# Adaptive Evolution of the Water Stress-Induced Gene *Asr2* in *Lycopersicon* Species Dwelling in Arid Habitats

*Nicolás Frankel,* Esteban Hasson,† Norberto D. Iusem,*[1] and María Susana Rossi*[1]*

*Laboratorio de Fisiología y Biología Molecular, Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, and †Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

The *Asr2* gene encodes a putative transcription factor that is up-regulated in leaves and roots of tomato plants exposed to water-deficit stress. This gene was first cloned and characterized in a cultivar of commercial tomato (*Lycopersicon esculentum* cv. Ailsa Craig). In this work, we report the complete coding sequences of the orthologous *Asr2* genes in six wild tomato lineages: *L. hirsutum*, *L. cheesmanii*, *L. esculentum* v. cerasiforme, *L. chilense*, *L. peruvianum* v. humifusum and *L. peruvianum* f. glandulosum. Estimates of the $K_a/K_s$ ratio (ω) in pairwise comparisons within the genus *Lycopersicon* were equal or greater than 1 (a signature of adaptive evolution) when involving *L. chilense* and *L. peruvianum* v. humifusum. Interestingly, these two species are distinct from the others in their adaptation to dry habitats. We also mapped the detected substitutions onto a phylogenetic tree of the genus *Lycopersicon*. Remarkably, there are two and three amino acid substitutions, which contrast with the absence of synonymous substitutions along the terminal branches leading to *L. chilense* and *L. peruvianum* v. humifusum, respectively. Likelihood ratio tests confirmed that ω values in the branches leading to these species are significantly different from the remaining branches of the tree. Moreover, inferred changes in the branches leading to these species that inhabit dry areas are nonconservative and may be associated with dramatic alterations in ASR2 protein conformation. In this work, we demonstrate accelerated rates of amino acid substitutions in the *Asr2* gene of tomato lineages living in dry habitats, thus giving support to the hypothesis of adaptive Darwinian evolution.

## Introduction

Water-deficit stress in plants is thought to have occurred over 400 MYA in organisms as primitive as bryophytes, identified as closely linked to the origin and evolution of early land plants (Qiu and Palmer 1999). Recently, Oliver, Tuba, and Mishler (2000) proposed that desiccation tolerance, the ability to recover from almost complete loss of protoplasmic water, was primitively present in bryophytes but lost in the evolution of tracheophytes. The phytohormone abscicic acid (ABA) is thought to have played a major role in the induction of the responses to water deficit in all plant lineages.

Nowadays, drought is a major agronomic problem, resulting in reduction in yields of crops exposed to chronic or sporadic periods of drought (Boyer 1982). Therefore, this type of abiotic stress has been the focus of considerable attention, revealing physiological mechanisms of adaptation (Bohnert and Sheveleva 1998). Much research of plant responses to water deficit has been directed towards the isolation of stress-inducible genes by exposing plants to dehydration regimes (Skriver and Mundy 1990; Iusem et al. 1993). The function of only a few such genes is known (Zhu 2002). Genes related to osmotic adjustment, degradation, repair, and structural protection are up-regulated during dehydration (Ingram and Bartels 1996; Shinozaki and Yamaguchi-Shinozaki 1997).

Our interest focuses on genes responsive to limiting water availability and their evolution in lineages exposed to extreme habitats. In this context, the *Asr* gene family is a good working model, as it is up-regulated in leaves and roots of water-stressed plants (Maskin et al. 2001). Since *Asr2* is the best-characterized member of the family, we started this study by examining this gene, originally cloned from a cultivar of commercial tomato (*L. esculentum* cv. Ailsa Craig) (Rossi and Iusem 1994). *Asr2* encodes a putative transcription factor likely to be involved in one of the signaling pathways of ABA (Finkelstein, Gampala, and Rock 2002). The other two members, *Asr1* and *Asr3* share a high sequence identity with *Asr2* (Maskin et al. 2001). In vitro studies showed that ASR1 has a zinc-dependant DNA-binding activity and is localized to both nuclei and cytoplasm (Kalifa and Bar-Zvi, personal communication). On the other hand, an ASR protein from *Vitis vinifera* (grape) was found to be associated with a hexose transporter promoter (Atanassova et al. 2003). This piece of evidence leads to the conclusion that the *Asr* gene family would play a role in transcriptional modulation of one or many genes related to carbohydrate mobilization and/or osmoregulation. The role of hexoses during water stress has been studied (Hare, Cress, and Van Staden 1998), but many questions still remain to be answered.

The genus *Lycopersicon* comprises nine species (Rick 1979) growing in western South America, from Ecuador to northern Chile. These wild tomatoes dwell in a variety of habitats, spanning a wide range of water availability (Rick 1973; Taylor 1986). All species are diploid ($2n = 24$) (Rick 1979). Breeding systems vary from self-incompatible, facultative self-compatible, to entirely self-compatible (Kondo et al. 2002). Tomato plants have a gametophytic type of self-incompatibility controlled by a single multi-allelic "S" locus, which enables styles to recognize and reject self-pollen (Kondo et al. 2002).

In this work, we analyze the coding sequence of the *Asr2* gene from several populations of tomato species living in dry, mesic, or humid habitats. *L. hirsutum*, *L. peruvianum* f. glandulosum, *L. cheesmanii*, *L. esculentum* cv. cerasiforme (likely to be the wild ancestor of cultivated

[1] These authors contributed equally to the work.

Key words: tomato, *Lycopersicon*, water stress, Asr genes, adaptive evolution.

E-mail: srossi@fbmc.fcen.uba.ar.

