DISSERTATION

Titel der Dissertation

Determinants of the initiation of meiotic recombination

Verfasser

Mag.rer.nat. Bernd Edlinger

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien, 2011

Table of Contents

1 Abstract

Spo11, first described in *Saccharomyces cerevisiae*, is a eukaryotic homolog of the archeal DNA topoisomerase VIA subunit and is needed for the formation of DNA double strand breaks (DSBs) during meiosis. DNA is preferentially cleaved at certain sites in the genome, called meiotic recombination hotspots. After DNA cleavage, the Spo11 protein is released from the cleavage site with a short DNA oligonucleotide covalently attached. This short ssDNA sequence corresponds to the exact meiotic DSB cleavage site. The *Arabidopsis thaliana* genome encodes, unlike the ones of mammals and yeast, where only one *Spo11* is present, three *Spo11* homologs, *AtSPO11-1*, *AtSPO11-2* and *AtSPO11-3*. Only *At*SPO11-1 and *At*SPO11-2 are essential for meiosis, whereas *At*SPO11-3 is needed for somatic endoreduplication. In *S. cerevisiae*, in addition to Spo11, at least 9 other proteins are essential for meiotic DSB formation. In *Arabidopsis* only three further proteins, *At*PRD1, *At*PRD2 and *At*PRD3, have been identified to be essential for SPO11 mediated meiotic DSB formation.

The main topic of this thesis was the development of a protocol for the identification and detailed analysis of meiotic DSB sites in the model plant *Arabidopsis thaliana*. First, a deep-sequencing based protocol was developed and optimized using an HRP protein, conjugated to a 30mer oligonucleotide, as a model substrate. After successful optimization of the individual reactions steps and the deep-sequencing pipeline, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* were used as model organisms to verify the method and to compare the results to known DSB maps. Results for both yeasts were very well reproducible and found very good overlap with genome-wide hotspot maps generated earlier by other methods.

To implement the method in the higher eukaryote *Arabidopsis*, a transgenic line expressing a functional, tagged version of *At*SPO11-1 (*At*SPO11-1/18xmyc) was generated. *At*SPO11-1/18xmyc could be successful immuno-precipitated from plant material containing meiotic cells and be used in the protocol mentioned above. The first two deep-sequencing runs, using material from *Arabidopsis* have been performed, yielding 3.5 million and 2.2 million sequence reads that mapped to the *Arabidopsis* genome.

An additional aim of this thesis was the targeted stimulation of meiotic recombination in *A. thaliana*. A specific DNA binding domain was fused to the C-terminus of *At*SPO11-1 (*At*SPO11-1/Gal4) and plant lines carrying corresponding cis-elements (5xUAS) were

generated. After crossing plants with aforementioned features it is anticipated that the tagged *At*SPO11-1 will be guided to the cis-elements and stimulate formation of meiotic DSBs.

Furthermore it was aimed to find *At*SPO11-1 and *At*SPO11-2 interacting proteins. For this purpose, different approaches like mass spectrometry, direct yeast two-hybrid interaction and co-immuno-precipitation were used.

2 Zusammenfassung

Spo11, zuerst in *Saccharomyces cerevisiae* beschrieben, ist ein eukaryotisches Homolog der archealen DNS Topoisomerase VIA Untereinheit und ist für die Bildung von DNS Doppelstrangbrüchen (DSB) während der Meiose notwendig. Die DNS wird bevorzugt an bestimmten Stellen im Genom, welche meiotische Rekombinations Hotspots genannt werden, geschnitten. Nach dem DNS Schnitt wird Spo11 mit einem kovalent gebundenen Oligonukleotid von der Bruchstelle gelöst. Diese kurze einzelsträngige DNS Sequenz entspricht genau dem Ort des Doppelstrangbruches. Das *Arabidopsis thaliana* Genom kodiert, anders als bei Säugern und Hefe, wo es nur ein *Spo11* gibt, für drei *Spo11* Homologe, *AtSPO11-1*, *AtSPO11-2* und *AtSPO11-3*. Nur *At*SPO11-1 und *At*SPO11-2 sind für die Meiose notwendig, während *At*SPO11-3 an der somatischen Endoreduplikation beteiligt ist. Bei *S. cerevisae* sind zusätzlich zu Spo11 zumindest 9 andere Proteine an der DSB Bildung beteiligt. Bei *Arabidopsis* sind bisher nur die drei Proteine *At*PRD1, *At*PRD2 und *At*PRD3, als essentiell für die SPO11 abhängige DSB Bildung, identifiziert worden.

Das Hauptthema dieser Arbeit war die Entwicklung eines Protokolls, um genomweit DSB Stellen in *Arabidopsis thaliana* zu identifizieren. Zuerst wurde ein Deep-Sequencing basierendes Protokoll unter Zuhilfenahme eines Modellsubstrates, einem HRP Protein gekoppelt an ein 30 Nukleotide langes Oligonukleotid, entwickelt und optimiert. Nach der erfolgreichen Optimierung der einzelnen Reaktionsschritte und der Sequenzierung wurden *Saccharomyces cerevisiae* und *Schizosaccharomyces pombe* verwendet, um die Methode zu testen und die Resultate mit denen von bereits publizierten DSB Karten zu vergleichen. Die Resultate für beide Hefen waren sehr gut reproduzierbar und es wurde eine gute Übereinstimmung mit den genomischen Hotspot Karten, die bereits mit anderen Methoden produziert wurden, gefunden.

Um diese Methode auf *Arabidopsis thaliana* zu übertragen, wurde eine transgene Linie, welche eine funktionelle, getaggte Version von *At*SPO11-1 (*At*SPO11-1/18xmyc) exprimiert, generiert. Das Protein wurde erfolgreich aus Pflanzenmaterial, welches meiotische Zellen enthält, präzipitiert und für das oben genannte Protokoll verwendet. Die ersten Deep-Sequencing Durchgänge, welche mit *Arabidopsis* Material durchgeführt wurden, brachten 3,5 Millionen und 2,2 Millionen Sequenzen, welche zum *Arabidopsis* Genom passen.

Ein weiteres Ziel dieser Arbeit war die gezielte Stimulation von meiotischer Rekombination in *A. thaliana.* Eine spezifische DNS Bindedomäne wurde an den C-Terminus von

*At*SPO11-1 fusioniert (*At*SPO11-1/Gal4) und es wurden Pflanzen hergestellt, die korrespondierende Cis-Elemente (5xUAS) tragen. Nach dem Kreuzen der Pflanzen mit den vorher genannten Features wird erwartet, dass *At*SPO11-1/Gal4 an die 5xUAS Stellen bindet, um in weiterer Folge die Bildung von meiotischen Doppelstrangbrüchen zu stimulieren.

Weiters wurde versucht, Interaktionspartner von *At*SPO11-1 und *At*SPO11-2 zu finden. Für diesen Zweck wurden verschiedene Strategien, wie Massenspektrometrie, direkte Yeast two-hybrid Interaktion und Co-Immunopräzipitation verwendet.

3   Introduction

3.1   Meiosis

In eukaryotes a specialized cell division, called meiosis, ensures the formation of generative cells. Meiosis is a two-step division, with homologous centromeres being segregated during the first, and sister centromers during the second division. As there is no intervening DNA replication between the two meiotic divisions, each of the final division products contains only half of the initial DNA content. For a given diploid organism the developing generative cells are then haploid. In contrast, mitosis produces daughter cells with a chromosome complement that is identical to that of the progenitor cell (see figure 3.1). It is important to note that during meiosis genetic information between maternal and paternal chromosomes is mutually exchanged, leading to novel combinations of genetic traits in the following generation. Two genetically diverse generative cells fuse during the process of fertilization, re-establish the organism's original genome content and constitute an individual with a unique genetic set-up (Page and Hawley, 2003; Zickler and Kleckner, 1999).
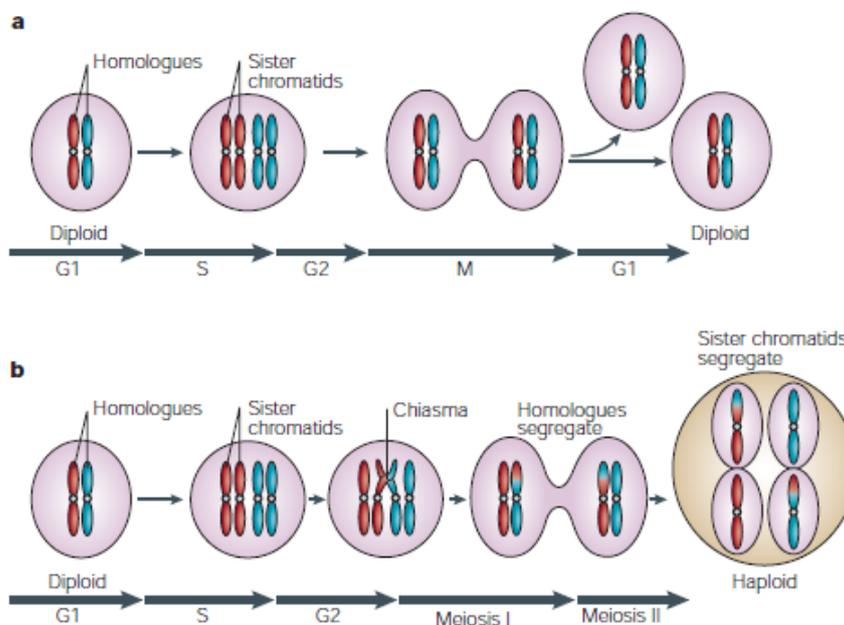
Figure 3.1: a | In mitosis, diploid cells replicate chromosomes during S phase and segregate sister chromatids during M phase, so that diploid daughter cells are produced. b | In meiosis, two chromosome-segregation phases, meiosis I and meiosis II, follow a single round of DNA replication during pre-meiotic S phase. During meiosis I, homologous chromosomes (shown in red and blue) are segregated to opposite poles. Sister chromatids segregate to opposite poles during meiosis II, which results in the formation of non-identical haploid gametes. Figure and text taken from (Marston and Amon, 2004).

Recombination between homologous chromosomes depends on the formation of DSBs. DSBs are formed by a protein complex, with Spo11 proteins representing the catalytically active subunits (see below). As an intermediate of the DNA cleavage process, Spo11 proteins remain covalently linked to the 5' termini of single stranded DNA (ssDNA) at the incision sites and have to be removed (Fig. 3.5C) (Bergerat *et al.*, 1997; Keeney *et al.*, 1997; Neale *et al.*, 2005). To release Spo11 from the DNA ends, DNA is nicked at a distance from the incision site by the MRX complex (Mre11/Rad50/Nbs1-Xrs2) in conjunction with Com1/Sae2 (Alani *et al.*, 1990; Cao *et al.*, 1990; Ivanov *et al.*, 1992; Longhese *et al.*, 2009; McKee and Kleckner, 1997; Mimitou and Symington, 2009a; Nairz and Klein, 1997; Prinz *et al.*, 1997). It is most likely the endonuclease activity of Mre11 that mediates the ssDNA nick formation and the 3'-5' exonuclease activity of Mre11 that resects the ssDNA towards the Spo11 protein. Spo11 is then released from the nascent cleavage site with a short DNA oligonucleotide remaining attached to the Spo11 protein (Fig. 3.5C). To yield long stretches of ssDNA that can probe for matching strands on homologous chromosomes (or in some cases on sister chromatids) three proteins have gained attention. It seems that the exonucleases Exo1 and Dna2 together with the helicase Sgs1 are instrumental for 5'-3' strand resection, starting at the Mre11-mediated ssDNA lesions to yield long single stranded overhangs (Cromie *et al.*, 2008; Longhese *et al.*, 2010; Manfrini *et al.*, 2010; Mimitou and Symington, 2008; Zhu *et al.*, 2008). The ssDNA is bound with high affinity by replication protein A (RPA) (reviewed in Broderick *et al.*, 2010; Fanning *et al.*, 2006), a pre-requisite for the loading of the strand exchange proteins Rad51 and Dmc1 (Fig. 3.5C). In yeast, this loading is mediated by Rad52 and accessory proteins, among them Rad54, Rad55, Rad57, Rad59 and Rdh54/Tid1 (reviewed in Krogh and Symington, 2004). In higher eukaryotes, the BRCA2 protein has been found essential for this step (reviewed in Longhese *et al.*, 2010). Specialized meiotic DNA repair proteins, together with other DNA repair factors mediate strand invasion, strand elongation by DNA synthesis, capture of the second DNA end and subsequent repair and ligation of two different DNA strands to yield novel allelic combinations (reviewed in Hunter, 2006).

3.2 Setting the stage for meiotic recombination

Once a cell has been committed to undergo meiosis and DNA replication has been licensed and initiated (Bergner *et al.*, 2010; Borde *et al.*, 2000; Boselli *et al.*, 2009; Chi *et al.*, 2009; Merritt and Seydoux, 2010; Ronceret *et al.*, 2009; Sanchez-Moran *et al.*, 2004), certain regions in the genome become disposed to being cleaved by the meiotic DSB

machinery. Until recently, the nature of *cis*-requirements for meiotic DSB formation remained elusive. The general understanding about what makes a certain genomic region more prone to receiving a break (generally referred to as "hotspots" of meiotic recombination) than another ("coldspots"), was the presence of an open chromatin status (Fan and Petes, 1996; Keeney and Kleckner, 1996; Ohta *et al.*, 1994). This status may be acquired by transcription factor binding (without a need for transcription) (White *et al.*, 1993), intrinsically open chromatin due to sequence constraints or open chromatin due to histone modification(s) (α-, β-, γ-hotspots; reviewed in Petes, 2001). Recently the Nicolas lab demonstrated that tri-methylation of lysine-4 on histone H3 (H3K4^{me3}) is enriched at meiotic DSB hotspots in promoter regions in *Saccharomyces cerevisiae* (see below for technical details how the DSB sites have been analyzed) (Borde *et al.*, 2009). Deletion of the *SET1* methyltransferase led to a genome-wide decrease of hotspot activity, with hotspots located in promoter regions (the majority of *S. cerevisiae* hotspots) being mostly affected. Interestingly, the H3K4^{me3} marks were present even before pre-meiotic DNA replication. Furthermore, it is worth mentioning that other histone modifications such as histone acetylation or ubiquitination influence hotspot activity as well and that some hotspots were not affected by the *set1* deletion (Borde *et al.*, 2009; Sollier *et al.*, 2004; Struhl, 1998; Yamashita *et al.*, 2004).

The yeast genome is small, and genes and their regulatory sequences are tightly packed, leaving hardly any space for heterochromatin (Zickler and Kleckner, 1999). The genomes of higher eukaryotes, in contrast, are often extremely large, and the co-ordination of DSB formation throughout the genome may even more rely on epigenetic marks. Work performed by the groups of de Massy, Myers and Petkov/Paigen shed new light on the question how certain chromosomal regions become more prone to receiving a meiotic DSB in mouse and human (Baudat *et al.*, 2009; Myers *et al.*, 2009; Parvanov *et al.*, 2009). In parallel, using different approaches, these groups found that the allelic status of a gene termed *PRDM9* was responsible for the efficient initiation of recombination at a certain locus in the mouse and human genomes (Baudat *et al.*, 2009; Myers *et al.*, 2009; Parvanov *et al.*, 2009). In a remarkable approach Myers and co-workers analyzed over 30000 human hotspots (deduced from linkage disequilibrium patterns) of certain well documented human populations for the occurrence of a consensus motif in close proximity to highly recombinogenic regions and found a degenerated 13-mer motif to be associated with about 40% of all human recombination hotspots (Myers *et al.*, 2005; Myers *et al.*, 2009; Myers *et al.*, 2008). In a second step, demonstrating the power of bioinformatics, a candidate list of zinc finger DNA-binding proteins that could specifically recognize the identified motif was established for the human genome, with PRDM9 as the

prime candidate. The de Massy group identified a genetic locus responsible for meiotic recombination activity in a certain region of the mouse genome. The activity determining locus on chromosome XVII was mapped between 12.2 and 16.8 Mb. This region contains the *Prdm9* gene, which encodes a protein with a SET-methyltransferase domain and a tandem array of twelve C2H2 zinc fingers (the human version contains 13 zinc fingers, with a tandem repeat structure similar to that observed in mice). Prdm9 tri-methylates H3K4 (that has already been mono- or di-methlyated) (Hayashi *et al.*, 2005) and is expressed specifically in germ cells during meiotic prophase. The zinc finger domain has been found to be variable in different mouse strains (and variable in different human populations). It is this variability that allows binding to certain genomic loci with higher or lower affinity, leading to more or less H3K4 tri-methylation. A high level of H3K4^{me3} was correlated with high recombination activity at a given recombination hotspot (Baudat *et al.*, 2009). Analysis of *Prdm9*$^{-/-}$ mice showed that *Prdm9* is essential for progression through meiotic prophase, but some DSBs were detected in *Prdm9*$^{-/-}$ spermatocytes, suggesting that *Prdm9* is not absolutely required for DSB formation (Baudat *et al.*, 2009). This is in line with the presence of a Prdm9 signature in only 40% of human recombination hotspots (see above) (Myers *et al.*, 2005; Myers *et al.*, 2009; Myers *et al.*, 2008).

In plants, no PRDM9 homolog has been identified so far. There are 176 genes encoding C2H2 zinc finger proteins (Ciftci-Yilmaz and Mittler, 2008; Englbrecht *et al.*, 2004) and about 47 genes encoding a SET methyltransferase domain (Ng *et al.*, 2007), but there is no clear homolog combining both features within one single reading frame (C. Uanschou, personal communication). The H3K4^{me3} modification seems to generally mark transcriptionally active chromatin, and plants interpret this signal in the same way (Borde *et al.*, 2009; Zhang *et al.*, 2009; Zhang *et al.*, 2006). It will be interesting to see, if plants use this mark in the meiotic context as well. Interestingly, a study published last year highlighted the impact of H3 histone acetylation on meiotic cross over (CO) formation in *Arabidopsis thaliana* (Perrella *et al.*, 2010). Hyperacetylation did not lead to a general increase in CO formation, but differentially to more COs on one chromosome, and to less on two others. This observation is in line with data from different organisms, indicating that histone acetylation has an effect on meiotic recombination (Merker *et al.*, 2008; Mieczkowski *et al.*, 2007; Yamada *et al.*, 2004) but that the modification has different consequences depending on the genomic context. In *Caenorhabditis elegans,* recent work suggests that chromatin modifications play an important role during meiosis. The *him-17* mutant with reduced H3K9^{me2} is defective for meiotic recombination and chromosome segregation due to a defect in DSB formation (Reddy and Villeneuve, 2004) and the *xnd-1*

mutant has elevated levels of H2AK5 acetylation and has an altered DSB and recombination landscape (Wagner *et al.*, 2010).

3.3 Deployment of meiotic DSB proteins

Meiotic DNA DSBs may only be introduced after DNA replication, and therefore the DSB forming machinery has to be connected to the cell cycle and replication control. In the yeast *S. cerevisae* this connection is provided by the S-phase cyclin dependent kinase Cdc28-Clb5 (CDK-S) and the Dbf4-dependent kinase Cdc7-Dbf4 (DDK), both needed for the initiation of (pre-meiotic) DNA replication. The protein Mer2, an essential factor for meiotic DSB formation, is phosphorylated by CDK-S (Henderson *et al.*, 2006). This phosphorylation primes Mer2 for an additional phosphorylation by DDK (Wan *et al.*, 2008). It was speculated that replication fork associated CDK and DDK may coordinate replication and hotspot maturation (Murakami and Keeney, 2008). The negative patch, generated on Mer2 by addition of the phosphate residues, allows the interaction with two further essential DSB complex members, Rec114 and Mei4 (Fig. 3.5A) (Matos *et al.*, 2008; Sasanuma *et al.*, 2008; Wan *et al.*, 2008).

Mer2/Rec107 has been independently described by Engebrecht et al. and Malone et al. in *S. cerevisiae* (Engebrecht *et al.*, 1991; Malone *et al.*, 1991). It is required for chromosome synapsis and initiation of meiotic recombination. A null mutation of *MER2* leads to meiotic lethality (Rockmill *et al.*, 1995). In yeast two-hybrid interaction assays it has been shown that Mer2 interacts with itself, Mei4, Xrs2 and Rec114 (Li *et al.*, 2006). A plant or mammalian counterpart for Mer2 has not been identified yet. The actual molecular link between Mer2 positioning (or the positioning of other DSB factors) and chromatin modifications (e.g. H3K4^{me3}, as outlined above) remains unknown. The challenging task for the future will be to find the factors that can recognize histone modifications and attract (or repel) the meiotic DSB machinery. Furthermore it seems reasonable to assume that similar mechanisms, as described above, are in place in plants. It is still open, whether targeted histone modifications govern DSB formation in plants with higher probability at certain loci. It seems undisputed that DNA replication and DSB formation are linked in plants. This is supported by reports on SPO11 deposition and DSB formation following DNA replication, visualized by BrdU incorporation, in *A. thaliana* (Sanchez-Moran *et al.*, 2008). In *A. thaliana* about 61 core cell cycle genes have been described (Inze and De Veylder, 2006; Vandepoele *et al.*, 2002). The distinct and shared roles of the 12 CDKs and of the at least 30 cyclins, and their impact on plant meiosis is still under investigation.

A CDC7 homolog has been found but its role during meiosis has not been characterized yet. Identifying the molecular mechanisms and factors that actually couple cell-cycle, DNA replication and meiotic DSB formation in plants will be an important task for future research.

Phosphorylated Mer2 attracts Mei4 and Rec114 (Fig. 3.5A). Rec114 and Mei4 have been first identified in two genetic suppressor screens and epistatic analysis suggested that they are needed together with Spo11. Furthermore mutations in either gene suppress genetic recombination (Malone *et al.*, 1991; Menees and Roeder, 1989). Later, it has been shown that Rec114, Mei4 and Mer2 associate with chromatin in the absence of other proteins essential for DSB formation, although Mei4 binding is reduced in *mer2Δ* (Li *et al.*, 2006). Rec114 is necessary for subsequent binding of Spo11 and Mre11 to future DSB regions in the genome and for Spo11 homo-dimer formation (Borde *et al.*, 2004; Prieler *et al.*, 2005; Sasanuma *et al.*, 2007). Interestingly, Rec114 over-expression suppresses DSB formation, suggesting a dual role for Rec114 (Bishop *et al.*, 1999), first as a scaffold protein of the DSB complex, second as a negative regulator of DSB formation. Rec114 shows sequence homology to the *Schizosaccharomyces pombe* Rec7 protein (Molnar *et al.*, 2001). Rec7 localizes to nuclei, associates with linear elements (LinEs, *S. pombe´s* rudimentary axial elements) of meiotic chromosomes and is required for DSB formation (Cervantes *et al.*, 2000; Davis and Smith, 2001; Lorenz *et al.*, 2006). Steiner et al. found that Rec7 interacts with Rec24, a *S. pombe* meiotic DSB protein, which is related to the *S. cerevisiae* Mei4 (Kumar *et al.*, 2010; Steiner *et al.*, 2010). Recently, the de Massy lab published the identification and characterization of the mouse orthologs of *Mei4* and *Rec114* (Kumar *et al.*, 2010). Murine *Mei4* and *Rec114* are expressed in testis and embryonic ovary and they interact with each other when expressed in HeLa cells. Cytological analysis showed that MEI4 is localized to the lateral elements of the synaptonemal complex, with the highest foci number in leptonema. MEI4 is not co-localizing with DMC1 and RPA and does not require SPO11 for localization. Greatly reduced γH2AX staining in *Mei4$^{-/-}$* mice meiocytes indicates a severe defect in DSB formation. Additionally, the *Mei4* knock-out mice are defective in homologous synapsis (Kumar *et al.*, 2010). Aligning the mouse *Mei4* and *Rec114* sequences with plant genomes, the already described *Arabidopsis AtPRD2* gene was recognized as the *Mei4* homolog and the already described *Arabidopsis AtPHS1* and maize *Zmphs1* genes as the *Rec114* homologs (De Muyt *et al.*, 2009; Pawlowski *et al.*, 2004; Ronceret *et al.*, 2009). The *PHS1* gene in maize and *Arabidopsis* is involved in pairing of homologous chromosomes. The maize mutants almost completely lack foci of the recombination protein RAD51 and at metaphase, maize *phs1* mutant alleles show univalents (Pawlowski *et al.*, 2004). These observations are indicative for a functional conservation of *Rec114*

homologs. Nevertheless, the authors of the study claim that DSB formation is not affected by mutations in the plant *PHS1* gene, but that *PHS1* is needed at the step of RAD51 nucleoprotein filament formation or RAD50 protein nuclear import, as broken DNA could be detected via TUNEL staining (Pawlowski *et al.*, 2004; Ronceret *et al.*, 2009). *Arabidopsis* plants lacking a functional *Mei4* homolog, the *Atprd2* mutants, have no vegetative growth defects, but show short siliques, and meiotic DSB formation is blocked. Cytological analysis of the male meiocytes shows univalents and no chiasma formation (De Muyt *et al.*, 2009). Further investigations are needed to characterize both, the plant Rec114 (PHS1) and Mei4 (PRD2) proteins and their impact on meiotic DSB formation.

Once Mer2, Mei4 and Rec114 are in place, they provide a platform for the binding of Spo11 (Fig. 3.5A). Spo11 is the catalytically active protein within the meiotic DSB complex (see below for details of the DNA cleavage reaction). Spo11, first described in *S. cerevisiae*, is related to the Top6A subunit of the archeal DNA topoisomerase type IIB (Bergerat *et al.*, 1997; Keeney *et al.*, 1997). This relationship (and other experiments discussed below) suggests that Spo11 needs to act as a dimer to catalyze DSB formation similar to type-II topoisomerases, by a transesterification reaction (Sasanuma *et al.*, 2007). The related type IIB topoisomerase consists of Top6A and Top6B subunits, but Top6B subunits have not been found in the genomes of eukaryotes with the exception of plants (Hartung *et al.*, 2002; Yin *et al.*, 2002). In plants, the Top6B homologs have no meiotic function (see below).
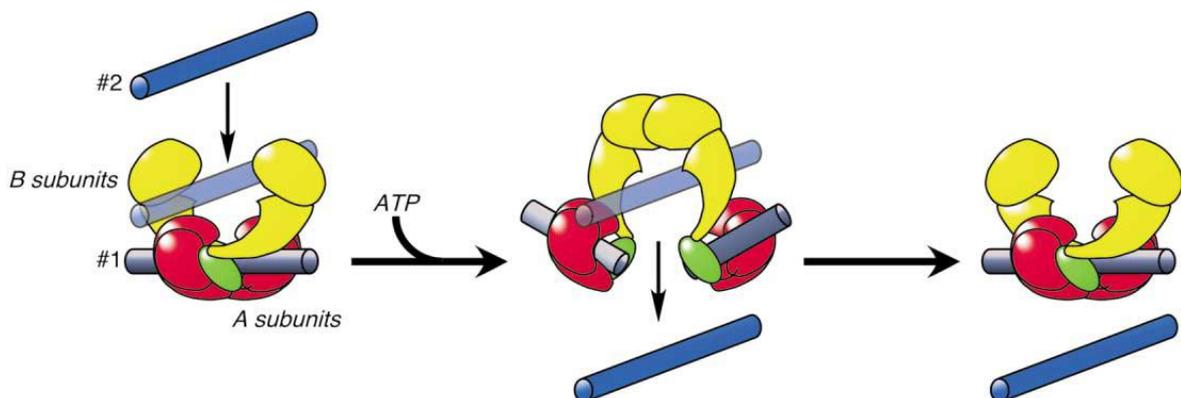
Figure 3.2: Proposed Reaction Mechanism for TopoVI. This model indicates how topoVI is thought to carry out DNA transport. Step 1: DNA segment #1 (gray) is bound by the topoVI A-subunits. Step 2: DNA segment #2 (blue) is trapped inside the enzyme upon closure of the ATP binding B subunit dimer. Concomitant with this capture, DNA segment #1 is cleaved and opened, and segment #2 is passed through the break. Step 3: DNA segment #1 is resealed and released, and the enzyme resets. Domains of the two subunits are colored as follows: the ATP binding B subunits are yellow, the A subunit CAP-like domains are green, and toprim domains are red. Figure and text taken from (Corbett and Berger, 2003).

This is consistent with the findings that the Top6B subunits provide ATPase activity, form a channel for DNA passage during the cleavage reaction and allow subsequent resealing of the original DNA strand at the break site. During meiosis, this kind of re-ligation is not desired, and the processing and repair of the SPO11 mediated DSBs is performed differently (see figure 3.2 and below) (Corbett *et al.*, 2007; Corbett and Berger, 2003, 2004).

Spo11 is conserved among eukaryotes (Malik *et al.*, 2007) and has been found, for instance, in *C. elegans* (Dernburg *et al.*, 1998), in *S. pombe* (Rec12) (Lin and Smith, 1994) and in *Drosophila melanogaster* (mei-W68) (McKim and Hayashi-Hagihara, 1998). It has been found in all plants analyzed so far. The *Arabidopsis* genome encodes, unlike the ones of mammals and yeast, where only one *Spo11* is present, three *Spo11* homologs, *AtSPO11-1*, *AtSPO11-2*, *AtSPO11-3* and additionally a *TOP6B* gene (Hartung and Puchta, 2000, 2001). Only *At*SPO11-1 and *At*SPO11-2 are essential for meiosis, whereas *At*SPO11-3 is needed for somatic endoreduplication and interacts with TOP6B (Grelon *et al.*, 2001; Hartung and Puchta, 2001; Stacey *et al.*, 2006). *AtSPO11-1* and *AtSPO11-2* are single copy genes and possess all five conserved motifs previously described for Spo11 proteins (sequence identity 26-41%) (Bergerat *et al.*, 1997). *Atspo11-1/Atspo11-2* double mutants do not differ from the single mutants, indicating that the two proteins work together (Stacey *et al.*, 2006). It is attractive to speculate that *At*SPO11-1 and *At*SPO11-2 may form an obligate hetero-dimer for meiotic DSB formation. This idea is supported by the finding that both, the active tyrosine of *At*SPO11-1 and of *At*SPO11-2 (*At*SPO11-1, Tyr 103; *At*SPO11-2, Tyr 124) need to be functional for successful DSB formation (Fig. 3.5B) (Hartung *et al.*, 2007).

In yeast, the interaction of Spo11 with Mer2/Rec114 requires the Ski8 protein (Fig. 3.5A). Ski8 has first been described as Rec103, found in a mutant screen overcoming the spore viability defects of a *rad52 spo13* haploid strain. Cloning and further characterization of Rec103 revealed that it is identical to Ski8 (Gardiner *et al.*, 1997). Ski8 seems to have a dual function, first in the mRNA decay pathway and second during meiosis (Arora *et al.*, 2004). Its multiple WD40 repeat structure (Evans *et al.*, 1997) allows multiple protein-protein interactions. It directly interacts with Spo11 and with Rec104, Rec114 and Mer2 as shown in yeast two-hybrid assays, apart from its non-meiosis-specific partners. Ski8 depends on Spo11 for nuclear entry and Ski8 stabilizes the association of Spo11 with meiotic chromosomes (Arora *et al.*, 2004). Interestingly, even though Ski8 is a conserved protein, functional differences have been observed. In *A. thaliana* Jolivet et al. identified a Ski8 homolog, also known as Vip3. They characterized two different allelic mutations (*vip3-2*, *vip3-3*). As in corresponding mutants in *S. pombe*, where the gene has been

named Rec14, or in *Sordaria macrospora*, these mutants grew poorly, but they displayed no meiotic defect (Jolivet *et al.*, 2006).

Rec102 and Rec104 were identified in a screen to isolate mutants defective in early steps of meiotic recombination (Malone *et al.*, 1991). They are meiosis specific, interact with and mutually depend on each other to localize to chromatin, suggesting that they act as a functional unit (Salem *et al.*, 1999). They are needed for localizing Spo11 to the nucleus and to chromatin and furthermore for Spo11 homo-dimer formation (Fig. 3.5A) (Kee *et al.*, 2004; Prieler *et al.*, 2005). No homologs of Rec102 and Rec104 have been identified outside of *S. cerevisiae* and closely related yeasts.

In the budding yeast three further proteins are essential for meiotic DSB formation (Fig. 3.5A). Mre11, Rad50 and Xrs2, forming the MRX complex, are conserved players in DNA repair. Mre11 was isolated in a screen for mutants with a defect in meiotic recombination (Ajimura *et al.*, 1993). Rad50 has first been described by Kupiec et al. and found to be needed for resistance to gamma-irradiation and methyl-methane-sulfonate (MMS) (Kupiec and Simchen, 1984). Xrs2 has first been described as a DNA repair gene and the study of Ivanov et al. showed that Xrs2 has also a meiotic function (Ivanov *et al.*, 1992). Homologs of Rad50, Mre11 and Xrs2/Nbs1 have been identified in all eukaryotes. In humans, plants, *S. pombe, Mus musculus and D. melanogaster* the interaction partner of MRE11 and RAD50 is NBS1, which displays only limited homology to Xrs2 (Akutsu *et al.*, 2007; Carney *et al.*, 1998; Ciapponi *et al.*, 2006; Ueno *et al.*, 2003; Vissinga *et al.*, 1999). In *A. thaliana* it was demonstrated that these proteins physically interact (Daoudal-Cotterell *et al.*, 2002; Gallego *et al.*, 2001; Waterworth *et al.*, 2007). A mutation in *AtRAD50* leads to meiotic defects, sterility and sensitivity against MMS (Bleuyard *et al.*, 2004; Gallego *et al.*, 2001). In vertebrates *Mre11* is an essential gene with roles in both, somatic and meiotic cells (D'Amours and Jackson, 2002), while in *A. thaliana MRE11* is non-essential. Depending on the *MRE11* mutant allele, plants are sterile, due to perturbed meiosis and severely affected in development or display only enhanced sensitivity to genotoxic agents (Bundock and Hooykaas, 2002; Puizina *et al.*, 2004). *Arabidopsis NBS1* mutant lines exhibit hypersensitivity to a DNA crosslinking reagent but no meiotic defects. Analysis of an *Arabidopsis nbs1-1/atm* double-mutant revealed a role for *At*NBS1 in meiotic recombination but no involvement in DSB formation. Interestingly, the requirement of the MRX complex for DSB formation is not universal. In other organisms than *S. cerevisiae* and *C. elegans,* like *A. thaliana, Tetrahymena* and *S. pombe,* the MRX complex is dispensable for meiotic DSB formation and only instrumental for meiotic DSB processing (Chin and Villeneuve, 2001; Lukaszewicz *et al.*, 2010; Puizina *et al.*, 2004; Young *et al.*, 2004).

Exhaustive genetic screens in *A. thaliana* identified novel genes that are essential for meiotic DSB formation but seem unrelated to known meiotic genes from other organisms. Primary screens for reduced fertility and secondary screens for the absence of meiotic DNA breaks identified the genes *AtPRD1*, *AtPRD2* and *AtPRD3* to be essential for SPO11 mediated meiotic DSB formation (Fig. 3.5B) (De Muyt *et al.*, 2009; De Muyt *et al.*, 2007). *AtPRD1* (*P*utative *R*ecombination initation *D*efect 1) has low similarity to mammalian *Mei1*. The mouse *Mei1* was characterized by Libby et al. and was isolated in a mutant screen for infertility (Libby *et al.*, 2002). The N-terminus of *At*PRD1 interacts with *At*SPO11-1 in a yeast two-hybrid assay, thereby identifying for the first time an interaction partner of a SPO11 protein in plants. The functional relevance of this interaction is unknown (De Muyt *et al.*, 2007). *AtPRD2* has later been recognized as a homolog of Mei4 (see above) (Kumar *et al.*, 2010). *At*PRD3 is a protein of unknown function, similar to the previously identified rice PAIR1 gene, but with no homologs outside the plant kingdom (De Muyt *et al.*, 2009; Nonomura *et al.*, 2004).

In different independent screens the plant SWI1 gene has been found to affect plant fertility. The data suggests that SWI1 is required for meiotic chromatin remodelling, sister chromatid cohesion, chromosome pairing, synapsis and recombination. No meiotic DSB breaks are formed in *swi1* mutants (Agashe *et al.*, 2002; Boateng *et al.*, 2008; Mercier *et al.*, 2003; Mercier *et al.*, 2001; Siddiqi *et al.*, 2000).

In the yeast *S. pombe* a couple of proteins needed for DSB have been identified that do not share sequence homology with proteins from *S. cerevisiae*, mammals or plants. Among them are Rec6, Rec15, Rec25, Rec27 and Mde2. All of these five proteins are very small and harbor no motifs to suggest any biological function. Rec6 and Rec15 are required for DSB formation (Cervantes *et al.*, 2000; Davis and Smith, 2001) and Rec15 interacts with Mde2 (Steiner *et al.*, 2010). Furthermore, Rec25 and Rec27 are important, but not absolutely essential for meiotic recombination in *S. pombe*. Both deletions exhibit a similar phenotype, namely aberrant asci with abnormal spore number and morphology, resulting from reduced meiotic recombination. Although no DSB formation has so far been observed in *rec25Δ* and *rec27Δ* mutants, recombination still occurs at low levels. Nevertheless, both Rec25 and Rec27 are absolutely necessary for Rec10 localization to LinEs and both co-localize with Rec10 during LinE formation (Davis *et al.*, 2008). Rec10, a homolog of the *S. cerevisiae* axial element protein Red1, is meiosis specific and localizes to LinEs (Lorenz *et al.*, 2004). In contrast to what has been observed in *S. cerevisiae* for Red1, a *rec10* deletion shows no DSB formation, whereas in Red1 mutants DSB formation still occurs, but is reduced to 20-60% of wild type levels (Hunter and Kleckner, 2001; Schwacha and Kleckner, 1997; Woltering *et al.*, 2000). A *rec10* deletion has the same recombination defect as a *rec12* mutant and the gene is indispensable for DSB

formation. So far no plant homolog of Red1 has been found. A dimerization partner of yeast Red1 is the HORMA domain protein Hop1 (de los Santos and Hollingsworth, 1999). The *C. elegans htp3* gene displays homology to yeast's *HOP1*, and is needed for DSB formation (Goodyer *et al.*, 2008). Deletion of the fission yeast Hop1 homolog reduces DSB frequency (Latypov *et al.*, 2010). In plants a gene displaying limited sequence similarity to Hop1 has been identified and named ASY1. ASY1 is one of the axial element proteins (like Hop1 in yeast), but has not been found to affect levels of meiotic DSB formation (Armstrong *et al.*, 2002; Caryl *et al.*, 2000; Sanchez-Moran *et al.*, 2008).

The *Drosophila* gene mei-P22 was isolated in a large scale P-element mutagenesis screen for mutants with a high frequency of X-chromosome non-disjunction in the female germline (McKim *et al.*, 1998). It is a 35.7 kDa protein, which cytologically localizes to meiotic chromosomes and it is necessary for the induction of DSB formation in the *Drosophila* female. Crossing over in such mutants can be restored to a level of 50% of wild type by gamma irradiation (Liu *et al.*, 2002).

3.4 Meiotic DNA DSB formation and processing

As outlined above, meiotic DSB formation is essential for subsequent recombination. Interestingly, an excess of breaks, compared to the number of reciprocal recombination products, is observed in most organisms analyzed. In the yeast *S. cerevisiae* it has been estimated that a single meiocyte forms approximately 140-170 (Buhler *et al.*, 2007) to 180-270 (Weiner and Kleckner, 1994) breaks but only about 90 to 95 COs (Chen *et al.*, 2008; Mancera *et al.*, 2008). Studies in *A. thaliana* established between 150 (Sanchez-Moran *et al.*, 2007) and 250 (Chelysheva *et al.*, 2007; Vignard *et al.*, 2007) breaks per meiocyte and about 10 COs (Higgins *et al.*, 2004; Mercier *et al.*, 2005; Wijeratne *et al.*, 2006). Interestingly, *C. elegans* seems to be an exception, with an estimation of about 12 breaks and 6 COs per meiocyte (Mets and Meyer, 2009).

As mentioned above, in *S. cerevisiae* Spo11 needs at least nine other proteins for catalysis of DSB formation (Fig. 3.5A; Mre11, Rad50, Xrs2, Rec102, Rec104, Rec114, Mei4, Mer2 and Ski8) (Keeney, 2001). The topology of the active cleavage complex has not been clarified yet. The meiotic chromatin is organized in loops and axes (Moens and Pearlman, 1988; Zickler and Kleckner, 1999), with cohesin molecules (Klein *et al.*, 1999) and meiosis specific axial element proteins (Smith and Roeder, 1997) forming the axis and chromatin loops emanating from there. The DNA sequences associated with cohesin were mapped with a resolution of ~ 1 kb (Blat and Kleckner, 1999; Glynn *et al.*, 2004). While the DSB proteins form foci on meiotic prophase chromatin, their localization has

only recently been analyzed by chromatin immuno-precipitation. Mre11 was found to localize to DSB hotspots in one study (Borde *et al.*, 2004) and equally strongly to hotspots and cohesin sites in another (Mendoza *et al.*, 2009). Spo11 was reported to bind to DSB hotspots and cohesin sites (Kugou *et al.*, 2009), while Mer2, Rec114 and Mei4 in fact avoid to bind to most strong DSB hotspots, while localizing to sites flanking the hotspots, usually coinciding with cohesin sites (Panizza and Klein, submitted). The latter observation supports a model in which Spo11 mediated cleavage occurs after loop sequences are transiently recruited to the DSB machine located at the chromosome axis (Panizza and Klein, submitted). This model is related to a series of models put forward first by (Zickler and Kleckner, 1999), who proposed that DSBs are made at the chromosome axis, then assuming that hotspot sequences are close to axis protein binding sites, which was later found not to be the case.

In plants, the distribution of DSB proteins on meiotic chromatin, and their interdependencies have not been analyzed yet. An *At*SPO11-1 specific antibody has been generated (Sanchez-Moran *et al.*, 2008), but so far it was only used to determine the massive enrichment of *At*SPO11-1 on meiotic chromatin in *Atmre11-3*, *Atrad50* and *Atcom1-1* mutants (Uanschou *et al.*, 2007), but not for precise localization of SPO11-1 and not for determining the dependencies of the meiotic cleavage complex in the context of chromatin and other protein factors.

A major component of meiotic chromosome axes are cohesins, which are better known for their role in sister chromatid cohesion (Orr-Weaver, 1999; van Heemst and Heyting, 2000) and which consist of four highly conserved proteins, namely Smc1, Smc3, Rec8 and Scc3. Scc1 is the kleisin of the mitotic cohesion complex and Rec8 is its meiotic paralog (Guacci *et al.*, 1997; Klein *et al.*, 1999; Michaelis *et al.*, 1997). In yeast, Hop1 and Red1 are two meiosis-specific axial element (AE) proteins, which are required for full levels of DSB formation and interhomolog-bias. The two proteins co-localize to AEs and interact as well in co-immunoprecipitation and in yeast two-hybrid experiments (de los Santos and Hollingsworth, 1999; Smith and Roeder, 1997). In the model plant *Arabidopsis*, 6 Scc1/Rec8 homologs have been identified with SYN1/DIF1 representing the ortholog of Rec8 (Bai et al., 1999; Bhatt et al., 1999). *syn1/dif1/rec8* mutants display defects in cohesion, chromosome condensation, and DNA repair but DSBs are still formed. ASY1, as described above, has been identified as a homolog of yeast's Hop1. To date, it is not clear whether plant cohesin and axial element proteins positively support meiotic DSB formation. It is unknown if less DSBs are formed in *Atrec8* and *asy1* mutants, respectively (Chelysheva et al., 2005).

Once the DSB complex has formed, Spo11 is anticipated to form a homo-dimer, in analogy to the defined structure of Top6A (Corbett and Berger, 2004; Nichols *et al.*, 1999). Furthermore, it is believed that Spo11 mediated DNA cleavage occurs via a transesterification mechanism, as described for type-IIB topoisomerases (Corbett and Berger, 2004).
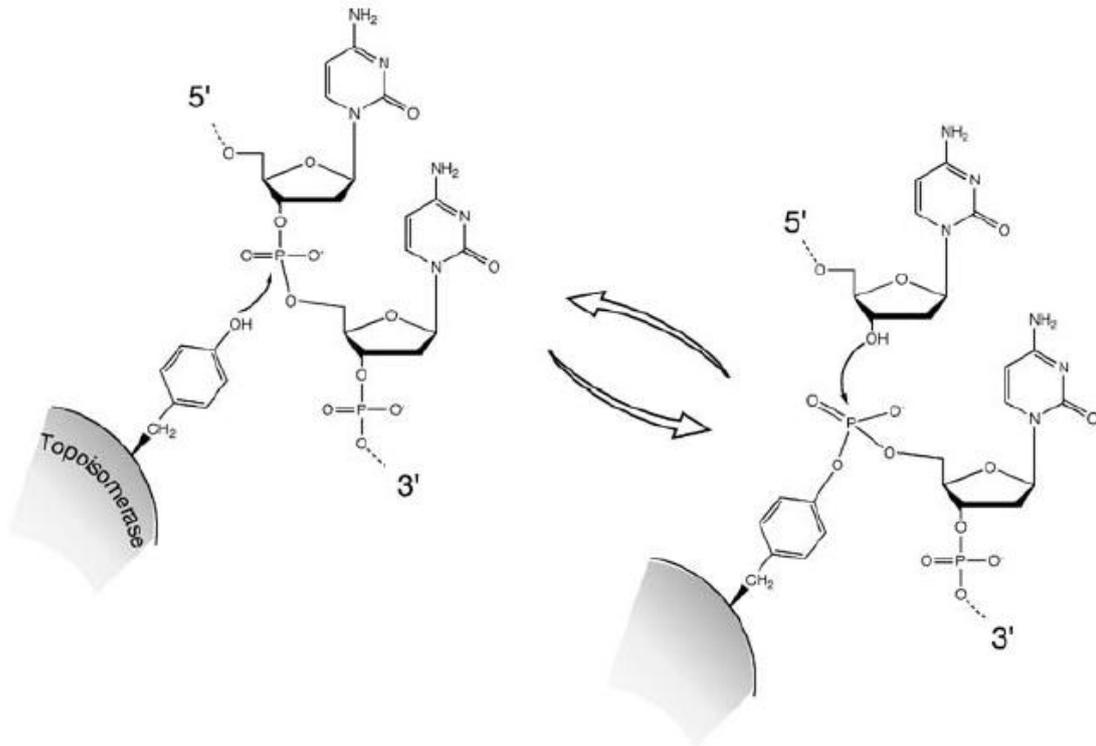
Figure 3.3: Chemistry of the DNA cleavage and religation reactions catalyzed by topoisomerases. A tyrosine side chain on the topoisomerase protein carries out a nucleophilic attack on the DNA phosphodiester backbone. This transesterase reaction severs the DNA backbone and covalently links the protein to the DNA end via a tyrosyl phosphodiester linkage. The protein-DNA linkage is reversed and the DNA is resealed by attack of the deoxyribose hydroxyl. In this example, the tyrosyl phosphodiester links the protein to the 5' end of the cleaved strand, as is seen in some type I and all type II topoisomerases. For a type II enzyme, two topoisomerase monomers work in concert to cleave the two strands of the duplex. Figure and text taken from (Keeney, 2007).

In yeast, it is the side chain of tyrosine 135 of Spo11 that carries out a nucleophilic attack on the DNA phosphodiester backbone. In the course of this reaction, the 5' phosphorus of the DNA becomes covalently linked to the tyrosine via a phosphodiester link, thereby generating a protein-DNA intermediate and a nick in the DNA strand. It is anticipated that the nucleophilic attacks occur simultaneously on both DNA strands thereby generating a DNA double strand break (see figure 3.3). As Top6A, the Spo11 DSB complex has been shown to generate breaks with a two-nucleotide 5' overhang (Liu *et al.*, 1995a). Whereas in the case of topoisomerases, the broken DNA ends are held together via the Top6B subunits and are resealed in a reversion of the transesterification described above (see

figure 3.2), the meiotic Spo11 containing DSB complex lacks analogs of Top6B subunits and the DSB is therefore differently processed and subsequently yields recombinogenic single stranded DNA strands (Corbett and Berger, 2004). An intermediate of the reaction represents Spo11 covalently attached to the 5' end of the DNA strand at the break site. Release of Spo11, attached to a short oligonucleotide derived from the DNA adjacent to the DSB site, determines the irreversibility of the cleavage reaction (see figures 3.3, 3.4 and 3.5C) (Neale *et al.*, 2005).
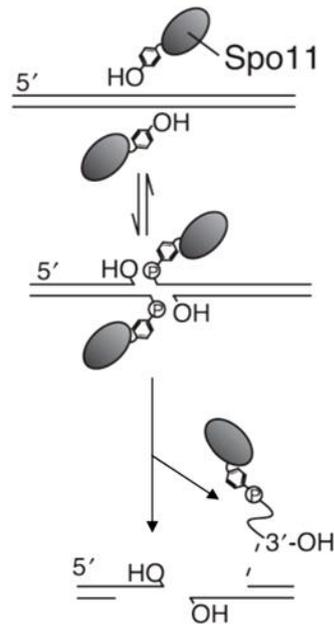
Figure 3.4: Endonucleolytic processing of covalent Spo11-DSB complexes. After DSB formation to initiate meiotic recombiantion, Spo11 remains covalently attached to 5' strand termini. In budding yeast DSBs are processed by endonucleolytic cleavage that releases Spo11 attached to an oligonucleotide with a free 3'-OH. Figure and text taken and modified from (Neale *et al.*, 2005).

The MRX complex together with Sae2/Com1 mediates the needed strand incision in a distance from the site of Spo11 activity (Neale *et al.*, 2005). To date it is not clear if the incision takes place close to the initial site of Spo11 activity, thereby generating the final size of the ssDNA oligonucleotide attached to Spo11 when released, or if the incision takes place further away from the break site with a subsequent need for 3'-5' exonuclease activity. Unpublished results from our lab indicate that the latter possibility is more likely to occur in yeast (Edlinger, Schlögelhofer, this thesis). In such a scenario, the endonucleolytic activity of MRX/Com1 would be exerted first in a distance. Starting from this nicked DNA site, the exonucleolytic 3'-5' activity of the MRX complex would work towards the Spo11 protein, and the 5'-3' exonucleolytic activity of Exo1 and other factors would be directed outwards away from the Spo11 mediated DSB. In *S. cerevisiae* Spo11

has been found with two different classes of oligonucleotides attached, approximately half with 7-12 nt in length and the other half between 21-37 nt in length (Neale *et al.*, 2005), whereas in *S. pombe* only one class of oligonucleotides could be detected (between 17-27 in length with an average of 23) (Milman *et al.*, 2009; Rothenberg *et al.*, 2009). This indicates that the meiotic DSB complex may be structurally different in various organisms, which is in line with the fact that the conservation of proteins involved in DSB formation is often low or absent (Bleuyard *et al.*, 2004; Jolivet *et al.*, 2006; Young *et al.*, 2004).

Following Spo11 mediated DSB formation and release of Spo11, with the help of MRX/Com1, 3' ssDNA becomes exposed. The length of the 3' ssDNA is determined by the activity of the resection machinery that exhibits exonucleolytic activity in the 5'-3' orientation and is mediated during meiosis most likely by Sgs1-Dna2 and Exo1 (Farah *et al.*, 2009; Mimitou and Symington, 2008; Zhu *et al.*, 2008). The resulting 3'ended ssDNA tails are believed to serve as probes to identify homologous partner chromosomes and to initiate D-loop formation and single end invasion followed by second end capture (Hunter *et al.*, 2001; Krogh and Symington, 2004; Mimitou and Symington, 2009b; Paques and Haber, 1999; San Filippo *et al.*, 2008).

Following the resection of DNA strands in budding yeast, Rad52 assembles the sequence-homology-dependent DSB-repair machinery (Gasior *et al.*, 1998). Rad52 orthologs are known in vertebrates (Van Dyck *et al.*, 1999) but not in *A. thaliana* (Ray and Langer, 2002). Either DSB repair is independent of Rad52 in *A. thaliana* or this protein has, thus far, not been detected in database searches. However, other protein factors involved in DNA strand exchange, such as proteins of the RecA-like recombinase protein family, have been identified in plants (Osakabe *et al.*, 2002). The completion of the *Arabidopsis* genome sequence revealed twelve genes with a conserved RecA domain. Five of them are closely related to the bacterial RecA protein (three of the five possess target sequences for mitochondria or chloroplasts). The other seven proteins have known homologs in man and have been analyzed in *A. thaliana*. Mutations in *AtRAD51*, *AtDMC1*, *AtXRCC3* and *AtRAD51C* lead to severe meiotic defects (Abe *et al.*, 2005; Bleuyard and White, 2004; Couteau *et al.*, 1999; Li *et al.*, 2004; Osakabe *et al.*, 2005). After DNA DSB formation and resection, Rad51 is loaded onto ssDNA. Rad51 plays a role in both, somatic and meiotic recombination, whereas the closely related strand exchange factor Dmc1 is exclusively loaded onto ssDNA during meiosis (Fig. 3.5C) (Dresser *et al.*, 1997; Paques and Haber, 1999; Sato *et al.*, 1995a; Sato *et al.*, 1995b; Sato *et al.*, 1995c) and is required together with other factors, for inter-homolog recombination (but not inter-sister recombination) (Schwacha and Kleckner, 1997) (reviewed in San Filippo *et al.*, 2008).

3.5 Methods to analyze sites of meiotic DSB formation

As outlined, meiotic DSB formation is essential for subsequent recombination to occur. Knowing the sites of meiotic DSB formation, allows analysis of determining factors and, in the future, may lead to novel plant breeding strategies by targeting meiotic recombination to desired loci in the genome of crop plants. Approximate locations of (a limited set of) DSB sites in the genome of a given organism can be deduced from the recombination products. This has been done extensively in various organisms including plants (Baudat and Nicolas, 1997; Drouaud *et al.*, 2006; Gerton *et al.*, 2000). With an increasing density of genetic markers, recombination maps in *Arabidopsis* now have the power to identify genetic exchange points in a window of about only 2 kb (Drouaud *et al.*, 2006). Recombination events can be monitored in the offspring (F1 generation) of two genetically distinct ecotypes by determining the exchange rate of known genetic markers. Historically, these markers have been phenotypic traits, but advances in molecular analysis and genome sequencing projects led to the discovery of a vast number of single nucleotide polymorphisms (SNPs) or small insertions or deletions (INDELS), of which many can be monitored simultaneously (Drouaud *et al.*, 2006).

