer to nucleosome-free sites than is expected, when compared to the distribution based on random sampling.

not higher than the *Arabidopsis* genome average of 35% (Arabidopsis_Genome_Initiative 2000).

Discussion

Our analysis presents the most detailed sequence based report of *Arabidopsis* CO sites so far. Using whole-genome sequencing of doubled haploids and intact spore tetrads we set out to answer how the ~120 to ~235 DSBs are repaired in *Arabidopsis* (Chelysheva *et al.* 2007; Sanchez-Moran *et al.* 2007; Vignard *et al.* 2007). We determined the tract lengths of COs and NCOs, found specific motifs that accompany CO sites, which gave insight into the extent to which new gene combinations are introduced from one generation to the next one.

We were able to map 113 CO positions and 28 NCOs to the nearest SNP. At the onset of this research we hypothesized that it might be difficult to separate NCOs from incidental closely spaced COs. The very small size of all recovered NCOs makes confusion with double crossovers very unlikely. Even though NCOs are more easily recovered using homozygous DH lines, meiotic tetrads provide more information that allowed the estimation of COCT lengths and allow the detection of three consensus sequences associated with COs.

CO and NCO tract lengths and DSB repair in *Arabidopsis*.

CO sites in *Arabidopsis* are accompanied by a conversion tract in 61% of events, which is similar to yeast (69.1%) (Mancera *et al.* 2008). We estimated the size of conversion tracts accompanying CO sites at 127-663 bp based on minimum and maximum converted tract lengths (**Figure 2**). However, such tracts occasionally are longer than 1kb. Our length estimate is in the same range of that mouse (500 bp) (Cole *et al.* 2010) and human (460 bp) (Jeffreys and May 2004), but significantly smaller than the 2Kb tracts of yeast (Mancera *et al.* 2008). A previous estimate for *Arabidopsis* of 558 bp (n=6) was based on "midpoint lengths", in which the COCT length were approximated by the minimal tract length plus half the lengths of flanking intervals (Lu *et al.* 2012). That approach likely introduces a bias to overestimate COCT lengths because of low SNP density in *Arabidopsis*.

COCTs occasionally span regions that show many polymorphisms, which concurs nicely with an earlier observation, in which a COCT spanning a 86 bp deletion was reported (Lu *et al.* 2012). When a CO-associated GC is formed we usually observe a full conversion of markers, in which the converted SNPs are derived from one chromatid. Few COCTs show complex conversion tracts (1.8%) in which adjacent converted markers are derived from two, instead of only one strand. This incidence is about 10-fold lower than the 11.1% reported for yeast (Mancera *et al.* 2008).

With a length of 153-282 bp, conversion tracts of NCOs are shorter than conversion tracts of COs. A previous reports in maize suggested that 17 Kb conversion tracts could have arisen through NCO formation (Yao *et al.* 2002), which would pose a large contrast to our data. The shorter lengths of NCOs in comparison to COCTs, compares well with yeast and mouse in which concurring ratios were reported (2000/1800 and ~500/16-117 bp. respectively) (Cole *et al.* 2010; Mancera *et al.* 2008). Four *Arabidopsis* NCOs were previously reported, one of which has a minimal length of 1799 bp (Lu *et al.* 2012). We revisited these data, and found two of these NCOs (including the one of 1599 bp) to represent duplications which are not present in the Col reference sequence, because they show a read surplus and imperfect alignments (data not shown). As such, these two NCOs (including the one of 1799 bp) reported by Lu *et al.* (2012) should be considered incorrect.

The tract length estimate we obtained for NCOs (153-282 bp) is a rough approximation based on nine observations only, while the lower limit is based on just one observation (Figure 4). The majority of observed NCOs (30 out of 39 = 77%) represent just 1 converted marker. The *Arabidopsis* SNP density in our analysis (*i.e.*, the SNPS that we could reliably call based on short reads) corresponds to about 1 SNP per 200 bp, which is simply too low for reliable assessment of these short tracts. Simulations predict ~2 recovered NCOs per genome when tract lengths measure 50 bp and 5 NCOs per genome at a tract length of 200 bp. This means that we face a 2-5 fold difference between our simulations and observed data. It is not easy to reconcile these differences. In our simulations we assumed that all mismatches would lead to detectable gene conversions. This might however not be the case. Random repair of heteroduplexes would restore the original genotype, and bring

our observations in concordance with our simulations. In COCTs we hardly ever observe random repair (when there are multiple SNPs). If random repair takes place in NCOs, these would be repaired following a different mechanism.

While we have narrowed down the NCO tract length size other uncertainties should also be considered , their exact lengths and numbers of formed DSBs numbers still have a considerable error margin. The observation that gene conversions can occur in maize centromeres (Shi *et al.* 2010) suggests that DSBs can also form in heterochromatic areas that we excluded from our simulations (*e.g.*, the regions spanned by the pericentromere heterochromatin). This may render our simulated numbers of DSBs too high. It might be possible that (some) heteroduplexes are not resolved until the first pollen mitosis, rendering some of them to go undetected in offspring. Alternatively, intersister repair might take place in regions in which the broken strand finds no homology. A hypothesis that proposes preferential intersister instead of interhomologue DSB repair in *Arabidopsis* (as was proposed by Lu *et al.* (2012) is at least less likely in the light of the recovery of short NCO tracts (and exclusion of a previously reported very long tract). In yeast, mouse and *Drosophila* all meiotic DSB repair is thought to be primarily directed through homologous recombination (Cole *et al.* 2010; Mancera *et al.* 2008; Schwacha and Kleckner 1997; Sturtevant 1925). Based on our observations we currently see no obvious reasons to suggest that this is different for plants. Finally, our results might also show a systematic error. The distance between a converted SNP and its flanking markers was found to be longer in COCTs than in NCOs. This could point to the possibility that we are less likely to detect NCOs in more diverged regions. If we preferentially detect NCOs in less complex regions, the likeliness of multiple SNPS being involved in NCOs becomes much smaller.

In a recent study Sun *et al.* (2012) based on fluorescent markers 1 gene conversion was observed per marker per 2,833 meiotic tetrads. This equals to a conversion rate of (1/11332=) 8.86×10^{5} per marker per genome. We used 7698191 markers to call CO and NCO events in 30 offspring plants (**Supplementary Tables S2 and S3**) and identified a total of 141 converted polymorphisms in CO and NCO conversion tracts (**Supplementary Tables S6 and S8**); note we excluded the indels that were not used for calling COs/NCOs). This equals a gene-conversion rate of 1.8×10^{5} per marker per genome. This is 4.9 fold lower than the estimate by Sun *et al.* (2012). We have missed some NCO events because we may have applied too stringent filters for NCO detection. However these differences might simply represent interlocus variation, already shown to exist by Sun *et al.* (2012). Our observations that CO events are overrepresented at sites of constitutively open chromatin suggests that transgenes used for NCO detection represent open chromatin themselves, and may therefore be more prone to gene conversion. Alternatively, the transgenes may have inserted in or near open chromatin and therefore have a higher chance of being subjected to gene conversion.

CO and NCO positioning in relation to DNA sequence motifs and H3K4me3

We looked for correlations of CO sites with a variety of genomic features previously suggested to correlate with CO positioning. The annotation of CO sites does not suggest a preference of COs to be directed to specific genomic regions. This contrasts to two reports that suggested CO hotspots to be preferentially located in transposable elements and intergenic regions (Horton *et al.* 2012; Kim *et al.* 2007). Different reasons could account for these differences. The presence of recombination hotspots in these studies were inferred from haplotype blocks in natural populations and their data suggest that CO-sites in transposable elements could be selected for in *Arabidopsis*. On the other hand, the method for hotspot-detection by Horton *et al.* (2012) was based on a SNP set in which TE-derived markers are more abundant (i.e. 20% of their SNPs were annotated as TE elements, whereas in our study this amounts to ~10%). The use of SNPs in transposable elements can only be done reliably when the SNP-calls are from identical sites within the genome and it could be hypothesized that transposition of transposable elements could introduce a bias in their data. During our analyses we noted that TE elements recurrently introduce false positives in calling COs and NCOs.

The occurrence of high GC content at hotspot sites in a variety of organisms has been suggested to result from biased mismatch repair, favoring repair of mismatches to GC pairs (Duret and Galtier 2009; Galtier *et al.* 2001). We found a GC content of 33-34% at our CO sites, which is slightly lower than the genome average of 35% (Arabidopsis_Genome_Initiative 2000). Sequences near recombination sites were thus not found to be enriched for GC content in our study, corroborating previous findings (Giraut *et al.* 2011; Horton *et al.* 2012) and model based predictions (Marais *et al.* 2004). About 50% of recovered recombination sites in our study is located in both introns and exons of genes. Coding regions in *Arabidopsis* were reported to have an average CG content of 44% (Arabidopsis_Genome_Initiative 2000), because of which we would expect to have found a higher GC content simply by correlation. Possibly, the association of CO sites with open chromatin (i.e. usually poly-A (-T) tracts), might causes the relatively low GC content at *Arabidopsis* recombination sites.

The importance of H3K4 trimethylation for mouse and human recombination initiation has been well established, but its importance for yeast is under debate (Tischfield *et al.* 2012). We found no obvious pattern of co-localization, when comparing H3K4 trimethylation (as determined in rosettes) with CO sites. A previous record of a single hotspot coinciding with a H3K4 site in *Arabidopsis* (Yelina *et al.* 2012) could be an exception rather than the rule. It should be noted that in mouse and human, H3K4 methylation is actively induced at prospective DSB sites through the methyltransferase PRDM9 (Grey *et al.* 2011). To determine whether H3K4 trimethylation marks recombination sites in *Arabidopsis*, establishing H3K4me3 sites in meiotic cells would be of highest importance.

Sequence motifs

A last genomic feature that has been associated with DSB formation is the accessibility of DNA/open chromatin. The accessibility of chromatin in meiosis of mouse and human appears to be actively regulated through the binding of the methyltransferase PRDM9 at specific sequences, leading to local chromatin remodeling (Berg *et al.* 2010; Grey *et al.* 2011). Research in yeast showed that 90% of DSBs localize to constitutively nucleosome free regions (Pan *et al.* 2011), which might be modified by the presence of specific transcription factors (Pan *et al.* 2011). We showed that nucleosome-free regions are also overrepresented near our CO-sites. This suggests the presence of structurally open chromatin provides a target for the *Arabidopsis* recombination machinery, as has previously been suggested for maize (Liu *et al.* 2009). We detected a poly-A like motif in 97% of the CO sites. Whether these are also regions of constitutively (nucleosome-free) open chromatin is currently unknown. The association of CO events and open chromatin might hold the key to observations that changing temperatures can influence recombination frequencies in *Arabidopsis* (Francis *et al.* 2007). These observations suggest that studies relating nucleosome occupancy to meiosis could bring a better understanding of plant meiosis. In line with our findings, a recent study showed the overrepresentation of uniform poly-A, as well as non-uniform poly-A motifs near *Drosophila* recombination sites (Comeron *et al.* 2012).

The palindromic consensus motif (CTTCTTCTTCT) present in 66% of CO sites was found to be very similar to the *Arabidopsis* TL1 binding motif, that commonly occurs in the promoter regions of genes important for systemic acquired resistance (Wang and Griffith 1996). This motif is the binding site of HSFB1, a heat shock factor like transcription factor (Pajerowska-Mukhtar *et al.* 2012). HSFB1 is known to interact with the histone acetylase HAC1, an *Arabidopsis* CREB binding protein orthologue and can induce transcriptional activation (Bharti *et al.* 2004) by recruiting HAC1 to DNA. This corresponds with the yeast *Saccharomyces cerevisieae* where transcription factors were suggested to promote recombination (Pan *et al.* 2011; White *et al.* 1993). In fission yeast, *Schizosaccharomyces pombe* (Kon *et al.* 1997; Steiner *et al.* 2002; Wahls and Smith 1994; Yamada *et al.* 2004), a protein complex able to recruit histone acetylases to recombination through of an ATF/CREB-family transcription factor complex suggests regulation in *Arabidopsis* in a similar way. However, we currently have no evidence that HSFB1 indeed is active during *Arabidopsis* meiosis.

Conclusions

Our analyses have shown the possibility of reliably detecting meiotic crossovers and gene conversion events using whole-genome sequencing. The use of appropriate filters was found to be of the highest importance, as alignment errors resulting from segmental

duplications or transposable elements easily give rise erroneous NCO calls. CO and NCO tracts were shown to be measure ~500 and ~153-282 bp respectively. These sizes suggest that Arabidopsis preferably compares to mouse in which COCTs and conversion tracts of NCOs are of roughly similar lengths. The question of how DSBs in *Arabidopsis* are repaired is not completely resolved, but these short tract lengths open up the possibility that we simply miss most of the NCOs because of very short lengths. Marker density in Arabdopsis surely poses strict limits to our ability to resolve this question in the future. Where our tract length estimates suggest similarities to mouse, the recombination pattern in Arabidopsis compares better to yeasts (*Saccharomyces cerevisieae* and *Schizosaccharomyces pombe*) than that of mouse. In the latter chromatin is known to be actively remodeled by PRDM9 that thereby creates open chromatin and provides a target for the recombination machinery. Recombination in *Arabidopsis* and *S. cerevisiae* primarily targets genomic regions that are constitutively free of nucleosomes.

Literature

Arabidopsis_Genome_Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796

Armstrong SJ, Jones GH (2003) Meiotic cytology and chromosome behaviour in wild-type *Arabidopsis thaliana*. Journal of Experimental Botany 54:1

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research 34:W369

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B (2010) PRDM9 Is a major determinant of meiotic recombination hotspots in humans and mice. Science 327:836

Berchowitz LE, Hanlon SE, Lieb JD, Copenhaver GP (2009) A positive but complex association between meiotic double-strand break hotspots and open chromatin in *Saccharomyces cerevisiae*. Genome Research 19:2245

Berg IL, Neumann R, Lam K-WG, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ (2010) PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nature Genetics 42:859

Bharti K, von Koskull-Döring P, Bharti S, Kumar P, Tintschl-Körbitzer A, Treuter E, Nover L (2004) Tomato heat stress transcription factor HsfB1 represents a novel type of general transcription coactivator with a histone-like motif Interacting with the plant CREB binding protein ortholog HAC1. The Plant Cell 16:1521

Buard J, Barthes P, Grey C, de Massy B (2009) Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. EMBO Journal 28:2616

Buhler C, Borde V, Lichten M (2007) Mapping meiotic single-strand dna reveals a new landscape of dna double-strand breaks in *Saccharomyces cerevisiae*. PLoS Biology 5:e324

Chelysheva L, Gendrot G, Vezon D, Doutriaux M-P, Mercier R, Grelon M (2007) Zip4/Spo22 is required for class I CO formation but not for synapsis completion in *Arabidopsis thaliana*. PLoS Genetics 3:e83

Cole F, Keeney S, Jasin M (2010) Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. Molecular Cell 39:700

Comeron JM, Ratnappan R, Bailin S (2012) The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genetics 8:e1002905

Copenhaver GP, Keith KC, Preuss D (2000) Tetrad analysis in higher plants. A budding technology. Plant Physiology 124:7

Dijk v, K., Ding Y, Malkaram S, Riethoven J-JM, Liu R, Yang J, Laczko P, Chen H, Xia Y, Ladunga I, Avramova Z, Fromm M (2010) Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in *Arabidopsis thaliana*. BMC Plant Biology 10:238

Duret L, Galtier N (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. Annual Review of Genomics and Human Genetics 10:285

Filippo JS, Sung P, Klein H (2008) Mechanism of eukaryotic homologous recombination. Annual Review of Biochemistry 77:229

Francis KE, Lam SY, Harrison BD, Bey AL, Berchowitz LE, Copenhaver GP (2007) From the cover: pollen tetrad-based visual assay for meiotic recombination in *Arabidopsis*. Proceedings of the National Academy of Sciences 104:3913

Galtier N, Piganeau G, Mouchiroud D, Duret L (2001) GC-Content evolution in mammalian genomes: The biased gene conversion hypothesis. Genetics 159:907

Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mézard C (2011) Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. PLoS Genetics 7:e1002354

Grey C, Barthès P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B (2011) Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLOS Biology 9:e1001176

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjalmsson BJ, Nordborg M, Borevitz JO, Bergelson J (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. Nature Genetics 44:212

Jang JK, Sherizen DE, Bhagat R, Manheim EA, McKim KS (2003) Relationship of DNA double-strand breaks to synapsis in *Drosophila*. Journal of Cell Science 116:3069

Jeffreys AJ, May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. 36:151

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 39:1151

Kon N, Krawchuk MD, Warren BG, Smith GR, Wahls WP (1997) Transcription factor Mts1/Mts2 (Atf1/Pcr1, Gad7/Pcr1) activates the M26 meiotic recombination hotspot in *Schizosaccharomyces pombe*. Proceedings of the National Academy of Sciences 94:13765

Liu S, Yeh C-T, Ji T, Ying K, Wu H, Tang HM, Fu Y, Nettleton D, Schnable PS (2009) Mu Transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genetics 5:e1000733

Lu P, Han X, Qi J, Yang J, Wijeratne AJ, Li T, Ma H (2012) Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg erecta and all four products of a single meiosis. Genome Research 22:508

Luykx P, Bajić IV, Khuri S (2006) NXSensor web tool for evaluating DNA for nucleosome exclusion sequences and accessibility to binding factors. Nucleic Acids Research 34:W560

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

Marais G, Charlesworth B, Wright SI (2004) Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. Genome Biology 5:1

Matys V, Fricke E, Geffers R, Gößling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos D-U, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E (2003) TRANSFAC®: transcriptional regulation, from patterns to profiles. Nucleic Acids Research 31:374

Mazón G, Mimitou EP, Symington LS (2010) Snapshot: homologous recombination in DNA double-strand break repair. Cell 142:648.e1

McMahill MS, Sham CW, Bishop DK (2007) Synthesis-dependent strand annealing in meiosis. PLoS Biololgy 5:e299

Mets DG, Meyer BJ (2009) Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. Cell 139:73

Miller DE, Takeo S, Nandanan K, Paulson A, Gogol MM, Noll AC, Perera AG, Walton KN, Gilliland WD, Li H, Staehling KK, Blumenstiel JP, Hawley RS (2012) A whole-chromosome analysis of meiotic recombination in *Drosophila melanogaster*. G3: Genes|Genomes|Genetics 2:249

Moens PB, Chen DJ, Shen Z, Kolas N, Tarsounas M, Heng HHQ, Spyropoulos B (1997) Rad51 immunocytology in rat and mouse spermatocytes and oocytes. Chromosoma 106:207

Muñoz-Fuentes V, Di Rienzo A, Vilà C (2011) Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes. PLos One 6:e25498

Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP (2009) Accelerated evolution of the PRDM9 speciation gene across diverse metazoan taxa. PLoS Genetics 5:e1000753

Osman K, Higgins JD, Sanchez-Moran E, Armstrong SJ, Franklin FCH (2011) Pathways to meiotic recombination in *Arabidopsis thaliana*. New Phytologist 190:523

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Research 18:2024

Pajerowska-Mukhtar Karolina M, Wang W, Tada Y, Oka N, Tucker Chandra L, Fonseca Jose P, Dong X (2012) The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. Current Biology 22:103

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau Hannah G, Tischfield Sam E, Zhu X, Neale Matthew J, Jasin M, Socci Nicholas D, Hochwagen A, Keeney S (2011) A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell 144:719

Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GAB (2012) Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biology and Evolution 4:787

Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A (2009) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Research

Preuss D, Rhee SY, Davis RW (1994) Tetrad analysis possible in *Arabidopsis* with mutation of the *QUARTET* (*QRT*) genes. Science 264:1458

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

Ravi M, Chan SWL (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464:615

Sanchez-Moran E, Santos J-L, Jones GH, Franklin FCH (2007) ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in *Arabidopsis*. Genes and Development 21: 2220

Schneeberger K, Hagmann J, Ossowski S, Warthmann N, Gesing S, Kohlbacher O, Weigel D (2009) Simultaneous alignment of short reads against multiple genomes. Genome Biology 10:R98

Schwacha A, Kleckner N (1995) Identification of double holliday junctions as intermediates in meiotic recombination. Cell 83:783

Schwacha A, Kleckner N (1997) Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. Cell 90:1123

Shi J, Wolf SE, Burke JM, Presting GG, Ross-Ibarra J, Dawe RK (2010) Widespread gene conversion in centromere cores. PLoS Biology 8:e1000327

Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV (2011) Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature 472:375

Stahl FW, Foss HM (2010) A two-pathway analysis of meiotic crossing over and gene conversion in *Saccharomyces cerevisiae*. Genetics 186:515

Steiner WW, Schreckhise RW, Smith GR (2002) Meiotic DNA breaks at the *S. pombe* recombination hot spot M26. Molecular Cell 9:847

Stevison L, Noor M (2010) Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. Journal of Molecular Evolution 71:332

Sturtevant AH (1925) The effects of unequal crossing over at the bar locus in *Drosophila*. Genetics 10:117

Sun Y, Ambrose JH, Haughey BS, Webster TD, Pierrie SN, Muñoz DF, Wellman EC, Cherian S, Lewis SM, Berchowitz LE, Copenhaver GP (2012) Deep genome-wide measurement of meiotic gene conversion using tetrad analysis in *Arabidopsis thaliana*. PLoS Genetics 8:e1002968

Suter B, Schnappauf G, Thoma F (2000) Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. Nucleic Acids Research 28:4083

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25

Tischfield SE, Keeney S, (2012) Scale matters: The spatial correlation of yeast meiotic DNA breaks with histone H3 trimethylation is driven largely by independent colocalization at promoters. Cell Cycle 11:1496

Vignard J, Siwiec T, Chelysheva L, Vrielynck N, Gonord F, Armstrong SJ, Schlögelhofer P, Mercier R (2007) The Interplay of RecA-related proteins and the MND1–HOP2 complex during meiosis in *Arabidopsis thaliana*. PLoS Genetics 3:e176

Wahls WP, Davidson MK (2010) Discrete DNA sites regulate global distribution of meiotic recombination. Trends in Genetics 26:202

Wahls WP, Smith GR (1994) A heteromeric protein that binds to a meiotic homologous recombination hot spot: correlation of binding and hot spot activity. Genes & Development 8:1693

Wang D, Weaver ND, Kesarwani M, Dong X (2005) Induction of protein secretory pathway is required for systemic acquired resistance. Science 308:1036

Wang YH, Griffith JD (1996) The [(G/C)3NN]n motif: a common DNA repeat that excludes nucleosomes. Proceedings of the National Academy of Sciences 93:8863

White MA, Dominska M, Petes TD (1993) Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in *Saccharomyces cerevisiae*. Proceedings of the National Academy of Sciences 90:6621

Wijnker E, van Dun K, de Snoo CB, Lelivelt CLC, Keurentjes JJB, Naharudin NS, Ravi M, Chan SWL, de Jong H, Dirks R (2012) Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. Nature Genetics 44:467

Wu T, Lichten M (1994) Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science 263:515

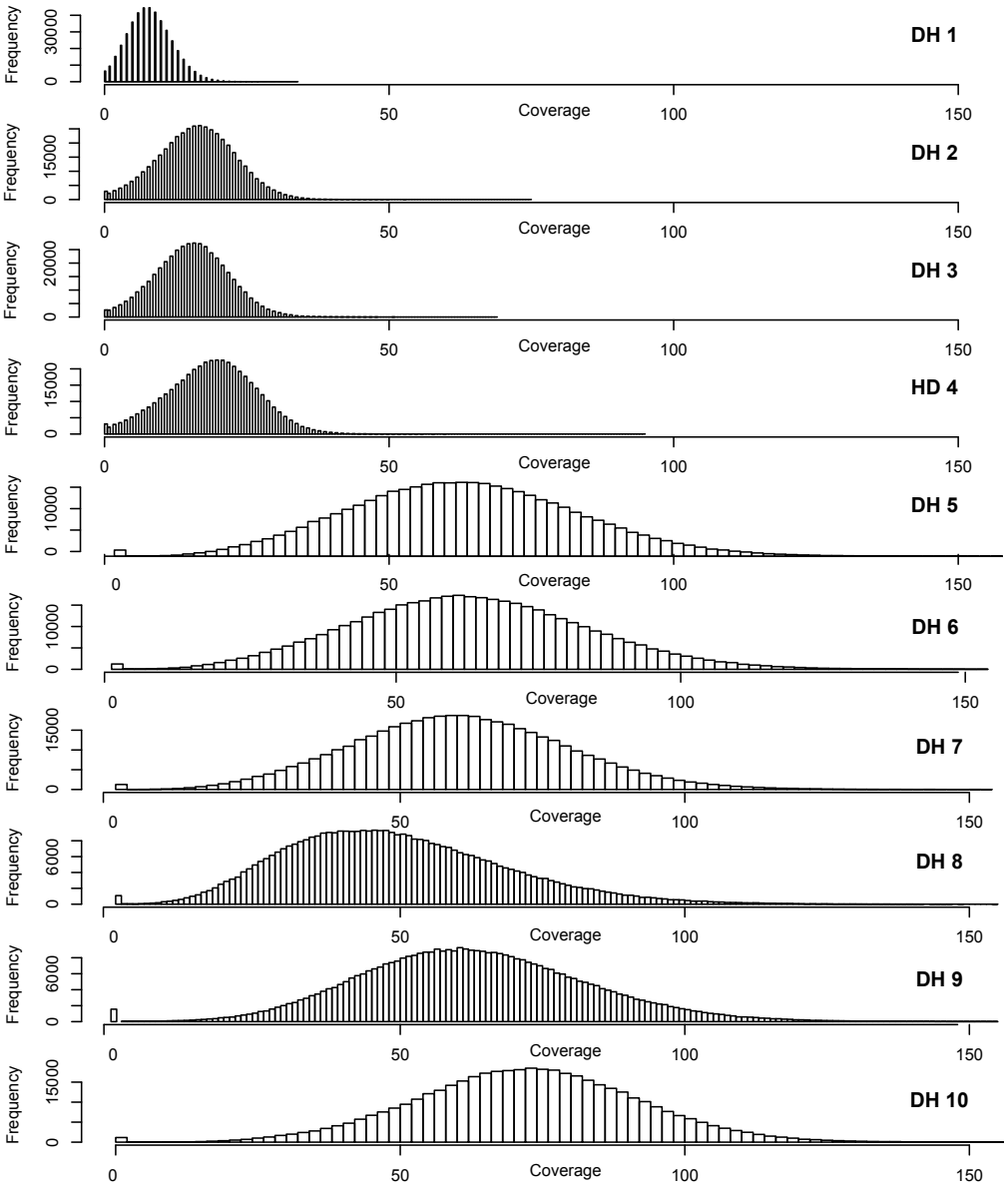
Yamada T, Mizuno K-i, Hirota K, Kon N, Wahls WP, Hartsuiker E, Murofushi H, Shibata T, Ohta K (2004) Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. EMBO Journal 23:1792

Yao H, Zhou Q, Li J, Smith H, Yandeau M, Nikolau BJ, Schnable PS (2002) Molecular characterization of meiotic recombination across the 140-kb multigenic *a1-sh2* interval of maize. Proceedings of the National Academy of Sciences of the United States of America 99:6157

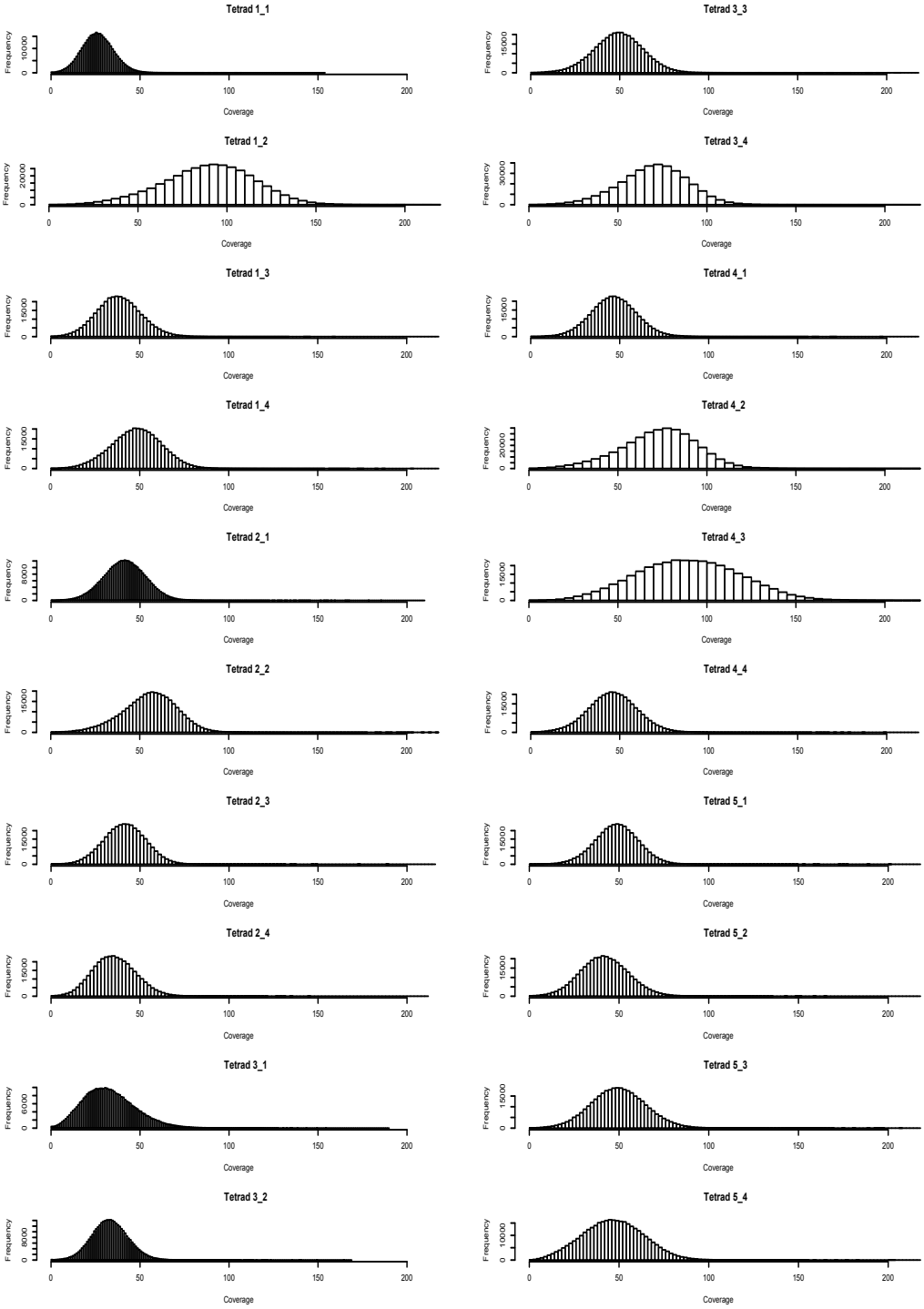
Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, Miller N, Drouaud J, Grelon M, Copenhaver GP, Mezard C, Kelly KA, Henderson IR (2012) Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. PLoS Genetics 8:e1002844

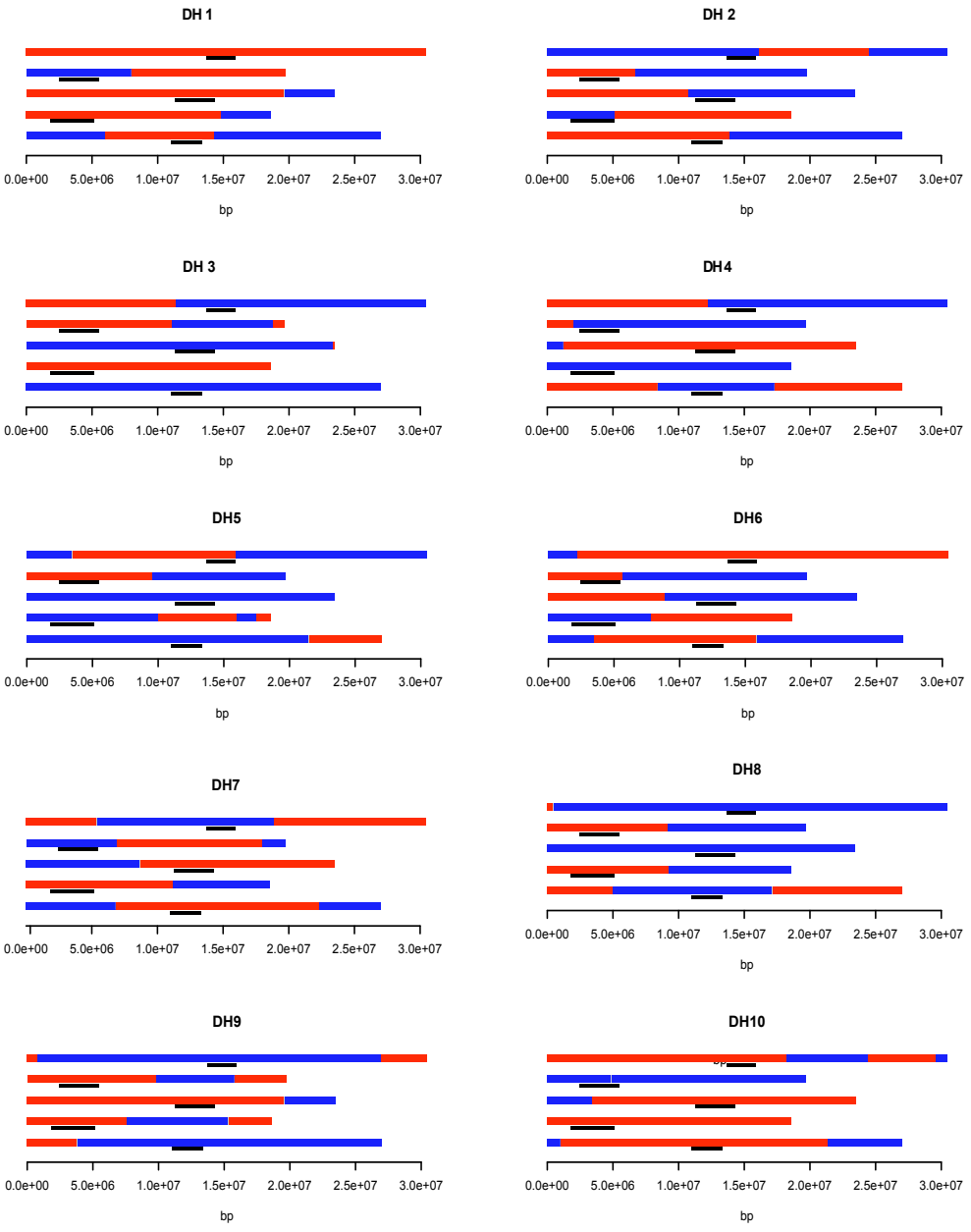
Zickler H (1934) Genetische Untersuchungen an einem heterothallischen Askomyzeten (*Bombardia lunata* nov. spec.). Planta 22:573

Supplementary Figures

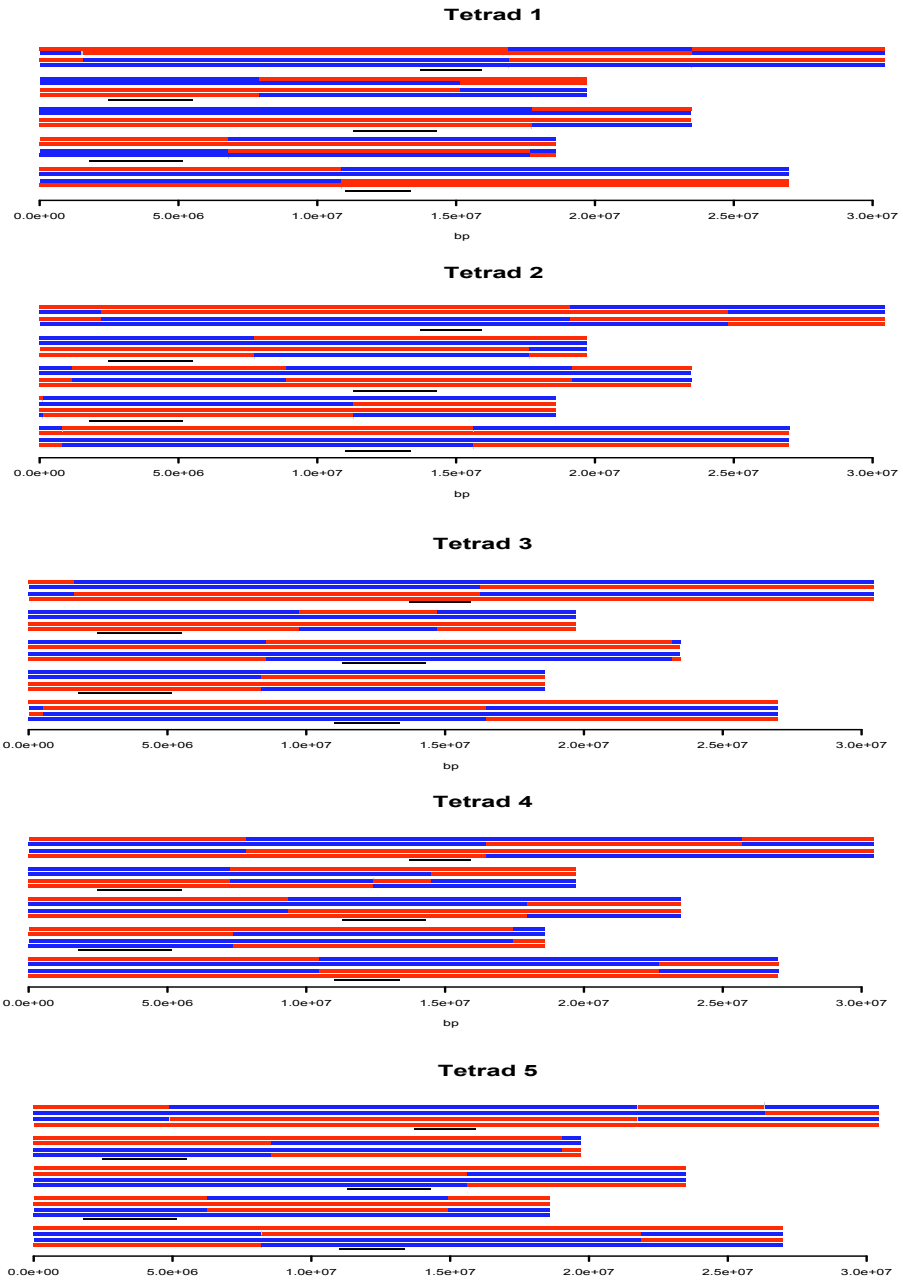


Supplementary Figure S1: Read coverage of sequenced DH lines and tetrads. Distributions shown read coverage (x-axis) and the number of reads per class.

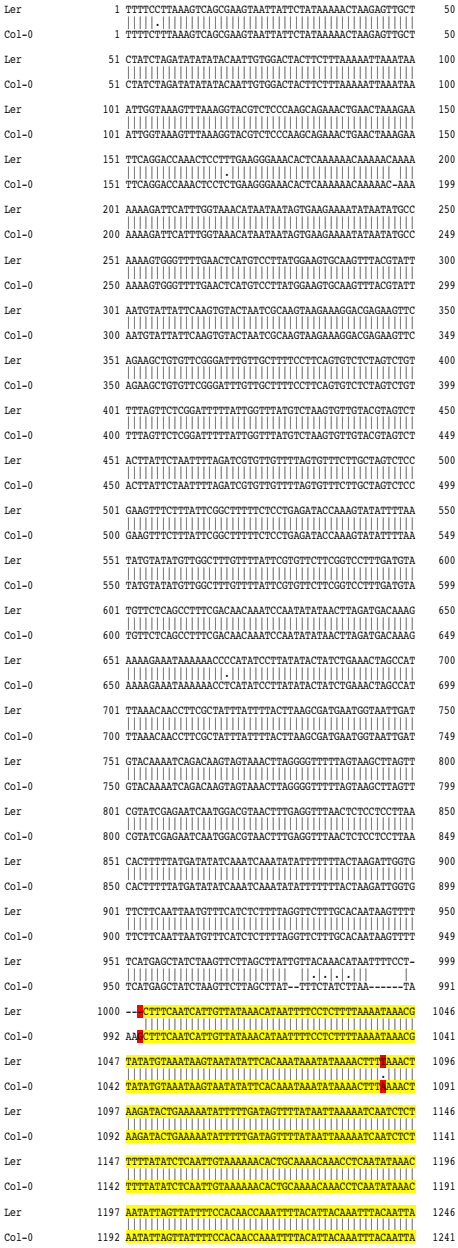

Tetrad 1_1
Tetrad 3_3
Tetrad 1_2
Tetrad 3_4
Tetrad 1_3
Tetrad 4_1
Tetrad 1_4
Tetrad 4_2
Tetrad 2_1
Tetrad 4_3
Tetrad 2_2
Tetrad 4_4
Tetrad 2_3
Tetrad 5_1
Tetrad 2_4
Tetrad 5_2
Tetrad 3_1
Tetrad 5_3
Tetrad 3_2
Tetrad 5_4
Frequency
Coverage



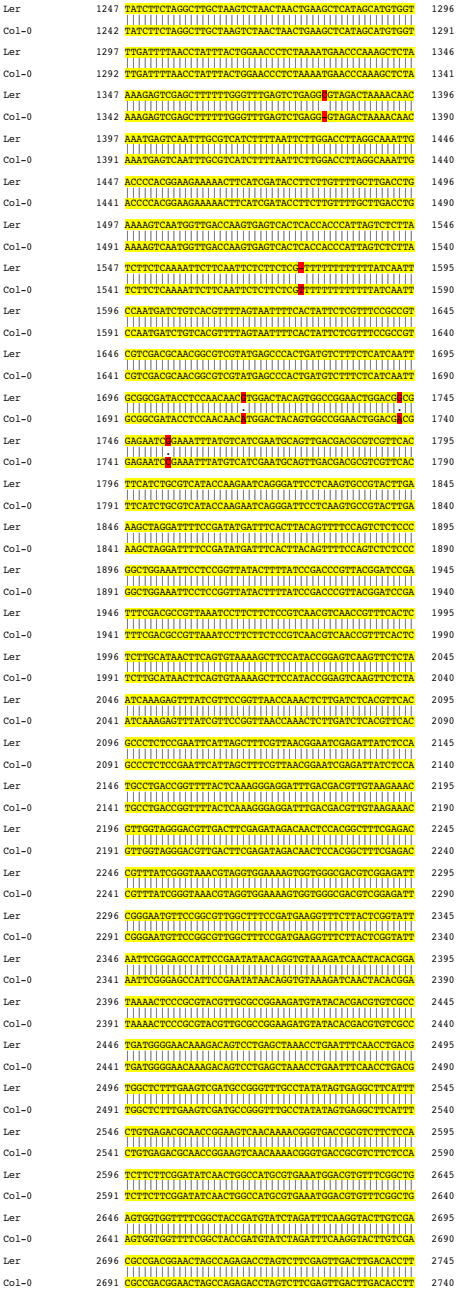
Supplementary Figure S2: Genome composition of DH offspring. Columbia sequences shown in red, Landsberg *erecta* sequences in blue. For each genome composition chromosomes are ordered from 1 (top) to 5 (bottom). Black short lines indicate centromere positions.

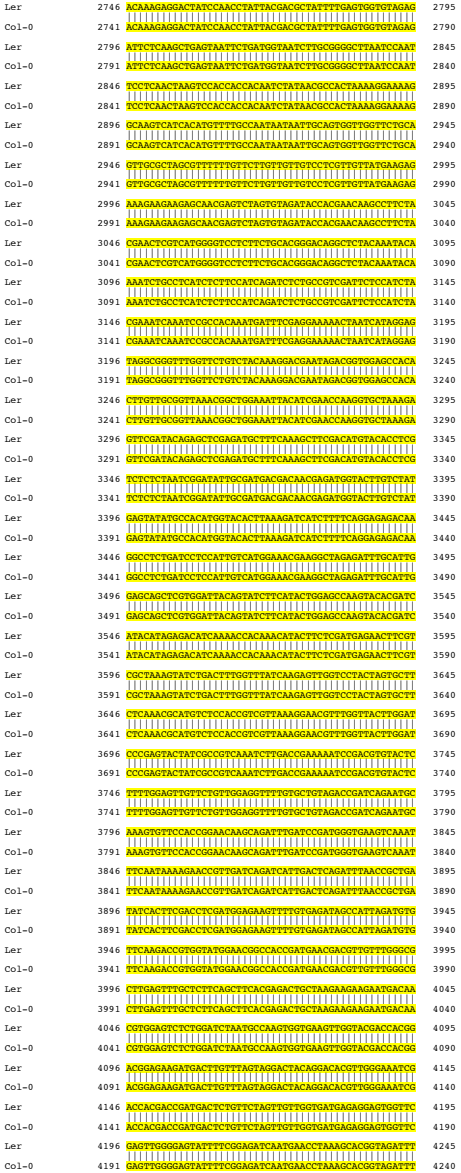


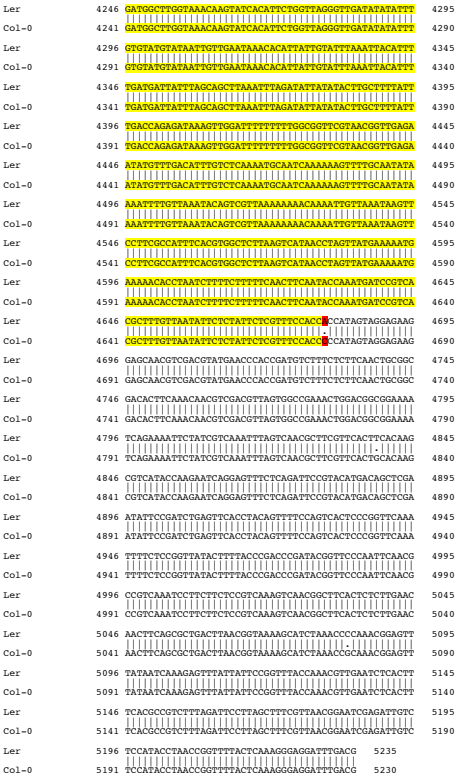
Supplementary Figure S3: Genome composition of tetrads. Note that each tetrad is represented by four spores, that are numbered 1 through 4 from top to bottom. Columbia derived sequences are shown in red, and Landsberg *erecta* sequences are shown in blue. Chromsomes are ordered from 1 (top) to 5 (bottom). Black short lines indicate centromere positions.

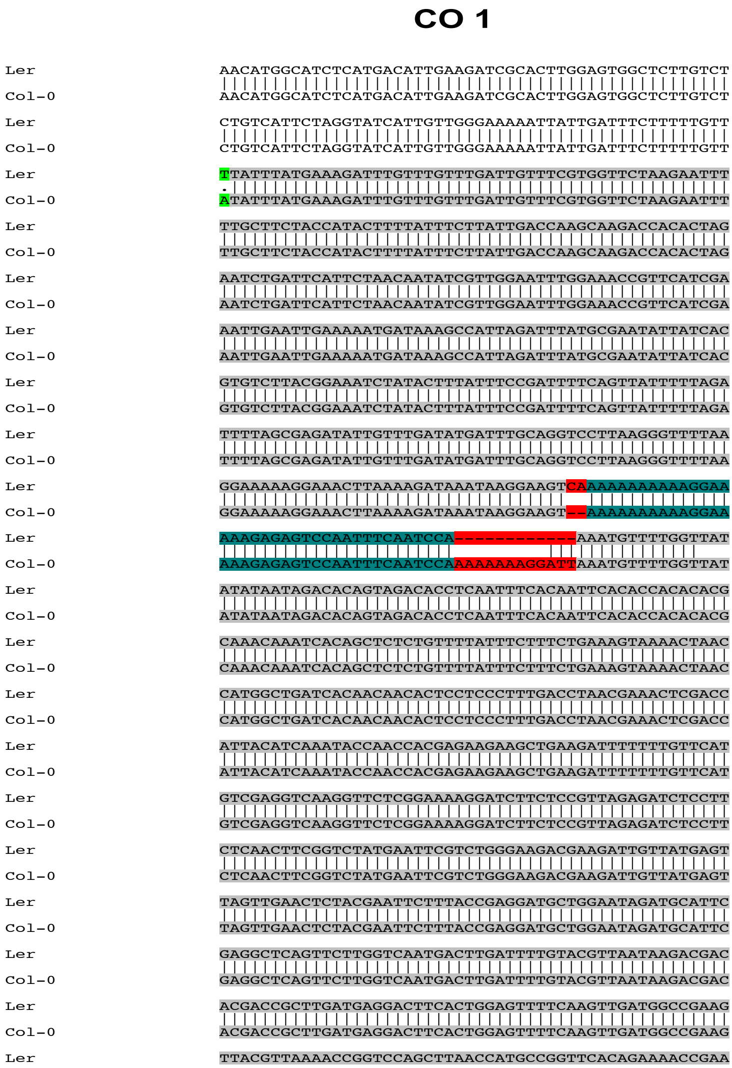


Supplementary Figure S4: Example of COCT with complex conversions. Alignment of the genomic sequence of L*er* and Col-0 of CO 21, tetrad 2 on chromosome 5: 15608224::15611905. All six polymorphisms within the conversion tract were fully converted to L*er*, except for a 1 bp deletion, that originates from the Col parent. Note that the COCT terminates near a highly polymorphic region (here at 977-1003 bp). Regions highlighted in yellow mark the maximum conversion tract. Red regions show polymorphisms within maximum conversion tract.


Ler 1247 TATCTTCTAGGCTTGCTAAGTCTAACTAACTGAAGCTCATAGCATGTGGT 1296
Col-0 1242 TATCTTCTAGGCTTGCTAAGTCTAACTAACTGAAGCTCATAGCATGTGGT 1291
Ler 1297 TTGATTTTAACCTATTTACTGGAACCCTCTAAAATGAACCCAAAGCTCTA 1346
Col-0 1292 TTGATTTTAACCTATTTACTGGAACCCTCTAAAATGAACCCAAAGCTCTA 1341
Ler 1347 AAAGAGTCGAGCTTTTTTGGGTTTGAGTCTGAGGGGTAGACTAAAACAAC 1396
Col-0 1342 AAAGAGTCGAGCTTTTTTGGGTTTGAGTCTGAGGGGTAGACTAAAACAAC 1390
Ler 1397 AAATGAGTCAATTTGCGTCATCTTTTAATTCTTGGACCTTAGGCAAATTG 1446
Col-0 1391 AAATGAGTCAATTTGCGTCATCTTTTAATTCTTGGACCTTAGGCAAATTG 1440
Ler 1447 ACCCCACGGAAGAAAAACTTCATCGATACCTTCTTGTTTTGCTTGACCTG 1496
Col-0 1441 ACCCCACGGAAGAAAAACTTCATCGATACCTTCTTGTTTTGCTTGACCTG 1490
Ler 1497 AAAAGTCAATGGTTGACCAAGTGAGTCACTCACCACCCATTAGTCTCTTA 1546
Col-0 1491 AAAAGTCAATGGTTGACCAAGTGAGTCACTCACCACCCATTAGTCTCTTA 1540
Ler 1547 TCTTCTCAAAATTCTTCAATTCTCTTCTCGTTTTTTTTTTTTTATCAATT 1595
Col-0 1541 TCTTCTCAAAATTCTTCAATTCTCTTCTCGTTTTTTTTTTTTTATCAATT 1590
Ler 1596 CCAATGATCTGTCACGTTTTAGTAATTTTCACTATTCTCGTTTCCGCCGT 1645
Col-0 1591 CCAATGATCTGTCACGTTTTAGTAATTTTCACTATTCTCGTTTCCGCCGT 1640
Ler 1646 CGTCGACGCAACGGCGTCGTATGAGCCCACTGATGTCTTTCTCATCAATT 1695
Col-0 1641 CGTCGACGCAACGGCGTCGTATGAGCCCACTGATGTCTTTCTCATCAATT 1690
Ler 1696 GCGGCGATACCTCCAACAACTTGGACTACAGTGGCCGGAACTGGACGGCG 1745
Col-0 1691 GCGGCGATACCTCCAACAACTTGGACTACAGTGGCCGGAACTGGACGGCG 1740
Ler 1746 GAGAATCTGAAATTTATGTCATCGAATGCAGTTGACGACGCGTCGTTCAC 1795
Col-0 1741 GAGAATCTGAAATTTATGTCATCGAATGCAGTTGACGACGCGTCGTTCAC 1790
Ler 1796 TTCATCTGCGTCATACCAAGAATCAGGGATTCCTCAAGTGCCGTACTTGA 1845
Col-0 1791 TTCATCTGCGTCATACCAAGAATCAGGGATTCCTCAAGTGCCGTACTTGA 1840
Ler 1846 AAGCTAGGATTTTCCGATATGATTTCACTTACAGTTTTCCAGTCTCTCCC 1895
Col-0 1841 AAGCTAGGATTTTCCGATATGATTTCACTTACAGTTTTCCAGTCTCTCCC 1890
Ler 1896 GGCTGGAAATTCCTCCGGTTATACTTTTATCCGACCCGTTACGGATCCGA 1945
Col-0 1891 GGCTGGAAATTCCTCCGGTTATACTTTTATCCGACCCGTTACGGATCCGA 1940
Ler 1946 TTTCGACGCCGTTAAATCCTTCTTCTCCGTCAACGTCAACCGTTTCACTC 1995
Col-0 1941 TTTCGACGCCGTTAAATCCTTCTTCTCCGTCAACGTCAACCGTTTCACTC 1990
Ler 1996 TCTTGCATAACTTCAGTGTAAAAGCTTCCATACCGGAGTCAAGTTCTCTA 2045
Col-0 1991 TCTTGCATAACTTCAGTGTAAAAGCTTCCATACCGGAGTCAAGTTCTCTA 2040
Ler 2046 ATCAAAGAGTTTATCGTTCCGGTTAACCAAACTCTTGATCTCACGTTCAC 2095
Col-0 2041 ATCAAAGAGTTTATCGTTCCGGTTAACCAAACTCTTGATCTCACGTTCAC 2090
Ler 2096 GCCCTCTCCGAATTCATTAGCTTTCGTTAACGGAATCGAGATTATCTCCA 2145
Col-0 2091 GCCCTCTCCGAATTCATTAGCTTTCGTTAACGGAATCGAGATTATCTCCA 2140
Ler 2146 TGCCTGACCGGTTTTACTCAAAGGGAGGATTTGACGACGTTGTAAGAAAC 2195
Col-0 2141 TGCCTGACCGGTTTTACTCAAAGGGAGGATTTGACGACGTTGTAAGAAAC 2190
Ler 2196 GTTGGTAGGGACGTTGACTTCGAGATAGACAACTCCACGGCTTCGAGAC 2245
Col-0 2191 GTTGGTAGGGACGTTGACTTCGAGATAGACAACTCCACGGCTTCGAGAC 2240
Ler 2246 CGTTTATCGGGTAAACGTAGGTGGAAAAGTGGTGGGCGACGTCGGAGATT 2295
Col-0 2241 CGTTTATCGGGTAAACGTAGGTGGAAAAGTGGTGGGCGACGTCGGAGATT 2290
Ler 2296 CGGGAATGTTCCGGCGTTGGCTTTCCGATGAAGGTTTCTTACTCGGTATT 2345
Col-0 2291 CGGGAATGTTCCGGCGTTGGCTTTCCGATGAAGGTTTCTTACTCGGTATT 2340
Ler 2346 AATTCGGGAGCCATTCCGAATATAACAGGTGTAAAGATCAACTACACGGA 2395
Col-0 2341 AATTCGGGAGCCATTCCGAATATAACAGGTGTAAAGATCAACTACACGGA 2390
Ler 2396 TAAAACTCCCGCGTACGTTGCGCCGGAAGATGTATACACGACGTGTCGCC 2445
Col-0 2391 TAAAACTCCCGCGTACGTTGCGCCGGAAGATGTATACACGACGTGTCGCC 2440
Ler 2446 TGATGGGGAACAAAGACAGTCCTGAGCTAAACCTGAATTTCAACCTGACG 2495
Col-0 2441 TGATGGGGAACAAAGACAGTCCTGAGCTAAACCTGAATTTCAACCTGACG 2490
Ler 2496 TGGCTCTTTGAAGTCGATGCCGGGTTTGCCTATATAGTGAGGCTTCATTT 2545
Col-0 2491 TGGCTCTTTGAAGTCGATGCCGGGTTTGCCTATATAGTGAGGCTTCATTT 2540
Ler 2546 CTGTGAGACGCAACCGGAAGTCAACAAAACGGGTGACCGCGTCTTCTCCA 2595
Col-0 2541 CTGTGAGACGCAACCGGAAGTCAACAAAACGGGTGACCGCGTCTTCTCCA 2590
Ler 2596 TCTTCTTCGGATATCAACTGGCCATGCGTGAAATGGACGTGTTTCGGCTG 2645
Col-0 2591 TCTTCTTCGGATATCAACTGGCCATGCGTGAAATGGACGTGTTTCGGCTG 2640
Ler 2646 AGTGGTGGTTTTCGGCTACCGATGTATCTAGATTTCAAGGTACTTGTCGA 2695
Col-0 2641 AGTGGTGGTTTTCGGCTACCGATGTATCTAGATTTCAAGGTACTTGTCGA 2690
Ler 2696 CGCCGACGGAACTAGCCAGAGACCTAGTCTTCGAGTTGACTTGACACCTT 2745
Col-0 2691 CGCCGACGGAACTAGCCAGAGACCTAGTCTTCGAGTTGACTTGACACCTT 2740

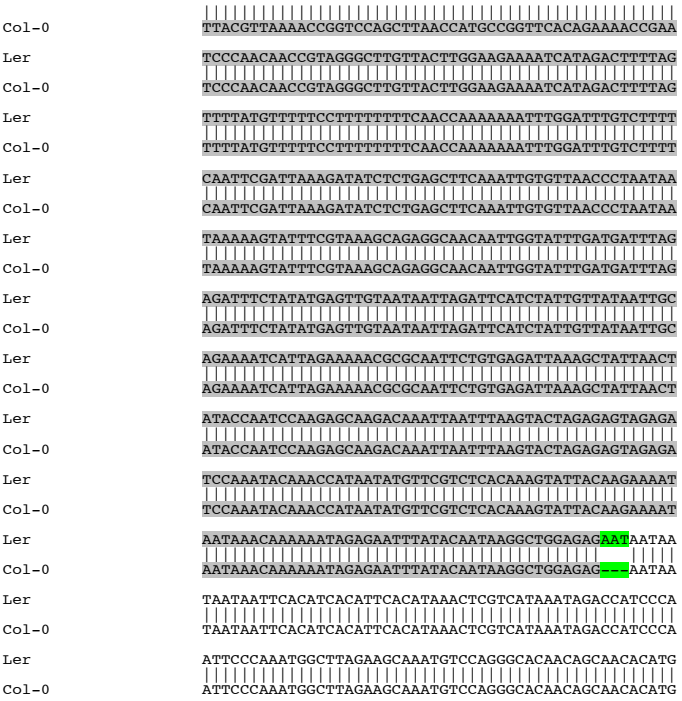

Ler 2746 ACAAAGAGGACTATCCAACCTATTACGACGCTATTTTGAGTGGTGTAGAG 2795
Col-0 2741 ACAAAGAGGACTATCCAACCTATTACGACGCTATTTTGAGTGGTGTAGAG 2790
Ler 2796 ATTCTCAAGCTGAGTAATTCTGATGGTAATCTTGCGGGGCTTAATCCAAT 2845
Col-0 2791 ATTCTCAAGCTGAGTAATTCTGATGGTAATCTTGCGGGGCTTAATCCAAT 2840
Ler 2846 TCCTCAACTAAGTCCACCACCACAATCTATAACGCCACTAAAAGGAAAAG 2895
Col-0 2841 TCCTCAACTAAGTCCACCACCACAATCTATAACGCCACTAAAAGGAAAAG 2890
Ler 2896 GCAAGTCATCACATGTTTTGCCAATAATAATTGCAGTGGTTGGTTCTGCA 2945
Col-0 2891 GCAAGTCATCACATGTTTTGCCAATAATAATTGCAGTGGTTGGTTCTGCA 2940
Ler 2946 GTTGCGCTAGCGTTTTTTGTTCTTGTTGTTGTCCTCGTTGTTATGAAGAG 2995
Col-0 2941 GTTGCGCTAGCGTTTTTTGTTCTTGTTGTTGTCCTCGTTGTTATGAAGAG 2990
Ler 2996 AAAGAAGAAGAGCAACGAGTCTAGTGTAGATACCACGAACAAGCCTTCTA 3045
Col-0 2991 AAAGAAGAAGAGCAACGAGTCTAGTGTAGATACCACGAACAAGCCTTCTA 3040
Ler 3046 CGAACTCGTCATGGGGTCCTCTTCTGCACGGGACAGGCTCTACAAATACA 3095
Col-0 3041 CGAACTCGTCATGGGGTCCTCTTCTGCACGGGACAGGCTCTACAAATACA 3090
Ler 3096 AAATCTGCCTCATCTCTTCCATCAGATCTCTGCCGTCGATTCTCCATCTA 3145
Col-0 3091 AAATCTGCCTCATCTCTTCCATCAGATCTCTGCCGTCGATTCTCCATCTA 3140
Ler 3146 CGAAATCAAATCCGCCACAAATGATTTCGAGGAAAAACTAATCATAGGAG 3195
Col-0 3141 CGAAATCAAATCCGCCACAAATGATTTCGAGGAAAAACTAATCATAGGAG 3190
Ler 3196 TAGGCGGGTTTGGTTCTGTCTACAAAGGACGAATAGACGGTGGAGCCACA 3245
Col-0 3191 TAGGCGGGTTTGGTTCTGTCTACAAAGGACGAATAGACGGTGGAGCCACA 3240
Ler 3246 CTTGTTGCGGTTAAACGGCTGGAAATTACATCGAACCAAGGTGCTAAAGA 3295
Col-0 3241 CTTGTTGCGGTTAAACGGCTGGAAATTACATCGAACCAAGGTGCTAAAGA 3290
Ler 3296 GTTCGATACAGAGCTCGAGATGCTTTCAAAGCTTCGACATGTACACCTCG 3345
Col-0 3291 GTTCGATACAGAGCTCGAGATGCTTTCAAAGCTTCGACATGTACACCTCG 3340
Ler 3346 TCTCTCTAATCGGATATTGCGATGACGACAACGAGATGGTACTTGTCTAT 3395
Col-0 3341 TCTCTCTAATCGGATATTGCGATGACGACAACGAGATGGTACTTGTCTAT 3390
Ler 3396 GAGTATATGCCACATGGTACACTTAAAGATCATCTTTTCAGGAGAGACAA 3445
Col-0 3391 GAGTATATGCCACATGGTACACTTAAAGATCATCTTTTCAGGAGAGACAA 3440
Ler 3446 GGCCTCTGATCCTCCATTGTCATGGAAACGAAGGCTAGAGATTTGCATTG 3495
Col-0 3441 GGCCTCTGATCCTCCATTGTCATGGAAACGAAGGCTAGAGATTTGCATTG 3490
Ler 3496 GAGCAGCTCGTGGATTACAGTATCTTCATACTCGGAGCCAAGTACACGATC 3545
Col-0 3491 GAGCAGCTCGTGGATTACAGTATCTTCATACTCGGAGCCAAGTACACGATC 3540
Ler 3546 ATACATAGAGACATCAAAACCACAAACATACTTCTCGATGAGAACTTCGT 3595
Col-0 3541 ATACATAGAGACATCAAAACCACAAACATACTTCTCGATGAGAACTTCGT 3590
Ler 3596 CGCTAAAGTATCTGACTTTGGTTTATCAAGAGTTGGTCCTACTAGTGCTT 3645
Col-0 3591 CGCTAAAGTATCTGACTTTGGTTTATCAAGAGTTGGTCCTACTAGTGCTT 3640
Ler 3646 CTCAAACGCATGTCTCCACCGTCGTTAAAGGAACGTTTGGTTACTTGGAT 3695
Col-0 3641 CTCAAACGCATGTCTCCACCGTCGTTAAAGGAACGTTTGGTTACTTGGAT 3690
Ler 3696 CCCGAGTACTATCGCCGTCAAATCTTGACCGAAAAATCCGACGTGTACTC 3745
Col-0 3691 CCCGAGTACTATCGCCGTCAAATCTTGACCGAAAAATCCGACGTGTACTC 3740
Ler 3746 TTTTGGAGTTGTTCTGTTGGAGGTTTTGTGCTGTAGACCGATCAGAATGC 3795
Col-0 3741 TTTTGGAGTTGTTCTGTTGGAGGTTTTGTGCTGTAGACCGATCAGAATGC 3790
Ler 3796 AAAGTGTTCCACCGGAACAAGCAGATTTGATCCGATGGGTGAAGTCAAAT 3845
Col-0 3791 AAAGTGTTCCACCGGAACAAGCAGATTTGATCCGATGGGTGAAGTCAAAT 3840
Ler 3846 TTCAATAAAAGAACCGTTGATCAGATCATTGACTCAGATTTAACCGCTGA 3895
Col-0 3841 TTCAATAAAAGAACCGTTGATCAGATCATTGACTCAGATTTAACCGCTGA 3890
Ler 3896 TATCACTTCGACCTCGATGGAGAAGTTTTGTGAGATAGCCATTAGATGTG 3945
Col-0 3891 TATCACTTCGACCTCGATGGAGAAGTTTTGTGAGATAGCCATTAGATGTG 3940
Ler 3946 TTCAAGACCGTGGTATGGAACGGCCACCGATGAACGACGTTGTTTGGGCG 3995
Col-0 3941 TTCAAGACCGTGGTATGGAACGGCCACCGATGAACGACGTTGTTTGGAGG 3990
Ler 3996 CTTGAGTTTGCTCTTCAGCTTCACGAGACTGCTAAGAAGAAGAATGACAA 4045
Col-0 3991 CTTGAGTTTGCTCTTCAGCTTCACGAGACTGCTAAGAAGAAGAATGACAA 4040
Ler 4046 CGTGGAGTCTCTGGATCTAATGCCAAGTGGTGAAGTTGGTACGACCACGG 4095
Col-0 4041 CGTGGAGTCTCTGGATCTAATGCCAAGTGGTGAAGTTGGTACGACCACGG 4090
Ler 4096 ACGGAGAAGATGACTTGTTTAGTAGGACTACAGGACACGTTGGGAAATCG 4145
Col-0 4091 ACGGAGAAGATGACTTGTTTAGTAGGACTACAGGACACGTTGGGAAATCG 4140
Ler 4146 ACCACGACCGATGACTCTGTTCTAGTTGTTGGTGATGAGAGGAGTGGTTC 4195
Col-0 4141 ACCACGACCGATGACTCTGTTCTAGTTGTTGGTGATGAGAGGAGTGGTTC 4190
Ler 4196 GAGTTGGGGAGTATTTTCGGAGATCAATGAACCTAAAGCACGGTAGATTT 4245
Col-0 4191 GAGTTGGGGAGTATTTTCGGAGATCAATGAACCTAAAGCACGGTAGATTT 4240

