Genome wide cloning of maize meiotic recombinase Dmc1 and its functional structure through molecular phylogeny

F. Etedali[1], B. Baghban Kohnehrouz[1], M. Valizadeh[1], A. Gholizadeh[2] and M.A. Malboobi[3]

[1]Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
[2]Research Institute for Fundamental Sciences, University of Tabriz, Tabriz, Iran
[3]National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

Corresponding author: B. Baghban Kohnehrouz
E-mail: bahramrouz@yahoo.com

ABSTRACT. The development of meiotic division and associated genetic recombination paved the way for evolutionary changes. However, the secondary and tertiary structure and functional domains of many of the proteins involved in genetic recombination have not been studied in detail. We used the human *Dmc1* gene product along with secondary and tertiary domain structures of *Escherichia coli* RecA protein to help determine the molecular structure and function of maize Dmc1, which is required for synaptonemal complex formation and cell cycle progression. The maize recombinase *Dmc1* gene was cloned and characterized, using rice *Dmc1* cDNA as an orthologue. The deduced amino acid sequence was used for

elaborating its 3-D structure, and functional analysis was made with the CDD software, showing significant identity of the *Dmc1* gene product in *Zea mays* with that of *Homo sapiens*. Based on these results, the domains and motives of WalkerA and WalkerB as ATP binding sites, a multimer site (BRC) interface, the putative ssDNA binding L1 and L2 loops, the putative dsDNA binding helix-hairpin-helix, a polymerization motif, the subunit rotation motif, and a small N-terminal domain were proposed for maize recombinase Dmc1.

Key words: Maize Dmc1; Functional structure; Meiotic recombinase; Phylogeny

INTRODUCTION

Increasingly knowledge on the molecular mechanisms like genetic recombination opens new insights on understanding the evolution and their implications. Proteins play a major part in such mechanisms and their conservation across various genomes and/or mutational changes to fit the environment highlights the role of importance of the protein in the functioning of the cell. The knowledge in public domain shows that different proteins may display quite varied patterns of similarity relationships due to a host of confounding factors (Smith et al., 1992), conflicting taxonomic schemes and unresolved changes concerning evolutionary relationships (Keeling and Doolittle, 1995). Since genetic recombination is considered as a common and necessary factor to introduce such changes and paving the way for evolution, understanding the mechanisms of recombination and the process will unravel new insights. The genes implicated in the recombination process that are well studied include bacterial *RecA*, archaeal *RadA* or *Rad51*, eukaryotic *Rad51*, and meiosis-specific *Dmc1*. The protein homology studies of archaebacteria RadA have demonstrated more similarities to eukaryotic Rad51 and Dmc1 rather than prokaryotic RecA (Sandler et al., 1996). It is commonly accepted that families of related proteins generate essentially the same 3-D structure (Pellegrini et al., 2002). The RecA family of proteins is generally involved in homologous recombination, maintenance of genomic stability by protecting against DNA double-strand breaks during mitosis and meiosis through non-homologous end-joining or by homologous recombination (Chen et al., 2007) and repairing reciprocal crossovers (Sun et al., 1991). Dmc1 as a requirement of synaptonemal complex formation and cell cycle progression (Bishop et al., 1992; Metkar et al., 2004) has been particularly reported as a single-strand DNA binding to form extensive helical nucleoprotein filaments called presynaptic complex with both ATPase and DNA strand exchange activities (Hong et al., 2001; Chen et al., 2007).

Plant Dmc1 orthologues have been identified in *Lilium longiflorum* (Kobayashi et al., 1993), *Arabidopsis thaliana* (Sato et al., 1995; De Muyt et al., 2009), *Oryza sativa* (Ding et al., 2001; Metkar et al., 2004; Deng and Wang, 2007; Sakane et al., 2008). Due to a lack of information on maize we targeted the cloning of maize Dmc1 3' end exon and used this sequence to validate the full-length Dmc1 cDNA obtained by *in silico* cloning. Deduced amino acid sequences were used in phylogenetic analysis across prokarya, archaebacteria and eukarya. We also discuss implications in terms of structure and function

of maize Dmc1 with emphasis on further necessary experiments to elucidate in detail the structure and function of Dmc1 in meiotic recombination in plants.

MATERIAL AND METHODS

Identification of the *Zea mays Dmc1* gene by EST database mining

We used the rice *Dmc1* sequences for homology search of maize expressed sequence tags (ESTs) available at the National Center for Biotechnology Information (NCBI, http://blast.ncbi.nlm.nih.gov/Blast.cgi), (http://www.Maizegdb.org) and (http://plantgdb.org) to access the related 3' and 5' ESTs. *In silico* analysis was conducted to assemble and identify the complete and non-redundant *Dmc1* cDNA. In this analysis, the relevant ESTs were aligned either pairwise or multiple by ClustalW2 (Thompson et al., 1994).

Partial *Dmc1* gene cloning

The partial 3' end of the *Dmc1* gene including last exon containing stop codon and poly-adenylation signal sequence was isolated from maize SC 704 by polymerase chain reaction (PCR) amplification using two specific primers, F1 (5'-gttttccaggttacatca ggtg-3') and R1 (5'-agaacagtgttatttatacgcaagca-3') under thermo-cycler program, which consisted of 5 min at 94°C for initial denaturing, 30 cycles for 1 min at 94°C, 30 s at 65°C, 1 min at 72°C, and finally 10 min at 72°C for extension. The obtained fragment was cloned into pGEM®-T-easy vector (Promega, Madison, WI, USA) and sequenced by T7 primer.