The physical distances between the analyzed markers define the resolution of the recombination map. In 2007 the Weigel and Nordborg labs published a set of around one million non-redundant single nucleotide polymorphisms for different accessions of *Arabidopsis*. To examine sequence variation in this model plant, they performed high density array re-sequencing of about 20 different accessions (ecotypes). They observed that approximately 4% of the genome is highly dissimilar or deleted relative to the reference genome (Clark *et al.*, 2007). In a following publication, using a similar, but larger data set, historic recombination events have been deduced from regions with linkage disequilibrium. Furthermore, it has been demonstrated that the historic hot regions correlate well with recent recombination events (Kim *et al.*, 2007). In another study, an in depth analysis of recombination utilized about 100 recombinant inbred lines. Genomic DNA from these lines was hybridized to microarrays representing open reading frames of *A. thaliana* ecotype Columbia. The initial cross of the mapping population has been performed with the ecotypes Columbia and Landsberg, with the latter genome hybridizing to many probes of the microarray sequences with lesser affinity. This allowed the genome-wide differentiation of the two initial genomes and interrogation of initial recombination events (Singer *et al.*, 2006).

Another approach utilized a microarray-based readout for recombination in *S. cerevisiae*. The innovative aspect was that not diploid offspring but haploid cells, the products of

meiosis, were analyzed. Furthermore, since the four cells from one individual meiosis (tetrads) were analyzed, cross-over events, non-crossover events, gene conversion and crossover interference could be studied (Chen *et al.*, 2008; Mancera *et al.*, 2008).

Analyzing recombination directly in the haploid products of meiosis is as well possible in higher organisms, via sperm or pollen typing techniques (Arnheim *et al.*, 2007; Cui and Li, 1998; Jeffreys *et al.*, 2004; Kauppi *et al.*, 2007; Li *et al.*, 1988; Tiemann-Boege *et al.*, 2006) (Drouaud and Mezard, INRA Versaille, personal communication). Because of the high number of post-meiotic cells that can be studied, this technique allows a more efficient determination of meiotic recombination, compared to pedigree analysis.

The drawbacks of the approaches outlined above are that first, polymorphic markers have to be present and known in the organism of interest, second that recombination not necessarily reflects DSB initiation sites and third that the polymorphisms needed for such studies may influence DSB formation and recombination distribution and frequency. These methods have been instrumental for recombination analysis, but they have not provided high-resolution maps to identify the actual underlying DSB sites and omitted all those DSB sites not leading to exchange of genetic information. However, detailed information on DSB sites is required to identify underlying *cis-* and *trans-* determinants of meiotic DSB formation. Below, a range of methods is outlined that are dedicated to directly analyze and identify meiotic DSB sites throughout the genome.

Work performed in *S. cerevisiae* in the labs of Simchen and Nicolas (Baudat and Nicolas, 1997; Zenvirth *et al.*, 1992) revealed meiotic DSB sites on yeast chromosomes. Chromosomes and chromosome fragments from synchronized yeast cultures, containing a mutation that enriches for meiotic DSBs, *rad50S*, were separated by pulsed-field gelelectrophoresis and detected via Southern blotting. The *rad50S* hypermorphic mutation allows DSB formation to occur but subsequent processing steps are blocked. With this direct approach for DSB detection, 76 DSB regions have been identified. Furthermore, these experiments revealed the existence of cold and hot domains with respect to DSB formation and the quantitative differences of various hotspots. Most DSBs in *S. cerevisiae* were found in intergenic promoter-containing intervals and some of the hot DSB sites were known to be as well hot recombination sites.

A much more refined technique has been published in 2000 by the Petes lab (Gerton *et al.*, 2000). In this study, a tagged Spo11 protein was immuno-precipitated from synchronized *rad50S* yeast cultures. The intermediate of meiotic DSB formation, Spo11 covalently attached to 5' ends of DNA, was thereby enriched and allowed analysis of the bound DNA (Fig. 3.5C). The Spo11-associated DNA was fragmented, amplified by PCR and labeled. The DNA samples were then applied to microarrays, comprised of about

~6400 DNA sequences representing yeast open reading frames (DeRisi *et al.*, 1997). 177 hotspots of DSB formation and 40 coldspots were identified. In more detail, each chromosome has at least one hotspot of DSB formation and there is a significant correlation between chromosome size and hotspot number. Large chromosomes have relatively few hotspots/kb as compared to small chromosomes. The average distance between hotspots was determined to be 54 kb, for intervals including the centromere, about 117 kb (Gerton *et al.*, 2000; Lichten and Goldman, 1995). As already found (Sharp and Lloyd, 1993) for chromosome III, hot regions show a positive correlation with high GC content (Gerton *et al.*, 2000). Caveats concerning this technique to map meiotic DSB are as follows: first, microarrays may contain a biased set of probes, as in the case described above (e.g.: ORFs only). Only an unbiased microarray, using genomic probes, with equal spacing and, preferentially, overlap of the probes, will yield a high resolution map. For organisms with larger genomes and with high sequence redundancy the quality of microarray-based assays will always depend on the available microarray platform. Although custom-made arrays are available, the standard genomic arrays for *Arabidopsis* provide on average a probe of 25 nucleotides in length every 35 nucleotides. As discussed below, deep-sequencing of immuno-precipitated DNA will most likely substitute microarray based techniques for many applications. Second, DSB mapping in the *rad50S* background needs fragmentation of DNA prior to immuno-precipitation of DNA and the resolution of DSB maps therefore crucially depends on thorough fragmentation of genomic DNA. The average fragment size will define the broadness of the hybridization signal. Third, and most importantly, the DSB mapping approach outlined above depends on the *rad50S* mutant allele. Later studies demonstrated that in *S. cerevisiae* DSB formation is reduced in *rad50S* mutants (Blitzblau *et al.*, 2007; Buhler *et al.*, 2007).

Similar studies have subsequently been performed in *S. pombe* (Cervantes *et al.*, 2000; Cromie *et al.*, 2007; Ludin *et al.*, 2008) and DSB maps based on electrophoretic separation of DSB-generated fragments and on immuno-precipitation of tagged Rec12 (Spo11) were found to correlate. DSB sites are separated by at least 65 kb, mostly situated within large intergenic regions and underrepresented in coding DNA regions. The intervening regions undergo almost no breakage (Cromie *et al.*, 2007). This is in contrast to the situation in *S. cerevisiae*, with most of the DSBs in promoter regions and much higher density of DSBs over the genome. Later, the Smith lab mapped meiotic DSB sites in *S. pombe* wild-type cultures (Hyppa *et al.*, 2008). Intriguingly, the locations of DSBs were found to be indistinguishable in *rad50*[+] and in *rad50S* strains. However, the signal intensity was lower in the *rad50*[+] strains, most likely due to ongoing DNA repair. It should be noted, that Rad50 is not needed for DSB formation in *S. pombe*, even so the *rad50S* allele has the same defect of inhibiting post-DSB processing events. It may therefore be

assumed, that in organisms with no need for the MRX complex for DSB formation, *rad50S* and *mre11S* mutant alleles may be potent tools to enrich for SPO11-DNA intermediates at meiotic DSB sites and to faithfully represent the DSB landscape of these organisms.

In principle, the outlined approaches could be performed as well in plants. It should be emphasized that in higher plants RAD50 (and neither MRE11) is not needed for meiotic DSB formation, only for ensuing processing, together with MRE11 and COM1/SAE2 (Bundock and Hooykaas, 2002; Gallego *et al.*, 2001; Puizina *et al.*, 2004). Mutations in either of the three corresponding genes lead to accumulation of *At*SPO11-1 on meiotic chromatin (Uanschou *et al.*, 2007), suggesting conservation of DSB processing. The *Atcom1-1* mutation seems best suited to be used, as plants do not display any somatic aberrations during unchallenged life. In contrast, *mre11-3* and *rad50* mutants display pleiotropic somatic defects under normal growth conditions (Bundock and Hooykaas, 2002; Gallego *et al.*, 2001; Puizina *et al.*, 2004). Tagged and functional *At*SPO11 proteins have been generated (Edlinger, Schlögelhofer, this thesis) and efforts are currently ongoing to set up the experimental frame for genome wide DSB identification in a higher plant.

A different approach took advantage of another intermediate of meiotic DNA repair. After DSB formation, break processing generates a long stretch of ssDNA, which serves as a probe for finding a DNA template for repair. As briefly outlined, meiotic DNA repair depends on recombinases like Rad51 and Dmc1. In the yeast *S. cerevisiae*, the turnover of these stretches of ssDNA is blocked in a *dmc1Δ* mutant strain. Therefore ssDNA, generated at exactly the positions of a former DSB sites, becomes enriched in a *dmc1Δ* mutant and is amenable to biochemical analysis (Fig. 3.5C). Buhler et al., and in a similar approach Blitzblau et al., developed a technique to isolate ssDNA from synchronized *dmc1Δ* meiotic cell cultures (Blitzblau *et al.*, 2007; Buhler *et al.*, 2007). They used Benzoyl naphtoyl DEAE (BND) cellulose to enrich ssDNA tracts, amplified, labeled and hybridized it to the Agilent 44k, a yeast whole genome oligonucleotide array. It turned out that in *S. cerevisiae* the DSB landscape is more subtle than anticipated from the *rad50S* maps. For instance, regions close to the centromeres and telomeres, previously thought to be devoid of DSBs, were found to contain DSB sites, even so these regions have a very low recombination rate (Baudat and Nicolas, 1997; Borde *et al.*, 2004; Gerton *et al.*, 2000; Robine *et al.*, 2007). Importantly, all of the hot regions previously identified in the *rad50S* background were as well found in *dmc1Δ*, and in total about five times more DSB sites were found in the *dmc1Δ* background. Caveats of this technique are that different ssDNA-containing intermediates may not be equally stable. Furthermore, 5' to 3' end resection continues over time in *dmc1Δ* mutants (Bishop *et al.*, 1992; Shinohara *et al.*, 1992), and

early-forming DSBs might therefore be associated with more ssDNA than late-forming DSBs (Buhler *et al.*, 2007).

In principle, meiotic ssDNA could be isolated from any organism that provides sufficient amounts of staged meiotic cells. However, it should be emphasized that in plants loss of DMC1 does not lead to accumulation of ssDNA and not to a block in meiotic progression, but to RAD51 dependent DSB repair via the sister chromatid (Couteau *et al.*, 1999; Siaud *et al.*, 2004). So-called accessory proteins, like MND1 and HOP2 (AHP2 in *Arabidopsis*) support DMC1 and to a lesser extend RAD51 during meiotic DNA repair (Kerzendorfer *et al.*, 2006; Schommer *et al.*, 2003; Vignard *et al.*, 2007). Mutations in MND1, for instance, lead to unrepaired DSBs and might be used to enrich for ssDNA.

Another intermediate of meiotic recombination are the meiotic nucleoprotein filaments, ssDNA associated with the recombinases RAD51 and/or DMC1 (reviewed in Sehorn and Sung, 2004). Antibodies against RAD51/DMC1 are commercially available for many organisms and can be used to immuno-precipitate these nucleoprotein filaments and analyze the associated DNA (Fig. 3.5C). This has been done in the Dernburg lab for the model organism *C. elegans* (A. Dernburg, personal communication), in the Petukhova lab for mouse (G. Petukhova, personal communication) and in the Pawlowski lab for maize (W. Pawlowski, personal communication). The detailed and genome-wide analysis in mouse provides the first genome wide map of meiotic DSB formation and furthermore confirmed the existence of a target motif for Prdm9 (see above). In *C. elegans*, the Dernburg lab generated the first genome wide DSB map and furthermore identified a DNA sequence motif in the center of hotspot regions, indicative for a conservation of the Prdm9-like mechanism, first described in mouse. The analysis for maize is still ongoing, but it is anticipated that only these high-resolution techniques will generate DSB maps with a resolution and density that will allow analysis of putative DNA motifs in the center of DSB hotspots.

The most recent advances in DSB detection methods turned back to the core enzyme of meiotic DSB formation. As outlined, two Spo11 proteins become covalently linked to 5' ends of DNA at either side of the duplex DNA at a given DSB site. DSB processing releases these two Spo11 proteins with a short DNA oligonucleotide attached to the active tyrosine residue (Bergerat *et al.*, 1997; Corbett *et al.*, 2007; Corbett and Berger, 2004; Keeney *et al.*, 1997; Neale *et al.*, 2005). The short DNA oligonucleotides exactly represent the regions of meiotic DSB activity, and moreover, the 5' ends of these oligonucleotides represent the nucleotide of Spo11 activity. High througput sequencing of these Spo11-associated oligonucleotides would allow establishing a high-resolution map of meiotic DSB sites (Fig. 3.5C). The Keeney lab (S. Keeney, personal communication) and our lab (Edlinger, Schlögelhofer, this thesis) established protocols to ligate, with high efficiency,

adapters to Spo11-associated oligonucleotides after immuno-precipitating Spo11 from synchronized *S. cerevisiae* or *S. pombe* cultures. The amplification products are sequenced using deep-sequencing platforms and the mapped reads represent the meiotic DSB landscape with nucleotide resolution. Chromatin immuno-precipitation followed by deep sequencing (ChIP-Seq) was one of the first applications for next generation sequencing and the first results were published in 2007 (Barski *et al.*, 2007; Johnson *et al.*, 2007; Mikkelsen *et al.*, 2007; Robertson *et al.*, 2007). Compared to ChIP-on-chip (Chromatin immuno-precipitation with subsequent hybridization to a microarray), ChIP-Seq has a higher resolution, generates fewer artefacts and a better coverage. Additionally, for ChIP-Seq only a very low amount of input DNA is needed, it has a better dynamic range and multiplexing is possible. Importantly, the technique does not require prior knowledge of DSB sites and moreover, reduces the danger of biased results (Park, 2009). Efforts are currently under way to implement the method described above in the model plant *A. thaliana* (Edlinger, Schlögelhofer, this thesis).

Yet another possibility to study *cis* and *trans* requirements of meiotic DSB formation is the generation of artificial hotspots. Certainly such a strategy is of interest for applications in (plant-) breeding efforts by establishing novel genetic combinations, which would have been difficult to achieve with conventional breeding strategies. Most of the pioneering work in this field was done in the labs of Alain Nicolas and Kunihiro Ohta (Fukuda *et al.*, 2008; Pecina *et al.*, 2002; Robine *et al.*, 2007). They generated a Spo11 protein fused to the Gal4 DNA binding domain. Gal4 is a transcriptional activator in *S. cerevisiae* and binds to the so-called upstream activator sequence (UAS) through its N-terminal domain (Johnston, 1987; Pecina *et al.*, 2002). The Spo11-Gal4 fusion protein was expressed under the control of the constitutive ADH1 promoter in a *spo11Δ* strain and found to be functional. It was furthermore established that many original hotspots of meiotic DSB formation were still activated by Spo11-Gal4. In addition, the Spo11-Gal4 fusion protein targeted meiotic DSB formation to the Gal4 binding sites (UAS), found in promoters of *GAL2*, *GAL7* and *GAL1/10* (Pecina *et al.*, 2002).

In a follow-up study, Robine et al. (Robine *et al.*, 2007) analyzed the genome-wide redistribution of meiotic DSBs in *S. cerevisiae* strains expressing Spo11-Gal4. They found that Gal4BD-Spo11 mediates DSB formation only in a subset of its possible binding sites, indicating that the association of Spo11 with chromatin is not sufficient for DSB formation, supporting the view that additional factors locally control the cleavage. It was, for instance, not possible to induce DSB formation by Gal4BD-Spo11 in close proximity to centromeres. Artificially introduced DSBs by Gal4BD-Spo11, inhibited further DSB

formation in a region covering over 60 kb. In a third study, Fukuda and colleagues induced meiotic DSBs using Gal4BD-Spo11 or an endonuclease (VDE). VDE is a member of the homing endonuclease proteins and generates DSBs at the VDE-recognition sequence (VRS). The main findings were that Gal4BD-Spo11 more efficiently introduces DSBs in naturally hot than in cold regions and that VDE cleaves reliably at its recognition site irrespective of its location in hot or cold regions, suggesting that in principle a cold region is not inaccessible for DSB formation. Interaction studies showed that self-association of Spo11 occurred at UAS sites in hot regions but not in cold ones, suggesting that Spo11 remains in an inactive configuration in cold regions (Fukuda *et al.*, 2008). For the plant community the targeted stimulation of meiotic recombination has inherent potential of facilitating breeding of crop plants in the future. In addition, these artificial hotspots would represent a valuable tool to analyze the function of meiotic recombination.

Most parts of the introduction were taken and modified from (Edlinger and Schlogelhofer, 2011).

A   *Saccharomyces cerevisae*

B   *Arabidopsis thaliana*

C
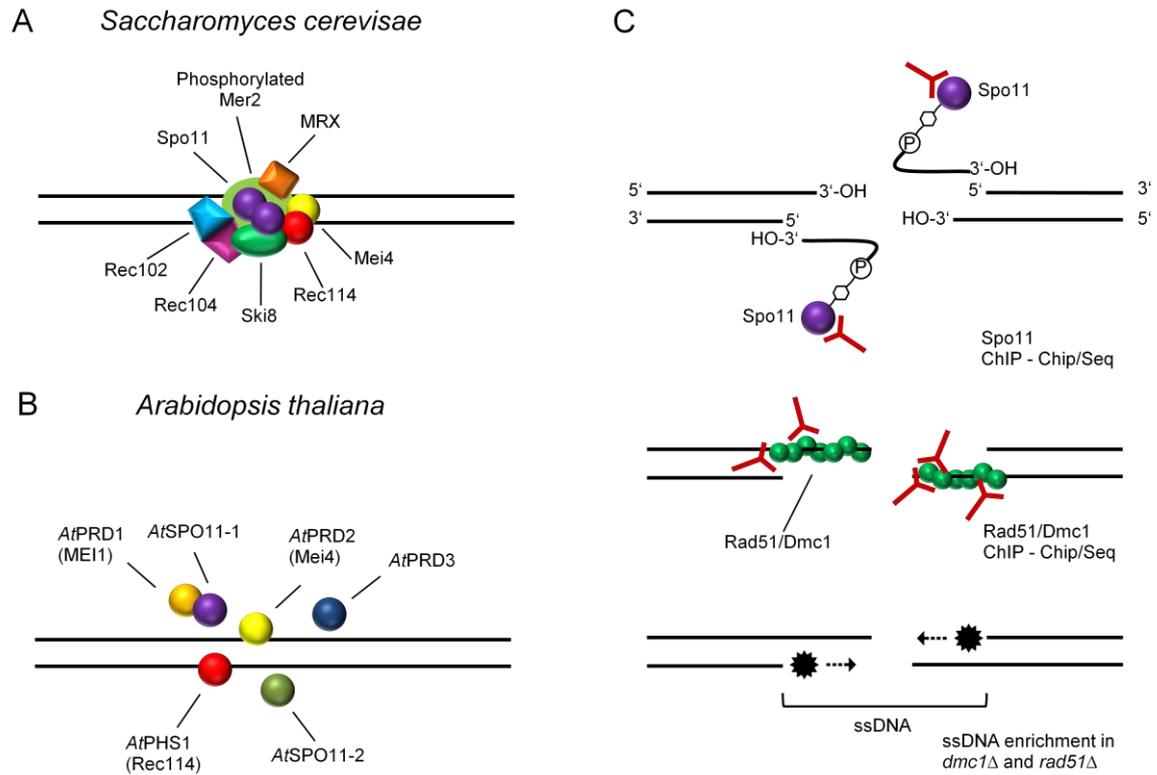
Figure 3.5: Proteins involved in meiotic DNA DSB formation (Related proteins are depicted in the same color). **A)** Schematic drawing of *S. cerevisiae* proteins essential for meiotic DSB formation. Please refer to the text for further explanations of spatial and temporal interactions of meiotic proteins. **B)** Schematic drawing of *A. thaliana* proteins known to be essential for meiotic DSB formation. The spatial distribution of proteins is hypothetical and not supported by experimental data. So far, only the interaction of SPO11-1 and PRD1 has been demonstrated (De Muyt *et al.*, 2007). **C)** Processing of SPO11-mediated DSBs. Different methods to identify meiotic DSBs are depicted next to the intermediates of meiotic recombination. Please refer to the text for further discussion. Figure and text taken and modified from (Edlinger and Schlogelhofer, 2011).

S. cerevisiae	S. pombe	A. thaliana	M. musculus	C. elegans	D. melanogaster	Function	References
Cdc28-Clb5* (CDK-S)						Kinase	(Matos et al., 2008; Sasanuma et al., 2008; Wan et al., 2008)
Cdc7-Dbf4* (DDK)						Kinase	(Matos et al., 2008; Sasanuma et al., 2008; Wan et al., 2008)
Mer2* (Rec107)						DSB formation	(Arora et al., 2004; Cool and Malone, 1992)
Ski8* (Rec103)	Rec14*	SKI8/VIP3				DSB formation	(Arora et al., 2004; Evans et al., 1997; Gardiner et al., 1997; Steiner et al., 2010; Tesse et al., 2003)
Rec102*						DSB formation	(De Muyt et al., 2009; Kee et al., 2004; Malone et al., 1991; Salem et al., 1999)
Rec104*						DSB formation	(Kee et al., 2004; Malone et al., 1991; Salem et al., 1999)
Rec114*	Rec7*	PHS1[?]				DSB formation	(Cervantes et al., 2000; Davis and Smith, 2001; Li et al., 2006; Malone et al., 1991; Prieler et al., 2005; Ronceret et al., 2009; Sasanuma et al., 2007; Steiner et al., 2010)
Mei4*	Rec24*	PRD2*	Mei4*			DSB formation	(De Muyt et al., 2009; Kumar et al., 2010; Li et al., 2006; Martin-Castellanos et al., 2005; Menees and Roeder, 1989; Prieler et al., 2005; Sasanuma et al., 2007; Steiner et al., 2010)
Spo11*	Rec12*	SPO11-1*/ SPO11-2*	Spo11*	SPO11*	mei-W68*	DSB formation	(Bergerat *et al.*, 1997; Dernburg *et al.*, 1998; Hartung and Puchta, 2000; Keeney *et al.*, 1997; Lin and Smith, 1994; McKim and Hayashi-Hagihara, 1998; Romanienko and Camerini-Otero, 1999; Steiner *et al.*)
Mre11*	Rad32	MRE11	Mre11	MRE11*	Mre11	DSB end processing	(Ajimura et al., 1993; Chin and Villeneuve, 2001; Ciapponi et al., 2004; Tavassoli et al., 1995; Xiao and Weaver, 1997)
Rad50*	Rad50	RAD50	Rad50	RAD50*	Rad50	DSB end processing	(Ciapponi et al., 2004; Colaiacovo et al., 2002; Gallego et al., 2001; Kupiec and Simchen, 1984; Luo et al., 1999)
Set1			Prdm9			Methyl-transferase	(Baudat et al., 2009; Mihola et al., 2009; Myers et al., 2009; Nislow et al., 1997; Parvanov et al., 2009)
	Rec6*					DSB formation	(Cervantes et al., 2000; Davis and Smith, 2001; Lin and Smith, 1994)

	Rec15*					DSB formation	(Cervantes *et al.*, 2000; Davis and Smith, 2001; Lin and Smith, 1994, 1995a; Steiner *et al.*, 2010)
	Rec25[?]					Linear Element component	(Davis et al., 2008; Martin-Castellanos et al., 2005)
	Rec27[?]					Linear Element component	(Davis et al., 2008; Martin-Castellanos et al., 2005)
	Mde2*					DSB formation	(Gregan et al., 2005; Steiner et al., 2010)
Red1	Rec10*					Axial Element, Linear Element protein	(DeVeaux and Smith, 1994; Lin and Smith, 1995b; Lorenz et al., 2004; Rockmill and Roeder, 1988)
		PRD1*	Mei1*			DSB formation	(De Muyt et al., 2007; Munroe et al., 2000)
		PRD3*				DSB formation	(De Muyt et al., 2009)
				mei-P22*		DSB formation	(Sekelsky et al., 1999)

Table 3.1: Compilation of proteins needed for meiotic DSB formation in various organisms and selected references for further reading. Proteins depicted with a "*" are essential for meiotic DSB formation. Proteins depicted with a "?" are suspected to be essential, but lack the experimental proof. Table taken from (Edlinger and Schlogelhofer, 2011).

4 Aim of PhD thesis

Meiotic recombination depends on the controlled formation of DNA DSBs mediated by Spo11, a homolog of the archaebacterial topoisomerase subunit Top6A (Bergerat *et al.*, 1997; Keeney *et al.*, 1997). This process is highly regulated and depends on the help of at least nine other proteins in budding yeast (Keeney, 2001; Prieler *et al.*, 2005; Smith and Nicolas, 1998). Meiotic recombination is initiated preferentially at certain sites in the genome, called hotspots. Hotspots of meiotic recombination have been shown to exist in yeast, plants, mice and humans (Nishant and Rao, 2006). Although many proteins involved in meiotic recombination are highly conserved, only five proteins have been identified to be essential for meiotic DSB formation in *Arabidopsis*. In addition to *At*SPO11-1 and *At*SPO11-2, *At*PRD1, *At*PRD2 and *At*PRD3 were identified to be essential for DSB formation (De Muyt *et al.*, 2009; De Muyt *et al.*, 2007; Grelon *et al.*, 2001; Stacey *et al.*, 2006). Remarkably, *Arabidopsis* is the only organism that possesses two meiotically active SPO11 proteins. Meiotic hotspots in plants have only been characterized genetically for chromosome IV of *A. thaliana* (Drouaud *et al.*, 2006), thereby firmly establishing the existence of meiotic hotspots, but the nature of these regions remains elusive.

This project aimed at the analysis of fundamental questions in plant meiosis, to learn more about the process of the initiation of meiotic recombination in the higher eukaryote *Arabidopsis thaliana*.

1) The first aim was to establish a method for the genome-wide identification of DSB sites in various organisms with the final aim to apply this technique for *A. thaliana*, to learn more about DSB distribution and regulation.

2) The second aim was the isolation of interacting partners of the two *Arabidopsis* SPO11 proteins using different biochemical approaches. We specifically wanted to know which further proteins are involved in DSB formation in *Arabidopsis*, how they interact in a complex and whether *At*SPO11-1 and *At*SPO11-2 form a heterodimer.

3) The third aim was the targeted stimulation of meiotic recombination (generation of artificial hotspots), with the inherent potential of facilitating breeding of crop plants in the future. These hotspots would represent a valuable tool to analyze the function of meiotic recombination.

5 Materials and Methods

Unless stated otherwise, all chemicals for growth media, buffers and solutions were purchased from Sigma, Roth, Fluka and Merck. The glass- and plastic-ware was purchased from VWR, Eppendorf and Sarstedt.

5.1 Growth Media

5.1.1 Bacterial Media

LB medium: 10 g/l peptone, 10 g/l NaCl, 5 g/l yeast extract, pH 7.4, autoclaved. For plates liquid LB was supplemented with 15 g/l Bacto Agar (Difco). If necessary, media were supplemented with antibiotics. Ampicilline [100 mg/l], kanamycine [50 mg/l], gentamycine [50 mg/l] (all Duchefa).

YEB medium: 5 g/l beef extract, 1 g/l yeast extract, 5 g/l peptone, 5 g/l sucrose, 2 mM MgSO$_4$. For plates liquid YEB medium was supplemented with 15 g/l Bacto Agar (Difco).

5.1.2 Plant Media

ARA plates: 4.3 g/l MS basal salt mixture, 0.5 g MES, 1 ml 1000x vitamine B5 stock, 6 g/l plant agar (all Duchefa), 10 g sucrose, pH 5.7, autoclaved. If necessary, media were supplemented with antibiotics. Hygromycin B (Roche), [25 mg/l], kanamycin (Duchefa), [25 mg/l].

5.1.3 Yeast Media

YPD (YPAD): 10 g/l yeast extract, 20 g/l peptone, 20 g/l dextrose, 40 g/l adenine sulphate, autoclaved.

YSD: 1.7 g/l yeast nitrogen base, 5 g/l potassium sulphate, 20 g/l glucose, up to 900 ml with dH$_2$O, pH 5.8, autoclaved. 100 ml sterile filtered 10x dropout mix was added after autoclaving. For the composition of 10x dropout mix see sections 1.1-1.7 in the yeast lab book of Nina Riehs (Riehs, 2006).

For plates all media were supplemented with 20 g agar per liter.

5.2 Plant Work

5.2.1 Cultivation of plants on soil or plates

After pre-germination in water for 2 days at 4°C without light, seeds were put on a soil (ED63, Einheitserde)/perlite (3:1) mixture and plant trays were covered with a lid until the four leaf status of the plants was obtained. Plants were grown under long day standard growth conditions (21°C, 16 hours light, 8 hours dark, 60% humidity).

Alternatively, for the selection of some transgenic markers like hygromycin and kanamycin, plants were grown on Ara plates (see section 5.1.2). Therefore seeds were sterilized using a 5% calcium hypochloride solution (5% $Ca(OCl)_2$, 0.02% Triton X-100, dH_2O). The seed sterilization solution was stirred at least for 1 hour, then centrifuged at 3000 rpm for 5 min and the supernatant was used for seed sterilization. The solution was added to approximately 100 µl of seeds, followed by 30 minutes rotation on a top-over wheel. After sedimentation seeds were washed three times with water in a sterile work bench. After two days of drying, seeds were applied to respective media plates under sterile conditions. Plates were sealed with Parafilm and put on 4°C without light for pre-germination. After two days, plates were transferred to a plant incubator and plants were grown under long day conditions (21°C, 16 hours light, 8 hours dark, 60% humidity). Prior to shooting, plants were transferred from plates on soil using a forcep.

5.2.2 Plant transformation

Arabidopsis plants were transformed with different transgenic constructs using a standard *Agrobacterium* mediated floral dip transformation. 3 ml LB medium (50 mg/l kanamycin, 50 mg/l gentamycin) were inoculated with a single colony of *A. tumefaciens* (strain GV3101), harboring the transgene of interest and grown at 28°C over night. On the following day 500 ml LB supplemented with antibiotics were inoculated with the pre-culture and grown at 28°C over night under vigorously shaking conditions. For the pGreenII0179 compatible *Agrobacterium* strain C58C9, bacteria were cultivated in YEB medium, containing rifampicine [50 mg/l] and tetracycline [12 mg/l], as described above.

After harvesting the cells (25 min, 5000 rpm), they were washed with infiltration medium (5% sucrose, dH_2O), centrifuged again and then resuspended in 200 ml infiltration medium supplemented with 40 µl Vac-In-Stuff (Silwet L-77, Lehle seeds).
After cutting the mature siliques, plants were dipped into the infiltration solution for 30 sec and then put in to plastic boxes covered with saran wrap. After two days, plants were taken out from the boxes and subjected to standard growth conditions.

5.2.3   Selection of transgenic plants

After transformation seeds were harvested and applied to soil or plates (T0) for the selection of positive transformants using 150 mg/l BASTA® (Glufosinate Ammonium Salt, Bayer Crop Science), 25 mg/l hygromycin B (Roche) or 25 mg/l kanamycin (Duchefa). Transformants were brought to the next generation (T1) and analyzed for a 3:1 Mendelian segregation, indicative for a heterozygous single locus insertion of the transgene of interest.

5.2.4   Crossing of plants

From the mother plants all mature siliques and open flowers were removed using a forcep. Under a binocular the meristem with the small buds was removed, resulting in 3-5 remaining buds per inflorescence which were used for crossing. These buds were opened with a forcep and after removal of petals, sepals and anthers, stigmata were pollinated using anthers covered with pollen from the father plants. Silique outgrowth was indicative for successful crossing.

5.2.5   Tetrad analysis of fluorescent pollen

Tetrad analysis was performed according to the protocol published by the inventers of this technique (Francis et al., 2007). I learned this technique directly in the lab of Gregory P. Copenhaver (University of North Carolina at Chapel Hill) during a two weeks visit in June 2008. Pollen screening was performed with an Axioskop Epifluorescence Microscope (Zeiss) using a tripleband filter CFP/YFP/DsRed (Tripleband Excitation Filter 422/503/572nm, Tripleband Beam Splitter 444/520/590nm, Tripleband Emission Filter 465/537nm/623nm, AHF Analysentechnik).

5.3   DNA Work

5.3.1   Genomic DNA preparation from *Arabidopsis thaliana*

For fast DNA isolation 1-3 small leafs were collected in 1.5 ml reaction tubes, 200 µl plant DNA extraction buffer (200 mM Tris HCl, pH 7.5, 250 mM NaCl, 0.5% SDS, 25 mM EDTA, pH 8) was added and leafs were disrupted using a small plastic pistil. After the addition of another 200 µl DNA extraction buffer, samples were centrifuged for 5 min at 13000 rpm. The supernatant was transferred to a new reaction tube and one volume of isopropanol was added. After light shaking and incubation for 5-10 min at room temperature, samples

were centrifuged again for 5 min at 13000 rpm. The supernatant was discarded and the pellet was washed with 70% ethanol. The supernatant was discarded, the pellet was dried in a heat-block and finally dissolved in 50 µl 1xTE (10 mM Tris HCl, pH 8, 1 mM EDTA, pH 8).

5.3.2 Small scale plasmid DNA preparation from *Escherichia coli*

3 ml LB medium supplemented with antibiotics were inoculated with a single colony of transformed cells and grown overnight at 37°C with shaking at 190 rpm. The culture was transferred to a 1.5 ml reaction tube and cells were harvested by centrifugation at 13000 rpm for 1 min. The pellet was resuspended in 200 µl solution I (50 mM glucose, 10 mM EDTA, pH 8, 25 mM Tris HCl, pH 7.5), then mixed with 200 µl solution II (0.2 N NaOH, 1% SDS) followed by the addition of 200 µl solution III (3 M potassium acetate/5 M acetic acid) and inverting several times. After centrifugation of the lysate at 13000 rpm at room temperature for 5 min the supernatant was transferred into a new sterile 1.5 ml reaction tube and thoroughly mixed with 600 µl isopropanol. After 15 min of centrifugation at 13000 rpm at room temperature, the supernatant was discarded and the pellet was washed with 70% ethanol. After drying the pellet, plasmid DNA was redissolved in dH$_2$O containing RNAse A [0.1 µg/µl].

5.3.3 Small scale plasmid DNA preparation from *Agrobacterium tumefaciens*

Agrobacterium cultures were transferred to 1.5 ml reaction tubes and cells were harvested by centrifugation at full speed for 1 min. The pellet was resuspended in 150 µl solution I (for 1 ml: 880 µl 1xTE, 100 µl proteinase K [10 mg/ml], 20 µl 25% n-lauroyl-sarkosyl), then incubated for 1 hour at 37°C without shaking. After addition of 200 µl solution II (0.2 N NaOH, 1% SDS) and 5 min incubation on ice, 150 µl solution III (3 M potassium acetate/5 M acetic acid) were added followed by a 15 min incubation on ice. After centrifugation of the lysate at 13000 rpm at 4°C for 5 min, the supernatant was transferred into a new sterile 1.5 ml reaction tube and thoroughly mixed with 1 ml 96% ethanol. After 15 min of centrifugation at 13000 rpm at 4°C, the supernatant was discarded and the pellet was washed with 70% ethanol. After drying the pellet, plasmid DNA was redissolved in 40 µl 1xTE (100 mM Tris HCl, pH 7.5, 10 mM EDTA, pH 8).

5.3.4 Precipitation of DNA

To a given volume of DNA sample, 1/10 volume of 3 M sodium acetate, pH 4.7 and 2.5 volumes of 100% ethanol (-20°C) were added and the mixture was kept at -20°C in a freezer for 30 min or on ice for 5 min followed by centrifugation at 4°C for 30 min at

13000 rpm. The supernatant was discarded and the pellet was washed with 400 µl 70% ethanol. After centrifugation at 4°C for 5 min at 13000 rpm the pellet was vacuum dried and redissolved in dH$_2$O.

5.3.5 Preparation of BAC DNA

Preparation of BAC DNA was performed using the Qiagen Plasmid Midi Kit according to the manufacturer's instructions.

5.3.6 Transformation of *Escherichia coli*

Chemically competent *E. coli* cells (XL-1 Blue, DH5α or Top10F') were prepared as described in Molecular cloning: A laboratory manual (Sambrook and Russell, 2001). For transformation, cell aliquots were thawn on ice, mixed with up to 1 µg of plasmid DNA and incubated for 30 min on ice. After 90 sec of heat shock at 42°C, cells were immediately put back on ice followed by the addition of 1 ml of prewarmed LB medium. After one hour incubation with shaking at 37°C, cells were collected by 30 sec centrifugation at 13000 rpm. 90% of the supernatant were discarded, cells were re-suspended in the remaining liquid and spread on LB plates, containing respective antibiotics, under sterile conditions. Plates were incubated over night at 37°C.

5.3.7 Transformation of *Agrobacterium tumefaciens*

Electro competent cells of the strain GV3101 were transformed as follows. Electroporation was performed with a Micro Pulser (Bio-Rad) according to the manufacturer's instructions (400 Ω, 25 µF, 2.5 kV for 0.2 cm cuvettes). After re-thawing on ice, cells were mixed with up to 1 µg plasmid DNA and the mixture was transferred to a pre-chilled 0.2 cm cuvette (Bio-Rad). The cuvette was shaken slightly in order to settle the cells to the bottom, wiped dry and placed into the sample chamber. After pulsing, 1 ml pre-warmed sterile LB medium was immediately added followed by a incubation at room temperature without shaking for 60 min. Aliquots were spread on LB plates containing respective antibiotics, followed by a two to three days incubation at room temperature or 28°C.

The chemically competent cells of the strain C58C9 were transformed as follows. 200 µl competent cells were re-thawn on ice, mixed with 2-5 µg plasmid DNA and kept in liquid nitrogen for 1 minute. After 5 minutes incubation at 37°C, 800 µl YEB medium containing rifampicine [50 mg/l] and tetracycline [12 mg/l] was added and cells were incubated at 28°C for 2 hours with shaking. 100 µl of the cell suspension was spread on a YEB plate with antibiotics and incubated for up to 5 days at RT or 28°C.

5.3.8 Polymerase chain reaction (PCR)

PCR was used for several experimental approaches like cloning, determination of genotypes and ChIP DNA amplification. Standard genotyping PCRs were performed in a volume of 20 µl, containing 1 µl forward primer [10 µM], 1 µl reverse primer [10 µM], 2 µl 2 mM dNTPs, 2 µl 10x Dream Taq buffer (including 25 mM $MgCl_2$, Fermentas), 0.2 µl Dream Taq polymerase (Fermentas), 12.8 µl dH_2O and 1 µl template DNA.

All detailed PCR compositions and conditions can be found in the respective sections. A list of all primers used in this work can be found in the appendix.

5.3.9 *Atspo11-1-2* genotype determination

The *Atspo11-1-2* allele (Grelon *et al.*, 2001) is an EMS allele which produces an additional *Vsp*I restriction site which can be used as a CAPS marker. For genotype determination, 1 µl genomic DNA of the candidate plants was subjected to PCR using primer spo11-1-2-MG52 and spo11-1-2-MG96 and the following PCR program: 94°Cx30 sec, 30 cycles (94°Cx20 sec, 60°Cx20 sec, 72°Cx1 min), 72°Cx5 min. The PCR product has a size of 955 bp and in the presence of the *Atspo11-1-2* mutant allele the product can be digested with *Vsp*I. For the VspI CAPS digest, 2.5 µl buffer O^+, 2 µl dH_2O and 0.5 µl *Vsp*I were added to 20 µl PCR product. The reaction mixture was incubated for at least 2 hours at 37°C and the samples were analyzed using standard agarose gel electrophoresis. Wild-type plants, which are lacking the CAPS marker, result in an uncleaved 955 bp amplicon, homozygous plants yield two co-migrating 471 and 484 bp fragments. Heterozygous plants were identified by the presence of all three fragments.

5.3.10 *Atspo11-2-3* genotype determination

The *Atspo11-2-3* mutant (GK 749C12) is a T-DNA insertion line (Hartung *et al.*, 2007). The determination of the genotype was performed using two allele specific primers (Spo11-2_GABI_down, Spo11-2_GABI_up) and one T-DNA specific (GABItest1) primer and the following program: 94°Cx15 sec, 40 cycles (94°Cx15 sec, 55°Cx15 sec, 72°Cx90 sec), 72°Cx5 min. The wild-type band has a size of ~1000 bp, the mutant allele results in three different bands with sizes of ~500, 750 and 1400 bp.

For a more reliable determination of the mutant allele, the two additional primers GABI LB BE and spo11-2 gk genomic were designed. The PCR conditions for this primer pair are as follows: 94°Cx30 sec, 40 cycles (94°Cx15 sec, 55°Cx15 sec, 72°Cx90 sec), 72°Cx5 min, resulting in an amplicon with a size of 840 bp.

5.3.11 TAIL (thermal asymmetric interlaced) PCR

For the characterization of transgene/insert junctions, TAIL (thermal asymmetric interlaced) PCR was used which is described in (Liu *et al.*, 1995b). A simplified version of the protocol, developed by Tuncay Baubec and Ortrun Mittelsten Scheid (GMI, Vienna) was applied.

Two sets of three specific primers for the T-DNA, 200 bp, 150 bp and 50 bp from the left border and oriented outwards, were designed. These are primer 5xuas tail A, B, C and 5xuas tail lb1, 2, 3. As a degenerated primer, AD7 (different from AD1-3 in the reference), with the sequence AWGCANGNCWGANATA (W = A+T, N=A+C+G+T) was used.

For the first PCR, 12 µl dH$_2$O were mixed with 2 µl 10x High Fidelity Enzyme Mix buffer (Fermentas), 2 µl 2 mM dNTPs, 1 µl 4 µM primer A, 1 µl 80 µM primer AD7, 1 µl DNA and 1 µl High Fidelity Enzyme Mix (5U/µl, Fermentas). Wild-type DNA and water were used as negative control and no template control.

The first PCR program was as follows:

Repeats	Temp 1	Time 1	Temp 2	Time 2	Temp 3	Time 3
1 x	95°C	2 min				
5 x	94°C	30 sec	62°C	1 min	72°C	2.5 min
1 x	94°C	30 sec	25°C →72°C	3 min 3 min	72°C	2.5 min
15 x 3 alternating cycles	94°C	10 sec	68°C	1 min	72°C	2.5 min
	94°C	10 sec	68°C	1 min	72°C	2.5 min
	94°C	10 sec	44°C	1 min	72°C	2.5 min
1 x	72°C	5 min				
pause	4°C					

The PCR products of the 1st PCR were diluted 1:50 in dH$_2$O and 1 µl was used as template for the 2nd PCR. Reactions were mixed as described for the 1st PCR, but instead of primer A, primer B was used.

The second PCR program was as follows:

Repeats	Temp 1	Time 1	Temp 2	Time 2	Temp 3	Time 3
12 x 3 alternating cycles	94°C	10 sec	64°C	1 min	72°C	2.5 min
	94°C	10 sec	64°C	1 min	72°C	2.5 min
	94°C	10 sec	44°C	1 min	72°C	2.5 min
1 x	72°C	5 min				
pause	4°C					

The PCR products of the second PCR were diluted 1:10 in dH$_2$O and 1 µl was used as template for the 3rd PCR. Reactions were mixed as described for the 1st PCR, but instead of primer A, primer C was used.

The third PCR program was as follows:

Repeats	Temp 1	Time 1	Temp 2	Time 2	Temp 3	Time 3
20 x	94°C	15 sec	44°C	1 min	72°C	2.5 min
1 x	72°C	5 min				
pause	4°C					

PCR products from all rounds were analyzed by standard agarose gel electrophoresis. While the primary PCR products usually have multiple, faint and unspecific bands, the secondary and tertiary PCR products should have distinct bands, with the size difference expected due to the position of the primer within the transgene. Products of choice were extracted from the gel and sequenced as described in materials and methods using the transgene specific primer closest to the left border. Transgenic insertion sites were identified by blasting the obtained sequence reads to the *Arabidopsis* genome.

5.3.12 PCR for the determination of transgene presence

For the presence of transgenes in different plant lines, standard PCR with specific primers for every line was applied. For details see table below.

Transgene	Primer 1	Primer 2	Amplicon length (bp)	Cycles, annealing temp
AtSPO11-1 wt	pNOS dn	bar_fw	770	30, 50°C
AtSPO11-1/18xmyc	gATSPOup	c-myc_new_dn	1198	30, 58°C
AtSPO11-1/24xHA	spo11-1 24xha forw	spo11-1 24xha rev	745	20, 55°C
AtSPO11-1/Gal4	gATSPOup	AtGal4dn	894	30, 58°C
AtSPO11-1/SF	gATSPOup	SF_forw	588	30, 58°C
AtSPO11-2 wt	pNOS dn	pNOS up	850	20, 60°C
AtSPO11-2/18xmyc	spo11-2 insert rev	spo11-2 18xmyc forw	1127	20, 55°C
AtSPO11-2/24xHA	spo11-2 insert rev	spo11-2 24xha forw	1127	20, 55°C
AtSPO11-2/Gal4	spo11-2 insert rev	spo11-2 gal4 forw	1219	20, 55°C

5.3.13 DNA sequencing

For standard sequencing reactions 300-400 ng plasmid DNA were mixed with 2 µl 2 µM sequencing primer, 1.5 µl 2.5x sequencing buffer, 1 µl Terminator Ready Reaction Mix (both part of Big Dye Terminator Cycle Sequencing Kit, Applied Biosystems) and filled up to 10 µl with dH_2O. The PCR program for the sequencing reaction was comprised of 2 min at 96°C initial denaturation followed by 25 cycles of 96°Cx30 sec, 45°Cx15 sec, 60°Cx4 min. The reaction products were precipitated as described in section 5.3.4.

Sequencing readout was performed with an ABI Prism 377 DNA Sequencer (Applied Biosystems) at the Institute of Botany, University of Vienna.

5.3.14 DNA agarose gel electrophoresis

Standard DNA agarose gel-electrophoresis was performed as described in (Sambrook and Russell, 2001). DNA samples were mixed with 6x Orange G loading dye (2.5% Ficoll 400, 11 mM EDTA, 3.3 mM Tris HCl, 0,15% Orange G, pH 8) to a final 1x concentration of the dye. To determine size and concentration of DNA fragments, two DNA markers were used, namely Gene Ruler^{TM} 1kb DNA ladder and Gene Ruler^{TM} Ultra Low Range DNA ladder (both Fermentas).

5.3.15 Restriction digest

Restriction digests were performed in a volume of 20 µl containing 1 - 5 µg plasmid DNA, 0.5 µl restriction enzyme (10 Units), 0.5 µl of a second restriction enzyme (if required), 2 µl of the appopriate 10 x buffer, and 2 µl 10 mg/ml BSA (if required). Unless stated otherwise, the reaction mix was incubated at 37°C for at least 2 hours.

5.3.16 DNA ligation

Standard ligation reactions for cloning purposes were performed in a volume of up to 20 µl, containing ~100 ng vector DNA, a 3-4 fold molecular excess of insert DNA, 1x T4 ligation buffer and 1 µl of T4 DNA ligase (Fermentas or Roche, 1U/µl). Ligations were performed either at RT for 2 hours or preferentially over night at 16°C. The reactions were heat-inactivated for 10 min at 65°C prior to transformation. For T/A cloning into the vector pCR®2.1 (Invitrogen) only 12.5 ng vector DNA were used.

5.3.17 Gel extraction

DNA fragments were cut out from agarose gels under UV light with a clean scalpel and after freezing and re-thawing, DNA was extracted using the E.Z.N.A.® gel extraction kit (Omega Bio-Tek) according to the manufacturer's instructions.

5.4   Protein Work

5.4.1   Preparation of *A. thaliana* cell extracts

For the preparation of *Arabidopsis* cell extracts, buds were collected, immediately frozen in liquid nitrogen and stored at -80°C. For protein extraction, 1 ml of buds was transferred to a Multi-beads shocker tube together with 400 µl 2x Ferrando extraction buffer (50 mM Tris HCl, pH 7.5, 10% glycerol, 0.5% NP-40, 1 mM EDTA, 1 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine, 2 µg/mL pepstatin, 0.5 µg/mL aprotinin, 0.5 µg/mL leupeptin, and 0.5 µg/mL antipain) and a steel bead (3mm diameter). Cells were disrupted in a pre-cooled Multi-beads shocker® (Yasui Kikai) at 2500 rpm for 600 seconds (30" on, 30" off). To get rid of cell debris, samples were centrifuged for 10 min at full speed at 4°C. The cleared supernatant was centrifuged for a second time and then used for downstream applications.

5.4.2   Preparation of *S. cerevisiae* and *S. pombe* cell extracts

The preparation of *S. cereviasie* and *S. pombe* cell extracts is described in detail in section 5.5.2.

5.4.3   Immuno-precipitation (IP)

Unless stated otherwise, protein G sepaharose™ 4 Fast Flow (GE Healthcare) was used for standard immuno-precipitation experiments. 20 µl beads/sample were washed 3x with NET2 buffer (50 mM Tris HCl, pH 7.5, 150 mM NaCl, 0.05% NP-40) and then incubated over night with respective antibodies at 4°C. On the next day, beads were washed again 3x with NET2 buffer, the protein extract was added and IP was performed on a top-over wheel at 4°C for 3-4 hours. After the IP beads were washed 3x with NET2 buffer, proteins were eluted by adding 20-30 µl 2xLämmli buffer and heating for 5 min at 94°C. For further analysis protein eluates were subjected to SDS PAGE and/or Western blot analysis.

5.4.4 SDS-PAGE

SDS-PAGE is a method for the separation of proteins on account of their molecular weights. The proteins become denatured and negatively charged while cooking with SDS. Therefore proteins are separated predominantly on account of their molecular weight and not of their own charge. SDS-PAGE was performed using a standard protocol with a 4.5% stacking gel and a 10% separating gel. For one separating gel 2.325 ml dH_2O, 1.275 ml 40% acrylamide/bisacrylamide mix, 1.3 ml 1.5 M Tris HCl, pH 8.8, 50 µl 10% SDS, 50 µl 10% ammonium persulfate and 3 µl TEMED were mixed. For polymerization the gel was overlaid with isopropanol. After polymerization, the isopropanol was removed and the stacking gel was casted. For 2 ml stacking gel 1.48 ml dH_2O were mixed with 0.25 ml 40% acrylamide/bisacrylamide mix, 0.25 ml 1.5 M Tris HCl, pH 6.8, 20 µl 10% SDS, 20 µl 10% ammonium persulfate and 2 µl TEMED. After polymerization slots were rinsed with dH_2O, protein samples were mixed with 2x Lämmli buffer and denatured at 94°C for 5 min. The samples were loaded on the gel and separated electrophoretically (20 mA/gel) in a Mini Protean® chamber (Biorad) using 1x PAGE buffer (5x PAGE, pH 8.6: 15.1 g Tris base, 72 g glycine, 50 ml 10% SDS). As a size marker 5 µl of PageRuler^{TM} Prestained Protein ladder (Fermentas) was loaded.

5.4.5 Western blot analysis

For Western blot analysis the transfer cassette was assembled as follows: black side down, foam sponge, filter paper, protein gel, membrane (PVDF, Millipore, equilibrated in methanol), filter paper, foam sponge, white side up). The transfer was performed in a standard blotting tank (Bio Rad) using transfer buffer (25 mM Tris base, 192 mM glycine, 20% methanol) at 400 mA for one hour. After the transfer, the sandwich was disassembled, the membrane was briefly washed in 1x TBST (150 mM NaCl, 10 mM Tris HCl, pH 7.5, 0.05% Tween-20) and blocked for 1 hour at RT in 1x TBST+3% non-fat dry milk, followed by washing 3x 5 min with 1x TBST. For detection of the transferred proteins, blots were incubated with respective primary antibodies diluted in 1x TBST+3% non-fat dry milk for at least 2 hours at RT or preferentially over night at 4°C on a shaking platform. After 3x 10 min washing with 1x TBST, the secondary HRP coupled antibodies were applied to the membrane diluted in 1x TBST for two hours of incubation at RT on a shaking platform. Then blots were washed 3x 10 min with 1x TBST and subjected to chemiluminescence detection (Pierce ECL Western Blotting Substrate). For this purpose 1 ml of detection reagent 1 was mixed with 1 ml of detection reagent 2 and pipetted onto the membrane. After one minute, the membrane was wrapped in cellophane and exposed

to a chemiluminescence film. Protein bands were visualized by developing and fixing the film (all Kodak).

5.4.6 SDS-PAGE, silver-staining and sample preparation for mass spectrometry

For the identification of *At*SPO11-1 interacting proteins by mass spectrometry, the following protocol was applied. For immuno-precipitation GammaBind Plus protein G sepharose beads (GE Healthcare) were washed 2x with 1xTBST (150 mM NaCl, 10 mM Tris HCl, pH 7.5, 0.04% Triton X-100). For all subsequent steps TBST with Triton X-100 instead of Tween-20 was used. After binding of the antibody as described in section 5.4.3, the antibody was crosslinked to the beads for 30 min at RT using 20 mM DMP in 0.2 M sodium borate, pH 9.2. Cross-linking was stopped by washing the beads 2x 10 min with 0.2 M Tris HCl, pH 8. For re-equilibration, beads were washed with 1xTBST twice, followed by pre-elution consisting of two washes with 0.1 M glycine. After a second re-equilibration step, cell extracts were prepared as described in section 5.4.1. Cell extracts were pre-cleared while incubating with GammaBind Plus protein G sepharose beads, pre-loaded and cross-linked to non-specific mouse IgG (Sigma), for 30 min at 4°C. Pre-cleared cell extracts were mixed with the antibody coupled beads and IP was performed for 30 min at 4°C on a top-over wheel. Beads were washed 3x with ice-cold IP buffer (20 mM Tris HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 2 mM EDTA, 0.1% NP-40), 3 times with ice-cold IP buffer with higher salt and detergent (0.45 M NaCl, 1% Triton X-100) and 3 times with ice-cold IP buffer without detergent.

Since elution of AtSPO11-1/18xmyc was not possible using glycine or hydrochloric acid, beads were mixed with 4x protein sample buffer (NuPAGE® LDS sample buffer, Invitrogen) and elution of potential protein complexes was performed by heating to 94°C for 5 min.

