Exon-intron structure of CuZn-superoxide dismutase isoform genes from the fern *Equisetum arvense* and the moss *Pogonatum inflexum* reveals their molecular evolution in the plant lineage

Sumio Kanematsu*, Shinpei Ueno and Misao Okayasu

Department of Food Science, Minami-Kyushu University, Miyazaki 880-0032, Japan

Received October 20, 2010; Accepted January 17, 2011

CuZn-superoxide dismutase (SOD) in plants consists of chloroplastic and cytosolic isoforms. To elucidate the adaptation of plants to the changing oxidative stress during the course of evolution, we investigated the molecular evolution of CuZn-SOD by cloning and characterizing their isoform genes from the fern *Equisetum arvense* and the moss *Pogonatum inflexum*.

Two chloroplastic and one cytosolic CuZn-SOD cDNAs were obtained from both *E. arvense* and *P. inflexum* by RT-PCR and RACE. This was the first demonstration of two types of chloroplastic isoform gene for the lower plants. The two chloroplastic genes were phylogenetically shown to be paralogs produced by gene duplication after divergence of the species. The chloroplastic and cytosolic CuZn-SOD cDNAs in the fern and moss encoded respective CuZn-SODs similar to those of higher plants.

Of the chloroplastic and cytosolic CuZn-SOD genomic genes, 18,343 bp and 16,593 bp were sequenced for *E. arvense* and 7,146 bp and 5,583 bp for *P. inflexum*, respectively. The chloroplastic CuZn-SOD gene of *P. inflexum* contained nine exons, while that of *E. arvense* contained eight. The first intron of the *P. inflexum* gene, which is also present in that of the green alga *Spirogyra* sp., was absent from that of *E. arvense* as well as those of higher plants, indicating that the corresponding intron was deleted prior to the divergence of pteridophytes in the plant lineage during the course of evolution. The cytosolic genes from the fern and moss exhibited the same exon-intron structures consisting of seven exons. From the present results we propose the molecular evolution of CuZn-SOD genes during the early phase of land plant evolution.

Key words: superoxide dismutase, molecular evolution, exon-intron structure, *Pogonatum inflexum*, *Equisetum arvense*.

INTRODUCTION

The landing of ancestral land plants on a terrestrial surface from water at Silurian, 450 million years ago, was a great challenge for organisms during the course of evolution (Bhattacharya and Medlin 1998, Becker and Marin 2009). For the first land plants, adaptation to a terrestrial environment was crucial for survival. For this adaptation, reinforcement of the protective system against reactive oxygen species (ROS) that have deleterious effects on all cells (Fridovich 1995) must have been one of the most important responses, because the new environment resulted in the production of much ROS including superoxide, H_2O_2, hydroxyl radical, and singlet oxygen due to dehydration, ultraviolet irradiation and high intensity of light (Asada 1999, 2006). Thus, the enzyme system of ROS-scavenging might have been primarily responsible for the adaptation of ancestral plants to severe ROS-producing environments.

Superoxide dismutase (SOD), which catalyzes the dismutation of superoxide to H_2O_2 and molecular oxygen at a diffusion-controlled rate and protects cells from oxidative damage caused by ROS (Fridovich 1995), is a key enzyme in the ROS-scavenging system since the superoxide is located furthest upstream in the cascade reactions of ROS production. SOD is a metalloenzyme and is classified into four isozymes in terms of prosthetic metals: CuZn-SOD, Mn-SOD, Fe-SOD and Ni-SOD. Ni-SOD, the latest found SOD (Youn et al. 1996), had been thought to be restricted to only some bacteria, but recently it was found in eukaryotic photoplankton, suggesting

*Corresponding author: E-mail, kanemats@nankyudai.ac.jp; Fax, +81-985-83-3521.

The abbreviations used are: *Esod, E. arvense* CuZn-SOD gene; ESOD, gene product of *Esod*; *Psod, P. inflexum* CuZn-SOD gene; PSOD, gene product of *Psod*; ROS, reactive oxygen species; SOD, superoxide dismutase.

Accession numbers of cDNAs and genomic genes are as follows: *Esod-1* cDNA, AB79540; *Esod-2* cDNA, AB79541; *Esod-3* cDNA, AB605764; *Esod-1*, AB193093; *Esod-2*, AB193094; *Psod-2* cDNA, AB079114; *Psod-1* cDNA, AB079115; *Psod-3* cDNA, AB079116; *Psod-2*, AB205008; *Psod-3*, AB205009. Part of the present results have been presented elsewhere (Kanematsu et al. 2003, Okayasu el al. 2004, Ueno and Kanematsu 2004, Kanematsu 2005).

the involvement of horizontal gene transfer (Dupont et al. 2008, Cuvelier et al. 2010). CuZn-SOD, Mn-SOD and Fe-SOD are distributed widely among organisms along with their phylogenetic relationships (Kanematsu and Asada 1994). In plants, CuZn-SOD consists of chloroplastic and cytosolic isoforms that are encoded by respective genes. Mn-SOD and Fe-SOD, which are products of paralogous genes, are localized in mitochondria and chloroplasts, respectively. Although the Fe-SOD gene is found in plant genomes, it is not necessarily expressed in many plants (Kanematsu and Asada 1994).

In the evolution of the monophyletic lineage of Viridiplantae (green algae and land plants), land plants (embryophytes) including bryophytes, pteridophytes and spermatophytes are thought to have evolved from charophycean green algae (charophytes) (Becker and Marin 2009). The first land plant is believed to have been an ancestor of bryophytes. Previously, we characterized the algal CuZn-SOD gene for the first time from the green alga *Spirogyra* sp. and suggested that the divergence of chloroplastic and cytosolic isoforms occurred prior to the divergence of chlorophyte algae (e. g., *Chlamydomonas*) and streptophyte algae (charophyte algae, e. g., *Spirogyra* and *Nitella*) (Kanematsu et al. 2010). Here, to clarify the adaptation of land plants to the changing oxidative stress in the course of evolution, we cloned and characterized cDNAs and genomic genes of chloroplastic and cytosolic CuZn-SODs from the fern *Equisetum arvense* and the moss *Pogonatum inflexum*.

In this work, we found two types of chloroplastic CuZn-SOD genes in both organisms, although only one type of chloroplastic CuZn-SOD has been found in higher plants so far. Furthermore, we determined for the first time the structures of CuZn-SOD genomic genes of lower plants. The exon-intron structures show that the *E. arvense* chloroplastic CuZn-SOD gene resembles those of higher plants whereas the *P. inflexum* chloroplastic SOD gene resembles that of the green alga *Spirogyra*, reflecting the order of appearance in the plant lineage.

MATERIALS AND METHODS

Materials. Isogen, DNeasy Plant Mini Kit and Quantum Prep Plasmid Miniprep Kit were purchased from Nippon Gene (Toyama, Japan), Qiagen (Valencia, CA, USA) and Bio-Rad (Hercules, CA, USA), respectively. SuperScript First-Strand Synthesis System for RT-PCR was from Life Technologies (Rockville, MD, USA). SMART RACE cDNA Amplification Kit and Universal GenomeWalker Kit are products of Clontech (Palo Alto, CA, USA). pGEM-T Easy Vector System I, Kilo-Sequence Deletion Kit and Epicurian Coli XL1-Blue MRF' Supercompetent Cells were purchased from Promega (Madison, WI, USA), Takara (Kyoto, Japan) and Stratagene (La Jolla, CA, USA), respectively.

Sterile and fertile shoots of the fern *Equisetum arvense* (horsetail) and leafy gametophytes of the moss *Pogonatum inflexum* were collected in the Takanabe campus of Minami-Kyushu University, Miyazaki, Japan. The materials were used on the same day of collection.

Isolation of poly(A)$^+$ RNA and genomic DNA. Total RNAs of *E. arvense* and *P. inflexum* were extracted from intact tissues using Isogen by a similar method to that described previously (Kanematsu and Fujita 2009). In the case of horsetail, both sterile and fertile shoots were used for RNA extraction. Poly (A)$^+$ mRNAs were purified from the total RNA as described previously (Kanematsu and Fujita 2009). Genomic DNA of *E. arvense* and *P. inflexum* were isolated from intact tissues using a DNeasy Plant Mini Kit according to the manufacturer's instructions. The genomic DNA of horsetail was obtained from green sterile shoots.

Amplification of partial cDNA fragments by RT-PCR. cDNA cloning for CuZn-SODs of *E. arvense* and *P. inflexum* was performed by a 5'- and 3'-RACE method using gene-specific primers after obtaining sequence information by RT-PCR using degenerated primer sets. The reverse transcription of poly(A)$^+$ mRNA and amplification of the core region of the genes by PCR were done as before (Kanematsu et al. 2010). The following degenerate primers were used: sense primer, 5'-CARGARGAYGAYGGNCCNAC-3' (SPGY-SENSE); and antisense primers, 5'-CCNCCYTTNCCAARRTCRTC-3' (SPGY-AS-A), 5'-CCNCCYTTNCCGARRTCRTC-3' (SPGY-AS-G), 5'-CCNCCYTTNCCTARRTCRTC-3' (SPGY-AS-T), and 5'-CCNCCYTTNCCCARRTCRTC-3' (SPGY-AS-C).

5'- and 3'-RACE for full-length cDNA cloning. The preparation of first-strand cDNA for 5'- and 3'-RACE, and amplification by 5'- and 3'-RACE using a SMART RACE cDNA Amplification Kit were conducted according to Kanematsu and Fujita (2009). The following gene-specific primers were used for 5'- and 3'-RACE of *E. arvense* SOD cDNAs: for *Esod-1* cDNA, EQ 5GSP#10 (5'-upstream antisense) 5'-AACACCATCAGGTCCCGCAACAAC-3' and EQ 3GSP#3 (3'-downstream sense) 5'-CTGGCAAGCACGGATTCCACCTC-3'; for *Esod-2* cDNA, EQ 5GSP#13 (5'-upstream antisense) 5'-GCCTTCCCATCATCTCCGACAGTG-3' and EQ 3GSP#2 (3'-downstream sense) 5'-TGTGGTTGGTTCTCTTAGCGGGCTTTC-3'.

