Genetica (2009) 136:13–25
DOI 10.1007/s10709-008-9295-1

Nucleotide polymorphism in the drought responsive gene *Asr2* in wild populations of tomato

Mariano I. Giombini · Nicolás Frankel · Norberto D. Iusem · Esteban Hasson

Received: 10 December 2007 / Accepted: 5 July 2008 / Published online: 19 July 2008

Abstract The *Asr* gene family (named after abscicic acid [ABA], stress, ripening), exclusively present in plant genomes, is involved in transcriptional regulation. Its members are up-regulated in roots and leaves of water- or salt-stressed plants. In previous work, evidence of adaptive evolution (as inferred from synonymous and nonsynonymous divergence rates) has been reported for *Asr2* in *Solanum chilense* and *S. arcanum*, two species dwelling in habitats with different precipitation regimes. In this paper we investigate patterns of intraspecific nucleotide variation in *Asr2* and the unlinked locus *CT114* in *S. chilense* and *S. arcanum*. The extent of nucleotide diversity in *Asr2* differed between species in more than one order of magnitude. In both species we detected evidence of non-neutral evolution, which may be ascribed to different selective regimes, potentially associated to unique climatic features, or, alternatively, to demographic events. The results are discussed in the light of demographic and selective hypotheses.

Keywords Polymorphism · *Asr* genes · *Solanum* · Drought · Selection

M. I. Giombini · N. Frankel · N. D. Iusem
Departamento de Fisiología, Biología Molecular y Celular and IFIByNE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, C1428EHA Buenos Aires, Argentina

E. Hasson (✉)
Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, C1428EHA Buenos Aires, Argentina
e-mail: ehasson@ege.fcen.uba.ar

Introduction

Plants have coped with water deficit since they colonized land environments about 400 MYA (Qiu and Palmer 1999). Thereafter, specific drought tolerance adaptations have evolved in Tracheophytes allowing them to expand their distribution range over drier areas. These adaptations consist of both morphological features and physiological responses. Although the molecular underlying mechanisms have been comprehensively studied (Hoekstra et al. 2001), the whole picture is still missing. Gaining a deeper insight into plant adaptations to water deficit is of great interest not only to agronomists but also to evolutionary biologists and population geneticists (Kane and Rieseberg 2007).

The desiccation stress response involves a complex ABA (abscicic acid)-mediated signal transduction pathway, probably initiated by membrane-bound osmosensors (Urao et al. 1999) in key tissues and culminates with the synthesis of protective molecules such as LEA (late embryogenesis abundant) proteins (Tunnacliffe and Wise 2007) and sugars (Peters et al. 2007). The *Asr* gene family (named after ABA, stress, ripening), which encodes transcription factors (Cakir et al. 2003), is present in most seed plants, and captures considerable attention as its members respond to ABA in roots and leaves of water- and salt-stressed plants (Iusem et al. 1993; Maskin et al. 2001; Kalifa et al. 2004a). Interesting for both mechanistic and biotechnological motivations, delivery of *Asr* transgenes happens to confer high salinity- and/or drought-tolerance to yeast (Bermudez-Moretti et al. 2006), maize (Jeanneau et al. 2002), tobacco (Kalifa et al. 2004b) and Arabidopsis (Yang et al. 2005).

For the above-mentioned reasons, the *Asr* family is a suitable and attractive model to study the evolutionary forces shaping polymorphism and divergence of its members and their contribution to adaptation of organisms to

threatening environmental demands. Recent comparative studies showed a significant excess of non-synonymous substitutions relative to synonymous substitutions in the coding sequence of *Asr2* in the branches of the phylogeny leading to *S. chilense* and *S. arcanum* (formerly—*L. peruvianum* v. humifusum) which inhabit arid and mesic habitats, respectively (Frankel et al. 2003). Moreover, the replacement substitutions observed included amino acid changes at nucleotide positions that are highly conserved even among distant genera. These results led the authors to reject the null hypothesis of neutral evolution and to conclude that *Asr2* might have been a target of natural selection. However, intraspecific patterns of nucleotide variation in this regulatory gene were not analyzed in Frankel et al. (2003). This issue might be relevant since population genetic theory makes specific predictions about patterns of intraspecific variation in the neighborhood of nucleotide sites under selection (Hudson 1990). Such patterns can be detected as departures from expectations under mutation-drift equilibrium (Kimura 1983).

In this study, we investigate the forces governing the evolutionary history of *Asr2* in two wild tomato species, *S. chilense* and *S. arcanum*, that inhabit areas with different rainfall regimes. The former lives in areas which are steadily dry along the year, while the latter inhabit areas with alternating dry and wet seasons along the year.

These sharply different climatic conditions may impose disparate environmental challenges to the plants. Extremely dry conditions may impose strong selective responses to drought responsive genes, whereas alternating dry and humid seasons may result in time-dependent selection that also represents a major force shaping molecular variation (Mustonen and Lässig 2007). Our hypothesis is that *S. arcanum* and *S. chilense* might have different patterns of sequence variation in the drought responsive gene *Asr2* as responses to such dissimilar climatic regimes.

To this end, we obtained *Asr2* sequences sampled in wild populations of *S. chilense* and *S. arcanum* to analyse patterns of intraspecific variation. However, since certain selective regimes and demographic events may leave similar footprints in patterns of intraspecific variation, we extended our study to a second non-linked gene region, *CT114*, as an attempt to disentangle the effects of demography and natural selection shaping variation in *Asr2*. The rationale of this experimental design is simple: demographic events are supposed to have a genome-wide effect, while selection is expected to affect only the target site and tightly linked regions (Wall et al. 2002).

Remarkably, patterns of variation in *Asr2* were non-compatible with neutral expectations and suggest that different selective forces govern nucleotide variation in both species, which, we argue, may be a consequence of adaptation to different regimes of rainfall. However, demographic events such as population bottlenecks and population admixture cannot be ruled out in *S. chilense* and *S. arcanum*, respectively

Materials and methods

Genes and populations

We analyzed nucleotide variation in the gene loci *Asr2* (Rossi and Iusem 1994) and *CT114* (Ganal et al. 1998). Both genes are situated in regions of low recombination. *Asr2* maps close to the centromere of chromosome 4 between *TG208* and *TG75A* markers (Rossi et al. 1996). *CT114*, which encodes a putative phospho-glycerate kinase (Ganal et al. 1998), is in chromosome 7 (see *Sol Genomics Network* at http://www.sgn.cornell.edu) and the estimated recombination rate for this gene region is $<0.001 \times 10^{-8}$ per site per generation (Roselius et al. 2005). *CT114* was chosen for comparative purposes as a potential neutral control, since patterns of nucleotide variation in this gene were shown to be consistent with neutral expectations in *S. peruvianum,* a close relative of *S. arcanum* (Städler et al. 2005). Unfortunately, there are no sequence data available for *S. arcanum*.

Seeds of *S. arcanum* from the populations of San Juan and Morochupa (accessions LA0385, 7°12′36″S/78°30′W and LA2151, 7°16′30″S/78°37′30″W, respectively) and *S. chilense* from Ayavirí (accession LA2884, 22°14′S/68°23′W) were obtained from the *Tomato Genetics Resource Center* (University of California, Davis; http://tgrc.ucdavis.edu). San Juan and Morochupa are about 15 km away from each other along an altitudinal gradient along Jequetepeque river valley, with no apparent geographical barrier between them. These localities are situated in a semiarid-mesic region of northwestern Perú with an average precipitation of ~600 mm per year and a marked seasonal regime. Rainfall, in this area is not evenly distributed along the year, and the climatic characteristic is the presence of a dry and a humid season (http://www.fao.org/ag/agl/swlwpnr/reports/y_lm/z_pe/pe.htm). In turn, Ayavirí is in the very desertic area of northern Chile with less than 100 mm of annual precipitations (http://www.serviu.cl/10/documentos/Urbanismoyconstrucción/Normativatécnica/aguaslluvias/capítulo3.pdf).

