Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Gene 378 (2006) 74–83

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Evolutionary history of the *Asr* gene family

Nicolás Frankel [a], Fernando Carrari [b], Esteban Hasson [c], Norberto D. Iusem [a,*]

[a] *Laboratorio de Fisiología y Biología Molecular, Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, and IFIBYNE-CONICET, Buenos Aires, Argentina*
[b] *Instituto de Biotecnología, CICVyA, INTA, Castelar, Buenos Aires, Argentina* [1]
[c] *Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina*

Received 26 October 2005; received in revised form 8 May 2006; accepted 11 May 2006
Available online 26 May 2006
Received by O. Clay

Abstract

The *Asr* gene family is widespread in higher plants. Most *Asr* genes are up-regulated under different environmental stress conditions and during fruit ripening. ASR proteins are localized in the nucleus and their likely function is transcriptional regulation. In cultivated tomato, we identified a novel fourth family member, named *Asr4*, which maps close to its sibling genes *Asr1–Asr2–Asr3* and displays an unshared region coding for a domain containing a 13-amino acid repeat. In this work we were able to expand our previous analysis for *Asr2* and investigated the coding regions of the four known *Asr* paralogous genes in seven tomato species from different geographic locations. In addition, we performed a phylogenetic analysis on ASR proteins. The first conclusion drawn from this work is that tomato ASR proteins cluster together in the tree. This observation can be explained by a scenario of concerted evolution or birth and death of genes. Secondly, our study showed that *Asr1* is highly conserved at both replacement and synonymous sites within the genus *Lycopersicon*. ASR1 protein sequence conservation might be associated with its multiple functions in different tissues while the low rate of synonymous substitutions suggests that silent variation in *Asr1* is selectively constrained, which is probably related to its high expression levels. Finally, we found that *Asr1* activation under water stress is not conserved between *Lycopersicon* species.
© 2006 Elsevier B.V. All rights reserved.

Keywords: *Lycopersicon*; Water stress; Gene conversion; Codon usage; Expression pattern

1. Introduction

Gene families are common in eukaryotic genomes. Divergence in function or expression pattern among family members has given plasticity and innovation to species evolution (Ohta, 1989). Cases of variable family size observed between species are explained by the birth-and-death model of molecular evolution, which posits that repeated duplication and subsequent loss or inactivation of duplicated genes determine a variable number of members (Quesada et al., 2005). In addition, gene family members can experience concerted evolution, with the consequent intraspecific gene copy homogeneization (Nei and Rooney, 2005). Comparisons within *Arabidopsis* show that most paralogous genes accumulate high levels of amino acid sequence divergence without appreciable differences in expression patterns (Hughes and Friedman, 2005).

Plants are sessile organisms that cope with abiotic stress. Water deficit is one of the most studied stresses, as it severely limits crop productivity and plant natural distribution. For this reason, efforts have been focused on understanding plant stress physiology. In this direction, several drought-responsive genes have been cloned and characterized. It was shown that some of them encode components of signal transduction pathways involved in

Abbreviations: ABA, abscicic acid; CDS, coding sequence; GC, guanosine+cytidine; Ka, replacement substitution rate; Ks, synonymous substitution rate; Myr, million years; NJ, neighbor-joining; TC, tentative consensus assembly of ESTs.

* Corresponding author. Laboratorio de Fisiología y Biología Molecular, Facultad de Ciencias Exactas y Naturales, Pab. II, Ciudad Universitaria, C1428EHA, Buenos Aires, Argentina. Tel.: +54 11 4576 3386; fax: +54 11 4576 3321.
E-mail address: norbius@fbmc.fcen.uba.ar (N.D. Iusem).

[1] partner group of the Max Planck Institute of Molecular Plant Physiology (Golm, Germany).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.05.010

adaptation to water stress. *Asr* genes (named after *A*bscicic acid, *s*tress, *r*ipening), as likely participants of the water stress transcriptional response, can be grouped among them. The expression of these genes was also detected during fruit ripening in several plant species (Maskin et al., 2001). *Asrs* are found in Spermatophyta forming small gene families. However, the model plant *Arabidopsis* lacks *Asr* genes and in grape, only one copy has been detected using Southern Blot hybridization (Cakir et al., 2003). In tomato (genus *Lycopersicon*), ASR1 protein is localized in the nucleus and binds a specific DNA sequence (Kalifa et al., 2004). Moreover, ASR proteins from other species are also localized in the nucleus where they regulate specific promoters (Wang et al., 2005; Cakir et al., 2003; Yang et al., 2005). These findings suggest that ASR proteins are transcription factors whose likely targets are hexose transporters and abscicic acid (ABA) responsive genes. Since many *Asr* genes are themselves induced by ABA and regulate genes involved in sugar transport and ABA response, a cross-talk between plant hormones and sugars responding to water stress can be speculated. In fact, ABA and sugars act in a concerted fashion during developmental processes in plants (Finkelstein and Gibson, 2002).

In a previous work (Frankel et al., 2003), we examined the evolution of *Asr2* –originally cloned from a cultivar of commercial tomato (Rossi and Iusem, 1994)– in tomato wild species that inhabit areas with different levels of water availability. We found that this gene was a target of positive selection in tomato species that live in dry areas. In this work, we wanted to gain insight into the evolution of the *Asr* gene family in plants. First, we analyzed a novel tomato *Asr* gene, named *Asr4*. This gene bears an extra region (absent in the other members of the family) encoding a domain with a repeated motif. This new tomato *Asr* was also detected by Dóczi et al. (2005), who reported its expression in water stressed leaf. We also built a phylogenetic tree of ASR proteins of higher plants and compared ASR families from tomato and rice. With this information we hypothesize on the evolution of the gene family. In addition, we expanded our previous study for *Asr2* and examined the forces governing the evolution of the four tomato *Asrs* by comparing synonymous and replacement rates within *Lycopersicon*. Finally, given the high degree of CDS conservation in *Asr1* we decided to investigate if this fact correlated with a conserved expression pattern in *Lycopersicon*. To this purpose, we analyzed the expression of *Asr1* and *Asr4* in water stressed leaf in seven tomato wild species from dry and humid climates.