The gene-specific primers used for RACE of *P. inflexum* SOD cDNAs were as follows: for *Psod-2* cDNA, PO 5GSP#11 (5'-upstream antisense) 5'-GGACCAGACAAAGGGATCTGATCGTCAG-3' and PO 3GPS#9 (3'-downstream sense) 5'-TTTGGGCAATGTTGTGGCTGGAC-3'; for *Psod-3* cDNA, PO 5GPS#29 (5'-upstream antisense) 5'-GACCAATGATGGAATCACACCCAGTGAG-3' and PO 3GPS#16 (3'-downstream sense) 5'-CCTAAGAATTTGGAGCATGGAGCCCC-3'; for *Psod-1* cDNA, PO 5GSP#14 (5'-upstream antisense) 5'-GTTCGTGGATGACAAATGCTCTACCGAC-3' and PO 3GSP#3 (3'-downstream sense) 5'-ATATCGACCGCCACGCAGGGGAC-3'. For amplified DNA fragments in 5'-RACE, the clones having the largest insert were selected by insert check with PCR after cloning into vectors and used for further analysis.

Amplification of the central portion of the genomic genes. The *E. arvense* and *P. inflexum* CuZn-SOD genomic genes were obtained by each combination of three sequences, i.e. central portion and two adjacent 5'-upstream and 3'-downstream portions of the gene as previously described (Kanematsu and Fujita 2009). A central portion of the gene was PCR-amplified with gene-specific primers based on the cDNA sequence, and 5'-upstream and 3'-downstream regions of the gene were obtained using the Universal GenomeWalker Kit (Clontech) with gene-specific primers designed from the sequence of the central portion of the gene.

The following primers were used for the amplification of the central portion of the *E. arvense* genes: sense and antisense primers for *Esod-1*, EQGL-F255 (sense) 5'-CGAGAAGAAGGCTGTCGCTGTCCTCAAAG-3', EQGL-F504 (sense) 5'-AGGGAATGTTGTTGCGGGACCTGATGG-3', EQGL-B483 (antisense) 5'-TTCATCTTCAGGGGCACCATGTGTAAGTCC-3' and EQGL-B774 (antisense) 5'-CTCTGCTTTACTGATGAAGCCAACCCACCCAAC-3'; sense and antisense primers for *Esod-2*, EQGY-F1 (sense) 5'-ACAGTTCGCCTGCGCTGTCTTGACCTAGGG-3', EQGY-F175 (sense) 5'-CCCTACGACTGTGGTTGGTTCTCTTAGCG-3', EQGY-F495 (sense) 5'-GCAAAACTACTGGAAATGCTGGAGGAAGACTTGC-3', EQGY-B309 (antisense) 5'-CCATGCACTTTGTTGGCAGGATTGTAG-3', EQGY-B529 (antisense) 5'-CGCAAGTCTTCCTCCAGCATTTCCAGTAGTTTTGC-3' and EQGY-B655 (antisense) 5'-TGACAGGTTCTCTATTTCACTGCCCTCGG-3'.

The following primers were used for the central region amplification of the *P. inflexum* genes: sense and antisense primers for *Psod-2*, POGLJ-F14 (sense) 5'-TTCATTCCCTGTGCGAAGAAGTTTACCATTGCC-3' and POGLJ-B853 (antisense) 5'-CATGCCAGTTGTACCTGGTCCCTAACCTCCAATC-3'; sense and antisense primers for *Psod-3*, POGY-F23 (sense) 5'-GCTCCTGCTCTTCTTCTCGGTGTCCAAACCC-3', POGY-F1709 (sense) 5'-TGGTGGGACACTGTTTATGGTTATTTGGTGCCG-3', POGY-B1742 (antisense) 5'-TGCGGCACCAAATACCATAAACAGTGTCCCAC-3' and POGY-B638 (antisense) 5'-GTCATTTCCACAGACTCTCCAGGCATCGCTG-3'.

The PCR conditions were the same as before (Kanematsu et al. 2010). The amplified DNA fragments were cloned and sequenced by a directional gene walking method along with the long sequences using gene-specific primers as described before (Kanematsu and Fujita 2009). At least three clones for each portion of the gene were used for sequencing. The highly repeated region of *Esod-1* was sequenced by using deleted clones in 3'-side with universal primers of a vector.

Generation of deleted clones from the central genomic DNA fragment of *E. arvense*. To avoid mispriming of a gene-specific sequence primer in a highly repeated sequence region of the central fragment of *Esod-1*, we conducted sequencing using deleted DNA fragments in pGEM-T EZ vector with vector-specific sequencing primers (-21M13 and M13RV). A series of clones differing in length was generated using Kilo-Sequence Deletion Kit (Takara) from *Esod-1* fragment (E2-E5) ligated in pGEM-T Easy vector after linearization by *Sac* I and *Sal* I.

Amplification of the upstream and downstream portions of the genomic genes. Construction of pools of adaptor-ligated genomic DNA fragments (GenomeWalker libraries) from which DNA fragments adjacent to the central sequence of the gene were amplified, and PCR of each library using the gene-specific primers were performed as before (Kanematsu et al. 2010). The following primers were used for 5'-upstream and 3'-downstream GenomeWalker amplification of the *E. arvense* genes. For 5'-upstream amplification: for *Esod-1*, EQL5GW1 (outer antisense) 5'-AAGTCCGTCTCTATTACCGCCGTCTTC-3' and EQL5GW2 (inner antisense) 5'-AAGGCTTGCGAGTCTTGTCGAGTGTACG-3'; for *Esod-2*, EQY5GW1 (outer sense) 5'-AGAAATGGACGACACCAGCGACTCCTG-3' and EQY5GW2 (inner sense) 5'-GAAAGCACTGCCACTGCCTTCAACCCC-3'. For 3'-downstream amplification: for *Esod-1*, EQL3GW1 (outer sense) 5'-TTGGGTGGGTTGGCTTCATCAGTAAAG-3' and EQL3GW2 (inner sense) 5'-AGTTGTAGATTTTTGAAGGCGGAGCCG-3'; for *Esod-2*, EQY3GW1 (outer sense) 5'-AGCTTGAAGGTGTACCCCCTATATGCC-3' and EQY3GW2 (inner sense) 5'-TCCGAGGGCAGTGAAATAGAGAACCTG-3'.

The following primers were used for 5'-upstream and 3'-downstream amplification of the *P. inflexum* genes. For 5'-upstream amplification: first walk for *Psod-2*, POCJ5GW1 (outer antisense) 5'- TTCAGGGTTTGGATTTGTCACAAGCAC-3' and POCJ5GW2 (inner antisense) 5'-TTGCTGCTATTCAAATTGCAGGATGCG-3'; second walk for *Psod-2*, POGWCJ5-2 (outer antisense) 5'-TCTTCGCACAGGGAATGAAAGGGCAAAG-3' and POGWCJ5-2N (inner antisense) 5'-ACGCACCTTTTCCCAGTCGCTGGTATGC-3'; for *Psod-3*, POCY5GW1 (outer antisense) 5'-TCGAATCTTTCATGCCTTCCCAACAAG-3' and POCY5GW2 (inner antisense primer) 5'-TCAGATTGACAGCGTTCCGAGTGTGAG-3'. For 3'-downstream amplification: for *Psod-2*, POCJ3GW1 (outer sense) 5'-TAAATGTGGGTTACACTGGCAGGAGAG-3' and POCJ3GW2 (inner sense) 5'-ACACCAAATTAAGAGCGGAAGGGATGC-3'; for *Psod-3*, POCY3GW1 (outer sense) 5'-AATGCCTGTGAGAGTTCTGGTTGATCC-3' and POCY3GW2 (inner sense) 5'-CCTAAATGCAATATGATGACGAGCAGC-3'.

Other methods. Cycle sequencing reaction and DNA sequencing, and other DNA and RNA manipulations were conducted as described previously (Kanematsu and Sato 2008, Kanematsu and Fujita 2009, Kanematsu et al. 2010). SOD activity assay, protein determination, Native-PAGE, SDS-PAGE, SOD activity staining, protein staining and immunoblotting were performed as before (Kanematsu and Asada 1990, Ueno and Kanematsu 2007).

RESULTS

cDNAs and genomic genes of the fern *E. arvense* CuZn-SOD isoforms

cDNAs. The cDNA cloning for CuZn-SOD isoforms of *E. arvense* was conducted by 5'- and 3'-RACE after elucidating the sequence in a core region of each cDNA with RT-PCR amplification. In RT-PCR, we employed the degenerate primers that were used for simultaneous amplification of both chloroplastic and cytosolic CuZn-SOD cDNAs from maize (Kanematsu and Fujita 2009) and for amplification of chloroplastic CuZn-SOD cDNA from *Spirogyra* (Kanematsu et al. 2010). By sequencing seven clones, RT-PCR product from poly(A)$^+$ mRNA of horsetail's green sterile stems contained only one type of 283-bp fragment (without primers) having a homology with those of chloroplastic CuZn-SOD. No fragment for cytosolic CuZn-SOD cDNA was found from the green tissue.