3'*Dmc1* cDNA cloning

To access *Dmc1* cDNA, the young leaves (as control), immature husk and tassel of maize (as developmental specific tissues) were used for total RNA extraction by the Trizol method. First-strand cDNA was synthesized using oligo d[T] and ReverTra (Fermentas GmbH, St. Leon-Rot, Germany). The products of these reactions were used as template for PCR amplification with maize *Dmc1* cDNA-specific primers F1 and R1 corresponding to the last exon containing stop codon and poly-adenylation signal sequence. The amplified fragments were cloned into pGEMT-easy vector (Promega) and sequenced by *T7* primer. To normalize the amount of cDNA input the amplified *tubulin TubA* cDNA was used as an internal control (Zhang et al., 2006).

Phylogenic sequence analysis

The *Dmc1* connected files were searched out and downloaded from NCBI and used for phylogeny analysis along with obtained full-length cDNA sequence from maize based on nucleotide and amino acid sequences. Preliminary results from BLASTn, tBLASTn, tBLASTx, and BLASTp searches showed that amino acid sequences could be more useful and productive than nucleotide sequences in clustering at high resolution level across

different genus and species of major kingdoms of living organisms (data not shown). We started with about 1000 Dmc1/RecA-like protein accession numbers and only a member from 81 genera with complete sequence (as appeared in footnote[1]) was chosen to carry out a multiple alignment among them using ClustalW2. A phylogenetic tree was constructed using the GeneBee Tree Top phylogenetic tree prediction server (http://www.genebee.msu.su/genebee.html). Our initial pairwise alignments were obtained using the Gonnet PAM250 amino acid substitution scoring matrix (Genebee program) and a significant threshold of 1% BLOSUM62 matrix of Henikoff and Henikoff (1992).

Comparative and functional structure of maize Dmc1 protein

To carry-out comparative studies on protein functional domains, maize Dmc1 amino acid sequence was applied to find out the most similar sequences deposited in the data bank using online softwares of http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi and http://www.sanger.ac.uk/Software/Pfam (the Pfam collection of protein families and domains). The secondary structural features of left-handed maize helical filament were predicted in PSIPrediction (http://www.bioinf.cs.ucl.ac.uk/psipred/) and Raswinsx programs. Tertiary structure of maize Dmc1 protein was also designed by the Raswinsx program. Prediction of signal peptide of maize Dmc1 was performed by SignalP 3.0 Server - prediction using neural networks (NN) and hidden Markov models (HMM) trained on eukaryote (Bendtsen et al., 2004). Another program that was used for homology search was signal blast (Frank and Sippl, 2008).

-----

[1]A1: *Danio rerio* (90112019), A2: *Henrardia persica* (11141585), A3: *Crithodium monococcum* (11141575), A4: *Crithopsis delileana* (11141555), A5: *Bromus sterilis* (11141543), A6: *Dreissena polymorpha* (21262969), A7: *Candida albicans* (1145716), A8: *Ciona intestinalis* (198436503), A9: *Hieracium caespitosum* (145952328), A10: *Coprinopsis cinerea* (6714639), A11: *Theileria parva* (71028324), A12: *Trichomonas vaginalis* (84784026), A13: *Hystrix duthiei* (166077981), A14: *Leucoagaricus* (110349609), A15: *Macaca mulatta* (109094191), A16: *Oreochromis niloticus* (63852084), A17: *Penaeus monodon* (183397237), A18: *Pleurodeles waltl* (82569454), A19: *Ricinus communis* (255552035), A20: *Anguilla japonica* (63852080), A21: *Australopyrum pectinatum* (11141579), A22: *Festucopsis serpentini* (11141569), A23: *Heteranthelium piliferum* (11141553), A24: *Bombyx mori* (2058711), A25: *Carassius auratus* (148727866), A26: *Cryptosporidium hominis* (67606014), A27: *Gallus gallus* (118082782), A28: *Taeniatherum caput* (11141573), A29: *Thinopyrum bessarabicum* (11141583), A30: *Triticum aestivum* (222154111), A31: *Ixodes scapularis* (242002736), A32: *Leymus triticoides* (166078005), A33: *Micromonas* (255086169), A34: *Oryza sativa* (14588672), A35: *Penicillium marneffei* (212546047), A36: *Pleurotus ostreatus* (13161942), A37: *Schizosaccharomyces pombe* (3176384), A38: *Hordeum vulgare* (7229683), A39: *Elymus hystrix* (166077985), A40: *Amblyopyrum muticum* (11141561), A41: *Dasypyrum villosum* (11141551), A42: *Homo sapiens* (40786809), A43: *Arabidopsis thaliana* (1790922), A44: *Aspergillus clavatus* (121709155), A45: *Chlamydomonas reinhardtii* (159469155), A46: *Canis familiaris* (73969089), A47: *Taeniopygia guttata* (224095169), A48: *Toxoplasma gondii* (237843305), A49: *Trypanosoma cruzi* (71663769), A50: *Leishmania major* (3132711), A51: *Lodderomyces elongisporus* (149239438), A52: *Monodelphis domestica* (126339552), A53: *Oryzias latipes* (199652322), A54: *Peridictyon sanctum* (11141563), A55: *Psathyrostachys fragilis* (11141601), A56: *Secale strictum* (11141571), A57: *Aegilops uniaristata* (82907643), A58: *Hordelymus europaeus* (171190312), A59: *Agropyron cristatum* (11141557), A60: *Eremopyrum distans* (11141547), A61: *Ajellomyces capsulatus* (154279522), A62: *Glycine max* (1518157), A63: *Entamoeba histolytica* (67482461), A64: *Giardia intestinalis* (33667818), A65: *Cynops pyrrhogaster* (10944306), A66: *Talaromyces stipitatus* (242824534), A67: *Tribolium castaneum* (91078458), A68: *Zea mays* (195626386), A69: *Lentinula edodes* (242389898), A70: *Lophopyrum elongatum* (11141567), A71: *Mus musculus* (111601086), A72: *Pan troglodytes* (114686387), A73: *Plasmodium vivax* (156097941), A74: *Pseudoroegneria spicata* (11141565), A75: *Hydra vulgaris* (11994855), A76: *Leucoagaricus gongylophorus* (110349629), A77: *Brugia malayi* (170591324), A78: *Ornithorhynchus anatinus* (149419771), A79: *Neosartorya fischeri* (119489209), A80: *Leucocoprinus fragilissimus* (110349631), A81: *Uncinocarpus reesii* (258576827).


