Genet. Res., Camb. (2002), **80**, *pp.* 89–98. *With* 3 *figures.* © 2002 Cambridge University Press
DOI: 10.1017/S0016672302005815 Printed in the United Kingdom

Nucleotide polymorphism at the *Atmyb2* locus of the wild plant *Arabidopsis thaliana*

TAKU KAMIYA, AKIRA KAWABE AND NAOHIKO T. MIYASHITA*
Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University, Sakyo-ku, Kyoto, 606-8502, Japan

(*Received 17 August 2001 and in revised form 20 March 2002*)

Summary

DNA variation was studied in a 2·2 kb region of the regulatory gene *Atmyb2* using 20 ecotypes of *Arabidopsis thaliana* and one accession each of *Arabis gemmifera* and *Arabidopsis himalaica*. Nucleotide diversity (π) in the region was 0·0027, which was lower than for other loci in *A. thaliana*. The MYB domain of the *Atmyb2* gene ($\pi = 0{\cdot}0036$) had a larger variation than the non-MYB region ($\pi = 0{\cdot}0013$). Tajima's test and Fu and Li's test did not give a significant result. In contrast to the low level of polymorphism, the degree of divergence of the *Atmyb2* region was higher between *A. thaliana and A. gemmifera* ($K = 0{\cdot}0730$) than for other loci. The MYB domain ($K = 0{\cdot}0436$) had smaller divergence than the non-MYB region ($K = 0{\cdot}0939$). The HKA test detected significant discordance in the ratio of polymorphism to divergence in some comparisons. The pattern of low polymorphism and high divergence, which is mainly observed in the non-MYB region of the gene, is inconsistent with the neutral mutation theory. Strong purifying selection after establishment of *A. thaliana* and a species-specific adaptive process could be invoked to account for this pattern of polymorphism and divergence of *Atmyb2*.

1. Introduction

MYB genes are developmental control genes that play a pivotal role in control of myeloid cell growth. Most MYB-related proteins regulate gene expression at the transcriptional level by binding regulatory regions of target genes; however, a few MYB proteins play a more structural role, binding telomeric regions without inducing gene expression. The targets of MYB proteins are quite diverse, including cell cycle, circadian clock, disease resistance, stress response, metabolism and development (Jin & Martin, 1999). The MYB DNA-binding domain is a helix–turn–helix structure $\sim$ 50 amino acids in length that is highly conserved in fungi, animals and plants (Romero *et al.*, 1998). The target of the MYB domain is a consensus sequence called the MYB binding site (MBS). There are three types of MBS: MBS1 (CNGTTR), MBS2 (GKTWGTTR) and MBS2G (GKTWGGTR) (Lüscher & Eisenman, 1990; Grotewold *et al.*, 1991).

* Corresponding author. Tel: +81 75 753 6138. Fax: +81 75 753 6486. e-mail: arabis@kais.kyoto-u.ac.jp

The MYB domain can be repeated in tandem up to three times and the repeats are called R1, R2 and R3 in the N- to C-terminal direction. MYB proteins with three repeated MYB domains are denoted R1R2R3–MYB proteins. DNA binding specificity is determined primarily by R2 and R3 (Jin & Martin, 1999); R1 is not thought to be important for DNA binding specificity. Some MYB proteins have two MYB domains and lack R1 (R2R3–MYB), and proteins with a single MYB domain are also known (R1–MYB, R2–MYB or R3–MYB). Therefore, MYB genes are classified into three groups based on the number of MYB domains. These three groups of MYB genes originated by a duplication of the R1–MYB domain, which is thought to have occurred more than a billion years ago 'in the last common ancestor of eukaryotes' (Lipsick, 1996).

Most plant MYB proteins are the R2R3 type (Jin & Martin, 1999). In *A. thaliana*, there are at least 85 and probably more than 100 R2R3–MYB proteins (Romero *et al.*, 1998). The *Atmyb2* locus is in the terminal region of the long arm of chromosome 2 (Urao *et al.*, 1993). ATMYB2 is 273 amino acids long

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

(aa) and has two distinct functional regions (Urao *et al*., 1996), the MYB domain (104 aa) and the acidic transactivating domain (30 aa). *Atmyb2* is transcriptionally induced in root cells by conditions of stress, including low oxygen (Hoeren *et al*., 1998), water stress, high salt, drought and abscisic acid (Urao *et al*., 1993). Target genes of ATMYB2 include the dehydration-responsive *rd* genes (Urao *et al*., 1993; Abe *et al*., 1997) such as *rd22*, *rd29A* and *rd29B* (Yamaguchi-Shinozaki *et al*., 1995; Abe *et al*., 1997). The MBSs in the *rd* genes are not clear. When ATMYB2 binds to the promoter of *rd22*, its transactivating domain interacts with MYC (myelocell)-related protein, which is encoded by *rd22BP1*, and transactivates the promoter of *rd22* (Urao *et al*., 1996). ATMYB2 also binds to the promoter of *Adh*, which is required under low-oxygen conditions. There are two MBSs in the *Adh* promoter, GCAACG and AAAACCAA, which are located approximately 290 bp and 230 bp upstream of *Adh*, respectively (Hoeren *et al*., 1998).