To obtain cytosolic CuZn-SOD cDNA fragments, RT-PCR was conducted using poly(A)$^+$ mRNA from the non-

Esod-1 cDNA (chloroplastic, major)

```
                                                                                                    50                                                                                                   100
AAGGAACAAG GGTCTATCTA ATCATCTTTC TCTCTCTCCC TCTCTCTCTT CTCTAGGTCT TCTTCTTGTT GATAATCCAT GGCTATGGCT ATGTCTGTAG
                                                                                                                                                                  M   A   M   A   M   S   V

                                                                                                   150                                                                                                   200
CTATGGCTTC CCCCTCCTCT TTGGCATTGC CCTCTTCTTC CTCCTCCTCC TCCCTCCTCG TCAAGTCCTC CTTCTCCGGA GCCCGCCTTG CCCTCCTGCC
A   M   A   S     P   S   S     L   A   L     P   S   S   S     S   S   S     S   L   L     V   K   S   S     F   S   G     A   R   L     A   L   L   P

                                                                                                   250                                                                                                   300
GGCTTGCTTT GTTTCCCCTC GCCTCCCCTC TCGCGCCCTC ACCGTCGTAG CCGCCGAGAA GAAGGCTGTC GCTGTCCTCA AAGGGACCTC CAATGTCGAA
  A   C   F     V   S   P     R   L   P   S     R   A   L     T   V   V     A   A   E   K     K   A   V     A   V   L     K   G   T   S     N   V   E
                                                                                                                A   E   K     K   A   V     A   V   L     K   G   T   S     N   V   E

                                                                                                   350                                                                                                   400
GGTGTCATCA ACTTGTTCCA GGAAGACGAC GGCCCCACAA CCGTGAAGGT CAAAATTTCC GGTTTAGCAC CTGGCAAGCA CGGATTCCAC CTCCACCAGT
  G   V   I     N   L   F   Q     E   D   D     G   P   T     T   V   K   V     K   I   S     G   L   A     P   G   K   H     G   F   H     L   H   Q
  G   V   I     N   L   F   Q     E   D   D     G   P   T     T   V   K   V     K   I   S     G   L   A     P   G   R   H     G   F   H     L   H   Q

                                                                                                   450                                                                                                   500
TTGGAGACAC CACCAATGGC TGCATGTCAA CAGGTCCTCA TTTTAACCCT CAGGGACTTA CACATGGTGC CCCTGAAGAT GAAGTTCGAC ATGCCGGTGA
F   G   D   T     T   N   G     C   M   S     T   G   P   H     F   N   P     Q   G   L     T   H   G   A     P   E   D     E   V   R     H   A   G   D
F

                                                                                                   550                                                                                                   600
TCTAGGGAAT GTTGTTGCGG GACCTGATGG TGTTGCAGAA GCCACCATAG TTGACTCACA AATTCCTTTG TCTGGACCTA ATTCCGTAAT TGGGAGAGCC
  L   G   N     V   V   A     G   P   D   G     V   A   E     A   T   I     V   D   S   Q     I   P   L     S   G   P     N   S   V   I     G   R   A

                                                                                                   650                                                                                                   700
TTTGTCATTC ATGAACTTGA GGACGATCTG GGAAAAGGAG GACACGAACT TAGCCCTACA ACTGGTAATG CTGGAGGTCG ACTTGCTTGT GGTATTGTGG
  F   V   I     H   E   L   E     D   D   L     G   K   G     G   H   E   L     S   P   T     T   G   N     A   G   G   R     L   A   C     G   I   V

                                                                                                   750                                                                                                   800
GTCTCTCTCC ATAAATTTTC TACAAGGAGA TATCAAGTTT AGTTGGGTGG GTTGGCTTCA TCAGTAAAGC AGAGTTGAAT ATATGTGATA AAGTGAAGTT
G   L   S   P

                                                                                                   850                                                                                                   900
GTAGATTTTT GAAGGCGGAG CCGAAAAAAA TAGAATGAGA ACAAGCATTG AATAAATAAA TTTTATTTAT TTGTGCGAGT CTCTTTTGAA ATTAAAAAAA

AAAAAAAAAA AAAAAAAAAA
```

Esod-2 cDNA (cytosolic)

```
                                                                                                    50                                                                                                   100
ACAGTTCGCC TGCGCTGTCT TGACCTAGGG TTTCTCTCTC TCCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCCCCCT GTTGTGTTGA GATGGGGTTG
                                                                                                                                                                  M   G   L

                                                                                                   150                                                                                                   200
AAGGCAGTGG CAGTGCTTTC CGGTAGCGCA GGAGTCGCTG GTGTCGTCCA TTTCTCTCAG GACACTCCCA ATGGCCCTAC GACTGTGGTT GGTTCTCTTA
  K   A   V     A   V   L   S     G   S   A     G   V   A     G   V   V   H     F   S   Q     D   T   P     N   G   P   T     T   V   V     G   S   L

                                                                                                   250                                                                                                   300
GCGGGCTTTC CCCTGGATTG CATGGATTTC ATGTCCATGC CCTTGGTGAC ACCACAAATG GATGCATGTC CACAGGTGCT CACTACAATC CTGCCAACAA
S   G   L   S     P   G   L     H   G   F     H   V   H   A     L   G   D     T   T   N     G   C   M   S     T   G   A     H   Y   N     P   A   N   K

                                                                                                   350                                                                                                   400
AGTGCATGGG GCTCCAGAGG ATGAAGATCG CCATGCTGGA GATTTAGGAA ATGTCACTGT CGGAGATGAT GGGAAGGCAC AGCTCTCAAT TACAGACTGC
  V   H   G     A   P   E     D   E   D   R     H   A   G     D   L   G     N   V   T   V     G   D   D     G   K   A     Q   L   S   I     T   D   C

                                                                                                   450                                                                                                   500
CAGATTCCTT TGGATGGACC TAATTCAATC ATTGGCCGTG CTGTTGTTGT ACATGCTGAT CCTGATGATC TAGGCAAGGG TGGTCATGAG TTGAGCAAAA
  Q   I   P     L   D   G   P     N   S   I     I   G   R     A   V   V   V     H   A   D     P   D   D     L   G   K   G     G   H   E     L   S   K

                                                                                                   550                                                                                                   600
CTACTGGAAA TGCTGGAGGA AGACTTGCGT GTGGTGTTAT TGGTCTTCAG GCTGCAGCTT GAAGGTGTAC CCCCTATATG CCTCTATAAG CTTTTTTGGT
T   T   G   N     A   G   G     R   L   A     C   G   V   I     G   L   Q     A   A   A

                                                                                                   650                                                                                                   700
CTTGTAATGA AGAAATAATA TATTTTCCGA GGGCAGTGAA ATAGAGAACC TGTCATACAA TCAAGCTGGG TTGCATACAT GTAGTTGCAA GCCAGAGAAT

                                                                                                   750                                                                                                   800
AAGTAGTTCA CATCTTTCTT GTTTTGAAAA GTGACAAGCA TATATGCAAA AATGACTGGT GGAATAAAGC AAGCTTGCCT TTATAGCTTT TAAGCTGTAA

                                                                                                   850
ATATGGATTG AAAACATTAC CCAGATTTTC TCATCAATGC TAATCTTCTA GTAATGTTAA AAAAAAAAAA AAAAAAAAAA AAAA
```

Esod-3 cDNA (chloroplastic, minor)

```
                                                                                                    50                                                                                                   100
CACTGTAACC CTGAAAATCA CCGGCCTCAC CCCTGGATTG CACGGATTCC ACTTGCATCA ATTTGGAGAT ACCACCAACG GATGCATGTC AACAGGCCCC
  T   V   T     L   K   I     T   G   L   T     P   G   L     H   G   F     H   L   H   Q     F   G   D     T   T   N     G   C   M   S     T   G   P

                                                                                                   150                                                                                                   200
CACTTCAACC CCAAGGGACT TACTCACGGT GCCCCCAGCG ACGAAATCAG ACACGCCGGC GATTTGGGAA ATCTTGTCGC GAATGATGAA GGTGTCGCCG
  H   F   N     P   K   G   L     T   H   G     A   P   S     D   E   I   R     H   A   G     D   L   G     N   L   V   A     N   D   E     G   V   A

                                                                                                   250
AAGCCACAAT CGTTGATTCA CAGATTCCCT TATCCGGCGA GAACTCTGTT GTCGGCAGGG CCTTTGTCGT CCATGAGCTT GAA
E   A   T   I     V   D   S     Q   I   P     L   S   G   E     N   S   V     V   G   R     A   F   V   V     H   E   L     E
```

Fig. 1. Nucleotide and deduced amino acid sequences of *Esod-1*, *Esod-2* and *Esod-3* cDNAs of *E. arvense*. The nucleotide sequences of 920 bp and 884 bp for *Esod-1* and *Esod-2* cDNAs, respectively, are presented with those of partial *Esod-3* cDNA. The deduced amino acid sequences are indicated in blue letters. In ESOD-1, a transit peptide is shown in green letters. The N-terminal amino acid sequence (50 residues) in red letters of horsetail chloroplastic CuZn-SOD (Kanematsu and Asada 1990) is aligned with the deduced sequence. One mismatched residue is indicated in light-blue letters.

