Evolution of the wild tomato species *Solanum chilense*: demography and natural selection

Dissertation
an der Fakultät für Biologie
der Ludwig-Maximilians-Universität
München

vorgelegt von
Katharina Barbara Böndel
aus Rinteln

München, April 2014

Dekan: Prof. Dr. Heinrich Leonhardt

1. Gutachter: Prof. Dr. Wolfgang Stephan

2. Gutachter: Prof. Dr. Wolfgang Enard

Tag der Abgabe: 15. April 2014

Tag der mündlichen Prüfung: 10. Juli 2014

Erklärung:

Diese Dissertation wurde im Sinne von §12 der Promotionsordnung von Prof. Dr. Stephan betreut. Ich erkläre hiermit, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

Eidesstattliche Erklärung:

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 15.04.2014

Katharina Böndel

Note

In this thesis I present the results of my doctoral research which I conducted from November 2010 until April 2014 in the area of evolutionary biology. This research comprises predominantly genetic analyses of the wild tomato species *Solanum chilense*, but also a phenotypic experiment, and was done in collaboration with several other scientists.

The project for the genetic part was designed by Mamadou Mboup, Aurélien Tellier, Wolfgang Stephan and me. The salt stress experiment was designed by Tetyana Nosenko and me.

All of the experimental and analytical work has been done by myself except for the following: Hilde Lainer did about half of the DNA and PCR product preparation for the *S. chilense* sequencing, the *S. chilense* sequencing itself (including library preparation) was done by the GATC Biotech AG in Konstanz, Hilde Lainer and Gaby Kumpfmüller did most of the outgroup sequencing, Pablo Duchén provided the PERL script to extract SNP information from the pileup file, Armin Scheben analysed the synonymous and nonsynonymous polymorphism and divergence and did the McDonald-Kreitman tests for half of the candidate genes as his IRT1 ('individual research training') project, Paula Brücher performed the population genetic analyses of the consensus sequence data set as part of her bachelor thesis project, Tetyana Nosenko did the phylogenetic analysis of the consensus sequence data set, and the salt stress treatment and documentation was done together with Tetyana Nosenko.

Contents

Summary

Demography and adaptation are important factors determining the evolution of plant species. Many plant species are substructured into populations or demes connected by migration (metapopulations). The spatial distribution of populations and migration patterns depend on the means of dispersal. Since plants are sessile organisms, they also have to cope with both biotic and abiotic stresses. Therefore adaptations to local environmental conditions are essential to ensure survival and duration of the species.

Wild tomato species (*Solanum* section *Lycopersicon*) are native to western South America. They occur in diverse and often extreme habitats including rain forests, coastal regions, high altitude habitats in the Andean Mountains and also hyperarid deserts in the Atacama Desert. Therefore, wild tomatoes are a good model system to study plant evolution and genomic bases for plant adaptation. This study focuses on the wild tomato species *Solanum chilense*, which exhibits a metapopulation structure with populations distributed from southern Peru to northern Chile. In its native range, *S. chilense* is confronted with different abiotic stresses including drought, cold and salinity. I sequenced 30 unlinked nuclear genes from 23 populations using next generation sequencing. 16 genes are involved in the abiotic stress response and serve as candidates for selection and adaptation. The remaining 14 genes are used as references to study the genomic average and species past demography.

In the first part of this study, I investigated the demographic history of the wild tomato species *Solanum chilense*. Genetic data analyses revealed a north-south cline. This cline includes 1) a decrease of genetic variation from north to south, 2) an increase in the strength of population expansion along the cline, and 3) an increase in genetic differentiation from other wild tomato species towards the south of the range. Results further revealed that the populations form four groups: a central group and three peripheral groups. Altogether the results suggest that *S. chilense* originated in the northern part of its current distribution and migrated to the south, via two routes, along the coast and higher up in the Andes. During this north-south colonization, at least three bottlenecks occurred.

In the second part of this study, I investigated natural selection and local adaptation in *S. chilense*. Signatures of selection and local adaptation were detected in the abiotic stress-related genes, for example signatures of positive selection in high altitude populations were found possibly indicating adaptation to low temperatures. Interestingly, signatures of balancing selection were detected as well in high altitude populations reflecting probable

adaptation to different types of abiotic stresses. The coastal populations showed a distinct pattern. Several genes involved in the salt stress response exhibited signatures of local adaptation. Performing a salt stress experiment, I revealed that low altitude populations cope better with such stress than populations from intermediate or high altitudes. The coastal populations also showed an accumulation of nonsynonymous and possibly deleterious genetic variation, which can be explained by extreme bottlenecks and potential occurrence of selfing in some populations. Signatures of selection and local adaptation in *S. chilense* were mainly detected in populations from the peripheral groups and not in the central region, in agreement with the hypothesis that local adaptation is associated with the colonization of new territories.

In summary, this study showed that demography plays an important role in the evolutionary history of *S. chilense* and that local adaptation for key abiotic stresses occurs more frequently in the marginal ranges of the species distribution.

Zusammenfassung

Demographie und Anpassung spielen eine wichtige Rolle in der Evolution von Pflanzen. Viele Pflanzenarten sind unterteilt in Populationen die durch Migration miteinander verbunden sind (Metapopulationsstruktur). Ihre räumliche Ausbreitung und die Migrationsraten sind abhängig von den Verbreitungsmechanismen. Da es sessile Lebewesen sind, können sie biotischen und abiotischen Stressfaktoren nicht ausweichen. Daher sind Anpassungsmechanismen an die lokalen Umweltbedingungen essentiell um das Überleben und Fortbestehen der Art zu gewährleisten.

Wildtomaten (*Solanum* sect. *Lycopersicon*) sind im Westen Südamerikas beheimatet. Sie kommen in unterschiedlichen und teilweise extremen Habitaten vor. Diese reichen von Regenwäldern, Küstenregionen, Bergregionen in den Anden zu den hyperariden Wüstenregionen der Atacama. Daher sind Wildtomaten geeignet um Evolution und die genomischen Grundlagen für die Anpassung an verschiedene Umweltbedingungen zu untersuchen. Diese Studie beschäftigt sich mit der Wildtomatenart *Solanum chilense*. Diese Art hat eine Metapopulationsstruktur und die einzelnen Populationen sind vom südlichen Peru bis ins nördliche Chile verteilt. In ihrem natürlichen Verbreitungsgebiet ist *S. chilense* mit verschiedenen abiotischen Stressfaktoren konfrontiert, u.a. mit Trockenheit, Kälte und hohem Salzgehalt im Boden. 30 Gene aus 23 Populationen wurden mit den neuen Sequenziertechnologien sequenziert. 16 dieser Gene sind Teil der abiotischen Stressantwort und dienen als Kandidatengene für Selektion und Anpassung. Die übrigen 14 Gene werden als Referenzgene verwendet um den genomischen Durchschnitt und die Demographie der Art zu untersuchen.

Der erste Teil dieser Studie beschäftigt sich mit der Demographie der Wildtomatenart *S. chilense*. Die Analyse von genetischen Daten zeigte einen Nord-Süd Gradienten. Dieser Gradient beinhaltet von Nord nach Süd 1) eine Abnahme von genetischer Variation, 2) einen Anstieg in der Stärke des Populationswachstums und 3) einen Anstieg von genetischer Differenzierung zu anderen Wildtomatenarten. Die Analyse zeigte zudem, dass die Populationen sich vier Gruppen zuordnen lassen: einer zentralen Gruppe und drei peripheren Gruppen. Zusammengenommen deuten die Ergebnisse darauf hin, dass der Ursprung *S. chilense*'s im nördlichen Teil ihres heutigen Verbreitungsgebietes liegt und dass die Art sich gen Süden ausgebreitet hat. Während dieser Nord-Süd Ausbreitung kam es zu mindestens drei Flaschenhalseffekten.

Der zweite Teil dieser Studie beschäftigt sich mit natürlicher Selektion und lokaler Anpassung in *S. chilense*. Anzeichen für Selektion und lokale Anpassung wurden in verschiedenen Genen, die in die abiotische Stressantwort involviert sind, gefunden. Einige Gene zeigten Anzeichen für positive Selektion in Bergregionen und damit möglicherweise Anpassung an niedrige Temperaturen. Interessanterweise wurden auch Anzeichen für balancierende Selektion in Bergpopulationen gefunden. Dies könnte eine Anpassung an verschiedene Stressfaktoren wiederspiegeln. Die Küstenpopulationen zeigten ein eigenes charakteristisches Muster. Einige Gene, die in die Salzstressantwort involviert sind, zeigten Anzeichen für lokale Anpassung. Ein Salzstressexperiment zeigte, dass Populationen von niedrigen Höhenlagen besser mit Salzstress zurechtkommen als Populationen von höheren Höhenlagen. Außerdem zeigten die Küstenpopulationen erhöhte nichtsynonyme und damit möglicherweise schädliche genetische Variation. Dies könnte durch einen starken Flaschenhalseffekt oder möglicherweise durch Selbstbestäubung erklärt werden. Insgesamt wurden Anzeichen für Selektion und lokale Anpassung häufiger in Populationen der peripheren Gruppen gefunden. Dieses Ergebnis stimmt mit der Hypothese überein, dass die Kolonialisierung eines neuen Gebietes mit lokaler Anpassung einhergeht.

Zusammenfassend zeigte diese Studie dass Demographie in der evolutionären Geschichte *S. chilense*'s eine wichtige Rolle spielt und dass lokale Anpassung an abiotische Stressfaktoren vermehrt in den Randgebieten vorkommt.

Abbreviations

a. a.	amino acid
ABA	abscisic acid
ABRE	ABA-responsive element
all	all genes
AP2/EREP	AP2 domain/ethylene responsive element-binding protein
AREB/ABF	ABA-response element binding factor
Asr	ABA/water stress/ripening induced
bp	base pair
bZIP	basic leucine zipper
can	candidate genes
CBF	C-repeat binding factor
cv.	cultivar
D	Tajima's *D*
dehydrin	dehydration induced
DNA	deoxyribonucleic acid
DRE	dehydration responsive element
DREB	dehydration responsive element binding protein
ERF	ethylene responsive factor protein
F_{ST}	genetic differentiation
FDR	false discovery rate
GB	giga base pairs
GTR	generalized time-reversible
JERF	jasmonate and ethylene responsive factor
LEA/lea	late embryogenesis abundant
LRR	leucine rich repeat
ML	maximum likelihood
mya	million years ago
n. a./na	not available
NADP	nicotinamide adenine dinucleotide phosphate
NS	nonsynonymous SNP
PCR	polymerase chain reaction
ref	reference genes
S	synonymous SNP
S	number of segregating sites
SC	self-compatible
SI	self-incompatible
SNP	single nucleotide polymorphism
SOL	Solanaceae
SoLyco	*S. lycopersicum*
T_m	annealing temperature
TF	transcription factor
TGRC	Tomato Genetics Resource Center
UTP	uridine triphosphate
UTR	untranslated region

List of figures

List of tables

CHAPTER 1: INTRODUCTION

1.1 Adaptation in plants

The theory of evolution was proposed by Charles Darwin in 1859 (Darwin 1859). Today this theory is widely accepted and selection and adaptation have been studied in all kinds of organisms. Well known examples are antibiotic resistances in bacteria (reviewed in Mazel & Davies 1999; Davies & Davies 2010; van Hoek *et al.* 2011), the morphological adaptation of the peppered moth (*Biston betularia*) to industrialism in Great Britain (reviewed in Cook 2003), or the adaptive radiation of the Darwin finches (*Passeriformes*) on the Galápagos Islands (reviewed in Grant 2003; Abzhanov 2010 and references therein).

Environmental conditions and environmental changes are strong selective forces which lead to adaptation and speciation. The influence of the environment on evolution is especially strong in plants. Plants are sessile organisms and cannot escape unfavourable conditions like animals. Nevertheless, plants occur in almost every region on earth and have undergone many adaptations to extreme environments. Changes in morphology are the most obvious adaptations. Leaf phenology, for example, evolved as a response to colder environments in angiosperms (Zanne *et al.* 2014).

At the molecular level, selection acts upon the genotype. New mutations appear in the genome and over time are either lost or increase in frequency and eventually become fixed. Mutations can occur in the coding region of genes where they may lead to amino acid changes and thus to an altered protein product. An altered protein product might not be able to fulfil its function anymore or it might be able to fulfil it better. Mutations in regulatory regions of genes may affect transcription factor binding motifs and therefore alter gene expression patterns. The fate of a new mutation depends on many factors including chance events (genetic drift) or selection. If the mutation is advantageous, selection will increase its frequency, and if not it will decrease it. These processes leave signatures in the genomes and can be detected by population genetic analyses. Patterns of local adaptation were reported in drought-related genes in Mediterranean pine species (Grivet *et al.* 2011). The *phytochrome A* (*PHYA*) locus shows signatures of local adaptation in a northern population of *Arabidopsis lyrata* (Toivainen *et al.* 2014). Some studies were able to link single mutations to phenotypic or physiological changes and thus to adaptations. In *Arabidopsis*, for example, one mutation in the Na^+ transporter gene, *AtHKT1;1*, leads to a

higher Na^+ accumulation capacity and thus to a better adaptation to saline soils (Baxter *et al.* 2010).

1.2 Wild tomatoes as a model system to study evolution

1.2.1 Natural habitat and adaptations

Wild tomato species (*Solanum* section *Lycopersicon*) are a good model system to study plant adaptation to abiotic stresses. They represent a young group of 13 species within the plant family Solanaceae (Peralta *et al.* 2008). Their diversification occurred about two million years ago (Sarkinen *et al.* 2013). Wild tomato species are distributed from Ecuador to northern Chile and comprise two endemic species on the Galápagos Islands (Spooner *et al.* 2005). Within these regions they occur in diverse habitats and encounter different abiotic stresses. Their natural habitats range from rain forests in Ecuador to hyperarid regions in the Chilean Atacama Desert and from sea level up to almost 4,000 m in the Andean Mountains (Moyle 2008; Chetelat *et al.* 2009). It has been shown that their geographic distribution is mainly determined by abiotic environmental conditions (Nakazato *et al.* 2010). The species are partly overlapping in their distribution (Moyle 2008; Peralta *et al.* 2008). This enables interspecific admixture and introgression. Several parts of the genome were found to be shared between the sister species *S. lycopersicum* and *S. pimpinellifolium* (Nakazato & Housworth 2011). Adaptive introgression of resistance gene alleles was found between *S. peruvianum* and *S. chilense* (Böndel 2010; Hörger 2011; Hörger *et al.* in preparation).

Most wild tomato species (*Solanum* section *Lycopersicon*) show distinct morphological features that reflect adaptation to their natural habitat (Moyle 2008 and references therein). Two wild tomato species, *S. pennellii* and *S. chilense*, developed morphological adaptations to arid environments. *S. pennellii* has cactus-like features which include *e.g.* succulent leaves and shallow spreading roots (reviewed in Moyle & Muir 2010; Nakazato *et al.* 2010). An extremely deep root system allows *S. chilense* to utilize groundwater accumulated at the deep bedrocks during occasional floodings (reviewed in Moyle 2008; Moyle & Muir 2010; Nakazato *et al.* 2010). Furthermore, comparative studies on seed germination demonstrated adaptation to low temperature conditions in a high-altitude population of *S. chilense*. Seed germination responses under low temperatures were better in this population than in several other wild tomato species and tomato cultivars (Scott & Jones 1982, 1985). Plants of

S. chilense also develop smaller and stiffer leaves when grown under low temperatures (T. Nosenko, unpublished data). These findings indicate that *S. chilense* is not only adapted to dry but also to cold environments.

1.2.2 Genomic features, mating system, and genetic variation

Wild tomato species are mostly diploid (2n = 24) and have different mating systems with predominantly self-compatible (SC) or self-incompatible (SI) species (Moyle 2008; Peralta *et al.* 2008). For several species known as SI, SC populations have been reported at the marginal ranges of the species distributions (*e.g. S. habrochaites* Rick *et al.* 1979; *S. pennellii* Rick & Tanksley 1981). For example, one population of the SI species *S. peruvianum* from the southern range of the species distribution is a facultative selfer (Graham *et al.* 2003). This indicates that mating system shifts are an ongoing process in wild tomatoes. The shift from SI to SC is possibly associated with bottlenecks and colonization of new areas.

Several studies analysed genetic variation and demography of wild tomato species. Between species, nucleotide variation depends strongly on the type of mating system. SI species exhibit significantly higher silent nucleotide diversity than SC species (Baudry *et al.* 2001; Roselius *et al.* 2005). Seed banks contribute to the high levels of variation observed in *S. peruvianum* and *S. chilense* (Tellier *et al.* 2011b). Many wild tomato species are substructured with populations scattered throughout their species distribution (*e.g.* Peralta *et al.* 2008; Chetelat *et al.* 2009). Significant isolation by distance was reported for some wild tomato species (*S. pimpinellifolium* and *S. lycopersicum* Nakazato & Housworth 2011; *S. peruvianum* Nakazato *et al.* 2012) as well as for two wild tomato related nightshade species, *S. lycopersicoides* and *S. sitiens*, which occur in sympatry with some wild tomato species in Chile (Albrecht *et al.* 2010). Speciation processes have been analysed and evidence for divergence with gene flow has been found for the sister species *S. chilense* and *S. peruvianum* (Stadler *et al.* 2005; Stadler *et al.* 2008).

1.2.3 The cultivated tomato

An important crop species, the cultivated tomato (*Solanum lycopersicum*), belongs to *Solanum* section *Lycopersicon* (Peralta *et al.* 2008). The production of crops is often limited by abiotic conditions like drought, heat or salinity as well as by combinations of abiotic conditions (*e.g.* Boyer 1982; Mittler 2006). In August 2000, the occurrence of a drought and

heat wave has caused a damage of more than US$ 4.2 billion to the US agriculture (Mittler 2006). The approaching climate change and the increasing human census population size will provide new challenges to the world wide crop production (Godfray *et al.* 2010; Tester & Langridge 2010). Several studies attempted to engineer stress tolerant crops by crossing in favourable genotypes or generating transgenic plants (reviewed in Flowers 2004; Hirayama & Shinozaki 2010; Golldack *et al.* 2011). In tomato, introgression from wild tomatoes is frequently used to improve the cultivated tomato (reviewed in Ranjan *et al.* 2012). Several introgression lines carrying chromosomal fragments of the drought-tolerant wild tomato species *S. pennellii* were shown to be more stress resistant than the cultivar (*e.g.* Gur & Zamir 2004; Gong *et al.* 2010). Therefore, a better understanding of how the wild relatives of tomato are adapted to their environments and the abiotic stresses associated with them, could help to improve crop production. Successfully identified genes or alleles that are involved in adaptation to abiotic stresses could be crossed or transferred into the cultivated tomato to generate more stress tolerant plants.

1.3 Abiotic stress response in plants

At the molecular level, plants respond to abiotic stresses by activating stress response pathways. Abiotic stress response pathways have been studied extensively in *Arabidopsis* and rice (*e.g.* Shinozaki & Yamaguchi-Shinozaki 2000; Xiong *et al.* 2002; Shinozaki *et al.* 2003; Shinozaki & Yamaguchi-Shinozaki 2007; Cutler *et al.* 2010). The ABA-dependent, ABA-independent and *CBF* pathways are among the best characterized and studied pathways in plants (Figure 1.1). The plant hormone abscisic acid (ABA) is accumulated in plant tissues during abiotic stress (reviewed in Wasilewska *et al.* 2008 and references therein). In the ABA-dependent pathway, ABA activates through a double negative regulatory system [PYR/PYL/RCAR --| PP2C --| SnRK2] transcription factors (reviewed in Umezawa *et al.* 2010). Examples are members of the AREB/ABF subfamily of basic leucine zipper transcription factors or MYB transcription factors (*e.g.* Miyazono *et al.* 2009; Cutler *et al.* 2010; Umezawa *et al.* 2010). These transcription factors or regulatory genes bind to *cis*-acting elements of responsive or functional genes and thereby activate their transcription. Well studied *cis*-elements are ABRE (ABA-responsive element) and DRE (dehydration responsive element) (reviewed in Yamaguchi-Shinozaki & Shinozaki 2005). A relatively well studied group of stress responsive genes is represented by genes encoding LEA (late embryogenesis abundant) proteins (reviewed in Allagulova Ch *et al.* 2003; Battaglia *et al.*

2008). One group of *lea* genes, dehydrins, were reported to accumulate during the last stages of embryogenesis as well as in response to drought, salt or low temperature stresses in many plant species. It has been shown that dehydrins might confer stress tolerance by functioning as chaperones in *Arabidopsis* (Kovacs *et al.* 2008a; Kovacs *et al.* 2008b). The ABA-dependent pathway is mainly known to be activated by drought or salt stress. However, *AREB1* (*ABA-response element binding factor 1*), one of the ABA-responsive members of the AREB/ABF subfamily of basic leucine zipper transcription factors, was recently reported to be also induced by low temperature in tomato (Yanez *et al.* 2009). This observation suggests a crosstalk between the cold sensoring genes and ABA-dependent activation of transcription factors. Drought and salt stress can also activate transcription factors in an ABA-independent manner. Interestingly, one of these transcription factors, DREB2 (*dehydration-responsive element binding protein 2*), binds to the same *cis*-regulatory element as the *DREB1*/*CBF* genes. The *CBF* (*C-repeat binding factor*) genes, which are the key enzymes of the *CBF* pathway, are mainly activated by low temperature stress (reviewed in Shinozaki & Yamaguchi-Shinozaki 2000; Shinozaki *et al.* 2003; Thomashow 2010). Additionally several other gene networks mediating abiotic stress responses were hypothesized as for example the NtNAK/NtCK25 pathway in tobacco (Kodama *et al.* 2009).