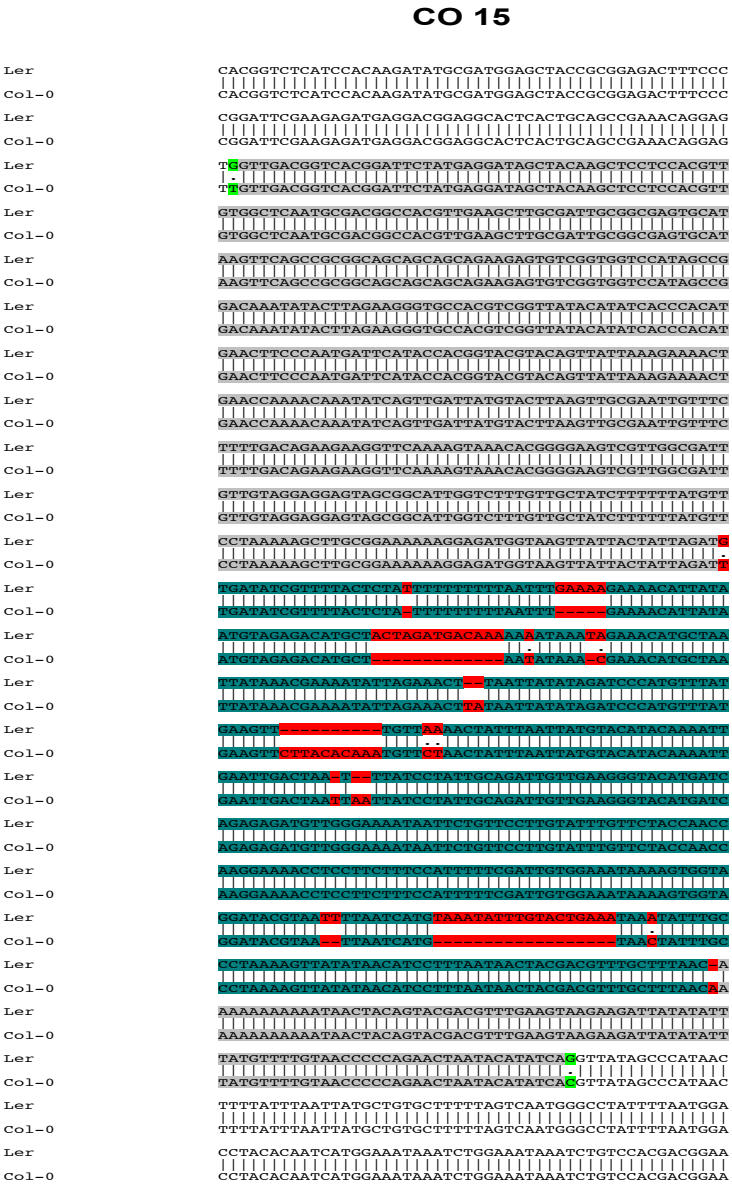

Ler 4246 GATGGCTTGGTAAACAAGTATCACATTCTGGTTAGGGTTGATATATATTT 4295
Col-0 4241 GATGGCTTGGTAAACAAGTATCACATTCTGGTTAGGGTTGATATATATTT 4290
Ler 4296 GTGTATGTATAATTGTTGAATAAACACATTATTGTATTTAAATTACATTT 4345
Col-0 4291 GTGTATGTATAATTGTTGAATAAACACATTATTGTATTTAAATTACATTT 4340
Ler 4346 TGATGATTATTTAGCAGCTTAAATTTAGATATTATATACTTGCTTTTATT 4395
Col-0 4341 TGATGATTATTTAGCAGCTTAAATTTAGATATTATATACTTGCTTTTATT 4390
Ler 4396 TGACCAGAGATAAAGTTGGATTTTTTTTTGGCGGTTCGTAACGGTTGAGA 4445
Col-0 4391 TGACCAGAGATAAAGTTGGATTTTTTTTTGGCGGTTCGTAACGGTTGAGA 4440
Ler 4446 ATATGTTTGACATTTGTCTCAAAATGCAATCAAAAAAGTTTGCAATATA 4495
Col-0 4441 ATATGTTTGACATTTGTCTCAAAATGCAATCAAAAAAGTTTGCAATATA 4490
Ler 4496 AAATTTTGTTAAATACAGTCGTTAAAAAAAACAAAATTGTTAAATAAGTT 4545
Col-0 4491 AAATTTTGTTAAATACAGTCGTTAAAAAAAACAAAATTGTTAAATAAGTT 4540
Ler 4546 CCTTCGCCATTTCACGTGGCTCTTAAGTCATAACCTAGTTATGAAAAATG 4595
Col-0 4541 CCTTCGCCATTTCACGTGGCTCTTAAGTCATAACCTAGTTATGAAAAATG 4590
Ler 4596 AAAAACACCTAATCTTTTCTTTTTCAACTTCAATACCAAATGATCCGTCA 4645
Col-0 4591 AAAAACACCTAATCTTTTCTTTTTCAACTTCAATACCAAATGATCCGTCA 4640
Ler 4646 CGCTTTGTTAATATTCTCTATTCTCGTTTCCACCA-CCATAGTAGGAGAAG 4695
Col-0 4641 CGCTTTGTTAATATTCTCTATTCTCGTTTCCACCA-CCATAGTAGGAGAAG 4690
Ler 4696 GAGCAACGTCGACGTATGAACCCACCGATGTCTTTCTCTTCAACTGCGGC 4745
Col-0 4691 GAGCAACGTCGACGTATGAACCCACCGATGTCTTTCTCTTCAACTGCGGC 4740
Ler 4746 GACACTTCAAACAACGTCGACGTTAGTGGCCGAAACTGGACGGCGGAAAA 4795
Col-0 4741 GACACTTCAAACAACGTCGACGTTAGTGGCCGAAACTGGACGGCGGAAAA 4790
Ler 4796 TCAGAAAATTCTATCGTCAAATTTAGTCAACGCTTCGTTCACTTCACAAG 4845
Col-0 4791 TCAGAAAATTCTATCGTCAAATTTAGTCAACGCTTCGTTCACTGCACAAG 4840
Ler 4846 CGTCATACCAAGAATCAGGAGTTTCTCAGATTCCGTACATGACAGCTCGA 4895
Col-0 4841 CGTCATACCAAGAATCAGGAGTTTCTCAGATTCCGTACATGACAGCTCGA 4890
Ler 4896 ATATTCCGATCTGAGTTCACCTACAGTTTTCCAGTCACTCCCGGTTCAAA 4945
Col-0 4891 ATATTCCGATCTGAGTTCACCTACAGTTTTCCAGTCACTCCCGGTTCAAA 4940
Ler 4946 TTTTCTCCGGTTATACTTTTACCCGACCCGATACGGTTCCCAATTCAACG 4995
Col-0 4941 TTTTCTCCGGTTATACTTTTACCCGACCCGATACGGTTCCCAATTCAACG 4990
Ler 4996 CCGTCAAATCCTTCTTCTCCGTCAAAGTCAACGGCTTCACTCTCTTGAAC 5045
Col-0 4991 CCGTCAAATCCTTCTTCTCCGTCAAAGTCAACGGCTTCACTCTCTTGAAC 5040
Ler 5046 AACTTCAGCGCTGACTTAACGGTAAAAGCATCTAAACCCCAAACGGAGTT 5095
Col-0 5041 AACTTCAGCGCTGACTTAACGGTAAAAGCATCTAAACCGCAAACGGAGTT 5090
Ler 5096 TATAATCAAAGAGTTTATTATTCCGGTTTACCAAACGTTGAATCTCACTT 5145
Col-0 5091 TATAATCAAAGAGTTTATTATTCCGGTTTACCAAACGTTGAATCTCACTT 5140
Ler 5146 TCACGCCGTCTTTAGATTCCTTAGCTTTCGTTAACGGAATCGAGATTGTC 5195
Col-0 5141 TCACGCCGTCTTTAGATTCCTTAGCTTTCGTTAACGGAATCGAGATTGTC 5190
Ler 5196 TCCATACCTAACCGGTTTTACTCAAAGGGAGGATTTGACG 5235
Col-0 5191 TCCATACCTAACCGGTTTTACTCAAAGGGAGGATTTGACG 5230

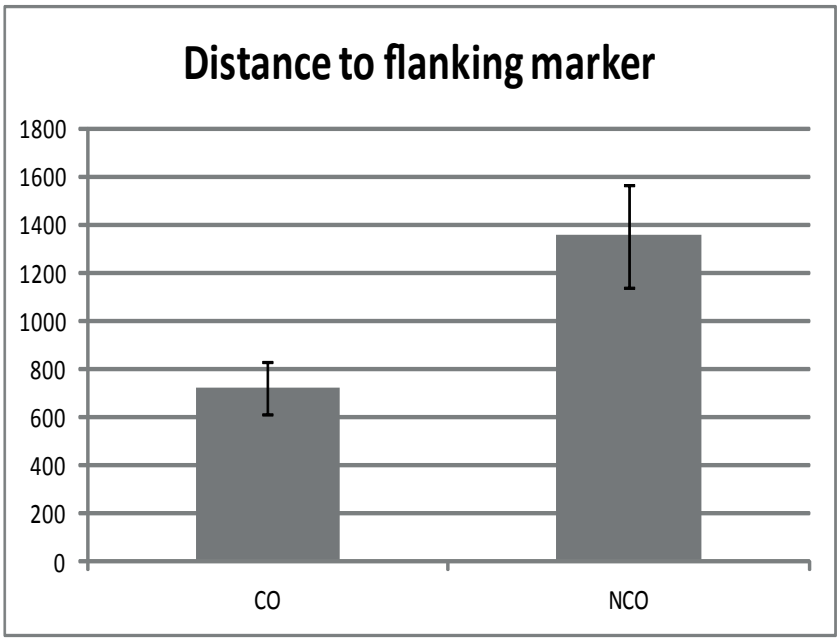


Supplementary Figure S5: Alignment of two COCTs covering complex genomic regions. Pairwise alignments of the genome sequence of L*er* and Col-0 of CO 1 (tetrad 1, chr. 1). This conversion tract shows full conversion to L*er* alleles. CO 15 (tetrad 2, chr. 3) shows full conversion to Col alleles.

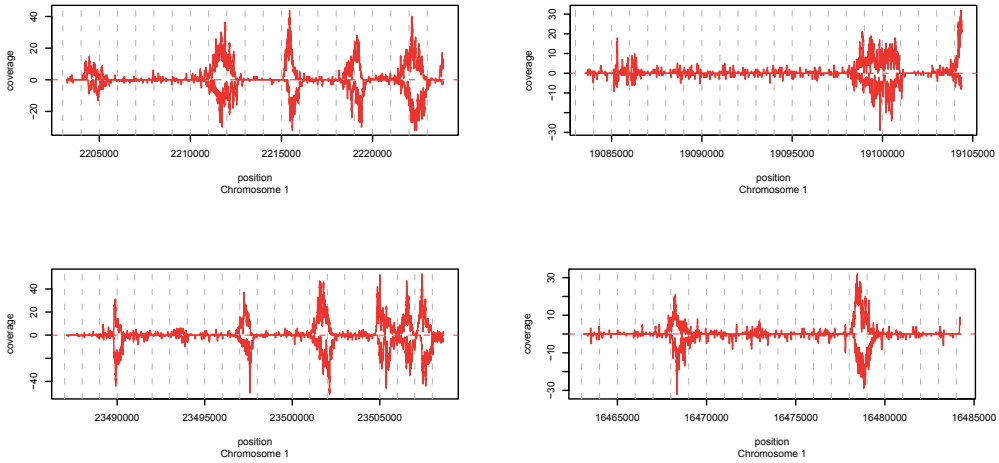
The COCT is highlighted in dark blue for conserved regions and in red for genomic differences between Col and L*er.* The region between the converted polymorphisms and the flanking markers are highlighted in grey. The flanking polymorphisms are highlighted in green. Both alignments show stark sequence divergence in the COCT.


Col-0 TTACGTTAAAACCGGTCCAGCTTAACCATGCCGGTTCACAGAAAACCGAA
Ler TCCCAACAACCGTAGGGCTTGTTACTTGGAAGAAAATCATAGACTTTTAG
Col-0 TCCCAACAACCGTAGGGCTTGTTACTTGGAAGAAAATCATAGACTTTTAG
Ler TTTTATGTTTTTCCTTTTTTTTCAACCAAAAAAATTTGGATTTGTCTTTT
Col-0 TTTTATGTTTTTCCTTTTTTTTCAACCAAAAAAATTTGGATTTGTCTTTT
Ler CAATTCGATTAAAGATATCTCTGAGCTTCAAATTGTGTTAACCCTAATAA
Col-0 CAATTCGATTAAAGATATCTCTGAGCTTCAAATTGTGTTAACCCTAATAA
Ler TAAAAAGTATTTCGTAAAGCAGAGGCAACAATTGGTATTTGATGATTTAG
Col-0 TAAAAAGTATTTCGTAAAGCAGAGGCAACAATTGGTATTTGATGATTTAG
Ler AGATTTCTATATGAGTTGTAATAATTAGATTCATCTATTGTTATAATTGC
Col-0 AGATTTCTATATGAGTTGTAATAATTAGATTCATCTATTGTTATAATTGC
Ler AGAAAATCATTAGAAAAACGCGCAATTCTGTGAGATTAAAGCTATTAACT
Col-0 AGAAAATCATTAGAAAAACGCGCAATTCTGTGAGATTAAAGCTATTAACT
Ler ATACCAATCCAAGAGCAAGACAAATTAATTTAAGTACTAGAGAGTAGAGA
Col-0 ATACCAATCCAAGAGCAAGACAAATTAATTTAAGTACTAGAGAGTAGAGA
Ler TCCAAATACAAACCATAATATGTTCGTCTCACAAAGTATTACAAGAAAAT
Col-0 TCCAAATACAAACCATAATATGTTCGTCTCACAAAGTATTACAAGAAAAT
Ler AATAAACAAAAAATAGAGAATTTATACAATAAGGCTGGAGAGAATAATAA
Col-0 AATAAACAAAAAATAGAGAATTTATACAATAAGGCTGGAGAG---AATAA
Ler TAATAATTCACATCACATTCACATAAACTCGTCATAAATAGACCATCCCA
Col-0 TAATAATTCACATCACATTCACATAAACTCGTCATAAATAGACCATCCCA
Ler ATTCCCAAATGGCTTAGAAGCAAATGTCCAGGGCACAACAGCAACACATG
Col-0 ATTCCCAAATGGCTTAGAAGCAAATGTCCAGGGCACAACAGCAACACATG

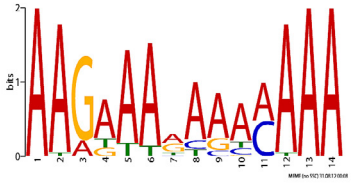

CO 15
Ler
Col-0



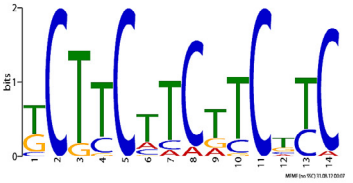
Supplementary Figure S6: Distance to flanking SNPs for COs and NCOs. The average distance between converted markers and flanking markers in CO-associated GCs (left) and NCOs (right) are shown in bp. The distance from a converted marker by an NCO to the nearest flanking marker is longer than those for CO-associated GCs. Error bars give standard error.



Supplementary Figure 7: Crossover positions in relation to H3K4me3 trimethylation. Each graph shows 10 Kb windows in which the CO is positioned in the middle. H3K4me3 trimethylation enrichment is shown in red. The graphs show representative images, in which H3K4 methylation does (left) or does not (right) coincide with CO-positions.



TGTCGTTCGT AAGAAAGAAAAAAA AGCAACCTAA
TTGCGTGAAA AACAAAGAAAAAAA CATATTAGAA
AAGAAGAAAC AACAAAAAAAAAAA CACATAGTAT
TCCTACAAAC AACAAAGCAACAAA GGTGACATCA
TTCTTTTTAA AAGGAAGAAAAAAA AATCAGAAAA
CCACTCACGT AAGAAAAAGACAAA TGAAGAAAGT
ATAACCTTGA AAGAAAAAGAAAAA GGATAT
TAATGTTTTG AAGGAAAAGACAAA ATCTATCCCT
AACCAAAACC AAGAAAAAATCAAA CCAAAACTCA
AATTTGCATA AAGTAAGAACCAAA CTAGAGGAGC
CTAATTGGAA AAGGAAAACACAAA TTGACCGACT
AACAAACTTT AAGAAAAAATAAAA AAAATTGGAC
CAGAAAATCC AAAAAAAAAACAAA GTCAGTTTCT
ATTTAAGTTA AAGAAAGAAGAAAA ACGATCACTA
TTTTTTAAAA AAGGAAACAAAAAA ACTTAGAAGT
GAACATAAAA AAGATAGCAACAAA GACCAATGCC
TTTTTTTCTG AAGTAAACAACAAA ATTAAAACTA
AAAGATAAAT AAGGAAGTAAAAAA AAAAGGAAAA
TTTTTTGCTA AAGAAATAGCAAAA GCTCCAAATG
CCAATCTGAA AAGAATTAAACAAA CAA
TTCCCACTGA AAGATTGAAACAAA GAAACATTTT
AACTCCAGAT AAAAAACAAACAAA ACTGCGCATT
TTATTATAGA AAAATAGAAACAAA AGCTTACATT
TTTGACTATT AAGTAATAACCAAA AAGAAAAAGT
GAAATGTTTT AAGAATGAATAAAA TTATTCCTGT
TTTGGGGAAA AAGTAAAAGTAAAA CTTGTGGCTG
CCCAAAAATC AAAAAACAGAAAAA TTACCAAACC
ACACAAAAGC ATGAAAAACAAAAA AAAACGAGAT
AAGCTTTATT AAGATAGAAAATAA GCTAAGAACT
GACAGTAACT AAGAAATTCCCAAA CTTGGAATCC
GTGAAACAAT AAATAAATAAAAAA TCCCCTTTAT



ACATCATCGG GCTTCTTCTTCTTC GGCTCCTCAG
TCGTCCCATT TCTTCTTCTTCTTC CTATAATCTA
ATCTTAATTC TCTTCTTCTTCTCC ATATTTTTTG
CTCAAACTCT TCTTCTTCATCTCC TAGAGTGTTT
ACCCTACCAA TCTCCTTCTTCCTC CTCCTCCTCC
GGGGAAAATA TCTTCTTCTTCATC TGTAACAAAC
TGTTAGATAG TCTTCTCCTTCGCC ACGGAACACA
GGCTGGGGAG GCTCCCTCGTCTCC CATAGCTACA
CTCAATCCAA GCTTCATCTTCCTC GAGGAACACT
TTTGTAATAA TCTTCTCCTCCTTC CGACAATCTC
TGGCTCTCCG GCTTCCTCATCGCC GATTTCTTCC
GCCGTTAAAT CCTTCTTCTCCGTC AACGTCAACC
GAACTTGGAT GCTGCTTCGTCTTC AGGGCACGTT
AATGGGCAAA TCTTCATCGTCCTC AAAGATAATC
ATTTTTGGAA GCTTCTTCGGCTTC CAGTACAATC
ATGCAATCAT TCGCCGTCTTCTTC CACAATTGTG
GGAGCTGCTG GCTCCACCTCCGCC ATCCTGAGGC
AAAAAAAAAA TCGTCTTATTCTTC TAAGATGATT
ATGGATGAGA TCTCCTACATCGTC TGGCAACTTC
AATAAAAGAT GCGTCTTAGTCTTC AGACATAACT
TTTTGTTTTG CCTTCCTCTCCACC TTAATCTCCA
ATTGGATGCA TCTTCAACATCTTC TTCCATTGAT
TTCCAACATG TCTTCTTCTTCACA GACACTTTCG

Supplementary Figure 8: Alignments of CO-sequences with consensus sequence. Sequences of COCTs are compared to the identified common motifs identified using MEME software. The poly-A like motif is shown on the left and the CTT-like motif is shown on the right.

Supplementary Tables

Supplementary Table S1: Read number and alignments.

Sample	Sample type	Sequenced read pairs	Read_ length	Aligned read pairs used*	Alignment target	Avg. nuclear genome coverage
Col	Parent	12940490	151	12599414	TAIR10	16.6
Ler	Parent	77936067	101	63518844	TAIR10	101.2
DH 1	Double Haploid	23403771	151	4336637	TAIR10	8.7
DH 2	Double haploid	24654019	151	8552413	TAIR10	18.2
DH 3	Double haploid	20065118	151	7961734	TAIR10	17
DH 4	Double haploid	21516922	151	9889499	TAIR10	21.1
DH5	Double haploid	66952956	101	49330593	TAIR10	71.1
DH6	Double haploid	55917926	101	47162497	TAIR10	71.8
DH7	Double haploid	52135752	101	45232912	TAIR10	68.5
DH8	Double haploid	49620686	101	38441915	TAIR10	58
DH9	Double haploid	55399712	101	46187638	TAIR10	70.7
DH10	Double haploid	65204463	101	53859721	TAIR10	79.6

Supplementary Table S1 – continued: Read number and alignments.

Sample	Sample type	Sequenced read pairs	Read_ length	Aligned read pairs used*	Alignment target	Avg. nuclear genome coverage
Col_tetrad	Parent	52078668	101	43303695	TAIR10	56.5
Ler_tetrad	Parent	46554725	101	34207322	TAIR10	49.4
Cvi_tetrad	Parent	56429489	101	41212019	TAIR10	63.2
1_1	tetrad	31566060	101	19820222	TAIR10	28.5
1_2	tetrad	28529159	101	22237059	TAIR10	41.0
1_3	tetrad	96662553	101	66079176	TAIR10	96.5
1_4	tetrad	53379191	101	35360375	TAIR10	51.9
2_1	tetrad	63055319	101	40605530	TAIR10	59.6
2_2	tetrad	37908618	101	25276663	TAIR10	38.6
2_3	tetrad	45485638	101	27041540	TAIR10	39.7
2_4	tetrad	48428343	101	32847111	TAIR10	47.8
3_1	tetrad	39470042	101	27004256	TAIR10	36.9
3_2	tetrad	48073711	101	35452303	TAIR10	52.0
3_3	tetrad	37102772	101	25158766	TAIR10	36.7
3_4	tetrad	76672532	101	53138435	TAIR10	75.6
4_1	tetrad	74332245	101	53442589	TAIR10	79.4
4_2	tetrad	45843587	101	34104496	TAIR10	49.4
4_3	tetrad	43763556	101	33451664	TAIR10	49.4
4_4	tetrad	1,02E+08	101	67772426	TAIR10	98.4
5_1	tetrad	50565207	101	34967278	TAIR10	51.6
5_2	tetrad	51571962	101	38639574	TAIR10	53.4
5_3	tetrad	45271453	101	31677562	TAIR10	46.4
5_4	tetrad	50373376	101	37322244	TAIR10	53.9

Supplementary Table S2: Markers used in DH analysis.

ID	Markers genotyped	Total number of markers used for calling NCO/CO	Number of markers used for calling NCO/CO			
			Col Background	L*er* Background	Ambiguous	Non-informative
DH 1	438915	406223	295708	110515	584	32108
DH 2	438915	426102	156412	269690	3290	9523
DH 3	438915	418899	199729	219170	9228	10788
DH 4	438915	428542	183681	244861	1190	9183
DH 5	438915	416177	117433	298744	19869	2869
DH 6	438915	426699	231652	195047	9453	2763
DH 7	438915	424753	266647	158106	11307	2855
DH 8	438915	420545	133381	287164	16136	2234
DH 9	438915	427817	188967	238850	8420	2678
DH 10	438915	421480	293546	127934	15165	2270

Supplementary Table S3: Markers used in tetrad analysis.

Sample	Markers genotyped (1)	Markers used for calling NCO/ CO (4)
1_1	354977	54694
1_2	354977	124721
1_3	354977	258500
1_4	354977	193783
2_1	354977	226621
2_2	354977	114762
2_3	354977	124921
2_4	354977	180583
3_1	354977	84909
3_2	354977	187477
3_3	354977	96861
3_4	354977	246457
4_1	354977	247892
4_2	354977	173739
4_3	354977	178677
4_4	354977	257029
5_1	354977	196294
5_2	354977	184519
5_3	354977	153040
5_4	354977	195475

Supplementary Table S4: Crossing over sites and their positions in DH offspring.

Sample	CO-ID	Chr	Begin	End
DH 1	16	Chr2	7995651	7996021
DH 1	17	Chr3	19659682	19661374
DH 1	18	Chr4	14820839	14821272
DH 1	19	Chr5	6045145	6045248
DH 1	20	Chr5	14294240	14294466
DH 2	1	Chr1	16130894	16131373
DH 2	2	Chr1	24497411	24498026
DH 2	3	Chr2	6694469	6694810
DH 2	4	Chr3	10770271	10770691
DH 2	5	Chr4	5165967	5166637
DH 2	6	Chr5	13901059	13901168
DH 3	7	Chr1	11379900	11380518
DH 3	8	Chr2	11095697	11095784
DH 3	9	Chr2	18809685	18809986
DH 3	10	Chr3	23360434	23361259
DH 4	11	Chr1	12256399	12256672
DH 4	12	Chr2	1994684	1995104
DH 4	13	Chr3	1230066	1230404
DH 4	14	Chr5	8408206	8408326
DH 4	15	Chr5	17302925	17303130
DH 5	21	Chr1	3473648	3474347
DH 5	22	Chr1	15922557	15922622
DH 5	23	Chr2	9579983	9580578
DH 5	24	Chr4	10007430	10007762
DH 5	25	Chr4	15988038	15991333
DH 5	26	Chr4	17512230	17512537
DH 5	27	Chr5	21481654	21482180

DH 6	28	Chr1	2245324	2246148
DH 6	29	Chr2	5660821	5661013
DH 6	30	Chr3	8880723	8880778
DH 6	31	Chr4	7837100	7837675
DH 6	32	Chr5	3480602	3480742
DH 6	33	Chr5	15878487	15879014
DH 7	34	Chr1	5407433	5408280
DH 7	35	Chr1	18914464	18915314
DH 7	36	Chr2	6947299	6947738
DH 7	37	Chr2	17981163	17982103
DH 7	38	Chr3	8686510	8687570
DH 7	39	Chr4	11220881	11222041
DH 7	40	Chr5	6832134	6832292
DH 7	41	Chr5	22337799	22337975
DH 8	42	Chr1	469655	473429
DH 8	43	Chr2	9171354	9171524
DH 8	44	Chr4	9240644	9241336
DH 8	45	Chr5	4979856	4980118
DH 8	46	Chr5	17157500	17158113
DH 9	47	Chr1	772303	772458
DH 9	48	Chr1	26948393	26949224
DH 9	49	Chr2	9837379	9838028
DH 9	50	Chr2	15787110	15787184
DH 9	51	Chr3	19595156	19595927
DH 9	52	Chr4	7574005	7574566
DH 9	53	Chr4	15314953	15321002
DH 9	54	Chr5	3824133	3824879
DH 10	55	Chr1	18216192	18216483
DH 10	56	Chr1	24451497	24453758
DH 10	57	Chr1	29578939	29579904
DH 10	58	Chr3	3427419	3427867
DH 10	59	Chr5	1010853	1010912
DH 10	60	Chr5	21377456	21378060

Supplementary Table S5: Crossing over sites and the positions of their accompanying COCT identified in the tetrad samples. Chromosomal locations refer to the positions in the reference sequence. CO-ID is a unique identifier for each CO event, which allows cross-referencing between other supplemental tables and figures. Maximal, midpoint and minimal length refer to three different approximations of the real conversion tract length (see main text). COCT describes the regions of the converted polymorphisms.

CO-ID	Gamete 1	Gamete 2	Chr	Flanking marker up-stream	Flanking marker down-stream	Maximal length	Midpoint length	Minimal length	COCT begin	COCT end	COCT length
1	1_3	1_2	Chr1	1557671	1559010	1338	380	48	1558004	1558051	48
2	1_3	1_1	Chr1	16895874	16896222	347	173	0	-	-	-
3	1_1	1_2	Chr1	23497124	23498622	1497	559	1	23498063	23498063	1
4	1_1	1_4	Chr2	7918187	7918663	475	237	0	-	-	-
5	1_3	1_2	Chr2	15154598	15156880	2281	676	1	15156204	15156204	1
6	1_1	1_4	Chr3	17728032	17728656	623	311	0	-	-	-
7	1_3	1_1	Chr4	6796827	6797695	867	433	0	-	-	-
8	1_3	1_4	Chr4	17653068	17653474	405	202	0	-	-	-
9	1_3	1_1	Chr5	10876774	10877520	745	398	119	10877054	10877172	119
10	2_3	2_2	Chr1	2213217	2213877	659	411	345	2213284	2213628	345
11	2_3	2_1	Chr1	19093540	19094388	847	604	467	19093678	19094144	467
12	2_2	2_4	Chr1	24809419	24810980	1560	1316	1229	24809664	24810892	1229
13	2_4	2_1	Chr2	7726474	7727253	778	389	0	-	-	-
14	2_4	2_3	Chr2	17645816	17647565	1748	611	1	17646954	17646954	1
15	2_1	2_3	Chr3	1160294	1161237	942	495	410	1160742	1161151	410
16	2_3	2_1	Chr3	8880855	8882075	1219	389	84	8881161	8881244	84
17	2_1	2_3	Chr3	19163980	19164563	582	291	0	-	-	-
18	2_4	2_1	Chr4	118032	118932	899	449	0	-	-	-

19	2_4	2_2	Chr4	11295554	11296431	876	438	0	-	-	-
20	2_1	2_4	Chr5	809355	810388	1032	481	404	809907	810310	404
21	2_1	2_4	Chr5	15608224	15611905	3680	754	663	15608316	15608978	663
22	3_1	3_3	Chr1	1643314	1645247	1932	922	65	1644172	1644236	65
23	3_2	3_3	Chr1	16271774	16273478	1703	887	539	16272591	16273129	539
24	3_4	3_1	Chr2	9739200	9740314	1113	556	0	-	-	-
25	3_4	3_1	Chr2	14700261	14701311	1049	335	4	14700976	14700979	4
26	3_4	3_1	Chr3	8557068	8558265	1196	633	569	8557632	8558200	569
27	3_4	3_1	Chr3	23161397	23162155	757	264	1	23161891	23161891	1
28	3_4	3_2	Chr4	8365173	8365440	266	133	0	-	-	-
29	3_2	3_3	Chr5	532301	532938	636	318	0	-	-	-
30	3_4	3_2	Chr5	16501638	16502018	379	189	0	-	-	-
31	4_3	4_1	Chr1	7842859	7843292	432	122	42	7843170	7843211	42
32	4_2	4_4	Chr1	16473123	16474232	1108	531	124	16473701	16473824	124
33	4_2	4_1	Chr1	25676514	25679256	2741	1370	0	-	-	-
34	4_3	4_1	Chr2	7252497	7253486	988	520	136	7252882	7253017	136
35	4_4	4_3	Chr2	12414849	12418705	3855	793	1	12417912	12417912	1
36	4_2	4_3	Chr2	14495108	14495835	726	363	0	-	-	-
37	4_1	4_3	Chr3	9335488	9337106	1617	616	141	9336490	9336630	141
38	4_4	4_2	Chr3	17951154	17951522	367	183	0	-	-	-
39	4_4	4_2	Chr4	7369829	7370740	910	455	0	-	-	-
40	4_1	4_3	Chr4	17463448	17464290	841	264	115	17464026	17464140	115
41	4_1	4_3	Chr5	10486315	10486602	286	54	1	10486369	10486369	1
42	4_2	4_3	Chr5	22704512	22704994	481	240	0	-	-	-
43	5_1	5_3	Chr1	4899961	4900846	884	442	0	-	-	-

44	5_1	5_3	Chr1	21763389	21764082	692	301	210	21763781	21763990	210
45	5_2	5_1	Chr1	26328821	26328949	127	63	0	-	-	-
46	5_1	5_4	Chr2	8548811	8549573	761	477	240	8549049	8549288	240
47	5_2	5_3	Chr2	19020888	19030799	9910	71	1	19020959	19020959	1
48	5_1	5_4	Chr3	15615636	15616191	554	427	324	15615740	15616063	324
49	5_2	5_3	Chr4	6257906	6260351	2444	1369	432	6258844	6259275	432
50	5_2	5_3	Chr4	14919739	14923965	4225	301	1	14920040	14920040	1
51	5_1	5_4	Chr5	8194916	8195394	477	229	1	8195145	8195145	1
52	5_1	5_3	Chr5	21896525	21897982	1456	406	284	21896648	21896931	284

Supplementary Table S6: Polymorphisms of the CO- associated GCs identified in the tetrad samples. Chromosomal locations of the genomic differences refer to the positions in the reference sequence. Polymorphisms that reside in conversion tracts, but nevertheless are inherited in a 2:2 fashion are highlighted in grey, which we observed in 2 CO-events. CO-ID is a unique identifier for each CO event, which allows cross-referencing between other supplemental tables.

CO-ID	Chr	Begin	End	Col-0 allele	Ler allele	Parental allele in gamete 1	Parental allele in gamete 2
1	Chr1	1558004	1558005	--	CA	Col-0	Col-0
1	Chr1	1558042	1558053	AAAAAAAGGATT	----------	Col-0	Col-0
3	Chr1	23498063	23498063	A	0	Col-0	Col-0
5	Chr2	15156204	15156204	T	A	Col-0	Col-0
9	Chr5	10877054	10877054	A	G	Col-0	Col-0
9	Chr5	10877169	10877169	G	T	Col-0	Col-0
9	Chr5	10877171	10877172	-	T	Col-0	Col-0
10	Chr1	2213284	2213284	A	T	Ler	Ler
10	Chr1	2213614	2213614	G	C	Ler	Ler
10	Chr1	2213628	2213628	T	C	Ler	Ler
11	Chr1	19093678	19093678	A	C	Col-0	Col-0
11	Chr1	19093683	19093683	C	G	Col-0	Col-0
11	Chr1	19093944	19093945	-	T	Col-0	Col-0
11	Chr1	19093986	19093986	G	C	Col-0	Col-0
11	Chr1	19094144	19094144	C	T	Col-0	Col-0
12	Chr1	24809664	24809664	T	C	Ler	Ler
12	Chr1	24809805	24809805	A	T	Ler	Ler
12	Chr1	24809909	24809909	T	C	Ler	Ler
12	Chr1	24809910	24809910	G	A	Ler	Ler
12	Chr1	24809952	24809952	T	A	Ler	Ler

12	Chr1	24810000	24810000	A	T	Ler	Ler
12	Chr1	24810006	24810006	A	T	Ler	Ler
12	Chr1	24810010	24810010	T	C	Ler	Ler
12	Chr1	24810044	24810044	A	T	Ler	Ler
12	Chr1	24810059	24810060	-	T	Ler	Ler
12	Chr1	24810069	24810069	A	T	Ler	Ler
12	Chr1	24810082	24810082	A	-	Ler	Ler
12	Chr1	24810139	24810139	T	C	Ler	Ler
12	Chr1	24810183	24810183	C	T	Ler	Ler
12	Chr1	24810448	24810448	C	A	Ler	Ler
12	Chr1	24810467	24810467	C	T	Ler	Ler
12	Chr1	24810476	24810476	C	A	Ler	Ler
12	Chr1	24810510	24810510	C	G	Ler	Ler
12	Chr1	24810514	24810514	T	G	Ler	Ler
12	Chr1	24810546	24810546	T	C	Ler	Ler
12	Chr1	24810593	24810593	C	T	Ler	Ler
12	Chr1	24810601	24810608	-------	ATATATAG	Ler	Ler
12	Chr1	24810628	24810628	A	-	Ler	Ler
12	Chr1	24810711	24810711	G	A	Ler	Ler
12	Chr1	24810731	24810732	GA	--	Ler	Ler
12	Chr1	24810735	24810749	TCCCTGGTTAAACCG	----------	Ler	Ler
12	Chr1	24810802	24810802	A	-	Ler	Ler
12	Chr1	24810820	24810820	A	T	Ler	Ler
12	Chr1	24810840	24810840	T	A	Ler	Ler
12	Chr1	24810892	24810892	T	C	Ler	Ler

14	Chr2	17646954	17646954	A	T	Ler	Ler
15	Chr3	1160742	1160742	T	G	Col-0	Col-0
15	Chr3	1160762	1160763	-	T	Col-0	Col-0
15	Chr3	1160777	1160778	-----	GAAAA	Col-0	Col-0
15	Chr3	1160802	1160803	----------	ACTAGATGACAAA	Col-0	Col-0
15	Chr3	1160806	1160806	T	A	Col-0	Col-0
15	Chr3	1160812	1160813	-	T	Col-0	Col-0
15	Chr3	1160813	1160813	C	A	Col-0	Col-0
15	Chr3	1160850	1160851	TA	--	Col-0	Col-0
15	Chr3	1160878	1160888	CTTACACAAA	---------	Col-0	Col-0
15	Chr3	1160893	1160893	C	A	Col-0	Col-0
15	Chr3	1160894	1160894	T	A	Col-0	Col-0
15	Chr3	1160934	1160934	T	-	Col-0	Col-0
15	Chr3	1160936	1160937	AA	--	Col-0	Col-0
15	Chr3	1161082	1161082	--	TT	Col-0	Col-0
15	Chr3	1161092	1161091	----------	TAAATATTTGTACTGAAA	Col-0	Col-0
15	Chr3	1161095	1161095	C	A	Col-0	Col-0
15	Chr3	1161151	1161151	A	-	Col-0	Col-0
16	Chr3	8881161	8881161	G	T	Col-0	Col-0
16	Chr3	8881194	8881194	A	-	Col-0	Col-0
16	Chr3	8881244	8881244	C	A	Col-0	Col-0
20	Chr5	809907	809907	A	T	Col-0	Col-0
20	Chr5	810310	810310	A	G	Col-0	Col-0
21	Chr5	15608316	15608316	A	T	Ler	Ler
21	Chr5	15608605	15608606	-	C	Col-0	Ler

21	Chr5	15608801	15608801	T	-	Ler	Ler
21	Chr5	15608941	15608941	A	G	Ler	Ler
21	Chr5	15608968	15608968	A	G	Ler	Ler
21	Chr5	15608978	15608978	C	G	Ler	Ler
22	Chr1	1644172	1644172	G	T	Ler	Ler
22	Chr1	1644236	1644236	G	A	Ler	Ler
23	Chr1	16272591	16272591	G	T	Col-0	Col-0
23	Chr1	16272669	16272669	G	A	Col-0	Col-0
23	Chr1	16272681	16272681	G	T	Col-0	Col-0
23	Chr1	16272920	16272920	T	C	Col-0	Col-0
23	Chr1	16273058	16273058	G	C	Col-0	Col-0
23	Chr1	16273073	16273073	A	T	Col-0	Col-0
23	Chr1	16273096	16273096	G	T	Col-0	Col-0
23	Chr1	16273112	16273112	C	T	Col-0	Col-0
23	Chr1	16273129	16273129	C	-	Col-0	Col-0
25	Chr2	14700978	14700978	A	G	Col-0	Col-0
26	Chr3	8557632	8557632	T	C	Col-0	Col-0
26	Chr3	8558197	8558200	TCTC	----	Col-0	Col-0
27	Chr3	23161891	23161891	T	C	Col-0	Col-0
31	Chr1	7843170	7843170	T	-	Col-0	Col-0
31	Chr1	7843198	7843198	C	G	Col-0	Col-0
31	Chr1	7843200	7843200	T	A	Col-0	Col-0
31	Chr1	7843203	7843203	A	C	Col-0	Col-0
31	Chr1	7843211	7843211	C	G	Col-0	Col-0
32	Chr1	16473701	16473701	C	A	Col-0	Col-0

32	Chr1	16473766	16473766	A	G	Col-0	Col-0
32	Chr1	16473776	16473776	A	-	Col-0	Col-0
32	Chr1	16473823	16473824	-	A	Col-0	Col-0
34	Chr2	7252882	7252882	A	C	Ler	Ler
34	Chr2	7253017	7253017	T	C	Ler	Ler
35	Chr2	12417912	12417912	A	T	Col-0	Col-0
37	Chr3	9336490	9336490	T	G	Col-0	Col-0
37	Chr3	9336630	9336630	G	T	Col-0	Col-0
40	Chr4	17464026	17464026	T	-	Ler	Ler
40	Chr4	17464140	17464140	T	A	Ler	Ler
41	Chr5	10486369	10486369	C	A	Ler	Ler
44	Chr1	21763781	21763781	T	A	Col-0	Ler
44	Chr1	21763990	21763990	T	A	Ler	Ler
46	Chr2	8549049	8549049	T	A	Ler	Ler
46	Chr2	8549070	8549070	A	T	Ler	Ler
46	Chr2	8549287	8549288	-	T	Ler	Ler
47	Chr2	19020959	19020959	T	-	Col-0	Col-0
48	Chr3	15615740	15615740	A	G	Ler	Ler
48	Chr3	15615988	15615988	T	A	Ler	Ler
48	Chr3	15616063	15616063	G	A	Ler	Ler
49	Chr4	6258844	6258844	C	G	Ler	Ler
49	Chr4	6258920	6258920	A	-	Ler	Ler
49	Chr4	6258924	6258924	T	-	Ler	Ler
49	Chr4	6259275	6259275	G	A	Ler	Ler
50	Chr4	14920040	14920040	T	A	Ler	Ler

51	Chr5	8195145	8195145	A	G	Col-0	Col-0
52	Chr5	21896648	21896648	C	G	Ler	Ler
52	Chr5	21896710	21896710	G	A	Ler	Ler
52	Chr5	21896931	21896931	T	C	Ler	Ler

Supplementary Table S7: Locations of the transition between parental alleles within the single plants. Nearly all crossing over events release genome sequences with a clear non-interrupted transition from one parental allele to the other one. This table list all transitions regions identified in the tetrad and DH lines, along with their location in respect to the reference.

Sample	CO-ID	Chr	Begin	End
Tetrad	1	Chr1	1557671	1558004
Tetrad	1	Chr1	1558051	1559010
Tetrad	2	Chr1	16895874	16896048
Tetrad	2	Chr1	16896049	16896222
Tetrad	3	Chr1	23497124	23498063
Tetrad	3	Chr1	23498063	23498622
Tetrad	4	Chr2	7918187	7918425
Tetrad	4	Chr2	7918426	7918663
Tetrad	5	Chr2	15154598	15156204
Tetrad	5	Chr2	15156204	15156880
Tetrad	6	Chr3	17728032	17728344
Tetrad	6	Chr3	17728345	17728656
Tetrad	7	Chr4	6796827	6797261
Tetrad	7	Chr4	6797262	6797695
Tetrad	8	Chr4	17653068	17653271
Tetrad	8	Chr4	17653272	17653474
Tetrad	9	Chr5	10876774	10877054
Tetrad	9	Chr5	10877172	10877520
Tetrad	10	Chr1	2213217	2213284
Tetrad	10	Chr1	2213628	2213877
Tetrad	11	Chr1	19093540	19093678
Tetrad	11	Chr1	19094144	19094388
Tetrad	12	Chr1	24809419	24809664
Tetrad	12	Chr1	24810892	24810980
Tetrad	13	Chr2	7726474	7726863
Tetrad	13	Chr2	7726864	7727253
Tetrad	14	Chr2	17645816	17646954
Tetrad	14	Chr2	17646954	17647565
Tetrad	15	Chr3	1160294	1160742
Tetrad	15	Chr3	1161151	1161237
Tetrad	16	Chr3	8880855	8881161
Tetrad	16	Chr3	8881244	8882075
Tetrad	17	Chr3	19163980	19164271
Tetrad	17	Chr3	19164272	19164563
Tetrad	18	Chr4	118032	118482

Tetrad	18	Chr4	118483	118932
Tetrad	19	Chr4	11295554	11295992
Tetrad	19	Chr4	11295993	11296431
Tetrad	20	Chr5	809355	809907
Tetrad	20	Chr5	810310	810388
Tetrad	21	Chr5	15608224	15608316
Tetrad	21	Chr5	15608978	15611905
Tetrad	22	Chr1	1643314	1644172
Tetrad	22	Chr1	1644236	1645247
Tetrad	23	Chr1	16271774	16272591
Tetrad	23	Chr1	16273129	16273478
Tetrad	24	Chr2	9739200	9739757
Tetrad	24	Chr2	9739758	9740314
Tetrad	25	Chr2	14700261	14700976
Tetrad	25	Chr2	14700979	14701311
Tetrad	26	Chr3	8557068	8557632
Tetrad	26	Chr3	8558200	8558265
Tetrad	27	Chr3	23161397	23161891
Tetrad	27	Chr3	23161891	23162155
Tetrad	28	Chr4	8365173	8365306
Tetrad	28	Chr4	8365307	8365440
Tetrad	29	Chr5	532301	532619
Tetrad	29	Chr5	532620	532938
Tetrad	30	Chr5	16501638	16501828
Tetrad	30	Chr5	16501829	16502018
Tetrad	31	Chr1	7842859	7843170
Tetrad	31	Chr1	7843211	7843292
Tetrad	32	Chr1	16473123	16473701
Tetrad	32	Chr1	16473824	16474232
Tetrad	33	Chr1	25676514	25677885
Tetrad	33	Chr1	25677886	25679256
Tetrad	34	Chr2	7252497	7252882
Tetrad	34	Chr2	7253017	7253486
Tetrad	35	Chr2	12414849	12417912
Tetrad	35	Chr2	12417912	12418705
Tetrad	36	Chr2	14495108	14495471

Tetrad	36	Chr2	14495472	14495835
Tetrad	37	Chr3	9335488	9336490
Tetrad	37	Chr3	9336630	9337106
Tetrad	38	Chr3	17951154	17951338
Tetrad	38	Chr3	17951339	17951522
Tetrad	39	Chr4	7369829	7370284
Tetrad	39	Chr4	7370285	7370740
Tetrad	40	Chr4	17463448	17464026
Tetrad	40	Chr4	17464140	17464290
Tetrad	41	Chr5	10486315	10486369
Tetrad	41	Chr5	10486369	10486602
Tetrad	42	Chr5	22704512	22704753
Tetrad	42	Chr5	22704754	22704994
Tetrad	43	Chr1	4899961	4900403
Tetrad	43	Chr1	4900404	4900846
Tetrad	44	Chr1	21763389	21763781
Tetrad	44	Chr1	21763990	21764082
Tetrad	45	Chr1	26328821	26328885
Tetrad	45	Chr1	26328886	26328949
Tetrad	46	Chr2	8548811	8549049
Tetrad	46	Chr2	8549288	8549573
Tetrad	47	Chr2	19020888	19020959
Tetrad	47	Chr2	19020959	19030799
Tetrad	48	Chr3	15615636	15615740
Tetrad	48	Chr3	15616063	15616191
Tetrad	49	Chr4	6257906	6258844
Tetrad	49	Chr4	6259275	6260351
Tetrad	50	Chr4	14919739	14920040
Tetrad	50	Chr4	14920040	14923965
Tetrad	51	Chr5	8194916	8195145
Tetrad	51	Chr5	8195145	8195394
Tetrad	52	Chr5	21896525	21896648
Tetrad	52	Chr5	21896931	21897982

Supplementary Table S8: NCO positions in DHs and tetrads. Positions of all converted SNPs are given in the table below. Note that there are NCOs represented by several SNPs.

NCO_ID	Sample	Chr	Locus	NCO	Back-ground	Back-ground allele	NCO allele	left_end	right_end	Left dis-tance	Richt Dis-tance	MinN-COsize	Max NCOsize
1	1_1	Chr5	16568231	Col	Ler	T	A	16568093	16568686	138	455	1	592
2	2_2	Chr3	22673645	Col	Ler	G	C	22671897	22673766	1748	121	1	1868
3	3_4	Chr1	3562592	Ler	Col	A	T	3562114	3564166	478	1562	13	2051
3	3_4	Chr1	3562604	Ler	Col	-	A						
4	3_4	Chr3	465914	Ler	Col	A	G	464063	467193	1851	1078	202	3129
4	3_4	Chr3	465976	Ler	Col	A	G						
4	3_4	Chr3	466115	Ler	Col	A	C						
5	3_2	Chr5	2831387	Ler	Col	T	G	2825460	2838442	5927	7055	1	12981
6	3_3	Chr3	654195	Col	Ler	T	A	653545	654801	650	606	1	1255
7	4_2	Chr1	29232758	Col	Ler	T	A	29231900	29233214	858	456	1	1313
8	4_1	Chr2	13925604	Ler	Col	A	T	13925085	13927255	519	1651	1	2169
9	4_3	Chr1	4090578	Col	Ler	G	A	4089908	4090851	670	237	37	942
9	4_3	Chr1	4090614	Col	Ler	C	T						
10	4_3	Chr2	16012275	Col	Ler	G	T	16011036	16012323	1239	42	7	1286
10	4_3	Chr2	16012281	Col	Ler	G	T						
11	4_3	Chr5	3784266	Col	Ler	A	T	3784231	3784704	35	420	19	472
11	4_3	Chr5	3784284	Col	Ler	C	A						
12	4_3	Chr3	10122998	Ler	Col	C	G	10121358	10125483	1640	2485	1	4124
13	5_1	Chr5	22718313	Col	Ler	A	C	22714374	22724170	3939	5576	282	9795
13	5_1	Chr5	22718594	Col	Ler	T	A						
14	5_1	Chr5	23233151	Col	Ler	C	T	23231108	23235819	2043	2668	1	4710

15	5_1	Chr4	16211203	Ler	Col	G	T	16210149	16212324	1054	1121	1	2174
16	5_1	Chr4	18208163	Ler	Col	T	G	18206405	18209100	1758	937	1	2694
17	5_2	Chr3	20571904	Ler	Col	A	G	20571630	20573677	274	1773	1	2046
18	5_3	Chr5	1036220	Col	Ler	T	G	1036027	1037208	193	988	1	1180
19	DH 1	Chr3	22842900	Col	Ler	G	T	22835558	22845852	7341	2951	1	10293
20	DH 2	Chr1	631901	Col	Ler	T	C	631363	633512	537	1610	1	2148
21	DH 5	Chr3	7537467	Col	Ler	C	T	7536833	7537919	633	451	1	1085
22	DH 5	Chr2	12471762	Col	Ler	C	T	12471348	12472207	413	359	86	858
23	DH 5	Chr2	12471847	Col	Ler	A	T						
23	DH 5	Chr3	19086067	Col	Ler	G	A	19085262	19086158	804	90	1	895
24	DH 5	Chr3	20583226	Col	Ler	C	T	20583054	20584318	171	1091	1	1263
25	DH 5	Chr5	24511119	Ler	Col	T	C	24510298	24511811	820	623	69	1512
25	DH 5	Chr5	24511131	Ler	Col	C	T						
25	DH 5	Chr5	24511187	Ler	Col	T	G						
26	DH 6	Chr4	13992357	Ler	Col	G	A	13991068	13992908	1288	550	1	1839
27	DH9	Chr5	6514168	Col	Ler	C	T	6514108	6514262	59	93	1	153
28	DH 10	Chr5	19158556	Ler	Col	C	G	19156118	19158701	2437	144	1	2582
29	DH 10	Chr3	20620584	Ler	Col	G	A	20620432	20622179	151	1594	1	1746

On nearly understanding randomness in meiosis…

God, vertwijfeld: 'Maar ik dacht…'
De mensen:'U moet niet denken! U moet dobbelen!'

Zij zwaaien met hun vingers, bedreigen hem
en bidden,
offeren hun verveling en hun wildste fantasie.
'Wij eisen gerechtigheid!" roepen zij.
Zij zien het bloed van zijn behoedzaamste daden,
horen hem talmen en aarzelen,
Maar hij dobbelt niet.

T. Tellegen

(In: 'Tijger onder de slakken', 1994)

Chapter 3

CDKA;1 shapes the recombination landscape in *Arabidopsis thaliana*

Authors:
T.G. (Erik) Wijnker[1]*, Hirofumi Harashima[2,3]*, Liudmila Chelysheva[4], Bastiaan de Snoo[5], Niko Dissmeyer[2,6], Martin Bayer[7], Mathilde Grelon[4], Hans de Jong[1], Arp Schnittger[2,3,8].

1) Laboratory of Genetics, Wageningen University, Wageningen, The Netherlands.
2) Department of Molecular Mechanisms of Phenotypic Plasticity, Institut de Biologie Moléculaire des Plantes du Centre National de la Recherche Scientifique, Université de Strasbourg, Strasbourg, France
3) Trinationales Institut für Pflanzenforschung
4) Institut Jean-Pierre Bourgin, INRA Centre de Versailles-Grignon, UR-254 Versailles, France
5) Rijk Zwaan R&D Fijnaart, Eerste Kruisweg 9, 4796 RS Fijnaart, the Netherlands
6) Present address: Leibniz Institute of Plant Biochemistry (IPB), D-06120 Halle (Saale), Germany
7) Max-Planck-Institut für Entwicklungsbiologie, Spemannstraße 35, 72076 Tübingen, Germany
8) For correspondence: arp.schnittger@ibmp-cnrs.unistra.fr
*) Equal contribution

Abstract

Crossover (CO) formation during meiosis is tightly regulated, with only small variation in crossover numbers and a tightly regulated crossover distribution. Such COs are not equally distributed along the chromosome axes, but keep some distance to each other (CO-interference). Remarkably, their number and distribution differ between male and female meiosis, leading to stable, sex specific recombination landscapes (Drouaud *et al*. 2007; Petkov *et al.* 2007). How these processes are controlled is nonetheless largely unknown. Using an allelic series of the central *Arabidopsis* cell-cycle regulator *CDKA*;1 that is homologous to *Cdk*1/*Cdc*2/*Cdc*28 we show that CDKA;1 controls chromosome condensation, chromosome pairing and CO formation at meiotic prophase I. At later stages, high levels of CDKA;1 activity are required for sister chromatid cohesion as well as a the execution of the second meiotic division. While low kinase activity during meiosis does not lead to viable spores, hypomorphic mutants with intermediate kinase activity produce plants with a wildtype phenotype, except for reduced fertility. Genetic analyses of male offspring of this mutant show a completely altered recombination landscape. COs are placed much closer to distal chromosome ends, and CO-interference is strongly reduced. Our data strongly suggest that CDKA;1 is a key constituent of different meiotic processes, as well regulator able to shape recombination landscape.

Introduction

Meiosis is a specialized cell cycle in which a single DNA replication phase (S phase) is followed by two consecutive cell divisions (meiosis I and meiosis II). The first meiotic division generally reduces the chromosome number of the cell by separating homologous chromosomes, and the second segregates sister chromatids, typically leading to the formation of four haploid cells. To ensure proper disjunction of homologues at meiosis I, formation of at least one crossover (CO) per chromosome pair (crossover assurance) is required (Shinohara *et al.* 2008). These COs then reciprocally exchange genetic material between non-sister chromatids, while at the same time provide physical links between the homologues until metaphase I.

In male meiosis CO formation starts with the formation of ~120-~235 double strand breaks (DSBs) in an early meiotic cell complement (Chelysheva *et al*. 2007; Sanchez-Moran *et al*. 2007; Vignard *et al*. 2007). This number is much higher than the number of COs that actually becomes manifested (~6.7 in female- and ~11.5 in male meiosis (Giraut *et al*. 2011). Most of these DSBs are probably resolved as non-crossovers (NCOs) (see chapter 2). The decision of which breaks are resolved as COs or NCOs is thought to be a crucial moment in shaping the recombination landscape. In yeast, this decision takes place as early as DSB formation (Bishop and Zickler 2004; Shinohara *et al*. 2003).

Once formed, COs are usually non-randomly distributed along the synapsed chromosomes, and well-spaced by a phenomenon called CO-interference (Sturtevant 1915). Recent studies in plant meiosis revealed that 85% of COs are sensitive to interference (known as class I COs), and form through the ZMM pathway (Higgins et al. 2004; Mercier

et al. 2005). This pathway was named after the diverse proteins that act in it: ZYP1, Mer3 (Chen *et al.* 2005), MSH4 and MSH5 (Higgins *et al.* 2004; Higgins *et al.* 2008b). A second pathway generates interference independent (class II) COs, mediated by other proteins like MUS81 (Higgins *et al*. 2008a) and EME1.

The mechanism that controls interference is largely unknown, but there is experimental evidence that CO-interference is highly correlated to the physical length of the synaptonemal complex (SC), a proteinaceous structure that - through its transverse filaments - joins the axes of homologous chromosomes during meiotic prophase I. The SC is completely assembled during pachytene, when chromosomes are said to "pair" along their entire length. Short SCs, as found in male mouse and female *Arabidopsis*, correlate with strong CO-interference and low CO numbers (Drouaud *et al.* 2007; Petkov *et al.* 2007; Vizir and Korol 1990). CO-interference is thus tightly linked to chromatin organization and both play an important role in shaping the "recombination landscape", *i.e.*, the specific stable distribution of COs along paired chromosomes.

In the nematode *Caenorhabditis elegans* specific subunits of the condensin protein complex were shown to control meiotic chromosome structure and concomitantly, guide DSB formation and shape the resulting CO-landscape (Mets and Meyer 2009). They showed that SC lengths increased in condensin mutants while simultaneously CO-interference strength decreased and the CO landscape changed when certain genetic intervals increased in length relative to others.

Progression through the meiotic cell cycle is controlled by cyclin dependent kinases (CDKs) which regulate a variety of processes (Malumbres and Barbacid 2005; Morgan 1997; Pines 1999; Satyanarayana and Kaldis 2009). Cdc28, the main CDK of budding yeast is essential for the generation of double strand breaks by phosphorylation of Mer2, a Spo11 ancillary protein (Henderson et al. 2006; Murakami and Keeney 2008). The use of conditional mutants (that allowed the knock-down of Cdc28 after DSB formation, led to impaired SC formation (Zhu *et al.* 2010) and showed the requirement of *Cdc2*8 at later processes. Similarly, Cdk2 is indispensable for meiosis in mammals, since knock-out mice were found to be female and male sterile, and did not form synaptonemal complexes (Berthet et al. 2003; Ortega et al. 2003). Mouse Cdk2 and Cdk4 could be visualized as distinct foci on SCs (Ashley et al. 2001), as was the case for Cdc28 in yeast (Zhu *et al.* 2010).

Bulankova et al. (2010) have shown the requirement of high CDKA;1 kinase activity during *Arabidopsis* meiosis I and II which is in line with with the observation that hypomorphic mutants of the *Cdk*1/*Cdk*2 homolog *CDKA*;1 are viable, but completely sterile due to aberrant microspore production (Dissmeyer *et al*. 2007; Dissmeyer *et al.* 2009; Nowack *et al.* 2012). CDKs depend for their kinase activity on binding to cyclins, two of which have shown to be essential for proper meiotic progression in *Arabidopsis*. SOLODANCERS (SDS) has similarities to A- and B-type cyclins and is crucial for CO formation. SDS forms active complexes with CDKA;1 and not with other *Arabidopsis* CDKs (Harashima and Schnittger 2012). The second one, TARDY ASYNCHRONOUS MEIOSIS (TAM, CYCLIN A1;2)

(Bulankova et al. 2010; Wang et al. 2010; Wang et al. 2004), also forms active complexes with CDKA;1 (Cromer *et al.* 2012) and is essential for the transition of meiosis I to meiosis II and functions in a complex regulatory network with late meiotic genes (Bulankova et al. 2010; d'Erfurth et al. 2010).

Additional evidence of the involvement of CDKs in plant meiosis comes from studies on Pairing homoeologous 1 (*Ph1*) in allohexaploid wheat, a locus that controls homologue recognition. When present, homologous chromosomes pair and recombine, whereas deletion of *Ph1* leads to pairing and CO formation between not only homologues, but also homoeologues of the A, B and D genome. Ph1 harbors a number of inactive *Cdk2*-like alleles (Griffiths *et al.* 2006), that suppress the activity of *Cdk*2 like genes on other chromosomes. The artificial upregulation of kinase activity phenocopies *ph1* like effects (Knight *et al.* 2010), suggesting that high Cdk2 activity leads to homoeologous pairing. Doing the opposite (*i.e.*, increasing *Ph1* dosages) was shown to lead to a loss of CO-formation and the presence of univalents at metaphase I (Feldman 1966). This suggests that the fidelity of CO formation is heavily dependent on CDK dosages.

We set out to elucidate the function of CDKA;1 in male *Arabidopsis* meiosis, at the cytological and molecular level. Using an allelic series of hypomorphic mutants, we show that CDKA;1 is required for chromosome condensation, CO formation and the transition from meiosis I to interkinesis and the completion of meiosis II. Importantly, in a hypomorphic allele in which meiosis is only mildly compromised, CO-interference is released and the recombination landscape profoundly changed towards a more distal CO-localization on chromosomes. Our data suggest that CDKA;1 is crucial for the imposition of CO-interference on class I crossovers and points to interference strength as a key component for shaping recombination landscapes. We also show results that suggest that CDKA;1 gives rise to differences between male and female genetic maps in *Arabidopsis*.

Materials and methods

Plant materials

We used three previously generated hypomorphic alleles of *CDKA;1*, called *CDKA;1*^{T161D} (*cdka*;1-D), *CDKA*;1$^{T14D;Y15E}$ (*cdka*;1-DE) and *PRO*$_{CDKA;1}$*:CDKB1;1* (Dissmeyer *et al*. 2007; Dissmeyer *et al*. 2009). A fourth allele *cdka;*1-DBD was constructed as follows. The construct was fused by PCR with Pfu polymerase (fermentas) to wild-type *Arabidopsis CDKA;1* (At0g12345) cDNA and *CYCLINB1;1* (At0g12345) genomic DNA. The fusion was flanked by Gateway attB1 and -2 sites and recombined in pDONR201 (Invitrogen). Primer combinations used for the *CDKA;1* fragment were ND10-ss_attB1:CDKcoreN and ND18-as_ CDKcoreC:DBovlpN and ND19- ss_DBcoreN:CDKovlpC and ND21-as_DBcoreC:attB2 for the *CYCLINB1;1* part (Table 1). The two fragments were fused in a final PCR with ND10-

ss_attB1:CDKcoreN and ND21-as_DBcoreC:attB2. The destruction box motif 30-RQVLG-DIGN-38 was changed by two substitutions (C/Arg>G/Gly and C/Leu>G/Val) to 30-Gxx-VxxIxN-38 according to (Weingartner *et al.* 2004) by site-directed mutagenesis on the previously described entry clone with PfuTurbo (Stratagene) using primers ND90-ss_B11_dead and ND91-as_B11_dead in order to generate the *CDKA;1*-DBD variant. After sequencing, the obtained Gateway entry clones were recombined with the binary Gateway destination vector pAM-PAT-GW-Pro*CDKA;1* (Nowack *et al.* 2006). Resulting expression vectors conferring phosphinothricin resistance were retransformed into *Agrobacterium tumefaciens* GV3101-pMP90RK (Koncz and Schell 1986) and transformed into heterozygous *cdka*;1+/- by floral dip (Nowack *et al.* 2006).