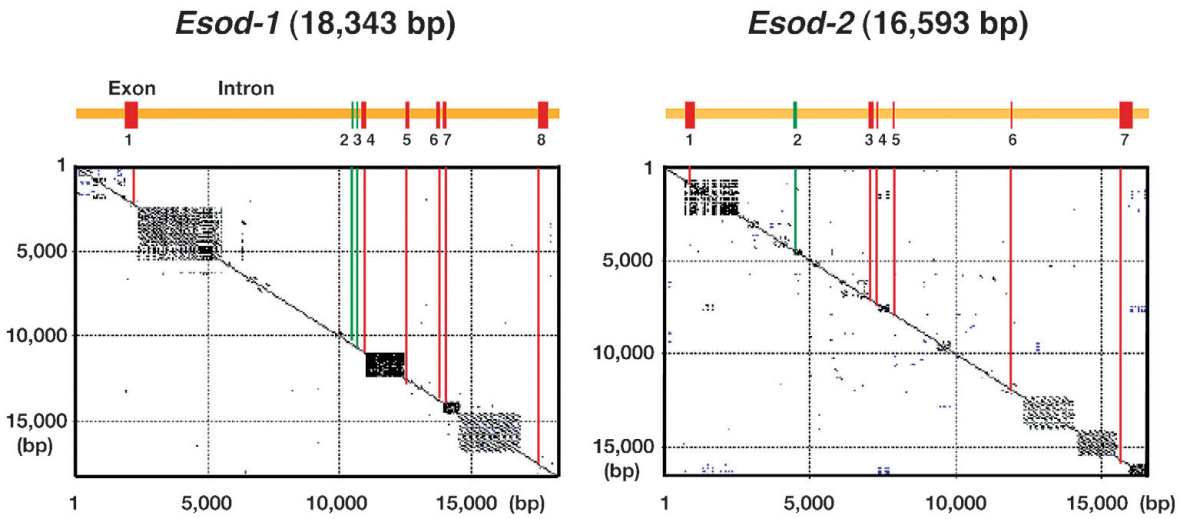

Fig. 2. Occurrence of repeated sequences in introns of *Esod-1* and *Esod-2* genomic genes from *E. arvense*. The repeated sequences in the genomic genes were analyzed by using the Pustell DNA matrix analysis program from the MacVector software with a DNA database matrix as a scoring matrix. The following conditions were used: window size = 30, minimum % score = 60, hash value = 6 and jump = 1. Dense areas of dots indicate where highly repeated sequences occur. The exon-intron structures of the genes are based on the results from Figure 8. The exon 2 and exon 3 in *Esod-1* are merged in exon 2 in *Esod-2*, which are shown in green color.

green fertile stems. Out of 10 clones analyzed, six clones were shown to be the same as those from the sterile stems (referred to as chloroplastic major type). In the remaining clones, three were classified as chloroplastic type (chloroplastic minor type) and one was cytosolic type. Thus, it is evident that *E. arvense* contains at least three genes for CuZn-SOD isoforms.

Complete cDNA sequences for the chloroplastic major type and cytosolic CuZn-SODs were obtained by 5'- and 3'-RACE using the adaptor primer (Clontech) and gene-specific primers designed to make an overlapping sequence. 0. 6-kb (5'-RACE) and 0. 7-kb (3'-RACE) fragments with overlapping regions of 164 bases were obtained for chloroplastic cDNA, and 0.5-kb (5'-RACE) and 0.7-kb (3'-RACE) fragments with overlapping 195 bp were obtained for cytosolic cDNA. By assembling the sequences, finally we obtained 920-bp chloroplastic major (*Esod-1*) and 884-bp cytosolic (*Esod-2*) CuZn-SOD cDNAs. Amplification in RACE for the chloroplastic minor (*Esod-3*) isoform cDNA was not successful. The cDNA nucleotide sequences of *Esod-1*, *Esod-2* and *Esod-3* genes, and their deduced amino acid sequences, are shown in Fig. 1.

Both *Esod-1* and *Esod-2* cDNAs contained the start and stop codons indicating their complete sequences in coding regions. *Esod-1* cDNA showed the sequences for a chloroplast transit peptide region and for the N-terminal region of mature chloroplastic CuZn-SOD protein (Kanematsu and Asada 1990). Thus, *Esod-1* cDNA was identified to be the cDNA for the previously purified chloroplastic CuZn-SOD. *Esod-2* cDNA, obviously devoid of the transit sequences region, was for the cytosolic SOD. The deduced amino acid sequences for ESOD-1 and ESOD-2 consisted of 211 and 156 amino acid residues, respectively. The mature ESOD-1 contained 153 residues, revealed by comparison with N-terminal sequences of the purified protein (Kanematsu and Asada 1990).

Chloroplastic major CuZn-SOD genomic gene (*Esod-1*). The genomic genes of chloroplastic major (*Esod-1*) and cytosolic (*Esod-2*) CuZn-SODs from *E. arvense* were obtained by the combination of three portions of the sequences amplified by PCR, i.e. central, upstream and downstream of the gene. PCR amplification for the central region of *Esod-1* using gene-specific primers based on the cDNA sequences gave a 15-kb fragment (primer set, EGGL-F255 vs. EGGL-B774). In the cloning processes, recovery of the 15-kb fragment (F255-B774) was unsuccessful due to its long size, but we obtained truncated fragments of 10 (clone #4-18) and 8 kb, whose 3'-ends were located in the intron between exon 5 and 6, and exon 3 and 4, respectively, while 5'-ends held complete sequences.

To characterize the sequence of the F255-B774 fragment, we divided it into two portions by PCR amplification from genomic DNA using two primer sets (EGGL-F255 vs. EGGL-B483 and EGGL-F504 vs. EGGL-B774). Thus, we obtained a 8.5-kb fragment (E2-E5) corresponding to the 5'-side portion of the 15-kb fragment, and a 7.5-kb fragment (E5-E9) corresponding to the 3'-side portion. Since the binding sites of primers of B483 and F504 were separated by 20 nucleotides in cDNA sequence, the fragments F255-B483 and F504-B774 amplified with those primers did not have an overlapping region. However, we confirmed that there was no intron in the 20-bp gap by using the truncated fragment (clone #4-18) from the F255-B774 fragment as a bridging clone.

The entire region of the F504-B774 fragment and one third of the F225-B483 fragment were sequenced by directional gene walking along the sequences of the

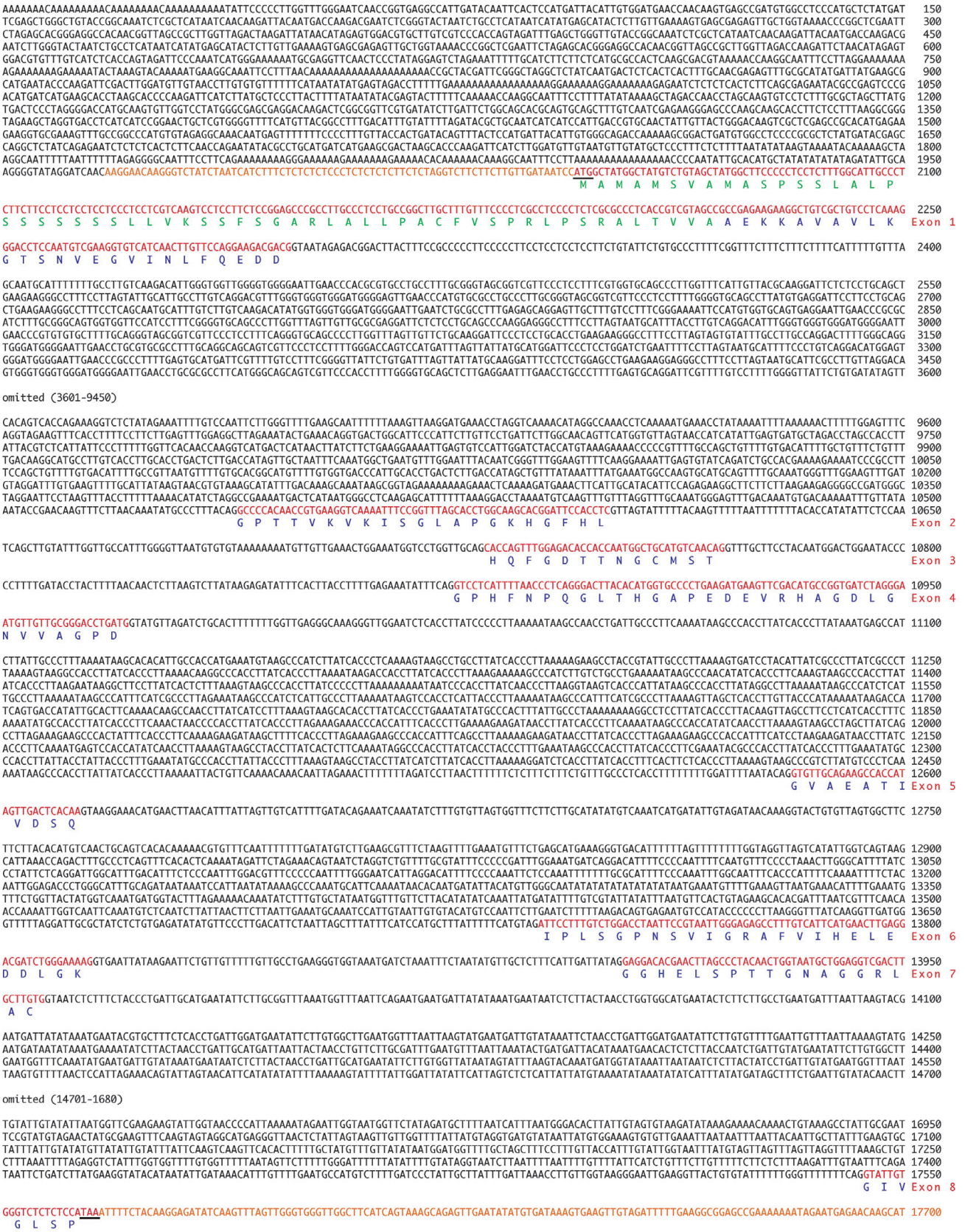

Fig. 3. Nucleotide sequence of *E. arvense* chloroplastic genomic gene (*Esod-1*). The nucleotide sequence of 18, 343 bp of the gene is presented. Exons and UTRs are shown in red and orange letters, respectively. Start and stop codons are underlined. The deduced amino acid sequence is indicated in blue letters, in which the chloroplast transit peptide is in green.