Thousands of abiotic stress-related genes have been identified in the genus *Solanum* using a large scale of transcriptomic approaches including cDNA libraries and microarray. Most of these data were generated for plants exposed to either abiotic stress or abscisic acid (*e.g.* Cohen & Bray 1990; Godoy *et al.* 1990; Chen *et al.* 1993; Wei & O'Connell 1996; Yanez *et al.* 2009) or by mutant screening (*e.g.* Burbidge *et al.* 1999; Borsani *et al.* 2001; Borsani *et al.* 2002). The majority of these studies focused on the cultivated tomato *S. lycopersicum* (*e.g.* Cohen & Bray 1990; Godoy *et al.* 1990; Zegzouti *et al.* 1997). A few studies used the wild tomato species *S. chilense* (Chen *et al.* 1993; Chen *et al.* 1994). Recently, several comparative studies have been conducted that included cultivated tomato, wild species of *Solanum* and introgression lines (*e.g.* Gong *et al.* 2010; Liu *et al.* 2012).

Population genetic analyses of these abiotic stress genes could help to understand how these genes evolve and to evaluate their potential for crop production improvement.

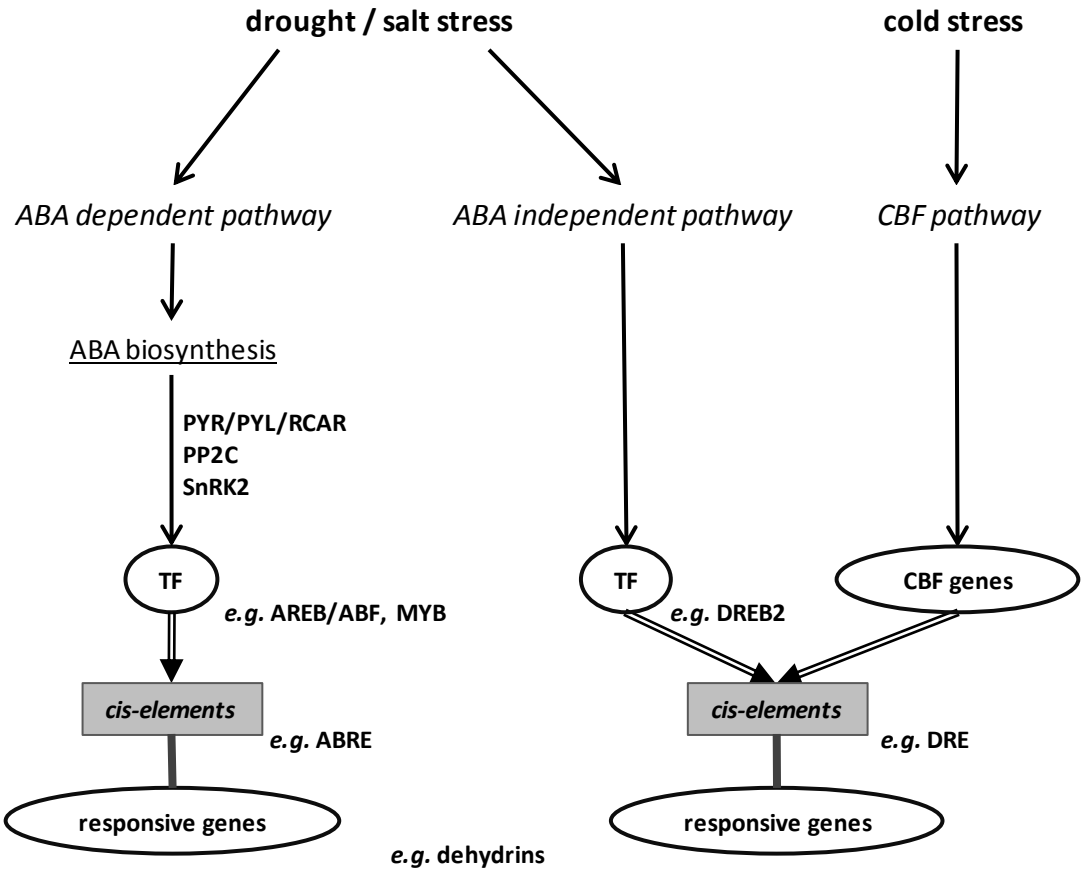

Figure 1.1: Schematic overview over the molecular abiotic stress response in plants. Drought, salt and cold stress activate the respective pathways. Transcription factors (TF) bind to *cis*-elements of responsive genes. The other abbreviations are explained in the main text.

1.4 Molecular signatures of local adaptation to abiotic stresses in wild tomato species

The evolution of several abiotic stress-related genes has been analysed in the drought-tolerant wild tomato species *S. chilense* and its closely related sister species *S. peruvianum* (Peralta *et al.* 2008; Stadler *et al.* 2008). Estimates of the divergence time between *S. peruvianum* and *S. chilense* range from 0.5 to 5 million years ago (≤ 0.55 mya Stadler *et al.* 2008; 0.73 (or 5.1) mya Naduvilezhath *et al.* 2011; 0.7 (or 4.6) mya Mathew *et al.* 2013; 0.74 mya Sarkinen *et al.* 2013). The two species overlap in their geographic distribution: *S. peruvianum* occurs from central Peru to northern Chile and *S. chilense* from southern Peru to northern Chile, but reaches farer to the south than *S. peruvianum* (Peralta *et al.* 2008).

The key gene of the ABA biosynthesis, *LeNCED1*, which encodes a 9-cis-epoxy-carotenoid dioxygenase (Thompson *et al.* 2000), was reported to evolve under purifying selection (Xia *et al.* 2010). Another gene, the dehydrin *pLC30-15*, exhibits a haplotypic pattern in a *S. chilense* population which was sampled in a relatively mesic environment in Quicacha in southern Peru (Xia *et al.* 2010). This pattern can be explained by diversifying selection and with one haplotype putatively originating from the sister species *S. peruvianum* (Xia *et al.* 2010). Arunyawat *et al.* (2007) reported a clinal pattern in nucleotide diversity for an alcohol dehydrogenase class III gene (*CT208*) in *S. chilense*. This pattern is consistent with an ongoing selective sweep scenario.

Two transcription factor gene families, the *Asr* (*ABA/water stress/ripening induced*) gene family and the *CBF* (*C-repeat binding factor*) gene family, which are involved in drought and cold stress response, respectively, have been analysed (Fischer *et al.* 2011; Mboup *et al.* 2012). One member of the *Asr* gene family, *Asr1*, was reported to evolve under strong purifying selection while another member, *Asr4*, showed a pattern consistent with local adaptation in a *S. chilense* population from an extremely dry environment near Tacna in southern Peru (Fischer *et al.* 2011). The *CBF* genes were also reported to have different evolutionary histories (Mboup *et al.* 2012). *CBF3* was found to evolve under purifying selection and *CBF2* under balancing selection. The *CBF2* gene further exhibits a trans-species polymorphism which could be linked to allele-specific gene expression (Mboup *et al.* 2012). These findings indicate that transcription factors although being involved in many interactions can show a dynamic evolutionary history. However, these observations may also be explained by the fact that these transcription factors, like many other transcription factors, form gene families. Gene family evolution is characterized by different selective constraints acting upon the different gene copies: as long as one copy maintains the original function, the other copies are free to evolve and thus can potentially acquire new functions (reviewed in Flagel & Wendel 2009; Magadum *et al.* 2013).

These studies have shown that abiotic stress-related genes can evolve under different selective constraints and that further studies including more genes will be required to fully understand how evolution is acting upon stress response pathways.

1.5 The aim of this study

Previous studies showed that the wild tomato species *S. chilense* is a valuable candidate to study demographic processes and local adaptation to abiotic stress conditions. This species is native to southern Peru and northern Chile where it exhibits a patchy population distribution across desert regions, at high altitudes and at the coast (Peralta *et al.* 2008; Chetelat *et al.* 2009). Therefore, different populations encounter different abiotic stresses including drought, cold, salinity and also combinations of these factors. In comparison to other wild tomato species, *S. chilense* can grow under the driest and coldest conditions (Moyle 2008). Signatures of local adaptation possibly associated with either drought or cold stress were found in abiotic stress-related genes (Xia *et al.* 2010; Fischer *et al.* 2011; Mboup *et al.* 2012). Differences in the degree of genetic variation were observed between populations (Arunyawat *et al.* 2007). Although being promising and interesting, these results were obtained using relative small data sets with either three or four populations and five to seven plants per population, as they were generated by traditional Sanger sequencing.

This study employed the advantages of the newly arisen sequencing technologies (reviewed in Mardis 2008; Shendure & Ji 2008; Metzker 2010) to generate a big data set. 30 genes from an exhaustive population sample of *S. chilense* were sequenced. These genes include abiotic stress response genes and genes that are not involved in any abiotic stress regulation pathway. This data set allowed me to further extend the analysis on demography and selection in the wild tomato species *S. chilense*. To my knowledge, this is one of the biggest population-based samples that were sequenced in a wild tomato species.

Demographic history of *Solanum chilense*

In the first part of this study, I investigated the demographic history of *S. chilense*. Previous work on demography in wild tomatoes showed that population structure and isolation by distance is common (*e.g.* Nakazato & Housworth 2011; Nakazato *et al.* 2012). Different degrees of genetic variation were observed in wild tomato populations (*e.g.* Arunyawat *et al.* 2007). Theory predicts that populations in the marginal ranges of the species distribution, in which they encounter sub-optimal conditions, tend to exhibit lower genetic variation than populations in the central distribution (reviewed in Eckert *et al.* 2008). Given the environments in which wild tomatoes occur, this scenario is likely for wild tomatoes. Based on genetic data, it was hypothesized that *S. chilense* could be derived from *S. peruvianum* or a *S. peruvianum*-like ancestor (Baudry *et al.* 2001). Since the species

distribution of *S. peruvianum* is north of *S. chilense* (overlapping in the Peruvian-Chilean border region), this would imply that *S. chilense* originated in the northern part of its current distribution and migrated to the south. This question was addressed with this data set.

Local adaptation to abiotic stresses in *Solanum chilense*

The second part of this study focused on local adaptation in *S. chilense*. Different approaches were applied to address this topic. The basic approach was to analyse the abiotic stress-related genes – the so called candidate genes – in comparison to the genes that are not related to the abiotic stress response – the so called reference genes. The reference genes represent the genomic average. Whenever a candidate gene differs from the genomic average, selection and adaptation may be assumed. A candidate gene approach was successfully used before in wild tomatoes (Xia *et al.* 2010; Fischer *et al.* 2011; Mboup *et al.* 2012). This approach allowed me to detect adaptation events within populations.

The abiotic stress-related genes are from different layers of the abiotic stress response. Some of them were reported to induce the expression of others and can therefore be linked in a gene network. According to pathway theory, 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 (reviewed in Olson-Manning *et al.* 2012). A recent study on the *Arabidopsis* protein interactome has shown 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 (Alvarez-Ponce & Fares 2012). Therefore, this data set enabled me to gain first insight into the evolution of pathways or gene classes in wild tomatoes.

Another aspect of my dissertation is to relate signatures of local adaptation to population structure. For species with a metapopulation structure, *i.e.* that consist of many populations with varying migration rates between them, contrasting theories exist in literature. One hypothesis predicts that small populations have a lower potential for adaptation than large populations (Willi *et al.* 2006). Therefore, local adaptation is supposed to be more frequent in large populations which usually occur in the centre of the species range. Since the effect of genetic drift is weaker in large populations, selection will dominate and decide upon the fate of new, possibly advantageous, mutations. A meta-analysis in plants showed that local adaptation is more common in large populations (Leimu & Fischer 2008). Another hypothesis predicts that populations that recently colonized a new territory should more likely show local adaptation events (*e.g.* Innan & Kim 2008). This is due to the fact that the new territory is different from the old one and that therefore founder

individuals are confronted with new conditions and only those that are adaptively predisposed will manage to establish there. Therefore, colonization of a new territory is associated with an initial local adaptation event. This has been shown in sticklebacks and their parallel colonization from marine water to freshwater (Colosimo *et al.* 2005). This sample of *S. chilense* comprises populations from the centre of the species distribution and from the marginal ranges. Therefore, this data set allowed me to reveal which theory applies to wild tomatoes.

Analysis of the consensus sequences

In the third part of the study, a new approach to analyse a pooled sequence data set was tested: the analysis of the consensus sequences. A consensus sequence has at each position the nucleotide that the majority of reads has at this position. Consensus sequences of pooled sequence data show what the majority of alleles have and thus can be regarded as average alleles for the populations. This analysis focused on three questions. First, gene evolution was investigated on the species level. This is similar to a species wide sampling approach in which one allele is randomly chosen per population (Stadler *et al.* 2009; Tellier *et al.* 2011b). Species wide samples can be used to analyse the collecting phase of the coalescence of a substructured species (Pannell 2003). Unlike the randomly chosen allele, the consensus sequence represents the average allele of the population and thus is a more accurate representative of the population. Second, gene trees of concatenated consensus sequences were constructed to infer population relationships. And third, the alignments were screened for nonsynonymous SNPs or amino acid changes. They represent high frequency nonsynonymous SNPs in the whole data set. These nonsynonymous SNPs can be considered as being advantageous and thus under positive selection, since they were not eliminated by purifying selection. Finally, the results of the consensus sequence analysis were compared to the analyses of the whole data set to evaluate this new approach.

Phenotypic responses to salt stress in *Solanum chilense*

Several *S. chilense* populations grow at the coast or near the salt lakes in the Atacama Desert (Peralta *et al.* 2008; Chetelat *et al.* 2009). Therefore, salinity tolerance is expected in these populations. This experiment tested if plants from habitats with high soil salinity perform better under salt stress than plants from other habitats. The observations of this experiment helped to understand some of the patterns observed on the molecular level.

CHAPTER 2: MATERIAL AND METHODS

2.1 Sequence evolution in *Solanum chilense*

2.1.1 Plant material and plant growing

The wild tomato species *Solanum chilense* is a diploid, self-incompatible perennial plant (Peralta *et al.* 2008). It is native to a broad range of different habitats in southern Peru and northern Chile (Peralta *et al.* 2008): from coast to high altitudes in the Andean mountains and from mesic to hyperarid desert regions in the Atacama Desert (Chetelat *et al.* 2009). The Tomato Genetics Resource Center (TGRC) at the University of California at Davis (http://tgrc.ucdavis.edu) has a comprehensive collection of wild tomato species and their relatives. Plant breeders at the TGRC try to maintain populations of about 50 plants in each generation and perform mass sib pollinations using a bulk pollen sample collected from all plants to obtain seeds (R. Chetelat, personal communication). This collection includes more than 100 populations of *S. chilense* (105 populations on Nov 20th 2013, 146 populations on Dec 31st 2011). Detailed information about the collection and the environmental conditions at the collection site are available for each population. Based on these information, 23 populations were chosen and seeds were obtained from the TGRC. These 23 populations are distributed over the whole species range and cover all the diverse habitats and climate conditions this species encounters (Figure 2.1, Table A1.1). This population sample comprises 13 populations from the centre of the species range in which they occur at altitudes ranging from 200 to 3400 m. Four populations were sampled close to the coast in the southern range of the species distribution (LA2750, LA2932, LA4108, LA4107). Four other populations were sampled in the San Pedro de Atacama high altitude region (2440 to 2980 m), which is also in the southern range of the species distribution (LA4332, LA4118, LA4119, LA2880). Two populations, LA1930 and LA3784, are from the most northern range of the species distribution. Therefore, this exhaustive population sample can be regarded as a good representation for the entire species. These populations have experienced between one and five cycles of regeneration *ex situ*, *i.e.* they were for one to five generations in the green houses of the TGRC (R. Chetelat, personal communication). Some of these populations are part of a species wide sampling approach applied in previous studies on seed banks and biotic stress resistance gene evolution (Böndel 2010; Hörger 2011; Tellier *et al.* 2011b;

Hörger *et al.* in preparation). For outgroup comparisons, one population of each *S. ochranthum* (LA2682) and *S. lycopersicoides* (LA2951) was chosen. The population from *S. ochranthum* is from a mesic environment in Cusco, Peru, while the *S. lycopersicoides* population comes from a dry region in Tarapaca, Chile, and grows in the same region as many *S. chilense* populations. According to a phylogenetic study based on a fragment of the granule-bound starch synthase I (*GBSSI*) gene, *S. ochranthum* together with *S. juglandifolium* are the closest outgroup to the wild tomatoes (*Solanum* section *Lycopersicon*), while *S. lycopersicoides* together with *S. sitiens* are basal to this group (Peralta & Spooner 2001). Recent estimations place the split between the wild tomatoes and *S. ochranthum*/*S. juglandifolium* to 5.59 million years ago and the split between the wild tomatoes and *S. lycoperisocides*/*S. sitiens* to 5.95 million years ago (Sarkinen *et al.* 2013).

Between 40 and 55 seeds per population were treated with 2.7 % NaOCl for 20 minutes to initiate germination. Germination rates were observed to vary between populations. Therefore, a sufficient number of seeds was required to obtain at least 25 plants of each population for the experiment. Treated seeds were incubated on wet filter paper in the dark at room temperature for up to two weeks. Approximately one week old seedlings (*i.e.* root > 2 cm, hypocotyl > 2 cm, leaves clearly visible) were planted into normal garden soil and grown at 18 – 22 °C with a 16 hour light and eight hour dark cycle for five to eight weeks until the plants had six leaves or more. Leaf material was sampled from every individual, shock frozen in liquid nitrogen and stored at -80 °C.

Figure 2.1: Map of southern Peru and northern Chile with the 23 *S. chilense* populations. Latitude and longitude information are from the Tomato Genetics Resource Center, UC Davis (TGRC, http://tgrc.ucdavis.edu/).

2.1.2 Choice of genes and primer design

In this study, 30 genes – 16 candidate genes and 14 reference genes – were sequenced and analysed.

The 16 candidate genes were chosen from the literature based on their involvement in the abiotic stress response and on the availability of sequence data from *Solanum* section *Lycopersicon* or related Solanaceae species like potato (*S. tuberosum*) or tobacco (*Nicotiana tabacum*) either on GenBank (http://www.ncbi.nlm.nih.gov/genbank/) or printed within the publication. 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 A1.2). This classification is common for abiotic stress response genes in plants (*e.g.* Shinozaki & Yamaguchi-Shinozaki 2007). Two of the regulatory genes, *AREB1* and *JERF1*, were reported to induce the expression of some of the functional genes in this study (Wu *et al.* 2007; Yanez *et al.* 2009; Orellana *et al.* 2010). Therefore, these genes can be linked in a gene network (Figure 2.2). 13 reference genes come from Tanksley *et al.* (1992), nine of them have been subject to previous studies on genetic variation and demography in wild tomato species (*e.g.* Baudry *et al.* 2001; Stadler *et al.* 2005; Tellier *et al.* 2011b). Note that in these previous studies they were called "reference loci", but for convenience they will be called "reference genes" in this study. The granule-bound starch synthase I (*GBSSI*) gene was used previously to examine the phylogenetic relationships of wild tomato species (Peralta & Spooner 2001). The nine previously used genes were chosen to allow comparisons with existing data, the other five to provide an even distribution of the genes over the entire genome, with at least one gene per chromosome (Table A1.3).

The primers for 16 genes (*dhn1*, *DREB3*, *ER5*, *JERF1*, *JERF3*, *His1*, *le25*, *LTP*, *NtC7*, *TAS14*, *TPP*, *TSW12*, *CT021*, *CT114*, *CT182*, *CT192*) were designed based on available sequences from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and the SOL genomics network (The Tomato Genome Sequencing Consortium 2012; http://solgenomics.net/; Table A1.4). These sequences include mRNA and genomic sequences from the cultivated tomato, *S. lycopersicum* cv. Heinz, contigs from the preliminary *S. pimpinellifolium* genome (The Tomato Genome Sequencing Consortium 2012; http://solgenomics.net/) and sequences from the *S. peruvianum de novo* transcriptome BLAST dataset (Park *et al.* 2012; http://solgenomics.net/). In case of SNPs between the species, the primer sequence was designed to match the *S. peruvianum* sequence since *S. peruvianum* is sister species to *S. chilense* (Peralta *et al.* 2008; Stadler *et al.* 2008). The primers for the candidate genes were

designed around the start and stop codons, the primers for the reference genes to bind in the coding region and to amplify fragments between 1500 base pairs (bp) and 1800 bp. All primers were evaluated using *Netprimer* (http://www.premierbiosoft.com/netprimer/). The primers for *AREB1* and *AREB2* are from Yanez *et al.* (2009) and Orellana *et al.* (2010), respectively, for *pLC30-15* from Steige (2011), and for *GBSSI* from (Peralta & Spooner 2001). The primers for the remaining ten reference genes were taken from previous population genetic studies (Arunyawat *et al.* 2007; Hörger *et al.* in preparation; 3' primer *CT208* T. Städler, W. Stephan unpublished work; Table A1.5). Based on the cultivated tomato genome, the length of the candidate genes varies between 520 bp and 4316 bp (mean 1376.38 bp) and between 760 bp and 1941 bp (mean 1467.69 bp) for the reference genes. The total sequence length of the candidate and reference genes is 22,022 bp and 20,542 bp, respectively. The total sequence length over all 30 genes is 42,564 bp. These numbers include noncoding and coding regions.

