Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 366 (2006) 256–265
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Molecular evolution of the AP2 subfamily

Mikao Shigyo [a], Mitsuyasu Hasebe [b,c], Motomi Ito [a,*]

[a] *Department of General System Studies, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan*
[b] *National Institute for Basic Biology, Okazaki 444-8585, Japan*
[c] *Department of Molecular Biomechanics, The Graduate University for Advanced Studies, Okazaki 444-8585, Japan*

Received 22 April 2005; received in revised form 1 July 2005; accepted 8 August 2005
Available online 4 January 2006
Recieved by Takashi Gojobori

Abstract

The AP2 (APETALA2)/EREBP (Ethylene Responsive Element Binding Protein) multigene family includes developmentally and physiologically important transcription factors. *AP2/EREBP* genes are divided into two subfamilies: *AP2* genes with two AP2 domains and *EREBP* genes with a single AP2/ERF (Ethylene Responsive Element Binding Factor) domain. Based on previous phylogenetic analyses, *AP2* genes can be divided into two clades, AP2 and ANT groups. To clarify the molecular evolution of the AP2 subfamily, we isolated and sequenced genes with two AP2 domains from three gymnosperms, *Cycas revoluta*, *Ginkgo biloba*, and *Gnetum parvifolium,*as well as from the moss *Physcomitrella patens*. Expressions of *AP2*-like genes, including *AP2*, in *Arabidopsis thaliana* are regulated by the microRNA *miR172*. We found that the target site of *miR172* is significantly conserved in gymnosperm *AP2* homologs, suggesting that regulatory mechanisms of gene expression using microRNA have been conserved over the three hundred million years since the divergence of gymnosperm and flowering plant lineages.

We inferred a phylogenetic relationship of these genes with the green alga *Chlamydomonas reinhardtii* and seed–plant genes available in public DNA databases. The phylogenetic tree showed that the AP2 subfamily diverged into the AP2 and ANT groups before the last common ancestor of land plants and after *C. reinhardtii* diverged from the land-plant lineage. The tree also indicated that each AP2 and ANT group further diverged into several clades through gene duplications prior to the divergence of gymnosperms and angiosperms.

Keywords: *AINTEGUMENTA*; *APETALA2*; Gymnosperms; *Physcomitrella patens*; MicroRNA

1. Introduction

The AP2 (APETALA2)/ERF (Ethylene Responsive Element Binding Factor) domain is a DNA-binding domain that consists of approximately 60 conserved amino acid residues (Jofuku et al., 1994; Hao et al., 1998). The genes containing a single or two AP2/ERF domains encode putative transcription factors and belong to the AP2/EREBP multigene family (Riechmann and Meyerowitz, 1998). Sakuma et al. (2002) found 145 AP2/EREBP genes in the genome of *Arabidopsis thaliana*. This multigene family is classified into two subfamilies: the EREBP (ethylene responsive element binding protein) subfamily, which has a single AP2/ERF domain, and the AP2 subfamily, which has two AP2 domains. Most members of the EREBP subfamily function in signal transduction pathways of biotic and environmental stress responses (reviewed by Riechmann and Meyerowitz, 1998), while some of them function in cambial tissue development (van der Graaff et al., 2000, 2003). The AP2 subfamily is further divided into two monophyletic groups: AP2 and ANT (Shigyo and Ito, 2004). Previously reported genes in the AP2 subfamily function as key developmental regulators in reproductive and vegetative organs (Riechmann and Meyerowitz, 1998), including the floral homeotic

Abbreviations: ABI3, ABSCISIC ACID-INSENSITIVE 3; AIC, Akaike information criterion; ANT, AINTEGUMENTA; AP2, APETALA2; *CrANTL1*, *Cycas revoluta ANT*-like gene 1; *CrAP2L1*, *Cycas revoluta AP2*-like gene 1; EREBP, Ethylene responsive element binding protein; ERF, Ethylene responsive element binding factor; *GbANTL1*, *Ginkgo biloba ANT*-like gene 1; *GbAP2L1*, *Ginkgo biloba AP2*-like gene 1; *GpANTL1*, *Gnetum parvifolium ANT*-like gene 1; *GpAP2L1*, *Gnetum parvifolium AP2*-like gene 1; ML, Maximum-likelihood; NJ, Neighbor-joining; RAV, Related to ABI3/VP1; RELL, Resampling-of-estimated-log-likelihood; VP1, VIVIPAROUS1.

* Corresponding author. Tel/fax: +81 3 5454 6638.
E-mail address: cmito@mail.ecc.u-tokyo.ac.jp (M. Ito).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.08.009

gene *APETALA2*(*AP2*) (Jofuku et al., 1994) and *AINTEGUMENTA*(*ANT*; Elliott et al., 1996; Klucher et al., 1996), which is involved in lateral organ development by controlling cell number and growth (Mizukami and Fischer, 2000; Mizukami, 2001).

Determining the phylogenetic relationships of the AP2/EREBP multigene family among plants is an important step in elucidating the evolution of this developmentally and physiologically important gene family. *AP2/EREBP* genes have been reported from several angiosperms and conifers (Vahala et al., 2001; Shigyo and Ito, 2004). Domains sharing sequence similarity with the AP2/ERF domain had not been found outside the plant kingdom, and AP2/ERF domain had been considered plant specific (Riechmann and Meyerowitz, 1998; Krizek, 2003). Recently, however, AP2/ERF-domain-containing genes have been reported from non-plants for the first time, the ciliate *Tetrahymena thermophila*(Wuitschick et al., 2004), the cyanobacterium *Trichodesmium erythraeum*, the virus *Enterobacteria phage Rb49*, and the virus *Bacteriophage Felix 01* (Magnani et al., 2004). The nonplant AP2-domain-containing-proteins each have a HNH domain adjacent to the AP2/ERF domain, and they are predicted HNH endonucleases characterized by the sequence motif 'His-Asn-His' (Shub et al., 1994; Dalgaard et al., 1997). Magnani et al. (2004) hypothesized that a horizontal transfer of an HNH endonuclease bearing AP2/ERF domain into plants might have occurred through the endosymbiosis of a cyanobacterium, viral infections, or other lateral gene transfer events. An AP2 gene of the green alga *Chlamydomonas reinhardtii* with two AP2 domains has been deposited in a public DNA database. Collectively, however, these plants represent a poor sample of the major clades in the phylogeny of land plants (Kenrick and Crane, 1997). Therefore, it is still not clear (1) when the AP2 and EREBP subfamilies diverged with a duplication of the AP2/ERF domain, (2) when AP2 subfamily diverged into the AP2 and ANT groups, and (3) when genes of the AP2 and ANT groups diverged during seed–plant evolution.