2. Materials and methods

2.1. Plant material

Seeds from *L. chilense* (accession no. LA2884), *L. esculentum v. cerasiforme* (accession no. LA1228), *L. hirsutum* (accession no. LA1223), *L. peruvianum* v. humifusum (accession no. LA0385), *L. peruvianum* f. glandulosum (accession no. LA1292), *L. cheesmanii* (accession no. LA0521) and *L. esculentum* cv. M-82 were kindly provided by the Tomato Genetics Resource Center (University of California, Davis). Plants were grown in soil pots in a growth chamber (250 μmol photons m^{-2} s^{-1} at 22 °C under a 16 h light/8 h dark regime).

2.2. Sequence and phylogenetic analysis

Sequences were aligned using CLUSTALX and the resulting alignments were visually inspected and manually edited. Identity between ASR tomato proteins was calculated with BioEdit (Hall, 1999). ASR protein sequences were retrieved from GenBank and Plant Unigene Indices (http://www.tigr.org/tdb/tgi/plant.shtml) using BLASTP and tBLASTn with tomato ASR1 as query. For Fig. 2A we selected all species with more than one *Asr* gene and representatives of different plant groups with only one known gene. The accession numbers for the corresponding tentative consensus (TC) and proteins are: ASR1, 2282019 and TC162125; ASR2, 584787 and TC165353; ASR3, 400471 (incomplete) and TC158596 (complete); ASR4, DQ058762 and TC124842; CI21A, 4098248; CI21B, 4098250; DS2, 23095773; RiceASR1, 50252229; RiceASR2, 15289937; RiceASR3, 20146221; RiceASR4, 20146222; RiceASR5, 2773154; RiceASR6, 38605916; MaizeASR1, TC262475; MaizeASR3, TC253974; MaizeASR5, TC258649; MaizeASR6, TC258651; MaizeASR2, TC250893; MaizeASR4, TC258618; LP3-0, 1519370; LP3-1, 1297089; LP3-2, 1401226; LP3-3, 1401234; Pummello, 624672; Lily, 6525055; Ice plant, 7484607; Apricot, 2677824; Grape, 14582465; Ginkgo, 38532363; Banana, 47575681and Pepper ASR1, AY496130. *Asr1*, *Asr3* and *Asr4* sequences obtained in this work were deposited in GenBank database with nos. DQ058744 through DQ058762. *Asr2* GenBank nos. are AY217009 through AY217014 and L20756. The tree of ASR proteins was constructed with MEGA 3.1 freeware (Kumar et al., 2001) using the Neighbor-joining algorithm with pairwise deletion option and the *p*-distances. The bootstrap test included 500 pseudoreplicates. A Poisson correction of the distances gave an identical tree with bootstrap support (>70%) to the same groups. Rice genome information was taken from NCBI (http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html).

Replacement vs. synonymous substitution rates (Ka/Ks=ω) were also calculated with MEGA 3.1 using the Nei–Gojobori algorithm (Nei and Gojobori, 1986). Alignment gaps were not considered for calculations. PAML (Yang, 1997) was used to estimate the number of synonymous and non-synonymous substitutions per branch and the total number of synonymous and replacement sites under a free-ratio model (Yang, 1998). This model assumes different Ka/Ks (ω) per branch. Codon equilibrium frequencies were calculated from average nucleotide frequencies at each of the three codon positions and the transition/transversion ratio (κ) was estimated from the data. In Table 1 the number of synonymous and non-synonymous substitutions (summed only for *Lycopersicon* branches or the branches in the whole tree, which includes *Solanum*) were normalized by the

Table 1
Synonymous (s) and non-synonymous substitutions (n) in *Asr* genes within the *Lycopersicon*/*Solanum* clade

	Asr1	*Asr2*	*Asr3*	*Asr4*
s (*Lycopersicon*/whole tree)	0.060/0.140	0.126/0.383	0.376/?	0.172/0.317
n (*Lycopersicon*/whole tree)	0.000/0.016	0.033/0.045	0.018/?	0.023/0.043

Note — in each cell: on the left, the normalized changes after tomato cladogenesis, on the right, normalized changes on the whole tree.

number of synonymous and non-synonymous sites in each gene. The number of synonymous/non-synonymous sites in each gene are: *Asr1*, 67.5/256.5; *Asr2*, 68/247; *Asr3*, 42.8/281.2; *Asr4*, 84.8/368.2. Differences between genes in the number of substitutions (synonymous or non-synonymous) were statistically tested with a 2×2 contingency table and a Fisher exact test. The accepted phylogeny of *Lycopersicon* (obtained with ribosomal ITS sequences) used in this work is described in Frankel et al. (2003).

Codon usage and GC content were calculated with DAMBE (Xia and Xie, 2001). Gene conversion statistical analysis was performed with Geneconv using default parameters (Sawyer, 1989). This software analyzes whether regions of a pair of sequences have more common identical silent polymorphic sites than would be expected by chance. *P*-values were Bonferroni corrected for multiple testing.

2.3. DNA amplification and sequencing

Genomic DNA extractions were performed as previously indicated (Frankel et al., 2003). The following primers were used in PCR reactions to amplify the coding sequence of *Asr1*, *Asr3* and *Asr4*. For *Asr1*, upper primer: 5′-GATAGATTTATTGTTTCAGATGGAG-3′, lower primer: 5′-CCATATTTGGTATAAGTGTTGTGTC-3′. For *Asr3*, upper primer: 5′-CAAAGCATAAATTGTCTATCGACGT-3′, lower primer: 5′-GTCCATAGATGGGTGAC-3′. For *Asr4*, upper primer: 5′-GTCACCAAAAAACATCATGGC-3′, lower primer: 5′-AGGATAGATTGATCAGCACAC-3′.

The primers used for sequencing *Asr1* were the following: 5′-CCTCTTTTGTTGGTGATAAT-3′ and 5′-AACAACACTTATTTTGAAACG-3′. The primers used for sequencing *Asr3* were the same as those used for PCR. The primer used for partial sequencing of *Asr4* was 5′-CGGAGGAGATAATACATATGG-3′. *Asr1* from *L. chilense* and *L. peruvianum* f. glandulosum was amplified and sequenced with a different set of primers from cDNA of water stressed leaf; upper: 5′-CAGATGGAGGAGGAG-3′, lower: 5′-TAGAAGAGATGGTGGTGTCCC-3′.