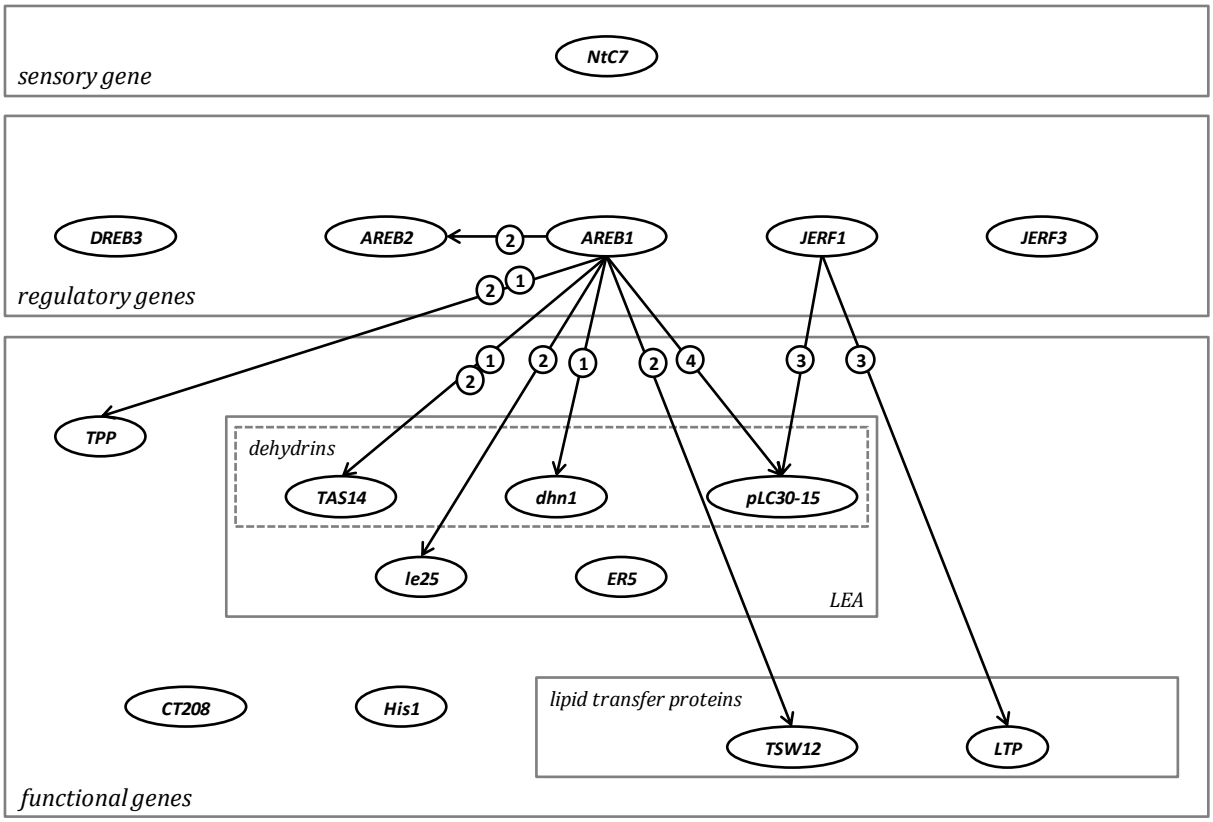

Figure 2.2: Schematic overview over the candidate genes. Candidate genes (Table A1.2) were classified as sensory, regulatory or functional genes according to Shinozaki and Yamaguchi-Shinozaki (2007). Boxes indicate the layers of the stress response (sensory genes, regulatory genes, functional genes) or the gene classes (dehydrins, LEA, lipid transfer proteins). Two transcription factors were reported to induce the expression of other genes: ① Yanez *et al.* (2009), ② Orellana *et al.* (2010), ③ Wu *et al.* (2007), and ④ SOL genomics network (http://solgenomics.net/).

2.1.3 Sequencing approach for *Solanum chilense*

The 30 genes were sequenced in a pooling approach for the 23 *S. chilense* populations with 25 individuals per population (Figure 2.3). Pooling approaches provide a cost effective approach to study a high number of populations by neglecting individual information and were successfully applied to detect selection in animals (*e.g.* Obbard *et al.* 2009; Amaral *et al.* 2011) and in plants (*e.g.* Turner *et al.* 2010; Zhou *et al.* 2011).

Genomic DNA was first isolated from each individual plant from frozen leaf material using the DNeasy Plant Mini Kit (*Qiagen*) following the manufacturer's protocol. DNA concentrations were quantified using the NanoDrop 1000 Spectrophotometer V3.7 (*Thermo Scientific*) and diluted to a concentration of approximately 10 ng/µl. For each population five genomic DNA pre-pools (each pre-pool with five individuals) were mixed. Each gene was amplified from each pre-pool by the polymerase chain reaction (PCR) using the Phusion® High-Fidelity DNA Polymerase kit (*Finnzymes*, distributed by *New England BioLabs, Inc.*). Annealing temperature varied depending on primer combination (Tables A1.4, A1.5). Agarose gel electrophoresis was performed after each PCR to confirm amplification success. For this purpose, 1 µl of the PCR products were mixed with 4 µl 1 x loading dye and loaded on a 1 % agarose gel containing ethidiumbromid. The five PCR product pre-pools per gene and population were brought together during PCR product purification with the MinElute® Gel Extraction Kit (*Qiagen*) following the manufacturer's protocol, resulting in one PCR product pool per population and gene. All 30 PCR product pools per population were mixed in equimolar quantities and concentrated to ≥ 200 ng/µl using the Amicon® Ultra-0.5 Centrifugal Filter Devices (*Millipore*) following the manufacturer's protocol. The final PCR product quantity and quality was assessed using the NanoDrop 1000 Spectrophotometer V3.7 (*Thermo Scientific*).

DNA library construction and high-throughput sequencing on a Genome Sequencer Illumina HiSeq2000 were performed by the GATC Biotech AG (Konstanz, Germany). For each population one library was constructed and tagged. All libraries were sequenced in one lane. The expected total amount of data was 36 GB and per library between five and nine million paired-end reads of 100 bp were expected. FastQ data files for each library, *i.e.* population, were obtained from the sequencing company.

In order to evaluate the whole sequencing approach, one population, namely LA1968, was sequenced two times starting from the five genomic pre-pools.

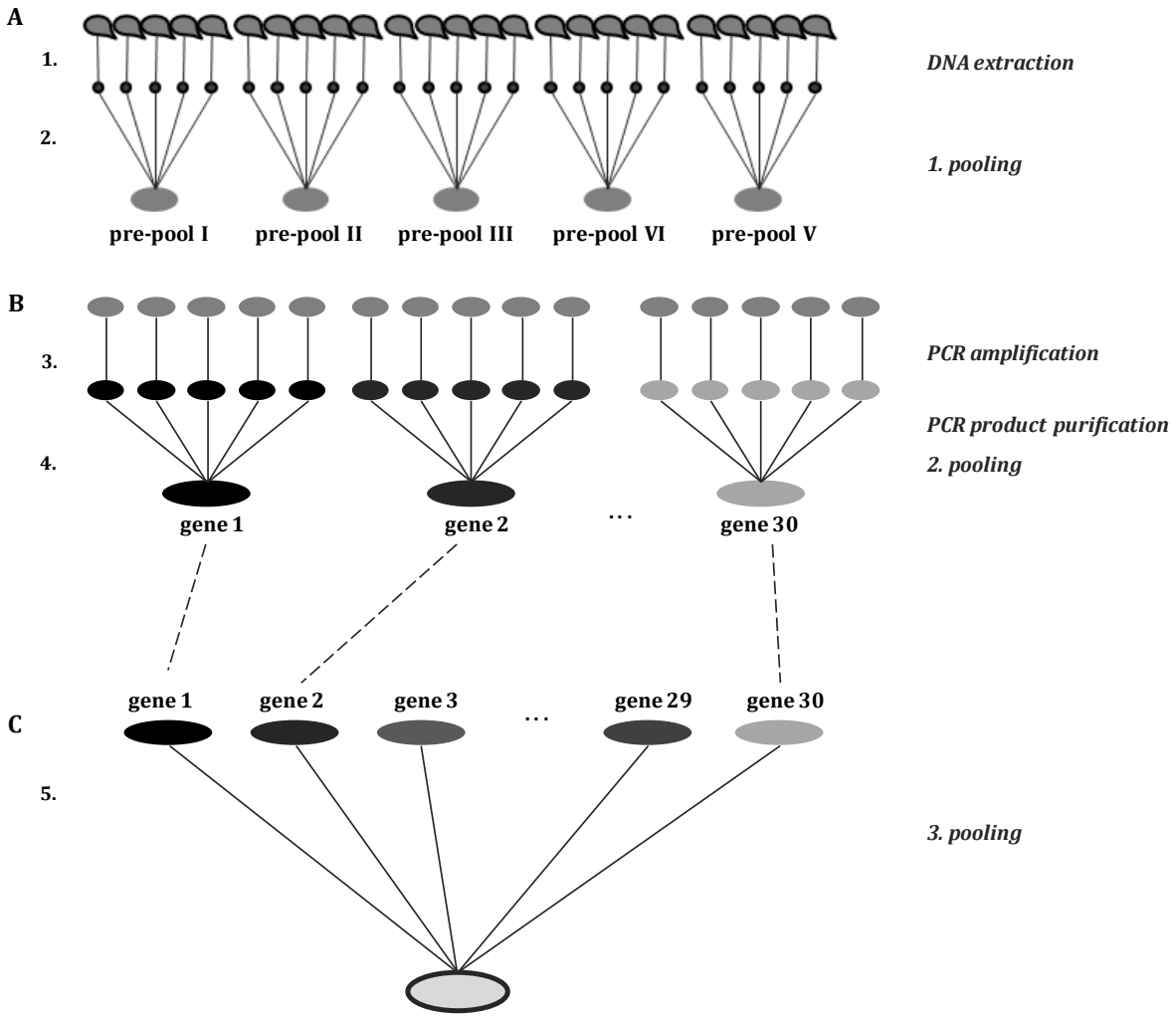

Figure 2.3: Schematic overview over the pooling procedure. A) Genomic DNA pooling of one population. DNA was extracted from each of the 25 individual plants and diluted to app. 5 ng/µl (1.). Five diluted DNA samples were pooled (2.) resulting in five DNA pre-pools per population. B) PCR amplification per gene. Each gene was amplified from each pre-pool (3.) resulting in five PCR products per gene. The five PCR products of each gene were brought together during PCR product purification (4.) resulting in one PCR product pool per gene. C) Pooling for 30 genes of one population. The 30 PCR product pools per accession, representing the 30 genes, were pooled in equimolar quantities (5.) before being sequenced.

2.1.4 Sequencing of the outgroup species *S. ochranthum* and *S. lycopersicoides*

Genomic DNA was isolated from one individual of each outgroup species as described for *S. chilense*. Genes were amplified from each individual as described for *S. chilense*. Except for *CT021* in *S. ochranthum*, the same primer combinations as for *S. chilense* were used (Tables A1.4 - A1.6). PCR products were purified with the enzyme mix ExoSAP-IT™ (*Amersham*) and directly sequenced with the primers from the original PCR reaction. Sequencing was performed according to the Sanger sequencing protocol using the DNA analyzer ABI 3730 (*Applied Biosystems & Hitachi*). Sequences were edited with the program Sequencher™ 4.8 (*©1991 – 2007 Gene Codes Corporation and its licensers*). In cases of observed heterozygosity, a cloning approach was applied. Cloning of PCR products was performed using the StrataClone™ Blunt PCR Cloning Kit (*Stratagene*). Multiple clones were used for reamplification of the allele of interest with the *Taq* DNA Polymerase Kit (*Invitrogen*) and subsequent sequencing (DNA analyzer ABI 3730, *Applied Biosystems & Hitachi*) using either the PCR primers or standard M13 primers until at least one allele could be successfully identified. For several genes, additional sequencing primers were designed based on the previously obtained sequences to obtain clean sequences covering the entire fragment (Table A1.6). Altogether, outgroup sequences of both species were obtained for 27 genes. Attempts to amplify *TAS14* from *S. ochranthum* (LA2862) and *His1* and *CT021* from *S. lycopersicoides* (LA2951) failed.

2.1.5 Sequence data analyses

2.1.5.1 Data assembly

The short paired-end reads of the *S. chilense* populations generated by Illumina sequencing were mapped with Stampy v1.0.20 (Lunter & Goodson 2011). Stampy was chosen since it showed a higher mapping accuracy than other programs for regions with high genetic variation in a comparative study (Nookaew *et al.* 2012) and also because it deals better with gap regions in *Drosophila melanogaster* (A. Catalan personal communication, Catalan *et al.* 2012). The default parameters for paired-end read mapping and a substitution rate of 0.01 towards the reference sequence were used. The substitution rate reflects the expected divergence from the reference sequence.

As reference sequence the corresponding fragments of the cultivated tomato genome (*S. lycopersicum* cv. Heinz, release SL2.40) from the SOL genomics network

(http://solgenomics.net/) were used. For three genes, *CT021*, *NtC7* and *TAS14*, *S. chilense* sequences were used since preliminary tests revealed different length in *S. chilense* compared to *S. lycopersicum* (K. Böndel, W. Stephan unpublished data). Additionally, for *TAS14* different lengths ranging between 600 and 800 bp between *S. chilense* populations were observed. Reference sequences for these genes were generated by PCR amplification, cloning and Sanger sequencing from the populations LA1968 (*CT021*), LA2747 (*NtC7*, *TAS14* long) and LA3786 (*TAS14* short) similar to the procedure described for the outgroup species (see above). The SAMtools program (Li *et al.* 2009) was used to generate a pileup file from the assembly. The pileup file summarizes the mapping results at each position. Pileup files were generated per population. The source code was altered to allow for a sequencing depth of 1,000,000, since the default value for this command is 8,000 which turned out to be not enough for the generated sequence data set. A set of PERL scripts was written to extract all positions with SNP information from the pileup file and to transform it into a table with single nucleotide polymorphism (SNP) information for each position. A polymorphic position was defined as a position with a sequencing depth of 2,000 or higher, a minimum base quality of 20, a minimum mapping quality of 20, and the mutation present in more than 1 % of the reads.

The resulting per population SNP tables were used directly for some of the population genetic analyses, but also to compile consensus sequences and so called 'artificial' alignments. Consensus sequences were compiled per gene and population and represent the average allele of the population, *i.e.* for each position the nucleotide that is present in the majority of the alleles in the population (frequency of > 0.5) is shown (Figure 2.4). These consensus sequences were used to analyse the genes on the species level, similar to a species wide sample (Stadler *et al.* 2009), and to obtain first indications for local adaptation events. For the 'artificial' alignments the alleles were generated based on the SNP information for each position: for example if position 10 is a singleton (49 A's and 1 T), then one allele receives the T and the other 49 alleles an A. If position 20 has a SNP in a frequency of 0.5 (25 C's and 25 G's), then 25 alleles receive a C and the other 25 alleles a G. These alignments do not give any information about phase, *i.e.* linkage between SNPs, but can be used for analyses that do not require phase information. Consensus sequence alignments and 'artificial' alignments were merged manually with the outgroup sequences (*S. ochranthum*, *S. lycopersicoides*, and *S. lycopersicum*) in the program Mesquite v2.74 (Maddison & Maddison 2010). If not noted otherwise, 'artificial' alignments and consensus sequence alignments were analysed with DnaSP v5.10 (Librado & Rozas 2009).

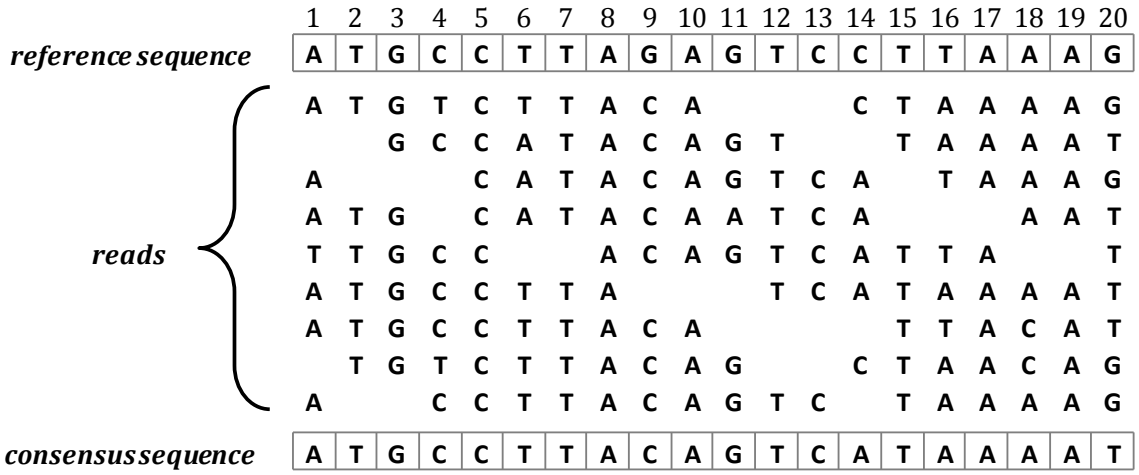

Figure 2.4: Schematic compilation of a consensus sequence. The example shows how a consensus sequence is compiled. For each nucleotide position the consensus sequence receives the nucleotide that the majority of reads carry. *E.g.* at position 4 five reads carry a C and two reads a T, therefore the consensus sequences has an C.

2.1.5.2 Summary statistics of the whole data set

DNA polymorphism and divergence

The DNA polymorphism for all sites was estimated directly from the SNP table with the Watterson estimator, Θ_W (Watterson 1975), which is based on the number of segregating sites observed in a given sample of size *n*, and with the nucleotide diversity π, which is the average number of pairwise nucleotide differences (Nei 1987). The 'artificial' alignments were used to calculate both estimators for silent sites, and to calculate π for synonymous (π_s) and nonsynonymous (π_a) sites. Further the ratio between nonsynonymous and synonymous nucleotide diversity was assessed (π_a/π_s).

The divergence, *K*, from the outgroup species *S. ochranthum* and *S. lycopersicoides* as well as the cultivated tomato, *S. lycopersicum*, was calculated from the 'artificial' alignments for all sites, synonymous (K_s) and nonsynonymous (K_a) sites.

To find outlier loci and populations, the density of each polymorphism and divergence value was plotted and the 2.5 % quantiles were computed using R (R Development Core Team 2005).

Neutrality test: Tajima's D

Neutrality tests were applied to find deviations from the standard neutral model of evolution. If the null hypothesis can be rejected, either some selective pressure or demography has acted on a gene of interest.

The neutrality test statistic Tajima's *D* (D; Tajima 1989) was calculated directly from the SNP table. This test is based on the two estimators of DNA polymorphism, Θ_W and π, and uses their different ways to estimate DNA polymorphism: while Θ_W takes only the number of segregating sites into account, π also considers their frequencies. Tajima's *D* assumes that these two estimators should be equal under neutrality. More precisely, Tajima's *D* takes the difference, *d*, between π and Θ_W and divides it by the square root of the variance of *d* to normalize the test statistic. Negative Tajima's *D* values indicate an excess of low or high frequency variants which are expected under negative (excess of low frequency variants) or positive (excess of high frequency variants) selection as well as for an expanding population (excess of low frequency variants). Positive Tajima's *D* values indicate an excess of intermediate frequency variants which indicates either balancing selection, substructure within the sample or a recent admixture event.

Outliers were determined computing the 2.5 % quantiles of the density distribution (R, R Development Core Team 2005).

2.1.5.3 Regressions and correlations

In order to test for correlations in this data set, regressions were performed and the Spearman's rank correlation ρ was calculated in R (R Development Core Team 2005). Positive ρ values show positive correlation and negative ρ values negative correlations. ρ values close to 0 indicate no correlation while ρ values close to 1 or -1 indicate correlation, *i.e.* values are close to being identical.

2.1.5.4 Population differentiation and isolation by distance

Population differentiation

The degree of differentiation between the *S. chilense* populations was assessed using the F_{ST} statistics calculated for each population pair directly from the SNP tables (Hudson *et al.* 1992). This statistics compares the total genetic variation of two populations with the variation within each population. In addition to the population differentiation among the *S. chilense* populations, F_{ST} was calculated for each *S. chilense* population compared to a

species wide sample of the sister species *S. peruvianum* (Tellier *et al.* 2011b; Hörger *et al.* in preparation) for the nine reference genes that are present in both data sets. These genes include *CT066*, *CT093*, *CT143*, *CT166*, *CT179*, *CT189*, *CT198*, *CT251*, and *CT268*. Note that *CT208*, although present in both data sets, was not included into this analysis due to the possible selective events in *S. chilense* (Arunyawat *et al.* 2007). F_{ST} was further calculated for each *S. chilense* population compared to pooled population samples of two other wild tomato species, *S. arcanum* and *S. habrochaites* (Tellier *et al.* 2011a) for the reference genes which are present in both data sets (*CT066*, *CT093*, *CT166*, *CT179*, *CT198*, *CT251*, and *CT268*). These interspecies F_{ST} values were calculated from the 'artificial' alignments, which were merged with the *S. peruvianum*, *S. arcanum* and *S. habrochaites* sequences previously generated by Sanger sequencing (Tellier *et al.* 2011a; Tellier *et al.* 2011b; Hörger *et al.* in preparation).

Isolation by distance

To assess isolation by distance in *S. chilense*, the pairwise F_{ST} values were plotted against the geographic distance, regressions were computed and a Mantel test was performed in R (Mantel 1967; R Development Core Team 2005). Geographic distances were inferred from the latitudinal and longitudinal decimals from the TGRC. This analysis was performed for the average pairwise F_{ST} values of all 30 genes and for the average pairwise F_{ST} values of the 14 reference genes. Further this analysis was performed for subsets of the whole data set. The same analyses were performed for pairwise F_{ST} values plotted against the pairwise distance in altitude. Altitudinal data was obtained from the TGRC.

2.1.5.5 Detecting selection in Solanum chilense

McDonald-Kreitman test statistic

The McDonald-Kreitman test (McDonald & Kreitman 1991) was calculated for the 30 genes from the 'artificial' alignments. This test assumes that under neutrality the ratio between nonsynonymous and synonymous fixed differences between species should be the same as the ratio between nonsynonymous and synonymous polymorphisms within species. An excess of nonsynonymous fixed differences should indicate positive selection while an excess of nonsynonymous polymorphisms should indicate balancing selection. If available, both outgroup species, *S. ochranthum* and *S. lycopersicoides*, were used.