Since for Mass Spec analysis the purest chemicals are required, precasted NuPAGE® Novex® Bis-Tris gels were used combined with the NuPAGE MES SDS running buffer (all Invitrogen). Samples were loaded and electrophoretically separated at 20 mA/gel in a special gel chamber for pre-casted gels (Invitrogen).

Silverstaining was performed as described in (Blum *et al.*, 1987). Stained gels were brought to the MFPL Mass Spectrometry facility for further processing. Lanes were cut in pieces, proteins were digested with trypsine and analyzed by mass spectrometry. Further data analysis was performed using Scaffold 2 software (www.proteomsoftware.com). For a detailed description of the work at the MFPL Mass Spec facility see section 5.4.7.

5.4.7 Enzymatic digest, LC-MS/MS analysis and data analysis

Silver stained gel bands were used for the nano-electrospray LC-MS/MS investigations. The whole lane was excised and cut into 30 slices. These bands were cut, chopped and washed several times with high quality water (Millipore). Proteins were reduced by dithiothreitol (DTT, Roche) and alkylated by iodoacetamide (Sigma-Aldrich). Trypsin (Roche, proteomics grade) was used as protease, the digest was carried out overnight at 37°C and was stopped by acidifying the aliquot with 10% aqueous formic acid (Merck) to an end concentration of approximately 1%. The HPLC used was an UltiMate™ system (Dionex Corporation) equipped with a PepMap C18 purification column (300µm x 5mm) and a 75µm x 150mm analytical column of the same material. 0.1% TFA (Pierce) was used on the Switchos module for the binding of the peptides and a linear gradient of acetonitrile (Chromasolv®, Sigma-Aldrich) and 0.1% formic acid in water was used for the elution. The gradient was (mobile phase A: 5% acetonitrile / 0.1% formic acid in water; mobile phase B: 80% acetonitrile / 0.1% formic acid in water): 0% B for 8 min, 50% B in 60 min, 95% B in 1 min, 100% B for 5 min, 0% B in 1 min, 0% B for 22 min. LC-MS/MS analyses were carried out with the UltiMate™ system interfaced to an LTQ (Thermo) linear ion trap mass spectrometer. The nanospray source of Proxeon (Odense) was used with the distal coated silica capillaries of New Objective (Woburn). The electrospray voltage was set to 1500V. Peptide spectra were recorded over the mass range of m/z 450-1600, MS/MS spectra were recorded in information dependent data acquisition, the default charge state was set to 3; the mass range for MS/MS measurements was calculated according to the masses of the parent ions. One full spectrum was recorded followed by 4 MS/MS spectra for the most intense ions, automatic gain control was applied and the collision energy was set to the arbitrary value of 35. Helium was used as collision gas. The instrument was operated in data dependent modus; fragmented ions were set onto an exclusion list for 20 seconds.

Raw spectra were interpreted by Mascot 2.2.04 (Matrix Science Ltd.) using Mascot Daemon 2.2.2. Peptide tolerance was set to +/- 2 Da, MS/MS tolerance was set to +/- 0.8 Da. Carbamidomethylcysteine was set as static modification, oxidation of M, was set as variable modifications. Each band was measured separately and the raw files belonging to the same lane were merged into one Mascot mgf file for the interpretation.

The database used for Mascot search was the nr protein data base of NIH (NCBI Resources), taxonomy was *A. thaliana*. Ion score cut off was set to 20 for Mascot. Mascot results were loaded into Scaffold (Version 2.01.01, Proteome Software Inc.) for rescoring and overview. Peptide identification threshold was set to 95 % protein identification

threshold was set to 99%. Protein identifications were accepted if they contained at least 2 identified peptides. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. This section was written by Edina Csaszar, Mass Spectrometry Facility, MFPL, Vienna.

5.5 Solexa Protocol

5.5.1 Basic overview of the ChIP-Seq protocol

Starting from 5×10^9 yeast cells (*S. cerevisiae* and *S. pombe*) the first step of the protocol is the preparation of cell extracts using glass beads to disrupt the cells in a Multi-beads shocker® (Yasui Kikai). For the samples, strains harboring either a myc-tagged Spo11 (*S. cerevisiae*) or Rec12 (*S. pombe*) were taken. After clearing the lysate in a 2-step centrifugation process, the protein extract was pre-cleared while incubating with Ademtech protein G beads without antibody. After pre-clearing, the protein extract was added to Ademtech protein G beads coupled to an anti-myc antibody for ChIP. After stringent washing the first Solexa adapter ligation was performed on beads followed by a unidirectional PCR reaction using a biotinylated primer. The PCR product was recovered by performing a pull-down with Ademtech Streptavidin beads. After the second Solexa adapter ligation and stringent washing, the adapter containing fragments were amplified in 30 cycles of a standard PCR reaction using adapter specific primers. The PCR products were loaded on a 3% agarose gel, cut out and gel-extracted. After re-amplification of the products in an identical PCR reaction as before, products were separated on a 3% agarose gel, cut out, gel-extracted and sent to the sequencing facility. The concentration of the DNA was determined using a Nanodrop or qPCR, prior sending samples to the sequencing facility.

S. cerevisae strains were generated and provided by Martin Xaver (Franz Klein lab). *S. pombe* strains were generated and provided by Anna Estreicher (Josef Loidl lab) and Silvia Polakova (Juro Gregan lab). All are members of the Department of Chromosome Biology, University of Vienna.

5.5.2 Preparation of cell extracts

1) Thaw cells on ice.

2) Resuspend cells in 1 ml lysis buffer + inhibitors (for 50 ml: 0.05 g sodium deoxycholate, 2.5 ml 1M HEPES KOH, pH 7.5, 7 ml 1M NaCl, 0.1 ml 0.5M EDTA, 0.5 ml 100% Triton X-100, one tablet of Complete Mini Protease Inhibitor from Roche and 0.5 ml 250 mM

PMSF) by gently pipetting with a 1 ml tip and transfer to two 2 ml Multi beads shocker tubes.

3) Add glass beads (0.4-0.6 mm, Sartorius, 90% of the volume of the cell suspension) and disrupt cells in a pre-cooled Multi-beads shocker® (Yasui Kikai) at 2500 rpm for 600 seconds (30" on, 30" off).

4) Check disruption of cells under the microscope. If the cells are not disrupted to an extend of 80-90%, repeat the Multi-beads shocker run.

5) Poke a hole into the bottom of two 0.5 ml reaction tubes (23G needle, B. Braun) and stack them into 2 ml tubes (caps cut off).

6) Transfer the cells lysate with glass beads completely into the poked tubes.

7) Centrifuge for 1 min at 4000 rpm at 4°C and transfer the cell extract to 1.5 ml reaction tubes.

8) Repeat this transfer step from the Multi-beads shocker tubes once to recover all of the disrupted cells.

9) Centrifuge the cell lysates at full speed for 5 min at 4°C and take the supernatant for chromatin-immunoprecipitation.

5.5.3 Pre-clearing, preparation of beads, chromatin immuno-precipitation

10) Wash 10 µl of Bio Adembeads protein G (Ademtech) 3x with NET2 buffer (50 mM Tris HCl, pH 7.5, 150 mM NaCl, 0.05% NP-40) for every sample.

11) Add the protein extract and perform pre-clearing for 1 hour at 4°C on a top-over wheel.

12) The day before couple your antibody to protein G beads as follows: Take 25 µl of Bio Adembeads protein G (Ademtech) for every sample, wash them 3x with NET2 buffer and add 300 µl of anti-myc antibody (9E11, mouse monoclonal, cell culture supernatant, IMP) and 300 µl of NET2 buffer. Incubate over night at 4°C on a top-over wheel.

13) During pre-clearing, wash your prepared beads from the day before 3 times with NET2 buffer.

14) Capture the beads after pre-clearing the protein extract and add pre-cleared protein extract to the washed antibody coupled beads.

15) Perform immuno-precipitation at 4°C on a top-over wheel for 3-4 hours.

16) Wash beads 2x with lysis buffer without inhibitor (for 50 ml: 0.05g sodium deoxycholate, 2.5 ml 1M HEPES KOH, pH 7.5, 7 ml 1M NaCl, 0.1 ml 0.5M EDTA, 0.5 ml 100% Triton X-100).

17) Wash beads 2x with lysis buffer +360 mM NaCl without inhibitors (the same as lysis buffer, except 25 ml of 1M NaCl instead of 7 ml).

18) Wash beads 2x with washing buffer (for 50 ml: 0.25 g sodium-deoxycholate, 0.5 ml 1M Tris HCl, pH 8, 2.5 ml 5M lithium acetate, 2.5 ml 10% NP-40, 0.1 ml 0.5M EDTA).

19) Wash beads 2x with 1-fold TE (100 mM Tris HCl, pH7.5, 10 mM EDTA, pH 8).

20) Wash beads 2x with 1-fold T4 ligase buffer (Roche).

5.5.4   Annealing of Solexa adapters

21) For annealing of the Solexa adapters take oligos gAdapter1/PCR_gAdapter1-3rd generation for Solexa adapter 1 and gAdapter2/PCR_gAdapter2-3rd generation for Solexa adapter 2. Mix 4 µl of the respective oligonucleotides (100 µM each) with 1 µl 100 mM Tris HCl, pH8 and 1 µl 1 M NaCl. Put the samples into a thermocycler and perform the reaction as follows: 95°C for 2 min, then 70 cycles at 95°C for 1 min (-1°C/cycle).

5.5.5   1st Solexa adapter ligation

22) Take the washed beads from step 20 and resuspend in 22 µl dH$_2$O and add 3 µl adapter 1, 3 µl 10x T4 ligase buffer and 2 µl T4 ligase (both Roche). Incubate overnight at 1000-1200 rpm in a thermoblock at 16°C.

23) Wash beads 2x with NET2 buffer.

24) Wash beads 2x with 1-fold High Fidelity enzyme mix buffer (Fermentas).

25) Resuspend beads in 34 µl dH$_2$O.

5.5.6   Unidirectional PCR

26) Take the re-suspended beads from step 25 and add 4 µl unidirectional PCR Primer_3rd generation [100 nM], 5 µl 10x High Fidelity buffer, 6 µl 0.5 mM dNTPs and 1 µl High Fidelty enzyme mix (Fermentas) to a final volume of 50 µl.

27) Run the following PCR program: denaturate for 1 min at 94°C, followed by 40 cycles of 20 sec at 94°C, 10 sec at 56°C, 10 sec at 72°C, followed by a final extension of 3 min at 72°C.

5.5.7   Isolation of newly synthesized, biotinylated ssDNA

28) Capture protein G beads with a magnetic device, the supernatant contains the unidirectional PCR product, which is used for further processing.

29) For every sample take 15 µl of Bio Adembeads Streptavidin Plus (Ademtech) and wash 3x with NET2 buffer.

30) Transfer the unidirectional PCR product to a new tube and heat up for 2 min at 94°C.

31) Put the tube on ice immediately after heating up.

32) Mix the unidirectional PCR product with 200 µl NET2 and add to the washed beads from step 29.

33) Incubate on a top-over wheel for 1 hour at room temperature.

34) Wash beads 2x with NET2 buffer.

35) Wash beads 2x with 1-fold T4 ligase buffer.

36) Resuspend beads in 22 µl dH$_2$O.

5.5.8   2nd Solexa adapter ligation

37) Take the adapter 2 from step 21 and beads from step 36 and add 3 µl adapter 2, 3 µl 10x T4 ligase buffer and 2 µl T4 ligase. Incubate overnight at 1000-1200 rpm in a thermoshaker at 16°C.

38) Wash beads 2x with NET2 buffer.

39) Wash beads 2x with 1-fold High Fidelity enzyme mix buffer.

40) Resuspend beads in 34.5 µl dH$_2$O.

5.5.9   Solexa amplification PCR

41) Take beads from step 40 and add the following: 5 µl 10x High Fidelity enzyme mix buffer, 5 µl 2 mM dNTPs, 2.5 µl gSOLEXA_primer_1 [10 µM], 2.5 µl gSOLEXA_primer_2 [10 µM] and 0.5 µl High Fidelity enzyme mix.

42) Run the following PCR program: denaturate for 30 sec at 94°C, followed by 30 cycles of 15 sec at 94°C and 15 sec at 72°C, followed by a final extension of 2 min at 72°C.

5.5.10 Re-amplification

43) Load samples on a 3% agarose gel, let run as long as possible and cut out the smear above the adapter/primer dimer band between 95-160 bp.

44) Make a gel extraction according to the protocol of the E.Z.N.A.® kit. Elute with 2x20 µl of dH$_2$O.

45) Take the whole eluate (also 20 µl is fine) and perform a re-amplification as described in section 5.5.9.

5.5.11 Sample preparation, sequencing and data analysis

46) Load samples on a 3% agarose gel, let run as long as possible and cut out the smear above the adapter band between 95-160 bp.

47) Make a gel extraction according to the protocol of the E.Z.N.A.® kit. Elute with 2x20 µl of dH$_2$O.

48) Determine the concentration of your DNA using Nanodrop or equivalent. For one Solexa run 15 ng of DNA is the minimum requirement.

49) Send samples to Solexa sequencing.

Sequencing was performed at the IMP Genomics Facilty by Andreas Sommer and co-workers on an Illumina Genome Analyzer. After base-calling and quality control, raw-sequence data was transferred to Fritz Sedlazeck (CIBIV, Vienna) for further processing. He clipped the adapter sequences, mapped the sequence reads to the respective

genomes using his own program NextGenTool (NGT) and produced coverage files and all other data output for further analysis.

The raw reads were clipped using NextGenTool (NGT). It detects the adapter sequence by searching form 3' to 5' end. A putative adapter sequence in a read was defined as either 10 consecutive characters matching to the adapter sequence or twelve or more characters matching with a maximum of two mismatches. After finding the putative adapter sequence, the respective subsequence was deleted. Sequence reads shorter than 13 nt were not included in further analysis due to mapping ambiguity.

The mapping of the clipped reads to the reference genome was done with NextGenMap (NGM). NGM computes a banded Smith-Waterman alignment and employs a graphic card to speed up calculations. A read is called a putatively mapped read if at least two sub-sequences of 13 bp are identical to the reference genome and occur close to each other. For all regions in the reference genome where a read putatively maps, a banded Smith Waterman alignment (allowing for maximal 10 consecutive insertion or deletions) was computed and the region(s) with the highest alignment score were selected. The actual alignment was calculated for these regions. If multiple regions provide the same highest score, the alignment was computed for each region. However, in the subsequent analyses all reads that map to multiple regions were discarded. If the alignment of a mapped read has an identity below 80% this read is also excluded from further analysis.

Consensus sequences and coverages were computed as follows: the most common nucleotide of all reads that map to the same position in the reference genome was taken for the consensus sequence. The number of reads that map to such a position was computed and called the total coverage of the genome at this position. Moreover reads that map to the plus strand (plus-coverage) and reads that map to the minus strand (minus-coverage) were counted. The total coverage at a position is the sum of plus-coverage and minus-coverage.

Further data analysis and visualization was mainly done by myself using the free statistic software "R" (www.r-project.org).

All oligonucleotides for the Solexa protocol were purchased from IDT with the highest purification grade (either PAGE or HPLC) and are listed below for the standard single end runs.

gAdaptor_1

5'-/5Phos/AGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG/3SpC3/ -3'

PCR_gAdaptor_1_3rd generation

5'-/5AmMC6/AAAAAAAACAAGCAGAAGACGGCATACGAGCTCTTCCGATCTNNN
NN/3SpC3/-3'

Unidirectional Primer_3rd generation

5'-/5BioTEG/AAAAAAAACAAGCAGAAGACGGCATACGAGCTCTTCCGATCT-3'

gAdaptor_2

5'-/5Phos/AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/ -3'

PCR_gAdaptor_2_3rd generation

5'-/5AmMC6/AAAAAAAAACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN
N/3SpC3/ -3'

gSOLEXA_primer_1

5'-CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT-3'

gSOLEXA_primer_2

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATCT-3'

For Paired End runs a modified adapter 1 with an inserted binding site for a Solexa
sequencing primer was necessary.

PCR_gAdaptor_1_PE

5'-/5AmMC6/AAAAAAAACGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNN
NN/3SpC3/-3'

gAdaptor_1_PE

5'-/5Phos/AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG/3SpC3/ -3'

Unidirectional Primer_PE

5'-/5BioTEG/AAAAAAAACGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT -3'

gSOLEXA_primer_1_PE

5'-
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCC
GATCT-3'

For Multiplexing runs, the adapter 2 was modified by inserting different 3 base barcodes at
the end (see below).

BAR1_CCT:

PCR_gAdaptor_2_BAR1_CCT

5'-
/5AmMC6/AAAAAAAAACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTNNNNN/3S
pC3/-3'

gAdaptor_2_BAR1_CCT

5'-/5Phos/AGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/-3'

BAR2_GGT

PCR_gAdaptor_2_BAR2_GGT

5'-
/5AmMC6/AAAAAAAAACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTNNNNN/3S
pC3/-3'

gAdaptor_2_BAR2_GGT

5'-/5Phos/ACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/-3'

BAR3_AAT

PCR_gAdaptor_2_BAR3_AAT

5'-
/5AmMC6/AAAAAAAAACACTCTTTCCCTACACGACGCTCTTCCGATCTAATNNNNN/3S
pC3/-3'

gAdaptor_2_BAR3_AAT

5'-/5Phos/ATTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/-3'

BAR4_CGT

PCR_gAdaptor_2_BAR4_CGT

5'-
/5AmMC6/AAAAAAAAACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTNNNNN/3S
pC3/-3'

gAdaptor_2_BAR4_CGT

5'-/5Phos/ACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/-3'

BAR5_ATT

PCR_gAdaptor_2_BAR5_ATT

5'-
/5AmMC6/AAAAAAAAACACTCTTTCCCTACACGACGCTCTTCCGATCTATTNNNNN/3S
pC3/-3'

gAdaptor_2_BAR5_ATT

5'-/5Phos/AATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/-3'

6 Results

6.1 Generation of tools for the analysis of the initiation of meiotic recombination

Both, *At*SPO11-1 and *At*SPO11-2 were identified to be the catalytic proteins, responsible for DNA DSB formation in *Arabidopsis thaliana* (Grelon *et al.*, 2001; Hartung and Puchta, 2000, 2001; Hartung *et al.*, 2007; Stacey *et al.*, 2006). The mechanistic mode of DSB formation seems conserved in all eukaryotes, but the initiation of meiotic recombination is different between *S. cerevisiae* and *A. thaliana*, concerning the involved protein complexes. In *S. cerevisiae*, in addition to Spo11, at least 9 other proteins are involved in DSB formation. In *A. thaliana* so far, only three additional factors, *At*PRD1, *At*PRD2 and *At*PRD3, were identified recently (De Muyt *et al.*, 2009; De Muyt *et al.*, 2007) and only one is related to a *S. cerevisiae* protein (see introduction).

Nevertheless, many open questions remain elusive. Are there other proteins involved in DSB formation in *Arabidopsis*, how do they interact in a complex and do *At*SPO11-1 and *At*SPO11-2 form a heterodimer. To answer such questions, *At*SPO11-1 fusion proteins were generated by Ondrej Krsicka during his diploma thesis. During my PhD thesis I created additional *AtSPO11-1* clones and started to generate a set of genomic *At*SPO11-2 fusion proteins, to increase the set of molecular tools with the aim to find new *At*SPO11-1 and *At*SPO11-2 interacting proteins and to determine their potential heterodimer formation. Additionally, the generated tools are instrumental for all other projects described in this thesis (see below).

AtSPO11-1 and *AtSPO11-2* perfectly match with the Spo11 family in an alignment of motifs I-V. The proteins are identical between 22 and 35% to their eukaryotic counterparts and 28% to each other (Hartung and Puchta, 2000). *AtSPO11-1* consists of 16 exons and 15 introns, with an open reading frame (ORF) of 1089 bp, coding for a protein consisting of 362 amino acids. *AtSPO11-2* contains 12 exons and 11 introns, with a length of 1152 bp, coding for a 383 amino acid protein. Both genes harbor an intron in the 3' untranslated region (UTR) which is not always spliced out. Different 3' RACE experiments revealed three polyadenylation sites for *AtSPO11-1*, one is located within the 3' UTR intron and two distal to the intron. 3' RACE for *AtSPO11-2* also revealed three different poly(A) sites, two of them are located in the 3' UTR intron and one was found to be located 245 bp away from the 3' intron (Fig. 6.1), (Hartung and Puchta, 2000). RT-PCR experiments of *AtSPO11-1* always resulted in a complex banding pattern of PCR products, suggesting

alternative splicing forms, since *AtSPO11-1* and *AtSPO11-2* are single copy genes in the *Arabidopsis* genome. Further analysis of the splice products revealed that all of the alternatively spliced mRNAs contain premature stop codons (Fig. 6.2). As alternatively spliced cDNAs are also present in somatic and generative tissue, the biological function of this phenomenon is not clear. One possibility would be that *AtSPO11-1* expression is regulated via the nonsense mediated decay (NMD) pathway (Gutierrez *et al.*, 1999). The NMD pathway is known to be responsible for a rapid decay of the respective mRNAs due to premature stop codons (Hartung and Puchta, 2000).
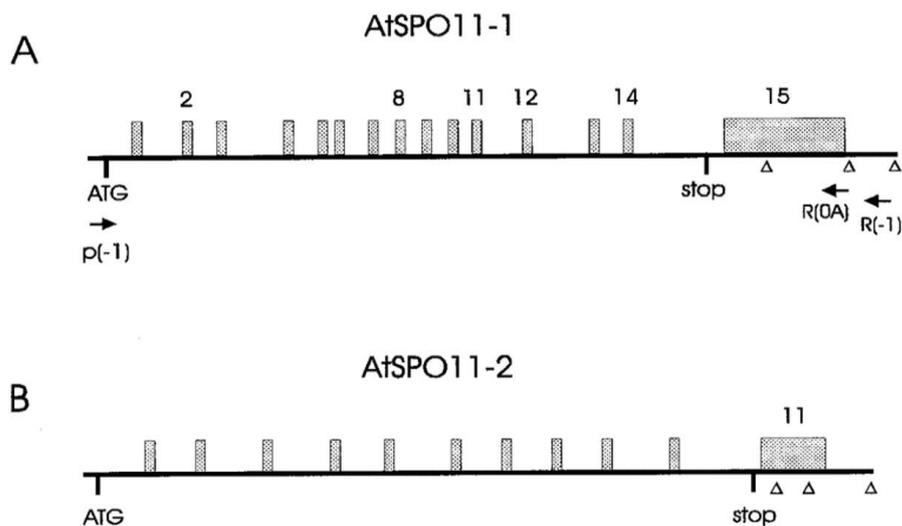
Figure 6.1: Schematic structure of both *A. thaliana* SPO11 homologous genes: Introns are shown as grey boxes above the sequence. Verified polyadenylation sites are represented by open triangles. **(A)** Primers used for RT–PCR analysis of AtSPO11-1 are shown as arrows below the schematic sequence. Figure and text taken from (Hartung and Puchta, 2000).
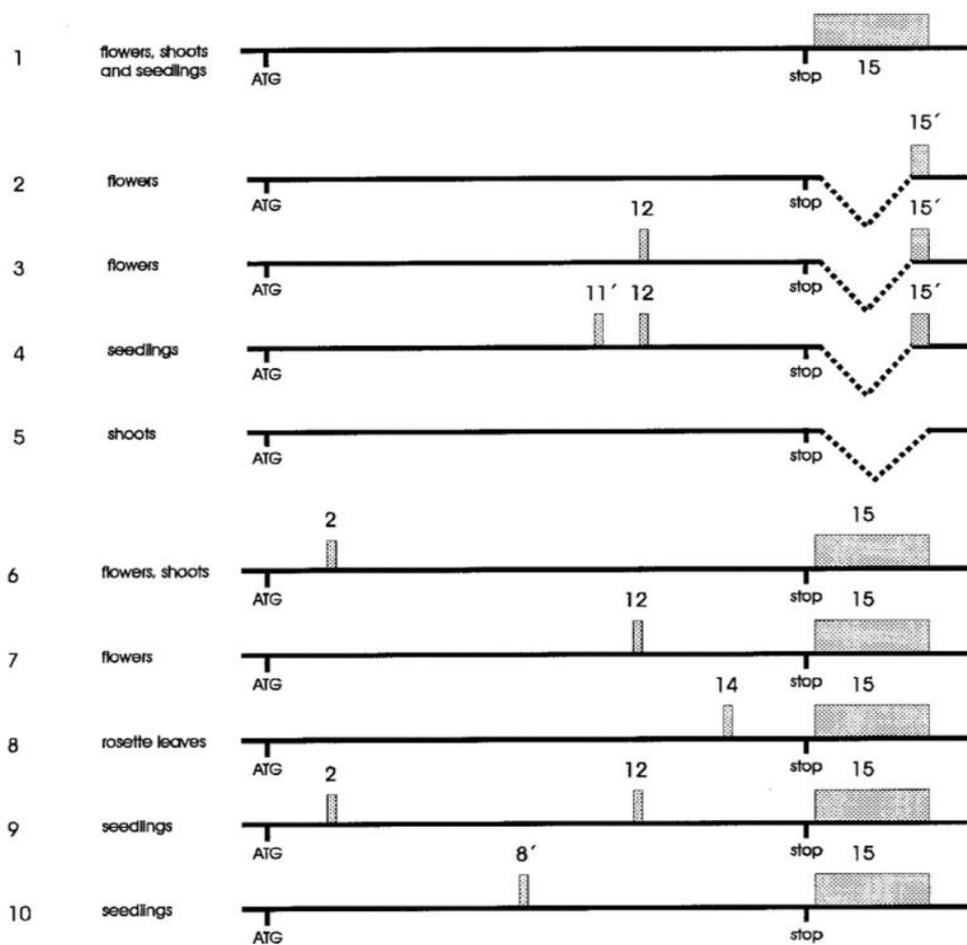
Figure 6.2: Analysis of alternative splicing products of *AtSPO11-1* by RT–PCR and sequencing. All splice products with resting introns would result in truncated proteins due to premature stop codons. Introns are represented by grey boxes and numbered in the same order as in the full gene. The partially spliced out intron 15 is represented by a curved dashed line. Partially remaining introns are numbered 8′, 11′ and 15′. The organs from which the cDNAs derived are shown on the left. In number five the intron 15 is slightly longer then in all other lanes because an alternative splice junction was found. Figure and text taken from (Hartung and Puchta, 2000).

In the past several attempts to generate *At*SPO11-1 fusion proteins for the expression in plants were performed. However, generated cDNAs were not successful to complement the *Atspo11-1* mutant alleles (Christine Mezard, Mathilde Grelon, INRA Versailles, personal communication).

To avoid such problems for *AtSPO11-1* and also *AtSPO11-2*, it was decided to generate full genomic clones for both, containing 5' and 3'UTRs, harboring promoter regions and potential regulatory downstream sequences.

6.1.1 Generation of *At*SPO11-1 fusion proteins

The generation of *At*SPO11-1 fusion proteins was initiated in the lab by Ondrej Krsicka and is described in detail in the diploma theses of Ondrej Krsicka and Michael Peter Janisiw (Janisiw, 2009; Krsicka, 2007).
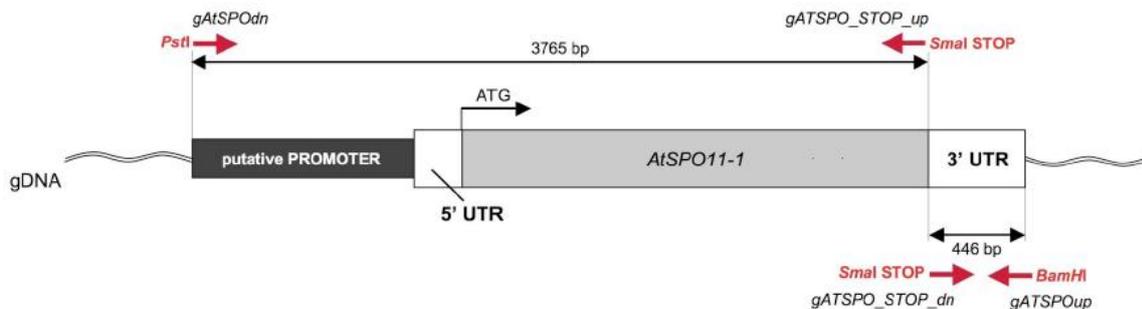
Figure 6.3: Schematic representation of cloned *At*SPO11-1 wild-type like transgene. Previous to this thesis the whole genomic region was cloned by PCR amplification of two separate fragments (bordered by vertical lines). Oligonucleotide primers were designed to amplify the first fragment from the stop codon of the upstream gene (At3g13175), including the putative promoter (dark grey), 5' UTR, and the full-length gene from the start codon (ATG) to the last nucleotide upstream of the stop codon (STOP). Primers *gATSPOdn* and *gATSPO_STOP_up* were designed to introduce artificial restriction endonuclease (RE) sites *Pst*I and *BamH*I, allowing downstream cloning of the amplified genomic region (indicated in red). The second fragment was amplified from the stop codon (STOP), artificially introducing a *Sma*I RE site in frame with and right upstream of the stop codon, down to the start codon of the downstream gene (At3g13160), using primers *gATSPO_STOP_dn* and *gATSPOup*. The *Sma*I RE site in frame with the stop codon of *At*SPO11-1 allows the introduction of protein tags at the C-terminus of *At*SPO11-1 and thereby the creation of artificial fusion proteins. Figure and text taken from (Janisiw, 2009).

For this purpose the genomic region of *AtSPO11-1* and its regulating upstream and downstream sequences were cloned as depicted in Fig. 6.3. The two 3765 and 446 bp fragments were amplified from *Arabidopsis* wild-type DNA, ecotype Columbia (Col-0), using a high fidelity, proofreading polymerase. By extension of the PCR primers, different restriction sites were introduced during the amplification steps. Flanking the whole construct, *Pst*I and *BamH*I were added, a *Sma*I restriction site was introduced directly in front of the stop codon. Both fragments were subcloned into the vector pCR®2.1 (Invitrogen). Successful cloning was verified by restriction analysis and further sequencing of the plasmids. After the successful assembly of both fragments in the pCR®2.1 vector via the *Sma*I restricition site, the whole construct was subcloned into the multiple cloning site (MCS) of the binary plant vector pCB302 (Xiang *et al.*, 1999) as a *Pst*I/*BamH*I fragment, which resulted in the untagged wild-type like clone *pCB302/AtSPO11-1* (the respective plant line was later depicted as *49DI*) (Janisiw, 2009; Krsicka, 2007).

As mentioned before, a *Sma*I restriction site was introduced directly in front of the stop codon with the intention to fuse different tags to the C-terminus of the SPO11-1 protein. During his work Ondrej Krsicka generated two more *At*SPO11-1 fusion proteins. He fused the 18xmyc epitope and the Gal4-2 binding domain to the C-terminus of *AtSPO11-1*, which resulted in the clones *pCB302/AtSPO11-1/18xmyc (8BI)* and *pCB302/AtSPO11-1/Gal4 (4CI)*. He started the selection of these transgenic plant lines and I took over the work at the beginning of my PhD thesis.

In addition to these three existing *AtSPO11-1* clones, I generated two further ones, harboring a 24xHA epitope and a SF tag, respectively resulting in *pCB302/AtSPO11-1/24xHA (3BE)* and *pCB302/AtSPO11-1/SF (9BJ)*.

The 24xHA tag was generated by fusing 2 times 12xHA, which was amplified from a plasmid named pPILY/15xHA (kindly provided by Thomas Potuschak, CNRS, Strasbourg). A *Sma*I and a *EcoR*V restriction site were added by primer extension during amplification to be compatible with the blunt end fusion to the *Sma*I site at the C-terminus of either *At*SPO11-1 or *At*SPO11-2. Primers pPILY 15xHA forw and pPILY 15xHA rev were used running following PCR program: 94°Cx15 sec initial denaturation, 30 cycles (94°Cx15 sec, 55°Cx15 sec, 72°Cx90 sec), final elongation at 72°C for 5 min.

The SF (Strep/Flag) tag was also amplified adding a *Sma*I and a *EcoR*V restriction site for blunt end fusion to a *Sma*I restriction site. The SF tag is very small (41 aa) compared to a standard TAP tag (202 aa). Due to the avoidance of overnight protease cleavage (desthiobiotin and Flag peptide elution) a very fast purification within 2.5 hours is possible (Gloeckner *et al.*, 2007; Gloeckner *et al.*, 2009a; Gloeckner *et al.*, 2009b). Amplification, cloning and sequencing of a SF-cloning cassette was done by Clemens Uanschou (Schlögelhofer lab, C. Uanschou, unpublished).

6.1.2 Selection and complementation analysis of transgenic lines

All of the transgenic constructs were transformed into plants of a population originating from a heterozygous *Atspo11-1-2* mutant plant and selected as described in materials and methods using the BASTA resistance marker of the pCB302 vector. Homozygous *Atspo11-1-2* mutants show a severe sterility phenotype, with shortened siliques, containing on average 1 seed/silique (compared to 40-50 in wild-type plants). After successful selection of transgenic lines, T1 transgenic lines with a single locus insertion

were propagated and monitored for homozygosity of the endogenous *Atspo11-1-2* mutant allele and the transgene by PCR and BASTA resistance.

In general, successful complementing transgenic lines should rescue the sterility defect phenotype of *Atspo11-1-2* mutants. In case the endogenous *Atspo11-1-2* mutant allele is homozygous and all plants are sterile, the transgene is not functional and not complementing the mutation. If the endogenous *Atspo11-1-2* mutant allele is homozygous and the transgene is heterozygous and functional, 75% of the plants are fertile and 25% are sterile. If the endogenous *Atspo11-1-2* mutant allele is homozygous and the transgene is homozygous and functional, then all plants should be fertile.

Monitoring the fertility of the transgenic plant lines listed in figures 6.4 and 6.5, all transgenes, except *pCB302/AtSPO11-1/24xHA* are complementing the *Atspo11-1-2* mutant phenotype, restoring wild-type silique size, suggesting a functional expression of the respective fusion proteins.
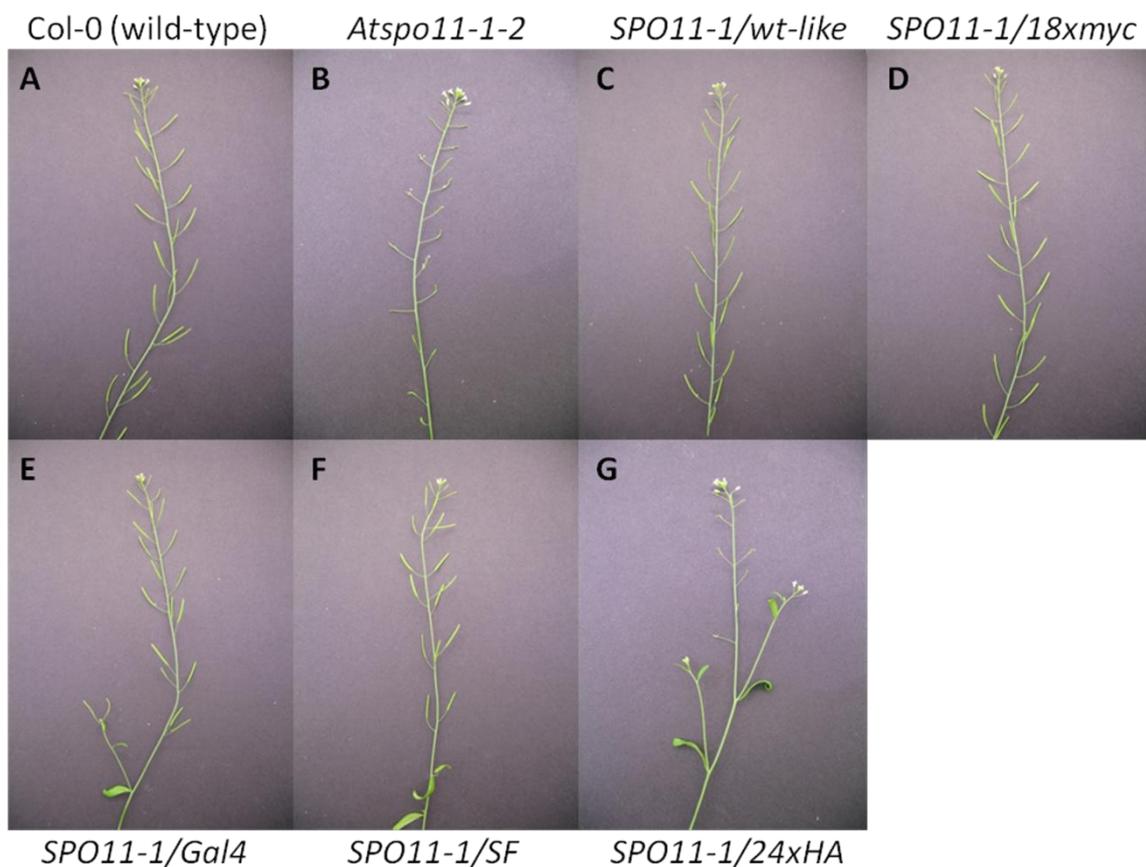
Figure 6.4: Transgenic *AtSPO11-1* lines. (A) Fertile wild-type plant, ecotype Columbia, (B) sterile *Atspo11-1-2* mutant plant, (C) complementing, fertile *SPO11-1/wt-like* plant, (D) complementing, fertile *SPO11-1/18xmyc* plant, (E) complementing, fertile *SPO11-1/Gal4* plant, (F) complementing, fertile *SPO11-1/SF* plant, (G) non-complementing, sterile *SPO11-1/24xHA* plant.

Additionally to this genotypic and phenotypic analysis, a seed count was performed for all transgenic plant lines to assess the level of complementation compared to 100% fertile wild-type plants. For all complementing transgenic lines wild-type seed levels were observed (see Fig. 6.5).

Figure 6.5: At SPO11-1 fusion proteins. The upper three (untagged, 18xmyc, Gal4-2 BD) were generated by Ondrej Krsicka during his diploma thesis (Krsicka, 2007), the latter two (24xHA, SF) were generated by myself. All constructs, except *pCB302/AtSPO11-1/24xHA*, display wild-type silique size and seed level, complementing a homozygous *Atspo11-1-2* mutant population, indicative for the expression of functional fusion proteins in plants. The seed number/silique for respective transgenic lines is displayed in the figure (n=500-1000 siliques counted).

6.1.3   Generation of *At*SPO11-2 fusion proteins

For this purpose the genomic region of *AtSPO11-2*, containing the putative promoter region, 5' and 3' UTR was cloned as depicted in Fig. 6.6. All restriction sites depicted in black are natural ones. *Sma*I, which is depicted in red, was introduced by primer extension directly in front of the stop codon for the fusion of different tags to the C-terminus of *At*SPO11-2. The two smaller fragments (259 bp and 1186 bp) were amplified from *Arabidopsis* wild-type DNA, ecotype Columbia (Col-0), using KOD DNA polymerase (Novagen) under following PCR conditions. 94C°x15 sec initial denaturation, 30 cycles (94°Cx15 sec, 54°Cx15 sec, 72°Cx75 sec), final elongation at 72°C for 5 min. For the *Eco91*I/*Sma*I fragment primers Spo11-2_Stop_up/Spo11-2_*Eco91*I_dn and for the *Sma*I/*Nhe*I fragment primers Spo11-2_Stop_dn/Spo11-2_*Nhe*I_up were used. Prior the final elongation, 1 µl of Taq polymerase (Fermentas) was added for A-tailing. PCR products were gel-purified and sub-cloned into the vector pCR2.1 (Invitrogen). Successful cloning was verified by restriction digest and further sequencing of the plasmids. Both fragments were assembled in the pCR2.1 vector via the *Sma*I restriction site, resulting in the *Eco91*I/*Nhe*I fragment and kept for further processing.
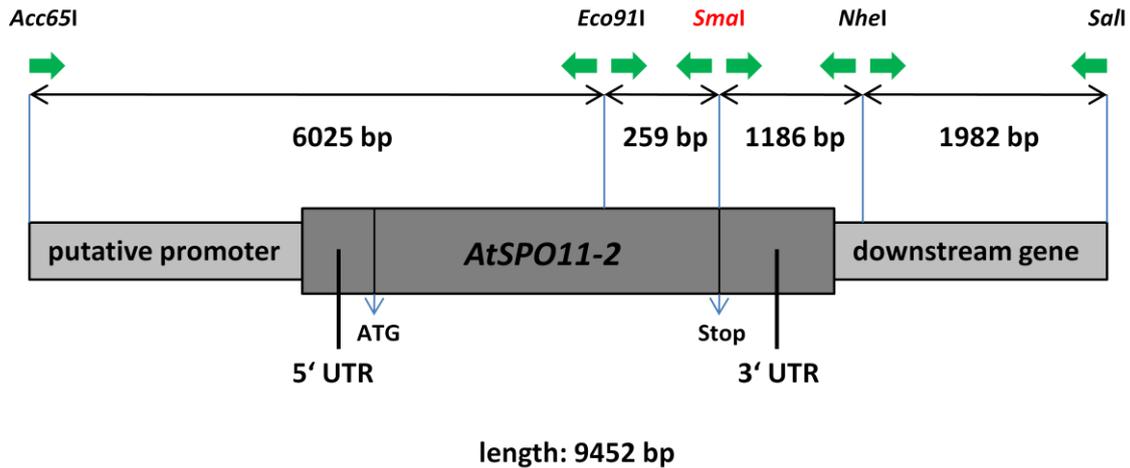
Figure 6.6: Strategy for the genomic cloning of *AtSPO11-2*. The genomic region of *AtSPO11-2* was cloned as displayed in the figure, including the putative promoter region, 5' and 3' UTR. The four fragments were amplified with either KOD polymerase (Novagen) or High Fidelity Enzyme Mix (Fermentas) and were cloned into the vector pCR2.1 (Invitrogen). Successful cloning was verified by restriction analysis and further sequencing. All four fragments were assembled and cloned into plant binary vector pBIB-Hyg (Becker, 1990) as a *Acc65*I/*Sal*I fragment which resulted in the clone *pBIB-Hyg/AtSPO11-2 (4BE)*. All restriction sites depicted in black are natural ones. *Sma*I, which is depicted in red, was introduced by primer extension directly in front of the stop codon for the fusion of different tags to the C-terminus of *At*SPO11-2. The green arrows represent the primer positions for the amplified fragments.

The two bigger fragments (1982 and 6025 bp) were amplified from BAC (F22C12) DNA, containing the genomic region of choice, using the High Fidelity Enzyme Mix (Fermentas). The *Acc65*I/*Eco91*I fragment was generated using primers Spo11-2_right_forw/Spo11-2_right_rev running the following conditions: 94°Cx1 min, 10 cycles (94°Cx30 sec, 54°Cx30 sec, 68°Cx6 min), 20 cycles (94°Cx30 sec, 54°Cx30 sec, 68°Cx6 min (+10 sec/cycle)), 68°Cx10 min. The *Nhe*I/*Sal*I fragment was amplified using primers Spo11-2_left_forw/Spo11-2_left_rev running the following protocol: 94°Cx1 min, 30 cycles (94°Cx30 sec, 55°Cx30 sec, 72°Cx150 sec), 72°Cx10 min. PCR products were gel-purified and sub-cloned into the vector pCR2.1. Successful cloning was verified by restriction digest and further sequencing of the plasmids.

These two fragments then were assembled with the already assembled *Nhe*I/*Sal*I fragment in the pCR2.1 vector. Finally, the full length construct was cloned into the MCS of the plant binary vector pBIB-Hyg (Becker, 1990) as a *Acc65*I/*Sal*I fragment which resulted in the clone *pBIB-Hyg/AtSPO11-2 (4BE)*. As mentioned before a *Sma*I restriction site was introduced directly in front of the stop codon with the intention to fuse different tags to the C-terminus of the SPO11-2 protein. This was done by generating the clones *pBIB-Hyg/AtSPO11-2/24xHA (5BF), pBIB-Hyg/AtSPO11-2/18xmyc (6BG) and pBIB-Hyg/AtSPO11-2/Gal4-2 (7BH)*. The 18xmyc epitope and the Gal4-2 BD were generated as described in (Krsicka, 2007). The 24xHA epitope was generated as described in

section 6.1.1. All tags were fused as *Sma*I/*Eco*R*V* blunt end fragments to the C-terminus of *At*SPO11-2.

6.1.4 Selection and complementation analysis of transgenic lines

All of the transgenes were transformed into plants of a population originating from a heterozygous *Atspo11-2-3* mutant plant and selected as described in materials and methods using the hygromycin resistance marker of the pBIB-Hyg vector. Homozygous *Atspo11-2-3* mutants show a severe sterility phenotype, with shortened siliques, containing on average 1 seed/silique (compared to 40-50 in wild-type plants). After successful selection of transgenic lines, T1 transgenic lines with a single locus insertion, were propagated and monitored for homozygosity of the endogenous *Atspo11-2-3* mutant allele and the transgene by PCR and hygromicin resistance.

In general, successful complementing transgenic plants should rescue the sterility defect phenotype of *Atspo11-2-3* mutants. In case, the endogenous *Atspo11-2-3* mutant allele is homozygous and all plants are sterile, the transgene is not functional and not complementing the mutation. If the endogenous *Atspo11-2-3* mutant allele is homozygous and the transgene is heterozygous and functional, 75% of the plants are fertile and 25% are sterile. If the endogenous *Atspo11-2*-3 mutant allele is homozygous and the transgene is homozygous and functional, then all plants should be fertile.

Monitoring the fertility of the transgenic plant lines listed in figure 6.7 and 6.8, the wild-type like transgene *pBIB-Hyg/AtSPO11-2* was complementing the *Atspo11-2-3* mutant phenotype, restoring wild-type silique size, suggesting a functional expression of the respective fusion protein. *pBIB-Hyg/AtSPO11-2/18xmyc* lines show shorter siliques as wild-type plants, indicative for a partial complementation. The other two constructs, harboring a 24xHA tag or a Gal4-2 binding domain show the same sterility defect as *Atspo11-2-3* mutants, indicative for a non-functional transgene.
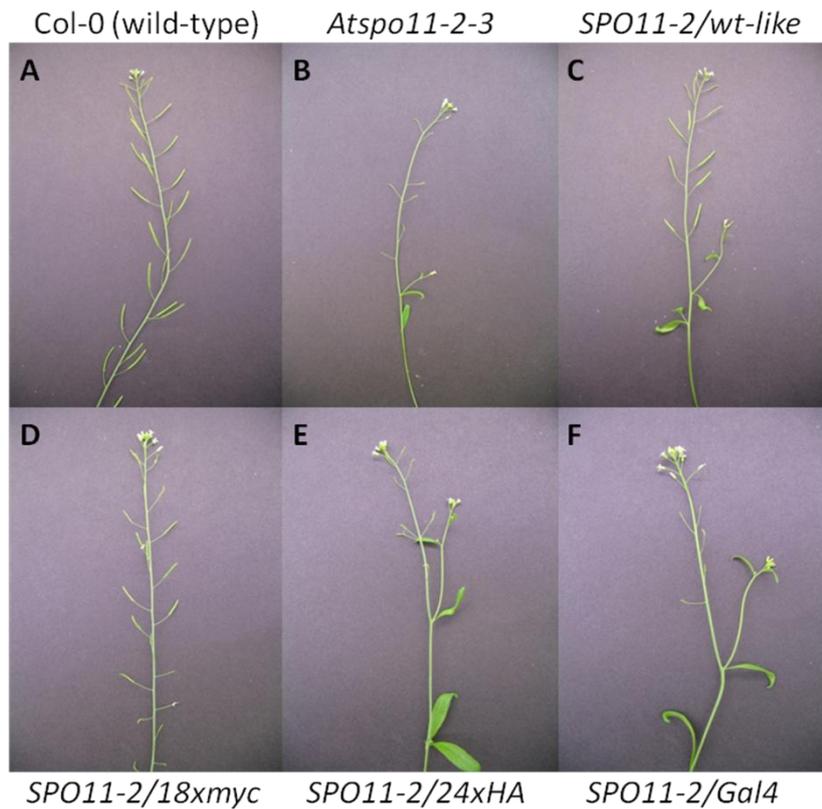
Figure 6.7: Transgenic *AtSPO11-2* lines. (A) fertile wild-type plant, ecotype Columbia, (B) sterile *Atspo11-2-3* mutant plant, (C) complementing, fertile *SPO11-2/wt-like* plant, (D) partial-complementing, semi-fertile *SPO11-2/18xmyc* plant, (E) non-complementing, sterile *SPO11-2/24xHA* plant, (F) non-complementing, sterile *SPO11-2/Gal4* plant.

Additionally to this genotypic and phenotypic analysis, a seed count was performed for all transgenic plant lines to assess the level of complementation compared to 100% fertile wild-type plants. For the *pBIB-Hyg/AtSPO11-2* transgenic lines, wild-type seed level was observed, *pBIB-Hyg/AtSPO11-2/18xmyc* lines showed a reduced seed number (12 seeds/silique), for the other two lines, the seed level was the same as for *Atspo11-2-3* mutant plants (see figure. 6.8).
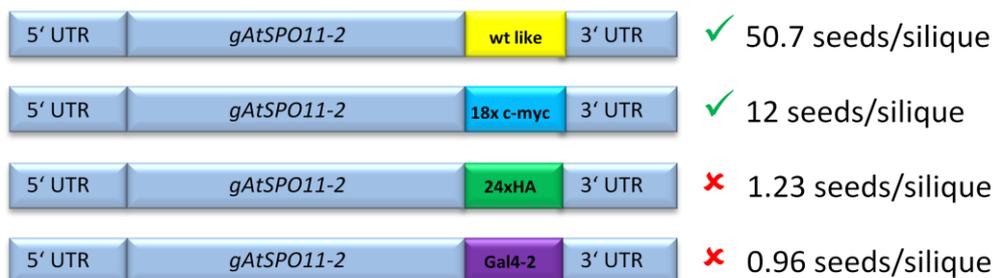
Figure 6.8: *At*SPO11-2 fusion proteins. Only the untagged transgenic line displays wild-type silique size and seed level, indicative for the expression of a functional fusion protein. The *pBIB-Hyg/AtSPO11-2/18xmyc* line exhibits a shorter silique size and contains only 12 seeds/silique, indicative for a partial complementation of the mutant phenotype. The other two lines (24xHA, Gal4-2) are sterile and therefore are not complementing the *Atspo11-2-3* mutant phenotype. The seed number/silique for respective transgenic lines is displayed in the figure (n=500-1000 siliques counted).

6.1.5 Biochemical detection of tagged *At*SPO11 fusion proteins

To determine the expression of the transgenes on protein level, immuno-precipitation experiments followed by Western blot analysis were performed as described in the materials and methods section. For the myc-tagged proteins the 9E11 anti-myc antibody (IMP, mouse monoclonal, cell culture supernatant) was used for the IP and the 9E10 clone (IMP, mouse monoclonal, cell culture supernatant) for Western blot detection. For the HA tagged proteins an anti-HA antibody was used (Covance, HA.11, clone 16B12, mouse monoclonal) for IP and Western blot detection. As a secondary antibody an anti-mouse HRP coupled antibody was used (Pierce, #31444, ImmunoPure, goat anti-mouse).

From previous studies it is known (Marc Behrlinger, personal communication), that the fusion of a protein with the long 18xmyc eptiope changes the running behavior of the protein on a SDS gel, resulting in a visible shift of up to 30 kDa. This was observed in the presented study for all 18xmyc tagged fusion proteins. In addition, this phenomenon was observed as well for proteins fused with the long 24xHA epitope.

In figure 6.9 (panel A) a Western blot performed for the *At*SPO11-1/18xmyc transgenic line is shown. The calculated molecular weight for the fusion protein is 68.9 kDa, the observed band which is marked with a red arrow has a size of about 90 kDa, which is consistent with the findings in previous studies mentioned above. The untagged sample and the mock IP sample (no protein extract added) show no bands with the same size, suggesting that the 90 kDa band represents the full length *At*SPO11-1/18xmyc protein (lane 1). The band above the 95 kDa marker band, which is present in all three lanes, may result as background from the antibodies or the beads.

Panel B shows the Western blot for the *At*SPO11-2/18xmyc transgenic line. Here, the calculated molecular weight of the protein is 70.2 kDa, the band marked with a red arrow, which is representing the full length *At*SPO11-2/18xmyc protein has a size of about 90 kDa (lane 3). The calculated molecular weight of *At*SPO11-2/18xmyc is 3 kDa higher than for *At*SPO11-1/18xmyc, which is not visible on the gel (compare lane 2+3) at this resolution.
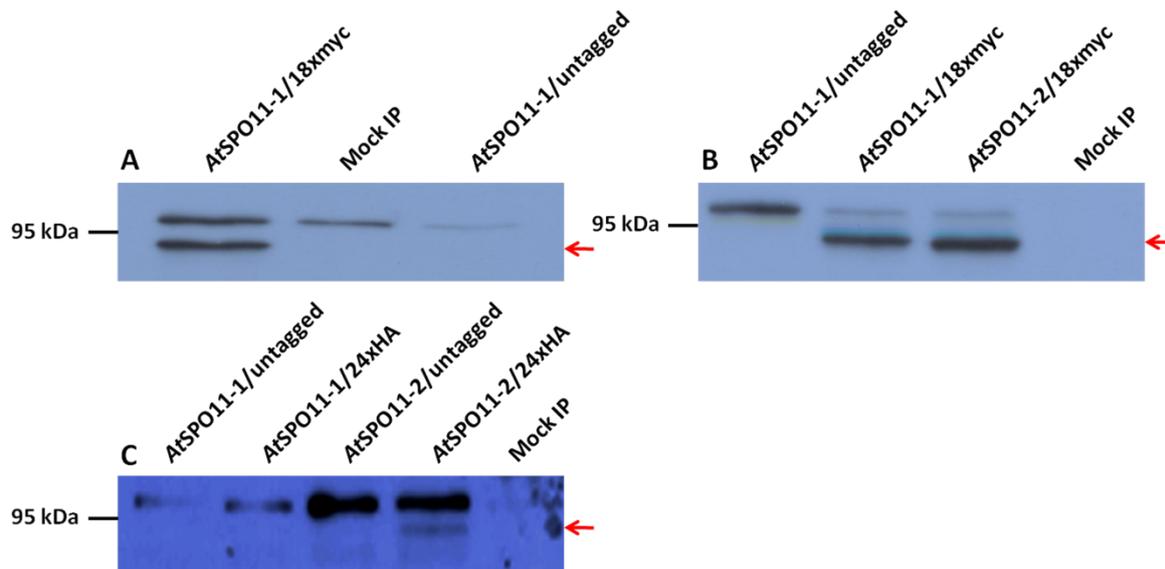
Figure 6.9: Biochemical detection of *At*SPO11-1 and *At*SPO11-2 fusion proteins. Immuno-precipitation from protein extracts, made from buds of transgenic plant lines, was performed using protein G sepharose coupled with either an anti-myc (9E11) or anti-HA antibody. Proteins were separated by SDS PAGE and transferred to a PVDF membrane. Detection was performed using either an anti-myc (9E10) or anti-HA antibody as primary antibody and a HRP linked anti-mouse secondary antibody. Bands were visualized as described in materials and methods using ECL. (A) Western blot for the *At*SPO11-1/18xmyc line. Lane 1: *At*SPO11-1/18xmyc, lane 2: mock IP (no protein extract added), lane 3: *At*SPO11-1/untagged. The red arrow shows the band, which represents the full length *At*SPO11-1/18xmyc protein with a molecular weight of about ~90 kDa. The upper band, which is present in all 3 lanes, may result as background from the antibodies or the beads (for all blots). (B) Shows the Western blot for the *At*SPO11-2/18xmyc line. Lane 1: *At*SPO11-1/untagged, lane 2: *At*SPO11-1/18xmyc, lane 3: *At*SPO11-2/18xmyc, lane 4: mock IP. The red arrow shows the band, which represents the full length *At*SPO11-1/18xmyc and the *At*SPO11-2/18xmyc proteins with a molecular weight of about ~90 kDa. (C) Shows the Western blot for both constructs harboring the 24xHA tag. Lane 1: *At*SPO11-1/untagged, lane 2: *At*SPO11-1/24xHA, lane 3: *At*SPO11-2/untagged, lane 4: *At*SPO11-2/24xHA, lane 5: mock IP. The red arrow shows the band, which represents the full length *At*SPO11-2/24xHA protein with a molecular weight of about ~90 kDa. *At*SPO11-1/24xHA was never detected. For all calculated molecular weights please refer to the text.