Genomic DNA extraction and amplification

We followed the protocol of Peralta and Spooner (2001) for DNA extraction starting with 8–10 mg of leaf tissue.

We PCR-amplified the coding sequences of *Asr2* and *CT114* using the following primers: forward 5′-AGAGA AGCAATACAATATGGCT-3′ and reverse 5′-TATTAGA CAAAACATAGAGTCC-3′ for *Asr2* (Frankel et al. 2003); and forward: 5′-ATATTGCTTAGGCGTCATCCA-3′ and

reverse: 5′-TTGAAACCAGCCGTTGC-3′ for *CT114* (Roselius et al. 2005). PCR amplifications were performed in a PTC-100 (M.J. Research) according to the following PCR profile: 35 cycles of denaturation (1 min at 92°C), annealing (1 min at 52°C for *Asr2*, and 58°C for *CT114*) and extension (1 min at 72°C). Amplification products were run in 1% agarose gels and bands of the expected size (~500 bp for *Asr2* and ~1.17 kbp for *CT114*) excised from the gel.

To avoid sequence ambiguities resulting from heterozygote sites, we cloned the amplified fragments into a plasmid before sequencing. Inserts of purified plasmids were sequenced by Macrogen Inc. (Korea) (http://www.macrogen.com) using *T3* and *T7* universal primers. Thus, each insert corresponds to a PCR amplified fragment coming from a single plant (i.e. one sequence per individual). Sequencing a single clone from a PCR amplified fragment could result in few spurious singletons (reflecting errors of polymerase) and/or in vitro recombination (Bradley and Hillis 1997). However these potential artifacts cannot substantially alter the results since they would tend to increase variation in *S. chilense* and to attenuate haplotype structure in *S. arcanum* (see results), and hence, in both cases, our inferences would be further supported. The exclusion of putative *Asr2* recombinants in *S. arcanum* may be controlling for the second possibility. Moreover, direct sequencing of PCR products revealed the presence of several identical alleles in *S. chilense* and in *S. arcanum* the presence of double peaks in some sites, due to heterozygosity, which were confirmed after cloning.

In addition, we retrieved eight *Asr2* sequences (Frankel et al. 2003; AY217009) and 10 *CT114* sequences (AY941646–AY941655) of *S. chilense,* and one *Asr2* sequence of *S. arcanum* (AY217011) from GenBank. All new sequences reported in this paper have been deposited in GenBank under accession numbers EF488609–EF488666.

On the whole, we obtained the entire coding sequence of *Asr2*, which includes an intron and two exons (~470 nt length, Figs. 1 and 2), from 17 individuals of *S. chilense* (Ayavirí), and 25 of *S. arcanum* (14 from San Juan and 11 from Morochupa) (Table 1). The amplified fragment of *CT114* (~1165 nt) encompassed two incomplete exons at the 5′ and 3′ ends of the amplified region plus two complete exons and three introns (Figs. 1 and 2). Twelve *CT114* sequences of *S. chilense* (Ayavirí) and 23 of *S. arcanum* (12 sequences from San Juan and 11 from Morochupa) were included in this study (Table 1).

Data analyses and tests of neutrality

Sequences were aligned using *ClustalW*. *DnaSP 4.1* (Rozas et al. 2003) was used to obtain estimates of nucleotide variation and for standard neutrality tests. We quantified nucleotide variation in terms of two common estimators of the neutral mutation parameter ($\theta = 4N_e\mu_0$, where N_e is the effective population size and μ_0 is the neutral mutation rate): the mean number of pairwise differences per site (π), also named nucleotide diversity, and Watterson's estimator of θ (θ_w) based on the observed number of segregating sites (Watterson 1975). Length polymorphisms (indels) were considered in these estimates since many of them were informative, unless it is explicitly stated. Each indel was counted as a single segregating site regardless of its length.

The recombination rate parameter (R), the minimum number of recombination events (R_m) and haplotype diversity (Hd) were also estimated using *DnaSP 4.1* for specific purposes that are explained subsequently. *DnaSP 4.1* was also employed to estimate genetic differentiation between populations and between *Asr2* haplotypes in *S. arcanum*, in terms of K_{st*} (Hudson 1992), D_{xy} (i.e. the average number of nucleotide differences per site between haplogroups), and the number of fixed differences between populations or haplogroups. Statistical significance of K_{st*} was assessed by means of permutation tests with 10,000 replications.

Gene genealogies were estimated by the statistical parsimony algorithm (Templeton et al. 1992) as implemented in *TCS 1.21* software (Clement et al. 2000). The *Asr2* sequence of *Solanum habrochaites* (GenBank accession numbers: AY217010) was used as a predefined outgroup (i.e. regardless of the outgroup weights given by TCS 1.21 software). For the only purpose of allowing the inclusion of this predetermined outgroup in the haplotype networks, parsimony confidence limits were sufficiently relaxed. However, except for the branch leading to the root, network structures were exactly the same as those obtained for 95% confidence limits (9 mutational steps). We applied the test proposed by Crandall and Templeton (1999) to evaluate the occurrence of intralocus recombination in the context of haplotype networks. This test evaluates whether mutations are spatially associated along the sequence of the putative recombinant haplotype, so that the ones occurring in one of the two branches that converge at the candidate recombinant are preferentially located at 5′ or 3′ region, while those present in the other branch occupy 3′ or 5′ region, respectively.

We employed Tajima's D (Tajima 1989) and Fu and Li's F (Fu and Li 1993) tests to investigate whether the frequency spectrum of nucleotide variation in our population samples were in agreement with neutral expectations. Tajima's D test is based on the standardized difference between π and θ_w. Similarly, Fu and Li's F is the standardized difference between π and the number of external mutations (i.e. singletons absent in the sequence of an outgroup) and takes into account whether variants are ancestral or derived. The logic of these tests is the same,

a

```
                EXON 1        INTRON        EXON 2
Asr2      1==========159--------279=========467


                   1 1                        4
               1   0 5                        4
               8   5 6                        2

SchN1          C   C C                        A
SchN2          .   . .                        .
SchN3          .   . .                        .
SchN4          .   . .                        .
SchN5          .   . .                        .
SchN6          .   . .                        .
SchN7          .   . .                        .
SchN8          .   . .                        .
Sch2           .   . .                        .
SchC           .   . .                        .
SchD           .   . .                        .
SchF           .   . .                        .
SchI           .   . .                        .
Schb           .   T T                        .
Sch1           .   . .                        .
SchB           .   . .                        .
Schce          T   . .                        G

S. habr.       .   . .                        .
```

b

```
                INTRON          EXON                INT.              EX.
CT114     1===17--------241=============558---------------727======//


                    1         3 3 4 4  5        5  6 6 6 7 7
                3   4         0 5 0 2  3        7  1 4 9 0 1
                5   6         3 3 4 6  7        4  0 5 0 7 2

Schce           C   C         C G C G  T        A  - T A G T
Schb            .   T         . A T .  .        .  . - . . .
AY941655        .   T         . A . .  .        .  . - . . .
AY941654        .   T         . A . .  .        .  . - . . .
AY941653        .   T         . A . .  .        .  . . . . .
AY941652        .   T         . A . .  .        .  . - . . .
AY941651        .   T         . A . .  .        .  . - . . .
AY941650        .   T         . A . .  .        .  . - . . .
AY941649        .   T         . A . .  .        .  . - . . .
AY941648        .   T         . A . .  .        .  . - . . .
AY941647        T   T         T A . A  C        G  A . G A C
AY941646        .   T         . A . .  .        .  . - . . .

S. habr.        .   T         T A . A  .        G  A . G . C
```