To study meiotic recombination in *cdka*;1-DBD, a Col-0 (heterozygous for *cdka*;1-1 and *cdka*;1-DBD) was crossed with a newly isolated *cdka;1* allele (*cdka*;1-3) in the Landsberg *erecta* (L*er*) background (heterozygous for *cdka*;1-3 and *cdka*;1-DBD). We obtained Col/L*er* F1 hybrids in a *cdka*;1-/- DBD background. Backcrosses to either Col or L*er* were made to generate *cdka;1*-DBD BC1 populations. We made control crosses with WT Col-L*er* F1's reciprocally to Col, generating both male and female derived BC1 populations. All crosses were made at the same moment, under similar growth conditions in a greenhouse under standard long day growth conditions. For construction of the *cdka*;1-1/*spo*11 double mutant, we used the *spo11-1-1* allele N646172. Plants were grown under standard greenhouse conditions under long day regimes (16 hrs day). Primer sequences used for *spo11* genotyping: N646172U and N646172L (Table 1).

ND10-ss_attB1:CDKcoreN	AACACAAGTTTGTACAAAAAAGCAGGCTTCAACAATGGATCAGTACGAGAAAG
ND18-as_CDKcoreC:DBovlpN	AGAAGTCATCATAGGCATGCCTCCAAGATCCT
ND19-ss_DBcoreN:CDKovlpC	GGAGGCATGCCTATGATGACTTCTCGTTCGATTGTTC
ND21-as_DBcoreC:attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAACAATCTTTTCTTTCT-GTTTCTTCT
ND90-ss_B11_dead	GTAGCGAAAGGAAGAAACGGTCAAGTTGTTGGTGATATCGGTAATGTTG
ND91-as_B11_dead	GAACAACATTACCGATATCACCAACAACTTGACCGTTTCTTCCTTTCGC
N646172U:	AATCGGTGAGTCAGGTTTCAG
N646172L:	CCATGGATGAAAGCGATTTAG

Table 1. Primer sequences used in this study

Cytogenetic analysis

Meiotic cell spreads and slides for Fluorescent *in situ* hybridization (FISH) were made from whole flower buds, using standard protocols (Ross *et al.* 1997). We used probes for the coding sequences of the 18S–25S rDNA plasmid PTa71 of *Triticum aestivum* (Gerlach

and Bedbrook 1979) and plasmid PCt 4.2 containing the *Arabidopsis* full coding sequence of 5S rDNA (GenBank: M65137.1). These were directly labeled with 7-Diethylaminocoumarin-3-carboxylic acid succinimidyl ester (DEAC; Perkin Elmer, http://www.perkinelmer.com) and Cyanine dye 3.5 (CY3.5; Amersham, http://www.gelifesciences.com/) respectively and hybridized using a standard protocol (Tang *et al.* 2008). Immunolabelling also followed standard protocols (Chelysheva *et al.* 2010). A Zeiss Axioplan microscope equipped with an epifluorescence filterset was used for analysis. Images were captured with a Photometrics Sensys 1305 x 1024 pixel CCD camera, and processed with ImageJ (http://rsbweb.nih.gov/ij/) and Adobe Photoshop.

Genetic analysis

BC1 offspring were genotyped using a previously described marker-set (Wijnker *et al.* 2012) based on co-dominant KASPar SNP probes (http://www.kbioscience.co.uk/reagents/KASP.html). Genetic maps were constructed with Joinmap (Stam 1993) using regression mapping and a fixed order of makers based on known map positions (http://www.kyazma.nl/index.php/mc.JoinMap/)

Statistics

A χ^2 goodness of fit test was used to see whether the observed number of COs per chromosome fits a Poisson distribution. We combined observed CO counts for the smallest classes to ensure that the expected Poisson estimate was at least 5. We used α=0.01, and applied a Bonferroni correction to correct for multiple testing.

Results

CDKA;1 localizes on chromosome axes during meiotic prophase

For a detailed understanding of the role of CDKA;1 in meiosis, we first analyzed the expression pattern of CDKA;1 using plants that express a *StrepIII-tag*-CDKA;1 fusion construct known to completely rescue the *cdka;1* mutant phenotype (Pusch *et al.* 2012). In figure 1 all chromosomes are fully stained with an Immunofluorescent signal, that appears to be continuous, except for the nucleolar organizer and pericentromere regions.

Meiotic progression in cdka;1 D/DE hypomorphs

The previously described hypomorphic alleles *cdka;1*-D and -DE have strongly reduced kinase activity and were found to exhibit identical meiotic phenotypes during male meiosis (see materials and methods). In wildtype (WT) meiosis chromosomes condense

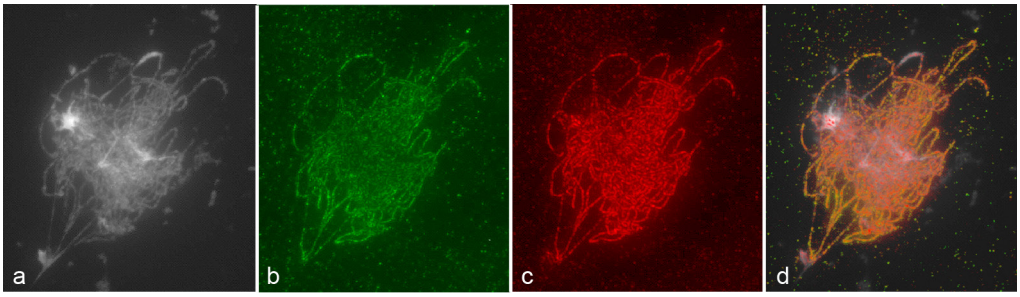


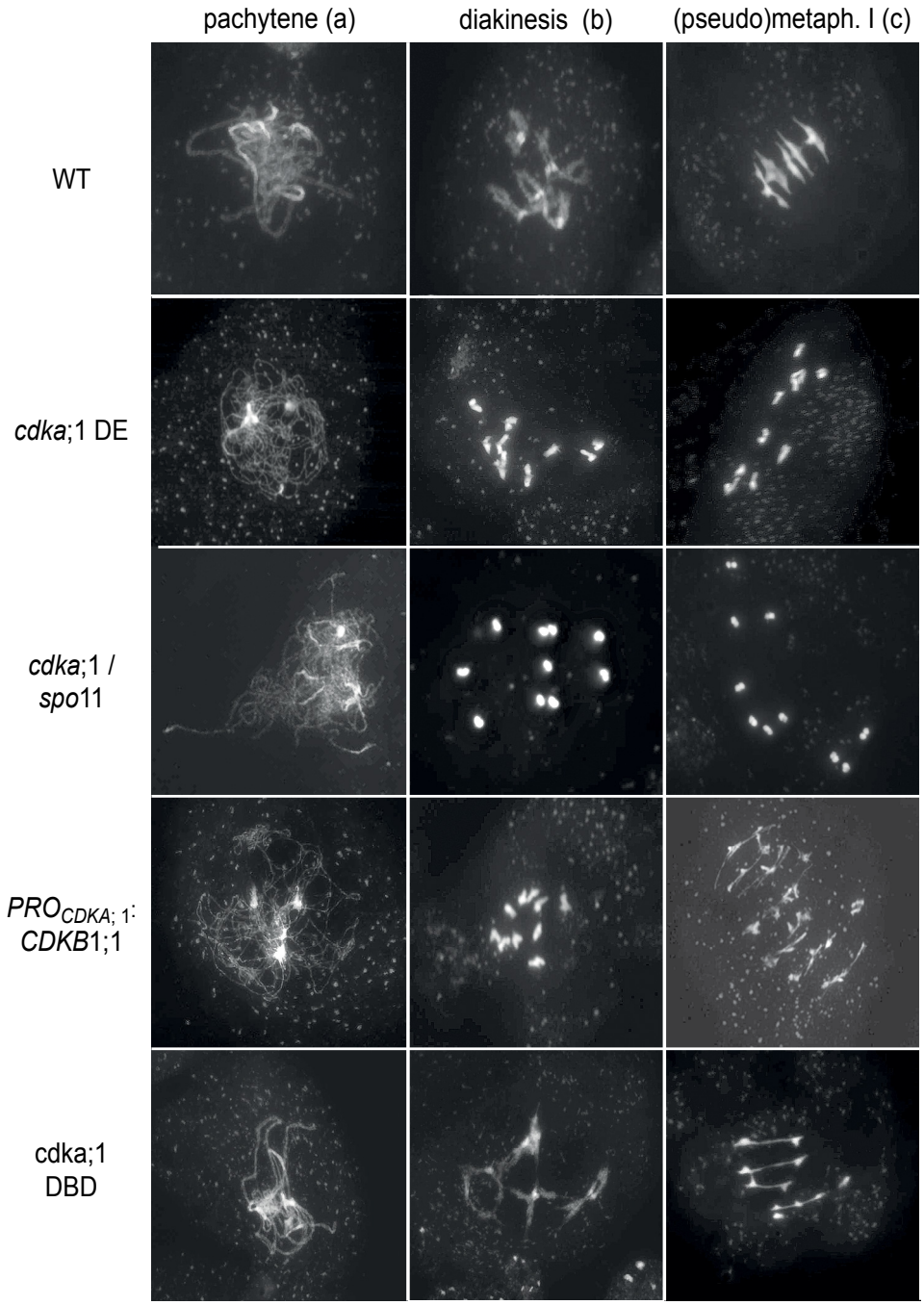
Figure 1. Immunolocalization of CDKA;1 on a WT male prophase nucleus. a) DAPI stained chromatin of a pachytene nucleus. Note the strongly stained centromeres and rDNA (a.o. lower left) that appear as bright white spots; b) Immunofluorescence of CDKA;1 shows all chromosome axes, except for centromeres and rDNA; c) ASY1, a protein marker for prophase chromosome axes co-localizes with DNA and centromeres; d) overlay of a, b and c.

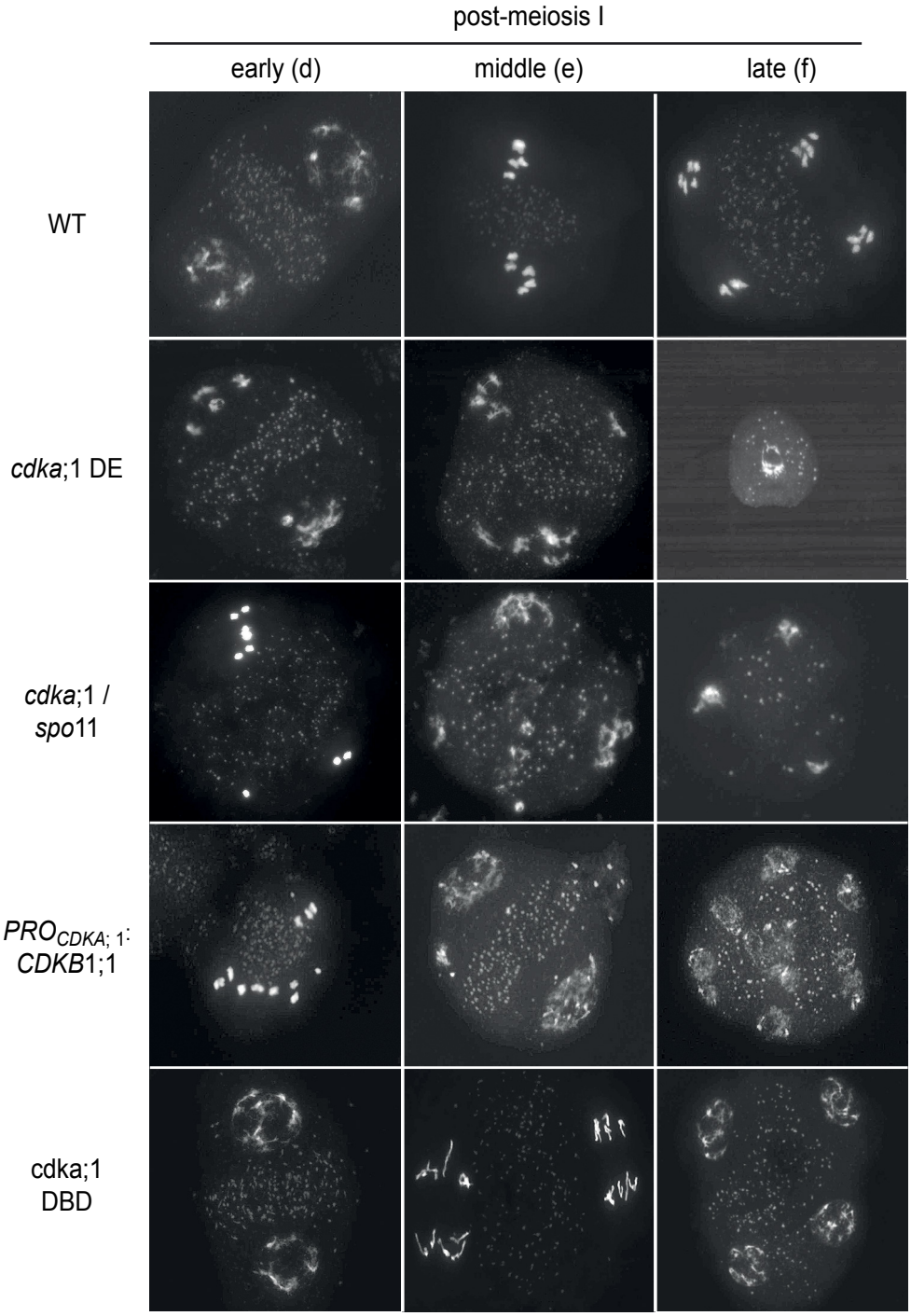
during early prophase, leading to full pairing during pachytene (Figure 2 WT: a). The chromosome structures then diffuse, followed by chromosome condensation towards diakinesis (Figure 2 WT: b). Five pairs of bivalents become visible, that align on a metaphase I plane, and homologues segregate to opposite poles at anaphase I, giving rise to two daughter nuclei separated by an organelle band (interkinesis). The nuclei then undergo a second meiotic division in which chromosome condensation is followed by the segregation of sister chromatids, generating a tetrad with four haploid spores (Figure 2 WT: c-f).

Chromosome morphology in the *cdka;1* hypomorphs resembles WT at early meiotic stages, but the typical pairing of a pachytene stage is absent (Figure 2 *cdka*;1 DE: a). Chromosomes then condense at diakinesis, but univalents become visible instead of bivalents (Figure 2). These are rod shaped and show fuzzy borders. The univalents in mutant late diakinesis cells are about twice as big as single chromatids in WT telophase II, suggesting that *cdka;1* chromosomes consist of two chromatids and did undergo premeiotic S-phase. In 4 out of 30 metaphase/anaphase I cells we observed sister-chromatid segregation at metaphase I.

Meiotic progression in *cdka*;1 hypomorphic mutants is highly disturbed after diakinesis, and hence made unequivocal interpretation of subsequent stages impossible. Chromosomes either form loosely organized groups or remain as single chromosomes following meiosis I that decondense into chromatin masses. At least part of the cells give

Figure 2 (next page)
Meiotic atlas of *cdka*;1 hypomorphs and double mutants. All images are DAPI stained nuclei of pollen mother cells. WT: d) interkinesis; e) metaphase II; f) Tetrad stage. *cdka*;1 DE: d) interkinesis, e) tetrad-like late meiocyte (note there are roughly 10 chromosome domains), f) typical microspore. *cdka*;1/*spo*11: e) telopase I or possibly metaphase II, f) tetrad like stage, g) typical image of microspore like stage in which three chromatin masses are visible. $PRO_{cdka;1}$:CDKA1;1: e) telophase I / metaphase II like stage, f) interkinesis, g) tetrad like stage. *cdkb1;*1-DBD e) interkinesis, f) anaphase II (note the long, decondensed chromosomes), g) tetrad stage.


pachytene (a)
diakinesis (b)
(pseudo)metaph. I (c)
WT
cdka;1 DE
cdka;1 / *spo*11
PRO$_{CDKA;1}$:*CDKB*1;1
cdka;1 DBD


post-meiosis I
early (d)
middle (e)
late (f)
WT
cdka;1 DE
cdka;1 / spo11
PRO CDKA; 1: CDKB1;1
cdka;1 DBD

rise to interkinesis-like stages where two daughter nuclei are separated by a clear organelle band. In these interkinesis nuclei, the chromosomes occupy apparent distinct domains within a nucleus like structure. A clear second meiotic division has not been observed. Occasionally, a phragmoplast becomes visible within the organelle band at interkinesis (5 out of 42 cells), indicating that cell division is in progress at this stage (Figure 2 d-f). We also find numerous polyads in which chromosomes display as about ten unstructured chromatin masses, after which cytokinesis commences. These cells do not display a clear well defined organelle band. The observation that microsporocytes contain about 10 chromosomes in loose aggregates suggests that sister chromatids do not separate. Moreover, we have no indications that these cells undergo a second meiotic division.

Asynapsis in *cdka;1* does not result from impaired function of SPO11 or SDS

To genetically disentangle the function of CDKA;1 we constructed the double mutant of hypomorphic *cdka*;1-D with *spo*11. Such mutants show a phenotype different from *cdka;1* (Figure 2, *cdka*;1/*spo*11). The meiotic prophase of the double mutant is asynaptic (like the *spo11* and *cdka*;1-D single mutants, leading to the formation of univalents at diakinesis). Univalents in *cdka;1/spo11* double mutants are indistinguishable from the *spo11* single mutant, in being rounded and having sharp edges, and differ from the more diffuse univalents in *cdka;1* hypomorphs. This implies that spo11 is epistatic over *cdka;1* and suggests that the poor condensation phenotype of the *cdka;1* single mutant results at least in part from DSB formation. Equally, the higher chromosome condensation in the *cdka;1 spo11* double mutant suggests that SPO11 is at least partially active in the *cdka;1* hypomorphs, which is consistent with the occasional fragments that we observe in *cdka;1* meiocytes. Homologues segregate at meiosis I, forming dyads or polyads after meiosis I. However, we did not find metaphase II cells and we did not observe cells with a phenotype similar to *cdka;1* hypomorphs with polyad-like cells showing loose aggregates of decondensed chromosomes.

To further analyze the requirement of CDKA;1, we performed immunofluorescence microscopy of ASY1 (Sanchez-Moran *et al.* 2007) and DMC1 on meiotic prophase I nuclei of *cdka;1*-D/DE hypomorphs. Continuous ASY1 foci are formed along the chromosomes (Figure 3), suggesting that high levels of CDKA;1 activity are not required for the formation of lateral elements and loading of ASY1 on the chromosomes. The observation that the pattern of DMC1 foci does not differ from WT meiosis indicates that DSBs are formed in a *cdka;1* hypomorphic mutant background and that SPO11 is functional at low CDKA;1 levels. Taken together, these data suggest that the requirement of CDKA;1 in meiosis lies downstream of SPO11, ASY1 and DMC1.

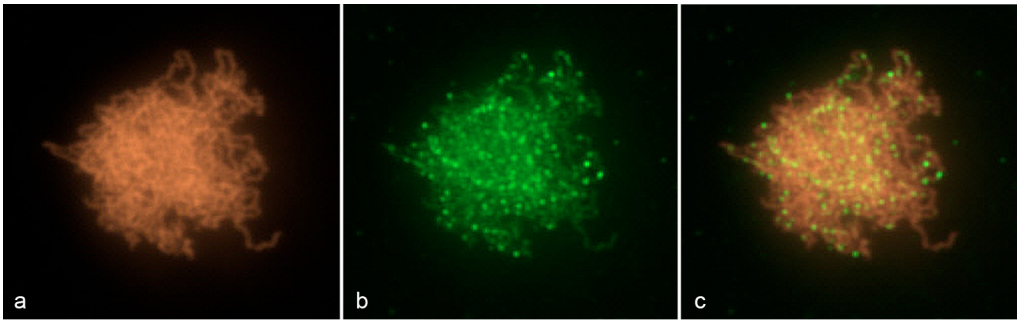


Figure 3. Immunolocalization of DMC1 and ASY1 in *cdka;*1 D/DE. a) DAPI stained chromatin. b) ASY1 localizes onto chromosome axes in D/DE, similar as in WT (cf. Figure 1); b) DMC1 shows a punctuate staining in cdka;1, displaying the loading of DMC1 onto chromosome axes; c) Overlay of the two images using the lighten blending mode in Adobe Photoshop CS6.

CDKA;1 function cannot be compensated by other kinases

Previous studies have shown that the function of CDKA;1 in mitosis can be partially substituted by CDKB1;1 (Nowack *et al.* 2012). To clarify this, we asked to what degree CDKB1;1 could also rescue cdka;1 null mutants in meiosis. We therefore analyzed homozygous *cdka*;1-/- mutants with a $PRO_{CDKA;1}$*:CDKB*1;1 expression construct, in which *CDKB*1;1 is driven from a *CDKA*;1 promoter. We can then also evaluate whether the D/DE hypomorphic transformants display mild or "leaky" phenotypes that are expected in these lines.

The phenotypes of these $PRO_{CDKA;1}$*:CDKB*1;1 hypomorphs are highly similar to those of the *cdka;*1 D/DE hypomorphs. The prophase is completely asynaptic, and gives rise to rod-shaped univalents at diakinesis with poorly defined "fuzzy" borders, similar to cdka;1 hypomorphs (Figure 2, $PRO_{CDKA;1}$*:CDKB*1;1: b). No fragments were observed. Like in *cdka;1* D/DE mutants, meiotic progression is highly disturbed beyond diakinesis. Homologues and sister-chromatids segregate at meiosis I giving rise to unbalanced meiosis I products. All or most of the sister chromatids segregate in 4 out of 20 anaphase I cells, giving rise to two daughter cells consisting of both chromosomes and chromatids. These cells then form polyad–like cells, similar to *cdka;1* D/DE hypomorphs. Metaphase II cells are uncommon and like in *cdka;1* D/DE hypomorphs we observed occasional progression of cytokinesis after meiosis I (Figure 2).

$PRO_{CDKA;1}$*:CDKB*1;1 plants exhibit meiotic phenotypes that are highly reminiscent of *cdka;*1 D/DE hypomorphs. Up to diakinesis their phenotypes are is similar showing complete achiasmatic meiosis, and rod-shaped univalents at metaphase I. A main difference is the regular segregation of chromatids at metaphase I (16 out of 20 events) in $PRO_{CDKA;1}$*:CDKB*1;1. Execution of meiosis II was only observed once. In 4 out of 25 events we observed that chromosomes segregated without segregating chromatids at metaphase I. In 4 out of >70 late meiotic stages we noted presence of chromosome fragments in the cytoplasm, which presumably originate from fragmentation at metaphase I. The segre-

gation of chromatids at metaphase I is more prevalent in $PRO_{CDKA;1}$:*CDKB*1;1 than in *cdka;1* D/DE hypomorphs, again suggesting a role for CDKA;1 in sister-chromatid cohesion.

Meiotic progression in *cdka*;1 DBD: a partial loss of function allele of *CDKA*;1

The inviability of D/DE spores precludes an analysis of CDKA;1 function beyond meiosis. Therefore, we scanned through other *cdka*;1 mutant alleles that were generated in our laboratory (unpublished data). Among these mutants we found one allele (DBD) of which the phenotype of which the sporophytic growth was hardly distinguishable from WT (in contrast to the smaller size of *cdka*;1 D/DE), with the exception of showing strongly reduced fertility (producing 0-5 seeds per fully grown plant).

Cytological analysis of *cdka*;1-DBD pollen mother cells revealed that the meiotic prophase is largely unaffected: we observed complete pairing of homologues at pachytene (Figure 2, *cdka;1*-DBD: a). The first aberrations become evident at metaphase I when 1/3 (15 out of 44) metaphase I cells show the presence of univalent pairs (Figure 4, a-c). In

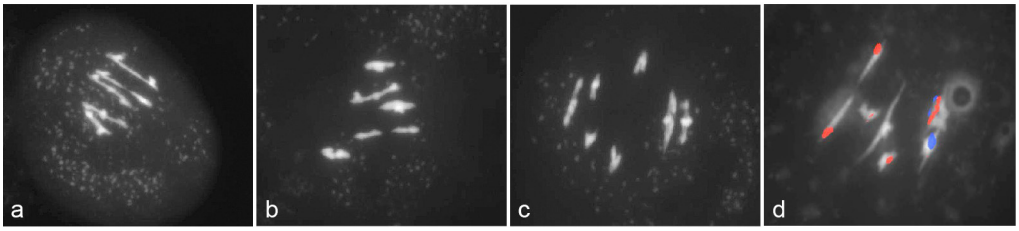


Figure 4. Metaphase I in *cdka;1*-DBD. The images a-c suggest occasional precocious homologue segregation in *cdka;1*-DBD; b) FISH using 5S (red) and 45S (blue) rDNA probes demonstrates that only chromosomes with identical FISH patterns (homologues) form crossovers.

addition, chromosomes appear less condensed at metaphase I, concordant with observations on the *cdka*;1 D/DE hypomorphs. Following a partial desynaptic prophase, meiotic progression is largely regular, with the occasional aberrations related to univalent segregation at metaphase I (*i.e.*, lagging chromosomes, unbalanced segregation and occasional polyads). We found cells at metaphase I in which homologues were held together by very thin chromatin threads, or in which homologue pairs segregated precociously (15 out of 44 events). This could suggest that some COs locate extremely distal on the chromosomes in *cdka*;1-DBD. Progression through metaphase II is regular, with the exception that anaphase chromosomes are clearly less condensed as in WT meiosis (Figure 2: *cdka;*1-DBD: e).

Upregulation of kinase activity during meiosis has previously been shown to affect the fidelity of homologue recognition in wheat and leads to homoeologous pairing (Knight *et al.* 2010). Such effects are thus not to be expected in *cdka*;1 hypomorphs in which *CDKA*;1 activity is thought to be low. Even though non-homologous pairing was not suggested by our observations of pachytene and metaphase I stages in DBD (the condensation patterns of homologues joined by chiasmata are identical), we verified the for-

mation of COs between homologues through fluorescent in-situ hybridization, using 5S and 45S rDNA probes that allow the identification of homologues at metaphase I, which shows that the *cdka*;1-DBD hypomorphs indeed form COs between homologous chromosomes (Figure 4d).

CDKA;1 shapes the recombination landscape

We made cross-pollinations between *cdka*;1-DBD and WT plants in order to assess spore viability. Infertility was found to result from female meiosis, as only DBD-pollen frequently gave rise to offspring. To further analyze the meiotic behavior of *cdka*;1-DBD, we generated F1 hybrids between the *Arabidopsis* accessions Columbia (Col) and Landsberg *erecta* (L*er*) in a *cdka*;1-DBD background (see M&M). These F1's were backcrossed as males using WT Col and L*er* as female parents. BC1 offspring were genotyped with 36 SNP markers spanning the entire genome. 754 Recombination events were detected in 147 WT offspring and 758 recombination events in 216 cdka;1-DBD offspring. In cdka;1-DBD we note an average number of 3.5 COs per BC1, against 5.1 in WT (Table 2), suggesting a decrease in recombination frequency. We then constructed genetic maps using Haldanes mapping function (see supplemental figure 1). We noted that genetic intervals at the distal ends of chromosomes are longer in *cdka;1*-DBD in comparison to WT. Apparently, the recombination landscape changed in DBD.

In *Arabidopsis* genetic maps one usually corrects for CO-interference to estimate the total genetic map length. But for a proper comparison of total genetic map lengths between WT and DBD it is important to verify whether CO-interference is affected in DBD. We thus asked whether observed changes in recombination frequencies could be attributed to differences in CO-interference. Since our SNP markers are near evenly spaced at roughly 4 Mb apart, the incidence of double crossovers (DCOs) (here defined as CO events that are present on either side of a specific SNP) can be used as a (rough) approximation of CO-interference (Table 2). With 0.20 (WT) and 0.19 (*cdka*;1-DBD) detected events per offspring plant, DCO incidence is similar for both populations. However, we noted a remarkable difference in the distribution of DCOs over the different chromosomes. *Arabidopsis* chromosomes 2 and 4 are the shortest chromosomes, and have the

Chromosome	WT (n=147)				DB_dead (n=216)			
	COs	Mean	SD	DCOs	COs	Mean	SD	DCOs
1	184	1.25	0.96	4	152	0.79	0.79	7
2	121	0.82	0.65	1	143	0.72	0.72	8
3	157	1.07	0.82	9	155	0.77	0.77	9
4	112	0.76	0.73	3	132	0.69	0.69	7
5	180	1.22	0.87	13	176	0.77	0.77	9
total	754	5.12		30	758	3.5		40

Table 2: Detected numbers of crossovers (COs) and double crossovers (DCOs) in WT and *cdka;*1-DBD BC1 (male) offspring. See text for details.

smallest genetic maps in male meiosis (Giraut *et al*. 2011; Wijnker *et al*. 2012). In our WT population we noted that DCO incidence on the smallest chromosomes is very small (4 events in 147 offspring) whereas in *cdka*;1-DBD DCOs are equally likely on all chromosomes. This suggests that CO-interference might be affected in cdka;1-DBD.

The fit of a CO distribution to a poisson distribution is a widely used indicator for the presence of interference (Higgins *et al*. 2004; Qi *et al*. 2009). Consistent with previous findings (Drouaud *et al*. 2006), the distribution of COs in WT was significantly different from a random distribution ($\alpha = 0.01$) (Figure 5). In contrast, the DB-dead population was not significantly different from a poisson distribution, indicating that COs in *cdka*;1-DBD are randomly placed onto chromosomes.

For a proper comparison of the total genetic map lengths of WT and *cdka*;1-DBD we used different approaches for calculating meiotic maps of both popolations. WT male total genetic map length was estimated using the Kosambi mapping function which corrects for CO-interference (supplemental figure 3). This WT map measures 532 cM, which is very close to previous reports, with largely similar segregation distortions (Supplemental Figure 2) (Wijnker *et al*. 2012). We used the Haldane mapping function to estimate the total genetic map length for *cdka*;1-DBD, because this does not correct for CO-interference. Since low levels of interference cannot be excluded, the map length might be slightly overestimated. The *cdka*;1-DBD total genetic map measures 411 cM, which is ~20% shorter than WT male meiosis (Supplemental Figure 3). We observed some differences in segregation distortions between *cdka*;1-DBD and WT (see Supplemental Figure 2).

The genetic map of *cdka*;1-DBD is not only shorter than that of WT meiosis, but also shows remarkable differences in the relative sizes of the genetic intervals. Distal intervals become longer, whereas intervals in the middle of chromosomes become smaller as compared to WT. Figure 6 illustrates the changes of genetic map length relative to the same interval in WT meiosis.

It was previously shown that *Arabidopsis* male meiosis differs from female meiosis by having more COs positioned at the chromosome ends, near telomeres (Drouaud *et al.* 2007; Giraut *et al.* 2011). Finding that low levels of CO-interference lead to an enrichment of COs at chromosome ends, and a depletion in the chromosome middle, we asked whether the CO landscape in female meiosis, in which interference strength is known to be stronger, might show a CO landscape opposite to that of *cdka*;1-DBD. To this end a mapping population was analysed for WT female meiosis analogous to the WT male mapping population, by fertilizing a WT Col-L*er* F1 hybrid with WT pollen (using the same F1 hybrids as used for generation of the male meiotic map). Hundred and six offspring were generated and genotyped using the same SNP markerset. A genetic map was generated (Kosambi mapping function) with a total length of 328 cM which is highly similar to previously generated maps (Giraut *et al.* 2011) (supplemental figure 3), and unusual segregation distortions were not observed (supplemental figure 2). The relative differences in

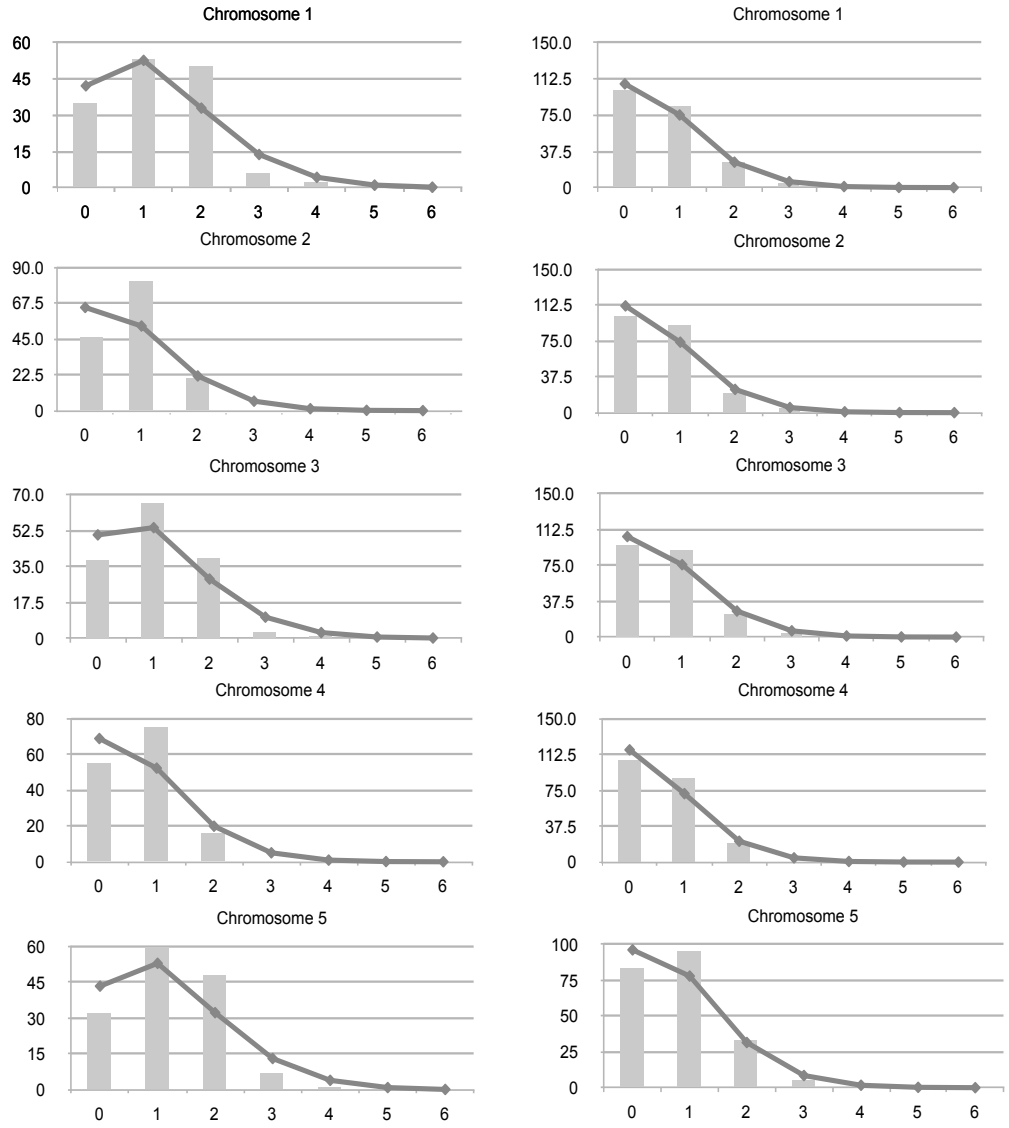


Figure 5. Observed numbers of COs fitted to a Poisson distribution in WT (left) and *cdka;*1-DBD (right). Observed numbers of COs per chromosome are given in grey bars. Expected numbers based on random distributions are estimated using a Poisson distribution (dark lines).

recombination rates in the various intervals are compared for the three populations in figure 6.

The comparison of the relative genetic interval lengths of WT male meiosis with *cdka*;1-DBD male and WT female meiosis shows that decreases in recombination in *cdka*;1-DBD are almost perfectly mirrored by relative increases in recombination in female meiosis. For *cdka;1*-DBD that show an increase in recombination relative to WT,

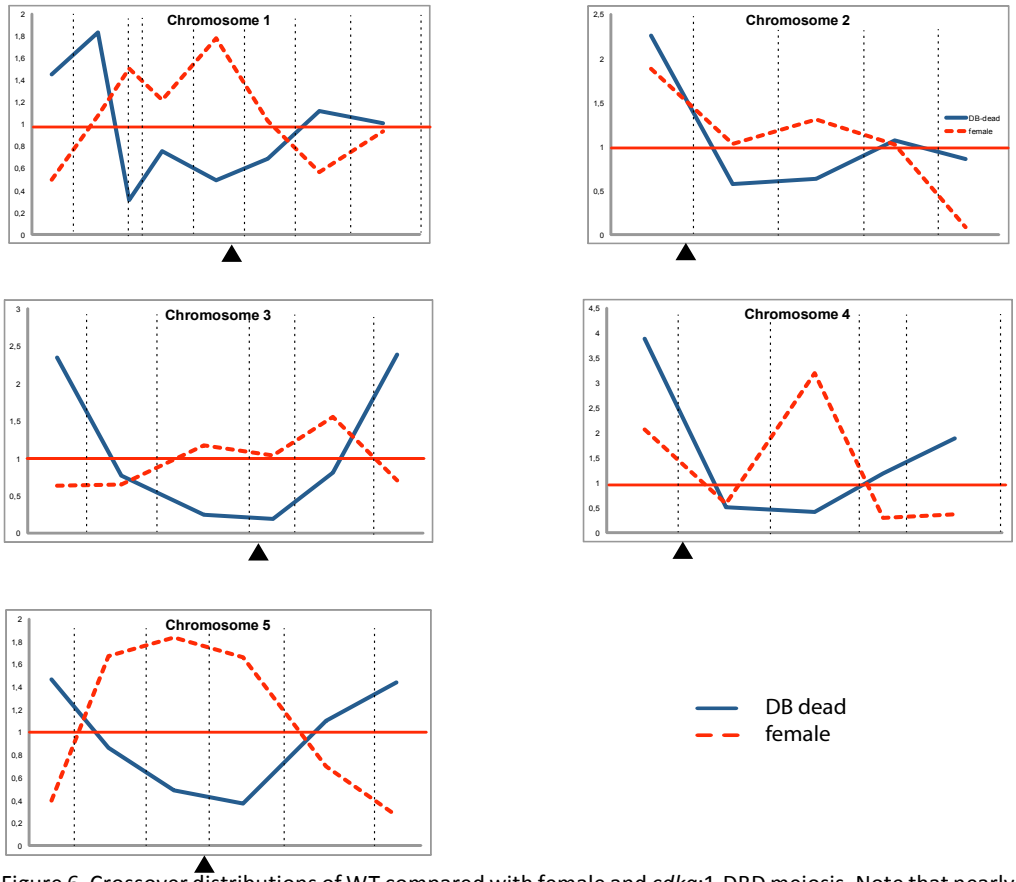


Figure 6. Crossover distributions of WT compared with female and *cdka*;1-DBD meiosis. Note that nearly all intervals that are longer in DBD than in WT male, are smaller in WT female, and vice versa. The red horizontal line represents the recombination in WT male (*i.e.*, everything is normalized there, setting rf's in male to 1). The thin dotted lines denote the marker positions. The black triangle shows the position of the centromere.

female meiosis shows almost invariably the inverse by increased recombination. The pattern of relative recombination increase and decrease for *cdka;1*-DBD is U-shaped for chromosomes 3, 4 and 5, but somewhat irregular on chromosome 1. The reasons for this are unclear, but might be related to unique chromatin dynamics for this chromosome (Supplemental discussion).

The intervals covering the complete short arms of chromosomes 2 and 4 (and to smaller extent) the distal end of the long arm of chromosome 2 do not show opposite recombination changes in *cdka*;1-DBD and WT female meiosis as compared to WT male meiosis. The two short arms flank the nuclear organizing regions (NORs) and the satellites harboring the 45S rDNA and show a relative recombination increase in both *cdka;1*-DBD and female meiosis. The short arm of chromosome 4 harbors a well known inver-

sion between L*er* and Col, which disrupts recombination. But since the genetic interval covering the chromosome 4 short arm also covers co-linear sequences and both chromosome 2 and 4 show the same pattern, we favor another explanation. Giraut *et al.* (2011) showed that there is a sharp rise in recombination frequency in male meiosis very close to the NORs in male meiosis. Our SNP markers are not placed at the most distal end, for which reason we may have missed COs near telomeres and the NOR. The use of additional markers more close to the NOR might show that WT male recombination in this distal interval is in fact higher, which would potentially correct the sharp rise of recombination frequencies in both female and *cdka*;1-DBD in figure 6.

Discussion

To investigate the requirement of the major cell cycle kinase *CDKA*;1 for *Arabidopsis* meiosis, we used an allelic series of hypomorphic mutants that allowed the study of an otherwise barely viable null mutant (Nowack *et al.* 2012). The *cdka*;1-D and *cdka*;1–DE hypomorphs have strongly reduced CDKA;1 kinase activity. A nearly similar phenotype was seen in plants rescued with a $PRO_{CDKA;1}$*:CDKB1;1* expression construct, in which CDKB1;1 is driven from a *CDKA*;1 promoter. A fourth allele is *cdka;1*-DBD in which kinase activity is higher, presumably intermediate between WT and the other hypomorphs.

Meiosis in an allelic series of CDKA;1 hypomorphs

In *cdka;1*-DBD, the *cdka;*1 allele showing the mildest phenotype, one would expect that only those processes that require the highest kinase activity are affected. Cytogenetic analyses pointed to aberrant chromosome morphology, as witnessed by rod-shaped, "fuzzy" univalents; compromised CO-assurance, as shown by a desynaptic prophase I following a fully paired pachytene; and we noted the loss of CO-interference as well as distinct changes in the recombination landscape as CO-events occur more frequently at distal chromosome ends.

At lower CDKA;1 kinase activity in the *cdka;1*-D/DE hypomorphs, chromosome pairing and CO-formation are completely lost. Chromosomes condense poorly at diakinesis as in *cdka;1*-DBD and display a loss of sister chromatid segregation at meiosis I. After meiotic prophase I only part of the meiocytes form clear interkinesis-like stages, but most form groups of with groups of more or less partly condensed chromatin masses, after which cytokinesis commences. it seems that meiosis stops at this stage as no later stages are observed. When CDKA;1 is completely removed, and CDKA;1 is substituted for CDKB1;1 (in $PRO_{CDKA;1}$*:CDKB1;1*), the phenotype is largely similar to *cdka;1*-D/DE hypomorps, although we note in addition to the above that sister-chromatid cohesion appears further compromised (as witnessed by the segregation of sister-chromatids in 3/4th of

anaphase I cells). Interestingly, we see that the second meiotic division is sometimes executed when CDKB1;1 replaces CDKA;1.

Overseeing the diversity of different phenotypes, CDKA;1 appears to be required for the fidelity of different processes. Table 3 lists various *Arabidopsis* meiotic proteins that are discussed here, and indicates the presence of predicted CDK phosphorylation target sites based on amino acid sequences that are characterized by a specific sequence of amino acids: [S/T]-P-X-[R/X]. It is worth noting that in most groups of proteins there is at least one protein highly dependent on phosphorylation, which does not a priory point to the disruption of a specific meiotic process in cdka;1 meiosis. The requirements of CDKA;1 for meiosis are discussed in the following.

protein	function	cyclin binding [R/K]-X-L	predicted CDK phospho sites	minimal [S/T]=P	maximal [S/T]-P-X-[R/K]
SPO11-1	DSB formatinon	3	2	2	-
ASY1	Axis protein	10	5	5	1
DMC1	Strand invasion	2	-	-	-
REC8	Sister chromatid cohesion	3	9	9	3
OSD1	APC/C inhibitior	3	7	7	3
RBR1	CDKA;1 inhibitor	10	16	16	8
TAM	Meiotic cyclin	2	5	5	-
SDS	Meiotic cyclin	7	6	6	-
ZYP1a	Class I COs / ZMM	10	4	4	-
ZYP1b	Class I COs / ZMM	11	3	3	-
MER3/RCK	Class I COs / ZMM	6	12	12	3
MSH4	Class I COs / ZMM	4	-	-	-
MSH5	Class I COs / ZMM	6	2	2	-
MUS81	Class II COs	4	3	3	-
EME1A	Class II COs	7	5	5	-
EME18	Class II COs	7	8	8	1
FANCM	Class II COs	14	23	23	3

Table 3. Predicted CDK phosporylation sites of meiotic proteins mentioned in the text

CDKA;1 requirement during early prophase

Immunolocalization of CDKA;1 on WT meiotic cells spreads shows that CDKA;1 co-localizes with euchromatin during zygotene, and we noted the absence of a fluorescent signal on centromeres and rDNA regions. At early pachytene no more signal is observed on chromosomes on paired regions. Since the disappearance of a fluorescent signal coincides with the loss of ASY1 (which has previously been shown to remain present throughout pachytene at paired regions), the disappearance of CDKA;1 could be an artifact rather than an indication of true absence. These data nevertheless suggest that

CDKA;1 physically interacts with chromatin during the early meiotic prophase. Interestingly, the presence of full staining of chromosomes that we observed is different from the localization pattern of Cdk2 and Cdk4 in mouse and Cdc28 in budding yeast, for both of which a distinct punctuate staining was reported (Ashley *et al.* 2001; Zhu *et al.* 2010). Apart from localization to chromatin in zygotene, also the loss of chromosome pairing in *cdka;1*-D/DE points to a requirement of CDKA;1 in early meiotic prophase. Our observations have shown that a variety of early meiotic processes are still properly executed: pre-meiotic S-phase, DSB formation, ASY1 and DMC1 loading onto chromosomes. All these processes are either independent of CDKA;1, require only very low CDKA;1 kinase activity or suggest there is functional redundancy with other kinases. Importantly, the absence of pachytene pairing we observed in *cdka;1*-D/DE is not a downstream effect of impaired SPO11 action and the loss of DSB formation. This marks a clear difference to yeast where the presence of Cdc28 activity is essential for the formation of DSBs (Henderson et al. 2006; Murakami and Keeney 2008).

As of yet we cannot conclude whether the absence of pachytene pairing in *cdka;1* points to the direct involvement of CDKA;1 in SC-formation, or whether it is a downstream effect of an earlier defect. The loss of CO-interference (in *cdka;1*-DBD) and SC formation (cdka;1-D/DE) are often associated with mutations in the ZMM pathway that forms class I COs. ZMM mutants usually show the loss of a SC (*zyp1*) (Higgins *et al.* 2005), usually coupled to a strongly (85%) reduced crossing over as in *ZIP4* (Chelysheva *et al.* 2007), *MSH4, MSH5* (Higgins *et al.* 2004; Higgins *et al.* 2008b) and *AtMER3* (Chen *et al.* 2005). Since in *cdka;1*-DBD we do not note such a 85% CO-reduction and *cdka;1*-D/DE shows the complete loss of CO-formation, it seems unlikely that *cdka;1* phenocopies ZMM mutants. Nevertheless, it is possible that CDKA;1 acts onto ZMM proteins though phosphorylation and while most classic ZMM proteins like ZYP1 or MSH4/ MSH5 are not highly packed with phosphosites, MER3 is a notable exception (see table 3).

So, the absence of chromosome pairing could well be a downstream effect of CDKA;1 function. Our observation that in a *cdka;1 spo11* double mutant the condensation phenotype of *cdka;1* disappears, suggests the direct involvement of CDKA;1 in DNA repair. Interestingly, such a role has previously been suggested for the meiotic cyclin SDS (De Muyt *et al.* 2009), the mutation of which leads to impaired loading of DMC1 onto chromosomes. We note however that asynaptic phenotype of cdka;1 is not a simple phenocpoy of the SDS mutation, since DMC1 localization in cdka;1 is not affected.

CDKA:1 is required in late meiosis

Following the asynaptc prophase in *cdka;1*-D/DE, meiotic progression after diakinesis/ anaphase I is strongly compromised. The same was observed for the *cdka;1/spo11* double mutant. Interkinesis is occasionally observed, but chromosome decondensation seems impaired, and in many cells chromosomes persist as partially decondensed chromatin masses until cytokinesis proceeds. These observations are very similar to observations

in *Caenorhabditis elegans cdk1* and *cdk7* mutants in which a diakinesis arrest takes place while chromosomes are present as chromatin masses (Wallenfang and Seydoux; Boxem *et al,* 1999).

Two proteins with many phosphorylation sites have before been implicated in meiotic cell cycle progression before and are known to interact with CDKA;1. These are RETINABLASTOMA RELATED (RBR) and OMISSION OF SECOND MEIOTIC DIVISION (OSD1) (Chen *et al.* 2011; d'Erfurth *et al.* 2009). RBR is a known cell cycle inhibitor, and presumably the low kinase activity is not enough to overcome the inhibitory action of RBR during meiosis. OSD has recently been shown to act as inhibitor of the anaphase promoting complex (APC/C) (Cromer *et al*. 2012) and it is well possible that also the compromised action of OSD might cause difficulties in execution of metaphase II.

CDKA;1, the recombination landscape and heterochiasmy

The construction of the cdka;1-DBD hypomorph provides a unique insight into the impact of lowering CDKA;1 kinase activity on recombination landscapes. Our data show that CO-interference at least in part got lost and that COs occur more distally than in WT meiosis. We also noted a decrease in total genetic map length, but the observation that COs localize more distally in *cdka;1*-DBD suggests that some COs go undetected and might underestimate the genetic map length.

Our data provide the first evidence of the involvement of a major CDK in CO-interference, but the molecular base for this relation is still unclear. In the case of CDKA;1, the loss of CO-interference does not lead to increases in recombination rates, suggesting that CO-interference is not the limiting factor in CO-number. A more striking correlation is the apparent inverse relation between the *cdka;1*-DBD and WT-female genetic maps in comparison to the WT male meiotic map (Table 4). Our data strongly suggest that the phenotypes of WT-female meiosis, WT-male meiosis and *cdka;1*-DBD male meiosis represent a continuum in which CO-interference strength decreases from WT female meiosis to meiosis in *cdka;1*-DBD. In intervals in the physical chromosome middle, recombination is relatively high in WT female meiosis, whereas it decreases in *cdka;1*-DBD.

The strength CO-interference apparently covaries with the distal or proximal localization of COs. When we consider previous reports of the higher CO-interference strength in *Arabidopsis* female meiosis (Drouaud *et al*. 2007). Our inability to examine female re-

	*Cdka;*1-DBD	WT male	WT female
CO-localisation	Distal	Intermediate	Proximal
CO-Interference	Low (or absent)	Medium	High
CO-number	Medium (or high?)	High	Low

Table 4: Summary of observations of CO-localization, CO-interference and CO-number in our three mapping populations. The interference strength for female meiosis is inferred from (Drouaud *et al.* 2007)

combination in a cdka;1-DBD background is unfortunate, as it precludes us of assessing whether the effect seen in male meiosis is similar in female meiosis. The use of yet new alleles that either show an even milder phenotype, or increases of CDKA;1 activity would be very helpful to see whether indeed CO-interference strength is directly controlled by CDKA;1 kinase activity. From our data we predict that increasing CDKA;1 activity in male *Arabidopsis* meiosis would induce a recombination landscape that is more close to female meiosis.

The activity of CDKs is tightly controlled during the cell-cycle and is kept stable though multiple regulatory mechanisms. This regulation is likely mediated through cyclins, the CDK chaperones that mediate CDK kinase activity. CDK activity could present a master switch to regulate the CO-landscape, both between species, but also within species. Sex specific recombination rates (heterochiasmy) have been shown to be pretty variable between species (Lenormand and Dutheil 2005). In *Brassica oleracea* for example, female meiosis shows the longer map whereas in *Arabidopsis* the male map is longer (Kearsey *et al.* 1996).

CDKA;1 might act through condensins

It has become clear in recent years that key regulators of chromosome architecture and chromosome dynamics are multi-subunit protein complexes known as condensins (Hirano 2005). Condensins are direct targets of mammalian CDK1 (Sánchez and Dynlacht 2005), and *Xenopus* Cdc2 (Kimura *et al.* 1998), and packed with phosphosites, although most of these sites do not follow the standard [S/T]-P-X-[R/X] consensus sites, suggesting that either CDKs show relaxed specificity for these substrates, or that there are other kinases that also act upon condensins (Bazile *et al*. 2010). Even though various kinases likely act upon condensin, it is the CDKA;1 homologs (CDK1, Cdc2 and others) that are specifically required for the initial activation of supercoiling action of condensing (Hagstrom *et al*. 2002; Kimura *et al.* 2001; Kimura *et al*. 1998; St-Pierre *et al*. 2009).

Condensins have been implicated in meiosis of *Arabidopsis* (Siddiqui *et al*. 2003), *S. cerevisiae* (Yu and Koshland 2003) and *C. elegans* (Mets and Meyer 2009). Mutant phenotypes include poor chromosome compaction, impaired SC assembly, chromosome missegregation and (DSB-dependent) anaphase bridges in yeast (Yu and Koshland 2003). In addition, mutants in specific condensin subunits in *C. elegans* have recently been shown to affect crossover placement and lead to reductions in CO-interference (Mets and Meyer 2009). Research on *Arabidopsis* condensins has not revealed such compelling similarities (Siddiqui *et al.* 2003), but this does not necessarily argue against condensin dependent action in cdka;1 hypomorphs. The presence of two different, likely redundant condensin complexes in *Arabidopsis* might well obscure a clear vision of condensin requirement in plant mitosis and meiosis (Siddiqui *et al.* 2003). The moment of action of condensins in *C. elegans* lies at the CO/NOC decision (Tsai *et al*. 2008). This directs attention (again) to the very early prophase (*i.e.* before SC formation), in which we suspected the critical ac-

tion of *cdka;1* D/DE hypomorphs would lie. As of yet, we can however not substantiate a CDKA;1 - condensin interactions in *Arabidopsis* with experimental data.

CDKA;1 mimics *Ph1*

Ph1 has previously been shown to be a suppressor of homoeologous pairing in wheat and the Ph1 locus allegedly ensures a constitutive low expression of Cdk2 like kinases, promoting pairing between homologues instead of homoeologous chromosomes. The precise way in which *Ph*1 acts is currently unknown, but there is a multitude of evidence that chromatin remodeling during meiotic prophase is of key importance (Colas *et al.* 2008). The presence of *Ph1* causes a constitutive low activity of mammalian *CDK*2 like genes (Yousafzai *et al*. 2010a; Yousafzai *et al*. 2010b). Increasing copy number would be expected to lower the CDK activity in these plants even more. Our *cdka;1*-DBD nicely phenocopies wheat lines with multiple doses of (*Ph*1) (Feldman 1966), in which increasing dosages of *Ph*1 led to the occurrence of univalents at metaphase I.

A model by Bazile *et al.* (2010) suggested that (mitotic) condensin action is most sensitive to low CDK activity and relatively insensitive to high levels of CDK activity. Such a hypothesis would nicely reconcile with *Ph1* action. Low CDK activity could mediate the differential compaction of homoeologous chromosomes, thereby promoting homologous chromosome pairing. Higher levels of CDK activity will push the cells through prophase much more quick, leaving less time to differentially remodel their chromatin, and find their true homologous partners. Direct evidence for differential chromatin remodeling in our data is few, but the observation that chromosome 1 behaves differently under low CDKA;1 activity (Supplemental Discussion), at least suggests that such dynamics might be at play.

Acknowledgements

We would like to thank Dóra Szinay (Wageningen University) for help with FISH experiments. Rob Dirks (Fijnaart) is gratefully acknowledged for warmly supporting the conduct of this research at the Rijk Zwaan breeding company. Maarten Koornneef and Joost Keurentjes (Wageningen University) are thanked for insightful discussions and critical reading of this manuscript.

Literature

Ashley T, Walpita D, de Rooij D (2001) Localization of two mammalian cyclin dependent kinases during mammalian meiosis. Journal of Cell Science 114:685

Bazile F, St-Pierre J, D'Amours D (2010) Three-step model for condensin activation during mitotic chromosome condensation. Cell Cycle 9:3243

Berthet C, Aleem E, Coppola V, Tessarollo L, Kaldis P (2003) Cdk2 knockout mice are viable. Current Biology 13:1775

Bishop DK, Zickler D (2004) Early decision: meiotic crossover interference prior to stable strand exchange and synapsis. Cell 117:9

Bulankova P, Riehs-Kearnan N, Nowack MK, Schnittger A, Riha K (2010) Meiotic progression in Arabidopsis is governed by complex regulatory interactions between SMG7, TDM1, and the meiosis I–specific cyclin TAM. The Plant Cell 22:3791

Chelysheva L, Gendrot G, Vezon D, Doutriaux M-P, Mercier R, Grelon M (2007) Zip4/Spo22 is required for class I CO formation but not for synapsis completion in *Arabidopsis thaliana*. PLoS Genetics 3:e83

Chelysheva L, Grandont L, Vrielynck N, le Guin S, Mercier R, Grelon M (2010) An easy protocol for studying chromatin and recombination protein dynamics during *Arabidopsis thaliana* meiosis: immunodetection of cohesins, histones and MLH1. Cytogenetic and Genome Research 129:143

Chen C, Zhang W, Timofejeva L, Gerardin Y, Ma H (2005) The *Arabidopsis ROCK-N-ROLLERS* gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. The Plant Journal 43:321

Chen Z, Higgins JD, Hui JTL, Li J, Franklin FCH, Berger F (2011) Retinoblastoma protein is essential for early meiotic events in Arabidopsis. EMBO Journal 30:744

Colas I, Shaw P, Prieto P, Wanous M, Spielmeyer W, Mago R, Moore G (2008) Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. Proceedings of the National Academy of Sciences 105:6075

Cromer L, Heyman J, Touati S, Harashima H, Araou E, Girard C, Horlow C, Wassmann K, Schnittger A, De Veylder L, Mercier R (2012) OSD1 promotes meiotic progression via APC/C inhibition and forms a regulatory network with TDM and CYCA1;2/TAM. PLoS Genetics 8:e1002865

d'Erfurth I, Cromer L, Jolivet S, Girard C, Horlow C, Sun Y, To JPC, Berchowitz LE, Copenhaver GP, Mercier R (2010) The CYCLIN-A CYCA1;2/TAM Is required for the meiosis I to meiosis II transition and cooperates with OSD1 for the prophase to first meiotic division transition. PLoS Genetics 6:e1000989

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R (2009) Turning meiosis into mitosis. PLoS Biology 7:e1000124

De Muyt A, Pereira L, Vezon D, Chelysheva L, Gendrot G, Chambon A, Lainé-Choinard S, Pelletier G, Mercier R, Nogué F, Grelon M (2009) A high throughput genetic screen identifies new early meiotic recombination functions in Arabidopsis thaliana. PLoS Genetics 5:e1000654

Dissmeyer N, Nowack MK, Pusch S, Stals H, Inzé D, Grini PE, Schnittger A (2007) T-Loop phosphorylation of Arabidopsis CDKA;1 is required for its function and can be partially substituted by an aspartate residue. The Plant Cell 19:972

Dissmeyer N, Weimer AK, Pusch S, De Schutter K, Kamei CLA, Nowack MK, Novak B, Duan G-L, Zhu Y-G, De Veylder L, Schnittger A (2009) Control of cell proliferation, organ growth, and DNA damage response operate independently of dephosphorylation of the *Arabidopsis* Cdk1 homolog CDKA;1. The Plant Cell 21:3641

Drouaud J, Camilleri C, Bourguignon P-Y, Canaguier A, Bérard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B, Quesneville H, Mézard C (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots". Genome Research 16:106

Drouaud J, Mercier R, Chelysheva L, bérard A, Falque M, Martin O, Zanni V, Brunel D, Mézard C (2007) Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4 PLoS Genetics 3:1096

Feldman M (1966) The effect of chromosomes 5B, 5D, and 5A on chromosomal pairing in *Triticum aestivum*. Proceedings of the National Academy of Sciences 55:1447

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Research 7:1869

Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mézard C (2011) Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. PLoS Genetics 7:e1002354

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature 439:749

Hagstrom KA, Holmes VF, Cozzarelli NR, Meyer BJ (2002) C. elegans condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. Genes & Development 16:729

Harashima H, Schnittger A (2012) Robust reconstitution of active cell-cycle control complexes from co-expressed proteins in bacteria. Plant Methods 8:in press

Henderson KA, Kee K, Maleki S, Santini PA, Keeney S (2006) Cyclin-dependent kinase directly regulates initiation of meiotic recombination. Cell 125:1321

Higgins JD, Armstrong SJ, Franklin FCH, Jones GH (2004) The *Arabidopsis* MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. Genes and Development 18:2557

Higgins JD, Buckling EF, Franklin FCH, Jones GH (2008a) Expression and functional analysis of AtMUS81 in *Arabidopsis* meiosis reveals a role in the second pathway of crossing-over. The Plant Journal 54:152

Higgins JD, Sanchez-Moran E, Armstrong SJ, Jones GH, Franklin FCH (2005) The *Arabidopsis* synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. Genes and Development 19:2488

Higgins JD, Vignard J, Mercier R, Pugh AGF, Franklin CH, Jones GH (2008b) AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in *Arabidopsis thaliana*, but is not required for synapsis. The Plant Journal 55:28

Hirano T (2005) Condensins: organizing and segregating the genome. Current Biology 15:R265

Kearsey M, Ramsay L, Jennings D, Lydiate D, Bohuon E, Marshall D (1996) Higher recombination frequencies in female compared to male meisoses in *Brassica oleracea*. TAG Theoretical and Applied Genetics 92:363

Kimura K, Cuvier O, Hirano T (2001) Chromosome condensation by a human condensin complex in *Xenopus* egg extracts. Journal of Biological Chemistry 276:5417

Kimura K, Hirano M, Kobayashi R, Hirano T (1998) Phosphorylation and activation of 13S condensin by Cdc2 in vitro. Science 282:487

Knight E, Greer E, Draeger T, Thole V, Reader S, Shaw P, Moore G (2010) Inducing chromosome pairing through premature condensation: analysis of wheat interspecific hybrids. Functional and Intergrative Genomics 10:603

Koncz C, Schell J (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Molecular and General Genetics MGG 204:383

Lenormand T, Dutheil J (2005) Recombination difference between sexes: a role for haploid selection. PLoS Biology 3:e63

Malumbres M, Barbacid M (2005) Mammalian cyclin-dependent kinases. Trends in Biochemical Sciences 30:630

Mercier R, Jolivet S, Vezon D, Huppe E, Chelysheva L, Giovanni M, Nogué F, Doutriaux M-P, Horlow C, Grelon M, Mézard C (2005) Two meiotic crossover classes cohabit in *Arabidopsis*: one is dependent on MER3, whereas the other one is not. Current Biology 15:692

Mets DG, Meyer BJ (2009) Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. Cell 139:73

Morgan DO (1997) Cyclin-dependent kinases: Engines, clocks, and microprocessors. Annual Review of Cell and Developmental Biology 13:261

Murakami H, Keeney S (2008) Regulating the formation of DNA double-strand breaks in meiosis. Genes & Development 22:286

Nowack MK, Grini PE, Jakoby MJ, Lafos M, Koncz C, Schnittger A (2006) A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. Nature Genetics 38:63

Nowack Moritz K, Harashima H, Dissmeyer N, Zhao XA, Bouyer D, Weimer Annika K, De Winter F, Yang F, Schnittger A (2012) Genetic framework of cyclin-dependent kinase function in *Arabidopsis*. Developmental Cell 22:1030

Ortega S, Prieto I, Odajima J, Martin A, Dubus P, Sotillo R, Barbero JL, Malumbres M, Barbacid M (2003) Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. 35:25

Petkov PM, Broman KW, Szatkiewicz JP, Paigen K (2007) Crossover interference underlies sex differences in recombination rates. Trends in Genetics 23:539

Pines J (1999) Four-dimensional control of the cell cycle. Nature Cell Biology 1:E73

Pusch S, Harashima H, Schnittger A (2012) Identification of kinase substrates by bimolecular complementation assays. The Plant Journal 70:348

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

Ross K, Fransz P, Armstrong S, Vizir I, Mulligan B, Franklin F, Jones G (1997) Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA-transformed lines. Chromosome Research 5:551

Sanchez-Moran E, Santos J-L, Jones GH, Franklin FCH (2007) ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in *Arabidopsis*. Genes and Development 21: 2220

Sánchez I, Dynlacht BD (2005) New insights into cyclins, CDKs, and cell cycle control. Seminars in Cell & Developmental Biology 16:311

Satyanarayana A, Kaldis P (2009) Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. Oncogene 28:2925

Shinohara M, Oh SD, Hunter N, Shinohara A (2008) Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. Nature Genetics 40:299

Shinohara M, Sakai K, Shinohara A, Bishop DK (2003) Crossover interference in *Saccharomyces cerevisiae* requires a TID1/RDH54- and DMC1-dependent pathway. Genetics 163:1273

Siddiqui NU, Stronghill PE, Dengler RE, Hasenkampf CA, Riggs CD (2003) Mutations in *Arabidopsis* condensin genes disrupt embryogenesis, meristem organization and segregation of homologous chromosomes during meiosis. Development 130:3283

St-Pierre J, Douziech M, Bazile F, Pascariu M, Bonneil É, Sauvé V, Ratsima H, D'Amours D (2009) Polo Kinase Regulates Mitotic Chromosome Condensation by Hyperactivation of Condensin DNA Supercoiling Activity. Molecular Cell 34:416

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join Map. The Plant Journal 3:739

Sturtevant AH (1915) The behavior of the chromosomes as studied through linkage. Molecular and General Genetics MGG 13:234

Tang X, Szinay D, Ramanna MS, Vossen EAGvd, Datema E, Klein Lankhorst RM, Boer JMd, Peters SA, Bachem CWB, Stiekema WJ, Visser RGF, Jong JHd, Bai Y (2008) Cross-species BAC-FISH painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180:1319

Tsai CJ, Mets DG, Albrecht MR, Nix P, Chan A, Meyer BJ (2008) Meiotic crossover number and distribution are regulated by a dosage compensation protein that resembles a condensin subunit. Genes & Development 22:194

Vignard J, Siwiec T, Chelysheva L, Vrielynck N, Gonord F, Armstrong SJ, Schlögelhofer P, Mercier R (2007) The Interplay of RecA-related proteins and the MND1–HOP2 complex during meiosis in *Arabidopsis thaliana*. PLoS Genetics 3:e176

Vizir I, Korol AB (1990) Sex difference in recombination frequency in *Arabidopsis*. Heredity 65:379

Wang Y, Jha AK, Chen R, Doonan JH, Yang M (2010) Polyploidy-associated genomic instability in Arabidopsis thaliana. Genesis 48:254

Wang Y, Magnard J-L, McCormick S, Yang M (2004) Progression through Meiosis I and Meiosis II in Arabidopsis Anthers Is Regulated by an A-Type Cyclin Predominately Expressed in Prophase I. Plant physiology 136:4127

Weingartner M, Criqui M-C, Mészáros T, Binarova P, Schmit A-C, Helfer A, Derevier A, Erhardt M, Bögre L, Genschik P (2004) Expression of a nondegradable cyclin B1 affects plant development and leads to endomitosis by inhibiting the formation of a phragmoplast. The Plant Cell 16:643

Wijnker E, van Dun K, de Snoo CB, Lelivelt CLC, Keurentjes JJB, Naharudin NS, Ravi M, Chan SWL, de Jong H, Dirks R (2012) Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. Nature Genetics 44:467

Yousafzai F, Al-Kaff N, Moore G (2010a) Structural and functional relationship between the <i>Ph1</i> locus protein 5B2 in wheat and CDK2 in mammals. Functional & Integrative Genomics 10:157

Yousafzai FK, Al-Kaff N, Moore G (2010b) The molecular features of chromosome pairing at meiosis: the polyploid challenge using wheat as a reference. Functional & Integrative Genomics 10:147

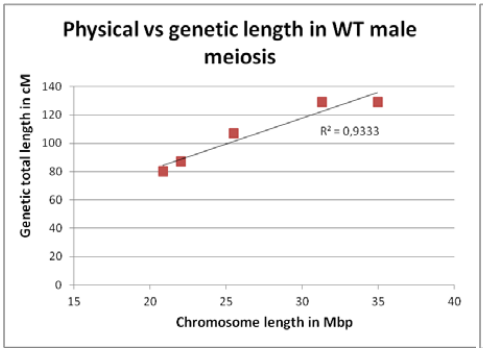
Yu H-G, Koshland DE (2003) Meiotic condensin is required for proper chromosome compaction, SC assembly, and resolution of recombination-dependent chromosome linkages. The Journal of Cell Biology 163:937

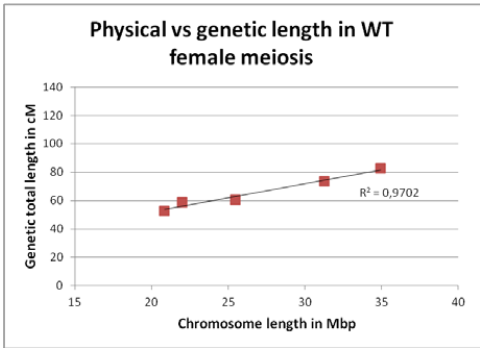
Zhu Z, Mori S, Oshiumi H, Matsuzaki K, Shinohara M, Shinohara A (2010) Cyclin-dependent kinase promotes formation of the synaptonemal complex in yeast meiosis. Genes to Cells 15:1036

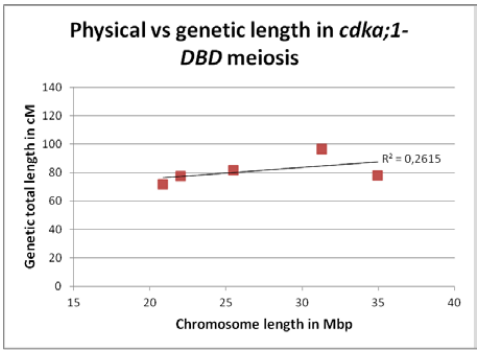
Supplementary discussion

As for most chromosomes in our mapping populations, chromosome 1 shows relative increases and decreases in recombination for *cdka;1*-DBD and WT-female of opposite direction (text figure 6). But the pattern on chromosome 1 is clearly more irregular than the recombination in- and decreases as seen in for example chromosome 5 (see main text). We therefore wondered whether this could be due to specific attributes of chromosome 1. For this we made graph (above) comparing the total genetic map length (in cM, y-axis) of all five chromosomes in relation to their physical lengths (x-axis, data from TAIR; http://www.Arabidopsis.org/). Map lengths for WT male and female populations were estimated using the kosambi mapping function, and *cdka;1*-DBD map was estimated using the Haldane mapping function.