*Mol. Biol. Evol.* 20(12):1955–1962. 2003
DOI: 10.1093/molbev/msg214
*Molecular Biology and Evolution*, Vol. 20, No. 12,

tomato) and *L. esculentum* cv. Ailsa Craig grow in mesic or humid habitats, whereas *L. peruvianum* v. humifusum and *L. chilense* dwell in dry habitats. *L. chilense* is particularly interesting since it inhabits the Atacama desert, exposed to one of the driest climates in the world. Our study strongly suggests that *Asr2* has been the target of positive selection during the evolution of *Lycopersicon* species dwelling in dry habitats.

## Materials and Methods

### Plant Material

Seeds from *L. chilense* (accession LA2884), *L. esculentum* v. cerasiforme (accession LA1228), *L. hirsutum* (accession LA1223), *L. peruvianum* v. humifusum (accession LA0385), *L. peruvianum* f. glandulosum (accession LA1292) and *L. cheesmanii* (accession LA0521) were kindly provided by the Tomato Genetics Resource Center (University of California, Davis). Plants were grown in soil pots under environmental light in a greenhouse.

### DNA Amplification and Sequencing

Genomic DNA extractions were performed following the protocol of Peralta and Spooner (2001). The following primers were used in PCR reactions to amplify the coding sequence of *Asr2*. Upper primer: 5′-AGAGAAGCAATACAATATGGCT-3′. Lower primer: 5′-TATTAGACAAAACATAGAGTCC-3′. Thirty-five cycles consisting of denaturation (95°C for 1 min), annealing (55°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 of approximately 520 bp (as expected for the known sequence of *Asr2* from *L. esculentum* cv. Ailsa Craig) were excised from the gel and purified using the Concert Kit (Gibco). The PCR products were sequenced using the same primers at the Biotechnology Resource Center (Cornell University) with an ABI 3700 sequencer. DNA sequences were determined on both strands. Sequences were edited using BioEdit (Hall 1999). Numbers of alleles analyzed for each species were *L. peruvianum* v. humifusum, n = 4; *L. peruvianum* f. glandulosum, n = 4; *L. chilense*, n = 8; *L. cheesmanii*, n = 6; *L. hirsutum*, n = 4; *L. esculentum* cv. Cerasiforme, n = 2; and *L. esculentum* cv. Ailsa Craig, n = 4. All the alleles within each species were identical.

Intron and exon boundaries were determined by comparing genomic and cDNA clones. Amino acid sequences were deduced from cDNA sequences. Deduced peptide sequences are legitimate since antibodies raised against synthetic peptides were able to recognize the ASR2 protein in Western blot experiments (N. Frankel and N. Iusem, unpublished data).

### Sequence Analysis

Sequences were aligned using ClustalX (Thompson et al. 1997). Replacement versus synonymous substitution rates ($K_a/K_s = \omega$) were calculated with MEGA version 2.1 freeware (Kumar et al. 2001) using the Nei-Gojobori algorithm (Nei and Gojobori 1986). Alignment gaps were not considered for $K_a/K_s$ calculations.

We also estimated the ($K_a/K_s$ ) ratio between *L. esculentum* and *L. chilense* for several loci available in databases: anonymous EST clone CT268 (GenBank accession number AA824988 and nonannotated, respectively), anonymous EST clone CT251 (GenBank accession number AA824968 and nonannotated, respectively), *sucr* or invertase (Elliot et al. 1993 and nonannotated, respectively), dehydrin (accession numbers BF097038 and M97211, respectively), H1-like (accession numbers Z11842 and AF253416, respectively), and class I acidic endochitinase (accession numbers Z15141 and L19342, respectively). The nonannotated sequences from *L. chilense* were kindly provided by Thomas Städler, University of Munich, Germany.

### Phylogenetic Analysis

We retrieved the complete ITS1 and ITS2 rDNA sequences from GenBank (accession numbers AJ300200, AJ300201, AJ300202, AJ300203, AJ300204, AJ300208, AJ300209, AJ300210, and AJ300215), reported by Marshall et al. (2001). In contrast to the analysis by Marshall et al. (2001), we only included fully aligned sites, eliminating trailing as well as internal gaps from the data matrix. The g1 value (−1.285007) computed from 10,000 random trees was significant ($P < 0.001$), indicating that this data matrix has a strong phylogenetic signal (Hillis and Huelsenbeck 1992). Data were run under PAUP* version 4.0 beta (Swofford 1998) using neighbor-joining, minimum-evolution, maximum-parsimony, and maximum-likelihood algorithms. *Nicotiana tabacum* was used as outgroup. Maximum parsimony and minimum evolution were run using heuristic search, branch-swapping tree-bisection-reconnection, and MulTrees option in effect. Bootstrap values for ingroup nodes were above 50% in maximum-parsimony–based, neighbor-joining–based, and minimum-evolution–based trees. In all cases, bootstrap values were obtained with 500 pseudoreplicates. Maximum-likelihood methods gave low bootstrap values for all nodes.

Ancestral states of variable positions were inferred by means of the distance-based Bayesian approach, using the software Anc-gene (Zhang and Nei 1997). The use of a Poisson or a JTT amino acid substitution matrix did not change the inference of ancestral sequences. PAML (Yang 1997) was used to estimate the number of synonymous and nonsynonymous substitutions per branch under a free-ratio model.

