What Comparative Genomics Tells Us About the Evolution of Eukaryotic Genes Involved in Recombination

F. Hartung and H. Puchta*

Botanical Institute II, University of Karlsruhe, Fritz-Haber Weg 4, D-76128 Karlsruhe, Germany

Abstract: The growing number of completely deciphered genomic sequences provides an enormous reservoir of data, which can be used for addressing questions related to functional and evolutionary biology. The wealth of this approach is documented by the fast growing numbers of recent publications in the field of evolutionary biology based on comparative genomics. Many proteins of the recombination machinery are conserved between plants, fungi and animals but some of them also show remarkable differences regarding their presence, copy number or molecular structure. For example, the protein responsible for double strand break (DSB) induction during meiosis, SPO11, which is related to the subunit A of the archaebacterial topoisomerase VI, is coded by a single gene in animals and fungi. In contrast, plants harbour three distantly related homologues, which seem to have non-redundant functions either in meiosis or in somatic cells and are indispensable for viability. Moreover, plants possess a homologue of the subunit B of the archaebacterial topoisomerase VI, not present in other eukaryotes. We also summarise the recent progress in the usage of genomic data to analyse the evolution of other DNA recombination factors. Finally, several recent studies report on a strong conservation of a reasonable number of intron positions between plants, animals and fungi. This kind of study provides a basis for comparative genomic analyses across kingdoms and demonstrates the existence of ancient introns, a topic of intensive debate.

Key Words: Comparative genomics, homologous recombination, non-homologous end-joining, intron positions.

INTRODUCTION

A single completely sequenced genome can tell you something about a specific organism but two genomes can give a lot more than just twice the information due to comparative genomics. The accelerated accumulation of completed genomic sequences provides an incredibly fast growing reservoir of data. The analysis of these data had just started but the number of recent publications extracted from the data provided by genomic sequencing efforts is increasing as fast as the data pool itself. Even incomplete genomic sequences are valuable for general evolutionary investigations like the analysis of macro- or microsynteny between different species. Most of the pioneer work of comparative genomics has been done using mammals [1] and *Drosophila melanogaster* but already during the sequencing and especially after the completion of the *Arabidopsis thaliana* and *Oryza sativa* genome a reasonable number of studies also concerned in the comparison of different plant species [2-4].

One of the values of a completed genome sequence is its usefulness to simplify homology based cloning of specific genes from related species whose genomes are not yet or will not be sequenced in the near future. Of course, homology based cloning is only suitable if there is enough backbone data available. The *Arabidopsis* genome alone can only help cloning genes of species, which show a reasonable degree of sequence homology or synteny. Several completed genomes (f. e. *Arabidopsis*, rice, tomato and poplar) will cover in future a much bigger range of synteny, and therefore a much broader range of plant species. Several macro- and micro-synteny studies already revealed a reasonable amount of collinearity between different plant species like *Arabidopsis*, tomato, soybean and to some extent also rice [3, 5-8]. Nevertheless, some of the limits of comparative genomics have also been demonstrated. In different comparisons between maize and *Arabidopsis* the sequence analyses of three maize loci exhibited a lack of reasonable collinearity [9]. Furthermore, on the expression level approximately 25,000 new transcripts are expected in the whole maize transcriptome based on the analysis of 40% of the transcriptome, which showed already 10,000 new ORFs (open reading frames) in comparison to *Arabidopsis* [9]. Most probably these new genes arose by duplication followed up by extensive functional diversification, a common evolutionary mechanism which unfortunately can impair the construction of synteny maps and therefore the cloning of genes based on such maps [9].

Another major outcome of completed genomes is their value for evolutionary biology. Comparative genomics and also smaller scaled sequence comparisons can help to elucidate various so far unsolved or unclear questions regarding the relationship of organisms in general. Furthermore, such comparisons can help to clarify the evolutionary development of gene regulation and the evolution of proteins or protein complexes involved in important cellular functions like recombination. Using comparative genomics we can identify also conserved noncoding sequences, intron positions and sequences which are important for the splicing machinery [10, 11]. Comparative genomics has shown its potentials by recent studies comparing the intron positions in

*Address correspondence to this author at the University of Karlsruhe, Botanical institute II, D-76128 Karlsruhe, Germany; Tel: +49-721-608-889; Fax: +49-721-608-4874; E-mail: holger.puchta@bio.uka.de

1389-2029/04 $45.00+.00

several sequences from animals, fungi and plants [12, 13]. Investigations of Fedorov *et al.* [14] and Roy *et al.* [15] using computer aided comparisons of a much larger set of intron containing genes from mammals, fungi and plants further sustained this approach. In such wide range studies one can not directly compare the DNA sequence data, because of their low nucleotide homology. However, it is possible to take the deduced protein data and plot them against their respective genomic exon-intron structure. In this way new and very exciting information about the evolution of specific genes or even gene families can be obtained [16, 17].

The basic assumption of all comparative genomics approaches is that the genomes under examination had a common ancestor [1]. Therefore, the presence of specific genetic information in two (or more) organisms can be explained as a combination of its presence in the original ancestral genome and the forces of evolution. In principle, evolution is a combination of two processes: the mutational forces, which generate random mutations (f.e. replication errors or incorrect repair of DNA damage) and the selection pressure. Three types of selection pressure can be envisaged: i) negative selection, which eliminates random mutations ii) neutral selection, which has no effect on the mutation or iii) positive selection, which results in a gain of fitness in the population [1]. Gene duplication, a major driving force of evolution, provides a substrate for random mutations to occur in a specific gene sequence, thereby allowing new genes and new functions to arise. Examples of this will be given below.

In this review we summarise and discuss the results of comparative genomics and smaller scaled sequence comparisons concerning the evolution of genes involved in recombination.

GENES OF THE RECOMBINATION MACHINERY

The completion and annotation of the sequence of several eukaryotic genomes like *Homo sapiens*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Oryza sativa* and *Chlamydomonas reinhardtii* made it possible to screen whole genomic sequences for common genes involved in recombination. Such comparisons showed that a reasonable part of the recombination and repair machinery in fungi, animals and plants is highly conserved. Several key factors of the homologous recombination (HR) and the non-homologous end-joining pathway (NHEJ) like the MRE11 complex, the DNA end binding proteins KU70 and 80, the Ligase 4, RAD51 and RECQ genes are maintained ubiquitously in all so far investigated eukaryotic organisms. This indicates that the recombination machinery is a very ancient complex, which has been conserved during the evolution from bacteria to eukaryotes. Searching for the most prominent genes of the recombination machinery in rice and *Arabidopsis* we were able to find every gene in both organisms in the same copy number pointing out the conservation of the key factors of recombination between mono- and dicotyl plants (Table **1**).

Nevertheless, also a number of differences could be found by comparing the completely sequenced genomes of *A. thaliana, O. sativa, C. reinhardtii, H. sapiens* and *S. cerevisiae* (Table **1**). Some of the important genes in DNA repair of animals like RAD52, the DNA-dependent protein kinase DNA-PK (which is involved in NHEJ) and p53 have not yet been found in any plant, neither by comparative genomics nor by functional studies. *Vice versa*, some of the genes like the topoisomerase VI subunit B and the DNA helicase SRS2 are not present in animals but play important roles in plants and/or yeast. *S. cerevisiae* does not possess homologues of 6 of the 22 genes listed in (Table **1**) what may point to a lesser complex mechanism of DNA repair in this unicellular organism. Interestingly, the green algae *C. reinhardtii* lacks besides a set of animal/yeast specific genes which are not present in plants, four more genes that are present in *Arabidopsis* as well as in rice. Among these genes are the checkpoint kinase ATM, and the XRCC4 protein, which is involved in concert with Ligase 4 in NHEJ. Furthermore, *Chlamydomonas* seems to have only one of three SPO11 homologues, which is different to higher plants. The *C. reinhardtii* genome sequencing project is just finished and the scaffolds are not totally assembled in the moment. Therefore, it is possible but unlikely that some of the mentioned genes are lacking because they are located in the gaps which will be filled up in near future solving this question.

The RecA/RAD51 Gene(s)

One of the most important genes for recombination, the RAD51 gene, can easily be traced back to the procaryotic RECA gene. RecA is a ssDNA binding protein which protects ssDNA during ongoing repair or recombination [19-21]. The RecA protein promotes identification and exchange of DNA homology regions [22]. Rad51, the main RecA homologue of eukaryotic organisms acts in principle like RecA by forming helical filaments on single- and double stranded-DNA but of an opposite polarity [23]. It can catalyse homologous DNA pairing as well as strand exchange in an ATP-dependent manner [24, 25]. Interestingly, this mechanism seems to be different in meiotic and mitotic cells as all sequenced eukaryotes possess a RAD51 homologue, the DMC1/LIM15 gene, which is specialised only in meiotic recombination [26]. Comparative genomics analysis demonstrates that DMC1 is always present as a single copy gene whereas RAD51 is present in several highly diverged copies depending on the respective organism (Table **1**). Obviously, the RECA gene must have been duplicated once to result in the two genes RAD51 and DMC1 and two studies using different means implicate, that this event most probably precedes the divergence of eukaryotes and archaebacteria approximately 1,000 mya [13, 27]. During the early evolution of eukaryotes, the RAD51 gene has been duplicated several times resulting in the different RAD51like family members, which can be found already in yeast and *Chlamydomonas*.