Note that the physical- and genetic map lengths correlate very well for all three populations, except for the largest chromosome 1, depicted by the three most right data points. In WT male and female populations, a larger physical length (in Mbp) strongly correlates with higher CO numbers (R2 values of 0.93 and 0.97 respectively), which is visible by a linear trend in the accompanying figure. Correlation is much lower in *cdka;1*-DBD (R^2 value of 0.26).



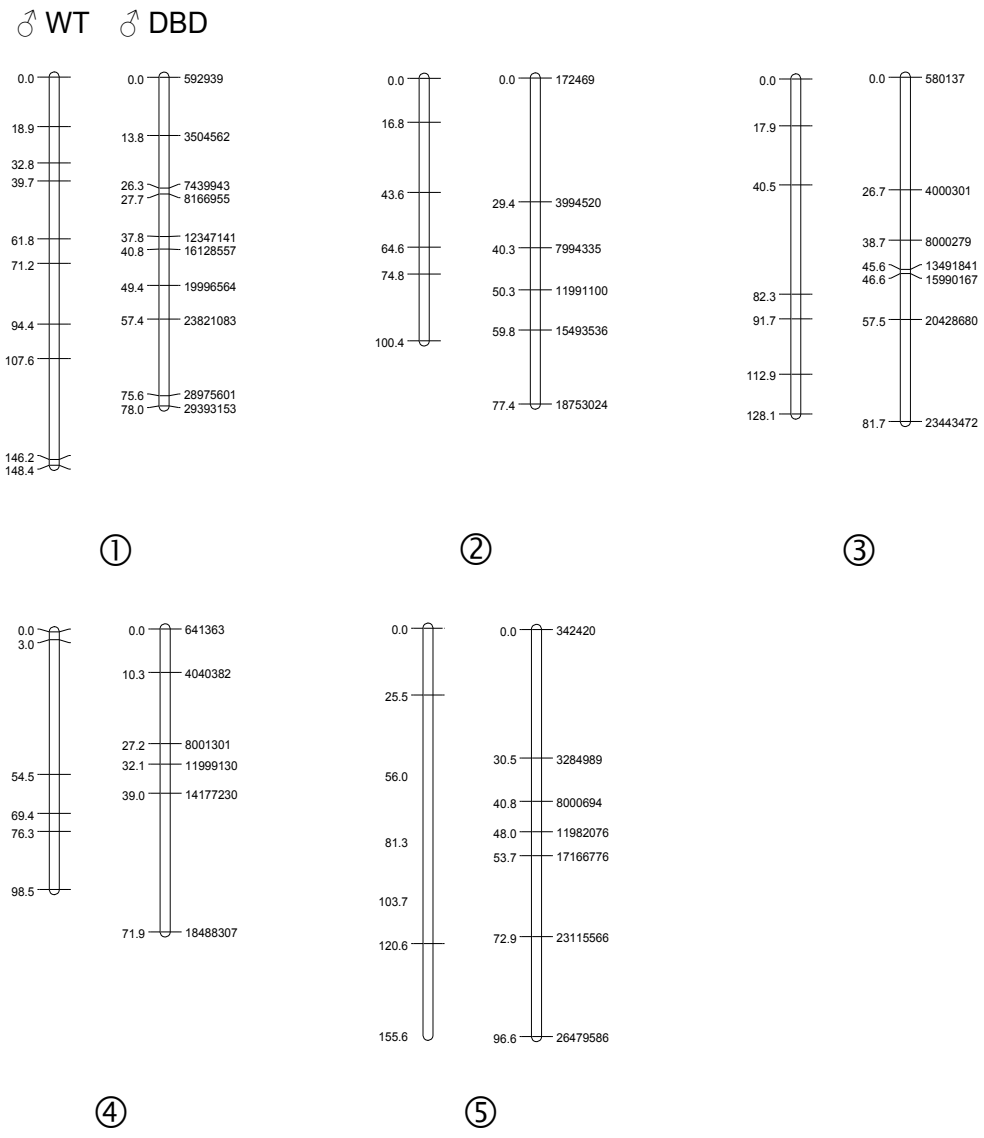




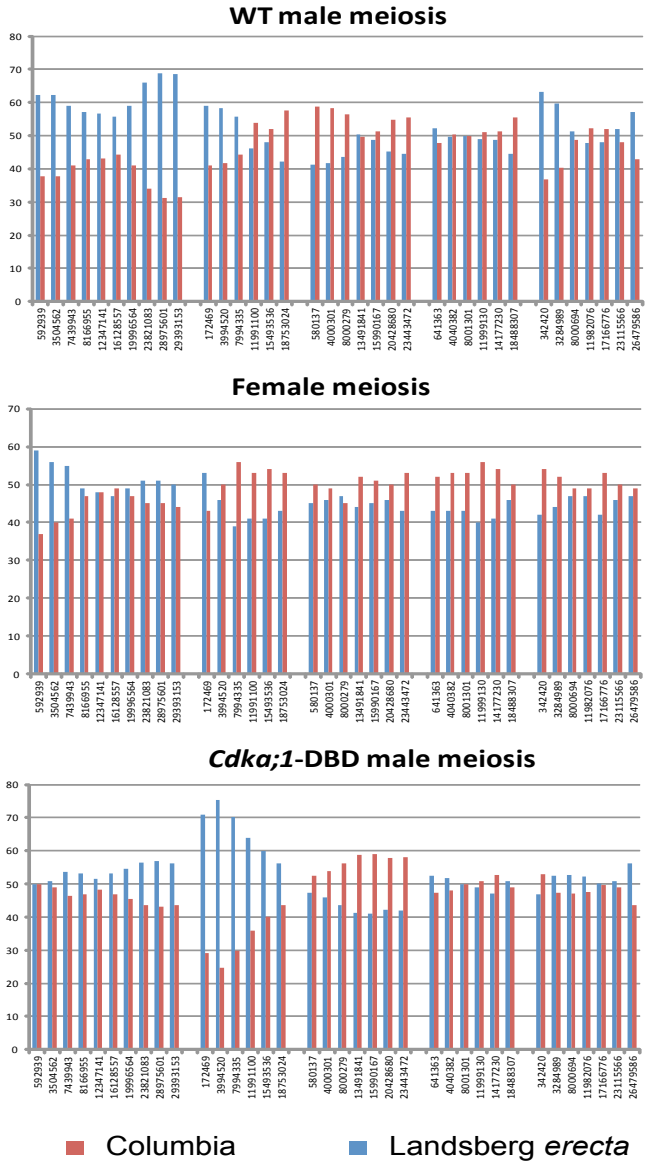
The poor correlation of genetic and physical map length in *cdka;1*-DBD is mainly caused by chromosome 1. Chromosome 1 is not well known for having notable physical features like 5S or 45S rDNA. There is one specific feature of chromosome 1 that might be of interest here, which is that chromosome 1 is a fusion product of two ancient chromosomes, and still harbours an ancient centromere (Hansson *et al.*, 2006). If chromatin dynamics underlie chromosome condensation, it might be possible that under certain conditions, the chromatin of *Arabidopsis* chromosome I might still react differently to specific kinase levels than chromatin that was not once part of a centromere

Hansson, B., Kawabe, A., Preuss, S., Kuittinen, H. & Charlesworth, D. Comparative gene mapping in *Arabidopsis lyrata* chromosomes 1 and 2 and the corresponding *A. thaliana* chromosome 1: recombination rates, rearrangements and centromere location. Genetics Research 87, 75-85 (2006).

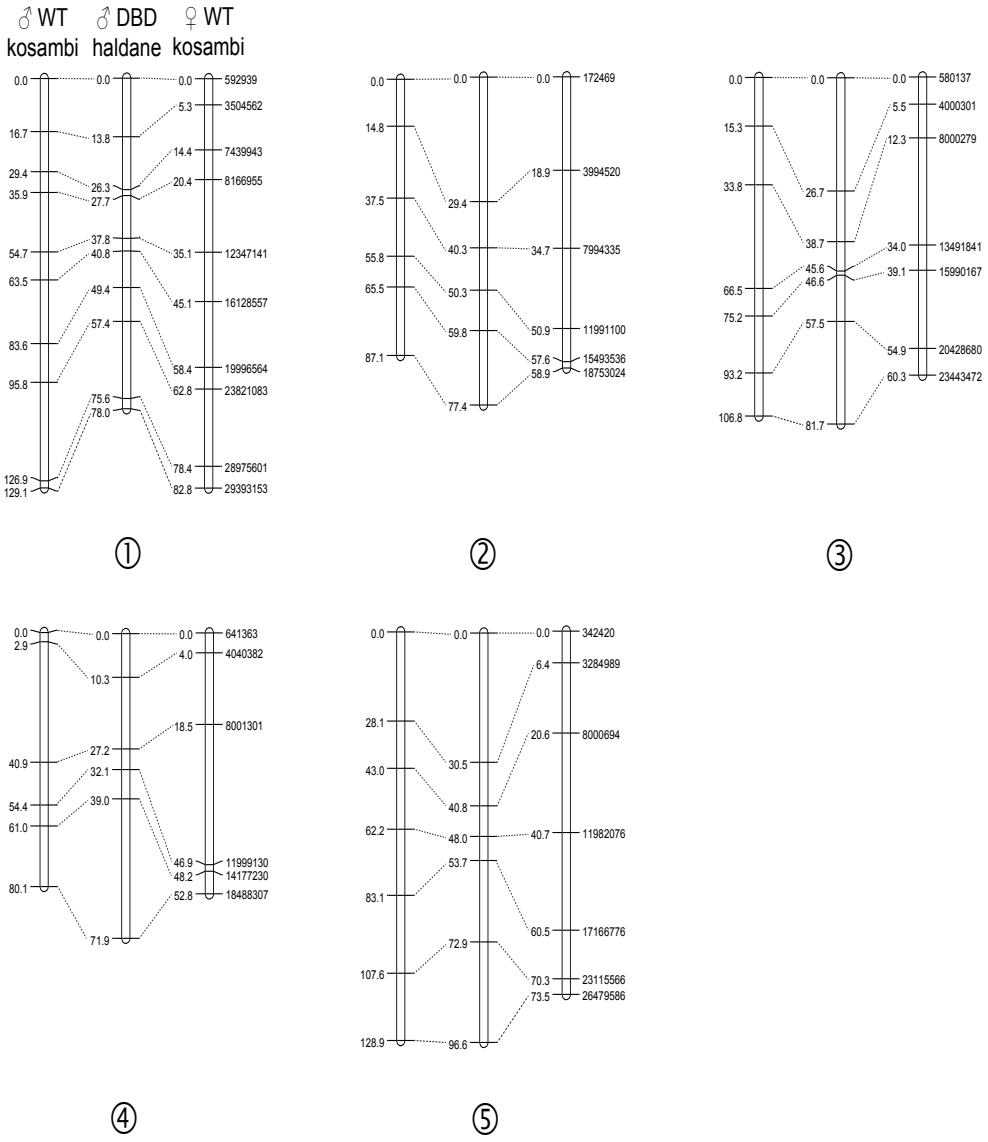
Supplementary Figures



Suppl. Figure 1. Joinmap linkage maps of the five *Arabidopsis thaliana* chromosomes. Genetic maps of WT male meiosis are shown left, and *cdka;1*-DBD male meiotic maps right. Recombination frequencies were estimated using Haldane mapping function.



Suppl. Figure 2. Segregation distortion in WT and *cdka*;1-DBD offspring. WT male data are shown on top. These show a slight difference with previous data, as in our WT cross where we observed an over-representation of Columbia chromosome 1 in offspring, whereas Wijnker *et al.* (2012) show an over-representation of only the long arm of chromosome 1. Segregation distortions for WT female are shown in the middle, and the lower graph represents *cdka;1*-DBD. Note in *cdka*;1-DBD that there is a strong segregation distortion on the top of chromosome 2, possibly due to segregation of the *cdka*;1-DBD rescue construct, which may have been present on the L*er* chromosome 2.



Suppl. Figure 3. Joinmap linkage maps of the five *Arabidopsis thaliana* chromosomes comparing estimates of WT male (left) and WT female (right), both corrected for interference using the Kosambi mapping function. The *cdka;1*-DBD male map is shown in the middle, and is not corrected for interference (by using Haldane mapping function).

On (our view of) evolution…

Ik droomde, dat ik langzaam leefde ….
langzamer dan de oudste steen.
Het was verschrikkelijk: om mij heen
schoot alles op, schokte of beefde,
wat stil lijkt. 'k Zag de drang waarmee
de bomen zich uit de aarde wrongen
terwijl ze hees en hortend zongen;
terwijl de jaargetijden vlogen
verkleurende als regenbogen …..
Ik zag de tremor van de zee,
zijn zwellen en weer haastig slinken,
zoals een grote keel kan drinken.
En dag en nacht van korte duur
vlammen en doven: flakkrend vuur.
- De wanhoop en welsprekendheid
in de gebaren van de dingen,
die anders star zijn, en hun dringen,
hun ademloze, wrede strijd ….
Hoe kón ik dat niet eerder weten,
niet beter zien in vroeger tijd ?
Hoe moet ik het weer ooit vergeten ?

M. Vasalis,

(In: 'Uit Parken en Woestijnen', 1940).

Chapter 4

Chromosome evolution in *Solanum* traced by cross-species BAC-FISH

Auteurs:
Dóra Szinay[1]*, T.G. (Erik) Wijnker[1]*, Ronald van den Berg[2], Richard G. F. Visser[3], Hans de Jong[1] and Yuling Bai[3,4]

1) Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands;
2) Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands;
3) Wageningen UR Plant Breeding, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands
4) Author for correspondence: bai.yuling@wur.nl
*) These authors contributed equally to this work.

New Phytologist (2012) 195(3), pp 688-698
doi: 10.1111/j.1469-8137.2012.04195.x

Summary

• Chromosomal rearrangements are relatively rare evolutionary events and can be used as markers to study karyotype evolution. This research aims to use such rearrangements to study chromosome evolution in *Solanum*.

• Chromosomal rearrangements between *Solanum* crops and several related wild species were investigated using tomato and potato bacterial artificial chromosomes (BACs) in a multicolour fluorescent in situ hybridization (FISH). The BACs selected are evenly distributed over seven chromosomal arms containing inversions described in previous studies. The presence / absence of these inversions among the studied Solanum species were determined and the order of the BAC-FISH signals was used to construct phylogenetic trees.

• Compared with earlier studies, data from this study provide support for the current grouping of species into different sections within Solanum; however, there are a few notable exceptions, such as the tree positions of S. etuberosum (closer to the tomato group than to the potato group) and *S. lycopersicoides* (sister to *S. pennellii*). These apparent contradictions might be explained by interspecific hybridization events and / or incomplete lineage sorting.

• This cross-species BAC painting technique provides unique information on genome organization, evolution and phylogenetic relationships in a wide variety of species. Such information is very helpful for introgressive breeding.

Introduction

Chromosome painting based on fluorescent *in situ* hybridization (FISH) can detect individual chromosomes in nuclei and cell complements, and represents one of the most common cytogenetic methods for establishing structural and numerical chromosomal variants in all eukaryotic model species, including yeast, *Arabidopsis* and man (Lysak *et al.*, 2001; Schubert *et al.*, 2001). The method is also very powerful in demonstrating large-scale chromosomal rearrangements that may be responsible for or accompanied by unique events leading to evolutionary divergence (e.g. Müller *et al.*, 2003). In plants chromosome painting has been a big challenge for a long time because repeats in pooled DNA probes from isolated chromosomes paint all chromosomes equally and thus do not allow such probes to detect individual chromosomes (Schubert *et al.,* 2001). This issue was overcome by Lysak *et al*. (2001) who selected repeat-poor bacterial artificial chromosomes (BACs) covering whole chromosome arm euchromatin regions as probes for FISH detection of individual *Arabidopsis thaliana* chromosomes. Later comparative painting studies of *Arabidopsis* BAC probes on chromosome complements of related *Brassica* species under lower stringency allowed comparative chromosome painting among related species, revealing the evolution of their karyotypes (Lysak *et al.*, 2003, 2007; Lysak & Lexer, 2006). Recently, the production and allocation of various tomato and potato BAC libraries has allowed adaptation of the cross-species BAC-FISH mapping technique to *Solanum* crops and wild species, and revealed cytogenetic evidence of known and novel

chromosomal rearrangements between tomato, potato and related *Solanum* species (Iovene *et al.*, 2008; Tang *et al*., 2008; Lou *et al.*, 2010).

Chromosome (Tomato (T))	wilde relative of tomato	potato (P)	Eggplant (E)	Pepper (C)	Citation[1]
5		5S inversion	5L + 12 L 5S and 5L inversion	5S and 5L inversion	a, b, c, d
6	*Mi*-homologues on 6S inverted in *S. peruvianum* PI 128657; 6S Inversion in *S. juglandifolium* LA2788	Upper[2]	6 inversion	Co-linear	a, f, e, j
7	7S Inversion in *S. pennellii* LA716		Two inverted segments	Upper[2] 7 scattered	a, d, i
9		9S inversion	Nine inversions	Nine inversions; additional rearrangements	a, b, c
10	10S Inversion in *S. juglandifolium* LA2788; 10L inversion in *S. sitiens* LA1974 and *S. Lycopersicoides* LA2951	10L inversion	5S + 12S + 10L Lower[2] 10 inversion	Lower[2] 10 inversion	a, b, c, f, h
11		11S inversion 11 inversion 4S + 11S	11S inversion Lower[2]	T11S = C12L 11S inversion (indication)	a, b
12	12S proximal inversion in *S. chilense* LA0458 Reciprocal translocation with 8S in *S. ochranthum* LA3650 and *S. Juglandifolium* LA2788	12S inversion 12S + 11L Lower[2] 12 inversions with T and P	Upper T = E = C	Upper[2] T=E=C Translocation 12S 11S	a, b, g, j

Tomato *(S. lycopersicum*) is used as the reference. Chromosome arms are named with numbers (chromosome number) followed by either S (short arm) or L (long arm).

1) a, Doganlar *et al.* (2002); b, Tanksley *et al.* (1992); c, Bonierbale *et al.* (1988); d, Livingstone *et al.* (1999); e, Seah *et al.* (2004); f, Canady *et al.* (2006); g, Stamova & Chetelat (2000); h, Pertuzé *et al.* (2002); i, van der Knaap *et al.* (2004); j, Albrecht & Chetelat (2009).

2) Description of the chromosome part is according to Doganlar *et al.* (2002)

Table 1. Overview of chromosomal rearrangements in *Solanum* and *Capsicum* based on comparative genetic mapping

The study of the genetic and cytogenetic relationship between tomato and potato has a long history. Gottschalk's (1954) pioneering work in chromosome morphology revealed a surprising similarity in overall chromosome morphology among tomato, potato and several other *Solanum* species. Later studies of the genetics and genomics of *Solanum*

crops revealed large-scale synteny (Bonierbale *et al.*, 1988; Tanksley *et al.*, 1992; Grube *et al.*, 2000; Doganlar *et al.*, 2002; Fulton *et al.*, 2002; Wu & Tanksley, 2010), but also demonstrated varying numbers of translocations and inversions between tomato (*Solanum lycopersicum*), potato (*S. tuberosum*), eggplant (*S. melongena*) and pepper (*Capsicum* spp.) (Table 1). Such major chromosomal rearrangements, which are supposedly relatively rare and independent events, are one of the dramatic processes that shape the genome and the karyotype, and thus can be used as phylogenetic markers for the study of chromosome evolution in the Solanaceae family.

Phylogenetic relationships between *Solanum* species have been the subject of various studies (e.g. Bohs & Olmstead, 1997; Weese & Bohs, 2007). Many studies have focused on the economically important species of section *Lycopersicon* (wild and cultivated tomatoes, *e.g.* Peralta & Spooner, 2001; Spooner *et al.*, 2005; Peralta *et al.*, 2008), section *Petota* (wild and cultivated potatoes, *e.g.* Spooner & Raul Castillo, 1997; Jacobs *et al.*, 2008) and

Species	Accession /introgression line /cultivar	Taxonomy[1]
Solanum peruvianum	LA2172	Section *Lycopersicon*
S. peruvianum	LA2157	
S. habrochaites	G1.1290	
S. habrochaites	G1.1560	
S. pimpinellifolium	G1.1554	
S. habrochaites f. glabratum	CGN.1561	
S. lycopersicum	Heinz 1706	
S. chilense	LA1969	
S. pennellii	LA716	
S. lycopersicoides	CGN90124 (PI255549 or PI365378)	Section *Lycopersicoides*
S. ochranthum	LA2166	Section *Juglandifolia*
S. etuberosum	PI558054	Section *Etuberosum*
S. bulbocastanum	PI275198	Section *Petota*
S. tarijense	CGN22729	
S. megistracrolobum	PI458396	
S. pinnatisectum	GLKS2268	
S. tuberosum	RH-89-039-16; van der Voort *et al.* (1997)	
S. melongena	Half Lange Violette and MM738	Section *Melongena*
1) The taxonomic classification is according to Peralta *et al.* (2008) except for *S. melongena* that belongs to subgenus *Leptostemonum*, section *Melongena*.		

Table 2. Plant materials used for this study.

the interrelationships between these groups (Spooner *et al.*, 1993; Rodriguez *et al.*, 2009). The different types of data used in these studies (morphology, AFLPs, sequences of chloroplast DNA, internal transcribed spacer (ITS), the GBSSI gene and COSII markers) have often resulted in conflicting phylogenetic reconstructions.

In this paper we present the analysis of inversion events as a parameter of chromosomal evolution in *Solanum* species, with an emphasis on species related to tomato and potato. We detected chromosomal rearrangements by cross-species BAC-FISH using tomato and potato BACs as probes on chromosome complements of various *Solanum* species of the tomato and potato clade. With *S. melongena* as the outgroup representative, we constructed a phylogenetic tree that was then compared with a generalized tree derived from a number of the studies mentioned above. We discuss the power of this chromosome painting technique in relation to other techniques such as comparative genetic mapping (Wu & Tanksley, 2010), DNA sequences comparison (Wang *et al.*, 2008; Wu *et al.*, 2009a,b) and chromosome paring analysis of F1 hybrids (de Jong *et al.*, 1993; Anderson *et al.*, 2010).

Materials and Methods

Plant materials

The plants used in this study were from section *Lycopersicon* (six species), section *Lycopersicoides* (one species), section *Juglandifolia* (one species), section *Etuberosum* (one species) and section *Petota* (five species) (Table 2). In addition, two eggplant cultivars were included (Table 2). Plants were grown in a glasshouse and young flower buds were collected in the morning and immediately fixed in freshly prepared acetic acid / ethanol (1 : 3) at 4°C. The next day, the flower buds were transferred to 70% ethanol for storage at 4°C. In total, 18 tomato and 17 potato BACs (Fig. 1) were selected covering seven chromosome arms (chromosomes 5, 6, 7, 9, 10, 11 and 12; five BACs per chromosome arm), where previous papers on comparative genetics suggested inversions. All BACs were repeat-poor, except for H146I19 that hybridized to the heterochromatin of several chromosomes of all species of the subsection *Lycopersicon*. This BAC was used because no repeat-poor BAC was available in the middle of the euchromatin on the short arm of chromosome 12.

Slide preparation

Slides were prepared according to the protocol of Szinay *et al.* (2008) with the following minor modifications. The standard enzyme mix containing 1% pectolyase Y23 (Sigma P-3026), 1% cellulase RS (Yakult 203033; Yakult Pharmaceutica, Tokyo, Japan) and 1% cytohelicase (Bio Sherpa 24970-014) was diluted 10 times with 10 mM sodium citric buffer

(pH 4.5) for most of the species. For species that have pollen mother cells with thick callose walls the dilution was less: *Solanum melongena* (three times dilution); *S. pennellii, S. lycopersicoides* and species of section *Petota* (two times diluted stock).

BAC labelling

BACs were isolated as described in Szinay *et al.* (2008). In some cases we used the High Pure Plasmid Isolation Kit (Roche 11754785001) (Szinay *et al*., 2008). BAC DNA was labelled by nick translation following the manufacturer's protocol (Roche (http://www.roche.com). The following direct labelling systems with dXTPs were used: Cy3-dUTP (Amersham, http://www.gelifesciences.com/), Cy3.5-dCTP (Amersham) and Diethylaminocoumarin-5-dUTP (DEAC; Perkin Elmer, http://www.perkinelmer.com), and two indirect labelling systems of dUTPs labelled with biotin and digoxigenin, respectively. For painting BACs on chromosome 12, we used Cot-100 DNA to block labelled repetitive sequences in BAC

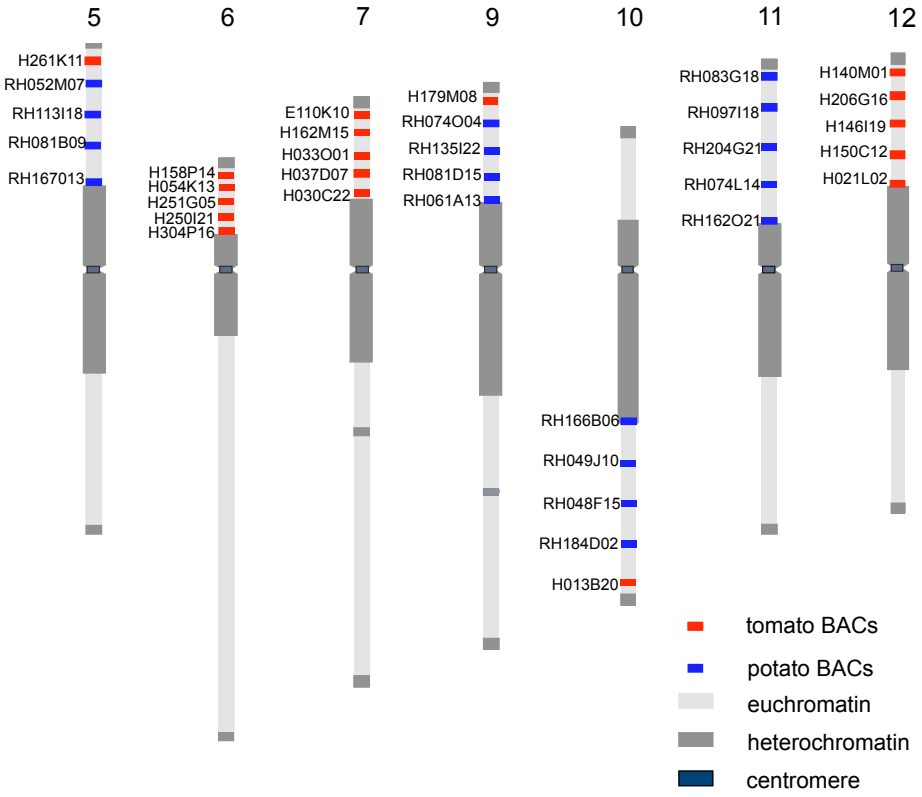


Figure 1. Chromosomal location of tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) BACs on the seven studied chromosome arms, based on BAC-FISH signals on the tomato chromosomes.

H146I19 from hybridization (Peterson *et al.*, 1998; Chang *et al.,* 2008). Cot-100 DNA was isolated according to Szinay et al. (2008).

FISH procedure and data analysis

FISH was performed according to the protocols of Rens *et al.* (2006) and Szinay *et al.* (2008) with the following modifications. Hybridization was carried out for 2 or 3 d, followed by a post-hybridization wash in a series from 82% to 64% formamide at 42°C (Schwarzacher & Heslop-Harrison, 2000) for 3 - 5 min. The biotin-labelled probes were amplified three times for 45 min with streptavidin conjugated with Cy5 (Invitrogen, http://www.invitrogen.com/site/us/en/home.html) and biotinylated.anti-streptavidin (Vector laboratories, http://www.vectorlabs.com/). The digoxigenin-labelled probes were amplified twice with anti.digoxigenin-FITC (Roche, http://www.roche.com) and anti-sheep-FITC (Invitrogen, http://www.invitrogen.com/site/us/en/home.html). Microscopy and FISH data interpretation were carried out as described by Szinay *et al. (2008).*

Phylogenetic analysis

The order of the BAC FISH signals was established on selected chromosome arms (Fig. 1), except for chromosome 5 and 12 on *S. melongena*. Doganlar *et al.* (2002) pointed out that the short arms of chromosome 5 and 12 of tomato translocated to chromosome 10 of *S. melongena*. We thus used missing values for BACs on chromosome 5 and 12 on *S. melongena*. In addition, BAC H037D07 on chromosome 7 and RH162O21 on chromosome 11 were not included in the phylogenetic analysis as a result of their insufficient hybridization on *S. etuberosum* and *S. melongena*.

We explored a number of coding strategies and methods of analysis. We used SoRT2 (Huang *et al.*, 2010) to infer a phylogenetic tree based on pairwise genome rearrangement distances. Treeview was used to visualize the output files of SoRT2 (text files in Newick format) and to root the tree with *S. melongena*. In another approach, we coded different BAC-orders as (unordered) character states for each chromosome and performed a maximum parsimony (MP) analysis with PAUP v4.0 (Swofford, 1999) (MP, chromosome coding). In a third approach, we adopted the method of Müller *et al.* (2003) to derive discrete characters by coding the presence or absence of adjacent chromosomal segments in a binary data matrix. A MP analysis on this data matrix was performed in PAUP v4.0 (Swofford, 1999) and Jackknife analyses were performed with 10 000 replicates (MP, segment coding). The most parsimonous evolutionary history of the events on the investigated chromosome arms was reconstructed and illustrated as a phylogenetic tree in Fig. 4 (left panel).

The topology of our phylogenetic reconstruction was compared with a generalized tree derived from published phylogenies of a similar set of materials (Spooner *et al.*,

1993, 2005; Peralta & Spooner, 2001; Rodriguez *et al.*, 2009). The material investigated in those studies included representatives of section *Etuberosum* (usually *S. etuberosum*, sometimes also *S. palustre*), section *Petota* (represented by a varying number of species), sec.tion *Lycopersicoides* (with *S. lycopersicoides* and *S. sitiens*), section *Juglandifolia* (*S. juglandifolium* and *S. ochranthum*) and section *Lycopersicon* (*S. lycopersicum* and related species). The trees in the papers by Peralta & Spooner (2001, Fig. 4), Spooner *et al*. (2005, Fig. 7) and Rodriguez *et al.* (2009, Fig. 4) show an identical topology, while the tree in the paper from Spooner *et al.* (1993) deviates slightly in placing *S. sitiens* (section *Lycopersicoides*) closest to *S. ochranthum* (section *Juglandifolia*), separated from the other species of section *Lycopersicoides* and *S. juglandifolium*. Although these branches have bootstrap support of only 51% and 52%, we give an informal 'consensus topology' in Fig. 4 (right panel).

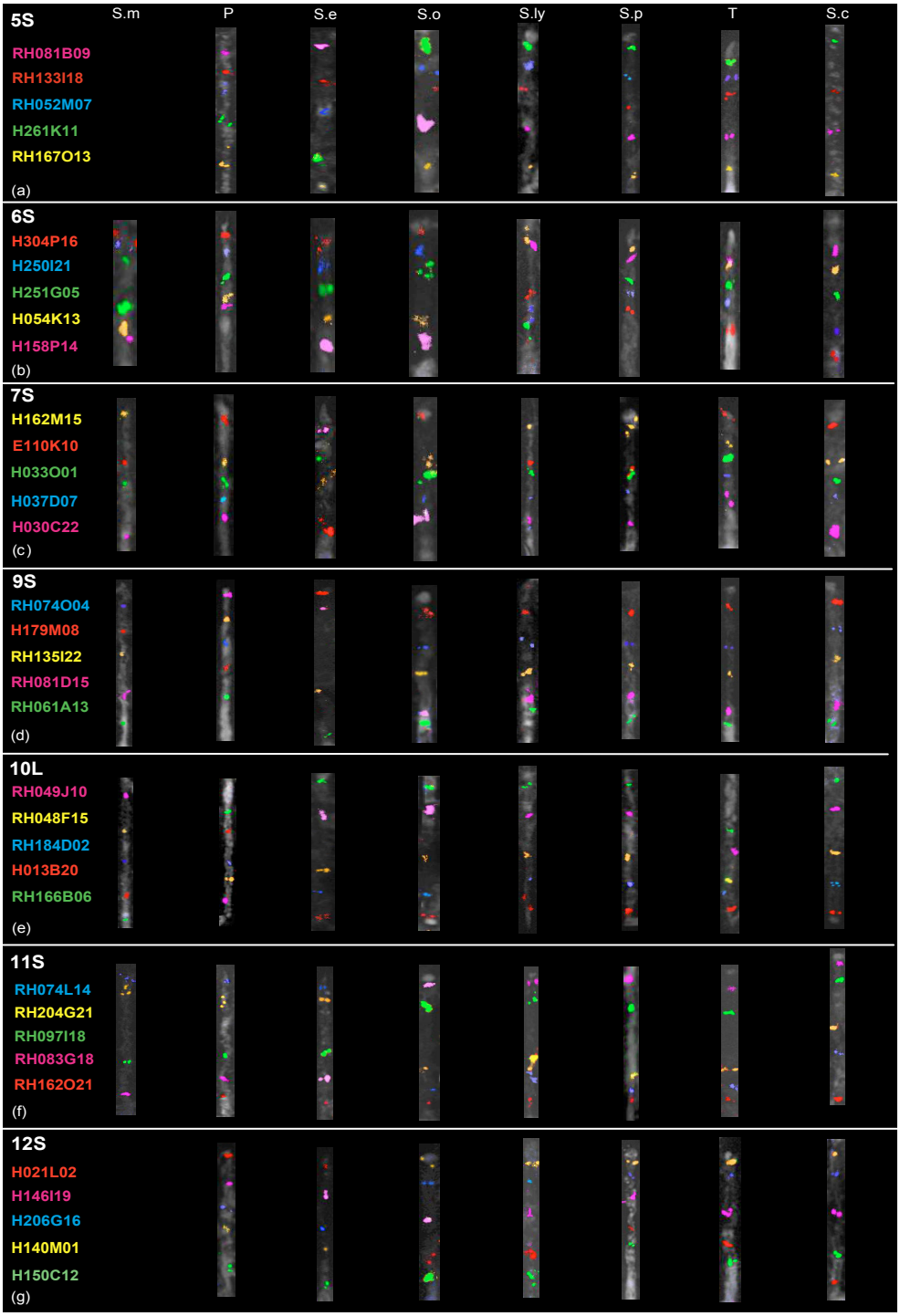
Results

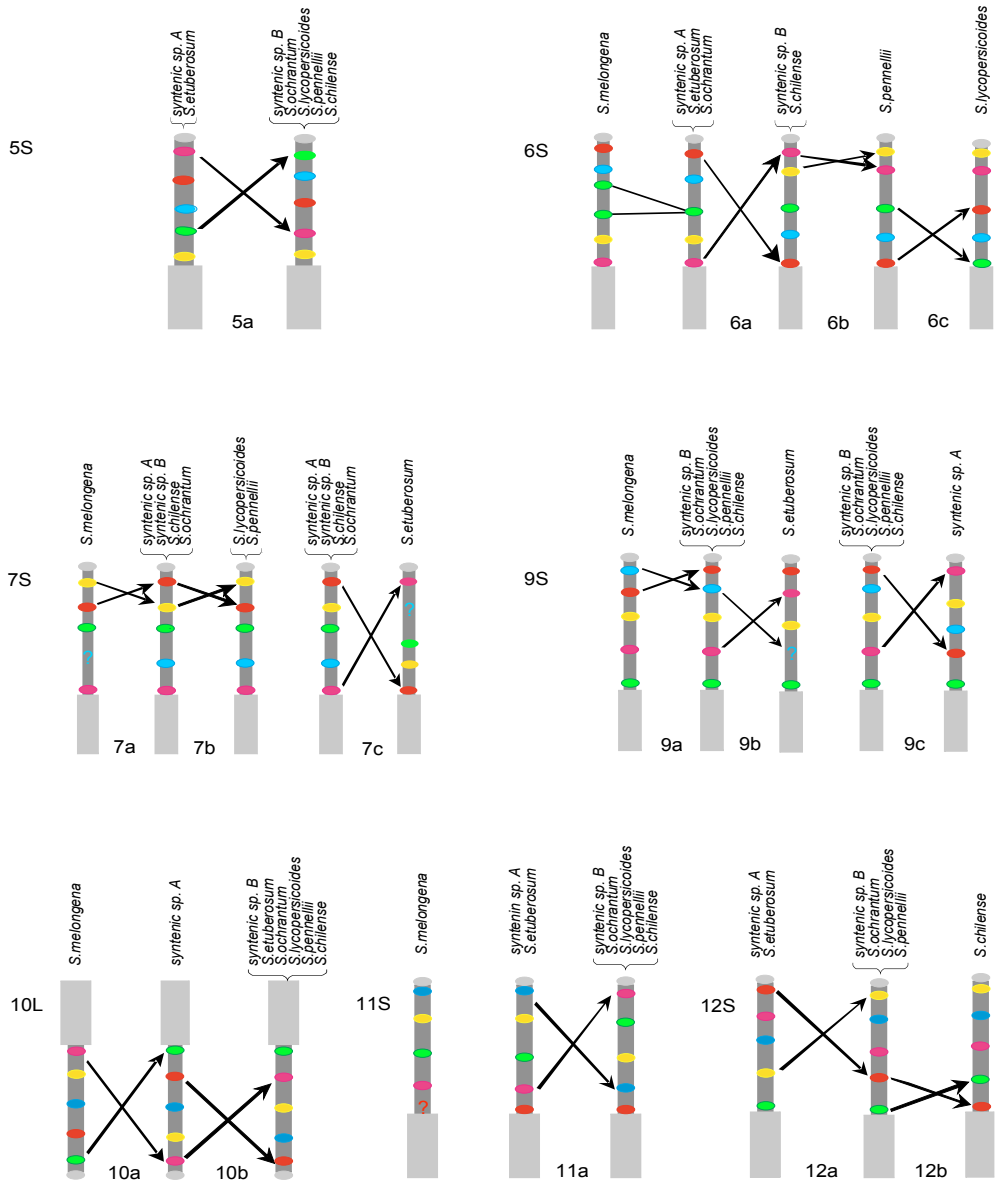
Chromosomal rearrangements revealed by FISH

For establishing chromosomal rearrangements in the *Solanum* species we selected BACs on seven chromosome arms (Figs 1, 2) that were known from comparative genetic studies of *Solanum* crops to contain inversions (Table 1). We use here the term 'syntenic species' to indicate those species with identical BAC-FISH patterns on all studied chromosome arms (Figs 2, 3). The first group of syntenic species (syntenic sp. A) includes potato (*S. tuberosum*) and its wild relatives *S. bulbocastanum*, *S. tarijense*, *S. megistacrolobum* and *S. pinnatisectum*; the second one (syntenic sp. B) comprises tomato (*S. lycopersicum*) and its wild relatives *S. peruvianum*, *S. habrochaites* and *S. pimpinellifolium*. These two syntenic species groups show different BAC orders on chromosomes 5, 6, 9, 10, 11 and 12 (see Fig. 3 for schematic representations, namely 5a, 6a, 9c, 10b, 11a and 12a).

In comparison to syntenic sp. B: two proximal BACs display an inverted order on chromosome 12S (S, short arm) of *S. chilense* (Fig. 2g; 12b in Fig. 3); all BACs on 6S are inverted in *S. ochranthum* (Fig. 2b, 6a in Fig. 3); two distal BACs on 6S (Fig. 2b; 6b in Fig. 3); and 7S (Fig. 2c; 7b in Fig. 3) is inserted in *S. lycopersicoides* and *S. pennellii*. Between *S. lycopersicoides* and *S. pennellii*, three proximal BACs on 6S are inverted (Fig. 2b; 6c in

Figure 2. Opposite page.
BAC-FISH images showing an overview of BACs on pachytene of selected chromosomes. The leftmost column shows information on the chromo.some arms with their number followed by S (short arm) or L (long arm). The other columns show BAC FISH images on chromosomes of *Solanum melongena* (S.m), *S. tuberosum* (P), *S. etuberosum* (S.e), *S. ochranthum* (S.o), *S. lycopersicoides* (S.ly), *S. pennellii* (S.p), *S. lycopersicum* (T, represented by Heinz 1706), *S. chilense* (S.c). BAC names and their hybridization images on a specific chromosome arm are labelled with different colours. There are no images for 5S and 12S of *S. melongena* due to hybridization difficulties. →


S.m
P
S.e
S.o
S.ly
S.p
T
S.c
5S
RH081B09
RH133I18
RH052M07
H261K11
RH167O13
(a)
6S
H304P16
H250I21
H251G05
H054K13
H158P14
(b)
7S
H162M15
E110K10
H033O01
H037D07
H030C22
(c)
9S
RH074O04
H179M08
RH135I22
RH081D15
RH061A13
(d)
10L
RH049J10
RH048F15
RH184D02
H013B20
RH166B06
(e)
11S
RH074L14
RH204G21
RH097I18
RH083G18
RH162O21
(f)
12S
H021L02
H146I19
H206G16
H140M01
H150C12
(g)



Syntenic species A: *S. tuberosum, S. tarijense, S. megistracrolobum, S. bulbocastanum, S. pinnatisectum*
Syntenic species B: *S. pinnatisectum* N, *S. lycopersicum, S. pimpinellifolium, S. habrochaites.*

Figure 3. Schematic overview of chromosomal inversions in *Solanum*. Arrows indicate likely directions of inversions. Below the arrows the inversion events are numbered with a small letter in alphabetic order distinguishing the different inversions within one chromosome. Lines indicate inversions with unknown directions. Question marks are weak or missing FISH signals. Signals of BACs are shown by coloured dots.

Fig. 3). BAC orders in *S. etuberosum* are very similar to that in syntenic sp. A, except for inverted orders of BACs on 7S, 9S and 10L (L, long arm) (Fig. 2c–e; 7c, 9b and 10b in Fig. 3). On chromosome 10L, the BAC order in *S. etuberosum* is similar to the order in syntenic sp. B.

We could not obtain interpretable FISH signals for chromosome arms 5S and 12S of S. melongena, which is presumably caused by translocations involving these chromosome arms (Doganlar *et al*., 2002; Wu *et al*., 2009a,b). In comparison to all other species, BACs on 7S, 9S and 10L are inverted in *S. melongena* (Fig. 2c–e; 7a, 9a and 10a in Fig. 3). The order of BACs on 7S is similar to both *S. lycopersicoides* and *S. pennellii* (Fig. 3). Interestingly, BAC H251G05 showed double signals on 6S of *S. melongena* (Figs 2b, 3), which can be interpreted as breakpoints in the chromosomal target of this BAC.

Phylogenetic analysis

The trees we made differ somewhat depending on the methods used (datasets and resulting trees are presented as Supporting Information Figs S1–S3; Tables S1–S4), but a general pattern is clear: *S. etuberosum* and the clade (polytomy) including *S. pinnatisectum*, *S. bulbocastanum*, *S. megistracrolobum*, *S. tarijense* and *S. tuberosum* (syntenic sp. A) are placed basal in all trees. The remaining species are joined in a clade in all trees, in which all identify a clade joining *S. pennellii* and *S. lycopersicoides*, a branch with *S. ochranthrum*, and a branch including *S. peruvianum*, *S. habrochaites*, *S. pimpinellifolium* and *S. lycopersicum* (syntenic sp. B).

Distance-based approaches (SoRT2 by Huang *et al*. (2010) in our study: Fig. S1; Table S1) rely on overall similarity instead of (syn-)apomorphic character states, which make us favour the trees based on MP. The most conservative (and least resolved) tree of our two parsimony approaches results from a simple coding method in which specific BAC-orders on chromosomes are considered as (unordered) character states (Fig. S2; Table S2). In another approach (proposed by Müller *et al*., 2003), the presence and absence of adjacent loci are used as characters, leading to decidedly higher resolution, as well as high jackknife values (82–98%) for all clades except for the combination of the *S. lycopersicoides /S. pennellii* clade with the tomato clade, which received only 74% jackknife support (Fig. S3; Tables S3, S4). This method, however, considers the adjacent segments as independent characters which they formally are not (*i.e.* one inversion typically leads to two new characters). Because of this, we will use the simplest (and most conservative) MP method (based on BAC-orders) as base for our phylogenetic reconstruction. Under this method a character state can change to any other state with equal chance.

Phylogenetic reconstruction

After deriving a phylogenetic tree using the conservative MP method based on BAC-patterns (Fig. S2), we used this tree as a starting point to resolve the remaining polyto-

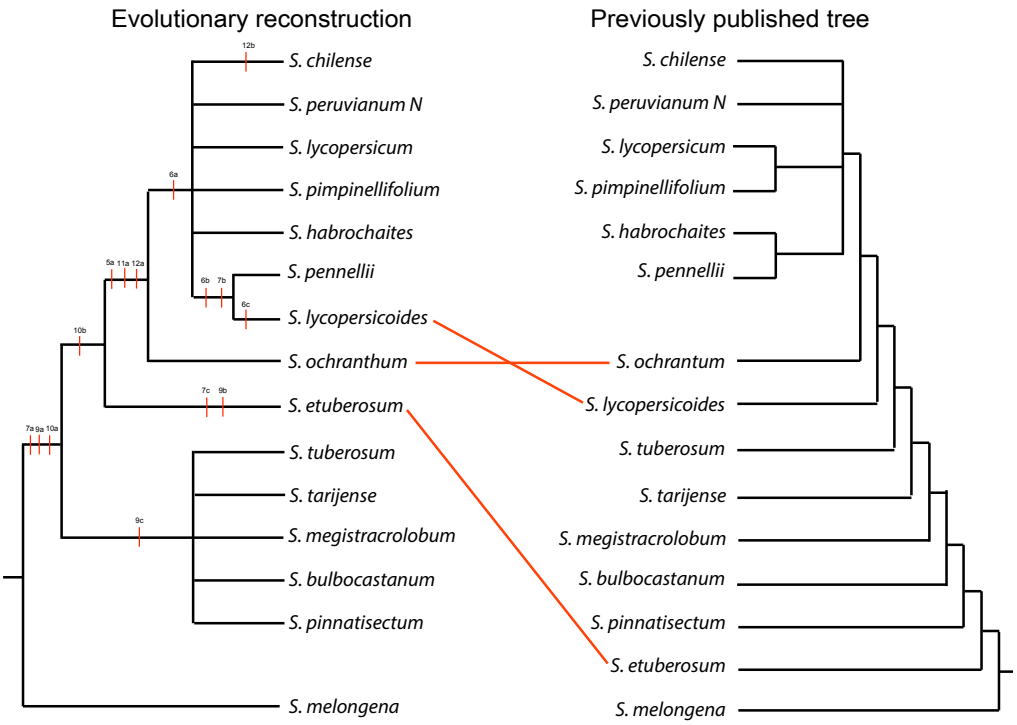


Figure 4. The topology of our phylogenetic reconstruction (left panel) in comparison with a generalized tree (right panel) derived from published phylogenies of a comparable set of material (see the Materials and Methods section; Spooner *et al.*,1993; Peralta & Spooner, 2001; Spooner *et al.*, 2005; Rodriguez *et al.*, 2009). The polytomy within the potato group in the generalized tree only indicates the lack of information about the relationships among these species in the used references. Red lines indicate the different positions of *Solanum lycopersicoides, S. ochranthum* and *S. etuberosum*. Inversion events that were numbered in Figure 3 are indicated on the tree (left panel).

mies by the analyses described below. The most likely way in which the karyotypes of the studied *Solanum* species evolved are presented in Fig. 3. By inferring ancestral and derived states with *S. melongena* as outgroup, we could place inversion events on the branches of our phylogenetic reconstruction (Fig. 4, left panel). This reconstruction is identical to the tree produced by MP tree based on segment coding (after Müller *et al.*, 2003).

The ancestral state for chromosome 5 could not be established because the order of BACs in the outgroup could not be deter.mined. Syntenic sp. A and *S. etuberosum* show a similar order of BACs, which differs from all other species that have the four distal BACs inverted. Recent literature (*e.g.* Peralta *et al.*, 2008) considers the potatoes (section *Petota*) as sisters to the tomatoes *s.l.* (sections *Lycopersicoides*, *Juglandifolia* and *Lycopersicon*) and section *Etuberosum* as sisters to the combined group of potatoes and tomatoes

s.l. Both *S. etuberosum* and section *Petota* can thus be considered basal compared to the tomatoes *s. str.* (section *Lycopersicon*). This leads us to consider the distal inversion (5a in Fig. 3) as a synapomorphy for the clade including the remaining species from *S. ochranthum* onward.

The ancestral BAC order for chromosome 6 is similar to our outgroup, as potatoes (syntenic sp. A) and *S. melongena* share the same order. However, there is a difference between the groups because the middle BAC shows two foci in *S. melongena*. This may point to a duplication /deletion event (a duplication in *S. melongena* or deletion in our ingroup) or a small interstitial inversion in which the chromosome breakpoint lies within the used BAC. Because *S. ochranthum* shows a similar BAC order to that of the potato clade and *S. etuberosum*, we consider *S. ochranthum* as the most basal of the wild tomatoes (and as such, we place *S. ochranthum* as a sister to the clade including *S. pennellii*, *S. lycopersicoides* and *S. lycopersicum*, Fig. 4). Looking at the remainder of species, the most parsimonious solution is to assume a whole arm euchromatin inversion giving rise to the order found in syntenic sp. B and *S. chilense*. A distal inversion then joins *S. pennellii* and *S. lycopersicoides* in a clade, and a proximal inversion separates *S. lycopersicoides* from *S. pennellii*. This hypothesis (placing *S. pennellii* and *S. lycopersicoides* as sister taxa) contrasts with all existing phylogenetic reconstructions as proposed by Spooner and collaborators (Spooner *et al.*, 1993, 2005; Peralta & Spooner, 2001; Rodriguez *et al.*, 2009) in which *S. lycopersicoides* is usually considered (together with *S. sitiens*) to be sister to a clade consisting of the species of section Juglandifolia and section *Lycopersicon* (including *S. pennellii*). If we follow previously proposed phylogenies, we must assume a complex rearrangement (in which the *S. lycopersicoides* BAC order is generated directly from the ancestral *S. melongena*/potato type), followed by two consecutive inversions giving rise to *S. pennellii* and later to syntenic sp. B. We here favour the simple inversion scenario for two reasons: first, inversions are more common than complex rearrangements that involve multiple breaks, as all inversions we found can be interpreted as simple inversions; and second, the top inversion on 7S supports the sister group relationship between *S. pennellii* and *S. lycopersicoides*.

The ancestral karyotype for chromosome 7 is easily established by the fact that *S. melongena* and *Capsicum annuum* (data not shown) share the same order of the BACs involved. Supporting evidence for this comes from comparative genomics based on COS-markers which also indicates that *S. melongena* is collinear with *C. annuum* (Doganlar *et al.,* 2002). A top inversion of the most distal BACs led to the BAC order found in most species. Three species show another order of BACs: *S. etuberosum* has the whole arm inverted, and *S. pennellii* and *S. lycopersicoides* have the ancestral BAC order. The most parsimonious explanation is based on the assumption that a reversal took place in these two species, joining these two species together in a clade separate from all other species. As mentioned in the Results section on chromosome 6, this concurs well with the most parsimonious solution for events that happened for 6S. If we assume that *S. lyco-*

persicoides and *S. pennellii* have the ancestral karyotype, we would introduce various inconsistencies to our tree: either various incidences of parallel (convergent) evolution in three independent lines, or the placement of *S. pennellii* and *S. lycopersicoides* in a position basal to the potato clade, which is very unlikely given all other information (and introducing many conflicts to the tree).

The evolutionary scenario on chromosome 9 is straightforward. As *Capsicum annuum* and eggplant are collinear (Wu et al., 2009a,b; Wu & Tanksley, 2010), we may assume that *S. melongena* represents the ancestral type. A top inversion gave rise to the BAC order found in most species (9a in Fig. 3). A subsequent interstitial inversion of three BACs gave rise to the *S. etuberosum* type, whereas a proximal inversion is unique to *S. chilense.*

Chromosome 10L provides a unique insight into the position of *S. etuberosum* with respect to the potato clade. *Solanum etuberosum* is usually placed basal to a combined potato–tomato clade (Fig. 4, right panel). All investigated species (except for our outgroup *S. melongena*) have the same (green) BAC placed proximally, suggesting that this is ancestral for the whole group. The question then is to determine what type (potato type or the type of the other species is ancestral). The use of *S. melongena* as out-group for chromosome 10 must be done with caution, as previous research showed that *S. melongena* experienced extensive rearrangements on chromosome 10. Eggplant 10L is a mosaic of tomato chromosomes 5, 10 and 12 (Wu *et al*., 2009a). The *S. melongena* type is nevertheless very close to the potato type, because it can be obtained by a single whole arm inversion from the potato type, as was also shown by the analysis by Wu *et al.* (2009a). From Wu et al. (2009b) and Wu & Tanksley (2010) it can be inferred that *Capsicum annuum* chromosome 10 has a similar order to the potato. This substantiates the hypothesis of the potato/*Capsicum*-type representing the ancestral state, and the tomato /*S. etuberosum* type being derived. We therefore place syntenic sp. A basal to *S. etuberosum*, which in turn becomes the sister of a group including all other species, including tomato.

The ancestral state on chromosome 11 is represented by *S. melongena*, because it shares its BAC order with syntenic sp. A and *S. etuberosum*. The inversion of the four most distal BACs joins *S. ochranthum* with the tomato and other species. Rearrangements on chromosome 12S are straightforwardly placed on the tree. Although due to translocations we have no data on eggplant, the proximal position of the green BAC is ancestral (*i.e.* present in all species, except *S. chilense*). If we are right in assuming that syntenic sp. A and *S. etuberosum* should be placed basally to the other investigated species, inversion 12a (Fig. 3) is a shared derived character for *S. ochranthum* and its sister clade (including *S. chilense*).

Discussion

The aim of our study is to use chromosomal rearrangements as phylogenetic markers for the study of chromosome evolution in the Solanaceae family. We show that the hybridiza-

tion of BACs on the chromosomes of crops and their related species enables us to confirm directly the genomic collinearity, to show inversions between their homeologues, and to reconstruct the most likely way in which the karyotypes of the studied species evolved.

Comparison to published phylogenies

The reconstruction of *Solanum* phylogeny has a long history, in which the proposed interrelationships between potato, tomato and their close relatives have changed many times when new data became available. To compare the topology of our reconstructed tree (Fig. 4, left panel) with existing data, we constructed a generalized tree derived from published phylogenies of a comparable set of material (Spooner *et al.*, 1993, 2005; Peralta & Spooner, 2001; Rodriguez *et al.*, 2009). In this generalized tree (Fig. 4, right panel), the following sistergroup relationships are present: after the outgroup consecutive branches lead to the representatives of sections *Etuberosum*, *Petota*, *Lycopersicoides*, *Juglandifolia* and *Lycopersicon*. We note three remarkable differences between our tree and the generalized tree: (1) the position of *S. etuberosum*, (2) the position of *S. ochranthum*, and (3) the presence of a clade containing *S. pennellii* and *S. lycopersicoides*. Based on the inversion in 10L that *S. etuberosum* shares with all species higher up in the tree, the clade containing potato and its wild relatives (section Petota) is placed basally to *S. etuberosum* in our tree, whereas the reverse is suggested by all earlier studies. The basal placement of *S. ochranthum* with respect to *S. lycopersicoides* is a consequence of the order of BACs on chromosome 6. The grouping of *S. pennellii* in a clade with *S. lycopersicoides* is supported by *S. pennellii* and *S. lycopersicoides* sharing an inversion on chromosome 7 and a presumed additional inversion on chromosome 6. This poses a strong contrast with almost all previously published trees where *S. pennellii* is firmly nested in a 'tomato group' (Peralta & Spooner, 2001; Spooner *et al*., 2005; Rodriguez *et al*., 2009), with one exception where *S. pennellii* was suggested to be closest to *S. lycopersicoides* (Zuriaga *et al*., 2009).

Our evolutionary hypothesis is exactly similar to the MP-segments tree that we generated following the method of Müller *et al.* (2003), and shows significant differences with the phylogenetic trees as previously published. Notably, our data are not the first chromosome data that challenge evolutionary relation.ships proposed earlier: a recent study on pairing configurations in spreads of synaptonemal complexes of tomato - wild species hybrids (Anderson *et al.*, 2010) casts doubt on the proposed phylogeny based on chromosome synteny. In that study *S. pennellii* and *S. habrochaites* (two species that are sister according to Rodriguez *et al.* (2009)) showed remarkably high levels of pairing irregularities, suggesting that both were equally far apart from one another as they are from tomato, which was thought to be more distantly related.

We also note that the data we obtained for chromosome 10 are not entirely in concordance with previously published data. Compared with the results reported by Doganlar *et al.* (2002) and Wu *et al.* (2009a), our data showed a reversed orientation of

BACs on *S. melongena* for the whole arm. We identified a large inversion comprising most of the euchromatic part between *S. lycopersicoides* and *S. tuberosum*; while Pertuzé *et al.* (2002) described marker synteny between these two species. In addition, we found that *S. lycopersicum* and *S. lycopersicoides* are collinear, while Chetelat *et al.* (2000) suggested chromosomal rearrangements between these two species. A possible explanation for all of these observed differences is that different accessions were used in the different studies and that chromosomal rearrangements exist among accessions within one species. Another possibility is that the genetic studies misinterpreted the inversions due to low marker density or suppression of recombination.

Previous trees were based on extensive datasets using a multitude of markers, sequences and morphological traits. How can we best explain these apparent inconsistencies? Chromosomal rearrangements behave no differently from other markers, and their correct interpretation might be obscured by similar 'noise' to which all characters are subjected such as homoplasy (parallel evolution or reversals) or complex modes of speciation. The possibility of interspecific hybridization or incomplete lineage sorting should surely be considered when studying the evolutionary history of *Solanum*, especially in the light of the documented crossability between various investigated species. Interspecific hybridization followed by backcrosses to either one of the parents could lead to introgression of new alleles into a recipient species. In case of inversions, there is the possibility of introgression of whole inversions into a new background, becuase recombination will unlikely happen in small inverted segments during meiosis (Verlaan *et al.,* 2011). The possibility of hybridization followed by introgression of inverted segments could account for, for example, the small distal inversions on 6S and 7S that join *S. lycopersicoides* and *S. pennellii* in a clade separate from other species.

Notes on rearrangements

It is remarkable that 75% (12 out of 15) of the inversions we studied involved the most distally placed BAC on chromosome arms. It is unknown whether these inversions are the result of complete arm inversions (*i.e.* including the telomere) or whether there was a breakpoint in the sub-telomere heterochromatin.

One inversion on chromosome 7S is remarkable, as it represents the only character reversal. Given our observation that 75% of inversions involve the distal-most BAC, the occurrence of a distal reversal might not be entirely unpredicted: any breakpoint in a similar interval as the inversion preceding it, will in the far majority of inversions result in a reversal. We also note that the reuse of chromosomal breakpoints has frequently been documented in other species (Wu & Tanksley, 2010).