Several models aimed to analyze whether certain branches of the tree have unusually high $K_a/K_s$ ($\omega$) ratios were compared by means of likelihood ratio tests. In these tests, codon equilibrium frequencies were calculated from average nucleotide frequencies at each of the three codon positions and the transition/transversion ratio ($\kappa$) was estimated from the data (about 1.2 for all models) (Goldman and Yang 1994). The natural logarithms of the likelihoods (LnL) associated to each one of the different models of interest (see *Results* for further explanation) were also calculated using PAML. The simplest model considered is one in which all branches have a background $\omega_o$

**Table 1**
***Lycopersicon* Species Used in This Work and Their Habitat Conditions Related to Water Availability**

| Species | Population | GenBank Accession Number | Habitat |
|---|---|---|---|
| *L. chilense* | Ayavirí, Antofagasta, Chile. | AY217009 | Desertic |
| *L. esculentum* cv. Ailsa Craig | — | L20756 | — |
| *L. esculentum* v. cerasiforme | Macas, Morona-Santiago, Ecuador | AY217012 | Humid |
| *L. hirsutum* | Alausí, Chimborazo, Ecuador | AY217010 | Mesic |
| *L. cheesmanii* | Fernandina, Galápagos Islands, Ecuador | AY217013 | Humid |
| *L. peruvianum* v. humifusum | San Juan, Cajamarca, Perú | AY217011 | Semiarid |
| *L. peruvianum* f. glandulosum | San Mateo, Lima, Perú | AY217014 | Mesic |

NOTE.—GenBank accession numbers are for the corresponding *Asr2* gene.

value (Ln L = –577.08). Other models of interest are those in which the $K_a/K_s$ ratio in the branches leading to *L. chilense* ($\omega_{ch}$) and *L. peruvianum* f. humifusum ($\omega_{ph}$) are equal but different from $\omega_o$ ($\omega_o \neq \omega_{ch} = \omega_{ph}$, Ln L = –569.18) or different to each other but one of the ω values equal to $\omega_o$ ($\omega_o = \omega_{ph} \neq \omega_{ch}$, Ln L = –574.82 or $\omega_o = \omega_{ch} \neq \omega_{ph}$, Ln L = –572.15); more complex models considering $\omega_o \neq \omega_{ph} \neq \omega_{ch}$ (Ln L = –569.18) and the free-ratio model (Ln L = –565.30) were also tested. Comparisons between models were performed by means of $\chi^2$ tests with one degree of freedom.

### Climatological Data

The climatological data were retrieved from the Dirección Meteorológica of Chile (http://www.meteochile.cl), the Servicio Nacional de Meteorología e Hidrología of Perú (http://www.senhami.gob.pe), and the Instituto Nacional de Meteorología e Hidrología of Ecuador (http://www.inhami.gov.ec).

## Results

### Amino Acid and Synonymous Substitution Rates

We determined the complete coding sequence of *Asr2* gene from seven tomato populations living in different habitats (table 1 and fig. 1*A*). The aligned sequences are approximately 345 bp long and include two exons. The differences in length are due to a trinucleotide repeat of variable size at the 3′ end generating a histidine tract at the carboxylic end of the protein (fig. 1*B*). Variables sites make up 7% of all nucleotide positions. Sequences from *L. cheesmanii* and both varieties of *L. esculentum* turned out to be identical. We included *Ci21b* in the alignment (fig. 1), the orthologous gene from *Solanum tuberosum* (Schneider, Salamini, and Gebhardt 1997). *Solanum* and *Lycopersicon* diverged about 10 MYA (Alba et al. 2000).

To investigate the evolutionary forces acting on *Asr2* genes, we calculated the $K_a/K_s$ ratio of all pairwise comparisons (table 2). Increasing trends of this ratio are usually indicative of relaxed purifying selection and/or positive selection events (Kreitman 2000), whereas values greater than 1 can be viewed as a strong evidence of positive selection. However, interpretation of this parameter is not straightforward, because it might overlook many cases of Darwinian selection occurring on individual amino acids, in long branches (Liberles et al. 2001), or during demographic events (Yang and Bielawski 2000).

Analysis of table 2 shows that $K_a/K_s$ ratios higher than 1 correspond to most of the pairwise comparisons involving *L. peruvianum* v. humifusum. The same situation is