RESULTS

In silico cloning of the *Zea mays Dmc1* cDNA and its validation by last exon

The results of the survey in public-domain collection of maize ESTs by the rice *Dmc1* gene and protein showed strong homologies (data not shown) with two groups of ESTs including 5' ESTs (BM501047; Co533960; AI854907; EE190345; DV531377; DY541248; DY621394; DV522737) and 3' ESTs (DV495313; EE294070; AI833692; DV531376; DV522736; Dy621393).

The assembling of the 5' and 3' ESTs was manually done by pairwise alignments for all combinations to resolve the longest nucleotide fragment. Consequently, one pair of EST, DV531376 and DV531377, was chosen to cover the complete open reading frame (ORF) as a single representative cDNA without any shift mutation containing 5' and 3' untranslated regions as well as poly-adenylate polymerase signal. The unique putative ORF achieved by aligning was translated to amino acid sequences. The homology search of cloned 3' end of gene and cDNA (data not shown) with the complete cDNA obtained from *in silico* cloning showed 99.6 and 100% identity for nucleotide and amino acid sequences, respectively. The fine structure of the maize *Dmc1* gene was studied through homology search of Maize *Dmc1* cDNA sequence with the maize high throughput genomic sequences (HTG). These studies demonstrated that the *Dmc1* gene located on chromosome 3 showed 15 exons and 14 introns (Figure 1) while the rice *Dmc1* gene showed 14 exons and 13 introns (Ding et al., 2001). In contrast, the first intron of maize *Dmc1* positioned within 5'UTR sequence before Kozak sequence (CAC) **(**Cavener, 1987; Kozak, 1987) appears as 3 bp upstream of the translational start codon (Figure 1).

Phylogenetic tree based on Dmc1 amino acid sequences

A multiple alignment across 81 genera spanning the major kingdoms of living organisms including maize using the Dmc1 protein sequence was performed to generate a phylogenetic tree (Figure 2). The tree is divided into three major clusters. The first cluster includes main branches of insects, fishes, mammals, and birds. The second cluster has got three main branches, the first contains some fungi, the second contains parasitic protozoan and the third branch contains plants, which are further subdivided into domesticated mono- and dicots. The wild monocot and grass plants from Poaceae constitute the third cluster. Maize Dmc1 shows 85% identity with wheat Dmc1 and 61% with *Arabidopsis* (Figure 2).
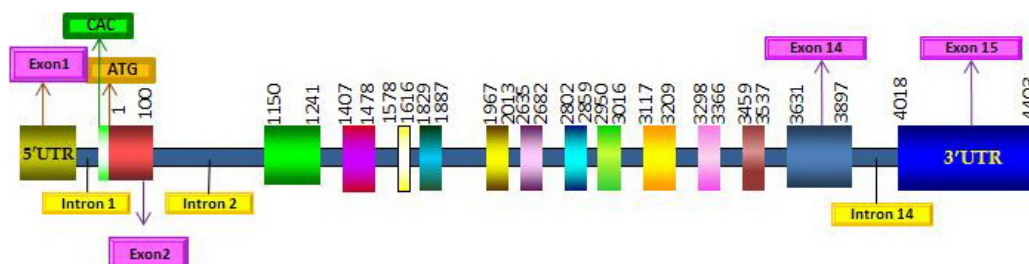
Figure 1. Maize *Dmc1* gene fine structure (15 exons + 14 introns) predicted by sequence alignments of maize *Dmc1* cDNA and maize genome sequence of chromosome 3.
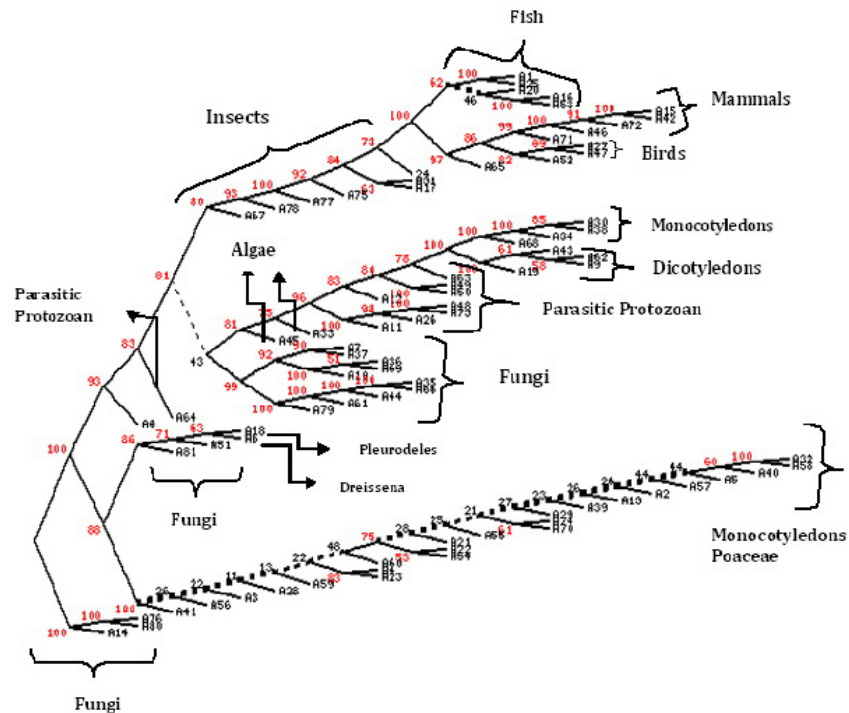
Figure 2. Phylogenetic tree of *Dmc1* sequences. Values indicate percent bootstrap support for each branch. For accession number and organisms, see Material and Methods.