This study has several goals. The first is to investigate the level and pattern of nucleotide variation in the regulatory gene *Atmyb2* in *A. thaliana* and the related species *Arabis gemmifera* and *Arabidopsis himalaica*. Most previous studies analysed nucleotide variation in *A. thaliana* genes that produce catalytic enzymes. Because ATMYB2 regulates the expression of genes that respond to environmental stress, ATMYB2 could play an important role in adaptation. Thus, the function of ATMYB2 might influence the level and pattern of DNA variation in *Atmyb2*, causing it to differ from other genes studied to date. DNA polymorphism was recently analysed in the regulatory MADS-box genes (Purugganan & Suddith, 1998, 1999), so it is possible to compare DNA variation in these two regulatory genes. The second goal of this study relates to the neutral mutation theory (Kimura, 1983), which suggests that levels of polymorphism and divergence are positively correlated. Because ATMYB2 has two functionally different regions, this study provides the opportunity to test this hypothesis using a single gene. The third goal of this study is to investigate a possible epistatic relationship between *Atmyb2* and *Adh*. Recent studies of the *Adh* region revealed that dimorphic variation (two divergent sequence types) is restricted to the 5′ regulatory and coding regions of *Adh* (Innan *et al*., 1996; Miyashita, 2001). Because ATMYB2 interacts physically with MBSs in the 5′ regulatory region of *Adh*, the dimorphic pattern in *Adh* might be related to regulatory variation of ATMYB2 and vice versa. If this is the case, there might also be a dimorphic pattern in the *Atmyb2* region. It is of interest to investigate whether variation in the MYB domain corresponds to variation in the 5′ flanking region of *Adh*.

2. Materials and methods

(i) *Plant species and growth conditions*

In this study, we used 19 ecotypes of *A. thaliana* and one accession each of *A. gemmifera* and *A. himalaica*. The ecotypes of *A. thaliana* included samples from locations distributed around the world (Table 1). Growth conditions for these plants were described in Kawabe *et al*. (1997).

(ii) *PCR amplification and sequencing*

Total DNA was extracted by a modified CTAB method, and used as PCR template. Approximately 2·2 kb of the *Atmyb2* region was amplified using PCR primers in the 5′ and 3′ flanking regions of *Atmyb2*. The PCR was performed by 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 2 min and extension at 72 °C for 3 min. Primer sequences were selected based on ecotype Columbia (GenBank accession number D14712) and were as follows: 5′-AGT GAT TGA CAG AGG AAA GA-3′ (sense) and 5′-ATT TTC GTT TTG TCG TTT TT-3′ (antisense). A second PCR reaction was carried out using the first PCR product as template. The second PCR product was purified using 1% agarose gel electrophoresis followed by extraction with glass powder. Sequencing reactions were carried out using the dideoxy chain

Table 1. *Plant materials used in this study*

Species	Name code	Accession code[a]	Origin
Arabidopsis thaliana	Aa-0	JA1	FRG
	Ag-0	JA2	France
	Ci-0	JA54	UK
	Dra-0	JA68	Czechoslovakia
	Es-0	JA76	Finland
	Gr-1	JA95	Austria
	Hau-0	JA100	Denmark
	In-0	JA110	Austria
	Ita-0	JA112	Morocco
	Kas-1	JA119	India
	Mr-0	JA155	Italy
	Mt-0	JA158	Libya
	Ost-0	JA179	Sweden
	Shokei	JW101	Japan
	Su-0	JA225	UK
	Ts-1	JA230	Spain
	Uk-2	JA242	FRG
	Ws-0	JA252	Russia
	Yo-0	JA262	USA
Arabis gemmifera	Ashibi56		Japan
Arabidopsis himalaica		JOS18	

[a] Accession number of the Sendal *Arabidopsis* Seed Stock Center (SASSC).

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

termination method and the Thermo sequenase Cy5 dye terminator kit (Amersham Pharmacia Biotech). Both strands were sequenced using a Pharmacia ALFred sequencer with sequencing primers at 400–600 bp intervals.

Total DNA of *A. gemmifera* and *A. himalaica* was used as PCR template. PCR primers for *A. gemmifera* were 5′-GCG ATA CTG TGT GGG AAT AA-3′ (sense) in the 5′ flanking region and 5′-ATG TCG TAT CGG GGC AGA AC-3′ (antisense) in exon 3. Primers for *A. himalaica* were 5′-GTC AAC TTC GTC TCT ATT CA-3′ (sense) in exon 1 and the same antisense primer as for *A. gemmifera*. PCR conditions were identical for the two relative species and *A. thaliana*. 5 μl of PCR product was analysed by 1% agarose gel electrophoresis. One band was observed in DNA from *A. gemmifera* and *A. himalaica*, and it was assumed that there is one copy of *Atmyb2* in these species. The PCR product was cloned into pUC18 and sequenced using the Thermo Sequenase fluorescently labelled cycle sequencing kit with 7-deaza-dGTP (Amersham Pharmacia Biotech). Three plasmid clones were sequenced for each species and a consensus sequence was determined from these clones. All sequence information was deposited in the DDBJ (accession numbers AB052230–AB052250).

