



Cite this article: Böndel KB, Nosenko T, Stephan W. 2018 Signatures of natural selection in abiotic stress-responsive genes of *Solanum chilense*. *R. Soc. open sci.* **5**: 171198. <http://dx.doi.org/10.1098/rsos.171198>

Received: 21 August 2017

Accepted: 4 December 2017

Subject Category:

Biology (whole organism)

Subject Areas:

evolution

Keywords:

abiotic stress, local adaptation, plant population genetics, *Solanum chilense*

Author for correspondence:

Katharina B. Böndel

e-mail: katharina.boendel@ed.ac.uk

[†]Present address: Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, Scotland, UK.

[‡]Present address: PGSB—Plant Genome and Systems Biology, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Neuherberg, Germany.

[§]Present address: Museum für Naturkunde, Berlin, Germany.

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3959986>.

Signatures of natural selection in abiotic stress-responsive genes of *Solanum chilense*

Katharina B. Böndel[†], Tetyana Nosenko[‡] and Wolfgang Stephan[§]

Section of Evolutionary Biology, Department of Biology II, Ludwig-Maximilian University of Munich, Planegg-Martinsried, Germany

KBB, 0000-0001-8919-729X

Environmental conditions are strong selective forces, which may influence adaptation and speciation. The wild tomato species *Solanum chilense*, native to South America, is exposed to a range of abiotic stress factors. To identify signatures of natural selection and local adaptation, we analysed 16 genes involved in the abiotic stress response and compared the results to a set of reference genes in 23 populations across the entire species range. The abiotic stress-responsive genes are characterized by elevated nonsynonymous nucleotide diversity and divergence. We detected signatures of positive selection in several abiotic stress-responsive genes on both the population and species levels. Local adaptation to abiotic stresses is particularly apparent at the boundary of the species distribution in populations from coastal low-altitude and mountainous high-altitude regions.

1. Introduction

Wild tomato species (*Solanum* sect. *Lycopersicon*) are native in South America and occur in diverse habitats, including rainforests in Ecuador, hyperarid regions in the Chilean Atacama Desert, coastal regions and high-altitude regions in the Andes [1,2]. Abiotic environmental conditions contribute to the geographical distribution of these species [3]. Many wild tomato species show distinct morphological adaptations to their natural habitat [2,3]. For instance, *S. chilense*, the wild tomato species that can grow under the driest and coldest conditions [2], is characterized by a deep root system that may allow the utilization of deep groundwater, enhancing drought tolerance [2,3].

Abiotic stress response genes and pathways have been studied extensively in *Arabidopsis* and rice [4,5]. Briefly, receptor or sensory genes receive the stress signal and forward it to

transcription factors, e.g. through the plant hormone abscisic acid (ABA)-dependent pathway. Transcription factors or regulatory genes bind to *cis*-acting elements of functional genes located downstream thereby activating their transcription. Although abiotic stress response pathways are not entirely understood in the genus *Solanum*, many abiotic stress-responsive genes have been identified using a range of transcriptomic approaches including cDNA libraries and microarrays. Many of these data were generated for plants exposed to either abiotic stress or ABA [6–8].

As abiotic stress-responsive genes are important to ensure survival, natural selection is expected to influence their evolution. In fact, signatures of natural selection were reported for several abiotic stress-responsive genes in the drought-tolerant wild tomato species *S. chilense* and its closely related sister species *S. peruvianum* [9–12]. Signatures suggesting positive selection were detected in the key transcription factor of the cold-signalling pathway, *ICE1* [11]. Members of transcription factor gene families (*Asr*, ABA/water stress/ripening induced; *CBF*, C-repeat binding factor) evolve under positive directional (*Asr4*), purifying (*Asr1*, *CBF3*) and balancing (*CBF2*) selection [9,10]. The key gene of the ABA biosynthesis, *LeNCED1* (9-*cis*-epoxy-carotenoid dioxygenase), was reported to evolve under purifying selection, while a functional gene, the dehydrin *pLC30-15*, exhibits a haplotypic pattern possibly explained by diversifying selection [12]. These signatures were found on both the species (*Asr1*, *CBF3*, *ICE1*, *LeNCED1*) and the population (*Asr4*, *CBF2*, *pLC30-15*) level, implying either species-wide or population-specific selection, i.e. local adaptation.

In the present study, we focus on the evolution of genes and pathways involved in the abiotic stress response in *S. chilense*. This species is an ideal candidate for adaptation studies as it exhibits a patchy distribution and occurs across a range of environmental conditions including hyperarid desert, cold high-altitude and high soil salinity coastal regions in southern Peru and northern Chile [1,13]. We use information from our previous study on the demography of *S. chilense*, in which we generated a multi-gene/multi-population dataset for abiotic stress-responsive genes (candidate genes) and reference genes [14]. Analyses of population averages revealed a clinal pattern indicating a north–south colonization and a substructure consisting of four population groups. In the present study, we employ standard population genetics to explore whether natural selection is acting upon these genes and whether there are differences between the different populations, i.e. whether we can detect signatures of local adaptation. Furthermore, we investigated if and how natural selection acts on the pathways the genes are organized in. Theory predicts that genes with higher connectivity in a gene network, i.e. with many interacting genes, should be under higher constraint than genes with fewer interactions, and thus are expected to evolve under purifying selection [15]. A recent study on the *Arabidopsis* protein interactome has supported this prediction; it showed that genes acting in the centre of a gene network are under stronger evolutionary constraint than genes acting in the periphery of a gene network [16]. On the other hand, one might also argue that a gene with many interactions has more influence and thus, if altered, might accelerate evolution more than a gene with only few interactions. We find an excess of nonsynonymous single-nucleotide polymorphisms (SNPs) and divergence in the abiotic stress-responsive genes in comparison to the reference genes. Furthermore, we identify genes that evolve under positive directional or balancing selection within populations from the coastal low-altitude region and also from high altitudes, suggesting local adaptation.

2. Material and methods

2.1. The *Solanum chilense* populations

Solanum chilense is a diploid ($2n = 24$), self-incompatible perennial plant, native to a multitude of different habitats in South America ranging from mesic areas in southern Peru to hyperarid areas in the Atacama Desert and the high altitudes in the Andes in northern Chile [1,2,13]. The Tomato Genetics Resource Center (TGRC) at the University of California at Davis (<http://tgrc.ucdavis.edu> (accessed 16 August 2017)) has a comprehensive collection of wild tomato species and their relatives. Based on available information about sampling locations and population and habitat characteristics, including population sizes and sample sizes, 23 populations (figure 1, electronic supplementary material, table S1) representing the entire species and habitat range were chosen for analyses. One population of *S. ochranthum*, LA2682, was chosen for outgroup comparisons. *Solanum ochranthum* is closely related to the wild tomato clade with an estimated split time of around 5.59 million years [17]. According to our previous study on the species demography these 23 *S. chilense* populations form four population groups: a northern group (NG; including LA1930 and LA3784), a southern high-altitude group (SHG; LA2880, LA4118, LA4119,