The MRE11 Complex

A very important protein complex, the MRE11 complex [28], is involved in both, HR and NHEJ. It is a rather ancient complex and its development can be traced back to a bipartite protein complex of eubacteria, the SbcCD complex [29]. The original bacterial SbcCD complex harbours a hairpin endonuclease and a 3`-5`-exonuclease activity and

Table 1. Prominent Genes Involved in Recombination and Repair

Gene name	*A. thaliana / O. sativa*	*C. reinhardtii*	*H. sapiens*	*S. cerevisiae*
Homologous recombination and non-homologous end-joining				
DNA-PK	-	-	1	-
KU70	1	1	1	1
KU80	1	1	1	1
Ligase 4	1	1	1	1
MRE11	1	1	1	1
RAD50	1	1	1	1
DMC1/LIM15	1	1	1	1
RAD51	1	1	1	1
RAD51 paralogs	5	4, no XRCC2	5	3
RAD52	-	-	1	1
RECQlike	7	4	5[1]	1 (SGS1)
SRS2	1	1	-	1
SPO11	3	1	1	1
TOP6B	1	1	-	-
TOP3	2	1 (TOP3B)	2	1
WRN	1	1	1	-
XRCC4	1	-	1	1 (LIF1)
DNA damage assessment and repair				
ATM	1	-	2 (ATM/ATR)	2 (MEC1/TEL1)
BRCA1	1	1	1	-
BRCA2	2 [2]	1	1	-
P53	-	-	1	-
RAD18	-	-	1	1

Table 1. shows the occurrence of important genes which are involved in recombination and repair in 5 different organisms. The numbers are representing the copy number of the respective gene in an organism. Frequently used synonymous gene names are given in brackets behind the copy number.

[1] The Rec5 gene exists in 3 isoforms due to alternative splicing [18].

[2] the two BRCA2 homologues of *A. thaliana* are a result of a recent duplication event (unpublished data).

most probably processes hairpins which are formed during DNA replication in *E. coli* [30, 31].

The genes of a very similar complex harbouring also an endonuclease and an ATPase function are called MRE11 and RAD50. These conserved genes have been found by comparative genomics or earlier by similarity search in yeast and all other so far completed eukaryotic genomic sequences [32-37]. Both genes show a conserved structure and in the case of MRE11 even more than 35% of the intron positions are identical between plants and animals [13]. Several functional studies of the MRE11/RAD50 protein complex have indeed demonstrated that it harbours a nuclease activity comparable to the SbcCD complex of *E. coli* and an ATPase function which is located in the RAD50 protein [38-42]. One of the main function of the complex is most probably to bridge DNA ends together during recombination events [43, 44]. Interestingly, a number of *in vitro* studies have also been performed with the MRE11/RAD50 complex of the archaebacterium *Pyrococcus furiosus* showing the structural and functional conservation of this recombination complex [41, 45, 46].

During the eukaryotic development a third protein joined the complex, which is named XRS2 in yeast and NBS1 in mammals [47-49]. The NBS1 protein is indispensable for MRE11 phosphorylation upon damage and is itself the primary target for phosphorylation by the "Ataxia telangiectasia mutated" ATM kinase involved in checkpoint control [50, 51]. Furthermore, NBS1 potentiates DNA

unwinding and endonuclease cleavage of the MRE11/RAD50 complex *in vitro* [49]. Most probably due to this third protein the *in vivo* functions of the whole MRE11/RAD50/NBS1 (MRN) complex (Fig. **1**) are more variable as in prokaryotes. The MRN complex is structurally involved in DSB repair via HR as well as NHEJ [28, 42, 52-54]. In mitotic cells the DSB is repaired via both the NHEJ and the HR pathways but in meiotic cells preferably HR occurs [39, 55-57]. Furthermore, it has been shown that the MRN complex is essential for telomere maintenance in yeast, animals and plants [58-62] and it is also involved in checkpoint control [48, 51, 63-69]. The various different functions of the MRN complex are summarised in (Fig. **1**).

Unfortunately, the XRS2/NBS1 protein is not conserved at the sequence level and revealed hardly any detectable homology between yeast and mammals [70]. Therefore, it has not yet been detected in plants which clearly demonstrates the limits of comparative sequence analysis. In summary the MRN complex is one of the most important key enzymes of recombination and repair and the core components MRE11/RAD50 can easily be identified by means of comparative genomics in all eukaryotes.

The RECQ Like Genes

The function of the RecQlike genes in eukaryotic recombination is not very well understood but these genes are also going back to an eubacterial ancestor gene named RECQ. In *E. coli* the RecQ protein was originally described as a suppressor of illegitimate recombination [71]. The evolutionary development of the single eubacterial RECQ gene in eukaryotes is quite amazing. The protein size of some RecQ homologues has increased from bacteria to yeast and *Neurospora crassa* more than three times from 600 to more than 1950 amino acids (Fig. **2**) [72, 73]. Furthermore, the gene must have been duplicated several times during evolution of higher eukaryotes. It turned out by comparative genomics that there are in *C. elegans* four, *D. melanogaster* and *H. sapiens* five and in plants seven different RecQlike genes [73, 74]. It is not yet known if all of this genes have acquired different functions but at least most of the genes in

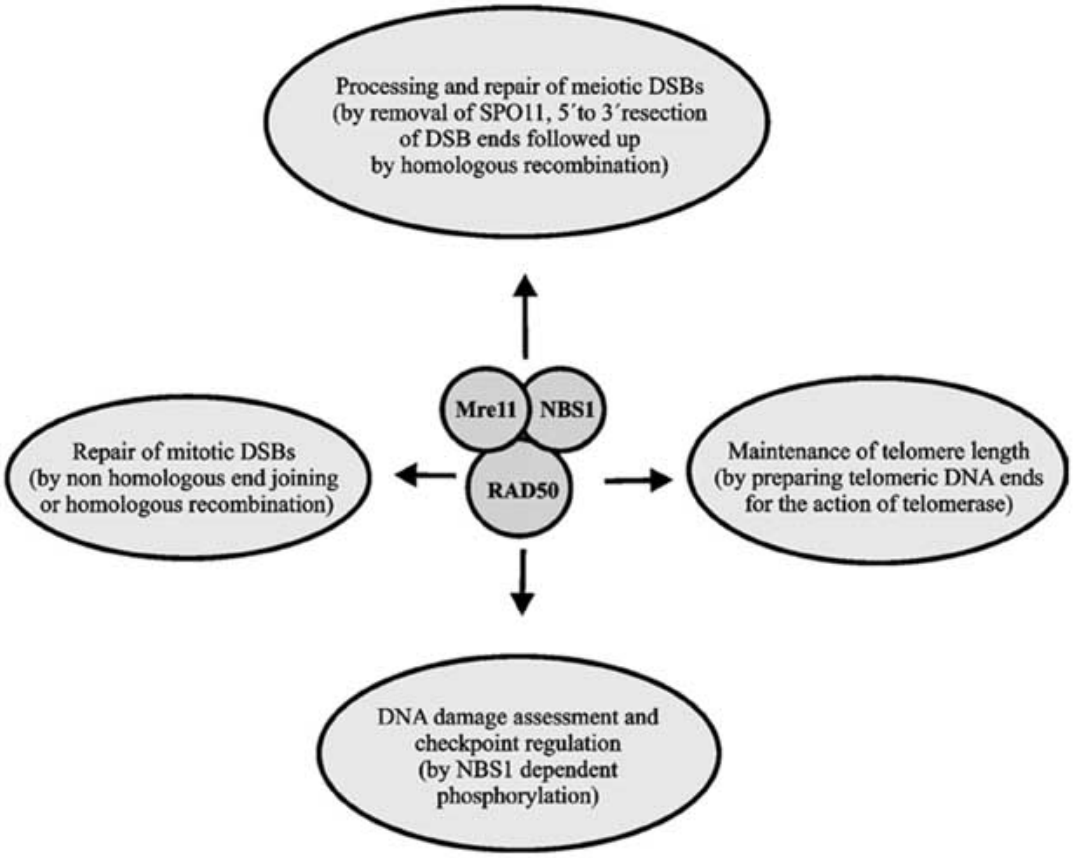

Fig. (1). Different functions of the MRN complex in eukaryotes.

One of the main functions of the MRN complex is to bridge and tether the DNA ends of a DSB since coordinated repair occurrs either by NHEJ or by HR [137]. This function seems to be essential for proper replication because it prevents the formation of non repaired DSBs during replication [66]. The MRN complex also processes the DNA ends to make them accessible for subsequent reactions like telomere maintenance and mitotic or meiotic recombination. Finally, the MRN complex is involved in DNA damage assessment via producing single stranded DNA which serves as a damage signal for checkpoint kinases (e.g. the ATM kinase) and repair proteins.