When the placement of rearrangements on the phylogenetic tree is considered, inversion events appear to cluster on the branch leading to *S. ochranthum* and its sister group. Such a clustering could point to an evolutionary 'bottleneck' in which a small population

fixed a number of inversions, radiating into many species at a later point. Alternatively, the presence of many rearrangements could be taken to indicate the passage of a long time period (assuming something like a constant 'inversion rate'). The absence of rearrangements within the potato clade among the studied wild and cultivated potato species suggests that these represent relatively recent splits. This is not surprising in the light of recent publications (Jacobs *et al.*, 2008, 2011). The split between the close relatives of tomato (*i.e.* commonly referred to as the section *Lycopersicon*) shows considerably more variation, suggesting that these species might have started diverging earlier than the potatoes.

Conclusions

Our study features advances in molecular cytogenetic tools that support plant genetics, genomics and breeding programmes. In general, our data support the current grouping of *Solanum* species into different sections: most previously defined sections are identified herein as having an unique order of chromosomal segments. There are nevertheless a few remarkable differences in our phylogenetic reconstruction compared to earlier studies. The apparent conflict between our hypothesis and previously proposed hypotheses points to the possibility that the evolutionary history of Solanum may have seen its share of reticulation: interspecific hybridization followed by introgression of inverted segments. Alternatively, incomplete lineage sorting may have played a role in the apparent complexity of Solanum phylogeny.

The progress of more and better sequencing practices will reveal novel and detailed information on chromosome rearrangements among species. The occurrence of chromosomal rearrangements stresses the important of a correct physical map in ordering scaffolds of related species that are being re-sequenced. The cross-species BAC-FISH method as presented here will remain an indispensible tool in comparative genomics as it reveals chromosomal rearrangements without a priori de novo sequencing of the species involved. In addition, cross-species BAC-FISH can detect chromosomal rearrangements involving heterochromatin areas and vice versa, and so may shed light on possible epigenetic changes in the chromosome regions under study. Finally, cross-species BAC-FISH is a powerful instrument for the detection of pairing failure between introgressed homeologous regions thus explaining the absence of crossovers, a phenomenon known as linkage drag (e.g. Verlaan et al., 2011).

Acknowledgements

The authors would like to thank Bert Essenstam, Michel Hoogendoorn, Fien Meijer-Dekens, Dirk Jan Huigen, Jack Vossen and Jan de Boer for their help in arranging plant materials and potato BACs. We also acknowledge our colleagues of the Centre for Genetic Resources, the Netherlands (CGN), Dr. Gerard M. van der Weerden of the *Solanum* collec-

tion of the Botanical and Experimental Garden of Nijmegen University, and Dr. Dörthe Dräger, RijkZwaan B.V. de Lier, the Netherlands for providing various *Solanum* species and accessions.

References

Albrecht E, Chetelat RT. 2009. Comparative genetic linkage map of *Solanum* sect. *Juglandifolia*: evidence of chromosomal rearrangements and overall synteny with the tomatoes and related nightshades. Theoretical and Applied Genetics 118: 831–847.

Anderson L, Covey P, Larsen L, Bedinger P, Stack S. 2010. Structural differences in chromosomes distinguish species in the tomato clade. Cytogenetic and Genome Research 129: 24–34.

Bohs L, Olmstead RG. 1997. Phylogenetic relationships in *Solanum* (Solanaceae) based on ndhF sequences. Systematic Botany 22: 5–17.

Bonierbale MW, Plaisted RL, Tanksley SD. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120: 1095–1103.

Canady MA, Ji Y, Chetelat RT. 2006. Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174: 1775–1788.

Chang SB, Yang TJ, Datema E, Van Vugt J, Vosman B, Kuipers A, Meznikova M, Szinay D, Lankhorst RK, Jacobsen E et al. 2008. FISH mapping and molecular organization of the major repetitive sequences of tomato. Chromosome Research 16: 919–933.

Chetelat RT, Meglic V, Cisneros P. 2000. A genetic map of tomato based on BC1 *Lycopersicon esculentum* - *Solanum lycopersicoides* reveals overall synteny but suppressed recombination between these homeologous genomes. Genetics 154: 857–867.

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD. 2002.A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics 161: 1697–1711.

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457–1467.

Gottschalk W. 1954. Die Chromosomenstruktur der Solanaceen unter Berücksichtigung phylogenetischer Fragestellungen. Chromosoma 6: 539–626.

Grube RC, Radwanski ER, Jahn M. 2000. Comparative genetics of disease resistance within the Solanaceae. Genetics 155: 873–887.

Huang YL, Huang CC, Tang CY, Lu CL. 2010. SoRT2: a tool for sorting genomes and reconstructing phylogenetic trees by reversals, generalized transpositions and translocations. Nucleic Acids Research 2010: 1–7.

Iovene M, Wielgus SM, Simon PW, Buell CR, Jiang J. 2008. Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato. Genetics 180: 1307–1317.

Jacobs MMJ, Smulders MJM, van den Berg RG, Vosman B. 2011. What's in a name: genetic structure in *Solanum* section *Petota* studied using population-genetic tools. BMC Evolutionary Biology 11: 42.

Jacobs MMJ, van den Berg RG, Vleeshouwers VGAA, Visser M, Mank R, Sengers M, Hoekstra R, Vosman B. 2008. AFLP analysis reveals a lack of phylogenetic structure within *Solanum* section *Petota*. BMC Evolutionary Biology 8: 145.

de Jong JH, Wolters AMA, Kok JM, Verhaar H, van Eden J. 1993. Chromosome pairing and potential for intergeneric recombination in some hypotetraploid somatic hybrids of *Lycopersicon esculentum* (+) *Solanum tuberosum*. Genome 36: 1032–1041.

van der Knaap E, Sanyal A, Jackson SA, Tanksley SD. 2004. High-resolution fine mapping and fluorescence *in situ* hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics Society of America 168: 2127–2140.

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK. 1999. Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. Genetics 152: 1183–1202.

Lou Q, Iovene M, Spooner DM, Buell CR, Jiang J. 2010. Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence in situ hybridization mapping. Chromosoma 119: 435–442.

Lysak M, Pecinka A, Schubert I. 2003. Recent progress in chromosome painting of *Arabidopsis* and related species. Chromosome Research 11: 195–204.

Lysak MA, Cheung K, Kitschke M, Bures P. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. Plant Physiology 145: 402–410.

Lysak MA, Fransz PF, Ali HB, Schubert I. 2001. Chromosome painting in *Arabidopsis thaliana*. Plant Journal 28: 689–697.

Lysak MA, Lexer C. 2006. Towards the era of comparative evolutionary genomics in Brassicaceae. Plant Systematics and Evolution 259: 175–198.

Müller S, Hollatz M, Wienberg J. 2003. Chromosomal phylogeny and evolution of gibbons (Hylobatidae). Human Genetics 113: 493–501.

Peralta IE, Spooner DM. 2001. Granule-bond starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* (Mill.) Wettst. Subsection *Lycopersicon*. American Journal of Botany 88: 1888–1902.

Peralta IE, Spooner DM, Knapp S. 2008. The taxonomy of tomatoes: a revision of wild tomatoes (*Solanum* section *Lycopersicon*) and their outgroup relatives in section *Juglandifolia* and *Lycopersicoides*. Systematic Botany Monographs 84: 1–186.

Pertuzé RA, Ji Y, Chetelat RT. 2002. Comparative linkage map of the *Solanum lycopersicoides* and *S. sitiens* genomes and their differentiation from tomato. Genome 45: 1003–1012.

Peterson DG, Pearson WR, Stack SM. 1998. Characterization of the tomato (*Lycopersicon esculentum*) genome using *in vitro* and *in situ* DNA reassociation. Genome 41: 346–356.

Rens W, Fu B, O'Brien PCM, Ferguson-Smith M. 2006. Cross-species chromosome painting. Nature Protocols 1: 783–790.

Rodriguez F, Wu F, Ane C, Tanksley S, Spooner DM. 2009. Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? BMC Evolutionary Biology 9: 191.

Schubert I, Fransz P, Fuchs J, De Jong J. 2001. Chromosome painting in plants. Methods in Cell Science 23: 57–69.

Schwarzacher T, Heslop-Harrison P. 2000. Practical *in situ* hybridization. Oxford, UK: BIOS Scientific Publishers Ltd. ISBN 1-85996-138-x, 1-203.

Seah S, Yaghoobi J, Rossi M, Gleason CA, Williamson VM. 2004. The nematode-resistance gene, *Mi*-1, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. Theoretical and Applied Genetics 108: 1635–1642.

Spooner DM, Anderson GJ, Jansen RK. 1993. Chloroplast evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). American Journal of Botany 80: 676–688.

Spooner DM, Peralta IE, Knapp S. 2005. Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. Section *Lycopersicon* (mill.) wettst.]. Taxon 54: 43–61.

Spooner DM, Raul Castillo T. 1997. Reexamination of series relationships of South American wild potatoes (Solanaceae: *Solanum* section *Petota*): evidence from chloroplast DNA restriction site variation. American Journal of Botany 84: 671–685.

Stamova BS, Chetelat RT. 2000. Inheritance and genetic mapping of cucumber mosaic virus resistance introgressed from *Lycopersicon chilense* into tomato. Theoretical and Applied Genetics 101: 527–537.

Swofford DL. 1999. PAUP 4.0: phylogenetic analysis using parsimony (and other methods). Sunderland, MA, USA: Sinauer Associates, Inc.

Szinay D, Chang SB, Khrustaleva L, Peters S, Schijlen E, Bai Y, Stiekema WJ, Van Ham RCHJ, de Jong H, Klein Lankhorst RM. 2008. High-resolution chromosome mapping of BACs using multicolour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. Plant Journal 56: 627–637.

Tang X, Szinay D, Lang C, Ramanna MS, Van Der Vossen EAG, Datema E, Lankhorst RK, De Boer J, Peters SA, Bachem C et al. 2008. Cross-species bacterial artificial chromosome-fluorescence *in situ* hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180: 1319–1328.

Tanksley SD, Ganal MW, Prince JP, Devicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB et al. 1992. High-density molecular linkage maps of tomato and potato genomes. Genetics 132: 1141–1160.

Verlaan MG, Szinay D, Hutton SF, de Jong H, Kormelink R, Visser RGF, Scott JW, Bai Y. 2011. Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene *Ty*-1. Plant Journal 68: 1093–4403.

van der Voort JR, Van Zandvoort P, Van Eck H, Folkertsma R, Hutten R, Draaistra J, Gommers F, Jacobsen E, Helder J, Bakker J. 1997. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Molecular and General Genetics 255: 438–447.

Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD. 2008. Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. Genetics, 180: 391–408.

Weese TL, Bohs L. 2007. A three-gene phylogeny of the genus *Solanum* (Solanaceae). Systematic Botany 32: 445–463.

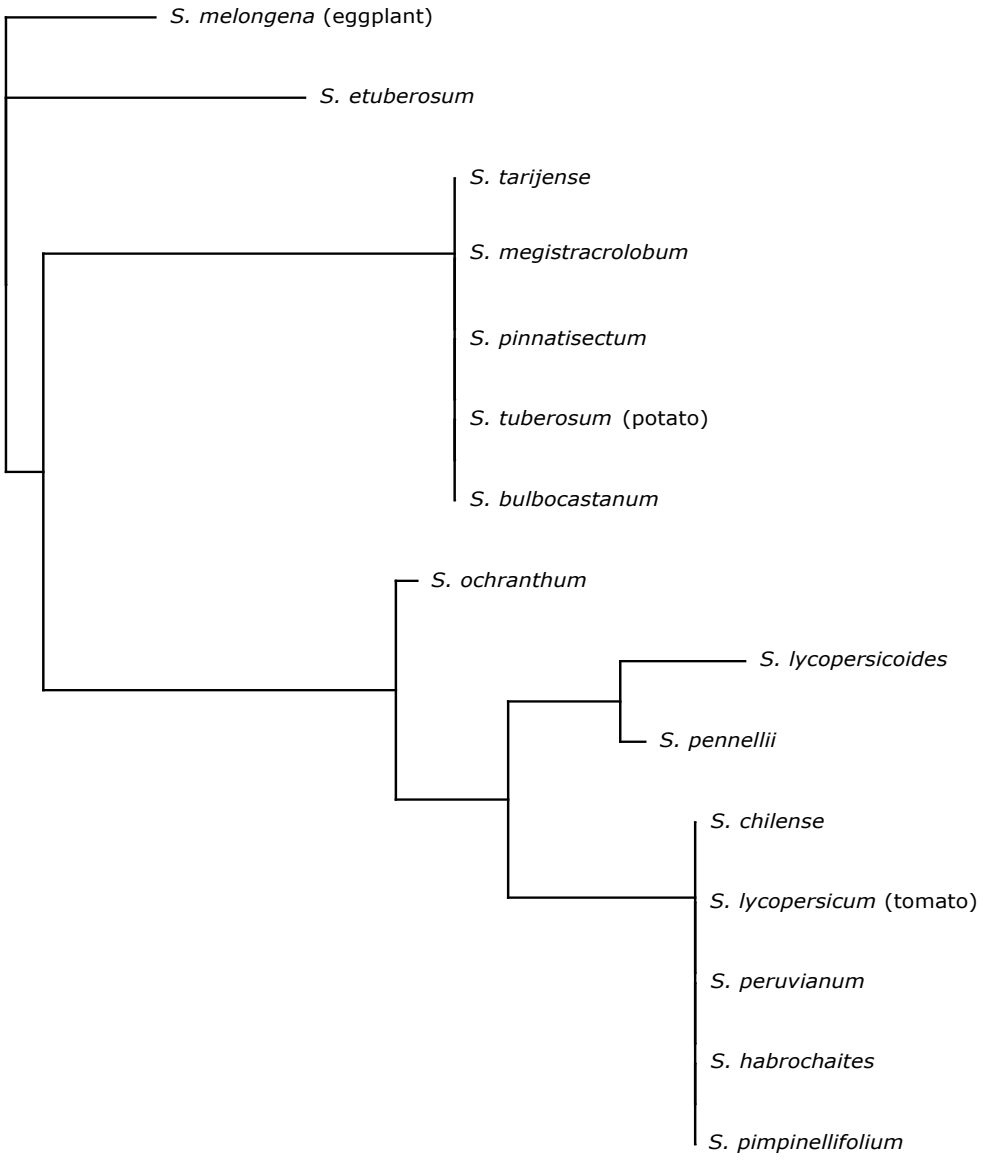
Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD. 2009b. A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. Theoretical and Applied Genetics 118: 1279–1293.

Wu F, Eannetta NT, Xu Y, Tanksley SD. 2009a. A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theoretical and Applied Genetics 118: 927–935.

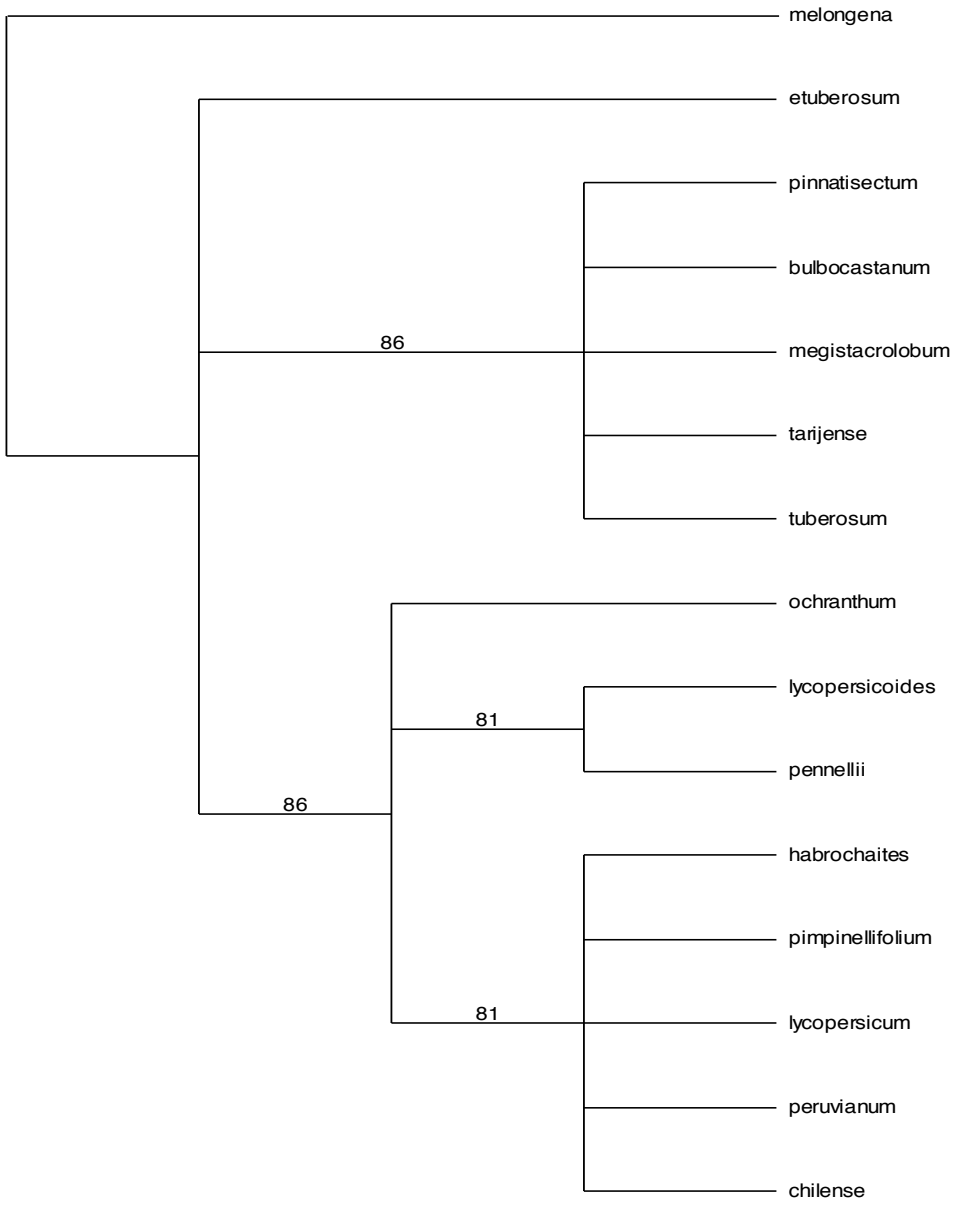
Wu F, Tanksley SD. 2010. Chromosomal evolution in the plant family Solanaceae. BMC Genomics 182: 1471–2164.

Zuriaga Z, Blanca J, Nuez F. 2009. Classification and phylogenetic relationships in *Solanum* section *Lycopersicon* based on AFLP and two nuclear gene sequences. Genetic Resources and Crop Evolution 56: 663–678.

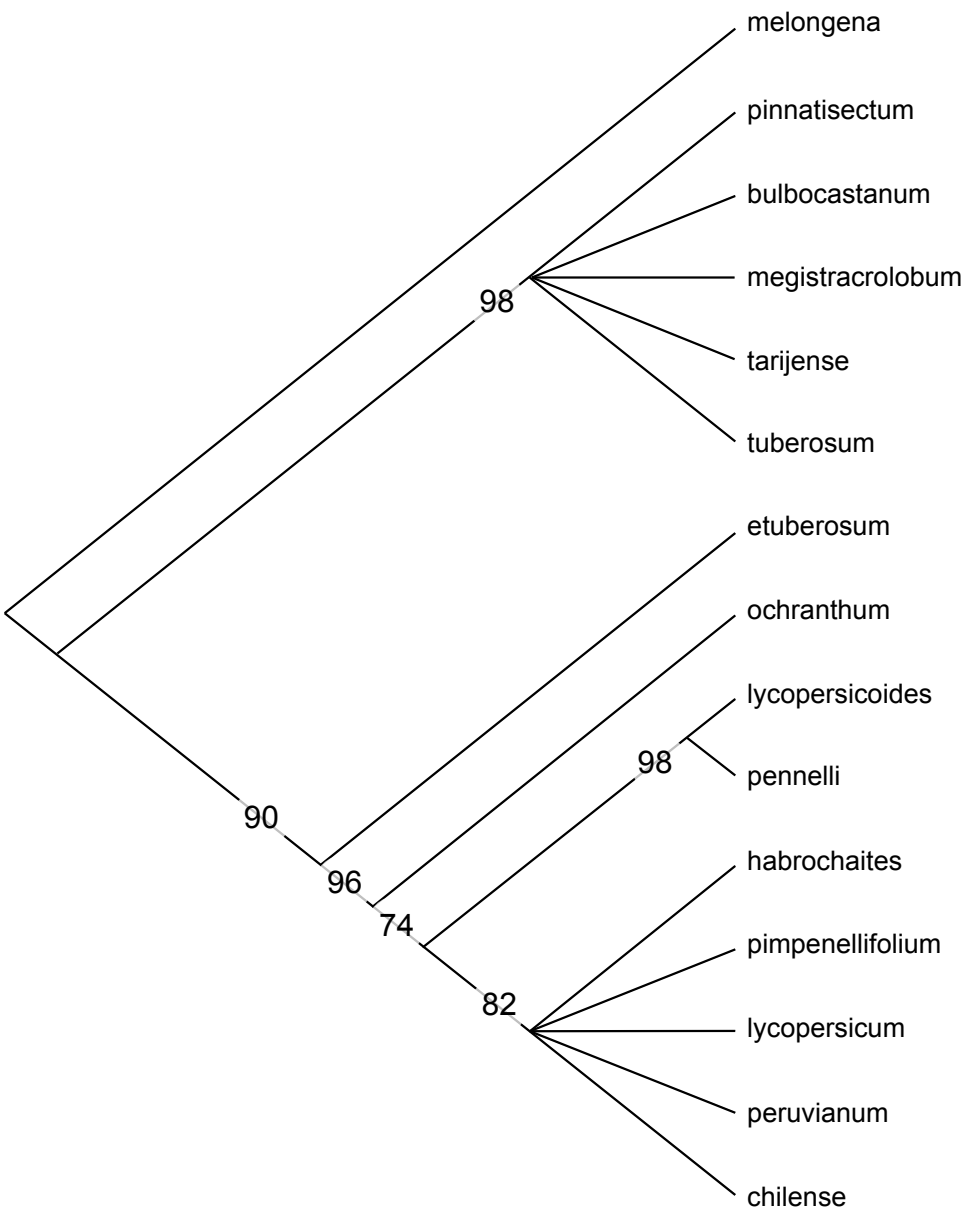
Supporting Information



Supplemenary figure S1: Phylogenetic tree made using SoRT2 (Huang *et al.,* 2010) based on pairwise genome rearrangement distances. The tree was rooted using *S. melongena*. See supporting Table S1 for the corresponding data matrix.



Supplemenary f igure S2: Maximum Parsimony (MP) tree where BAC - orders were coded as (unordered) character states for each chromosome . An MP analysis was executed with PAUP 4.0 (Swofford, 1999) . For the corresponding Nexus file, see Supporting Table S2.



Supplementary figure S3: Maximum Parsimony (MP) analysis based on a method by Müller *et al.* (2003) in which a binary data matrix was generated using the absence or presence of adjacent chromosome segments (*i.e.* BACs). The analysis was performed in PAUP 4.0 (Swofford, 1999) and Jackknife analysis was performed with 10,000 replicates (values given on the tree). For the corresponding Nexus file see Supporting Table S3; for details on characters and coding see Supporting Table S4.

This is the sort file for chromosomes 6S, 7S ,9S, 10L and 11S.
Data of chromosome 5 and 12 are missing due to translocations
For clarity: the 1 2 3 4 5 $ represent the five BACs on chromosome 6S
blue was left from analysis on chromosomes 7,9 and red from 11
The content of this file can be pasted to: http://bioalgorithm.life.nctu.edu.tw/SORT2/

> S. melongena.
1 2 3 4 5 $ 6 7 8 9 $ 10 11 12 13 $ 14 15 16 17 18 $ 19 20 21 22 $
> S. tuberosum.
1 2 3 4 5 $ 7 6 8 9 $ 12 11 10 13 $ 18 17 16 15 14 $ 19 20 21 22 $
> S. bulbocastanum.
1 2 3 4 5 $ 7 6 8 9 $ 12 11 10 13 $ 18 17 16 15 14 $ 19 20 21 22 $
> S. pinnatisectum.
1 2 3 4 5 $ 7 6 8 9 $ 12 11 10 13 $ 18 17 16 15 14 $ 19 20 21 22 $
> S. megistacrolobum.
1 2 3 4 5 $ 7 6 8 9 $ 12 11 10 13 $ 18 17 16 15 14 $ 19 20 21 22 $
> S. tarijense.
1 2 3 4 5 $ 7 6 8 9 $ 12 11 10 13 $ 18 17 16 15 14 $ 19 20 21 22 $
> S. etuberosum.
1 2 3 4 5 $ 9 8 6 7 $ 10 12 11 13 $ 18 14 15 16 17 $ 19 20 21 22 $
> S. ochranthrum.
1 2 3 4 5 $ 7 6 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $
> S. lycopersicoides.
4 5 1 2 3 $ 6 7 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $
> S. pennellii.
4 5 3 2 1 $ 6 7 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $
> S. chilense.
5 4 3 2 1 $ 7 6 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $
> S. lycper.
5 4 3 2 1 $ 7 6 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $
> S. peruvianum.
5 4 3 2 1 $ 7 6 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $
> S. habrochaites.
5 4 3 2 1 $ 7 6 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $
> S. pimpinellifolium.
5 4 3 2 1 $ 7 6 8 9 $ 10 11 12 13 $ 18 14 15 16 17 $ 22 21 20 19 $

Table S1: Data file used to generate the SoRT2 tree (after Huang *et al.*, 2010). For the corresponding tree see Supporting Fig. 1S.

```
#NEXUS

begin data;
	dimensions ntax=15 nchar=7;
	format missing=? datatype=standard symbols="A B C D";

matrix

melongena
C	A	A	A	A	A	D
etuberosum
A	A	B	B	B	A	A
pinnatisectum
A	A	C	C	C	A	A
bulbocastanum
A	A	C	C	C	A	A
megistacrolobum
A	A	C	C	C	A	A
tarijense
A	A	C	C	C	A	A
tuberosum
A	A	C	C	C	A	A
ochranthum
B	A	C	D	B	B	B
lycopersicoides
B	B	A	D	B	B	B
pennellii
B	C	A	D	B	B	B
habrochaites
B	D	C	D	B	B	B
pimpinellifolium
B	D	C	D	B	B	B
lycopersicum
B	D	C	D	B	B	B
peruvianum
B	D	C	D	B	B	B
chilense
B	D	C	D	B	B	C;
end;
```

Table S2: Nexus file used to generate the MP tree given in Supporting Fig. S2.

```
#NEXUS


begin data;
dimensions ntax=15 nchar=49;
format missing=? datatype=standard;
matrix
melongena
? ? ? ? 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 1 1 0 0 ? ? ? ? ? ? ? ? ?
etuberosum
1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 0 1 0 1 1 1 1 0 0 1 1 1 1 0 0 0 0
pinnatisectum
1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 0 0 1 1 1 1 0 0 0 0
bulbocastanum
1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 0 0 1 1 1 1 0 0 0 0
megistracrolobum
1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 0 0 1 1 1 1 0 0 0 0
tarijense
1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 0 0 1 1 1 1 0 0 0 0
tuberosum
1 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1 0 0 1 1 1 1 0 0 0 0
ochranthum
0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 1 0 1 1 0 0
lycopersicoides
0 0 1 1 0 0 0 0 0 1 0 1 1 1 1 1 0 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 1 0 1 1 0 0
pennellii
0 0 1 1 0 0 0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 1 0 1 1 0 0
habrochaites
0 0 1 1 0 1 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 1 0 1 1 0 0
pimpinellifolium
0 0 1 1 0 1 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 1 0 1 1 0 0
lycopersicum
0 0 1 1 0 1 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 1 0 1 1 0 0
peruvianum
0 0 1 1 0 1 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 1 0 1 1 0 0
chilense
0 0 1 1 0 1 0 1 1 0 0 0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 1 0 0 0 0 1 1 1 1;
end;
```

Table S3: Nexus file used to generate the MP tree given in Supporting Fig. S3. Coding was executed by identifying all adjacent chromosome segments. See Supporting Table S4 for details on characters and coding.

Chapter 5:

On writing a first review paper…

[…]

'Nee, voor mij alleen is het al niet genoeg. Hu!' 'Ach, jij', zegt de prins, 'krent die je daar bent!' En bij die woorden gaat pats boem zijn vrouw tegen de vlakte! Vrouwlief heeft haar hele tronie bezeerd, ze ligt op de vloer en huilt. En de prins hulde zich in een wijde mantel en keerde terug naar zijn toren, daar had hij kooien staan. U moet weten dat hij daar kippen fokte. De prins komt dus aan in de toren, en de kippen gaan vreselijk tekeer, ze willen eten. Een van de kippen is zelfs aan het hinniken geslagen. 'Hé jij daar', zegt de prins tegen haar, 'Chanteclair! Stil jij, anders krijg je op je bliksem!' De kip begrijpt niet wat hij zegt en hinnikt verder. Het slot van het liedje was dus dat er op de toren een kip zat te hinniken, dat de prins liep te vloeken als een ketter en dat zijn vrouw beneden op de vloer lag – kortom een waar Sodom.

Dat was dan het verhaal dat Andrej Andrejevitsj verzonnen had. Uit dit verhaal valt wel op te maken dat Andrej Andrejevitsj een enorm talent is. Andrej Andrejevitsj is een heel intelligent man. Heel intelligent en heel goed!

Daniil Charms

(In: 'Een nieuwe talentvolle schrijver', 1938;
Published in: Ik zat op het dak, Uitgeverij Atlas, Amsterdam/Antwerpen, 1999)

Chapter 5

Managing meiotic recombination in plant breeding

Authors:
T.G. (Erik) Wijnker and Hans de Jong
Laboratory of Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands

Trends in Plant Science, 13 (12), 640-646.

Abstract

Crossover recombination is a crucial process in plant breeding because it allows plant breeders to create novel allele combnations on chromosomes that can be used for breeding superior F1 hybrids. Gaining control over this process, in terms of increasing crossover incidence, altering crossover positions on chromosomes or silencing crossover formation, is essential for plant breeders to effectively engineer the allelic composition of chromosomes. We review the various means of crossover control that have been described or proposed. By doing so, we sketch a field of science that uses both knowledge from classic literature and the newest discoveries to manage the occurrence of crossovers for a variety of breeding purposes.

The plant breeders' desire to control crossovers

Plant breeding attempts to combine valuable traits from different parents in new elite varieties. These traits are encoded for by genes on chromosomes. The success of a breeding program depends on the ability of plant breeders to bring the desired alleles together in one hybrid, both by constructing desired combinations of alleles on chromosomes and by designing the right combination of chromosomes. Meiotic recombination has a pivotal role in successful plant breeding because the reshuffling of homologues and chromosome segments takes place only during meiosis. The maximum obtainable amount of meiotic recombination is determined by two factors: the number of chromosomes of a plant (random chromosome assortment) and the number and positions of crossovers on the pairs of homologous chromosomes (crossover recombination). Plant breeders have no direct control over the number of chromosomes, except perhaps by adding artificial chromosomes [1], and therefore look for means of imposing control over that other part of meiotic recombination: cross-over formation.

Glossary

Crossover recombination: meiotic recombination resulting from crossovers between chromatid segments in a chromosome pair.
Doubled haploids (DHs): diploid plants grown from (haploid) spores in which genome duplication resulted in diploidy. Chromosomes are identical. DHs are commonly used to directly fix the genotype of meiotic spores.
Homoeologous chromosomes: chromosomes from different species that show a higher degree of sequence divergence than homologous chromosomes do and display less or no pairing at meiotic prophase I. Differences might be of a higher order of magnitude, showing minor structural rearrangements such as inversions, translocations or differences in repetitive sequences. Sequence divergence can be so large that crossover formation during meiosis is impaired.
Homologous chromosomes: chromosomes that are sufficiently similar for regular meiotic pairing but show a limited (allelic) degree of sequence divergence.
Random chromosome assortment: meiotic recombination resulting from an independent assortment of chromosomes.
Univalent: a single chromosome that is not bound to another by a chiasma at anaphase I.

Box 1. Tetrad analysis in plants

The possibility for tetrad analysis in plants emerged with the discovery in *Arabidopsis thaliana* of the *quartet* mutant, in which the four meiotic spores remain together [56,57]. In combination with fluorescent markers expressed by a pollen-specific promotor, this mutant directly displays the consequences of crossover recombination in pollen grains [58].

Species producing pollen tetrads are widespread among plants [59] and, with the details of the *quartet* mutation known, such phenotypes are possibly inducible in crops [60] or can be identified directly by mutant screens. Insert lines with fluorescent markers are currently being used for *Arabidopsis*, but the construction of such lines for other species would require a considerable investment. However, these insert lines could potentially result in methods for evaluating the effects of treatments for altering recombination frequencies in crops, which would be of enormous benefit, especially since different crops can react differently to certain treatments (as discussed in the main text). A rather similar technique was developed for the direct observation of crossovers in seeds [61] using insertion lines of fluorescent proteins expressed by seed-specific promotors. Although this technique does not require a specific phenotype (*e.g. quartet*), its application in crops might be somewhat more limited because the production of seeds in crops is generally much lower than production of pollen.

In the past, the lack of practical tools for establishing crossover frequencies hampered systematic studies on crossover management in crops. Determining specific crossover frequencies was costly and laborious and was mostly confined to model species. Nowadays, modern methods for high-throughput genotyping and the development of dense sets of markers provide the tools for efficiently determining crossover frequencies and crossover positions [2]. Research will surely benefit as well from new technologies, such as tetrad analysis in a *quartet* background (Box 1).

Because they are no longer constrained by technical issues, the interest of plant breeders in control over crossover formation is larger than ever before. We explore the possibilities for controlling crossover formation and describe how several methods have considerable potential. We show how plant breeders can exert influence over crossover frequencies, crossover position and crossover allocation to homoeologous regions. It is even possible to suppress crossover formation completely and reduce the complexity of meiotic recombination to only the random assortment of whole chromosomes. This opens up opportunities to extract and fix whole chromosomes from heterozygous complements (*i.e.* F1 hybrids). In this paper we will focus on mechanistic aspects of crossover control and point out methods that can be utilized even without a precise understanding of the processes underlying crossover formation (Box 2).

Controlling crossover incidence

The number and distribution of crossovers during meiosis is tightly constrained. There is typically at least one crossover per chromosome pair to ensure proper segregation at metaphase I, known as 'crossover assurance' [3]. The total number and relative position of crossovers on each chromosome is limited to generally one, or perhaps two, per chro-

Box 2. Controlling crossovers: managing the unknown

Crossover formation is a complex process that is regulated at multiple levels, and factors governing crossover formation are still not well understood [62]. It partly depends on the homology search that follows double-strand break (DSB) formation in plants [18]. Mismatch repair proteins, which are involved in homology search, prohibit recombination between non-homologous segments [63,64]. The coordinated remodelling of chromatin affects pairing and recombination affinities between chromosomes [34,65], and the placement of crossovers in a pair of homologues is tightly regulated (as discussed in the main text). Such processes complicate plant breeders' efforts to engineer chromosome structure. Although high-throughput screening of large populations is sometimes an option for obtaining rare crossovers, recombinants will not always be found. More efficient recombination might be required when, for example, alleles in trans of closely linked loci need to be combined. This is especially difficult in regions where recombination is suppressed or absent. In yet other cases, recombination might be required between chromosomes that do not even pair in meiosis.

mosome arm by interference, a phenomenon distributing crossovers in a non-random, semi-uniform pattern [4]. On a smaller scale, crossovers preferentially occur in certain areas called 'recombination hotspots' [5,6], and the areas with almost no crossovers are called 'recombination cold spots'. In the following section we explore variability in crossover incidence and discuss how this can be influenced.

Internal factors, such as genetic background and morphological and developmental differences, can have a considerable impact on crossover incidence. Related genotypes can have significantly different crossover frequencies, and up to a 30% difference was reported among barley (*Hordeum vulgare*) cultivars [7]. Similarly, differences of 17% were found among *Arabidopsis* accessions [8]. More strikingly, recurrent selection for high and low recombining individuals starting from a single heterozygous plant in an F2 population of lima bean (*Phaseolus lunatus*) led to a threefold difference in recombination frequencies in the F6 generation [9].

There is ample evidence for differences between the sexes in crossover frequencies, both in plants and animals [10], and this phenomenon is likely to be caused by differing compaction states of chromatin during meiotic prophase in male and female meiosis [11,12]. In addition, crossovers typically locate more distally at male meiosis in *Arabidopsis* [11]. The physical position of a flower can influence crossover incidence: in *Arabidopsis*, anthers on secondary or tertiary branches have up to 16% more crossovers than those on primary branches [13]. Such effects are species specific: barley and rye (*Secale cereale*), for example, show no variation in crossover incidence in relation to flower position [14].

Some reports have shown how external factors, such as environmental influences and chemical treatments, can alter crossover frequencies. Random environmental variation was shown to result in a twofold difference in recombination frequency between two linked loci in lima bean [9]. In more controlled experiments in which only temperature was varied, recombination rates increased with higher temperature in *Arabidopsis* (up to 18%), as had previously been described for species like *Hordeum vulgare* and *Vicia*

faba, whereas high temperatures decreased recombination frequencies in *Allium ursinum* [13]. Crossover frequency in barley, as in rye, is less susceptible to environmental influence [14].

Recombination frequencies can be increased artificially by the use of various chemical agents or physical stress, such as temperature shock or UV exposure [15]. The feasibility of this approach was originally shown by a study in *Hordeum*, where actinomycin D, as well as diepoxubytane, was shown to lead to a threefold increase in recombination frequency between linked markers [16]. A survey study using various chemical agents also showed large (roughly two- to sevenfold) increases in recombination frequencies by the use of various chemical agents, a fourfold increase by heat shock and threefold increase by UV radiation in *Arabidopsis* [15]. These data, however, were based on a relatively small number of plants and might be limited to the specific (pericentromere) genomic regions that were assessed [17].

In recent years, many proteins involved in crossover formation have been identified [18,19]. The possible impact of genetic regulation on crossover formation is illustrated by the uncharacterized X-ray sensitive4 (xrs4) mutant of *Arabidopsis*, in which recombination frequency in certain regions increased over twofold [20]. Such mutants fuelled the idea that either overexpression or silencing of such proteins could modulate recombination frequencies [21]. However, only a few studies on this topic were published, and the extent of their analysis was limited. In tomato (*Solanum lycopersicum*), overexpression of *MutL homolog1* (*MLH*1, which encodes a mismatch repair protein) led to a 10% increase in chiasma frequency [22], whereas a genetic interval in *Arabidopsis* showed a twofold increase of recombination frequency upon overexpression of *RADIATIONSENSITIVE*51 (*RAD*51, a gene involved in DNA repair) [23].

Crossovers follow changes in chromosome structure

The location of crossovers along the chromosome field (*i.e.* proximal versus distal events) is tightly regulated. Whereas in a species like Welsh onion (*Allium fistulosum*) crossovers localize proximally, those in a close relative, *Allium cepa*, localize distally [24], and such distal localization is also found in species such as barley (*Hordeum vulgare*), maize (*Zea mays*) and wheat (*Triticum aestivum*) [25]. The occurrence of crossovers is in part determined by higher order chromosome structure, and they are less frequent in pericentromere areas [26] (Figure 1a). The presence of tandem repeats in distal heterochromatin blocks in *A. fistulosum* was suggested to move chiasmata away from the ends (L. Khrustaleva, personal communication) (Figure 1b). This strong regulation of crossover placement poses strong constraints on the extent to which the allelic content of a chromosome can be changed by crossover recombination. Because some regions are not subjected to crossover recombination, loci remain tightly or completely linked, which limits breeding potential. In the following section we explore possibilities for altering the positioning of crossover events along the chromosome axis.

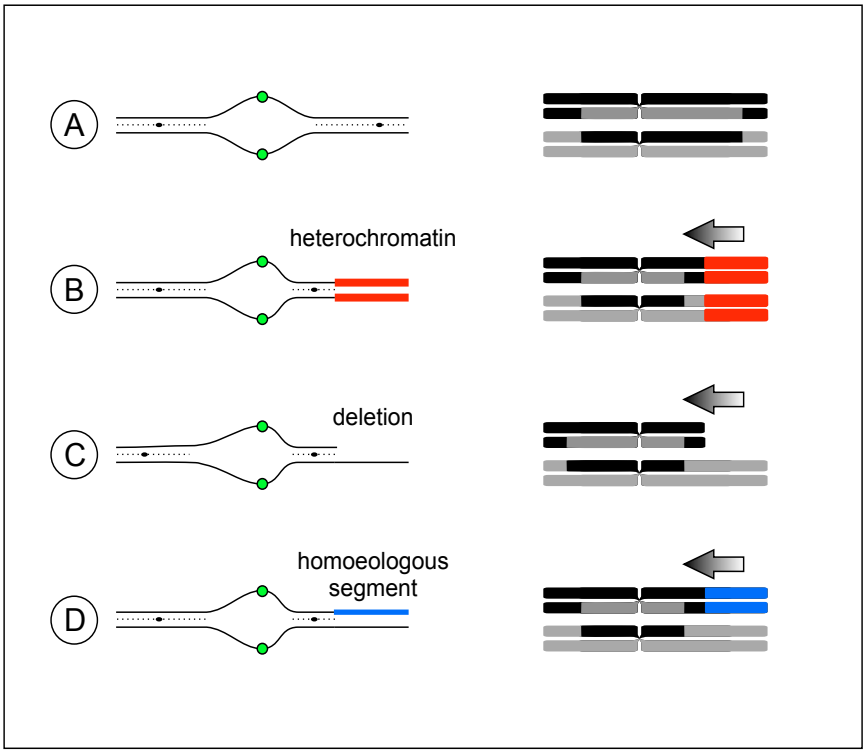


Figure 1. Examples of how structural changes can lead to crossover shifts. Drawings on the left represent a chromosome pair at mid-prophase of meiosis I. Solid lines represent chromosome axes; dotted lines represent the proteinaceous structure keeping paired homologues together during meiotic prophase (the synaptonemal complex); and small black spheres mark the crossover sites that will later form the chiasmata. Green spheres represent the centromere regions. The drawings on the right represent the corresponding recombinant chromosomes at anaphase I. (a) Chromosome pairing and crossover recombination of a normal chromosome pair. The region around the centromere is the pericentromere, which is known to synapse later and is poor in crossover events. In these examples we assume that there are two crossovers per chromosome pair and that crossovers occur between only two chromatids (in actuality the number of crossovers and the number of recombining chromatids can vary). (b) Pairing and recombination in a chromosome pair containing a recently formed large distal heterochromatic block (represented in red) in one of the chromosome arms. (c) Pairing and recombination in a pair in which one partner has the distal end of one arm deleted. (d) Pairing and recombination for a chromosome pair in which one partner has a homoeologous chromosome segment (represented in blue).

It is known that even a short terminal deletion in one of the pairing chromosomes can severely reduce crossover formation in the affected arm [27]. This is likely to be due to a disturbed pairing initiation that generally (but not exclusively) starts at distal chromosome ends [28,29]. In one experiment, radiation induced the deletion of the terminal end of the short arm of chromosome VI in *Petunia hybrida* [30]. Crossover formation in the truncated arm was severely reduced but, interestingly, there was an up to sevenfold increase of crossovers in intervals on the long arm of that same chromosome (Figure 1c).

To obtain such deletions, one can use pollen irradiation and then select for the loss of dominant distal markers.

Other types of structural variants that change crossover positions are translocations. Effects of translocations, which also comprise the skewed transmission of alleles due to chromosome deficiencies in gametes, can be well observed in chiasma configurations during meiotic prophase. When a chromosome carries a distal translocation, the translocation introduces strong heterology at the chromosome end in the chromosome pair. This results in a shift of crossovers to interstitial chromosome segments (the area between the centromere and the translocation site) [31]. Although such translocations might be undesired in breeding programs, they illustrate the mechanism by which changes in chromosome structure reallocate crossovers to the homologous sites between the pairing partners.

A related case of chromosome structure alterations has been described for tomato lines carrying introgressed homoeologous segments of related wild species. It was shown that the presence of such segments in otherwise homologous chromosome pairs affected crossover localization: crossover frequencies increased strongly in adjacent homologous sequences [32] (Figure 1d). Comparable observations were made in *Lolium-Festuca* hybrids [27,28,33,34] using allotriploid offspring comprising two homologues of one species in addition to one homologue of the other. Crossovers in the *Lolium* and *Festuca* parents preferentially form in the distal chromosome regions. In the allotriploids, homologous chromosomes behaved similarly, preferentially forming distal crossovers. However, when their homoeologous partner joined in trivalents, it formed crossovers mostly in proximal regions [33,35] because homoeology disturbs crossover formation at the distal chromosome ends. In wheat and *Triticeae* species, patterns of homoeologous recombination were shown to vary between different species and homoeologous recombination can, like the *Lolium-Festuca* hybrids, be highly localized (reviewed in [36]).

Approaches for altering crossover localization using transgenic approaches are very scarce. Nicolas *et al.* [36] designed a method for recruitment of SPORULATION-DEFICIENT11 (SPO11) a key protein for crossover initiation, to selected sequences of DNA by designing an artificial fusion protein comprised of SPO11 and a DNA-binding domain. The DNA-binding domain recruits the fusion protein to binding sites on the DNA and induces crossover formation at these sites. This technique has been shown to work well in yeast [37], although the induction of double-strand breaks occurs mostly in binding sites in open chromatin regions and not, or less, in natural cold spots, such as centromere areas [38]. This method, which has been proposed for use in plants, would provide a powerful tool for induction of site-specific crossovers. SPO11 could be fused to a variety of different DNA binding domains, thus resulting in a suite of fusion proteins that could, in theory, recruit SPO11 to various sites on the DNA.

Crossovers in homoeologous regions

The mechanisms that control crossovers between homologues can be frustrating to plant breeders in their attempts to integrate valuable traits through introgressive hybridization. Examples of such traits are genes for resistance or drought tolerance that might be found in related species. Typically, a cross is made between a recipient crop and a related taxon carrying a trait of interest. This is followed by recurrent backcrosses to the crop in which the introgressed homoeologous region is narrowed down, keeping the gene of interest and removing 'wild' undesired genes (linkage drag). Because crossovers are generally suppressed or absent between the introgressed segment and its homoeologous counterpart, it is imperative to find the mechanisms and genes that control the pairing between homoeologous segments.

To induce crossovers in introgressed segments, one can reallocate crossovers to the homoeologous region by making the other regions in the chromosome pair even more homoeologous [32]. If, for example, crossovers are to be induced in one chromosome arm carrying an introgressed segment, one could provide a pairing partner that carries an introgressed segment of a more distantly related taxon on the opposite chromosome arm. Crossovers then reallocate to the least homoeologous sites. One can predict that any rearrangement could be used to direct crossovers to homoeologous regions of interest. In the same study, it was shown that in tomato hybrids, larger homoeologous segments have a higher incidence of crossovers than shorter segments. This led to the suggestion that in introgression breeding, plant breeders should initially search for those plants with the largest alien insert and then select for single crossovers close to the locus of interest. Crossing two recombinant lines with crossovers on different sides of the locus would then reduce linkage drag to a minimum [39].

Different genes have been identified that influence homoeologous recombination. The best known is *Pairing homoeologous*1 (*Ph*1) [40,41], which inhibits homoeologous pairing between wheat chromosomes. In the absence of *Ph1*, pairing and recombination between homoeologous chromosomes is frequent, which greatly facilitates introgressive hybridization [42]. However, the constitutive deletion of *Ph1* can over time lead to rearranged chromosomes in the genome, which can later interfere with further breeding efforts. The use of *Ph1* in plant breeding would greatly benefit from means of temporarily switching off the locus [43]. A gene, Pairing regulator in *Brassica napus* (*PrBn*), with a comparable function was also identified in *Brassica* [44], but it has not been characterized at the molecular level. It was further hypothesized that during meiosis, the knockdown of genes, such as *MutS homolog*2 (*MSH*2) or *MSH*3, that encode mismatch repair proteins might promote homoeologous recombination [45,46]. Knockdown of these genes could be achieved, for example, by RNA interference (RNAi) or dominant negative suppression, in which a truncated gene disrupts the functioning of protein complexes [47].

Preserving elite genotypes

Most of the cultivars produced by breeders are heterozygous F1 hybrids, which are bred for their unique combination of alleles and outperform their parents by hybrid vigour. Controlled creation of elite heterozygosity is achieved by simply crossing two carefully selected homozygous parents. Doing the reverse (creating homozygous parents for any heterozygous F1) is, on the contrary, an almost unfeasible task. The allele combinations that give the F1 its unique character are broken apart by recombination when the F1 is used in crosses (Figure 2). The answer to preserving these combinations during meiosis lies in technologies and strategies that reduce the complexity of meiotic recombination; one such strategy is reverse breeding (schematically summarized in Figure 2) [48].

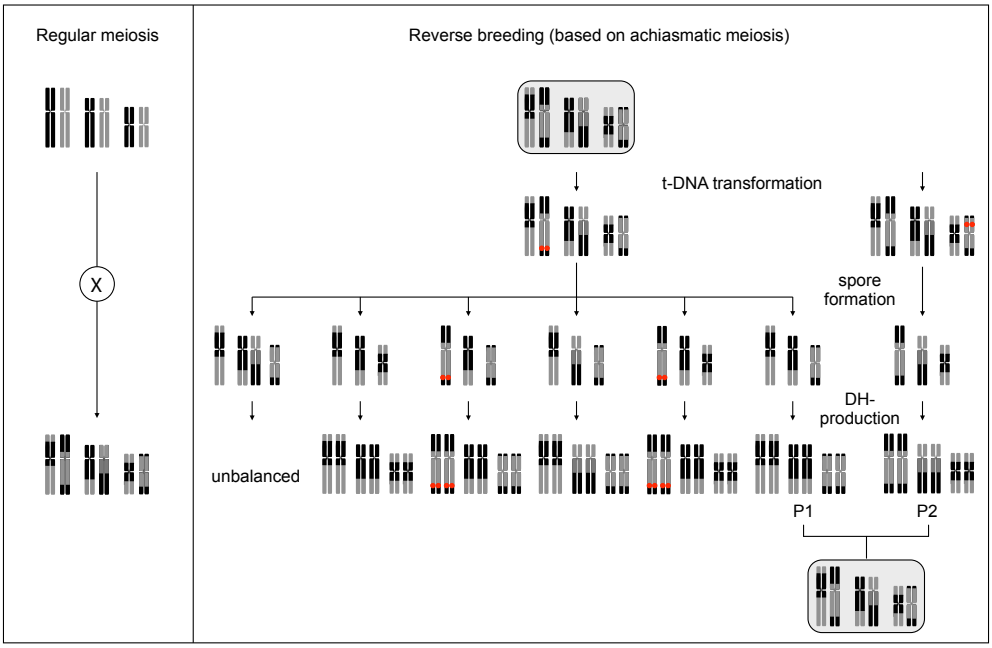


Figure 2. Schematic representation of (a) regular meiosis (the selfing of a heterozygote) and (b) the reverse breeding technique. We consider a fictive F1 with three chromosome pairs (2n = 2x = 6). During regular meiosis, chromosomes and chromosome segments recombine, giving rise to an infinite number of genotypically different offspring (a). In an alternative approach (b), crossover recombination is prevented by transgenic suppression (RNAi, etc.) of one of the genes essential for crossing over. The red dots represent the transgene. Achiasmatic meiosis gives rise to spores. Note that spores carry non-recombinant chromosomes. Most spores are unbalanced (one possibility drawn), but some spores are balanced (several possibilities drawn). Doubled haploids (DHs) are produced from balanced spores, giving rise to homozygous diploids. Among the DHs produced, reciprocal genotypes (P1, P2) can be recruited that, upon crossing, exactly reconstitute the original F1. These are the homozygous parental lines for the F1 hybrid. Note that P2 is derived from a second transformant carrying the transgene on a different chromosome.

Reverse breeding is based on suppression of crossover formation by RNAi or comparable gene silencing techniques. Studies have shown that RNAi silencing of essential early meiotic genes, such as *DISRUPTED MEIOTIC CDNA*1 (*DMC*1), can lead to (almost) complete suppression of crossover formation [49,50]. Consequently, homologues are not joined by chiasmata (the physical manifestation of crossing over) during meiotic prophase I and remain as univalents at anaphase I. These univalents (non-recombinant parental chromosomes) then segregate randomly to daughter cells during the first meiotic division [51]. Most resulting spores will be unbalanced, containing either none, one or two copies of a given chromosome. However, balanced spores, containing one copy of each chromosome, will be formed at a probability of (1/2)x, where x equals the basic chromosome number. Consequently, the chance of obtaining balanced spores decreases exponentially with the chromosome number and seems feasible for species in which the chromosome number equals 12 or less [48].

In reverse breeding, any given elite heterozygote is transformed using an RNAi construct targeting a gene that encodes a protein that mediates the formation of crossovers. The resulting plant is expected to produce low numbers of viable balanced haploid spores that are then regenerated into doubled haploid, perfectly homozygous plants. Other spores with an unbalanced chromosome number will, if they are still viable, produce aneuploid individuals with poor regeneration rate and vigour. Among the doubled haploids, parents with complementary genotypes can be recruited that, upon crossing, will reconstitute the exact genotype of the elite hybrid again (Figure 2).

In *Arabidopsis*, various mutants that lack crossovers (almost) exclusively produce univalents [18,19], although their chromosome behaviour during meiotic prophase can be different. It was recently shown that in univalent-producing mutants (*desynaptic*1 [*dsy*1], *meiotic prophase amonipeptidase*1 [*mpa*1]) in which chromsomes pair normally during prophase, univalents segregate preferentially to opposite poles during metaphase I [52]. This suggests that pairing, even without chiasma formation, to some extent orients homologues to opposite poles. Targeting such genes for reverse breeding might be fruitful because the chance of recovering balanced gametes increases substantially. As such, genes such as *PARTING DANCERS* (*PTD*), for which the mutant shows complete pairing and few residual crossovers next to high levels of univalents [53], might also be of interest to reverse breeding. The benefit of its regular segregation might well outweigh the downside of few remaining crossovers.

Reverse breeding provides plant breeders with new possibilities for further breeding. When one transforms a hybrid for which the parents are known, one can directly select chromosome substitution lines from among the produced doubled haploids. These chromosome substitutions have various potential applications, as for example in the generation of near isogenic lines by recurrent backcrosses. Such lines are extremely valuable for mapping quantitative trait loci (QTL) and for advanced forms of markerassisted breeding [54,55].

Conclusions

The improvement of crop species relies on the possibility to select and carefully produce new allele combinations. Over the last decade, plant breeding practice has been revolutionized by the advent of high-density marker collections that enable high-throughput screening in breeding selection schemes. The ease of genotyping shifted the focus of breeding to marker-assisted breeding, which greatly increased the predictability of breeding efforts, in which crossovers are and will remain crucial.

In spite of the plethora of genes known to be involved in crossover control, few studies have been published on the practical applications of suchgenes. This is incontrast to the various patents for crossover control that have been filed, indicating that methods for crossover control have the attention of many researchers and that the economic value of such methods is acknowledged. As we see it, research is progressing along several lines. On the one hand, we expect a revival of classic meiotic research: variability within crops, within the plants or induced by internal and external factors might be evaluated using high-throughput marker technology. On the other hand, we foresee that an increasing knowledge on the molecular control of meiosis might create new applications for plant breeding.

Acknowledgements

We express our gratitude to G.P. Copenhaver, P.J. van Dijk, R.H.G. Dirks, C.M.P. van Dun, E. Jacobsen, J.B. Keurentjes, M. Koornneef, C.L.C. Lelivelt and J. Sybenga for valuable comments on the manuscript. We likewise highly appreciate the useful suggestions made by our anonymous referees.

References

1 Carlson, S.R. et al. (2007) Meiotic transmission of an in vitro-assembled autonomous maize minichromosome. PLoS Genet. 3, e179
2 Gupta, P.K. et al. (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101, 5–18
3 Shinohara, M. et al. (2008) Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. Nat. Genet. 40, 299–309
4 Sturtevant, A.H. (1915) The behavior of the chromosomes as studied through linkage. Z. Indukt. Abstamm. Vererbungsl. 13, 234–287
5 Drouaud, J. et al. (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination 'hot spots'. Genome Res. 16, 106–114
6 Mezard, C. (2006) Meiotic recombination hotspots in plants. Biochem. Soc. Trans. 34, 531–534
7 Sall, T. et al. (1990) Genetic control of recombination in barley: variation in recombination frequency measured with inversion heterozygotes. Hereditas 112, 157–170
8 Sanchez-Moran, E. et al. (2002) Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. Genetics 162, 1415–1422

9 Allard, R.W. (1963) Evidence for genetic restriction of recombination in the lima bean. Genetics 48, 1389–1395
10 Lenormand, T. and Dutheil, J. (2005) Recombination difference between sexes: a role for haploid selection. PLoS Biol. 3, e63
11 Drouaud, J. et al. (2007) Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. PLoS Genet. 3, e106
12 Petkov, P.M. et al. (2007) Crossover interference underlies sex differences in recombination rates. Trends Genet. 23, 539–542
13 Francis, K.E. et al. (2007) From the cover: pollen tetrad-based visual assay for meiotic recombination in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. 104, 3913–3918
14 Sall, T. (1990) Genetic control of recombination in barley: variation in linkage between marker genes. Hereditas 112, 171–178
15 Preuss, D. and Copenhaver, G.P. University of Chicago. Chemical and physical treatment that stimulate recombination, WO/2000/054574
16 Sinha, R.P. and Helgason, S.B. (1969) The action of actinomycin D and diepoxybutane on recombination of two closely linked loci in *Hordeum*. Can. J. Genet. Cytol. 11, 745–751
17 Copenhaver, G.P. et al. (1999) Genetic definition and sequence analysis of *Arabidopsis centromeres*. Science 286, 2468–2474
18 Ma, H. (2006) A molecular portrait of *Arabidopsis* meiosis. In The *Arabidopsis* Book (Somerville, C.R. and Meyerowitz, E.M., eds), pp. 1–39, American Society of Plant Biologists
19 Mercier, R. and Grelon, M. (2008) Meiosis in plants: ten years of gene discovery. Cytogenet. Genome Res. 120, 281–290
20 Masson, J.E. and Paszkowski, J. (1997) *Arabidopsis thaliana* mutants altered in homologous recombination. Proc. Natl. Acad. Sci. U. S. A. 94, 11731–11735
21 Rozwadowski, K.L. and Lydiate, D.J. Her Majesty the Queen in Right of Canada as represented by the Minister of Agriculture and Agrifood Canada. Modulation of meiotic recombination, WO/2002/022811
22 Wittich, P.E. et al. Keygene N.V. Homologous recombination in plants, WO/2007/030014
23 Betzner, A., Stefan, et al. RhoBio. Composition and method for increased meiotic recombination in plants, WO/2002/008432
24 Albini, S.M. and Jones, G.H. (1987) Synaptonemal complex spreading in *Allium cepa* and *A. fistulosum*. Chromosoma 95, 324–338
25 Mezard, C. et al. (2007) The road to crossovers: plants have their say. Trends Genet. 23, 91–99
26 Anderson, L.K. and Stack, S.M. (2005) Recombination nodules in plants. Cytogenet. Genome Res. 109, 198–204
27 Jones, E. et al. (2002) The effect of a deficiency and a deletion on recombination in chromosome 1BL in wheat. Theor. Appl. Genet. 104, 1204–1208
28 Qi, L.L. et al. (2002) A strategy for enhancing recombination in proximal regions of chromosomes. Chromosome Res. 10, 645–654
29 Corredor, E. et al. (2007) Terminal regions of wheat chromosomes select their pairing partners in meiosis. Genetics 177, 699–706
30 Gerats, A.G.M. et al. (1984) Recombination behaviour and gene transfer in *Petunia hybrida* after pollen irradiation. Mol. Gen. Genet. 198, 57–61
31 Sybenga, J. (ed.) (1975) Meiotic Configurations, Springer-Verlag

32 Canady, M.A. et al. (2006) Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174, 1775–1788

33 Kosmala, A. et al. (2006) Chromosome pairing in triploid intergeneric hybrids of *Festuca pratensis* with *Lolium multiflorum*, revealed by GISH. J. Appl. Genet. 47, 215–220

34 Colas, I. et al. (2008) Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. Proc. Natl. Acad. Sci. U. S. A. 105, 6075–6080

35 Zwierzykowski, Z. et al. (1999) The pattern of homoeologous recombination in triploid hybrids of *Lolium multiflorum* with *Festuca pratensis*. Genome 42, 720–726

36 Nicolas, A. et al. Centre National de la Recherche Scientifique, Institut National de la Sante et de la Recherche Medicale, Institut National de la Recherche Agronomique and Institut Curie. Methods for inducing targeted stimulation of meiotic recombination for performing said methods, WO/2004/016795

37 Pecinña, A. et al. (2002) Targeted stimulation of meiotic recombination. Cell 111, 173–184

38 Robine, N. et al. (2007) Genome-wide redistribution of meiotic doublestrand breaks in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 27, 1868–1880

39 Sears, E.R. (1981) Transfer of alien genetic material to wheat. In Wheat Science – Today and Tomorrow (Evans, L.T. and Peacock, W.J., eds), pp. 75–89, Cambridge University Press

40 Griffiths, S. et al. (2006) Molecular characterization of *Ph*1 as a major chromosome pairing locus in polyploid wheat. Nature 439, 749–752

41 Al-Kaff, N. et al. (2008) Detailed dissection of the chromosomal region containing the *Ph*1 locus in wheat *Triticum aestivum*: with deletion mutants and expression profiling. Ann. Bot. (Lond.) 101, 863–872

42 Qi, L. et al. (2007) Homoeologous recombination, chromosome engineering and crop improvement. Chromosome Res. 15, 3–19

43 Martinez-Perez, E. and Moore, G. (2008) To check or not to check? The application of meiotic studies to plant breeding. Curr. Opin. Plant Biol. 11, 222–227

44 Jenczewski, E. et al. (2003) *PrBn*, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics 164, 645–653

45 Doutriaux, M-P. et al. Aventis Cropscience S.A.. Isolated DNA that encodes an *Arabidopsis thaliana* MSH3 protein involved in DNA mismatch repair and a method of modifying the mismatch repair system in a plant transformed with the isolated DNA, United States Patent 6734019

46 Nogue, F. et al. Institut National de la Recherche Agronomique and Cellectis. Homeologous recombination in *MSH*2 inactivated plants or cells thereof, WO/2006/134496

47 Rozwadowski, K.L. and Lydiate, D.J. Modifying DNA recombination and repair, United States Patent 20060160222

48 Dirks, R.H.G. et al. Rijk Zwaan Zaadteelt en Zaadhandel B.V.. Reverse breeding, WO/2003/017753

49 Siaud, N. et al. (2004) Brca2 is involved in meiosis in *Arabidopsis thaliana* as suggested by its interaction with Dmc1.EMBOJ. 23, 1392–1401

50 Higgins, J.D. et al. (2004) The *Arabidopsis MutS* homolog *AtMSH4* functions at an early step in recombination: evidence for two classes of recombination in *Arabidopsis*. Genes Dev. 18, 2557–2570

51 Couteau, F. et al. (1999) Random chromosome segregation without meiotic arrest in both male and female meiocytes of a dmc1 mutant of *Arabidopsis*. Plant Cell 11, 1623–1634

52 Pradillo, M. et al. (2007) An analysis of univalent segregation in meiotic mutants of *Arabidopsis thaliana:* a possible role for synaptonemal complex. Genetics 175, 505–511

53 Wijeratne, A.J. et al. (2006) The *Arabidopsis thaliana PARTING DANCERS* gene encoding a novel protein is required for normal meiotic homologous recombination. Mol. Biol. Cell 17, 1331–1343
54 Keurentjes, J.J.B. et al. (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics 175, 891–905
55 Peleman, J.D. and van der Voort, J.R. (2003) Breeding by design. Trends Plant Sci. 8, 330–334
56 Preuss, D. et al. (1994) Tetrad analysis possible in *Arabidopsis* with mutation of the *QUARTET* (QRT) genes. Science 264, 1458–1460
57 Copenhaver, G.P. et al. (2000) Tetrad analysis in higher plants. A budding technology. Plant Physiol. 124, 7–16
58 Berchowitz, L.E. and Copenhaver, G.P. (2008) Fluorescent *Arabidopsis* tetrads: a visual assay for quickly developing large crossover and crossover interference data sets. Nat. Protoc. 3, 41–50
59 Copenhaver, G.P. (2005) A compendium of plant species producing pollen tetrads. J. North Carolina Acad. Sci. 12, 17–35
60 Francis, K.E. et al. (2006) Separation of *Arabidopsis* pollen tetrads is regulated by *QUARTET*1, a pectin methylesterase gene. Plant Physiol. 142, 1004–1013
61 Melamed-Bessudo, C. et al. (2005) A new seed-based assay for meiotic recombination in *Arabidopsis thaliana.* Plant J. 43, 458–466
62 Page, S.L. and Hawley, R.S. (2003) Chromosome choreography: the meiotic ballet. Science 301, 785–789
63 Li, L. et al. (2006) The impact of sequence divergence and DNA mismatch repair on homeologous recombination in *Arabidopsis*. Plant J. 45, 908–916
64 Svetlanov, A. and Cohen, P.E. (2004) Mismatch repair proteins, meiosis, and mice: understanding the complexities of mammalian meiosis. Exp. Cell Res. 296, 71–79
65 Prieto, P. et al. (2004) Homologue recognition during meiosis is associated with a change in chromatin conformation. Nat. Cell Biol. 6, 906–908

On germplasm ownership…

September
en ik plukte een roos
en iemand riep:
'Dat mag niet! Die zijn niet van jou…'
[…]

T. Tellegen

(In: 'September'. Een dansschool, 1992).