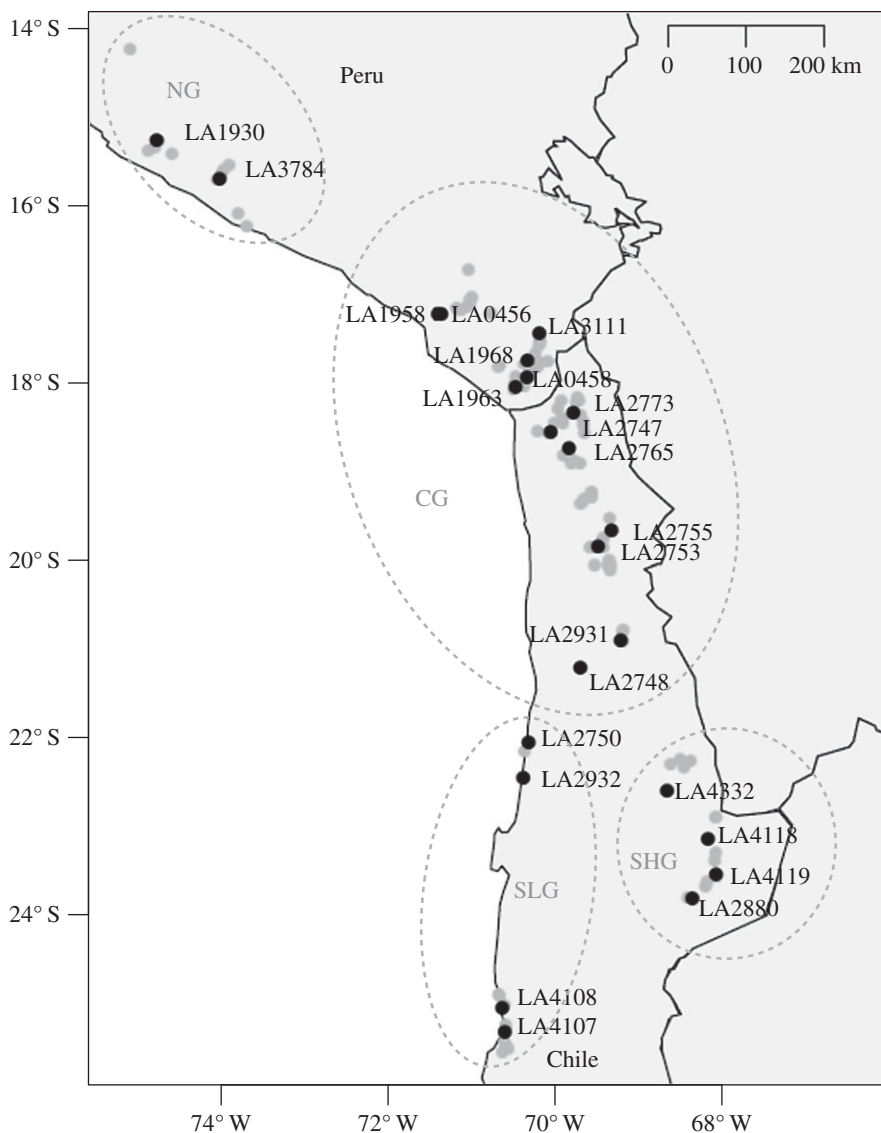


Figure 1. Geographical distribution of the *S. chilense* populations. Map with all *S. chilense* populations sampled by the TGRC (grey circles), the 23 *S. chilense* populations of this study (black circles), and the four population groups that were recently identified (dotted circles; NG, northern group; CG, central group; SLG; southern low-altitude group; SHG, southern high-altitude group) [14].

LA4332), a southern low-altitude group (SLG; LA2750, LA2932, LA4107, LA4108), and a central group (CG; LA0456, LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2753, LA2755, LA2765, LA2773, LA2931, LA3111; figure 1 [14]). The populations of SHG are located in the Andes; the populations of SLG are scattered along the Pacific coast in the southern range of the species distribution. CG covers populations from low and high altitudes. For convenience, we will refer to this grouping throughout this study. For more details on sampling sites, habitat characteristics and breeding procedures at the TGRC, see the electronic supplementary material, table S1, Böndel *et al.* [14] and the TGRC website (<http://tgrc.ucdavis.edu> (accessed 16 August 2017)).

2.2. Choice of genes and sequencing approach

We have recently re-sequenced a set of 30 genes from these 23 *S. chilense* populations in a pooling approach [14]. Sixteen of the genes were previously reported to be involved in the abiotic stress response in *Solanum* sect. *Lycopersicon* or related Solanaceae species like potato or tobacco, and serve as candidate genes for natural selection and local adaptation. They comprise five regulatory genes (*AREB1*, *AREB2*, *JERF1*, *JERF3*, *DREB3*), ten functional genes (*CT208*, *dhn1*, *ER5*, *His1*, *le25*, *LTP*, *pLC30-15*, *TAS14*, *TPP*, *TSW12*) and one sensory gene (*NtC7*; table 1). This classification is common for abiotic stress-responsive

Table 1. Abiotic stress-responsive (candidate) genes included in this study.

gene ^a	locus and coordinates ^b	gene description
sensory gene		
<i>NtC7</i>	Solyc03g083480; SL2.50ch03:53361485..53362716	receptor-like membrane protein; induced by salt and osmotic stress in <i>Nicotiana tabacum</i> [18]
regulatory genes		
<i>AREB1</i>	Solyc04g078840; SL2.50ch04:63504038..63500731	AREB/ABF subfamily of basic leucine zipper (bZIP) transcription factor; induced by salt, drought, low temperature and ABA in tomato [8]; induces the expression of <i>dhn1</i> , <i>pLC30-15</i> , <i>TAS14</i> , <i>le25</i> , <i>TSW12</i> and <i>TPP</i> [8,18] (Sol Genomics Network, https://solgenomics.net , accessed 16 August 2017)
<i>AREB2</i>	Solyc11g044560; SL2.50ch11:32458669..32463971	AREB/ABF subfamily of basic leucine zipper (bZIP) transcription factor; induced by drought, salt and ABA in tomato, and <i>AREB1</i> [19]
<i>JERF1</i>	Solyc06g063070; SL2.50ch06:39819912..39817161	ethylene responsive factor (ERF) protein; induced by salt and ABA in tomato [20]; enhances tolerance and growth under salinity and cold temperature in transgenic tobacco [21]; induces expression of <i>pLC30-15</i> and <i>LTP</i> [21]
<i>JERF3</i>	Solyc03g123500; SL2.50ch03:70333120..70335563	ERF protein; induced by cold, salt and ABA in tomato [22]
<i>DREB3</i>	Solyc04g072900; SL2.50ch04:59868088..59868882	AP2/ERF transcription factor family; induced by salt, drought and low temperature in tomato [23]
functional genes		
<i>dhn1</i> (<i>ERD10B</i>)	Solyc02g084840; SL2.50ch02:47921771..47920803	group 2 LEA (dehydrin); induced by low temperature and ABA in potato [24]; induced by <i>AREB1</i> [8]
<i>pLC30-15</i>	Solyc04g082200; SL2.50ch04:65958867..65957495	group 2 LEA (dehydrin); induced by drought, cold and ABA in <i>S. chilense</i> [25,26]; induced by <i>JERF1</i> [21]; signatures of local adaptation in <i>S. chilense</i> [12]
<i>TAS14</i> (<i>le4</i>)	Solyc02g084850; SL2.50ch02:47926954..47925819	group 2 LEA (dehydrin); induced by salt, drought and ABA in tomato [6,7]; induced by <i>AREB1</i> [8,19]
<i>ER5</i>	Solyc01g095140; SL2.50ch01:86522427..86523239	atypical hydrophobic group of LEA; induced by drought and ABA in tomato [27], not upregulated by overexpressing <i>AREB1</i> [8]
<i>le25</i>	Solyc10g078770; SL2.50ch10:60482744..60481971	group 4 LEA; induced by drought and ABA in tomato [6]; expression in <i>S. cerevisiae</i> resulted in higher salt and freezing tolerance [28]; induced by <i>AREB1</i> [19]
<i>LTP</i> (<i>NtLTP1</i>)	Solyc10g075100; SL2.50ch10:58811212..58810582	lipid transfer protein; induced by salt in tomato (GenBank entry DQ073079); induced by <i>JERF1</i> [21]
<i>TSW12</i>	Solyc10g075110; SL2.50ch10:58833096..58832488	lipid transfer protein; induced by salt in tomato [29]; induced by <i>AREB1</i> [19]
<i>CT208</i>	Solyc09g064370; SL2.50ch09:61552580..61557226	alcohol dehydrogenase class III; signature of selection in <i>S. chilense</i> [30]
<i>His1</i> (<i>le20</i>)	Solyc02g084240; SL2.50ch02:47397430..47398592	H1 histone gene; induced by drought and ABA in tomato and <i>S. pennellii</i> [31,32]; induced by drought in <i>S. chilense</i> [32]
<i>TPP</i>	Solyc03g083960; SL2.50ch03:53898914..53896143	trehalose-6-phosphate phosphatase; induced by <i>AREB1</i> [8,19]