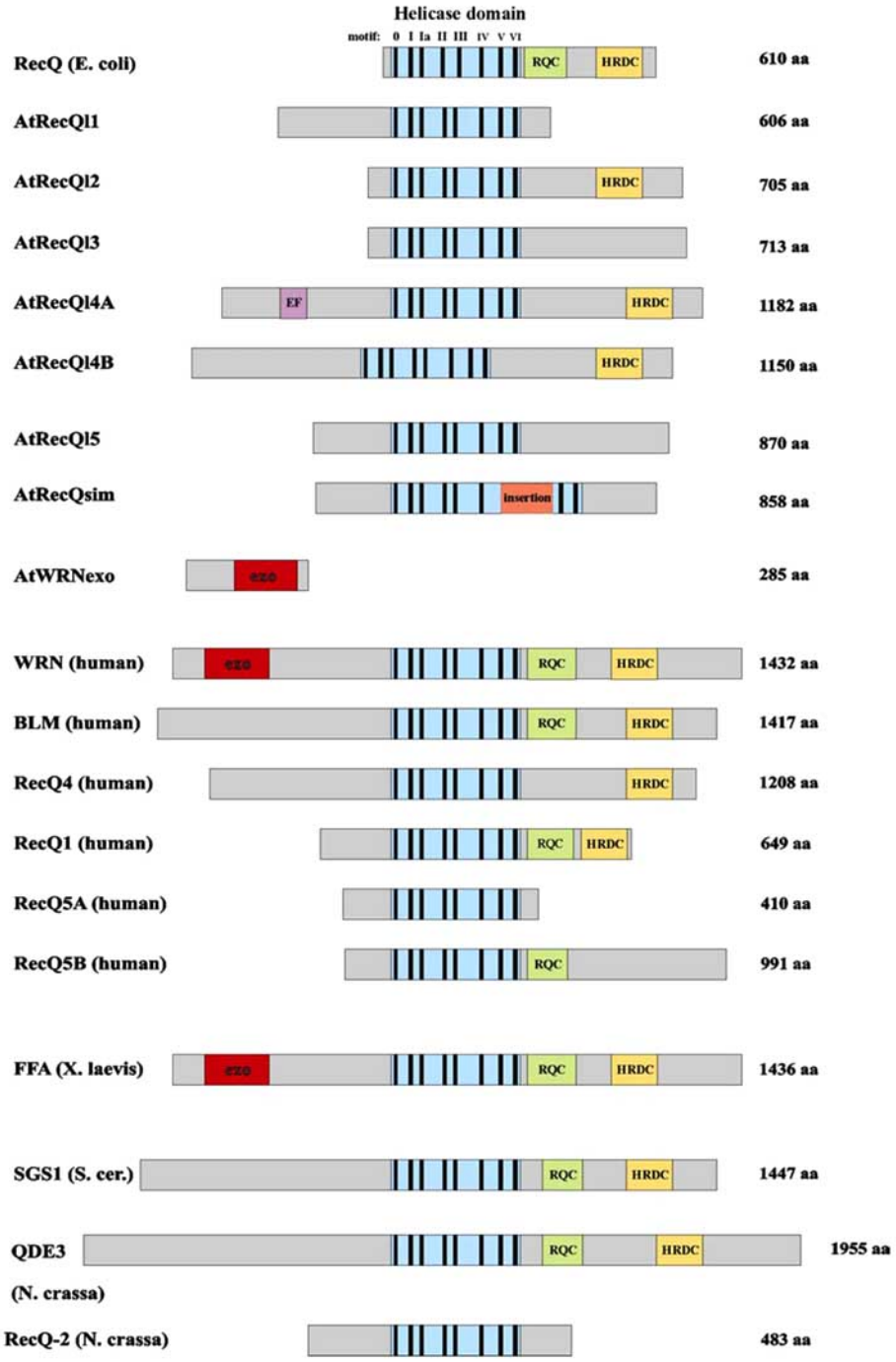

Fig. (2). Schematic structure of the RecQlike protein family.

The structure of several RecQlike proteins from plants, animals and fungi as well as the original *E. coli* protein are depicted. In general all RecQlike proteins contain the conserved helicase domain but they can also harbour additional conserved areas like the RQC and/or HRDC domain which are already present in the RecQ protein of *E. coli*. One RecQlike protein (WRN/FFA) contains a new domain which codes for an exonuclease function and is most probably the result of a gene fusion event during animal evolution. Colour code; blue with black stripes: the conserved helicase domain with the 8 highly conserved motifs; green box: the also conserved RQC domain; light yellow: the HRDC domain; red: the exonuclease domain of the human WRN protein, the *Xenopus laevis* FFA and the WRNexo protein of *A. thaliana*; purple: the EF-hand domain which was only found in AtRecQl4A; orange: the insertion in the helicase domain of AtRecQsim. The size of the respective proteins is given on their right side and their respective names on the left side.

mammals seem to differ functionally. Three of the five RecQlike genes in human exhibit severe diseases like the Blooms, the WRN and the Rothmund-Thomson syndrome [75-77]. All these diseases have in common an increased genetic instability like chromosomal rearrangements and an increase in sister chromatid exchange [78]. There is no natural mutation known from the other two human RecQlike genes [RECQl1 and RECQl5) but chicken DT40 cell lines containing mutated forms of one or both genes exhibited no significant difference to the wild-type cells [79]. However, in the same study double mutants of RecQl1 or RecQl5 and the BLM protein showed an increased rate of death cells indicating that RecQl1 and 5 are somehow involved in cell viability under conditions of an impaired function of the BLM protein [79].

All RecQlike proteins discovered so far harbour a helicase domain which consists of seven motifs (Fig. **2**, I to VI), which are indispensable for the helicase function itself and an eight motif (motif 0) which has recently been characterised as essential for transient DNA binding [80]. Two other domains have been described for a number of RecQlike proteins. First, the RecQ conserved domain (RQC), which is preserved in most of the RecQlike proteins but interestingly not in plants. No specific function has been addressed to this region so far [80, 81]. Second, the HRDC domain, which is found in different RecQlike proteins of all organisms but also in the RNase D protein and its eukaryotic homologues [81]. All three domains are conserved in several of the known RecQlike proteins regardless of the size of the protein in the respective organism (Fig. **2**).

The RecQlike proteins of the fungi *S. cerevisiae* and *Neurospora crassa* (*sgs1* and *qde3*) possess the typical RecQlike structure containing all three domains. However, in contrast to yeast a second RecQlike gene has been identified in *N. crassa* named RECQ-2 and this might already reflect the difference in complexity between unicellular and multicellular organisms. Interestingly, the RECQ5 gene of *H. sapiens* and *Drosophila melanogaster* exists in three different splicing forms resulting in two nearly identical short and one long variant of this protein (Fig. **2**) [18, 82]. Most probably the three different isoforms of *recQ5* possess different functions *in vivo,* which still have to be analysed. It has been shown so far that already the small isoform of *D. melanogaster recQ5* is capable to perform all the major helicase functions *in vitro* [83, 84]. One is tempted to speculate that putative functional variations are triggered rather by interactions with other proteins than by the RecQ5 protein itself.

A new function has been adopted by the human WRN syndrome and the *Xenopus laevis* FFA (focus forming activity) gene [76, 85]. In addition to the three already described domains, an exonuclease domain located in the N-terminal part of the protein (Fig. **2**) enables the protein to fulfil in addition to the helicase function also an exonuclease function, both of which can be separated to different domains within the protein [86-88]. This kind of composed protein is neither found in yeast nor in plants. However, by means of comparative genomics we were able to find in *A. thaliana* (and also in *O. sativa* and *C. reinhardtii*) in addition to the seven RecQlike genes a very small open reading frame named AtWRNexo, which had a striking similarity to the N-terminal exonuclease domain of the human WRN protein (Fig. **2** [74]). This finding gives a hint for the evolutionary mechanism of proteins harbouring domains of different function. Most probably two independent proteins existed (like in plants nowadays) that participated in some common complexes. During the animal development a fusion event occured between one of the duplicated RecQlike genes and the exonuclease gene resulting in a new protein, the human WRN syndrome gene. The fact that the AtWRNexo protein exhibited two hybrid interactions with several RECQlike genes of plants ([74] and unpublished results) strongly supports this hypothesis. The postulated fusion event enabled the large WRN protein to perform its function more precisely and faster than the two separated proteins in plants as its expression is not any longer depending on two different chromosomal loci. The selective advantage of such a fusion event seems to be valuable enough to enable the WRN protein to persist during animal evolution. *In vitro* work on the AtWRNexo protein indicates that its exonuclease activity shows indeed strong similarities compared to the human WRN protein [86, 89-92, 138].

A second new type of RecQlike gene has been found in plants, named the RECQsim gene [74, 93]. In this gene a new DNA sequence coding for an unusual high number of the amino acids glutamic and aspartic acid has been inserted in the conserved helicase domain between the motifs IV and V (Fig. **2**). By means of comparative genomics one can find this RecQlike gene harbouring the same type of insertion at the same location also in *O. sativa* but not in *C. reinhardtii*. Thus the insertion must have occurred more than 160 mya (the assumed divergence time of mono- and dicotyledonous plants) and less than 500 mya (the divergence time of green algae from land plants). Obviously, this new type of RECQlike gene achieved a special function, otherwise it would have been mutated to a pseudogene or already eliminated during the last 160 million years simply by forces of selection. Supporting the functional conservation hypothesis, the AtRECQsim gene has the potential to partially complement a yeast mutant of SGS1, the unique RECQlike gene in this organism, and therefore has at least conserved some of its original RecQlike function [93].