## Maize Dmc1 conserved sequences, 2-D, 3-D structure and functions

Outputs of the Conserved Domain DNA (CDD) online software demonstrate the significant similarities on conserved domains and functional motifs in initial 10 files (Figure 3), which was in agreement with results obtained by the Pfam collection of protein families and domains. As indicated in Figure 3 some of the previously reported functional motifs appear in the maize Dmc1 amino acid sequence. These include the putative dsDNA binding helix-hairpin-helix (HhH) motif in smaller NTD region, the putative ssDNA binding L1 and L2 loops, the ATP binding WalkerA and WalkerB motifs, the polymerization motif (PM), and the subunit rotation motif (SRM).

The resulting secondary structure of the maize Dmc1 protein consisted of 12 α-helices and 11 β-strands (Figure 4A and B) showing some similarities and differences with other Dmc1 protein sequences.

Maize Dmc1 nucleotide binding (NB) domain, A site (WalkerA)

The GEFRSGKT segment called ATP binding WalkerA from positions 133 to 140 is almost totally conserved relative to the aggregate of 10 sequences (Figures 3 and 5). The two exceptions are S-137 (serine at position 137), identical in 2 sequences but changing to T, C, V, and A in other groups, and E-134, identical in 8 sequences but changing to A and P.

There are two other conserved segments, which appeared as pre-WalkerA and post-WalkerA starting from positions 120 to 130 and 141 to 153, respectively. The pre-WalkerA segment was primarily buried in the structure, containing two invariant glycine positions and several conserved hydrophobic residues. The segment of post-Walker with the some hydrophobic residues is similarly buried but less conserved. The hydrophobic environment presumably enhances structural stability at the A site (Figures 3 and 5).
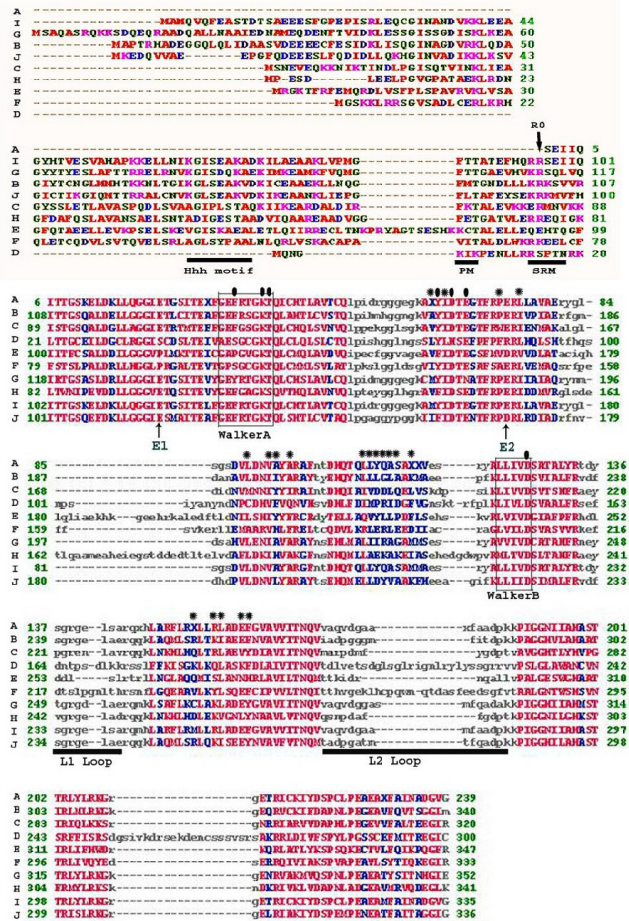
Figure 3. Sequence alignment of RecA family proteins. **A.** 1 NOW_ A, *Homo sapiens*, chain A, crystal structure of a Rad51-Brca2 Brc repeat complex; **B.** GI [195626386] *Zea mays*; **C.** GI [14286170] *Sulfolobus solfataricus*, DNA repair and recombination protein RadA; **D.** GI [15242137] *Arabidopsis thaliana*, XRCC3; ATP binding/ damaged DNA binding/protein binding/single-stranded DNA binding; **E.** GI [3914534] *Homo sapiens*, DNA repair protein RAD51 homolog; **F.** GI [30354420] *Danio rerio*, meiotic recombinase Dmc1; **G.** GI [25145771] *Caenorhabditis elegans*, RecA/Rad51/Dmc1-like protein; **H.** GI [13878695] *Halobacterium salinarum*, DNA repair and recombination protein radA; **I.** GI [585770] *Gallus gallus* (chicken), DNA repair protein RAD51 homolog 1; **J.** GI [2500099] *Homo sapiens*, meiotic recombination protein DMC1/LIM15 homolog. All these RecA-like strand exchange proteins have similar N-terminal domains. The C-terminal RecA domains have been deleted for clarity. Functional motifs are indicated under their corresponding amino acid sequences: the putative ssDNA binding L1 and L2 loops; the ATP binding WalkerA and WalkerB motifs; ATP binding sites (blots) and multimer site (BRC) interface (asterisks), the polymerization motif (PM), the subunit rotation motif (SRM), and position of the R0-E1-E2.