Thirty-five cycles consisting of denaturation (95 °C for 1 min), annealing (53 °C for 1 min) and extension (72 °C for 1 min) were programmed in a PTC-100 thermocycler (M.J. Research). The amplification products were run on 1% LMP agarose gels. Fragments were excised from the gel and purified using the Concert Kit (Gibco). The PCR products were sequenced at the Biotechnology Resource Center (Cornell University) with an ABI 3700 sequencer. Sequences were edited using BioEdit (Hall, 1999).

2.4. Expression analysis

Plants were grown on MS medium and after 15 days transferred to soil pots and maintained in a phytotron with controlled light, humidity and temperature. At 45 days, the plants were removed from the pots and let dry during 6 or 24 h on the laboratory bench. Leaves were then detached, immediately immersed in liquid nitrogen and stored at −80 °C until RNA extraction. Controls were obtained from unstressed plants. To confirm that the severity of the stress was similar to all the species, we measured the relative water content of the leaf. The values at 6 and 24 h did not differ significantly between species (data not shown).

Total RNA was obtained from leaves according to the Trizol procedure (Gibco BRL). RNAs from three different plants were pooled for the assay. Northern Blots were performed following standard procedures. The use of specific probes that cover mostly non-coding regions combined to stringent hybridization conditions (65 °C, 0.1× SSC, 0.1% SDS) avoided cross detection of other family members. Probes were obtained through amplification of cDNA (Prime-a-Gene Labeling System, Promega) from cultivated tomato with the following primers: For *Asr1*, upper: 5′-TCGATAGATTTATTGTTTCAGATGGAG-3′, lower: 5′-GACACAACACTTATACCAAATATGG-3′. For *Asr4*, upper: 5′-GCATTCCATGAACATCATCAG-3′, lower: 5′-AGGATAGATTGATCAGCACAC-3′.

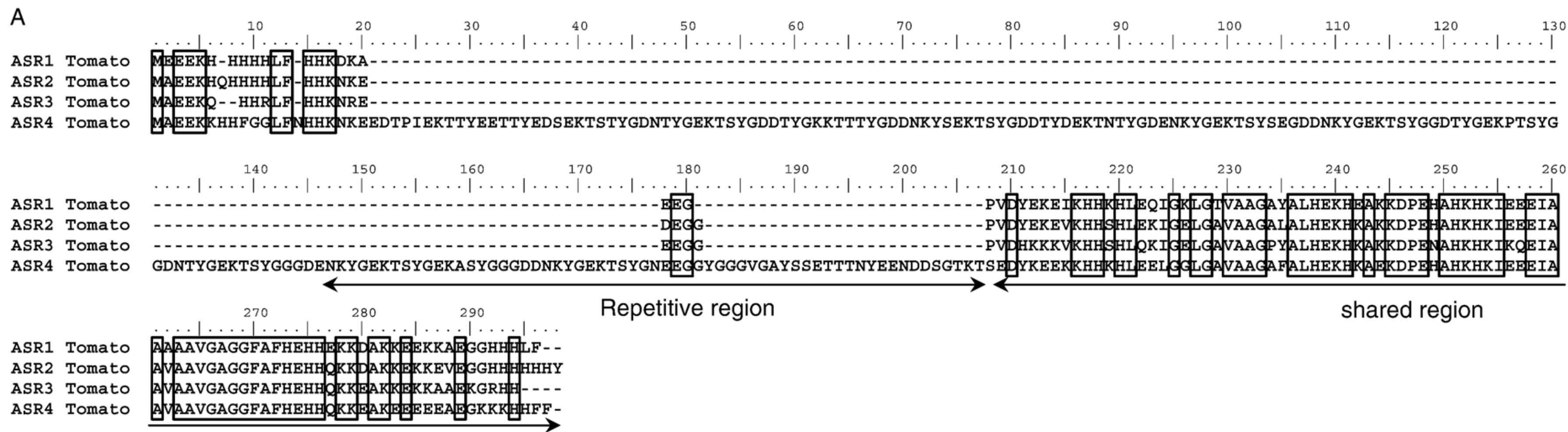
3. Results

3.1. A new member of the Asr gene family in tomato

Up to now, three members of the tomato *Asr* gene family were known (Rossi et al., 1996). Searching the tomato EST database (TIGR, www.tigr.org) for eventual novel members, we found the TC #124842 (defined as Tentative Consensus sequence, created by assembling ESTs into contigs), which exhibited sequence identity to known *Asr* genes. In particular, this contig happened to show high similarity to DS2 gene from potato – genus *Solanum* – (GenBank accession no. CAC86102) and the seashore plant *Calystegia soldanella* (GenBank accession no. BAB19963). We were first puzzled because this TC had STOP codons in the putative coding sequence, what might represent a pseudogene mark. To solve this question, we designed PCR primers to amplify the presumptive coding sequence from cDNA and genomic DNA. From sequence analysis, we reasoned that this TC corresponds to an unspliced transcript and named this gene *Asr4*. When the manuscript of this work was in preparation, Dóczi et al. (2005) reported that the same tomato TC is expressed in water stressed leaf.

The alignment of the four ASR proteins from tomato shows a high number of totally conserved positions (Fig. 1A). In addition, ASR4 protein reveals an unshared domain of 186 amino acids that includes ten imperfect 13-amino acid repeats composed of the consensus sequence YGEKTSYGG(G/D)D(T/N)K. The pairwise percentage identity of ASR proteins is shown in a table under the alignment (Fig. 1B; only the common regions were included in the calculations).

In order to determine if tomato *Asr4* maps close to the other *Asr* genes in chromosome IV (Rossi et al., 1996), we used a line of *L. esculentum* with part of its genome replaced with homologous genomic regions from *L. pennellii*. In particular, we used line IL 4-3-2, which bears a single *pennellii* introgression in a zone of chromosome 4. This 19 cM region (flanked by markers TG182 and CD55; Eshed et al., 1992,http://www.sgn.cornell.edu/maps/pennellii_il/pennelli_i_il_map.html), includes the tightly linked *Asr1*–*Asr2*–*Asr3* (Rossi et al., 1996). For mapping purposes, we sequenced *Asr4* from the two parental species and the



B

Protein	ASR1	ASR2	ASR3	ASR4
ASR1	--	82%	72%	71%
ASR2		--	79%	75%
ASR3			--	67%
ASR4				--

Fig. 1. A) Alignment of the *Lycopersicon esculentum* ASR family. Boxes show totally conserved sites. The two ASR4 regions sequenced from wild species are underlined (with arrowheads at their ends). B) Protein identity between the four members of the ASR family for the common region.

introgressed line, discovering that *Asr4* sequences from IL 4–3–2 and *pennellii* are identical (data not shown). Therefore, *Asr4* maps approximately to the same location of the other members of the family, within a segment not larger than 19 cM.