MicroRNAs have recently been found to be important regulators at the translational level in both the plant and the animal kingdom (Lagos-Quintana et al., 2001, 2002; Lau et al., 2001; Lee and Ambros, 2001; Llave et al., 2002; Mourelatos et al., 2002; Park et al., 2002; Reinhart et al., 2002). The microRNA *miR172*with 21-nucleotide non-coding RNA was reported to down-regulate several *Arabidopsis* genes in the AP2 subfamily (Aukerman and Sakai, 2003; Chen, 2004). Recently, it was also reported that maize *AP2*-like gene *glossy15* (Moose and Sisco, 1996) was down-regulated by *miR172* (Lauter et al., 2005). The conifer genes of the AP2 subfamily, *PtAP2L1*and *PtAP2L2,* contain nucleotides that are >85% identical to *miR172*, and are likely regulated in a manner similar to *Arabidopsis AP2* genes (Shigyo and Ito, 2004). It is unclear whether such regulation with microRNA is conserved in other lineages of seed plants.

To address these questions, at the beginning, we isolated both *AP2*-like and *ANT*-like genes from three of the four major extant lineages of gymnosperms, *Cycas*, *Ginkgo*, and *Gnetum*, as well as an *ANT*-like gene from the moss *Physcomitrella patens*. Based on phylogenetic analyses of these genes in conjunction with previously reported *AP2/EREBP* genes, we discuss the molecular evolution of the AP2/EREBP multigene family.

2. Materials and methods

2.1. Plant materials

Young female reproductive organs of *Cycas revoluta* and *Ginkgo biloba* were collected at Chiba University, Japan, in April. Reproductive organs of *Gnetum parvifolium* were collected from plants cultivated in a greenhouse of the Botanical Gardens, Graduate School of Science, University of Tokyo, from early March to April. All materials were frozen in liquid nitrogen.

2.2. Cloning of two-AP2-domain-containing genes

Total RNAs of *C. revoluta* and *G. biloba* were extracted as described by Shigyo and Ito (2004). RNA of *G. parvifolium* was extracted as described by Shindo et al. (1999). To clone the two-AP2-domain-containing genes of the three gymnosperms, we followed the procedure of Shigyo and Ito (2004). We synthesized three AP2-domain-specific primers based on an alignment of amino acid sequences of previously reported AP2-domain-containing genes: dAP2-SENSE1 (SQYRGV; 5′-CUACUACUACUAWSNCARTAY-MGNGGNGT-3′), dAP2-ANTI1 (AVTNF; 5′-CAUCAUCA-UCAUTCRAARTTNGTNACNGC-3′), AP2-1 (WDCGKQ; 5′-CAUCAUCAUCAUTTGGGACTGTGGGAAACAAGT-3′), and ANT-1 (NSFKKE; 5′-CAATAGTTTCAAGAAG-GAAG-3′), where M, N, R, S, W, and Y follow the IUPAC code.

Nishiyama et al. (2003) constructed cDNA libraries and EST databases from auxin-treated, cytokinin-treated, and untreated gametophytes of the moss *P. patens*, and we searched for ESTs with similarity to the AP2 domains of *Arabidopsis AP2*.

2.3. Phylogenetic analysis

The deduced amino acid sequences of *AP2/EREBP* genes were obtained from the NCBI DNA database. Sequences were aligned with the program CLUSTAL X, version 1.83 (Thompson et al., 1997), and then revised manually. All of the alignment gaps were eliminated. Phylogenetic analyses were performed with MOLPHY version 2.3b3 (Adachi and Hasegawa, 1996). To construct a maximum-likelihood (ML) tree, we used a neighbor-joining (NJ) tree as a start tree for a local rearrangement search (Adachi and Hasegawa, 1996). An NJ tree (Saitou and Nei, 1987) was obtained with NJdist program based on the ML distance in the JTT model (Jones et al., 1992). The likelihoods of trees were calculated using ProtML program under the JTT model, and the trees were sorted according to their Akaike information criterion (AIC) values (Adachi and Hasegawa, 1996). The local bootstrap probability of each branch was estimated by using the resampling-of-estimated-log-likelihood (RELL) method (Kishino et al., 1990; Hasegawa and Kishino, 1994).

3. Results

3.1. Isolation of AP2 subfamily genes from three gymnosperms

Candidate *AP2* and *ANT* homologs of *C. revoluta*, *G. biloba*, and *G. parvifolium* were obtained from young unpollinated female reproductive organs using the 3′RACE system with AP2-domain-specific primers based on the alignment of genes in both AP2 and ANT groups. Two of these primers corresponded to conserved amino acid residues located in the 5′ region of the AP2 repeat-1 domain (dAP2-SENSE1) and in the 3′ region of the AP2 repeat-2 domain (dAP2-ANT1).