Figure 4. Secondary structural features of the maize Dmc1 helical filament as indicated in pink (α-helices) and yellow (β-strands) colors. **A.** Raswin program result; **B.** PSIpred result.

Figure 5. Maize Dmc1 tertiary structure: WalkerA (blue), WalkerB (orange), ATP binding (red and dark blue) and multimer interface (green).

Maize Dmc1 B site (WalkerB)

The B motif of the Walker box comprises four successive hydrophobic residues followed by aspartate at positions 224 to 228. The key residue of the B site is D-228. The secondary structure of this segment is a β-strand that can anchor residue D-228 with a dense hydrophobic milieu. Intriguingly, the immediately carboxyl segment from positions 229 to 233 (SVIAL) was invariant across all RecA sequences (Figures 3 and 5).

Maize Dmc1 ATP binding sites

There are some regions between WalkerA and WalkerB as ATP binding sites, which

are almost invariant over all sequences including F-135, K-139 to T-140, Y-165, D-167, E-169, and D-228 residues (Figures 3 and 5).

Maize Dmc1 DNA binding domains

The disordered loops L1 (239 to 246) featuring three acidic and two glycine residues and L2 (276 to 289) have been proposed as DNA binding domains. The L2 stretch is substantially conserved only in D-287, P-288 and K-289. In particular, the boundary residue I-276 was variant, as seen in positions of D-278, G-282, I-285, and T-286.

Another DNA binding motif detected using alignment of maize Dmc1 with the yeast RecA protein (Chen et al., 2007) as an HhH motif KGLSEAKV (70 to 79) in the N-terminal domain mediates dsDNA binding (Figure 3).

Maize Dmc1 multimer site (BRC) interface

Another region proposed to interact with DNA is contained in the BRC interface in the segment from positions: A-164, I-166, P-174, I-177, L-192, I-195 to I-196, A-198, LLLGLA (209 to 214), K-216, R-257, K-260 to I-261, and E-264 to F-265. This region, also involved in Monomer-Monomer (M-M) interactions, includes three highly conserved cationic residues I-166, P-174 and L-177. Another PM contained residues 92 to 94 in maize Dmc1 (Figures 3 and 5).

R0 amino acid inside the SRM, which is evolutionarily conserved in almost RecA family proteins (Figure 3), is also present here and probably acts as PM for filament assembly.

Maize Dmc1 subunit rotation motif

A hinge region KRKSV (amino acid residues 101-105), required for conformational transition changes of the Dmc1 protein located immediately after the sixth α-helix (amino acid residues 94-100) (Figure 3), designated the SRM.

Maize Dmc1 E1-R0-E2 triad

The E1-R0-E2 (E-117, R-96 and E-175) triad acts as a clip to fasten two AMP-PNP binding surfaces at the neighboring core ATPase domains. As reported earlier the E1-R0-E2 triad may control DNA binding as well as DNA release.

Signal peptide detection for the maize Dmc1 protein

Results obtained from the neural network algorithm showed three scores. The high S-score indicates a signal peptide at the beginning, i.e., 20 amino acids at the first sequence; the C-score indicates a cleavage site between amino acids 19 and 24, so that the mature protein possibility starts at (and includes) position 24, and the true cleavage site is assigned from the Y-score in A-19 (Figure 6).

Blast homology for the maize Dmc1 signal peptide showed maximum S-value of 100%, bit score of 84.3% (Table 1) and high identity of 62% with the soybean Dmc1 signal peptide (Figure 7).
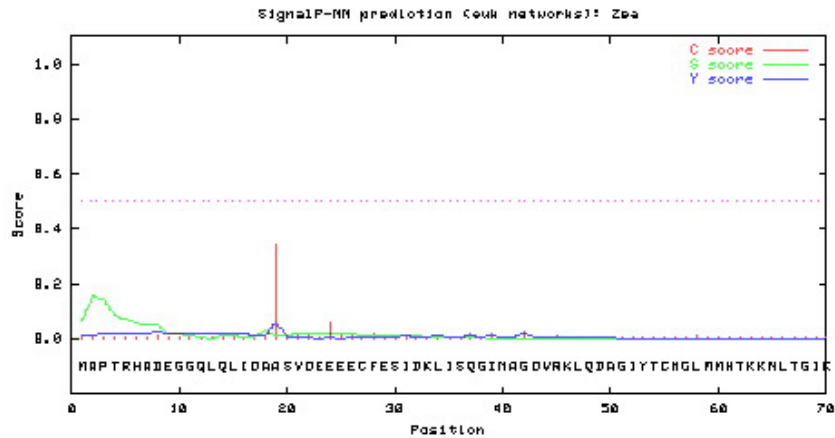
Figure 6. Signal peptide prediction for the maize Dmc1 protein using neural networks (NN) by SignalP 3.0 server.