Chapter 6

Reverse breeding: a novel breeding approach based on engineered meiosis

Rob Dirks[1], Kees van Dun[1], C. Bastiaan de Snoo[1], Mark van den Berg[1], Cilia L. C. Lelivelt[1], William Voermans[1], Leo Woudenberg[1], Jack P. C. de Wit[1], Kees Reinink[1], Johan W. Schut[1], Eveline van der Zeeuw[1], Aat Vogelaar[1], Gerald Freymark[1], Evert W. Gutteling[1], Marina N. Keppel[1], Paul van Drongelen[1], Matthieu Kieny[1], Philippe Ellul[1], Alisher Touraev[2], Hong Ma[3,4], Hans de Jong[5] and T.G. (Erik) Wijnker[5,*]

1) Rijk Zwaan Breeding BV, Fijnaart, The Netherlands
2) Max F. Perutz Laboratories, Department of Plant Molecular Biology, Vienna University, Vienna, Austria
3) Department of Biology, The Huck Institutes of the Life Sciences, The Pennsylvania State University, PA, USA
4) School of Life Sciences, Institute of Plant Biology, Fudan University, Shanghai, China
5) Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands

Plant Biotechnology Journal (2009) 7, pp. 837–845
doi: 10.1111/j.1467-7652.2009.00450.x

Summary

Reverse breeding (RB) is a novel plant breeding technique designed to directly produce parental lines for any heterozygous plant, one of the most sought after goals in plant breeding. RB generates perfectly complementing homozygous parental lines through engineered meiosis. The method is based on reducing genetic recombination in the selected heterozygote by eliminating meiotic crossing over. Male or female spores obtained from such plants contain combinations of non-recombinant parental chromosomes which can be cultured in vitro to generate homozygous doubled haploid plants (DHs). From these DHs, complementary parents can be selected and used to reconstitute the heterozygote in perpetuity. Since the fixation of unknown heterozygous genotypes is impossible in traditional plant breeding, RB could fundamentally change future plant breeding. In this review, we discuss various other applications of RB, including breeding per chromosome.

Introduction

One of the most important insights in plant breeding was the observation that hybrid (F1) progeny typically are superior in size, growth characteristics and yield in comparison to their homozygous parents, a phenomenon known as heterosis. Its underlying driving mechanisms may be multiple and are unfortunately poorly understood (Springer and Stupar, 2007; Stupar *et al*., 2008; Fernandez-Silva *et al*., 2009; Wei *et al*., 2009). The unpredictable nature of heterosis confronts breeders with considerable difficulties: how does one optimize the performance of crop varieties when the constituents for success are unknown? Breeders can evaluate heterosis by controlled crosses of inbred lines (i.e. by *apriori* selection and combination of unknown alleles). The hit-or-miss nature of this approach makes it difficult to optimize the effects of heterosis. Here, we propose an alternative strategy based on the reversal of crop selection: the generation of defined populations with high levels of heterozygosity and random variation. These populations are then assessed in a variety of environmental conditions (latitude, salinity, humidity, etc.) and the best performing heterozygous germplasm is selected for further breeding.

A barrier to achieving high levels of variation in current plant breeding programs is that uncharacterised heterozygotes are difficult—if not impossible—to reproduce by seeds. Favourable allele combinations of the elite heterozygote are lost in the next generation due to segregation of traits. Because of this difficulty, the development of methods for easy preservation of heterozygous genotypes is one of the greatest challenges in plant breeding. Apomixis has repeatedly been proposed as a way to preserve heterozygous phenotypes, but has not yet led to breeding applications (Perotti *et al.*, 2004).

In this paper, we show how a new technique, reverse breeding, meets the challenge of fixation of complex heterozygous genomes by constructing complementing homozygous lines (Dirks *et al*., 2003). This is accomplished by the knockdown of meiotic crossovers and the subsequent fixation of non-recombinant chromosomes in homozygous doubled haploid lines (DHs). The approach not only allows fixation of uncharacterized germplasm but provides breeders with a breeding tool that, when applied to plants of known back-

grounds, allows the rapid generation of chromosome substitutions that will facilitate breeding on an individual chromosome level. After a brief introduction to the RB breeding scheme, we first elaborate on the basis of RB: the unique character of achiasmatic meiosis. Thereafter, we show how the technique may be implemented in crops followed by a discussion of its main applications.

Reverse breeding

Reverse breeding comprises two essential steps: the suppression of crossover recombination in a selected plant followed by the regeneration of DHs from spores containing non-recombinant chromosomes. Figure 1 shows an idealized crossing scheme that employs RB. It depicts the generation of a segregating population (in this case a segregating F2), from which a genotypically uncharacterized plant with a favourable combination of traits is selected. Crossing over is suppressed in this plant and achiasmatic gametes are collected, cultured, and used to generate DHs. The DH lines can then be used to recapitulate the elite heterozygote on a commercial scale.

In another application, RB can be applied to plants of known background (Figure 2). If crossing over is eliminated in the F1 hybrid rather than the F2 generation, RB can be used to generate chromosome substitution lines. These lines contain one or more chromosomes from one parent in the background of the other parent. By backcrossing the chromosome substitution lines to the original parental lines, one can obtain populations that segregate only for the heterozygous chromosome(s). Reverse breeding, in theory, allows the re-shuffling of chromosomes between two homozygous plants in all possible ways.

Reverse breeding relies on achiasmatic meiosis

On the function of crossovers

In flowering plants, the formation of crossovers during meiotic prophase I relies on synapsis, the extensive and stable interaction between homologous chromosomes, mediated by a complex proteinacious structure called the synaptonemal complex (Moses, 1956). During crossing over, two homologues become physically joined when the distal end of one chromatid is attached to the proximal end of a non-sister chromatid and vice versa (Figure 3). The resulting intermediate of joined homologues is called a bivalent. Cross-over sites are visible as cross-like structures after synaptonemal complex disassembly, the chiasmata. They are usually maintained until metaphase I / anaphase I, when homologues segregate to opposite poles. In most plants, a chromosome pair typically has one or two crossovers. Many mutants have been described that reduce or eliminate crossovers (reviews in Hamant *et al*., 2006; Noyes *et al*., 2006; Roeder, 1990; Zickler and Kleckner, 1999).

Achiasmatic chromosomes (chromosomes that did not form crossovers) remain as univalents (Figure 3). Chiasmata, that in bivalents promote segregation of homologues

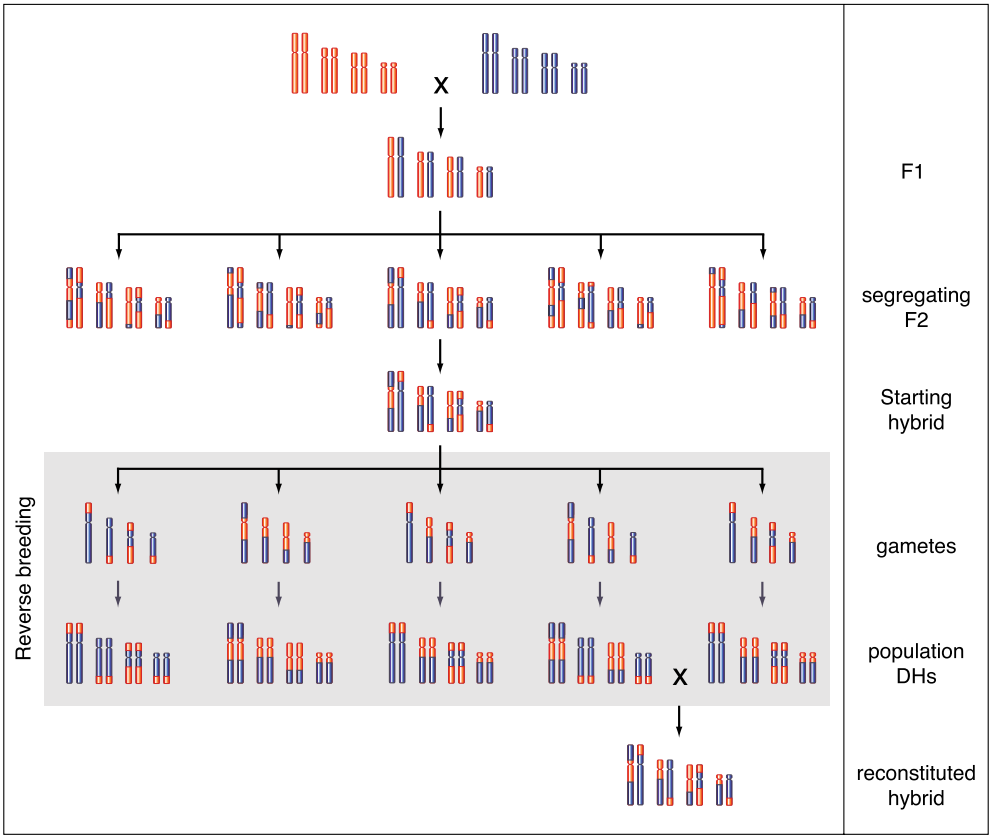


Figure 1. Reverse breeding can be used to fix unknown heterozygotes. Crossing two homozygous parents (red and blue bars) creates a heterozygous F1. When selfed, the F1 produces a segregating F2 population. A starting hybrid of unknown genetic constitution is selected for its desireable characteristics, and subjected to the two steps of reverse breeding (grey box). By knocking down meiotic crossing over, whole parental chromosomes are transmitted through spores, without rearrangement. Note, in this example the four chromosomes in the hybrid can generate 16 different combinations in the gametes—only five are shown for convenience. The achiasmatic gametes are then used produce doubled haploid (DH) lines using in vitro culture techniques. From this population, complementary parents can be chosen that when crossed perfectly reconstitute the starting hybrid. The DH lines then serve as a permanent library that can be used to predictably generate a wide variety of defined hybrids.

to opposite poles (regular disjunction), are absent in univalents and the homologues may segregate to the same pole instead (non-disjunction). This leads to unbalanced chromosome numbers (aneuploidy) in the spores. Consequently, achiasmatic plants are highly sterile (Couteau *et al.*, 1999; Hartung *et al.*, 2007). The more univalents are present, the more aneuploid pollen are formed. Assuming that each univalent has an equal chance of moving to either pole, the probability of a spore with a normal chromosome complement is $½^x$, with x equalling the haploid chromosome number of the species. Hence, for

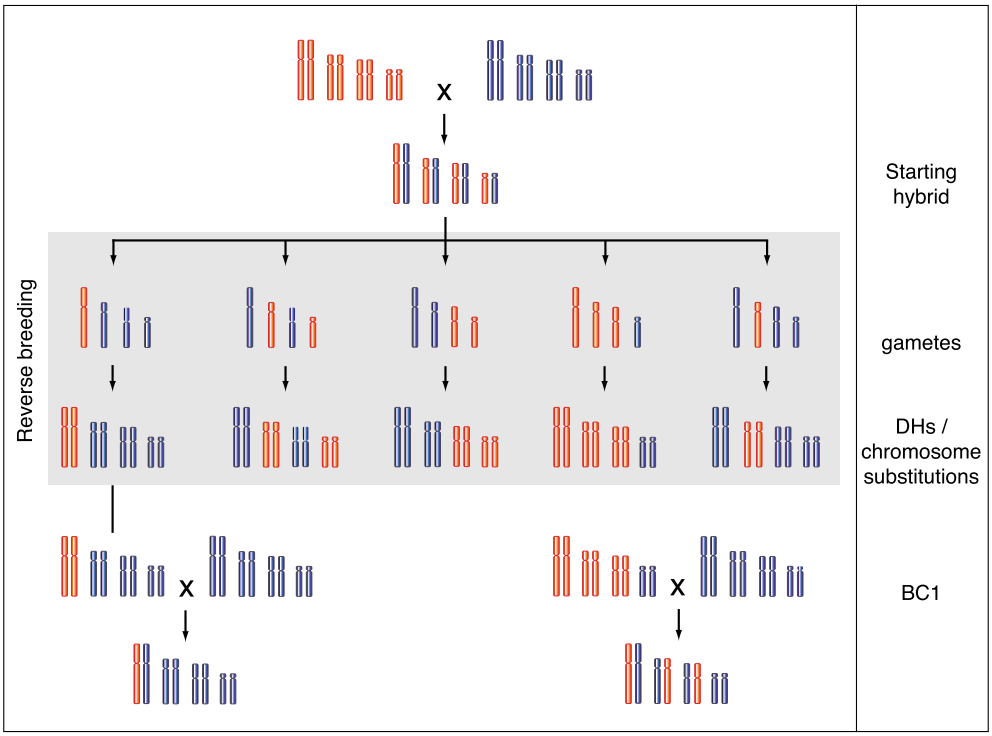


Figure 2 Reverse breeding can be used as advanced breeding tool. As starting hybrid for a reverse breeding experiment, a fully heterozygous F1 is chosen, resulting from a cross between two homozygous parents. Application of reverse breeding (grey box) leads to a population of doubled haploids. Note that among those DHs, there are chromosome substitution lines of one of the starting parents into the backgroud of the other. Lower left: a chromosome substitution line for a red chromosome in the blue parent can be backcrossed with the fully blue parent to create a hybrid that is heterozygous for just one chromosome. Such hybrids serve as starting point for breeding per chromosome (explained in text). Lower right: a chromosome substitution line for a blue chromosome in the red parent can be backcrossed with the fully black parent to create a hybrid that is homozygous for just one chromosome. Such hybrids are starting points for studying background interactions (explained in text).

Arabidopsis (2n = 2x = 10), the frequency of balanced spores is one in 32 (3%). For plants with a basic chromosome number exceeding 12, the chances of finding balanced gametes get very (perhaps too) small (with only one out of over 4000 spores being balanced). This has to be evaluated for individual crops. In the case of *Petunia* with an estimated number of 30,000 microspores per anther (Kapoor *et al.*, 2002) and seven chromosome pairs, we expect that the number of euploid spores per anther will be 235. *Arabidopsis* by comparison only produces 2800 spores per flower on average (Noyes *et al.*, 2006) and will generate 88 euploid spores assuming no bivalent formation and random chromosome distribution.

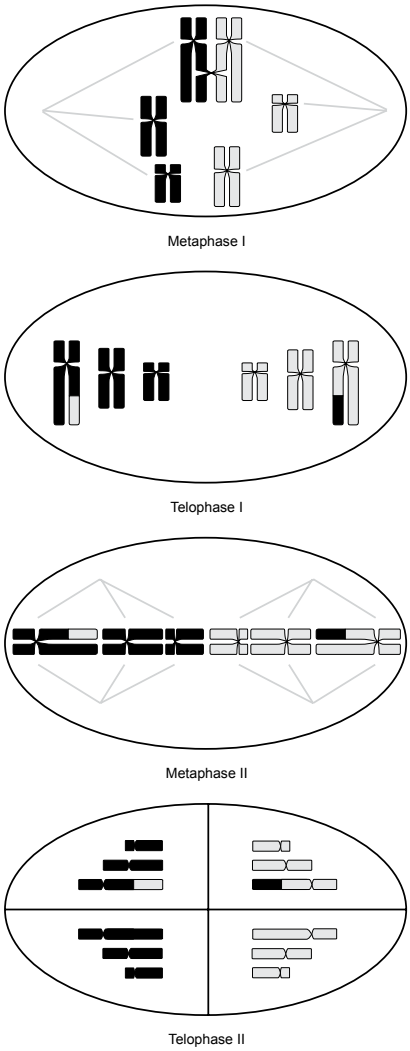


Figure 3 The presence of a single crossover in a chromosome pair does not affect the utility of reverse breeding. The figure depicts four cells at different stages of meiosis. At metaphase I, a single crossover is present in one chromosome pair (a bivalent pair) whereas other homologues remain as univalents. The homologues joined by a chiasma segregate to opposite poles and—in this example—the univalents segregate randomly to opposite poles, giving rise, in this case, to a balanced dyad (at telophase I). Meiosis then proceeds through metaphase II, separating sister chromatids, and at telophase II four gametes are formed. Half of these gametes contain a recombinant chromosome (upper two), whereas the other half contain non-recombinant chromosomes (lower two) and are useful for reverse breeding.

Chances of finding complementing parents

The maximum number of different DHs obtained from a heterozygous diploid in a RB experiment equals 2x, with x being the basic chromosome number. The probability that two DHs form a pair of 'complementary' parents (as shown in Figure 2) equals $2^x / (2x)^2 = (½)^x$, and the probability that they, upon crossing, do not reconstruct the original genotype is $1-(½)^x = (2^x-1) / 2^x$. The number of combinations between different DHs, presuming that reciprocal crosses result in the same phenotype, is $n(n-1) / 2$.

In the case of n DHs, the probability of not finding a complementary pair of lines is therefore $[(2^x-1) / 2^x]^{n(n-1) / 2}$ and the probability of at least one complementary combination of two DHs is given by the formula $[(2^x-1) / 2^x]^{n(n-1)} / 2 = 0.01$ (P = 99%). This equation can be

Table 1. Number of non-recombinant DHs required for reconstructing the original starting plant at different probability levels in various species.

haploid chromosome number	Probably				Model species / crop
	0.90	0.95	0.99	1.00	
5	13	14	18	47	*Arabidopsis*
6	18	20	25	67	Spinach, corn salad
7	25	28	35	94	Cucumber, barley, rye
8	35	40	49	133	Onion
9	49	56	69	188	Carrot, sugarbeet, most vegetable Brassicas, lettuce
10	69	79	98	266	Maize, *Sorghum, Asparagus*, cacao
11	98	111	138	377	Banana, watermelon, celery, fennel, common bean
12	138	157	195	532	Tomato, pepper, melon, rice, egg plant

solved for different values of x. The number *(n)* of DHs that must be generated for finding a complementary match is highly dependent on the haploid chromosome number (x) and is given in Table 1.

The technical realization of reverse breeding

Effective suppression of recombination

Reverse breeding relies on the effective suppression of meiotic crossovers. Therefore, genes that are essential in crossover formation but leave the chromosome structure intact are particularly useful. Examples are the *Arabidopsis ASY*1 and the rice *ASY*1 homologue *PAIR*2, the mutants of which display univalents at metaphase I (Ross *et al.*, 1997; Caryl *et al.*, 2000; Nonomura *et al.*, 2004). Other mutants with similar phenotypes are *dmc*1, *sds*, *ptd* and *spo*11 (Couteau *et al.*, 1999; Azumi *et al.*, 2002; Stacey *et al.*, 2006; Wijeratne *et al.*, 2006).

The knockdown of gene expression, essential for RB, can be achieved by targeting genes using RNA interference (RNAi) (as shown by Siaud *et al.*, 2004; Higgins *et al.*, 2004) or siRNAs, which will result in predominantly post-transcriptional gene silencing (PTGS). Alternatively, dominant-negative mutations of the target gene can be used. The human meiotic recombination protein DMC1 forms octomeric rings (Kinebuchi *et al.*, 2004), but is fully defective in both ssDNA and dsDNA binding activities, when an amino terminal deletion lacking 81 amino acids is made (Kinebuchi *et al.*, 2005). Similar dominant-negative alleles of *DMC*1 resulted in loss of male meiotic recombination in mice (Bannister *et al.*, 2007).

In crops in which stable transformation is difficult or impossible to achieve, other techniques could be applied. Virus-induced gene silencing (VIGS) was shown to be an

effective technique for induction of PTGS. A plant then is infected with a virus that was modified to include a target gene RNA sequence. In a defence reaction, the plant will break down the viral RNA using siRNA, targeting simultaneously the plants' endogenous mRNA (Ruiz *et al*., 1998; Baulcombe, 2004). Alternatively, target genes may be silenced by silencing molecules delivered by graft transmission (Shaharuddin *et al*., 2006). Shoots of the plant in which genes are to be silenced would be grafted on transgenic rootstocks. In this case, only few transgenic rootstocks would be required to routinely apply RB in many crops. Another more recent approach is based on a forward chemical genetic screen that identified 'mirin' as an inhibitor for the Mre11-Rad50-Nbs1 complex (Dupré *et al*., 2008). Exogenous application of compounds that cause inhibition or omission of recombination during meiosis would speed up the application of RB enormously.

A major advantage for using chemicals that repress crossovers or graft transmission of silencing molecules is that the resultant RB products (DHs) are free of transgenes. This is important, because the RB products are destined to be used in further breeding schemes, and should not have a achiasmatic phenotype. Perhaps contrary to intuition, DHs produced by transgene-mediated methods can be transgene free. If a dominant knock-down construct is present in hemizygous state, half of the spores that are formed will not carry the transgene and, hence, are non-transgenic. Multiple transgenic lines with knockdown constructs on different chromosomes can be used to generate a full array of complementary DHs that do not carry transgenes (Wijnker and de Jong, 2008). Crossover suppression need not be complete to be useful for RB. It can be explained that a single residual crossover may still occur in any chromosome pair(s). A single crossover causes regular segregation of the homologues involved (thereby increasing the chance of obtaining a balanced gamete twofold). A crossover also generates two recombinant chromatids, which are not useful for RB. But since a crossover affects only half of the chromatids of the bivalent pair, the other two chromatids are non-recombinant, and useful. Consequently, half of the resulting spores are potentially useful for RB (Figure 3). In short, residual crossovers (provided there is only one per bivalent) increase the incidence of DHs carrying recombinant chromosomes, but still produce 50% of spores carrying non-recombinant chromosomes. These non-recombinant spores can be selected for by using molecular markers.

Doubled haploids

Doubled haploid plants resulting from achiasmatic meiosis can be obtained from unfertilized ovules (gynogenesis) or from microspore and anther cultures (androgenesis), according to well-established protocols that have been developed for a variety of plant species including crops (Jain *et al*., 1996). The efficiency of DH formation from haploid spores is species dependent (Forster *et al.*, 2007). The unique characteristic of DHs made from spores produced through achiasmatic meiosis is explained in Figure 1: they contain non-recombinant parental chromosomes. Note however that aneuploid unfertile spores,

which are in fact most prevalent, were not depicted. Selection of the required euploid spores is in part automatically achieved since only spores containing at least one copy of all chromosomes can pass through all developmental stages, from cell division and embryogenesis to plant regeneration. Hyperploid offspring could be selected against using co-dominant markers or flow cytometry.

Development of RB is limited to those crops where DH technology is common practice. For the great majority of crop species this technology is well established and professional breeding companies routinely use such techniques in their breeding programs (Maluszynski *et al.*, 2003; Forster *et al.*, 2007). There are, however, some notorious exceptions such as soybean, cotton, lettuce and tomato where doubled haploid plants are rarely formed or not available at all (Croser *et al.*, 2006; Segui-Simarro and Nuez, 2007; Zhang *et al.*, 2008). Genotyping of DHs by molecular markers is routine practice in contemporary plant breeding (De Vienne, 2003) and is also indispensable for RB. In the complete absence of meiotic recombination one polymorphic molecular marker per chromosome would suffice to genotype every DH since the entire chromosome would behave as a single linkage block. In the presence of any residual crossovers, two markers (as distally located as possible) are required per chromosome.

Reverse breeding applications

Reconstruction of heterozygous germplasm

For crops where an extensive collection of breeding lines is still lacking, RB can accelerate the development of varieties. In these crops, superior heterozygous plants can be propagated without prior knowledge of their genetic constitution (also see Figure 1). Table 1 shows the number of doubled haploid plants that are necessary to reconstruct the starting plant at different levels of probability. The number of DHs that is required is surprisingly low. For instance in maize (x = 10) just 98 DHs are expected to contain a set of two reciprocal genotypes (P = 99%).

Breeding on the single chromosome level

Many interesting characteristics in crops are based on polygenic gene interactions, very often located on different chromosomes. These quantitative traits are therefore not easy to breed on. Figure 2 explains how chromosome substitution lines can be obtained when RB is applied to an F1 hybrid of known parents. These homozygous chromosome substitution lines provide novel tools for the study of gene interactions. When crossed with one of the original parents, hybrids can be formed in which one of the chromosomes is homozygous (Figure 2, lower right), whereas it is also possible to produce hybrids in which just one chromosome is heterozygous (Figure 2, lower left). The former allows the study of epistatic interactions between the background and genes contributed by the substitution chromosome. Offspring of plants in which just one chromosome is heterozygous, will segregate for traits present on that chromosome only. Selfing plants that

carry a substituted chromosome (or using recurrent backcrosses) will allow breeders to fine-tune interesting characteristics on a single chromosome scale. This could bring forth improved breeding lines carrying introgressed traits. The few examples were shown here demonstrate that RB presents breeders with full control over homo- or heterozygosity at the single chromosome level.

Note that finding specific substitution lines may be difficult, since they are rare occurrences. Depending on the efficiency of the DH system, especially crops with high chromosome numbers may pose problems. In these cases backcrossing a DH line carrying the desired substitution in addition to another (undesired substitution) with one of the original parents may be helpful. Using marker assisted breeding the desired chromosome substitution can be obtained with relative ease.

Reverse breeding and marker assisted breeding

Especially in combination with (high throughput-) genotyping, reverse breeding becomes a versatile tool. Evidently, high throughput genotyping speeds up the process of identification of complementing parents in populations of DHs in early stages. But perhaps more powerful is its use in the study of gene interactions of the various heterozygous inbred families (HIFs) that can be produced by crossing and backcrossing the products of RB (as was explained above). The screening of populations that segregate for traits on a single chromosome allow the quick identification of QTLs, when genotyping is combined with –for example- transcriptome or metabolome profiling. Such HIFs further aid the generation of chromosome specific linkage maps and the fine mapping of genes and alleles. RB can as such provide highly valuable insights into the nature of heterotic effects.

Backcrossing in CMS back ground

In several vegetable crops such as cabbages and carrots, breeders make use of cytoplasmic male sterility (CMS) (Chase, 2007). In these systems, the presence of male sterility presents a special challenge to RB. In these cases, gynogenesis rather than androgenesis can be used to obtain DH plants. This is perfectly compatible with RB in the sense that the chromosomes from the maintainer line can be recovered directly in the cytoplasm of the sterile line in one step. Gynogenesis has been described in several crops such as *Brassica*, maize, sugar beet, cucumber, melon, rice, onion, sunflower, and barley (Keller and Korzun, 1996). However, the development or improvement of the protocol for many species was often abandoned when anther and microspore culture techniques were developed. In cases where the efficiency of gynogenesis is too low, it is possible to cross the male sterile (A) lines with maintainer lines (B) that carry one copy of a restorer gene. The AB combination will be fertile and RB can be performed. In rice, restorer genes have been successfully transformed (Wang *et al*., 2006). It should therefore be possible

to use a restorer gene and a gene for crossover suppression in the same vector (both transgenes) and perform RB in a 'double suppressed' (CMS and crossover) background.

Conclusions

The combination of crossover suppression, followed by the regeneration of haploid spores into DHs results in novel and powerful breeding applications. One important application is the production of complementary homozygous lines that can be used to generate specific F1 hybrids. Additionally, when RB is applied to F1 heterozygotes, it is possible to generate chromosome substitution lines that allow targeted breeding on the single chromosome scale. RB is fully compatible with commercial CMS lines that are frequently used in modern agriculture.

The technique however is limited to crops with a haploid chromosome number of 12 or less and in which spores can be regenerated into DHs. In polyploids or species with high chromosome numbers, another reconstruction method has been proposed that is based on the omission of the second meiotic division, leading to unreduced second division restitution (SDR) spores. The use of these SDR spores enables the near reconstruction of desired phenotypes, and also provides the possibility of obtaining chromosome substitution lines (Van Dun and Dirks, 2006).

There is growing interest in the development of plant breeding techniques that are based in modifications of meiosis (Wijnker and de Jong, 2008). However, most techniques are merely extensions of the 'classic' plant breeding practice aimed at more efficient introgression of traits from alien backgrounds into crops. Pivotal for understanding the expected impact of germplasm fixation on plant breeding should be the realization that plant breeding relies heavily on the human eye for the selection of breeding lines. It is not difficult to imagine that selection for (overdominant) complex traits or QTLs is a daunting task. Visual selection is therefore always accompanied by extensive testcrosses aimed at control avoiding the loss of valuable traits during selection. Methods that allow the fixation of elite germplasm (apomixis and reverse breeding) provide alternatives to this selection process. Though reverse breeding may appear more complex than apomixis at a first glance, it does not suffer from the drawback of the current knowledge of apomixis where the three mechanisms essential for apomixis (apomeiosis, parthenogenesis and endosperm formation) have to be operational and synchronized (Koltunow and Grossniklaus, 2003). As a plant breeding tool, reverse breeding may be regarded more versatile as its controlled deconstruction of complex genotypes into homozygous parental lines allows the further improvement of these lines by classic breeding methods.

Acknowledgements

We would like to thank Greg Copenhaver for valuable comments during the completion of our manuscript. The first author, is greatly indebted to Coert van Eijnsbergen for his continuous sturdy support.

References

Azumi, Y., Liu, D., Zhao, D., Li, W., Wang, G., Hu, Y. and Ma, H. (2002) Homolog interaction during meiotic prophase I in *Arabidopsis* requires the *SOLO DANCERS* gene encoding a novel cyclin-like protein. EMBO J. 21, 3081–3095.

Bannister, L.A., Pezza, R.J., Donaldson, J.R., Rooij, D.Gd., Schimenti, K.J., Camierini- Otero, R.D. and Schimenti, J.C. (2007) A dominant, recombination-defective allele of *DMC*1 causing male-specific sterility. PLoS Biol. 5, 1016–1025.

Baulcombe, D. (2004) RNA silencing in plants. Nature, 431, 356– 363.

Caryl, A.P., Armstrong, S.J., Jones, G.H. and Franklin, F.C.H. (2000) A homologue of the yeast *HOP*1 gene is inactivated in the *Arabidopsis* meiotic mutant *asy*1. Chromosoma, 109, 62– 71.

Chase, C.D. (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial nuclear interactions. Trends Genet. 23, 81–90.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D. and Doutriaux, M.-P. (1999) Random chromosome segregation without meiotic arrest in both male and female meiocytes of a dmc1 mutant of *Arabidopsis*. Plant Cell, 11, 1623–1634.

Croser, J., Lülsdorf, M., Davies, P., Clarke, H., Bayliss, K., Mallikarjuna, N. and Siddique, K. (2006) Toward doubled haploid production in the Fabaceae: progress, constraints, and opportunities. Crit. Rev. Plant Sci. 25, 139–157.

De Vienne, D. (2003) Molecular Markers in Plant Genetics and Biotechnology. Enfield, NH, USA: Science Publishers, Inc.

Dirks, R.H.G., van Dun, C.M.P. and Reinink, K. (2003) Rijk Zwaan Zaadteelt en Zaadhandel B.V. Reverse Breeding, WO/ 2003 / 017753.

Dupré, A., Boyer-Chatenet, L., Sattler, R.M., Modi, A.P., Lee, J.-H., Nicolette, M.L., Kopelovich, L., Jasin, M., Raer, R., Paull, T.T. and Gautier, J. (2008) A forward chemical genetic screen reveals an inhibitor of the mre11-rad50-nbs1 complex. Nat. Chem. Biol. 4, 119–125.

Fernandez-Silva, I., Moreno, E., Eduardo, I., Arus, P., Alvarez, J.M. and Monforte, A.J. (2009) On the genetic control of heterosis for fruit shape in melon (*Cucumis melo* L.). J. Hered. 100, 229–235.

Forster, B.P., Heberle-Bors, E., Kasha, K.J. and Touraev, A. (2007) The resurgence of haploids in higher plants. Trends Plant Sci. 12, 368–375.

Hamant, O., Ma, H. and Cande, W.Z. (2006) Genetics of meiotic prophase I in plants. Annu. Rev. Plant Biol. 57, 267–302.

Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S. and Puchta, H. (2007) The catalytically active tyrosine residues of both *SPO*11-1 and *SPO*11-2 are required for meiotic double-strand break induction in *Arabidopsis*. Plant Cell, 19, 3090– 3099.

Higgins, J.D., Armstrong, S.J., Franklin, F.C.H. and Jones, G.H. (2004) The *Arabidopsis MutS* homolog *AtMSH*4 functions at an early step in recombination: evidence for two classes of recombination in *Arabidopsis*. Genes. Dev, 18, 2557–2570.

Jain, S., Sopory, S.K. and Veilleux, R.E. (1996) In vitro haploid production in higher plants. In Current Plant Science and Biotechnology in Agriculture, Dordrecht, Boston, London: Kluwer Academic Publishers.

Kapoor, S., Kobayashi, A. and Takatsuji, H. (2002) Silencing of the tapetum-specific zinc finger gene *TAZ*1 causes premature degeneration of tapetum and pollen abortion in *Petunia*. Plant Cell, 14, 1–15.

Keller, E.R.J. and Korzun, L. (1996) Ovary and ovule culture for haploid production. In *In vitro* Haploid Production in Higher Plants (Jain, S.M., Sopory, S.K. and Veilleux, R.E., eds), pp. 217–235, Dordrecht: Kluwer Acad Publ.

Kinebuchi, T., Kagawa, W., Enomoto, R., Tanaka, K., Miyagawa, K., Shibata, T., Kurumizaka, H. and Yokoyama, S. (2004) Structural basis for octameric ring formation and DNA interaction of the human homologous-pairing protein *DMC*1. Mol. Cell, 14, 363–374.

Kinebuchi, T., Kagawa, W., Kurumizaka, H. and Yokoyama, S. (2005) Role of the N-terminal domain of the human *DMC*1 protein in octamer formation and DNA binding. J. Biol. Chem. 280, 28382–28387.

Koltunow, A.M. and Grossniklaus, U. (2003) Apomixis: a developmental perspective. Annu. Rev. Plant Biol. 54, 547–574.

Maluszynski, M., Kasha, K.J. and Szarejko, I. (2003) Published protocols for other crop plant species. In Doubled Haploid Production in Crop Plants: A Manual (Maluszynski, M., Kasha, K.J., Forster, B.P. and Szarejko, I., eds), pp. 309–336, Dordrecht, the Netherlands: Kluwer Academic Publishers.

Moses, M.J. (1956) Chromosomal structures in crayfish spermatocytes. J. Biophys. Biochem. Cytol. 2, 215–218.

Nonomura, K.I., Nakano, M., Murata, K., Miyoshi, K., Eiguchi, M., Iayao, A., Hirochika, H. and Kurata, N. (2004) An insertional mutation in the rice *PAIR*2 gne, the ortholog of *Arabidopsis ASY*1, results in a defect in homologous chromosome pairing during meisosis. Mol. Genet. Genomics, 271, 121–129.

Noyes, R.D., Baker, R. and Mai, B. (2006) Mendelian segregation for two-factor apomixis in *Erigeron annuus* (Asteraceae). Heredity, 98, 92–98.

Perotti, E., Grimanelli, D., John, P., Hoisington, D. and Leblanc, O. (2004) Why is transferring apomixis to crops still a dream? In New Directions for a Diverse Planet: Proceedings for the 4th International Crop Science Congress (Fischer, T., Turner, N., Angus, J., McIntyre, L., Robertson, M., Borrell, A. and Lloyd, D., eds), Brisbane, Australia: URL: http://www.cropscience.org.au/ icsc2004/poster/3/2/1/1367_perottie.htm.

Roeder, G.S. (1990) Chromosome synapsis and genetic recombination: their roles in meiotic chromosome segregation. Trends Genet. 6, 385–389.

Ross, K., Fransz, P., Armstrong, S., Vizir, I., Mulligan, B., Franklin, F. and Jones, G. (1997) Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA transformed lines. Chromosome Res. 5, 551–559.

Ruiz, M.T., Voinnet, O. and Baulcombe, D.C. (1998) Initiation and maintenance of virus-induced gene silencing. Plant Cell, 10, 937–946.

Segui-Simarro, J.M. and Nuez, F. (2007) Embryogenesis induction, callogenesis, and plant regeneration by in vitro culture of tomato isolated microspores and whole anthers. J. Exp. Bot. 58, 1119–1132.

Shaharuddin, N.A., Yuanhuai, H., Hongying, L. and Grierson, D. (2006) The mechanism of graft transmission of sense and antisense gene silencing in tomato plants. FEBS Lett. 580, 6579–6586.

Siaud, N., Dray, E., Gy, I., Gérard, E., Takvorian, N. and Doutriaux, M.-P. (2004) Brca2 is involved in meiosis in *Arabidopsis thaliana* as suggested by its interaction with Dmc1. EMBO J 233, 1392–1401.

Springer, N.M. and Stupar, R.M. (2007) Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Res. 17, 264–275.

Stacey, N.J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K. and Sugimoto-Shirasu, K. (2006) *Arabidopsis SPO*11-2 functions with *SPO*11-1 in meiotic recombination. Plant J. 48, 206–216.

Stupar, R., Gardiner, J., Oldre, A., Haun, W., Chandler, V. and Springer, N. (2008) Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biol. 8, 33.

Van Dun, C.M.P. and Dirks, R.H.G. (2006) Rijk Zwaan Zaadteelt en Zaadhandel B.V. Near Reverse Breeding, WO/ 2006 / 094773.

Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., Long, Y., Zhong, Y. and Liu, Y.-G. (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell, 18, 676–687.

Wei, G., Tao, Y., Liu, G., Chen, C., Luo, R., Xia, H., Gan, Q., Zeng, H., Lu, Z., Han, Y., Li, X., Song, G., Zhai, H., Peng, Y., Li, D., Xu, H., Wei, X., Cao, M., Deng, H., Xin, Y., Fu, X., Yuan, L., Yu, J., Zhu, Z. and Zhu, L. (2009) A transcriptomic analysis of superhybrid rice *LYP*9 and its parents. PNAS, 106, 7695– 7701.

Wijeratne, A.J., Chen, C., Zhang, W., Timofejeva, L. and Ma, H. (2006) The *Arabidopsis thaliana PARTING DANCERS* gene encoding a novel protein is required for normal meiotic homologous recombination. Mol. Biol. Cell 17, 1331–1343.

Wijnker, E. and de Jong, H. (2008) Managing meiotic recombination in plant breeding. Trends Plant Sci. 13, 640–646.

Zhang, X, Jin, S, Liu, D, Lin, Z, Zeng, F, Zhu, L, Tu, L and Guo, X (2008) Cotton biotechnology: challenge the future for cotton improvement. In Advances in Plant Biotechnology (Rao, G.P., Zhao, Y., Radchuk, V.V. and Bhatnagar, S.K., eds), pp. 241– 301. Houston, Texas: Studium Press LLC.

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

For Simon Chan (1974-2012)

We turn away to face the cold, enduring chill
As the day begs the night for mercy love
The sun so bright it leaves no shadows
Only scars
Carved into stone
On the face of earth
The moon is up and over One Tree Hill
We see the sun go down in your eyes

You run like a river, on like a sea
You run like a river runs to the sea

[…]

U2

(In: 'One Tree Hill'. The Joshua Tree, 1987)

Chapter 7

Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant

T.G. (Erik) Wijnker[1,7], Kees van Dun[2,7], C Bastiaan de Snoo[2], Cilia L C Lelivelt[2], Joost J B Keurentjes[1,3], Nazatul Shima Naharudin[4], Maruthachalam Ravi[5], Simon W L Chan[5,6], Hans de Jong[1] & Rob Dirks[2]

1) Laboratory of Genetics, Wageningen University, Wageningen, The Netherlands.
2) Rijk Zwaan R&D Fijnaart, Fijnaart, The Netherlands.
3) Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands.
4) Department of Crop Science, Faculty of Agriculture, Universiti Putra Malaysia, Selangor, Malaysia.
5) Department of Plant Biology, University of California, Davis, Davis, California, USA.
6) Howard Hughes Medical Institute, University of California, Davis, Davis, California, USA.
7) These authors contributed equally to this work.
Correspondence should be addressed to E.W. (erik.wijnker@wur.nl).

Nature Genetics 44, 467–470 (2012), doi:10.1038/ng.2203

Summary

Traditionally, hybrid seeds are produced by crossing selected inbred lines. Here we provide a proof of concept for reverse breeding, a new approach that simplifies meiosis such that homozygous parental lines can be generated from a vigorous hybrid individual. We silenced DMC1, which encodes the meiotic recombination protein DISRUPTED MEIOTIC cDNA1, in hybrids of *A. thaliana*, so that non-recombined parental chromosomes segregate during meiosis. We then converted the resulting gametes into adult haploid plants, and subsequently into homozygous diploids, so that each contained half the genome of the original hybrid. From 36 homozygous lines, we selected 3 (out of 6) complementing parental pairs that allowed us to recreate the original hybrid by intercrossing. In addition, this approach resulted in a complete set of chromosome-substitution lines. Our method allows the selection of a single choice offspring from a segregating population and preservation of its heterozygous genotype by generating homozygous founder lines.

Introduction

Hybrid vigor is essential to produce high-yielding varieties in many crops[1]. However, a favorable heterozygous genotype cannot be stably propagated through seeds because parental chromosomes will recombine before being passed on to progeny. Recombination of alleles is influenced by two interrelated events during meiosis I: crossover recombination and the orientation of homologous chromosomes on the metaphase plate. Crossover recombination leads to new allele combinations by reciprocal exchange of chromosome segments, whereas random orientation during metaphase I generates novel combinations of parental chromosomes. If there were no crossover recombination, the only factor diversifying genetic information would be chromosome orientation, and, consequently, intact parental chromosomes would segregate to the gametes.

Because hybrids cannot be stably maintained, breeders recreate elite hybrids afresh through crossing homozygous parental lines. Such lines simultaneously provide the means to improve hybrid performance by improving its parents. The inability of breeders to easily establish breeding lines for uncharacterized heterozygotes is a major obstacle to adopting elite heterozygotes from outbreeding (or other segregating) populations into their hybrid breeding programs. Clonal propagation (or apomixis[2]) allows for the preservation of the parental genotype, but prevents its further improvement through adapting parental lines.

We previously introduced an approach termed reverse breeding[3] in which meiotic recombination is suppressed and gametes are directly converted into adult plants. Here we show that reverse breeding can construct homozygous parental lines that, when mated, perfectly reconstitute the selected heterozygous genotype. These homozygous parents can be propagated indefinitely and crossed at will, in contrast to a heterozygote, which will lose its desirable genotype if propagated sexually.

RESULTS

Silencing of DMC1 suppresses crossover recombination

The first step in reverse breeding is to produce gametes from the desired heterozygote without crossover recombination. This is best achieved by dominantly suppressing one of several genes required for meiotic recombination4; complete knockout of a gene by a recessive mutation is not suitable for this purpose as it would reintroduce the same mutation into the reverse-breeding offspring. We therefore used RNA interference (RNAi) to knock down the function of the *RecA* homolog *DMC*1, a meiosis-specific recombinase essential for the formation of crossovers. As RNAi is genetically dominant, it is easy to obtain progeny devoid of the RNAi cassette that would otherwise cause sterility phenotypes among reverse-breeding offspring. RNAi silencing is easy to implement in many crops, and a single cassette targeting a well-conserved meiotic gene can be used across multiple crop species. We used the *Brassica carinata DMC1* gene (91.1% identity to *A. thaliana DMC1*) to silence *A. thaliana DMC*1 (**Supplementary Fig. 1**).

Among 50 RNAi-transformed plants, we observed a range of fertility phenotypes. These ranged from plants with siliques of the normal wild-type length to almost sterile plants (ten plants, 20% of the total) that harbored few or no seeds in their siliques (**Supplementary Fig. 2**) and produced irregularly sized pollen (**Supplementary Fig. 3**). Severely RNAi-transformed plants thus phenocopied the semi-sterile phenotype of A. thaliana dmc1 mutants5. Transcriptional analyses of the RNAi cassette and the endogenous *DMC1* gene further confirmed the effective knockdown of the meiotic recombination pathway (Supplementary Fig. 4). For our experiments we selected those transformants that showed the most sterile phenotypes (that is, those carrying the shortest siliques).

Balanced gametes can be produced in the absence of chiasmata

By physically linking homologous chromosomes, chiasmata ensure their proper segregation during meiosis I. In the absence of crossovers (achiasmatic meiosis), non-recombinant chromosomes segregate randomly. This segregation is generally unbalanced, leading to aneuploidy in gametes and hence explaining the sterile phenotype of achiasmatic plants6. However, by chance balanced gametes are also produced at frequencies that depend on the plant's chromosome number[3]. *A. thaliana* (n = 5) theoretically produces 3.25% balanced gametes ($2^{-5} = 1/32$). Such numbers predict the ample production of viable pollen in the absence of crossovers, even for species with higher chromosome numbers.

We examined meiotic cell spreads in more than 100 diakinesis cells to confirm the absence of pachytene pairing and the presence of only univalents, which explained the absence of crossovers in our transformants (**Fig. 1**). Later meiotic stages also showed con-

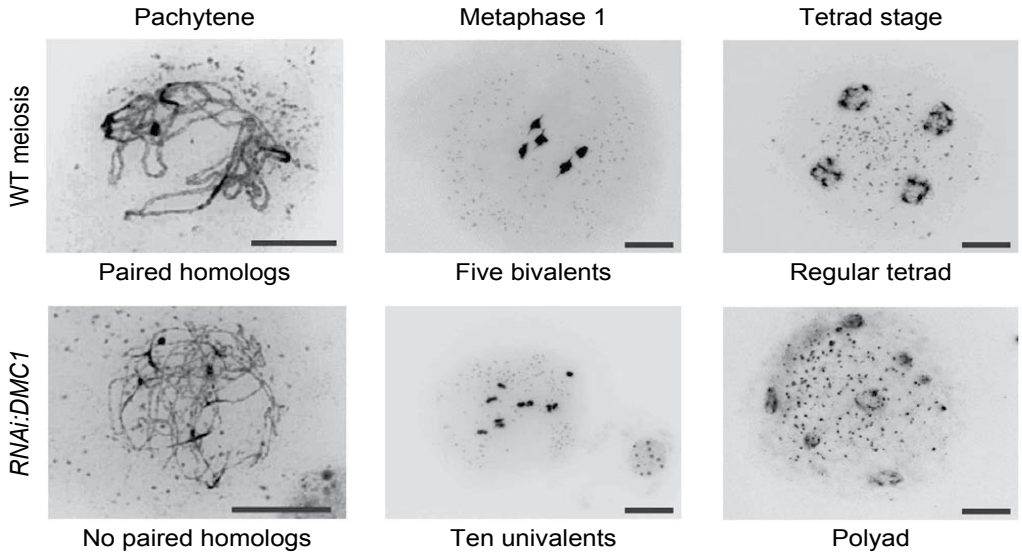


Figure 1. Meiosis in wild-type (WT; above) and *RNAi:DMC1* transformants (below). In wild-type meiosis, chromosomes pair at pachytene stage after which five bivalents are formed in metaphase 1. This results in tetrads showing four regular nuclei. In *RNAi:DMC1* transformants, tetrads are generally unbalanced, showing polyads, owing to unbalanced univalent segregation at metaphase 1. Suppression of DMC1 also disrupts pairing of chromosomes at pachytene. Scale bars, 10 µm.

figurations that support the random segregation of chromosomes at meiosis I and meiosis II. We observed unbalanced tetrads and polyads predominantly at the end of meiosis. Analysis of 85 dyad-stage meiocytes using FISH showed occasional meiotic irregularities such as lagging chromosomes or chromatid segregation at meiosis I (**Supplementary Fig. 5**), and only two instances in which there was at least one balanced meiosis I product. This result seems close to the expected 3.25%. In addition, the low numbers of viable seeds suggest low-frequency production of balanced gametes, although exact numbers vary between independent T1 plants, possibly because of variation in the expression of the RNAi transgene (**Supplementary Table 1**).

Conversion of balanced gametes into adult plants

To examine the effect of crossover silencing in gametes for reverse-breeding purposes, we constructed two near-isogenic *A. thaliana* hybrids, differing only in the presence of the *RNAi:DMC*1 transgene. We crossed accessions Columbia (Col-0) and Landsberg *erecta* (L*er*-0) to create a wild-type F1 (WT F1). We also crossed Col-0 carrying an *RNAi:DMC1* transgene with L*er*, creating a reverse-breeding F1 (RB F1).

The second step in reverse breeding is to convert haploid gametes, carrying non-recombined chromosomes, into homozygous diploid adults. This can be achieved through different methods depending on the plant species[7]. In *A. thaliana*, haploid plants can be

produced by centromere-mediated genome elimination[8]. We crossed both WT F1 and RB F1 pollen to the predominantly male sterile haploid inducer, creating wild-type and

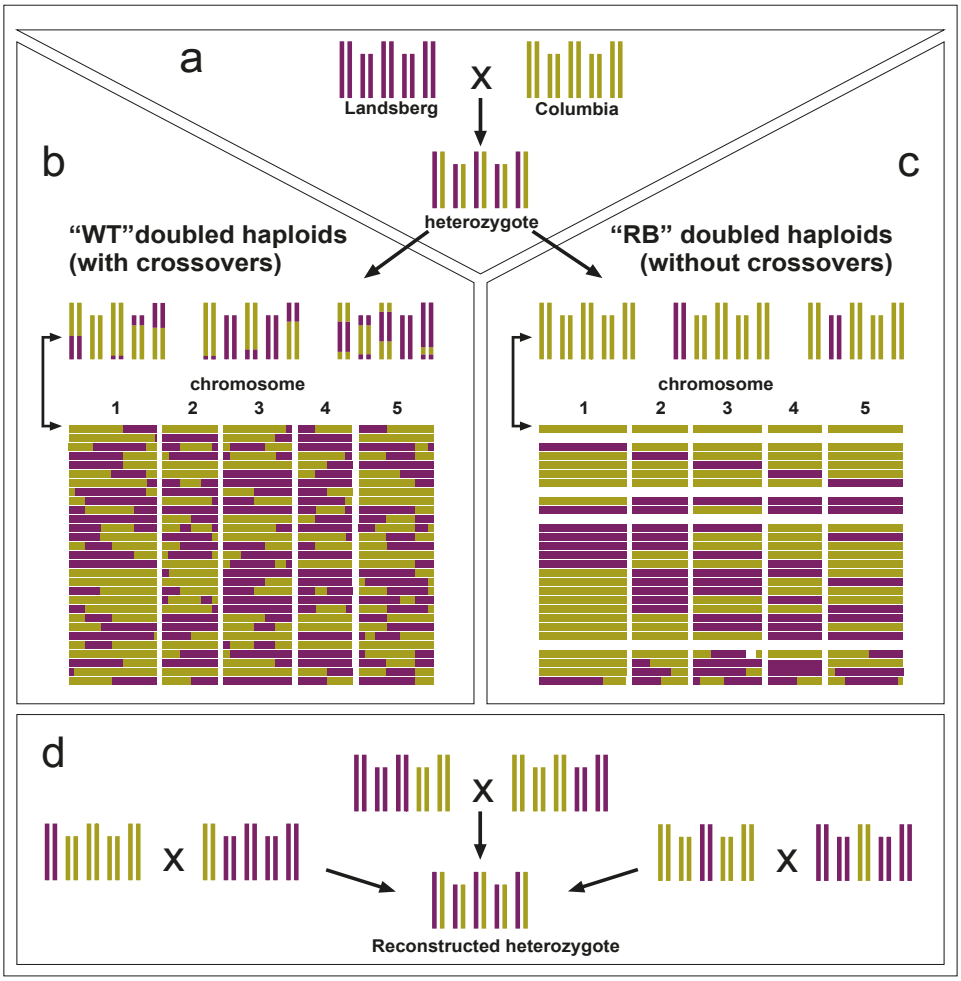


Figure 2. Reverse-breeding strategy and genotypes of wild-type (WT) and reverse-breeding (RB) doubled-haploid offspring. (a) Reverse breeding starts with a heterozygote in which meiotic recombination can be suppressed. (b) Genotype of 29 randomly selected wild-type doubled haploids. Three individuals are shown with 'classic' vertical chromosomes, but others as horizontal lines only. Each line represents chromosomes 1–5 for an individual plant. Note the presence of non-recombinant chromosomes. (c) We recovered 21 different genotypes in which no crossovers occurred from among 36 reverse-breeding doubled haploids. The first row represents the genotype of one of the recovered original parents; the next seven genotypes represent chromosome-substitution lines and the remainder are mosaics of Col and L*er* chromosomes. The last four represent genotypes of haploid offspring that showed crossovers (**Supplementary Note**). (d) Three pairs of reverse-breeding doubled haploids were crossed to recreate the initial hybrid; they have the RNAi transgene.

reverse-breeding haploids. We converted these haploids into fertile diploids (doubled haploids) by collecting the rare seeds following self-pollination of the haploids (see Online Methods).

We genotyped wild-type and reverse-breeding haploids and corresponding doubled haploids using evenly spaced SNP markers at approximately 4-Mb intervals (**Supplementary Fig. 6 and Supplementary Table 2**) to identify recombination events in parental meiosis (**Fig. 2**). For our calculations of recombination frequencies, we used the larger data set of recovered haploids. The vast majority of chromosomes in wild-type haploids were recombinant, with an average of 1.1 crossovers detected per chromosome in any given doubled haploid (**Supplementary Table 3**). A haploid typically shows half the number of the crossovers that occurred during meiosis, and it will occasionally receive non-recombined chromatids from a bivalent. These non-recombinant chromosomes were usually restricted to a single chromosome, and no single plant was non-recombinant for all five chromosomes. As such, crossover recombination in our doubled-haploid population was comparable to that in previously published data[9–12] in terms of genetic map length (crossover frequencies) as well as slight segregation distortions (see Online Methods).

Genetic analysis of 65 reverse-breeding haploids and their derived doubled haploids (that is, those plants resulting from achiasmatic meiosis) showed the complete absence of recombination, consistent with our cytological analyses (**Fig. 1**). Intact parental chromosomes segregated independently, leaving random assortment as the only process creating genetic variation among doubled haploids (Fig. 2c). Four additional reverse-breeding haploids did show crossovers (Fig. 2c), which might have resulted from accidental cross pollination or from incomplete silencing of DMC1 by the RNAi construct (**Supplementary Note**).

The main objective of reverse breeding is to generate homozygous parental lines that can be mated to recreate a desired heterozygous genotype. In our set of non-recombinant reverse-breeding doubled haploids we identified 21 of the 32 (25) possible genotypes, including the original Col-0 parent (**Fig. 2c**). Notably, we could identify six sets of complementing parents—that is, genotypes that would reconstitute the initial hybrid when crossed. These complementing pairs are genetically distinct, and also differ from the original Col-0 and L*er* parents. To complete reverse breeding, we made crosses between three pairs of selected reverse-breeding doubled haploid progeny to reconstitute the starting heterozygous parent (Fig. 2d). These crosses gave rise to perfectly heterozygous plants that were genetically identical to the achiasmatic Col/L*er* hybrid parent (**Supplementary Table 2**).

Reverse breeding creates chromosome-substitution lines

Segregation of intact parental chromosomes also creates chromosome-substitution lines, in which a single chromosome is substituted by the corresponding homolog from a

different line. Chromosome-substitution lines are valuable tools in many breeding applications, such as trait mapping, the study of epistatic interactions and targeted inbreeding[13]. Chromosome-substitution lines in *A. thaliana* have been generated through traditional crossing[14]; however, that method requires more generations of crossing and extensive genotyping to identify a particular chromosome-substitution line. With reverse breeding, one can obtain all possible chromosome-substitution lines in a short period of time. In two generations, we obtained a complete set of L*er* chromosome substitutions in the Col-0 background, as well as two substitutions of a Col-0 chromosome into a L*er* background from our population of only 36 doubled-haploid plants (**Fig. 2c**).

DISCUSSION

Reverse breeding allows any desired heterozygote to be selected from a large population and be propagated indefinitely as F1. This is crucial, because it alleviates one of the limitations of traditional breeding in which hybrids are generated by controlled crossing using few founder lines. The genomes of uncharacterized heterozygous plants can now be fixed in complementing immortal lines without knowledge of their provenance. This ability to fix heterozygous genomes resembles apomixis, in which heterozygotes reproduce clonally through seeds[2], but is fundamentally different. If apomixis were to be engineered for crops[15], the clonally propagated line would represent a dead end, as it could not be improved further by conventional hybrid breeding techniques. Because reverse breeding generates homozygous parental lines for the selected heterozygote, it allows the improvement of the heterozygote through improvement of the individual parental lines by traditional breeding methods such as backcrossing, mutagenesis and so on.

The technical feasibility of reverse breeding in *A. thaliana* suggests that it might be possible to apply this technology in crop improvement. Crucially, genes governing meiotic recombination are widely conserved[16], and haploid generation methods are available for many crops[7]. Although the probability of recovering balanced gametes decreases with increasing chromosome numbers, many agronomically important crops have 12 chromosomes or fewer, and are within the reach of successful reverse breeding[3]. Notable examples include cucumber, onion, broccoli, cauliflower, sugar beet, maize, pea, sorghum, (water-) melon, tomato, pepper, rice, eggplant and so on. Reverse breeding might be difficult to adopt in crops with higher chromosome numbers (for example, soybean) as well as in polyploids such as canola, cotton, wheat and potato.

The use of a dominant RNAi transgene means that 50% of reverse-breeding offspring carry the transgene, rendering them semi-sterile. The use of inducible RNAi constructs could circumvent this caveat. Alternatively, one could select two independent transformants of the same heterozygote, such that the RNAi constructs are inserted on different chromosomes. Pairs of transgene-free parents can then be obtained. In our case, reverse-breeding doubled-haploid lines 17 and 61 do not contain the *RNAi:DMC1* transgene

and would generate a transgene-free reconstructed heterozygote when crossed (**Supplementary Table 2**).

The expected low production of balanced gametes in crops with higher chromosome numbers is a possible bottleneck in reverse breeding. However, successful reverse breeding does not require the complete knockdown of crossovers, and the occurrence of a few crossovers might be beneficial, as non-recombinant chromatids are still present in bivalent pairs with a single crossover[3]. In rice (n = 12), for example, allowing three crossovers would already increase the chance of finding a balanced gamete from 2–12 = 1/4,096 to 2–9 = 1/520, at the expense of a lower recovery rate for true non-recombinant reverse-breeding offspring (1 out of 8 [2–3]). Our observation that different transformants showed varying degrees of sterility suggests that incomplete knockdown by RNAi could yield a desired level of crossover suppression. Segregation distortion in our population of haploid offspring (see Online Methods) was not extreme and was similar to that in previously described RIL populations[11]. All chromosomes were still frequently transmitted to offspring.

Our ability to exert control over the complex outcomes of meiosis, together with the ever increasing need for new methods for crop improvement[17], advocates for the rapid development of reverse breeding in crops. We therefore envision that new possibilities for the selection and improvement of favorable genotypes by reverse breeding may contribute to increasing future crop production.

URLs. *Arabidopsis* information resource, http://www.arabidopsis.org/; Joinmap, http://www.kyazma.nl/; KASPar SNP genotyping system, http://www.kbioscience.co.uk/reagents/KASP.html; MSQT, http://msqt.weigelworld.org/; pKANNIBAL vector, http://www.pi.csiro.au/RNAi/vectors.htm

Methods

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information (Suppl table 2) is available on the Nature Genetics website.

Acknowledgments

We thank T. Stoker, G. Stunnenberg, C. Pillen and H. Blankestijn for help in the greenhouse, J. van Ooijen for advice in using Joinmap, and L. Mlynarova, Y.-F. Lin and D.Z. Agudelo for their help with real-time PCR. We thank M. Koornneef, B. Zwaan and P.J. Flood for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

K.v.D., C.L.C.L., H.d.J. and R.D. conceived the research. E.W. performed cytology and crosses, C.B.d.S. constructed vectors and performed genotyping and N.S.N. performed cytology and FISH. E.W. analyzed data with the help of C.B.d.S., J.J.B.K., M.R. and S.W.L.C. E.W. wrote the manuscript and made the figures with substantial contributions by J.J.B.K., M.R., S.W.L.C. and H.d.J. All authors except N.S.N. were involved in planning and design of experiments, and read and improved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

References

1. Chen, Z.J. Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci. 15, 57–71 (2010).
2. van Dijk, P. & van Damme, J. Apomixis technology and the paradox of sex. Trends Plant Sci. 5, 81–84 (2000).
3. Dirks, R. *et al*. Reverse breeding: A novel breeding approach based on engineered meiosis. Plant Biotechnol. J. 7, 837–845 (2009).
4. Ma, H. A molecular portrait of *Arabidopsis* meiosis. *Arabidopsis* Book 4, e0095 (2006).
5. Couteau, F. *et al.* Random chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc*1 mutant of *Arabidopsis*. Plant Cell 11, 1623–1634 (1999).
6. Hartung, F. *et al.* The catalytically active tyrosine residues of both SPO11–1 and SPO11–2 are required for meiotic double-strand break induction in Arabidopsis. Plant Cell 19, 3090–3099 (2007).
7. Forster, B.P., Heberle-Bors, E., Kasha, K.J. & Touraev, A. The resurgence of haploids in higher plants. Trends Plant Sci. 12, 368–375 (2007).
8. Ravi, M. & Chan, S.W.L. Haploid plants produced by centromere-mediated genome elimination. Nature 464, 615–618 (2010).
9. Drouaud, J. *et al.* Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. PLoS Genet. 3, e106 (2007).
10. Toyota, M., Matsuda, K., Kakutani, T., Terao Morita, M. & Tasaka, M. Developmental changes in crossover frequency in *Arabidopsis*. Plant J. 65, 589–599 (2011).
11. Alonso-Blanco, C. *et al.* Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. Plant J. 14, 259–271 (1998).
12. Hauge, B.M. *et al.* An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. Plant J. 3, 745–754 (1993).
13. Nadeau, J.H., Singer, J.B., Matin, A. & Lander, E.S. Analysing complex traits with chromosome substitution strains. Nat. Genet. 24, 221–225 (2000).

14. Koumproglou, R. *et al.* STAIRS: A new genetic resource for functional genomic studies of *Arabidopsis*. Plant J. 31, 355–364 (2002).
15. Marimuthu, M.P.A. *et al.* Synthetic clonal reproduction through seeds. Science 331, 876 (2011).
16. Mercier, R. & Grelon, M. Meiosis in plants: Ten years of gene discovery. Cytogenet. Genome Res. 120, 281–290 (2008).
17. Tester, M. & Langridge, P. Breeding technologies to increase crop production in a world. Science 327, 818–822 (2010).

ONLINE METHODS

Plant material. *A. thaliana* plants were grown under standard conditions in a greenhouse. WT F1 was obtained by L*er*-0 (CS20) (female) × Col-0 (ABRC stock CS60000) (male). RB F1 plants were made by crossing male Ler-0 to two semi-sterile *RNAi:DMC*1 transformants: T39 and T62 (in Col-0). For a third RB F1 we used a male sterile (*ms1* $^{-/-}$) L*er* (CS261) (for convenient crossing) pollinated with T62. WT F1 and RB F1 plants were crossed as the male to *cenh3-1 GFP-tailswap* females to generate haploids[8]. By using GFP-tailswap as a female, we aimed for the recovery of reverse-breeding gametes through microspores, as these greatly outnumber egg cells (and therefore increases the probability of recovering balanced gametes).

Our crosses of both WT F1 and RB F1 to GFP-tailswap females yielded populations of diploid, haploid and aneuploid offspring, the last group resulting from incomplete genome elimination. Haploids were identified by their homozygous genotype and vegetative phenotype (semi-sterile flowers, smaller rosette size and narrow leaves compared to diploids)[8]. In the offspring of the RB F1 plants we also recovered semi-sterile diploids as a result of the RNAi construct. Aneuploids were discarded based on their aberrant growth phenotypes. The WT F1 to *GFP-tailswap* crosses yielded 73% of haploids, whereas the RB F1 to *GFP-tailswap* cross resulted in 42% of haploids. Recovery rates for aneuploids were low (~3%) in both populations.

We obtained no seeds for 56% and 55% of doubled haploids from the wild-type and reverse-breeding haploid populations, respectively. This was presumably in part due to growth conditions (on soil). We have found later that, when haploids are grown to larger sizes (on rock wool), haploid plants produce more seeds (data not shown).