The proportion of adaptive amino acid changes

The proportion of adaptive amino acid changes, α, was estimated based on the π_a/π_s and K_a/K_s ratios (Rand & Kann 1996; Fay *et al.* 2001). α is calculated as $1 - ((\pi_a/\pi_s)/(K_a/K_s))$. Adaptive evolution is present when $K_a/K_s > \pi_a/\pi_s$ and thus, $\alpha > 0$. α was estimated with divergence from both outgroup species, *S. ochranthum* and *S. lycopersicoides*.

BayeScan

The program BayeScan v2.1 (Foll & Gaggiotti 2008) was used to detect outlier SNPs which are candidates of natural selection in the data set. This program uses allele frequencies and calculates SNP specific F_{ST} coefficients. These F_{ST} coefficients consist of a population specific component and a SNP specific component, which is α. If the SNP specific component is required to explain the observed pattern, then selection acting upon this SNP can be assumed. A positive value of α indicates diversifying, *i.e.* positive, selection while a negative value of α either indicates balancing or purifying selection. BayeScan was run once on all SNPs, including multiple hits, and once on all SNPs excluding multiple hits. The latter was done, since the underlying model of BayeScan assumes the infinite sites model. BayeScan further calculates a q value for each SNP. A false discovery rate (FDR) of 5 % was applied to identify the outlier SNPs.

2.1.5.6 Analysis of the consensus sequence data set

The consensus sequence data set was analysed for three purposes: 1) to assess gene evolution on the species level, 2) to infer the relationships between the populations and 3) to identify candidates for local adaptation. For the first purpose, the same summary statistics as for the whole data set (*i.e.* polymorphism, divergence, neutrality test) were calculated for the consensus sequence data set. For more details see also Brücher (2013). For the second purpose, phylogenetic trees of concatenated consensus sequence alignments were constructed of all 30 genes, of the 14 reference genes and of the 16 candidate genes. The outgroup species of *S. ochranthum* and *S. lycopersicoides*, and the cultivated tomato, *S. lycopersicum*, were included into this analysis. Concatenations were generated using FASconCAT (Kuck & Meusemann 2010). Maximum likelihood (ML) trees were constructed for each of the three concatenated alignments using RaxML v7.2.7 (Stamatakis *et al.* 2005) under the Generalized time-reversible (GTR) model of sequence evolution (Tavaré 1986). Node support for the ML trees was evaluated with 100 rapid bootstrap replicates (Stamatakis *et al.* 2008). For the third purpose, the alignments were screened for amino acid

changes. If a population has a nonsynonymous SNP, *i.e.* an amino acid change, in its consensus sequence, then this SNP has a frequency of > 0.5 in the population. A nonsynonymous SNP with a frequency of > 0.5 could indicate an evolutionary advantage since selection did not eliminate it.

2.2 Salt stress experiment

The salt stress response of young *S. chilense* plants was examined to address experimentally local adaptation to saline environments. Several *S. chilense* populations grow at the coast or near the salt lakes in the Atacama Desert (Peralta *et al.* 2008; Chetelat *et al.* 2009). The aim of this experiment was to test for differences in salinity tolerance between *S. chilense* populations, *i.e.* if plants from habitats with high soil salinity perform better under salt stress than plants from other habitats.

A total of 86 *S. chilense* plants were used in this experiment (Table A2.1). The choice of this plant set was based on plant availability. Nevertheless, this sample includes plants from coastal populations, plants from populations near the salt lakes in the Atacama Desert and plants from regions that are distant from any salt providing source. Furthermore, they can be grouped according to altitude. Given the topography of southern Peru and northern Chile, altitude corresponds to distance from the coast. Approximately ten day old seedlings were planted into common garden soil and grown at 18 – 22 °C with a 16 hour light and eight hour dark cycle for approximately five weeks. Plants were watered with increasing salt concentrations (50 mM to 400 mM; Table 2.1). After symptoms of stress were observed in every plant, plants were allowed to recover under normal conditions. Plant survival was documented four weeks after the last salt application according to the following scheme: 1 = no signs of stress, *i.e.* completely recovered, 2 = slight symptoms of stress, *i.e.* partly wilting, and 3 = severe symptoms of stress, *i.e.* completely wilting. A generalized linear model and a Wilcoxon rank sum test (R, R Development Core Team 2005) were applied to analyse this data set.

Table 2.1: Experimental procedure for the salt stress experiment.

	Timepoint	**Treatment**
start salt stress	day 1	10 ml of 50 mM NaCl solution per plant
	day 6	10 ml of 100 mM NaCl solution per plant
	day 9	10 ml of 200 mM NaCl solution per plant
	day 13	10 ml of 400 mM NaCl solution per plant
	day 14	10 ml of 400 mM NaCl solution per plant
	day 15	soaking with 400 mM NaCl solution
start recovery phase	day 17	start watering with H_2O
	day 44	final documentation

CHAPTER 3: RESULTS

3.1 Sequence data

A pooling approach was used to sequence 30 genes from 23 populations of the wild tomato species *Solanum chilense*. It was not possible to amplify the gene *AREB2* from all individuals of LA2750 and the gene *CT179* from all individuals of LA4332. Therefore, these two data points are not present in the study. Furthermore, only a short fragment of the gene *CT208* could be amplified from LA3784.

Between 6,328,429 (LA3784) and 14,601,371 (LA2880) paired-end reads per population were obtained (Table B1.1). Between 81.40 % (LA4119) and 97.41 % (LA2773) of the reads were mapped. The number of reads and the number of mapped reads are highly correlated ($R^2 = 0.981$, Spearman's rank correlation $\rho = 0.983$, p-value < 0.001). Tests for correlations with genetic variation (θ_W and π) were performed to exclude any effect of number of reads or number of mapped reads on the results. No correlation was observed with either of the two estimators (number of reads with θ_W: $R^2 = 0.002$, $\rho = 0.025$, p-value = 0.907; number of reads with π: $R^2 = 0.014$, $\rho = 0.061$, p-value = 0.777; number of mapped reads with θ_W: $R^2 = 0.007$, $\rho = 0.063$, p-value = 0.768; number of mapped reads with π: $R^2 = 0.029$, $\rho = 0.106$, p-value = 0.621).

In order to evaluate the repeatability of this pooling approach one population (LA1968) was sequenced two times. Significant correlations were found for the two repetitions of this population (θ_W: $R^2 = 0.786$, $\rho = 0.864$, p-value < 0.001; π: $R^2 = 0.982$, $\rho = 0.976$, p-value < 0.001; Tajima's *D*: $R^2 = 0.804$, $\rho = 0.880$, p-value < 0.001; *S*: $R^2 = 0.982$, $\rho = 0.982$, p-value < 0.001). This confirms the repeatability of this approach.

The 30 genes were also sequenced from two outgroup species, *S. ochranthum* and *S. lycopersicoides*. All attempts failed to amplify *TAS14* from *S. ochranthum* and *His1* and *CT021* from *S. lycopersicoides*. A putative *NtC7* sequence could be obtained from both outgroup species, but in both cases the divergence is on average more than four times higher than for the other genes (*e.g.* Tables B3.11 - B3.14). Furthermore, the *NtC7* sequence of *S. ochranthum* is pseudogenized (7 bp deletion in exon 1 resulting in a frameshift and premature stop codon). Therefore mean divergence values for the populations were calculated excluding these four genes.

3.2 Demographic history of *Solanum chilense*

In order to assess the demographic history of the wild tomato species *S. chilense*, 30 genes of 23 populations which are distributed over the whole species range were sequenced (Figure 2.1). Analyses were performed twice, once with the values averaged over all 30 genes and once with the values averaged only over the 14 reference genes. The first data set is larger and has therefore more statistical power, but comprises genes, which are possibly under environmental selection and could bias the results. The second data set is smaller but comprises only genes that are not expected to evolve under any environmental constraint. Furthermore, the analyses were performed for all sites, because almost perfect correlations were observed between all sites and silent sites variation and Tajima's *D* (θ_W: $R^2 = 0.833$, $\rho = 0.867$, p-value < 0.001; π: $R^2 = 0.987$, $\rho = 0.975$, p-value < 0.001; Tajima's *D*: $R^2 = 0.974$, $\rho = 0.951$, p-value < 0.001).

3.2.1 Within population levels of variation and population averages of Tajima's *D*

To assess the variation of each population, the Watterson estimator, θ_W, and the nucleotide diversity, π, were calculated and averaged over all 30 genes ($\theta_{W, all}$, π_{all}) and over the 14 reference genes ($\theta_{W, ref}$, π_{ref}; Table 3.1). Differences in genetic variation can be observed between the populations. The values of $\theta_{W, all}$ range from 0.00448 (LA4107) to 0.01185 (LA2765) and for $\theta_{W, ref}$ from 0.00356 (LA2880) to 0.01073 (LA2765). Two most northern populations LA1930 and LA3784 are among the populations with the highest θ_W values. Three of the coastal populations are among the populations with the lowest θ_W values, namely LA4107, LA4108, and LA2932, while the fourth coastal population, LA2750, has the second highest $\theta_{W, all}$ and $\theta_{W, ref}$ values. Two of the southern high altitude populations are also among the populations with the lowest values, namely LA2880 and LA4118. The other two southern high altitude populations, LA4332 and LA4119, have higher values, but lie still in the lower half of the distribution. Among the populations from the central region of the species distribution, LA0458 has the lowest θ_W value and LA2765 and LA1963 have the highest θ_W values. The π_{all} values range from 0.00206 (LA4107) to 0.01012 (LA1930) and the π_{ref} values range from 0.00238 (LA4107) to 0.00980 (LA1930). The two northern populations, LA1930 and LA3784, show the highest π_{all} and π_{ref} values. The four coastal populations (LA4107, LA4108, LA2932, and LA2750), three of the southern high altitude populations (LA2880, LA4119, and LA4118) and LA0458 have low π_{all} and π_{ref} values.

Plotting variation against latitude revealed correlations between variation and latitude. Genetic variation decreases from north to south. This 'north-south-cline' is present for θ_W and π for averages over all 30 genes ($\theta_{W, all}$: $R^2 = 0.336$, $\rho = 0.492$, p-value = 0.018; π_{all}: $R^2 = 0.639$, $\rho = 0.712$, p-value < 0.001; Figure 3.1A, B, Table 3.1) and also for averages over the 14 reference genes ($\theta_{W, ref}$: $R^2 = 0.378$, $\rho = 0.546$, p-value = 0.008; π_{ref}: $R^2 = 0.621$, $\rho = 0.708$, p-value < 0.001; Figure 3.1A, B, Table 3.1). The same pattern can be observed for synonymous nucleotide diversity and to a less extent for nonsynonymous nucleotide diversity (data not shown). Furthermore, the ratio between nonsynonymous and synonymous nucleotide diversity increases from north to south ($\pi_{a, all}/\pi_{s, all}$: $R^2 = 0.447$, $\rho = -0.565$, p-value = 0.006; $\pi_{a, ref}/\pi_{s, ref}$: $R^2 = 0.399$, $\rho = -0.622$, p-value < 0.001, Figure 3.1C, Table 3.1). However, this correlation is mainly due to the four coastal populations LA2750, LA2932, LA4107, and LA4108, which have the highest π_a/π_s ratios (Table 3.1).

The test statistic Tajima's *D* was also calculated for the average over all 30 genes (D_{all}) and the average over the 14 reference genes (D_{ref}) for every population (Table 3.1). Four populations have positive D_{all} values (LA1958, LA2773, LA2755, and LA0458). LA2773 and LA2755 are high altitude populations. The remaining populations have negative D_{all} values ranging from -0.008 (LA2747) to -1.763 (LA2750). Among the five populations with the lowest D_{all} values, *i.e.* $D < -0.9$, are the four coastal populations LA2750, LA4107, LA4108, and LA2932. Nine populations have positive D_{ref} values ranging from 0.011 (LA2753) to 0.411 (LA2755). The other populations have negative Tajima's *D* values ranging from -0.029 (LA2931) to -1.770 (LA2750). The four coastal populations are among the populations with the lowest D_{ref}. One of the southern high altitude populations, LA4119, is also among the populations with the lowest Tajima's *D* values. The Tajima's *D* values further decrease from north to south with most negative values observed in the southern range of the species distribution (D_{all}: $R^2 = 0.410$, $\rho = 0.525$, p-value = 0.011; D_{ref}: $R^2 = 0.330$, $\rho = 0.432$, p-value = 0.041; Figure 3.1D, Table 3.1).

Table 3.1: Mean genetic variation and mean Tajima's *D* values for the *S. chilense* populations. Mean Watterson estimator, θ_W, nucleotide diversity π, π_a/π_s ratio, and Tajima's *D* for the *S. chilense* populations.

Population	$\theta_{W, all}$	$\theta_{W, ref}$	π_{all}	π_{ref}	$\pi_{a, all}/\pi_{s, all}$	$\pi_{a, ref}/\pi_{s, ref}$	D_{all}	D_{ref}
LA0456	0.00680	0.00585	0.00678	0.00633	0.315	0.125	-0.065	0.017
LA0458	0.00490	0.00444	0.00468	0.00451	0.387	0.567	0.030	0.180
LA1930	0.01090	0.01008	0.01012	0.00980	0.634	0.310	-0.285	-0.243
LA1958	0.00938	0.00836	0.00996	0.00902	0.292	0.124	0.241	0.204
LA1963	0.01045	0.01006	0.00919	0.00896	0.323	0.132	-0.426	-0.444
LA1968	0.00892	0.00875	0.00811	0.00806	0.471	0.207	-0.415	-0.392
LA2747	0.00795	0.00752	0.00786	0.00781	0.297	0.132	-0.008	0.118
LA2748	0.00840	0.00755	0.00762	0.00768	0.367	0.132	-0.242	-0.107
LA2750	0.01120	0.01049	0.00540	0.00556	0.934	0.736	-1.763	-1.770
LA2753	0.00980	0.00885	0.00948	0.00917	0.422	0.221	-0.078	0.011
LA2755	0.00729	0.00694	0.00772	0.00813	0.344	0.188	0.124	0.411
LA2765	0.01185	0.01073	0.00908	0.00899	0.266	0.137	-0.682	-0.512
LA2773	0.00815	0.00758	0.00861	0.00839	0.234	0.131	0.204	0.307
LA2880	0.00497	0.00356	0.00460	0.00351	0.788	0.290	-0.280	-0.207
LA2931	0.00952	0.00848	0.00882	0.00848	0.307	0.205	-0.196	-0.029
LA2932	0.00558	0.00518	0.00391	0.00308	1.096	0.602	-0.935	-1.324
LA3111	0.00852	0.00714	0.00750	0.00725	0.288	0.130	-0.234	0.035
LA3784	0.01116	0.01045	0.00957	0.00967	0.415	0.173	-0.465	-0.335
LA4107	0.00448	0.00415	0.00206	0.00238	1.161	1.257	-1.659	-1.409
LA4108	0.00494	0.00396	0.00267	0.00260	0.970	0.501	-1.319	-1.184
LA4118	0.00575	0.00511	0.00508	0.00506	0.480	0.484	-0.488	-0.289
LA4119	0.00818	0.00685	0.00495	0.00426	0.689	0.596	-1.369	-1.317
LA4332	0.00790	0.00627	0.00685	0.00700	0.372	0.210	-0.410	0.051

Note: "all" averaged over all 30 genes, "ref" averaged over the 14 reference genes

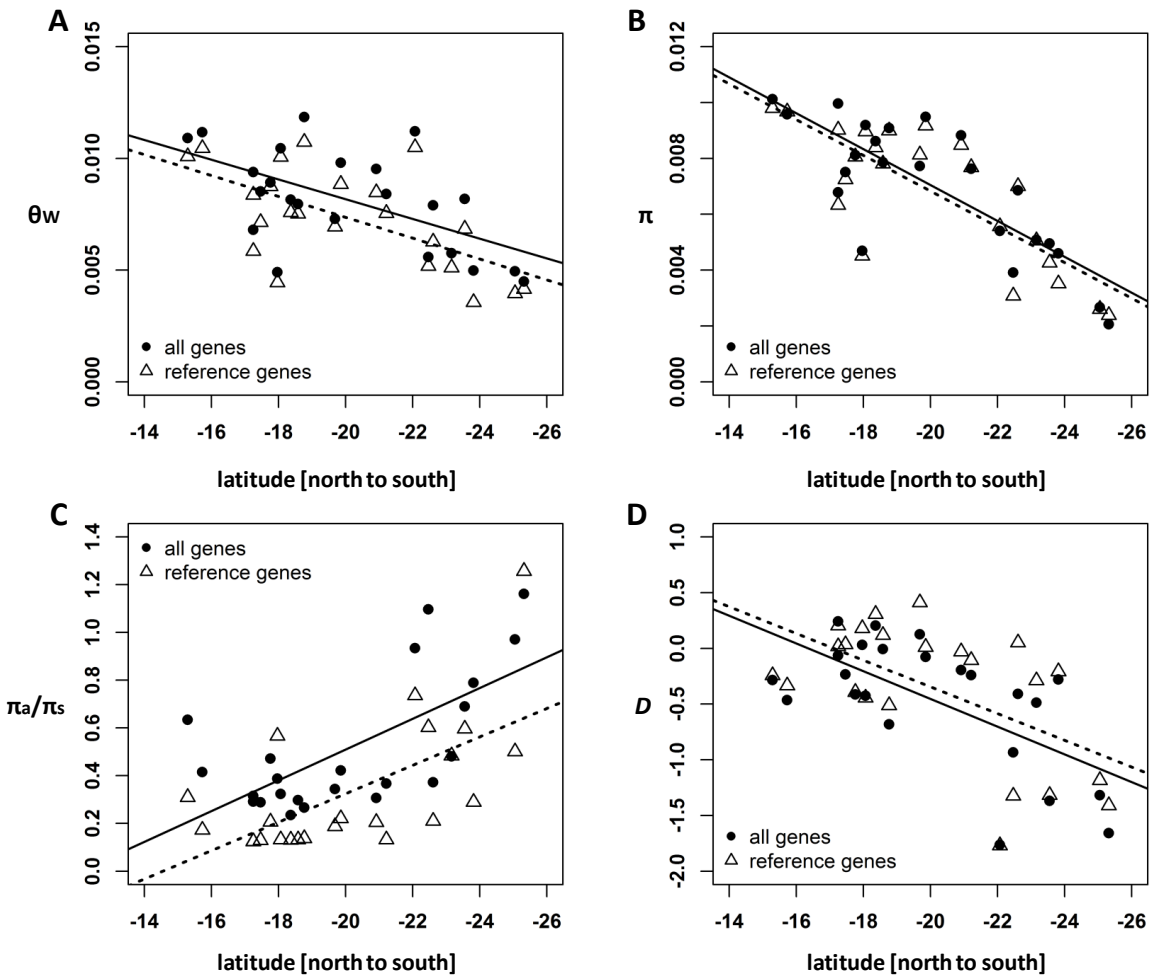

Figure 3.1: Correlation between genetic data and latitude. Genetic data of all genes (black points) and reference genes (triangles) plotted against latitude. Solid line: regression for all genes, dashed line: regression for reference genes. A) Watterson estimator θ_W, B) nucleotide diversity π, C) π_a/π_s ratio, and D) Tajima's *D*.

3.2.2 Population differentiation and isolation by distance

The genetic differentiation between each population pair was assessed with F_{ST}, the pairwise genetic differentiation between populations, and averaged over all 30 genes (F_{STall}) and over the 14 reference genes (F_{STref}). F_{STall} ranges between 0.056 (population pair LA2747 – LA2773) and 0.467 (LA3784 – LA4107) with a mean of 0.228. LA2747 and LA2773 are populations from the same region east of Arica, northern Chile, and are separated by approximately 40 km. LA3784 is one of the most northern populations of this sample and LA4107 is the most southern population. These two populations are separated by over 1100 km. F_{STref} ranges between 0.050 (LA2747 – LA2773) and 0.511 (LA2880 – LA2932) with 0.231 as mean. LA2880 and LA2932 are both in the southern range of the

species distribution and separated by approximately 250 km. LA2932 is from the coast and LA2880 is from high altitudes. The overall patterns of F_{STall} and F_{STref} are similar (Figure 3.2A, Table B2.1). The highest F_{ST} values (F_{STall} and F_{STref}) are between the two northern populations (LA1930, LA3784) and all other populations, between the coastal populations (LA2750, LA2932, LA4108, LA4107) and all other populations and between the southern high altitude populations (LA4118, LA4119, LA2880) and all other populations (Figure 3.2A, Table B2.1). The population LA4332 is geographically a southern high altitude population, but has relatively low F_{ST} values compared with any of the other 22 populations. This indicates that LA4332 is intermediate between the southern high altitude populations and the other populations. The populations LA0456 and LA0458 from the central region have relatively high F_{ST} values compared with other populations.

The average F_{STall} in comparison with all other populations was calculated for each population. Lowest values are found in the centre of the species distribution. LA1963 and LA2931 are the populations with the lowest values (Figure 3.2B). Among the populations with the highest values are the two most northern and the two most southern populations. These observations are consistent with the pairwise F_{ST} values (Figure 3.2A, Table B2.1).

These findings indicate that the 23 populations can be clustered into four groups: the northern group (LA1930, LA3784), the coastal group (LA2750, LA2932, LA4108, LA4107), the southern high altitude group (LA4118, LA4119, LA2880), the central group with all remaining populations, and with LA4332 being the link between the central group and the southern high altitude group.