By means of comparative sequence analysis we also found a recent duplication of one of the plant RECQlike genes, which resulted in the genes AtRECQl4A and 4B as shown in (Fig. **3**). Both genes are located on chromosome 1 of *A. thaliana*. The duplication event is not only restricted to the RECQlike genes but consists of an area of approximately 35kb containing 7-8 genes from which are 5 conserved in position and sequence (Fig. **3**). Interestingly, 3 of the 5 conserved genes changed their orientation with respect to the two loci. This change from sense to complementary orientation of the Ser/Thr Kinase, AtRECQl4B and the P4,5-Kinase most probably occurred after the duplication event and therefore very recently. Sequence comparisons of AtRECQl4A and 4B with the same genes of a near relative of *A. thaliana* named *Capsella rubella* placed the duplication time point approximately 15 million years ago (unpublished results). The separation of *Arabidopsis* and *Capsella* from a common ancestor is traced back 8-10 mya [94]. Due to this

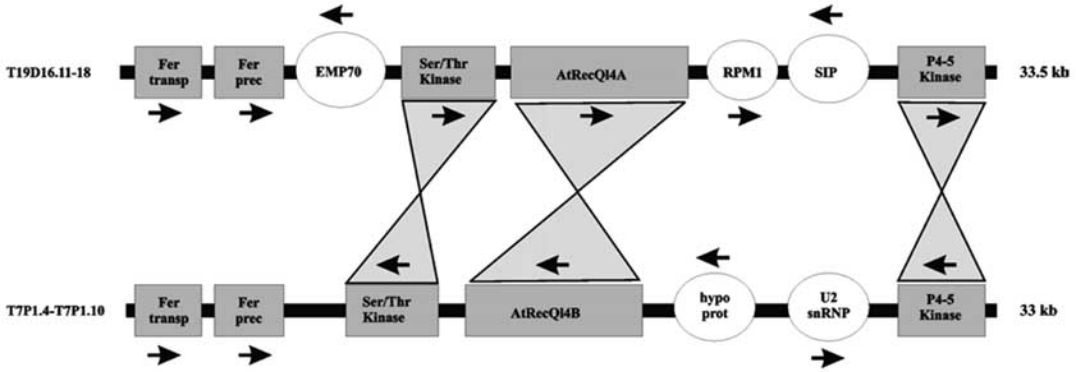

Fig. (3). Duplication of the AtRECQl4 locus.

A region spanning approximately 33 kb has been duplicated recently on Chr. 1 of *Arabidopsis thaliana*. The conserved genes in the duplicated region are shown as grey boxes, other genes as white bubbles. The complementary orientated connection of three of the genes is shown as grey shaded triangles. The gene names are given on the left and the size of the duplicated area on the right side. Abbreviations: Fer tra = iron regulated transporter; Fer prec. = ferredoxin precursor; EMP70 = endomembrane protein EMP70 precusor; RPM1 = disease resistance protein RPM1; SIP = membrane-associated salt-inducible protein; P4,5-Kinase = phosphatidylinositol-4-phosphate 5-kinase; U2 snRNP = similar to U2 snRNP auxiliary factor, large subunit.

fact the duplication cannot be linked to one of the more ancient polyploidisation events in an *Arabidopsis* ancestor [5, 95, 96]. Therefore, the duplication event has to be regarded as rather recent and locally restricted. Interestingly, the AtRecQl4A protein harbours an EF-hand domain which is neither present in AtRecQl4B nor in other RecQlike proteins (Fig. **2**). The EF-hand domain is found in Ca2+-binding proteins that are involved in cell signal transduction [93] and might enable AtRecQl4A to perform a different function than AtRecQl4B.

In general, multiple changes occurred during evolution of the RECQlike gene(s). Especially the copy number and the size of several homologous RecQ proteins is nowadays varying dramatically in eukaryotes. Nevertheless, the original protein structure and the helicase function of the proteins has been well preserved. As it has been shown in several investigations, the general conserved function of the RecQlike protein seems to be required for the resolution of various uncommon DNA structures like holiday junctions, DNA triple helices and G4 DNA tetraplexes [97-101]. Such structures are common in various aspects of cellular processes like recombination, DNA repair, transcription, replication and telomere maintenance and RecQlike proteins have been shown to be involved in all of these processes [102-108].

The described functions of RecQlike genes are obviously indispensable for the proper development of cells. The increased number of RecQlike genes in higher eukaryotes is most probably not the result of redundancy. More likely this means that each of the various RecQlike genes has become specialised to act on certain DNA substrates and that this specialisation is conferred by specific protein-protein interactions. Finally, the more complex structure of the RecQ family in multicellular organisms compared to unicellular ones is pointing to a more urgent need to resolve various uncommon DNA structures in these organisms than in the latter ones.

SPO11, AN ARCHAEBACTERIAL TOPOISOMERASE AND MEIOTIC RECOMBINATION

The SPO11 gene of yeast was originally found in a sporulation defective yeast mutant [109]. S*po11* is the factor introducing double strand breaks (DSBs) to initiate meiotic recombination in yeast [110]. It turned out to be the homologue of one part of a bipartite archaebacterial topoisomerase VI [111]. This archaebacterial topoisomerase consists of two subunits (A and B) and SPO11 is the homologue of the subunit A. The function of the archaebacterial topoisomerase VI consists of cleaving the phosphodiester backbone of a DNA duplex through the formation of covalent phosphotyrosyl linkages with the 5´ends of the broken strands (subunit A). Subsequently, the topoisomerase captures a second DNA duplex, passages it through the break and reseals the DSB of the first DNA after passage (subunit B) [112]. During evolution, the function of subunit A has been conserved in the protein SPO11 but now this protein remains covalently linked to the DSB and due to the lack of subunit B has to be removed by a different protein [110]. By means of comparative genomics one can find only one homologue of SPO11 and furthermore rule out the existence of a subunit B in all the completely sequenced eukaryotic genomes except in plants. However, in *Arabidopsis* and also in the rice genome we were able to find not only three different SPO11 genes but also a homologue for the subunit B [113, 114]. The three SPO11 genes seem to fulfil different functions in *Arabidopsis*. For two of them this can be documented by their different mutant phenotypes [115, 116]. The mutant plants of AtSPO11-1 are nearly sterile, showing an abnormal meiosis resulting in chromosomal missegregation with only a few surviving normal seeds. Furthermore, a dramatic decrease of meiotic recombination was observed by microsatellite marker analysis [115]. Nevertheless, some of the seeds can survive and propagate. In strong contrast, the phenotype of knockout plants of AtSPO11-3 as well as AtTOP6B is totally different. Both

mutant lines were lethal or at least severely impaired [116-118]. The plants showed strong growth retardation, enhanced DNA damage and a limited ability of endoreduplication, which stopped before the second endoreduplication cycle from 4C to 8C was finished [116, 118]. To our knowledge, there is no AtSPO11-2 mutant available and neither a knockout phenotype nor an *in vivo* function of this gene has been described. Nevertheless, the fact that the phenotypes of both mentioned AtSPO11 mutants (of no. 1 and 3) are severe but different indicates that there is no reasonable redundancy between the three SPO11 genes. Therefore, we can speculate that the function of AtSPO11-2 differs from AtSPO11-1 and 3, otherwise it should have been lost during plant evolution.

Most interestingly, the AtSPO11-3 and the AtTOP6B proteins were shown to interact with each other in a two hybrid assay [114]. This finding and the identical knockout phenotypes demonstrated the existence of a conserved and functional Topoisomerase VI in plants [119]. Surprisingly *C. reinhardtii* exhibits only one homologue of the AtSPO11-3 gene in its nearly finished genome (www.biology.duke.edu/chlamy_genome/). This organism belonging to the green algae, which separated almost 500-600 mya from land plants possesses also the subunit B in its genome but no AtSPO11-1 and 2 homologues could be found so far. This is somehow peculiar because the AtSPO11-1 gene is indispensable for initiation of meiotic recombination as has been shown extensively in fungi, animals and plants [110, 115, 120-123].

To summarise, one part of the topoisomerase VI, the subunit A (SPO11) has retained its physical function to produce DSBs from archaebacteria throughout all investigated eukaryotic organisms, and therefore resembles a highly conserved and important gene in recombination. In higher plants this gene has been duplicated and retained together with the topoisomerase subunit B, due to its indispensable role in endoreduplication. As no mutant has been isolated till today the function of the third SPO11 gene (AtSPO11-2) in plants is still elusive.

SEQUENCE COMPARISONS OVER BIG EVOLUTIONARY DISTANCES

A further big advantage of completed or ongoing genomic sequences is the possibility of structural comparisons like the exon/intron structure of certain genes or their regulatory sequences like promoters and 3´-UTRs [10, 11, 124]. The exon/intron structure of important genes turned out to be very well conserved. Two recent studies addressing these kinds of investigations showed a 15-25% conservation over nearly 1 billion years of the intron positions in a number of genes [13, 14]. The former study was done on several genes of the recombination machinery and revealed that 25% of the intron positions were conserved between animals, fungi and plants. The latter study was done in a broader comparative genomics attempt using several thousand genes and revealed an average of 14-18% of conserved intron positions in plants, animals and fungi respectively [14]. In both cases the idea of early introns was strongly supported. Nevertheless, such studies will become much more meaningful as more genomes of different kinds of organisms like mosses, ferns, gymnosperms or algae will be sequenced due to their intermediate positions in the time scale of evolution.

Fig. (**4**) shows as an example the intron position comparison of the MRE11 gene from a fungi, an animal and a plant [13]. The position of each verified intron was plotted onto the aligned protein sequences of all three organisms. Resulting identical intron positions between plants and animals/fungi were regarded as ancient. Intron positions identical between fungi and animals were not regarded as ancient because the divergence of animals and fungi occurred after the separation of animals and plants and so it could be an intron gain event before separation of animals and fungi [125, 126]. All introns occurring only in one organism were considered as recent intron gain events. An ancient intron means that it was already present before the divergence of plants and animals and therefore existed already in their last common ancestor 1,000 million years ago. Interestingly, the basidiomycete fungus *Coprinus cinereus* has lost 6 of the 10 ancient introns during his evolution, which points to a preferred intron loss mechanism in this organism.