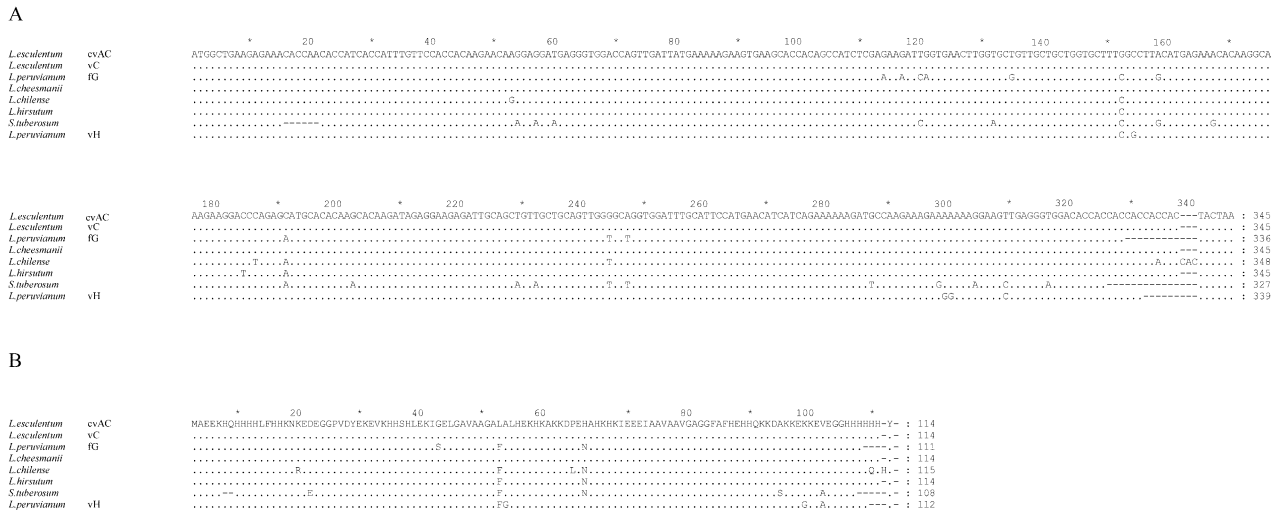


FIG. 1.—Nucleotide (*A*) and amino acid (*B*) sequence alignment of *Lycopersicon Asr2*/ASR2 and the orthologous *Ci21b*/CI21B (GenBank accession number U76611) from potato. Abbreviations: cvAC = cv. Ailsa Craig, fg = f. glandulosum, vH = v. humifusum, and vC = v. cerasiforme.

**Table 2**
**$K_a$, $K_s$, and ω Values for All Pairwise Comparisons Involving *Lycopersicon* Species and the Outgroup *Solanum Tuberosum***

| | *L. chilense* | *L. cheesmanii* | *L. es.* cerasiforme | *L.* es. Ailsa Craig | *L. pe.* humifusum | *L. hirsutum* | *L. pe.* glandulosum | *S. tuberosum* |
|---|---|---|---|---|---|---|---|---|
| *L. chilense* | | **1.002** | **1.002** | **1.002** | 1.773 | 0.245 | 0.116 | 0.083 |
| *L. cheesmanii* | 0.0161<br>0.0160 | | NA | NA | ∞ | 0.495 | 0.098 | 0.076 |
| *L. esculentum* v. cerasiforme | 0.0161<br>0.0160 | 0<br>0 | | NA | ∞ | 0.495 | 0.098 | 0.076 |
| *L. esculentum* cv. Ailsa Craig | 0.0161<br>0.0160 | 0<br>0 | 0<br>0 | | ∞ | 0.495 | 0.098 | 0.076 |
| *L. peruvianum* f. humifusum | 0.0283<br>0.0160 | 0.0201<br>0 | 0.0201<br>0 | 0.0201<br>0 | | **1.251** | 0.199 | 0.092 |
| *L. hirsutum* | 0.0080<br>0.0325 | 0.0080<br>0.0161 | 0.0080<br>0.0161 | 0.0080<br>0.0161 | 0.0201<br>0.0161 | | 0.028 | 0.042 |
| *L. peruvianum* f. glandulosum | 0.0120<br>0.1027 | 0.0120<br>0.1215 | 0.0120<br>0.1215 | 0.0120<br>0.1215 | 0.0242<br>0.1212 | 0.0040<br>0.1411 | | 0.065 |
| *S. tuberosum* | 0.0210<br>0.2415 | 0.0201<br>0.2643 | 0.0201<br>0.2643 | 0.0201<br>0.2643 | 0.0242<br>0.2635 | 0.0120<br>0.2884 | 0.0160<br>0.2447 | |

NOTE.—$K_a$ (top) and $K_s$ (bottom) values are displayed in each square below the diagonal. ω values ($K_a/K_s$ ratios) are shown above the diagonal. Ratios higher than 1 are in bold. Abbreviations: *L. es.* = *L. esculentum*; *L. pe.* = *L. peruvianum*; NA = not applicable.

observed in the comparison *between L. chilense* and *L. peruvianum* v. humifusum. In addition, the $K_a/K_s$ ratios between *L. chilense* and *L. cheesmani* and both varieties of *L. esculentum* are slightly higher than 1. The rest of the species show $K_a/K_s$ values lower than 1. Interestingly, *L. chilense* and *L. peruvianum* v. humifusum inhabit dry habitats, whereas the rest grow in mesic or humid environments (table 1).

We also investigated whether other genes exhibit analogous departures from the usual $K_a \ll K_s$ trend by comparing our $K_a/K_s$ estimates for *Asr2* with those for genes available in databanks: CT251, CT268, *sucr*, dehydrin, histone H1-like, and vacuolar acidic endochitinase (table 3). This analysis was possible only for *L. chilense*, as there is a lack of sequence information for *L. peruvianum* f. humifusum, the other species living in dry habitats. The outcome of these analyses revealed that $K_a/K_s$ values were lower than 1 for five out of six of these "control" genes. The exception was that encoding class I endochitinase, previously proved to have undergone adaptive evolution in the genus *Arabis* (Bishop, Dean, and Mitchell-Olds 2000).

Our preliminary conclusion from these data is that there exists an apparent acceleration in the rate of amino acid substitution in the *Asr2* gene in the lineages presently exposed to limited water availability.

## Phylogenetic Analysis

To analyze the pattern of nucleotide substitution in the *Asr2* gene of tomato in a phylogenetic context, we first searched for the published phylogenies of the genus *Lycopersicon*. Two recent molecular phylogenetic studies on *Lycopersicon* were performed by Peralta and Spooner (2001), based on the sequence of a gene encoding the enzyme granule-bound starch synthase (GBSSI) and Marshall et al. (2001), based on ITS1-ITS2 rDNA. A third study by Miller and Tanksley (1990) reported dendrograms constructed using RFLP markers.