^a Alternative gene names are given in brackets.

^b Locus name and coordinates in the genome of the cultivated tomato (*S. lycopersicum* cv. Heinz, SL2.50; Sol Genomics Network, <https://solgenomics.net> (accessed 16 August 2017 [33])).

genes in plants [5]. Two of the regulatory genes (*AREB1*, *JERF1*) were reported to induce the expression of other genes included in this study [8,19,21], which allows us to link some of our candidate genes into a gene network (electronic supplementary material, figure S1). To our knowledge these are the only direct interactions between our candidate genes that have been reported in *Solanum* sect. *Lycopersicon*. The remaining 14 genes serve as reference genes in this study (electronic supplementary material, table S2). Thirteen reference genes are single-copy cDNA markers [34]; nine of them have been subject to previous studies on genetic variation and demography in wild tomatoes [35–38]. The granule-bound starch synthase I (*GBSSI*) gene was used previously to examine the phylogenetic relationships of wild tomato species [39]. According to our previous analyses, these reference genes are suitable to represent the genomic average of each population [14].

The sequence data are publicly available on GenBank (BioProject PRJNA313142, accessions SAMN04524139–SAMN04524161). To obtain this data, genomic DNA of 25 individual plants per population was extracted using the DNeasy Plant Mini Kit (Qiagen). A two-step pooling approach was applied for each population. First, DNA of five individuals was mixed to pre-pools prior to gene amplification by polymerase chain reaction (PCR) using the Phusion[®] high-fidelity DNA Polymerase kit (Finnzymes, distributed by New England BioLabs, Inc.). Second, the five PCR product pre-pools per gene were brought together during PCR product purification with the MinElute[®] Gel Extraction Kit (Qiagen), resulting in one PCR product pool per gene. Finally, all 30 PCR product pools per population were mixed in equimolar quantities and concentrated to ≥ 200 ng μl^{-1} using the Amicon[®] Ultra-0.5 Centrifugal Filter Devices (Millipore). DNA library construction and high-throughput sequencing (paired-end, 100 bp) on a Genome Sequencer Illumina HiSeq2000 were performed by GATC Biotech AG (Konstanz, Germany). Consequently, this *S. chilense* dataset consists of 575 individual genotypes (25 genotypes from each of the 23 populations) and 1150 alleles (50 alleles from each of the 23 populations) for each gene.

2.3. Sequence data assembly and analyses

The *S. chilense* short reads were mapped against sequences from the cultivated tomato genome (*S. lycopersicum* cv. Heinz, release SL2.40) from the SOL genomics network (<http://solgenomics.net> (accessed 16 August 2017)) using STAMPY [40] with the default parameters for paired-end read mapping and a substitution rate of 0.01. We chose Stampy because it showed a higher mapping accuracy than other programs for regions with high genetic variation [41] and gap regions [42]. Three genes (*CTO21*, *NtC7* and *TAS14*) were mapped against *S. chilense* sequences because preliminary tests revealed different length in *S. chilense* compared to *S. lycopersicum*. Pileup files were generated with SAMTOOLS [43] and transformed into tables with sequence depth and SNP information for each position. We conducted all data analyses based on these tables. Positions with a sequencing depth below 2000, a base quality and/or mapping quality below 20 were masked and excluded from all analyses. Mutations present in less than 1% of the reads were considered as sequencing errors and therefore excluded from all analyses. Indel polymorphisms were also excluded. The exact numbers of SNPs per gene and population are shown in the electronic supplementary material, table S3.

We previously assessed several statistics for this dataset [14]. We calculated the nucleotide diversity π , which is the average number of pairwise nucleotide differences per site [44], and the test statistic Tajima's *D* [45] per gene and population. Tajima's *D* is based on the difference between the Watterson estimator, Θ_W , and π . Tajima's *D* is expected to be 0 under standard neutral expectations. An excess of either low or high frequency polymorphisms leads to a negative Tajima's *D* and indicates purifying or positive directional selection. An excess of intermediate frequency polymorphisms leads to a positive Tajima's *D* and indicates balancing selection. An excess of low frequency polymorphisms and intermediate frequency polymorphisms can further indicate population expansion and population substructure, respectively. We also calculated divergence, *K*, to the outgroup *S. ochranthum*. The *S. ochranthum* sequences were generated previously [14] and are available on GenBank (accessions KY978852–KY978879). We further assessed both π and *K* for synonymous (π_s and K_s) and nonsynonymous (π_a and K_a) sites separately. We also calculated the pairwise genetic differentiation (F_{ST}) between each pair of populations for each gene [46].

To address the evolution of gene classes and networks on the species level, we summarized π , π_s , π_a , Tajima's *D*, *K*, K_s and K_a over all corresponding genes and populations and compared these statistics for (i) reference and candidate genes, (ii) regulatory and functional genes and (iii) subsets of candidate genes that were reported to be linked within networks to reference genes by applying the Wilcoxon rank sum test in R [47]. Each of the two analysed gene networks contains one regulatory and two or five functional genes reported to be induced by the respective regulatory gene: *AREB1* with *dhn1*, *le25*, *pLC30-15*, *TAS14*

and *TPP* and *JERF1* with *LTP* and *pLC30-15* (electronic supplementary material, figure S1). We did the same comparisons across populations to investigate whether the observed patterns on the species level are also present in the individual populations.

Then we investigated natural selection and local adaptation by comparing for each population the 16 abiotic stress-responsive genes (candidate genes) with the reference genes, which are assumed to represent the genome average.