The high sequence conservation among tomato ASRs prompted us to seek for homologues in plants other than tomato and to inspect the evolution of these genes. Thus, we searched public databases for plant ASRs.

3.2. Asr genes in higher plants

We found ASR proteins in a wide taxonomic range within Spermatophyta. With ASR sequences, we performed a phylogenetic analysis with the aim of understanding the history of the family. Fig. 2A depicts a NJ tree of ASR proteins (rooted with the gymnosperms Ginkgo/Pine) from families and representatives of different groups of Spermatophyta. The tree allows to observe some orthology relationships between potato and tomato genes, namely *Asr1*/*ci21A*, *Asr2*/*ci21B* and *Asr4*/*DS2*. The absence of an *Asr3*-like transcript from the large potato EST databases together with our fruitless attempts to amplify this gene in potato suggest that this member is unique to tomato. A close inspection of the tree (Fig. 2A) allows to envisage some orthology relationships between maize and rice genes.

The topology of the tree shows a pattern of intra-group identity, as ASR proteins from tomato/potato and pine form different clusters. This observation is further corroborated by the calculations of protein identity between and within groups. Within tomato, ASR pairwise identities range from 70 to 80% (Fig. 1B) while within pine identities run from 60 to 70% (not shown). However the identities drop to about 50% in comparisons between pine and tomato (not shown). Some of the maize/rice genes also group together, but this cluster is not supported by bootstrap. As can be seen, ASR2 and ASR6 from maize/rice are highly divergent in sequence. To gain insight into the evolutionary history of the family we made use of the genomic information of rice.

3.3. Asr genes in rice and tomato

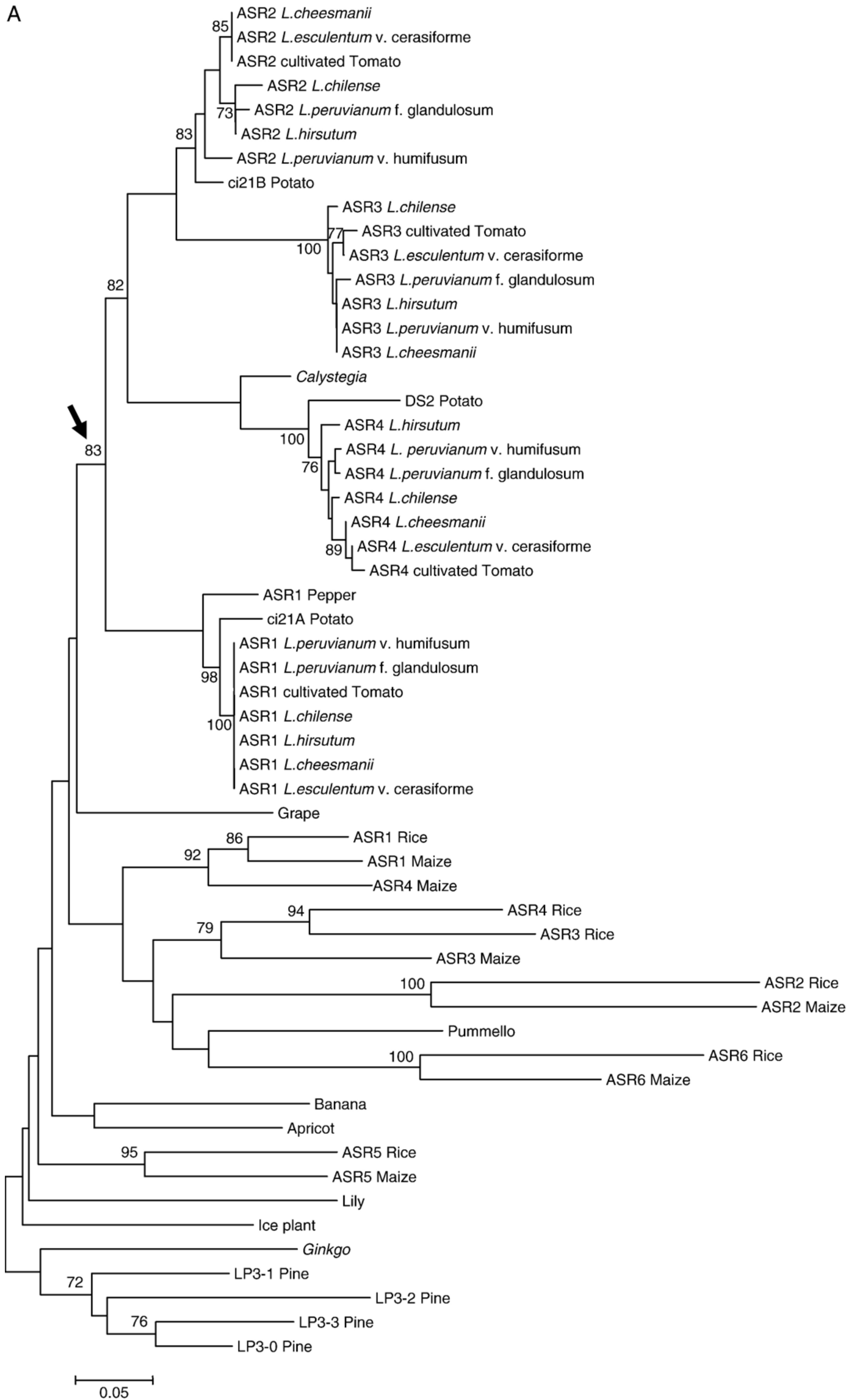
By an *in silico* screening of the rice genome, we had found six *Asr* genes. In the tomato genome, the total number of *Asr*s is likely to be definitive, as the large EST database of this plant (www.tigr.org) includes just four different *Asr* transcripts. Fig. 2B gives a comparison between tomato and rice *Asr* genes. As mentioned earlier, all four tomato *Asr* genes are located close to each other in chromosome IV while rice *Asr* genes are scattered on different chromosomes. In addition, all rice and tomato genes have associated ESTs that confirm their expression. Rice *Asr1* has four alternatively spliced mRNA isoforms that give rise to four different proteins. Rice *Asr3* and *Asr4* are located in tandem and have the same intron length (which probably represents a recent gene duplication) while rice *Asr2* is more than 200 kb away from the pair *Asr3*–*Asr4* in chromosome I. Notably, Rice ASR6 is highly divergent from other ASRs (Fig. 2A) but shares with tomato ASR4 the position of an N-terminal insertion that makes a larger protein (Fig. 2B). The alignment of these two insertions shows traces of identity (Fig. 2D). Moreover, when the C-termini of tomato ASR4 and rice ASR6 are carefully inspected substantial identity emerges between these two proteins (Fig. 2C). This identity fades not far from the C-termini and tomato ASR4 becomes more similar to the other *Lycopersicon* members (in Fig. 2C only tomato ASR2 is shown for simplicity). Indeed, when we apply a statistical test for gene conversion (Sawyer, 1989) among tomato *Asr*s, an identical fragment of 62-bp between *Asr4* and *Asr3* yields a significant value (corrected $P = 0.012$).