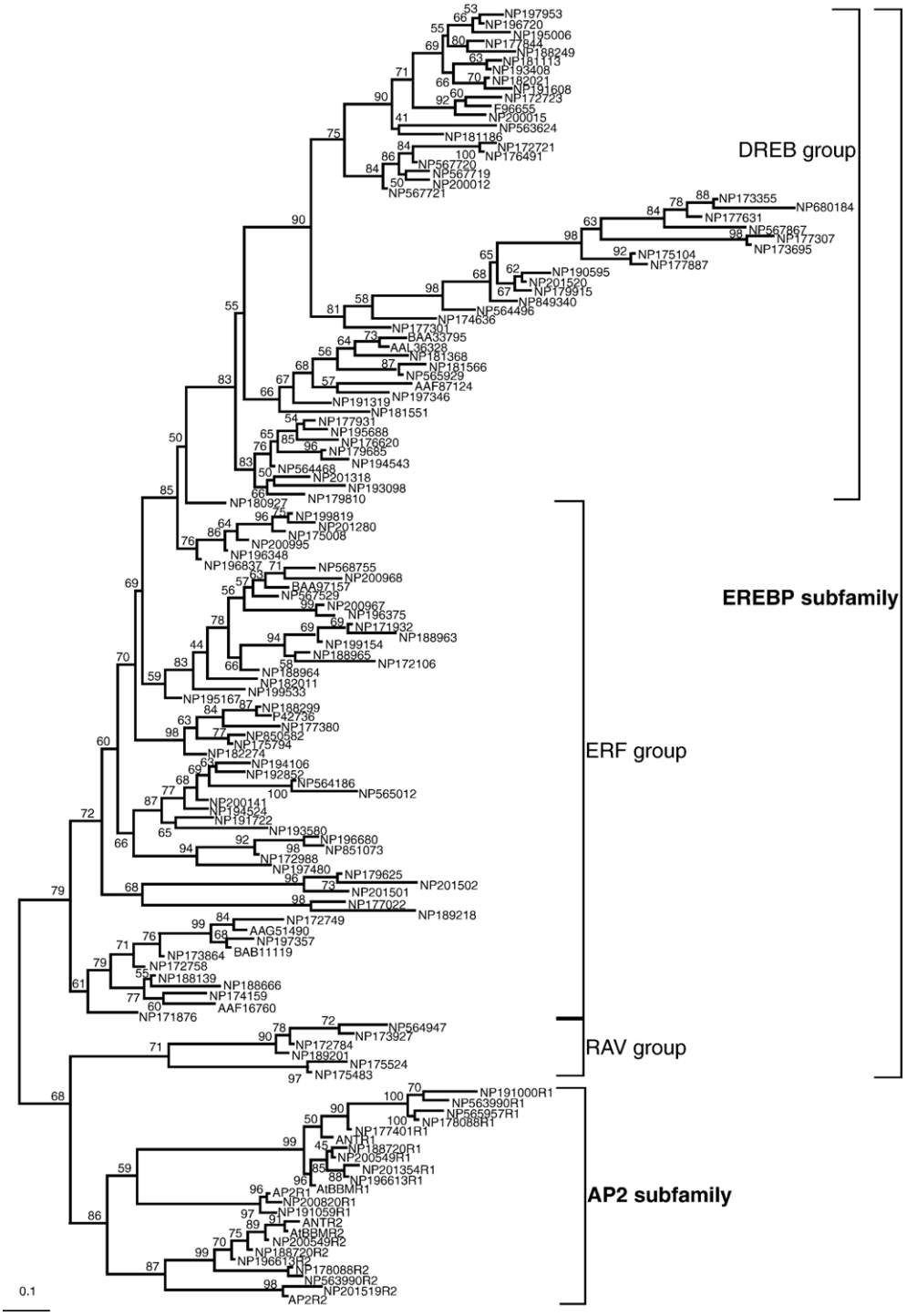

Fig. 1. Unrooted maximum-likelihood tree of the AP2/EREBP multigene family. All *Arabidopsis thaliana* genes of the AP2/EREBP multigene family were used. Local bootstrap probabilities are shown on or below branches. Scale bar corresponds to 0.1 amino acid substitutions per residue. Two AP2 domains of the AP2 subfamily genes, the AP2 repeat-1 domain and the AP2 repeat-2 domain, were treated separately as two OTUs. In the AP2 subfamily, R1 indicates the AP2 repeat-1 domain, and R2 indicates the AP2 repeat-2 domain.

Using these primers, we obtained partial fragments of the putative *C. revoluta AP2* homolog, the putative *G. biloba AP2* homolog, and the putative *G. parvifolium AP2* and *ANT* homologs. These genes of the AP2 subfamily were distinguished based on the presence (*ANT* homolog) or absence (*AP2* homolog) of ten amino acid residues in AP2 repeat-1 domain. Furthermore, *ANT*-specific primers were synthesized based on the alignment of genes in the ANT group. These primers corresponded to amino acid residues located in the 5′ region of the AP2 repeat-1 domain (ANT-1) and in the 3′ region of the AP2 repeat-2 domain (dAP2-ANTI1). We obtained partial fragments of the putative *C. revoluta ANT* homolog and the putative *G. biloba ANT* homolog. The remaining parts of these DNA fragments were cloned using 3′ and 5′ RACE systems. Primers used in these experiments were deposited in a DNA database with each gene sequence. The *AP2* and *ANT* homologs were named *C. revoluta AP2*-like gene *1* (*CrAP2L1*) (GenBank accession number is AB195242), *C. revoluta ANT*-like gene *1* (*CrANTL1)*(GenBank accession number is AB195243), *G. biloba AP2*-like gene *1* (*GbAP2L1*) (GenBank accession number is AB195244), *G. biloba ANT*-like gene *1* (*GbANTL1*) (GenBank accession number is AB195245), *G. parvifolium AP2*-like gene *1* (*GpAP2L1)*(GenBank accession number is AB195246), and *G. parvifolium ANT*-like gene *1* (*GpANTL1*) (GenBank accession number is AB195247). To confirm that the 5′ and 3′ RACE products were from the same cDNA, we synthesized gene-specific primers based on regions located close to the 3′ and 5′ ends of each cDNA according to the nucleotide sequences of the RACE products. To exclude PCR errors from the resulting sequences, at least two independently amplified PCR products were sequenced for each gene.

3.2. P. patens ANT gene homolog

Nishiyama et al. (2003) constructed a full-length enriched cDNA library from the gametophytic tissues of the moss *P. patens*, and 15,883 putative transcripts were catalogued based on EST analyses. Through similarity searches of this EST database together with a public DNA database, cDNA clones

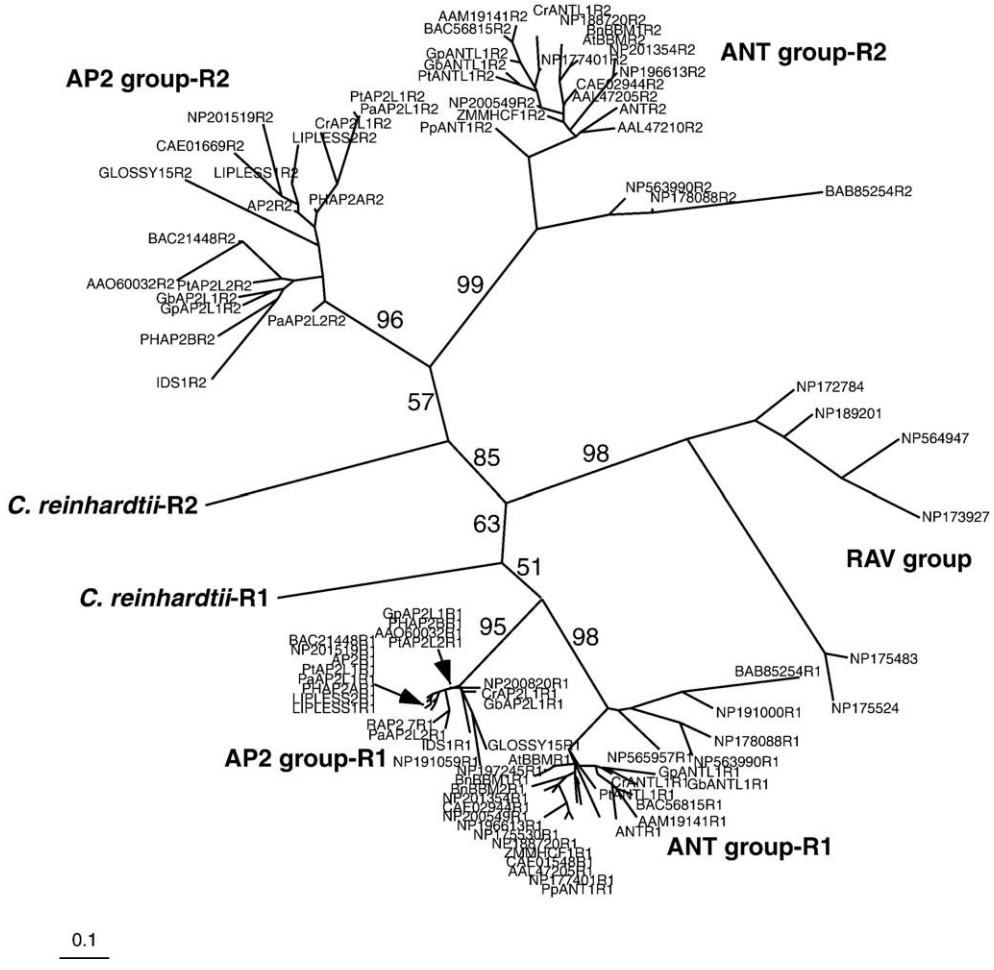

Fig. 2. Maximum-likelihood tree of the AP2 subfamily using separated AP2 domains. We used 60 amino acid residues of the AP2/ERF domain region. Local bootstrap probabilities are shown on or below the main branches of the tree. Scale bar corresponds to 0.1 amino acid substitutions per residue. We used the genes isolated in this study and all genes of *Arabidopsis thaliana*, *Antirrhinum majus*, *Brassica napus*, *Petunia hybrida, Oryza sativa*, *Zea mays*, and *Picea abies* that were registered in the NCBI database in November 2003 as well as *Chlamydomonas reinhardtii* AV622151. The *A. thaliana* RAV group was used as the outgroup. Two AP2 domains of the AP2 subfamily genes, the AP2 repeat-1 domain and the AP2 repeat-2 domain, were treated separately as two OTUs. In the AP2 subfamily, R1 indicates the AP2 repeat-1 domain, and R2 indicates the AP2 repeat-2 domain.