The work about the exon/intron structure of genes relevant for recombination demonstrated, for the set of genes used, that there are (in spite of their limited level of macrosynteny [3]) not many differences between mono- and dicotyledonous plants on the level of homologous genes. The comparison of *A. thaliana* and *O. sativa* with respect to their intron positions exhibited only one difference out of 51 intron positions [13]. Regarding the separation time of mono- and dicots around 160-180 mya this was a really astonishing result as it indicates that in the last 160 million years only 2% of the intron positions changed with respect to the structure of the investigated genes. Of course this is not always the case as the analysed genes belong to a rather conserved class. In fact, comparative genomic approaches demonstrated that there are indeed significant differences between rice and *Arabidopsis* [3]. Nevertheless, 160 million years seem not to be a very long time for changes on the level of the individual gene structure and the evolutionary course might occur rather in nonlinear jumps and stops than in a linear fashion.

First attempts of comparative genomics/transcriptomics between *Arabidopsis* and a more distantly related lower plant, the moss *Physcomitrella patens*, have already been started and showed that around 66% of the transcribed genes of moss have homologues in the *Arabidopsis* genome [127]. This investigations were done using the EST data (transcriptome) from moss in comparison to several finished genomes including the one of *Arabidopsis*. The moss transcriptome used in this study was not saturated, and therefore the real amount of homologues between *Arabidopsis* and *Physcomitrella* is estimated to be even higher than 66%. In a recent study two RAD51 genes of *Physcomitrella* have been isolated and compared to other plant RAD51 genes [128]. Astonishingly, both *Physcomitrella* RAD51 genes have a totally different gene structure than the *Arabidopsis* genes, they possess no introns. In contrast, all known RAD51 or RAD51like genes of plants and animals possess several introns, which are sometimes even differentially spliced [129]. The intronless RAD51 genes are quite surprising out of two reasons: a) introns are a common feature in *Physcomitrella* genes [128] and b) the comparison of the intron positions between RAD51 and DMC1 genes of animals, fungi and plants revealed several highly conserved

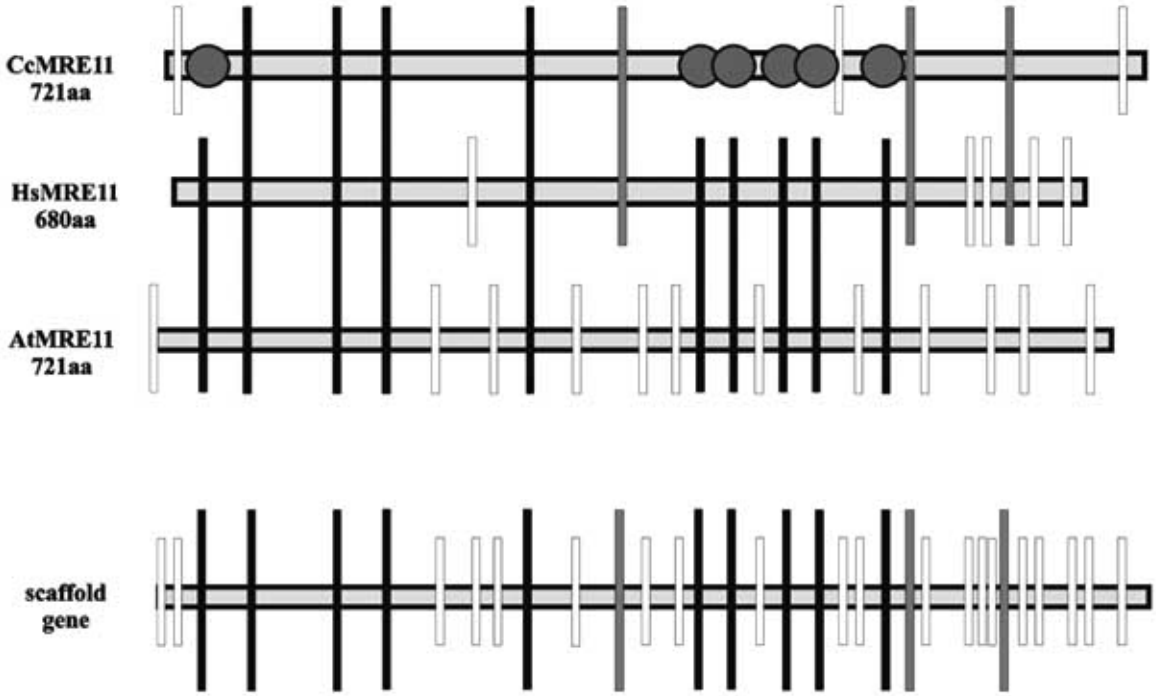

Fig. (4). Intron position comparison across three kingdoms. The intron positions of the MRE11 gene from a fungi (*Coprinus cinereus* = Cc), an animal (*Homo sapiens* = Hs) and a plant (*Arabidopsis thaliana* = At) is shown schematically. The respective intron positions of the genomic sequence of each organisms gene has been transferred to an alignment of the protein sequences. If an intron is located in the same position in all three organisms or plants and animals only it is regarded as ancient (shown as black bar). Ancient means that the intron was already existing before the divergence of animals and plants approximately 1.000 mya. If an intron is present at the same position in animals and fungi only it is considered as semi-ancient (shown as grey bar). An intron which is only present in one of the three organisms is defined as a recent gain event (shown as white bar). An intron lacking in the fungus but present in animals and plants demonstrates an intron loss event and is shown as grey circle. Below the three genes a constructed scaffold gene composed out of all existing intron positions is depicted. 30% of all intron positions in the MRE11 scaffold gene are regarded as ancient (modified, according to [13]).

intron positions during evolution [13]. Therefore, it is most likely that an intronless mRNA was after reverse transcription reintegrated into the moss genome and the intron bearing copy was lost subsequently by a gene conversion event [128]. This is a common mechanism for intron-loss but it does not explain why otherwise highly conserved introns of the RAD51 genes are absent and therefore unnecessary in *Physcomitrella patens*.

In contrast to higher plants it has been shown that *Physcomitrella* has an unusual high efficiency of homologous recombination (HR) in *Physcomitrella* [130-133]. Whether the enhanced HR is due to specific additional moss genes or due to specific genes of the NHEJ pathway missing in moss is a very interesting question, which can be addressed best by comparative genomic analysis. It has been speculated that the enhanced HR frequency in the moss is due to the specific synchronised cell cycle stage of the moss culture. The protoplasts, which are obtained from these cultures and always used for transformation are nearly all arrested at G2/M [134]. This arrest might indeed be correlated with a specific pattern of expression of HR genes in this stage. Whether this speculation is correct or a qualitative difference(s) in the recombination machinery of *Physcomitrella* is the reason for enhanced HR frequencies in moss can only be shown by a full genome comparison including a transcriptome analysis of the respective cell stage.

GENES NOT FOUND BY COMPARATIVE GENOMICS

Finally, several of the genes which are important for recombination or repair have still not been found in some organisms, neither by comparative genomics nor by other methods. For example, in plants, homologs of p53, RAD52, RAD18 and DNA-PK have not been found. Of course this could mean that these genes have been lost in one organism during evolution but it can also mean that the function of these genes has been invented "twice". Therefore, functional homologues without conservation of their sequence might be present in the respective genomes. The evolution of topoisomerases is a good example for the latter case. During evolution the type II topoisomerases probably have been invented twice and persist in nowadays living organisms [135]. Secondly, the yeast XRS2 and human NBS1 protein are both important factors of the MRE11/RAD50 complex in yeast and humans but show hardly any sequence homology to each other except a domain at the N-terminus known as forkhead associated domain [136]. In such cases it is nearly impossible to discover these genes by genomic comparisons. However, it might be possible to find such non conserved

genes by interaction studies or functional mutant screens as it has been done in mammals and yeast [48].

CONCLUSIONS

The above chosen examples demonstrated very nicely the potential of comparative genomics even across long evolutionary distances like they are between *Physcomitrella* and *Arabidopsis*. With this method we can learn a lot about evolution and structural conservation of genes. But this can only be the starting point for more precise investigations addressing the question why the gene of interest has been conserved during evolution. Of course a high degree of evolutionary conservation is a good argument that a specific gene has an important function in a given pathway.

Coming back to the initial question what comparative genomics tells us about the genes of the recombination machinery, we can say that most of the genes, which are involved in recombination can be found easily by means of comparative genomics in different organisms because this pathway is a very old and conserved one. Finally, a very exciting and important output of the genome sequence data and the deduced protein sequences is their usefulness for the identification of the parts of a protein, which specify its function, "the functional core" of a protein. To elaborate this, we can expect to define by sequence comparisons in each position of a protein one of the following types of amino acids, i) absolutely conserved; ii) structurally conserved and iii) varying amino acids. Most amino acids of a protein are not important for its function and will therefore be of type iii). The opportunity to align a growing number of homologous sequences enables us to identify more precisely the really conserved and therefore functional important regions of a protein. The same kind of sequence analysis gives us a very nice idea about the structural requirements of a protein, which is often correlated with the presence of synonymous amino acids in the primary sequence. Therefore, we can draw the obvious conclusion that, the more genomes (especially such ones with a reasonable evolutionary distance) will be completely sequenced the more complete the identification and understanding of the conserved and variable areas of specific proteins will be.