Furthermore, within and between group average F_{ST} values were calculated (LA4332 was considered to belong to the southern high altitude group; Figure 3.3, Table B2.2). The central group has the lowest within group F_{ST}, followed by the southern high altitude group. The coastal group has the highest within group F_{ST}. The lowest between group F_{ST} is observed between the central group and the southern high altitude group. The genetic differentiation between the central group and the coastal group is similar to the one between the central group and the northern group. The highest genetic differentiation is between the northern group and the coastal group. The genetic differentiation between the northern group and the southern high altitude group is similar to the one between the coastal group and the southern high altitude group. This is remarkable given the fact that the geographic distance between the southern high altitude group and the coastal group is approximately 200 – 300 km while the geographic distance between the coastal group and the northern group is approximately 800 – 1200 km.

Genetic distance (F_{ST}) was plotted against the geographic distance to test for isolation by distance (Figure 3.4A, Figure B2.1A). A significant pattern of isolation by distance is observed for F_{STall} and F_{STref} (all genes: $R^2 = 0.382$, Mantel test p-value < 0.001; reference genes: $R^2 = 0.332$, Mantel test p-value < 0.001). Since the F_{ST} values indicated that the 23 *S. chilense* populations of this data set fall into four groups, the groups were tested for isolation by distance separately. This revealed that isolation by distance is not present in the central group (all genes: $R^2 = 0.009$, Mantel test not significant; reference genes: $R^2 = 0.002$, Mantel test not significant; Figure 3.4B, Figure B2.1B). A trend for isolation by distance is present in the coastal group (all genes: $R^2 = 0.777$, Mantel test not significant; reference genes: $R^2 = 0.710$, Mantel test not significant) and in the southern high altitude group (all genes: $R^2 = 0.990$, Mantel test not significant; reference genes: $R^2 = 0.671$, Mantel test not significant). Genetic distance was also plotted against altitudinal distance, but no correlation was observed ($R^2 < 0.07$).

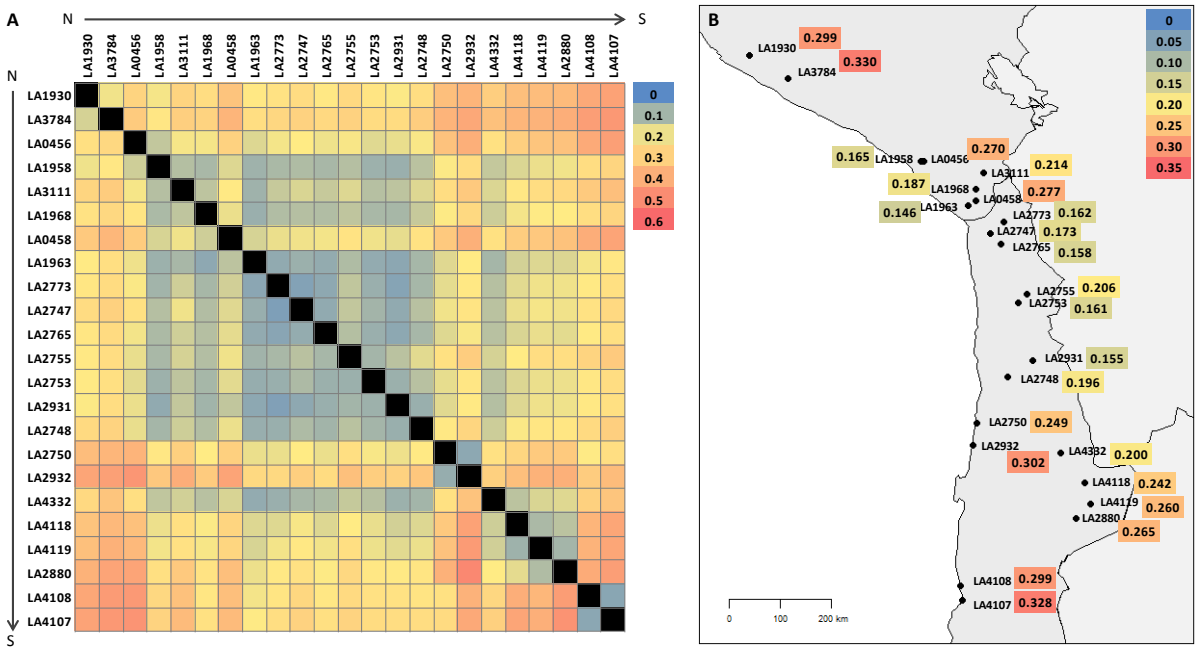

Figure 3.2: Pairwise population genetic differentiation and mean population genetic differentiation. A) Pairwise population genetic differentiation matrix. Populations are sorted from north to south (left to right and top to bottom), each cell represents the pairwise genetic differentiation (F_{ST}) between two populations, colour of the cell corresponds to the F_{ST} value (blue: $F_{ST} = 0$, red: $F_{ST} = 0.6$). Cells above the diagonal represent mean F_{ST} values of all 30 genes and cells below the diagonal represent mean F_{ST} values of the 14 reference genes. B) Mean population genetic differentiation. Map with the 23 *S. chilense* populations. Each population is labelled with a box. Value within the box is the mean F_{ST} value (over all 30 genes) of this population compared to the other 22 populations, colour of the box corresponds to the F_{ST} value (blue: $F_{ST} = 0$, red: $F_{ST} = 0.35$).

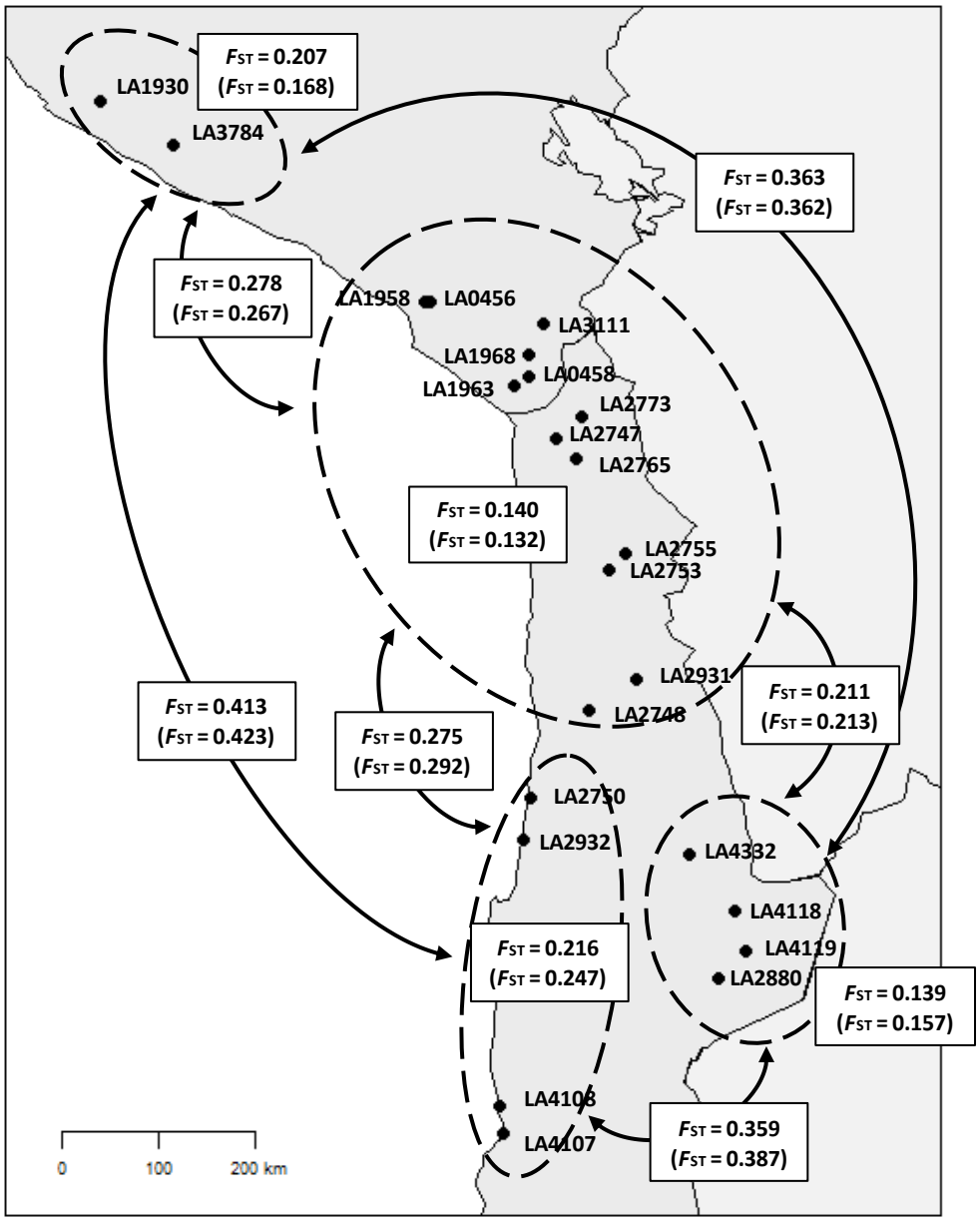

Figure 3.3: Mean within and between group genetic differentiation. Map with the *S. chilense* populations. Dashed line circles indicate population groups. Mean genetic differentiation (F_{ST}) in white boxes, upper values: mean F_{ST} values for all 30 genes, lower values in brackets: mean F_{ST} values for the 14 reference genes. Arrows indicate between group F_{ST} values. For minimum and maximum F_{ST} values within and between groups see Table B2.2.

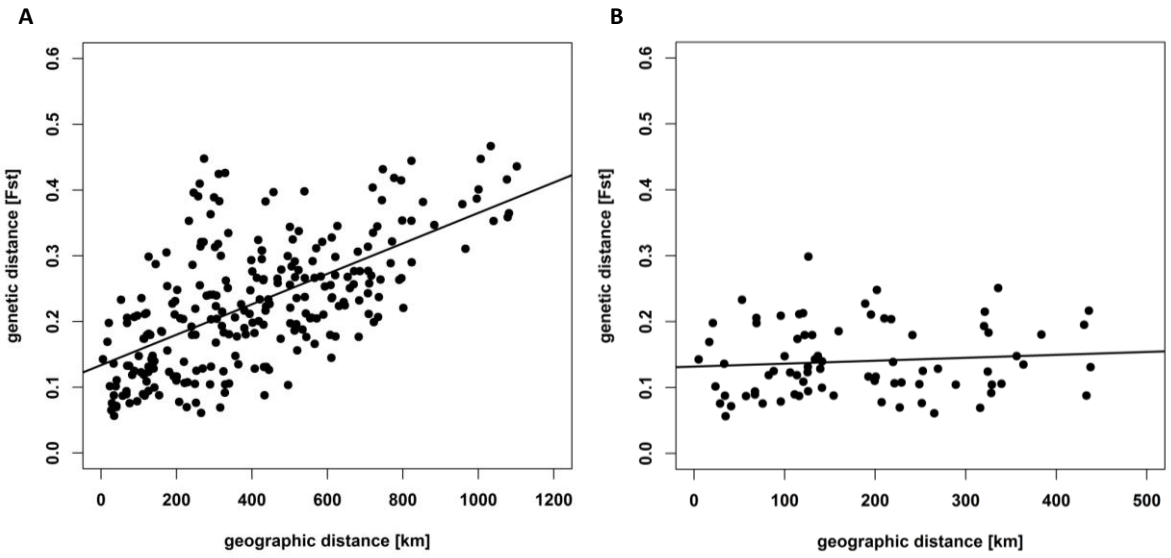

Figure 3.4: Isolation by distance for the average over all genes. Genetic distance (F_{ST}) between populations plotted against geographic distance (km). A) A pattern of isolation by distance is observed for all 23 *S. chilense* populations (R^2 = 0.382, Mantel test p-value < 0.001). B). A pattern of isolation by distance is not observed for the 13 *S. chilense* populations from the central group (R^2 = 0.009, Mantel test not significant).

3.2.3 Genetic differentiation and divergence from other Solanaceae species

F_{ST} values for each *S. chilense* population were calculated in comparison with other wild tomato species (Table B2.3). Two data sets that share some of the reference genes are available: a species wide sample of *S. peruvianum* (Tellier *et al.* 2011b; Hörger *et al.* in preparation) and pooled population samples of *S. arcanum* and *S. habrochaites* (Tellier *et al.* 2011a). The overall mean F_{ST} is lowest in comparison with *S. peruvianum* (0.379) and highest with respect to *S. habrochaites* (0.611) while *S. arcanum* (0.488) is intermediate. The *S. chilense* populations that show the lowest genetic differentiation from *S. peruvianum* are LA2765 and LA2753 closely followed by LA1963 and LA3784. Except for LA3784, which is from the northern group, these populations are from the central group. Highest interspecies F_{ST} values are found for three of the coastal populations, LA4107, LA4108, and LA2932, followed by the southern high altitude populations, the fourth coastal population LA2750 and LA0458. Concerning the two other wild tomato species, *S. arcanum* and *S. habrochaites*, the lowest interspecies F_{ST} values are with LA2765 and LA3784 and the highest with LA4107 and LA4108. Interspecies F_{ST} values were plotted against the latitudinal coordinates of the *S. chilense* populations (Figure 3.5A). Interspecies F_{ST} values increase from north to south for all three species (*S. peruvianum*: R^2 = 0.544, ρ = -0.574, p-value = 0.005;

S. arcanum: R^2 = 0.588, ρ = -0.605, p-value = 0.003; *S. habrochaites*: R^2 = 0.623, ρ = -0,658, p-value = 0.001).

Sequences for both outgroup species, *S. ochranthum* and *S. lycopersicoides*, are available for 26 genes. The average divergence of each *S. chilense* population for all genes (K_{all}) and for the reference genes (K_{ref}) was calculated (Table B2.4). The overall mean K_{all} and K_{ref} in comparison with *S. ochranthum* are 0.0351 and 0.0329, respectively. In comparison with *S. lycopersicoides* the overall mean K_{all} is 0.0337 and the mean K_{ref} is 0.0318. The *S. chilense* populations have similar divergence from both outgroup species. The lowest K_{all} values compared to *S. ochranthum* are for LA2880 and LA4107 while the lowest K_{ref} values are for LA4332, LA1963 and LA2753. The highest K_{all} values are for LA3784, LA0456 and LA1930 and the highest K_{ref} values for LA2750, LA2932 and LA3784. In comparison with *S. lycopersicoides* the lowest K_{all} values are for LA1963 and LA2748. LA1963 has also the lowest K_{ref} value. The highest K_{all} values are for LA3784 and LA1930 and the highest K_{ref} values for LA3784 and LA2750. Interestingly, the northern population LA3784 is always among the populations with the highest *K* values. K_{all} and K_{ref} were also plotted against the latitudinal coordinates of the *S. chilense* populations (Figure 3.5B), but for neither of them any correlation between divergence and latitude can be observed.

Mean divergence for all genes (K_{all}) and for the reference genes (K_{ref}) was further calculated in comparison to the cultivated tomato, *S. lycopersicum*, with K_{all} = 0.0183 and K_{ref} = 0.0176 (Table B2.4). The populations with the lowest K_{all} are LA0456, LA2765, and LA2931 and the populations with the lowest K_{ref} are LA0456 and LA1930. The populations with the highest K_{all} are LA2750 and LA3784 while the populations with the highest K_{ref} are LA2750 and LA2932. No correlation between K_{all} and K_{ref} and the latitudinal coordinates of the *S. chilense* populations were observed (Figure 3.5B).

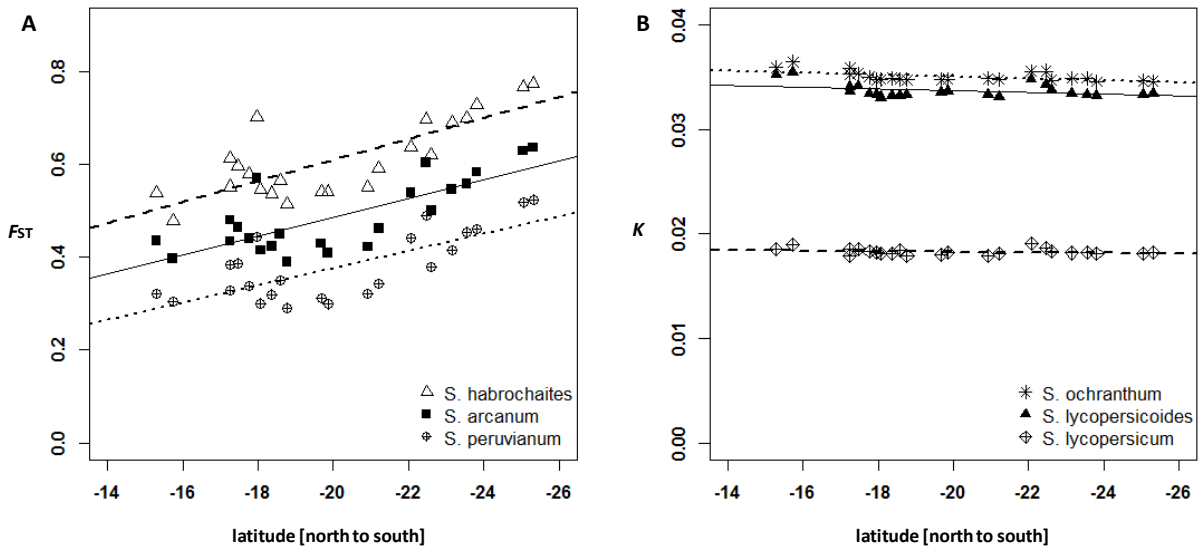

Figure 3.5: Interspecies genetic differentiation and divergence. A) Genetic differentiation (F_{ST}) from other wild tomato species. Triangle: *S. habrochaites*, black square: *S. arcanum*, circle: *S. peruvianum*. Dashed line: regression for *S. habrochaites* ($R^2 = 0.623$), solid line: regression for *S. arcanum* ($R^2 = 0.588$), dotted line: regression for *S. peruvianum* ($R^2 = 0.544$). B) Divergence from other Solanaceae species. Stars: *S. ochranthum*, black triangles: *S. lycopersicoides*, diamonds: *S. lycopersicum*. Dotted line: regression for *S. ochranthum* ($R^2 = 0.306$), solid line: regression for *S. lycopersicoides* ($R^2 = 0.148$), dashed line: regression for *S. lycopersicum* ($R^2 = 0.040$).

3.3 Genes under selection in *Solanum chilense*

3.3.1 Candidate genes vs. reference genes

The genetic variation of the 16 candidate genes was assessed with θ_W and π and compared to the 14 reference genes, which represent the genomic average. π was also calculated for synonymous (π_s) and nonsynonymous (π_a) sites (Tables B3.1 – B3.8).

The mean θ_W of the candidate genes ($\theta_{W, can}$) is for every population higher than the mean of the reference genes ($\theta_{W, ref}$; Table 3.2). This difference is especially profound for LA4332, LA2880, LA3111, and LA4119. These populations are all from altitudes of above 2500 m, three of them are from the southern high altitude group. The population with the lowest difference is LA1968.

A different pattern was observed for π. Six populations have a lower average π_{can} than π_{ref} (LA2755, LA4107, LA2750, LA4332, LA2748; Table 3.2). The populations with the largest difference between their mean π_{can} and π_{ref} values are LA2880, LA1958, LA2932, LA4119. Only four populations have a higher $\pi_{s, can}$ than $\pi_{s, ref}$ (LA4119, LA4118, LA2880, LA2747; Table 3.2). Three of these populations are from the southern high altitude group. However, all populations have a higher $\pi_{a, can}$ than $\pi_{a, ref}$ (Table 3.2). The differences are especially high for LA1958, LA1963, and LA2773.

Next synonymous and nonsynonymous nucleotide diversity is averaged for each gene over all populations. This separates candidate and reference genes as shown in Figure 3.6. All seven genes in the wedge between the lower line and the $\pi_a/\pi_s = 1$ line are candidate genes. In addition, *le25* that is above the $\pi_a/\pi_s = 1$ line is a candidate gene. These eight genes (*i.e.* half of the candidate genes) have the property that both π_a/π_s and π_a are relatively high. In contrast, the reference genes typically exhibit higher π_s and lower π_a/π_s values. However, there are some exceptions that will be discussed below.

The Tajima's *D* test statistic was applied to the data set (Tables B3.9, B3.10). Candidate genes have on average lower Tajima's *D* values (D_{can}) than the reference genes (D_{ref}) in most of the populations (Table 3.2). Four populations have higher D_{can} than D_{ref} (LA2750, LA1963, LA1958, LA2932). This difference is quite small for LA2750, LA1963, and LA1958, but for LA2932 the difference is large.

All sites (K), synonymous sites (K_s) and nonsynonymous sites (K_a) divergence in comparison to the other three species (*S. ochranthum*, *S. lycopersicoides*, *S. lycopersicum*) was calculated (Tables B3.11 – B3.28). Almost all populations have higher mean K and K_a for the candidate genes than for the reference genes. The only exception is the population

LA2932, which has a higher *K* for the reference genes in respect to the cultivated tomato. Six populations (LA1968, LA2880, LA2932, LA4107, LA4108, LA4119) have lower K_s compared to *S. ochranthum* for the candidate genes than for the reference genes. Further all populations have lower K_s compared to *S. lycopersicoides* for the candidate genes than for the reference genes and about half of the populations have lower K_s compared to the cultivated tomato for the candidate genes than for the reference genes.

Table 3.2: Comparison of candidate genes and reference genes. Mean genetic variation (Watterson estimator θ_W, nucleotide diversity π, synonymous nucleotide diversity π_s, and nonsynonymous nucleotide diversity π_a) and Tajima's *D* values for the *S. chilense* populations.