pphb45d02 and pphb21g12 were found to be similar to the two-AP2-domain-containing genes. The 5′ end sequences of pphb21g12 and pphb45d02 were registered in the EST database, and the 558-bp-long-sequence of pphb21g12 included the same sequence as the 555-bp-long-sequence of pphb45d02. The clone pphb21g12 was sequenced and found to be an *ANT* homolog with two AP2 domains, named *PpANT1*. This is the first reported *ANT* homolog isolated from non-vascular plant.

3.3. C. reinhardtii gene belonging to the AP2 subfamily

An EST sequence of *C. reinhardtii*, AV622151, which was similar to AP2 domains of *Arabidopsis AP2*, was found in the NCBI DNA database. A deduced amino acid sequence of AV622151 contained two AP2 domains. AV622151 lacked the 5′ region of the AP2 repeat-1 domain. A genomic sequence that was 99% identical to the AV622151 cDNA sequence was found in the *C. reinhardtii* genome database of the Kazusa DNA

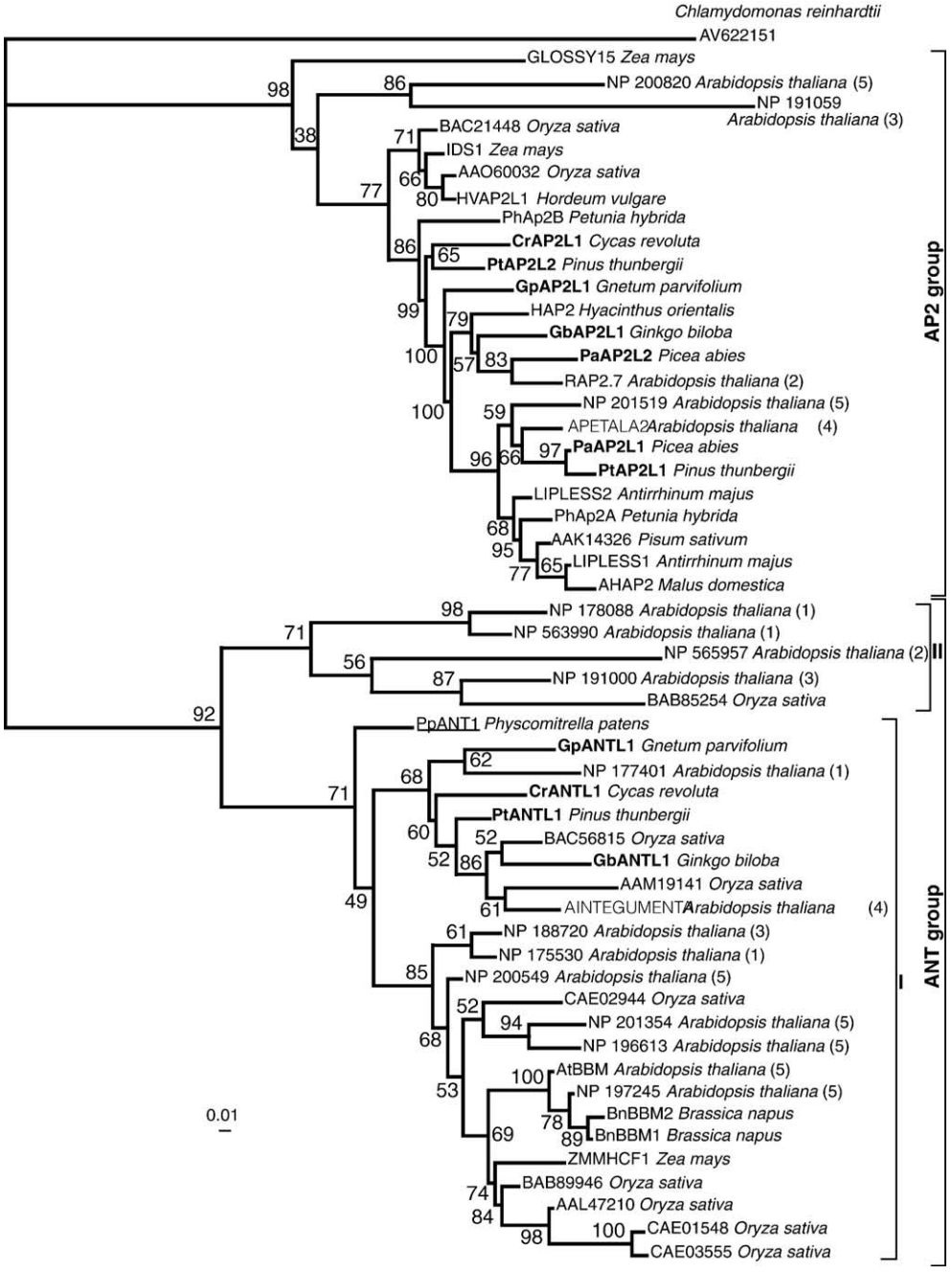

Fig. 3. Maximum-likelihood tree of the AP2 subfamily using the AP2 domain-linker region. We used 86 amino acid residues of the AP2 domain-linker region of the AP2 subfamily genes. Local bootstrap probabilities are shown on or below the branches. Scale bar corresponds to 0.01 amino acid substitution per residue. We used the genes isolated in this study and all genes of *Arabidopsis thaliana*, *Antirrhinum majus*, *Brassica napus*, *Malus domestica*, *Petunia hybrida*, *Pisum sativum*, *Hordeum vulgare*, *Hyacinthus orientalis*, *Oryza sativa*, *Zea mays*, and *Picea abies* that were registered in the NCBI database in November 2003 as well as *Chlamydomonas reinhardtii* AV622151. I indicates group I, II indicates group II. The chromosome map positions of each *A. thaliana* gene are indicated with numbers.

Research Institute (http://www.kazusa.or.jp/), and this genomic sequence was considered to correspond to AV622151 cDNA. The missing region of the AP2 repeat-1 domain of AV622151 was assigned as SSQYRGVTRHRRSGRWEA based on the genomic sequence.