Table 1. Signal peptide blast.

Significant alignments: name	Bit score	S-value
DMC1_SOYBN	84.3	100.0
DMC1_ARATH	79.3	3.5
DMC1_HUMAN	42.7	0.0
DMC1_MOUSE	41.2	0.0
RAD51_MOUSE	38.1	0.0
RAD51_HUMAN	37.4	0.0
RAD51_CANFA	37.4	0.0
RAD51_BOVIN	37.4	0.0
RAD51_RABBIT	37.0	0.0
DLH1_CANAL	36.6	0.0


```
Score=84.3, Bits (207), Expect=7e-18, Identities=37/59 (62%), Positives = 49/59 (83%)

Query: 1 MAPTRHADEGGQLQLIDAASVDEEEECFESIDKLISQGINAGDVRKLQDAGIYTCNGLMM 60
         + A + +       GQLQL++  +  D++E+ FE+IDKLI+QGINAGDV+KLQDAGIYTCNGLM
Subject: 2 LATLKSEESSGQLQLVEREDIDDDEDLFEAIDKLIAQGINAGDVKKLQDAGIYTCNGLM 60
```

Figure 7. Results of the maize Dmc1 signal peptide blast with soybean signal peptide.

DISCUSSION

Maize as a food and industrial monoecious plant is almost produced by hybrid seeds, which need to be revolutionized by cost-effective new seed production strategies like apomixes. It seems that gathering molecular data on meiosis division and genetic recombination not only improve food production, but also might be exploited for any ongoing manipulations to recover genetic disabilities in animal and human through generative stem cell culture and differentiation in the near future. In this context, Dmc1 recombinase as a RecA-like family protein in major kingdoms of living organisms targeted for structural and functional analysis. Due

to availability of 5' and 3' ESTs from generative organs and high throughput genomic sequence data of maize on the one hand and full-length rice *Dmc1* cDNA as an efficient orthologue on the other hand, genomic wide *in silico* cloning of maize *Dmc1* and its validation by 3' end of gene and cDNA was carried out (Figure 8). Figure 1 demonstrates that the first intron in this gene was inserted at 5'UTR immediately before Kozak sequence CAC, which may potentially play a role at pre-translation regulation of *Dmc1* expression. Despite the above discussed first intron, the remaining exons and introns are in the same order as seen in rice *Dmc1* with some variation in nucleotide number and sequences.
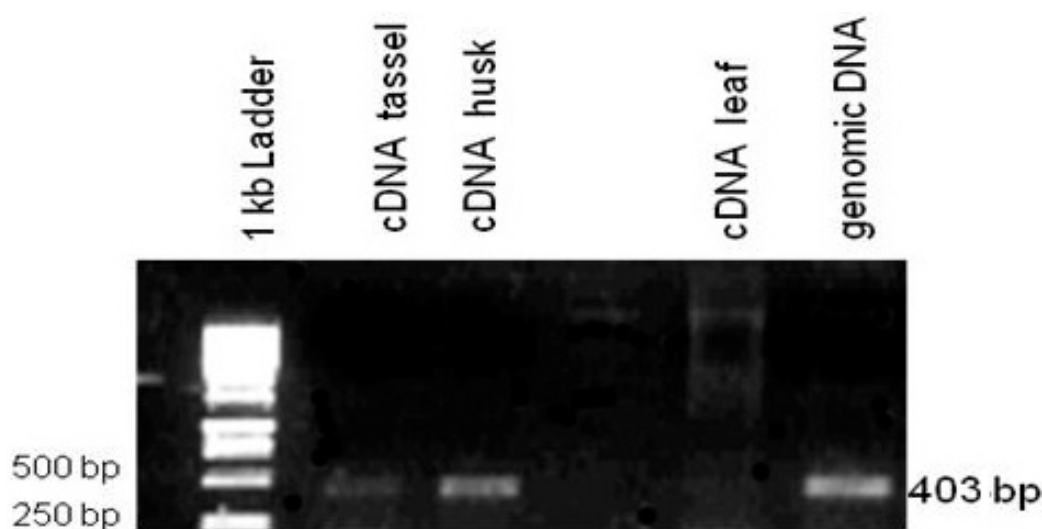
Figure 8. PCR amplification of maize Dmc1 from cDNA of tassel, husk, leaf (negative control), and maize genomic DNA (positive control) as template.

Because of high levels of conservation of the *Dmc1* exon nucleotides and amino acid sequences across prokarya, archae and eukarya (Sandler et al., 1996; Brendel et al., 1997), the phylogenic tree of the 81 RecA-like/Dmc1 protein sequences spanning various living organisms including the maize was constructed by different softwares including Mega4, Phylip, Treecon, Bioedit, Genebee, etc., resulting in various results. The phylogenic tree produced by Genebee demonstrated in Figure 2 was mainly divided into three major clusters of non-plant species as first, some domesticated monocots (including maize, rice, wheat, and barley) and dicot plants plus fungi/protists appeared in branches as second cluster and wild monocots and grasses from Poaceae as third.