The position of *L. peruvianum* v. humifusum in the tree of Marshall et al. (2001) was different from those reported by Miller and Tanksley (1990) and Peralta and Spooner (2001). Given this conflicting issue, we decided to reanalyze the ITS1-ITS2 data set from Marshall et al. (2001) and to construct phylogenetic trees using alternative methods. The resulting topology using maximum-parsimony, minimum-evolution, and neighbor-joining algorithms supported the position of *L. peruvianum* v. humifusum as sister species of the *esculentum/cheesmanii* clade (fig. 2). The position of *Lycopersicon peruvianum* v. humifusum in all trees obtained using rDNA sequences is consistent with that obtained by Peralta and Spooner (2001) and by Miller and Tanksley (1990), indicating that *L. peruvianum* does not conform to a clade. The node that joins *L peruvianum* v. humifusum with the *L. esculentum/L. cheesmanii* complex (fig. 2) was supported by 94%, 74%, and 53% bootstrap values when analyzed under Neighbor-Joining, minimal-evolution and maximum-parsimony algorithms, respectively.

**Table 3**
**$K_a$, $K_s$ and ω Values for *Asr2* and Other Available Genes in Pairwise Comparisons Between *L. Chilense* and *L. Esculentum***

| Gene[a] | $K_a$ | $K_s$ | ω |
|---|---|---|---|
| *CT268* | 0.00736 | 0.0273 | 0.269 |
| *CT251* | 0.0368 | 0.0916 | 0.347 |
| *sucr* | 0.0081 | 0.0210 | 0.385 |
| *dehydrin* | 0.0157 | 0.0266 | 0.590 |
| *H1-like* | 0.0134 | 0.0290 | 0.462 |
| *endochitinase* | 0.0269 | 0.0189 | 1.423 |
| *Asr2* | 0.0161 | 0.0160 | 1.002 |

[a] GenBank Accession numbers or source of sequences are stated in *Materials and Methods*.

The split of *L. peruvianum* v. humifusum from *L. peruvianum* complex is consistent with the fact that the former taxa is morphologically distinct and has a high degree of cross incompatibility to the other subspecies of *L. peruvianum* (Taylor 1986).

### Substitutions Along the Branches of the Lycopersicon Tree

We estimated the number of replacement ($n$) and synonymous ($s$) substitutions on the branches of the phylogenetic tree by different approaches. A distance-based Bayesian method (Zhang and Nei 1997) was used to infer the ancestral nucleotide sequences that allowed us to map the substitutions along the branches of the tree. Finally, we estimated the number of replacement and synonymous changes per branch using a maximum-likelihood procedure (Goldman and Yang 1994) under a "free-ratio" model (Yang 1998), which assumes different $K_a/K_s$ ($\omega$) per branch. Maximum likelihood and Bayesian gave almost identical results, showing an acceleration in the rate of replacement/synonymous substitution in the terminal branches of *L. chilense* and *L. peruvianum* v. humifusum (fig. 2).

To determine if these accelerations are significant, we compared the results of likelihood ratio tests under models with different assumptions regarding branch-to-branch variation in the ratio of replacement to synonymous substitution rates (Yang 1998). Since our original aim was to explore whether lineages dwelling in arid habitats experienced accelerated rates of amino acid replacement, we *a priori* assumed three likely different $\omega$ rates: $\omega_{ch}$ (for the terminal branch leading to *L. chilense*), $\omega_{ph}$ (for the terminal branch leading to *L. peruvianum* v. humifusum), and $\omega_o$ (background ratio for the remaining branches of the tree). The one-ratio model ($\omega_o = \omega_{ch} = \omega_{ph}$) was contrasted with a two-ratio model ($\omega_o \neq \omega_{ch} = \omega_{ph}$) to test the hypothesis of acceleration in these terminal branches. The analysis indicates that $K_a/K_s$ in these two branches is significantly different from the background ratio ($\chi^2 = 15.8$, 1df, $P < 0.0001$). To explore whether only one or both branches exhibited accelerated $\omega$, we performed additional tests in which only one of the ratios is compared with the background ratio ($\omega_o$) while the other ratio is either allowed to vary freely or constrained to be equal to $\omega_o$. The estimated $\omega_{ch}$ is significantly different from $\omega_o$ when $\omega_{ph}$ is constrained to be equal to $\omega_o$ ($\omega_o = \omega_{ph} \neq \omega_{ch}$ versus $\omega_o = \omega_{ph} = \omega_{ch}$, $\chi^2 = 4.52$, 1df, $P < 0.05$) or allowed to vary freely ($\omega_o \neq \omega_{ph} \neq \omega_{ch}$ versus $\omega_o = \omega_{ph} \neq \omega_{ch}$, $\chi^2 = 11.28$, 1df, $P < 0.001$). Analogously, $\omega_{ph}$ is significantly different from $\omega_o$ when $\omega_{ch}$ is either constrained to be equal to $\omega_o$ ($\omega_o = \omega_{ch} \neq \omega_{ph}$ versus $\omega_o = \omega_{ph} = \omega_{ch}$, $\chi^2 = 9.8$, 1df, $P < 0.01$) or allowed to vary freely ($\omega_o \neq \omega_{ph} \neq \omega_{ch}$ versus $\omega_o = \omega_{ch} \neq \omega_{ph}$, $\chi^2 = 5.94$, 1df, $P < 0.05$). This statistical evaluation of our data validates the hypothesis of an accelerated replacement substitution rate in *Asr2* in *L. chilense* and *L. peruvianum* v. humifusum.

The Bayesian method permitted the mapping of each amino acid substitution onto the tree (fig. 2). We found one conservative substitution in the lower branch leading to the ancestor of the *peruvianum* (*esculentum*/*cheesmanii*) clade.