3.4. Gene tree of the AP2/EREBP multigene family

To analyze the phylogenetic relationships among genes in the AP2 and EREBP subfamilies, we used all AP2/ERF-domain-containing genes of *A. thaliana*. Two AP2 domains of genes belonging to the AP2 subfamily were treated separately as two operational taxonomic units (OTUs): AP2 repeat-1 domain (R1) and AP2 repeat-2 domain (R2). Forty-one amino acid residues corresponding to positions 284–289, 294–303, 320–340, and 344–347 in R1 and positions 386–391, 396–405, 414–434, and 438–441 in R2, counting from the initial methionine of the *Arabidopsis* ANT protein, and positions 62–67, 71–80, 87–107, and 111–114 in the EREBP proteins, counting from the initial methionine of *Arabidopsis* RAV1 (RAV, Related to ABI3/VP1; ABI3, ABSCISIC ACID-INSENSITIVE 3; VP1, VIVIPAROUS1) protein (Kagaya et al., 1999), were used for phylogenetic analyses with MOLPHY version 2.3b3 (Adachi and Hasegawa, 1996). Because NP193040 and R2 of NP191059, NP200820, NP191000, NP565957, and RAP2.7 appeared to have specific gaps, they were excluded to keep the available amino acid residues for gene tree construction. Because the distances between NP178173 and other AP2/EREBP multigene family genes were huge, NP178173 was removed. A gene tree (Fig. 1) suggested that AP2 subfamily was closely related to RAV group (Kagaya et al., 1999). This was supported by NJ tree also (data not shown).

3.5. Gene tree of the AP2 subfamily using AP2 domain regions

To analyze the phylogenetic relationships between AP2 subfamily genes of land plants and the green alga *C. reinhardtii AV622151*, a gene tree (Fig. 2) was constructed based on 60 amino acid sequences around the AP2 domain region. Sixty amino acid residues corresponding to positions 170–177, 181–192, 205, 209–233, 235–242, and 244–249 in R1 and positions 272–279, 283–294, 299, 303–327, 329–336, and 338–343 in R2, counting from the initial methionine of the conifer *Pinus thunbergii* PtANTL1 protein, and positions 59–66, 69–80, 83, 87–111, 113–120, and 122–127, counting from the initial methionine of the RAV1 protein, were also used for phylogenetic analyses. We used the genes isolated in this study, in addition to *A. thaliana*, *Antirrhinum majus*, *Petunia hybrida*, *Oryza sativa*, *Zea mays*, *Picea abies, P. thunbergii,* and *C. reinhardtii* that were registered in the NCBI database in November 2003. *A. thaliana* genes of the RAV group of the EREBP subfamily were used to represent the outgroup. We describe about this selection of outgroup later. A gene tree (Fig. 2) showed that AP2 group-R1 and ANT group-R1 formed a sister group with 51% bootstrap support, and *C. reinhardtii* R1 was sister to this group with 63% bootstrap support. AP2 group-R2 and ANT group-R2 formed a sister group with 57% bootstrap support, and *C. reinhardtii* R2 was sister to this group with a bootstrap value of 85%. The phylogenetic relationships among AP2 group-R1, AP2 group-R2, ANT group-R1, ANT group-R2, *C. reinhardtii* R1, and *C. reinhardtii* R2 were supported by NJ tree also (data not shown).

miR172	3'-UACGUCGUAGUAGUUCUAAGA-5'
AP2 mRNA	5'-CUGCAGCAUCAUCAGGAUUCU-3'
AHAP2 mRNA	5'-CUGCAGCAUCAUCAGGAUUCC-3'
GLOSSY15 mRNA	5'-CUGCAGCAUCAUCAGGAUUCC-3'
HAP2 mRNA	5'-CUGCAGCAUCAUCAGGAUUCG-3'
HVAP2L1 mRNA	5'-CUGCAGCAUCAUCAGGAUUCU-3'
IDS1 mRNA	5'-CUGCAGCAUCAUCAGGAUUCU-3'
LIP1 mRNA	5'-CUGCAGCAUCAUCAGGAUUCC-3'
LIP2 mRNA	5'-CUGCAGCAUCAUCAGGAUUCC-3'
PhAp2A mRNA	5'-CUGCAGCAUCAUCAGGAUUCU-3'
PhAp2B mRNA	5'-CUGCAGCAUCAUCAGGAUUCG-3'
RAP2.7 mRNA	5'-CAGCAGCAUCAUCAGGAUUCU-3'
AAK14326 mRNA	5'-CUGCAGCAUCAUCAGGAUUCC-3'
BAC21448 mRNA	5'-CUGCAGCAUCAUCAGGAUUCU-3'
NP 200820 mRNA	5'-AUGCAGCAUCAUCAGGAUUCU-3'
NP 201519 mRNA	5'-UGGCAGCAUCAUCAGGAUUCU-3'
CrAP2L1 mRNA	5'-CAGCAGCAUCAUCAGGAUUCU-3'
GbAP2L1 mRNA	5'-CUGCAGCAUCAUCAGNAUUCU-3'
GpAP2L1 mRNA	5'-UUGCAGCAUCAUCAGGAUUCU-3'
PaAP2L1 mRNA	5'-GUGCAGCAUCAUCAGGAUUCU-3'
PaAP2L2 mRNA	5'-GAUCCCCUGCAUCAGGAUUCU-3'
PtAP2L1 mRNA	5'-GUGCAGCAUCAUCAGGAUUCU-3'
PtAP2L2 mRNA	5'-CAGCAGCAUCAUCAGGAUUCU-3'

Fig. 4. Putative *miR172* target sites in mRNAs of *AP2* homologs. The underlined nucleotides are not complementary to *miR172*.

3.6. Gene tree of the AP2 subfamily using the two-AP2 domain and a linker region

To analyze the phylogenetic relationships of the AP2 subfamily in detail, an ML tree was constructed based on 86 amino acid sequences of R1, R2, and linker regions corresponding to positions 188–191, 204–205, 209–280, 286–290, 292, and 294–295, counting from the initial methionine of PtANTL1 (Fig. 3). The same OTUs used in Fig. 2 were analyzed, with the exception of *RAV* genes, and a *Chlamydomonas* gene was used as the outgroup. Genes were classified into AP2 or ANT groups, and putative *AP2* and *ANT* homologs obtained in this study were confirmed to be homologs of each gene. This was supported by NJ tree also (data not shown).

3.7. Site complementary to miR172 in AP2 homologs of gymnosperms

We searched for sites that were complementary to the microRNA *miR172* in seven gymnosperm *AP2* homologs used in this study. Every gymnosperm gene contained a sequence corresponding to *miR172* with an average similarity of approximately 84.4% (Fig. 4).