Fig. 1 Summary of nucleotide polymorphism in the aligned sequences of *Asr2* (**a**) and *CT114* (**b**) from *Solanum chilense* (LA2884). Location of exons and introns in the amplified region of each locus are shown at the top. Each row corresponds to an individual sequence, columns to polymorphic sites, and dots represent identity relative to the first aligned sequence. Insertion/deletions are indicated by short dashes. The last aligned sequence (*S. habr.*) corresponds to the outgroup *Solanum habrochaites*. Non-synonymous variable sites are shaded

since all statistics are unbiased estimators of the neutral parameter, they are expected to be equal under neutrality, whereas negative and positive significant values of the test statistics reflect departures from the neutral model. We also employed Fu's F_s test (Fu 1997), which examines the significance of the probability S' of having no fewer than k_0 haplotypes conditional on $\theta = \pi$. Finally, we applied Fay and Wu H test (Fay and Wu 2000) that is particularly efficient in distinguishing departures from the neutral standard model caused by positive selection from those caused by demography. This test is based on the difference between π and θ_H (another estimator of θ) which measures the relative number of high frequency derived variants among polymorphic sites. Neutrality tests were computed for variation at all, non-synonymous, silent (i.e. synonymous plus noncoding) and synonymous sites. For Fu and Li's and Fay and Wu′s tests, which require an outgroup, we used *Asr2* and *CT114* sequences of *S. habrochaites*

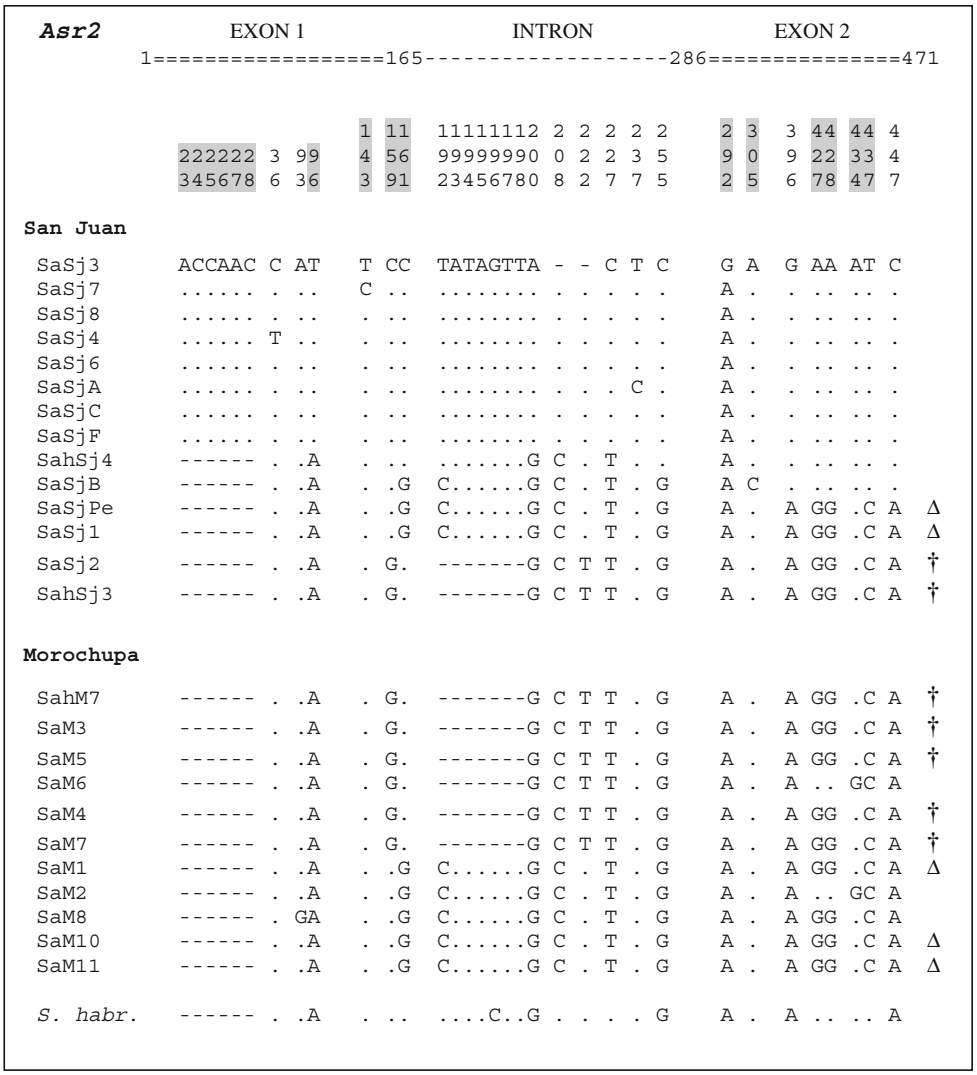

Fig. 2 Summary of nucleotide polymorphism in the aligned sequences of *Asr2* from *Solanum arcanum* (LA0385, San Juan and LA2151, Morochupa). See Fig. 1 for further explanation. Empty triangles (Δ) and black daggers (†) on the right indicate two types of haplotypes shared between populations

Table 1 Summary statistics of nucleotide heterozygosity in *S. chilense* and *S. arcanum*

	S. chilense		*S. arcanum*	
	Asr2	*CT114*	*Asr2*	*CT114*
n	17	12	25	23
L(nt)	466	1165	471	1169
S_{all}	4	13	22	32
syn	3	3	4	7
noncod	0	8	7	18
nonsyn	1	2	11	7
υ	3	4	12	17
θ_w (SD)	0.0025 (0.0015)	0.0037 (0.0017)	0.0126 (0.0048)	0.0074 (0.0028)
π (SD)	0.0010 (0.0006)	0.0020 (0.0012)	0.0172 (0.0013)	0.0034 (0.0011)

n: number of analyzed sequences; *L*(nt): sequence length (number of nucleotides); *S*: number of all (all), synonymous (syn), noncoding (noncod) and non-synonymous (nonsyn) segregating sites; υ: number of distinct haplotypes; θ_w: Watterson's estimator; π: average number of pairwise differences per site (nucleotide diversity); SD: standard deviation

(GenBank accession numbers: AY217010 and AY941663, respectively).

Statistical significance of standard tests of neutrality was assessed with one-sided tests, shifting side according to the sign of the observed value, and based on the empirical distribution of test statistics generated with *DnaSP 4.1* by means of coalescent simulations (10,000 replicates). We conditioned on S (the number of segregating sites) and assumed no recombination (except for Fu's F_s Test) to be conservative. Fu's F_s test is not conservative to recombination, which may generate additional recombinant haplotypes. Hence, we conditioned on an overestimate of the recombination (per gene) parameter (R) when we applied this test to *CT114* in *S. arcanum*, for which more extreme negative values were observed. We used an unlikely high value of R (=5) that largely exceeds the value estimated for our dataset ($R = 0.001$). We employed the *SCANMS* software (Ardell 2004) to correct for multiple testing when Tajima's D test was conducted in a sliding-window mode. We computed values of maxima for Tajima's test statistic D obtained using coalescent simulations (1,000 replicates, conditioned on S), and then tested the significance of positive values by means of a one-sided test.

We also conducted coalescent simulations to investigate whether the observed haplotype structure observed in *Asr2* in *S. arcanum* (see results section) is compatible with the equilibrium neutral model. In other words, what we want to know is the probability of recovering, under neutrality, a haplotype structuring at least as marked as that found in our sample. We employed the *ms* software (Hudson 2002) to run coalescent simulations (1,000 replicates) under the standard neutral model (Hudson 1990), conditioned on sample size and the number of silent segregating sites recorded in our sample. We selected those replicates in which the difference in the number of alleles included in the larger and smaller haplogroups splitting from the common ancestor in the genealogy of the entire sample was equal to or less than the observed (i.e. those cases with a haplotype structure at least as pronounced as the observed in our data) and recorded the number of fixed differences and the average number of nucleotide differences (D_{xy}) between the haplogroups. Then, we contrasted the observed number of silent fixed differences between haplotypic lineages and D_{xy} (for silent sites) with the empirical distribution obtained in the simulations.