We further studied if and how individual SNPs influence the genetic diversity in their immediate neighbourhood; i.e. we were interested in small-scale hitchhiking effects. The motivation for this analysis was that we recently discovered multiple sites characterized by narrow valleys of reduced nucleotide diversity, suggesting small-scale hitchhiking effects, within another abiotic stress-responsive gene, *ICE1*, in *S. chilense* [11]. We followed here the analysis of Lee *et al.* [48] in *Drosophila melanogaster*, because outcrossing tomato species like *S. chilense* show similar characteristics as *D. melanogaster* with regard to recombination rates and diversity levels [35,49,50]. We calculated π for 40-bp windows around each SNP. We excluded SNPs when more than three nucleotides within these 40-bp windows were masked and also species-wide fixed differences to *S. ochranthum*, as they likely represent events predating the demographic history of the *S. chilense* populations. We then summarized our results across genes and populations and compared reference genes to candidate genes and the two gene networks, as well as different SNP types (synonymous, nonsynonymous, intronic/intergenic) with each other by applying the Wilcoxon rank sum test.

3. Results and discussion

3.1. Signatures of adaptive evolution in gene classes and gene networks

We first compared candidate, i.e. abiotic stress-responsive, and reference genes by averaging π , π_s , π_a , K , K_s , K_a and Tajima's D across all populations. The main finding of this analysis was a highly significant excess of nonsynonymous nucleotide diversity and nonsynonymous divergence in the candidate genes in comparison with the reference genes. Both π_a and K_a values were highly significantly elevated ($p < 0.001$) in the candidate genes compared with the reference genes (figure 2). We also found π to be significantly elevated ($p < 0.05$) and both π_s and Tajima's D to be significantly reduced ($p < 0.05$) in the candidate relative to the reference genes (figure 2, electronic supplementary material, figure S2). The observed excess of nucleotide diversity, especially at nonsynonymous sites, and nonsynonymous divergence suggests that the candidate genes maintain more adaptive variation and thus may evolve faster. This is in agreement with the observation of an excess of nonsynonymous variation at adaptive genes in *Arabidopsis* [51]. If new environmental conditions are encountered, either because of environmental change or colonization of new habitats, standing genetic variation could become advantageous [52,53]. In fact, in our previous study we discovered possible local adaptation events from standing genetic variation in *S. chilense* [14].

The 16 candidate genes belong to different pathways and different layers of the abiotic stress response (electronic supplementary material, figure S1) and can be classified into sensory, regulatory and functional genes following e.g. Shinozaki & Yamaguchi-Shinozaki [5]. As we have only one sensory gene (*NtC7*), we compared regulatory to functional genes. Functional genes have significantly higher π_a than regulatory genes ($p < 0.001$; electronic supplementary material, figure S3), suggesting that they contribute most to the candidate versus reference gene π_a pattern described above. As functional genes are usually located more downstream in the abiotic stress response pathways than regulatory genes, this observation is in agreement with the analyses of the anthocyanin pathway, which showed that upstream genes evolve under stronger constraint than downstream genes [54,55]. This pattern may be explained with a higher connectivity of the upstream genes in gene networks [54,55] and stronger evolutionary constraints acting upon the higher connectivity genes [15].

High ratios of nonsynonymous to synonymous nucleotide diversity (π_a/π_s) or divergence (K_a/K_s) relative to the genomic average are usually considered to indicate adaptive evolution ([56], chapters 7 and 8). Candidate gene averages, especially functional gene averages, are above or closer to a π_a/π_s ratio of 1 than the averages of the reference genes (figure 3a). A similar pattern can be observed for the K_a/K_s ratios (figure 3b). These plots also show variation among genes from the same gene class. The π_a values of reference and regulatory genes are relatively similar to each other, while the functional genes differ more in π_a (electronic supplementary material, figure S4a). Genes from all classes vary in K_a (electronic supplementary material, figure S4b). The individual genes will be addressed later in this paper.

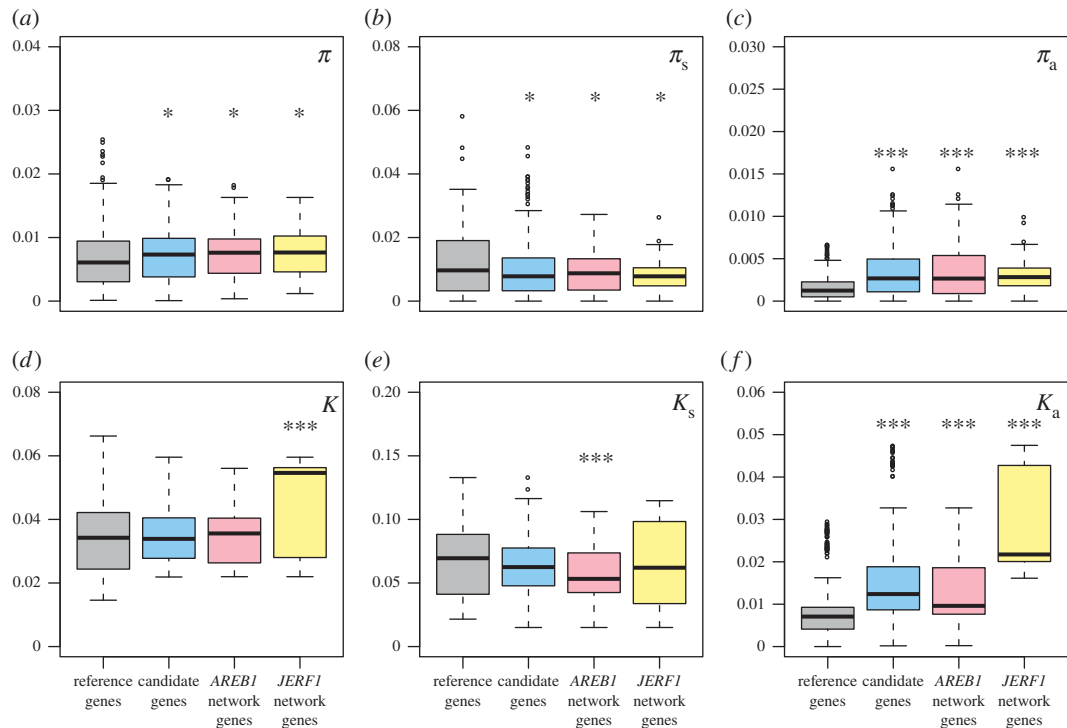


Figure 2. Nucleotide diversity and divergence of the gene classes and networks. The y-axis shows values of nucleotide diversity π and divergence K to *S. ochranthum* summarized over all populations for all sites (a,d), synonymous sites (b,e) and nonsynonymous sites (c,f) for reference genes (grey), candidate genes (blue), the *AREB1* network genes (light pink) and the *JERF1* network genes (yellow). Asterisks indicate significant differences to the reference genes (Wilcoxon rank sum test; * $p < 0.05$, *** $p < 0.001$).