Panel C shows the Western blot for both transgenic lines harboring the 24xHA tag. Both lines are not complementing the respective mutant phenotypes, indicative for missing expression or non-functionality of the fusion proteins. Although *At*SPO11-1/24HA (calculated molecular weight: 71.9 kDa) could not be detected on the protein level, a band with a size of ~90 kDa is visible in lane 4, which may represent the full length *At*SPO11-2/24xHA protein. The calculated molecular weight is 73.3 kDa.

Although complementing respective mutant phenotypes, *At*SPO11-1/Gal4-2 and *At*SPO11-1/SF were never detected on protein level. *At*SPO11-2/Gal4-2 is not complementing the *Atspo11-2-3* mutation and was never detected on the protein level (data not shown).

6.1.6   Generation of 5xUAS lines

During his diploma thesis, Ondrej Krsicka started to generate the plant binary vector pGreenII0179, harboring five UAS tandem repeats as described (Krsicka, 2007). I took over the plasmid pGreenII0179/5xUAS short, performed an *Agrobacterium* transformation, plasmid DNA isolation and re-transformation in *E. coli*, followed by restriction digest analysis, to check whether the plasmid remained unmodified in the *Agrobacterium* strain. Since an unexpected banding pattern, suggesting a deletion of sequences within the plasmid, was observed, I started to reclone the construct from the plasmid version pGreenII0179/5xUAS long. The plasmid was digested with *Xba*I and religated, to eliminate a dispensable sequence stretch of 196 bp, which was amplified from the original 5xUAS plasmid and was necessary for the cloning strategy. This again resulted in the plasmid pGreenII0179/5xUAS short. After successful sequencing of the plasmid, *Agrobacterium* transformation, plasmid DNA isolation, re-transformation in *E. coli* and restriction digest analysis was performed. Since successful cloning was verified, *Agrobacterium* mediated plant transformation in *Arabidopsis* wild-type plants, ecotype Columbia, was performed. All methods are described in the materials and methods section.

Seeds from outgrowing siliques were harvested (sT0) and sown onto hygromycin containing Ara plates to identify positive transformants (T0). 68 individual transformants were identified in this generation, transferred onto soil and seeds harvested separately for further selection (sT1). To identify 3:1 segregants, indicative for a heterozygous single locus insertion, lines were grown in the T1 generation and the ratio hygromycin resistant/hygromycin sensitive was examined. 19 3:1 segregating lines were found, which were further applied to TAIL PCR to map the transgenic insertion sites. These lines are: #1, 4, 6, 14, 15, 19, 20, 21, 23, 34, 42, 43, 50, 51, 54, 55, 58, 59 and 66.

6.1.7   Mapping of insertion sites

It is necessary to know the sites of the transgenic insertions for choosing appropriate FTL (fluorescent-tagged line) marker for the artificial hotspot project, which is described later. Mapping of insertion sites was performed using TAIL (thermal asymmetric interlaced) PCR which is described in (Liu *et al.*, 1995b). A simplified version of the protocol, which was developed by Tuncay Baubec and Ortrun Mittelsten Scheid (GMI, Vienna) was used, which is described in the materials and methods section.

For the lines 5xUAS#1, 6, 19, 20 and 21 the transgenic insertion sites could be determined using TAIL PCR. BLAST results can be found in the appendix. For confirmation of the mapping results, PCR using T-DNA specific and genomic insertion site specific primers, was performed. For a detailed summary see table and figure below.

5xUAS line	Chromosome/ Insertion site	T-DNA primer	Genomic primer	Amplicon size (bp)	Ann. Temp.
#1	Chr. II, 16363755	5xuas T-DNA confirm	5xUAS#1 cnfrm	1068	57°C
#6	Chr. III, 10401824	5xuas T-DNA confirm	5xuas#6gDNA	1208	57°C
#19	Chr. III, 17966362	5xuas T-DNA confirm	5xUAS#19B cnfrm	906	56°C
#20	Chr. II, 13006968	5xuas T-DNA confirm	5xuas#20gDNA	1159	57°C
#21	Chr. I, 3560629	5xuas T-DNA confirm	5xuas#21gDNA	910	57°C

Table 6.1: Mapped insertion sites and PCR parameter for 5xUAS lines. Insertion sites for successfully mapped 5xUAS lines #1, 6, 19, 20 and 21 are listed, including primerpairs and PCR conditions which were used to confirm the mapping results. For all insertions, 30 PCR cycles were applied.

Figure 6.10 shows the amplicons from the PCR made to confirm the mapping results for the 5xUAS lines #1, 6, 19, 20 and 21 after TAIL PCR. All fragments coincide with the expected size which is listed in table 6.1. For 5xUAS#6, 20 and 21, fragments with the size of 1208 bp, 1159bp and 910 bp were observed (panel A), for the 5xUAS#1 line a 1068 bp amplicon is visible on the gel (panel B) and for 5xUAS#19 the expected size of 906 bp was examined (panel C). In panel A, some unspecific bands are visible in the wild-type lanes for 5xUAS#6 and #20, suggesting that wild-type DNA produces some unspecific products with respective primer combinations.
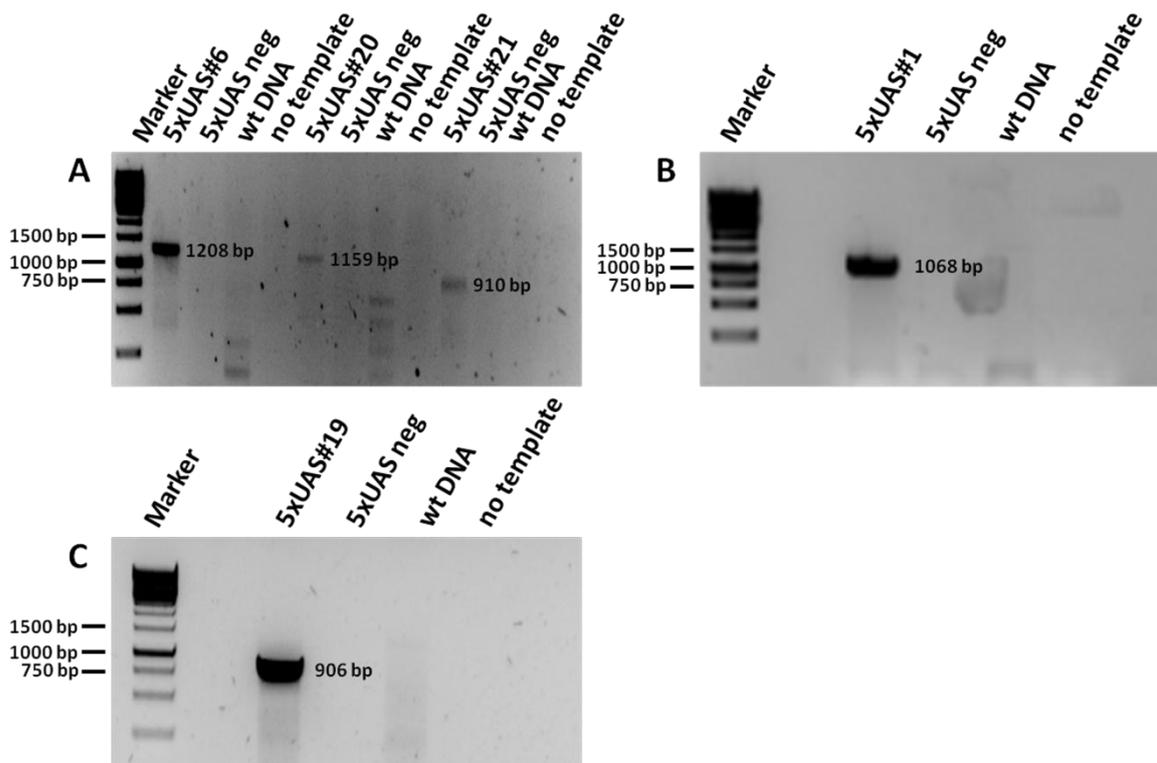
Figure 6.10: Confirmation of the mapping results for 5xUAS lines #1, 6, 19, 20 and 21 by PCR. All fragments, separated by standard agarose gel electrophoresis, coincide with the expected size which is listed in table 6.1. For 5xUAS#6, 20 and 21, fragments with the size of 1208 bp, 1159bp and 910 bp were observed (panel A), for the 5xUAS#1 line a 1068 bp amplicon is visible on the gel (panel B) and for 5xUAS#19 the expected size of 906 bp was examined (panel C). As a marker the 1kb ladder (Fermentas) was used. 5xUAS neg means that the DNA from a 5xUAS line with a different insertion site, was used as a negative control and therefore should not be amplified with respective primers.

6.2 Genome-wide identification of DNA DSB sites using ChIP-Seq

Since the biological relevance to study interactions of proteins with DNA in the context of different aspects has increased in the last decade, chromatin immunoprecipitation (ChIP) of DNA bound proteins has become a method of interest. In the past, the DNA sequences bound to the proteins were exclusively analyzed by hybridizing processed and labeled DNA to microarrays (ChIP-on-chip). Due to the higher resolution, the huge amount of data output to answer biological questions and the decreasing costs, deep sequencing of the DNA fragments is becoming a more and more powerful tool to answer such questions.

Nevertheless, as a first attempt to identify DSB sites in *Arabidopsis thaliana*, a ChIP-on-chip approach was started. As input material 1 g buds of the transgenic plant lines *AtSPO11-1/18xmyc* (sample) and *AtSPO11-1/untagged* (control) was used. All experimental steps, like DNA fragmentation, ChIP, labeling and amplification steps worked properly and there was enough DNA for hybridization to the *Arabidopsis* tiling array (Affymetrix). Two microarray hybridizations (biological repeats) were made in the lab of Martin Bilban (Medical University of Vienna, AKH), one hybridization (technical repeat) was performed in the European Affymetrix Service facility (NASC, Nottingham). The raw data processing and visualization of the data was performed by Anne Kupczok (CIBIV, Vienna). The same results were obtained for all three experiments, mainly no hybridization signals representing DSB sites in *Arabidopsis thaliana*, but background signals distributed over all five chromosomes for the sample and the control. In total, only three regions with signals above the background level were identified. The data for these regions was not reproducible between the single experiments. These regions are a 2000 bp region on chromosome I (around position 23,748,000), a 4000 bp region on chromosome III (7,266,000) and one single spike, representing a hybridization at one probe, on chromosome V (1,964,000). The best explanation for these results is that the amount of input material was too low (read more in the discussion).

Considering these results and the advantages of ChIP-Seq experiments, we decided to start a ChIP-Seq approach. After the generation of DSBs, Spo11 proteins are removed from DNA covalently attached to one short oligonucleotide. The 5' ends of these oligonucleotides precisely represent the positions where DSBs are made by the catalytic tyrosine residue of Spo11. Direct sequencing of these oligonucleotides in a high-throughput manner will result in the sequence information to generate genome-wide DSB maps and will help to gain much more insight into the process of Spo11-dependent DSB formation.

The aim of this project was to establish a technique for the genome-wide identification of DNA DSB sites in various organisms. The final aim is to establish the protocol for the model organism *Arabidopsis thaliana*. Until the end of this thesis, the protocol was developed and evaluated using a HRP protein conjugated with oligonucleotides as a model substrate and further applied to the two yeasts *S. cerevisiae* and *S. pombe*. For these two organisms results are highly reproducible and perfectly correlate with already published hotspot maps (Lichten, Hochwagen, Smith labs). In two experiments for *Arabidopsis thaliana*, the first genome-wide hotspot maps at nucleotide resolution were generated.

For a basic overview of the ChIP-Seq strategy please refer to figure 6.11, the text below and the detailed protocol described in the materials and methods section. The development and optimization of the protocol with the model substrate is described in section 6.2.1.
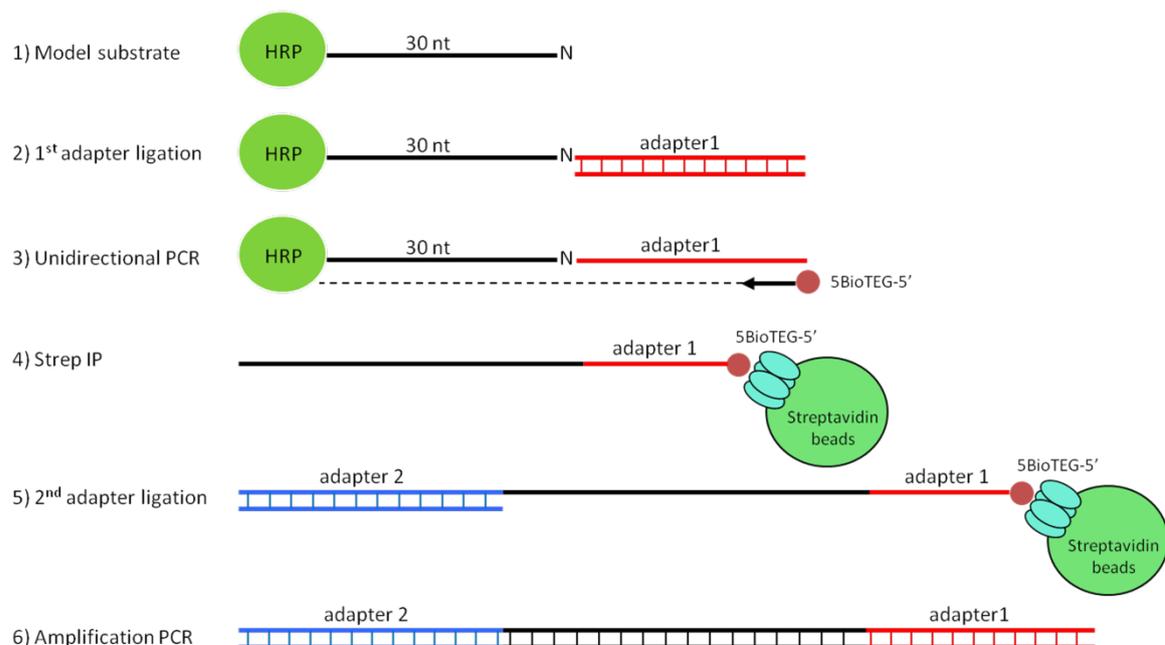
Figure 6.11: Basic overview of the ChIP-Seq strategy. Using a HRP protein coupled with a 30mer oligonucleotide of known sequence (position 30 is "N" to see whether there is a bias in the first adapter ligation) we established the experimental procedure. After ligation of the 1st Solexa adapter to the model substrate, the ligation product was amplified in an unidirectional PCR using a primer conjugated with a biotin label at the 5' end. Recovery of the PCR product was managed by interaction of the biotinylated amplicons with streptavidin coupled magnetic beads. The ligation of the 2nd Solexa adapter was performed on beads, the next step was the exponential amplification of the sequences using two adapter specific primers. PCR products were purified and sent to the sequencing facility.

For the model substrate no chromatin immuno-precipitation (ChIP) was necessary, for all other model organisms ChIP was performed to isolate Spo11 attached to one short

oligonucleotide. The crucial steps of the protocol are the ligations of both Solexa adapters. Both adapters are compatible with the Solexa system and allow the binding of the amplified products on the flow cell matrix for cluster formation. In addition, the second adapter contains a sequence stretch, which allows the binding of the Solexa sequencing primer. After the ligation of the first Solexa adapter, a unidirectional PCR reaction using a biotinylated primer was performed. The PCR product was recovered by performing a pull-down with Ademtech Streptavidin beads. After the second Solexa adapter ligation and stringent washing the adapter containing fragments were amplified in 30 cycles of a standard PCR reaction using adapter specific primers. The PCR products were loaded on a 3% agarose gel, cut out and gel-extracted. After purification of the PCR products, samples were sent to the Solexa sequencing facility for further processing.

6.2.1 The model substrate

As a model substrate a HRP protein coupled with up to three 30mers was used (Eurogentech, HRP-5'-ACTAGCCGACCTTGCTGACTATGCCATGCN-3'). Position 30 of the oligonucleotides is a "N", to see whether there is a ligation efficiency bias towards one nucleotide during the first adapter ligation. At the beginning of this project different strategies for the ligation of the Solexa adapters to the single stranded oligonucleotides coupled to the HRP protein were tested. Since approaches using Prokaria single strand ligase were not successful, a different protocol was developed and evaluated. Using terminal transferase and rGTPs, a controlled addition of Gs is possible, since it is known from (Schmidt and Mueller, 1996) that at a rGTP concentration of 0.5 µM, 2 or 3 Gs are added to an extent of 98%. After purification of the reaction using Vivaspin 500 columns (Sartorius), the first adapter ligation was performed using a double stranded adapter with a "CCN" overhang which is compatible with the 2 or 3 rGTPs. All steps of the protocol were experimentally tested and worked very efficiently. But in terms of the "real" experiment using different model organisms we thought about a much more efficient way for adapter ligation. The final protocol, which is described in detail in the materials and methods section, uses Solexa adapters with a five "N" overhang which is directly compatible for ligation without adding rGTPs. The use of these adapters has the advantage to get rid of two enzymatic reactions with subsequent column purification steps, avoiding the potential loss of input material.

In contrast to the protocol described in the materials and methods section, 2 µl HRP oligonucleotide [13,6 pmol/µl] were used as input for the first adapter ligation. After

purifying the first ligation using Vivaspin columns, the protocol is identical to the described one.

Since the length of the model substrate oligonucleotide and the Solexa primer is known, the expected size of the final amplicon, containing both Solexa adapters, was 125 bp. As negative controls samples were amplified using either Solexa primer 1 or 2, no primers or no template with both primers.

Figure 6.12 shows the final amplification for the model substrate. Only in lane 2, where the PCR reaction was performed with both Solexa primers, the expected amplicon with a size of 125 bp could be observed. Since all negative controls were negative, the band in lane 2 was excised from the gel, purified and the samples were delivered for sequencing.
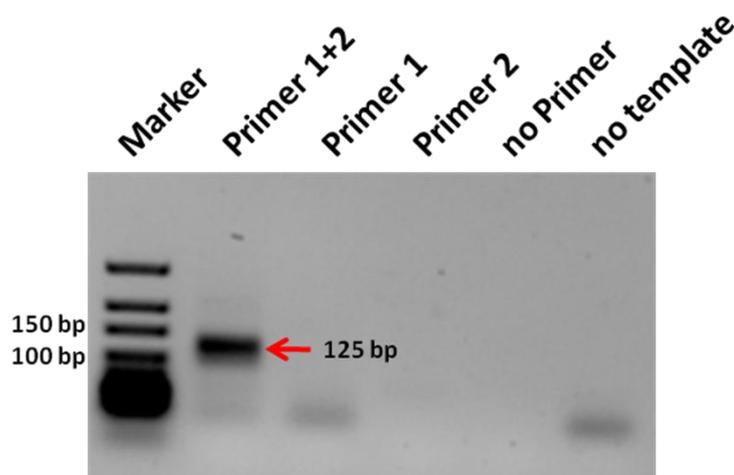
Figure 6.12: Solexa amplification using the model substrate. After ligation of both Solexa adapters, the model substrate was amplified using adapter specific primers. As negative controls, reactions containing either one of the primer, no primer or no template were performed. In lane 2, where the amplification was performed with both primers, an amplicon with the expected size of 125 bp is visible. After purification of the PCR products, samples were sent to the Solexa sequencing facility for further processing.

For this Solexa run 14.5 million total reads were obtained, 12.3 million perfectly matching the model substrate sequence, indicative for a very well developed protocol. A ligation efficiency bias towards one nucleotide (A: 20%, C: 25%, T: 40%, G: 15%) was observed, which may have different reasons. Maybe the bias is the result of an unbalanced synthesis of the model substrate oligonucleotides. Therefore the composition of the used model substrate should be checked by mass spectrometry. Prior to all other speculations, to exclude experimental discrepancies, the experiment has to be repeated, to see whether the same bias is obtained in a second experiment.

6.2.2 *Saccharomyces cerevisiae* ChIP-Seq

To apply the developed protocol to a "real" model organism, we have chosen *S. cerevisiae* due to different reasons. First, *S. cerevisiae* is easy to handle and a huge amount of synchronized cells at the accurate meiotic stage can be cultivated in a short period of time. Second, *S. cerevisiae* gives us the possibility to evaluate the data, since different DSB maps for data-comparison are already published. As described in materials and methods, 5×10^9 cells of a SK1 strain harboring a 18xmyc tagged Spo11 (MATa/alpha, ho::LYS2, ura3, leu2::hisG, trp1::hisG, SPO11-Myc18::TRP, 4 hours after induction of meiosis) were used as input for the identification of genome-wide DSB sites using ChIP followed by deep sequencing. After ChIP, ligation of both adapters and the amplification PCR, products were visualized by agarose gel electrophoresis. The marked area in panel A, above the dominant band, which results from a dimer of the unidirectional primer and the second adapter, was excised from the gel, the DNA was extracted and re-amplified as described. In contrast to the untagged control, for the tagged sample a band-shift was visible, indicating the amplification of sequences which were attached to Spo11 and containing both Solexa adapters (panel B). The marked areas were cut out, DNA was extracted and sent to the sequencing facilty. We performed one 72 bp and one 36 bp single end run (technical replicates). DSB profiles for both data sets were plotted at chromosome resolution, overlaid and compared, resulting in a perfect overlap of both data sets at this resolution. This means that at this resolution signals from the one data set were always accompanied by signals of the other data set.
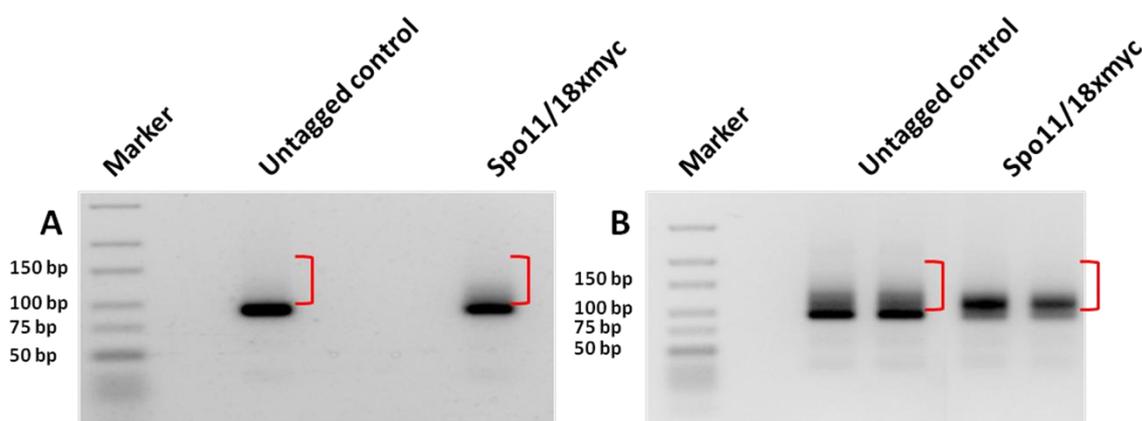
Figure 6.13: Solexa amplification (A) and re-amplification (B) with adapter specific primers. After ChIP, ligation of both adapters and the amplification PCR, products were visualized by agarose gel electrophoresis. The marked area in panel A, above the dominant band which results from a dimer of the unidirectional primer and the second adapter, was cut out from the gel the, DNA was extracted and re-amplified as described. In contrast to the untagged control, for the tagged sample a band-shift was visible, indicating the amplification of sequences which were attached to Spo11 and containing both Solexa adapters (panel B). The marked areas were cut out again, DNA was purified and sent to the sequencing facilty.

In addition, some regions were randomly chosen, and data-comparison at nucleotide resolution was performed, also resulting in a perfect overlap between both data sets.

From the 36 bp run, 26 million total reads were obtained for the sample. 7.9 million were used for mapping, 5.5 million mapped to the *S. cerevisiae* genome, thereof 4.2 million uniquely. For the control 28 million sequences reads were obtained and 8.6 million used for mapping. 8.3 million mapped to the *S. cerevisiae* genome, 1.1 million uniquely. The mapping statistic for the 72 bp run, where only a tagged sample was processed, is as follows: 8.4 million total reads, 3.3 million were used for mapping, 1.8 million mapped to the *S. cerevisiae* genome, thereof 1.5 million uniquely. It has to be mentioned that the number for mapped reads consists of the number of mapped reads plus the number of positions where the reads mapped more than once. Since the SK1 genome assembly is incomplete, data was mapped to the S288C genome. For all further analysis (except long oligonucleotide distribution) the unique data set of the 36 bp run was used.
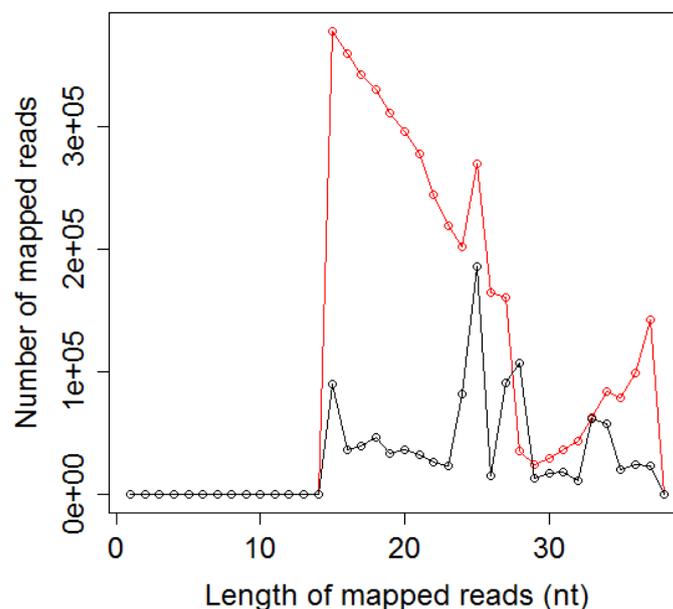
Figure 6.14: Read length distribution of uniquely mapped sequence reads for *S. cerevisiae*. In a 36 bp run about 4.2 million reads uniquely mapped for the sample (red), 1.1 million for the control (black).

Figure 6.14 shows the unique read length distribution of a 36 bp single end run for *S. cerevisiae*. With this data set, looking on the read length distribution profile, the existence of two oligonucleotide classes attached to Spo11 could be suggested. Combining these results one can say that we observe two classes of oligonucleotides attached to Spo11, one class is between 15 and 25 nt in length, the other class between 30 and 37 nt, which is consistent with the findings published by (Neale *et al.*, 2005). They also reported the

existence of two oligonucleotide classes, one half between 7-12 nt, the other half between 21-37 in length. Sequence reads shorter than 13 nt were not included in our analysis due to mapping ambiguity. The profile of the read length distribution for the 72 bp run shows the same pattern as described for the 36 bp run. In addition, the existence of Spo11-attached oligonucleotides, mapping to the *S. cerevisiae* genome, with a length up to 72 bp could be detected. For a detailed analysis of the long oligonucleotide distribution please refer to section 6.2.2.6.

6.2.2.1   A genome-wide DSB map at nucleotide resolution

For further analysis the unique data set of the 36 bp run was used. Coverage files for all *S. cerevisiae* chromosomes were produced as described in materials and methods. Since the coverage values for the background control were very low and rarely spread over the genome compared to the sample, the control was subtracted from the sample. The coverages were plotted using the free statistical software "R" to produce genome-wide DSB maps. Here we present the first genome-wide DSB map at nucleotide resolution for *S. cerevisiae*. An example for chromosome I is shown in the figure below, the maps for all other chromosomes can be found in the appendix.

S. cerevisiae
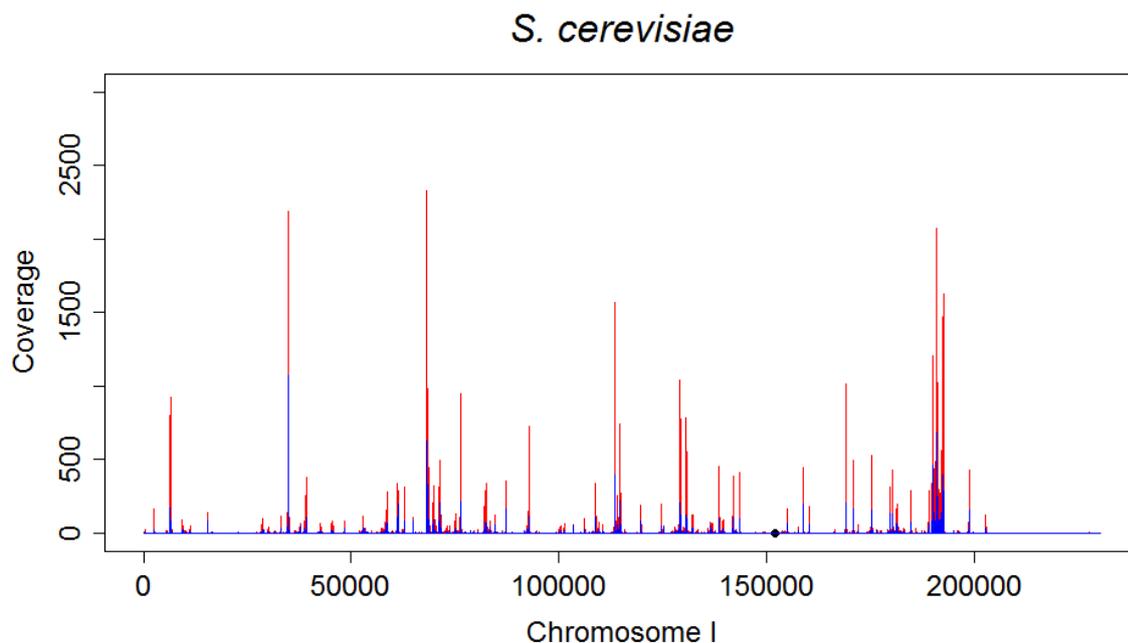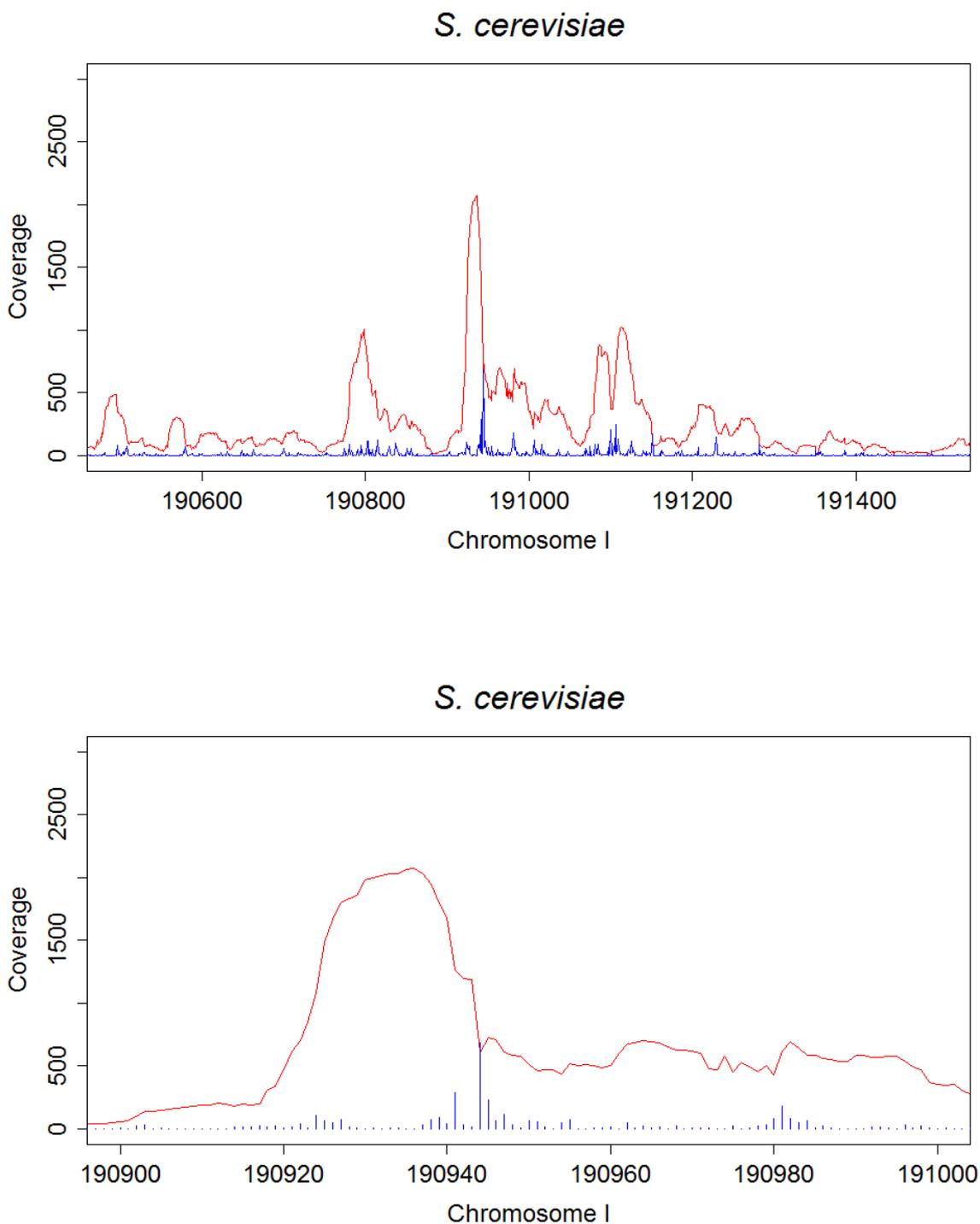
S. cerevisiae

Figure 6.15: DSB distribution for chromosome I. A 36 bp single end Solexa run was performed using 5×10^9 *S. cerevisiae* cells from a diploid strain with a 18xmyc tagged SPO11 (MATa/alpha, ho::LYS2, ura3, leu2::hisG, trp1::hisG, SPO11-Myc18::TRP, 4 hours after induction of meiosis). As control 5×10^9 cells of the untagged version were processed according to the protocol described in materials and methods. After sequencing, base-calling was performed, the adapter sequences were clipped and the sequence reads were mapped to the S288C genome using NextGen. Coverage files for the chromosomes were generated, and the data was plotted using the free statistic program "R". The red signals represent the full length oligonucleotides. In blue the exact 5' end positions for every mapped oligonucleotide are plotted. The coverage is displayed in number of mapped nucleotides/position. The chromosomes on the x-axis are displayed in bp. The black filled circle indicates the centromere position. The first plot shows the whole chromosome I, the other two plots show zooms.

Figure 6.15 shows an example for a DSB map of chromosome I. In red the mapped full length reads are plotted, in blue the 5' ends of the mapped reads are plotted, which exactly represent the DSB sites, where Spo11 catalytically acts. It is visible that the 5' end signals always merge with the full length data. The last plot shows a zoom at nucleotide resolution, where every blue signal represents one DSB site. As expected the coverage for the mapped 5' ends is not that high as for the total mapped reads, because every oligonucleotide produces just the information for one position (5' end), resulting in no overlap of one oligonucleotide with another at more than one position.

6.2.2.2 *Comparison to published DSB maps*

For further evaluation, our data was compared with the genome-wide DSB map published by the Lichten lab (Buhler *et al.*, 2007). They developed a method to map genome-wide, single-strand DNA (ssDNA)-associated DSBs that accumulate in the processing-capable, repair-defective *dmc1Δ* mutant. As visible in figure 6.16, a high degree of overlap was observed between both data sets. For all other chromosomes also a high degree of overlap was examined (data not shown). For this analysis our datasets including the coverage values for the full length mapped reads were used.
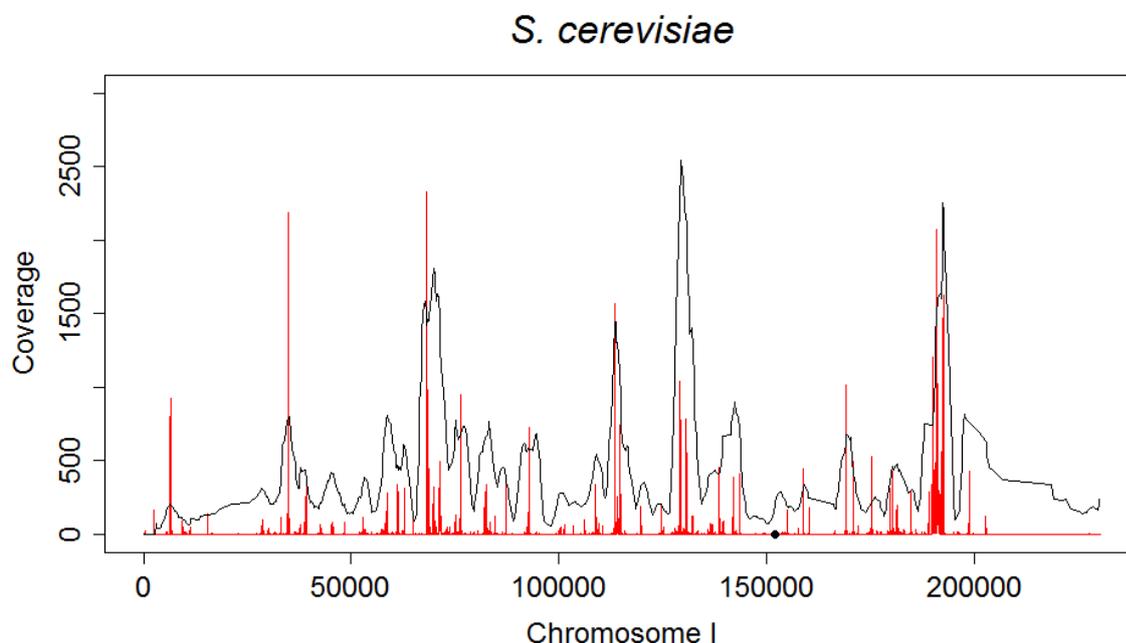
Figure 6.16: Comparison of own data with Lichten data. The own full length oligonucleotide ChIP-Seq data (red) was compared to the de-noised *dmc1Δ* ratios (black) published by (Buhler *et al.*, 2007). For reasons of easier visualization the peak intensities of the *dmc1Δ* data were multiplied by 100. Here an example for chromosome I is shown. The coverage for the ChIP-Seq data is displayed in number of mapped nucleotides/position. The chromosome on the x-axis is displayed in bp. The black filled circle indicates the centromere position.

For a more detailed comparison of both data-sets, a bioinformatical association study was performed by Stefanie Tauber (CIBIV, Vienna). The 95% quantile was calculated for the ChIP-Seq data (for each chromosome separately), this means that for the ChIP-Seq data only 5% of the positions with the highest coverage were included in the analysis. In this case, the cut-off value for the 95% quantile corresponds to a coverage-value of 17.63. For the ChIP-on-chip data, positions with a signal/background ratio of >5 were taken. These values were taken as cut-off for exhibiting a significant coverage value in the ChIP-Seq case or showing a significant ratio in the ChIP-on-chip case. Then the maximum within 2 kb windows (side by side, along the chromosome) was calculated for both and overlaps were detected. For the ChIP-Seq data, 4158 significant signals were detected, for the ChIP-on-chip data only 2499. 2117 were detectable in both data sets, which corresponds to an overlap of 84.71%. For a detailed overview please see table 6.2.

Chromosome	ChIP Seq and ChIP-on-chip	ChIP-Seq only	ChIP-on-chip only	ChIP-Seq total	ChIP-on-chip total	Overlap %
ChrI	34	35	8	69	42	80.95
ChrII	142	147	29	289	171	83.04
ChrIII	51	52	10	103	61	83.61
ChrIV	270	269	48	539	318	84.91
ChrV	97	89	22	186	119	81.51
ChrVI	44	46	7	90	51	86.27
ChrVII	194	177	29	371	223	87.00
ChrVIII	91	122	14	213	105	86.67
ChrIX	64	88	11	152	75	85.33
ChrX	117	118	34	235	151	77.48
ChrXI	126	104	23	230	149	84.56
ChrXII	165	195	26	360	191	86.39
ChrXIII	178	150	31	328	209	85.17
ChrXIV	160	129	26	289	186	86.02
ChrXV	204	178	35	382	239	85.36
ChrXVI	180	142	29	322	209	86.12
Total	2117	2041	382	4158	2499	84.71

Table 6.2: Genome-wide correlation of ChIP-Seq data with Lichten data. The 95% quantile was calculated for the ChIP-Seq data (for each chromosome separately), this means that for the ChIP-Seq data only 5% of the positions with the highest coverage were included in the analysis. In this case, the cut-off value for the 95% quantile corresponds to a coverage-value of 17.63. For the ChIP-on-chip data, positions with a signal/background ratio of >5 were taken. These values were taken as cut-off for exhibiting a significant coverage value in the ChIP-Seq case or showing a significant ratio in the ChIP-on-chip case. Then the maximum within 2 kb windows (side by side, along the chromosome) was calculated for both and overlaps were detected. For the ChIP-Seq data, 4158 significant signals were detected, for the ChIP-on-chip data only 2499. 2117 were detectable in both data sets, which corresponds to an overlap of 84.71%.

6.2.2.3 Most of the hotspots are intergenic rather than intragenic

Next, the occurrence of DSBs with regard to gene coding regions was examined. 56 hot regions, distributed all over the genome, were chosen and their location was analyzed. The requirement for the selection of a region was the clustering of DSB sites without any gaps of cold regions bigger than 50 bp between. The regions had a size between 50 and 3000 bp with an average of 590 bp and with an average coverage maximum of 3197. By definition, a hotspot in *S. cerevisiae* has a maximum width of 500 bp, therefore the larger regions contain more than one hotspot, but were taken as one hot region. From these 56 regions, 47 were determined to be intergenic, lying in promoters, only 9 were intragenic, which correlates with the findings published by (Blitzblau *et al.*, 2007). They examined the ssDNA enrichment (the more ssDNA enrichment, the hotter a region) across the largest 226 hotspot-associated genes and found that 80% of the genes exhibited higher ssDNA signal in their promoters relative to the coding regions. Together, these findings are consistent with the model that the majority of DSBs occur in intergenic regions containing promoters (Blitzblau *et al.*, 2007).

6.2.2.4 Substantial DSB formation in pericentromeric regions

Hotspot maps for *rad50S* like mutants, published by (Gerton *et al.*, 2000), showed that DSBs are absent within the pericentromeric region (centromere +/- 25 kb). Although DSB formation is repressed in such cohesin-protected regions, following studies detected the occurrence of pericentromeric DSBs within 5-10 kb of the core centromere (Blitzblau *et al.*, 2007; Buhler *et al.*, 2007). With our method we were able to detect substantial DSB activity within 1-5 kb of the core centromeres for all chromosomes. For chromosomes V, VI, IX, and XI substantial DSB activity was observed within 1 kb of the core centromeres, for chromosomes II, III, IV, VII, XV and XVI within 2 kb of the core centromeres, for chromosomes I, VIII, XII, XIII, and XIV within 3 kb of the core centromeres and for chromosome X within 5 kb of the core centromere. Although the hotness of DSBs decreases, the closer to the centromere they are located, the frequency of break sites exhibits chromosome average level also in these regions. As shown in figure 6.17, 17 hotspots were detected in the pericentromeric region of chromosome I, which perfectly correlates with the expectations from a model of random distribution (Blitzblau *et al.*, 2007).
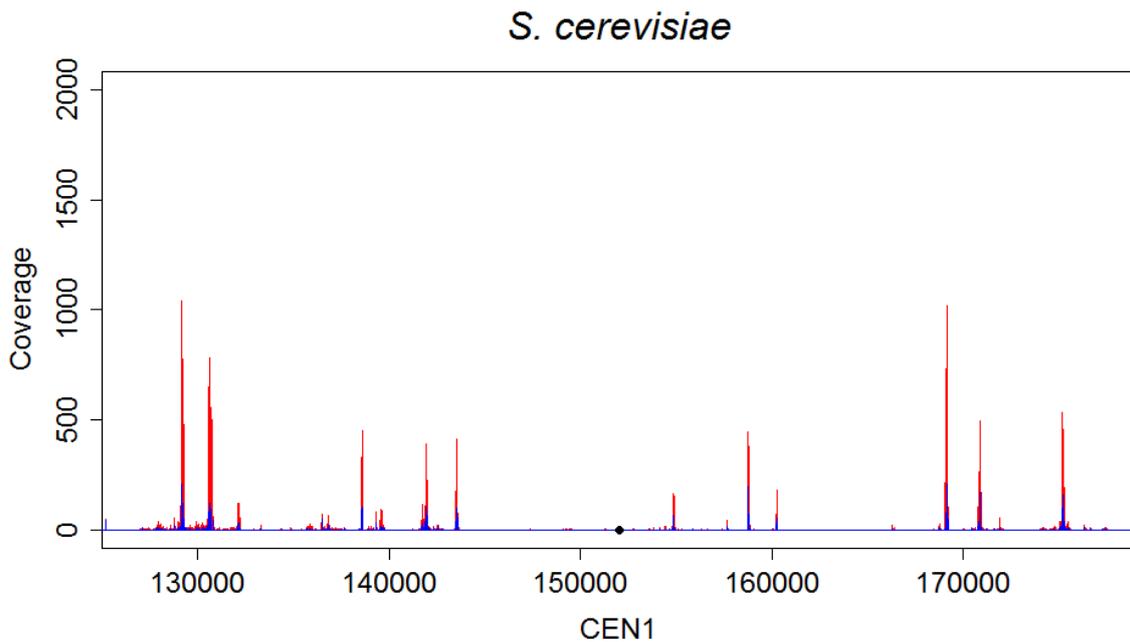
Figure 6.17: Occurrence of pericentromeric DSBs in *S. cerevisiae.* Figure 6.17 shows the pericentromeric region of *S. cerevisiae* chromosome I (centromere +/- 25 kb plotted). Although DSB formation is repressed in the cohesin-protected regions round the core centromeres, many DSBs were detected in the pericentromeric region. The same was also shown for all other chromosomes. The red signals represent the mapped full length oligonucleotides. In blue the exact 5' end positions for every mapped oligonucleotide are plotted. The coverage is displayed in number of mapped nucleotides/position. The chromosome on the x-axis is displayed in bp. The black filled circle indicates the centromere position.

6.2.2.5 Negative correlation of mapped Spo11 oligos with Rec8 binding sites

In meiosis, the cohesin Rec8 plays important roles in chromosomal metabolism. In *S. pombe,* Rec8 mutants exhibit a loss of proper segregation of homologous chromosomes and mono-oriented kinetochores. In addition, a reduction of meiosis-specific DNA breakage by Rec12 (Spo11) was observed in several intervals of the genome, but less in others (Ellermeier and Smith, 2005). In *S. cerevisiae rec8Δ* mutants, meiotic DSB formation and the binding of Spo11 to DSB sites are severely impaired at selective domains of many chromosomes. The results from Kugou and colleagues suggest that Rec8 at centromeres and cohesion sites choreographs the distribution of Spo11 to DSB sites during pre-meiotic DNA replication (Kugou *et al.*, 2009). As shown in figure 6.18, mapped Spo11 oligonucleotides negatively correlate with the binding sites of the meiosis-specific cohesin Rec8, as expected from prior studies (Kugou *et al.*, 2009). Especially at the centromere of chromosome IV, where high signals for Rec8 binding could be observed (green), only very low signals for Spo11 binding are present (red). For further information please refer to the discussion.
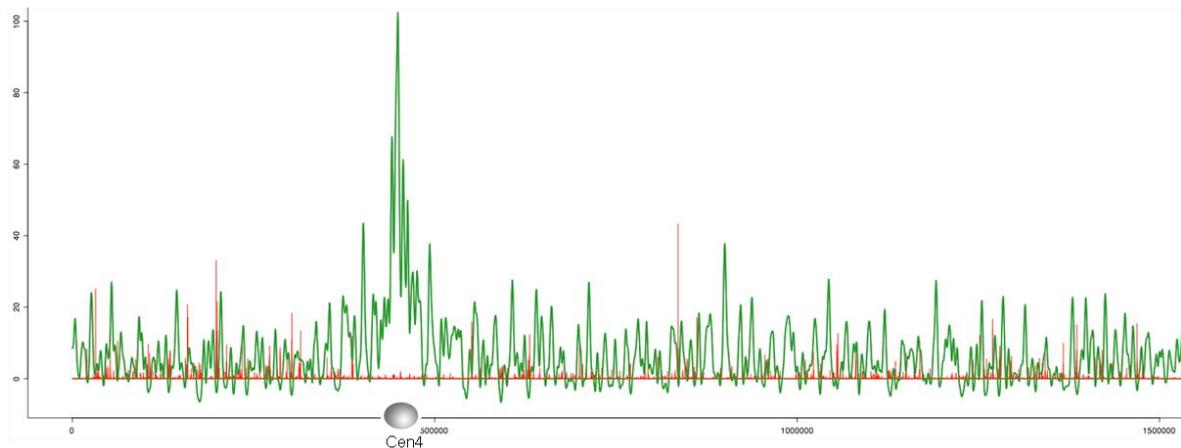
Figure 6.18: Negative correlation of mapped Spo11 oligonucleotides with Rec8 binding sites for chromosome IV. Data from our ChIP-Seq experiments (red) was compared with ChIP-on-chip data for Rec8 (green) from the Franz Klein lab. Data and plot kindly provided by Franz Klein, Department of Chromosome Biology, University of Vienna.

6.2.2.6   A longer species of Spo11 attached oligonucleotides

One of the most striking observations was the detection of Spo11 attached oligonucleotides with a length up to 72 bp, which mapped to the *S. cerevisiae* genome. Since 72 bp was the read length limitation of the Solexa run, we suggest the existence of oligonucleotides longer than 72 bp. The presented data is consistent with the idea (M. Neale, personal communication) that resection of DSBs starts up to 270-300 nt away from the DSB site. Resection starts at these points by Com1/Sae2 endonucleolytically nicking the DNA. The MRX complex is responsible for resection towards the DSB site, Exo1 is moving out to the other direction. Due to a steric inhibition of the MRX complex by the proteins at the DSB site, resection comes to a hold at various distances to the DSB site, which results in different lengths of oligonucleotides attached to Spo11. For more details please refer to the introduction.

For the 72 bp run, 1.34 million reads uniquely mapped to the *S. cerevisiae* genome. 536000 (40%) were between 14-20 nt, 762460 (56.9%) between 21-40 nt, 33500 (2.5%) between 41-60 nt and 8040 (0.6%) more than 60 nt long.

The distribution of Spo11 attached oligonucleotides was examined in detail for 10 regions, five very hot ones and five colder ones, to see whether the long oligonucleotides also map to hotspots of DSB formation. For two of the 10 regions the percentage of oligonucleotides longer than 60 nt correlates well with the genome-wide data of the mapped oligonucleotides mentioned before. For five regions oligonucleotides longer than 60 nt were detected to a lesser extent, for three not (see table 6.3).

Chromosome	Region	Region size (bp)	Number of total reads	> than 50 bp	> than 50 in % of total	> than 60 bp	> than 60 in % of total
I	92311 92966	655	157	0	0	0	0
II	490631 491337	706	221	7	3.17	1	0.45
III	211126 213596	2470	5381	25	0.46	4	0.07
IV	699745 700360	615	339	22	6.49	2	0.59
IV	835478 835703	225	116	1	0.86	0	0
VII	609715 610809	1094	8156	79	0.97	5	0.06
VIII	189316 191250	1934	3686	9	0.24	1	0.03
X	572426 573974	1548	175	0	0	0	0
XIII	387971 388914	943	3100	25	0.81	4	0.13
XVI	463965 464460	495	1671	16	0.96	1	0.06

Table 6.3: Distribution of long mapped oligonucleotides for 10 selected regions. For two of the 10 regions the percentage of oligonucleotides longer than 60 nt correlates well with the genome-wide data of the mapped oligonucleotides mentioned before. For five regions oligonucleotides longer than 60 nt were detected to a lesser extent, for three not.

To analyze the full length of the oligonucleotides attached to Spo11, a 76 bp paired end Solexa run was performed. This means that the amplified products are sequenced from both sites, resulting in the sequence information to obtain the full length sequence of the Spo11 attached oligonucleotides. During the writing of this thesis, samples were processed, sequenced and delivered for bioinformatical processing.