Since simulations were generated assuming no recombination, putative recombinant sequences were excluded from the analysis. Conversely, many recombination events would pass undetected and the test is conservative to those cases. In the simulations we considered the indels in sites 208 and 222. This strategy should not necessarily lead to a non-conservative test, since one of these sites would not count for fixed differences between haplogroups. In addition site 192, which would add an extra fixed difference if the indel was considered as a fifth state, was excluded from the analysis since it does not meet the infinite-sites model.

Results

Four polymorphic sites (1 non-synonymous) were detected in *Asr2* and 13 (2 non-synonymous) in *CT114* of *S. chilense* (Fig. 1, Table 1). Interestingly, all polymorphic variants were singletons in both genes. In addition, three indels were found in the intron of *CT114*. Three haplotypes were detected in our *Asr2* sample of *S. chilense* (Table 1) with a frequency distribution of 15:1:1 (Fig. 1a). Polymorphic variants in *CT114* were not randomly distributed among sequences. Indeed, 9 variants were exclusively associated to one haplotype that matched the ancestral state (as compared to the outgroup) in 6 out of the 13 polymorphic sites recorded in this gene (Fig. 1b). In consequence, haplotype frequencies in *CT114* were very skewed (Fig. 1b, Table 1) with a major haplotype and the others represented only once (9:1:1:1). Such skewness in the haplotype frequency spectrum of both genes resulted in values of haplotype diversity that were significantly lower than expected under neutrality ($Hd = 0.228$ and $Hd = 0.455$, for *Asr2* and *CT114*, respectively; $P < 0.02$ for both loci), as assessed by coalescent simulations conditioning on S and assuming no recombination.

Asr2 and *CT114* exhibited contrasting patterns of variation in *S. arcanum*. Nucleotide heterozygosity as measured by Watterson's estimator was 1.7 times greater in *Asr2* than in *CT114*. Twenty-two out of 471 sites (4.7%) were polymorphic in the former, while 32 out of 1,169 sites (2.7%) in the latter (Table 1). Under neutrality, this difference may be accounted for by differences in mutation rates between genes or simply by stochastic coalescent variance. However, the apparent difference in the proportion of polymorphic sites between genes was accompanied by clear differences in the frequency spectrum of variants segregating in each gene. In particular, 22 out of 32 segregating sites were singletons in *CT114*, whereas most variants in *Asr2* were in intermediate frequency. This difference is consistently reflected in the estimates of nucleotide diversity which was 5 times greater in *Asr2* than in *CT114* (Table 1).

Estimates of nucleotide diversity (π) in *Asr2* in *S. arcanum* were 17 times greater (pooling across sampling localities) than in *S. chilense*, whereas estimates of nucleotide diversity in *CT114* were more alike between species (Table 1). Likewise, estimates of θ_w for *Asr2* and *CT114* were 5 and 2 times greater in *S. arcanum* than in *S. chilense* (Table 1). However, the HKA test, which is effective in detecting departures from the neutral model due to

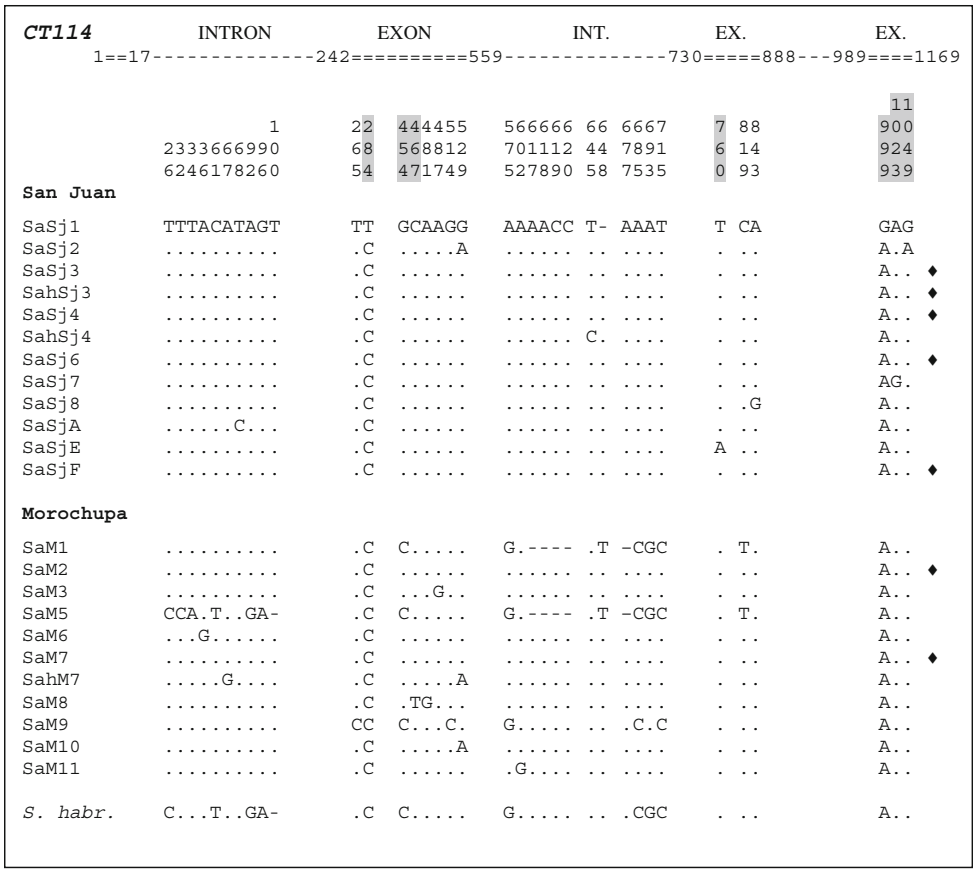

Fig. 3 Summary of nucleotide polymorphism in the aligned sequences of *CT114* from *Solanum arcanum* (LA0385, San Juan and LA2151, Morochupa). See Fig. 1 for further explanation. Black diamonds (◆) on the right indicate haplotypes shared between populations

selection and/or differences in mutation rates between the two genes, did not yield significant results neither in *S. arcanum* nor *S. chilense*.

Besides the 22 nucleotide polymorphisms found in the *Asr2* gene in *S. arcanum*, we detected the presence of 4 indels, one in the first exon and the rest in the intron (Fig. 2). Eleven nucleotide polymorphisms, including three singletons, were replacement changes (the indel affecting two amino acids was considered as a single non-synonymous polymorphic site in further analyses) in this species. Polymorphic variants in *S. arcanum* appear to be clustered into two haplotypic groups and two presumptive recombinant haplotypes (*SaSjB* and *SahSj4*) which exhibit combinations of variants of the main haplogroups (Fig. 2).

Populations of *S. arcanum* exhibited a certain degree of genetic differentiation for both loci due to differences in the frequency spectrum of polymorphic variants ($K_{st*} = 0.169$, $P < 0.001$ and $K_{st*} = 0.028$, $P < 0.05$, for *Asr2* and *CT114*, respectively) which might be responsible for the haplotype structure observed in *Asr2*. However, we did not find fixed differences between populations and several completely identical haplotypes were common to both populations (Figs. 2 and 3). These two features could be ascribed to ancestral shared polymorphisms and/or to moderate levels of gene flow.

Neutrality tests

Tajima's *D* and Fu and Li's *F* were negative and significant in *Asr2* in *S. chilense*, mainly due to synonymous variation (Table 2). These tests were also negative for *CT114*, though Tajima's *D* was significant only for all and silent variation (Table 2). These results point to an overall excess of rare variants in both genes. Fu's F_s was not significant neither in *Asr2* nor in *CT114* (Table 2), suggesting that the number of haplotypes conditional on the number of segregating sites observed, did not depart from neutral expectations. In contrast, Fay and Wu's test yielded significant results for *CT114*, except for non-synonymous sites, pointing to an excess of high frequency derived silent variants. The non-significant value of this test for non-synonymous sites, together with the negative values of Tajima's *D* and Fu and Li's *F* tests, may be accounted for by two singletons derived replacement changes (Fig. 1b).