(iii) *Sequence analysis*

20 sequences of *A. thaliana Atmyb2*, including the sequence of the Columbia ecotype, and one sequence each of *A. gemmifera* and *A. himalaica* were analysed. The region analysed for *A. thaliana* is between nucleotide positions 101 (5′ flanking region) and 2381 (3′ flanking region) of the nucleotide sequence of ecotype Columbia. The region analysed for *A. gemmifera* is between positions 533 (5′ flanking region) and, 1957 (36 bp upstream of the end of exon 3) and for *A. himalaica*, the region is between positions 1099 (120 bp downstream of the initiation site in exon 1) and, 1957. The coding region of *Atmyb2* was analysed in two distinct parts, the MYB domain and the non-MYB region, according to their different functions. Analysis of intra- and interspecies variation was performed using DnaSP 3.00 (Rozas & Rozas, 1999). Nucleotide diversity (π: Tajima & Nei, 1984) and θ ($4Ne\mu$: Watterson, 1975) were estimated after excluding indels. Neutrality was examined by four statistical tests: Tajima's test (1989) and Fu & Li's test (1993) were used for *A. thaliana*, and the MK test (McDonald & Kreitman, 1991) and HKA test (Hudson *et al.*, 1987) were used for interspecies comparison. The degree of divergence between species was estimated by the method of Jukes & Cantor (1969) for total (K), synonymous (K_s) and replacement sites (K_a). Published nucleotide sequences of *Adh*, *ChiA*, *ChiB* and *PgiC* of *A. thaliana*, *A. gemmifera* and *A. himalaica* were used to estimate genetic distance (Miyashita *et al.*, 1996; Innan *et al.*, 1996; Kawabe *et al.*, 1997; Miyashita *et al.*, 1998; Kawabe & Miyashita, 1999; Kawabe *et al.*, 2000), except for *ChiA* of *A. himalaica* (A. Kawabe, unpublished).

3. Results

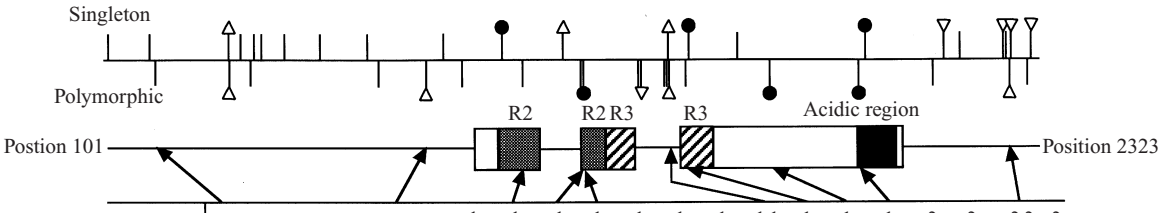
(i) *Nucleotide polymorphism at* Atmyb2 *in* A. thaliana

The nucleotide polymorphism in *Atmyb2* included 31 nucleotide changes (15 singletons) and 32 indels (23 singletons) for a total of 63 variations (Fig. 1). Variations were detected throughout the sequenced region but there was no evidence of a dimorphic pattern, which has been detected in other regions of the *A. thaliana* genome. In the *Atmyb2* coding region, there were six replacement changes (three singletons) and four synonymous changes (one singleton). One nonconservative replacement change was observed in exon 3 at nucleotide position 1689, which corresponds to changing the 172th codon from AGT (Ser) to AAT (Asn) (Miyata *et al.*, 1979). This nonconservative change is associated with a change in molecular weight, suggesting that the Ser-to-Asn substitution might change interaction with MYC in the non-MYB region of ATMYB2.

The level of nucleotide polymorphism in *Atmyb2* was analysed (Table 2). The level of variation is fairly constant over different functional regions of *Atmyb2*, but intron 2 has a relatively high level of variation. No nucleotide polymorphism was detected in intron 1 but there was a 16 bp singleton deletion. The results show that nucleotide variation is higher in the MYB domain than in the non-MYB region. In the MYB domain, synonymous variation is higher than replacement variation, whereas both synonymous and replacement variation are low in the non-MYB region. In addition, noncoding regions have a low level of variation except for intron 2. Tajima's test (1989) and Fu & Li's test (1993) were applied to the *Atmyb2* region (Table 2). None of the tests showed significant deviation from neutrality. Tajima's D and Fu and Li's D^* values were all negative, indicating that the observed number of rare variations exceeds the expected number in an equilibrium population. Sliding-window analysis was conducted and the results indicate no clear peak of polymorphism except for a small peak in intron 2 (data not shown).

Data on nucleotide variation are available for several *A. thaliana* nuclear genes, and these levels are compared in Table 3. *Atmyb2* has the lowest level of nucleotide polymorphism of the *A. thaliana* nuclear genes that have been studied but similar levels of variation are observed in *Atmyb2* and *CHI* (Kuittinen & Aguadé, 2000). Coincidentally, dimorphic patterns