4. Discussion

4.1. Sequence analysis of the AP2 subfamily of seed plants

In this study, AP2 domains were well conserved among flowering plants, gymnosperms, and the moss (Fig. 5). The

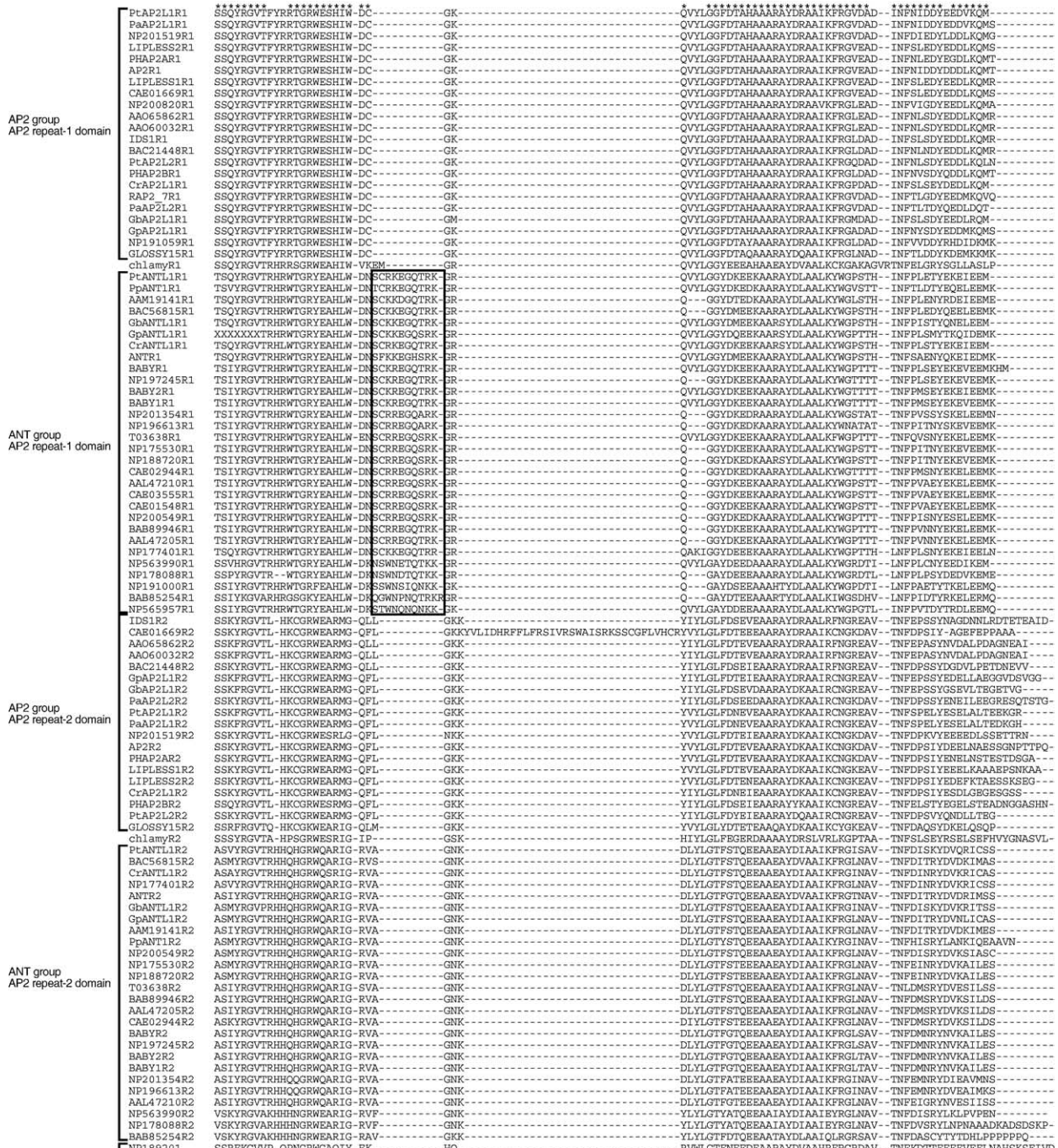

Fig. 5. Alignment of amino acid sequences of *AP2* subfamily genes and *RAV* genes used in this study. Dashes indicate gaps. Asterisks above the alignment indicate the 60 amino acids used in the phylogenetic analysis. The ten specific amino acid residues in the AP2 repeat-1 domain of the *ANT* homologs are outlined.

putative nuclear localization signal near the R1 domain of ANT (Klucher et al., 1996) was conserved in CrANTL1, GbANTL1, and PtANTL1 (data not shown). Another amino acid motif outside the two-AP2 domain and linker regions was conserved among ANT, CrANTL1, GbANTL1, and PtANTL1. This motif, QSLTLSMSPGSQ (Shigyo and Ito, 2004), corresponds to positions 113–124, counting from the initial methionine of PtANTL1. The putative nuclear localization signal adjacent to the R1 domain of AP2 (Jofuku et al., 1994) is conserved in all gymnosperm AP2 homologs, including PaAP2L1 and PaAP2L2. This signal is also conserved in the other AP2 homologs. The amino acid motifs corresponding to positions 109–116, 341–349, and 451–459 (counting from the initial methionine of PtAP2L1) outside the conserved two-AP2 domain and linker regions that have been identified in several AP2-like proteins (Vahala et al., 2001), are also conserved in the isolated

regions of CrAP2L1, GbAP2L1, and GpAP2L1. This suggests that gymnosperm *AP2* and *ANT* genes are likely to function as transcription factors.

Nucleotide sequences of the microRNA *miR172* binding site are well conserved in gymnosperm *AP2* homologs. Each *miR172* complementary site of *AP2* and gymnosperm *AP2* homologs is in the conserved region mentioned by Vahala et al. (2001), and located between the AP2 repeat-2 domain and the 3′ terminus. This suggests that the expression of *AP2* homologs is regulated by microRNA, as in seed plants.

4.2. Phylogenetic relationships among AP2 subfamily genes of gymnosperms, moss, and other land plants

Our analysis showed that AP2 subfamily genes, except for *C. reinhardtii* AV622151, consist of two clades, each of which corresponds to AP2 and ANT groups (Fig. 3). Genes of ANT group have the distinctive ten amino acid residues in AP2 repeat-1 domain (Fig. 5). In the AP2 group, gymnosperm *AP2* homologs do not form a clade. Extant gymnosperms consist of four groups: Coniferophyta, Cycadophyta, Ginkgophyta, and Gnetophyta. Whether extant gymnosperms are paraphyletic or monophyletic has been a topic of controversy. Morphological, fossil (Arber and Parkin, 1907, 1908; von Wettstein, 1907; Retallack and Dilcher, 1981; Cronquist, 1988), and molecular (Chase et al., 1993; Doyle, 1996) data have long supported the paraphyly of extant gymnosperms, albeit with low statistical support (Doyle, 1996). In contrast, almost all recent phylogenetic analyses of molecular data strongly supported the monophyly of extant gymnosperms (Hasebe et al., 1992; Chaw et al., 1997, 2000; Bowe et al., 2000; Gugerli et al., 2001). If we accept the monophyly hypothesis of extant gymnosperms, the gymnosperm *AP2* homologs used in this study are likely to be parologous, although *PtAP2L1* and *PaAP2L1* are probably orthologous based on the tree shown in Fig. 3, as well as *CrAP2L1* and *PtAP2L2*. *GbAP2L1* and *PaAP2L2* might be orthologous, as well as *CrANTL1*, *GbANTL1*, and *PtANTL1*, when the topology of the gene tree is not correct. We outlined three questions about the evolution of AP2/EREBP multigene family as described before. In the question (3): when genes of the AP2 and ANT groups diverged during seed–plant evolution, some possibilities are suggested. In the AP2 group, there are at least three moderately highly supported clades that include both angiosperm and gymnosperm *AP2* homologs, indicating that the last common ancestor of seed plants possibly had at least three *AP2* homologs. In the ANT group, gymnosperm ANT genes do not form a clade. The *Ginkgo GbANTL1* forms a clade with *A. thaliana* and *O. sativa* genes, which is supported by a bootstrap value of 86%. Phylogenetic relationships of other gymnosperm *ANT* genes are ambiguous because of low bootstrap supports for their branches. At least two *ANT* genes existed in the last common ancestor between extant gymnosperms and angiosperms.