Fig. 4. Nucleotide sequence of *E. arvense* cytosolic genomic gene (*Esod-2*). The nucleotide sequence of 16,593 bp is presented. The notation used is the same as in Figure 3.

Psod-1 cDNA (chloroplastic, minor)

```
                                                                                                    50                                                                                                   100
ATCCCATTTG CCCCTTCTCA TTATTCTCCT CTCTCACTCT CTTCGCCTCC ATTGCCTTCT TCTTACTCAG CTTTCCTTCA GGGCGTGATT TTTTATGCAA
                                                                                                                                                                                          M  Q

                                                                                                   150                                                                                                   200
GCGTTGGTAG CGGCGATGGC GGCGACGGCA ATGACAATGT CCACTTTGTC GACGTCTCTG GTGGCACCGG CCTCACAGTC CGCGTTTCAA GGCGTGGCTG
 A  L  V   A  A  M  A   A  T  A   M  T  M   S  T  L  S   T  S  L   V  A  P   A  S  Q  S   A  F  Q   G  V  A

                                                                                                   250                                                                                                   300
TGCAGCGCCA TTATGTGCCC ATGCTGGCCA TGGCCAAGGG CCGGTCGCTC ACCATCACGG CCATGGCCAC CAAGAAGGCC GTGGCTGTGC TCAAGGGCAC
V  Q  R  H   Y  V  P   M  L  A   M  A  K  G   R  S  L   T  I  T   A  M  A  T   K  K  A   V  A  V   L  K  G  T

                                                                                                   350                                                                                                   400
CTCCAACGTC GAGGGCGTCG TCACTCTTCT CCAGGAAGAC GATGGTCCAA CGACAGTCAG TGTGAAGATA TCTGGCCTCA CTCCTGGAAA ACATGGGTTC
  S  N  V   E  G  V   V  T  L  L   Q  E  D   D  G  P   T  T  V  S   V  K  I   S  G  L   T  P  G  K   H  G  F

                                                                                                   450                                                                                                   500
CATCTTCATC AGTTTGGTGA CACGACCAAT GGCTGCATGT CGACTGGTCC TCATTTTAAT CCGGAGGGAA AAACACACGG TGCCCCAGGG GATATCGACC
 H  L  H   Q  F  G  D   T  T  N   G  C  M   S  T  G  P   H  F  N   P  E  G   K  T  H  G   A  P  G   D  I  D

                                                                                                   550                                                                                                   600
GCCACGCAGG GGACCTGGGC AATGTGGTGG CTGGTGACAA TGGTGTTGTC GAGGTAGAAC TCTCGGACTC TCAGATCCCT CTGTCTGGTC CTAACTCTGT
R  H  A  G   D  L  G   N  V  V   A  G  D  N   G  V  V   E  V  E   L  S  D  S   Q  I  P   L  S  G   P  N  S  V

                                                                                                   650                                                                                                   700
TGTCGGTAGA GCATTTGTCA TCCACGAACT CGAAGATGAT CTTGGAAAGG GTGGCCATGA GCTCAGCTCC ACCACAGGAA ATGCAGGAGG CCGTCTGGCT
  V  G  R   A  F  V   I  H  E  L   E  D  D   L  G  K   G  G  H  E   L  S  S   T  T  G   N  A  G  G   R  L  A

                                                                                                   750                                                                                                   800
TGCGGTGTTG TGGGACTGAC ACCACTGTAA GCACTTGAGC TGTAAGGGCA CTACTTCATC TCAGCTTTCT TATATATGAA TGGGGAGTTG CGTTGTCGGG
 C  G  V   V  G  L  T   P  L

                                                                                                   850                                                                                                   900
CGGCATGTAC ATTGTGGATG GAGGTCACAT GGAGTAAATA ATCTCCACCT TTTTGCCGGA GTTTTAGTTC GCTTGAATGG AATCCTGACT GGCACCGAGG

                                                                                                   950                                                                                                  1000
GAGTGAGGGA GTGATGGTAG GCCTATTTTG TAGAGATTCG CAAGTATTTG CACTGTACTT TCAACGTGAA CGGCAATATC ATTCATTTTG CTTAGTTGCT

                                                                                                  1050
TGCGTACCAA ACTATTCTAG TTTTTGCCAA TGTATCTGTC ATTTTCAATA TTCTTTCCCA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
```

Psod-2 cDNA (chloroplastic, major)

```
                                                                                                    50                                                                                                   100
ATTGCCTTTG CCTTTCATTC CCTGTGCGAA GAAGTTTACC ATTGCCTTCT CACACATTCT AACGTAGTCT CATCCCATAG TCTGCCTTCT CATCTTCTTC

                                                                                                   150                                                                                                   200
GTCACCCATT CTTCTTCAAC CTCAAACCAA ACCAAGGAGG TAACCATGGC AGCCACATCT ATGGCGTTGT CATCATCCCT TGTCACTCCC ACTGTGGCAG
                                                                    M  A   A  T  S   M  A  L   S  S  S  L   V  T  P   T  V  A

                                                                                                   250                                                                                                   300
CTTCAAAGTC TGGATTCCAG GGAGTTGCTG TGCGAGTCAG CCATGTGCCA ATGATGGCCA AGAGCAACAA CCACCGCAAT CTTACCATTG TGGCTGCTAC
A  S  K  S   G  F  Q   G  V  A   V  R  V  S   H  V  P   M  M  A   K  S  N  N   H  R  N   L  T  I   V  A  A  T

                                                                                                   350                                                                                                   400
CAAGAAGGCT GTGGCTGTGC TCAAGGGTAC ATCCAATGTG GAGGGTGTCG TCAACCTTAT CCAGGAAGAT GATGGTCCTA CAACTGTGAA TGTGAAAATC
  K  K  A   V  A  V   L  K  G  T   S  N  V   E  G  V   V  N  L  I   Q  E  D   D  G  P   T  T  V  N   V  K  I

                                                                                                   450                                                                                                   500
ACTGGCCTTA CTCCTGGCAA GCATGGATTC CATCTTCATG AATTTGGTGA CACGACCAAT GGCTGCATCT CCACTGGTCC TCACTTCAAT CCTAAGGGCA
 T  G  L   T  P  G  K   H  G  F   H  L  H   E  F  G  D   T  T  N   G  C  I   S  T  G  P   H  F  N   P  K  G

                                                                                                   550                                                                                                   600
AAACCCATGG TGCTCCAGGA GACGAAATTC GTCATGCTGG AGATTTGGGC AATGTTGTGG CTGGACAAGA TGGTGTAGTG GAGGTGACAC TCACTGACGA
K  T  H  G   A  P  G   D  E  I   R  H  A  G   D  L  G   N  V  V   A  G  Q  D   G  V  V   E  V  T   L  T  D  D

                                                                                                   650                                                                                                   700
TCAGATCCCT TTGTCTGGTC CTACTTCTGT GGTGGGAAGG GCGTTTGTCA TCCACGAACT TGAGGATGAC CTTGGAAAGG GTGGGCACGA ACTCAGCTCC
  Q  I  P   L  S  G   P  T  S  V   V  G  R   A  F  V   I  H  E  L   E  D  D   L  G  K   G  G  H  E   L  S  S

                                                                                                   750                                                                                                   800
ACGACTGGAA ATGCAGGTGG CCGCCTTGCT TGTGGTGTGG TGGGGTTGAC ACCAAATTAA GAGCGGAAGG GATGCATATC CATGAGCTTG TTCACTTAAC
 T  T  G   N  A  G  G   R  L  A   C  G  V   V  G  L  T   P  N

                                                                                                   850                                                                                                   900
TGATTTTGAC ACCAAGATTG ATTGGAGGTT AGGGACCAGG TACAACTGGC ATGCGAAACG ACATGGAGAG TTCTGTCCGA AGTAAATAAC ATTCACCTCT

                                                                                                   950                                                                                                  1000
TTCTCTTTAA TGTATGGTGG GGTTCTGTAA ATGTGTTATA ATATCTGCTG TACTTGTGCA TTGTGGGGAC GTTAATATCA TTCAACCTTA CTGTTCACTT

                                                                                                  1050
ATCTTCAAGG TGGTGACTGA TGTTCTGCAA ATTATTATGA GCCCATTCAT CTTTATTTCG TAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA
```

Fig. 5. Nucleotide and deduced amino acid sequences of *Psod-1*, *Psod-2* and *Psod-3* cDNAs of *P. inflexum*. The nucleotide sequences of 1,090 bp, 1,093 bp and 865 bp for *Psod-1*, *Psod-2* and *Psod-3* cDNAs, respectively, are presented. The deduced amino acid sequences are indicated in blue letters.