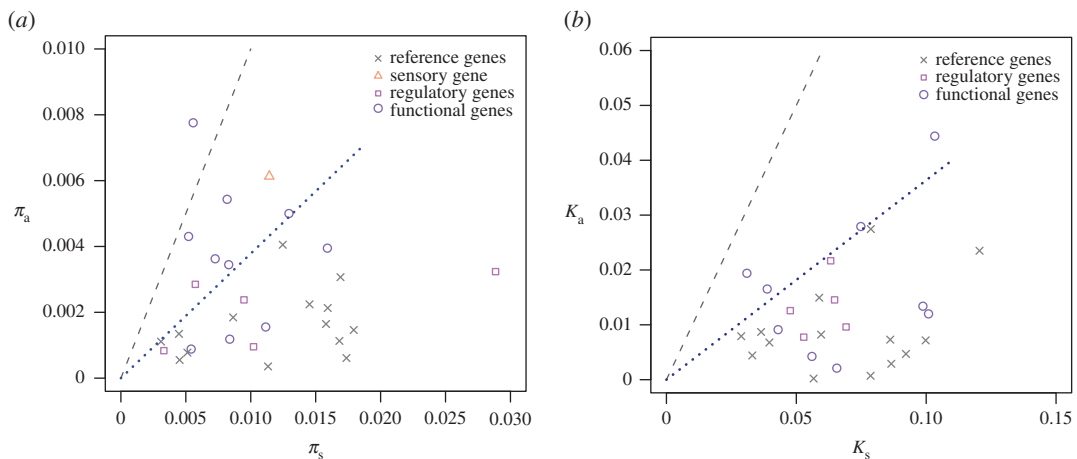


Figure 3. Nonsynonymous and synonymous nucleotide diversity and divergence. (a) Nucleotide diversity; the dashed line indicates a π_a/π_s ratio of 1, the dotted line indicates the upper boundary of the reference genes. (b) Divergence to *S. ochranthum*; the dashed line indicates a K_a/K_s ratio of 1. Gene classes are indicated as follows: grey crosses, reference genes; orange triangle, sensory gene; pink boxes, regulatory genes; purple circles, functional genes.

We also analysed two gene networks in comparison to the reference genes: *AREB1* with *dhn1*, *le25*, *pLC30-15*, *TAS14* and *TPP*, and *JERF1* with *LTP* and *pLC30-15*. Both *AREB1* and *JERF1* were reported to induce the respective regulatory genes ([8,19,21]; SOL Genomics Network, <https://solgenomics.net/> (accessed 16 August 2017)). Although additional direct interactions between our candidate genes cannot be excluded, to our knowledge, no other interactions have been reported for *Solanum* sect. *Lycopersicon* so far. Similar to the candidate gene averages, both networks show significantly elevated π , reduced π_s (both $p < 0.05$), and highly elevated ($p < 0.001$) π_a in comparison to the reference genes (figure 2). This

may not be surprising as the networks are subsets of the candidate genes. However, the two networks differ in their divergence patterns. The genes of the *JERF1* network show highly significantly elevated K ($p < 0.001$) and the genes of the *AREB1* network have significantly decreased K_s ($p < 0.05$). Like the candidate genes, both have significantly elevated K_a ($p < 0.001$), this being especially profound for the *JERF1* network genes. The transcription factor *JERF1* and the two genes downstream of *JERF1*, the dehydrin *pLC30-15* and the lipid transfer protein *LTP*, are among the four genes with highest K_a of all the candidate genes (electronic supplementary material, figure S4b), suggesting that this network may be undergoing adaptive evolution in *S. chilense*.

Furthermore, we calculated the aforementioned statistics for each population separately to investigate if the pattern observed on the species level is also present in the individual populations (electronic supplementary material, tables S4–S10). In all populations, averages of π_a , K and K_a are higher for the candidate than for the reference genes; the same holds for K and K_a for the *JERF1* network genes and for π_a and K_a for the *AREB1* network genes, while K and K_s are lower for the *AREB1* network genes. The functional genes in comparison to the regulatory genes show higher averages of π_a , K_s and K_a in all populations. All other statistics do not show a consistent pattern across populations in the respective gene class or network comparisons. The fact that the patterns observed on the species level are also largely present across populations indicates that the north–south colonization [14] had no major impact on the evolution of gene classes and networks.

3.2. Patterns of genetic diversity in the immediate neighbourhood of single-nucleotide polymorphisms

We have recently discovered elevated nucleotide diversity and Tajima's D in the transcription factor *ICE1* in *S. chilense*, which could be explained by interactions of several positively selected sites within the gene [11]. Since we also observed elevated nucleotide diversity, especially at nonsynonymous sites, in the present dataset, we decided to investigate the patterns around individual SNPs to assess the possibility of interactions between several positively selected variants. This analysis revealed that synonymous and nonsynonymous SNPs, i.e. exonic SNPs, show reduced π in the immediate neighbourhood in comparison to intronic and intergenic SNPs (figure 4). This reduction is highly significant ($p < 0.001$) in reference genes, candidate genes and also in the genes of the two networks. Wilcoxon tests comparing reference genes to candidate genes and reference genes to each of the two networks revealed also significant differences in π for each of the three SNP types (intronic/intergenic, synonymous and nonsynonymous). Furthermore, the reduction of π around nonsynonymous SNPs in the two networks appears to be less severe than in all candidate genes. These different degrees of reduction of genetic diversity around nonsynonymous SNPs may be explained by different strengths of selection acting on the candidate genes. Interaction between selected variants may lead to evolutionary traffic [57], which—in turn—is expected to reduce the strength of selection and increase nucleotide diversity between and around selected sites [58]. According to the results of our analyses, the frequency of the SNPs does not seem to greatly influence π in the immediate SNP neighbourhood in both candidate and reference genes although the scatter of π values is larger in the neighbourhood of SNPs fixed within populations than in the neighbourhood of SNPs in intermediate to high frequencies (electronic supplementary material, figure S5).

3.3. Species-wide signatures of adaptation

Analyses of individual genes on the population level identified several abiotic stress-responsive genes that show consistent trends across most populations, which might indicate that similar selective forces are acting upon these genes across the species distribution, i.e. across various environments. Four of these genes, the transcription factors *AREB1* and *AREB2*, the dehydrin *dhn1*, and the alcohol dehydrogenase *CT208*, are characterized by rather low nucleotide diversity, especially at nonsynonymous sites, in comparison to the reference gene averages (electronic supplementary material, tables S4–S6). Further, K is reduced in *dhn1*, K_s in *AREB1*, *AREB2* and *dhn1* and K_a in *dhn1* and *CT208* (electronic supplementary material, tables S8–S10). This may suggest that these genes are rather conserved in *S. chilense*. However, a selective sweep scenario has been hypothesized previously for the alcohol dehydrogenase *CT208* in *S. chilense* previously, which also agrees with the low nucleotide diversity observed in our study. Arunyawat *et al.*'s [30] main finding to support this hypothesis was a haplotype that decreased in frequency from south to north. Our analysis of pairwise genetic differentiation (F_{ST}) between populations

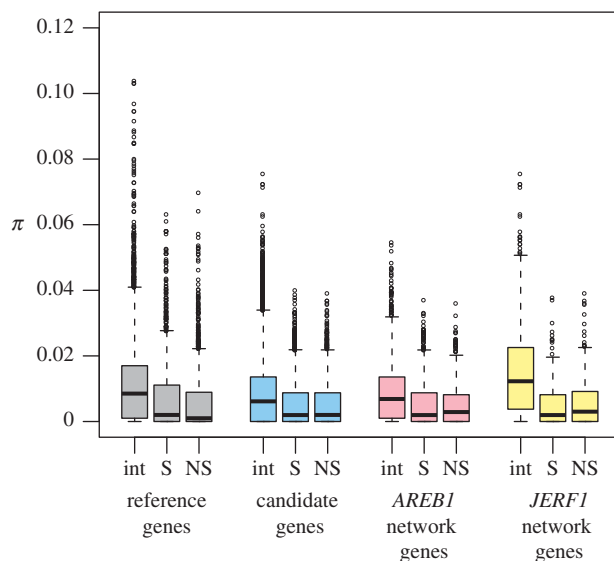


Figure 4. Patterns of nucleotide diversity surrounding SNPs. Values of π for 40-bp windows around intronic/intergenic (int), synonymous (S) and nonsynonymous (NS) SNPs in reference genes (grey), candidate genes (blue) and genes of the *AREB1* (light pink) and *JERF1* (yellow) networks.

shows relatively low F_{ST} between populations from the southern and/or central regions and elevated F_{ST} between the two most northern populations and populations from either central or southern regions (electronic supplementary material, figure S9b), which would agree with the haplotype pattern detected by Arunyawat *et al.* [30] and support the hypothesis of a selective sweep scenario across a large part of the species distribution.