ACKNOWLEDGEMENTS

This work was partly funded by grants of the Deutsche Forschungsgemeinschaft and the European Union to H. P. The authors would like to thank Anja Ipsen and Willem Reidt for critical reading of the manuscript.

REFERENCES

[1] Ureta-Vidal, A., Ettwiller, L., Birney, E. Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat. Rev. Genet.*, **2003**, *4*: 251-262.

[2] Bennetzen, J.L. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell*, **2000,** *12*: 1021-1029.

[3] Liu, H., Sachidanandam, R., Stein L. Comparative genomics between rice and Arabidopsis shows scant collinearity in gene order. *Genome Res.*, **2001**, *11*: 2020-2026.

[4] Ramakrishna, W., Dubcovsky, J., Park, Y.J., Busso, C., Emberton, J., SanMiguel, P., Bennetzen, J.L. Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. *Genetics*, **2002**, *162*: 1389-1400.

[5] Grant, D., Cregan, P., Shoemaker, R.C. Genome organization in dicots: Genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*: 4168-4173.

[6] Ku, H.M., Vision, T., Liu, J., Tanksley, S.D. Comparing sequenced segments of the tomato and Arabidopsis genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*: 9121-9126.

******[7] Salse, J., Piegu, B., Cooke, R., Delseny, M. Synteny between Arabidopsis thaliana and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Res.*, **2002**, *30*: 2316-2328.
This is the first broad scaled comparison of the rice and Arabidopsis genome showing a significant decrease in microcollinearity compared to earlier studies.

[8] Babula, D., Kaczmarek, M., Barakat, A., Delseny, M., Quiros, C.F., Sadowski, J. Chromosomal mapping of Brassica oleracea based on ESTs from Arabidopsis thaliana: complexity of the comparative map. *Mol. Genet. Genomics*, **2003**, *268*: 656-665.

*****[9] Brendel, V., Kurtz, S., Walbot, V. Comparative genomics of Arabidopsis and maize: prospects and limitations. *Genome Biol.* **2002**, *3*: reviews 1005.1-1005.6.
This review summarises the occuring problems of comparative genomics between maize and Arabidopsis.

[10] Hardison, R.C. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.*, **2000**, *16*: 369-372.

[11] Ludwig, M.Z. Functional evolution of noncoding DNA. *Curr. Opin. Genet. Dev.*, **2002**, *12*: 634-639.

[12] Boudet, N., Aubourg, S., Toffano-Nioche, C., Kreis, M., Lecharny, A. Evolution of intron/exon structure of DEAD helicase family genes in Arabidopsis, Caenorhabditis, and Drosophila. *Genome Res.*, **2001**, *11*: 2101-2114.

*****[13] Hartung, F., Puchta, H. Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery. *Nucleic Acids Res.*, **2002**, *30*: 5175-5181.
A concise study of the conservation of the exon/intron structure between plants animals and fungi.

*****[14] Fedorov, A., Merican, A.F., Gilbert, W. Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*: 16128-16133.
A broad scaled analysis of intron positions in plants animal and fungi showing an average of 17% conserved intron positions.

[15] Roy, S.W., Fedorov, A., Gilbert, W. Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*: 7158-7162.

[16] Trapp, S.C., Croteau, R.B. Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics*, **2001**, *158*: 811-832.

[17] Lecharny, A., Boudet, N., Gy, I., Aubourg, S., Kreis, M. Introns in, introns out in plant gene families: a genomic approach of the dynamics of gene structure. *J. Struct. Funct. Genomics*, **2003**, *3*: 111-116.

[18] Sekelsky, J.J., Brodsky, M.H., Rubin, G.M., Hawley, R.S. Drosophila and human RecQ5 exist in different isoforms generated by alternative splicing. *Nucleic Acids Res.*, **1999**, *27*: 3762-3769.

[19] West, S.C. Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.*, **1992**, *61*: 603-640.

[20] Clark, A.J., Sandler, S.J. Homologous genetic recombination: the pieces begin to fall into place. *Crit. Rev. Microbiol.*, **1994**, *20*: 125-142.

[21] Camerini-Otero, R.D., Hsieh, P. Homologous recombination proteins in prokaryotes and eukaryotes. *Annu. Rev. Genet.*, **1995**, *29*: 509-552.

[22] Kowalczykowski, S.C. Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange. *Annu. Rev. Biophys. Biophys. Chem.*, **1991**, *20*: 539-575.

[23] Sung, P., Robberson, D.L. DNA strand exchange mediated by a RAD51-ssDNA nucleoprotein filament with polarity opposite to that of RecA. *Cell*, **1995**, *82*: 453-461.

[24] Ogawa, T., Yu, X., Shinohara, A., Egelman, E.H. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science*, **1993**, *259*: 1896-1899.

[25] Sung, P. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science*, **1994**, *265*: 1241-1243.

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

[27] Stassen, N.Y., Logsdon, J.M. Jr., Vora, G.J., Offenberg, H.H., Palmer, J.D., Zolan, M.E. Isolation and characterization of rad51 orthologs from Coprinus cinereus and Lycopersicon esculentum, and phylogenetic analysis of eukaryotic recA homologs. *Curr. Genet.*, **1997**, *31*: 144-157.

[28] Haber, J.E. The many interfaces of Mre11. *Cell*, **1998**, *95*: 583-586.

[29] Sharples, G.J., Leach, D.R. Structural and functional similarities between the SbcCD proteins of Escherichia coli and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. *Mol. Microbiol.*, **1995**, *17*: 1215-1217.

[30] Connelly, J.C., Kirkham, L.A., Leach, D.R. The SbcCD nuclease of Escherichia coli is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc. Natl. Acad. Sci. USA*, **1998**, *95*: 7969-7974.

[31] Connelly, J.C., de Leau, E.S., Leach, D.R. DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. *Nucleic Acids Res.*, **1999**, *27*: 1039-1046.

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

[33] Petrini, J.H., Walsh, M.E., DiMare, C., Chen, X.N., Korenberg, J.R., Weaver, D.T. Isolation and characterization of the human MRE11 homologue. *Genomics*, **1995**, *29*: 80-86.

[34] Dolganov, G.M., Maser, R.S., Novikov, A., Tosto, L., Chong, S., Bressan, D.A., Petrini, J.H. Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. *Mol. Cell Biol.*, **1996**, *16*: 4832-4841.

[35] Hartung, F., Puchta, H. Isolation of the complete cDNA of the Mre11 homologue of Arabidopsis (accession no. AJ243822) indicates conservation of DNA recombination mechanisms between plants and other eukaryotes (PGR-99-132). *Plant Physiol.*, **1999**, *121*: 312.

[36] Daoudal-Cotterell, S., Gallego, M.E., White, C.I. The plant Rad50-Mre11 protein complex. *FEBS Lett.*, **2002**, *516*: 164-166.

*****[37] Bundock, P., Hooykaas, P. Severe developmental defects, hypersensitivity to DNA-damaging agents, and lengthened telomeres in Arabidopsis MRE11 mutants. *Plant Cell*, **2002**, *14*: 2451-2462.
The first description of the severe but viable phenotype of a knockout plant for MRE11, showing also its involvement in telomere turnover.

[38] Usui, T., Ohta, T., Oshiumi, H., Tomizawa, J., Ogawa, H., Ogawa, T. Complex formation and functional versatility of Mre11 of budding yeast in recombination. *Cell*, **1998**, *95*: 705-716.

[39] Furuse, M., Nagase, Y., Tsubouchi, H., Murakami-Murofushi, K., Shibata, T., Ohta, K. Distinct roles of two separable *in vitro* activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J.*, **1998**, *17*: 6412-6425.

[40] Paull, T.T., Gellert, M. The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol. Cell.*, **1998**, *1*: 969-979.

[41] Hopfner, K.P., Karcher, A., Craig, L., Woo, T.T., Carney, J.P., Tainer, J.A. Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell*, **2001**, *105*: 473-485.

******[42] D'Amours, D., Jackson, S.P. The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat. Rev. Mol. Cell. Biol.*, **2002**, *3*: 317-327.
The authors summarise and discuss in depth the various functions of the MRE11 complex.

[43] de Jager, M., van Noort, J., van Gent, D.C., Dekker, C., Kanaar, R., Wyman, C. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol. Cell.*, **2001**, *8*: 1129-1135.

[44] Chen, L., Trujillo, K., Ramos, W., Sung, P., Tomkinson, A.E. Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. *Mol. Cell.*, **2001**, *8*: 1105-1115.

[45] Hopfner, K.P., Karcher, A., Shin, D., Fairley, C., Tainer, J.A., Carney, J.P. Mre11 and Rad50 from Pyrococcus furiosus: cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine. *J. Bacteriol.*, **2000**, *182*: 6036-6041.

[46] Hopfner, K.P., Karcher, A., Shin, D.S., Craig, L., Arthur, L.M., Carney, J.P., Tainer, J.A. Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell*, **2000**, *101*: 789-800.

[47] Ivanov, E.L., Korolev, V.G., Fabre, F. XRS2, a DNA repair gene of Saccharomyces cerevisiae, is needed for meiotic recombination. *Genetics*, **1992**, *132*: 651-664.

[48] Carney, J.P., Maser, R.S., Olivares, H., Davis, E.M., Le Beau, M., Yates, J.R.. 3rd, Hays, L., Morgan, W.F. and Petrini, J.H. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell*, **1998**, *93*: 477-486.