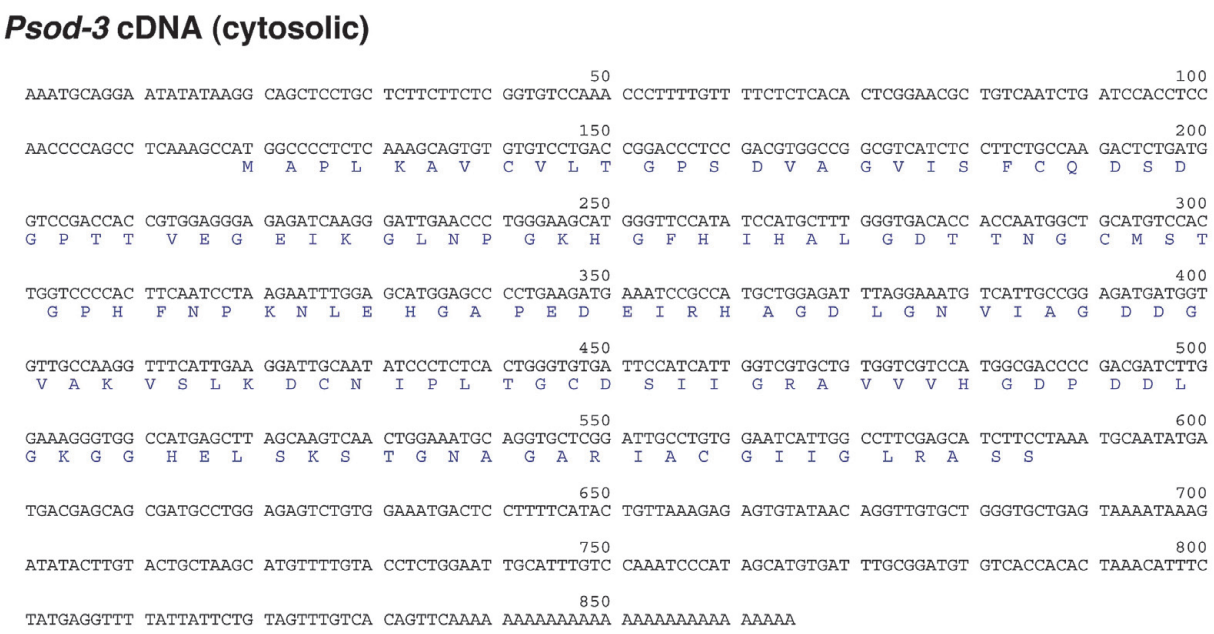

Fig. 5. ***Continued.***

clones using gene-specific sequence primers. However, the central portion of F225-B483 contained highly repeated sequences (Fig. 2) and the walking method was not applicable. Thus, this region was analyzed using a series of 3'-end deleted clones of F225-B483 and vector-specific sequence primers. Finally, we characterized the whole sequence of the 15-kb fragment.

Upstream and downstream fragments adjacent to the central region of *Esod-1* were amplified from uncloned, adaptor-ligated genomic DNA fragment pools treated by restriction enzyme (referred to as GenomeWalker library) and gene-specific primers based on both end sequences of the central fragment. For the chloroplastic SOD gene, we finally selected a fragment of 2.3 kb amplified from the *Dra* I GenomeWalker library for the 5' upstream region, and a fragment of 0.5 kb from the *Stu* I library for the 3' downstream region. Thus, we obtained 18,343 bp of a genomic gene (*Esod-1*) encoding chloroplastic CuZn-SOD by combination of 5'-side, central and 3'-side DNA fragments (Fig. 3).

Cytosolic CuZn-SOD genomic gene (*Esod-2*). A 10-kb DNA fragment for the central region of *Esod-2* was amplified using gene-specific primers (EQGY-F1 vs. EQGY-B529) based on the *Esod-2* cDNA sequence. To analyze the whole region of the 10-kb fragment, two shorter fragments were generated by PCR from the genomic DNA. We obtained fragments of 7 kb and 8 kb using the primer sets, EQGY-F1 vs. EQGY-B309 and EQGY-F175 vs. EQGY-B529, respectively, with an overlapping region of 5 kb. Since the 10-kb fragment lacked the region of the last exon, a 4-kb fragment corresponding to the region between the last exon and its neighboring one was obtained using EQGY-F495 and EQGY-B655. The 5'-upstream (0.8 kb) and 3'-downstream (0.9 kb) fragments of *Esod-2* were amplified from the *Stu* I and *Dra* I GenomeWalker libraries, respectively. Thus, by assembling 5'-upstream, central and 3'-downstream sequences, we obtained a cytosolic CuZn-SOD genomic gene (*Esod-2*) of 16,593 bp (Fig. 4).

cDNA and genomic genes of the moss *P. inflexum* CuZn-SOD isoforms

cDNAs. cDNA cloning for chloroplastic and cytosolic CuZn-SOD isoforms in *P. inflexum* was conducted by methods similar to those used for *E. arvense* SOD genes as described above. RT-PCR using the degenerate primers gave 283-bp DNA fragments (without primers) encoding chloroplastic or cytosolic CuZn-SOD. After being cloned, the analyzed nine clones were classified into three types: chloroplastic (*Psod-2*, termed chloroplastic major, five clones), another chloroplastic (*Psod-1*, termed chloroplastic minor, one clone) and cytosolic (*Psod-3*, three clones) type of CuZn-SOD gene. Thus, it is clear that *P. inflexum* contains two types of chloroplastic CuZn-SOD gene in accordance with the case of *E. arvense*.

Complete cDNA sequences for each type of SOD isoform were obtained by 5'- and 3'-RACE. For *Psod-2* cDNA, 0.62-bp and 0.52-bp fragments were amplified in 5'- and 3'-RACE, respectively. Similarly, for *Psod-1* and *Psod-3*, a set of 0.62-bp and 0.52-bp fragments and of 0.46-bp and 0.45-bp fragments were obtained, respectively. Finally, we assembled 1,093-bp, 1,090-bp and 865-bp fragments for *Psod-2*, *Psod-1* and *Psod-3* cDNAs, respectively. Figure 5 shows the nucleotide and deduced amino acid sequences of *P. inflexum* CuZn-SOD cDNAs. The deduced amino acid sequences for PSOD-2, PSOD-1 and PSOD-3 consisted of 204, 211 and 156 amino acid residues, respectively. The mature PSOD-2 and PSOD-1 were estimated to have 154 residues by amino acid sequence alignments with other chloroplastic SODs.

Fig. 6. Nucleotide sequence of *P. inflexum* chloroplastic genomic gene (*Psod-2*). The nucleotide sequence of 7,146 bp is presented. The notation used is the same as in Figure 3.

Chloroplastic major CuZn-SOD genomic gene (*Psod-2*). Amplification of the central region of *Psod-2* was performed using gene-specific primers (POGLJ-F14 vs. POGLJ-B853). A fragment of 4.5 kb was obtained and the whole region was sequenced by directional gene walking after cloning into vector. At least three clones were used for sequencing.

The sequences adjacent to the central region of the gene were analyzed by the GenomeWalker DNA walking method (Clontech) as described above. In the first walk to find the adjacent sequence to 5'-end of the central sequence of *Psod-2* using the gene-specific primers POCJ5GW1 and POCJ5GW2, we obtained only a 0.6-kb fragment from the *Stu* I GenomeWalker library. To obtain a longer sequence, we conducted a second walk using gene-specific primers derived from the first walk sequence, POGWCJ5-2 and POGWCJ5-2N. Fragments containing the upstream sequence of *Psod-2*, 3 kb and 1.8 kb, were obtained from the *Dra* I and *Eco* RV libraries, respectively. Thus, the 1.8-kb fragment was fully sequenced. The 3'-downstream sequences were found in a 0.7-kb fragment from the *Pvu* II library, 4 kb from the *Eco* RV library, and 5 kb from the *Stu* I library. The downstream sequence of 580 bp was analyzed with the 4-kb fragment. Finally, we sequenced 7,146 bp for *Psod-2* (Fig. 6).

Cytosolic CuZn-SOD genomic gene (*Psod-3*). Amplification for the central sequence of *Psod-3* with the primer set POGY-F23 vs. POGY-B638 gave a DNA fragment of 3.6 kb. In spite of the effort to obtain more clones for sequencing, only one clone was obtained by cloning. Therefore, the base on the sequence of this clone, two shorter fragments, 4.7 kb on the 5'-side and 1.9 kb on the

Fig. 7. Nucleotide sequence of *P. inflexum* cytosolic genomic gene (*Psod-3*). The nucleotide sequence of 5,583 bp is presented. The notation used is the same as in Figure 3.

3'-side, that cover the whole 3.6-kb fragment were amplified from the genome using the primer sets POGY-F23 vs. POGY-B1742, and POGY-F1709 vs. POGY-B638, respectively. Thus, the 3.6-kb sequence for the central region of *Psod-3* was analyzed using more than three clones.

We found a 2.7-kb fragment containing the upstream sequence adjacent to the central sequence from the *Eco* RV GenomeWalker library using primers POCY5GW1 and POCY5GW2. A downstream DNA fragment of 3 kb was obtained from the *Eco* RV library with primers POCY3GW1 and POCY3GW2. We sequenced 1.7 kb in the 3'-side region of the 2.7-kb upstream fragment and 0.5 kb in the 5'-side region of the 4-kb downstream fragment. Finally, a 5,583-bp sequence in total for *Psod-3* was characterized (Fig. 7).

DISCUSSION

cDNAs of higher plant CuZn-SODs have been obtained from many sources, and data on the whole genome of seed plants including *Arabidopsis thaliana* and rice are also available in public databases. Furthermore, the number of whole genomes of higher plants that have been completely sequenced is rapidly growing. However, there is not as much sequence data available for CuZn-SOD genes in lower plants; when we started this analysis of CuZn-SOD genomic genes from ferns and mosses, cDNA sequences had been reported only for several mosses (Tanaka et al. 1998). In this work, we characterized the cDNAs and genomic genes of CuZn-SOD isoforms from the fern *E. arvense* and the moss *P. inflexum* to clarify the response of SOD isoform genes to environmental changes in the early phase of land plant evolution. To the best of our knowledge, elucidation of the structure of *E. arvense* genomic genes may be the first instance for fern CuZn-SOD genes. Data on the whole genome of the moss *Physcomitrella patens* (Rensing et al. 2008) is available in a database.