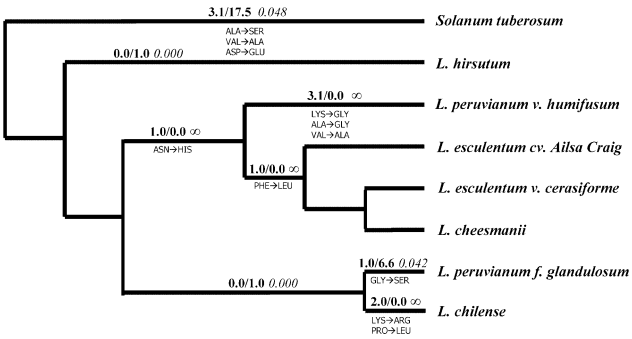


FIG. 2.—Phylogeny of the genus *Lycopersicon* obtained by maximum parsimony using ITS1-IT2 sequences. Above each branch: maximum-likelihood estimates of the number of replacement/synonymous substitutions (bold) and $\omega$ (italics) in *Asr2*/*Ci21b* estimated under a "free-ratio" model. The estimates of $\omega$ for a two-ratio model ($\omega_o \neq \omega_{ch} = \omega_{ph}$) (not shown in the figure) are the following: $\omega_o = 0.065$; $\omega_{ch} = \infty$; $\omega_{ph} = \infty$. A distance-based Bayesian method was used to infer ancestral nucleotide sequences and map amino acid changes (below the corresponding branches). As the ITS sequence of *L. peruvianum* f. glandulosum was not available, we used the ITS sequences of two subspecies of *L. peruvianum* closely related to *L. peruvianum* f. glandulosum (Miller and Tanksley 1990). Both appeared as sister taxa of *L. chilense*.

Another conservative change is present in the *esculentum*/*cheesmanii* clade. Remarkably, the terminal branches of *L. chilense* and *L. peruvianum* v. humifusum exhibit two and three amino acid substitutions in the water stress–inducible ASR2 protein, respectively, while being devoid of synonymous substitutions (according to the two methods used).

The changes in the branch leading to *L. chilense* are Pro63→Leu and Lys18→Arg. The former is a radical substitution probably associated with alterations of protein spatial conformation. One of the three changes in the branch of *L. peruvianum* v. humifusum can be considered conservative: Ala52→Gly, whereas the second Val104→Ala implies a size change in the residue (fig. 2). The third one is a nonconservative change of the uncharged and small glycine by the positively charged lysine.

## Discussion

The *Asr* gene family arose relatively early in plant genomes. *Asr* genes have been described in gymnosperms as well as in several angiosperms, and their expression is correlated to different abiotic stresses (Maskin et al. 2001). The present work shows that *Asr2*, one of the members of this gene family in tomato that is expressed under water-stress conditions, has experienced an accelerated rate of replacement substitutions in *L. chilense* and *L. peruvianum* v. humifusum, wild tomato lineages dwelling in dry habitats. Pairwise $K_a/K_s$ show values equal or greater than 1 (a hallmark of genes undergoing adaptive evolution) in comparisons involving these taxa. The likelihood ratio tests provide conclusive evidence favoring the hypothesis of the acceleration of the rate of replacement substitution in the terminal branches leading to *L. chilense* as well as to *L. peruvianum* v. humifusum. These significant increases in the $\omega$ values are a strong evidence for positive selection in these lineages. However, the high amino acid substitution rate observed in *L. chilense* and *L. peruvianum* v.

humifusum lineages could be the result of the fixation of slightly deleterious alleles in populations that experienced severe reductions in size (bottlenecks). The signature of such demographic events would be a genome-wide reduction of DNA variation. In this sense, a recent paper by Baudry et al. (2001), studying the same population of *L. chilense*, found a considerable degree of sequence polymorphism in several genes, thus ruling out the alternative of nonadaptive factors in the evolution of *Asr2* in this species. In addition, the sequence of the eight *L. chilense Asr2* alleles were identical (data not shown), including the 118-bp intron, suggesting that a recent and rapid fixation of a selectively favored allele swept most intraspecific sequence variation.

It is noteworthy that the branches of *L. chilense* and *L. peruvianum* v. humifusum, the only species analyzed here that grow in dry environments, display nonconservative amino acid substitutions in *Asr2*. None of the terminal branches leading to species living in mild conditions showed any radical amino acidic change (figs. 1*B* and 2). A strong piece of evidence supporting our conclusion of adaptive evolution is one radical amino acid substitution (Pro63→Leu) found only in *L. chilense*, a species from the Atacama Desert, a habitat where rainfalls are extremely infrequent. Interestingly, the proline in position 63 is conserved in almost all Asr-like proteins known thus far, even in primitive lineages such as gymnosperms (Maskin et al. 2001). The same radical amino acid change either in position 102 or 105 of the human PrPC gene is the cause of a conformational shift associated to a prionlike disease (Prusiner 1997). Analogously, the substitution Val→Ala that occurred in the terminal branch leading to *L. peruvianum* v. humifusum is the cause of a similar neurodegenerative phenotype if present in the PrPC protein (Prusiner 1997). The Gly101→Lys substitution in ASR2 from *L. peruvianum* v. humifusum implies a change in the net charge of this small protein. Any or all of the different and nonconservative amino acid changes occurring in the *L. chilense* and in the *L. peruvianum* v. humifusum branches would generate proteins with analogous functions relevant to water-stress responses.

Van der Hoeven et al. (2002) compared a large data set of *Lycopersicon* ESTs with the *Arabidopsis* genome and concluded that genes encoding transcription factors are the fastest evolving in these two lineages, which diverged 150 MYA. This trend could be valid for plants in general. Moreover, the present work shows that *Asr2*, a putative transcription factor, would be clear example of such rapidly evolving genes.