6.2.3 *Schizosaccharomyces pombe* ChIP-Seq

To apply the developed method to a second model organism, we have chosen *S. pombe* due to different reasons. First, *S. pombe* is easy to handle and a huge amount of synchronized cells at the accurate meiotic stage can be cultivated in a short period of time. Second, *S. pombe* gives us the possibility to evaluate the data, since DSB maps for data-comparison are already published. A 36 bp single end Solexa run was performed using 5×10^9 *S. pombe* cells from a haploid strain with a 13xmyc tagged Rec12 (pat1-114, rec12+::13xMYC, 4.5 hours after induction of meiosis). As a control 5×10^9 cells of the untagged version were processed according to the protocol described in materials and methods. The gels for the amplification and re-amplification looked identical. No shift of the dominant band after re-amplification, as seen for *S. cerevisiae,* was observed (see figure 6.19 and compare to figure 6.13).
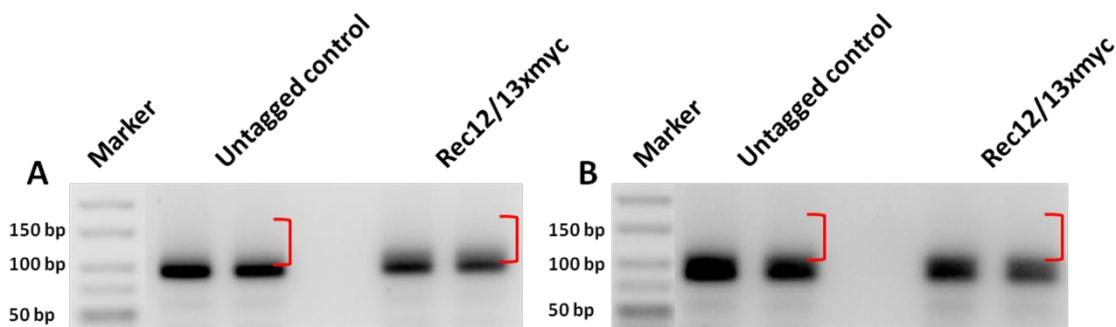
Figure 6.19: Solexa amplification (A) and re-amplification (B) with adapter specific primers. After ChIP, ligation of both adapters and the amplification PCR, products were visualized by agarose gel electrophoresis. The marked area in panel A, above the dominant band which results from a dimer of the unidirectional primer and the second adapter, was cut out from the gel the, DNA was extracted and re-amplified as described. After re-amplification (panel B) the marked areas were cut out again, DNA was purified and sent to the sequencing facilty.

For the tagged sample 14.5 million total reads were obtained, of which 2 million were used for mapping. 1.2 million sequence-reads mapped to the *S. pombe* genome, thereof 612000 uniquely. For the untagged control 2.7 million out of 18 million reads were used for mapping. 806000 mapped to the *S. pombe* genome, 274000 uniquely. For further analysis the unique dataset was used. The read length distribution of the mapped oligonucleotides (Fig. 6.20) shows that the majority of the mapped reads has a length between 15 and 20 nucleotides which is consistent with the findings published by (Milman *et al.*, 2009), where a single class ob Rec12 attached oligonucleotides was observed.
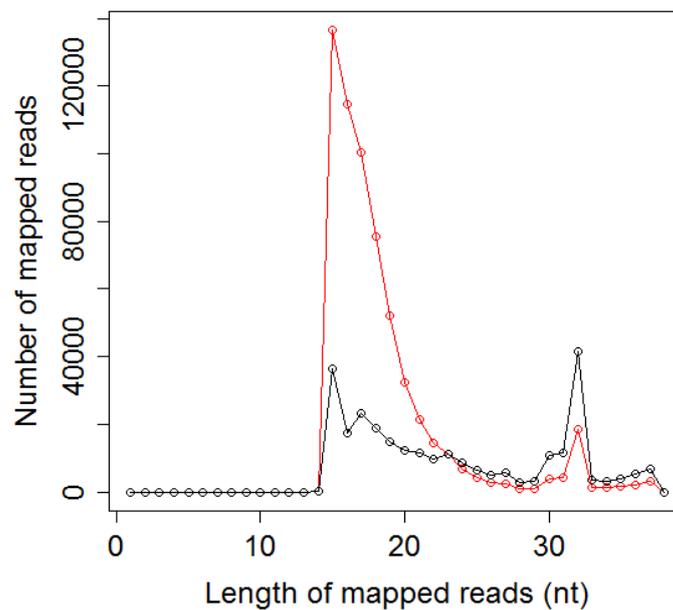
Figure 6.20: Read length distribution of uniquely mapped sequence reads for *S. pombe*. In a 36 bp run about 612000 reads uniquely mapped for the sample (red), 274000 for the control (black).

Figure 6.21 shows plots for all three *S. pombe* chromosomes. Since the coverage values for the background control were very low and rarely spread over the genome compared to the sample, the control was subtracted from the sample. In red the mapped full length reads are plotted, in blue the 5' ends of the mapped reads are plotted, which exactly represent the DSB sites, where Rec12 catalytically acts. It is visible that the 5' end signals always merge with the full length data. As expected, the coverage for the mapped 5' ends is not that high as for the total mapped reads, because every oligonucleotide produces just the information for one position (5' end), resulting in no overlap of one oligonucleotide with another at more than one position.

The data was not compared in that detail to *S. pombe* DSB maps as it was done for *S. cerevisiae*, since no comparable data was accessible in a suitable form. However, DSB profiles were compared with the printed versions of the published maps from the Smith lab (Hyppa *et al.*, 2008) and a good overlap was found. In addition to the DSB sites found in both data sets, a substantial number of DSB sites was found in our data-set only (data not shown).

S. pombe
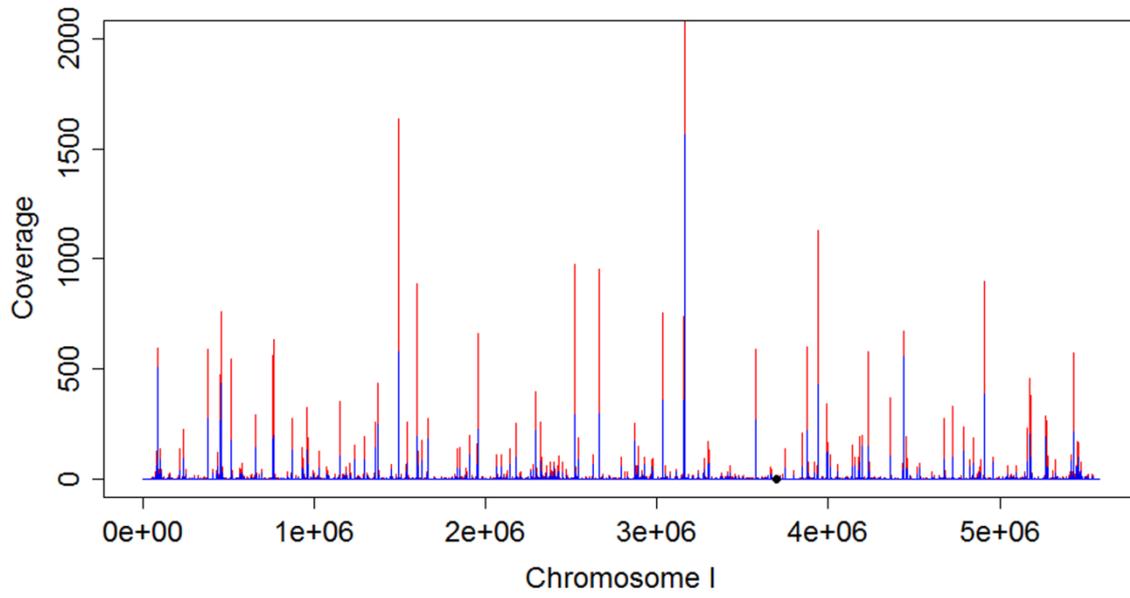
S. pombe
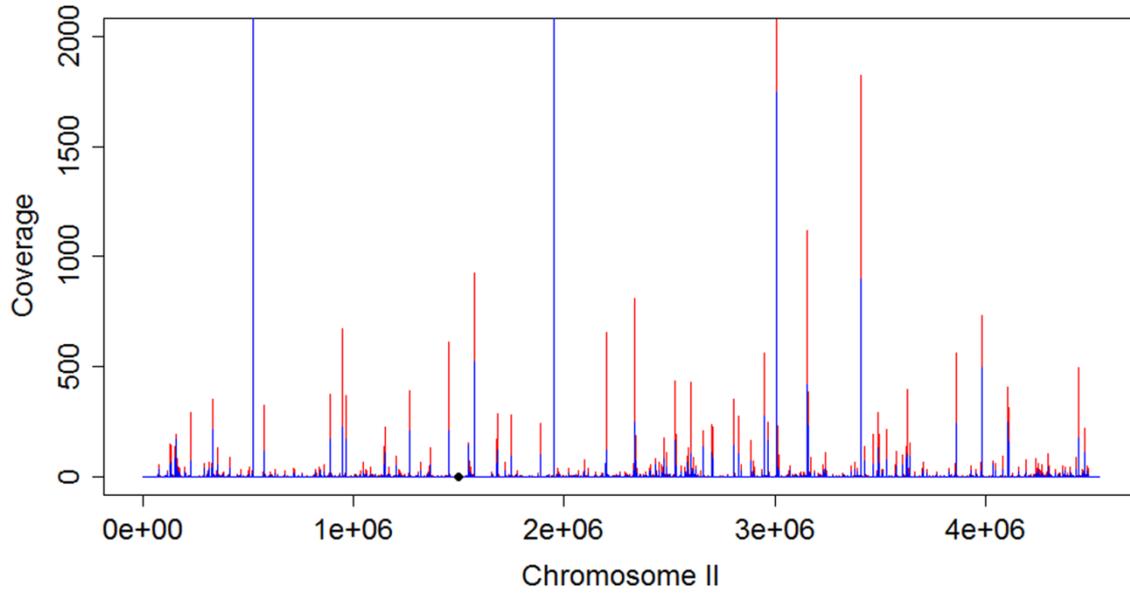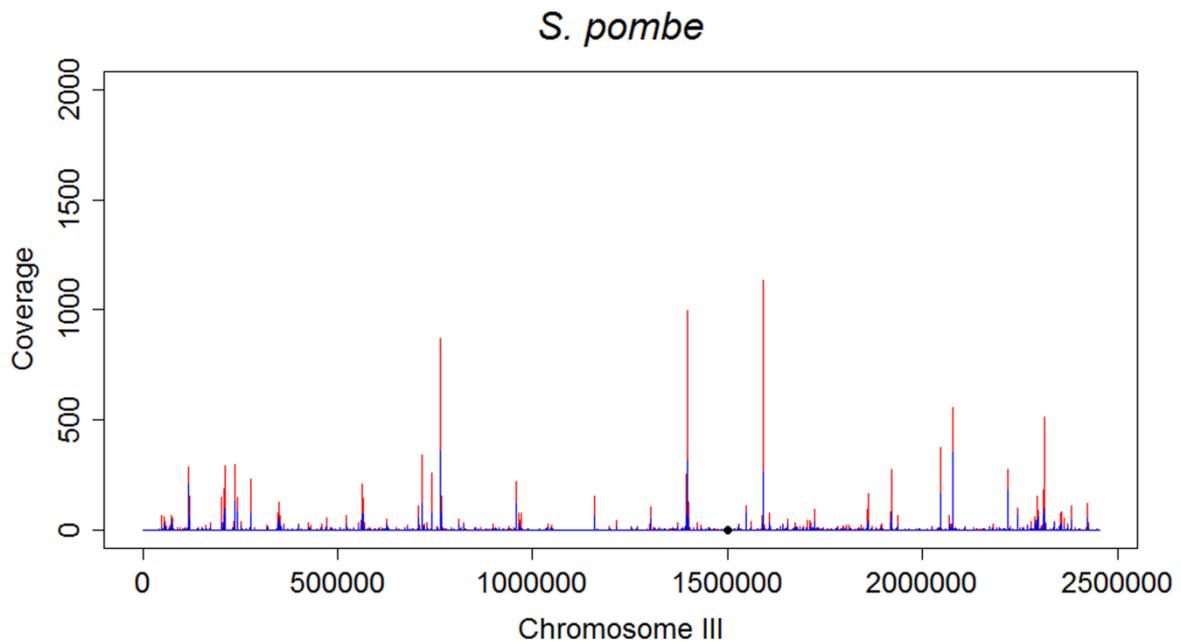
Figure 6.21: Genome-wide DSB distribution in *S. pombe*. A 36 bp single end Solexa run was performed using 5x10^9 *S. pombe* cells from a haploid strain with a 13xmyc tagged Rec12 (pat1-114, rec12+::13xMYC, 4.5 hours after induction of meiosis). As control 5x10^9 cells of the untagged version were processed according to the protocol described in materials and methods. After sequencing, base-calling was performed, the adapter sequences were clipped and the sequence reads were mapped to the *S. pombe* genome using NextGen. Coverage files for the chromosomes were generated, and the data was plotted using the free statistic program "R". The red signals represent the full length oligonucleotides. In blue the exact 5' end positions for every mapped oligonucleotide are plotted. The coverage is displayed in number of mapped nucleotides/position. The chromosomes on the x-axis are displayed in bp. The black filled circles indicate the centromere positions.

For a better evaluation of the data, a comparison with the region containing the *S. pombe* hotspots *mbs1+2* on chromosome I was performed. As visible in figures 6.22 and 6.23, the peak distribution in the *mbs1* and *mbs2* hotspot region is highly overlapping when comparing our ChIP-Seq data to the ChIP-on-chip data published by (Cromie *et al.*, 2007). All of the relevant positions show signals in both experiments. Additionally, we could identify some more small distinct peaks, which do not appear in the data from the Smith lab.
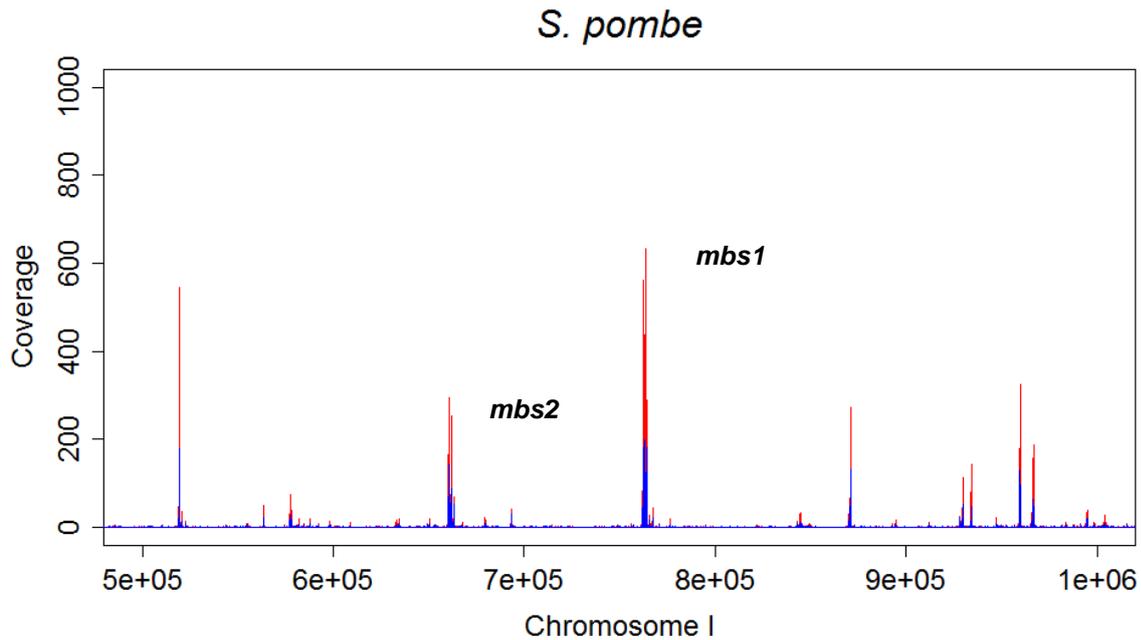
Figure 6.22: ChIP-Seq data of the *mbs1* and *mbs2* hotspot region. Mapped full length oligonucleotides (red) and 5' ends (blue) were plotted. The figure shows the *mbs1* and *mbs 2* hotspot region on chromosome I, which is highly comparable to the data visualized in figure 6.23. The coverage is displayed in number of mapped nucleotides/position. The chromosomes on the x-axis are displayed in bp.
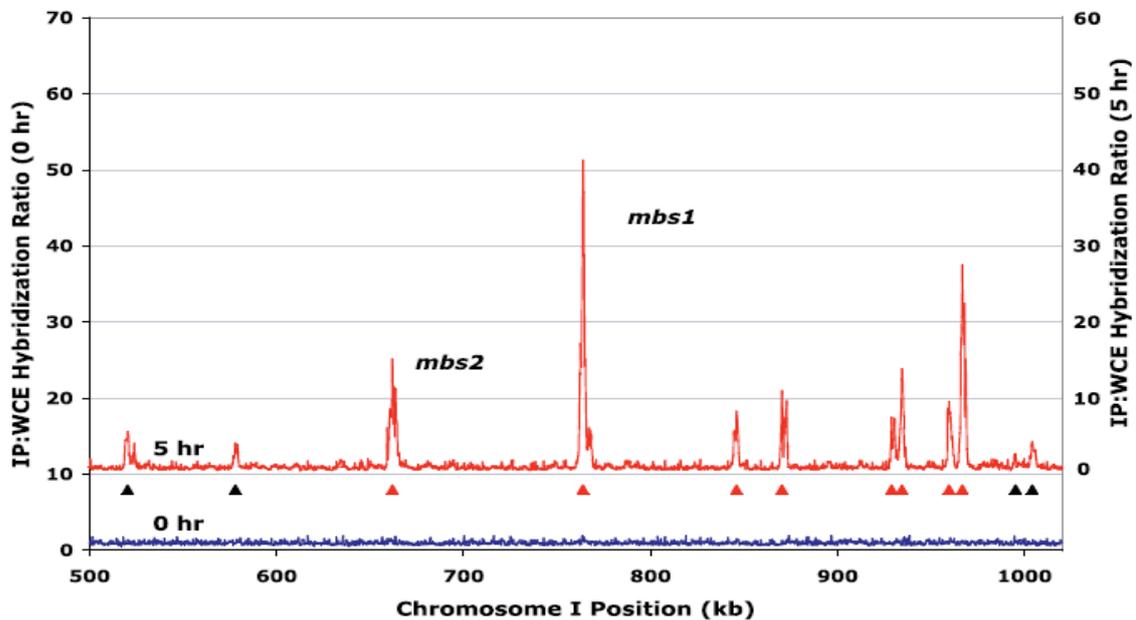
Figure 6.23: ChIP-on-chip data of the *mbs1* and *mbs2* hotspot region. Haploid strain GP6013 bearing the rec12–201::6His-2FLAG allele was analyzed for Rec12-DNA linkage by microarray analysis before (0 h; blue line) and 5 h (red line; mean of two experiments) after meiotic induction. Data are the median-normalized IP:WCE ratios for each oligonucleotide, spaced 300 bp apart. High values within contiguous groups (hotspot peaks) in the 5-h data were omitted if they were also spuriously high in the 0-h data. Carets indicate weak (black) and prominent (red) Rec12 peaks identified by PeakFinder. Left and right vertical axes are offset to allow separation of the 0- and 5-h datasets. Figure and text taken from (Cromie *et al.*, 2007).

In addition, to analyze the full length of the oligonucleotides attached to Rec12, as it was done for *S. cerevisiae*, a 76 bp paired end Solexa run was performed. This means that the amplified products are sequenced from both sites, resulting in the sequence information to obtain the full length sequence of the Rec12 attached oligonucleotides. During the writing of this thesis, samples were processed, sequenced and delivered for bioinformatical processing.

6.2.4 *Arabidopsis thaliana* ChIP-Seq

For *Arabidopsis thaliana*, two 36 bp single end Solexa runs (representing two technical repeats) were performed using the protocol described in the material and methods section. As input material, ~1 g of buds from *AtSPO11-1/18xmyc* and *AtSPO11-1/untagged* plants was used. Cell extracts were made as described in materials and methods. The gels for the amplification and re-amplification were identical to the gels for *S. pombe*, no shift of the dominant band was visible for the sample after re-amplification, as observed for *S. cerevisiae* (see figure 6.24 and compare to figure 6.13).
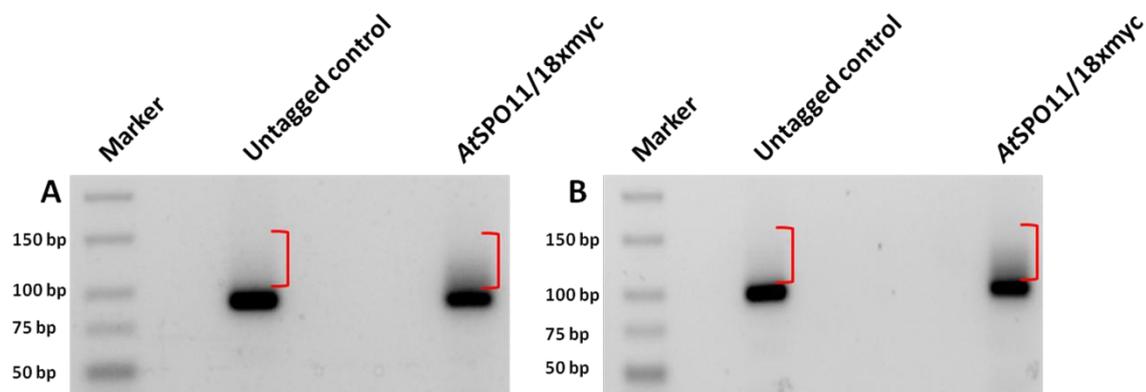
Figure 6.24: Solexa amplification (A) and re-amplification (B) with adapter specific primers. After ChIP, ligation of both adapters and the amplification PCR, products were visualized by agarose gel electrophoresis. The marked area in panel A, above the dominant band which results from a dimer of the unidirectional primer and the second adapter, was cut out from the gel the, DNA was extracted and re-amplified as described. After re-amplification (panel B), the marked areas were cut out again, DNA was purified and sent to the sequencing facilty.

In the first run, for the sample 16.8 million total reads were obtained, 13 million were used for mapping, 3.5 million mapped to the *Arabidopsis* genome, thereof 2.6 million uniquely. For the control 15.7 million total reads were obtained, 12.3 million were used for mapping, 3.5 million mapped to the *Arabidopsis* genome, thereof 2.6 million uniquely. For the second run sample 10.5 million reads were obtained, 9.2 million were used for mapping, 2.2 million mapped, 1.7 million uniquely. From the 11 million control reads, 9.4 million were used for mapping, thereof 2 million mapped to the *Arabidopsis* genome, 1.4 million uniquely.

In both runs, for sample and control nearly the same number of reads mapped to the *Arabidopsis* genome and the reads for both were distributed over the chromosomes, showing no cluster formation (hotspots) for the sample. Due to a high degree of overlap between sample and control, it was not possible to identify potential DSB sites when

plotting both data sets. Therefore only unique sample coverages were included in further analysis from positions where the coverage for the control was zero.

Compared to *S. cerevisiae* and *S. pombe*, only 40-60 single peaks/chromosome were detectable throughout the *Arabidopsis* genome. This means that on average only every 500 kb a single peak was detectable (see figure 6.25). Deeper analysis at nucleotide resolution showed that all of the peaks are really single peaks, resulting from one sequence read which mapped to the respective region (see figure 6.25, last plot). As showed for *S. cerevisiae* and *S. pombe*, one would expect a Gaussian distribution of the reads mapping to a hotspot, resulting from different overlapping sequence reads, mapping to such a region.

Moreover, the results of the first run were not reproducible in the second run (technical repeat), detecting only two signals at nearly the same position.

The most probable explanation for these results is the low amount of input material which was used for the Solexa experiments. Since for yeast 5×10^9 synchronized meiocytes at the right stage are used, one would need approximately 2.5 million inflorescences to obtain the same number of meiocytes for *Arabidopsis*. A second possible explanation is the influence of the growth conditions (temperature, humidity) on hotspot distribution (see discussion). To combine both aspects, for an improvement of the technique in *Arabidopsis*, we have grown a huge population of 20000 plants under constant conditions (16 hours day, 8 hours night, day temperature 19°C (+/-1°C), night temperature 21°C (+/-1°C), 60% humidity (+/-5%), 6500 Lux (2 lamps, 1x Osram cool white 840, 1x Osram warm white 827)) in the new GMI growth chambers. Buds from the sample and control population were harvested, processed and sent for sequencing during the writing of this thesis.

A. thaliana

A. thaliana
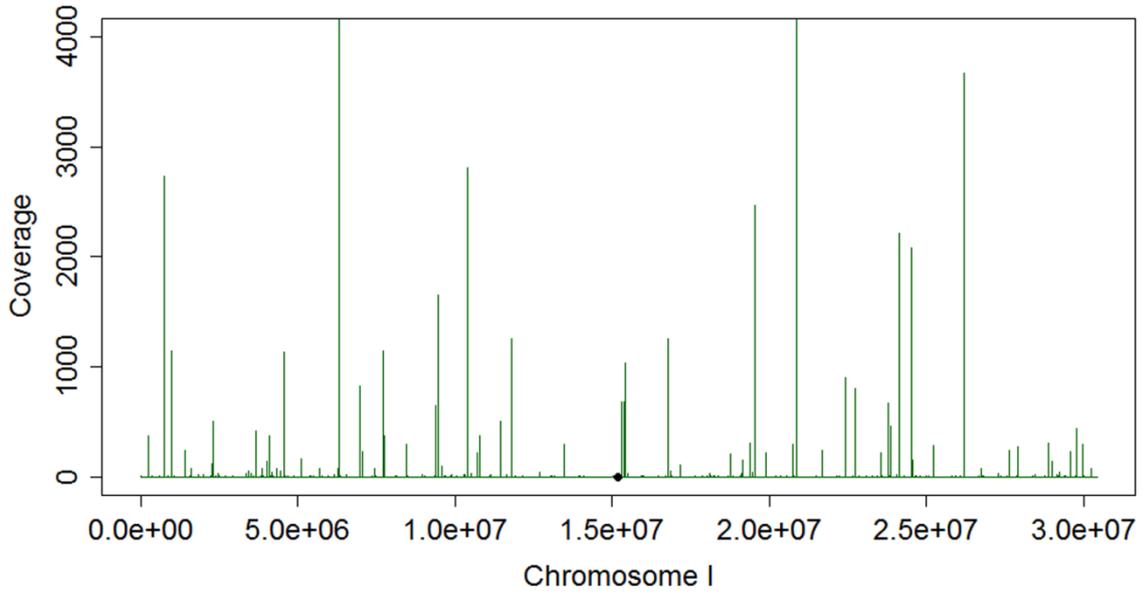
A. thaliana
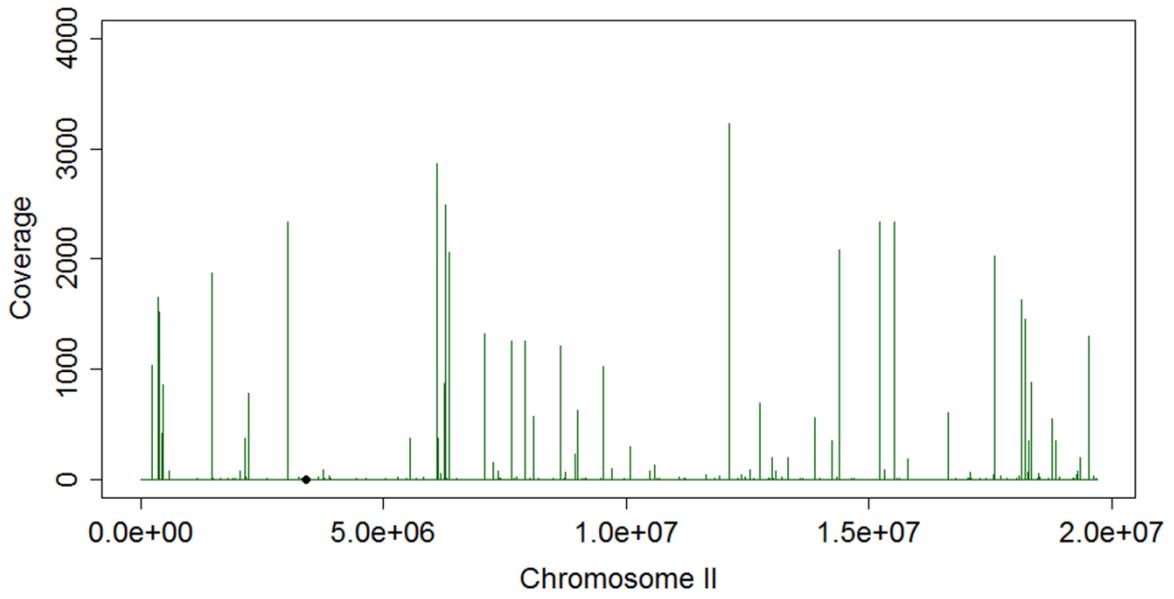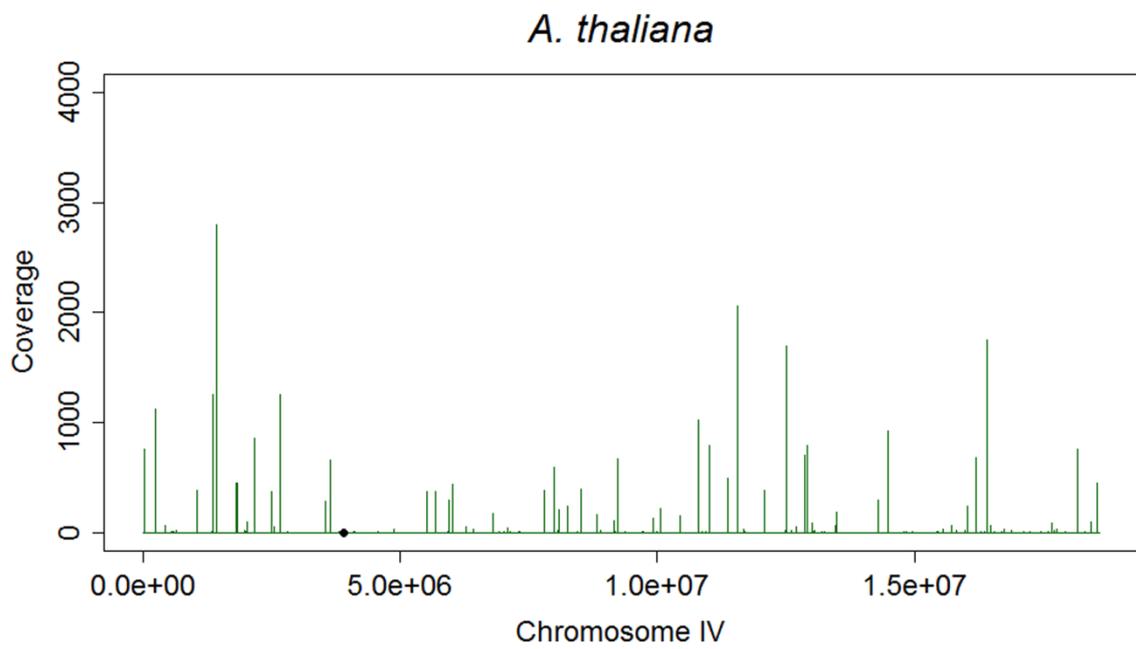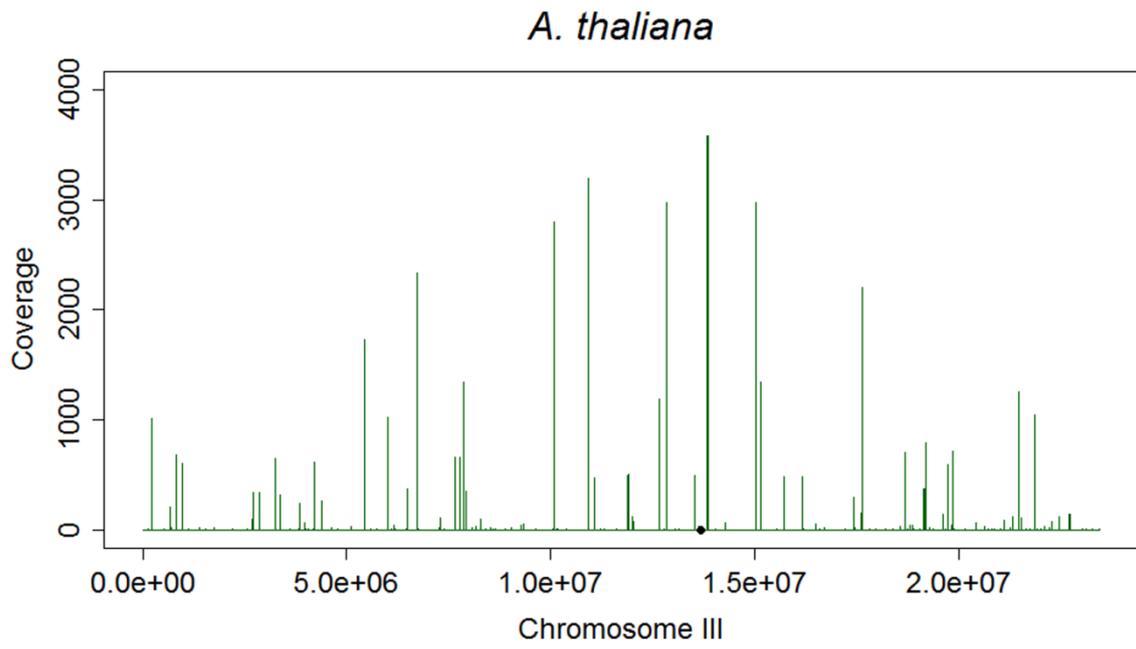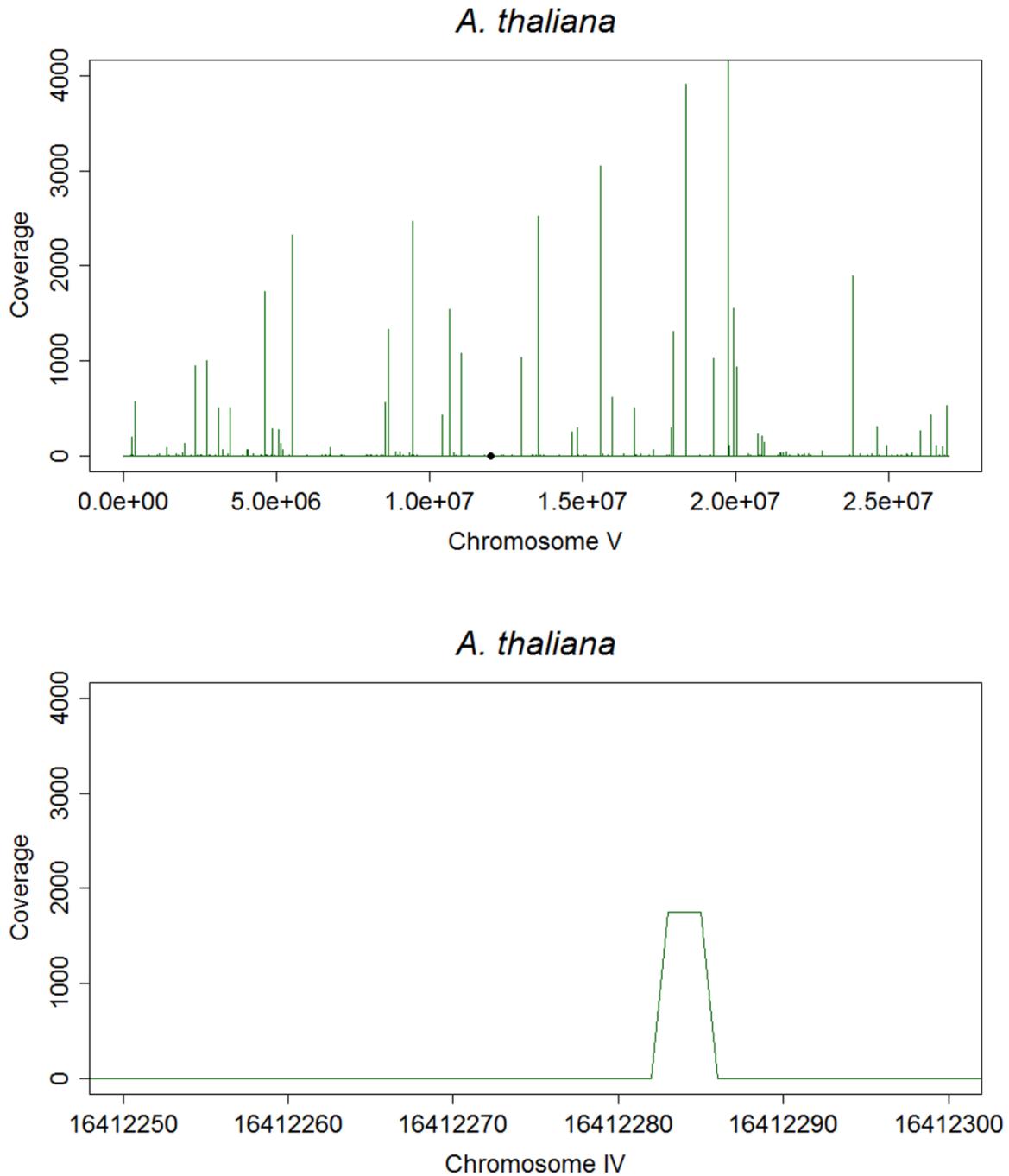
A. thaliana

Figure 6.25: Genome-wide DSB distribution in *Arabidopsis thaliana*. Using 1 g of buds from *AtSPO11-1/18xmyc* plants ChIP according the protocol described in materials and methods was performed. As a control the biological material from an untagged transgenic plant line was used. After Solexa sequencing the adapter sequences were clipped and the sequence reads were mapped to the *Arabidopsis* genome using NextGen. Coverage files for the chromosomes were generated, and the data was plotted using the free statistic program "R". Due to a high background noise, the sample values were plotted only at the genomic positions where the coverage of the control was zero. The first five plots show the results from the second experiment for the *Arabidopsis* chromosomes I-V. The last plot shows a zoom to a 50 bp region on chromosome IV. The coverage is displayed in number of mapped nucleotides/position. The chromosomes on the x-axis are displayed in bp. The black filled circles indicate the centromere positions.

6.3 Identification of *At*SPO11-1 and *At*SPO11-2 interacting proteins

The mechanistic mode of DSB formation is well conserved in all eukaryotes, but the initiation of meiotic recombination is different between *S. cerevisiae* and *Arabidopsis*, concerning the involved protein complexes (see introduction). In *S. cerevisiae*, in addition to Spo11, at least 9 other proteins are involved in DSB formation. In *Arabidopsis*, primary screens for reduced fertility and secondary screens for the absence of meiotic DNA breaks identified the genes *AtPRD1*, *AtPRD2* and *AtPRD*3 to be essential for SPO11 mediated meiotic DSB formation (De Muyt *et al.*, 2009; De Muyt *et al.*, 2007). *AtPRD1* (*P*utative *R*ecombination initation *D*efect 1) has low similarity to mammalian *Mei1*. The mouse *Mei1* was characterized by Libby et al. and was isolated in a mutant screen for infertility (Libby *et al.*, 2002). The N-terminus of *At*PRD1 interacts with *At*SPO11-1 in a yeast two-hybrid assay, thereby identifying for the first time an interaction partner of a SPO11 protein in plants. The functional relevance of this interaction is unknown (De Muyt *et al.*, 2007). *AtPRD2* has later been recognized as a homolog of Mei4 (Kumar *et al.*, 2010). *At*PRD3 is a protein of unknown function, similar to the previously identified rice PAIR1 gene, but with no homologs outside the plant kingdom (De Muyt *et al.*, 2009; Nonomura *et al.*, 2004). Text taken from (Edlinger and Schlogelhofer, 2011).

Nevertheless, many open questions remain elusive. Are there other proteins involved in DSB formation in *Arabidopsis*, how do they interact in a complex and do *At*SPO11-1 and *At*SPO11-2 form a heterodimer.

Therefore different approaches were undertaken to answer these questions. The first approach, which is mainly described in this thesis, is a protein complex purification via the myc tagged *At*SPO11-1/18xmyc construct from transgenic plant lines followed by LC-MS/MS analysis. Two other approaches, which are only mentioned in this section, were I) a direct yeast two-hybrid interaction study with *At*SPO11-1 and *At*SPO11-2 as bait proteins and II) co-immunoprecipitation experiments with *At*SPO11-1 and *At*SPO11-2 expressed in wheat germ extract.

6.3.1 *At*SPO11-1/18xmyc purification followed by LC-MS/MS analysis

Mass spectrometric analysis of proteins has contributed tremendously to our understanding of biological systems. Identification of protein complexes by mass spec analysis has allowed investigators to connect cellular pathways and to describe the dynamics of protein complexes. Importantly, to get a genuine picture of the in vivo situation, it is essential to avoid any changes in protein modification or protein complex

composition that might occur during the purification procedure. These approaches require a high degree of purification of proteins. Therefore, two step purification strategies have been proven to be very effective in reducing non-specific background, which is particularly important for the analysis of complex protein samples (Tagwerker *et al.*, 2006). In this section, a one-step purification with the 18xmyc tag is described. Since prior to the start of this thesis, one goal was the cytological localization of the *Arabidopsis* SPO11 proteins, various strategies for biochemical *At*SPO11-1 detection were addressed. An *AtSPO11-1* antibody (Courtesy of Chris Franklin, School of Biosciences, University of Birmingham, United Kingdom) was available but failed in *At*SPO11-1 IP experiments (Peter Schlögelhofer, personal communication). It might be possible that accessory proteins within a multi-protein complex hinder the immunological detection of *At*SPO11-1 by masking the immunogenic epitopes (Janisiw, 2009). Therefore the long 18xmyc epitope was chosen to avoid such problems. Although it is only a one step purification, the tag is very promising, since Zachariae and colleagues purified the anaphase promoting complex in *S. cervisiae* using this 18xmyc epitope (Zachariae *et al.*, 1998; Zachariae *et al.*, 1996).

The experimental set-up is described in detail in the materials and methods section. Proteins extracts were made from 1 g of buds of *AtSPO11-1/18xmyc* and *AtSPO11-1/untagged* lines followed by immuno-precipitation using protein G sepharose beads coupled with an anti-myc antibody. For the mock IP only protein extraction buffer was added instead of protein extract. After washing, 4xLDS sample dye was added to the beads and proteins were eluted by heating at 95°C for 5 min, since no other elution method was successful. Elution conditions, which were unsuccessful due to the strong interaction of the antibody with the 18xmyc tag, are pH-dependent elution with 0.2 M glycine, pH 1.5, 0.2 M glycine, pH 1.5 + 1M MgCl2 and 200 mM HCl, salt-dependent elution with 3 M thiocyanat and competitive elution with a myc peptide. Samples were loaded on a NuPAGE® Novex® Bis-Tris gel and separated electrophoretically using NuPAGE MES SDS running buffer.

Silverstaining was performed as described in (Blum *et al.*, 1987) and stained gels were brought to the MFPL Mass Spec facility for further processing. The total lanes were cut in pieces, proteins were digested with trypsine and analyzed by Mass Spec. Further data analysis was performed using Scaffold 2 software (www.proteomsoftware.com).
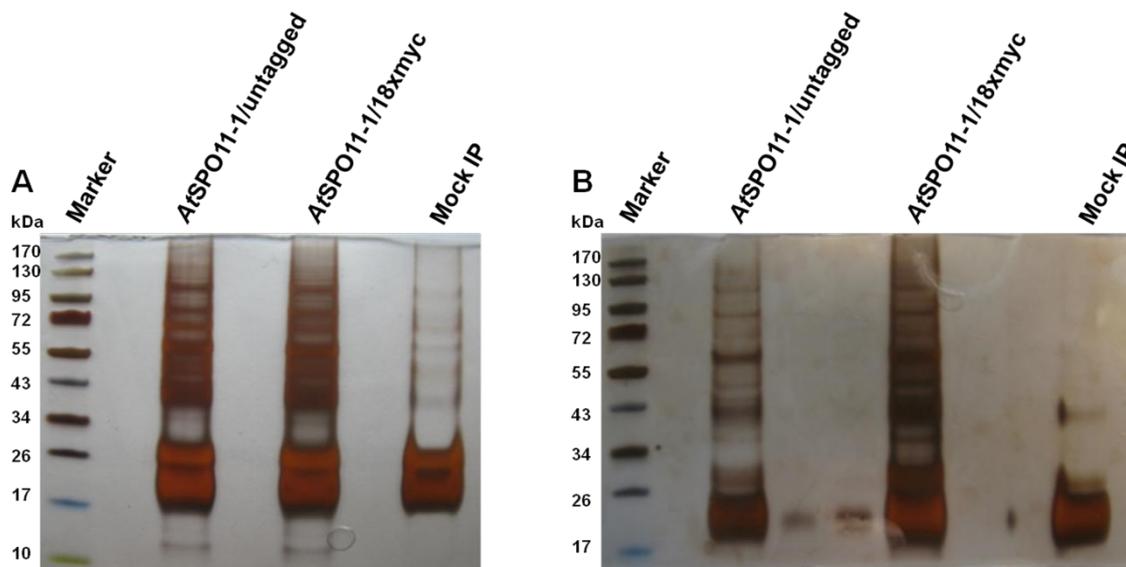
Figure 6.26: Silver stained gels from two experiments. (A) Experiment 1, (B) Experiment 2. Proteins extracts were made from 1 g of buds of *AtSPO11-1/18xmyc* and *AtSPO11-1/untagged* lines followed by immune-precipitation using protein G sepharose beads coupled with an anti-myc antibody. After washing, 4xLDS sample dye was added to the beads and proteins were eluted by heating at 95°C for 5 min. Samples were loaded on a NuPAGE® Novex® Bis-Tris gel and separated electrophoretically using NuPAGE MES SDS running buffer. Silverstaining was performed as described in (Blum *et al.*, 1987) and stained gels were brought to the MFPL Mass Spec facility for further processing. The total lanes were cut in pieces, proteins were digested with trypsine and analyzed by Mass Spec. Further data analysis was performed using Scaffold 2 software. The sample order for both gels is the same. Marker, *At*SPO11-1/untagged, *At*SPO11-1/18xmyc, Mock IP.

Figure 6.26 shows the silverstained gels from two experiments (A, B, biological replicates). In panel A the loaded protein amount seems to be the same for the untagged and the tagged sample. Much less silver stained protein is visible in the mock IP lane, which is expected since no protein extract was added. In panel B there is much more protein visible for the tagged sample compared to the untagged one. Since the amount of potential proteins of interest is very low compared to the total input, one would assume that the visible protein amount should be almost the same for the tagged and the untagged sample, except some bands representing specific proteins of interest. So maybe there is the possibility that less protein was loaded for the untagged sample in the second experiment (panel B). The very dominant area between 17-26 kDa may have its origin from the antibody, since this staining is also visible in the mock IP lane.

In the first experiment for the tagged sample 519, for the untagged 500 and for the mock IP 32 proteins were detected with LC-MS/MS anaylsis. In the second experiment for the untagged sample only 240, for the tagged sample 500 and for the mock IP 6 proteins were identified. The expectation is that for tagged and untagged sample almost the same amount of proteins should be present, since the unspecific background should be also

detected in the tagged sample. Most probably, in the second experiment less protein was loaded for the untagged sample.

Identifier	Description
gi|15239643	heavy-metal-associated domain-containing protein
gi|15238559	GS2 (GLUTAMINE SYNTHETASE 2
gi|15229631	60S ribosomal protein L26
gi|15223516	SAR1 (SECRETION-ASSOCIATED RAS); GTP binding
gi|2789660	eIF3c
gi|15242822	GSA1 (GLUTAMATE-1-SEMIALDEHYDE-2,1-AMINOMUTASE)
gi|1254996	serine/threonine protein phosphatase type 2A regulatory subunit A
gi|13937244	At2g20140/T2G17.6 [Arabidopsis thaliana]
gi|15221636	4-coumarate--CoA ligase family protein
gi|15232210	ATGDI2 (RAB GDP DISSOCIATION INHIBITOR 2)
gi|18396193	subtilase family protein
gi|18416870	unknown protein
gi|15237422	KAS I (3-KETOACYL-ACYL CARRIER PROTEIN SYNTHASE I)
gi|15217605	chalcone and stilbene synthase family protein
gi|15238284	CBS domain-containing protein
gi|30685661	chaperonin, putative
gi|18421544	eIF4-gamma/eIF5/eIF2-epsilon domain-containing protein
gi|15218869	isocitrate dehydrogenase, putative
gi|15218845	ATS9 (19S PROTEOSOME SUBUNIT 9)
gi|15233429	TUB9 (tubulin beta-9 chain); structural molecule
gi|15809972	AT4g37930/F20D10_50
gi|15237159	RPT3 (root phototropism 3); ATPase
gi|18395103	GRF10 (GENERAL REGULATORY FACTOR 10)
gi|18404382	malate dehydrogenase (NAD), mitochondrial
gi|6715645	T25K16.8
gi|15230435	PAD1 (20S PROTEASOME ALPHA SUBUNIT PAD1)
gi|15233851	CER2 (ECERIFERUM 2); transferase
gi|3063661	nucleoside diphosphate kinase Ia
gi|10177207	proteasome regulatory subunit-like
gi|15218860	POR C (PROTOCHLOROPHYLLIDE OXIDOREDUCTASE)
gi|22326950	strictosidine synthase family protein
gi|13449329	NADH dehydrogenase subunit 7
gi|145333783	PGR5-LIKE A
gi|15220146	PTAC17 (PLASTID TRANSCRIPTIONALLY ACTIVE17)
gi|25082754	Unknown protein
gi|15239078	ASP3 (ASPARTATE AMINOTRANSFERASE 3)
gi|15237348	dimethylmenaquinone methyltransferase family protein
gi|18410886	phenylalanyl-tRNA synthetase class IIc family protein
gi|7525019	ATP synthase CF0 B subunit
gi|15226754	UREG (urease accessory protein G)
gi|15233715	ATMKK2 (MAP KINASE KINASE 2)
gi|15232420	ATPUMP1; binding / oxidative phosphorylation uncoupler
gi|15220930	RPT1A (regulatory particle triple-A 1A); ATPase
gi|15234063	mitochondrial substrate carrier family protein
gi|7525042	acetyl-CoA carboxylase beta subunit
gi|18394608	ATARP4 (ACTIN-RELATED PROTEIN 4)

Table 6.4: List of 46 proteins which were present in both Mass Spec experiments in the tagged sample only.

Protein lists from both experiments were compared and analyzed, searching for proteins which are present in the tagged sample only. Since no potentially interesting candidate, for being a interaction partner of AtSPO11-1, was found in one of the single experiments, proteins were selected as potentially interesting candidate if detectable in both experiments in the tagged sample only (see table 6.4).

Since the bait protein AtSPO11-1/18xmyc was not detectable in the sample and one would assume that potentially interacting proteins should be present in approximately the same amount as the bait, depending on the stoichiometry of the protein complex, the conclusion can be made that most probably no real AtSPO11-1 interaction partner was identified. Additionally, no known interaction partner like AtPRD1 was found, therefore no deeper analysis of the proteins listed in table 6.4 was performed.

To be more successful in identifying AtSPO11-1 interaction partners, a SF tag was fused to the C-terminus of AtSPO11-1 (described in section 6.1). With this tag the achievement of a higher degree of protein purification should be possible and hopefully this tag is suitable for direct elution of the protein from the beads using a FLAG peptide and/or desthiobiotin. The transgene harboring the SF tag is complementing Atspo11-1-2 mutant plants, therefore these plants are ready for future experiments.

6.3.2   Interaction studies using direct yeast two-hybrid and Co-IP approaches

Direct yeast two-hybrid mating interaction studies were performed using all available clones in the Schlögelhofer lab (Spo11-1, Spo11-2, Mre11, Rad50, Nbs1, Com1, Mnd1, Dmc1, Rad51, FancD2, Ahp2, Mip1, Brca1, Brca2 IV, Brca2 V, Xrcc3 and Bard1). As bait proteins AtSPO11-1 and AtSPO11-2 cloned into the pGBKT7 and pGADT7 vectors were used (cloning was performed by Stefan Mevius, Schlögelhofer lab). They were mated in both directions with all other clones and themselves and then selected on YSD –LTH +2 mM 3-AT and YSD –LTA plates. No interaction between one of the tested proteins was observed (results not shown).

To perform interaction studies for AtSPO11-1 and AtSPO11-2 in all directions, the yeast two-hybrid clones were taken for protein expression in the TNT® T7 Coupled Wheat Germ Extract System from Promega using S35 methionine as labeled amino acid. Protein expression was confirmed by Western blot analysis and for the Co-IP experiments protein extracts were mixed in all possible combinations including the appropriate negative

controls. No interactions were observed in two replicates of these experiments (results not shown).