3.4. Evolution of the Asr gene family in the genus Lycopersicon

In a previous article (Frankel et al., 2003), we examined the CDS of *Asr2* in seven wild tomato species that inhabit areas with different levels of water availability. To extend the analysis to the whole family, we sequenced *Asr1*, *Asr3* and part of *Asr4* (see Fig. 1A for details) in these taxa of *Lycopersicon.* Protein sequences of these genes were used to construct the tree of Fig. 2A. With the nucleotide sequences, we mapped, for each gene, the number of replacement and synonymous changes per branch in the accepted *Lycopersicon*/*Solanum* phylogeny (see Materials and methods section and Supplementary material for details). A relevant conclusion emerging from these analyses is that the only member of the family that shows signs of positive selection is *Asr2* (Supplementary material). In Table 1 we present the synonymous and replacement changes per site for each gene (see Materials and methods section for details). A remarkable fact is the lack of replacement substitutions in *Asr1* after *Lycopersicon* cladogenesis (Table 1). This rate in *Asr1* is significantly lower (Fisher exact test, $P < 0.05$) compared to any of the rates in *Asr2*, *Asr3* and *Asr4*; whereas the rates in these latter genes are not different from each other. Despite this extreme conservation in *Lycopersicon*, there are four amino acid differences between ASR1s from *Lycopersicon* and *Solanum*. In order to determine whether amino acid changes occurred in the tomato lineage (prior to its speciation) or in the potato lineage, we retrieved ASR1 sequence from *Capsicum annum* (pepper), a close outgroup within the Solanaceae. Comparison of ASR1 sequence between these species (data not shown) indicates that tomato amino acids are identical to the ones in pepper when looking at diverged sites between potato and tomato, suggesting that the amino acid changes took place in the potato lineage. Hence, ASR1 remained unchanged in *Lycopersicon* since the divergence of tomato and potato 10 Myr ago.

Another remarkable fact is the low number of synonymous substitutions on *Asr1* as compared to the other members (Table 1). The number of synonymous changes (considering all branches in the tree) normalized by the total number of synonymous sites are 0.140, 0.383 and 0.317 for *Asr1*, *Asr2* and *Asr4*, respectively. *Asr1* rate is significantly lower (Fisher exact test, $P < 0.01$) compared to *Asr2* or *Asr4,* while rates of *Asr2* and *Asr4* do not differ from each other (*Asr3* is not taken into account here because there is no *Solanum* sequence). The number of synonymous substitutions per site within *Lycopersicon* are 0.060, 0.126, 0.376 and 0.172 for *Asr1*, *Asr2*, *Asr3* and *Asr4*, respectively (Table 1).

The observation of different synonymous substitution rates among *Asr* genes is puzzling. Rates should be fairly similar across genes under the assumption of selective neutrality. Nevertheless,


A
ASR2 L.cheesmanii
ASR2 L.esculentum v. cerasiforme
ASR2 cultivated Tomato
ASR2 L.chilense
ASR2 L.peruvianum f. glandulosum
ASR2 L.hirsutum
ASR2 L.peruvianum v. humifusum
ci21B Potato
ASR3 L.chilense
ASR3 cultivated Tomato
ASR3 L.esculentum v. cerasiforme
ASR3 L.peruvianum f. glandulosum
ASR3 L.hirsutum
ASR3 L.peruvianum v. humifusum
ASR3 L.cheesmanii
Calystegia
DS2 Potato
ASR4 L.hirsutum
ASR4 L. peruvianum v. humifusum
ASR4 L.peruvianum f. glandulosum
ASR4 L.chilense
ASR4 L.cheesmanii
ASR4 L.esculentum v. cerasiforme
ASR4 cultivated Tomato
ASR1 Pepper
ci21A Potato
ASR1 L.peruvianum v. humifusum
ASR1 L.peruvianum f. glandulosum
ASR1 cultivated Tomato
ASR1 L.chilense
ASR1 L.hirsutum
ASR1 L.cheesmanii
ASR1 L.esculentum v. cerasiforme
Grape
ASR1 Rice
ASR1 Maize
ASR4 Maize
ASR4 Rice
ASR3 Rice
ASR3 Maize
ASR2 Rice
ASR2 Maize
Pummello
ASR6 Rice
ASR6 Maize
Banana
Apricot
ASR5 Rice
ASR5 Maize
Lily
Ice plant
Ginkgo
LP3-1 Pine
LP3-2 Pine
LP3-3 Pine
LP3-0 Pine
85
73
83
100
77
82
83
100
76
89
98
100
86
92
94
79
100
100
95
72
76
0.05

B

	chromosome	intron size (bp)	protein size (aa)	EST	in tandem with
Rice					
Asr1	II	alternative splicing	63/71/91/105	yes	--
Asr2	I	440	182	yes	--
Asr3	I	131	105	yes	4
Asr4	I	131	96	yes	3
Asr5	XI	119	138	yes	--
Asr6	IV	84	229	yes	--
Tomato					
Asr1	IV	687	110	yes	?
Asr2	IV	111	114	yes	?
Asr3	IV	476	108	yes	?
Asr4	IV	111	297	yes	?