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press



Position	2 1 7	3 9 2	4 4 3	7 5 0	8 8 6 6 5 6	9 5 3	1 0 9 9	1 2 3 2	1 2 3 3	1 3 6 3	1 3 8 8	1 4 3 1	1 4 4 1	1 1 4 4 2 5	1 4 8 9	1 6 8 9	1 8 9 2	2 0 8 0	2 2 5 6	2 2 5 6 7 5	2 3 0 8	No. of uniques
Consensus	T	A	A	C	– –	C	C	C	G	C	A	T	T	4T	G	G	G	T	A	9A	T	
Col-0	*	–	*	A	A T	*	T	G	C	*	*	*	*	*	*	*	*	*	*	*	C	6
Uk-2	G	*	*	A	A T	*	T	G	C	*	*	*	*	*	*	*	*	*	*	*	C	0
Su-0	*	*	*	A	A T	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	C	0
Ts-1	G	–	*	A	A T	*	T	*	*	*	*	*	*	*	*	*	*	*	*	*	C	3
Yo-0	*	*	*	A	A T	*	*	*	*	*	*	*	*	*	*	*	*	*	*	8A	C	0
Ag-0	*	*	*	*	* *	*	*	*	*	*	_	*	*	*	A	*	*	*	*	*	*	0
Gr-1	*	*	*	*	* *	*	*	*	*	*	_	*	*	*	A	*	*	*	*	*	*	0
Hau-0	*	*	*	*	* *	*	*	*	*	*	_	*	*	*	A	*	*	*	*	*	*	1
Ws-0	*	*	C	*	* *	*	*	*	*	*	*	*	*	5T	*	A	*	*	*	*	*	0
Kas-1	*	*	C	*	* *	*	*	*	*	*	*	*	*	*	*	A	*	*	*	8A	*	0
Ost-0	*	*	C	*	* *	*	*	*	*	*	*	*	*	5T	*	A	*	*	*	*	*	1
In-0	*	*	*	*	* *	*	*	*	*	*	*	C	C	6T	*	*	A	*	*	*	*	0
Dra-0	*	*	*	*	* *	*	*	*	*	*	*	C	C	6T	*	*	A	*	*	*	*	1
Mt-0	*	*	*	*	A T	*	*	*	*	*	*	C	C	6T	*	*	*	*	*	11A	*	3
Ita-0	*	*	*	*	* *	T	*	*	*	T	*	C	C	5T	*	*	*	C	*	*	*	16
Mr-0	*	*	*	*	* *	T	*	*	*	T	*	C	C	5T	*	*	*	C	*	8A	*	0
Ci-0	*	*	*	*	* *	*	*	*	*	*	*	C	C	6T	*	*	*	*	*	*	*	2
Aa-0	*	*	*	*	* *	T	*	*	*	*	*	*	*	3T	*	*	*	*	C	*	*	2
Shokei	*	*	*	*	* *	*	*	*	*	*	*	*	*	5T	*	*	*	*	C	10A	*	2
Es-0	*	*	*	*	* *	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	1

Fig. 1. Polymorphism at the *Atmyb2* locus of *A. thaliana*. Vertical bar indicates DNA variation. Filled circle indicates replacement substitution and triangle indicates indel. Shaded regions in exon are DNA-binding domains. Variations found more than once are summarized at the bottom. *, identical to consensus.

Table 2. *Level of nucleotide variation in the* Atmyb2 *region of* A. thaliana

$n = 20$	Number of sites	S	π	θ	Talima's D	Fu and Li's D^*
Entire region	2198	31	0·0027	0·0040	−1·30 NS	−1·05 NS
Coding region	819	10	0·0022	0·0034	−1·30 NS	−0·49 NS
Synonymous	181·23	4	0·0046	0·0062	−0·77 NS	0·17 NS
Replacement	637·77	6	0·0015	0·0027	−1·40 NS	−0·86 NS
MYB domain	312	6	0·0036	0·0054	−1·08 NS	−0·15 NS
Synonymous	71·50	3	0·0102	0·0118	−0·44 NS	0·87 NS
Replacement	240·50	3	0·0016	0·0035	−1·14 NS	−0·59 NS
Non-MYB region	507	4	0·0013	0·0022	−1·21 NS	−0·76 NS
Synonymous	109·73	1	0·0009	0·0026	−1·16 NS	−1·54 NS
Replacement	397·27	3	0·0014	0·0021	−0·91 NS	−0·12 NS
Noncoding region	1379	21	0·0030	0·0043	−1·19 NS	−1·22 NS
5′ flanking	877	13	0·0023	0·0042	−1·63 NS	−1·98 NS
Intron 1	75	0	0	0	NA	NA
Intron 2	99	3	0·0109	0·0085	0·72 NS	1·01 NS
3′ flanking	328	5	0·0030	0·0043	−0·95 NS	−0·41 NS

S, number of segregating sites; NS, nonsignificant ($P > 0·05$); NA, not applicable.

of DNA variation were detected in neither *CHI* nor *Atmyb2*. In the absence of dimorphism (two divergent sequence types), nucleotide variation is expected to be low. As noted above, in the non-MYB region of *Atmyb2*, the level of polymorphism is exceptionally low especially for synonymous sites ($\pi = 0·0009$). Because the level of replacement variation at *Atmyb2* is comparable to other loci, it can be concluded that low synonymous variation in the non-MYB region contributes greatly to the overall low level of nucleotide variation at this locus.

(ii) *Recombination at* Atmyb2

It has been reported that the level of polymorphism and the recombination rate are positively correlated in genes from *Drosophila* (Moriyama & Powell, 1996). Thus, recombination parameters at *Atmyb2* were

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

Table 3. *Comparison of level of nucleotide variation between* Atmyb2 *and other nuclear genes of* A. thaliana