6.4 Targeted stimulation of meiotic recombination in *A. thaliana*

As discussed in the introduction, targeted stimulation of meiotic recombination is a well established method in the yeast *S. cerevisiae* (Fukuda *et al.*, 2008; Pecina *et al.*, 2002; Robine *et al.*, 2007). The aim of this project is to generate such artificial hotspots in *Arabidopsis thaliana*, by fusing a specific DNA binding domain (BD) to the C-terminus of *At*SPO11-1 (*At*SPO11-1/Gal4) and introducing corresponding cis-elements (5xUAS) into the genome. 5xUAS lines and transgenic lines harboring *At*SPO11-1 fused to the Gal4 BD were generated as described in section 6.1. After crossing plants with aforementioned features it is anticipated that the tagged *At*SPO11-1 will be guided to the cis-elements and stimulate formation of meiotic DSBs, resulting in meiotic recombination events (see figure 6.28). To monitor recombination events, a fluorescent pollen based marker system, which was developed in the lab of Gregory P. Copenhaver (UNC, Chapel Hill) was used (Francis *et al.*, 2007). The method is based on the phenotype of different *qrt* mutant alleles. Plants homozygous for this mutation fail to degrade the pollen mother cell wall, therefore the four meiotic products are held together as a tetrad, allowing all four products of a single meiotic event to be studied (Copenhaver *et al.*, 2000; Rhee and Somerville, 1998). The so called FTLs (fluorescent-tagged line) all harbor a T-DNA construct containing either ECFP, EYFP or DsRed driven by the pollen-specific LAT52 promoter and a selectable marker like kanamycin (*npt II*) which makes selection of plants and visualization of pollen tetrads possible (see figure 6.27) (Francis *et al.*, 2007).
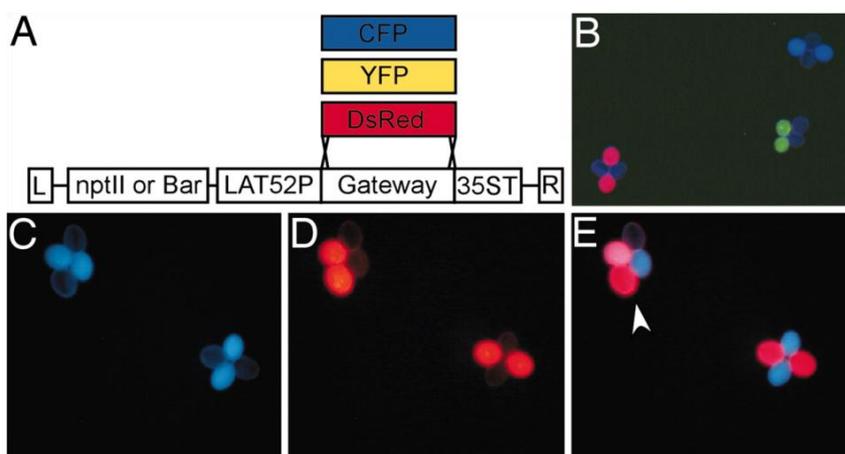
Figure 6.27: Fluorescent markers in pollen tetrads. (A and B) An Agrobacterium T-DNA construct (A) containing ECFP, EYFP, or DsRed driven by the LAT52 promoter and a selectable marker conferring resistance to either kanamycin (npt II) or glufosonate (Bar) was used to transform qrt1 Arabidopsis seed resulting in T1 plants expressing the fluorescent protein in their pollen (B shows pollen from three different plants). (C–E) Crossing lines with differently colored transgenes on the same chromosome (C and D) enables detection of CO events in the interval between the transgenes (E, merge of C and D with arrow indicating recombinant tetrad). Figure and text taken from (Francis *et al.*, 2007).
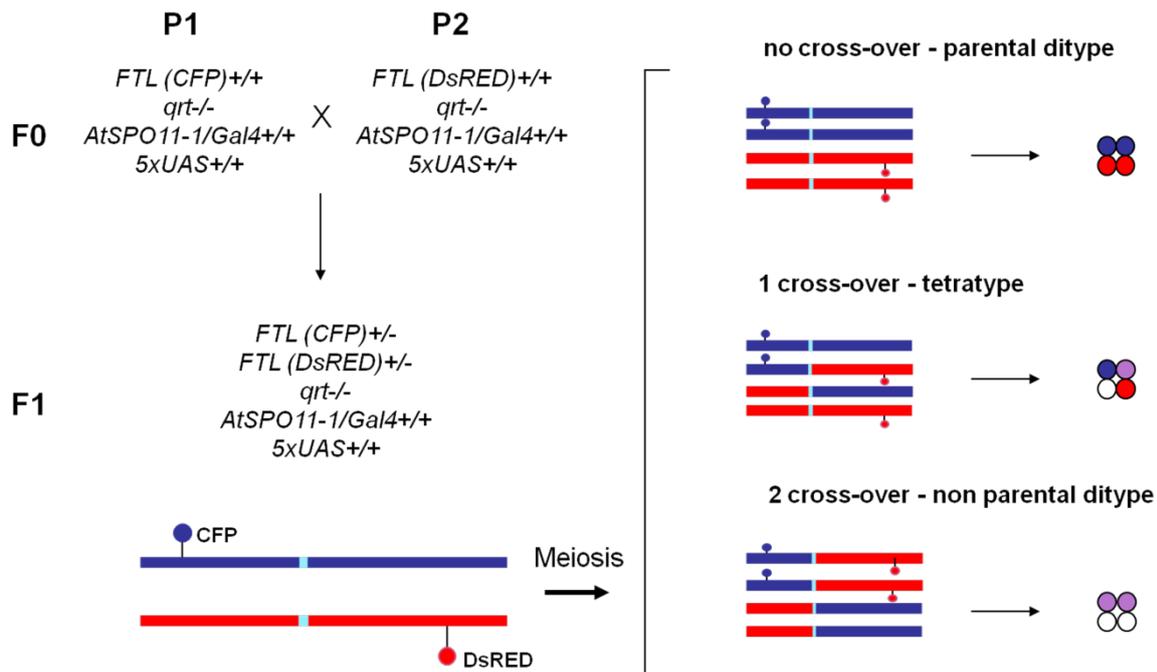
Figure 6.28: Crossing and read out strategy for the artificial hotspot project. The two parental plant lines (P1 and P2) are crossed together, directly resulting in the F1 tester line which is ready for tetrad analysis. The possible outcomes are parental ditype (PD), tetratype (TT) or non parental ditype (NPD). Recombination rates are compared to baseline recombination rates from plants harboring no *AtSPO11-1/Gal4*. The map distance between two markers is calculated using the Perkins formula: $[100 \times (1/2 \times TT + 3 \times NPD)/n]$.

For three out of five mapped 5xUAS lines (see section 6.1.7) suitable FTL marker lines were available which are listed in table 6.5.

5xUAS line	Chr./insertion site	FTL left/position	FTL right/position	Estimated marker distance
#1	Chr. II, 16363755	FTL965-red/14675407	FTL2913-blue/16387996	8.56 cM
#19	Chr. III, 17966362	FTL1268-red/17242947	FTL2201-blue/18319139	5.38 cM
#20	Chr. II, 13006968	FTL2269-blue/8276753	FTL965-red/14675407	32 cM

Table 6.5: FTL lines and map distances used in this study. The estimate for the physical distance was 5 cM/Mb.

The three 5xUAS lines listed in table 6.5 were used to start the crossing process described below. For a better understanding it has to be remembered that FTL lines harbor a kanamycin resistance, *AtSPO11-1/Gal4 and AtSPO11-1/untagged* lines a Basta resistance and 5xUAS lines a hygromycin resistance.

The crossing strategy to get the parental lines P1 and P2 depicted in figure 6.28 with the genotype *FTL +/+, qrt -/-, AtSPO11-1/Gal4 +/+, 5xUAS +/+, Atspo11-1-2 -/-* is as follows. The crossings were done separately for both FTL markers.
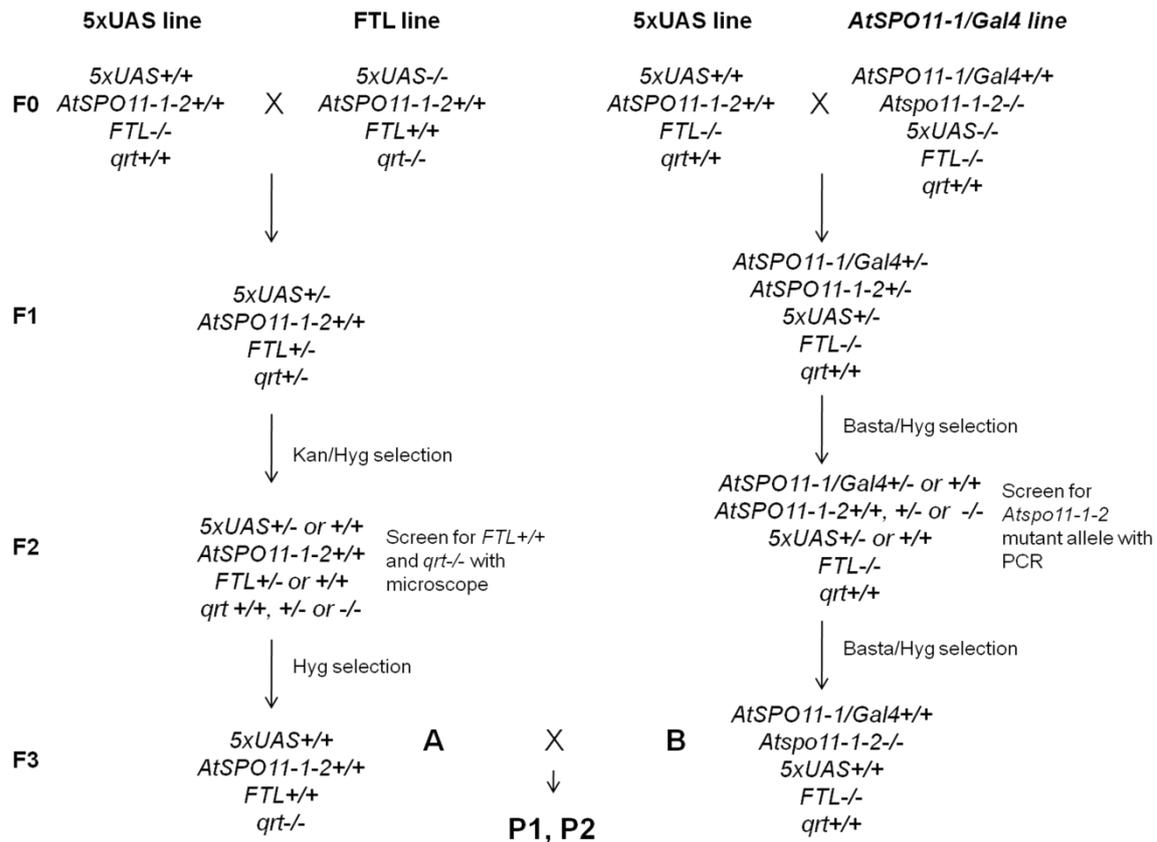
Figure 6.29: Crossing strategy to obtain parental lines P1 and P2. To get the plant named "A" with the genotype *5xUAS +/+, FTL +/+, qrt -/-, AtSPO11-1-2 +/+*, a homozygous 5xUAS line (*5xUAS +/+, AtSPO11-1-2 +/+, FTL -/-, qrt +/+*) is crossed with a homozygous FTL marker line (*5xUAS -/-, AtSPO11-1-2 +/+, FTL +/+, qrt -/-*) for one site of the 5xUAS insertion, resulting in 100% heterozygous plants for *5xUAS, FTL and qrt* in F1. The plants are brought to F2 on kanamycin/hygromycin plates, resulting in plants which are heterozygous or already homozygous for 5xUAS and the FTL marker. These plants are screened under a fluorescence microscope for *FTL +/+* and *qrt -/-*. If such plants are found, they go through a selection on hygromycin plates in F3 to find *5xUAS +/+* individuals. Sometimes one or more additional generation of marker segregation is necessary to end up with plant "A" with the genotype *5xUAS +/+, FTL +/+, qrt -/-, AtSPO11-1-2 +/+*. To get the plant named "B", with the genotype *5xUAS +/+, AtSPO11-1/Gal4 +/+, Atspo11-1-2 -/-, FTL -/-, qrt +/+*, a homozygous 5xUAS line (*5xUAS +/+, AtSPO11-1-2 +/+, FTL -/-, qrt +/+*) is crossed with a homozygous *AtSPO11-1/Gal4* plant (*AtSPO11-1/Gal4 +/+, Atspo11-1-2 -/-, 5xUAS -/-, FTL -/-, qrt +/+*), resulting in 100% heterozygous plants for *AtSPO11-1/Gal4, AtSPO11-1-2 and 5xUAS* in F1. After several rounds of segregation, the plant "B" is obtained.

Crossing of plants "A" and "B", results in the following genotype: *5xUAS +/+, AtSPO11-1/Gal4 +/-, AtSPO11-1-2 +/-, FTL +/-, qrt +/-*. After several rounds of segregation the parental plants P1 and P2 with the genotype *5xUAS +/+, AtSPO11-1/Gal4 +/+, Atspo11-1-2 -/-, FTL +/+, qrt -/-* are obtained. Then two P1 and P2, one with the left and one with the right FTL marker, are crossed together and can be directly screened in the F1 for recombination events (see figure 6.28). As controls, to monitor baseline recombination, crossings of respective FTL lines alone, FTL with 5xUAS or with 5xUAS and untagged *At*SPO11-1 are used. The map distance is calculated using the Perkins formular: [100x(1/2xTT + 3xNPD)/n].

6.4.1   Current plant crossing status

5xUAS#20 insertion

5xUAS#20xFTL2269-blue sF5: plants #33, 34, 39 and 44

5xUAS#20xFTL965-red sF5: plants #85, 86

All plants have the genotype *5xUAS +/+, FTL +/+, qrt -/-, AtSPO11-1-2 +/+*, being "A" plants ready to cross to "B" plants.

5xUAS#20xAtSPO11-1/Gal4 sF3:

21 plants are harvested, which have the genotype *5xUAS +/- or +/+, AtSPO11-1/Gal4 +/- or +/+, AtSPO11-1-2 +/- or -/-, FTL -/- and qrt +/+.* These plants are ready for further selection to make the first 3 alleles homozygous to get plant "B".

5xUAS#20xAtSPO11-1/untagged sF3:

22 plants are harvested, which have the genotype *5xUAS +/- or +/+, AtSPO11-1/untagged +/- or +/+, AtSPO11-1-2 +/- or -/-, FTL -/- and qrt +/+.* These plants are ready for further selection to make the first 3 alleles homozygous to get plant "B".

5xUAS#1 insertion

5xUAS#1xFTL965-red sF3:

Plants #24 and 25 have the genotype *5xUAS +/-, FTL +/+, qrt +/+ or +/-, AtSPO11-1-2 +/+*

5xUAS#1xFTL2913-blue sF2:

All plants have following genotype after crossing: *5xUAS +/-, FTL +/-, qrt +/-, AtSPO11-1-2 +/+* and are ready for further selection

5xUAS#1xAtSPO11-1/Gal4 sF3:

16 plants are harvested, which have the genotype *5xUAS +/- or +/+, AtSPO11-1/Gal4 +/- or +/+, AtSPO11-1-2 +/- or -/-, FTL -/- and qrt +/+.* These plants are ready for further selection to make the first 3 alleles homozygous to get plant "B".

5xUAS#1xAtSPO11-1/untagged sF3:

6 plants are harvested, which have the genotype *5xUAS +/- or +/+, AtSPO11-1/untagged +/- or +/+, AtSPO11-1-2 +/- or -/-, FTL -/- and qrt +/+.* These plants are ready for further selection to make the first 3 alleles homozygous to get plant "B".

5xUAS#19 insertion

5xUAS#1xFTL1268-red sF3:

Plants #16 and 38 have the genotype *5xUAS +/-, FTL +/+, qrt +/+ or +/-, AtSPO11-1-2 +/+*

5xUAS#1xFTL2201-blue sF2:

All plants have following genotype after crossing: *5xUAS +/-, FTL +/-, qrt +/-, AtSPO11-1-2 +/+* and are ready for further selection

5xUAS#19xAtSPO11-1/Gal4 sF3:

7 plants are harvested, which have the genotype *5xUAS +/- or +/+, AtSPO11-1/Gal4 +/- or +/+, AtSPO11-1-2 +/- or -/-, FTL -/- and qrt +/+.* These plants are ready for further selection to make the first 3 alleles homozygous to get plant "B".

5xUAS#19xAtSPO11-1/untagged sF3:

6 plants are harvested, which have the genotype *5xUAS +/- or +/+, AtSPO11-1/untagged +/- or +/+, AtSPO11-1-2 +/- or -/-, FTL -/- and qrt +/+.* These plants are ready for further selection to make the first 3 alleles homozygous to get plant "B".

7   Discussion and Outlook

7.1   Generation of *At*SPO11-1/2 fusion proteins

The generation of *At*SPO11-1/2 fusion proteins prior and during this study was a pre-requisite for all projects discussed in this thesis. Since no antibodies raised against the two *At*SPO11 proteins were available five years ago, we asked the question, whether we should produce antibodies and/or generate tagged fusion proteins of *At*SPO11-1 and *At*SPO11-2. We decided to start with the generation of different tagged fusion proteins, to have tools, to examine the roles of *At*SPO11-1 and *At*SPO11-2 in *Arabidopsis thaliana*. In the past, several attempts to generate *At*SPO11-1 fusion proteins for the expression in plants have been performed. But a common technique in the plant research community, namely to use cDNA-based vectors for protein expression and complementation, was not successful for *At*SPO11-1. The corresponding transgenic lines were not complementing an endogenous *At*spo11-1 mutation and resulted in sterile plants (Christine Mezard, Mathilde Grelon, INRA Versailles, personal communication). To avoid such problems for *AtSPO11-1* and also *AtSPO11-2*, it was decided to generate full-length genomic clones for both, containing long regions up- and down-stream of the gene, harboring 5' and 3' untranslated regions (UTRs), promoter regions and potential regulatory downstream sequences. In the meantime antibodies for *At*SPO11-1 and *At*SPO11-2 were produced by other lab members of the Schlögelhofer lab. The antibodies were tested for functionality in a series of experiments and seem to be very promising tools for the future (C. Uanschou, E. Altendorfer, M. P. Janisiw, P. Schlögelhofer, unpublished results).

One reason to generate a 18xmyc tagged version of *At*SPO11-1 and *At*SPO11-2 was the aim to perform a cytological localization of the proteins. Since from *S. cerevisiae* it is known that at least nine other proteins are involved in DSB formation, we thought of the possibility that *At*SPO11-1 and *At*SPO11-2 induce DSBs, embedded within a multiprotein complex. So it might be possible that accessory proteins hinder the immunological detection of *At*SPO11-1/2 by masking the immunogenic epitopes (Janisiw, 2009). To exclude this kind of pitfall, we have chosen the 18xmyc epitope, which seemed to be long enough for not being masked by other proteins of the *At*SPO11 cleavage complex. The second long tag was the 24xHA epitope, which was fused to *At*SPO11-1 and *At*SPO11-2 due to the same reasons described above. On one hand, such long tags are very risky because they may interfere with the structure and the folding of a protein, resulting in the expression of non-functional or truncated proteins. On the other hand, the 18xmyc tag is

very promising, as Zachariae and colleagues purified the anaphase promoting complex in *S. cervisiae* using this 18xmyc epitope (Zachariae *et al.*, 1998; Zachariae *et al.*, 1996). Other reasons for the generation of fusion proteins have been I) to have the possibility to perform Co-IP experiments using either differently tagged *At*SPO11-1 or 2 proteins, II) to perform ChIP experiments followed by mass spectrometry for the identification of SPO11 interacting proteins and III) to perform ChIP-on-chip and ChIP-Seq experiments for the identification of DSB sites. Such approaches are most of the time much more efficient using an antibody raised against a tag than using an antibody against the whole protein or a peptide.

AtSPO11-1/untagged, *AtSPO11-1/18xmyc* and *AtSPO11-1/Gal4* were generated by Ondrej Krsicka during his diploma thesis and were further processed by myself during my PhD thesis. All of these transgenic lines are expressing functional proteins, since full complementation of the severe sterility phenotype of *Atspo11-1-2* mutants was observed.

In addition to these three existing *AtSPO11-1* clones, I generated two other ones, harboring a 24xHA epitope and a SF tag, respectively. For *At*SPO11-2, I generated an untagged version, and clones harboring a 18xmyc tag, a 24xHA tag and a Gal4 binding domain. *At*SPO11-1/SF, *At*SPO11-2/untagged and *At*SPO11-2/18xmyc are functionally expressed. *At*SPO11-1/SF and *At*SPO11-2/untagged show full complementation of the endogenous mutations, whereas for *At*SPO11-2/18xmyc a reduced seed level was observed (12 seeds/silique vs. 45-50 in wild type), indicative for a partial complementation. Both 24xHA clones and the *AtSPO11-2/Gal4* clone are not complementing respective mutant populations, indicative for missing expression or non-functionality of the fusion proteins. This leads to the assumption that the 24xHA tag interferes with the protein expression of either *At*SPO11-1 or *At*SPO11-2. Maybe it is the length of the tag and the amino acid composition which change the net-charge and the hydrophobicity of the protein, resulting in non-functionality or missing protein expression. Although 18xmyc and Gal4 fused to *At*SPO11-1 show full complementation, this is not the case for *At*SPO11-2. *At*SPO11-2/18xmyc shows partial complementation, whereas *At*SPO11-2/Gal4 plants are sterile, indicative for missing expression or non-functionality of the fusion protein. Most probably, the C-terminus of *At*SPO11-2 doesn't allow the fusion of a tag without negative influence on the expression of proteins, since from three tagged lines only one shows partial complementation. The clone itself was fine, since the untagged version restored the seed level of a *Atspo11-2-3* mutant population to wild-type level. Therefore the fusion of tags with the N-terminus of *At*SPO11-2 should be considered for the future (see section 6.1 for all details).

7.2 Biochemical detection of tagged *At*SPO11 fusion proteins

For the biochemical detection of the tagged *At*SPO11 fusion proteins, immuno-precipitation experiments followed by Western blot analysis were performed using antibodies directed against the different tag epitopes. The first experiments were performed with biological material from *AtSPO11-1/18xmyc* transgenic lines. The calculated molecular weight for the corresponding fusion protein is 68.9 kDa, the observed size on the blot was ~90 kDa. A possible explanation is known from previous studies (Marc Behrlinger, personal communication). The fusion of a protein with the long 18xmyc eptiope changes the running behavior of the protein on a SDS gel, resulting in a visible shift of up to 30 kDa. This was observed in the presented study for all 18xmyc tagged fusion proteins. In addition, this phenomenon was observed as well for proteins fused with the long 24xHA epitope (see figure 6.9).

*At*SPO11-1/24xHA was not complementing the endogenous mutation and not detectable on protein level, indicative for no protein expression or the expression of a non-functional and truncated protein without tag. The two other *AtSPO11-1* clones harboring a SF tag or a Gal4 BD were functional but not detectable. Most probably the antibodies used to detect the fusion proteins were not optimal, since also positive controls failed to give results (see figure 6.9).

Of the three tagged *At*SPO11-2 fusion proteins, two could be detected by Western blot analysis; the partially complementing *At*SPO11-2/18xmyc (calculated molecular weight of 70.2 kDa) and the non-functional expressed *At*SPO11-2/24xHA (calculated molecular weight of 71.9 kDa). Both were detectable with a size of ~90 kDa, which is consistent with the band shift mentioned above. It is interesting that in contrast to *At*SPO11-1/24xHA, the 24xHA tagged version of *At*SPO11-2 is detectable on protein level, although both are not functionally expressed in plants, not complementing respective mutant populations. Reviewing these results, one can assume that *At*SPO11-2/24xHA is expressed as a non-functional full length protein. Since the tag is at the C-terminus directly in front of the stop codon, the protein could only be detected as a full length version. Since *At*SPO11-1/24xHA is not detectable on protein level, one can assume that the protein is not expressed at all or only as a truncated, non-functional version without tag (see figure 6.9).

To go into more detail, the expression of all transgenes should be analyzed on mRNA level by RT-PCR.

7.3   Genome-wide identification of DNA DSB sites using ChIP-Seq

As mentioned in the results section, chromatin immuno-precipitation followed by deep sequencing (ChIP-Seq) is a powerful tool to study interactions between DNA and proteins. In earlier studies ChIP DNA was exclusively analyzed by qPCR, Southern blotting or hybridization to genomic DNA microarrays (ChIP-on-chip). Although widely in use, ChIP-on-chip approaches have several drawbacks, including a low signal-to-noise ratio and a need for replicates to build statistical power to support putative binding sites (Mardis, 2008). Advantages of ChIP-Seq are the single nucleotide resolution, the unlimited coverage, the low amount of input material, the dynamic range and the high number of sequence reads obtained from one single experiment. Additionally, multiplexing experiments using different barcoded adapters are possible, thereby decreasing the costs (Park, 2009).

The final aim of this project was the genome-wide identification of DSB sites in *Arabidopsis thaliana*. For this purpose a ChIP-on-chip approach was started. As input material buds of the transgenic plant lines *AtSPO11-1/18xmyc* (sample) and *AtSPO11-1/untagged* (control) were used. All experimental steps, like DNA fragmentation, ChIP, labeling and amplification steps worked properly and there was enough DNA for hybridization to the *Arabidopsis* tiling array (Affymetrix). Two microarray hybridizations (biological repeats) were made in the lab of Martin Bilban (Medical University of Vienna, AKH), one hybridization (technical repeat) was performed in the European Affymetrix Service facility (NASC, Nottingham). For all three experiments the same results were obtained, mainly no hybridization signals representing DSB sites in *Arabidopsis thaliana*, but background signals distributed over all five chromosomes for the sample and the control. In total, only three regions with signals above the background level were identified for the tagged sample. The best explanation for these results is that the amount of input material was too low. Considering these results and the advantages of ChIP-Seq experiments, we decided to develop a ChIP-Seq protocol for the identification of DSB sites. After the generation of DSBs, Spo11 proteins are removed from DNA covalently attached to short oligonucleotides. The 5' ends of these oligonucleotides precisely represent the positions where DSBs are made by the catalytic tyrosine residue of Spo11. Direct sequencing of these oligonucleotides in a high-throughput manner will result in the sequence information to generate genome-wide DSB maps and will help to gain much more insight into the process of Spo11-dependent DSB formation (for all details see section 6.2).

7.3.1 The model substrate

To develop a ChIP-Seq protocol for the genome-wide identification of DSB sites in various organisms, the first step was to establish and optimize the method using a model substrate, a HRP protein conjugated with up to three 30mer oligonucleotides of known sequence. Position 30 of the oligonucleotides was a "N", to see whether there is a ligation efficiency bias towards one nucleotide during the adapter ligation (see figure 6.11).

At the beginning of this project different strategies for the ligation of the Solexa adapters to the single stranded oligonucleotides coupled to the HRP protein were tested. Since approaches using Prokaria single strand ligase were not successful, a different protocol was developed and evaluated. Using terminal transferase and rGTPs, a controlled addition of Gs is possible, since it is known from (Schmidt and Mueller, 1996) that at a rGTP concentration of 0.5 µM, 2 or 3 Gs are added to an extent of 98%. After purification of the reaction using Vivaspin 500 columns, the first adapter ligation was performed using a double stranded adapter with a "CCN" overhang which is compatible with the 2 or 3 rGTPs. All steps of the protocol were experimentally tested and worked very efficiently. But in terms of the "real" experiment using different model organisms we thought about a much more efficient way for adapter ligation. The final protocol which is described in detail in the materials and methods section, uses Solexa adapters with a five "N" overhang which is directly compatible for ligation without adding rGTPs. The use of these adapters has the advantage to get rid of two enzymatic reactions with subsequent column purification steps, avoiding the potential loss of input material.

For this Solexa run 14.5 million total reads were obtained, 12.3 million perfectly matching the model substrate sequence, indicative for a very well developed protocol. A ligation efficiency bias towards one nucleotide (A: 20%, C: 25%, T: 40%, G: 15%) was observed, which may have different reasons. Maybe the bias is the result of an unbalanced synthesis of the model substrate oligonucleotides. Therefore the composition of the used model substrate should be checked by mass spectrometry. Prior to all other speculations, to exclude experimental discrepancies, the experiment has to be repeated, to see whether the same bias is obtained in a second experiment (see section 6.2.1).

To evaluate the technique, we started with *S. cerevisiae*, because it is easy to cultivate a large amount of synchronized cells at the accurate meiotic stage and results can be compared to already published DSB maps.

7.3.2 *Saccharomyces cerevisiae* ChIP-Seq

For *S. cerevisiae* 5×10^9 cells of a SK1 strain harboring a 18xmyc tagged Spo11 were used as input for the ChIP-Seq experiment. After the first amplification PCR with adapter specific primers and separation of the products by agarose gel electrophoresis, a very dominant band (95 bp) with a smear above was observed (see figure 6.13). The smear contains amplicons which may have their origin from the amplification of oligonucleotides which were attached to Spo11. The dominant band has its origin from the ligation of the second Solexa adapter to the biotinylated unidirectional PCR primer. This may happen, since after the unidirectional PCR, the biotinylated PCR product is captured from the reaction mix, which still contains the non-elongated biotinylated primer, using streptavidin beads. Therefore the primer is also binding to the streptavidin beads. In the following on-bead ligation, the second adapter is ligated to the unidirectional PCR product and to the unidirectional primer. Efforts were made to get rid of the unidirectional primer after the unidirectional PCR and the primer/adapter dimer after the first PCR amplification using different column purification methods. In the end, neither method was successful to avoid the primer/adapter dimer formation or to get rid of the dimer. The magnitude of such an adapter dimer formation can be measured by counting the proportion of reads representing the sequence of any of the adapters used. The problem can be minimized by optimizing the ratio of sample to adapter molecules in the ligation steps (Kaufmann *et al.*, 2010). This strategy was applied and additionally we decided to make the separation between the primer/adapter dimer and the smear by slicing out the smear from the gel. This is not a 100% separation of both species, since the dominant primer/adapter dimer and the smear are close together without any gap. Since we thought that the amount of amplified DNA in the smear could be too little input for the deep sequencing, a re-amplification step with both adapter specific primers was performed. The results were very promising, since after the re-amplification a band-shift of the dominant band was observed for the tagged sample but not for the untagged control, indicating the amplification of sequences which were attached to Spo11 and containing both Solexa adapters. To discuss the re-amplification step it has to be mentioned that the analysis of the distribution of the number of reads representing the same sequence is important to detect any artefact due to the PCR amplification. Loss of complexity due to PCR amplification leads to a higher number of reads having the same sequence. Such artefacts can be minimized by reducing the cycle number for PCR amplifications during the sample preparation (Kaufmann *et al.*, 2010). Although the results for our ChIP-Seq experiments are very promising and highly comparable to already published DSB maps, we should consider a

reduction of the PCR cycle number in the future to obtain as less PCR artefacts as possible.

From the 36 bp run, 26 million total reads were obtained for the tagged sample. 7.9 million were used for mapping, 5.5 million mapped to the *S. cerevisiae* genome, thereof 4.2 million uniquely. For the control, 28 million sequences reads were obtained and 8.6 million used for mapping. 8.3 million mapped to the *S. cerevisiae* genome, 1.1 million uniquely. The mapping statistic for the 72 bp run, where only a tagged sample was processed, is as follows: 8.4 million total reads, 3.3 million were used for mapping, 1.8 million mapped to the *S. cerevisiae* genome, thereof 1.5 million uniquely. It has to be mentioned that the number for mapped reads consists of the number of mapped reads plus the number of positions where the reads mapped more than once. Since the SK1 genome assembly is incomplete, the data was mapped to the S288C genome. For all further analysis (except long oligonucleotide distribution) the unique data set of the 36 bp run was used.

This data represents the first genome-wide DSB map at nucleotide resolution for *S. cerevisiae*. As shown in the results section, we have the sequence coverage information for every position in the genome, and plotting only the 5' ends of the mapped oligonucleotides results in maps containing the exact positions of the Spo11 catalyzed cleavage reaction of DSB formation. For the future this sequence information will be a powerful tool to learn more about the distribution of DSBs and their regulation by determining sequence and structural motifs responsible for DSB formation and processing (see figure 6.15).

For a detailed evaluation of the data, it was compared to already published DSB maps from the Lichten lab, which were generated by ChIP-on-chip experiments using long tracts of enriched single stranded DNA, associated to DSBs accumulating in *dmc1Δ* mutants, as input material (Buhler *et al.*, 2007). Plotting both data sets, a high degree of overlap was observed (see figure 6.16). For a more detailed comparison, a bioinformatical association study was performed, which is described in the results section. Using 2 kb windows, 4158 significant signals were detected in our data set and 2499 in the ChIP-on-chip data. 2117 were detectable in both data sets, which corresponds to an overlap of 84.71% (see table 6.2).

Furthermore our data contains the sequence information for both, the Watson and the Crick strand, for future evaluation of asymmetry in early DSB processing steps according to the ideas of (Neale *et al.*, 2005). Matthew Neale and colleagues proposed three models to explain the 1:1 ratio of short and long oligonucleotides observed for *S. cerevisiae*. Either DSBs are processed asymmetrically resulting in one short and one long

oligonucleotide or they are processed symmetrically resulting in two short or two long oligonucleotides, with the same genome-wide likelihood, or resection starts close and far away from the DSB site, resulting in long and short oligonucleotides to the same extent.

One of the most striking observations of the presented study was the detection of Spo11 attached oligonucleotides with a length up to 72 bp, which mapped to the *S. cerevisiae* genome. 2.5% of the oligonucleotides were between 41-60 nt in length, 0.6% were longer than 60 nt. Since 72 bp was the read length for the long single end Solexa run, we suggest the existence of oligonucleotides longer than 72 bp. To determine the maximum oligonucleotide length, paired end samples were prepared. This means that the amplified products are sequenced from both sites, resulting in the sequence information to obtain the full length sequence of the Spo11 attached oligonucleotides. During the writing of this thesis, samples were processed, sequenced and delivered for bioinformatical processing. The presented data is consistent with the idea (M. Neale, personal communication) that the incision for DSB resection takes place further away from the break site with a subsequent need for 3'-5' exonuclease activity. In such a scenario, the endonucleolytic activity of MRX/Com1 would be exerted first in a distance. Starting from this nicked DNA site, the exonucleolytic 3'-5' activity of the MRX complex would work towards the Spo11 protein, and the 5'-3' exonucleolytic activity of Exo1 is moving out to the other direction. Due to a steric inhibition of the MRX complex by the proteins at the DSB site, resection comes to a hold at various distances from the DSB site, resulting in two different length classes of oligonucleotides attached to Spo11.

To analyze the occurrence of DSBs with regard to gene coding regions, 56 hot regions, distributed all over the genome, were chosen and their location was analyzed. 47 were determined to be intergenic, lying in promoters, only nine were intragenic, which correlates with the findings published by (Blitzblau *et al.*, 2007). They examined the ssDNA enrichment (the more ssDNA enrichment, the hotter a region) across the largest 226 hotspot-associated genes and found that 80% of the genes exhibited higher ssDNA signal in their promoters relative to the coding regions. Together, these findings are consistent with the model that the majority of DSBs occur in intergenic regions containing promoters (Blitzblau *et al.*, 2007). Additionally we found substantial DSB formation in pericentromeric regions within 1-5 kb of the core centromeres for all chromosomes, which has never been reported before (see section 6.2.2.4). Earlier ChIP-on-chip studies (Gerton *et al.*, 2000) in the *rad50S* background showed that DSBs are absent in the pericentromeric region (centromere +/- 25 kb), Blitzblau and colleagues detected DSB within 5-10 kb of the core centromere (Blitzblau *et al.*, 2007). A possible explanation for

our results is that the resolution of methods for DSB detection has increased, therefore the size for regions with suppressed DSB formation has decreased.

A further finding was the negative correlation of mapped Spo11 oligonucleotides with Rec8 binding sites as expected from prior studies (Kugou *et al.*, 2009). Especially at the centromeres, where a high signal for Rec8 binding could be observed, only very low DSB activity is present (see figure 6.18). Kugou and colleagues found that Spo11 initially accumulates to the centromeres and then relocates to arm regions during progression of pre-meiotic S-phase. During this stage a substantial proportion of Spo11 binds to Rec8 binding sites. Since binding of Spo11 and further DSB formation is severely affected in *rec8Δ* mutants, they speculate that Rec8 may provide some molecular landmarks along the meiotic chromosome to ensure the proper distribution of Spo11/DSB sites. For example, Rec8 at centromeres may provide origins for loading of Spo11 onto the chromosome arms, whereas Rec8 at cohesion sites may provide positional and geographical information (e.g. positions of loops and axes), to coordinate loading of Spo11 onto canonical DSB sites. Rec8 at both sites may be crucial for the proper targeting of Spo11 to canonical DSB sites. Indeed, Spo11 might eventually shift from the Rec8 binding sites to chromosome loop regions including DSB sites. Such spatial transition of Spo11 may prevent recombination between sister chromatids to facilitate inter-homolog recombination (Kugou *et al.*, 2009).

To answer biological questions and to shed light on the determinants of DSB distribution with regard to different mutant backgrounds in *S. cerevisiae*, a multiplexing experiment was performed. 5×10^9 cells of the mutants *red1Δ, hop1Δ, mek1Δ, red1Δ/hop1Δ*, *rec8Δ and rec8Δ/scc1* were processed as described to compare the pattern of DSB distribution to that of wild-type cells. During the writing of this thesis samples were processed and the raw data is currently in the bioinformatical pipeline. Red1, Hop1 and Mek1 are axial element components of the synaptonemal complex and necessary for chromosome segregation during the first meiotic division. All three mutants exhibit reduced levels of interhomolog recombination and produce no viable spores (Blat *et al.*, 2002; Hollingsworth and Ponte, 1997; Klein *et al.*, 1999; Page and Hawley, 2004; Smith and Roeder, 1997). Rec8 is a cohesin component and in the deletion mutant meiotic DSB formation and the binding of Spo11 to DSB sites are severely impaired at selective domains of many chromosomes (Kugou *et al.*, 2009). Scc1 is the kleisin of the mitotic cohesion complex and Rec8 is its meiotic paralog (Guacci *et al.*, 1997; Klein *et al.*, 1999; Michaelis *et al.*, 1997).

7.3.3 *Schizosaccharomyces pombe* **ChIP-Seq**

5×10^9 cells from a haploid strain with a 13xmyc tagged Rec12 were taken for the Solexa experiment and processed as described in previous sections. For the tagged sample 14.5 million total reads were obtained, of which 2 million were used for mapping. 1.2 million sequence-reads mapped to the *S. pombe* genome, thereof 612000 uniquely. For the untagged control 2.7 million out of 18 million reads were used for mapping. 806000 mapped to the *S. pombe* genome, 274000 uniquely. These are much lower numbers compared to the *S. cerevisiae* experiment. Although the *S. cerevisiae* genome and the *S. pombe* genome have the same size and one would assume that the same number of reads will be obtained for one sequencing run, it seems likely, as *S. pombe* DSB numbers are not that high than in *S. cerevisiae*. Additionally, we observed a lower coverage for single *S. pombe* hotspots than for *S. cerevisiae*, maybe resulting from a lower amount of processed cells due to a lower efficiency during the cell lysis. A second possible explanation is the lower degree of cell synchrony obtained for *S. pombe* compared to *S. cerevisiae*. Therefore it might be that for *S. pombe* the percentage of cells (from a total input of 5×10^9 cells) at the accurate meiotic stage is lower than for *S. cerevisiae*.

The data for *S. pombe* was not compared in that detail to *S. pombe* hotspot maps as it was done for *S. cerevisiae*, since no comparable data was accessible in a suitable form. When looking on the length distribution of the mapped reads, we observed that the majority of the mapped reads has a length between 15 and 20 nt which is consistent with the findings published by (Milman *et al.*, 2009), where a single class of Rec12 attached oligonucleotides was detetced. Nevertheless, DSB profiles were compared with already published DSB maps from the Smith lab and a remarkable overlap was found (Hyppa *et al.*, 2008). Due to the better resolution of our technique we found a substantial number of DSB sites which were not detected by the ChIP-on-chip experiments mentioned before. As proof of principle, the *mbs1* and *mbs2* hotspot region on chromosome I was taken for a detailed peak comparison. As shown in figure 6.22 and 6.23, we found a remarkable overlap of our data and the Smith data.

Since the *S. pombe* Solexa run was a 36 bp run and we want to know more about the maximum length of the Rec12 attached oligonucleotides, paired end samples for the determination of the maximum oligonucleotide length were prepared and are currently delivered for sequencing.

To analyze the dependencies of DSB formation with regard to different mutant backgrounds, a multiplexing experiment was performed. 5×10^9 cells of the mutants *hop1Δ*,

mek1Δ, rec10-155, rec27Δ and *mug20Δ* were processed as described to compare the pattern of DSB distribution to that of wild-type cells. During the writing of this thesis, samples were processed and the raw data is currently in the bioinformatical pipeline. It is anticipated to gain more information about the global, genome-wide DSB pattern of these mutants, since only the information for several genetic intervals is available so far or DSBs were not detectable due to technical limitations. Hop1 and Mek1 are both necessary for meiotic chromosome pairing and recombination partner choice in *S. pombe*. For both mutants, intergenic (crossover) and intragenic (conversion) recombination is reduced in certain genetic intervals and DSB levels are lowered (Latypov *et al.*, 2010). Rec10 is a LinE protein and is absolutely required for their development (Lorenz *et al.*, 2004). Whereas a *rec10* null mutant is defective in DSB formation and meiotic recombination, the *rec10-155* mutant is not. Rec10-155 is a C-terminal truncated protein, does not form LinEs but exhibits DSB formation and meiotic recombination (Wells *et al.*, 2006). Rec27 is also a component of the LinEs and is important but not absolutely necessary for meiotic recombination. Although meiotic DSBs are not detectable in the chromosomal regions tested in *rec27Δ* mutants, recombination is not reduced to the same extend as in other *rec* mutants (Davis *et al.*, 2008). Mug20 was co-purified with Rec10 in a TAP approach followed by LC-MS/MS analysis (Spirek *et al.*, 2010). It cytologically co-localizes with Rec10 and the deletion mutant exhibits aberrant LinE formation (Anna Estreicher, Josef Loidl, Department of Chromosome Biology, University of Vienna, unpublished results).

7.3.4   *Arabidopsis thaliana* ChIP-Seq

To identify DSB sites in *Arabidopsis thaliana*, a ChIP-on-chip approach was started using buds of the *AtSPO11-1/18xmyc* line as input material. Two biological and one technical repeat were performed, but for all three experiments the same unsatisfying results were obtained. No hybridization signals representing DSB sites were observed, but background signal distributed over the whole genome. Two possible explanations for these results are I) the low amount of input material and II) the unconstant growth conditions for the plants (both aspects are discussed later in the ChIP-Seq part). Considering these results, we decided to start a ChIP-Seq approach for the identification of DSB sites in *A. thaliana*.

After establishing the technique for a model substrate, *S. cerevisiae* and *S. pombe*, two 36 bp runs (one technical repeat) were perfomed, using buds from the *AtSPO11-1/18xmyc* line as input material. As a control the untagged line was used.

In both runs, for sample and control nearly the same number of reads mapped to the *Arabidopsis* genome and the reads for both were distributed over the chromosomes, showing no cluster formation (hotspots) for the sample. Due to a high degree of overlap between sample and control, it was not possible to identify potential DSB sites when plotting both data sets. Therefore only unique sample coverages were included in further analysis from positions where the coverage for the control was zero.

Compared to *S. cerevisiae* and *S. pombe* only 40-60 single peaks/chromosome were detectable for *Arabidopsis*. This corresponds to an average of only one peak/500 kb. Deeper analysis at nucleotide resolution showed that all peaks are really single peaks, resulting from one sequence read which mapped to the respective region. As shown for yeast, one would expect a Gaussian distribution of the reads mapping to a hotspot, resulting from different overlapping sequence reads, mapping to such a region (see figure 6.25).

Furthermore the results of the first run were not reproducible in the second run (technical repeat), detecting only two peaks at nearly the same position.

Nevertheless, the data was compared with already published recombination maps for *A. thaliana*. Drouaud and colleagues published a study dealing with the variation of cross over rates across chromosome IV, revealing the presence of meiotic recombination hotspots (Drouaud *et al.*, 2006). They genotyped 71 single nucleotide polymorphisms (SNPs) covering chromosome IV of *Arabidopsis* on 702 F2 plants, representing 1404 meioses and allowing the detection of 1171 COs, to study the CO localization in a higher plant (Drouaud *et al.*, 2006). We were able to detect only one potential signal overlap, which was located between the knob and NOR region at the beginning of chromosome IV. It has to be mentioned that we are working in a homogenous Columbia background and the work published by (Drouaud *et al.*, 2006) was performed in a heterogenous Columbia/Landsberg background, which could be a good explanation for the low percentage of overlap. It is known from other studies that recombination patterns between different *Arabidopsis* ecotypes can vary greatly.

Furthermore, our data was compared with data from genome-wide association studies performed in the Nordborg lab (Kim *et al.*, 2007), but no overlaps with the observed recombination events were detectable. Kim and colleagues described the genome-wide pattern of linkage disequilibrium (LD) in a sample of 19 *Arabidopsis thaliana* accessions using 341,602 non-singleton SNPs. The extent of LD is highly variable, and they could find clear evidence of historic recombination hotspots, which correlate well with recent

recombination events. Moreover, it has to be mentioned that we are looking on DSB sites and not on recombination. Since not all DSBs result in recombination events, one has to be careful in comparing DSB and recombination data.

In general, the most probable explanation for these unsatisfying results is the low amount of input material which was used for the Solexa experiments. Since for yeast 5×10^9 synchronized cells at the accurate meiotic stage are used, one would need approximately 2.5 million inflorescences to obtain the same number of meiocytes for *Arabidopsis* (not staged). A second possible explanation is the negative influence of the plant growth conditions (temperature, humidity) on the hotspot distribution and the reproducibility of the results. As published in (Francis *et al.*, 2007; Molinier *et al.*, 2006) plants are very sensitive to environmental changes concerning their recombination control. Francis and colleagues measured the sensitivity of recombination frequencies to temperature, growing plants at 20°C under long day conditions until they began to bolt. Then they shifted temperature between 19°C and 28°C under constant light and assayed recombination frequencies. For all intervals tested, they observed a significant increase of recombination with increasing temperature (Francis *et al.*, 2007). Considering these aspects it could be possible on the one hand that we produce somehow a "zero line" of DSBs, using unconstant growth conditions for one plant population. On the other hand, unconstant growth conditions may interfere with reproducibility of the results between different populations.

To combine both aspects for improving the technique in *Arabidopsis*, we have grown a huge population of 20000 plants under constant growth conditions in the new GMI growth chambers. For the detailed conditions please refer to section 6.2.4. During writing of this thesis, buds of the sample and control population were harvested, processed and samples were delivered for sequencing.

7.4   Identification of *At*SPO11-1 and *At*SPO11-2 interacting proteins

In *S. cerevisiae*, in addition to Spo11, at least nine other proteins are involved in DSB formation. In *Arabidopsis*, only three proteins have been identified to be essential for DSB formation together with *At*SPO11-1 and *At*SPO11-2.   In primary screens for reduced fertility and secondary screens for the absence of DSBs, the genes *AtPRD1*, *AtPRD2* and *AtPRD3* were identified recently (De Muyt *et al.*, 2009; De Muyt *et al.*, 2007). The only

direct interaction partner for *At*SPO11-1, which was identified in a yeast two-hybrid assay, is *At*PRD1, which interacts with its N-terminus with the meiotic key player *At*SPO11-1.

Compared to other organism like *S. cerevisiae* and *S. pombe,* not much is known about the initiation of meiotic recombination in *Arabidopsis*, therefore many questions remain elusive. Are there other proteis involved in DSB formation, how do they interact in a complex and do *At*SPO11-1 and *At*SPO11-2 form a heterodimer.

To identify *At*SPO11-1 and *At*SPO11-2 interacting proteins, different strategies were chosen. The first approach, mainly described in this thesis, is a protein complex purification followed by LC-MS/MS analysis of co-purifying partners, using the *At*SPO11-1/18xmyc fusion protein as bait. Two biological repeats were performed, but in both experiments the bait protein was not found by mass spectrometry analysis, resulting also in no identification of promising candidates for being interaction partners of *At*SPO11-1. Protein lists for both experiments were compared and analyzed, searching for proteins which are present in the tagged sample only, but not in the untagged control. Proteins were stated as potentially interesting, if detectable in both experiments in the tagged sample only (see section 6.3.1).

Possible explanations for these results are I) the low amount of input material, II) the strong interaction of the 18xmyc tag with the antibody and III) the single step purification instead of a TAP purification. To be more efficient in the future, we have to scale up the amount of input material to reach comparable levels of meiocytes, which are used with other organisms. At the moment a huge population of *Arabidopsis* plants (~20000 plants) is growing under constant growth conditions which can be used for such experiments. The second aspect is the strong interaction of the 18xmyc tag with the myc antibody. It was not possible to elute the bait protein including potential interaction partners using conventional methods which are compatible with direct mass spectrometry analysis. Elution was only successful, when adding LDS protein sample buffer and heating at 95°C. This method has the disadvantage that not only the proteins of interest are eluted with a little background, but also all of the proteins unspecifically bound to the beads. The third point includes the way of purification. We used only a single step purification instead of a TAP purification. One the one hand the 18xmyc tag is very promising, as Zachariae and colleagues purified the anaphase promoting complex in *S. cervisiae* using this 18xmyc epitope (Zachariae *et al.*, 1998; Zachariae *et al.*, 1996), on the other hand TAP purifications are state-of-the-art, almost exclusively used for protein complex purifications.

The analysis of protein-protein interactions under native conditions has been a challenge ever since immuno-precipitation became a common methodology. For a long time low yields and non-specific binding of proteins have been associated with IP (Gloeckner *et al.*, 2009b). Identification of protein complexes by mass spectrometry has allowed investigators to connect cellular pathways and to describe the dynamics of protein complexes (Aebersold and Mann, 2003; Yates, 2004). These approaches typically require a high degree of purification. To get a genuine picture of the *in vivo* situation it is essential to avoid any changes in protein modification or protein complex composition that might occur during the purification procedure (Tagwerker *et al.*, 2006).

Two-step purification strategies have been proven to be very effective in reducing non-specific background, which is important for the analysis of complex protein samples (Puig *et al.*, 2001). The first widely and successfully used tandem affinity tag was the TAP tag, which consists of the immunoglobulin interacting domain of Protein A, a calmodulin binding peptide (CBP) and a TEV cleavage site, which allows a sequential purification based on ProteinA/IgG-agarose and CBP/calmodulin bead affinities (Rigaut *et al.*, 1999). Other tandem affinity tags include a modified version of the TAP tag in which the CBP part is replaced by the S-tag (Cheeseman *et al.*, 2001) and combinations of multihistidine tags with FLAG (Denison *et al.*, 2005; Hannich *et al.*, 2005) or Myc tags (Graumann *et al.*, 2004). Parts of the text were taken from (Tagwerker *et al.*, 2006). Several features of such first generation tags remain suboptimal, such as the high mass of 21 kDa for a standard TAP tag, the dependency on the proteolytic cleavage, and CBP, which may interfere with calcium signaling within eukaryotic cells (Gloeckner *et al.*, 2009b).

To switch to such a tandem purification strategy, I generated a *At*SPO11-1 fusion protein harboring a SF tag. This Strep/FLAG tag was developed by (Gloeckner *et al.*, 2007) and combines a tandem Strep-tag II and a FLAG tag, resulting in a small 4.6 kDa tag. Both moieties have a medium affinity and avidity to their immobilized binding partners. Therefore the tagged proteins and their binding partners can be recovered under native conditions without the need for time-consuming proteolytic cleavage. In the first step, desthiobiotin is used for the elution of the SF-TAP fusion protein from the Strep-Tactin matrix. In the second step, the FLAG octapeptide is used for elution of the protein from the anti-FLAG M2 affinity matrix (Gloeckner *et al.*, 2009b). As mentioned before, the *At*SPO11-1/SF fusion protein is functionally expressed, rescuing the *Atspo11-1-2* mutant phenotype. Until the end of this thesis we were not able to detect the fusion protein by Western blot analysis, most probably due to problems during the detection with the anti-FLAG antibody which is not working properly. Establishing the protocol for the purification

of *At*SPO11-1 interacting proteins using this SF tag seems to be very promising, since this tag combines all features which are state-of-the-art for protein complex purification.

Two other approaches for the identification of *At*SPO11-1/2 interacting proteins were direct yeast two-hybrid mating studies with *At*SPO11-1 and *At*SPO11-2 as bait proteins and CoIP experiments with *At*SPO11-1 and *At*SPO11-2 expressed in wheat germ as described in the results. For the direct yeast two-hybrid mating studies no interaction between the two proteins was observed. It seems to be non-trivial, since until now no *At*SPO11 interaction partner was initially identified with this method. Although it was published that the N-terminus of *At*PRD1 interacts with *At*SPO11-1 in a yeast two-hybrid screen (De Muyt *et al.*, 2007), we were not able to reproduce these results in our lab (Elisabeth Altendorfer, Peter Schlögelhofer, personal communication).

After the expression of both *At*SPO11 proteins in wheat germ extract, using S35 methionine as labeled amino acid, CoIP experiments were performed. Protein expression was confirmed by Western blot analysis and for the Co-IP experiments protein extracts were mixed in all possible combinations including the appropriate negative controls. No interactions were observed in two replicates of these experiments.

In general, it seems to be difficult to handle the two *At*SPO11 proteins for interaction studies. With the knowledge from *S. cerevisiae*, one can assume that for the identification of SPO11 interacting proteins in *Arabidopsis*, *in vivo* conditions are necessary. Maybe it is essential that other factors and components of the DSB forming complex are present *in vitro*, to obtain binding of one of them with each other or with other complex partners. Other factors and components could be the presence of proteins involved in DSB formation, phosphorylation of proteins or the presence of DNA to stimulate *At*SPO11 in the search for partners.

7.5 Targeted stimulation of meiotic recombination in *Arabidopsis thaliana*

Meiotic recombination in *S. cerevisiae* is dependent on the formation of DSBs formed by the Spo11 protein. DSBs usually occur in intergenic regions with open chromatin accessibility, but many other determinants that control their distribution and frequency remain elusive. The targeted stimulation of meiotic recombination allows to study factors that control the genome-wide distribution of such sites and to examine hotspot behavior in

regions of choice by answering questions like: Why is a hot region hot? Can I turn a cold region into a hot one? Can I redistribute DSB patterns over the whole genome?

The more applied view in the plant community is that such artificial hotspots would bring benefits to crop breeders in establishing novel genetic combinations for selective breeding, which would have been difficult to obtain with conventional strategies.

In this study, to generate artificial hotspots in *Arabidopsis*, the 5xUAS/Gal4 system was used, which has been described to be functional in *S. cerevisiae* (Fukuda *et al.*, 2008; Pecina *et al.*, 2002; Robine *et al.*, 2007). Transgenic lines harboring a 5xUAS insertion and lines expressing an *At*SPO11-1/Gal4 BD fusion protein were generated as described. To monitor meiotic recombination, a fluorescent pollen based marker system (Francis *et al.*, 2007) was used, which is described in the results section. In a laborious and time-consuming process all the single features are crossed together with the aim that the *At*SPO11-1/Gal4 fusion protein binds to the introduced 5xUAS sites and thereby introducing DSBs, resulting in meiotic recombination events.

At the moment the crossing for three 5xUAS insertion sites is still in progress and it is too early to have evidence for the functionality of this system in *A. thaliana*.