Assignment of gene products of CuZn-SOD isoform genes from *E. arvense* and *P. inflexum*. We found three CuZn-SOD isoform genes in *E. arvense* by RT-

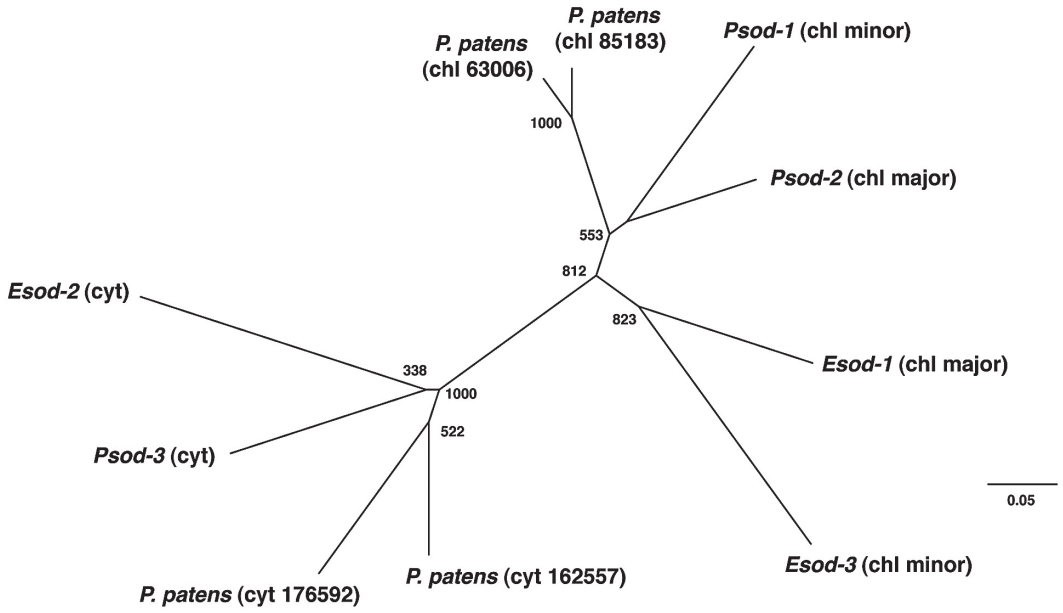

Fig. 8. Phylogenetic tree of two chloroplastic CuZn-SOD genes of *E. arvense, P. inflexum* and *P. patens* with their cytosolic genes. The 283-bp nucleotide sequences from the CuZn-SOD cDNAs of the fern and moss were aligned by the ClustalW method. Since *Esod-3* was a 283 bp DNA fragment, the other cDNAs were truncated to the same sizes for ClustalW comparison. The phylogenic tree was constructed with the neighbor-joining method using a distance matrix of Kimura and bootstrapping of 1,000 replications. An unrooted tree is represented. The bootstrap values are indicated at each node.

chl, chloroplastic; cyt, cytosolic. *Esod-1* (chl major) and *Esod-3* (chl minor), *E. arvense* chloroplastic SOD genes; *Psod-1* (chl minor) and *Psod-2* (chl major), *P. inflexum* chloroplastic SOD genes; *P. patens* (chl 85183) and *P. patens* (chl 63006), *P. patens* chloroplastic SOD genes with protein IDs (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html). *Esod-2* (cyt) and *Psod-3* (cyt), *E. arvense* and *P. inflexum* cytosolic SOD genes, respectively; *P. patens* (cyt 162557) and *P. patens* (cyt 176592); *P. patens* cytosolic SOD genes with protein IDs.

PCR in accordance with the previous results of native-PAGE, in which three cyanide-sensitive activity bands were identified to be CuZn-SOD (Kanematsu and Asada 1989). In the native-PAGE (see Fig. 1 and 7 in Kanematsu and Asada 1989), the fast-moving band was ascribed to chloroplastic CuZn-SOD by its purification. In the present work, it was confirmed that the purified chloroplastic SOD is a gene product of *Esod-1* (Fig. 1). The gene for the cytosolic SOD corresponding to the middle band, which was immunologically estimated, might be *Esod-2*. The top band of CuZn-SOD activity may be derived from *Esod-3*, the chloroplastic minor SOD gene, although the band exhibited weak cross-reactivity with anti-spinach cytosolic CuZn-SOD serum rather than with anti-spinach chloroplastic SOD serum. However, we could not exclude the possibility that the translation of *Esod-3* is repressed and the activity of *Esod-3* product is below the limit of detection.

P. inflexum chloroplastic CuZn-SOD showed a major activity band in native-PAGE and was immunologically identified (Ueno and Kanematsu 2007). Therefore, this SOD might be a product of *Psod-2*, a chloroplastic major SOD gene. However, the separation of other SOD isoforms was poor due to unknown reasons. Thus, assignment of the gene products of *Psod-1* and *Psod-3* was unsuccessful.

Occurrence of two chloroplastic CuZn-SOD genes in ferns and mosses. Chloroplastic CuZn-SOD plays an important role in the water-water cycle of chloroplasts, where excess photon energy is dissipated through the generation of superoxide, then disproportionation of superoxide to hydrogen peroxide, and reduction of hydrogen peroxide to water by several enzymes including SOD (Asada 2006). To facilitate the conversion of superoxide to hydrogen peroxide, CuZn-SOD is weakly associated with thylakoid membranes at the site of superoxide generation (Ogawa et al. 1995). Since the amino acid substitution rate in chloroplastic CuZn-SODs is slower than that of cytosolic CuZn-SODs (Kanematsu and Asada 1990), the chloroplastic CuZn-SOD seems to have remained under pressure from the requirements for membrane attachment, which is in accordance with the case of cytochrome *c* in mitochondria (Dickerson 1971).

As to chloroplastic CuZn-SOD, *E. arvense* and *P. inflexum* contain two types of chloroplastic CuZn-SOD gene, while higher plants contain one type of the gene. Recent genome analysis of the moss *Physcomitrella patens* (Rensing et al. 2008) also confirmed the presence of two chloroplastic CuZn-SOD genes in addition to two cytosolic SOD genes. Thus, it seems likely that ferns and mosses possess two chloroplastic CuZn-SOD genes. Phylogenetic analysis clearly showed that the two chloroplastic SOD genes in each organism were generated by gene duplication after separation of each species in pteridophytes and bryophytes (Fig. 8). Since

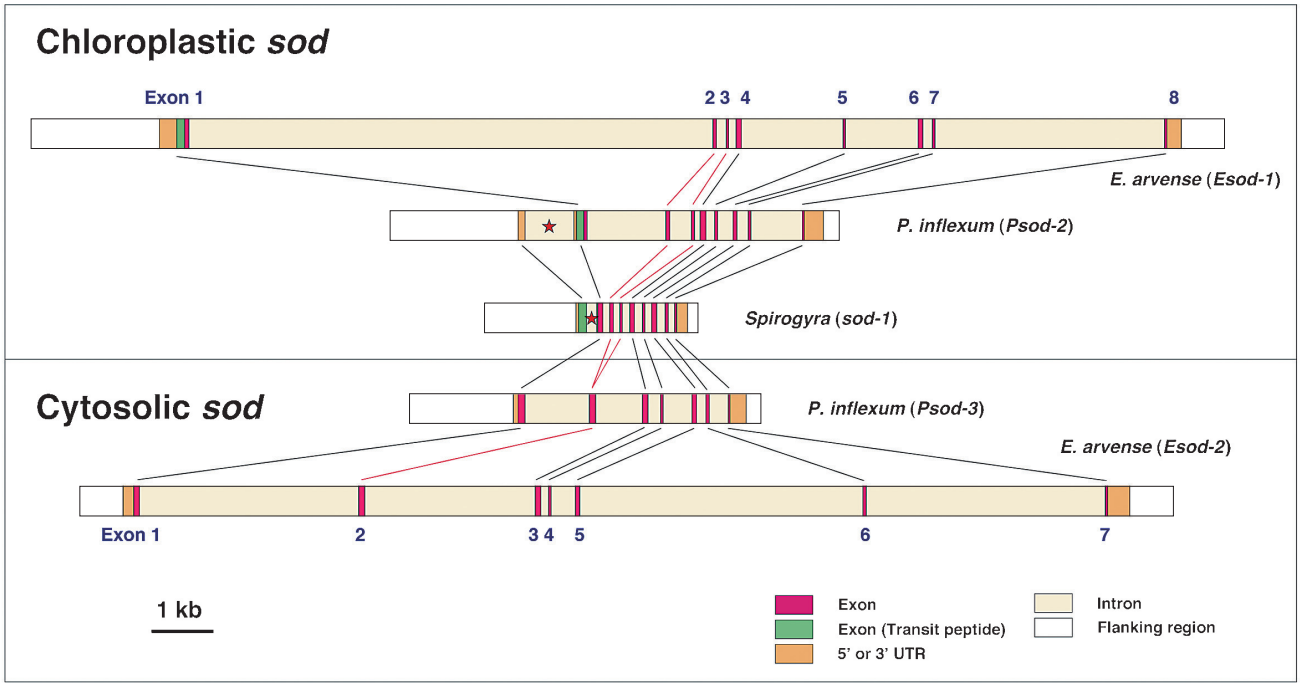

Fig. 9. Comparison of exon-intron structure of chloroplastic and cytosolic CuZn-SOD genes of *E. arvense*, *P. inflexum* and *Spirogyra*. *E. arvense Esod-1*, chloroplastic major gene; *E. arvense Esod-2*, cytosolic gene; *P. inflexum Psod-2*, chloroplastic major gene; *P. inflexum Psod-3*, cytosolic gene; *Spirogyra sod-1*, chloroplastic gene (AB098508, Kanematsu et al. 2010). The exon number is indicated on each schematic drawing of the *E. arvense* chloroplastic and cytosolic CuZn-SOD genes. The solid line shows the corresponding exon and the red line indicates merged exons. The stars in red indicate the introns that are absent in fern and higher plant chloroplastic CuZn-SOD genes.

the seed plants contain only one type of chloroplastic CuZn-SOD gene, duplication of the gene might not have occurred during the divergence of spermatophytes. The difference in physiological function of these isoforms in chloroplasts is not known. However, it is interesting to note that the rates of evolution of *Psod-1* and *Esod-3* were faster than those of *Psod-2* and *Esod-1* (Fig. 8), indicating that the latter is associated with thylakoid membrane.