			Coding region					
	Entire region		Total		Synonymous		Replacement	
Locus	π	θ	π	θ	π	θ	π	θ
Atmyb2	0·0027	0·0040	0·0022	0·0034	0·0046	0·0062	0·0015	0·0027
MYB domain			0·0036	0·0054	0·0102	0·0118	0·0016	0·0035
Non-MYB region			0·0013	0·0022	0·0009	0·0026	0·0014	0·0021
Adh[a]	0·0080	0·0093	0·0056	0·0052	0·0192	0·0152	0·0022	0·0021
ChiA[b]	0·0104	0·0202	0·0067	0·0140	0·0163	0·0345	0·0039	0·0081
ChiB[c]	0·0091	0·0112	0·0049	0·0062	0·0173	0·0209	0·0009	0·0015
CHI[d]	0·0040	0·0053	0·0015	0·0018	0·0028	0·0030	0·0011	0·0014
PgiC[e]	0·0038	0·0073	0·0024	0·0051	0·0051	0·0115	0·0016	0·0032
Cal[f]	0·0075	0·0122	0·0053	0·0109	0·0049	0·0123	0·0055	0·0106
Ap3[g]	0·0064	0·0132	0·0046	0·0115	0·0073	0·0168	0·0076	0·0147
Pl[g]	0·0053	0·0100	0·0032	0·0077	0·0041	0·0098	0·0030	0·0072

a, Innan *et al*. (1996); *b*, Kawabe *et al*. (1997); *c*, Kawabe & Miyashita (1999); *d*, Kuittinnen & Aguadé (2000); *e*, Kawabe *et al*. (2000); *f*, Purugganan & Suddith (1998); *g*, Purugganan & Suddith (1999).

Table 4. *Recombination parameters of nuclear genes in* Arabidopsis thaliana

Locus	*Rm*	*S*	*C*	*C* per site	θ per gene	cM Mbp^{-1}
Atmyb2	1	31	51·4	0·0234	5·8	3·75
Adh	6	75	3·3	0·0014	18·7	2·40
ChiA	2	119	0·0001	0·0000	19·0	12·00
ChiB	5	82	1·3	0·0006	19·6	2·37
CHI	3	20	132·0	0·0688	7·4	17·80
Cal	5	90	67·9	0·0309	16·3	5·32
Ap3	1	77	38·9	0·0233	10·7	155·4
Pl	1	67	> 1000	–	10·7	2·75

Table 5. *Genetic divergence of nuclear genes between* A. thaliana *and its relatives*

Locus	*K*	K_s	K_a	K_a/K_s
Between *A. thaliana* and *A. gemmifera*				
Atmyb2	0·0730	0·2334	0·0330	0·1414
MYB domain	0·0436	0·1758	0·0084	0·0478
Non-MYB region	0·0939	0·2779	0·0500	0·1799
Adh	0·0607	0·1924	0·0162	0·0842
ChiA	0·0568	0·1385	0·0334	0·2412
ChiB	0·0537	0·1568	0·0245	0·1563
PgiC	0·0438	0·1716	0·0093	0·0542
Between *A. thaliana* and *A. himalaica*				
Atmyb2	0·1039	0·2742	0·0620	0·2261
MYB domain	0·0760	0·2998	0·0208	0·0694
Non-MYB region	0·1213	0·2583	0·0875	0·3388
Adh	0·0790	0·3240	0·0200	0·0617
ChiA	0·0868	0·2656	0·0433	0·1630

estimated (Table 4). The estimated minimum number of recombination events (*Rm*: Hudson & Kaplan, 1985) at *Atmyb2* was 1. This small *Rm* value at *Atmyb2* suggests that few recombination events were detectable, which could be due to the low level of polymorphism in the gene. By contrast, *C* (Hudson,

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

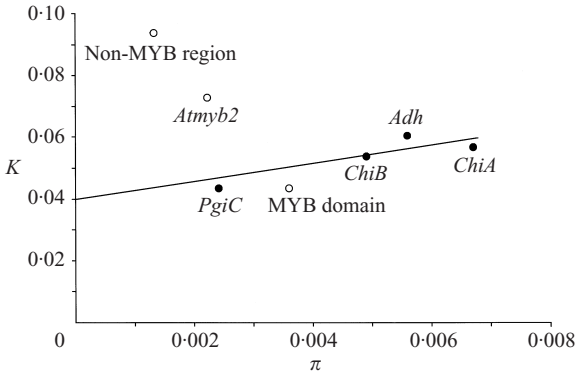


Fig. 2. Correlation between nucleotide diversity (π) and divergence (K) in the coding region of nuclear genes in *A. thaliana* and *A. gemmifera*. Regression line was determined for the values of *Adh*, *PgiC*, *ChiA* and *ChiB*. The coefficient of determination (R^2) was 0·794.

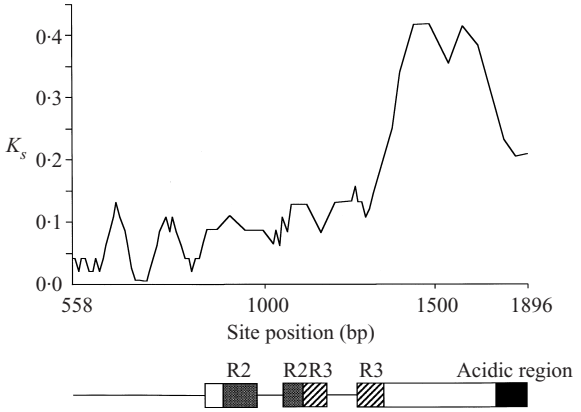


Fig. 3. Sliding-window analysis for K_s between *A. thaliana* and *A. gemmifera*. Window size is 50 bp with a 10 bp increment.

1987) was estimated to be 51·4 at *Atmyb2*, which is larger than for other loci. It is difficult to find a clear correlation between level of variation (Table 3) and these parameters. Genetic distance was estimated in centiMorgans per Mbp (cM Mbp^{-1}) using data from a recombination inbred line (http://www.arabidopsis.org/servlets/mapper). The estimated genetic length of the region that includes *Atmyb2* is 3·57 cM Mbp^{-1}, which is intermediate among selected loci. These results indicate that there is no clear association between the level of polymorphism and recombination parameters in *Atmyb2*, and that recombination is not likely to be the reason for low nucleotide polymorphism in this region.