It is well established that related proteins displaying a certain extent of structural diversity usually show functional differences that may have a strong impact on fitness. It is widely accepted that a protein co-opted for an emerging new function often experiences an episode of rapid sequence evolution driven by positive selection (Wallis 2001). Adaptive evolution after gene duplication has been reported in several gene families (Hughes 2002). In this context, *Asr2*, a member of a gene family, suggestively experienced an acceleration of the nonsynonymous rate in the two tomato lineages adapted to dry habitats. This pattern of evolution is in sharp contrast to that of other genes also known to be induced under drought in tomato, such as dehydrin and histone H1-like genes (Chen et al. 1993; Wei and O'Connell 1996). The only other available gene that displayed signs of such a type of selection was that encoding class I acidic endochitinase, a gene that is turned on by drought (Chen et al. 1994) but also involved in the defense to pathogenic fungi (Bishop, Dean, and Michael-Olds 2000).

The number of genes subjected to positive selection has been estimated to be as low as 0.5 % of the totality of genes in comprehensive DNA sequence databases (Endo, Ikeo, and Gojobori 1996). Other well-documented examples of such rapidly evolving genes are those encoding surface antigens of parasites with short generation times (Endo, Ikeo, and Gojobori 1996), ribonuclease genes in colobine monkeys (Zhang, Zhang, and Rosenberg 2002), mammal protein hormones (Wallis 2001), and proteins involved in gamete recognition (Vacquier, Swanson, and Lee 1997).

The present study also allows us to envisage possible biotechnological endeavors toward the improvement of crop yields in dry soils. A conceivable strategy to achieve that goal might well be the introduction of *Asr* genes from tolerant species in cultivated tomato by genetic engineering. In this regard, the task would not be straight forward because of the genetic complexity underlying the physiological response (Zhu 2002). However, it would be worth the effort, as there is evidence on the ability of certain transgenic plants overexpressing a single master transcription factor to acquire water-stress tolerance (Kasuga et al. 1999; Hsieh et al. 2002).

In summary, on grounds of the data reported in this work, we hypothesize that *Asr2* genes of the wild *L. chilense* and *L. peruvianum* v. humifusum species underwent adaptive changes that might be associated to success in colonizing arid habitats. Adaptation to such stringent conditions would depend on multiple physiological and genetic factors. For instance, at the physiological level, an overt adaptive phenotype is the development of a deep root system in *L. chilense* able to locate water trapped in the rocky soil (Rick 1973). Other genes, as well as their patterns of evolution, are to be investigated to gain full insight into the molecular adaptation mechanisms of plants to dryness.

## Acknowledgments

The authors wish to thank Spencer Muse and the anonymous reviewers for constructive criticisms and suggestions. This work was supported by grants from the Universidad de Buenos Aires and the Agencia Nacional de Promoción Científica y Técnica, Argentina. N.F. holds a fellowship from Fundación Antorchas. We thank the Tomato Genetics Resource Center for providing the seeds of the tomato wild populations. E.H., N.D.I. and M.S.R. are members of the Carrera del Investigador Científico, CONICET, Argentina.

## Literature Cited

Alba, R., P. M. Kelmenson, M. M. Cordonnier-Pratt, and L. H. Pratt. 2000. The phytochrome gene family in tomato and the

rapid differential evolution of this family in Angiosperms. Mol. Biol. Evol. **17**:362–373.

Atanassova, R., M. Leterrier, C. Gaillard, A. Agasse, E. Sagot, P. Coutos-Thévenot, and S. Delrot. 2003. Sugar-regulated expression of a putative hexose transport gene in grape. Plant Physiol. **131**:326–334.

Baudry, E., C. Kerdelhué, H. Innan, and W. Stephan. 2001. Species and recombination effects on DNA variability in the tomato genus. Genetics **158**:1725–1735.

Bishop, J. G., A. M. Dean, and T. Mitchell-Olds. 2000. Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. Proc. Natl. Acad. Sci. USA **97**:5322–5327.

Bohnert, H. J., and E. Sheveleva. 1998. Plant stress adaptations—making metabolism move. Curr. Opin. Plant Biol. **1**:267–274.

Boyer, J. S. 1982. Plant productivity and environment. Science **218**:444–448.

Chen, R. D., N. Campeau, A. F. Greer, G. Bellemare, and Z. Tabaeizadeh. 1993. Sequence of a novel abscisic acid- and drought-induced cDNA from wild tomato (*Lycopersicon chilense*). Plant Physiol. **103**:301.

Chen, R. D., L. X. Yu, A. F. Greer, H. Cheriti, and Z. Tabaeizadeh. 1994. Isolation of an osmotic stress- and abscisic acid-induced gene encoding an acidic endochitinase from *Lycopersicon chilense*. Mol. Gen. Genet. **245**:195–202.

Elliott, K. J., W. O. Butler, C. D. Dickinson, Y. Konno, and T. S. Vedvick. 1993. Isolation and characterization of fruit vacuolar invertase genes from two tomato species and temporal differences in mRNA levels during fruit ripening. Plant Mol. Biol. **21**:515–524.

Endo, T., K. Ikeo, and T. Gojobori. 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13**:685–690.

Finkelstein, R. R., S. S. Gampala, and C. D. Rock. 2002. Abscisic acid signaling in seeds and seedlings. Plant Cell **14**(Suppl.):S15–45.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

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.

Hare, P. D., W. A. Cress, and J. Van Staden. 1998. Dissecting the roles of osmolyte accumulation during stress. Plant Cell Environ. **21**:535–553.