Repeated sequences in exons of chloroplastic and cytosolic CuZn-SOD genes of *E. arvense*. It is well known that the chromosome numbers of ferns are significantly larger than those of angiosperms (e. g., horsetail 2n = 216), although the number is not necessarily proportional to the size of genome. The length of chloroplastic *Esod-1* and cytosolic *Esod-2* is three to four times longer than those of *Psod-2* and *Psod-3*, indicating the presence of longer introns in the *E. arvense* genes. We found the presence of highly repeated sequences in several introns, and this causes the longer sequence of the genes (Fig. 2). The repetitious region was composed of mainly short tandem repeats. Neither LINE nor SINE was observed in these areas. On the other hand, the repeated sequences found in the introns of *Esod-1* and *Esod-2* were not detected in those of *Psod-2* and *Psod-3*.

Comparison of exon-intron structure of chloroplastic and cytosolic CuZn-SOD genes from *E. arvense* and *P. inflexum*. Chloroplastic *Esod-1* and *Psod-2* are composed of eight and nine exons, respectively (Fig. 9). The intron corresponding to the first one in *Psod-2* is absent in *Esod-1*, whereas the other exons of both the genes are the same in respect of their length and splicing sites. The exon-intron structure of *Esod-1* is the same as those of higher plants including maize chloroplastic CuZn-SOD gene (Kanematsu and Fujita 2009). On the contrary, the structure of *Psod-2* resembles that of *Spirogyra* chloroplastic CuZn-SOD gene (Kanematsu et al. 2010) in terms of the splicing sites except the position of the first intron (Fig. 9): the first intron is located in 5'-UTR of *Psod-2* whereas it is located in the coding region for the transit peptide of the *Spirogyra* gene. Thus, the deletion and insertion of an intron in the 5'-UTR and CDS region might have been involved in the migration of the insertion site of the first intron of algal and fern SOD genes during their evolution. The two chloroplastic CuZn-SOD genes of *P. patens* also contain an intron in 5'-UTR which corresponds to the first intron of *Psod-2* (Rensing et al. 2008). However, in both the *P. patens* genes the intron between exon 6 and 7 (termed *P. inflexum*) is absent. Since the chloroplast genes of ferns and higher plants contain the corresponding intron, this intron might have been lost prior to duplication of the chloroplastic gene in *P. patens*. On the other hand, the exon-intron structures of cytosolic *Esod-2* and *Psod-3*, consisting of seven exons, are the same among mosses and ferns (Fig. 9).

Concluding remarks. To study the molecular evolution of genes, it is important not only to align the nucleic acid or amino acid sequences of target genes to construct a phylogenetic tree but also to compare their exon-intron structures, because in some cases, evolution leaves its footprint in an exon-intron structure. In this work, we characterized the cDNAs and genomic genes of CuZn-SOD isoforms from ferns and mosses, and estimated their molecular evolution by analyzing their exon-intron structures.

Chloroplastic and cytosolic CuZn-SOD genes are paralogs: thus, when did they separate and how did they evolve? Previously, we estimated that both genes had diverged from a common ancestor, probably in an ancestor of prasinophyte algae, at a very early phase in the divergence of streptophyte and chlorophyte algae (Kanematsu et al. 2010). The exon-intron structure of the prasinophyte algal chloroplastic CuZn-SOD gene (*Ostreococcus lucimarinus*) greatly differs from that of the green algal SOD gene (*Spirogyra*). The former has two introns in the CDS region while the latter has nine and exhibits a similar structure to that of *Psod-2*, as revealed by the present work. These results suggest that intron acquisition by CuZn-SOD genes occurred mostly during the evolution of prasinophytes and prior to separation to the chloroplastic and cytosolic isoform genes.

From the present data, we propose the molecular evolution of CuZn-SOD genes during the early phase of land plant evolution as follows: after separation into chloroplastic and cytosolic genes, which had a similar exon-intron structure, both genes might have evolved independently. In case of the chloroplastic gene, when mosses appeared, the deletion and insertion of the first intron occurred. Then, when ferns diverged, the first intron was lost. Thereafter, the structure of ferns' genes has been basically conserved in the higher plants. In the case of the cytosolic gene, one intron in the CDS region was removed, resulting in the merging of two exons when mosses emerged (Fig. 9) and this structure has remained up to those of seed plants.

We obtained sequence information for the promoter regions of CuZn-SOD genes from ferns and mosses. The regulation of the SOD genes upon environmental stresses, with which organisms in the early phase of land plant evolution adapted to the high ROS-producing environment, remains to be elucidated.

REFERENCES

Asada, K. (1999) The water-water cycle in chloroplasts: scavenging of active oxygens and dissipation of excess photons. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**: 601-639.

Asada, K. (2006) Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiol.* **141**: 391-396.

Becker, B. and Marin, B. (2009) Streptophyte algae and the origin of embryophytes. *Ann. Bot.* **103**: 999-1004.

Bhattacharya, D. and Medlin, L. (1998) Algal phylogeny and the origin of land plants. *Plant Physiol*. **116**: 9-15.

Cuvelier, M. L. et al. (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic photoplankton. *Proc. Natl. Acad. Sci. USA* **107**: 14679-14684.

Dickerson, R.E. (1971) The structure of cytochrome *c* and the rates of molecular evolution. *J. Molec. Evolution* **1**: 26-45.

Dupont, C. L., Neupane, K., Shearer, J. and Palenik, B (2008) Diversity, function and evolution of genes coding for putative Ni-containing superoxide dismutases. *Environ. Microbiol.* **10**: 1831-1843.

Fridovich, I. (1995) Superoxide radical and superoxide dismutases. *Annu. Rev. Biochem.* **64**: 97-112.

Kanematsu, S. (2005) Characterization of superoxide dismutase isozyme genes from a green alga, moss and fern: implication for their molecular evolution. *In* Proceedings of the 13th International Congress on Genes, Gene Families and Isozymes. Edited by Xue, G. Y., Zhu, Z. Y. and Wan, Y. L. pp. 67-70. Medimond S.r.l., Bologna.

Kanematsu, S. and Asada, K. (1989) CuZn-superoxide dismutases from the fern *Equisetum arvense* and the green alga *Spirogyra* sp.: Occurrence of chloroplast and cytosol types of enzyme. *Plant Cell Physiol.* **30**: 717-727.

Kanematsu, S. and Asada, K. (1990) Characteristic amino acid sequences of chloroplast and cytosol isozymes of CuZn-superoxide dismutase in spinach, rice and horsetail. *Plant Cell Physiol.* **31**: 99-112.

Kanematsu, S. and Asada, K. (1994) Superoxide dismutase. *In* Molecular Aspects of Enzyme Catalysis. Edited by Fukui, T. and Soda, K. pp. 191-210. Kodansha/VHC, Tokyo.

Kanematsu, S. and Fujita, K. (2009) Assignment of new cDNA for maize chloroplastic CuZn-superoxide dismutase (SOD-1) and structural characterization of *sod-1* genes. *Bull. Minamikyushu Univ.* **39A**: 89-101.

Kanematsu, S., Fujita, K., Shinpei, U. and Asada, K. (2003) Exon-intron structure of the chloroplastic CuZn-SOD genes in eukaryotic alga and land plants: its implication in the molecular evolution. *In* Abstracts of Plant Biology 2003. American Society of Plant Biologists, USA. Oxidative stress, Abstr. #114.

Kanematsu, S., Iriguchi, N. and Ienaga, A. (2010) Characterization of CuZn-superoxide dismutase gene from the green alga *Spirogyra* sp. (Streptophyta): Evolutionary implications for the origin of the chloroplastic and cytosolic isoforms. *Bull. Minamikyushu Univ.* **40A**: 65-77.

Kanematsu, S. and Sato, S. (2008) Cloning of Fe-superoxide dismutase gene from the diazotroph *Azotobacter vinelandii* and the stimulation of its expression under anaerobic conditions in *Escherichia coli. Bull. Minamikyushu Univ.* **38A**: 7-18.

Ogawa, K., Kanematsu, S., Takebe, K. and Asada, K. (1995) Attachment of CuZn-superoxide dismutase to thylakoid membranes at the site of superoxide generation (PSI) in spinach chloroplasts: detection by immuno-gold labeling after rapid freezing and substitution method. *Plant Cell Physiol.* **36**: 565-573.

Okayasu, M., Ueno, S., Tokuyama, Y., Nakamura, M. and Kanematsu, S. (2004) Chloroplastic CuZn-SOD gene from a moss possesses the exon-intron structure of algal type. *Plant Cell Physiol.* **45** (supplement): 807.

Rensing, S. A. et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**: 64-69.

Tanaka, K., Takio, S., Yamamoto, I. and Satoh, T. (1998) Characterization of a cDNA encoding CuZn-superoxide dismutase from the liverwort *Marchantia paleacea* var. *diptera. Plant Cell Physiol.* **39**: 235-240.

Ueno, S. and Kanematsu, S. (2004) Gene structure of CuZn-superoxide dismutase isozymes from the fern *Equisetum arvense*. *In* Abstracts of the 77th Annual Meeting of the Japanese Biochemical Society. The Japanese Biochemical Society, Tokyo, Japan, Abstr. p. 874.

Ueno, S. and Kanematsu, S. (2007) Immunological and electrophoretic characterization of proteins exhibiting superoxide dismutase activity in the moss *Pogonatum inflexum. Bull. Minamikyushu Univ.* **37A**: 1-9.

Youn, H.-D., Kim, E.-J., Roe, J.-H., Hah, Y.C. and Kang, S.-O. (1996) A novel nickel-containing superoxide dismutase from *Streptomyces* spp.. *Biochem. J.* **318**: 889-896.