Pop.	$\theta_{W, can}$	$\theta_{W, ref}$	π_{can}	π_{ref}	$\pi_{s, can}$	$\pi_{s, ref}$	$\pi_{a, can}$	$\pi_{a, ref}$	D_{can}	D_{ref}
LA0456	0.0076	0.0059	0.0072	0.0063	0.0077	0.0128	0.0033	0.0013	-0.136	0.017
LA0458	0.0053	0.0044	0.0041	0.0045	0.0069	0.0076	0.0021	0.0011	-0.100	0.180
LA1930	0.0116	0.0101	0.0111	0.0098	0.0136	0.0174	0.0039	0.0027	-0.321	-0.243
LA1958	0.0103	0.0084	0.0101	0.0090	0.0167	0.0183	0.0045	0.0017	0.274	0.204
LA1963	0.0108	0.0101	0.0089	0.0090	0.0144	0.0161	0.0043	0.0019	-0.411	-0.444
LA1968	0.0091	0.0087	0.0072	0.0081	0.0097	0.0139	0.0036	0.0018	-0.435	-0.392
LA2747	0.0083	0.0075	0.0072	0.0078	0.0123	0.0123	0.0036	0.0014	-0.118	0.118
LA2748	0.0091	0.0075	0.0075	0.0077	0.0104	0.0143	0.0031	0.0014	-0.360	-0.107
LA2750	0.0119	0.0105	0.0051	0.0056	0.0061	0.0086	0.0039	0.0023	-1.757	-1.770
LA2753	0.0106	0.0088	0.0093	0.0092	0.0114	0.0161	0.0049	0.0026	-0.156	0.011
LA2755	0.0076	0.0069	0.0068	0.0081	0.0117	0.0153	0.0037	0.0019	-0.127	0.411
LA2765	0.0128	0.0107	0.0087	0.0090	0.0126	0.0164	0.0039	0.0018	-0.832	-0.512
LA2773	0.0086	0.0076	0.0085	0.0084	0.0141	0.0143	0.0038	0.0015	0.113	0.307
LA2880	0.0062	0.0036	0.0054	0.0035	0.0058	0.0047	0.0025	0.0005	-0.343	-0.207
LA2931	0.0104	0.0085	0.0083	0.0085	0.0138	0.0147	0.0037	0.0019	-0.341	-0.029
LA2932	0.0059	0.0052	0.0044	0.0031	0.0049	0.0050	0.0023	0.0010	-0.594	-1.324
LA3111	0.0097	0.0071	0.0068	0.0072	0.0113	0.0132	0.0027	0.0018	-0.470	0.035
LA3784	0.0118	0.0105	0.0109	0.0097	0.0120	0.0197	0.0048	0.0028	-0.578	-0.335
LA4107	0.0048	0.0042	0.0018	0.0024	0.0016	0.0036	0.0020	0.0008	-1.876	-1.409
LA4108	0.0058	0.0040	0.0030	0.0026	0.0026	0.0040	0.0015	0.0007	-1.437	-1.184
LA4118	0.0063	0.0051	0.0048	0.0051	0.0071	0.0055	0.0024	0.0010	-0.663	-0.289
LA4119	0.0094	0.0068	0.0056	0.0043	0.0073	0.0053	0.0031	0.0009	-1.415	-1.317
LA4332	0.0092	0.0063	0.0064	0.0070	0.0102	0.0117	0.0031	0.0016	-0.784	0.051

Note: "can" averaged over the 16 candidate genes, "ref" averaged over the 14 reference genes

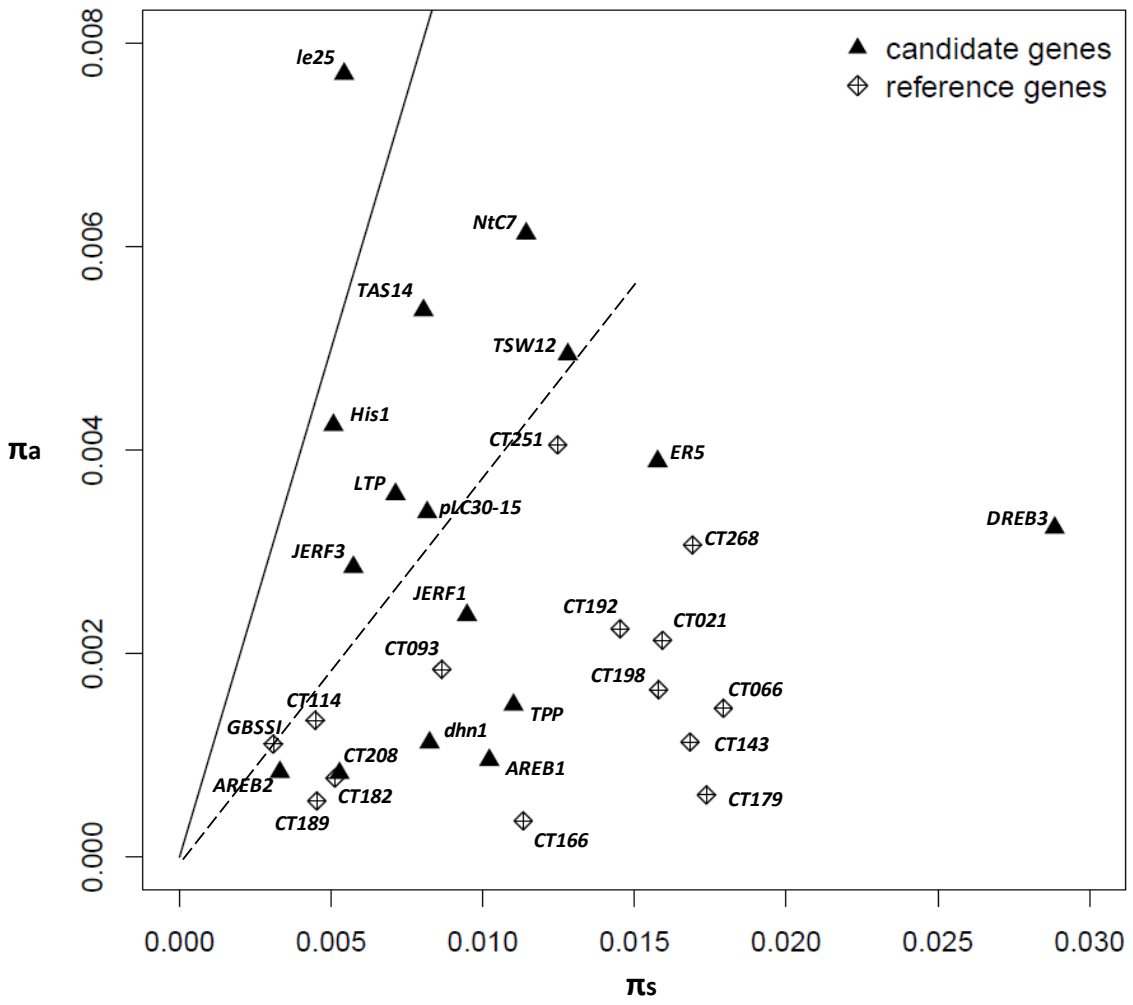

Figure 3.6: Mean nonsynonymous and synonymous nucleotide diversity. Values of π_a and π_s were averaged for each gene over all *S. chilense* populations and plotted against each other. Triangles: candidate genes; diamonds: reference genes. Solid line: $\pi_a/\pi_s = 1$, lower dashed line: upper boundary of reference genes.

3.3.2 Regulatory genes vs. functional genes

Next the patterns of genetic variation are described for the candidate genes. The 16 candidate genes of the data set come from different levels of the abiotic stress response pathways and can be classified into three groups: sensory genes (*NtC7*), regulatory genes (*AREB1*, *AREB2*, *JERF1*, *JERF3*, *DREB3*), and functional genes (*dhn1*, *pLC30-15*, *TAS14*, *ER5*, *le25*, *LTP*, *TSW12*, *CT208*, *His1*, *TPP*). To compare the regulatory genes with the functional genes, diversity statistics (θ_W, π, π_s, π_a), divergence from *S. ochranthum*, *S. lycopersicoides*, and *S. lycopersicum*, and Tajima's *D* were averaged over these two groups (Table 3.3, 3.4). The functional genes have on average higher variation at the amino acid level and a more

negative Tajima's *D* value than the regulatory genes, while the synonymous variation is comparable in these two groups. The difference between these two groups is also observable in Figure 3.6. Six functional genes, namely *le25*, *TAS14*, *TSW12*, *His1*, *LTP*, and *pLC30-15*, but only one regulatory gene, namely *JERF3*, are among the candidate genes with the highest π_a and π_a/π_s. The comparison of the divergence values between regulatory and functional genes shows a similar pattern.

Five functional genes belong to the late embryogenesis abundant (*lea*) genes of which three are dehydrins (*lea* type 2), namely *dhn1*, *pLC30-15*, and *TAS14*. These three dehydrins differ greatly. *TAS14* is the most polymorphic dehydrin and has also the highest Tajima's *D* value of the dehydrins. *TAS14* has further the highest *K* and K_s compared to *S. lycopersicoides*. In contrast, *dhn1* is the least variable dehydrin, has low divergence and the most negative Tajima's *D* value of the dehydrins. The third dehydrin, *pLC30-15*, is intermediate between *TAS14* and *dhn1*. The other two *lea* genes, *ER5* (atypical *lea*) and *le25* (*lea* type 4), have the most negative Tajima's *D* values among all functional genes and are also among the most polymorphic genes. Although *le25* has the highest θ_W and π_a values, it has one of the lowest π_s values. This also leads to a $\pi_a/\pi_s > 1$ (Figure 3.6). *ER5* has the highest π_s value of all functional genes.

Two other functional genes, *LTP* and *TSW12*, encode lipid transfer proteins. Both are also among the candidate genes with relatively high π_a and π_a/π_s (Figure 3.6). Overall *TSW12* has higher variation than *LTP* and has also a more negative Tajima's *D* value. *LTP* has the highest *K* and K_a of all functional genes. The remaining three functional genes, the alcohol dehydrogenase *CT208*, the histone gene *His1* and the trehalose-6-phosphate phosphatase *TPP*, exhibit relatively low values of variation and divergence. However, nonsynonymous variation and divergence are relatively high for *His1* and *TPP* has a relatively high π_s. Besides, *TPP* is the only functional gene with a positive Tajima's *D* value.

Two regulatory genes belong to the AREB/ABF subfamily of bZIP transcription factors. *AREB1* has higher θ_W, π, and π_s values than *AREB2*, while the π_a values are comparable. *AREB2* has a more negative Tajima's *D* value and overall lower divergence. Two other regulatory genes, *JERF1* and *JERF3*, encode ERF proteins. Similar to *AREB1* and *AREB2* they also show a different pattern. *JERF1* has higher values of Tajima's *D*, θ_W, π, and π_s, but a lower π_a than *JERF3*. *JERF1* has also the highest *K* compared to all three species, and the highest K_a compared to *S. ochranthum* and *S. lycopersicum*. The fifth regulatory gene, *DREB3*, is the most polymorphic gene among the regulatory genes and also has the most negative Tajima's *D* value (*D* = -0.918) of all regulatory genes. *DREB3* has further the highest π_s of all genes (Figure 3.6).

The only sensory gene *NtC7* is among the most polymorphic genes. It is also among the candidate genes with the highest π_a and π_a/π_s (Figure 3.6) and has the highest Tajima's *D* value (*D* = 0.204). The divergence in respect to *S. ochranthum* and *S. lycopersicoides* is extremely high (> 0.1).

Table 3.3: Mean genetic variation and Tajima's *D* for the candidate genes. Values of θ_W, π, π_s, π_a, and Tajima's *D* were averaged over all *S. chilense* populations.

Gene	θ_W	π	π_s	π_a	Tajima's *D*
sensory gene					
NtC7	0.01065	0.01160	0.01143	0.00613	0.204
regulatory genes					
AREB1	0.00656	0.00592	0.01021	0.00095	-0.439
AREB2	0.00380	0.00324	0.00331	0.00084	-0.713
JERF1	0.00947	0.00943	0.00948	0.00238	-0.083
JERF3	0.00820	0.00663	0.00573	0.00285	-0.772
DREB3	0.01195	0.00862	0.02885	0.00324	-0.918
mean	0.00800	0.00677	0.01152	0.00205	-0.585
functional genes					
dhn1	0.00684	0.00425	0.00824	0.00113	-1.103
pLC30-15	0.00817	0.00665	0.00817	0.00339	-0.644
TAS14	0.00983	0.00961	0.00804	0.00537	-0.065
ER5	0.01089	0.00644	0.01576	0.00389	-1.244
le25	0.01587	0.00762	0.00543	0.00770	-1.399
LTP	0.00776	0.00686	0.00712	0.00357	-0.295
TSW12	0.01213	0.01033	0.01280	0.00494	-0.496
CT208	0.00536	0.00483	0.00527	0.00082	-0.572
His1	0.00698	0.00575	0.00508	0.00425	-0.438
TPP	0.00658	0.00685	0.01101	0.00150	0.072
mean	0.00904	0.00692	0.00869	0.00366	-0.618

Table 3.4: Mean divergence for the candidate genes. All sites (K), synonymous sites (K_s) and nonsynonymous sites (K_a) divergence from *S. ochranthum*, *S. lycopersicoides*, and *S. lycopersicum*. Values were averaged over all *S. chilense* populations.

Gene	*S. ochranthum*			*S. lycopersicoides*			*S. lycopersicum*		
	K	***K*s**	***K*a**	***K***	***K*s**	***K*a**	***K***	***K*s**	***K*a**
sensory gene									
NtC7	0.1482	0.2781	0.1071	0.1540	0.2900	0.1014	0.0247	0.0400	0.0192
regulatory genes									
AREB1	0.0389	0.0529	0.0078	0.0400	0.0688	0.0118	0.0109	0.0144	0.0041
AREB2	0.0318	0.0477	0.0126	0.0185	0.0240	0.0072	0.0135	0.0062	0.0058
JERF1	0.0552	0.0632	0.0217	0.0687	0.0700	0.0192	0.0217	0.0228	0.0112
JERF3	0.0337	0.0647	0.0146	0.0341	0.0422	0.0244	0.0202	0.0302	0.0079
DREB3	0.0252	0.0692	0.0096	0.0219	0.0771	0.0050	0.0125	0.0352	0.0040
mean	0.0370	0.0595	0.0132	0.0367	0.0564	0.0135	0.0157	0.0218	0.0066
functional genes									
dhn1	0.0266	0.0553	0.0039	0.0368	0.0916	0.0007	0.0108	0.0167	0.0007
pLC30-15	0.0266	0.0303	0.0191	0.0325	0.0343	0.0172	0.0184	0.0295	0.0088
TAS14	n. a.	n. a.	n. a.	0.0625	0.2357	0.0147	0.0258	0.0104	0.0060
ER5	0.0347	0.1002	0.0116	0.0277	0.0512	0.0076	0.0198	0.0658	0.0115
le25	0.0402	0.0741	0.0276	0.0392	0.0260	0.0337	0.0147	0.0247	0.0125
LTP	0.0567	0.1026	0.0440	0.0449	0.0622	0.0436	0.0366	0.0628	0.0358
TSW12	0.0524	0.0981	0.0130	0.0339	0.0346	0.0042	0.0189	0.0456	0.0042
CT208	0.0351	0.0649	0.0018	0.0331	0.0568	0.0030	0.0156	0.0129	0.0005
His1	0.0283	0.0381	0.0162	n. a.	n. a.	n. a.	0.0236	0.0298	0.0161
TPP	0.0269	0.0423	0.0088	0.0315	0.0422	0.0099	0.0153	0.0305	0.0010
mean[a]	0.0374	0.0710	0.0162	0.0349	0.0499	0.0150	0.0199	0.0329	0.0097

Note: "n. a." not available
[a]: calculated excluding *His1* and *TAS14*

3.3.3 Detection of local adaptation

To detect local adaptation in the data set, the genes need to be analysed in a gene- and population-specific way. This leads to numerous tests. These approaches were applied to all 30 genes and the results will be summarized in the following.

McDonald-Kreitman test statistic

The McDonald-Kreitman test statistic compares synonymous and nonsynonymous polymorphism to synonymous and nonsynonymous divergence in comparison to an outgroup species. The McDonald-Kreitman test was performed for all genes with the outgroup species, *S. ochranthum* and *S. lycopersicoides*, separately. This approach identified

several genes with significant test results (Tables B3.29 – B3.32). In total, significant McDonald-Kreitman tests were obtained for 11 of the candidate genes (one sensory, four regulatory, six functional) and 11 of the reference genes in at least one of the *S. chilense* populations and with at least one of the outgroup species. Only five candidate genes and nine reference genes are significant in the same population with both outgroup species (Table 3.5). The reference gene *CT268* is significant due to an excess of nonsynonymous divergence and all other genes are significant due to an excess of nonsynonymous polymorphism. The candidate genes without a significant test result are *AREB2*, *pLC30-15*, *LTP*, *His1*, and *TPP*. The reference genes without a significant test result are *CT021*, *CT166*, and *CT198*.

The coastal populations have the highest number of genes with a significant McDonald-Kreitman test. The sensory gene *NtC7*, four of the regulatory genes (*AREB1*, *JERF1*, *JERF3*, *DREB3*), and six of the functional genes (*dhn1*, *TAS14*, *ER5*, *le25*, *TSW12*, *CT208*) have a significant McDonald-Kreitman test in at least one population from the coastal group with at least one of the outgroup species. Also ten of the reference genes have a significant McDonald-Kreitman test in at least one population from the coastal group and with at least one of the outgroup species. Ten genes have significant McDonald-Kreitman tests in at least one population from the southern high altitude group with at least one outgroup species. These genes include the functional genes *dhn1*, *TAS14*, *le25*, and *CT208*, and six reference genes. The populations from the northern group have also some genes with a significant McDonald-Kreitman test with at least one of the outgroup species. These genes include the regulatory gene *DREB3*, the two functional genes *TAS14* and *CT208*, and four reference genes. Nine genes are significant in populations from the central group with at least one outgroup species. These genes include two regulatory genes (*JERF3*, *DREB3*), four functional genes (*dhn1*, *TAS14*, *ER5*, *CT208*) and three reference genes. Interestingly, this group has the lowest proportion of significant reference genes.

Table 3.5: Genes with significant McDonald-Kreitman test statistics with both outgroup species in the same population. Significances (p-value < 0.05) according to Fisher's exact test with both outgroup species, *S. ochranthum* and *S. lycopersicoides*, given.

Gene	Population	Fisher's exact test, p-value	
		S. ochranthum	*S. lycopersicoides*
candidate genes			
AREB1	LA2750	0.00137	0.00024
AREB1	LA4107	0.00511	0.01057
AREB1	LA4108	0.00670	0.00836
DREB3	LA2750	0.01943	0.00540
DREB3	LA4107	0.04718	0.00371
dhn1	LA1968	0.04739	0.00022
dhn1	LA2750	0.02345	0.00009
dhn1	LA4119	0.02767	0.00012
dhn1	LA4332	0.03186	0.00041
ER5	LA2750	0.00832	0.04412
CT208	LA1968	0.00112	0.01099
CT208	LA2750	0.00002	0.00027
CT208	LA4107	0.00081	0.00398
CT208	LA4108	0.00130	0.00432
CT208	LA4118	0.01961	0.03922
CT208	LA4119	0.00016	0.00095
reference genes			
CT066	LA1930	0.03000	0.01418
CT066	LA2750	0.00716	0.00299
CT066	LA2880	0.00435	0.00251
CT066	LA4107	0.00012	0.00001
CT066	LA4108	0.00016	0.00007
CT066	LA4119	0.00046	0.00017
CT066	LA4332	0.04540	0.02125
CT093	LA2750	0.03699	0.00133
CT093	LA4107	0.02374	0.00590
CT093	LA4119	0.01150	0.00191
CT143	LA2750	0.00110	0.01405
CT143	LA4119	0.02778	0.04762
CT179	LA2750	0.04429	0.04429
CT182	LA2750	0.01522	0.03497
CT192	LA2750	0.00131	0.00088
CT251	LA2750	0.01747	0.00081
CT268	LA0456	0.03229	0.00339
GBSSI	LA2750	0.00368	0.02619

The proportion of adaptive amino acid changes

The proportion of adaptive amino acid changes, α, per gene was estimated based on the π_a/π_s and K_a/K_s ratios (Tables B3.33 – B3.36). A positive α is supposed to indicate adaptive evolution. The mean α over all candidate genes is -2.556 for *S. ochranthum* and -9.425 for *S. lycopersicoides*. The mean α over all reference genes is -3.092 for *S. ochranthum* and -5.681 for *S. lycopersicoides*.

The α values of each population were averaged over the candidate genes and over the reference genes. The populations have on average negative α values. Only the population LA0456 has a positive α value for the mean of the reference genes.

The α values of each gene were further averaged over all populations. Two candidate genes, *pLC30-15* and *TPP*, and one reference gene, *CT268*, have positive α values with both outgroup species (Tables B3.33 – B3.36). Three other genes have a positive α value only with one outgroup species: the candidate gene *JERF1* and the reference gene *CT021* with *S. ochranthum* and the candidate gene *LTP* with *S. lycopersicoides*. These six genes have also positive α values in the majority of the populations. The candidate genes *AREB1* and *DREB3* (only with *S. ochranthum*) have also many populations with positive α values, but on average they have negative α values.

BayeScan

This method is based on the allele frequencies between populations. Among the identified outlier SNPs, the α value determines which type of selection is most likely to explain the pattern. BayeScan was run twice on the data set: once including multiple hits (5880 SNPs) and once excluding multiple hits (5390 SNPs).