In *S. arcanum*, Tajima's and Fu and Li's tests were not significant in *Asr2*, though the test statistics were invariably positive (Table 2) even when both localities (San Juan and Morochupa) were analyzed separately (not shown). The highest value of Tajima's *D*, which corresponds to silent variation, was marginally significant ($P < 0.06$), pointing

Table 2 Tests of neutrality for *Asr2* and *CT114* in *S. chilense* and *S. arcanum*

	S. chilense		*S. arcanum*	
	Asr2	*CT114*	*Asr2*	*CT114*
Tajima's *D*				
D_{all}	−1.843*	−1.970**	1.344	−2.089**
D_{syn}	−1.706	−1.629	0.257	−1.765**
D_{sil}	−1.706	−1.911**	1.512	−1.996**
D_{nonsyn}	−1.164	−1.451	0.960	−1.896**
Fu and Li's *F*				
F_{all}	−2.958*	−1.591	0.502	−2.478*
F_{syn}	−2.661*	−0.490	−0.729	−2.482*
F_{sil}	−2.661*	−1.174	0.536	−1.901
F_{nonsyn}	−1.674	−2.100	0.326	−3.125**
Fu's F_s				
F_s^{all}	−0.445	1.272	0.131	−10.287**
F_s^{syn}	−0.963	1.054	0.589	−3.922*
F_s^{sil}	−0.963	2.269	0.825	−5.373*
F_s^{nonsyn}	−0.748	−1.325	−0.361	−4.322**
Fay and Wu's *H*				
H_{all}	0.441	−8.485**	0.853	−13.787**
H_{syn}	0.331	−3.182**	0.487	0.767
H_{sil}	0.331	−8.788**	−0.140	−12.941**
H_{nonsyn}	0.110	0.303	0.993	−0.846

Test statistics are given for all (all), synonymous (syn), synonymous plus noncoding (sil) and non-synonymous (nonsyn) variation

* $P < 0.05$; ** $P < 0.02$; *** $P < 0.001$

to an excess of silent variants in intermediate frequency. When Tajima's test was applied to *Asr2* using a sliding window approach (100 bp of window length and 40 bp each step size), two regions (161–260 and 361–460) exhibited a significant *D* (all variable sites considered) (Fig. 4). However, after applying the correction for multiple tests (as implemented in *SCANMS* software) Tajima's *D* for these regions became marginally significant. Interestingly, these regions contain 3 non-synonymous polymorphisms segregating in intermediate frequency and two of them (161 and 427–428) are polymorphic in both populations (Fig. 2). However, the first candidate fragment is far from being centered on the replacement substitution.

In contrast, Tajima's and Fu and Li's tests statistics were negative and significant in *CT114*, pointing to an overall excess of singletons (Table 2). Similar results were obtained for each population separately (not shown). Fu's F_s was also significant reflecting an excess of haplotypes relative to the expectations under mutation-drift equilibrium (Table 2). As in *S. chilense*, Fay and Wu's test yielded a negative and significant value for *CT114* (mostly due to variation in noncoding regions), indicating an excess of derived variants in high frequency. This result seems to be particularly caused by the frequency spectrum at sites 26, 61, 92, 96, 100, 454, 575, 685, 693 and 715, for which haplotype *SaM5* fully matches the ancestral states (Table 2, Fig. 3).

Haplotype structure in *Asr2*

Genealogical relationships of *Asr2* haplotypes, illustrated as a haplotype network, are shown in Fig. 5. The skewed frequency of *Asr2* haplotypes in *S. chilense* is evident in the network structure (Fig. 5a). The major haplotype is represented by 15 sequences. Two branches, with two mutations (singletons) each, connect the major haplotype to the two remaining sequences of the sample. It is worth to note that three fixed changes, which affect the encoded *ASR2* protein (two replacements and one indel), were

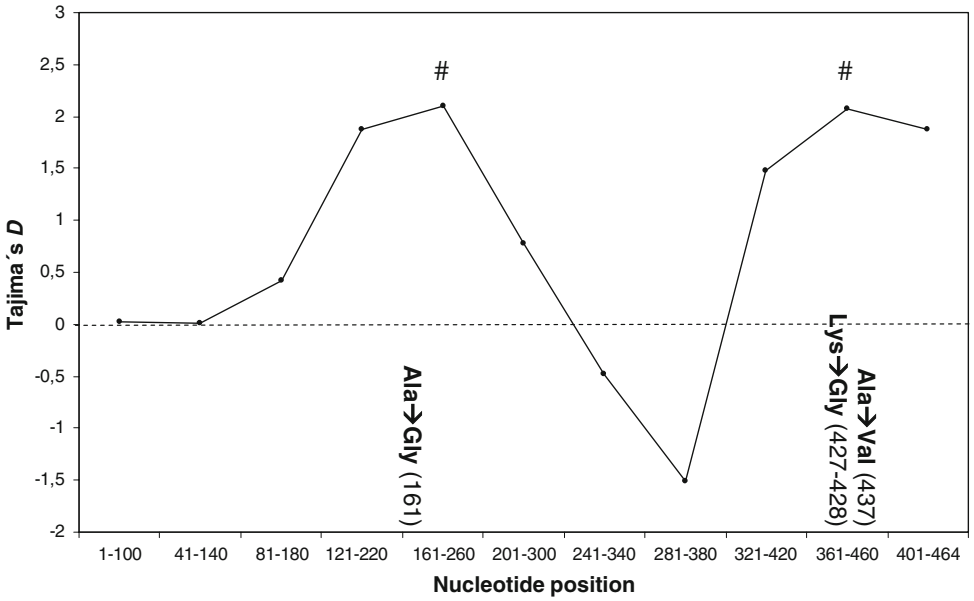

Fig. 4 A sliding window plot of Tajima's *D* test applied to all variation of *Asr2* sequences from *S. arcanum*. Window size spans 100 nt and each step size is 40 nt. $D_{all} = 2.100$ and $D_{all} = 2.079$ for regions comprised between sites 161–260 and 361–460, respectively ($^{\#}$ $0.09 > P > 0.05$, after correction for multiple tests). The amino acid replacements are showed in the corresponding gene regions (the exact nucleotide positions of those replacements are indicated in parentheses)