Although targeted stimulation of meiotic recombination is well studied in *S. cerevisiae*, the relative timing at which the factors of the DSB machinery come into play in detail remains elusive. It has been shown that there is a temporal and regional correlation between replication and DSB formation. A plausible hypothesis is that the progression of replication creates a DNA or chromatin substrate that allows Spo11 to be recruited to future DSB sites. Since Spo11-Gal4 BD cannot introduce DSBs in a *clb5 clb6* double mutant, which is defective in pre-meiotic replication and in the chromatin transition this is more evidence that Spo11 can only cleave DNA that is replicated in the meiotic context and provides a mechanistic link between meiotic replication and the initiation of recombination (Pecina *et al.*, 2002).

In conclusion one can say that the targeted stimulation of meiotic recombination provides an alternative strategy with distinct features and practical advantages to study the distribution and regulation of DNA DSB formation. Cleavage remains under normal physiological control with respect to trans-acting determinants, and as a consequence DSBs are repaired by the homologous recombination pathway, allowing for the production of viable gametes and the recovery of novel recombinants (Pecina *et al.*, 2002). We therefore anticipate to utilize the future artificial hotspots in plants for such studies.

Furthermore, *At*SPO11-zinc finger (ZF) fusion proteins will allow specific targeting of nearly any desired locus in the plant genome to mediate exchange of genetic information there. These artificial meiotic nucleases have the potential to target meiotic recombination in crop plants to certain loci, and hence control the efficiency of recombination between certain plant traits. Generating an artificial hotspot of meiotic recombination, either via the GAL4 (Bernd Edlinger, Peter Schlögelhofer, this thesis) or ZF domains (Michael Peter Janisiw, Peter Schlögelhofer, unpublished results), will allow analysis of the nature of DSB formation and processing events in great detail, and will provide the plant meiosis research community with a valuable tool.

During this thesis, valuable tools and techniques for the future were generated and developed, to gain more insight into the process of DSB formation and meiotic recombination, especially for the higher eukaryote *A. thaliana*. The technically challenging experiments seem justified, judging from the wealth of information that has been and still is gained from detailed analysis of DSBs, performed in other organisms like *S. cerevisiae* and *S. pombe*. In the plant field it is, for instance, still unknown which kind of *trans* and *cis* acting factors determine the "hotness" of a certain genomic region. So far, the research has been driven by genuine interest, but plant breeders are getting more and more interested in efficiently exploiting naturally occurring beneficial traits of crop plants. Understanding meiotic recombination may provide the base to develop the tools to modify recombination rates at desired loci in crop genomes (Edlinger and Schlogelhofer, 2011).

8 References

Abe K, Osakabe K, Nakayama S, Endo M, Tagiri A, Todoriki S, Ichikawa H, Toki S. 2005. Arabidopsis RAD51C gene is important for homologous recombination in meiosis and mitosis. *Plant Physiol* **139**, 896-908.

Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. *Nature* **422**, 198-207.

Agashe B, Prasad CK, Siddiqi I. 2002. Identification and analysis of DYAD: a gene required for meiotic chromosome organisation and female meiotic progression in Arabidopsis. *Development* **129**, 3935-3943.

Ajimura M, Leem SH, Ogawa H. 1993. Identification of new genes required for meiotic recombination in Saccharomyces cerevisiae. *Genetics* **133**, 51-66.

Akutsu N, Iijima K, Hinata T, Tauchi H. 2007. Characterization of the plant homolog of Nijmegen breakage syndrome 1: Involvement in DNA repair and recombination. *Biochem Biophys Res Commun* **353**, 394-398.

Alani E, Padmore R, Kleckner N. 1990. Analysis of wild-type and rad50 mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. *Cell* **61**, 419-436.

Armstrong SJ, Caryl AP, Jones GH, Franklin FC. 2002. Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. *J Cell Sci* **115**, 3645-3655.

Arnheim N, Calabrese P, Tiemann-Boege I. 2007. Mammalian meiotic recombination hot spots. *Annu Rev Genet* **41**, 369-399.

Arora C, Kee K, Maleki S, Keeney S. 2004. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *Mol Cell* **13**, 549-559.

Bai X, Peirson BN, Dong F, Xue C, Makaroff CA. 1999. Isolation and characterization of SYN1, a RAD21-like gene essential for meiosis in Arabidopsis. *Plant Cell* **11**, 417-430.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2009. PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science* **327**, 836 - 840.

Baudat F, Nicolas A. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc Natl Acad Sci U S A* **94**, 5213-5218.

Becker D. 1990. Binary vectors which allow the exchange of plant selectable markers and reporter genes. *Nucleic Acids Res* **18**, 203.

Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**, 414-417.

Bergner LM, Hickman FE, Wood KH, Wakeman CM, Stone HH, Campbell TJ, Lightcap SB, Favors SM, Aldridge AC, Hales KG. 2010. A novel predicted bromodomain-related protein affects coordination between meiosis and spermiogenesis in Drosophila and is required for male meiotic cytokinesis. *DNA Cell Biol* **29**, 487-498.

Bhatt AM, Lister C, Page T, Fransz P, Findlay K, Jones GH, Dickinson HG, Dean C. 1999. The DIF1 gene of Arabidopsis is required for meiotic chromosome segregation and belongs to the REC8/RAD21 cohesin gene family. *Plant J* **19**, 463-472.

Bishop DK, Nikolski Y, Oshiro J, Chon J, Shinohara M, Chen X. 1999. High copy number suppression of the meiotic arrest caused by a dmc1 mutation: REC114 imposes an early recombination block and RAD54 promotes a DMC1-independent DSB repair pathway. *Genes Cells* **4**, 425-444.

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

Blat Y, Kleckner N. 1999. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* **98**, 249-259.

Blat Y, Protacio RU, Hunter N, Kleckner N. 2002. Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. *Cell* **111**, 791-802.

Bleuyard JY, Gallego ME, White CI. 2004. Meiotic defects in the Arabidopsis rad50 mutant point to conservation of the MRX complex function in early stages of meiotic recombination. *Chromosoma* **113**, 197-203.

Bleuyard JY, White CI. 2004. The Arabidopsis homologue of Xrcc3 plays an essential role in meiosis. *EMBO J* **23**, 439-449.

Blitzblau HG, Bell GW, Rodriguez J, Bell SP, Hochwagen A. 2007. Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr Biol* **17**, 2003-2012.

Blum H, Beier H, Gross HJ. 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**, 93-99.

Boateng KA, Yang X, Dong F, Owen HA, Makaroff CA. 2008. SWI1 is required for meiotic chromosome remodeling events. *Mol Plant* **1**, 620-633.

Borde V, Goldman AS, Lichten M. 2000. Direct coupling between meiotic DNA replication and recombination initiation. *Science* **290**, 806-809.

Borde V, Lin W, Novikov E, Petrini JH, Lichten M, Nicolas A. 2004. Association of Mre11p with double-strand break sites during yeast meiosis. *Mol Cell* **13**, 389-401.

Borde V, Robine N, Lin W, Bonfils S, Geli V, Nicolas A. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *Embo J* **28**, 99-111.

Boselli M, Rock J, Unal E, Levine SS, Amon A. 2009. Effects of age on meiosis in budding yeast. *Dev Cell* **16**, 844-855.

Broderick S, Rehmet K, Concannon C, Nasheuer HP. 2010. Eukaryotic single-stranded DNA binding proteins: central factors in genome stability. *Subcell Biochem* **50**, 143-163.

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 Biol* **5**, e324.

Bundock P, Hooykaas P. 2002. Severe developmental defects, hypersensitivity to DNA-damaging agents, and lengthened telomeres in Arabidopsis MRE11 mutants. *Plant Cell* **14**, 2451-2462.

Cao L, Alani E, Kleckner N. 1990. A pathway for generation and processing of double-strand breaks during meiotic recombination in S. cerevisiae. *Cell* **61**, 1089-1101.

Carney JP, Maser RS, Olivares H, Davis EM, Le Beau M, Yates JR, 3rd, Hays L, Morgan WF, Petrini JH. 1998. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**, 477-486.

Caryl AP, Armstrong SJ, Jones GH, Franklin FC. 2000. A homologue of the yeast HOP1 gene is inactivated in the Arabidopsis meiotic mutant asy1. *Chromosoma* **109**, 62-71.

Cervantes MD, Farah JA, Smith GR. 2000. Meiotic DNA breaks associated with recombination in S. pombe. *Mol Cell* **5**, 883-888.

Cheeseman IM, Brew C, Wolyniak M, Desai A, Anderson S, Muster N, Yates JR, Huffaker TC, Drubin DG, Barnes G. 2001. Implication of a novel multiprotein Dam1p complex in outer kinetochore function. *J Cell Biol* **155**, 1137-1145.

Chelysheva L, Diallo S, Vezon D, Gendrot G, Vrielynck N, Belcram K, Rocques N, Marquez-Lema A, Bhatt AM, Horlow C, Mercier R, Mezard C, Grelon M. 2005. AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. *J Cell Sci* **118**, 4621-4632.

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

Chen SY, Tsubouchi T, Rockmill B, Sandler JS, Richards DR, Vader G, Hochwagen A, Roeder GS, Fung JC. 2008. Global analysis of the meiotic crossover landscape. *Dev Cell* **15**, 401-415.

Chi YH, Cheng LI, Myers T, Ward JM, Williams E, Su Q, Faucette L, Wang JY, Jeang KT. 2009. Requirement for Sun1 in the expression of meiotic reproductive genes and piRNA. *Development* **136**, 965-973.

Chin GM, Villeneuve AM. 2001. C. elegans mre-11 is required for meiotic recombination and DNA repair but is dispensable for the meiotic G(2) DNA damage checkpoint. *Genes Dev* **15**, 522-534.

Ciapponi L, Cenci G, Ducau J, Flores C, Johnson-Schlitz D, Gorski MM, Engels WR, Gatti M. 2004. The Drosophila Mre11/Rad50 complex is required to prevent both telomeric fusion and chromosome breakage. *Curr Biol* **14**, 1360-1366.

Ciapponi L, Cenci G, Gatti M. 2006. The Drosophila Nbs protein functions in multiple pathways for the maintenance of genome stability. *Genetics* **173**, 1447-1454.

Ciftci-Yilmaz S, Mittler R. 2008. The zinc finger network of plants. *Cell Mol Life Sci* **65**, 1150-1160.

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen H, Frazer KA, Huson DH, Scholkopf B, Nordborg M, Ratsch G, Ecker JR, Weigel D. 2007. Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. *Science* **317**, 338-342.

Colaiacovo MP, Stanfield GM, Reddy KC, Reinke V, Kim SK, Villeneuve AM. 2002. A targeted RNAi screen for genes involved in chromosome morphogenesis and nuclear organization in the Caenorhabditis elegans germline. *Genetics* **162**, 113-128.

Cool M, Malone RE. 1992. Molecular and genetic analysis of the yeast early meiotic recombination genes REC102 and REC107/MER2. *Mol Cell Biol* **12**, 1248-1256.

Copenhaver GP, Keith KC, Preuss D. 2000. Tetrad analysis in higher plants. A budding technology. *Plant Physiol* **124**, 7-16.

Corbett KD, Benedetti P, Berger JM. 2007. Holoenzyme assembly and ATP-mediated conformational dynamics of topoisomerase VI. *Nat Struct Mol Biol* **14**, 611-619.

Corbett KD, Berger JM. 2003. Emerging roles for plant topoisomerase VI. *Chem Biol* **10**, 107-111.

Corbett KD, Berger JM. 2004. Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annu Rev Biophys Biomol Struct* **33**, 95-118.

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

Cromie GA, Hyppa RW, Cam HP, Farah JA, Grewal SI, Smith GR. 2007. A discrete class of intergenic DNA dictates meiotic DNA break hotspots in fission yeast. *PLoS Genet* **3**, e141.

Cromie GA, Hyppa RW, Smith GR. 2008. The fission yeast BLM homolog Rqh1 promotes meiotic recombination. *Genetics* **179**, 1157-1167.

Cui X, Li H. 1998. Determination of gene organization in individual haplotypes by analyzing single DNA fragments from single spermatozoa. *Proc Natl Acad Sci U S A* **95**, 10791-10796.

D'Amours D, Jackson SP. 2002. The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol* **3**, 317-327.

Daoudal-Cotterell S, Gallego ME, White CI. 2002. The plant Rad50-Mre11 protein complex. *FEBS Lett* **516**, 164-166.

Davis L, Rozalen AE, Moreno S, Smith GR, Martin-Castellanos C. 2008. Rec25 and Rec27, novel linear-element components, link cohesin to meiotic DNA breakage and recombination. *Curr Biol* **18**, 849-854.

Davis L, Smith GR. 2001. Meiotic recombination and chromosome segregation in Schizosaccharomyces pombe. *Proc Natl Acad Sci U S A* **98**, 8395-8402.

de los Santos T, Hollingsworth NM. 1999. Red1p, a MEK1-dependent phosphoprotein that physically interacts with Hop1p during meiosis in yeast. *J Biol Chem* **274**, 1783-1790.

De Muyt A, Pereira L, Vezon D, Chelysheva L, Gendrot G, Chambon A, Laine-Choinard S, Pelletier G, Mercier R, Nogue F, Grelon M. 2009. A high throughput genetic screen identifies new early meiotic recombination functions in Arabidopsis thaliana. *PLoS Genet* **5**, e1000654.

De Muyt A, Vezon D, Gendrot G, Gallois JL, Stevens R, Grelon M. 2007. AtPRD1 is required for meiotic double strand break formation in Arabidopsis thaliana. *Embo J* **26**, 4126-4137.

Denison C, Rudner AD, Gerber SA, Bakalarski CE, Moazed D, Gygi SP. 2005. A proteomic strategy for gaining insights into protein sumoylation in yeast. *Mol Cell Proteomics* **4**, 246-254.

DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686.

Dernburg AF, McDonald K, Moulder G, Barstead R, Dresser M, Villeneuve AM. 1998. Meiotic recombination in C. elegans initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell* **94**, 387-398.

DeVeaux LC, Smith GR. 1994. Region-specific activators of meiotic recombination in Schizosaccharomyces pombe. *Genes Dev* **8**, 203-210.

Dresser ME, Ewing DJ, Conrad MN, Dominguez AM, Barstead R, Jiang H, Kodadek T. 1997. DMC1 functions in a Saccharomyces cerevisiae meiotic pathway that is largely independent of the RAD51 pathway. *Genetics* **147**, 533-544.

Drouaud J, Camilleri C, Bourguignon PY, Canaguier A, Berard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B, Quesneville H, Mezard C. 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.

Edlinger B, Schlogelhofer P. 2011. Have a break: determinants of meiotic DNA double strand break (DSB) formation and processing in plants. *J Exp Bot*.

Ellermeier C, Smith GR. 2005. Cohesins are required for meiotic DNA breakage and recombination in Schizosaccharomyces pombe. *Proc Natl Acad Sci U S A* **102**, 10952-10957.

Engebrecht JA, Voelkel-Meiman K, Roeder GS. 1991. Meiosis-specific RNA splicing in yeast. *Cell* **66**, 1257-1268.

Englbrecht CC, Schoof H, Bohm S. 2004. Conservation, diversification and expansion of C2H2 zinc finger proteins in the Arabidopsis thaliana genome. *BMC Genomics* **5**, 39.

Evans DH, Li YF, Fox ME, Smith GR. 1997. A WD repeat protein, Rec14, essential for meiotic recombination in Schizosaccharomyces pombe. *Genetics* **146**, 1253-1264.

Fan QQ, Petes TD. 1996. Relationship between nuclease-hypersensitive sites and meiotic recombination hot spot activity at the HIS4 locus of Saccharomyces cerevisiae. *Mol Cell Biol* **16**, 2037-2043.

Fanning E, Klimovich V, Nager AR. 2006. A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* **34**, 4126-4137.

Farah JA, Cromie GA, Smith GR. 2009. Ctp1 and Exonuclease 1, alternative nucleases regulated by the MRN complex, are required for efficient meiotic recombination. *Proc Natl Acad Sci U S A* **106**, 9356-9361.

Francis KE, Lam SY, Harrison BD, Bey AL, Berchowitz LE, Copenhaver GP. 2007. Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis. *Proc Natl Acad Sci U S A* **104**, 3913-3918.

Fukuda T, Kugou K, Sasanuma H, Shibata T, Ohta K. 2008. Targeted induction of meiotic double-strand breaks reveals chromosomal domain-dependent regulation of Spo11 and interactions among potential sites of meiotic recombination. *Nucleic Acids Res* **36**, 984-997.

Gallego ME, Jeanneau M, Granier F, Bouchez D, Bechtold N, White CI. 2001. Disruption of the Arabidopsis RAD50 gene leads to plant sterility and MMS sensitivity. *Plant J* **25**, 31-41.

Gardiner JM, Bullard SA, Chrome C, Malone RE. 1997. Molecular and genetic analysis of REC103, an early meiotic recombination gene in yeast. *Genetics* **146**, 1265-1274.

Gasior SL, Wong AK, Kora Y, Shinohara A, Bishop DK. 1998. Rad52 associates with RPA and functions with rad55 and rad57 to assemble meiotic recombination complexes. *Genes Dev* **12**, 2208-2221.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **97**, 11383-11390.

Gloeckner CJ, Boldt K, Schumacher A, Roepman R, Ueffing M. 2007. A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *Proteomics* **7**, 4228-4234.

Gloeckner CJ, Boldt K, Schumacher A, Ueffing M. 2009a. Tandem affinity purification of protein complexes from mammalian cells by the Strep/FLAG (SF)-TAP tag. *Methods Mol Biol* **564**, 359-372.

Gloeckner CJ, Boldt K, Ueffing M. 2009b. Strep/FLAG tandem affinity purification (SF-TAP) to study protein interactions. *Curr Protoc Protein Sci* **Chapter 19**, Unit19 20.

Glynn EF, Megee PC, Yu HG, Mistrot C, Unal E, Koshland DE, DeRisi JL, Gerton JL. 2004. Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. *PLoS Biol* **2**, E259.

Goodyer W, Kaitna S, Couteau F, Ward JD, Boulton SJ, Zetka M. 2008. HTP-3 links DSB formation with homolog pairing and crossing over during C. elegans meiosis. *Dev Cell* **14**, 263-274.

Graumann J, Dunipace LA, Seol JH, McDonald WH, Yates JR, 3rd, Wold BJ, Deshaies RJ. 2004. Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics* **3**, 226-237.

Gregan J, Rabitsch PK, Sakem B, Csutak O, Latypov V, Lehmann E, Kohli J, Nasmyth K. 2005. Novel genes required for meiotic chromosome segregation are identified by a high-throughput knockout screen in fission yeast. *Curr Biol* **15**, 1663-1669.

Grelon M, Vezon D, Gendrot G, Pelletier G. 2001. AtSPO11-1 is necessary for efficient meiotic recombination in plants. *Embo J* **20**, 589-600.

Guacci V, Koshland D, Strunnikov A. 1997. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell* **91**, 47-57.

Gutierrez RA, MacIntosh GC, Green PJ. 1999. Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control. *Trends Plant Sci* **4**, 429-438.

Hannich JT, Lewis A, Kroetz MB, Li SJ, Heide H, Emili A, Hochstrasser M. 2005. Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. *J Biol Chem* **280**, 4102-4110.

Hartung F, Angelis KJ, Meister A, Schubert I, Melzer M, Puchta H. 2002. An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Curr Biol* **12**, 1787-1791.

Hartung F, Puchta H. 2000. Molecular characterisation of two paralogous SPO11 homologues in Arabidopsis thaliana. *Nucleic Acids Res* **28**, 1548-1554.

Hartung F, Puchta H. 2001. Molecular characterization of homologues of both subunits A (SPO11) and B of the archaebacterial topoisomerase 6 in plants. *Gene* **271**, 81-86.

Hartung F, Wurz-Wildersinn R, Fuchs J, Schubert I, Suer S, Puchta H. 2007. 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.

Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438**, 374-378.

Henderson KA, Kee K, Maleki S, Santini PA, Keeney S. 2006. Cyclin-dependent kinase directly regulates initiation of meiotic recombination. *Cell* **125**, 1321-1332.

Higgins JD, Armstrong SJ, Franklin FC, Jones GH. 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.

Hollingsworth NM, Ponte L. 1997. Genetic interactions between HOP1, RED1 and MEK1 suggest that MEK1 regulates assembly of axial element components during meiosis in the yeast Saccharomyces cerevisiae. *Genetics* **147**, 33-42.

Hunter N. 2006. Homologous Recombination. *Topics in Current Genetics.*

Hunter N, Borner GV, Lichten M, Kleckner N. 2001. Gamma-H2AX illuminates meiosis. *Nat Genet* **27**, 236-238.

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

Hyppa RW, Cromie GA, Smith GR. 2008. Indistinguishable Landscapes of Meiotic DNA Breaks in rad50 and rad50S Strains of Fission Yeast Revealed by a Novel rad50 Recombination Intermediate. *PLoS Genet* **4**, e1000267.

Inze D, De Veylder L. 2006. Cell cycle regulation in plant development. *Annu Rev Genet* **40**, 77-105.

Ivanov EL, Korolev VG, Fabre F. 1992. XRS2, a DNA repair gene of Saccharomyces cerevisiae, is needed for meiotic recombination. *Genetics* **132**, 651-664.

Janisiw MP. 2009. Targeted Meiotic Recombination in Arbidopsi thaliana. *Diploma Thesis.*

Jeffreys AJ, Holloway JK, Kauppi L, May CA, Neumann R, Slingsby MT, Webb AJ. 2004. Meiotic recombination hot spots and human DNA diversity. *Philos Trans R Soc Lond B Biol Sci* **359**, 141-152.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502.

Johnston M. 1987. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. *Microbiol Rev* **51**, 458-476.

Jolivet S, Vezon D, Froger N, Mercier R. 2006. Non conservation of the meiotic function of the Ski8/Rec103 homolog in Arabidopsis. *Genes Cells* **11**, 615-622.

Kaufmann K, Muino JM, Osteras M, Farinelli L, Krajewski P, Angenent GC. 2010. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat Protoc* **5**, 457-472.

Kauppi L, Jasin M, Keeney S. 2007. Meiotic crossover hotspots contained in haplotype block boundaries of the mouse genome. *Proc Natl Acad Sci U S A* **104**, 13396-13401.

Kee K, Protacio RU, Arora C, Keeney S. 2004. Spatial organization and dynamics of the association of Rec102 and Rec104 with meiotic chromosomes. *Embo J* **23**, 1815-1824.

Keeney S. 2001. Mechanism and control of meiotic recombination initiation. *Curr Top Dev Biol* **52**, 1-53.

Keeney S. 2007. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome Dyn Stab* **2**, 81-123.

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

Keeney S, Kleckner N. 1996. Communication between homologous chromosomes: genetic alterations at a nuclease-hypersensitive site can alter mitotic chromatin structure at that site both in cis and in trans. *Genes Cells* **1**, 475-489.

Kerzendorfer C, Vignard J, Pedrosa-Harand A, Siwiec T, Akimcheva S, Jolivet S, Sablowski R, Armstrong S, Schweizer D, Mercier R, Schlogelhofer P. 2006. The Arabidopsis thaliana MND1 homologue plays a key role in meiotic homologous pairing, synapsis and recombination. *J Cell Sci* **119**, 2486-2496.

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. *Nat Genet* **39**, 1151-1155.

Klein F, Mahr P, Galova M, Buonomo SB, Michaelis C, Nairz K, Nasmyth K. 1999. A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell* **98**, 91-103.

Krogh BO, Symington LS. 2004. Recombination proteins in yeast. *Annu Rev Genet* **38**, 233-271.

Krsicka O. 2007. Generation of molecular tools for the analysis of meiotic recombination in Arabidopsis thaliana. *Diploma Thesis*.

Kugou K, Fukuda T, Yamada S, Ito M, Sasanuma H, Mori S, Katou Y, Itoh T, Matsumoto K, Shibata T, Shirahige K, Ohta K. 2009. Rec8 guides canonical Spo11 distribution along yeast meiotic chromosomes. *Mol Biol Cell* **20**, 3064-3076.

Kumar R, Bourbon HM, de Massy B. 2010. Functional conservation of Mei4 for meiotic DNA double-strand break formation from yeasts to mice. *Genes Dev* **24**, 1266-1280.

Kupiec M, Simchen G. 1984. Cloning and mapping of the RAD50 gene of Saccharomyces cerevisiae. *Mol Gen Genet* **193**, 525-531.

Latypov V, Rothenberg M, Lorenz A, Octobre G, Csutak O, Lehmann E, Loidl J, Kohli J. 2010. Roles of Hop1 and Mek1 in meiotic chromosome pairing and recombination partner choice in Schizosaccharomyces pombe. *Mol Cell Biol* **30**, 1570-1581.

Li HH, Gyllensten UB, Cui XF, Saiki RK, Erlich HA, Arnheim N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335**, 414-417.

Li J, Hooker GW, Roeder GS. 2006. Saccharomyces cerevisiae Mer2, Mei4 and Rec114 form a complex required for meiotic double-strand break formation. *Genetics* **173**, 1969-1981.

Li W, Chen C, Markmann-Mulisch U, Timofejeva L, Schmelzer E, Ma H, Reiss B. 2004. The Arabidopsis AtRAD51 gene is dispensable for vegetative development but required for meiosis. *Proc Natl Acad Sci U S A* **101**, 10596-10601.

Libby BJ, De La Fuente R, O'Brien MJ, Wigglesworth K, Cobb J, Inselman A, Eaker S, Handel MA, Eppig JJ, Schimenti JC. 2002. The mouse meiotic mutation mei1 disrupts chromosome synapsis with sexually dimorphic consequences for meiotic progression. *Dev Biol* **242**, 174-187.

Lichten M, Goldman AS. 1995. Meiotic recombination hotspots. *Annu Rev Genet* **29**, 423-444.

Lin Y, Smith GR. 1994. Transient, meiosis-induced expression of the rec6 and rec12 genes of Schizosaccharomyces pombe. *Genetics* **136**, 769-779.

Lin Y, Smith GR. 1995a. An intron-containing meiosis-induced recombination gene, rec15, of Schizosaccharomyces pombe. *Mol Microbiol* **17**, 439-448.

Lin Y, Smith GR. 1995b. Molecular cloning of the meiosis-induced rec10 gene of Schizosaccharomyces pombe. *Curr Genet* **27**, 440-446.

Liu H, Jang JK, Kato N, McKim KS. 2002. mei-P22 encodes a chromosome-associated protein required for the initiation of meiotic recombination in Drosophila melanogaster. *Genetics* **162**, 245-258.

Liu J, Wu TC, Lichten M. 1995a. The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. *EMBO J* **14**, 4599-4608.

Liu YG, Mitsukawa N, Oosumi T, Whittier RF. 1995b. Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J* **8**, 457-463.

Longhese MP, Bonetti D, Guerini I, Manfrini N, Clerici M. 2009. DNA double-strand breaks in meiosis: Checking their formation, processing and repair. *DNA Repair (Amst)*.

Longhese MP, Bonetti D, Manfrini N, Clerici M. 2010. Mechanisms and regulation of DNA end resection. *EMBO J* **29**, 2864-2874.

Lorenz A, Estreicher A, Kohli J, Loidl J. 2006. Meiotic recombination proteins localize to linear elements in Schizosaccharomyces pombe. *Chromosoma* **115**, 330-340.

Lorenz A, Wells JL, Pryce DW, Novatchkova M, Eisenhaber F, McFarlane RJ, Loidl J. 2004. S. pombe meiotic linear elements contain proteins related to synaptonemal complex components. *J Cell Sci* **117**, 3343-3351.

Ludin K, Mata J, Watt S, Lehmann E, Bahler J, Kohli J. 2008. Sites of strong Rec12/Spo11 binding in the fission yeast genome are associated with meiotic recombination and with centromeres. *Chromosoma* **117**, 431-444.

Lukaszewicz A, Howard-Till RA, Novatchkova M, Mochizuki K, Loidl J. 2010. MRE11 and COM1/SAE2 are required for double-strand break repair and efficient chromosome pairing during meiosis of the protist Tetrahymena. *Chromosoma* **119**, 505-518.

Luo G, Yao MS, Bender CF, Mills M, Bladl AR, Bradley A, Petrini JH. 1999. Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic development, and sensitivity to ionizing radiation. *Proc Natl Acad Sci U S A* **96**, 7376-7381.

Malik SB, Ramesh MA, Hulstrand AM, Logsdon JM, Jr. 2007. Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Mol Biol Evol* **24**, 2827-2841.

Malone RE, Bullard S, Hermiston M, Rieger R, Cool M, Galbraith A. 1991. Isolation of mutants defective in early steps of meiotic recombination in the yeast Saccharomyces cerevisiae. *Genetics* **128**, 79-88.

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

Manfrini N, Guerini I, Citterio A, Lucchini G, Longhese MP. 2010. Processing of meiotic DNA double strand breaks requires cyclin-dependent kinase and multiple nucleases. *J Biol Chem* **285**, 11628-11637.

Mardis ER. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9**, 387-402.

Marston AL, Amon A. 2004. Meiosis: cell-cycle controls shuffle and deal. *Nat Rev Mol Cell Biol* **5**, 983-997.

Martin-Castellanos C, Blanco M, Rozalen AE, Perez-Hidalgo L, Garcia AI, Conde F, Mata J, Ellermeier C, Davis L, San-Segundo P, Smith GR, Moreno S. 2005. A large-scale screen in S. pombe identifies seven novel genes required for critical meiotic events. *Curr Biol* **15**, 2056-2062.

Matos J, Lipp JJ, Bogdanova A, Guillot S, Okaz E, Junqueira M, Shevchenko A, Zachariae W. 2008. Dbf4-dependent CDC7 kinase links DNA replication to the segregation of homologous chromosomes in meiosis I. *Cell* **135**, 662-678.

McKee AH, Kleckner N. 1997. Mutations in Saccharomyces cerevisiae that block meiotic prophase chromosome metabolism and confer cell cycle arrest at pachytene identify two new meiosis-specific genes SAE1 and SAE3. *Genetics* **146**, 817-834.

McKim KS, Green-Marroquin BL, Sekelsky JJ, Chin G, Steinberg C, Khodosh R, Hawley RS. 1998. Meiotic synapsis in the absence of recombination. *Science* **279**, 876-878.

McKim KS, Hayashi-Hagihara A. 1998. mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes Dev* **12**, 2932-2942.

Mendoza MA, Panizza S, Klein F. 2009. Analysis of protein-DNA interactions during meiosis by quantitative chromatin immunoprecipitation (qChIP). *Methods Mol Biol* **557**, 267-283.

Menees TM, Roeder GS. 1989. MEI4, a yeast gene required for meiotic recombination. *Genetics* **123**, 675-682.

Mercier R, Armstrong SJ, Horlow C, Jackson NP, Makaroff CA, Vezon D, Pelletier G, Jones GH, Franklin FC. 2003. The meiotic protein SWI1 is required for axial element formation and recombination initiation in Arabidopsis. *Development* **130**, 3309-3318.

Mercier R, Jolivet S, Vezon D, Huppe E, Chelysheva L, Giovanni M, Nogue F, Doutriaux MP, Horlow C, Grelon M, Mezard C. 2005. Two meiotic crossover classes cohabit in Arabidopsis: one is dependent on MER3,whereas the other one is not. *Curr Biol* **15**, 692-701.

Mercier R, Vezon D, Bullier E, Motamayor JC, Sellier A, Lefevre F, Pelletier G, Horlow C. 2001. SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. *Genes Dev* **15**, 1859-1871.

Merker JD, Dominska M, Greenwell PW, Rinella E, Bouck DC, Shibata Y, Strahl BD, Mieczkowski P, Petes TD. 2008. The histone methylase Set2p and the histone deacetylase Rpd3p repress meiotic recombination at the HIS4 meiotic recombination hotspot in Saccharomyces cerevisiae. *DNA Repair (Amst)* **7**, 1298-1308.

Merritt C, Seydoux G. 2010. The Puf RNA-binding proteins FBF-1 and FBF-2 inhibit the expression of synaptonemal complex proteins in germline stem cells. *Development* **137**, 1787-1798.

Mets DG, Meyer BJ. 2009. Condensins regulate meiotic DNA break distribution, thus crossover frequency, by controlling chromosome structure. *Cell* **139**, 73-86.

Michaelis C, Ciosk R, Nasmyth K. 1997. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35-45.

Mieczkowski PA, Dominska M, Buck MJ, Lieb JD, Petes TD. 2007. Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **104**, 3955-3960.

Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373-375.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560.

Milman N, Higuchi E, Smith GR. 2009. Meiotic DNA double-strand break repair requires two nucleases, MRN and Ctp1, to produce a single size class of Rec12 (Spo11)-oligonucleotide complexes. *Mol Cell Biol* **29**, 5998-6005.

Mimitou EP, Symington LS. 2008. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770-774.

Mimitou EP, Symington LS. 2009a. DNA end resection: many nucleases make light work. *DNA Repair (Amst)* **8**, 983-995.

Mimitou EP, Symington LS. 2009b. Nucleases and helicases take center stage in homologous recombination. *Trends Biochem Sci* **34**, 264-272.

Moens PB, Pearlman RE. 1988. Chromatin organization at meiosis. *Bioessays* **9**, 151-153.

Molinier J, Ries G, Zipfel C, Hohn B. 2006. Transgeneration memory of stress in plants. *Nature* **442**, 1046-1049.

Molnar M, Parisi S, Kakihara Y, Nojima H, Yamamoto A, Hiraoka Y, Bozsik A, Sipiczki M, Kohli J. 2001. Characterization of rec7, an early meiotic recombination gene in Schizosaccharomyces pombe. *Genetics* **157**, 519-532.

Munroe RJ, Bergstrom RA, Zheng QY, Libby B, Smith R, John SW, Schimenti KJ, Browning VL, Schimenti JC. 2000. Mouse mutants from chemically mutagenized embryonic stem cells. *Nat Genet* **24**, 318-321.

Murakami H, Keeney S. 2008. Regulating the formation of DNA double-strand breaks in meiosis. *Genes Dev* **22**, 286-292.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321-324.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, Macfie TS, McVean G, Donnelly P. 2009. Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. *Science* **327**, 876-879.

Myers S, Freeman C, Auton A, Donnelly P, McVean G. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet* **40**, 1124-1129.

Nairz K, Klein F. 1997. mre11S--a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. *Genes Dev* **11**, 2272-2290.

Neale MJ, Pan J, Keeney S. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053-1057.

Ng DW, Wang T, Chandrasekharan MB, Aramayo R, Kertbundit S, Hall TC. 2007. Plant SET domain-containing proteins: structure, function and regulation. *Biochim Biophys Acta* **1769**, 316-329.

Nichols MD, DeAngelis K, Keck JL, Berger JM. 1999. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *Embo J* **18**, 6177-6188.

Nishant KT, Rao MR. 2006. Molecular features of meiotic recombination hot spots. *Bioessays* **28**, 45-56.

Nislow C, Ray E, Pillus L. 1997. SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol Biol Cell* **8**, 2421-2436.

Nonomura K, Nakano M, Fukuda T, Eiguchi M, Miyao A, Hirochika H, Kurata N. 2004. The novel gene HOMOLOGOUS PAIRING ABERRATION IN RICE MEIOSIS1 of rice encodes a putative coiled-coil protein required for homologous chromosome pairing in meiosis. *Plant Cell* **16**, 1008-1020.

Ohta K, Shibata T, Nicolas A. 1994. Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J* **13**, 5754-5763.

Orr-Weaver TL. 1999. The ties that bind: localization of the sister-chromatid cohesin complex on yeast chromosomes. *Cell* **99**, 1-4.

Osakabe K, Abe K, Yamanouchi H, Takyuu T, Yoshioka T, Ito Y, Kato T, Tabata S, Kurei S, Yoshioka Y, Machida Y, Seki M, Kobayashi M, Shinozaki K, Ichikawa H, Toki S. 2005. Arabidopsis Rad51B is important for double-strand DNA breaks repair in somatic cells. *Plant Mol Biol* **57**, 819-833.

Osakabe K, Yoshioka T, Ichikawa H, Toki S. 2002. Molecular cloning and characterization of RAD51-like genes from Arabidopsis thaliana. *Plant Mol Biol* **50**, 71-81.

Page SL, Hawley RS. 2003. Chromosome choreography: the meiotic ballet. *Science* **301**, 785-789.

Page SL, Hawley RS. 2004. The genetics and molecular biology of the synaptonemal complex. *Annu Rev Cell Dev Biol* **20**, 525-558.

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

Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669-680.

Parvanov ED, Petkov PM, Paigen K. 2009. Prdm9 Controls Activation of Mammalian Recombination Hotspots. *Science* **327**, 835.

Pawlowski WP, Golubovskaya IN, Timofejeva L, Meeley RB, Sheridan WF, Cande WZ. 2004. Coordination of meiotic recombination, pairing, and synapsis by PHS1. *Science* **303**, 89-92.

Pecina A, Smith KN, Mezard C, Murakami H, Ohta K, Nicolas A. 2002. Targeted stimulation of meiotic recombination. *Cell* **111**, 173-184.

Perrella G, Consiglio MF, Aiese-Cigliano R, Cremona G, Sanchez-Moran E, Barra L, Errico A, Bressan RA, Franklin FC, Conicella C. 2010. Histone hyperacetylation affects meiotic recombination and chromosome segregation in Arabidopsis. *Plant J* **62**, 796-806.

Petes TD. 2001. Meiotic recombination hot spots and cold spots. *Nat Rev Genet* **2**, 360-369.

Prieler S, Penkner A, Borde V, Klein F. 2005. The control of Spo11's interaction with meiotic recombination hotspots. *Genes Dev* **19**, 255-269.

Prinz S, Amon A, Klein F. 1997. Isolation of COM1, a new gene required to complete meiotic double-strand break-induced recombination in Saccharomyces cerevisiae. *Genetics* **146**, 781-795.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218-229.

Puizina J, Siroky J, Mokros P, Schweizer D, Riha K. 2004. Mre11 deficiency in Arabidopsis is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. *Plant Cell* **16**, 1968-1978.

Ray A, Langer M. 2002. Homologous recombination: ends as the means. *Trends Plant Sci* **7**, 435-440.

Reddy KC, Villeneuve AM. 2004. C. elegans HIM-17 links chromatin modification and competence for initiation of meiotic recombination. *Cell* **118**, 439-452.

Rhee SY, Somerville CR. 1998. Tetrad pollen formation in quartet mutants of Arabidopsis thaliana is associated with persistence of pectic polysaccharides of the pollen mother cell wall. *Plant J* **15**, 79-88.

Riehs N. 2006. Yeast lab book.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**, 1030-1032.

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**, 651-657.

Robine N, Uematsu N, Amiot F, Gidrol X, Barillot E, Nicolas A, Borde V. 2007. Genome-wide redistribution of meiotic double-strand breaks in Saccharomyces cerevisiae. *Mol Cell Biol* **27**, 1868-1880.

Rockmill B, Engebrecht JA, Scherthan H, Loidl J, Roeder GS. 1995. The yeast MER2 gene is required for chromosome synapsis and the initiation of meiotic recombination. *Genetics* **141**, 49-59.

Rockmill B, Roeder GS. 1988. RED1: a yeast gene required for the segregation of chromosomes during the reductional division of meiosis. *Proc Natl Acad Sci U S A* **85**, 6057-6061.

Romanienko PJ, Camerini-Otero RD. 1999. Cloning, characterization, and localization of mouse and human SPO11. *Genomics* **61**, 156-169.

Ronceret A, Doutriaux MP, Golubovskaya IN, Pawlowski WP. 2009. PHS1 regulates meiotic recombination and homologous chromosome pairing by controlling the transport of RAD50 to the nucleus. *Proc Natl Acad Sci U S A* **106**, 20121-20126.

Rothenberg M, Kohli J, Ludin K. 2009. Ctp1 and the MRN-complex are required for endonucleolytic Rec12 removal with release of a single class of oligonucleotides in fission yeast. *PLoS Genet* **5**, e1000722.

Salem L, Walter N, Malone R. 1999. Suppressor analysis of the Saccharomyces cerevisiae gene REC104 reveals a genetic interaction with REC102. *Genetics* **151**, 1261-1272.

Sambrook J, Russell D. 2001. Molecular cloning: A laboratory manual. *Cold Spring Harbor Laboratory Press* **3rd edition**.

San Filippo J, Sung P, Klein H. 2008. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* **77**, 229-257.

Sanchez-Moran E, Jones GH, Franklin FC, Santos JL. 2004. A puromycin-sensitive aminopeptidase is essential for meiosis in Arabidopsis thaliana. *Plant Cell* **16**, 2895-2909.

Sanchez-Moran E, Osman K, Higgins JD, Pradillo M, Cunado N, Jones GH, Franklin FC. 2008. ASY1 coordinates early events in the plant meiotic recombination pathway. *Cytogenet Genome Res* **120**, 302-312.

Sanchez-Moran E, Santos JL, Jones GH, Franklin FC. 2007. ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. *Genes Dev* **21**, 2220-2233.

Sasanuma H, Hirota K, Fukuda T, Kakusho N, Kugou K, Kawasaki Y, Shibata T, Masai H, Ohta K. 2008. Cdc7-dependent phosphorylation of Mer2 facilitates initiation of yeast meiotic recombination. *Genes Dev* **22**, 398-410.

Sasanuma H, Murakami H, Fukuda T, Shibata T, Nicolas A, Ohta K. 2007. Meiotic association between Spo11 regulated by Rec102, Rec104 and Rec114. *Nucleic Acids Res* **35**, 1119-1133.

Sato S, Hotta Y, Tabata S. 1995a. Structural analysis of a recA-like gene in the genome of Arabidopsis thaliana. *DNA Res* **2**, 89-93.

Sato S, Kobayashi T, Hotta Y, Tabata S. 1995b. Characterization of a mouse recA-like gene specifically expressed in testis. *DNA Res* **2**, 147-150.

Sato S, Seki N, Hotta Y, Tabata S. 1995c. Expression profiles of a human gene identified as a structural homologue of meiosis-specific recA-like genes. *DNA Res* **2**, 183-186.

Schmidt WM, Mueller MW. 1996. Controlled ribonucleotide tailing of cDNA ends (CRTC) by terminal deoxynucleotidyl transferase: a new approach in PCR-mediated analysis of mRNA sequences. *Nucleic Acids Res* **24**, 1789-1791.

Schommer C, Beven A, Lawrenson T, Shaw P, Sablowski R. 2003. AHP2 is required for bivalent formation and for segregation of homologous chromosomes in Arabidopsis meiosis. *Plant J* **36**, 1-11.

Schwacha A, Kleckner N. 1997. Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. *Cell* **90**, 1123-1135.

Sehorn MG, Sung P. 2004. Meiotic recombination: an affair of two recombinases. *Cell Cycle* **3**, 1375-1377.

Sekelsky JJ, McKim KS, Messina L, French RL, Hurley WD, Arbel T, Chin GM, Deneen B, Force SJ, Hari KL, Jang JK, Laurencon AC, Madden LD, Matthies HJ, Milliken DB, Page SL, Ring AD, Wayson SM, Zimmerman CC, Hawley RS. 1999. Identification of novel Drosophila meiotic genes recovered in a P-element screen. *Genetics* **152**, 529-542.

Sharp PM, Lloyd AT. 1993. Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. *Nucleic Acids Res* **21**, 179-183.

Shinohara A, Ogawa H, Ogawa T. 1992. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. *Cell* **69**, 457-470.

Siaud N, Dray E, Gy I, Gerard E, Takvorian N, Doutriaux MP. 2004. Brca2 is involved in meiosis in Arabidopsis thaliana as suggested by its interaction with Dmc1. *EMBO J* **23**, 1392-1401.

Siddiqi I, Ganesh G, Grossniklaus U, Subbiah V. 2000. The dyad gene is required for progression through female meiosis in Arabidopsis. *Development* **127**, 197-207.

Singer T, Fan Y, Chang HS, Zhu T, Hazen SP, Briggs SP. 2006. A high-resolution map of Arabidopsis recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genet* **2**, e144.

Smith AV, Roeder GS. 1997. The yeast Red1 protein localizes to the cores of meiotic chromosomes. *J Cell Biol* **136**, 957-967.

Smith KN, Nicolas A. 1998. Recombination at work for meiosis. *Curr Opin Genet Dev* **8**, 200-211.

Sollier J, Lin W, Soustelle C, Suhre K, Nicolas A, Geli V, de La Roche Saint-Andre C. 2004. Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. *EMBO J* **23**, 1957-1967.

Spirek M, Estreicher A, Csaszar E, Wells J, McFarlane RJ, Watts FZ, Loidl J. 2010. SUMOylation is required for normal development of linear elements and wild-type meiotic recombination in Schizosaccharomyces pombe. *Chromosoma* **119**, 59-72.

Stacey NJ, Kuromori T, Azumi Y, Roberts G, Breuer C, Wada T, Maxwell A, Roberts K, Sugimoto-Shirasu K. 2006. Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J* **48**, 206-216.

Steiner S, Kohli J, Ludin K. 2010. Functional interactions among members of the meiotic initiation complex in fission yeast. *Curr Genet* **56**, 237-249.

Struhl K. 1998. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* **12**, 599-606.

Tagwerker C, Flick K, Cui M, Guerrero C, Dou Y, Auer B, Baldi P, Huang L, Kaiser P. 2006. A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. *Mol Cell Proteomics* **5**, 737-748.

Tavassoli M, Shayeghi M, Nasim A, Watts FZ. 1995. Cloning and characterisation of the Schizosaccharomyces pombe rad32 gene: a gene required for repair of double strand breaks and recombination. *Nucleic Acids Res* **23**, 383-388.

Tesse S, Storlazzi A, Kleckner N, Gargano S, Zickler D. 2003. Localization and roles of Ski8p protein in Sordaria meiosis and delineation of three mechanistically distinct steps of meiotic homolog juxtaposition. *Proc Natl Acad Sci U S A* **100**, 12865-12870.

Tiemann-Boege I, Calabrese P, Cochran DM, Sokol R, Arnheim N. 2006. High-resolution recombination patterns in a region of human chromosome 21 measured by sperm typing. *PLoS Genet* **2**, e70.

Uanschou C, Siwiec T, Pedrosa-Harand A, Kerzendorfer C, Sanchez-Moran E, Novatchkova M, Akimcheva S, Woglar A, Klein F, Schlogelhofer P. 2007. A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. *Embo J* **26**, 5061-5070.

Ueno M, Nakazaki T, Akamatsu Y, Watanabe K, Tomita K, Lindsay HD, Shinagawa H, Iwasaki H. 2003. Molecular characterization of the Schizosaccharomyces pombe nbs1+ gene involved in DNA repair and telomere maintenance. *Mol Cell Biol* **23**, 6553-6563.

Van Dyck E, Stasiak AZ, Stasiak A, West SC. 1999. Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* **398**, 728-731.

van Heemst D, Heyting C. 2000. Sister chromatid cohesion and recombination in meiosis. *Chromosoma* **109**, 10-26.

Vandepoele K, Raes J, De Veylder L, Rouze P, Rombauts S, Inze D. 2002. Genome-wide analysis of core cell cycle genes in Arabidopsis. *Plant Cell* **14**, 903-916.

Vignard J, Siwiec T, Chelysheva L, Vrielynck N, Gonord F, Armstrong SJ, Schlogelhofer P, Mercier R. 2007. The interplay of RecA-related proteins and the MND1-HOP2 complex during meiosis in Arabidopsis thaliana. *PLoS Genet* **3**, 1894-1906.

Vissinga CS, Yeo TC, Woessner J, Massa HF, Wilson RK, Trask BJ, Concannon P. 1999. Identification, characterization, and mapping of a mouse homolog of the gene mutated in Nijmegen breakage syndrome. *Cytogenet Cell Genet* **87**, 80-84.

Wagner CR, Kuervers L, Baillie DL, Yanowitz JL. 2010. xnd-1 regulates the global recombination landscape in Caenorhabditis elegans. *Nature* **467**, 839-843.

Wan L, Niu H, Futcher B, Zhang C, Shokat KM, Boulton SJ, Hollingsworth NM. 2008. Cdc28-Clb5 (CDK-S) and Cdc7-Dbf4 (DDK) collaborate to initiate meiotic recombination in yeast. *Genes Dev* **22**, 386-397.

Waterworth WM, Altun C, Armstrong SJ, Roberts N, Dean PJ, Young K, Weil CF, Bray CM, West CE. 2007. NBS1 is involved in DNA repair and plays a synergistic role with ATM in mediating meiotic homologous recombination in plants. *Plant J* **52**, 41-52.

Weiner BM, Kleckner N. 1994. Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. *Cell* **77**, 977-991.

Wells JL, Pryce DW, Estreicher A, Loidl J, McFarlane RJ. 2006. Linear element-independent meiotic recombination in Schizosaccharomyces pombe. *Genetics* **174**, 1105-1114.

White MA, Dominska M, Petes TD. 1993. Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **90**, 6621-6625.

Wijeratne AJ, Chen C, Zhang W, Timofejeva L, 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.

Woltering D, Baumgartner B, Bagchi S, Larkin B, Loidl J, de los Santos T, Hollingsworth NM. 2000. Meiotic segregation, synapsis, and recombination checkpoint

functions require physical interaction between the chromosomal proteins Red1p and Hop1p. *Mol Cell Biol* **20**, 6646-6658.

Xiang C, Han P, Lutziger I, Wang K, Oliver DJ. 1999. A mini binary vector series for plant transformation. *Plant Mol Biol* **40**, 711-717.

Xiao Y, Weaver DT. 1997. Conditional gene targeted deletion by Cre recombinase demonstrates the requirement for the double-strand break repair Mre11 protein in murine embryonic stem cells. *Nucleic Acids Res* **25**, 2985-2991.

Yamada T, Mizuno K, 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 J* **23**, 1792-1803.

Yamashita K, Shinohara M, Shinohara A. 2004. Rad6-Bre1-mediated histone H2B ubiquitylation modulates the formation of double-strand breaks during meiosis. *Proc Natl Acad Sci U S A* **101**, 11380-11385.

Yates JR, 3rd. 2004. Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* **33**, 297-316.

Yin Y, Cheong H, Friedrichsen D, Zhao Y, Hu J, Mora-Garcia S, Chory J. 2002. A crucial role for the putative Arabidopsis topoisomerase VI in plant growth and development. *Proc Natl Acad Sci U S A* **99**, 10191-10196.

Young JA, Hyppa RW, Smith GR. 2004. Conserved and nonconserved proteins for meiotic DNA breakage and repair in yeasts. *Genetics* **167**, 593-605.

Zachariae W, Shevchenko A, Andrews PD, Ciosk R, Galova M, Stark MJ, Mann M, Nasmyth K. 1998. Mass spectrometric analysis of the anaphase-promoting complex from yeast: identification of a subunit related to cullins. *Science* **279**, 1216-1219.

Zachariae W, Shin TH, Galova M, Obermaier B, Nasmyth K. 1996. Identification of subunits of the anaphase-promoting complex of Saccharomyces cerevisiae. *Science* **274**, 1201-1204.

Zenvirth D, Arbel T, Sherman A, Goldway M, Klein S, Simchen G. 1992. Multiple sites for double-strand breaks in whole meiotic chromosomes of Saccharomyces cerevisiae. *Embo J* **11**, 3441-3447.

Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. 2009. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol* **10**, R62.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189-1201.

Zhu Z, Chung WH, Shim EY, Lee SE, Ira G. 2008. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134**, 981-994.