A. thaliana NP 178088, NP 563990, NP 565957, NP 191000, and *O. sativa* BAB85254 genes have the ANT-group-specific ten amino acid residues in AP2 repeat-1 domain, but their sequences differ significantly from other ANT genes (Fig. 5). These five genes form a different clade from the other ANT genes in the phylogenetic tree in Fig. 3. Here we term this clade ANT group II, while the other is ANT group I. Extant bryophytes consist of three groups: mosses, liverworts, and hornworts. Whether extant bryophytes are paraphyletic or monophyletic has been also widely debated and is still not resolved. Nishiyama et al. (2004) proposed bryophyte monophyly as the currently best hypothesis based on their phylogenetic analysis using 51 genes from the entire chloroplast genome sequences of 20 representative green plant species. In their phylogenetic analysis, a clade of extant bryophyte is supported with very high bootstrap value. The *ANT* homolog *PpANT1* of the moss *P. patens* clusters with ANT group I, indicating that the last common ancestor of mosses and vascular plants had at least one *ANT* gene for each group, I and II, if we accept the monophyly hypothesis of extant bryophytes.

The functions of only a few genes are characterized in the AP2 and ANT groups; therefore, it is hard to speculate on the evolution of the functions of this gene family. When we consider that previously characterized *AP2* and *ANT* genes are involved in several developmental processes and that the number of these groups increased during land-plant evolution, the evolution of this gene family via gene duplication and subsequent functional diversification is likely to be related to the evolution of developmental processes in land plants, similar to other transcription factors (Carroll et al., 2001). MADS-box genes encode transcription factors. Nam et al. (2004) found 107 functional MADS-box genes in the genome of *A. thaliana*. Members of MADS-box gene family are involved in several aspects of development in plants, and their evolution has been fundamental for morphological diversification in plants, especially in the reproductive organs (Hasebe, 1999; Theissen et al., 2000). According to phylogenetic analyses based on both morphological and molecular data, the closest relatives of land plants are freshwater green algae charophyceans (reviewed by Graham et al., 2000). Recently, MADS-box genes were isolated from three charophycean green algae, and the expression analyses of these genes suggested that the precursors of land plant MADS-box genes originally functioned in haploid reproductive cell differentiation and that the haploid MADS-box genes were recruited into a diploid generation during the evolution of land plants (Tanabe et al., 2005).

4.3. Molecular evolution of the AP2 subfamily

Based on the number of AP2/ERF domains, the AP2/EREBP multigene family is divided into two groups. Because the gene tree in Fig. 1 is unrooted, we cannot specify whether the common ancestor of AP2 and EREBP genes had one or two AP2/ERF domains. Magnani et al. (2004) suggested that plant AP2/ERF-domain-containing genes were derived from bacterial or viral endonucleases, and that the common ancestor of *AP2* and *EREBP* genes had one AP2/ERF domain. The green alga *C. reinhardtii* has a gene with two AP2 domains. *C. reinhardtii* belongs to the Chlorophyta lineage, and the Clorophyta lineage is sister to the Streptophyta lineage (Charophyceae and land plants) (Karol et al., 2001). In the question (1): when the AP2

and EREBP subfamilies diverged with a duplication of the AP2/ERF domain, it is suggested that genes in the AP2 and EREBP subfamilies are diverged before the Chlorophyta lineage is diverged from the Streptophyta lineage.

The EREBP subfamily has been classified into four groups based on the general similarity of amino acid sequences in the AP2/ERF domain: RAV, DREB, ERF, and "others" (Sakuma et al., 2002). The "others" group lacks a WLG motif (Sakuma et al., 2002) in the AP2/ERF domain, and this group was excluded to keep the available amino acid residues for gene tree construction. Phylogenetical relationship of AP2/EREBP multigene family was examined using NJ trees (Sakuma et al., 2002; Alonso et al., 2003; Magnani et al., 2004; in this study) and ML tree (Fig. 1). In all five trees, RAV group is closely related to AP2 subfamily, although bootstrap value for the association of AP2 subfamily and RAV group is not very high (68%) in the gene tree (Fig. 1). We selected RAV group as an outgroup to infer the phylogenetic relationships between AP2 subfamily genes of land plants and the green alga *C. reinhardtii AV622151*. In the question (2): when AP2 subfamily diverged into the AP2 and ANT groups, the gene tree in Fig. 2 indicates that the AP2 subfamily is divided into two groups, AP2 and ANT, following the divergence of *C. reinhardtii*and the land-plant lineage. As *P. patens*has an *ANT* ortholog, the AP2 and ANT groups diverged before mosses and land plants diverged, if we accept the monophyly hypothesis of extant bryophytes. It is very interesting to analyze two-AP2-domain-containing genes using charophycean green algae. To investigate further about the question (2), we are trying to isolate two-AP2-domain-containing genes from charophycean green algae now. It is necessary to analyze two-AP2-domain-containing genes using other plants, especially ferns and algae for the further understanding of the evolution of AP2 subfamily.

Acknowledgements

We thank S. Shindo for critical support in cloning *GpAP2L1* and *GpANTL1*. This work was partly supported by Monbusho Scientific Research Grant-in–Aid No. 09640822 to MI.

References

Adachi, J., Hasegawa, M., 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Comput. Sci. Monogr. 28, 1–150.

Alonso, J.M., et al., 2003. Science 301, 653–657 (36 co-authors).

Arber, E.A.N., Parkin, J., 1907. On the origin of angiosperms. J. Linn. Soc. Lond., Bot. 38, 29–80.

Arber, E.A.N., Parkin, J., 1908. Studies on the evolution of the angiosperms. The relationship of the angiosperms to Gnetales. Ann. Bot. (Lond.) 22, 489–515.

Aukerman, M.J., Sakai, H., 2003. Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. Plant Cell 15, 2730–2741.

Bowe, L.M., Coat, G., dePamphilis, C.W., 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. PNAS 97, 4092–4097.

Carroll, S.B., Grenier, J.K., Weatherbee, S.D., 2001. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design. Blackwell Scientific, Malden, MA.

Chase, M.W., et al., 1993. DNA sequence phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. 80, 528–580 (42 co-authors).

Chaw, S.-M., Zharkikh, A., Sung, H.-M., Lau, T.-C., Li, W.-H., 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol. Biol. Evol. 14, 56–68.

Chaw, S.-M., Parkinson, C.L., Cheng, Y., Vincent, T.M., Palmer, J.D., 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc. Natl. Acad. Sci. U. S. A. 97, 4086–4091.