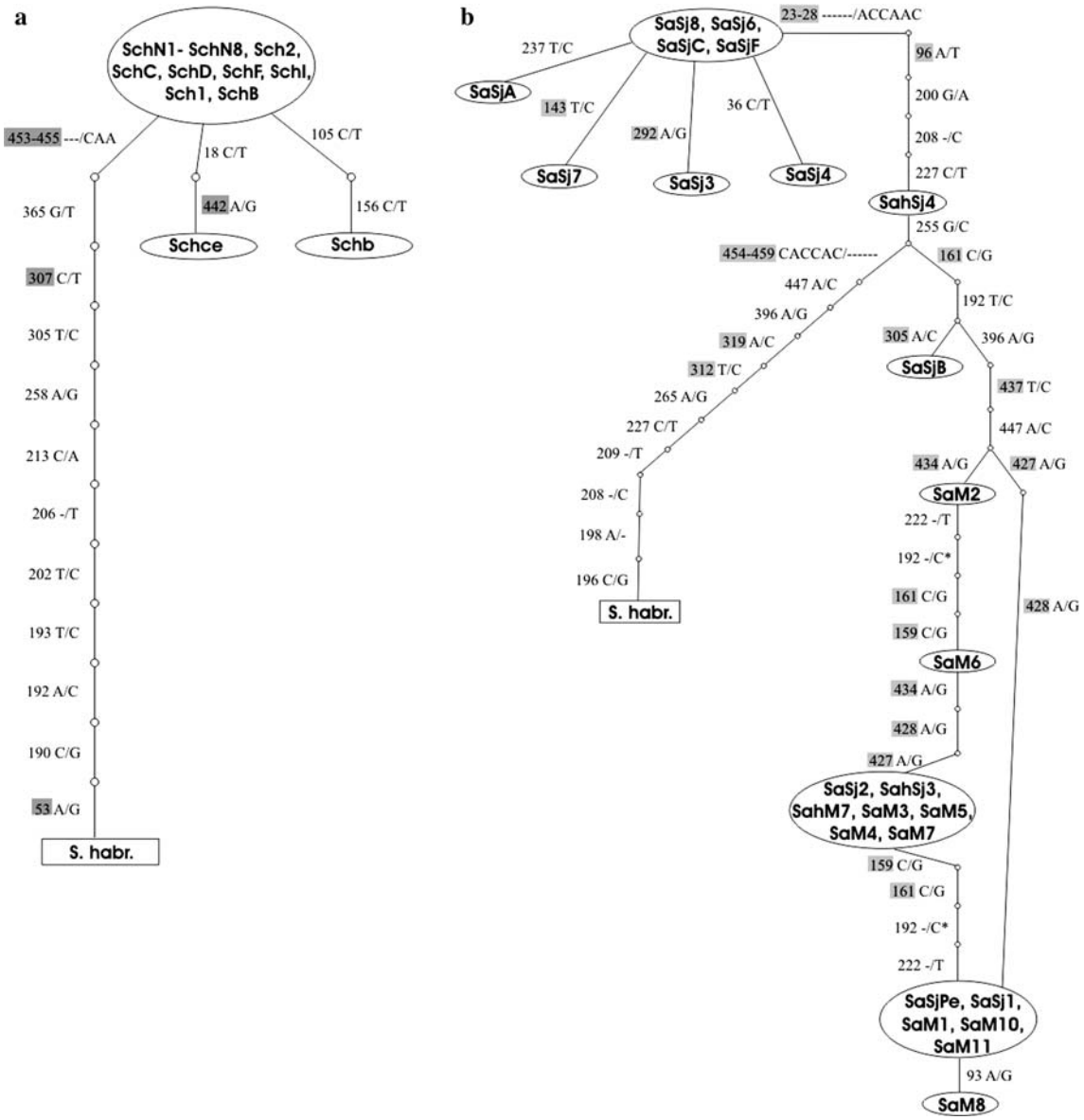

Fig. 5 Haplotype networks for *Asr2* sequences of *S. chilense* (**a**) and *S. arcanum* (**b**) inferred by the statistical parsimony method. Each branch denotes a single mutational step. Ovals correspond to one or several identical haplotypes. Empty circles represent hypothetical haplotypes or "missing intermediates". The outgroup (*S. habrochaites*) is indicated by a rectangle. Each nucleotide change, preceded by its position, is shown along branches in the form of ancestral/derived states according to the outgroup (* change polarity not determined). For branches with more than a single mutational step, changes were ordered arbitrarily. Non-synonymous changes are grey shaded

detected as mutations occurring in the branch connecting to the outgroup (Fig. 5a).

Segregating variants in *Asr2* do not seem to be evenly distributed among sequences in *S. arcanum* (Fig. 2) as mentioned earlier. In effect, a close examination of the genealogy presented in Fig. 5b reflects the presence of two fairly well differentiated haplotypic groups, diverging from a common node that connects to the outgroup. Consistently, a neighbor-joining tree (using *p-distance* algorithm) yielded a qualitatively equivalent structure with two major haplogroups (not shown). Indeed, K_{st*} statistic (Hudson et al. 1992) yielded a highly significant value ($K_{st*} = 0.389$, $P < 0.001$), just confirming the presence of two clearly defined haplogroups. Segregating sites 96, 200, 208, 227, 255, 396, 427–428, 437, 447 and the indel spanning positions 23–28, seem to be responsible for such haplotypic structure (Fig. 2). In contrast, and as expected from the observed distribution of polymorphic variants (Fig. 3), the genealogy of *CT114* did not reveal such a haplotypic structure (not shown).

There are two important features in the generalogy of *Asr2* haplotypes. First, one of the main *Asr2* haplogroups is composed exclusively by haplotypes sampled in San Juan (Fig. 5b), and, second, the rest of the haplotypes do not cluster according to their geographic origin. In fact, sequences sampled in San Juan and Morochupa appear, in general, intermingled (Fig. 5b). The latter observation is not unexpected given the extensive sharing of completely identical haplotypes between sampling sites (Fig. 2).

As mentioned earlier, the combinations of allelic variants present in *SaSjB* and *SahSj4* haplotypes are presumably the product of intragenic recombination between haplogroups (Fig. 2). In fact, at least one recombination event between sites 161 and 395 ($R_m = 1$) can be detected using the four-gamete test (Hudson and Kaplan 1985).

We also examined the structure of the haplotype network searching for recombination signals following Crandall and Templeton (1999). Putative recombinants are located closer to the node coming from the outgroup than the rest of the haplotypes, and connected to both haplogroups by rather long branches (Fig. 5b), so that if confidence limits are further constrained up to 99% (3 mutational steps) both haplotypes are the only ingroup sequences that segregate from the others as isolated networks. We further tested for physical clustering of mutations present in each pair of branches coming together at each putative recombinant haplotype (Crandall and Templeton 1999). This analysis yielded a hypergeometric tail probability of $P_{[k=3]} = 0.029$ for *SaSjB*, indicating a significantly non-random spatial pattern along the sequence. Although the spatial pattern of *SahSj4* is also quite suggestive, a non-significant value of hypergeometric tail probability ($P_{[k=3]} = 0.2651$) was obtained.

Coalescent simulations

From the empirical distribution of random generated genealogies conditioned on the number of sequences ($n = 23$) and the number of silent segregating sites ($S = 10$), we found that the probability of having two lineages of $n - x$ and x sequences each (with x varying from 8 to 11) with 6 or more fixed differences between lineages was 0.044. Likewise, the probability of obtaining a value of D_{xy} equal to or higher than the observed ($D_{xy} = 6.85$) was also unexpected under neutrality $\left(P_{[D_{xy} \geq 6.85]} = 0.049\right)$.

In order to control for population structure effects that may produce or accentuate the dimorphic haplotypic structure observed in *S. arcanum*, we conducted similar coalescent simulations considering only the sequences sampled in San Juan, since only in this locality we found the two main haplogroups. The probability of having two lineages of $n' - x'$ and x' sequences (where $x' = 4$–6) with the observed number (=6) or more fixed differences between lineages was 0.073 and the probability of obtaining a D_{xy} value equal to or higher than the observed (=6.75) was 0.064. In this case probability values are low though only marginally significant. Thus we may conclude that, although still suggestive, the observed haplotypic structure is compatible with neutrality if we only consider individuals sampled in San Juan.

Discussion

In a previous work, we have shown that *Asr2*, one of the members of the *Asr* family in tomato that is expressed under water-stress conditions, has experienced an acceleration in the rate of replacement substitutions in the lineages leading to tomato species known to live in dry and semiarid-mesic environments, namely *S. chilense* and *S. arcanum*, respectively (Frankel et al. 2003). However, the artifactual effects of false positives, potentially obtained through the underlying PAML method, cannot be completely ruled out. Hence, our analysis of polymorphism distribution in this drought-responsive gene offers a complementary approach to test for adaptation at the molecular level in these species.