(iii) *Divergence between* A. thaliana *and its relative species*

Under the neutral mutation theory (Kimura, 1983), degree of divergence is expected to be low when there is a low level of polymorphism. The degree of divergence was estimated between *A. thaliana* and related species in the coding region of *Atmyb2* and other loci (Table 5). The results clearly show that *Atmyb2* has larger divergence than other loci. Furthermore, the degree of divergence in the non-MYB region of *Atmyb2* is larger than in the MYB domain, indicating that the non-MYB region is responsible for the large degree of divergence in the coding region of *Atmyb2*. These results are contrary to the prediction of the neutral mutation theory.

There is a good correlation between polymorphism in *A. thaliana* and divergence between *A. thaliana* and *A. gemmifera* (Fig. 2); however, the number of loci was small and so the regression coefficient is not statistically significant. Nevertheless, there is a low level of polymorphism and large degree of divergence in the non-MYB region of *Atmyb2*. By contrast, in the MYB domain of *Atmyb2*, the levels of polymorphism and divergence are in concordance with other loci. Larger divergence in the non-MYB region than in the MYB domain was also detected in the other interspecies comparisons (Table 5). In addition, the ratio of K_a and K_s is higher in the non-MYB region than in the MYB domain for all interspecies comparisons.

Sliding window analysis was applied to interspecies data for *A. thaliana* and *A. gemmifera* (Fig. 3). This analysis reveals that the MYB domain has the lowest divergence and that the level of divergence increases greatly in the 3′ terminal region, with the exception of the acidic region. This result indicates that large divergence occurs in the sequence of unknown function in the non-MYB region, which is between the MYB domain and the acidic region. BLAST homology searches were conducted for nucleotide or protein sequences corresponding to the region of the high divergence, and these searches did not reveal the presence of any functional sequence in this region.

(iv) *Relationship between polymorphism and divergence*

The HKA test was used to compare the relationship between polymorphism and divergence in the coding regions of *Atmyb2* and other loci (data not shown). The results indicate that the ratio of polymorphism to divergence in the coding region of *Atmyb2* was smaller than for other loci, although significance was detected in a few comparisons. This pattern was more evident in the non-MYB region of *Atmyb2* (Table 6), but the pattern was abolished when the MYB domain was considered (i.e. no significance was detected in the HKA test; data not shown). Thus, only the non-MYB region of *Atmyb2* has a significant result in the HKA test.

The HKA test was also used to compare the difference in the ratio of polymorphism and divergence

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

Table 6. *Results of the HKA test on the non-MYB region of* Atmyb2 *and other nuclear genes between* A. thaliana *and* A. gemmifera

Region	Number of segregating sites[a] Within	Between	χ^2	P
Non-MYB region ($n = 20$)				
Total	4 (0·008)	40·25 (0·088)		
Synonymous	1 (0·009)	22·98 (0·231)		
vs *Adh* ($n = 17$)				
Total	19 (0·017)	44·21 (0·039)	4·20	0·040*
Synonymous	13 (0·051)	30·06 (0·119)	4·56	0·033*
vs *ChiA* ($n = 17$)				
Total	44 (0·049)	48·82 (0·054)	8·56	0·003**
Synonymous	25 (0·123)	25·94 (0·126)	8·50	0·004**
vs *ChiB* ($n = 17$)				
Total	22 (0·022)	52·47 (0·052)	4·08	0·043*
Synonymous	17 (0·071)	34·00 (0·141)	5·18	0·023*
vs *PgiC* ($n = 21$)				
Total	31 (0·018)	70·33 (0·042)	4·06	0·044*
Synonymous	16 (0·041)	58·62 (0·153)	2·67	0·102 NS

[a] The per-site value is shown in parenthesis.
*, $P < 0·05$; **, $P < 0·01$; NS, non-significant.

Table 7. *Results of HKA test on MYB domain and non-MYB region between* A. thaliana *and* A. gemmifera

Region ($n = 20$)	Number of segregating sites[a] Within	Between	χ^2	P
MYB domain				
Total	6 (0·019)	13·20 (0·042)		
Synonymous	3 (0·042)	11·20 (0·158)		
Replacement	3 (0·012)	2·00 (0·008)		
Non-MYB region				
Total	4 (0·008)	40·25 (0·088)	4·15	0·042*
Synonymous	1 (0·009)	22·98 (0·231)	2·56	0·109 NS
Replacement	3 (0·008)	17·28 (0·048)	4·39	0·036*

[a] The per-site value is shown in parenthesis.
*, $P < 0·05$; NS, non-significant.

in the MYB domain and the non-MYB region of *Atmyb2* (Table 7). Using *A. gemmifera* as a reference species, the test revealed a significant difference in the ratios of polymorphism and divergence for replacement sites in the two regions. Although significance was not detected for synonymous sites, the ratio of polymorphism and divergence ($0·009/0·231 = 0·04$) in the non-MYB region is much smaller than in the MYB domain ($0·042/0·158 = 0·27$). Significance was not obtained when *A. himalaica* was used for interspecies comparison (data not shown). This is probably because less sequence information is available for *A. himalaica* than for *A. gemmifera*. However, lower synonymous polymorphism was still observed in the non-MYB region.