C

```
Rice ASR6    (aa 153-229) : MYERHQAKKDPENAQRHRIEEGVAAAAALGSGGFAFHEHHDKKEAKQAAKDAEEEAEEESGSGARGGEGKKKHHLFG
Tomato ASR4  (aa 237-297) : LHEKHKAEKDPEHAHKHKIEEEIAAVAAVGAGGFAFHEHHQKKEAK------EEE-EE-----AEG---KKKHHFF-
Tomato ASR2   (aa 54-114) : LHEKHKAKKDPEHAHKHKIEEEIAAVAAVGAGGFAFHEHHQKKDAK------KEK-KE-----VEGG--HHHHHHY-
```

D

```
Rice ASR6    (aa 20-134): -------YESGGYN---RSSSGGADEYAAGRS-GR---AQKPVX--DASKRFTK-SR------RRATTYGXRRR---RV-------NK
Tomato ASR4 (aa 22-209): TPIEKTTYEETTYEDSEKTSTYGDNTYGEKTSYGDDTYGKKTTTYGDDNKYSEKTSYGDDTYDEKTNTYGDENKYGEKTSYSEGDDNK

Rice ASR6    : SGPRAS---DS--------GXNN----RS--GA---NRSTAATSPARRVQRRXG-----AE----ADEE--Y---VDGLSSRAPGEVQEG---G----E
Tomato ASR4  : YGEKTSYGGDTYGEKPTSYGGDNTYGEKTSYGGGDENKYGEKTSYGEKASYGGGDDNKYGEKTSYGNEEGGYGGGVGAYSSETTTNYEENDDSGTKTSE
```

Fig. 2. A) Neighbor-joining tree of ASR proteins from different plant species. The numbers in nodes are bootstrap values (only values >70% are shown). Bar length indicates *p*-distance. The arrow points at the tomato/potato group. The tree is rooted with the Ginkgo-Pine group (gymnosperms). B) Comparison of several features between the *Asr* family from rice (monocot) and tomato (dicot). C) Alignment of the C-termini of tomato ASR4, rice ASR6 and tomato ASR2. The coloring highlights positions where two proteins have identical aminoacids (grey) and the other protein has a different one (black). D) Alignment of the N-terminal insertions of tomato ASR4 and rice ASR6. Black boxes indicate identical amino-acids.

Asr1 showed a considerably lower rate of synonymous substitutions, suggesting that synonymous sites in this gene may be more constrained than in the other members of the family. Therefore, we decided to analyze the patterns of codon usage in the tomato *Asr* gene family.

3.5. Codon usage in the tomato Asr family

The GC content at third codon positions [(GC_3)] is a simple and commonly used measure of codon usage, insensitive to rare amino acids, reliable in the case of small genes (as in the present case) and fairly correlated to other measures of codon usage (Chiapello et al., 1998). Hence, we calculated the GC content at each codon position for all genes (Table 2). *Asr1* turned out to have the most divergent $(GC)_3$ content (50%) among *Asr* genes. *Asr2*, *Asr3* and the common part of *Asr4* exhibited a lower $(GC)_3$ content; close to the average 38% obtained for 1173 CDSs involving 510958 codons in *L. esculentum* (Codon Usage Database, www.kasuza.or.jp/codon) and the 42.3% calculated for *Arabidopsis thaliana* (this percentage is based on 70,492 CDSs, 27,825,665 codons). Glu, Gly, Ala, Lys and His are the most abundant residues in ASRs, together accounting for more than 75% of the protein. Glu, Gly, Ala and His codons exhibit the highest $(GC)_3$ in *Asr1* (see Supplementary material). Surprisingly, *Asr1* has the lowest $(GC)_3$ for Lys. When we re-calculate $(GC)_3$ content in codons for Glu, Gly, Ala and His, the difference between *Asr1* (55%) and the other *Asr*s (34% on average) or other tomato genes (32% for 1173 CDS) is even greater (Table 1). The high $(GC)_3$ in *Asr1* is not a matter of local nucleotide composition since all introns in the four *Asrs* have essentially the same GC content (Table 2).

Given the high degree of CDS conservation and differential codon usage in *Asr1*, we decided to investigate if this correlated with a conserved expression pattern in *Lycopersicon*. Similar expression patterns might also indicate conservation in the promoter *cis*-acting elements and *trans*-acting factors.

3.6. Asr expression analysis in the genus Lycopersicon

As *Asr1* and *Asr4* showed high levels of induction upon water stress in leaf (Maskin et al., 2001; Dóczi et al., 2005), we chose this treatment to compare their expression in the genus *Lycopersicon*. We analyzed mRNA levels at 0, 6 and 24 h of water stress (see Section 2.4 for details). Northern Blot analysis revealed that *Asr1* and *Asr4* mRNAs are detected at 24 h in

Table 2
Percent GC at the different codon positions and intron for the Asr *L. esculentum* genes

	1st	2nd	3rd	3rd E+A+H+G [a]	Intron
Asr1	71	25	50	56	26
Asr2	71	24	44	37	23
Asr3	67	29	38	35	28
Asr4 (minus insertion) [b]	68	25	40	32	28

[a] "E+A+H+G" refers to the subset of codons coding for glutamate, alanine, histidine and glycine.
[b] "Minus insertion" remarks that the N-terminal insertion in *Asr4* was not included for calculations.

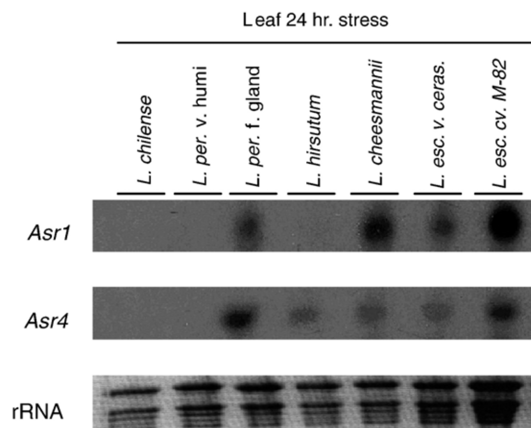


Fig. 3. *Asr1* and *Asr4* leaf mRNA levels after 24 hr-water stress in *Lycopersicon* species. Both *Lycopersicon chilense* and *Lycopersicon peruvianum* v. humifusum inhabit dry areas.

species from humid areas (except for *Asr1* in *L. hirsutum*) but not in *L. chilense* and *L. peruvianum* v. humifusum, which inhabit dry environments (Fig. 3). At 0 and 6 h of stress, both mRNAs were undetectable (data not shown). We also tried to measure *Asr2* and *Asr3*, but these mRNAs are not detectable by Northern Blot in leaves of any species under water stress (data not shown).