Other genes are characterized by rather high levels of nucleotide diversity and/or divergence in comparison to the reference gene averages (electronic supplementary material, tables S4–S6, S8–S10): the sensory gene *NtC7*, the transcription factors *JERF1* and *DREB3*, the *lea* genes *ER5* and *le25*, and the two lipid transfer protein encoding genes *LTP* and *TSW12*. *NtC7* and *JERF1* are further characterized by elevated Tajima's *D* values (electronic supplementary material, table S7). Elevated Tajima's *D*, i.e. an excess of intermediate frequency polymorphisms, may be indicative of balancing selection. The *NtC7* protein was reported to have a receptor-like function and to show structural similarities to tomato Cf-9, which is involved in the biotic stress response [18]. Originally, *NtC7* was identified as a wound-response gene [59] and shown to be induced by osmotic stresses [18]. These findings could imply that *NtC7* plays a role in both abiotic and biotic stress response. Balancing selection acting upon biotic stress-responsive genes is not uncommon in wild tomatoes [60,61]. Transcription factors could be involved in crosstalk between abiotic and biotic pathways [62]. The signature of balancing selection for *JERF1* could be explained by an involvement of this gene in different pathways in tomato as suggested by expression analyses [20]. The excess of (nonsynonymous) nucleotide diversity and divergence, which is observed for most abiotic stress-responsive genes, indicates evidence for adaptive evolution as discussed in §3.1.

3.4. Signatures of local adaptation

Some abiotic stress-responsive genes show signatures of selection across certain parts of the species range that are characterized by specific environments and may thus suggest local adaptation. For example, consistent signatures of positive selection are found in several genes of the SLG populations LA2750, LA2932, LA4107 and LA4108, whose habitat is influenced by the Pacific coast and the Atacama Desert. The SLG populations show reduced F_{ST} for the genes *dhn1*, *le25*, and *JERF3* in comparison to the reference gene averages (electronic supplementary material, figures S7 and S8). These three genes are known to be involved in the salt stress response. *JERF3* is induced by salt stress [61]. The expression of *dhn1* and *le25* is induced by the transcription factor *AREB1*, which is induced by salt stress [8,19]. *AREB1* does not show reduced genetic differentiation within this group (electronic supplementary material, figure S6b). However, we previously identified three outlier SNPs for positive selection in *AREB1* and all of them are either in high frequency or fixed in the SLG populations [14]. A positive selection scenario

for *AREB1* in the SLG populations is further supported by reduced Tajima's D values in comparison to the reference genes (electronic supplementary material, table S7). The Tajima's D values of *dhn1*, *le25* and *JERF3* are also reduced in the SLG populations (electronic supplementary material, table S7). However, the low nucleotide diversity and divergence of *dhn1* may suggest purifying instead of positive selection (see §3.3). *JERF3* and *le25*, on the other hand, show elevated K_a in SLG in comparison to the respective reference gene averages (electronic supplementary material, table S10) which is in agreement with positive selection. These findings imply that salinity may have been an important selective force acting upon these genes.

Different selection regimes seem to be prevalent in high-altitude populations. The high-altitude habitats of the CG populations are characterized by low temperatures; the SHG populations are exposed to low temperatures and also drought. *JERF3* has reduced Tajima's D and low F_{ST} in comparison to the reference gene averages in all SHG populations (electronic supplementary material, table S7 and figure S7a), which suggests positive selection and is in agreement with three previously identified outlier SNPs [14]. As *JERF3* is induced by cold stress [22], adaptation to low temperatures may be possible. However, because *JERF3* also showed signatures of positive selection in the SLG populations, this could imply that the observed pattern is specific to the southern range of the species distribution and must not necessarily be due to selection. However, we find that F_{ST} between the SLG and SHG populations is not reduced for *JERF3* (electronic supplementary material, figure S7a). This suggests that selection in these two population groups may act on different molecular variants.

The dehydrin *pLC30-15* is characterized by reduced Tajima's D , elevated K_a and F_{ST} in comparison to the reference gene averages in one high-altitude population from the northern CG range, namely LA3111 (electronic supplementary material, tables S7, S10 and figure S7d). These findings and one previously identified outlier SNP [14] may suggest that *pLC30-15* evolves under positive selection in LA3111. Previously *pLC30-15* was reported to evolve under diversifying selection in another *S. chilense* population [12] from the northern range of the distribution of the CG populations. Also other populations of this area, e.g. LA0456, LA1958 or LA1968, have reduced Tajima's D and elevated K_a for *pLC30-15* in comparison to the respective reference gene averages (electronic supplementary material, tables S7 and S10), suggesting that *pLC30-15* may be under positive selection in this area. Since the strongest signature is observed in LA3111, which exists in one of the coldest habitats, this could imply that *pLC30-15* is involved in adaptation to low temperatures. In fact, *pLC30-15* was reported to be upregulated under low-temperature stress [25].

The populations LA2880 and LA4118 (both SHG) are characterized by elevated Tajima's D and also by elevated π for the transcription factor *AREB1* and the dehydrin *TAS14* compared to the respective reference gene averages (electronic supplementary material, tables S4 and S7), which may suggest that balancing selection might be acting on these two genes in these two high-altitude populations. Furthermore, *TAS14* has elevated π_a in both populations (electronic supplementary material, table S6). Interestingly, *TAS14* is induced by *AREB1* [8,19] indicating that selection may be acting on both genes in the same way. The occurrence of balancing selection may be unexpected for abiotic stress-responsive genes as it has primarily been found for genes involved in the biotic stress response [63]. As the SHG populations are confronted with low temperatures, drought and partly also with high salt concentrations in the soil [1] and both genes were reported to be involved in more than one abiotic stress response [6–8], varying climatic and/or environmental conditions could explain the balancing selection signature. For example, *CBF2*, a transcription factor from the cold-signalling pathway, was reported to evolve under balancing selection in wild tomato species [10]. Expression analysis revealed that one allele was induced by cold stress and another one by drought stress. Such a scenario may be possible for *AREB1* and *TAS14* as well. Recently, patterns consistent with balancing selection were also reported for *ICE1*, the key gene of the cold-signalling pathway [11]. In this case, however, balancing selection was excluded and the observed pattern was explained by interactions of positively selected SNPs leading to intermediate frequency polymorphisms between the selected sites [58]. Such a scenario may also be possible. Our results as well as the two aforementioned studies indicate that signatures consistent with balancing selection may be more frequent in abiotic stress-responsive genes than expected and that this may be an interesting direction for future research on the evolution of abiotic stress-responsive genes.