The MK test was applied separately to the *Atmyb2* coding region, the MYB domain and the non-MYB region (data not shown). None of the tests showed significance, although the ratio of synonymous to replacement changes is higher in polymorphism data in any regions. However, the ratio of fixed synonymous and replacement substitutions is significantly different for the MYB domain and the non-MYB region (Table 8). A high frequency of fixed replacements was detected in the non-MYB region when interspecies comparisons were carried out with *A. gemmifera* and *A. himalaica*. This indicates that the non-MYB region has experienced more qualitative change than the MYB domain during speciation. Approximately 40 % or 43 % of the fixed replacements were drastic amino

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

Table 8. *Results of* G *test on MYB domain and non-MYB region between* A. thaliana *and its relatives*

	A. gemmifera		*A. himalaica*	
Outgroup species	Synonymous	Replacement	Synonymous	Replacement
Fixed in MYB domain	10	1	14	2
Fixed in non-MYB region	22	17	20	26
G (William's correction)	4·95		9·98	
P	0·026*		0·002**	

*, $P < 0{\cdot}05$; **, $P < 0{\cdot}01$.

acid changes when comparing *A. gemmifera* or *A. himalaica*, respectively. None of these drastic amino acid substitutions were in the acidic region of *Atmyb2*.

4. Discussion

(i) *Nucleotide polymorphism in the* Atmyb2 *region of* A. thaliana

This study shows that the level of nucleotide polymorphism in the *Atmyb2* region is lower than for other loci in *A. thaliana*, and that dimorphism (two divergent sequence types) is not detected at this locus. The lack of dimorphism implies that there might not be any sequence variation in *Atmyb2* that corresponds to the dimorphism at *Adh*. Because ATMYB2 is involved in the transcriptional regulation of many genes, *Atmyb2* cannot be specific to *Adh*. This suggests that epistasis between *Adh* and *Atmyb2* is unlikely.

In the *CHI* region of low polymorphism, a dimorphic pattern was also not observed (Kuittinen & Aguadé, 2000). It is not unexpected that the level of polymorphism is low if two divergent sequences are absent. Kuittinen & Aguadé (2000) explained the low level of polymorphism at *CHI* by the hitchhiking effect, referring to Tajima's negative *D* and the presence of one peak in the distribution of pairwise nucleotide differences. However, these features could reflect the recent expansion of *A. thaliana*. *Atmyb2* also has Tajima's negative *D* (Table 2) and one peak in the distribution of pairwise nucleotide differences (data not shown). In this case, one should also consider that the level of variation differs between the MYB domain and non-MYB region of the gene. It seems unlikely that hitchhiking influences the two adjacent regions differently. Thus, hitchhiking is probably not responsible for the low level of nucleotide variation at *Atmyb2*. It is also possible that the low level of polymorphism in *Atmyb2* is caused by purifying selection owing to a functional constraint. Because the MYB domain binds to MBSs in many genes (Yamaguchi-Shinozaki *et al.*, 1995; Abe *et al.*, 1997; Hoeren *et al.*, 1998) and the non-MYB region includes the acidic region that interacts with MYC-related proteins (Urao *et al.*, 1996), the *Atmyb2* coding region is likely to be under complex functional constraints. In addition, the MYB domain and non-MYB region of ATMYB2 might be differently constrained, and the non-MYB region might be subject to stronger selection than the MYB domain. Non-significance in the tests of neutrality on *Atmyb2* might be related to the strength of the purifying selection on this locus. If purifying selection were strong, the level of polymorphism would be low, so the tests would fail to detect significant deviation from neutrality (Tajima, personal communication). These ideas are consistent with the results presented here.

(ii) *Comparison between* Atmyb2 *and MADS-box genes*

The MADS-box genes are regulatory genes that are under strong purifying selection, and there are both similarities and differences between these genes and *Atmyb2*. Dimorphism is present at the MADS-box loci but not at *Atmyb2*. It was thought that dimorphism is under balancing selection at several loci (Hanfstingl *et al.*, 1994; Innan *et al.*, 1996; Kawabe *et al.*, 2000) but there is no clear evidence of balancing selection for the MADS-box genes (Purugganan & Suddith, 1998, 1999). Instead, neutrality tests produced significantly negative test statistics, reflecting an excess of singleton polymorphisms. This result indicates the presence of purifying selection on the MADS-box genes and suggests that purifying selection might act on many regulatory genes because of their functional importance. However, regulatory genes might have different levels of variation because of differences in selection intensity. In addition, it is possible that the dimorphism at the MADS-box genes could decline or eventually be eliminated by purifying selection.

(iii) *Divergence in the* Atmyb2 *region*

This study shows a low level of polymorphism and a high degree of divergence in the coding region of *Atmyb2*. This result is inconsistent with the neutral mutation theory. However, the high divergence is

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

mainly localized to the non-MYB region of *Atmyb2*. In addition, significant results of the HKA test for *Atmyb2* and other loci were specific to the non-MYB region. HKA tests also indicate that the MYB domain and the non-MYB region have different ratios of polymorphism and divergence. In other words, the MYB domain is relatively highly conserved and the non-MYB region is not. This result agrees with a previous study indicating that the C-terminal region of the non-MYB region of plant MYB proteins is less well conserved than the MYB domain (Jin & Martin, 1999). This study also showed that the K_a/K_s ratio in the non-MYB region was higher than other genic regions in all interspecies comparisons. The high divergence in the non-MYB region suggests species-specific changes in its amino acid sequence. However, drastic amino acid changes were not detected in the acidic region itself, and sliding window analysis (Fig. 3) showed a high peak of divergence in a region between the MYB domain and acidic region. Therefore, the acidic region might not have undergone any large functional changes, but the surrounding sequence might have influenced its function.