For the run with multiple hits, 265 outlier SNPs were discovered with a false discovery rate of 5 %. 39 of them have a positive α value, which indicates positive selection (Table 3.6). The run without multiple hits identified 244 outlier SNPs and 37 of them had a positive α value. All of these 37 SNPs were also identified in the run with multiple hits. Most of these outlier SNPs are either in high frequency in the northern populations, in all populations except for the northern populations or in the coastal populations (Table 3.6). Only 13 of these SNPs are in candidate genes. Four of them are in the regulatory gene *AREB1* and mainly in high frequency in the coastal populations. Five SNPs are in the regulatory gene *JERF3*. Two of them are in high frequency in the four southern high altitude populations. The other three SNPs in *JERF3*, the SNP in the regulatory gene *DREB3*, and the three SNPs in the functional gene *pLC30-15* are all in high frequency in some of the populations from the northern range of the central group and/or in populations from the

northern group. Except for one SNP in *pLC30-15*, all outlier SNPs are either synonymous or intronic.

The remaining 26 outlier SNPs for positive selection are in reference genes. 16 of them are in the reference gene *CT189*, which encodes a 40S ribosomal protein. The outlier SNPs in *CT021*, *CT189*, and *GBSSI* are either in high frequency in northern populations or in all populations except for the northern populations. The SNPs in *CT143*, *CT251*, and *CT268* are mainly in high frequency in coastal populations and the SNP in *CT166* is in high frequency in the two northern populations and in all four coastal populations. Two SNPs in *CT268* and one SNP in *GBSSI* are nonsynonymous, all other outlier SNPs in the reference genes are either synonymous or in noncoding regions.

Outlier SNPs with a negative α value are assumed to be SNPs under either purifying or balancing selection. SNPs with intermediate frequencies are assumed to be more likely under balancing than purifying selection. In the run with multiple hits 226 outlier SNPs with a negative α value were identified and in the run without multiple hits 207 SNPs. 29 SNPs in the run with multiple hits and 20 in the run without multiple hits are in intermediate frequency in at least one population (Table B3.37). Most of them have intermediate frequency in several of the populations from the central region of the species distribution as well as in populations from the coast, the north and/or the southern high altitude group. Ten of the SNPs are in candidate genes and 19 in reference genes. The reference gene *CT192* (ribosomal protein S6 kinase alpha-3) has six outlier SNPs which is the highest number any gene has. 11 of the SNPs are nonsynonymous and are found in the candidate genes *AREB1*, *ER5*, *JERF3*, *le25*, and *LTP*, and in the reference genes *CT021*, *CT093*, *CT192*, *CT251*, *CT268*, and *GBSSI*.

Table 3.6: Outlier SNPs for positive selection identified with BayeScan. List of the outlier SNPs (FDR 5 %, q value < 0.05) with positive α value from the run with multiple hits.

SNP	Gene	Type	α value	q value	Frequency > 0.5
175	*AREB1*	intron	1.187	0.0268	LA4107, LA4108
181[a]	*AREB1*	intron	0.718	0.0253	LA2750, LA2932, LA4107, LA4108
217	*AREB1*	intron	0.765	0.0274	LA2750, LA2932, LA4107, LA4108
295	*AREB1*	intron	0.614	0.0396	LA2750, LA2932, LA3111, LA4107, LA4108, LA4119, LA4332
808	*DREB3*	S	1.512	0.0057	LA1958, LA3784
1422	*JERF3*	S	1.251	0.0013	LA2880, LA4118, LA4119, LA4332
1465	*JERF3*	S	1.243	0.0247	LA3784
1603	*JERF3*	intron	1.000	0.0403	LA1930, LA3784
1608	*JERF3*	S	1.305	0.0073	LA1930, LA3784
1614	*JERF3*	S	0.874	0.0383	LA2880, LA4118, LA4119, LA4332
2156	*pLC30-15*	S	1.022	0.0416	LA1930, LA3784
2164	*pLC30-15*	NS	1.246	0.0030	LA3111, LA3784
2179	*pLC30-15*	intron	0.780	0.0377	LA1963, LA3111, LA3784
3067	*CT021*	intron	0.807	0.0160	LA0456, LA0458, LA1958, LA1963, LA2747, LA2748, LA2750, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
3957	*CT143*	3‘ UTR	1.537	0.0002	LA2750, LA2932
4148	*CT166*	intron	0.643	0.0237	LA1930, LA3784, LA2750, LA2932, LA4107, LA4108
4554	*CT189*	S	1.026	0.0084	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
4557	*CT189*	S	1.156	0.0024	LA0456, LA1930, LA3784
4558[a]	*CT189*	S	1.178	0.0018	LA0456, LA1930, LA3784
4559	*CT189*	S	1.336	0.0005	LA0456, LA1930, LA3784
4568	*CT189*	intron	1.329	0.0003	LA0456, LA1930, LA3784
4579	*CT189*	intron	1.321	0.0002	LA0456, LA1930, LA3784
4582	*CT189*	intron	0.747	0.0328	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
4583	*CT189*	intron	0.753	0.0317	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
4670	*CT189*	intron	1.196	0.0016	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118 LA4119, LA4332
4688	*CT189*	intron	1.196	0.0014	LA0456, LA1930, LA3784
4694	*CT189*	intron	0.787	0.0222	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
4706	*CT189*	intron	0.803	0.0194	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
4715	*CT189*	intron	1.183	0.0018	LA0456, LA1930, LA3784

Table 3.6: continued.

4717	*CT189*	intron	0.897	0.0086	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
4721	*CT189*	intron	0.894	0.0109	LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332
4728	*CT189*	intron	0.868	0.0323	LA0456, LA1930, LA3784
5327	*CT251*	S	0.799	0.0127	LA2748, LA2750, LA2932, LA4107, LA4108
5548	*CT268*	S	1.195	0.0171	LA2750, LA2932
5626	*CT268*	NS	1.537	0.0004	LA2750, LA2932
5657	*CT268*	NS	1.198	0.0152	LA2750, LA2932
5727	*GBSSI*	NS	1.394	0.0027	LA1930, LA3784
5752	*GBSSI*	intron	1.209	0.0078	LA1930, LA3784
5761	*GBSSI*	S	1.191	0.0101	LA0456, LA0458, LA1958, LA1963, LA1968, LA2747, LA2748, LA2750, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA2932, LA3111, LA4107, LA4108, LA4118, LA4119, LA4332

Note: "S" synonymous SNP, "NS" nonsynonymous SNP, "UTR" untranslated region
[a]: SNPs that were not identified in the run without multiple hits

Genetic differentiation within groups

A significant pattern of isolation by distance is present in the whole data set and a trend for isolation by distance was observed in the coastal group and in the southern high altitude group (3.2.2). This characteristic of the data set was employed to identify genes under selection in the two most extreme environmental groups: the coastal group and the southern high altitude group. The F_{ST} values for a gene under identical selection pressure should not increase with geographic distance, but stay rather low.

The mean F_{ST} value between populations from the coastal group is 0.216 for all genes and 0.247 for the reference genes (Table 3.7). Four candidate genes and two reference genes have reduced F_{ST} values (*i.e.* < 50 % of $F_{ST, all}$). These genes are *JERF3*, *dhn1*, *le25*, *CT208*, *CT166*, and *CT192*. Mean $F_{ST, all}$ between populations from the southern high altitude group is 0.139 and mean $F_{ST, ref}$ is 0.157 (Table 3.8). Four candidate genes and two reference genes have reduced F_{ST} values. These genes are *AREB2*, *JERF3*, *DREB3*, *CT208*, *CT093*, *CT182*.

Table 3.7: Genetic differentiation within the coastal group. Pairwise genetic differentiation, F_{ST}, for the *S. chilense* populations from the coastal group.

Gene	Population pair						Mean
	LA2750/ LA2932	LA2750/ LA4108	LA2750/ LA4107	LA2932/ LA4108	LA2932/ LA4107	LA4108/ LA4107	
Ø all genes	0.070	0.223	0.262	0.314	0.363	0.065	0.216
Ø reference genes	0.082	0.256	0.271	0.390	0.416	0.067	0.247
candidate genes							
NtC7	0.079	0.239	0.247	0.298	0.316	0.094	0.212
AREB1	0.028	0.294	0.396	0.439	0.555	0.205	0.320
AREB2	n. a.	n. a.	n. a.	0.450	0.504	0.050	0.335
JERF1	0.105	0.222	0.228	0.441	0.437	0.021	0.242
JERF3	0.028	0.060	0.097	0.105	0.150	0.086	0.088
DREB3	0.139	0.132	0.143	0.218	0.262	0.055	0.158
dhn1	0.032	0.033	0.041	-0.001	0.027	0.025	0.026
pLC30-15	0.079	0.152	0.172	0.328	0.369	0.028	0.188
TAS14	0.065	0.384	0.384	0.259	0.258	-0.003	0.224
ER5	0.018	0.228	0.219	0.207	0.200	-0.008	0.144
le25	0.006	0.005	0.006	0.001	-0.006	-0.005	0.001
LTP	0.054	0.365	0.350	0.284	0.271	0.018	0.224
TSW12	0.134	0.084	0.061	0.324	0.336	0.050	0.165
CT208	0.032	0.065	0.075	0.029	0.061	0.050	0.052
His1	0.024	0.290	0.604	0.225	0.517	0.144	0.301
TPP	0.053	0.343	0.782	0.362	0.806	0.202	0.424
reference genes							
CT021	0.319	0.164	0.186	0.622	0.695	0.033	0.336
CT066	0.119	0.317	0.353	0.264	0.321	0.091	0.244
CT093	0.086	0.293	0.333	0.337	0.393	0.045	0.248
CT114	0.040	0.295	0.318	0.359	0.375	0.223	0.268
CT143	0.043	0.228	0.216	0.340	0.316	0.010	0.192
CT166	0.053	0.017	0.021	0.125	0.109	0.009	0.056
CT179	0.046	0.338	0.404	0.460	0.549	0.078	0.312
CT182	0.016	0.560	0.540	0.570	0.549	0.048	0.380
CT189	0.276	0.129	0.104	0.666	0.698	0.167	0.340
CT192	0.035	0.148	0.115	0.154	0.119	-0.003	0.095
CT198	0.043	0.404	0.389	0.673	0.636	0.023	0.361
CT251	0.012	0.106	0.135	0.138	0.171	0.099	0.110
CT268	0.058	0.487	0.500	0.610	0.628	0.097	0.397
GBSSI	0.004	0.094	0.187	0.138	0.270	0.023	0.119

Table 3.8: Genetic differentiation within the southern high altitude group. Pairwise genetic differentiation, F_{ST}, for the *S. chilense* populations from the southern high altitude group.

Gene	Population pair						Mean
	LA4332/ LA4118	LA4332/ LA4119	LA4332/ LA2880	LA4118/ LA4119	LA4118/ LA2880	LA4119/ LA2880	
Ø all genes	0.133	0.174	0.181	0.111	0.133	0.101	0.139
Ø reference genes	0.158	0.161	0.225	0.093	0.190	0.118	0.157
candidate genes							
NtC7	0.200	0.191	0.246	-0.006	0.059	0.038	0.121
AREB1	0.186	0.498	0.352	0.318	0.124	0.303	0.297
AREB2	0.048	0.095	0.128	0.004	0.020	-0.002	0.049
JERF1	0.044	0.044	0.072	0.037	0.167	0.102	0.078
JERF3	0.017	0.012	0.017	0.034	0.044	-0.007	0.019
DREB3	0.080	0.037	0.055	0.076	0.148	0.004	0.067
dhn1	0.082	0.123	0.225	0.131	0.237	-0.001	0.133
pLC30-15	0.311	0.276	0.323	0.038	0.041	0.002	0.165
TAS14	0.121	0.140	0.093	0.037	0.015	0.046	0.075
ER5	0.062	0.083	0.163	0.100	0.079	0.126	0.102
le25	0.023	0.533	0.182	0.498	0.102	0.284	0.270
LTP	0.219	0.128	0.116	0.040	0.042	-0.008	0.089
TSW12	0.141	0.140	0.065	0.141	0.098	0.067	0.109
CT208	0.030	0.019	0.021	0.018	0.021	-0.007	0.017
His1	0.086	0.473	0.102	0.515	0.037	0.433	0.275
TPP	0.149	0.153	0.168	0.039	0.090	0.013	0.102
reference genes							
CT021	0.121	0.119	0.106	0.000	0.173	0.150	0.111
CT066	0.239	0.047	0.235	0.273	0.718	0.146	0.276
CT093	0.010	0.100	0.094	0.058	0.026	0.099	0.064
CT114	0.361	0.333	0.315	-0.004	0.153	0.110	0.211
CT143	0.041	0.209	0.140	0.196	0.121	0.221	0.155
CT166	0.345	0.218	0.402	0.111	0.077	0.068	0.204
CT179	n. a.	n. a.	n. a.	0.177	0.311	0.118	0.202
CT182	0.080	0.076	0.065	-0.004	0.011	0.017	0.041
CT189	0.154	0.039	0.234	0.129	0.415	0.154	0.187
CT192	0.074	0.191	0.108	0.085	0.000	0.128	0.098
CT198	0.104	0.198	0.168	0.122	0.107	0.023	0.120
CT251	0.142	0.112	0.118	0.078	0.042	0.000	0.082
CT268	0.375	0.465	0.572	0.075	0.158	0.072	0.286
GBSSI	0.004	-0.006	0.376	0.006	0.344	0.339	0.177

3.3.4 Single gene evolutionary histories

The summary statistics including divergence were calculated and several methods to identify local adaptation events in the dataset were applied. In the following the interesting findings for each candidate gene are summarized to further understand how each gene evolves in *S. chilense*.

Sensory genes

***NtC7*:** The sensory gene *NtC7* is one of the most polymorphic genes in the data set. Overall it has the highest π and second highest π_a (Table 3.3, B3.3, B3.7). LA1963 lies for θ_W, π, and π_a in the upper 2.5 % of the density distributions, LA1958 and LA1968 for π and π_a, and LA2931 and LA4332 for π (Tables B3.1, B3.3, B3.7). *NtC7* has further on average the highest Tajima's *D* value of all candidate genes (Table 3.3, B3.9, B3.10). Especially high Tajima's *D* values in comparison to the mean of the reference genes are found in high altitude populations (LA2880, LA2931, LA3111, LA4118, LA4119, LA4332), in two populations close to a high altitude population (LA1968, LA0458), and in the coastal population LA2932. The Tajima's *D* value of LA0458 lies further in the upper 2.5 % of the density distribution. BayeScan detected one outlier SNP for balancing selection in the intron of *NtC7* (Table B3.37). This SNP is in intermediate frequency in two coastal populations, LA2750 and LA2932, three populations from high altitudes, LA2773, LA2931, and LA4332, and in five other populations (LA0458, LA1958, LA1963, LA2747, LA2753). The high altitude population LA2755 has a positive α value with both outgroup species while the high altitude population LA2773, the coastal population LA4107, and LA1958 and LA2748 have a positive α value with *S. ochranthum* (Tables B3.33, B3.35). However, it has to be noted that the last two results have to be taken with caution, because the *NtC7* sequences of *S. ochranthum* and *S. lycopersicoides* are highly diverged (*e.g.* Tables B3.11 - B3.14) and the *S. ochranthum* sequence is even pseudogenized.

Regulatory genes

***AREB1*:** The coastal populations have low θ_W, very low π, and negative Tajima's *D* values for the transcription factor *AREB1* (Tables B3.1, B3.3, B3.9). LA2750 lies in the lower 2.5 % of the Tajima's *D* density distribution. Several high altitude populations have elevated θ_W and/or π, and positive Tajima's *D* values (Tables B3.1, B3.3, B3.9). *AREB1* has on average the second lowest divergence compared to the cultivated tomato (Tables 3.4, B3.15). BayeScan identified four outlier SNPs for positive selection in intronic regions (Table 3.6).

All of them are either in high frequency or fixed in the coastal populations. One outlier SNP for balancing selection was also identified (Table B3.37). This SNP is synonymous and in intermediate frequency in LA2747, LA2748, LA2753 and two high altitude populations (LA2773, LA4332). Three coastal populations have a significant McDonald-Kreitman test with both outgroup species (Table 3.5). Further 18 populations have a positive α value with both outgroup species, among them are all high altitude populations, but none of the coastal populations (Tables B3.33, B3.35).

***AREB2*:** The transcription factor *AREB2* has on average the lowest θ_W, π, π_s, K and K_s compared to *S. lycopersicoides*, K_s compared to the cultivated tomato, and the second lowest π_a (Tables 3.3, 3.4, B3.1, B3.3, B3.5, B3.7, B3.13, B3.19). Especially some of the high altitude populations, namely LA2755, LA2880, LA4118, LA4119, have low genetic variation and some of them lie in the lower 2.5 % of the density distributions. Contrasting patterns are present for Tajima's *D* (Table B3.9). The high altitude populations have either extremely high (LA2773, LA2931) or low (LA2755, LA2880, LA3111, LA4118, LA4119, LA4332) values. LA4119 lies also in the lower 2.5 % of the density distribution. Among the coastal populations LA2932 has a high and LA4107 and LA4108 have low Tajima's *D* values. Several populations have positive α values with either both outgroup species or only with *S. lycopersicoides* (Tables B3.33, B3.35). Among them are the two northern populations, LA1930 and LA3784, the coastal population LA2932, and some high altitude populations (LA2755, LA2773, LA3111, LA4119). The populations from the southern high altitude group have reduced genetic differentiation for *AREB2* (Table 3.8).

***JERF1*:** Overall, *JERF1* has the highest divergence compared to *S. lycopersicoides* and the second highest divergence compared to *S. ochranthum* (Tables 3.4, B3.11, B3.13). Several populations lie in the upper 2.5 % of the density distribution for *K* compared to *S. lycopersicoides*. Among them are all coastal populations, many high altitude populations (LA2755, LA2773, LA2880, LA2931, LA3111) and one of the northern populations (LA1930). LA1930 and LA2755 are the populations with the highest divergence in respect to the other species. Some populations have high π_a values (Table B3.7). These populations include also three high altitude populations (LA2755, LA2765, LA2773) and one of the northern populations (LA3784). The Tajima's *D* value of LA2880 lies in the upper 2.5 % of the density distribution (Table B3.9). The coastal population LA2750 has a significant McDonald-Kreitman test with *S. lycopersicoides* due to an excess of nonsynonymous polymorphism (Table B3.31). The α value of 13 populations is positive with both outgroup species and four other populations have a positive α with *S. ochranthum* (Tables B3.33,

B3.35). Among them are all coastal populations, the two northern populations and some of the high altitude populations.

***JERF3*:** Some populations have high π_a values for *JERF3* (Table B3.7). These populations include two high altitude populations (LA2755, LA3111), but also one northern population (LA1930) and one coastal population (LA4107). The northern population LA3784 has the lowest divergence compared to both outgroup species and the cultivated tomato (Tables B3.11, B3.13, B3.15). The divergence of LA2765 compared to the cultivated tomato instead, is very high (Table B3.15). The Tajima's *D* values of LA3784 and LA4119 lie in the lower 2.5 % of the density distribution (Table B3.9). BayeScan identified five outlier SNPs for positive selection and two outlier SNPs for balancing selection (Tables 3.6, B3.37). Two of the outlier SNPs for positive selection are in high frequency in the southern high altitude populations. The other three SNPs for positive selection are in high frequency in the northern populations. One of the balancing selection SNPs is nonsynonymous and in intermediate frequency in six populations. Three of them are high altitude populations (LA2755, LA2765, LA3111). This nonsynonymous SNP leads to an amino acid change from threonine to serine (ACT $\rightarrow$ TCT). The other balancing selection SNP is in the intron and is in intermediate frequency in LA2755 and in two coastal populations (LA2750, LA2932). Seven populations have a significant McDonald-Kreitman test with *S. ochranthum* due to an excess of nonsynonymous polymorphism (Table B3.29). These populations include three coastal populations (LA2750, LA2932, LA4107) and two high altitude populations (LA2755, LA3111). Another high altitude population, LA4118, has a positive α value with both outgroup species, while ten other populations have a positive α only with *S. lycopersicoides* (Tables B3.33, B3.35). Among them are also five high altitude populations. The populations from the coastal group and the populations from the southern high altitude group have reduced genetic differentiation (Tables 3.7, 3.8).

***DREB3*:** *DREB3* has overall the highest π_s values (Tables 3.3, B3.5, B3.6). Several populations are in the upper 2.5 % of the π_s density distribution. Many of them are high altitude populations. Some of the high altitude populations have also very negative Tajima's *D* values in comparison to the reference genes (Tables B3.9, B3.10). The coastal population LA2750 lies in the upper 2.5 % of the Θ_W density distribution (Table B3.1). The coastal populations have the lowest divergence compared to *S. lycopersicoides* while the northern populations have the highest divergence compared to *S. lycopersicoides* (Table B3.19). BayeScan identified one outlier SNP for positive selection (Table 3.6). This SNP is synonymous and in high frequency in LA1958 and fixed in LA3784. The McDonald-Kreitman test is significant for two coastal populations, LA2750 and LA4107, with both

outgroup species, and for the two other coastal populations, LA2932 and LA4108, as well as for LA0458 and LA3784 with *S. lycopersicoides* (Table 3.5, B3.29, B3.31). In all cases the test is significant due to an excess of nonsynonymous polymorphism. LA2748 and LA2765 have a positive α value with both outgroup species and 14 other populations have a positive α with *S. ochranthum* (Tables B3.33, B3.35). Altogether, all high altitude populations have a positive α with *S. ochranthum*. Furthermore, the two northern populations are among the populations that have a positive α value. The populations from the southern high altitude group have reduced genetic differentiation (Table 3.8).