4. Conclusion

The wild tomato species *S. chilense* is exposed to extremely variable and extreme environmental conditions. It is, therefore, reasonable to assume that frequent adaptations have occurred in the past.

In this study, we focused on a set of 16 abiotic stress-responsive genes. Our analyses revealed elevated nucleotide diversity and divergence, especially at nonsynonymous sites, which may suggest that these genes have evolved under positive selection. We also found interesting results for several individual genes, which may provide a basis for future studies. Some genes (e.g. *NtC7*, *dhn1*) were detected to have patterns that are consistent across the majority of populations, which may indicate that selection pressures are homogeneous across most of the species range. However, we also identified genes (e.g. *AREB1*, *JERF3*, *pLC30-15*) with signatures specific to certain environments, which may suggest local adaptation. Interestingly, two genes, *JERF3* and *AREB1*, may be involved in different adaptation processes because selection seems to be acting on different molecular variants within the genes: adaptation to coastal as well as to high-altitude conditions. Another interesting finding is that two interacting genes, *AREB1* and *TAS14*, show similar signatures in two populations. Some of our candidate genes are members of the same gene families (e.g. *JERF1* and *JERF3*) and our results suggest different evolutionary histories for the individual paralogs. We did not address gene family evolution further here but anticipate that wild tomatoes are suitable for future studies in this direction. Although our dataset is relatively limited with only 16 genes, it provides a first step to the evolutionary genetic analyses of whole abiotic stress pathways in wild tomatoes.

Data accessibility. The dataset used in this study is available on GenBank (pooled *S. chilense* sequences: accessions SAMN04524139–SAMN04524161, BioProject PRJNA313142; *S. ochranthum* sequences accessions KY978852–KY978879).

Authors' contributions. K.B.B. designed the study, analysed the data and wrote the manuscript. T.N. contributed to data analysis and helped write the manuscript. W.S. participated in the design of the project and helped write the manuscript. All authors gave final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by the German Research Foundation (DFG) grant no. STE 325/15 to W.S.

Acknowledgements. The authors thank Mamadou Mboup, Aurélien Tellier and the Munich evolutionary biology group for valuable discussions and three anonymous reviewers for valuable comments on the manuscript.

References

- Chetelat RT, Pertuzé RA, Faúndez L, Graham EB, Jones CM. 2009 Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. *Euphytica* **167**, 77–93. (doi:10.1007/s10681-008-9863-6)
- Moyle LC. 2008 Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect. *Lycopersicon*). *Evolution* **62**, 2995–3013. (doi:10.1111/j.1558-5646.2008.00487.x)
- Nakazato T, Warren DL, Moyle LC. 2010 Ecological and geographic modes of species divergence in wild tomatoes. *Am. J. Bot.* **97**, 680–693. (doi:10.3732/ajb.0900216)
- Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. 2014 The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* **5**, 170. (doi:10.3389/fpls.2014.00170)
- Shinozaki K, Yamaguchi-Shinozaki K. 2007 Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* **58**, 221–227. (doi:10.1093/jxb/erl164)
- Cohen A, Bray EA. 1990 Characterization of three mRNAs that accumulate in wilted tomato leaves in response to elevated levels of endogenous abscisic acid. *Planta* **182**, 27–33. (doi:10.1007/BF00239979)
- Godoy JA, Pardo JM, Pintor-Toro JA. 1990 A tomato cDNA inducible by salt stress and abscisic acid: nucleotide sequence and expression pattern. *Plant Mol. Biol.* **15**, 695–705. (doi:10.1007/BF0016120)
- Yañez M, Cáceres S, Orellana S, Bastías A, Verdugo I, Ruiz-Lara S, Casareto JA. 2009 An abiotic stress-responsive bZIP transcription factor from wild and cultivated tomatoes regulates stress-related genes. *Plant Cell. Rep.* **28**, 1497–1507. (doi:10.1007/s00299-009-0749-4)
- Fischer I, Camus-Kulandaivelu L, Allal F, Stephan W. 2011 Adaptation to drought in two wild tomato species: the evolution of the *Asr* gene family. *New Phytol.* **190**, 1032–1044. (doi:10.1111/j.1469-8137.2011.03648.x)
- Mboup M, Fischer I, Lainer H, Stephan W. 2012 Trans-species polymorphism and allele-specific expression in the *CBF* gene family of wild tomatoes. *Mol. Biol. Evol.* **29**, 3641–3652. (doi:10.1093/molbev/mss176)
- Nosenko T, Böndel KB, Kumpfmüller G, Stephan W. 2016 Adaptation to low temperatures in the wild tomato species *Solanum chilense*. *Mol. Ecol.* **25**, 2853–2869. (doi:10.1111/mec.13637)
- Xia H, Camus-Kulandaivelu L, Stephan W, Tellier A, Zhang Z. 2010 Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. *Mol. Ecol.* **19**, 4144–4154. (doi:10.1111/j.1365-294X.2010.04762.x)
- Peralta IE, Spooner DM, Knapp S. 2008 Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Syst. Bot. Monogr.* **84**, 1–186.
- Böndel KB, Lainer H, Nosenko T, Mboup M, Tellier A, Stephan W. 2015 North–south colonization associated with local adaptation of the wild tomato species *Solanum chilense*. *Mol. Biol. Evol.* **32**, 2932–2943. (doi:10.1093/molbev/msv166)
- Olson-Manning CF, Wagner MR, Mitchell-Olds T. 2012 Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat. Rev. Genet.* **13**, 867–877. (doi:10.1038/nrg3322)
- Alvarez-Ponce D, Fares MA. 2012 Evolutionary rate and duplicability in the *Arabidopsis thaliana* protein–protein interaction network. *Genome Biol. Evol.* **4**, 1263–1274. (doi:10.1093/gbe/evs101)
- Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013 A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* **13**, 214. (doi:10.1186/1471-2148-13-214)
- Tamura T, Hara K, Yamaguchi Y, Koizumi N, Sano H. 2003 Osmotic stress tolerance of transgenic tobacco expressing a gene encoding a membrane-located receptor-like protein from tobacco plants. *Plant Physiol.* **131**, 454–462. (doi:10.1104/pp.102.011007)
- Orellana S, Yañez M, Espinoza A, Verdugo I, González E, Ruiz-Lara S, Casareto JA. 2010 The transcription factor SIAREB1 confers drought, salt stress tolerance and regulates biotic and abiotic stress-related genes in tomato. *Plant Cell Environ.* **33**, 2191–2208. (doi:10.1111/j.1365-3040.2010.02220.x)
- Zhang H *et al.* 2004 The ethylene-, jasmonate-, abscisic acid- and NaCl-responsive tomato transcription factor JERF1 modulates expression of GCC box-containing genes and salt tolerance in tobacco. *Planta* **220**, 262–270. (doi:10.1007/s00425-004-1347-x)
- Wu L, Chen X, Ren H, Zhang Z, Zhang H, Wang J, Wang XC, Huang R. 2007 ERF protein JERF1 that transcriptionally modulates the expression of abscisic acid biosynthesis-related gene enhances the tolerance under salinity and cold in tobacco.