Our comparative study revealed qualitatively and quantitatively different patterns of variation in *Asr2* between *S. chilense* and *S. arcanum*. These results suggest, at first sight, that this gene may have been the target of different selective regimes in these species. On one hand, the extremely low values of nucleotide variation in *Asr2* in *S. chilense*, which dwells in the desert of Atacama (one of the driest regions of the world), are compatible with the expectations for a gene that has passed through a relatively recent selective sweep (Maynard Smith and Haigh 1974). Nevertheless, this result alone cannot be considered as conclusive evidence of positive selection since a reduction in the effective population size might also cause a concomitant genome-wide reduction of variation. In fact, patterns of nucleotide variation in the non-linked *CT114* gene point in this direction, as it also exhibited low levels of variation and a significant excess of rare variants.

To distinguish between positive selection and a severe bottleneck we need information of nucleotide variation in unlinked loci. Estimates of nucleotide polymorphism have been recently reported in *S. chilense* for several genes (Roselius et al. 2005; Städler et al., 2005; Arunyawat et al. 2007). The results reported so far show that levels of within population variation vary considerably across loci, and that the population studied in this paper (Ayavirí) exhibits the lowest estimates of heterozygosity. Thus, it may be argued that this population has suffered a bottleneck or it has been isolated as a deme for a relatively long time with a subsequent loss of genetic diversity and pronounced changes in the polymorphism frequency spectrum (Arunyawat et al. 2007 and present paper).

Our survey revealed that patterns of nucleotide variation in *Asr2* (and also *CT114*) in the *S. chilense* population of Ayavirí differ greatly from genome-wide trends reported in other studies of this species. For instance, studies of several loci yielded positive and significant values of Tajima's and Fu and Li's tests (Roselius et al. 2005; Städler et al. 2005; Arunyawat et al. 2007) pointing to a genome-wide excess of variants in intermediate frequency, irrespective of the level of variation detected in each locus, which is in sharp contrast with the negative Tajima's and Fu and Li's values obtained for *Asr2* in the present study (Table 2). Despite these differences might be attributed to different sampling schemes, a historical bottleneck or positive selection cannot be ruled out as plausible explanations of the patterns observed in Ayavirí.

Although the results of both Tajima's and Fu and Li's tests were compatible with a selective sweep in *Asr2*, the test of Fay and Wu, which is particularly suitable in detecting positive selection, failed to reject the null hypothesis of selective neutrality. Nevertheless, if an advantageous mutation is driven to rapid fixation producing a complete sweep of linked neutral variation, the test statistic (*H*) would be non-significant since ancestral variants would pass largely undetected in the ingroup (Fay and Wu 2000). Even, if the selected allele did not reach complete fixation but increased considerably in frequency, linked neutral derived variants would be dragged so that ancestral variants of those sites would be less likely to be sampled. Alternatively, if low mutation rates are assumed the beneficial mutation could have been associated with no derived neutral alleles during the sweep, thus hitchhiking only ancestral variants. To assess these possibilities we looked for fixed derived variants in *Asr2* between *S. chilense* and the outgroup. We found two aminoacid substitutions (Fig. 5a) between *S. chilense* and *S. habrochaites* at positions 53 (Lysine/Arginine), and 307 (Proline/Leucine), those mapped to the terminal branch of *S. chilense* (Frankel et al. 2003). The latter implies a radical substitution of a highly conserved amino acid site and would probably result in relevant functional differences.

Our survey of nucleotide variation in *Asr2* in *S. arcanum* revealed a fairly high number of intermediate frequency variants when sequences sampled in both populations were analyzed together (Fig. 4, Table 2). In addition, the genealogy uncovered the presence of two well defined haplotypic classes (Fig. 5b). These features contrast sharply with the pattern detected in *CT114*. Thus, a single process with genome-wide effects cannot simultaneously account for these contrasting patterns; however different selective regimes could indeed produce such difference.

Contrary to our suspicion, based on a study in *S. peruvianum* (Städler et al. 2005), a close relative of *S. arcanum*, showing that nucleotide variation in *CT114* did not depart from neutrality, our study revealed deviations from neutral expectations in *S. arcanum*. In fact, patterns of variation in *CT114* exhibited signatures compatible with positive selection as suggested by the results of Tajima's and Fay and Wu's tests. These results are not surprising given the potentiality of selection on the gene's function. Albeit the negative and significant Tajima's test statistic could reflect a population expansion, Fay and Wu test is less sensitive to demographic changes and false positives could only be expected from other factors such as mis-orientation of ancestral states resulting from multiple hits in the outgroup (Baudry and Depaulis 2003). In any case, these results cast doubts on the use of this locus as a potential control for demographic effects. Unfortunately, the lack of studies of genetic variation in *S. arcanum* limits our possibilities to accurately infer the evolutionary forces that shaped variation in *Asr2*.

Tajima's test uncovered a considerable number of intermediate frequency variants, although it was only marginally significant (after correction for multiple tests), in two regions of *Asr2* in *S. arcanum* when it was applied using a sliding window approach (Fig. 4). These regions involve replacement changes, which, incidentally, are in intermediate frequency, at positions 161, 427–428 (same codon), and 437. These three amino acid polymorphisms seem to be responsible for the accelerated rate of non-synonymous substitutions in the terminal branch of *S. arcanum* as reported in Frankel et al. (2003). One of these replacements (Alanine/Valine) implies a size change that may lead to a functionally altered protein (Frankel et al. 2003). Likewise, a drastic modification in the *ASR2* protein may result from a Lysine/Glycine change which brings about an alteration of the residue polarity and the charge of the polypeptide. These non-conservative amino acid replacements may be considered as likely candidates for targets of balancing selection.

The presence of a haplotypic structure has often been interpreted as the signature of balancing selection even when standard tests of neutrality do not reject the null hypothesis. For instance, two significantly differentiated allelic classes were interpreted as evidence of balancing selection in the *Adh* locus of *Arabidopsis thaliana* (Hanfstingl et al. 1994) and the *PgiC* locus of *Leavenworthia* (Filatov and Charlesworth 1999). In both studies, the authors noted that the lack of statistical power may be responsible for the failure to reject the null hypothesis. Since the tests employed assume no recombination, a reduced statistical power is expected in regions where intralocus recombination is likely to occur, as in the cases of *PgiC* and *Adh* (Filatov and Charlesworth 1999; Hanfstingl et al. 1994). Therefore, the lack of significance of standard neutrality tests in *Asr2* in *S. arcanum*, in which evidence of recombination was found, cannot be considered as strong evidence against balancing selection,

particularly for Fu's test which is based on the number of haplotypes. In fact, Tajima's *D* (considering all variation) for *Asr2* in *S. arcanum*, conditioned on the number of segregating sites and the value of the recombination parameter estimated from our dataset ($R = 5.4$), was significant at the 0.05 level.

However, non selective explanations can also account for dimorphic patterns in haplotype structure. For instance, Aguadé (2001) detected a similar pattern of variation in *FAH1* and *F3H* in *A. thaliana*. This model species is essentially self-compatible and heterozygous individuals are notably scarce, making effective recombination extremely rare. In this context, Aguadé (2001) showed that a dimorphic haplotypic structure could also be consistent with a neutral model without recombination in this species.

Interestingly, coalescent simulations showed that the observed number of fixed differences and D_{xy} between haplogroups of *Asr2* in *S. arcanum* are unlikely under an equilibrium neutral model with no recombination for the entire dataset (pooling San Juan and Morochupa populations). However, when only individuals sampled in San Juan were considered coalescent simulations yielded only marginally significant results, suggesting that the observed haplotypic structure may be accounted for by a "species-wide sampling" effect, like a natural population admixture of formerly isolated populations. In fact, the two diverged haplotypes only coexist in one of the populations sampled (San Juan) while in the other (Morochupa) only one of the haplogroups has been detected, suggesting that the former may be in the area of secondary contact.