Chen, X., 2004. A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. Science 303, 2022–2025.

Cronquist, A., 1988. The Evolution and Classifications of Flowering Plants. The New York Botanical Garden, New York.

Dalgaard, J.Z., Klar, A.J., Moser, M.J., Holley, W.R., Chatterjee, A., Mian, I.S., 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res. 25, 4626–4638.

Doyle, J.A., 1996. Seed plant phylogeny and the relationships of Gnetales. Int. J. Plant Sci. 157, S3–S39.

Elliott, C.R., et al., 1996. AINTEGUMENTA, an APETALA2-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. Plant Cell 8, 155–168.

Graham, L.E., Cook, M.E., Busse, J.S., 2000. The origin of plants: body plant changes contributing to a major evolutionary radiation. Proc. Natl. Acad. Sci. U. S. A. 97, 4535–4540.

Gugerli, F., et al., 2001. The evolutionary split of Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. Mol. Phylogenet. Evol. 21, 167–175.

Hao, D., Ohme-Takagi, M., Sarai, A., 1998. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J. Biol. Chem. 273, 26857–26861.

Hasebe, M., 1999. Evolution of reproductive organs in land plants. J. Plant Res. 112, 463–474.

Hasebe, M., Kofuji, R., Ito, M., Kato, M., Iwatsuki, K., Ueda, K., 1992. Phylogeny of gymnosperms inferred from *rbcL* gene sequences. Bot. Mag. Tokyo 105, 673–679.

Hasegawa, M., Kishino, H., 1994. Accuracies of the simple methods for estimating the bootstrap probability of a maximum-likelihood tree. Mol. Biol. Evol. 11, 142–145.

Jofuku, D.K., Boer, D.W.G.B., Montagu, V.M., Okamuro, K.J., 1994. Control of *Arabidopsis* flower seed development by the homeotic gene *APETALA2*. Plant Cell 6, 1211–1225.

Jones, D.T., Taylor, R.W., Thornton, M.J., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Kagaya, Y., Ohmiya, K., Hattori, T., 1999. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucleic Acids Res. 27, 470–478.

Karol, K.G., McCourt, R.M., Cimino, M.T., Delwiche, C.F., 2001. The closest living relatives of land plants. Science 294, 2351–2353.

Kenrick, P., Crane, P.R., 1997. The origin and early evolution of plants on land. Nature 389, 33–39.

Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplants. J. Mol. Evol. 30, 151.

Klucher, M.K., Chow, H., Reiser, L., Fischer, L.R., 1996. The *AINTEGUMENTA* gene of *Arabidopsis* required for ovule and female gametophyte development is related to the floral homeotic gene *APETALA2*. Plant Cell 8, 137–153.

Krizek, B.A., 2003. AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain. Nucleic Acids Res. 31, 1859–1868.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschi, T., 2001. Identification of novel genes coding for small expressed RNAs. Science 294, 853–858.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., Tuschi, T., 2002. Identification of tissue-specific microRNAs from mouse. Curr. Biol. 12, 735–739.

Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294, 858–862.

Lauter, N., Kampani, A., Carlson, S., Goebel, M., Moose, S.P., 2005. microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. Proc. Natl. Acad. Sci. U. S. A. 102, 9412–9417.

Lee, R.C., Ambros, V., 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294, 862–864.

Llave, C., Kasschau, K.D., Rector, M.A., Carrington, J.C., 2002. Endogenous and silencing-associated small RNAs in plants. Plant Cell 14, 1605–1619.

Magnani, E., Sjölander, K., Hake, S., 2004. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. Plant Cell 16, 2265–2277.

Mizukami, Y., 2001. A matter of size: developmental control of organ size in plants. Curr. Opin. Plant Biol. 4, 533–539.

Mizukami, Y., Fischer, L.R., 2000. Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. Proc. Natl. Acad. Sci. U. S. A. 97, 942–947.

Moose, S.P., Sisco, P.H., 1996. *Glossy15*, an *APETALA2*-like gene from maize that regulates leaf epidermal cell identity. Genes Dev. 10, 3018–3027.

Mourelatos, Z., et al., 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. 16, 720–728.

Nam, J., Kim, J., Lee, S., An, G., Ma, H., Nei, M., 2004. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proc. Natl. Acad. Sci. U. S. A. 101, 1910–1915.

Nishiyama, T., et al., 2003. Comparative genetics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc. Natl. Acad. Sci. U. S. A. 100, 8007–8012.

Nishiyama, T., et al., 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol. Biol. Evol. 21, 1813–1819.

Park, W., Li, J., Song, R., Messing, J., Chen, X., 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr. Biol. 12, 1484–1495.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P., 2002. MicroRNAs in plants. Genes Dev. 16, 1616–1626.

Retallack, G., Dilcher, D.L., 1981. Arguments for a glossopterid ancestry of angiosperms. Paleobiology 7, 54–67.

Riechmann, J.L., Meyerowitz, E.M., 1998. The AP2/EREBP family of plant transcription factors. Biol. Chem. 379, 633–646.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K., Yamaguchi-Shinozaki, K., 2002. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. Biochem. Biophys. Res. Commun. 290, 998–1009.

Shigyo, M., Ito, M., 2004. Analysis of gymnosperm two-AP2-domain-containing genes. Dev. Genes Evol. 214, 105–114.

Shindo, S., Ito, M., Ueda, K., Kato, M., Hasebe, M., 1999. Characterization of MADS genes in gymnosperm *Gnetum parvifolium* and its implication for the evolution of reproductive organs in seed plants. Evol. Dev. 3, 180–190.

Shub, D.A., Goodrich-Blair, H., Eddy, S.R., 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. Trends Biochem. Sci. 19, 402–404.

Tanabe, Y., et al., 2005. Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. Proc. Natl. Acad. Sci. U. S. A. 102, 2436–2441.

Theissen, G., et al., 2000. A short history of MADS-box genes in plants. Plant Mol. Biol. 42, 115–149.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Vahala, T., Oxelman, B., von Arnold, S., 2001. Two *APETALA2*-like genes of *Picea abies* are differentially expressed during development. J. Exp. Bot. 52, 1111–1115.

van der Graaff, E., den Dulk-Ras, A., Hooykaas, P.J.J., Keller, B., 2000. Activation tagging of the *LEAFY PETIOLE* gene affects leaf petiole development in *Arabidopsis thaliana*. Development 127, 4971–4980.

van der Graaff, E., Nussbaumer, C., Keller, B., 2003. The *Arabidopsis thaliana* rlp mutations revert the ectopic leaf blade formation conferred by activation tagging of the *LEP* gene. Mol. Gen. Genomics 270, 243–252.

von Wettstein, R.R., 1907. Handbuch der Systematischen Botanik. II. Band. Franz Deuticke, Leipzig, Wien.

Wuitschick, J.D., Lindstrom, P.R., Meyer, A.E., Karrer, K.M., 2004. Homing endonucleases encoded by germ line-limited genes in *Tetrahymena thermophila* have APETALA2 DNA binding domains. Eukaryot. Cell 3, 685–694.