4. Discussion

The first conclusion drawn from this work is that some ASR proteins display intra-group identity. Those from tomato and potato (Solanaceae, dicots) form a consistent group, and so do those from the gymnosperm *Pinus taeda*. Such a pattern of intra-group identity could be caused by either concerted evolution or birth and death of genes (considering an ancestral *Asr* family). Another possibility is the occurrence of independent duplications after the divergence of these major lineages (only one ancestral *Asr* gene), or it may well be a combination of processes. In other gene families, their members are already divergent in sequence and function before cladogenesis, leading to a gene tree defining orthology relationships between members from different species (as in the case of the phytochromes; Alba et al., 2000).

In addition, the likely absence of *Asr3* in potato and *Asr4* in tobacco (Dóczi et al., 2005) and a recent duplication in rice (that originated *Asr3* and *Asr4*) strengthen the hypothesis of a historically dynamic gene family. The presence of only one *Asr* in grape (Cakir et al., 2003; Goes da Silva et al., 2005) and the absence of this family in *Arabidopsis* point in the same direction. However, rice does not carry *Asr* pseudogenes in its genome, which would be expected under a birth and death model of evolution. Nevertheless, this should be taken with caution, since residence time of pseudogenes in genomes may be short.

Within tomato, pairwise identities are very similar in different comparisons (Fig. 1B). Moreover, the four genes share many blocks of conserved aminoacids (Fig. 1A) and form a consistent cluster in the NJ tree (Fig. 2A). Despite the unquestionable orthology of tomato *Asr4* and rice *Asr6* (that supports the existence of more than one ancestral *Asr*), the former now looks like the other tomato family members. In fact, a statistical test supports the occurrence of a gene conversion event between *Asr4* and *Asr3*. But we have no additional conclusive evidence on gene conversion. It is likely that the evolution of the family in tomato was affected by more than one process. The biological significance of the observed protein sequence homogeneity in tomato is difficult to hypothesize. Perhaps what gives *Asr* identity in tomato is the divergent expression pattern. However, different selective forces act on tomato *Asr* genes, casting doubt on the presumptive functional homogeneity. Similarly, pine genes form a consistent cluster in the NJ tree and have high identity in pairwise comparisons. Conversely, rice *Asr* genes are more divergent among each other and lie scattered on different chromosomes.

Notably, ASR1 sequence is totally conserved within *Lycopersicon* (Fig. 2A and Table 1); this protein has remained changeless for 10 Myr in different tomato lineages. This surprising sequence conservation in a transcription factor might be associated with its ubiquitous expression and the concomitant existence of several promoter targets. Thus, changes in the protein might disrupt any of its multiple binding activities. *Asr1* is expressed in fruit, flower, leaf, root, stem and seed under normal conditions and behaves like a housekeeping gene. In contrast, the other members do not display such a widespread pattern of expression under normal physiological conditions, being restricted to a specific organ (Maskin et al., 2001; unpublished data; Tomato EST database). A similar situation has been recently reported for the *Skp1* gene family, in which a more widespread pattern of expression is correlated with low non-synonymous substitution rates (Kong et al., 2004).

Asr1 is expressed at high levels in flower (Tomato EST databasehttp://www.tigr.org/tdb/tgi/) and during tomato fruit development (Bovy et al., 2002). In addition, the representation of *Asr1* in all tomato EST libraries is much higher than that of the other three members: 74, 5, 2 and 9 are the number of ESTs reported for *Asr1*, *Asr2*, *Asr3* and *Asr4*, respectively. These ESTs come from several libraries representing expressed genes in different organs (for a complete description of the libraries and *Asr* ESTs, see the Tomato EST databasehttp://www.tigr.org/tdb/tgi). So, is this level of expression of *Asr1* causally related to its low rate of synonymous substitutions? If all synonymous sites are indeed selectively neutral, rates of evolution should be roughly similar among genes (Kimura, 1983). However, our data revealed that the rate of synonymous substitution in *Asr1* is significantly lower when compared to the other members of the family. One possible explanation for the different rates is that synonymous codons in *Asr1* are not selectively equivalent, causing evolution at synonymous sites to be governed by a balance between selection and drift rather than drift alone (Ohta, 1973).

Claims against neutrality of silent changes appeared with the discovery that synonymous codons for a certain amino acid do not occur with equal frequencies (see Akashi and Eyre-Walker (1998) for a review). The idea that synonymous codon choice may affect fitness is supported by the observation that favored codons are those that match the most abundant tRNAs available for each amino acid in highly expressed genes (reviewed in Akashi (2001) and references therein). The relationship between codon usage

bias and level of expression in unicellular organisms and model eukaryotes has been confirmed by laboratory experiments (Akashi, 2001). Thus, codon usage bias in a particular gene may reflect a balance between the strength of selection on translation accuracy/efficiency and mutation. Specifically, Chiapello et al. (1998) have noticed the presence of two groups of genes in the model plant *A. thaliana*. One group contains mainly photosynthetic and housekeeping genes and exhibits a high $(GC)_3$. A second group shows a high $(AT)_3$ and very specific patterns of expression. In tomato, *Asr1* has features of the first group, a high $(GC)_3$ content (compared to the average in tomato) and high expression. The absence of data of relative abundance of tRNAs in tomato precludes any interpretation of differences among genes in $(GC)_3$ content in terms of translational efficiency, since we cannot assert which of the *Asr* genes have the optimal GC content at third codon positions. However, on the basis of present evidence, we hypothesise that differences in $(GC)_3$ content and rates of synonymous substitutions among tomato *Asr* genes may be causally related to differences in expression levels and differential strength of selection for translational accuracy/efficiency.

Our expression data on *Asr1* and *Asr4* in leaf from *L. esculentum* are in agreement with previous reports by Maskin et al. (2001) and Dóczi et al. (2005), respectively. *Asr1* coding sequence is highly conserved in all *Lycopersicon* species, but its induction upon water stress is not phylogenetically conserved. This difference might be caused by changes in the *Asr1* promoter region(s) responsive to water stress signals or in the trans-factors that could activate *Asr1*. Remarkably, *Asr1* expression is not observed in species that inhabit dry areas. The similar trend observed for *Asr4* makes us think that there might be a different stress response in species inhabiting arid habitats.