Plant transformation. A 293-bp sequence of the *B. carinata* cDNA (**Supplementary Fig. 1**) was PCR amplified to clone the *DMC1* coding sequence, of which both sense and antisense orientation were cloned into a pKANNIBAL hairpin RNAi vector (CSIRO). The vector was subsequently cloned in an pART27 binary vector[18] and transformed into Col-0 using floral dip[19].

Quantitative RT-PCR. The nucleic acid was extracted from 10 mg of unopened flower buds pooled from several inflorescences. Total RNA was isolated using the RNeasy Mini Kit (Qiagen), and cDNA was synthesized using the iScript cDNA Synthesis Kit (Bio-Rad). Real-Time PCR was done using iQ SYBR Green Supermix (Bio-Rad) on the CFX96 Real-Time PCR Detection System (Bio-Rad). The ΔCt was calculated using SAND and UBC as endogenous controls[20]. We calculated relative quantification values (RQ) by the $2^{-\Delta Ct}$ method ($RQ = 2^{-\Delta Ct}$); results represent the average of five biological replicates. Oligonucleotides used for real-time PCR are given in Supplementary Table 4). For more details about the RNAi construct and the coordinates of AtDMC1 oligos used for real-time PCR, please refer Supplementary Figure 1.

Microscopy and fluorescence *in situ* hybridization. Pollen preparations were stained with lactophenol acid fuchsin. Meiotic preparations and FISH followed standard protocols[21–23]. The probes used and their position on *A. thaliana* chromosomes are shown in Supplementary Figure 5. Slides were examined under a Zeiss Axioplan 2 imaging photomicroscope equipped with epifluorescence illumination and filters.

Genetic analysis. SNPs were selected using MSQT[24] and matched to the Columbia reference genome using the *Arabidopsis* information resource. Genotyping was done using the KASPar SNP genotyping system. Joinmap version 4.1 (ref. 25) was used for calculation of recombination frequencies, and $\chi 2$ tests were used for segregation distortions. Recombination frequencies were estimated using the regression mapping algorithm with 'independence LOD' as grouping parameter and the Kosambi mapping function to convert recombination frequencies to map distances.

Marker segregation in wild-type and reverse-breeding haploids. During marker analysis, we noticed clear segregation distortions in both wild-type and reverse-breeding haploids. Among wild-type haploids, Col alleles are overrepresented at the lower end of chromosome 1, the top and bottom of chromosome 2, the middle of chromosome 4 and at the top of chromosome 5 (**Supplementary Fig. 7**). Segregation distortions have previously been studied in a Col/L*er* recombinant inbred line (RIL) population[26,27]. The overrepresentation of Col alleles at the lower end of chromosome 1 in that population is concurrent with our observations, possibly having similar causes. Notably, the Col alleles on chromosomes 2 and 5 were overrepresented in our haploid population, in contrast to the overrepresentation of the L*er* alleles in the RIL population. A previous report described a genetic incompatibility leading to segregation distortions at the lower end of chromosome 1 and the top of chromosome 5 in *A. thaliana*[28]. Nonetheless, this is not expected to be causing the segregation distortions found in these populations, as Col and L*er* share the same alleles for these loci.

Because, in reverse breeding, linkage is absolute for all loci on a chromosome, any distorting locus would affect transmittance of the whole chromosome. The overrepresentation of Col chromosomes 1, 4 and 5 in the reverse-breeding haploids could be the direct effect of such distorting loci. For these chromosomes it seems plausible that this is caused by the same loci causing the preferential transmittance of Col alleles to the wild-type haploids. Notably, the overrepresentation of Col alleles at the top of chromosome 2 is not matched by an overrepresentation of that chromosome in reverse-breeding haploids. This might be caused by a 'balancing' mechanism, as at the lower end of chromosome 2 there seems to be a locus of which the Col allele is transmitted less frequently than its L*er* counterpart.

We hypothesize that the increased genetic linkage imposed upon the alleles on one chromosome in reverse breeding may strongly affect the transmittance of alleles. When linkage becomes absolute, distorting alleles may act either directionally or in a balancing manner, depending on the sign of their effect.

The differences in segregation distortions between the Col/L*er* RILs and the haploids may also lie in the methods used for the construction of these populations. Our haploids were produced using only male meiosis, whereas, in the RILs, gametes produced by female meiosis also contribute to observed effects. Second, the haploids underwent a process of genome elimination, and it is possible that there are alleles that favor either their genesis or the survival of haploid embryo.

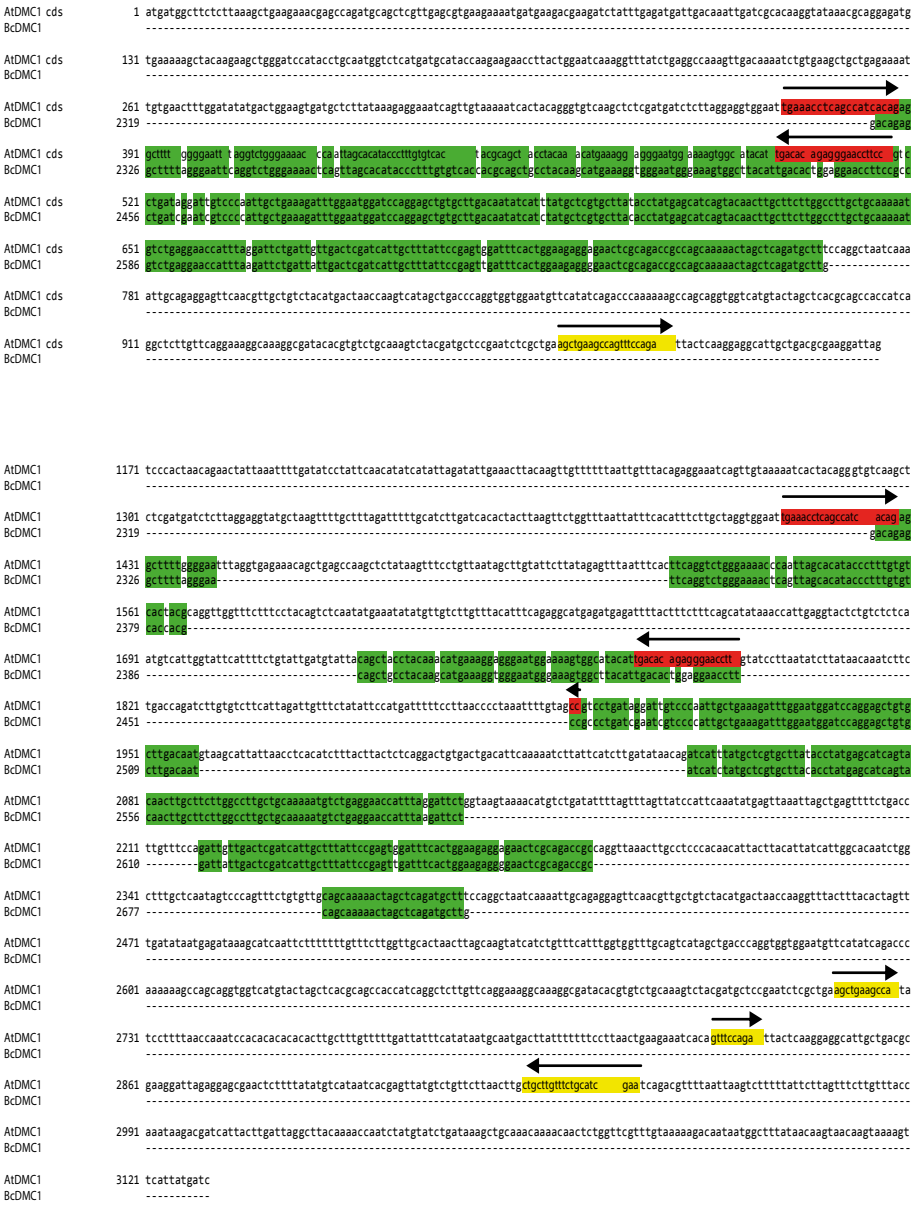
Additional references

18. Gleave, A.P. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Mol. Biol. 20, 1203–1207 (1992).
19. Clough, S.J. & Bent, A.F. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. 16, 735–743 (1998).
20. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. & Scheible, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol. 139, 5–17 (2005).
21. Ross, K.J., Fransz, P. & Jones, G.H. A light microscopic atlas of meiosis in *Arabidopsis thaliana*. Chromosome Res. 4, 507–516 (1996).
22. Ross, K.J. *et al.* Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA–transformed lines. Chromosome Res. 5, 551–559 (1997).
23. Tang, X. *et al.* Cross-species bacterial artificial chromosome–fluorescence *in situ* hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180, 1319–1328 (2008).
24. Warthmann, N., Fitz, J. & Weigel, D. MSQT for choosing SNP assays from multiple DNA alignments. Bioinformatics 23, 2784–2787 (2007).

25. Stam, P. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J. 3, 739–744 (1993).
26. Lister, C. & Dean, C. Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J. 4, 745–750 (1993).
27. Singer, T. *et al.* A high-resolution map of *Arabidopsis* recombinant inbred lines by whole-genome exon array hybridization. PLoS Genet. 2, e144 (2006).
28. Bikard, D. *et al.* Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science 323, 623–626 (2009).

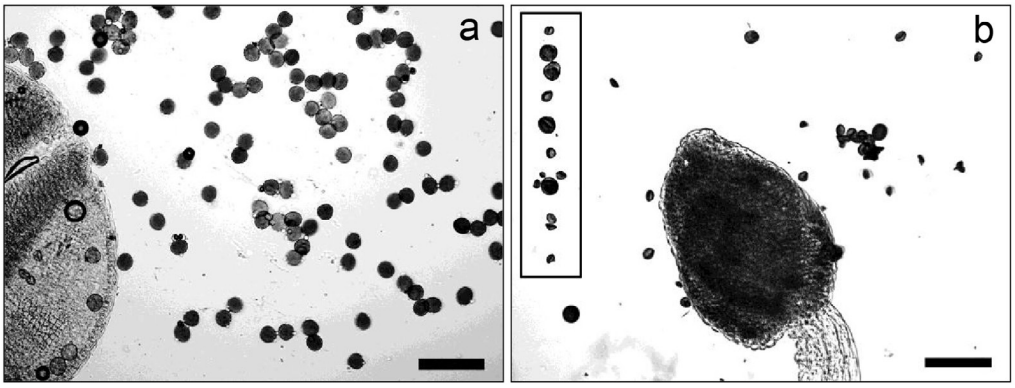
Supplementary Figures, Tables and Note

Contents

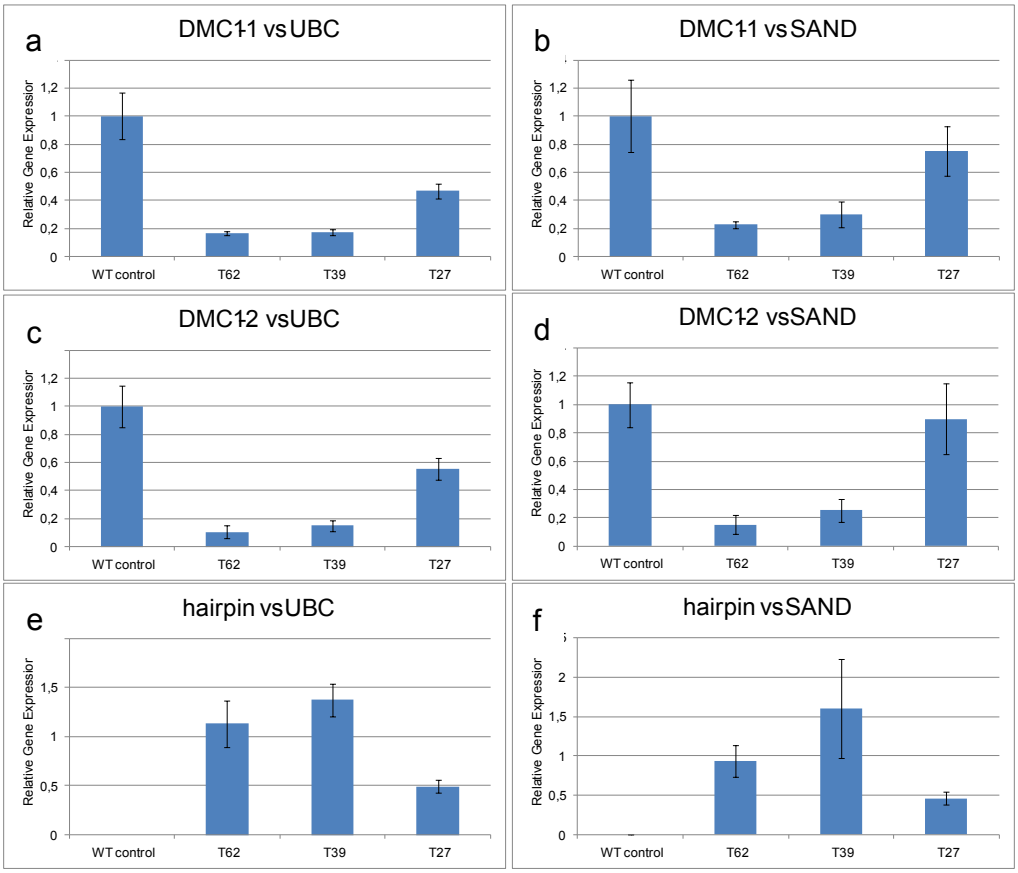


Supplementary Figure 1: Sequence alignment of *BcDMC*1 and *AtDMC*1. Alignment of *Brassica carinata DMC*1 (*BcDMC*1) with *DMC*1 of *Arabidopsis thaliana* (*AtDMC*1). The alignment was made against the coding sequence (above), as well as to the genomic sequence (below). Matching positions are in green colour. Positions of primer pairs used for real-time PCR are indicated in red (*DMC*1-1) and yellow (*DMC*1-2).

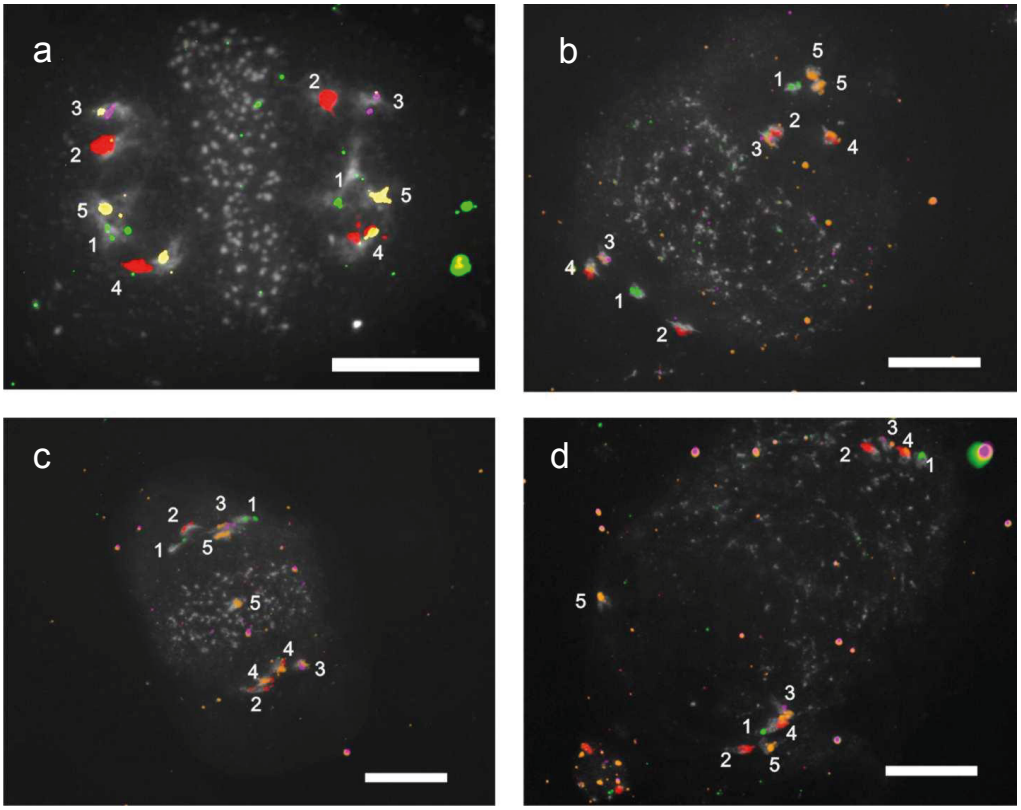
Supplementary Figure 2: Phenotype of wild-type (WT) and *BcDMC*1 RNAi transformed *Arabidopsis thaliana*. (a) Wild-type (Col-0) inflorescence with, well elongated (normal) siliques. *RNAi:DMC*1 transformant T27 (b) shows reduced fertility, as inferred from its reduced silique length. Two examples of the highly sterile *BcDMC*1 RNAi transformants T39 (c) and T62 (d). Note the very short siliques that harbor few or no seeds. T39 and T62 were used in our experiments. *DMC*1 expression for these four lines are shown in Supplementary Fig. 4.



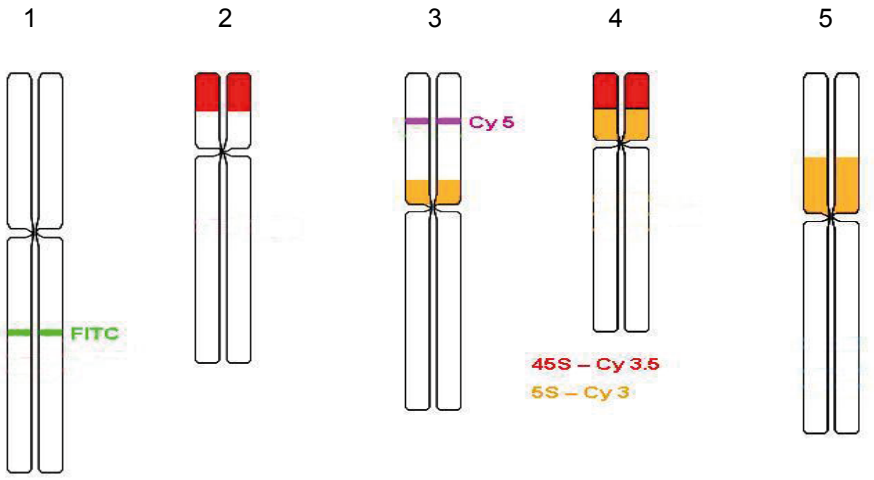
Supplementary Figure 3: Pollen morphology of WT F1 (Col x L*er*) and RB F1 (Col x L*er BcDMC*1 RNAi). The WT F1 (a) shows uniformly sized pollen. Silencing of crossovers leads to mostly small sized and irregular shaped pollen in RB F1 (b). The inset in b shows more examples of randomly photographed pollen grains. Scale bar = 100 µm.



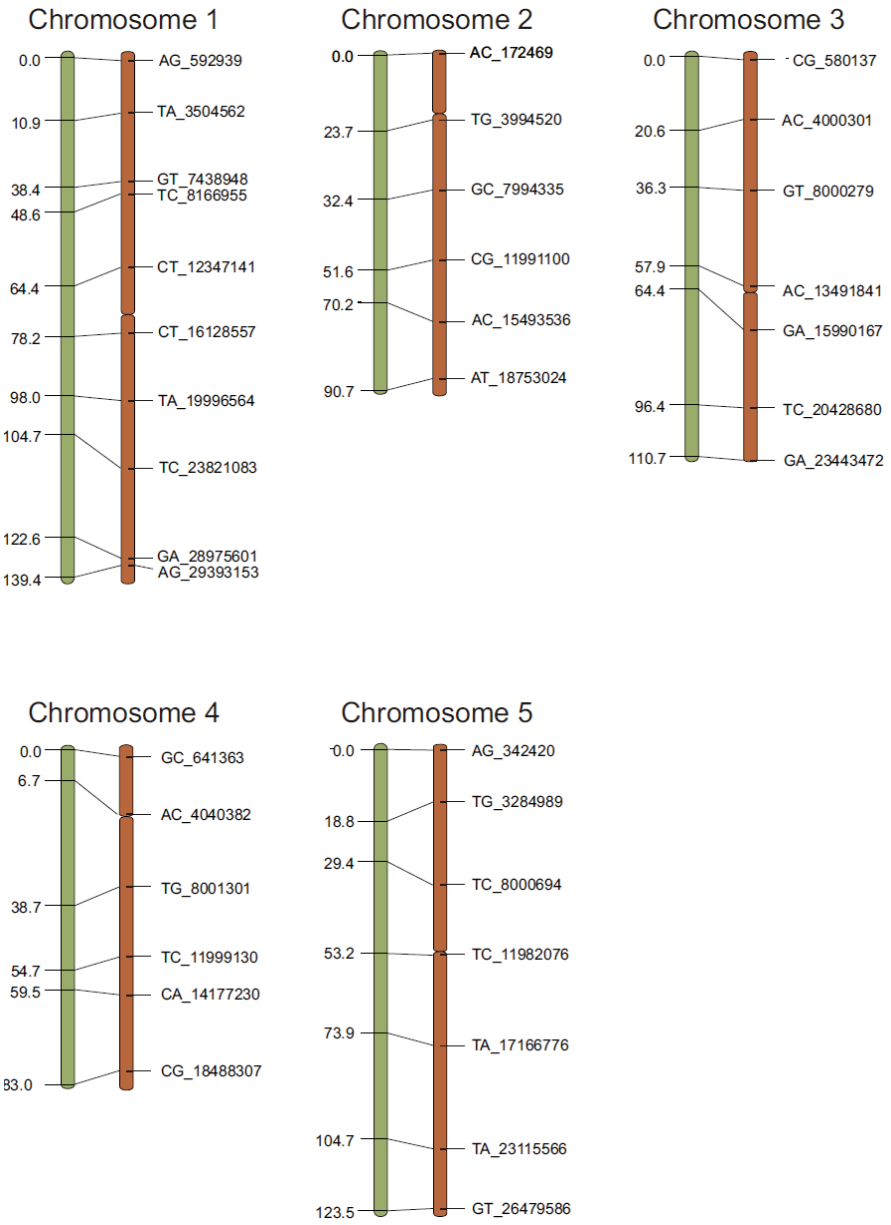
Supplementary Figure 4: *AtDMC*1 is down-regulated in *BcDMC*1 RNAi plants. Bar charts depicting the relative expression of *A. thaliana DMC*1 normalized to expression of housekeeping genes *UBC* and *SAND*. This is shown for two different primer pairs *DMC*1-1 (a, b) and *DMC*1-2 (c, d). *AtDMC*1 mRNA is down-regulated in the *BcDMC*1 RNAi lines in comparison to wildtype Col. T1 plants T39 and T62 were used for generating RB haploids. T27 is one of the primary transformants (T1) that shows intermediate fertility in comparison to WT and highly sterile transformants T39 and T62 (Supplementary Fig. 2). Charts (e) and (f) show transcription of the RNAi construct, where a relative gene expression of 1 refers to the average for the tree transformants. Error bars show the standard error for five biological replicates.



Supplementary Figure 5: Fluorescent *in situ* hybridization shows unbalanced (random) segregation of chromosomes in BcDMC1 RNAi transformants.
I) Fluorescent *in situ* hybridization of meiotic cells in WT Col (a) and RNAi transformed (b, c, d) plants. Each homolog is labelled with its respective chromosome number as identified by mulitcolor FISH, outlined in part II of this figure. Scale bar = 10 µm. Wildtype meiotic cell (a, interkinesis) showing balanced segregation of homologous chromosomes. Achiasmatic meiosis in *BcDMC*1 RNAi plants (b, c, d) showing unbalanced segregation of chromosomes during meiosis. (b) Metaphase II cell showing non disjunction of chromsome 5. The cell on the lower left lacks chromosome 5. (c) Metaphase II cell showing a lagging chromosome 5 in the organelle band. Note that the upper cell - although containing five chromosomes- is unbalanced, for it has two copies of chromosome 1 and lacks chromosome 4. (d) A telophase I cell showing that achiasmatic meiosis can lead to the formation of balanced gametes (the daughter cell in the middle, below). Also here, a single chromosome 5 shows missegregation, and might form a micronucleus (left).

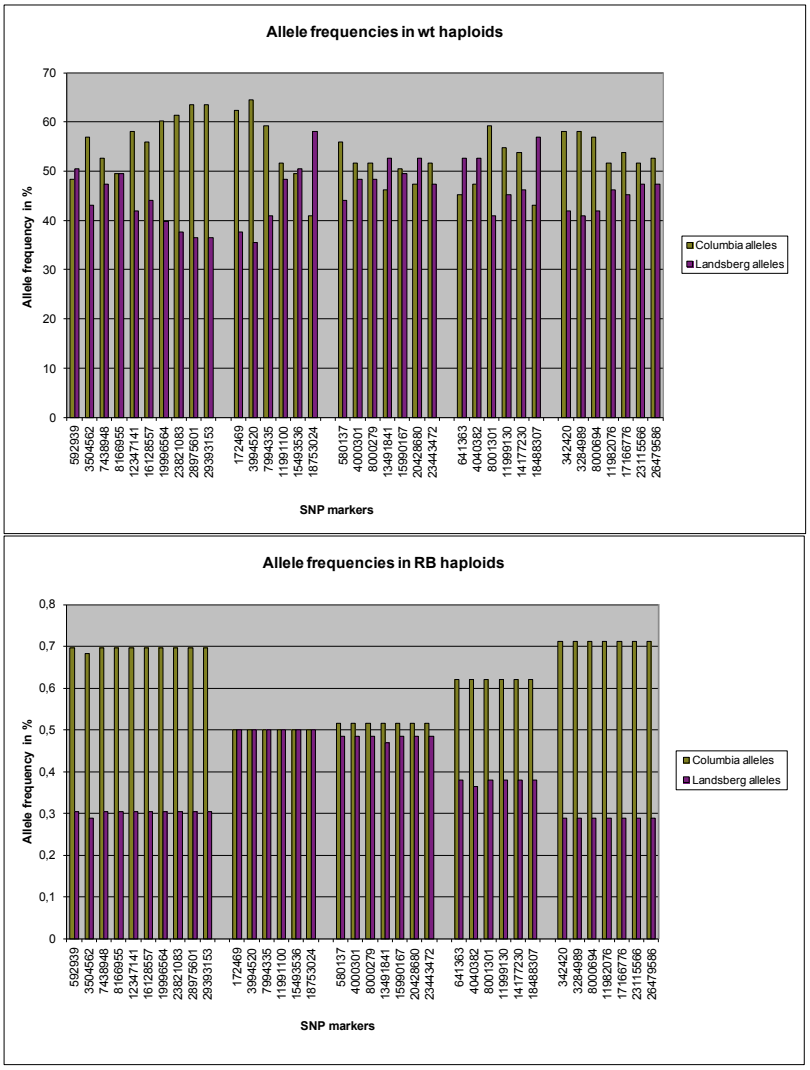


Supplementary Figure 5 (continued): Fluorescent *in situ* hybridization shows unbalanced (random) segregation of chromosomes in BcDMC1 RNAi transformants.
II) Chromosome painting scheme and approximate location of the probes used for FISH.
Chromosome 1 painted with BAC F13N6,
Chromosome 2 painted with plasmid pTa71 (45S rDNA),
Chromosome 3 painted with BAC MWL2 and plasmid pCT4.2 (5S rDNA),
Chromosome 4 painted with plasmid pTa71 and pCT4.2,
Chromosome 5 painted with pCT4.2.



Supplementary Figure 6: Map positions of SNP markers used for genotyping haploids/doubled haploids.

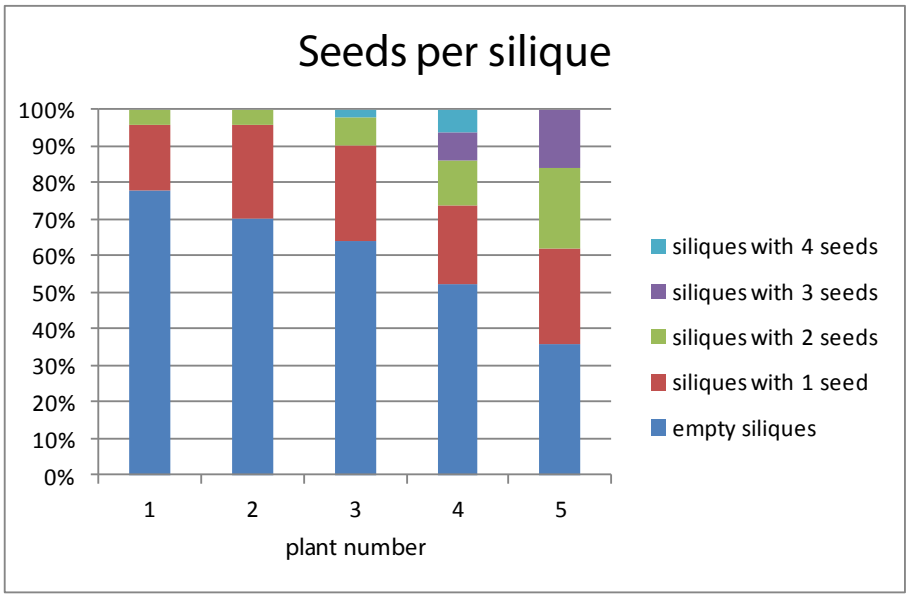
Genetic maps of all five chromosomes are presented in green (on the left) with genetic distances in centiMorgan (cM) besides it (on the left). Physical map (red) on the right, with locus names consisting of two letters (indicating nucleotide type in Col and L*er* respectively) followed by the co-ordinates of the respective SNP location on the Col-0 reference physical map (TAIR; http://www.arabidopsis.org/).



Supplementary Figure 7: Allele frequencies in WT and RB doubled haploids.

Allele frequencies for all SNP markers are shown for WT haploids (above) and RB haploids (below). Markers are ordered from top chromosome 1 (left) to end of chromosome 5 (right). There are loci that are preferentially transmitted to offspring. This is also reflected in the segregation pattern of the RB offspring where Col chromosomes 1 and 5 are significantly ($p \leq 0.05$) overrepresented among the haploid offspring (Bonferroni corrected $\chi 2$).

	plant number				
	1	2	3	4	5
empty siliques	39	35	32	26	18
siliques with 1 seed	9	13	13	11	13
siliques with 2 seeds	2	2	4	6	11
siliques with 3 seeds	0	0	0	4	8
siliques with 4 seeds	0	0	1	3	0
total	50	50	50	50	50
number of aborted seeds	9	12	18	31	46
number of WT seeds	4	5	7	16	13
Percentage of WT seeds	31	29	28	34	22
seeds/silique	0,26	0,34	0,5	0,94	1,18



Supplementary Table 1: Variation in fertility among sibs from a *BcDMC*1 RNAi T1 plant T39.

To examine the stability of the phenotype of our transformants, five offspring plants were grown from a primary transformant T39. From each plant 50 random siliques were examined under a dissecting microscope. The above data indicate that sibs show variation in seedset, which we assume is a result of differences in expression of the *BcDMC*1 RNAi transgene.

		chr1	chr2	chr3	chr4	chr5
WT haploids (n=93)	Detected crossovers per chromosome (SE)	1.31 (0.09)	0.88 (0.08)	1.04 (0.09)	0.81 (0.08)	1.24 (0.10)
	Fraction recombinant chromosomes	0.85	0.63	0.69	0.58	0.73
	Fraction non-recombinant chromosomes	0.15	0.37	0.31	0.42	0.27
	Genetic map length (cM)	139.4	90.7	110.7	83.0	123.5

Supplementary Table 3: Recombination data in WT meiosis.
The table shows recombination data as observed in WT haploids. "Detected crossovers per chromosome" refers to genotype changes between two adjacent markers that were detected on chromosomes in the WT haploids. The standard error is given in parentheses. The fraction of recombinant and non-recombinant chromosomes are also based on observed genotype changes in WT haploids. Genetic map lengths were calculated using Joinmap® 4.

	Forward primer 5' - 3'	Reverse primer 5' - 3'
UBC	CTGCGACTCAGGGAATCTTCTAA	TTGTGCCATTGAATTGAACCC
SAND	AACTCTATGCAGCATTTGATCCACT	TGATTGCATATCTTTATCGCCATC
DMC1-1	TGAAACCTCAGCCATCACAG	GGAAGGTTCCCTCTGTGTCA
DMC1-2	AGCTGAAGCCAGTTTCCAGA	TTCGATGCAGAAACAAGCAG
hairpin	CGCCTATGATCGCATGATATT	GGCGGTAAGGATCTGAGCTA

Supplementary Table 4: Primer combinations used in Real-time PCR.
Table showing forward and reverse primers for oligos used for Real-time PCR.

		Genotype haploid offspring															
		chromosome 1										chromosome 2					
plant number	Genotype F1 parent	592939	3504562	7438948	8166955	12347141	16128557	19996564	23821083	28975601	29393153	172469	3994520	7994335	11991100	15493536	18753024
1	RNAi:DMC1 Col-0 T62 x L*er*	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2	ms1 -/- L*er* x RNAi:DMC1 Col-0 T62	A	A	A	A	A	A	A	A	A	A	B	B	A	A	A	A
3	RNAi:DMC1 Col-0 T39 x L*er*	A	A	A	A	A	A	A	A	A	A	B	B	B	B	A	A
4	RNAi:DMC1 Col-0 T39 x L*er*	B	B	B	B	B	B	B	A	A	A	B	B	B	A	A	A

		Genotype haploid offspring																				
		chromosome 3							chromosome 4						chromosome5							
plant number	Genotype F1 parent	580137	4000301	8000279	13491841	15990167	20428680	23443472	641363	4040382	8001301	11999130	14177230	18488307	342420	3284989	8000694	11982076	17166776	23115566	26479586	PCR result for RNAi
1	RNAi:DMC1 Col-0 T62 x L*er*	A	A	B	B	B	x	A	A	A	A	A	A	A	A	A	A	A	B	B	B	-
2	ms1 -/- L*er* x RNAi:DMC1 Col-0 T62	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	-
3	RNAi:DMC1 Col-0 T39 x L*er*	B	B	B	B	B	A	A	B	B	B	B	B	B	A	B	B	B	B	B	B	-
4	RNAi:DMC1 Col-0 T39 x L*er*	B	A	A	B	B	B	B	B	B	B	A	A	A	A	A	B	B	B	B	A	-

Supplementary Note: The genotypes of four recombinant RB haploids.

B = Landsberg *erecta* allele, A = Columbia allele

"+" = RNAi construct present, "-" = RNAi construct absent

"x" = missing data

Four haploids were recovered from RB F1 pollen that reflect the presence of recombinant gametes, as shown in the table above. Because WT F1, RB F1 and the haploid inducer plants were grown together, we cannot completely exclude the possibility of accidental cross-pollination. For instance, although the insertion of the RNAi transgene (T62) in RB F1 was mapped to chromosome four, a PCR test for the presence of the construct was negative for plant number 1, carrying a Col genotype at chromosome four, indicating that this plant is indeed the result of WT F1 pollination. In addition, the low number of crossovers in plant number two and three suggests that crossover silencing can be incomplete even in the presence of the construct.

On a finished thesis…

Ik lig hier het bevlogene van zwaluwen te bestuderen.
Soms wordt er een hele zwerm overgesmeten, ik probeer te begrijpen
wie dit doet, de wind niet, maar ook zij niet, ze hebben geen wij,
alleen gewirwar. Ze proberen uit hun staart weg te vliegen,

knip-knip door het heelal, hier waren we al,
hier waren we al. Ik lig te snorkelen
aan mijn luchtpijp. Ik zie dat het goed is.
Ik wil er mijn handtekening wel onder zetten.

H. de Coninck

(In: 'Nu, dus', 1995)

CHAPTER 8

General discussion

Meiosis and the choice of Arabidopsis for breeding research

Research of meiotic recombination in higher plants is a diverse, highly dynamic and rapidly developing field, supported by unsurpassed technological and conceptual advances. It is especially the concerted efforts of cutting-edge science in plant breeding, (cyto-) genetics, bioinformatics and molecular biology that together generate powerful strategies to address hitherto unanswered questions. The implementation of such multidisciplinary strategies not only depends on novel research tools, but also greatly benefits from the integration of knowledge through comparative genomics. In my thesis I describe a wide variety of techniques and approaches related to meiotic recombination and chromosome biology. These include highdepth sequencing of *Arabidopsis* doubled haploids and meiotic tetrads, the construction of doubled haploids in *Arabidopsis thaliana*, advanced cross-species BAC FISH and the construction of a series of hypomorphic alleles that allowed otherwise lethal mutations to be studied.

A proper foundation for the understanding of molecular processes in plant meiosis was laid down by the many reports on meiotic mutants from the last 15 years (reviewed in Mercier and Grelon (2008), most of them using the power and versatility of *Arabidopsis*. The available resources and databases for this model plant are unique in their size and completeness with its small genome, its short generation time and full genetic and physical maps has shown its enormous potential for the development of new breeding applications (d'Erfurth *et al*. 2009; Marimuthu *et al*. 2011; Olmedo-Monfil *et al.* 2010; Ravi and Chan 2010; Ravi *et al*. 2008; Wijnker *et al.* 2012). A review on the possible applications of meiosis research (Chapter 4 of this thesis) was published in 2008 and was one of the first reviews specifically addressing the question of how breeders can manage and exploit meiotic variation for breeding (Wijnker and de Jong 2008). Four years later, this review serves as an excellent starting point for this discussion.

This thesis describes a multitude of leads for further research, both in fundamental as well as applied (breeding related-) research. As the relations between the topics are not always that obvious, I will try to clarify these in the text below with special attention for breeding related research. In the first part, focus lies on recombination landscapes and how these can be studied. Thereafter the meiotic cell cycle and chromatin dynamics are addressed that appear linked through CDKA;1 action. The potential applications of reverse breeding for constructing new mapping populations and novel designs for breeding programs will be dealt with in the last section.

How to navigate recombination landscapes

Crossover recombination events are non-randomly distributed over the genome, which together give rise to a specific recombination landscape. To answer the intriguing question of how such a landscape is formed, these landscapes need to be assessed in the highest detail. The experimental and conceptual approaches range widely, but an (almost)

fully sequenced genome is near indispensable to anchor genetic markers to a physical map and to allow the construction of accurate recombination maps.

Tools for assessing recombination frequencies based on the segregation of fluorescent markers in seeds (Melamed-Bessudo *et al.* 2005) and that of pollen tetrads (Francis *et al*. 2007) were gradually adopted in the scientific community (Berchowitz *et al.* 2007; Crismani *et al.* 2012; Melamed-Bessudo and Levy 2012; Yelina *et al.* 2012). These suffer however from requiring specific mutants or transgenic backgrounds, like a quartet background to study marker segregation in pollen tetrads or fluorescent markers to assess marker segregation in seeds. An additional drawback is that these methods provide only information of specific chromosome intervals. Though being great tools for establishing recombination frequencies, these approaches do not provide a genome-wide view on recombination.

Biotechnological innovations in recent years have lead to the development of powerful and efficient strategies to collect recombination data in plants. SNP marker assays can now be designed in which tens- to hundreds of markers can be placed at specific genome positions (Giraut *et al*. 2011; Salome *et al*. 2012; Wijnker *et al.* 2012; Yelina *et al*. 2012). For the comparison of studies over time, the adoption of a shared marker set between different labs would be the best choice for this kind of studies. The SNP marker set that was developed and used throughout these and one other study (Yelina *et al.* 2012) could provide the basis for such a set.

The smart use of sequencing provides new ways to genotype populations in high throughput and at low coverage (Huang *et al.* 2009). Key to the efficient use of high-throughput sequencing lies in finding ways of pooling multiple samples (usually through barcoding) and enriching for specific sequences to be sequenced. This can be done by using specific restriction enzymes as done in RAD-mapping (Baird *et al.* 2008), possibly followed by further enrichment steps like selection for specific restriction fragment sizes in reduced representation shotgun sequencing (Altshuler *et al.* 2000; Seymour *et al.* 2012). Such methods currently allow the genotyping of 96 samples per Illumina lane, but technical improvement would preferably push this to hundreds of samples per lane.

The whole-genome-sequencing (WGS) approach in Chapter 2 was chosen for obtaining high resolution of few meiotic offspring rather than looking at recombination at a populationwide level. The sequencing of offspring from (preferably homozygous) individuals brings unsurpassed information of the extent of recombination between parental genomes through crossovers and non-crossovers (COs and NCOs) and provides insights into the topography of CO placement (Chapter 2 and Henson *et al*. (2012); Lu *et al.* (2012) and Qi *et al.* (2009)). The potential of using WGS in combination with a population-oriented approach was recently presented by Wang *et al.* (2012) in which DNA derived from single sperm cells was amplified and genotyped using microarrays or sequenced to construct a personal recombination map. Such a method could also be developed for plants in which trinuleate pollen grains conveniently contain three instead of one hap-

loid nucleus (and thus have more DNA). Such a method would allow assessing genome-wide recombination landscapes in specific parts of the plants such as single flowers, or can precisely establish the effects of environmental stress.

When a recombination landscape is experimentally cestablished, this does not bring full understanding of how this landscape was actually shaped. This is because some recombination events (NCOs) that result from meiotic double strand breaks (DSBs) do not always leave detectable footprints on the DNA (Chapter 2), hence the relation between DSB-formation and (N)CO formation is thus difficult to study. It was recently shown that DSB formation can also be studied through high-throughput sequencing and that DSB-maps can be generated (Pan *et al.* 2011). This approach was based on the fact that the meiotic DSB inducing protein SPO11 remains covalently bound to DNA after DSB formation. Immuno-precipitated SPO11- oligonucleotide complexes were isolated and the sequenced reads mapped onto a reference genome for obtaining a genome-wide topography of recombination initiation. The method demonstrated that a recombination landscape is the result of a hierarchy of interacting factors, regulated by the accessibility of chromatin up to higher order chromosome structure (Pan *et al.* 2011). Even more insightful would be to know how DSB landscapes change concomitantly with recombination landscapes in mutant backgrounds, like for example in *cdka*;1 (Chapter 3). In the following we will explore some leads for future work on this cyclin-dependent kinase.

Chromatin defines homo(eo)logy

To study the requirement of CDKA;1 in meiosis, we designed a unique allelic series of engineered *CDKA;*1 mutants with different kinase activities, identifying this protein as a key regulator of the *Arabidopsis* meiotic cell cycle, involved in chromosome condensation and crossover interference, thereby acting as a major determinant of shaping the recombination landscape (Chapter 3). Chromatin condensation, meiotic homo(eo)logous recombination and CO-interference are all meiotic processes that result from complex interactions at different hierarchical levels, and the recovery of CDKA;1 as a mediator of these processes makes it promising for fundamental research as well as breeding applications. Especially as it might help to improve the transfer of loci from one species into the other through meiotic recombination. Most crops, including tomato, peas and cucumber, have a narrow genetic base (Esquinas-Alcazar 2005; The_Tomato_Genome_Consortium 2012), and traits missing in breeding lines, like disease resistances, need to be transferred by introgression from wild relatives (Canady *et al.* 2006; Gill *et al.* 2011). Methods that facilitate such introgression are of the highest interest and importance. The major bottleneck in such strategies is the incorporation of alien DNA by CO recombination. The best known gene that can control homoeologous crossovers is Ph1 in wheat that allows introgressive hybridization between the three parental genomes (A, B and D) of this hexaploid crop, but also allows transfer of chromosome parts of rye and other cereals to wheat (Moore 1998; Moore and Shaw 2009).

Recent studies have suggested the molecular machinery of *Ph1* to act through *CDK*s. When *Ph1* is present, wheat has a constitutive low activity of Cdk2, the wheat CDKA;1 orthologue. Under this condition the homoeologous chromosomes of wheat condense differentially, leading to crossover formation exclusively between homologues. In the absence of *Ph*1, the increased *CDK* activity equalizes the otherwise differential chromatin condensation of homoeologues, and hence allowing them to form crossovers (Colas *et al.* 2008; Greer *et al.* 2012). Even though in *Arabidopsis* there are no homoeologues, we were able to demonstrate in Chapter 3 that weak *cdka*;1 alleles display a phenotype similar to wheat plants with multiple dosages of *Ph*1 (with presumed low *Cdk*2 activity) (Feldman 1966).

The *Ph1* locus is not the only locus know to be involved in homoeologous pairing; a locus named *PrBn* was identified in *Brassica napus* that mediates crossover formation between homoeologous chromosomes in (di)haploids of this allopolyploid species (Jenczewski *et al*. 2003). The molecular base of this mutation is unknown, but the approach taken to test whether certain alleles induce homoeologous pairing (by examining (di) haploids of an allopolyploid species) surely provides a very promising approach for further studies into the roles of CDKs.

A next step in engineering homoeologous recombination should be the increase CDK activity in interspecific hybrids (or the above mentioned dihaploids). This can be achieved in different ways. Treatment with the protein serine/threonine phosphatase inhibitor okadaic acid (Yamashita *et al*. 1990) has been shown to achieve such an effect in wheat (Knight *et al.* 2010), which holds promises for other species as well. Alternatively, homoeologous recombination may also be possible if the kinase activity of major cyclin dependent kinases is increased. Important to realize is that CDK activity is continuously kept under control by a multitude of regulators like cyclins and various inhibitors (Inzé and De Veylder 2006; Nowack *et al.* 2012). As such, simple CDKA;1 overexpression might not initially work. The best approach might lie in the upregulation of CDKA;1 while simultaneously downregulating CDK inhibitors. In order to develop such an approach and to identify crucial inhibitors, it would be extremely useful to use the proper cytogenetic tools for monitoring meiotic progression.

The introgression of traits may be impaired by a variety of factors. Chapter 3 serves as a nice illustration of the existence of chromosome inversions between related *Solanum* species, which by far pose the biggest barrier to introgressive hybridization. Unless inversions are large enough to allow the formation of pairing loops in inversion heterozygotes (and allow for rare 2-strand double crossovers to be formed), it is impossible to introgress traits from wild relatives into a desired background without the considerable linkage drag of the whole inversion. The observation that NCO tracts are very short (Chapter 2) suggests that NCOs are no viable alternative for double crossovers. For the moment, inversions pose big difficulties to introgression, for which transgenic or cisgenic approaches may be the only solution for transfer of traits from alien donors to crops.

But also in the absence of inversions, the introgression of traits from related species is far from easy. This has convincingly been demonstrated in a study of Canady and colleagues on recombination rates in tomatoes that were heterozygous for an alien introgression. The recombination rates in homoeologous segments decreases when phylogenetic distances between parental lines increase (Canady *et al.* 2006). Although a clear explanation for this phenomenon is lacking, considerable evidence exists that sequence divergence is not the only restriction. For example, our observations in Chapter 2 indicate that crossovers can very well form near highly diverged sequences and the observations in *Ph*1 and *PrBn* mutants clearly show that sequence divergence does not prohibit crossover formation. Rather, it seems more likely that higher order organization of DNA in chromatin domains is crucial in the pairing process. Highly instructive images of SC-spreads of interspecific crosses of tomato with wild relatives were published showing that homologous sequences do not necessarily pair in SCs (Anderson *et al*. 2010). In addition, excessive pairing and chiasma formation between *Lolium* x *Festuca* hybrids with a markedly differentiated repeat content of both parental genomes as shown by genome painting provide clear evidence that sequences are conserved enough to facilitate regular pairing partner recognition and crossing over (Kopecký *et al.* 2008).

At the moment, we still lack the means to modify the pairing behavior in crops. The current best approach for increasing the efficiency of introgressive hybridization (*i.e.* through reducing linkage drag) might therefore lie in simply increasing the recombination frequency. The recent discovery of FANCM as a suppressor of the class II CO pathway holds great promises (Crismani *et al.* 2012), as plants in a fancm background display increased levels of CO recombination. It will nevertheless be important that homologous sequences find one another. A combined approach of increasing the recombination frequency through dominant down regulation of FANCM, with an increase of CDK activity may increase desired recombination in homoeologous regions.

In our study on DSB repair in *Arabidopsis* (Chapter 2) it was suggested that plants use interhomologous recombination as the main repair pathway during meiotic prophase I, which suggests that there are apparent switches that control the propensity of DSBs to be repaired through inter-homologue repair. Elucidating such factors could help in understanding why genetargeting in plants is so much more difficult in plants than it is in other organisms (Puchta 2002; Puchta 2005). Even though we currently have no clue as to what this switch might be, the constitutive high CDK levels of CDKA;1 activity might be of high interest here.

Reverse breeding approaches

Chapters 6 and 7 explain and describe the feasibility of reverse breeding, an anticipated breeding technique based on the generation of gametes with non-recombinant chromosomes by suppressing crossovers in the preceding meiosis. Spores with non-recombinant chromosomes are then formed, which can be regenerated as homozygous doubled

haploids from among which homozygous parents for a starting heterozygote can be selected. The additional power of the technology is that it can also used for the construction of chromosome substitution lines. The possibility of applying reverse breeding in Arabidopsis paves the way for its application in crops, provided that a suitable protocol for the regeneration of spores is present. The applicability of reverse breeding in other species will however also depend on the basic chromosome number, the number of available spores and the efficiency of making doubled haploids in that species. Low chromosome number crops like cucumber (7 pairs) and barley (7 pairs) provide outstanding possibilities for making doubled haploids, and make it most likely that reverse breeding can be applied immediately. In maize, harboring 10 chromosome pairs, the chance of finding a balanced spore with a complete set of non-recombinant chromosomes will be decidedly lower ($½^{10}$=~0.1%). "Inducer stocks", that are known to produce DHs through a genome elimination mechanism, similar to that in *Arabidopsis*, can be used to make as many as 500,000 DHs per year, some breeding companies claim. If the efficiency of making a doubled haploid is 10%, then at least 10,240 spores must be induced to obtain a reverse breeding DH population. In this case, the scale of making DHs is in maize is so vast, that it is well possible that the available scale for DH induction compensates for the low recovery of balanced gametes. However, this still has to be demonstrated in practice.

In other crops the DH techniques will be less efficient and plants may produce less spores while chromosome numbers may be higher. For a crop like eggplant, the chance of obtaining a balanced gamete from an achiasmatic meiosis is as low as 0.02%, which is, considering the low DH-production efficiency, a poor starting point for successful re-

chromosome number	CO=0	CO=1	CO=2	CO=3	CO-4
5	3.13	6.25	12.50	25.00	50.00
6	1.56	3.13	6.25	12.50	25.00
7	0.78	1.56	3.13	6.25	12.50
8	0.39	0.78	1.56	3.13	6.25
9	0.20	0.39	0.78	1.56	3.13
10	0.10	0.20	0.39	0.78	1.56
11	0.05	0.10	0.20	0.39	0.78
12	0.02	0.05	0.10	0.20	0.39
% of RB offspring (efficiency)	100	50	25	12.5	6.25

Table 1: Reverse breeding with residual crossovers. The expected number of balanced gametes (%) is given as a function of chromosome number and residual crossovers. The column CO=0 describes the situation under complete crossover suppression, whereas CO=4 describes the situation when 4 crossovers are present. We here assume a simple model in which only one CO is formed per chromosome pair. The efficiency (last row) describes the number of offspring that will be non-recombinant (*i.e.* true reverse breeding offspring). Note that in column CO=4, there is a 16 fold increase of balanced gametes, but only 6,25% of reverse breeding offspring will be useful for breeding.

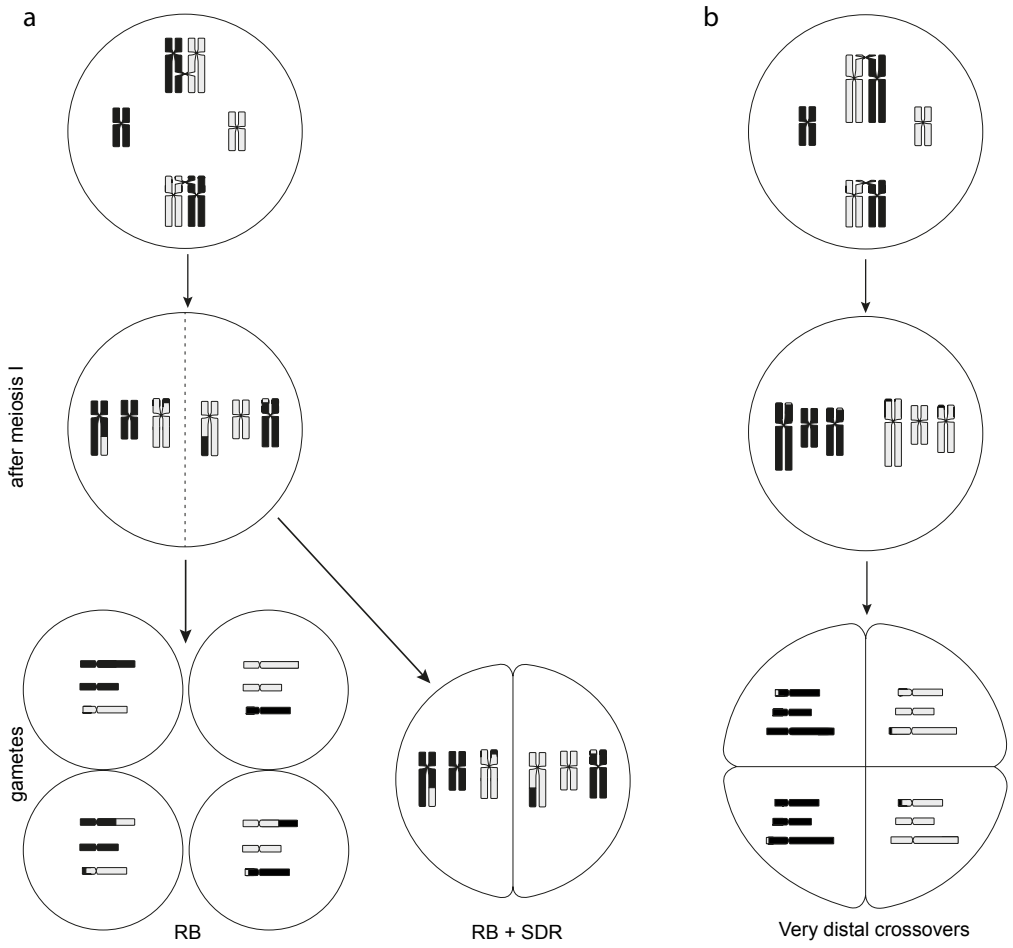


Figure 1: proposed modifications to reverse breeding to increase its efficiency. a) A reverse breeding approach in shown with residual crossovers. Reverse breeding leads to four gametes (lower left) in which some still contain non-recombinant chromosomes. These can be regenerated as DHs and lead to chromosome substitution lines. In an alternative approach, partial crossover suppression is combined with the production of SDR gametes (in which the second meiotic division is omitted. Note that the SDR gametes, consist of largely homozygous chromosome pairs, heterozygous for only those segments distal from crossover positions. Such gametes can be regenerated as diploid plants and in a subsequent DH-step, chromosome substitution lines can be obtained. b) A second reverse breeding approach relies on the distal localization of residual crossovers. The resulting gametes that can be regenerated as DHs show segregation of traits at distal chromosome ends only.

verse breeding. For such a crop the development of reverse breeding will benefit from tailored approaches, as will be discussed hereafter.

Reverse breeding for high chromosome number crops

At higher chromosome numbers the chances of obtaining balanced gametes decrease exponentially, because each additional chromosome pair will diminish the population of balanced spores with another 50% (Table 1) (Dirks *et al.* 2009; Wijnker *et al.* 2012). To increase the number of balanced gametes, an approach could be engineered in which not all, but only part of the COs are suppressed, which is likely feasible by knocking down the expression of genes like *MSH*4 or *MSH*5 (Higgins *et al.* 2008) using RNAi technology. All chromosome pairs with a crossover will segregate in a balanced manner and random segregation is limited to only the chromosome pairs without crossovers (see Figure 1 or Chapter 6). The number of balanced spores will be considerably higher, although this approach comes at a cost: crossovers introduce recombinant chromosomes into the gamete population (and resulting DHs), rendering potentially large portions of the DH population not useful for reverse breeding (Table 1).

It is possible to circumvent this drawback. In recent years, several mutations were described that lead to so-called second division restitution or SDR gametes: diploid (2n) spores that have omitted the second meiotic division (d'Erfurth *et al.* 2010; d'Erfurth *et al.* 2009). A similar effect can be through treatments with spindle inhibitors like nitrous oxide or cold shock (Barba-Gonzalez *et al.* 2006; Okazaki *et al.* 2005; Dewitte *et al.* 2011; De Storme *et al.* 2012). If one partially knocks-down COs as proposed above, while simultaneously inducing SDR meiosis, a (2n, diploid) restitution gamete is obtained which to a greater extent is homozygous (Figure 1a). These 2n gametes can be regenerated as diploid plants using existing techniques for haploid production (Marimuthu *et al.* 2011). The diploid offspring will be largely homozygous, except for those chromosome pairs that experienced a CO event and thus display some "residual" heterozygosity. In a second step, these largely homozygous plants would be used as parental lines for the generation of the required reverse breeding offspring (through DH production or selfing). The required chromosome substitution lines can then be selected from among the offspring using markers. During this step there will no longer be the need to suppress CO formation, since only few segments will segregate. The use of SDR gametes, albeit at WT levels of recombination, has previously been described as "near reverse breeding" (Van Dun and Dirks 2006).

The anticipated low levels of balanced spore formation during achiasmatic meiosis in higher chromosome number crops can thus be overcome by partial suppression of COs. And the drawback of recovering recombinant DH offspring resulting from residual COs can in turn be dealt with by the induction of SDR gametes. The greatest benefit of "SDR-reverse-breeding-withincomplete-CO-suppression" is the generation of offspring that can potentially all be used in further breeding. The residual heterozygosity in these

plants imposes the need of an extra generation for production of the desired reverse breeding lines. However, the required extra time will likely be made up for by the increased efficiency of obtaining reverse breeding lines in a much more targeted manner.

For many crops the genetic base is so small, that only part of the genome is in fact truly heterozygous. Residual crossovers in homozygous chromosome arms do not introduce new variation, which makes way for other reverse breeding approaches. If few crossovers were to be allowed in such plants, the frequency of recombinant chromosomes will increase less quick than shown in Table 1.

The detrimental effect of residual COs can also be lowered in yet another approach in which residual COs are directed to the distal ends of chromosomes (Figure 1b). In this case, the resulting gametes will be subjected to COs recombination, but the extent to which the genomes have been reshuffled is only a fraction as would have been the case of random CO placement and full heterozygosity. Only when distal chromosome ends harbor alleles that are important to the plants' phenotype, the resulting offspring will be useless. If COs could be directed to even the most extreme chromosome ends (close to the genetically empty chromosome regions containing only telomere and subtelomere repeats), reverse breeding would even be feasible in the presence of COs. To date, no mutants have been reported that place COs at distal chromosome ends, but the *cdka*;1DBD we described in Chapter three suggests that constructs targeting CDKA;1 expression could potentially induce such phenotypes.

In an ideal scenario, genes would be identified that induce distributive pairing in heterozygotes of interest. This is a mechanism well known from insects in which male fruitflies and female silkworm do not form crossovers, but nevertheless possess a mechanism to ensure balanced chromosome segregation (McKee 2009; Wolf 1993). Interestingly, such a mechanism has also been described for higher plants in *Fritillaria japonica* (Liliaceae) (Ito *et al.* 1998; Noda 1975). If true, this at least suggests that such phenotypes could be induced in plants. Even though we are still at the start of the implementation of reverse breeding in crops, there is a multitude of possibilities to further develop the efficiency of reverse breeding, in which there is space to tailor reverse breeding to the specific requirements of different crops.

Reverse breeding for mapping and breeding

In its essence, reverse breeding reduces a genome into its elementary physical parts: its chromosomes. These segregate as single units in reverse breeding offspring that typically are mosaics of complete (non-recombinant) parental chromosomes. Chromosome substitution lines, in which a chromosome is substituted for another chromosome of different background, have been constructed already for different species (e.g. wheat, *Arabidopsis* and mouse) and were recognized as potentially powerful mapping populations (Koumproglou *et al.* 2002; Singer *et al*. 2004; Snape *et al.* 1977). The construction of a complete set of all possible chromosome combinations, a chromosome substitution

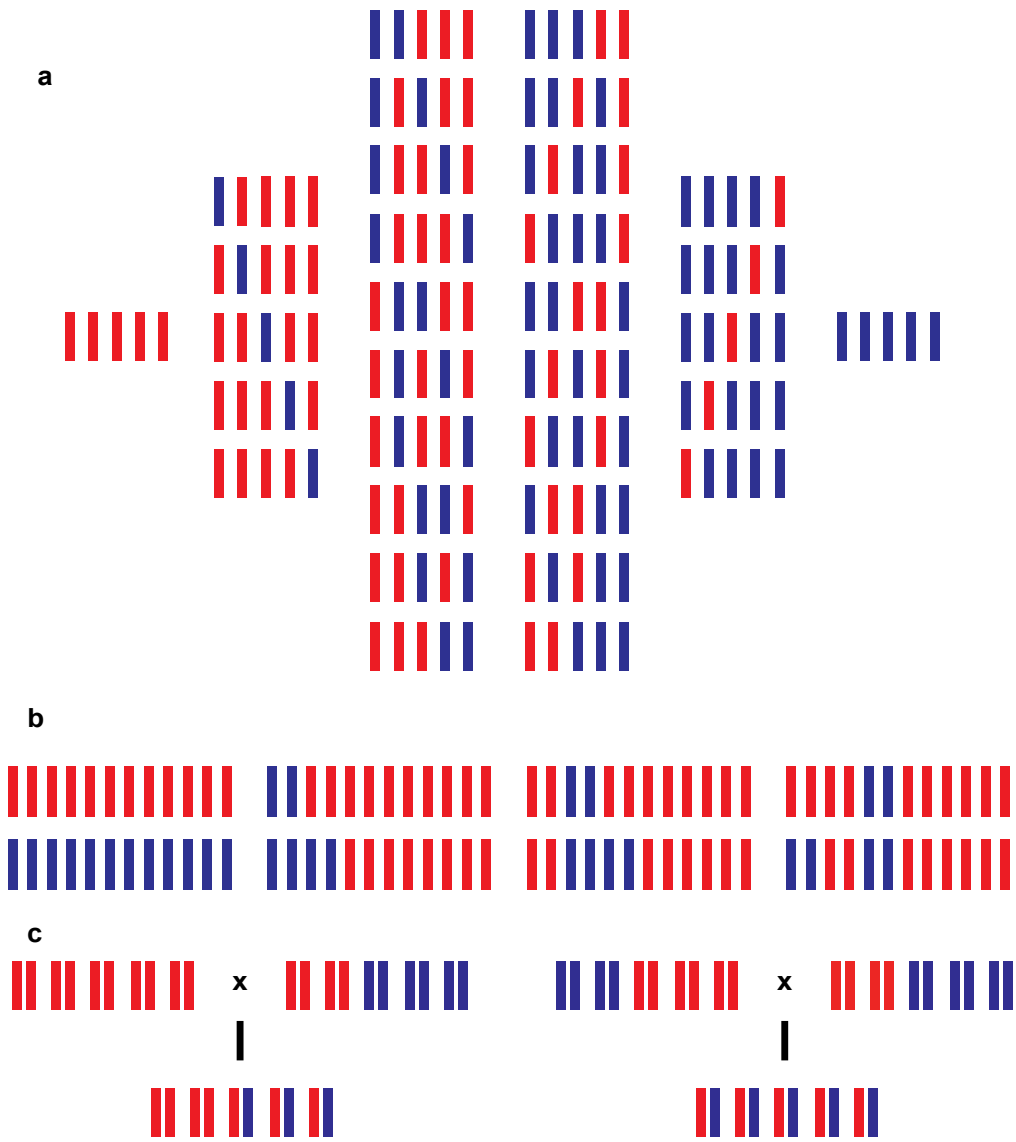


Figure 2: Chromosome substitution libraries. a) An *Arabidopsis* CLS shown above. On the left and right to parental lines are shown. The columns in the middle show all 30 possible chromosome combinations that can be made using the parental lines. One homologue is shown per homologue pair. b) Two parental lines of a tomato CSL are shown on the left (red and blue). The other lines represent randomly chosen lines from a population in which chromosomes segregate two by two. A CSL size can be downscalesd from 4096 to 64 lines. One homologue is shown per chromosome pair. c) A partial *Arabidopsis* hybrid can be constructed by intercrossing chromosome substitution lines. Chromosome pairs are represented by two homologues. Comparison of a partial hybrid (left) with a full hybrid (right) quantifies the contribution of chromosomes 1 and 2 to hybrid performance.

library (CSL) has however not been attempted for plants. In *Arabidopsis*, a CSL comprises a population of 32 different lines only, in which all the genetic variation of the original F1 is distributed (Figure 2a). The relative ease of designing such populations in plants through reverse breeding, prompts for an evaluation of their possible uses in breeding.