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

9 Abbreviations

A

AE - axial element

B

BAC - bacterial artificial chromosome

BD - binding domain

BLAST - basic local alignment search tool

BND - benzoyl naphtyl DEAE

BSA - bovine serum albumine

C

CAPS - cleaved amplified polymorphic sequence

CDK - cyclin dependent kinase

CFP - cyan fluorescent protein

ChIP - chromatin immuno-precipitation

ChIP-Seq - chromatin immuno-precipitation followed by sequencing

CIBIV - Center for Integrative Bioinformatics Vienna

CO - cross over

D

DEAE - diethylaminoethyl

DMP - dimethyl pimelimidate

DNA - deoxyribonucleic acid

DNS - Desoxyribonukleinsäure

DSB - double strand break

dsDNA - double stranded DNA

DTT - dithiothreitol

dUTP - deoxyuridine triphosphate

E

ECL - enhanced chemiluminescence

EDTA - ethylene-diamine-tetraacetic acid

F

FTL - fluorescent-tagged line

G

GMI - Gregor Mendel Institute of Molecular Plant Biology, Vienna

H

H3 - histone 3

H3K4 - histone 3 lysine 4

HA - hemagglutinin

HCl - hydrochloric acid

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPLC - high-performance liquid chromatography

HRP - horse-radish peroxidase

I

IMP - Institute of Molecular Pathology, Vienna

INDEL - insertion / deletion

INRA - French National Institute for Agricultural Research, Versaille

IP - immuno-precipitation

K

Kb - kilobase

kDa - kilodalton

L

LB - luria broth

LB - left border

LC - liquid chromatography

LD - linkage disequilibrium

LDS - lithium dodecyl sulphate

LinE - linear element

M

^{me3} - tri-methylation

Mb - megabase

MCS - multiple cloning site

MES - 2-(n-morpholino) ethanesulfonic acid

MMS - methyl methanesulfonate

mRNA - messenger RNA

MRX - protein complex consisting of Mre11/Rad50/Xrs2

MS - mass spectrometry

MS - Murashige and Skoog

N

NaCl - sodium chloride

NCO - non cross over

NMD - nonsense mediated decay

O

Oligo - oligonucleotide

ORF - open reading frame

O/N - over night

P

PAGE - polyacrylamide gel electrophoresis

PCR - polymerase chain reaction

PMSF - phenylmethylsulfonyl fluoride

PVDF - polyvinylidene fluoride

R

RACE - rapid amplification of cDNA ends

RPA - replication protein A

RT - room temperature

RT - reverse transcriptase

S

SDS - sodium dodecyl sulphate

ssDNA - single stranded DNA

SF - Strep/FLAG

SNP - single nucleotide polymorphism

T

TAIL - thermal asymmetric interlaced

TAP - tandem affinity purification

TE - Tris EDTA

TEMED - tetramethylethylenediamine

TUNEL - terminal deoxynucleotidyl transferase dUTP nick end labelling

U

UAS - upstream activator sequence

UTR - untranslated region

UV - ultraviolet

Y

YFP - yellow fluorescent protein

10 Appendix

10.1 Primer list

Primer name	Sequence 5'-3'	Internal number
spo11-1-2-MG52	GGATCGGGCCTAAAAGCCAACG	295
spo11-1-2-MG96	CTTTGAATGCTGATGGATGCATGTAGTAG	296
Spo11-2_GABI_down	GCTCGTGGAAGATCGTGTGTTC	405
Spo11-2_GABI_up	CCTGCATAGGAAAGTGGAGATTAGGAC	406
GABItest1	CCCATTTGGACGTGAATGTAGACAC	109
GABI LB BE	CGACAGAGGTGTGATGTTAGGCC	1018
spo11-2 gk genomic	GATGAGCACAACTTTTCCTGATCTGCC	1019
5xuas tail A	ATTGCGTTGCGCTCACTGC	555
5xuas tail B	GGTTTGCGTATTGGGCGC	556
5xuas tail C	CGCAGGAAAGAACATGAAGGCC	557
5xuas tail lb1	CCTTACGCTTGGTGGCACAG	566
5xuas tail lb2	TTCGGTACGCTGAAATCACC	567
5xuas tail lb3	CATGTGTTGAGCATATAAGAAACCC	568
AD7	AWGCANGNCWGANATA	554
pNOS dn	GGGTTTCTGGAGTTTAATGAGCT	152
bar_fw	AGCTGATCCTCTAGGGGTCATC	408
gATSPOup	CGCGGATCCGCTGAAGCTGAAGTTGCCACG	66
c-myc_new_dn	ATCCATGGCTCCCGGGTCCGGTTCTGCTGCTAGTGG	216
spo11-1 24xha forw	GATTACGCTGATGGGTGAAGATCTGC	548
spo11-1 24xha rev	GGTCAATGATGACCTGGTGC	549
AtGal4dn	ATGAAGCTCCTGTCCTCCATC	70
SF_forw	AGCTGGAGCCACCCTCAGTTCGAGAAGG	856
pNOS up	CAGAAACCCGCGGCTGAGTGGCT	154
spo11-2 insert rev	GCTGTTCTTGGAAAAGAGGTTGGG	550
spo11-2 18xmyc forw	GACTTGAACGGAGATGGGTAACAC	551
spo11-2 24xha forw	CCAGATTACGCTGATGGGTAACAC	552
spo11-2 gal4 forw	CCGTGGAGACCGACATGC	553
pPILY 15xHA forw	AGAGGACAGCCCAAGCTTCC	447
pPILY 15xHA rev	TAATACTAGTTCAGATATCAGCGTAATCTGGAACGTCATATGGATAGG	448
Spo11-2_Stop_up	TAATCCCGGGTATGTATTTGCCTTGCACGATCTTGGTAGC	387
Spo11-2_Eco91I_dn	TGTGAAGTGGATTGGTCTCCGAGGAG	388
Spo11-2_Stop_dn	ATAAACCCGGGTAACACTTAAAACACCAACATGTAAGGTCTTCCACC	390
Spo11-2_NheI_up	GTTTGTCTCGGGTGGTCTTCTGG	389
Spo11-2_right_forw	ATCTTAGAATTGCTTCGGCTTGACC	436
Spo11-2_right_rev	ACTCACCAGCATCAGAGATACCAC	437
Spo11-2_left_forw	ATCCCACCGATGATCAGAGGATACC	434
Spo11-2_left_rev	AGCCTGCAAAGAACATACCTGAACC	435
6kb seq1	ATACCACTGACTACATTGC	456
6kb seq2	TAGATTGACCTTGGACAC	457
6kb seq3	ACTACGAGATTCGACAGTG	458
6kb seq4	AGTTGAATCGGAAAGTTC	459
6kb seq5	TGCGAAAACGAAATGAAGG	460
6kb seq6	TGATCGGTTTATATGAAACATG	461
6kb seq7	TGTTAACCCTAGAAGAAGG	462
6kb seq8	TGGTTTATCCAAACCAATCC	463

6kb seq9	ACATACCACACATGCAACG	464
2kb seq1	ACTCTTCTCTTCCGCACC	465
2kb seq1	TGTTGGAGCAGGTGTTGG	466
6kb seqA	GAGGACAATAAGCTCTTAGC	471
6kb seqB	ACAACTATTGCTTCTCTTCC	472
6kb seqC	TGCATGCATGAGTTTGC	473
6kb seqD	TCTATAAACAACGACAAGGATG	474
6kb seqE	TGGGTTGAAGTACATTTGAC	475
6kb seqF	AGGAGAGATGGGAGGATTAG	476
6kb seqG	AATTATACCTCGTGAAGTCG	477
5xuas T-DNA confirm	TCGCCAACATCTTCTTCTGGAGGC	610
5xUAS#1 cnfrm	TGGGGACAGAAGAAACCAAACGACGATAG	868
5xuas#6gDNA	AGGTGGCGATTGTTATTGGGTTTGACC	611
5xUAS#19B cnfrm	ACGGTTTAAGGATGGTGTTTGCAACG	870
5xuas#20gDNA	TGGCTAGATTGACGGCGACTCAAC	612
5xuas#21gDNA	TGGTCTCATTCCTACCTCCCCTG	613
M13 forward	GTAAAACGACGGCCAG	391
M13 reverse	CAGGAAACAGCTATGAC	392
T7	GTAATACGACTCACTATAGGGC	97

All primers for the ChIP-Seq project are listed in the materials and methods section.

10.2 Vectors and Plasmids

All plasmid maps, except for pCR®2.1, were generated with Vector NTI (Invitrogen)

Map and Features of pCR®2.1

Map of pCR®2.1 The map of the linearized vector, pCR®2.1, is shown below. The arrow indicates the start of transcription for the T7 RNA polymerase. **The complete sequence of pCR®2.1 is available from our Web site (www.invitrogen.com) or by contacting Technical Service (page 18).**
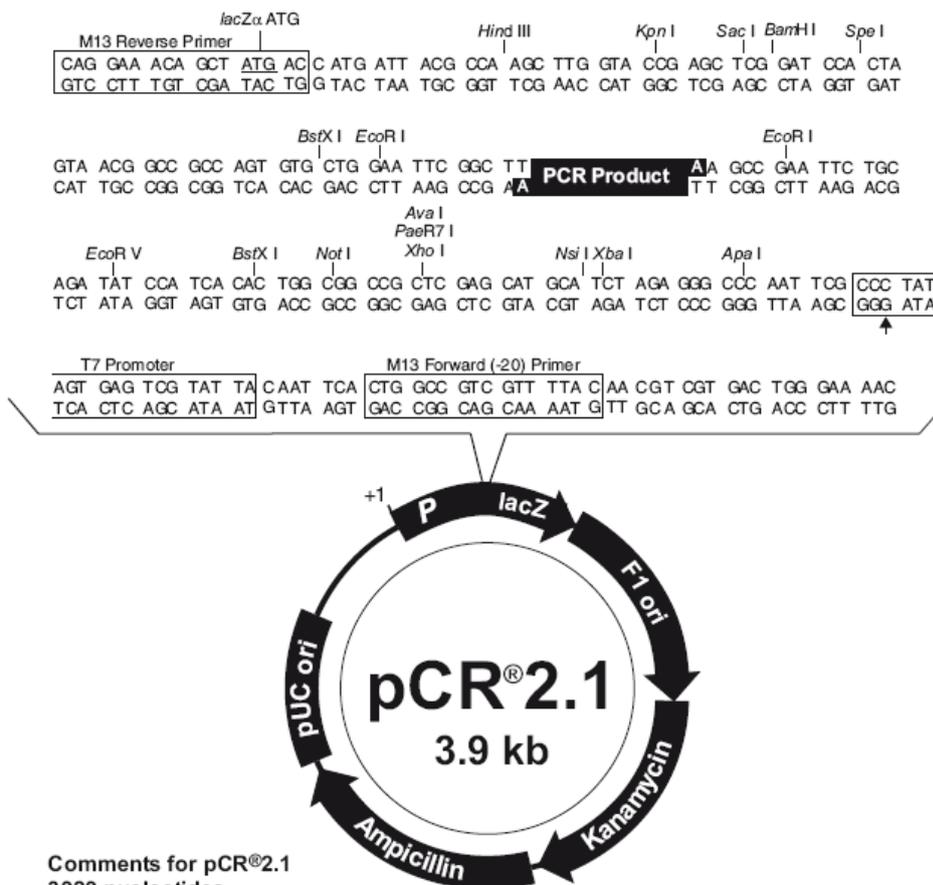
Comments for pCR®2.1
3929 nucleotides

LacZα gene: bases 1-545
M13 Reverse priming site: bases 205-221
T7 promoter: bases 362-381
M13 (-20) Forward priming site: bases 389-404
f1 origin: bases 546-983
Kanamycin resistance ORF: bases 1317-2111
Ampicillin resistance ORF: bases 2129-2989
pUC origin: bases 3134-3807

pCB302-AtSPO11-1/wt-like
9262 bp

LB-1
LB repeat
Pnos-2
similar to transposable element Tn5 [Split]
Pnos-3
Pnos-4
Pnos-1
bar-3
bar
bar-1
bar-2
:frag 3373 to 3435 of pCB301
frag 3373 to 3435 of pCB301 [Split]
3'NOS
:frag 2323 to 2546 of pD991 right border
frag 86 to 105 of nos terminator
Tnos-1
Tnos-3
Tnos-2
gATSPOdn
SpoSeq5UTR2up
SpoSeq5UTRdn
SpoSeq5UTRup
AtSPOGAL4dn
SpoSeq1dn
SpoSeq1up
SPO11-1
SpoSeq2dn new
RK2 oriV
TrfA
T-DNA right border
RB repeat
RB-1
3´UTR

pCB302-AtSPO11-1/18xmyc
9997 bp

LB-1
LB repeat
Pnos-2
similar to transposable element Tn5 [Split]
Pnos-3
Pnos-4
Pnos-1
bar-3
bar
bar-1
bar-2
:frag 3373 to 3435 of pCB301
frag 3373 to 3435 of pCB301 [Split]
3'NOS
:frag 2323 to 2546 of pD991 right border
frag 86 to 105 of nos terminator
Tnos-1
Tnos-3
Tnos-2
gATSPOdn
SpoSeq5UTR2up
SpoSeq5UTRdn
SpoSeq5UTRup
AtSPOGAL4dn
SpoSeq1dn
SpoSeq1up
SPO11-1
SpoSeq2dn new
RK2 oriV
TrfA
T-DNA right border
RB repeat
RB-1
3´UTR
18xc-myc

pCB302-AtSPO11-1/SF
9412 bp

Labels (clockwise from top):
- LB-1
- LB repeat
- Pnos-2
- similar to transposable element Tn5 [Split]
- Pnos-3
- Pnos-4
- Pnos-1
- bar-3
- bar
- bar-1
- bar-2
- :frag 3373 to 3435 of pCB301
- frag 3373 to 3435 of pCB301 [Split]
- 3'NOS
- :frag 2323 to 2546 of pD991 right border
- frag 86 to 105 of nos terminator
- Tnos-1
- Tnos-3
- Tnos-2
- gATSPOdn
- SpoSeq5UTR2up
- SpoSeq5UTRdn
- SpoSeq5UTRup
- AtSPOGAL4dn
- SpoSeq1dn
- SpoSeq1up
- SPO11-1
- SpoSeq2dn new
- RK2 oriV
- TrfA
- T-DNA right border
- RB repeat
- RB-1
- 3´UTR
- SF tag

pGreen II 0179-5xUAS short
5220 bp

Labels:
- left border
- potential left border end after sequencing
- 5xuas tail lb3
- 5xuas tail lb2
- 5xuas tail lb1
- right border
- 5xuas tail C
- 5xuas tail B
- 5xuas tail A
- 5xuas length rev
- 5xUAS-short
- 5xuas length forw
- T-DNA

152

pBIB-Hyg

AtSPO11-2 full genomic/wt-like for pBIB-Hyg

9458 bp

AtSPO11-2 full genomic/24xHA for pBIB-Hyg

10238 bp

AtSPO11-2 full genomic/18xc-myc for pBIB-Hyg

10193 bp

AtSPO11-2 full genomic/Gal4-2 BD for pBIB-Hyg

9905 bp

10.3 BLAST results for 5xUAS mapping

5xUAS#1

>☐ ref|NC_003071.7| 🟩E🟩D Arabidopsis thaliana chromosome 2, complete sequence
Length=19698289

```
Features flanking this part of subject sequence:
   3958 bp at 5' side: MLO12 (MILDEW RESISTANCE LOCUS O 12); calmodulin binding
   2427 bp at 3' side: nodulin family protein

Score =  196 bits (106),  Expect = 5e-50
Identities = 106/106 (100%), Gaps = 0/106 (0%)
Strand=Plus/Plus

Query  1         ATAATTAGTAGATGAATATCAGTGTATATGATGTTTTTAAAATTTAAGTACTTAAGTTAA  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16363755  ATAATTAGTAGATGAATATCAGTGTATATGATGTTTTTAAAATTTAAGTACTTAAGTTAA  16363814

Query  61        CATTTGACCAATTCTTAGGCATTAAATTGATATAAAATTTTGGTAT  106
                 ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  16363815  CATTTGACCAATTCTTAGGCATTAAATTGATATAAAATTTTGGTAT  16363860
```

5xUAS#6

>☐ ref|NC_003074.8| 🟩E🟩D Arabidopsis thaliana chromosome 3, complete sequence
Length=23459830

```
Features flanking this part of subject sequence:
   4626 bp at 5' side: pectinesterase family protein
   2292 bp at 3' side: enzyme inhibitor/ pectinesterase

Score =  200 bits (108),  Expect = 4e-51
Identities = 108/108 (100%), Gaps = 0/108 (0%)
Strand=Plus/Plus

Query  1         CGAGACATAGCGGTTAAGCCGTGAACTCGGTATCAATTTTTGTGTTTGTTTCTCTGACAA  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  10400487  CGAGACATAGCGGTTAAGCCGTGAACTCGGTATCAATTTTTGTGTTTGTTTCTCTGACAA  10400546

Query  61        GCAACCGTTCAGTTTCGTTTTCTCTACAGAACCAAACGGCTAAGGCTA  108
                 ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  10400547  GCAACCGTTCAGTTTCGTTTTCTCTACAGAACCAAACGGCTAAGGCTA  10400594
```

5xUAS#19

>☐ ref|NC_003074.8| 🟩E🟩D Arabidopsis thaliana chromosome 3, complete sequence
Length=23459830

```
Features flanking this part of subject sequence:
   888 bp at 5' side: RNA binding
   1206 bp at 3' side: hypothetical protein

Score =  124 bits (67),  Expect = 1e-28
Identities = 67/67 (100%), Gaps = 0/67 (0%)
Strand=Plus/Minus

Query  1         TTGTTAGAATCTTTTAGTATTACTTAATCTTCCGTTACTATTGTTAATGACCTAACTGAT  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  17966362  TTGTTAGAATCTTTTAGTATTACTTAATCTTCCGTTACTATTGTTAATGACCTAACTGAT  17966303

Query  61        CAGAAGC  67
                 |||||||
Sbjct  17966302  CAGAAGC  17966296
```

5xUAS#20

>☐ ref|NC_003071.7| **E D** Arabidopsis thaliana chromosome 2, complete sequence
Length=19698289

 Features in this part of subject sequence:
   kinase interacting family protein

 Score = 265 bits (143), Expect = 2e-70
 Identities = 143/143 (100%), Gaps = 0/143 (0%)
 Strand=Plus/Plus

```
Query  1         CTGAGAAACCAAAACAAGAACAAGAAAGAAAACATCAGATTTTGTTCTTTTTGAGTAAGT  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12999756  CTGAGAAACCAAAACAAGAACAAGAAAGAAAACATCAGATTTTGTTCTTTTTGAGTAAGT  12999815

Query  61        GGAGATCTCAATACTAAGTGAAGCAACATACTTTCTAGATTCTCTGCAAGCCATTTTGAG  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  12999816  GGAGATCTCAATACTAAGTGAAGCAACATACTTTCTAGATTCTCTGCAAGCCATTTTGAG  12999875

Query  121       TTCTTGGGACAATTGTGACTATC  143
                 |||||||||||||||||||||||
Sbjct  12999876  TTCTTGGGACAATTGTGACTATC  12999898
```

5xUAS#21

>☐ ref|NC_003070.9| **E D** Arabidopsis thaliana chromosome 1, complete sequence
Length=30427671

 Features in this part of subject sequence:
   hypothetical protein

 Score = 568 bits (307), Expect = 3e-161
 Identities = 307/307 (100%), Gaps = 0/307 (0%)
 Strand=Plus/Minus

```
Query  1        ATATCTAGTTTATAACCATATTCTCTAAATTCGAAACTGGAAGTTCTCGAATTTCACTGT  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3560943  ATATCTAGTTTATAACCATATTCTCTAAATTCGAAACTGGAAGTTCTCGAATTTCACTGT  3560884

Query  61       TTGCAAATTATTAAACATATTCCCTAAAATCAAAACAGGAAGTTCTTGAATTTAGTTTCC  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3560883  TTGCAAATTATTAAACATATTCCCTAAAATCAAAACAGGAAGTTCTTGAATTTAGTTTCC  3560824

Query  121      AAAACCATTGTGACTTCAAAAATTATGGTCTCATTCCTACCTCCCCTGTTTACAGATCAT  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3560823  AAAACCATTGTGACTTCAAAAATTATGGTCTCATTCCTACCTCCCCTGTTTACAGATCAT  3560764

Query  181      TGCAGGCAGTCGAAATTTTCATATCATTTTTATCAAAAATGAAAAGGAAATGTAATATAG  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3560763  TGCAGGCAGTCGAAATTTTCATATCATTTTTATCAAAAATGAAAAGGAAATGTAATATAG  3560704

Query  241      CAAAATCTAAGAAAATCTTCGTATAATATATGAGTATTGTGGAAGCTAATAAACACATGC  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3560703  CAAAATCTAAGAAAATCTTCGTATAATATATGAGTATTGTGGAAGCTAATAAACACATGC  3560644

Query  301      TTTCACC  307
                |||||||
Sbjct  3560643  TTTCACC  3560637
```

10.4 *S. cerevisiae* strains

Strain	Genotype	Internal number
wt untagged	Mat a/alpha, ho::LYS2, lys2, ura3, leu2::hisG, trp1::hisG, his3::hisG	3108
wt SPO11-18xmyc	MATa/alpha, ho::LYS2, ura3, leu2::hisG, trp1::hisG, SPO11-Myc18::TRP	3555
hop1	MATa/alpha, ho::LYS2, ura3, leu2::hisG, trp1::hisG, SPO11-Myc18::TRP1, hop1::LEU2, his4, (his4X or his4B not determined)	3039
red1	MATalpha/a, ho::LYS2, lys2, ura3, leu2::hisG, red1::LEU2, SPO11-Myc18::TRP1	4519
mek1	MAT a/alpha, ho::LYS2, mek1::LYS2, ura3, trp1, his4B::LEU2, ade2-Bgl2/ADE2, SPO11-Myc18::TRP1, (lys2/LYS2 not determined)	4479
red1/hop1	MATa/alpha, ho::LYS2, lys2, ura3, leu2::hisG, trp1::hisG, his4, SPO11-Myc18::TRP1, hop1::LEU2, red1::LEU2, pYCp50(URA3)::HO	4477
rec8Δ	MATa/alpha, ho::LYS2, ura3, leu2, rec8Δ::KanMX4, SPO11-Myc18::TRP1, trp1::hisG, (lys2/LYS2 not determined)	4565
rec8Δ/scc1	MATa/alpha ho::LYS2, lys2, leu2, ura3, SPO11-Myc18::TRP1, scc1-73, rec8Δ::KanMX4, (6x back crossed to SK1)	4516

10.5 *S. pombe* strains

Strain	Genotype
wt untagged	h$^+$ pat1-114
wt rec12-13xmyc	h$^+$ pat1-114 rec12+::13xmyc
hop1Δ	h$^+$ pat1-114 hop1-1::kanMX6 rec12+::13xmyc
mek1Δ	h$^+$ pat1-114 mek1::kanMX6 rec12+::13xmyc
mug20Δ	h$^+$ pat1-114 mug20::natMX6 rec12+::13xmyc
rec10-155	h$^+$ pat1-114 leu1-32 rec10-155::LEU2 rec12+::13xmyc
rec27Δ	h$^+$ pat1-114 rec27-184::kanMX6 rec12+::13xmyc

10.6 DSB maps for all *S. cerevisiae* chromosomes

The coverage data was plotted using the free statistic program "R". The red signals represent the mapped full length oligonucleotides. In blue the exact 5' end positions for every mapped oligonucleotide are plotted. The coverage is displayed in number of mapped nucleotides/position. The chromosomes on the x-axis are displayed in bp. The black filled circles indicate the centromere positions

S. cerevisiae
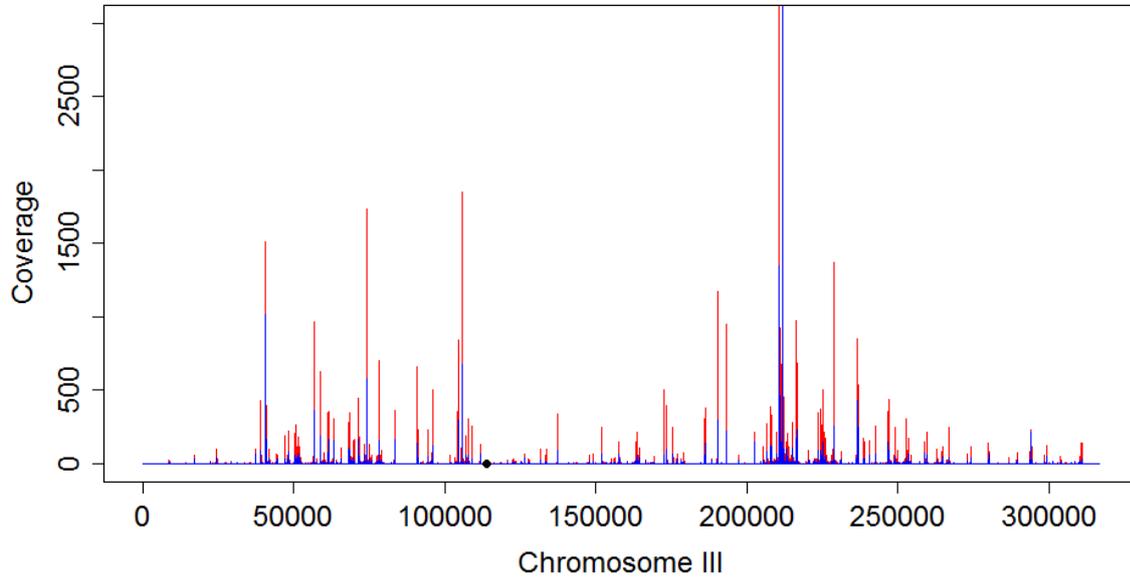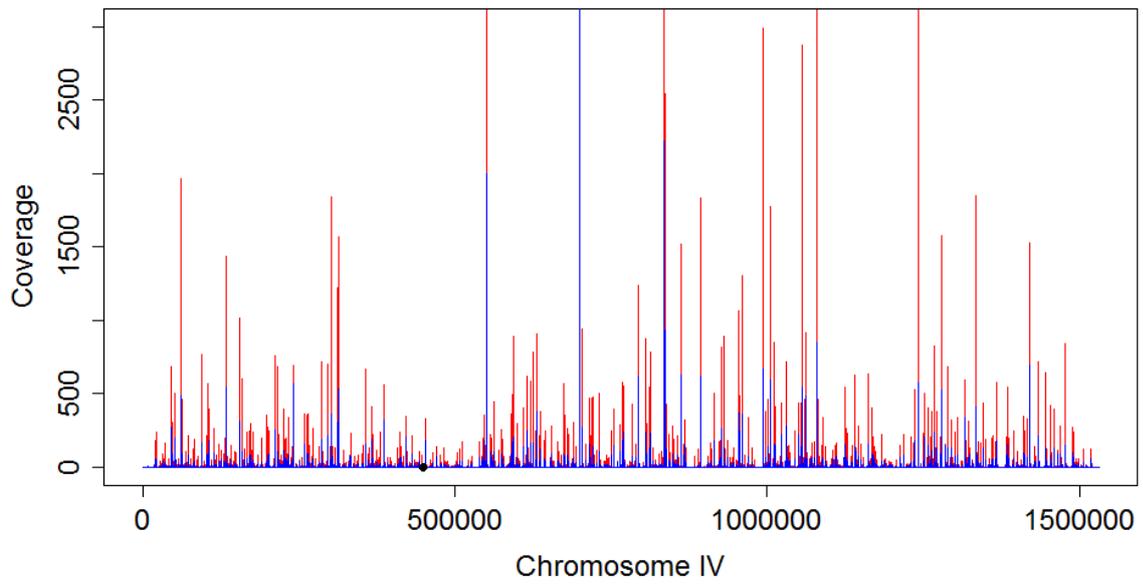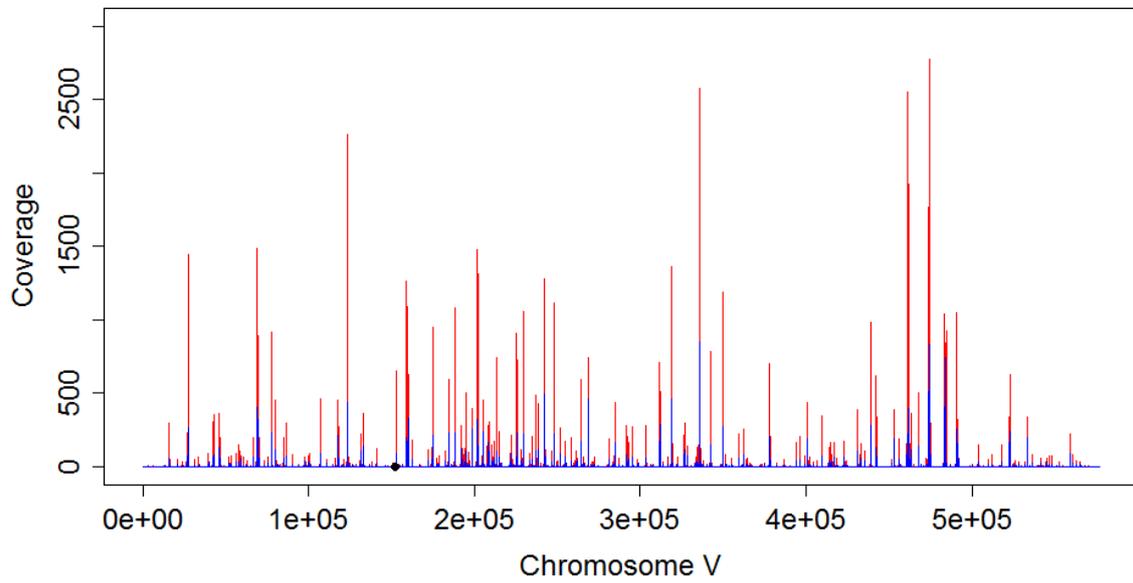
Coverage — Chromosome III

S. cerevisiae

Coverage — Chromosome IV

S. cerevisiae

S. cerevisiae

S. cerevisiae

Chromosome VII

S. cerevisiae

Chromosome VIII

S. cerevisiae
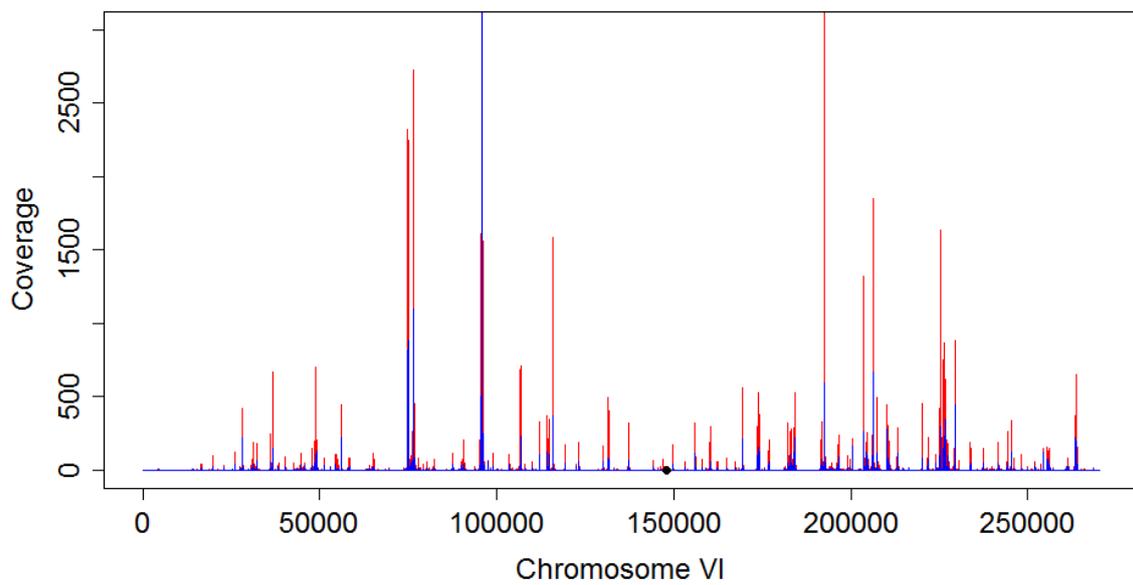
S. cerevisiae
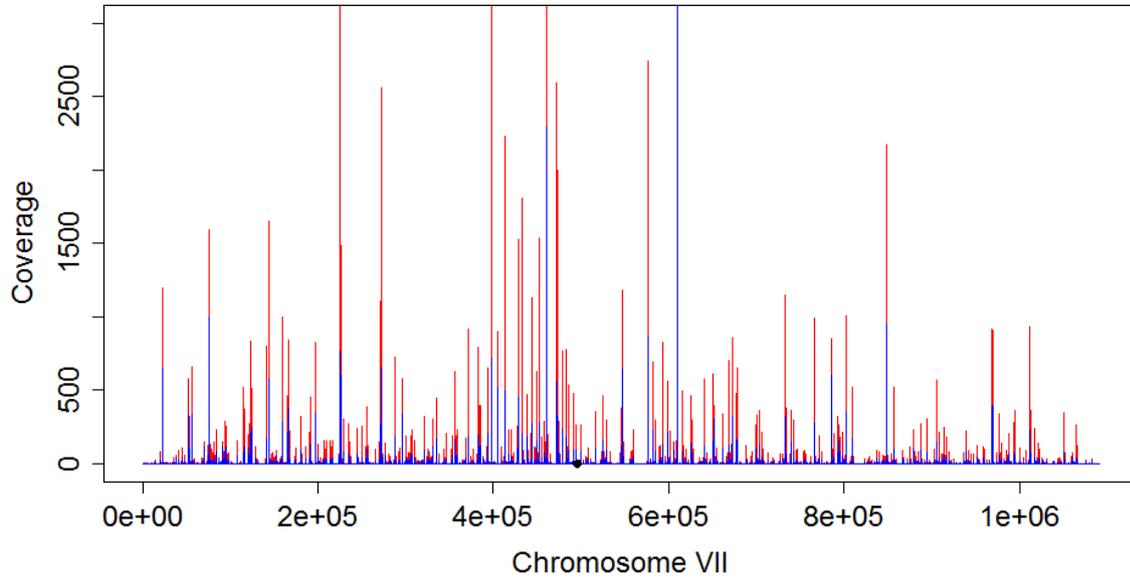
S. cerevisiae
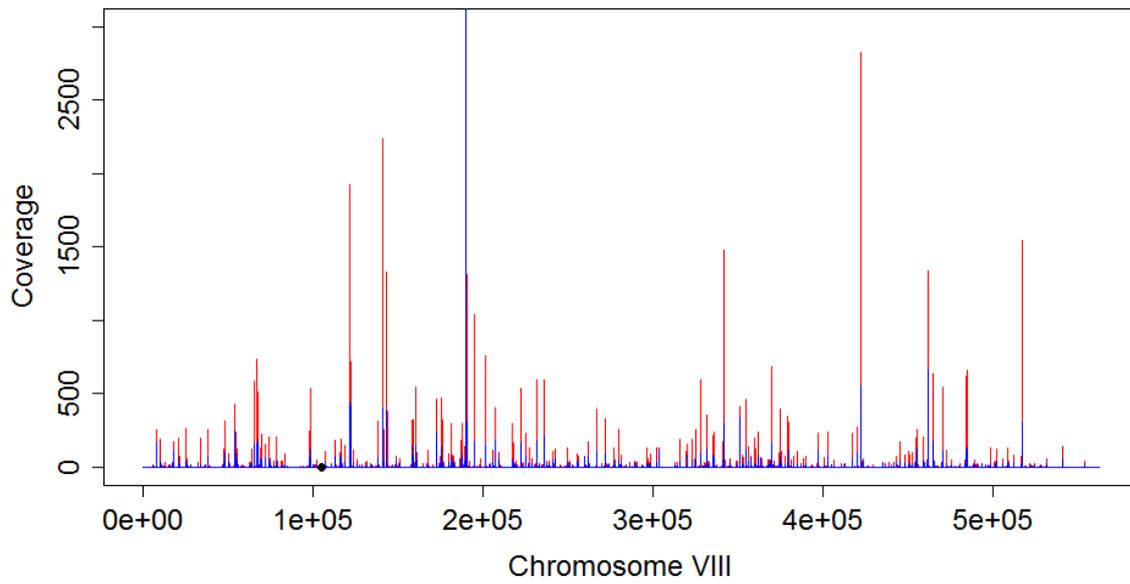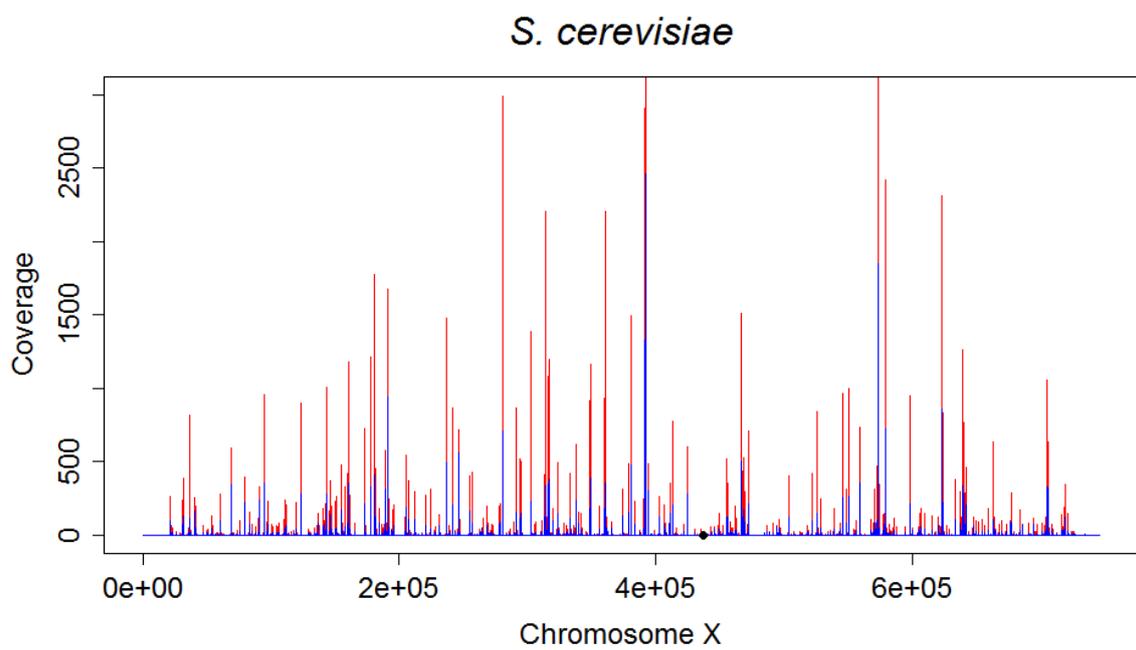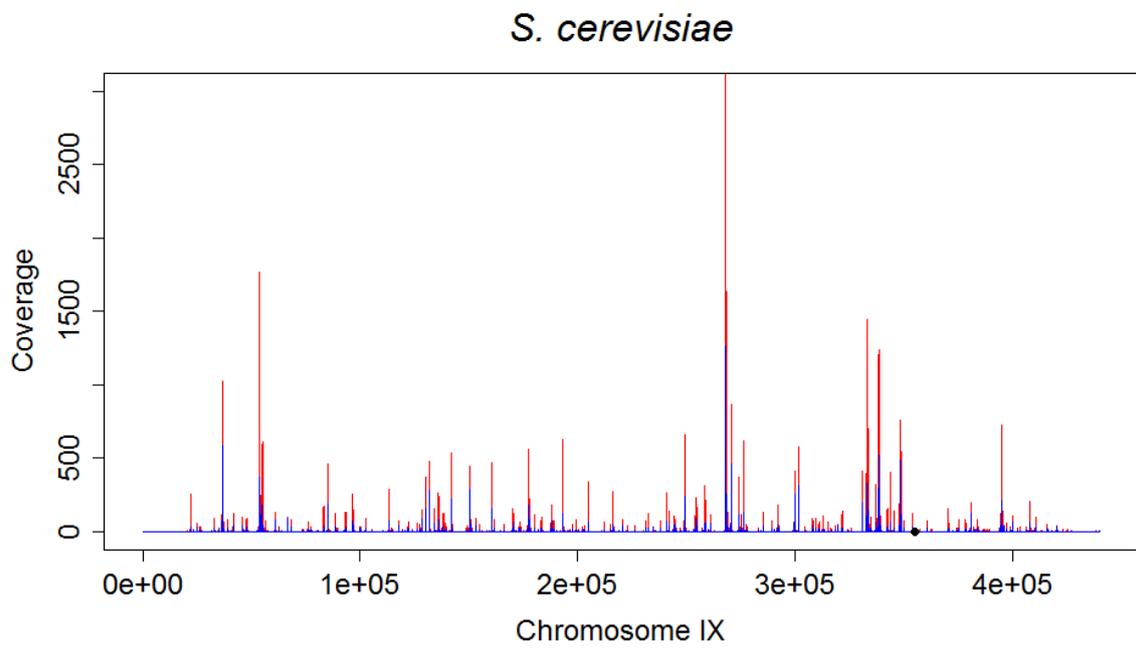
S. cerevisiae
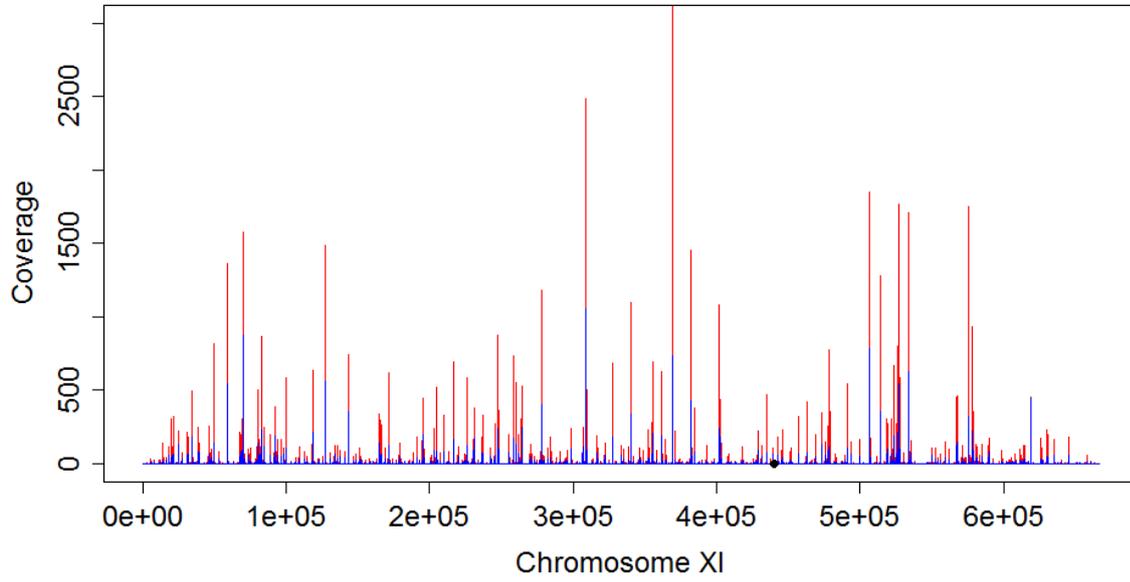
S. cerevisiae
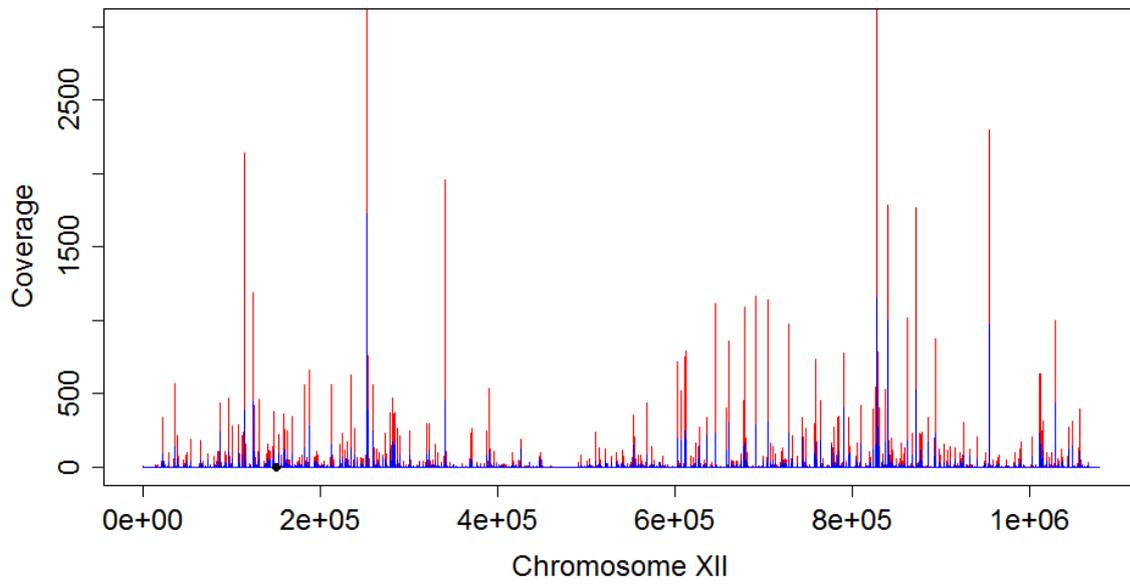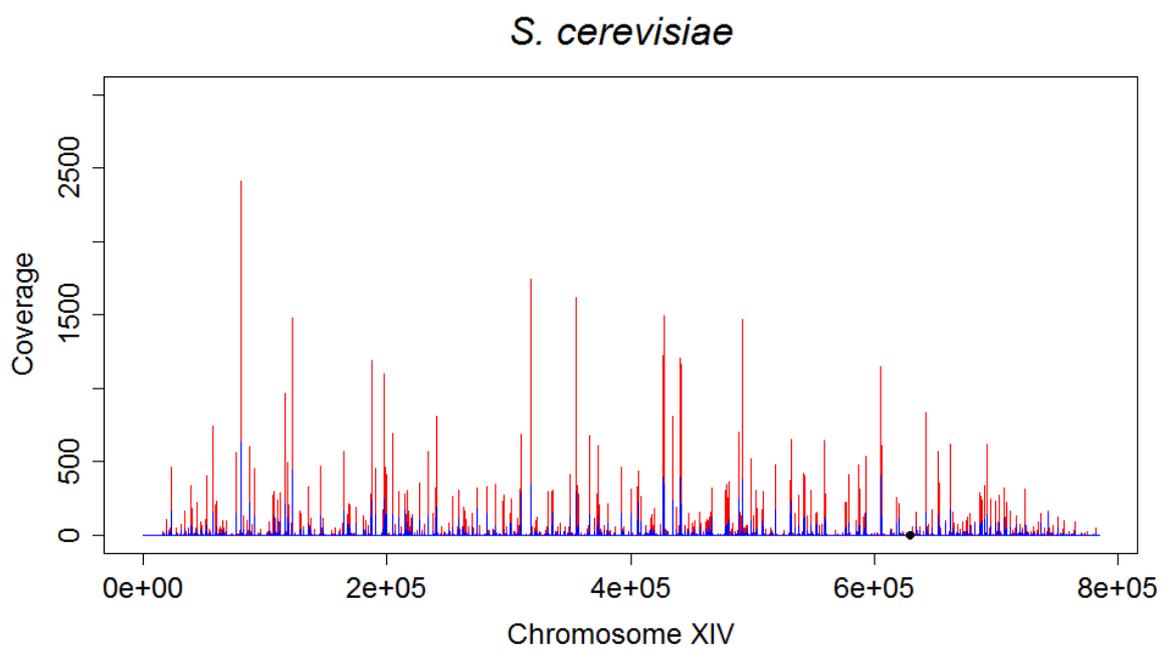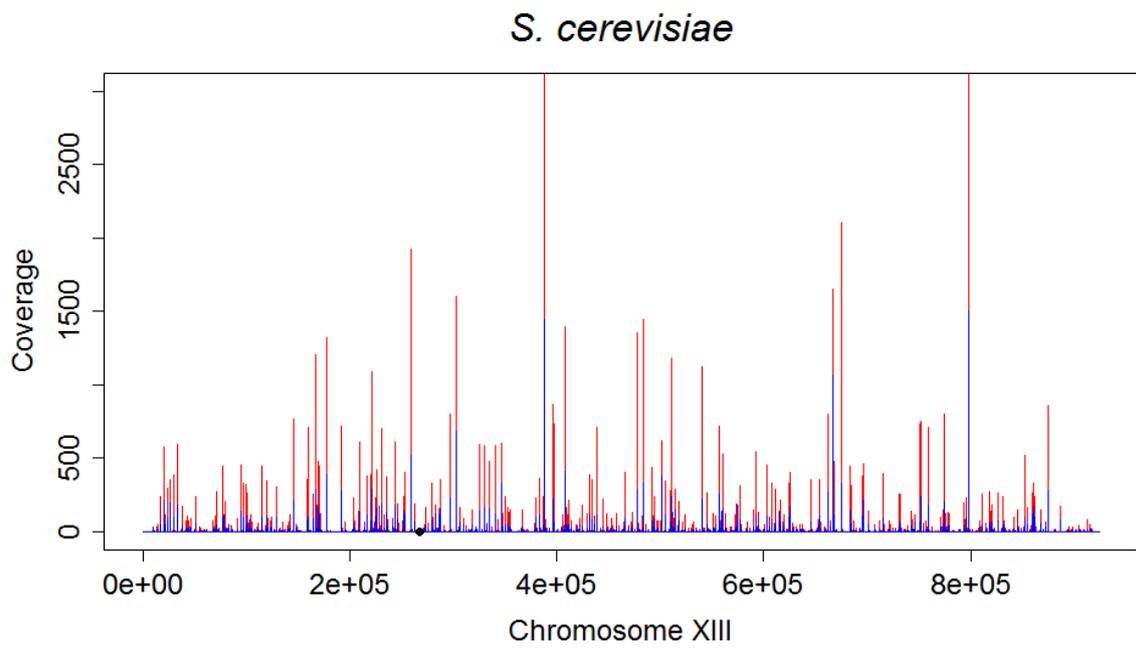
S. cerevisiae
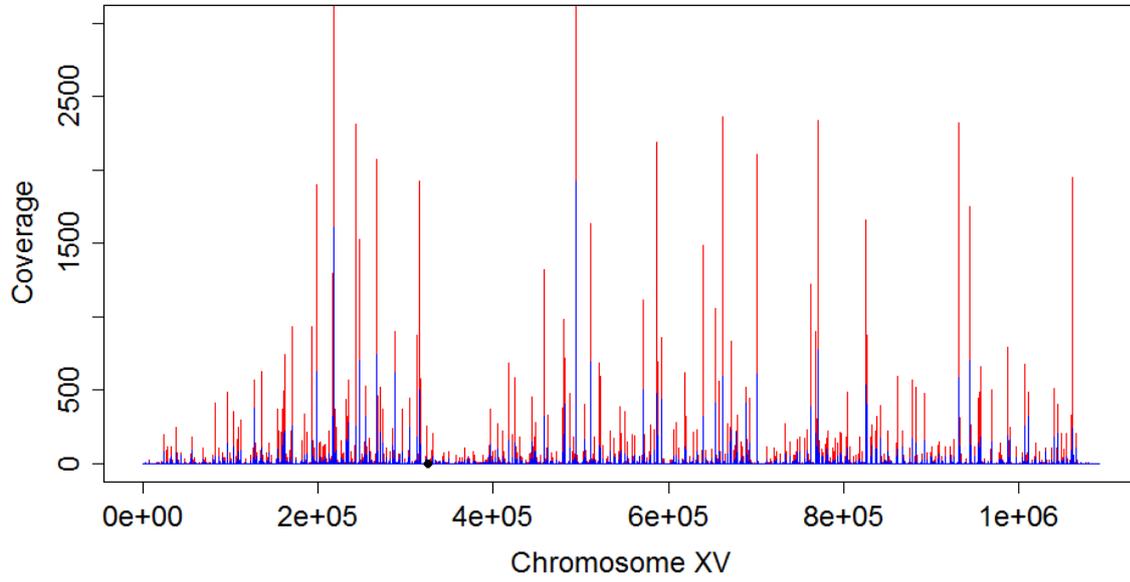
S. cerevisiae

S. cerevisiae

.
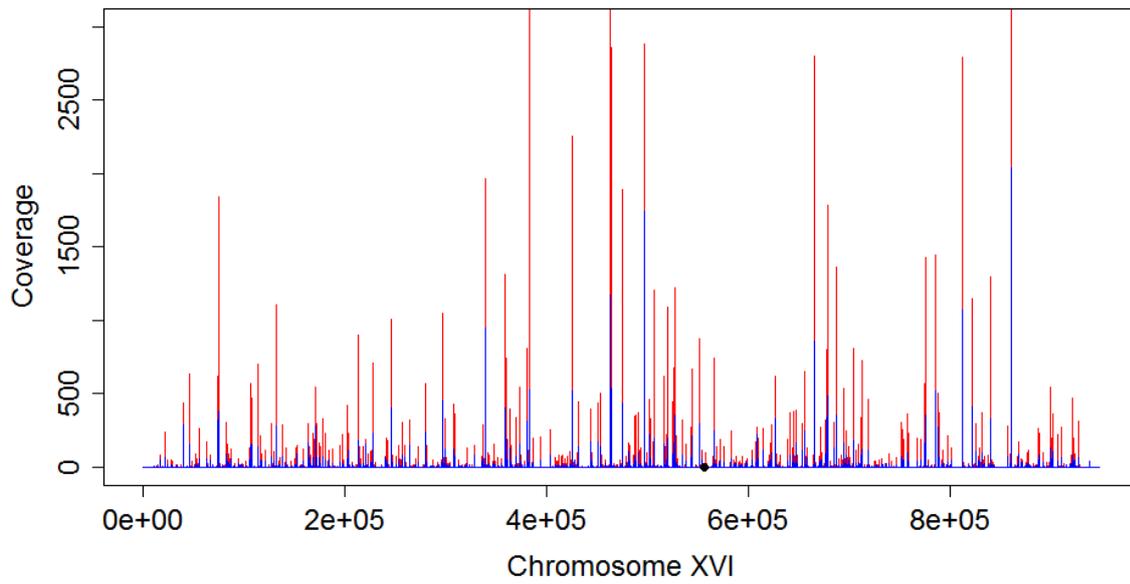
11 Acknowledgements / Danksagung

Herzlichen Dank auch meinen Eltern. Danke, dass Ihr mich während meiner ganzen Ausbildung unterstützt habt und immer zu mir gehalten habt. Danke dafür, dass vor allem in den letzten vier Jahren zu Hause immer ein Platz für mich frei war, um im Urlaub zu entspannen und die innere Ruhe zu finden. Danke!

Der größte Dank gilt meiner Freundin Anna. Danke, dass Du in den letzten vier Jahren immer für mich da warst. Danke dafür, dass du mich in schwierigen Zeiten unterstützt hast, mir Kraft gegeben hast und es Dir immer gelungen ist, mich wieder zu motivieren. Danke für die wunderbare Zeit mit Dir, ob zu Hause oder irgendwo anders auf der Welt. Mein Schatz, ich könnte mir nichts Schöneres vorstellen, als mit Dir zu leben. Du bist das Wichtigste und Wertvollste in meinem Leben. Ich liebe Dich!

12 Curriculum vitae

Personal Data

Mag. Bernd Edlinger

Spengergasse 25A/10

A-1050 Wien

Date of Birth: May, 2nd, 1980

Birthplace: Villach, Austria

Nationality: Austria

Scientific Education

10/2006 - 3/2011	PhD thesis at the Department of Chromosome Biology, MFPL, University of Vienna, Austria
	Supervisor: Dr. Peter Schlögelhofer
	Project title: Determinants of the initiation of meiotic recombination
1/2006 – 5/2006	Scientific work at the Center for Anatomy and Cell Biology, Neuromuscular Research Department, Medical University of Vienna, Austria
4/2004 – 7/2005	Diploma thesis at the Institute of Molecular Biosciences, Department of Biochemistry, Karl-Franzens University of Graz, Austria
	Supervisor: O. Univ.-Prof. Dr. Rudolf Zechner
	Project title: Overexpression and Characterization of Potential Monoglyceride Lipases in the Mouse Genome
10/1999 – 7/2005	Studies of Microbiology, Karl-Franzens University of Graz, Austria

Conferences and Poster Presentations

6/2010	Poster Presentation at the Gordon Meiosis Research Conference, New London, New Hampshire, USA
9/2009	Poster Presentation at the 6th Tri-National Arabidopsis Meeting, Cologne, Germany
9/2007	Poster Presentation at the 4th Tri-National Arabidopsis Meeting, Vienna, Austria

Research Visits

6/2008	2 weeks research visit in the laboratory of Gregory P. Copenhaver, University of North Carolina at Chapel Hill, USA

Other activities

Summer term 2009	Student assistant in the laboratoty course "Molekularbiologische Übungen IB"
Summer term 2008	Student assistant in the laboratoty course "Molekularbiologische Übungen IB"
Summer term 2004	Student assistant in the laboratory course "Biocomputing I"

Publications

Edlinger B and Schlögelhofer P, Have a break: determinants of meiotic DNA double strand break (DSB) formation and processing in plants, Review, 2011, Journal of Experimental Botany (PMID: 21220780).