To elaborate Dmc1 3-D and functional structure, the CDD analysis showed the significant identity of *Zea mays* with *Homo sapiens*, which has been studied in detail, and due to consensus amino acid sequences within proposed domains in Triticeae (Petersen and Seberg, 2002) we reproduced the maize Dmc1 relevant domains and motifs. Since the several key function and structure regions have been deduced from the crystal structure of the *Escherichia coli* RecA protein (Story and Steitz, 1992; Story et al., 1992; Karlin and Brocchieri, 1996) and *H. sapiens* Dmc1 (Passy et al., 1999; Kinebuchi et al., 2004; Okorokov et al., 2010) the

domains and motifs of WalkerA and WalkerB as ATP binding site; BRC interface; the putative ssDNA binding L1 and L2 loops (Figures 3 and 5) (Story et al., 1992; Story and Steitz, 1992; Brendel et al., 1997; Petersen and Seberg, 2002; Chen et al., 2007; Sakane et al., 2008), the putative dsDNA binding hairpin HhH, PM, SRM, and small NTD have been proposed (Figure 3) (Chen et al., 2007).

Among these regions two loops of L1 and L2 are considered to be involved in binding of mostly single- and less double-stranded DNA, the latter is proposed to be the primary binding site (Story et al., 1992; Karlin and Brocchieri, 1996). Both domains are fairly conserved among bacterial RecA proteins (Karlin and Brocchieri, 1996), similar to maize Dmc1. The sequence conservation, particular of the L2 region, is less pronounced among the broader group of recA-like proteins. The L1 domain is highly conserved in the angiosperms, whereas variation can be observed in four amino acid residues in the L2 domain (Petersen and Seberg, 2002). Intriguingly, the structural transition is accompanied by concerted movements of the L1, L2 and HhH motifs from filament center to filament surface and caused homologous pairing between ssDNA and dsDNA (Chen et al., 2007). In addition, a recent study indicated that the transition from ATP-bound state to an intermediate RadA (RecA family) involved extensive allosteric effects spanning the ATPase site and the L2 putative DNA-binding loop (Qian et al., 2005).

According to Chen et al. (2007), the axial rotation of SRMs controls quaternary structures of RecA family proteins and regulates their enzymatic functions too. Another report suggested that SRM is important for coupling *Saccharomyces cerevisiae* RadA's ssDNA binding and ATPase activities to its strand exchange activity similar to the rotary motor F1-ATPase (Qian et al., 2005). In this relationship, the hinge region exhibited sequential conformational changes corresponding to structural transitions from a closed ring (Shin et al., 2003), to the AMP-PNP right-handed filament (Wu et al., 2004), to the over wound right-handed filament (Ariza et al., 2005), and finally to the left-handed filament. These conformational changes appear to be caused mainly by the progressive rotations of their polypeptide backbones, as the torsion angles in this hinge region change stepwise during his clockwise axial rotation (Chen et al., 2007). The maize Dmc1 amino acid sequence shows a Ro site within the SRM, which functions as the polymerization motif for filament assembly (Conway et al., 2004), which is evolutionarily conserved in almost all RecA family proteins (Figure 3) (Chen et al., 2007).

The reports on polymeric structure of RecA protein in *E. coli* revealed that their monomers form either a polymeric helical filament or hexameric rings (Story et al., 1992) whereas human Dmc1 monomers form octameric rings, only (Passy et al., 1999). The octameric rings may be assembled pairwise in long stacks with DNA binding taking place in the central channel formed by the stacked rings with monomer-monomer contact (Passy et al., 1999; Kinebuchi et al., 2004; Okorokov et al., 2010).

Usually, the ATPase domain is preceded by a short β-strand polymerization motif (Pellegrini et al., 2002). A hydrophobic residue (phenylalanine in RADA, Rad51 and DMC1, isoleucin in RecA) in polymerization motif docked into a hydrophobic pocket on the neighboring core ATPase domain (Chen et al., 2007). For RecA proteins, the transition from closed rings to helical filaments could be flexible or regulated. Archaeal RadA proteins form both closed rings and right-handed helical filaments (Wu et al., 2004). *S. cerevisiae* Dmc1 (ScDmc1) and human Dmc1 proteins, in the presence of ssDNA, switch from rings to helical filaments in a Ca^{2+}-dependent manner (Chang et al., 2005) Conceivably, all members of the RecA family form similar right-handed nucleoprotein filaments in which the ssDNA lies close to the filament axis.

These nucleoprotein filaments are thought to be catalytically active in mediating strand exchange reactions (Chang et al., 2005). Due to lack of any data and information on protein crystal structure in plants and according to the presence of required polymeric regions and ATPase domain in maize Dmc1 amino acid sequences it is presumable that maize Dmc1 proteins may self-polymerize into right/left-handed helical filaments *in vitro* or form polymeric stacks similar to human Dmc1 and even could be flexible in the sense of regulated transition from closed rings to helical filaments, which is strongly need to be studied in detail in plants.

Regarding targeting the *Dmc1* gene product into the nucleus it is noteworthy to mention that the structural studies showed that about 19 to 24 amino acids are discarded during transportation and the homology search with an expanding array of peptide signaling groups in plants showed higher similarity with the soybean Dmc1 signal peptide (Figure 7 and Table 1).

ACKNOWLEDGMENTS

REFERENCES

Ariza A, Richard DJ, White MF and Bond CS (2005). Conformational flexibility revealed by the crystal structure of a crenarchaeal RadA. *Nucleic Acids Res.* 33: 1465-1473.

Bendtsen JD, Nielsen H, von Heijne G and Brunak S (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340: 783-795.