Testing for physical clustering of mutations (Crandall and Templeton 1999) gave strong support in favor of the occurrence of intralocus recombination along the sequence of *SaSjB*, while the evidence was not so clear for the other putative recombinant *SahSj4,* in spite of exhibiting a suggestive spatial pattern. However, this result could be explained by gene conversion or, alternatively, by a double recombination event (involving sites 159 and 161; Fig. 2). This possibility is not unlikely since recombinants can, not only, arise naturally but also as a PCR artifact (Bradley and Hillis 1997). Indeed, no additional mutations, i.e. other than those occurring in the putative parental branches, leading to *SahSj4* are present (Fig. 5b, Crandall and Templeton 1999).

In summary, we present an example of two closely related tomato species with different patterns of nucleotide variation in a gene that plays an important role to alleviate restricted water availability. In both species we observed departures from neutral expectations. However, further research is necessary to elucidate whether these apparent signatures of selection may be ascribed to selective forces associated to unique climatic features: one climate is rather constant whereas the other is seasonally fluctuating, or to demographic explanations based on population bottlenecks in *S. chilense* and population admixture of divergent populations in *S. arcanum*.

Acknowledgements We thank the Tomato Genetics Resource Center (University of California, Davis, USA) for providing the seeds of the tomato wild populations and Gustavo Gudesblat for his valuable help with the experimental work in the laboratory. We also wish to thank Romina Piccinali and David Ardell for their help with technical issues of software usage, and two anonymous reviewers for their constructive comments and criticisms on a previous version of this manuscript. This work was supported by grants from Universidad de Buenos Aires (UBA), Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) and Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Argentina. M.G. held a fellowship from UBA. N.F. held a fellowship from CONICET. N.D.I. and E.H. are members of the Carrera del Investigador Científico, CONICET, Argentina.

References

Aguadé M (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the FAH1 and F3H genes, in *Arabidopsis thaliana*. Mol Biol Evol 18:1–9

Ardell DH (2004) SCANMS: adjusting for multiple comparisons in sliding window neutrality tests. Bioinformatics 20:1986–1988. doi:10.1093/bioinformatics/bth187

Arunyawat U, Stephan W, Städler T (2007) Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. Mol Biol Evol 24:2310–2322. doi:10.1093/molbev/msm162

Baudry E, Depaulis F (2003) Effect of misoriented sites on neutrality tests with outgroup. Genetics 165:1619–1622

Bermudez-Moretti M, Maskin L, Gudesblat G, Correa-García S, Iusem ND (2006) Asr1, a stress-induced tomato protein, protects yeast from osmotic stress. Physiol Plant 127:111–118. doi:10.1111/j.1399-3054.2006.00664.x

Bradley RD, Hillis DM (1997) Recombinant DNA sequences generated by PCR amplification. Mol Biol Evol 14:592–593

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. doi:10.1105/tpc.013854

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1659. doi:10.1046/j.1365-294x.2000.01020.x

Crandall KA, Templeton AR (1999) Statistical methods for detecting recombination. In: Crandall KA (ed) The evolution of HIV. The Johns Hopkins University Press, pp 153–176

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Filatov DA, Charlesworth D (1999) DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia PgiC* locus. Genetics 153:1423–1434

Frankel N, Hasson E, Iusem ND, Rossi MS (2003) Adaptive evolution of water stress-induced gene *Asr2* in *Lycopersicon* species dwelling in arid habitats. Mol Biol Evol 20:1955–1962. doi:10.1093/molbev/msg214

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Ganal MW, Czihal R, Hannappel U, Kloos DU, Polley A, Ling HQ (1998) Sequencing of cDNA clones from the genetic map of tomato (*Lycopersicon esculentum*). Genome Res 8:842–847

Hanfstingl U, Berry A, Kellogg EA, Costa JT, Rudiger W, Ausubel FM (1994) Haplotypic divergence coupled with lack of diversity at the Arabidopsis thaliana alcohol dehydrogenase locus: roles for both balancing and directional selection? Genetics 138: 811–828

Hoekstra FA, Golovina EA, Buitink J (2001) Mechanisms of plant desiccation tolerance. Trends Plant Sci 6:431–438. doi: 10.1016/S1360-1385(01)02052-0

Hudson RR (1990) Gene genealogies and the coalescent process. Oxf Surv Evol Biol 7:1–44

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338. doi: 10.1093/bioinformatics/18.2.337

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Hudson RR, Boos D, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Biol Evol 9:138–151

Iusem ND, Bartholomew DM, Hitz WD, Scolnik PA (1993) Tomato transcript induced in water stress and ripening. Plant Physiol 102:1353–1354. doi:10.1104/pp. 102.4.1353

Jeanneau M, Gerentes D, Foueillassar X, Zivy M, Vidal J, Toppan A et al (2002) Improvement of drought tolerance in maize: towards the functional validation of the Zm-Asr1 gene and increase of water use efficiency by over-expressing C4-PEPC. Biochimie 84:1127–1135. doi:10.1016/S0300-9084(02)00024-X

Kalifa Y, Gilad A, Konrad Z, Zaccai M, Scolnik PA, Bar-Zvi D (2004a) The water- and salt-stress regulated Asr1 gene encodes a zinc-dependent DNA-binding protein. Biochem J 381:373–378. doi:10.1042/BJ20031800

Kalifa Y, Perlson E, Gilad A, Konrad Z, Scolnik PA, Bar-Zvi D (2004b) Over-expression of the water and salt stress-regulated Asr1 gene confers an increased salt tolerance. Plant Cell Environ 27:1459–1468. doi:10.1111/j.1365-3040.2004.01251.x

Kane NC, Rieseberg LH (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. Genetics 175:1823–1834. doi: 10.1534/genetics.106.067728

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK

Maskin L, Gudesblat GE, Moreno JE, Carrari FO, Frankel N, Sambade A et al (2001) Differential expression of the members of the *Asr* gene family in tomato (*Lycopersicon esculentum*). Plant Sci 161:739–746. doi:10.1016/S0168-9452(01)00464-2

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35

Mustonen V, Lässig M (2007) Adaptations to fluctuating selection in *Drosophila*. Proc Natl Acad Sci USA 104:2277–2282. doi: 10.1073/pnas.0607105104

Peralta IE, Spooner DM (2001) Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes. Am J Bot 88:1888–1902. doi:10.2307/3558365

Peters S, Mundree SG, Thomson JA, Farrant JM, Keller F (2007) Protection mechanisms in the resurrection plant Xerophyta viscosa (Baker): both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit. J Exp Bot 58:1947–1956. doi:10.1093/jxb/erm056

Qiu YL, Palmer JD (1999) Phylogeny of early land plants: insights from genes and genomes. Trends Plant Sci 4:26–30. doi: 10.1016/S1360-1385(98)01361-2

Roselius K, Stephan W, Städler T (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics 171:753–763. doi:10.1534/genetics.105.043877

Rossi MM, Iusem ND (1994) Tomato genomic clone homologous to a gene encoding an ABA-induced protein. Plant Physiol 104:1073–1074. doi:10.1104/pp. 104.3.1073

Rossi MM, Lijavetzky D, Bernacchi D, Hopp HE, Iusem ND (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

Rozas J, Sánchez-Del Barrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497. doi: 10.1093/bioinformatics/btg359

Städler T, Roselius K, Stephan W (2005) Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. Evol Int J Org Evol 59:1268–1279

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis. Genetics 123:585–595

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619–633

Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften 94:791–812. doi: 10.1007/s00114-007-0254-y

Urao T, Yakubov B, Satoh R, Yamaguchi-Shinozaki K, Seki M, Hirayama T et al (1999) A transmembrane hybrid-type histidine kinase in Arabidopsis functions as an osmosensor. Plant Cell 11:1743–1754

Wall JD, Andolfatto P, Przeworski M (2002) Testing models of selection and demography in Drosophila simulans. Genetics 162:203–216

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:188–193

Yang CY, Chen YC, Jauh GY, Wang CS (2005) A Lily ASR protein involves abscisic acid signaling and confers drought and salt resistance in Arabidopsis. Plant Physiol 139:836–846