[49] Paull, T.T., Gellert, M. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev.*, **1999**, *13*: 1276-1288.

[50] Dong, Z., Zhong, Q., Chen, P.L. The Nijmegen breakage syndrome protein is essential for Mre11 phosphorylation upon DNA damage. *J. Biol. Chem.*, **1999**, *274*: 19513-19516.

[51] Petrini, J.H. The Mre11 complex and ATM: collaborating to navigate S phase. *Curr. Opin. Cell Biol.*, **2000**, *12*: 293-296.

[52] Paull, T.T. New glimpses of an old machine. *Cell*, **2001**, *107*: 563-565.

[53] Connelly, J.C., Leach, D.R. Tethering on the brink: the evolutionary conserved Mre11-Rad50 complex. *Trends Biochem. Sci.*, **2002**, *27*: 410-418.

[54] Gherbi, H., Gallego, M.E., Jalut, N., Lucht, J.M., Hohn, B., White, C.I. Homologous recombination in planta is stimulated in the absence of Rad50. *EMBO Rep.*, **2001**, *2*: 287-291.

[55] Moore, J.K., Haber, J.E. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **1996**, *16*: 2164-2173.

[56] Bressan, D.A., Baxter, B.K., Petrini, J.H. The Mre11-Rad50-Xrs2 protein complex facilitates homologous recombination-based double-strand break repair in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **1999**, *19*: 7681-7687.

[57] Gerecke, E.E., Zolan, M.E. An mre11 mutant of Coprinus cinereus has defects in meiotic chromosome pairing, condensation and synapsis. *Genetics*, **2000**, *154*: 1125-1139.

[58] Nugent, C.I., Bosco, G., Ross, L.O., Evans, S.K., Salinger, A.P., Moore, J.K., Haber, J.E., Lundblad, V. Telomere maintenance is dependent on activities required for end repair of double-strand breaks. *Curr. Biol.*, **1999**, *8*: 657-660.

[59] Ritchie, K.B., Petes, T.D. The Mre11p/Rad50p/Xrs2p complex and the Tel1p function in a single pathway for telomere maintenance in yeast. *Genetics*, **2000**, *155*: 475-479.

[60] Zhu, X.D., Kuster, B., Mann, M., Petrini, J.H., de Lange, T. Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nat. Genet.*, **2000**, *25*: 347-352.

[61] Lombard, D.B., Guarente, L. Nijmegen breakage syndrome disease protein and MRE11 at PML nuclear bodies and meiotic telomeres. *Cancer Res.*, **2000**, *60*: 2331-2334.

[62] Gallego, M.E., White, C.I. RAD50 function is essential for telomere maintenance in Arabidopsis. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*: 1711-1716.

[63] Chamankhah, M., Xiao, W. Formation of the yeast Mre11-Rad50-Xrs2 complex is correlated with DNA repair and telomere maintenance. *Nucleic Acids Res.*, **1999**, *27*: 2072-2079.

[64] Wilson, S., Warr, N., Taylor, D.L., Watts, F.Z. The role of Schizosaccharomyces pombe Rad32, the Mre11 homologue, and other DNA damage response proteins in non-homologous end joining and telomere length maintenance. *Nucleic Acids Res.*, **1999**, *27*: 2655-2661.

[65] Zhong, Q., Chen, C.F., Li, S., Chen, Y., Wang, C.C., Xiao, J., Chen, P.L., Sharp, Z.D., Lee, W.H. Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. *Science*, **1999**, *285*: 747-750.

[66] Costanzo, V., Robertson, K., Bibikova, M., Kim, E., Grieco, D., Gottesman, M., Carroll, D., Gautier, J. Mre11 protein complex prevents double-strand break accumulation during chromosomal DNA replication. *Mol. Cell.*, **2001**, *8*: 137-147.

[67] Desai-Mehta, A., Cerosaletti, K.M., Concannon, P. Distinct functional domains of nibrin mediate Mre11 binding, focus formation, and nuclear localization. *Mol. Cell Biol.*, **2001**, *21*: 2184-2191.

[68] Mirzoeva, O.K., Petrini, J.H. DNA damage-dependent nuclear dynamics of the Mre11 complex. *Mol. Cell. Biol.*, **2001**, *21*: 281-288.

[69] Maser, R.S., Mirzoeva, O.K., Wells, J., Olivares, H., Williams, B.R., Zinkel, R.A., Farnham, P.J., Petrini, J.H. Mre11 complex and DNA replication: linkage to E2F and sites of DNA synthesis. *Mol. Cell. Biol.*, **2001**, *21*: 6006-6016.

[70] Matsuura, S., Tauchi, H., Nakamura, A., Kondo, N., Sakamoto, S., Endo, S., Smeets, D., Solder, B., Belohradsky, B.H., Der Kaloustian, V.M., Oshimura, M., Isomura, M., Nakamura, Y., Komatsu, K. Positional cloning of the gene for Nijmegen breakage syndrome. *Nat. Genet.*, **1998**, *19*: 179-181.

[71] Hanada, K., Ukita, T., Kohno, Y., Saito, K., Kato, J., Ikeda, H. RecQ DNA helicase is a suppressor of illegitimate recombination in Escherichia coli. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*: 3860-3865.

[72] Cogoni, C., Macino, G. Posttranscriptional gene silencing in Neurospora by a RecQ DNA helicase. *Science*, **1999**, *286*: 2342-2344.

******[73] Hickson, I.D. RecQ helicases: caretakers of the genome. *Nat. Rev. Cancer*, **2003**, *3*: 169-178.
A concise reflection on the function of mammalian RecQ homologues in the maintenance of genomic stability.

[74] Hartung, F., Plchova, H., Puchta, H. Molecular characterisation of RecQ homologues in Arabidopsis thaliana. *Nucleic Acids Res.*, **2000**, *28*: 4275-4282.

[75] Ellis, N.A., Groden, J., Ye, T.Z., Straughen, J., Lennon, D.J., Ciocci, S., Proytcheva, M., German, J. The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell*, **1995**, *83*: 655-666.

[76] Yu, C.E., Oshima, J., Fu, Y.H., Wijsman, E.M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G.M., Mulligan, J., Schellenberg, G.D. Positional cloning of the Werner's syndrome gene. *Science*, **1996**, *272*: 258-262.

[77] Kitao, S., Shimamoto, A., Goto, M., Miller, R.W., Smithson, W.A., Lindor, N.M., Furuichi, Y. Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. *Nat. Genet.*, **1999**, *22*: 82-84.

[78] Shen, J.C., Loeb, L.A. The Werner syndrome gene: the molecular basis of RecQ helicase-deficiency diseases. *Trends Genet.*, **2000**, *16*: 213-220.

*****[79] Wang, W., Seki, M., Narita, Y., Nakagawa, T., Yoshimura, A., Otsuki, M., Kawabe, Y., Tada, S., Yagi, H., Ishii, Y., Enomoto, T. Functional relation among RecQ family helicases RecQL1, RecQL5, and BLM in cell growth and sister chromatid exchange formation. *Mol. Cell. Biol.*, **2003**, *23*: 3527-3535.
A very interesting work about the two RecQlike genes which alone do not exhibit any impairment if they are mutated but might act in the same pathway as the *blm* protein.

******[80] Bernstein, D.A., Keck, J.L. Domain mapping of Escherichia coli RecQ defines the roles of conserved N- and C-terminal regions in the RecQ family. *Nucleic Acids Res.*, **2003**, *31*: 2778-2785.
An excellent in depth investigation of the *E. coli* RecQ protein domains, revealing a new motif in the helicase domain. This area named motif 0 seems to be responsible for transient DNA binding.

[81] Morozov, V., Mushegian, A.R., Koonin, E.V., Bork, P. A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases. *Trends Biochem. Sci.*, **1997**, *22*: 417-418.

[82] Shimamoto, A., Nishikawa, K., Kitao, S., Furuichi, Y. Human RecQ5beta, a large isomer of RecQ5 DNA helicase, localizes in the nucleoplasm and interacts with topoisomerases 3alpha and 3beta. *Nucleic Acids Res.*, **2000**, *28*: 1647-1655.

[83] Kawasaki, K., Maruyama, S., Nakayama, M., Matsumoto, K., Shibata, T. Drosophila melanogaster RECQ5/QE DNA helicase: stimulation by GTP binding. *Nucleic Acids Res.*, **2002**, *30*: 3682-3691.

[84] Ozsoy, A.Z., Ragonese, H.M., Matson, S.W. Analysis of helicase activity and substrate specificity of Drosophila RECQ5. *Nucleic Acids Res.*, **2003**, *31*: 1554-1564.

[85] Yan, H., Chen, C.Y., Kobayashi, R., Newport, J. Replication focus-forming activity 1 and the Werner syndrome gene product. *Nat. Genet.*, **1998**, *19*: 375-378.

[86] Shen, J.C., Gray, M.D., Oshima, J., Kamath-Loeb, A.S., Fry, M., Loeb, L.A. Werner syndrome protein. I. DNA helicase and dna exonuclease reside on the same polypeptide. *J. Biol. Chem.*, **1998**, *273*: 34139-34144.

[87] Machwe, A., Xiao, L., Theodore, S., Orren, D.K. DNase I footprinting and enhanced exonuclease function of the bipartite Werner syndrome protein (WRN) bound to partially melted duplex DNA. *J. Biol. Chem.*, **2002**, *277*: 4492-4504.