We thank R. Terauchi and S. Nasuda for comments and suggestion on the early version of this manuscript. This is contribution number 564 from the Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University.

References

Abe, H., Shinozaki, K. Y., Urao, T., Iwasaki, T., Hosokawa, D. & Shinozaki, K. (1997). Role of *Arabidopsis* MYC and MYB homologues in drought- and abscisic acid-regulated gene expression. *The Plant Cell* **9**, 1859–1867.

Fu, Y. X. & Li, W. H. (1993). Statistical tests of neutrality of mutation. *Genetics* **133**, 693–709.

Grotewold, E., Athma, P. & Peterson, T. (1991). Alternatively spliced products of the maize *P* gene encode proteins with homology to the DNA binding domain of Myb-like transcription factors. *Proceedings of the National Academy of Sciences of the USA* **88**, 4587–4591.

Hanfstingl, U., Berry, A., Kellog, E. A., Costa Iii, J. T., Rudiger, W. & Ausubel, M. (1994). Haplotype divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: role for both balancing and directional selection? *Genetics* **138**, 811–828.

Hoeren, F. U., Dolferus, R., Wu, Y., Peacock, W. J. & Dennis, E. S. (1998). Evidence for a role for AtMYB2 in the induction of the *Arabidopsis* alcohol dehydrogenase gene (*ADH1*) by low oxygen. *Genetics* **149**, 479–490.

Hudson, R. R. (1987). Estimating the recombination parameter of a finite population model without selection. *Genetical Research* **50**, 245–250.

Hudson, R. R. & Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Innan, H., Tajima, F., Terauchi, R. & Miyashita, N. T. (1996). Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. *Genetics* **143**, 1441–1452.

Jin, H. & Martin, C. (1999). Multifunctionality and diversity within the plant *MYB*-gene family. *Plant Molecular Biology* **41**, 577–585.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism* (ed. H. Munro), pp. 21–132. New York, USA: Academic Press.

Kawabe, A., Innan, H., Terauchi, R. & Miyashita, N. T. (1997). Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. *Molecular Biology and Evolution* **14**, 1303–1315.

Kawabe, A. & Miyashita, N. T. (1999). DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. *Genetics* **153**, 1445–1453.

Kawabe, A., Yamane, K. & Miyashita, N. T. (2000). DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. *Genetics* **156**, 1339–1347.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press.

Kuittinen, H. & Aguadé, M. (2000). Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. *Genetics* **155**, 863–872.

Lipsick, J. S. (1996). One billion years of MYB. *Oncogene* **13**, 223–235.

Lüsher, B. & Eisenman, R. (1990). New light on Myc and Myb. Part 2. Myb. *Genes and Development* **4**, 7092–7096.

Mcdonald, J. H. & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in Drosophila. *Nature* **351**, 652–654.

Miyashita, N. T. (2001). DNA variation in the 5′ upstream region of the *Adh* locus of the wild plants *Arabidopsis thaliana* and *Arabis gemmifera*. *Molecular Biology and Evolution* **18**, 164–171.

Miyashita, N. T., Innan, H. & Terauchi, R. (1996). Intra- and interspecific variation in the alcohol dehydrogenase locus region of wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. *Molecular Biology and Evolution* **13**, 433–436.

Miyashita, N. T., Kawabe, A., Innan, H. & Terauchi, R. (1998). Intra- and interspecific DNA variation and codon bias of alcohol dehydrogenase (*Adh*) locus in *Arabis* and *Arabidopsis* species. *Molecular Biology and Evolution* **15**, 1420–1429.

Miyata, T., Miyazawa, S. & Yasunaga, T. (1979). Two types of amino acid substitutions in protein evolution. *Journal of Molecular Evolution* **12**, 219–236.

Moriyama, E. & Powell, J. R. (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* **13**, 261–277.

Purugganan, M. D. & Suddith, J. I. (1998). Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. *Proceedings of the National Academy of Sciences of the USA* **95**, 8130–8134.

Purugganan, M. D. & Suddith, J. I. (1999). Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. *Genetics* **151**, 839–848.

Romero, I., Fuertes, A., Benito, M. J., Malpica, J. M., Leyva, A. & Paz-Ares, J. (1998). More than 80 R2R3–MYB regulatory genes in the genome of *Arabidopsis thaliana*. *The Plant Journal* **14**, 273–284.

Rozas, J. & Rozas, R. (1999). DnaSP version 3.00, a novel software package for extensive molecular population genetics analysis. *Computer Applications in the Biosciences* **13**, 307–311.

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press

Tajima, F. (1989). Statistical test for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Tajima, F. & Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* **1**, 269–285.

Urao, T., Shinozaki, K. Y., Urao, S. & Shinozaki, K. (1993). An *Arabidopsis myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *The Plant Cell* **5**, 1529–1539.

Urao, T., Noji, M., Shinozaki, K. Y. & Shinozaki, K. (1996). A transcriptional activation domain of ATMYB2, a drought-inducible *Arabidopsis* Myb-related protein. *The Plant Journal* **10**, 1145–1148.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–276.

Yamaguchi-Shinozaki, K., Urao, T. & Shinozaki, K. (1995). Regulation of genes that are induced by drought stress in *Arabidopsis thaliana*. *Journal of Plant Research* **108**, 127–136.

https://doi.org/10.1017/S0016672302005815 Published online by Cambridge University Press