The construction of CSLs

The generation of chromosome substitution lines as was done in Chapter 7 (Wijnker *et al.* 2012) relied solely on the random segregation of (non-recombinant) homologues at metaphase I, in which each chromosome combination has a chance of 1/32 to occur. It is nevertheless possible to design a more targeted approach for generation of chromosome substitution lines. After a first round of reverse breeding, specific chromosome substitution lines can be selected in which half of the chromosomes is derived from one parent and the other half from another parent. If such a line is backcrossed to one of the original parents, one obtains a partial heterozygote in which only half of the chromosomes are heterozygous. Subjecting that partial heterozygote to a second round of reverse breeding greatly increases the chances of obtaining more desired chromosome combinations. Note that in this case partial crossover suppression, as discussed in the previous section, would enhance the reverse breeding success manifold, as half of the residual crossovers would occur between identical homologues. This approach of stepwise reverse breeding would especially be useful for higher chromosome number crops like pepper, eggplant (n=12) and rice (n=10). A second round of reverse breeding in eggplant increases the chance of finding a specific chromosome combination from 1/4096 to 1/64. Recent experiments in *Arabidopsis* show that also for this species this is the fastest route to obtaining a complete CSL (data not shown).

CSLs as mapping populations

Trait discovery usually starts with the identification of lines that differ in specific traits, from which then mapping populations are constructed for discovery of loci of interest. Different experimental mapping populations can be constructed, ranging from recombinant inbred lines (RILs), near isogenic lines (NILs), F2's, doubled haploid populations, backcross populations and heterogeneous inbred families (HIFs). The choice for any of these then depends on trait characteristics like number of segregating loci, effect size, dominance, greenhouse space, epistatic interactions, etc. (Schneider 2005). In practice it is very difficult to a priory select the best mapping population and often the best experimental design becomes evident after initial trials when the complexity of a trait becomes evident (J.J.B. Keurentjes, pers. comm.).

If one were to design CSL populations for a crop of interest, like pepper or eggplant (n=12), parental lines can be chosen such that these cover a large proportion of genetic variation in the crop. To answer the initial key questions as mentioned above (i.e. whether genetic variation exists, number of QTLs, epistasis, etc.), one could initially screen a

CSL in which traits of interest are expected to segregate. Note that for such a first inventory, there is no need to screen all the lines in a 12 chromosome CSL which consists of (212=) 4096 lines. Initially, one could screen only single chromosome substitution lines (24 lines), or a subset of 64 lines in which chromosomes are substituted 2 by 2 (see Figure 2b).

In subsequent steps the trait can be further explored. Main effect QTLs can be easily fine mapped by constructing mapping populations that segregate for a single chromosome of interest. Figure 3 shows four DH lines that were generated in our reverse breeding experiment of Chapter 7 (Wijnker *et al.* 2012). QTLs for early flowering and epistatic interactions can be easily assigned to specific chromosomes using these lines (see Figure 3 legend). Anyone interested in fine mapping the early flowering locus on chromosome 5, would cross the two plants on the left (a and b), and fine map the locus in subsequent generations. Note that in this case the population will segregate for traits of only one chromosome, allowing fine mapping in a highly uniform background.

Mapping populations should be developed to most efficiently meet desired purposes. While in certain mapping experiments the detection of main effect loci is sufficient, mapping efforts in crops would benefit most from approaches in which all variation can be systematically detected, also because major effect loci have probably already been

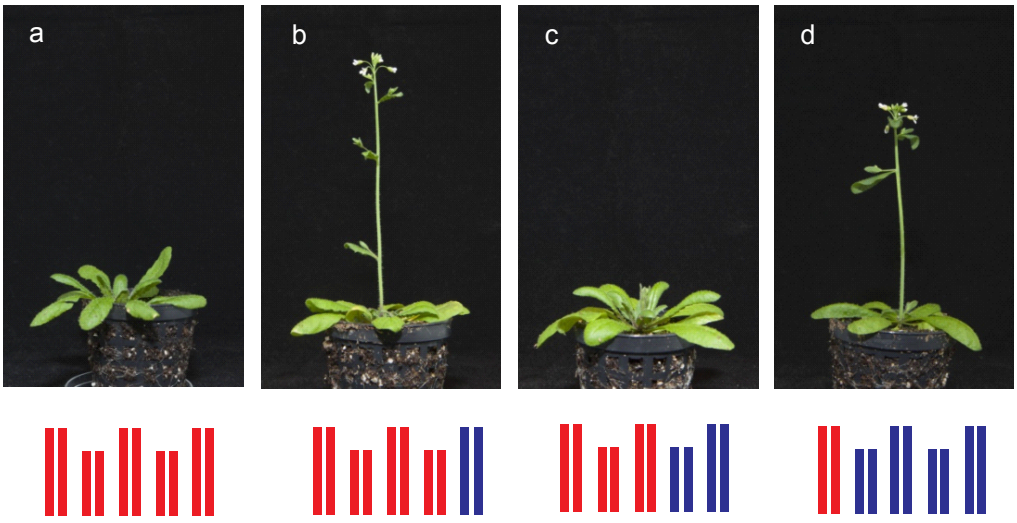


Figure 3: Epistatic interactions in reverse breeding lines. On the left an *Arabidopsis* Columbia plant is shown 27 days after sowing. An RB line (a) has its chromosome 5 substituted by a Landsberg *erecta* chromosome 5. A locus inducing early flowering is present on this chromosome. When also Col chromosome 4 is substituted (c), the early flowering is suppressed (an epistatic interaction). When also chromosomes 2 and 3 are substituted (d), then the early flowering phenotype re-appears (a second epistatic interaction). These interactions can be finemapped by making intercrosses. Crossing lines b with c, generates a plant heterozygous for chromosome 4, that will segregate for loci on chromosome 4 only.

fixed in breeding lines. In such cases, traditional mapping populations like RILs, F2s or DH populations have specific drawbacks. Consider a complex trait regulated by unlinked 12 loci in a RIL or DH population. To reliably detect their interactions, this would require a population of at least 212 (= 4096) plants, while if one required considerable certainty of detecting all loci, one requires a population about three times as large. Such large populations are near impossible to construct in breeding programs for vegetable crops, in which F2 populations of 100 or less plants are rule rather than exception (R. Dirks, pers. comm.).

If one is interested in assessing all variation in available germplasm, one must devise a way of exploring variation systematically, which CSLs can provide. In an initial screen traits can be assigned to specific chromosomes, where after in backcrosses populations can be made that segregate for any chromosome of interest. A CSL has a clear initial drawback in providing poor resolution, since it segregates for whole chromosomes only. If indeed 12 loci segregate, these are unlikely to be detected all at once. The CSL will however allow the subsequent systematic dissection of all loci in subsequent controlled crosses in which only specific chromosomes will segregate. The systematic screening that is achieved by such a stepwise approach will however ensure that no loci are excluded that were not recovered in the initial screen. RILs, F2 or DH populations have the additional drawback of segregating for many loci during backcrosses to one of the parent and show a continuously changing background during introgression. This makes it more likely that specific loci are missed in the mapping approach, or go lost in phenotypic noise. The possibility of constructing targeted mapping populations that segregate for one or more specific chromosomes, can subsequently be used to fine map traits in a highly uniform background, which is a unique property of a CSL. If breeding lines are part of a CSL, it would allow targeted introgression of the detected locus into the desired background in just few steps. Other than when using an F2, DH or RIL population, as CSL will not be the endpoint for experimental mapping, but could itself become the starting point for breeding programs: the building blocks of new varieties. Such benefits will outweigh the considerable investment of the CSL construction itself.

Conceptually CSLs can thus be positioned in between the F1 and any further generation. In the F1, no mapping can be done as no segregation occurs, while in subsequent generations an immediate scattering of the different alleles of all chromosomes occurs because of recombination and independent chromosome assortment. No matter how big a mapping population is set up, the original variation between the parental lines of the starting F1 will never be captured (in terms of practical numbers). A CSL introduces an intermediate step in which uncontrolled scrambling is prevented while ensuring that no variation is lost. Mapping can then be done in subpopulations in which only part of the genome segregates. A CSL can at any point be used to generate introgression line populations, in which chromosome segments are introgressed into a known background, as has been done for example Arabidopsis, tomato and rice (Eshed and Zamir 1995; Koum-

proglou *et al.* 2002; Xu *et al.* 2010) An interesting question will be whether reverse breeding will work with exotic germplasm. Interspecific crosses are known to lead to high levels of infertility in offspring, which makes the construction of RIL populations near impossible(Eshed and Zamir 1995). In such cases, reverse breeding could provide a way out, since it allows the construction of single chromosome substitution lines in one step, possibly overcoming sterility commonly encountered in spores of interspecific hybrids.

CSLs for the study of heterosis

Heterosis, in which a hybrid outperforms the underlying parental lines is well known by plant breeders (Shull 1952). The genetic basis of this phenomenon is far from resolved: the most common theories mention 'overdominance', in which heterozygotes are superior over the homozygotes and 'dominance', in which recessive alleles are complemented in the hybrid by dominant loci, leading to better performance (Birchler *et al.* 2010). A variety of further factors have over the years been shown to influence hybrid performance: maternal effects (Roach and Wulff 1987), paternal effects (House *et al.* 2010), cytoplasm (Fujimoto *et al.* 2012), micro RNAs (Groszmann *et al.* 2011) and epigenetic effects (Shen *et al*. 2012). To study these various effects, F1 hybrids are usually generated through reciprocal crosses of two parents, in which male and female derived offspring show specific differences. It is important to realize that in such crosses, all these factors change simultaneously and are near impossible to study apart: reciprocal F1s will differ in the (maternally) inherited cytoplasm, while maternal effects (*e.g.* endosperm composition) as well as genetically encoded effects like paternal all change simultaneously.

Using an *Arabidopsis* CSL, no less than 16 different crosses can be made to generate the genetically exact same F1 hybrid (and 32 combinations if reciprocal crosses are included). Every parental pair is unique and if genetics are the sole determining factor contributing to the hybrid phenotype, then all hybrids will be phenotypically identical. If the phenotypes are not identical, it matters how the alleles where exposed in their parental genetic backgrounds. The comparison between the (iso-genic) hybrids that are similar and dissimilar in phenotype even allows the identification of the chromosome(s) that causes the difference. Such studies will help in understanding the extent to which hybrids can vary as a result of epigenetic effects.

CSLs might well become prime resources for studying heterosis. By intercrossing different chromosome substitution lines, one can systematically generate "partial hybrids" or "sub-hybrid" families, such that only specific chromosome pairs become heterozygous (Figure 2c). The contribution of heterozygosity of a specific chromosome pair to a hybrid phenotype can as such be mapped on a chromosome-by-chromosome basis. Such studies are within reach for Arabidopsis in the very near future, but once reverse breeding is available for crops- the systematic study of heterosis will be applicable to those.

In crops like maize, breeders make use of so called heterotic pools, groups of germplasm that are genetically distant (Reif *et al.* 2005). Parental lines in maize are selected

from different pools and crossed to evaluate hybrid performance. CSLs and sub-hybrid families would allow the mapping of hybrid effects on a chromosome by chromosome basis, and could help in the identification of chromosome pairs that contribute most to hybrid vigour between specific heterotic groups. If chromosome combinations between different pools are found to contribute differentially to heterosis, new chromosome substitution lines could be constructed that combine the best performing chromosome combinations from different parental pairs into one new pair. Reverse breeding could as such help in the controlled construction of multi-parentderived hybrids a concept called 'Line Design' (Dirks 2012, filed patent application).

Conclusions

The development of new breeding techniques based on modifications of meiosis is still in its infancy. Various approaches have been discussed in the above that could improve reverse breeding through modifications. The rising increased interest in recombination landscapes holds promise that in future years we will start to understand how chromatin dynamics, sequence divergence and crossover interference together control the localization of crossovers. Also the regulation of the meiotic cell cycle by CDKA;1 presents various leads that could help in understanding and controlling homo(eo)logous recombination. As discussed, the development of advanced screening methods (possibly at single cell level), or the use of haploids in studying the induction of homoeologous recombination are promising leads. But perhaps most surprising are the promises held by complexity reduction through crossover suppression. The feasibility of reverse breeding in Arabidopsis strongly suggests that such techniques can indeed be developed for crops and in the above various strategies for improving reverse breeding efficiency have been suggested. Apart from the possibility of fixing uncharacterized heterozygous into breeding lines, reverse breeding offers the possibility of generating CSLs: unique mapping populations that can serve both for the identification of traits, but also as points of departure for the development of new breeding lines. The unique properties of CSLs, that allow the full and stepwise dissection of complex traits in homozygotes as well as heterozygotes and could prove to be unparalleled tools for the dissection of heterosis in years to come.

Literature

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513

Anderson LK, Covey PA, Larsen LR, Bedinger P, Stack SM (2010) Structural differences in chromosomes distinguish species in the tomato clade. Cytogenetic and Genome Research 129:24

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376

Barba-Gonzalez R, Miller C, Ramanna M, van Tuyl J (2006) Nitrous oxide (N2O) induces 2 n gametes in sterile F1 hybrids between Oriental × Asiatic lily (*Lilium*) hybrids and leads to intergenomic recombination. Euphytica 148:303

Berchowitz LE, Francis KE, Bey AL, Copenhaver GP (2007) The role of *AtMUS81* in interference insensitive crossovers in *A. thaliana*. PLoS Genetics 3:e132

Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA (2010) Heterosis. The Plant Cell 22:2105

Canady MA, Ji Y, Chetelat RT (2006) Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174:1775

Colas I, Shaw P, Prieto P, Wanous M, Spielmeyer W, Mago R, Moore G (2008) Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. Proceedings of the National Academy of Sciences 105:6075

Crismani W, Girard C, Froger N, Pradillo M, Santos JL, Chelysheva L, Copenhaver GP, Horlow C, Mercier R (2012) FANCM limits meiotic crossovers. Science 336:1588

d'Erfurth I, Cromer L, Jolivet S, Girard C, Horlow C, Sun Y, To JPC, Berchowitz LE, Copenhaver GP, Mercier R (2010) The CYCLIN-A CYCA1;2/TAM Is required for the meiosis I to meiosis II transition and cooperates with OSD1 for the prophase to first meiotic division transition. PLoS Genetics 6:e1000989

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R (2009) Turning meiosis into mitosis. PLoS Biology 7:e1000124

De Storme N, Copenhaver G and Geelen D (2012). Production of diploid male gametes in *Arabidopsis* by cold-induced destabilization of post-meiotic radial microtubule arrays. Plant physiology (in press)

Dewitte A, Van Laere K and van Huylenbroeck J (2011). Use of 2n Gametes in Plant Breeding. In plant breeding, Abdurakhmonov, I Y (ed) 59-86. InTech, Rijeka, Croatia

Dirks R, Van Dun K, De Snoo CB, Van Den Berg M, Lelivelt CLC, Voermans W, Woudenberg L, De Wit JPC, Reinink K, Schut JW, Van Der Zeeuw E, Vogelaar A, Freymark G, Gutteling EW, Keppel MN, Van Drongelen P, Kieny M, Ellul P, Touraev A, Ma H, De Jong H, Wijnker E (2009) Reverse breeding: a novel breeding approach based on engineered meiosis. Plant Biotechnology Journal 7:837

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147

Esquinas-Alcazar J (2005) Protecting crop genetic diversity for food security: political, ethical and technical challenges. Nature Reviews Genetics 6:946

Feldman M (1966) The effect of chromosomes 5B, 5D, and 5A on chromosomal pairing in *Triticum aestivum.* Proceedings of the National Academy of Sciences 55:1447

Francis KE, Lam SY, Harrison BD, Bey AL, Berchowitz LE, Copenhaver GP (2007) From the cover: pollen tetrad-based visual assay for meiotic recombination in *Arabidopsis*. Proceedings of the National Academy of Sciences 104:3913

Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES (2012) Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. Proceedings of the National Academy of Sciences 109:7109

Gill BS, Friebe BR, White FF (2011) Alien introgressions represent a rich source of genes for crop improvement. Proceedings of the National Academy of Sciences 108:7657

Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mézard C (2011) Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. PLoS Genetics 7:e1002354

Greer E, Martín AC, Pendle A, Colas I, Jones AME, Moore G, Shaw P (2012) The *Ph*1 locus suppresses *Cdk*2-type activity during premeiosis and meiosis in wheat. The Plant Cell 24:152

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES (2011) Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. Proceedings of the National Academy of Sciences

Henson S, Bishop RP, Morzaria S, Spooner PR, Pelle R, Poveda L, Ebeling M, Küng E, Certa U, Daubenberger CA, Qi W (2012) High-resolution genotyping and mapping of recombination and gene conversion in the protozoan *Theileria parva* using whole genome sequencing. BMC Genomics 13:in press

Higgins JD, Vignard J, Mercier R, Pugh AGF, Franklin CH, Jones GH (2008) AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in *Arabidopsis thaliana*, but is not required for synapsis. The Plant Journal 55:28

House C, Roth C, Hunt J, Kover PX (2010) Paternal effects in Arabidopsis indicate that offspring can influence their own size. Proceedings of the Royal Society B: Biological Sciences

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Research 19:1068

Inzé D, De Veylder L (2006) Cell cycle regulation in plant development. Annual Review of Genetics 40:77

Ito M, Takegami MH, Noda S (1998) Achiasmate meiosis in the *Fritillaria japonica* group. II. Formation of synaptinemal complexes in microsporocytes. Japanese Journal of Genetics 58:388

Jenczewski E, Eber F, Grimaud A, Huet S, Lucas MO, Monod H, Chevre AM (2003) *PrBn*, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics 164:645

Knight E, Greer E, Draeger T, Thole V, Reader S, Shaw P, Moore G (2010) Inducing chromosome pairing through premature condensation: analysis of wheat interspecific hybrids. Functional and Intergrative Genomics 10:603

Kopecký D, Lukaszewski AJ, Doležel J (2008) Cytogenetics of *Festulolium* (*Festuca* × *Lolium* hybrids). Cytogenetic and Genome Research 120:370

Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ, Kearsey MJ (2002) STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. The Plant Journal 31:355

Lu P, Han X, Qi J, Yang J, Wijeratne AJ, Li T, Ma H (2012) Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg erecta and all four products of a single meiosis. Genome Research 22:508

Marimuthu MPA, Jolivet S, Ravi M, Pereira L, Davda JN, Cromer L, Wang L, Nogué F, Chan SWL, Siddiqi I, Mercier R (2011) Synthetic clonal reproduction through seeds. Science 331:876

McKee B (2009) Homolog pairing and segregation in Drosophila meiosis. In: Benavente R, Volff J-N (eds) Genome Dynamics, pp. 56-68

Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover ates in euchromatic but not in heterochromatic regions in *Arabidopsis*. Proceedings of the National Academy of Sciences 109:981

Melamed-Bessudo C, Yehuda E, Stuitje AR, Levy AA (2005) A new seed-based assay for meiotic recombination in *Arabidopsis thaliana*. The Plant Journal 43:458

Mercier R, Grelon M (2008) Meiosis in plants: ten years of gene discovery. Cytogenetics and Genome Research 120:281

Moore G (1998) To pair or not to pair: chromosome pairing and evolution. Current Opinion in Plant Biology 1:116

Moore G, Shaw P (2009) Improving the chances of finding the right partner. Current Opinion in Genetics & Development 19:99

Noda S (1975) Achiasmate meiosis in the *Fritillaria japonica* group. Heredity 34:373

Nowack Moritz K, Harashima H, Dissmeyer N, Zhao XA, Bouyer D, Weimer Annika K, De Winter F, Yang F, Schnittger A (2012) Genetic framework of cyclin-dependent kinase function in *Arabidopsis*. Developmental Cell 22:1030

Okazaki K, Kurimoto K, Miyajima c, Enami A, Mizuochi H, Matsumoto Y, Ohya H (2005) Induction of 2n pollen in tulips by arresting the meiotic process with nitrous oxide gas. Euphytica 143:101

Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada J-P (2010) Control of female gamete formation by a small RNA pathway in *Arabidopsis*. Nature 464:628

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau Hannah G, Tischfield Sam E, Zhu X, Neale Matthew J, Jasin M, Socci Nicholas D, Hochwagen A, Keeney S (2011) A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell 144:719

Puchta H (2002) Gene replacement by homologous recombination in plants. Plant Molecular Biology 48:173

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. Journal of Experimental Botany 56:1

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

Ravi M, Chan SWL (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464:615

Ravi M, Marimuthu MPA, Siddiqi I (2008) Gamete formation without meiosis in *Arabidopsis*. Nature 451:1121

Reif JC, Hallauer AR, Melchinger AE (2005) Heterosis and heterotic patterns in maize. Maydica 50:215

Roach DA, Wulff RD (1987) Maternal effects in plants. Annual Review of Ecology and Systematics 18:209

Salome PA, Bomblies K, Fitz J, Laitinen RAE, Warthmann N, Yant L, Weigel D (2012) The recombination landscape in *Arabidopsis thaliana* F2 populations. Heredity 108:447

Schneider K (2005) Mapping populations and principles of genetic mapping. In: Meksem K, Kahl G (eds) The handbook of plant genome mapping. Genetic and physical mapping. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, pp. 1-22

Seymour DK, Filiault DL, Henry IM, Monson-Miller J, Ravi M, Pang A, Comai L, Chan SWL, Maloof JN (2012) Rapid creation of *Arabidopsis* doubled haploid lines for quantitative trait locus mapping. Proceedings of the National Academy of Sciences

Shen H, He H, Li J, Chen W, Wang X, Guo L, Peng Z, He G, Zhong S, Qi Y, Terzaghi W, Deng XW (2012) Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. The Plant Cell 24:875

Shull GH (1952) Beginnings of the heterosis concept. In: Gowen JW (ed) Heterosis. Iowa State College Press, Ames, pp. 14-48

Singer JB, Hill AE, Burrage LC, Olszens KR, Song J, Justice M, O'Brien WE, Conti DV, Witte JS, Lander ES, Nadeau JH (2004) Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304:445

Snape JW, Law CN, Worland AJ (1977) Whole chromosome analysis of height in wheat. Heredity 38:25

The_Tomato_Genome_Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635

Van Dun CMP, Dirks RHG (2006) Rijk Zwaan Zaadteelt en Zaadhandel B.V.. Near Reverse Breeding, WO/2006/094773

Wang J, Fan HC, Behr B, Quake Stephen R (2012) Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. Cell 150:402

Wijnker E, de Jong H (2008) Managing meiotic recombination in plant breeding. Trends in Plant Science 13:640

Wijnker E, van Dun K, de Snoo CB, Lelivelt CLC, Keurentjes JJB, Naharudin NS, Ravi M, Chan SWL, de Jong H, Dirks R (2012) Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. Nature Genetics 44:467

Wolf KW (1993) How meiotic cells deal with non-exchange chromosomes. Bioessays 16:107

Xu J, Zhao Q, Du P, Xu C, Wang B, Feng Q, Liu Q, Tang S, Gu M, Han B, Liang G (2010) Developing high throughput genotyped chromosome segment substitution lines based on population wholegenome re-sequencing in rice (*Oryza sativa* L.). BMC Genomics 11:656

Yamashita K, Yasuda H, Pines J, Yasumoto K, Nishitani H, Ohtsubo M, Hunter T, Sugimura T, Nishimoto T (1990) Okadaic acid, a potent inhibitor of type 1 and type 2A protein phosphatases, activates cdc2/H1 kinase and transiently induces a premature mitosis-like state in BHK21 cells. EMBO Journal 9:4331

Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, Miller N, Drouaud J, Grelon M, Copenhaver GP, Mezard C, Kelly KA, Henderson IR (2012) Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. PLoS Genetics 8:e1002844

Summary

Meiosis is a specialized cell division leading to the production of gametes and the transmission of genetic information in the form of chromosomes. Before being transmitted via gametes, chromosomes recombine with their homologous partners to give rise to new allele combinations. These new combinations can in turn give rise to new phenotypes. This makes a thorough understanding of meiosis pivotal for plant breeders, that rely on meiosis to generate new allele combinations for the development of new crop varieties. This thesis aims at unravelling, understanding and managing the processes that lead to meiotic recombination and asks how this knowledge can be utilized to design new and more efficient plant breeding strategies.

In **chapter 1** various important meiotic processes and their regulation are described and it is shown in what way these processes have links to plant breeding practice. Important strategies for the study of meiotic recombination are also introduced and discussed.

Chapter 2 presents a detailed investigation into the question of to what extent information is exchanged between homologous chromosomes during meiosis in *Arabidopsis thaliana*, a popular model plant in plant genetic research. This is done by the most detailed comparison yet of the genome sequence of offspring with their parents. It is described in detail how homologous chromosomes reciprocally exchange segments through crossover recombination and how at specific places sequences of one homologues are changed into those of another (gene-conversions). We show that gene-conversions in *Arabidopsis* are rare events that mainly occur at crossover breakpoints. We also show that the recombination proteins that during meiosis initiate recombination primarily target easily accessible DNA and finally that there are specific DNA sequences that are associated with crossover sites. One of these sequences shows high similarity to a known transcription factor binding site, that may promote the formation of crossovers.

In **chapter 3** focus shifts to the regulation and coordination of the different meiotic processes during the *Arabidopsis* cell cycle. *Arabidopsis* has one crucial cell-cycle regulator: the protein CDKA;1. The activity of this protein determines what processes take place during precise stages of the cell cycle. By artificially replacing this gene by less functional versions, it can be studied how this affects meiosis and thus reveals the requirement of the cyclin dependent kinase CDKA;1 for meiotic progression. CDKA;1 is required for chromosome condensation, crossover recombination and the second meiotic division. These diverse requirements suggest that CDKA;1 functions as master regulator during meiosis, just as it does in mitosis. A surprise was the observation in a specific mutant that even a slightly lower activity leads to significant changes in the positioning of crossovers onto chromosomes, that localize more distally in the mutant. This could point to a hitherto unknown regulatory mechanism that could explain the differences in crossover positioning between male and female meiosis through differential CDKA;1 activity. Apart from that we find important clues that the functioning of CDKA;1 might

be of high interest for the induction of crossover formation in interspecific hybrids, and hence for plant breeding practice.

Wild relatives of crops can possess traits, like disease resistances, that breeders want to introgress into cultivated crop varieties. Whether that is possible depends to a large extent on structural differences between chromosomes of crops and their wild relatives. **Chapter 4** focuses on these structural differences between the related species tomato, potato and their closes relatives in the genus *Solanum*. Through comparison of the order of sequences on chromosomes, with the use of fluorescent markers, we show the presence of chromosome inversions that occurred during the evolution and divergence of these species. Studying these inversions on the one hand helps understanding the phylogenetic relationships between these species and on the other hand brings valuable knowledge on the presence of inversions that is of the highest importance to plant breeding. The presence of inversions may severely complicate the introgression of traits when these are located in an inverted region, by causing linkage drag.

Chapter 5 is a literature review in which an overview is presented of all the methods that have been described so far that can influence meiotic recombination in plants. This overview can help in the development of strategies to modify recombination in a targeted way to help breeding. Strategies can be subdivided into methods that increase or decrease the total level of recombination or that affect the placement of crossover events on chromosomes. Important techniques for managing recombination include the application of physical stress like heat- or cold shock, chemicals, UV-radiation or the targeted disruption of gene function through transgenes.

Chapter 6 and 7 are both devoted to reverse breeding, an anticipated new breeding technique that was developed to provide breeders with new tools to better breed with complex genomes. In their breeding programs, breeders may encounter superior, unique heterozygous plants. When offspring are produced from such a heterozygote, meiotic recombination will ensure that the favorable allele combinations of the heterozygote go lost in subsequent generations. Reverse breeding is based on the thought that by eliminating crossover recombination, the inheritance pattern of alleles is greatly simplified. Chromosomes will not recombine, but are transmitted as non- recombinant chromosomes to gametes. By regenerating these gametes as homozygous plants (as so called doubled haploids), one can produce parental lines for the chosen heterozygote. When these new parental lines are crossed, the elite heterozygote can be reconstructed. In chapter 6 the theory behind this technique is explained, whereas chapter 7 explains how this technique can be applied in *Arabidopsis*.

In **chapter 8** the knowledge gained through research described in this thesis is evaluated from a plant breeding perspective. It is discussed how new sequencing techniques can be used for studying recombination landscapes: the pattern of crossover placement on chromosomes. The possibilities of changing these recombination landscapes are discussed, in part based on our observations in CDKA;1. Attention is given to the

possibilities of inducing recombination in interspecific hybrids by modifying CDKA;1 expression. Finally the possible applications of reverse breeding for plant improvement are addressed. Reverse breeding not only allows breeders to fix complex heterozygous genotypes in homozygous parents, but also allows the construction of novel mapping populations in plants: chromosome substitution lines. These plants have a remarkable genome constitution since they consist of non-recombinant chromosomes. It is expected that such lines are extremely useful for mapping traits and the study of gene interactions that give rise to heterosis: the observation that hybrids outperform their homozygous parents in terms of yield.

Samenvatting

Meiosis is een gespecialiseerde celdeling speciaal gericht op het produceren van geslachtscellen en het doorgeven van genetische informatie in de vorm van chromosomen. Voordat chromosomen via gameten worden doorgegeven, recombineren homologe chromosomen met elkaar waarbij nieuwe allelcombinaties ontstaan. Deze nieuwe allelcombinaties kunnen in de nakomelingen leiden tot nieuwe phenotypes. Een goed begrip van de meiose is daarmee van cruciaal belang voor plantenveredelaars, omdat daar de variatie ontstaat waarmee zij nieuwe rassen kunnen ontwikkelen. Deze thesis richt zich op het ontrafelen, begrijpen en sturen van processen die leiden tot meiotische recombinatie en hoe met deze kennis nieuwe, meer efficiënte strategieën ontwikkeld kunnen worden voor de veredeling van planten.

In **hoofdstuk 1** wordt een kort overzicht geschetst van meiotische processen, hun regulatie en waar deze mechanismen raken aan plantenveredeling. Tevens worden belangrijke nieuwe strategieën besproken die kunnen helpen bij de bestudering van meiotische recombinatie.

Hoofdstuk 2 beschrijft het tot nog toe meest gedetailleerde onderzoek ooit naar de vraag in welke mate tijdens de meiose informatie wordt uitgewisseld tussen homologen in *Arabidopsis thaliana*, een veel gebruikte modelplant voor genetisch plantenonderzoek. Daarvoor vergelijken we de precieze genoom sequentie van nakomelingen met die van hun ouders. Er wordt in detail beschreven hoe homologen chromosoom stukken uitwisselen door middel van crossovers en hoe op bepaalde plaatsen sequenties van de ene homoloog veranderen in de sequentie van de andere (genconversies). We laten zien dat genconversies in *Arabidopsis* uiterst zeldzame gebeurtenissen zijn, die vooral optreden bij crossover breekpunten. We laten daarnaast zien dat recombinatie vooral plaatsvindt op die plaatsen in het genoom waar het DNA eenvoudig toegankelijk is voor recombinatie eiwitten. Ten slotte laten we zien dat bepaalde DNA sequenties meer dan gemiddeld voorkomen vlakbij recombinatie plaatsen. Een van de gevonden sequenties lijkt verrassend veel op een bekende bindingsplaats voor een transcriptiefactor, die mogelijk crossovers bevordert.

In **hoofdstuk 3** wordt dieper ingegaan op de regulatie en coördinatie van de verschillende meiotische processen gedurende de meiotische celcyclus. *Arabidopsis* heeft één cruciale regulator van die celcyclus: de van-cycline-afhankelijke kinase CKDA;1. De activiteit van dat eiwit bepaalt welke processen op welk moment plaatsvinden. Door dit gen kunstmatig te vervangen door minder goed functionerende varianten, konden we afleiden dat CDKA;1 voor een groot aantal verschillende processen belangrijk is tijdens de meiose: voor chromosoom condensatie, voor crossover recombinatie en voor de tweede meiotische celdeling. Dat wijst erop dat CDKA;1 als lijkt te functioneren als een zogenaamde “master” regulator. Een verrassing was de vaststelling in een bepaalde mutant dat al bij een lichte afname in de activiteit van dit eiwit er grote veranderingen

optreden met betrekking tot de plaatsing van crossovers op de chromosomen, die daarin naar de chromosoom uiteindes verschuiven. Dit wijst mogelijk op een tot op heden onbekend regulatie mechanisme waardoor verschillen in crossover plaatsing tussen mannelijke en vrouwelijke meiose verklaard kunnen worden door verschillen in CDKA;1 activiteit. Daarnaast vinden we aanwijzingen die suggereren dat de activiteit van CDKA;1 van groot belang zou kunnen zijn voor het induceren van crossovers in soortskruisingen. Op termijn zou dat belangrijk kunnen zijn voor de veredelingspraktijk.

In wilde verwanten van gewassen kunnen eigenschappen voorkomen, zoals resistenties, die veredelaars willen inkruisen in bijvoorbeeld bestaande tomatenrassen. De mate waarin dat lukt hangt goeddeels af van bestaande structurele verschillen tussen chromosomen. In **hoofdstuk 4** ligt de focus op deze structurele verschillen tussen chromosomen van aardappel, tomaat en hun meest nauwe verwanten uit het genus *Solanum*. Door het vergelijken van de volgorde van sequenties op chromosomen door middel van fluorescente markers, laten we zien dat chromosomen tussen verschillende soorten aanzienlijk kunnen verschillen in ordening van genen, doordat in de evolutie chromosoom inversies zijn opgetreden. Door het vergelijken van deze inversies kunnen enerzijds verwantschappen tussen soorten worden vastgesteld maar anderzijds is kennis van deze inversies ontzettend belangrijk voor de veredelingspraktijk. Wanneer een veredelaar eigenschappen van een verwante soort in een tomaat wil inkruisen, is dat door inversies soms eenvoudigweg onmogelijk zonder tegelijkertijd alle andere genen en allelen in de inversie in te kruisen.

Hoofdstuk 5 beschrijft de een literatuuronderzoek waarin een overzicht geschetst wordt van de manieren waarop in het verleden is aangetoond hoe meiotische recombinatie beïnvloed kan worden in planten. Daarmee kunnen verschillende strategieën ontwikkeld worden voor het doegericht beïnvloeden van recombinatie in de veredeling van planten. Daarbij kan onderscheid gemaakt worden in strategieën die de hoeveelheid recombinatie beïnvloeden, of juist de plaatsing van recombinatie gebeurtenissen op de chromosomen. Belangrijke technieken zijn de toepassing van physiologische stress als hitte of koude shock, chemicaliën, UV-straling of het doelgericht uitschakelen van bepaalde genen met behulp van transgenen.

Hoofdstukken 6 en 7 zijn beide gewijd aan reverse breeding, een nieuwe veredelingstechniek die werd ontwikkeld om veredelaars een nieuw gereedschap in handen te geven om met complexe genomen te kunnen veredelen. Veredelaars komen in hun veredelingsprogrammas soms planten tegen die heterozygoot zijn: planten waarvan de beide homologe chromosomen verschillend zijn. Wanneer zulke planten bijzondere, unieke eigenschappen hebben, is het in de praktijk uiterst moeilijk om zulke planten te recreëren: Wanneer uit deze planten nakomelingen worden gekweekt, gaan door meiotische recombinatie de unieke allelcombinaties van de heterozygoot onherroepelijk verloren. Reverse breeding is een nieuwe veredelingstchniek waarin de meiotische recombinatie grotendeels wordt uitgeschakeld. De consequentie daarvan is dat in de gameten

veel minder variatie wordt aangetroffen dan in traditioneel gerecombineerde gameten. Chromosomen wisselen in reverse breeding geen stukken meer uit door crossover recombinatie, en chromosomen erven als niet-gerecombineerde chromosomen over naar nakomelingen. Door deze gameten niet te kruisen, maar door speciale technieken direct op te laten groeien als planten (zogenaamde dubbele haploiden), kunnen ouderlijnen voor de heterozygote plant gemaakt worden. Wanneer deze planten vervolgens gekruisd worden, kan de veredelaar de unieke plant weer recreëren. In hoofdstuk 6 wordt de theorie en mogelijke toepassingen van deze techniek besproken, terwijl in hoofdstuk 7 wordt beschreven hoe deze techniek in *Arabidopsis* kan worden toegepast.

In **hoofdstuk 8** worden de resultaten van de verschillende beschreven onderzoeken besproken vanuit het perspectief van de planten veredeling. Er wordt ingegaan op de mogelijkheden van het gebruik van nieuwe (sequencing-) technieken voor het onderzoeken van recombinatielandschappen: van het patroon waarin recombinatie plaatsvindt op chromosomen. Mogelijkheden voor het veranderen van dit recombinetielandschap, en mogelijke andere toepassingen van het gebruik van onze nieuwe kennis van CDKA;1 worden ook besproken. Daarbij veel aandacht voor de mogelijkheden die er zouden kunnen liggen voor het induceren van recombinatie van chromosomen die afkomstig zijn van verschillende soorten. Ten slotte worden nieuwe toepassingen van reverse breeding besproken. Reverse breeding stelt veredelaars niet alleen in staat om ouderplanten te creëren voor heterozygote planten, maar biedt ook mogelijkheden voor het maken van een nieuw soort plantenpopulaties: chromosoom substitutielijnen. Deze planten hebben een bijzondere genoomopbouw: ze bestaan immers uit niet-gerecombineerde chromosomen. Naar verwachting bieden deze bijzondere mogelijkheden voor het vinden van genen die coderen voor bepaalde eigenschappen en de studie van interacties in heterozygote genomen. Dat laatste is van het grootste belang, omdat de meeste commerciële rassen bestaan uit heterozygote planten, die een veel hogere opbrengst geven dan homozygote planten (een effect dat bekend staat als heterosis).

Acknowledgements

Almost to my surprise, this thesis here finds its final form and shape. I look back on this PhD with the greatest pleasure and it is great to sit down for a moment and oversee the time that passed to this moment. Although there is some sort of redemption in spending hours in solitude behind a microscope or sniffing the sweet smell of acetic acid that emanates from a hot plate once you cook a slide to perfection, these do not constitute the majority of my memories. These go out to those that I worked with that make the time spent so much more pleasurable and valuable. I thus happily accept this opportunity to direct some words to those that helped me out, and worked with me during these years. You are all so connected to the time I worked on my thesis, that it would not be complete without you in it.

First of all, I wanted to thank my supervisor and promotor Hans, whom I first met during a masters course. As it turns out, your enthusiasm and love for genetics planted a seed that would have a profound impact on all years thereafter. Following a thesis project during my masters, I found my way back to your office years later, wondering whether I should pursue a PhD in the field of which you showed me the elegant and challenging complexity. What struck me most during the time we worked together were continuous efforts in helping me find my way though science and to grow. You helped me challenge myself, sent me off to conferences to explore what science could be like and have provided me with the greatest support throughout. Your devotion to- and joy in introducing science to new students is something that I experienced first hand as a student and later we got to share this when discussing the possible ways of presenting and teaching complex ideas. I think we truly found one another in our joy of communicating science to others. I have great memories of all the times we drove back and forth from Wageningen to Fijnaart, discussing results and plans on the way there and reflecting on life and science on the way back home. Thanks a lot for all your help.

My greatest appreciation also goes out to all my fellow researchers at the Rijk Zwaan breeding company. You have provided me with amazing support throughout my studies and our collaboration turned out to be highly fruitful. Rob, my co-promotor, your enthusiasm is contagious and your constant flow of ideas is a joy for any discussion. I remember one of the first days Hans and I visited Rijk Zwaan for a day full of discussions. I was lost in complete confusion when driving back to Wageningen and realizing I had been unable to make a single useful comment to help the discussion forward. I made it my challenge never to leave discussions like that thereafter, which, luckily, has changed through time. The confusion is gone, but when I now leave, new ideas and insights still tend to keep me occupied for hours afterwards. That's just great. I have really enjoyed our working together.

But there is another part to the Rijk Zwaan story that was most valuable to me. During my PhD work I have felt like a part of a team, devoted to get the reverse breeding

story running further. My meetings with Rob, Kees, Cilia, Bas en Hans were a constant factor during my PhD. Kees, thank you very much for the valuable discussions, bringing clarity in the occasional complex world of breeding and keeping my eye on the feasibility and details of our research. Cilia, you share that eye for detail and your knowledge on crop cytogenetics keeps turning up great surprises. You were all so important in keeping the communication lines between the Fijnaart and the Wageningen offices running smoothly. Bas, working with you while keeping the experiments going on both sides of the phone line has been a true pleasure. We got to share so many good moments on the phone, discussing both disappointing as well as great results, that at times I felt you were part of our office in Wageningen. Thank you all very much for the great times we had together.

My thanks also go out to Taco, whom at times patiently and enthusiastically manned the Fijnaart-Wageningen hotline and played a not-to-underestimate part in the interdepartmental trafficking services. More thanks go to Dilly, a stronghold in our research infrastructure, to Leo for all the interesting insights in the existing GMO regulations, to Paul and Ilse for your help with nitrous oxide experiments and to Gerald, Aat, Jack and all the others that got involved at any point. Even though not directly involved, I thank the directors of the Rijk Zwaan breeding company, especially Kees Reinink, for giving us the opportunity to make this collaboration the success it is. Our close collaborations gave me invaluable insights into the interests of breeding companies and the questions they have when it comes to crop breeding. I sincerely hope that we will continue working together on new questions and ideas in the future.

There were others as well that were a great support during my PhD. Among those, I wanted to thank my promotor Maarten Koornneef who got more and more involved during the last years of my research. I have really appreciated your help over these years that included a lot of fun discussions and helping out with papers and research questions. I have with joy seen and experienced what great things can happen when you set your network to work and bring new people together. As well, I wanted to thank my external supervisor Tom Gerats. While things were usually running rather smoothly with respect to my PhD progress, I appreciated the moments when we could reflect on the PhD as a whole and appreciated your help in bringing directions to the next steps to take. You share that great quality with Evert Jacobsen. Evert, also with you I spent memorable moments pondering the life of a scientist. I have the best memories of was our rather recent visit to Korea, in which you literally showed me the wider profound impact of working in science and with students from all over the world. My thanks here also go out to Douwe Zuidema at EPS who has been a great help throughout my PhD as well as to Roland Feenstra (transitie-instituut, Wageningen) for your great help in disentangling and sorting some intricate questions regarding science and life.

Science is a great challenge that can make one go to work with a big smile in the morning. In this respect I need to thank the last of strongholds in my PhD infrastructure:

Joost. Your presence in our group has illustrated the importance- and given a completely new meaning to the concept "coffee break". Developing new research ideas at times seemed to turn into a daily routine. You showed me that in developing research the sky is the limit and that proper thinking can occasionally really get you around those last barriers as well. Either as colleague, shadow supervisor or friend, your help in getting some projects off the ground was really invaluable; I look forward in bringing all those plans we have cooking to new great publications!

In Hans' group, Dora and myself ran rather parallel in working our way through the occasional hiccups that present themselves when working through a PhD. Dora, you steadily beat me to finishing that PhD, but I feel that we both turned out pretty well in the end. We both learned so much from that *Solanum* paper we did together and I think back with great pleasure of the time we spent together. I wish you a lot of success in the times to come and your new job. In our group I have also dear memories to working with our queen of crosses: Hetty. Thank you for all your help in making the sheer endless list of crosses that were made these last years. I have slept so much better having you looking after our babies in the greenhouse! My thanks go as well to Jose, who by taking firm control amidst rising chaos, has given our lab a soft landing in tranquility and organized structure. Thank you for your help throughout, and I look forward to getting our first paper done together (and hitching a ride on your publication list!). Laurens, you just joined Hans' group a year ago, but I have the best memories of our first explorations into the wonder world of reverse breeding applications. You are man who can do it all… the best of luck in your new job! Xianwen, good luck with that last upcoming year, and Leila, the same to you with finishing your studies. Xiaomin, Sarut, Guy, Henny, Penka and all the others that were around in the group, I have enjoyed these years with you.

The help I have gotten in and around the lab has been truly great. Corrie, while at times convincingly playing your role in strictly keeping the genetics lab free from disorderly conduct, your luckily have great difficulties hiding the helpful colleague you are behind your veil of strictness. I enjoyed the times we worked together in the lab or greenhouses. Frank, what can I say? Working with you has been great fun and the walls around your desk would be covered in special certificates if those were awarded for all the times you saved the day. Thanks for your help so far and I hope we have not reached the end of it.

When it comes to invaluable help, I could not forget to acknowledge those who have with great care looked after all the plants we have grown these years. Bert Essenstam, Casper Pillen, Taede Stoker, Gerrit Stunnenberg and Michel Hagendoorn have all helped me and us out many times during these years. Thank you very much for all your care in and around the greenhouses.

A pillar supporting the research-infrastructure in our lab has rather recently joined our group: our new professor Bas. After you joined our group some years ago, your efforts to bring the group together and in stimulating us to find common ground in our

experiments meant a true boost for the atmosphere in our group. I have enjoyed our discussions and comments on the research we were doing, but also much appreciate your help in the more practical problems of the last months. My thanks as well to Mark, Fons, Arjen, Duur, Klaas and Bart Pannebakker with all of whom I have had great discussions, both on science and other stuff. You have been very helpful and a true pleasure to work with. There are some plans up there in the air on which I hope we could develop some of those in the future. Marijke and Bertha, as well as all those in the group of Joost (including Miranda), I enjoyed the time we discussed and spent together.

One of the great things of being a PhD student is the opportunity to help master students finding their way into science and further. Shima, I enjoyed working with you, and it is no surprise your excellent work found your way into one of our publications. Fu Yu, your lively presence in our group was a real joy, you will have taken to you current lab. Daniel, you would be surprised to know how many times your name came up in conversations on the though side of science, where sometimes precision and repetitive work is required to lay down the base of the best discoveries. If I would have half your ability for precision and keeping records, my future as scientist would be that of near certain glory. Jan, it was great to have you around and sharing the sheer fun in learning new things. Xiaoqian, I'm sure you'll find your way in science and thanks for the best compliment someone made me during all these years.

The fact that I have just placed a nice cool beer next to my laptop, is indicative of me diverting some thoughts the less scientific parts of science. Although… That distinction line can be all too thin, as perhaps best illustrated by Padraic with whom great science and great fun quickly morphs into a continuum that is relentlessly celebrated in lab or bar (just as it suits the time of day). Julio, with whom science near naturally grows to NASA-like proportions, was the devoted loyal second member of the make-sure-Erik-doesn't-go-home-and drinks-more-beers taskforce. Guys, if *Arabidopsis* collecting were ever to be olympic, I know we will all happily change careers immediately to devote our life to outright and endless pleasure. "To party like there is no tomorrow" has been given a lasting meaning throughout these years. I do already miss having you close by.

It was great fun to work together with Wytske and Bart to spent time getting activities going in the "feestcommissie" (with the occasional great help of Siemen, Frank, Julio and others). So many good memories I have (e.g. the great Christmas parties that we had at Rolfs place, the sports days with Arabidop-THIS!... I will need to mine all the photos and movies on Charles' harddisk to remind me of all!). I do already miss the Friday drinks at our coffee corner. It made for many good laughs with Wilko, Leo, Ronnie, Leonie, Ralph and many others of plant Physiology. I above briefly mentioned the *Arabidopsis* collecting initiative. Admittedly, it may have gone out of hand a little bit (since we, ehh, I mean Frank, are still busy processing these seeds), but one day will we'll have all those plants properly bagged. Joost, Rik, Frank, Padraic, Julio and Maarten… and all those others collecting from Maastricht to Eemshaven… We had some great momentum

going there (among others Wouter). For sure, one of us will find Wilco at his desk one day, asking for the samples… And we will give him, of course… But make him sweat and beg first!

Canoetrips, BBQs, propositions pubs or the many the many Friday evenings in de Zaaier with Bart (my surely best dressed colleague), Felipe (I loved those goose eggs!), Charles, Natalia, Manickam, Mohamed, Florien, Benyamin, Parisa, Aina, Tina, Cecilia, Daniela, Anna, Jenny, Merijn, Eric. I liked that a lot. And to all of those in genetics who were around: Ana Carolina, Mihal, Paola, Zeshan, Pingping, Marieke, Johanna, Myriam, Vicenzio, Roxanne, Martijn, Mina, Claudio, Yanli, Jelle, Jimeng, Wessel, Ya-Fen and Diana: thank you all very much. My last special thanks in this group go out to Wytske. Your office is sort of a sunny resort, or perhaps safe haven where one can get that friendly, comforting chat that helps one overcome the tiny hiccups in the system of science.

I wanted to thank a few more colleagues at other groups as well. It was great to share our lab with the nice colleagues of molecular Biology. Thanks Marijke for all your help in and outside the lab, as well as Ton, Olga and Ludmila. Kerstin, Anneke and Kim at PRI Biosciences have been a great help various experiments over these years. Sander Peters and Paul Fransz, I enjoyed those discussions whenever you visited our lab, as well as those with Fred van Eeuwijk and Martin Boer with whom I hope we will keep fun research going.

To my pleasure, I was able to visit a variety of labs throughout the world that to me was both very inspiring and presented the opportunity to and set up some lasting collaborations. I wanted to thank Greg Copenhaver for hosting me in his lab and for help thereafter. My various visits to the lab of Arp Schnitger have nearly re-defined hospitality. Arp and Maren, my stays made me near literally feel like one of the family and the collaborations with Hiro, Anni, Nico and the others in the lab there were as inspiring as they were pleasant. We have so many fun data lying around, that I can hardy wait to see what it boils down to in a year from now. The collaboration with Korbinian Schneeberger at Cologne has been a true journey of its own: the work de did with Joost, Maarten, Geo and others has occasionally been like pressure cooker in which great fun, challenges and puzzling results made way for some great and special science. To conclude my virtual trip around the globe, my staying and later working closely together and discussing science during nightly skype meetings with Maruthachalam Ravi at UCDavis was a true joy. Also with you I hope we have not seen that last of projects together.

During my PhD, I have had the privilege to work together with some of the most inspiring scientists in my field. With pain I saw two of them pass away during my studies: Piet Stam (Wageningen) and Simon Chan (UCDavis). Piet was one of the most endearing and charming persons to work with. His calm appearance and sharp mind made for good moments when, with great humor, he analyzed the humor in the revolving nature of science. His passion for science was equaled by that of Simon, who in the short time he had, lived close to his dream of devoting his life to his passion: science. The fun of thinking of

new challenging ideas could radiate off of him like light from fire, and I will dearly miss his presence at the many conferences to come. To both I keep dear memories.

I find myself here looking back on the time that past from university to today. The route leading up to this moment where I now write these finishing words has been very, very rewarding. Not all too straightforward, though. While distant in time, the moment at which a scientific career seemed near impossible to pursue sometimes seems remarkably close. Dear parents, Theo and Mieke, I have come to know you as my warmest supporters throughout my life. From the earliest days of picking apart owl pellets, from the days of overloading our family car during holidays with the thousands of fossils we all collected, from the days of stuffing your freezer with precious roadkills, from those days to now. To this rather strange moment where it dawns upon me, that it curiously feels like I've actually done it, that I've become an actual scientist. Isn't that weird? The opportunities you showed me and created for me now create themselves though all those I introduced to you above. Isn't that amazing? Thanks so much for showing me that path as a little kid and even more all the precious moments when you patiently and lovingly manned the mental "ravitaillering" along the route.

Many thanks as well to my brother and sister Andries and Winnifred. The various times we met together in Utrecht, together with Saraya and Leendert, or just together, were precious moments in which you pulled me from the small town that Wageningen is into the big city life where it is great to reflect over a nice dinner. Amies, your stories I can nearly dream after all the numerous moments that I read them during my studies. They still serve as the most powerful remedy to "reset" my mind at times when I am out of inspiration. Winnifred, the cover of this thesis really looks very neat! Thanks a lot for that finishing touch.

A few words here also to all my friends in Utrecht and around, since the countless evenings when you provided me with the necessary diversions from my everyday life of science that appears so omnipresent at times, was so valuable to be. Aniek, your number was always among the most frequently chosen whenever Wageningen got too small or close. The many cups of tea and glasses of beer we had, as well as our canoeing or walks in Rhijnauwen were the good escapes that I value greatly. Pondering life (and limbs) in the Utrecht bars with Louk, in DBs with Louk and Wim or the dark German forests while enjoying cool beer and blackened bratwurst with Louk and Wim and Martijn and Barend: more literally could the escape from daily life hardly be! The inspiring evenings in which poetry was served along with great dishes meant yet again other welcome diversions with Rienk, Angelique, Geertien, Ingmar (sport, sport, sport!), Louk (who else?), Eva and others. "Uitwaaien" was taken most literally when Jaap organized the invaluable weeks on the MD3 to set our minds straight and back on track while sailing the Waddenzee.

Then some last words of thanks to all those others that played crucial roles in helping me otherwise occupied. Harold and Danielle, thanks for looking after my cat Poes during my stays abroad. Pepijn, it has taken me some time to come to terms with you

re-infecting me with the fossil virus (and making me question whether Genetics or Palaeontology is more fun to study). The question has luckily been resolved now and you more than made up for it since (with for example all the great fish we cooked or tasted). Eddy en Vilmar, John de Vos, Frank Wesseling, Storrs Olson, Adiel Klompmaker, Pieter de Schutter, Freek Bakker, Ronald van den Berg, you did very well in keeping fossils and systematics close enough for me not to forget about them.

Dear all, it's been great. Hope to see you soon again.

Erik

Curriculum Vitae

Theoderik Gerardus (Erik) Wijnker was born on 31 December, 1974 in Utrecht, the Netherlands. After obtaining his highschool degree from the "Herman Jordan Lyceum" in Zeist, the Netherlands, he moved to Utrecht in 1994 to start his studies at the University of Utrecht at the faculty of Social Sciences. He completed two years of the four-year curriculum of the masters degree programme Algemene Sociale Wetenschappen (General Social Sciences). In 1997 he moved to Wageningen, the Netherlands and enrolled in the masters degree programme Biology at Wageningen University.

He followed the curriculum of theoretical and mathematical Biology. In this first MSc thesis he studied the attachment pattern of seeds of the parasitic weed *Striga hermonthica* onto host roots through 3D-computer modeling of Sorghum root systems. In a second MSc thesis at the Netherlands Institute of Ecology (NIOO) in Heteren, the Netherlands he studied apomixis (clonal reproduction through seeds) in Dandelion (*Taraxacum officinale*). For this he phenotyped and genotyped diploid and triploid offspring and genetically mapped apomictic traits in a segregating mixed diploid/triploid population. For a third MSc thesis he worked at Naturalis, the National Museum of Natural History (Leiden, the Netherlands) and the Smithsonian institution, National Museum of Natural History (Washington, USA) and identified 35 species (of which 6 new to science) in a collection of marine Neogene birds from the Netherlands. After his graduation –*cum laude*- in 2005, he shortly worked as a research assistant at Wageningen University.

In 2007, Erik started working as a PhD student at the Laboratory of Genetics under the supervision of Prof. Hans de Jong in close collaboration with the Rijk Zwaan breeding company, the results of which are described in this thesis. The research focused on meiotic recombination and how this can be modified to develop new plant breeding strategies. Meanwhile, he published four papers in peer reviewed journals on Geology and Palaeontology. Currently he works as a researcher in the group of Dr. Arp Schnittger on the regulation of meiosis by CYCLIN-DEPENDENT KINASE A;1 (CDKA;1) and its interacting partners at the IBMP-CNRS, University of Strasbourg, France.

Publications

D Szinay, E Wijnker, R van den Berg, R G F Visser, H de Jong &Y Bai, 2012. Chromosome evolution in *Solanum* traced by cross-species BAC-FISH. *New Phytologist* 195(3):688-98.

N E Yelina, K Choi, L Chelysheva, M Macaulay, C B de Snoo, E Wijnker, N Miller, J Drouaud, M Grelon, G P Copenhaver, C Mezard, K A Kelly & I R Henderson, 2012. Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. *PLoS Genetics* 8(8):e1002844.

E Wijnker, K van Dun, C B de Snoo, C L C Lelivelt, J J B Keurentjes, N S Naharudin, M Ravi, S W L Chan, H de Jong & R Dirks, 2012. Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. *Nature Genetics* 44(4):467-70.

P J de Schutter & E Wijnker, 2012. Large *Centrophorus* (Chondrichthyes, Squaliformes) of the Belgian Neogene continental shelf. *Geologica Belgica* 15:26-36.

R Kooke, E Wijnker & J J B Keurentjes, 2012. Backcross populations and near isogenic lines. *Methods in Molecular Biology* 871:3-16.

E Wijnker & S L Olson, 2009. A revision of the fossil genus *Miocepphus* and other Miocene Alcidae (Aves: Charadriiformes) of the Western North Atlantic Ocean. *Journal of Systematic Palaeontology* 7(4):471-487.

R Dirks, K van Dun, C B de Snoo, M van den Berg, C L C Lelivelt, W Voermans, L Woudenberg, J P C de Wit, K Reinink, J W Schut, E van der Zeeuw, A Vogelaar, G Freymark, E W Gutteling, M N Keppel, P van Drongelen, M Kieny, P Ellul, A Touraev, H Ma, H de Jong & E Wijnker, 2009. Reverse breeding: a novel breeding approach based on engineered meiosis. *Plant Biotechnology Journal* 7(9):837-845.

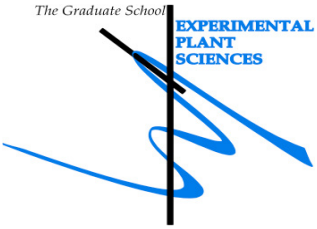
G De Capdeville, M T Souza, D Szinay, L, E C Diniz, E Wijnker, R Swennen, G J H Kema & H de Jong, 2009. The potential of high-resolution BAC-FISH in banana breeding. *Euphytica* 166(3):431-443.

E Wijnker & H de Jong 2008. Managing meiotic recombination in plant breeding. *Trends in Plant Science* 13(12):640-646.

E Wijnker, T J. Bor, F P. Wesselingh, D K. Munsterman, H Brinkhuis, A W. Burger, H B. Vonhof, K Post, K Hoedemakers, A C. Janse & N Taverne, 2008. Neogene stratigraphy of the Langenboom locality (Noord-Brabant, the Netherlands). *Netherlands Journal of Geosciences* 87(2):165-180.

J de Vos & E Wijnker 2006. A deer (*Cervus rhenanus*) from the Early Pliocene of Langenboom, Noord-Brabant (the Netherlands). *Cainozoic Research* 5:107-110.

Education Statement of the Graduate School



Experimental Plant Sciences

Issued to: Theoderik Gerardus Wijnker
Date: 6 February 2013
Group: Laboratory of Genetics, Wageningen University and Research Centre

1)	Start-up phase	date
►	**First presentation of your project**	
	Reverse breeding: a proposed new plant breeding tool	Oct 08, 2007
►	**Writing or rewriting a project proposal**	
	Cytogenetics of reverse breeding technology	Mar 2008
►	Writing a review or book chapter	
	Wijnker and de Jong, 2008, Trends in Plant Science 13(12):640-646	Dec 2008
►	**MSc courses**	
►	**Laboratory use of isotopes**	
	Subtotal Start-up Phase	*5.5 credits**
2)	**Scientific Exposure**	
►	**EPS PhD Student Days**	
	PhD student day 2007, Wageningen	Sep 13, 2007
	PhD student day 2009, Leiden	Feb 26, 2009
►	**EPS Theme Symposia**	
	Theme 4 symposia 2007	Dec 07, 2007
	Theme 4 symposia 2008	Dec 12, 2008
	Theme 4 symposia 2009	Dec 11, 2009
	Theme 4 symposia 2010	Dec 10, 2010
►	**NWO Lunteren days and other National Platforms**	
	ALW-NWO 'Experimental Plant Sciences', Lunteren	Apr 7-8, 2007
	ALW-NWO 'Experimental Plant Sciences', Lunteren	Apr 6-7, 2008
	ALW-NWO 'Experimental Plant Sciences', Lunteren	Apr 2-3, 2012
►	**Seminars (series), workshops and symposia**	
	Cursus 1 kwantit. Genetica, Rijk Zwaan, door Piet Stam	Apr 28-29, 2009
	Cursus 2 kwantit. Genetica, Rijk Zwaan, door Piet Stam	Oct 12-13, 2009
	Cursus 3 kwantit. Genetica, Rijk Zwaan, door Piet Stam	Dec 14-15, 2009
►	**Seminar plus**	
►	**International symposia and congresses**	
	16th International Chromosome conference	Aug 26-29, 2007

Gordon conference meiosis 2008	Jun 08-13, 2008
EMBO meiosis conference, Isle de Sorgue, France	Sep 19-23, 2009
Plant and Animal Genome XVIII Conference, San Diego, CA, USA	Jan 09-13, 2010
Symposium on Contributions from scientific research to the risk assessment of GMO, Brussels, Belgium	Oct 21-22, 2010
Plant and Animal Genome XX Conference, San Diego, CA, USA	Jan 13-18, 2012
EMBO Conference Genetic stability and change, Roscoff, France	May 02-05, 2012
43st annual meeting of the Korean Society of Breeding Science, Daegu, South Korea	Jul 05-06, 2012

► **Presentations**

Presentation: 16th International Chromosome conference, Amsterdam, Netherlands	Aug 29, 2007
Presentation: Group Greg Copenhaver, Univ. of North Carolina at Chapel Hill, NC, USA	Jun 04, 2008
Poster: Gordon Meiosis Conference, New Londen, New Hampshire	Jun 10, 2008
Presentation: EPS theme 4 symposium, Wageningen, Netherlands	Dec 12, 2008
Presentation: Group of Mathilde Grelon, INRA, Versailles, France	Jan 29, 2009
Poster: EMBO Meiosis Conference, Isle de Sorgue, France	Sep 19-23, 2009
Presentation: EPS theme 4 symposium, Nijmegen, Netherlands	Dec 11, 2009
Presentation: Group of Arp Schnittger, IBMP-CNRS, Univ. de Strasbourg, France	Mar 16, 2009
Presentation: PAG XVIII Conference, San Diego, CA, USA (invited speaker)	Jan, 10 2010
Presentation: Tomato Seq. Buisiness Meeting at PAG XVIII, San Diego, CA, USA	Jan 12, 2010
Presentation: Group of Simon Chan, UC Davis, CA, USA.	Jan 14, 2010
Presentation: Group of Plant Breeding, Wageningen University, Netherlands	Sept 13, 2010
Presentation: Symposium on Contributions from Scientific Research to the Risk Assessment of GMO, Brussels, Belgium (invited speaker)	Oct 21-22, 2009
Presentation: KeyGene, Wageningen, Netherlands	Nov 02, 2010
Presentation: 'Nieuwe Veredelingstechnieken en Regelgeving', Plantum NL, Zoetermeer, Netherlands (invited speaker)	Feb 25, 2011
Presentation: Syngenta Symposium "The Frontiers of Breeding; Creating New Variation," Enkhuizen, Netherlands (invited speaker)	Sep 21, 2011
Poster: Plant and Animal Genome XX Conference, San Diego, CA, USA	Jan 13, 2012
Presentation: PAG XX Conference, session Recombination Mechanisms, San Diego, CA, USA (invited speaker)	Jan 15, 2012
Presentation: Plant and Animal Genome XX Conference, session Plant	

Cytogenetics, San Diego, CA, USA (invited speaker)		Jan 15, 2012
Presentation: Group of Arp Schnittger, Université de Strasbourg, France		Mar 16, 2012
Presentation: NWO Lunteren Days, Lunteren, Netherlands (plenary session)		Apr 02, 2012
Presentation: EMBO Conference Genetic Stability and Change, Roscoff, France (invited speaker)		May 04, 2012
Presentation: 43st annual meeting of the Korean Society of Breeding Science, Daegu, South Korea (invited speaker)		Jul 05, 2012
Presentation: Max Planck Institute for Plant Breeding Research, Cologne, Germany		Jul 23, 2012
► **IAB interview**		Dec 04, 2009
► **Excursions**		
	Subtotal Scientific Exposure	*39.4 credits**
3) **In-Depth Studies**		
► **EPS courses or other PhD courses**		
PhD Workshop 'Natural variation in plants'		Aug 26-29, 2008
PhD workshop 'Molecular phylogenies: reconstruction & interpretation'		Oct 19-23, 2009
► **Journal club**		
► **Individual research training**		
	Subtotal In-Depth Studies	*3.0 credits**
4) **Personal development**		
► **Skill training courses**		
PhD Competence Assessment		2007
Guide to digital scientific artwork		Dec 21-23, 2008
15 sessions career planning with personal coach		Jul 2011 - Mar-2012
► **Organisation of PhD students day, course or conference**		
► **Membership of Board, Committee or PhD council**		
	Subtotal Personal Development	*2.7 credits**
TOTAL NUMBER OF CREDIT POINTS*		51.0

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

*) A credit represents a normative study load of 28 hours of study. The whole programme must contain a minimum of 30 ECTS credits.

The research described in this thesis was carried out at the Laboratory of Genetics at Wageningen University, Wageningen, the Netherlands, and was financially supported by the Rijk Zwaan breeding company, Fijnaart, the Netherlands.

Cover by Winnifred Wijnker (winnifred.wijnker@gmail.com).
Strand invasion following a DNA double strand break.

Thesis layout by Hans de Jong.

Printed by CPI Koninklijke Wöhrmann.