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

Brendel V, Brocchieri L, Sandler SJ, Clark AJ, et al. (1997). Evolutionary comparisons of RecA-like proteins across all major kingdoms of living organisms. *J. Mol. Evol.* 44: 528-541.

Cavener DR (1987). Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acids Res.* 15: 1353-1361.

Chang YC, Lo YH, Lee MH, Leng CH, et al. (2005). Molecular visualization of the yeast Dmc1 protein ring and Dmc1-ssDNA nucleoprotein complex. *Biochemistry* 44: 6052-6058.

Chen LT, Ko TP, Chang YC, Lin KA, et al. (2007). Crystal structure of the left-handed archaeal RadA helical filament: identification of a functional motif for controlling quaternary structures and enzymatic functions of RecA family proteins. *Nucleic Acids Res.* 35: 1787-1801.

Conway AB, Lynch TW, Zhang Y, Fortin GS, et al. (2004). Crystal structure of a Rad51 filament. *Nat. Struct. Mol. Biol.* 11: 791-796.

De Muyt A, Pereira L, Vezon D, Chelysheva L, et al. (2009). A high throughput genetic screen identifies new early meiotic recombination functions in *Arabidopsis thaliana*. *PLoS Genet.* 5: e1000654.

Deng ZY and Wang T (2007). OsDMC1 is required for homologous pairing in *Oryza sativa*. *Plant Mol. Biol.* 65: 31-42.

Ding ZJ, Wang T, Chong K and Bai S (2001). Isolation and characterization of *OsDmc1*, the rice homologue of the yeast *Dmc1* gene essential for meiosis. *Sex. Plant Reprod.* 13: 285-288.

Frank K and Sippl MJ (2008). High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics* 24: 2172-2176.

Henikoff S and Henikoff JG (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* 89: 10915-10919.

Hong EL, Shinohara A and Bishop DK (2001). *Saccharomyces cerevisiae* Dmc1 protein promotes renaturation of single-strand DNA (ssDNA) and assimilation of ssDNA into homologous super-coiled duplex DNA. *J. Biol. Chem.* 276: 41906-41912.

Karlin S and Brocchieri L (1996). Evolutionary conservation of *RecA* genes in relation to protein structure and function. *J. Bacteriol.* 178: 1881-1894.

Keeling PJ and Doolittle WF (1995). Archaea: narrowing the gap between prokaryotes and eukaryotes. *Proc. Natl. Acad.*

Sci. U. S. A. 92: 5761-5764.

Kinebuchi T, Kagawa W, Enomoto R, Tanaka K, et al. (2004). Structural basis for octameric ring formation and DNA interaction of the human homologous-pairing protein Dmc1. *Mol. Cell* 14: 363-374.

Kobayashi T, Hotta Y and Tabata S (1993). Isolation and characterization of a yeast gene that is homologous with a meiosis-specific cDNA from a plant. *Mol. Gen. Genet.* 237: 225-232.

Kozak M (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15: 8125-8148.

Metkar SS, Sainis JK and Mahajan SK (2004). Cloning and characterization of the *Dmc1* genes in *Oryza sativa*. *Curr. Sci.* 87: 353-357.

Okorokov AL, Chaban YL, Bugreev DV, Hodgkinson J, et al. (2010). Structure of the *hDmc1*-ssDNA filament reveals the principles of its architecture. *PLoS One* 5: e8586.

Passy SI, Yu X, Li Z, Radding CM, et al. (1999). Human Dmc1 protein binds DNA as an octameric ring. *Proc. Natl. Acad. Sci. U. S. A.* 96: 10684-10688.

Pellegrini L, Yu DS, Lo T, Anand S, et al. (2002). Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* 420: 287-293.

Petersen G and Seberg O (2002). Molecular evolution and phylogenetic application of Dmc1. *Mol. Phylogenet. Evol.* 22: 43-50.

Qian X, Wu Y, He Y and Luo Y (2005). Crystal structure of *Methanococcus voltae* RadA in complex with ADP: hydrolysis-induced conformational change. *Biochemistry* 44: 13753-13761.

Sakane I, Kamataki C, Takizawa Y, Nakashima M, et al. (2008). Filament formation and robust strand exchange activities of the rice Dmc1A and Dmc1B proteins. *Nucleic Acids Res.* 36: 4266-4276.

Sandler SJ, Satin LH, Samra HS and Clark AJ (1996). *recA*-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast *Saccharomyces cerevisiae. Nucleic Acids Res.* 24: 2125-2132.

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

Shin DS, Pellegrini L, Daniels DS, Yelent B, et al. (2003). Full-length archaeal *Rad51* structure and mutants: mechanisms for *RAD51* assembly and control by *BRCA2. EMBO J.* 22: 4566-4576.

Smith MW, Feng DF and Doolittle RF (1992). Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* 17: 489-493.

Story RM and Steitz TA (1992). Structure of the recA protein-ADP complex. *Nature* 355: 374-376.

Story RM, Weber IT and Steitz TA (1992). The structure of the *E. coli* recA protein monomer and polymer. *Nature* 355: 318-325.

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

Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

Wu Y, He Y, Moya IA, Qian X, et al. (2004). Crystal structure of archaeal recombinase RADA: a snapshot of its extended conformation. *Mol. Cell* 15: 423-435.

Zhang L, Tao J, Wang S, Chong K, et al. (2006). The rice *OsRad21*-4, an orthologue of yeast *Rec8* protein, is required for efficient meiosis. *Plant Mol. Biol.* 60: 533-554.