Asr4 deserves a separate comment. The biochemical function of the encoded extra domain is intriguing as *in silico* searching of data bases for consensus motifs did not render any conclusive clue. This domain contains the motif Ser–Tyr–Gly within a repeat of 13 amino acids. This motif is also present several times in proteins responding to stress in *S. cerevisiae* (Treger and McEntee, 1990). Likewise, this tripeptide moiety is found at least ten times in proteins of unknown function from the fungus *Ashbya gossypii* (GenBank accession no.NM210779) and *Plasmodium falciparum* (GenBank accession no. AAC47854).

Despite the observed sequence homogeneity among tomato *Asr* members, *Asr1* seems to be unique in tomato. Its coding sequence is remarkably conserved, a situation likely caused by its ubiquitous and high expression. Its multiple targets and the need for an efficient and accurate translation may be the constraints that determine the slow evolution of this gene. In contrast, the other family members might be more free to change, a situation that presumably conferred *Asr2* the potential to evolve under positive selection during adaptation of plants to dry conditions.

Acknowledgments

We thank the Tomato Genetics Resource Center for providing the seeds of the tomato wild populations. This work was supported by grants from the Universidad de Buenos Aires. N.F. holds a fellowship from CONICET, Argentina. F.C., E.H. and N.D.I. are members of the Carrera del Investigador Científico, CONICET, Argentina.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2006.05.010.

References

Akashi, H., 2001. Gene expression and molecular evolution. Curr. Opin. Genet. Dev. 11, 660–666.

Akashi, H., Eyre-Walker, A., 1998. Translational selection and molecular evolution. Curr. Opin. Genet. Dev. 8, 688–693.

Alba, R., Kelmenson, P.M., Cordonnier-Pratt, M.M., Pratt, L.H., 2000. The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms. Mol. Biol. Evol. 17, 362–373.

Bovy, A., et al., 2002. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. Plant Cell 14, 2509–2526.

Cakir, B., Agasse, A., Gaillard, C., Saumonneau, A., Delrot, S., Atanassova, R., 2003. A grape ASR protein involved in sugar and abscisic acid signaling. Plant Cell 15, 2165–2180.

Chiapello, H., Lisacek, F., Caboche, M., Henaut, A., 1998. Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. Gene 209, GC1–GC38.

Dóczi, R., Kondrak, M., Kovacs, G., Beczner, F., Banfalvi, Z., 2005. Conservation of the drought-inducible DS2 genes and divergences from their ASR paralogues in solanaceous species. Plant Physiol. Biochem. 43, 269–276.

Eshed, Y., Abu-Abied, M., Saranga, Y., Zamir, D., 1992. *Lycopersicon esculentum* lines containing small overlapping introgressions from *L. pennellii*. Theor. Appl. Genet. 83, 1027–1034.

Finkelstein, R.R., Gibson, S.I., 2002. ABA and sugar interactions regulating development: cross-talk or voices in a crowd? Curr. Opin. Plant Biol. 5, 26–32.

Frankel, N., Hasson, E., Iusem, N.D., Rossi, M.S., 2003. Adaptive evolution of water-stress induced gene *Asr2* in *Lycopersicon* species dwelling in arid habitats. Mol. Biol. Evol. 20, 1955–1962.

Goes da Silva, F., et al., 2005. Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple vitis species and development of a compendium of gene expression during berry development. Plant Physiol. 139, 574–597.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hughes, A.L., Friedman, R., 2005. Expression patterns of duplicate genes in the developing root in *Arabidopsis thaliana*. J. Mol. Evol. 60, 247–256.

Kalifa, Y., Gilad, A., Konrad, Z., Zaccai, M., Scolnik, P.A., Bar-Zvi, D., 2004. The water- and salt-stress-regulated Asr1 (abscisic acid stress ripening) gene encodes a zinc-dependent DNA-binding protein. Biochem. J. 381, 373–378.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Kong, H., Leebens-Mack, J., Ni, W., dePamphilis, C.W., Ma, H., 2004. Highly heterogeneous rates of evolution in the SKP1 gene family in plants and animals: functional and evolutionary implications. Mol. Biol. Evol. 21, 117–128.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Arizona State University, Tempe, Arizona.

Maskin, L., et al., 2001. Differential expression of the members of the *Asr* gene family in tomato (*Lycopersicon esculentum*). Plant Sci. 161, 739–746.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nei, M., Rooney, A.P., 2005. Concerted and Birth-and-Death Evolution of Multigene Families. Annu. Rev. Genet. 39, 121–152.

Ohta, T., 1973. Slightly deleterious substitutions in evolution. Nature 246, 96–98.

Ohta, T., 1989. Role of gene duplication in evolution. Genome 31, 304–310.

Quesada, H., Ramos-Onsins, S.E., Aguade, M., 2005. Birth-and-death evolution of the Cecropin multigene family in *Drosophila*. J. Mol. Evol. 60, 1–11.
Rossi, M.M., Iusem, N.D., 1994. Tomato genomic clone homologous to a gene encoding an ABA-induced protein. Plant Physiol. 104, 1073–1074.
Rossi, M.M., Lijavetzky, D., Bernacchi, D., Hopp, H.E., Iusem, N.D., 1996. *Asr* genes belong to a tomato gene family of at least three closely linked loci located to chromosome 4. Mol. Gen. Genet. 252, 489–492.
Sawyer, S.A., 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526–538.
Treger, J.M., McEntee, K., 1990. Structure of the DNA damage-inducible gene DDR48 and evidence for its role in mutagenesis in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 10, 3174–3184.
Wang, H.J., Hsu, C.M., Guang, Y.J., Wang, C.S., 2005. A lily pollen ASR protein localizes to both cytoplasm and nuclei requiring a nuclear localization signal. Physiol. Plant. 123, 314–320.
Xia, X., Xie, Z., 2001. DAMBE: data analysis in molecular biology and evolution. J. Heredity 92, 371–373.
Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.
Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15, 568–573.
Yang, C.Y., Chen, Y.C., Jauh, G.Y., Wang, C.S., 2005. A Lily ASR Protein Involves Abscisic Acid Signaling and Confers Drought and Salt Resistance in Arabidopsis. Plant Physiol. 139, 836–846.