- Planta* **226**, 815–825. (doi:10.1007/s00425-007-0528-9)
22. Wang H, Huang Z, Chen Q, Zhang Z, Zhang H, Wu Y, Huang D, Huang R. 2004 Ectopic overexpression of tomato *JERF3* in tobacco activates downstream gene expression and enhances salt tolerance. *Plant Mol. Biol.* **55**, 183–192. (doi:10.1007/s11003-004-0113-6)
 23. Islam MS, Wang MH. 2009 Expression of dehydration responsive element-binding protein-3 (*DREB3*) under different abiotic stresses in tomato. *BMB Rep.* **42**, 611–616. (doi:10.5483/BMBRep.2009.42.9.611)
 24. Baudou MM, Meza-Zepeda LA, Palva ET, Heino P. 1996 Induction of homologous low temperature and ABA-responsive genes in frost resistant (*Solanum commersonii*) and frost-sensitive (*Solanum tuberosum* cv. Bintje) potato species. *Plant Mol. Biol.* **30**, 331–336. (doi:10.1007/BF00020118)
 25. Fischer I, Steige KA, Stephan W, Mboup M. 2013 Sequence evolution and expression regulation of stress-responsive genes in natural populations of wild tomato. *PLoS ONE* **8**, e78182. (doi:10.1371/journal.pone.0078182)
 26. Chen RD, Campeau N, Greer AF, Bellemare G, Tabaeizadeh Z. 1993 Sequence of a novel abscisic acid- and drought-induced cDNA from wild tomato (*Lycopersicon chilense*). *Plant Physiol.* **103**, 301. (doi:10.1104/pp.103.1.301)
 27. Zegzouti H, Jones B, Marty C, Lelièvre JM, Latché A, Pech JC, Bouzayen M. 1997 ERS, a tomato cDNA encoding an ethylene-responsive LEA-like protein: characterization and expression in response to drought, ABA and wounding. *Plant Mol. Biol.* **35**, 847–854. (doi:10.1023/A:1005860302313)
 28. Imai R, Chang L, Ohta A, Bray EA, Takagi M. 1996 A lea-class gene of tomato confers salt and freezing tolerance when expressed in *Saccharomyces cerevisiae*. *Gene* **170**, 243–248. (doi:10.1016/0378-1119(95)00868-3)
 29. Torres-Schumann S, Godoy JA, Pintor-Toro JA. 1992 A probable lipid transfer protein gene is induced by NaCl in stems of tomato plants. *Plant Mol. Biol.* **18**, 749–757. (doi:10.1007/BF00020016)
 30. 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)
 31. Kahn TL, Fender SE, Bray EA, O'Connell MA. 1993 Characterization of expression of drought- and abscisic acid-regulated tomato genes in the drought-resistant species *Lycopersicon pennellii*. *Plant Physiol.* **103**, 597–605. (doi:10.1104/pp.103.2.597)
 32. Wei T, O'Connell MA. 1996 Structure and characterization of a putative drought-inducible H1 histone gene. *Plant Mol. Biol.* **30**, 255–268. (doi:10.1007/BF00020112)
 33. The Tomato Genome Consortium. 2012 The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641. (doi:10.1038/nature11119)
 34. Tanksley SD *et al.* 1992 High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**, 1141–1160.
 35. Baudry E, Kerdellhué C, Innan H, Stephan W. 2001 Species and recombination effects on DNA variability in the tomato genus. *Genetics* **158**, 1725–1735.
 36. Städler T, Arunyawat U, Stephan W. 2008 Population genetics of speciation in two closely related wild tomatoes (*Solanum* section *Lycopersicon*). *Genetics* **178**, 339–350. (doi:10.1534/genetics.107.081810)
 37. Städler T, Roselius K, Stephan W. 2005 Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. *Evolution* **59**, 1268–1279. (doi:10.1554/04-722)
 38. Tellier A, Laurent SJY, Lainer H, Pavlidis P, Stephan W. 2011 Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proc. Natl Acad. Sci. USA* **108**, 17 052–17 057. (doi:10.1073/pnas.1111266108)
 39. Peralta IE, Spooner DM. 2001 Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). *Am. J. Bot.* **88**, 1888–1902.
 40. Lunter G, Goodson M. 2011 Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939. (doi:10.1101/gr.111120.110)
 41. Nookaew I, Papini M, Pornputtpong N, Scalcinati G, Fagerberg L, Uhlén M, Nielsen J. 2012 A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **40**, 10 084–10 097. (doi:10.1093/nar/gks804)
 42. Catalán A, Hutter S, Parsch J. 2012 Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genomics* **13**, 654. (doi:10.1186/1471-2164-13-654)
 43. Li H *et al.* 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
 44. Nei M, Li WH. 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273. (doi:10.1073/pnas.76.10.5269)
 45. Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
 46. Hudson RR, Slatkin M, Maddison WP. 1992 Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589.
 47. R Development Core Team. 2005 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
 48. Lee YCG, Langley CH, Begun DJ. 2014 Differential strengths of positive selection revealed by hitchhiking effects at small physical scales in *Drosophila melanogaster*. *Mol. Biol. Evol.* **31**: 804–816. (doi:10.1093/molbev/mst270)
 49. 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)
 50. Stephan W, Langley CH. 1998 DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**, 1585–1593.
 51. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowicz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011 Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83–86. (doi:10.1126/science.1209244)
 52. Barrett RD, Schluter D. 2008 Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44. (doi:10.1016/j.tree.2007.09.008)
 53. Innan H, Kim Y. 2008 Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* **179**, 1713–1720. (doi:10.1534/genetics.108.086835)
 54. Rausher MD, Lu Y, Meyer K. 2008 Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *J. Mol. Evol.* **67**, 137–144. (doi:10.1007/s00239-008-9105-5)
 55. Rausher MD, Miller RE, Tiffin P. 1999 Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.* **16**, 266–274. (doi:10.1093/oxfordjournals.molbev.a026108)
 56. Li WH. 1997 *Molecular evolution*. Sunderland, MA: Sinauer Associates.
 57. Kirby DA, Stephan W. 1996 Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. *Genetics* **144**, 635–645.
 58. Chevin LM, Billiard S, Hospital F. 2008 Hitchhiking both ways: effect of two interfering selective sweeps on linked neutral variation. *Genetics* **180**, 301–316. (doi:10.1534/genetics.108.089706)
 59. Hara K, Yagi M, Koizumi N, Kusano T, Sano H. 2000 Screening of wound-responsive genes identifies an immediate-early expressed gene encoding a highly charged protein in mechanically wounded tobacco plants. *Plant Cell Physiol.* **41**, 684–691. (doi:10.1093/pcp/41.6.684)
 60. Hörger AC, Ilyas M, Stephan W, Tellier A, Van Der Hooft RA, Rose LE. 2012 Balancing selection at the tomato *RCR3* Guardee gene family maintains variation in strength of pathogen defense. *PLoS Genet.* **8**, e1002813. (doi:10.1371/journal.pgen.1002813)
 61. Rose LE, Grzeskowiak L, Hörger AC, Groth M, Stephan W. 2011 Targets of selection in a disease resistance network in wild tomatoes. *Mol. Plant Pathol.* **12**, 921–927. (doi:10.1111/J.1364-3703.2011.00720.x)
 62. Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K. 2006 Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.* **9**, 436–442. (doi:10.1016/j.pbi.2006.05.014)
 63. Delph LF, Kelly JK. 2014 On the importance of balancing selection in plants. *New Phytol.* **201**, 45–56. (doi:10.1111/nph.12441)