[88] Xue, Y., Ratcliff, G.C., Wang, H., Davis-Searles, P.R., Gray, M.D., Erie, D.A., Redinbo, M.R. A minimal exonuclease domain of WRN forms a hexamer on DNA and possesses both 3'- 5' exonuclease and 5'-protruding strand endonuclease activities. *Biochemistry*, **2002**, *41*: 2901-2912.

[89] Huang, S, Li, B., Gray, M.D., Oshima, J., Mian, I.S., Campisi, J. The premature ageing syndrome protein, WRN, is a 3'-->5' exonuclease. *Nat. Genet.*, **1998**, *20*: 114-116.

[90] Shen, J.C., Gray, M.D., Oshima, J., Loeb, L.A. Characterization of Werner syndrome protein DNA helicase activity: directionality, substrate dependence and stimulation by replication protein A. *Nucleic Acids Res.*, **1998**, *26*: 2879-2885.

[91] Kamath-Loeb, A.S., Shen, J.C., Loeb, L.A., Fry, M. Werner syndrome protein. II. Characterization of the integral 3' --> 5' DNA exonuclease. *J. Biol. Chem.*, **1998**, *273*: 34145-34150.

[92] Suzuki, N., Shiratori, M., Goto, M., Furuichi, Y. Werner syndrome helicase contains a 5'-->3' exonuclease activity that digests DNA and RNA strands in DNA/DNA and RNA/DNA duplexes dependent on unwinding. *Nucleic Acids Res.*, **1999**, *27*: 2361-2368.

[93] Bagherieh-Najjar, M.B. de Vries, O.M.M., Kroon, J.T.M., Wright, E.L., Elborough, K.M., Hille, J., Dijkwel, P.P. *Arabidopsis* RecQsim, a plant-specific member of the RecQ helicase family, can suppress the MMS hypersensitivity of the yeast *sgs1* mutant. *Plant Mol. Biol.*, **2003**, *52*: 273-284.

[94] Acarkan, A., Rossberg, M., Koch, M., Schmidt, R. Comparative genome analysis reveals extensive conservation of genome organisation for Arabidopsis thaliana and Capsella rubella. *Plant J.*, **2000**, *23*: 55-62.

[95] Ku, H.-M., Vision, T. Liu, J., Tanksley, S. D. Comparing sequenced segments of the tomato and Arabidopsis genomes: Large- scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*: 9121-9126.

[96] Vision, T.J., Brown, D.G., Tanksley, S.D. The origins of genomic duplications in Arabidopsis. *Science*, **2000**, *290*: 2114-2117.

[97] Harmon, F.G., Kowalczykowski, S.C. RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev.*, **1998**, *12*: 1134-1144.

[98] Sun, H., Karow, J.K., Hickson, I.D., Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.*, **1998**, *273*: 27587-27592.

[99] Fry, M., Loeb, L.A. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.*, **1999**, *274*: 12797-12802.

[100] Constantinou, A., Tarsounas, M., Karow, J.K., Brosh, R.M., Bohr, V.A., Hickson, I.D., West, S.C. Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep.*, **2000**, *1*: 80-84.

[101] Brosh, R.M. Jr., Majumdar, A., Desai, S., Hickson, I.D., Bohr, V.A., Seidman, M.M. Unwinding of a DNA triple helix by the Werner and Bloom syndrome helicases. *J. Biol. Chem.*, **2001**, *276*: 3024-3030.

[102] Lebel, M., Spillare, E.A., Harris, C.C., Leder, P. The Werner syndrome gene product co-purifies with the DNA replication complex and interacts with PCNA and topoisomerase I. *J. Biol. Chem.*, **1999**, *274*: 37795-37799.

[103] Lee, S.K., Johnson, R.E., Yu, S.L., Prakash, L., Prakash, S. Requirement of yeast SGS1 and SRS2 genes for replication and transcription. *Science*, **1999**, *286*: 2339-2342.

[104] Gangloff, S., Soustelle, C., Fabre, F. Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases. *Nat. Genet.*, **2000**, *25*: 192-194.

[105] Huang, P., Pryde, F.E., Lester, D., Maddison, R.L., Borts, R.H., Hickson, I.D., Louis, E.J. SGS1 is required for telomere elongation in the absence of telomerase. *Curr. Biol.*, **2001**, *11*: 125-129.

[106] Johnson, F.B., Marciniak, R.A., McVey, M., Stewart, S.A., Hahn, W.C., Guarente, L. The Saccharomyces cerevisiae WRN homolog

Sgs1p participates in telomere maintenance in cells lacking telomerase. *EMBO J.*, **2001**, *20*: 905-913.

[107] Saintigny, Y., Makienko, K., Swanson, C., Emond, M.J., Monnat, R.J. Jr. Homologous recombination resolution defect in werner syndrome. *Mol. Cell. Biol.*, **2002**, *22*: 6971-6978.

[108] Adams, M.D., McVey, M., Sekelsky, J.J. Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. *Science*, **2003**, *299*: 265-267.

[109] Atcheson, C.L., DiDomenico, B., Frackman, S., Esposito, R.E., Elder, R.T. Isolation, DNA sequence, and regulation of a meiosis-specific eukaryotic recombination gene. *Proc. Natl. Acad. Sci. USA*, **1987**, *84*: 8035-8039.

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

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

[112] Nichols, M.D., DeAngelis, K., Keck, J.L., Berger, J.M. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J.*, **1999**, *18*: 6177-6188.

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

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

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

*****[116] Hartung, F., Angelis, K.J., Meister, A., Schubert, I., Melzer, M., Puchta, H. An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Curr. Biol.*, **2002**, *12*: 1787-1791.
This study, together with [118], shows the involvement of topoisomerase VI in plant endoreduplication and its urgent need for proper cell growth.

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

*****[118] Sugimoto-Shirasu, K., Stacey, N.J., Corsar, J., Roberts, K., McCann, M.C. DNA topoisomerase VI is essential for endoreduplication in Arabidopsis. *Curr. Biol.*, **2002**, *12*: 1782-1786.
See comment on [116].

[119] Corbett, K.D., Berger, J.M. Emerging Roles for Plant Topoisomerase VI. *Chem. Biol.*, **2003**, *10*: 107-111.

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

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

[122] Romanienko, P.J., Camerini-Otero, R.D. The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol. Cell.*, **2000**, *6*: 975-987.

[123] Celerin, M., Merino, S.T., Stone, J.E., Menzie, A.M., Zolan, M.E. Multiple roles of Spo11 in meiotic chromosome behavior. *EMBO J.*, **2000**, *19*: 2739-2750.

[124] Hardison, R.C. A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*: 5675-5679.

[125] Sogin, M.L. Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.*, **1991**, *1*: 457-63.

[126] Dacks, J.B., Doolittle, W.F. Reconstructing/Deconstructing the Earliest Eukaryotes: How Comparative Genomics Can Help. *Cell*, **2001**, *107*: 419-425.

*****[127] Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y., Hasebe, M. Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: Implication for land plant evolution. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*: 8007-8012.
The first broad comparison study between moss and Arabidopsis exemplifies the feasibility of comparative studies between this distantly related plants.

[128] Markmann-Mulisch, U., Hadi, M.Z., Koepchen, K., Alonso, J.C., Russo, V.E., Schell, J., Reiss, B. The organization of Physcomitrella patens RAD51 genes is unique among eukaryotic organisms. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*: 2959-2964.

[129] Kawabata, M., Saeki, K. Multiple alternative transcripts of the human homologue of the mouse TRAD/R51H3/RAD51D gene, a member of the recA/rad51 gene family. *Biochem. Biophys. Res. Commun.*, **1999**, *257*: 156-162.

[130] Schaefer, D.G., Zryd, J.P. Efficient gene targeting in the moss Physcomitrella patens. *Plant J.*, **1997**, *11*: 1195-1206.

[131] Strepp, R., Scholz, S., Kruse, S., Speth, V., Reski, R. Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proc. Natl. Acad. Sci. USA*, **1998**, *95*: 4368-4373.

[132] Schaefer, D.G., Zryd, J.P. The moss Physcomitrella patens, now and then. *Plant Physiol.*, **2001**, *127*: 1430-1438.

[133] Puchta, H. Gene replacement by homologous recombination in plants. *Plant Mol. Biol.*, **2002**, *48*: 173-182.

[134] Schween, G., Gorr, G., Hohe, A., Reski, R. Unique tissue-specific cell cycle in Physcomitrella. *Plant Biology*, **2003**, *5*: 50-58.

******[135] Gadelle, D., Filee, J., Buhler, C., Forterre, P. Phylogenomics of type II DNA topoisomerases. *Bioessays*, **2003**, *25*: 232-242.
An excellent review concerning about the evolution of topoisomerases.

[136] Tauchi, H., Kobayashi, J., Morishima, K., Matsuura, S., Nakamura, A., Shiraishi, T., Ito, E., Masnada, D., Delia, D., Komatsu, K. The forkhead-associated domain of NBS1 is essential for nuclear foci formation after irradiation but not essential for hRAD50/hMRE11/NBS1 complex DNA repair activity. *J. Biol. Chem.*, **2001**, *276*: 12-15.

[137] Cromie, G.A., Connelly, J.C., Leach, D.R. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol. Cell*, **2001**, *8*: 1163-1174.

[138] Plchova, H., Hartung, F., Puchta, H. Biochemical characterization of a exonuclease from Arabidopsis thaliana reveals similarities to the DNA exonuclease of the human Werner protein. *J. Biol. Chem.*, **2003**, *278*: 44128-44138.