Hillis, D. M., and J. P. Huelsenbeck. 1992. Signal, noise and reliability in molecular phylogenetic analysis. J. Hered. **83**:189–195.

Hsieh, T. H., J. T. Lee, Y. Y. Charng, and M. T. Chan. 2002. Tomato plants ectopically expressing Arabidopsis CBF1 show enhanced resistance to water deficit stress. Plant Physiol. **130**:618–26.

Hughes, A. L. 2002. Adaptive evolution after gene duplication. Trends Genet. **18**:433–434.

Ingram, J., and D. Bartels. 1996. The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47**:377–403.

Iusem, N. D., D. M. Bartholomew, W. D Hitz, and P. A. Scolnik. 1993. Tomato transcript induced in water stress and ripening. Plant Physiol. **102**:1353–1354.

Kasuga M., Q. Liu, S. Miura, K. Yamaguchi-Shinozaki, and K. Shinozaki. 1999. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat. Biotech. **17**:287–291.

Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. **01**:539–559.

Kondo, K., M. Yamamoto, D. P. Matton, T. Sato, M. Hirai, S. Norioka, T. Hattori, and Y. Kowyama. 2002. Cultivated tomato has defects in both S-RNase and HT genes required for stylar function of self-incompatibility. Plant J. **29**:627–636.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Distributed by the authors, Arizona State University, Tempe, Ariz.

Liberles, D. A., D. R. Schreiber, S. Govindajaran, S. G. Chamberlin, and S. A. Benner. 2001. The adaptive evolution database (TAED). Genome Biol. **2**:RESEARCH0028.1–0028.6.

Marshall, J. A., S. Knapp, M. R. Davey, J. B. Power, E. C. Cocking, M. D. Bennett, and A. V. Cox. 2001. Molecular systematics of *Solanum* section *Lycopersicum* (*Lycopersicon*) using the nuclear ITS rDNA region. Theor. App. Genet. **103**:1216–1222.

Maskin, L., G. E. Gudesblat, J. E. Moreno, F. O. Carrari, N. Frankel, A. Sambade, M. Rossi, and N. D. Iusem. 2001. Differential expression of the members of the *Asr* gene family in tomato (*Lycopersicon esculentu*m). Plant Sci. **161**:739–746.

Miller, J. C., and S. D. Tanksley. 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. **80**:437–448.

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

Oliver, M. J., Z. Tuba, and B. D. Mishler. 2000. Evolution of dessication tolerance in land plants. Plant Ecol. **151**:85–100.

Peralta, I. E., and D. M. Spooner. 2001. Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (Solanum L. Secton *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). Am. J. Bot. **88**:1888–1902.

Prusiner, S. B. 1997. Prion diseases and the BSE crisis. Science **278**:245–251.

Qiu, Y. L., and J. D. Palmer. 1999. Phylogeny of early land plants: insights from genes and genomes. Trends Plant Sci. **4**:26–30.

Rossi, M. M., and N. D. Iusem. 1994. Tomato genomic clone homologous to a gene encoding an ABA-induced protein. Plant Physiol. **104**:1073–1074.

Rick, C. M. 1973. Potential genetic resources in tomato species: clues from observations in native habitats. Pp. 255–269 *in* A. M. Srb, ed. Genes, enzymes, and populations. Plenum, New York.

———. 1979. Biosystematic studies in *Lycopersicon* and closely related species of Solanum. Pp. 667–677 *in* J. G. Hawkes, R. N. Lester, and A. D. Skelding, eds. The biology and taxonomy of Solanaceae. Linnean Society Symposium Series 7. Academic Press, New York.

Schneider, A., F. Salamini, and C. Gebhardt. 1997. Expression patterns and promoter activity of the cold-regulated gene ci21A of potato. Plant Physiol. **113**:335–345.

Shinozaki, K., and K. Yamaguchi-Shinozaki. 1997. Gene expression and signal transduction in water-stress response. Plant Physiol. **115**:327–334.

Skriver, K., and J. Mundy. 1990. Gene expression in response to abscisic acid and osmotic stress. Plant Cell **2**:503–512.

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0 beta. Sinauer Associates, Sunderland, Mass.

Taylor, I. B. 1986. Biosystematics of the tomato. Pp. 1–34 *in* J. G. Atherton and J. Rudich, eds. The tomato crop: a scientific basis for improvement. Chapman and Hall, London.

Thompson, J. D., T. J. Gibson, F. Pleuniak, E. Janemougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24**:4876–4882.

Vacquier, V. D., W. J. Swanson, and Y-H. Lee. 1997. Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? J. Mol. Evol. **44**(Suppl. 1):S15–S22.

Van der Hoeven R., C. Ronning, J. Giovannoni, G. Martin, and S. Tanksley. 2002. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell **14**:1441–1456.

Wallis, M. 2001. Episodic evolution of protein hormones in mammals. J. Mol. Evol. **53**:10–18.

Wei, T., and M. A. O'Connell. 1996. Structure and characterization of a putative drought-inducible H1 histone gene. Plant Mol. Biol. **30**:255–268.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**:555–556.

———. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

Yang, Z., and P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **1512**:496–501.

Zhang, J., and M. Nei. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. **44**(Suppl. 1):S139–S146.

Zhang, J., Y. Zhang, and H. F. Rosenberg. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat. Genet. **30**:411–415.

Zhu, J.-K. 2002. Salt and drought stress signal transduction in plants. Annu. Rev. Plant Biol. **53**:247–273.

Spencer Muse, Associate Editor

Accepted June 22, 2003