Functional genes

***dhn1*:** Overall, *dhn1* has very low sequence variation and divergence in respect to the other three species (Tables 3.3, 3.4). LA2880 lies in the lower 2.5 % of the π and π_s density distributions (Tables B3.3, B3.5). LA4108 lies also in the lower 2.5 % of the π density distribution. LA4118 lies in the lower 2.5 % of the π_a and K_a compared to *S. lycopersicoides* and cultivated tomato density distributions (Tables B3.7, B3.25, B3.27). LA4119 lies in the lower 2.5 % of the Tajima's *D* values density distribution (Table B3.9). The two northern populations have the highest *K* in comparison to all other species (Tables B3.11, B3.13, B3.15). Two high altitude populations (LA2755, LA2765) have also high divergence values. BayeScan identified one outlier SNP for balancing selection in LA2748 (Table B3.37). This SNP is synonymous. Four populations have a significant McDonald-Kreitman test with both outgroup species due to an excess of nonsynonymous polymorphism, eight populations only with *S. lycopersicoides* (Table 3.5, B3.29, B3.31). Among these populations are all coastal populations and many high altitude populations. Furthermore, seven populations have a positive α value with *S. ochranthum* (Table B3.33). Among them are also some high altitude populations, but not the same than with a significant McDonald-Kreitman test. The populations from the coastal group have reduced genetic differentiation (Table 3.7).

***pLC30-15*:** Overall, the dehydrin *pLC30-15* has high π_a values and low divergence from *S. ochranthum* (Tables 3.3, 3.4, B3.7, B3.11). The population LA1963 lies in the upper 2.5 % of the π_a density distribution. The population LA3111 has a very negative Tajima's *D* value and has the highest K_a in comparison to the two outgroup species and the cultivated tomato (Tables B3.9, B3.23, B3.25, B3.27). The two northern populations, LA1930 and LA3784, have also quite high divergence values compared to the other species (Tables B3.11, B3.13, B3.15). BayeScan identified three outlier SNPs for positive selection (Table 3.6). The first one is synonymous and in high frequency in the two northern populations LA1930 and LA3784. The second one is nonsynonymous and leads to an amino acid change from

asparagine to lysine (AAC → AAG). This SNP is present in low frequency in three populations (LA1930, LA1963, LA2931), in LA3111 in high frequency and is fixed in LA3784. The third SNP is in the intron and in low frequency in five populations, in high frequency in two populations (LA1963, LA3111) and fixed in LA3784. Interestingly, the *pLC30-15* gene was reported to be under diversifying selection in a *S. chilense* population that is close to LA3784 (Xia *et al.* 2010). A total of 12 populations have a positive α value with both outgroup species (Tables B3.33, B3.35). Among them are all high altitude populations. The average α over all populations is also positive with both outgroup species. Eight other populations have a positive α with *S. ochranthum*.

***TAS14*:** Except for LA0458, LA4107, and LA4108, the populations have high π_a values for *TAS14* (Table B3.7). The π_a values of LA3784 and LA4332 lie in the upper 2.5 % of the density distribution. LA3784 lies also in the upper 2.5 % of the Θ_W and π density distributions (Tables B3.1, B3.3). The other northern population, LA1930, lies also in the upper 2.5 % of the π density distribution and the coastal population LA4108 lies in the lower 2.5 % of the π density distribution (Tables B3.3). The Tajima's *D* values for the southern high altitude populations are highly elevated in comparison to the mean of the reference genes. LA2880 and LA4118 lie also in the upper 2.5 % of the Tajima's *D* density distribution (Table B3.9). Nine populations have a significant McDonald-Kreitman test with *S. lycopersicoides* due to an excess of nonsynonymous polymorphism (Table B3.31). Among them are the two northern populations and one of the southern high altitude populations, namely LA4332. Further LA0458 has a positive α value (Table B3.35). But it has to be noted that the recovered *TAS14* sequence of *S. lycopersicoides* is quite diverged from the *S. chilense* sequences (*e.g.* Table B3.19). Therefore these results should be taken with caution.

***ER5*:** Overall, *ER5* has high π_s values and the second lowest Tajima's *D* value (Tables 3.3, B3.5, B3.9). LA1958 is the only population with a positive Tajima's *D* value. LA2750 and LA4119 are in the lower 2.5 % of the Tajima's *D* density distribution. Two populations (LA1958, LA1963) are in the upper 2.5 % and two populations (LA2880, LA4107) are in the lower 2.5 % of the π_s density distribution. The northern population LA3784 is in the upper 2.5 % of the π_a density distribution (Table B3.7). The two northern populations have further the highest divergence in comparison to the two outgroup species and the cultivated tomato (Tables B3.11, B3.13, B3.15). Both are also in the upper 2.5 % density distribution for K_s compared to *S. ochranthum* and the cultivated tomato (Tables B3.17, B3.19). Several other populations are also in the upper 2.5 % of these density distributions. BayeScan identified two outlier SNPs for balancing selection (Table 3.37). The first one is in the intron and in intermediate frequency in the northern population LA1930

and the coastal population LA2932. The second SNP is nonsynonymous leading to an amino acid change from threonine to alanine (ACA → GCA) and in intermediate frequency in LA2753. LA2750 has a positive McDonald-Kreitman test with both outgroup species and LA0456 with *S. ochranthum* (Tables 3.5, B3.29, B3.31). Both are significant due to an excess of nonsynonymous polymorphism. Furthermore, LA1958 has a positive α value with both outgroup species and LA1930 with *S. lycopersicoides* (Tables B3.33, B3.35).

***le25*:** Overall, *le25* has the highest Θ_W and π_a and the lowest Tajima's *D* values (Tables 3.3, B3.1, B3.7, B3.9). Except for the neighbouring populations LA2747 and LA2773 all populations have negative Tajima's *D* values, which are for most populations more negative than the mean of the reference genes (Table B3.9, B3.10). The difference between *le25* and the reference genes is especially large for many high altitude populations. Several populations lie in the upper 2.5 % of the Θ_W and/or π_a density distributions (Tables B3.1, B3.7). Some other populations lie in the lower 2.5 % of the π_s and/or K_s compared to *S. lycopersicoides* density distributions (Tables B3.5, B3.19). The populations LA2750 and LA3111 lie further in the lower 2.5 % of the Tajima's *D* density distribution (Table B3.9). The high altitude populations LA3111 and LA4332 have the highest K_a all other species (Tables B3.23, B3.25, B3.27). BayeScan identified one outlier SNP for balancing selection which is in intermediate frequency in LA4119 (Table B3.37). This SNP is nonsynonymous and leads to an amino acid change from glutamic acid to alanine (GAG → GCG). Three populations have a significant McDonald-Kreitman test with *S. ochranthum* due to an excess of nonsynonymous polymorphism (Table B3.29). Two of these populations are coastal (LA2932, LA4107) and the third is from high altitudes (LA2880). The northern population LA1930 has a positive α value with both outgroup species, and the populations LA1963, LA2750, LA2755, LA2765, LA3111, and LA4332 have a positive α with *S. lycopersicoides* (Tables B3.33, B3.35). Among the latter are four populations from high altitudes. The populations from the coastal group have reduced genetic differentiation (Table 3.7).

***LTP*:** The lipid transfer protein encoding gene *LTP* has overall the highest divergence in respect to the other species (Table 3.4). It has *e.g.* the highest K_a compared to both outgroup species and compared to the cultivated tomato. Most populations and also the averages over all populations lie in the upper 2.5 % of the three K_a density distributions (Tables B3.23, B3.25, B3.27). Two coastal populations, LA4108 and LA4107, lie in the lower 2.5 % of the π_s and π_a density distributions, respectively (Tables B3.5, B3.7). LA0456 lies further in the upper 2.5 % of the π_a density distribution. Several other populations have very high π_a values in respect to the average over the reference genes. These populations include the two northern populations and six of the high altitude populations (LA2765, LA2773, LA2880,

LA4118, LA4119, LA4332). Two high altitude populations, LA2773 and LA2931, have very negative Tajima's *D* value in comparison to the reference genes while two other high altitude populations, LA2880 and LA4332, have a very high Tajima's *D* value (Tables B3.9, B3.10). BayeScan identified one outlier SNP for balancing selection which is in intermediate frequency in ten populations (Table B3.37). Six of these populations are from high altitude (LA2755, LA2765, LA2773, LA2880, LA4119, LA4332). This SNP is nonsynonymous and leads to an amino acid change from serine to alanine (TCT → GCT). Ten populations have a positive α value with both outgroup species (Tables B3.33, B3.35). Among them are a northern populations (LA1930), three coastal populations (LA2750, LA2932, LA4107), and three high altitude populations (LA2755, LA2931, LA3111). Furthermore, eight populations have a positive α with *S. lycopersicoides*. The average over all populations has also a positive α with *S. lycopersicoides*.

***TSW12*:** *TSW12* has on average high variation (2nd highest Θ_W and π, 3rd highest π_s; Table 3.3) and high divergence (*K* compared to *S. ochranthum* and *S. lycopersicoides*, K_s compared to *S. ochranthum*; Table 3.4). Except for LA0458 and LA4107 all populations have a very high π_a (Table B3.7). These two populations lie also in the lower 2.5 % of the π_s density distribution (Table B3.5). LA1958 lies in the upper 2.5 % of the π_a density distribution. Two populations from the southern high altitude group, LA2880 and LA4119, have very high Θ_W values and LA2880 has further the highest *K* compared to the cultivated tomato (Tables B3.1, B3.15). The two northern populations, LA1930 and LA3784, have high divergence in comparison to both outgroup species and the cultivated tomato (Tables B3.11, B3.13, B3.15). LA2750 has a significant McDonald-Kreitman test with *S. ochranthum* due to an excess of nonsynonymous polymorphism (Table B3.29).

***CT208*:** Overall, *CT208* is one the least polymorphic genes among the candidate genes (Table 3.3). π_a is very low across all populations except for LA1958 (Table B3.7). Several populations lie in the lower 2.5 % of the density distributions for the variation values: LA4118 for Θ_W, π, and π_s, LA2880 for Θ_W, π, and π_a, LA0458 for Θ_W, LA4119 for π, and LA2747 for π_a (Tables B3.1, B3.3, B3.5, B3.7). Three of these populations are from high altitudes in the south. Four populations have positive Tajima's *D* values that are elevated in respect to the reference genes (Tables B3.9, B3.10). One of these populations, LA2748, lies in the upper 2.5 % of the Tajima's *D* density distribution. BayeScan identified one outlier SNP for balancing selection (Table B3.37). This SNP is intronic and in intermediate frequency in eight populations. Among them are both northern populations and three populations from high altitudes. Six populations have a significant McDonald-Kreitman test with both outgroup species and eight populations only with *S. ochranthum* (Tables 3.5, B3.29, B3.31).

All populations are significant due to an excess of nonsynonymous polymorphism. Among them are three coastal populations and six populations from high altitudes. Furthermore, three populations have a positive α value with both outgroup species and two with *S. lycopersicoides* (Tables B3.33, B3.35). Three populations from high altitude are among them. The populations from the coastal group and from the southern high altitude group have reduced genetic differentiation (Tables 3.7, 3.8). The genetic differentiation between the groups is also low. This result is consistent with previous analysis of *CT208* that showed a clinal pattern with almost no variation in the most southern population of their *S. chilense* sample (Arunyawat *et al.* 2007).

***His1*:** Overall, the histone gene *His1* has low π_s values and K_s compared to *S. ochranthum* (Tables 3.3, 3.4). The population LA0458 lies in the lower 2.5 % of the density distribution of π_s (Table B3.5). Four populations have Tajima's *D* values that are highly elevated in comparison to the mean of the reference genes (Tables B3.9, B3.10). These populations are LA0456, LA0458, LA2773, and LA2932. The coastal population LA2750 has low Θ_W and π values as well as the lowest *K* compared to *S. ochranthum* (Tables B3.1, B3.3, B3.11). Another coastal population, LA4107, has the highest K_s compared to *S. ochranthum* and the cultivated tomato, a very high π_s, lies in the upper 2.5 % of the π_a density distribution and in the lower 2.5 % of the Tajima's *D* density distribution (Tables B3.5, B3.7, B3.9, B3.17, B3.21). Four high altitude populations have a positive α value with *S. ochranthum* (LA2773, LA2880, LA4118, LA4332; Table B3.33).

***TPP*:** The trehalose-6-phosphate phosphatase *TPP* has on average a slightly positive Tajima's *D* value (Table 3.3). Tajima's *D* is for 12 populations elevated in respect to the mean of the reference genes (Tables B3.9, B3.10). These populations include the two northern populations, LA1930 and LA3784, and some of the high altitude populations (LA2755, LA2773, LA2931, LA4332). Concerning the Tajima's *D* values, LA0458 lies in the upper 2.5 % and LA4107 lies in the lower 2.5 % of the density distribution. LA4107 lies also in the lower 2.5 % of the π density distribution and has the highest *K* compared to *S. lycopersicoides* and the highest K_s compared to both outgroup species and the cultivated tomato (Tables B3.5, B3.3.13, B3.17, B3.19, B3.21). Four populations have an excess of π_a and one of them, LA1958, has also the highest K_a compared to both outgroup species and the cultivated tomato (Tables B3.7, B3.23, B3.25, B3.27). The average α value is also positive with both outgroup species and 15 populations have positive α values with both outgroup species (Tables B3.33, B3.35). These populations include the two northern populations, LA1930 and LA3784, two of the coastal populations (LA2932, LA4108) and seven of the

high altitude populations (LA2765, LA2773, LA2880, LA2931, LA3111, LA4119, LA4332). Another high altitude population, LA2755, has a positive α with *S. ochranthum*.

3.4 Analysis of the consensus sequence data set

3.4.1 Gene evolution on the species level

Consensus sequences of every gene were compiled for each population. These consensus sequences represent average alleles for each population. It was anticipated that these consensus sequences can be used to understand gene evolution on the species level similar to a species wide sampling approach (Stadler *et al.* 2009). This data set was used to assess polymorphism, divergence and to perform the neutrality test statistic Tajima's *D* of each gene.

The polymorphism (θ_W, π, π_s, π_a), divergence (*K*, K_s, K_a, compared to *S. ochranthum*, *S. lycopersicoides*, *S. lycopersicum*) and Tajima's *D* values were averaged over the reference genes and over the candidate genes (Tables 3.9, 3.10). On average the candidate genes have higher θ_W, π and π_a values than the reference genes. The candidate genes have further a higher divergence from the three other species and have also a more negative Tajima's *D* value. Among the candidate genes *le25* has the highest θ_W and π, *DREB3* has the highest π_s and *NtC7* has the highest π_a. *CT208* and *AREB2* have the lowest genetic variation. Overall *NtC7*, *le25*, *TAS14* and *ER5* have the highest divergence and *AREB2* and *TPP* have the lowest divergence. All candidate genes have negative Tajima's *D* values. *ER5* has the most and *TPP* the least negative one.

A comparison between the regulatory genes and the functional genes revealed that the functional genes have on average higher values for θ_W, π and π_a than the regulatory genes. The functional genes have also a higher divergence in comparison the three other species and have a more negative Tajima's *D* value (Tables 3.9, 3.10).

Table 3.9: Summary statistics of the consensus sequence data set.

Gene	θ_W	π	π_s	π_a	Tajima's *D*
sensory gene					
NtC7	0.0230	0.0185	0.0221	0.0096	-0.777
regulatory genes					
AREB1	0.0121	0.0094	0.0139	0.0014	-0.875
AREB2	0.0060	0.0053	0.0052	0.0008	-0.489
JERF1	0.0143	0.0123	0.0128	0.0029	-0.550
JERF3	0.0129	0.0091	0.0130	0.0046	-1.174
DREB3	0.0091	0.0071	0.0297	0.0007	-0.827
regulatory genes mean	0.0109	0.0086	0.0149	0.0021	-0.783
functional genes					
dhn1	0.0087	0.0054	0.0062	0.0024	-1.384
pLC30-15	0.0166	0.0116	0.0126	0.0055	-1.164
TAS14	0.0140	0.0093	0.0194	0.0056	-1.289
ER5	0.0130	0.0074	0.0161	0.0033	-1.611
le25	0.0426	0.0306	0.0097	0.0045	-1.133
LTP	0.0152	0.0122	0.0146	0.0054	-0.750
TSW12	0.0122	0.0100	0.0074	0.0075	-0.664
CT208	0.0077	0.0049	0.0007	0	-1.403
His1	0.0087	0.0066	0.0046	0.0040	-0.901
TPP	0.0118	0.0105	0.0148	0.0007	-0.424
functional genes mean	0.0150	0.0108	0.0106	0.0039	-1.072
candidate genes mean	0.0142	0.0106	0.0127	0.0037	-0.964
reference genes					
CT021	0.0255	0.0170	0.0216	0.0019	-1.334
CT066	0.0060	0.0076	0.0283	0.0008	0.965
CT093	0.0021	0.0021	0.0070	0.0004	-0.036
CT114	0.0050	0.0034	0.0039	0.0011	-1.218
CT143	0.0239	0.0225	0.0205	0.0003	-0.240
CT166	0.0089	0.0101	0.0152	0	0.487
CT179	0.0138	0.0117	0.0227	0	-0.590
CT182	0.0159	0.0092	0.0050	0.0007	-1.673
CT189	0.0102	0.0074	0.0130	0	-1.096
CT192	0.0111	0.0090	0.0121	0.0013	-0.767
CT198	0.0082	0.0062	0.0102	0.0012	-0.933
CT251	0.0081	0.0075	0.0122	0.0040	-0.298
CT268	0.0073	0.0080	0.0237	0.0033	0.400
GBSSI	0.0026	0.0013	0.0032	0.0004	-1.774
reference genes mean	0.0106	0.0088	0.0142	0.0011	-0.579

Table 3.10: Divergence for the consensus sequence data set. All sites (*K*), synonymous sites (K_s) and nonsynonymous sites (K_a) divergence from *S. ochranthum*, *S. lycopersicoides*, and *S. lycopersicum*.

	S. ochranthum			***S. lycopersicoides***			***S. lycopersicum***		
	K	***Ks***	***Ka***	***K***	***Ks***	***Ka***	***K***	***Ks***	***Ka***
<u>sensory gene</u>									
NtC7	0.2070	0.3466	0.1824	0.2165	0.3633	0.1751	0.0261	0.0437	0.0205
<u>regulatory genes</u>									
AREB1	0.0384	0.0528	0.0081	0.0397	0.0693	0.0127	0.0105	0.0125	0.0038
AREB2	0.0339	0.0447	0.0119	0.0181	0.0265	0.0067	0.0145	0.0053	0.0054
JERF1	0.0542	0.0736	0.0200	0.0715	0.0189	0.0189	0.0199	0.0217	0.0115
JERF3	0.0327	0.0555	0.0146	0.0333	0.0360	0.0223	0.0206	0.0302	0.0093
DREB3	0.0250	0.0771	0.0102	0.0217	0.0752	0.0065	0.0106	0.0348	0.0038
mean	0.0368	0.0607	0.0130	0.0369	0.0452	0.0134	0.0152	0.0209	0.0068
<u>functional genes</u>									
dhn1	0.0108	0.0519	0.0043	0.0381	0.0944	0.0013	0.0108	0.0112	0.0040
pLC30-15	0.0282	0.0260	0.0176	0.0345	0.0398	0.0174	0.0178	0.0257	0.0081
TAS14	n. a.	n. a.	n. a.	0.0654	0.2026	0.0290	0.0255	0.0303	0.0044
ER5	0.0324	0.1031	0.0077	0.0323	0.0608	0.0111	0.0178	0.0428	0.0077
le25	0.0575	0.0731	0.0289	0.0778	0.0310	0.0456	0.0674	0.0250	0.0081
LTP	0.0616	0.1153	0.0471	0.0482	0.0777	0.0427	0.0480	0.0653	0.0531
TSW12	0.0536	0.0876	0.0175	0.0325	0.0273	0.0091	0.0258	0.0393	0.0161
CT208	0.0396	0.0649	0	0.0321	0.0649	0	0.0162	0.0084	0
His1	0.0259	0.0403	0.0138	n. a.	n. a.	n. a.	0.0234	0.0326	0.0138
TPP	0.0277	0.0373	0.0077	0.0307	0.0370	0.0088	0.0163	0.0298	0.0028
mean[a]	0.0389	0.0699	0.0164	0.0408	0.0541	0.0170	0.0269	0.0310	0.0118
mean[b]	0.0381	0.0664	0.0151	0.0393	0.0507	0.0156	0.0232	0.0287	0.0108
<u>reference genes</u>									
CT021	0.0671	0.1124	0.0224	n. a.	n. a.	n. a.	0.0412	0.0320	0.0040
CT066	0.0279	0.0947	0.0057	0.0293	0.1000	0.0061	0.0136	0.0457	0.0031
CT093	0.0138	0.0313	0.0064	0.0149	0.0445	0.0064	0.0075	0.0203	0.0037
CT114	0.0189	0.0253	0.0070	0.0281	0.0395	0.0139	0.0102	0.0022	0.0012
CT143	0.0340	0.0725	0.0001	0.0319	0.0430	0.0001	0.0207	0.0148	0.0001
CT166	0.0293	0.0520	0	0.0326	0.0520	0	0.0155	0.0100	0
CT179	0.0390	0.0796	0.0024	0.0324	0.0779	0.0047	0.0267	0.0475	0
CT182	0.0429	0.0768	0.0034	0.0321	0.0451	0.0004	0.0176	0.0124	0.0004
CT189	0.0369	0.0289	0.0038	0.0434	0.0790	0	0.0250	0.0474	0
CT192	0.0515	0.0760	0.0060	0.0480	0.0784	0.0053	0.0179	0.0255	0.0008
CT198	0.0381	0.0527	0.0111	0.0356	0.0658	0.0072	0.0265	0.0263	0.0033
CT251	0.0464	0.0859	0.0299	0.0398	0.0810	0.0208	0.0222	0.0374	0.0105
CT268	0.0230	0.0538	0.0137	0.0181	0.0351	0.0130	0.0125	0.0335	0.0062
GBSSI	0.0232	0.0391	0.0079	0.0292	0.0391	0.0125	0.0101	0.0141	0.0033
mean[c]	0.0327	0.0591	0.0075	0.0319	0.0600	0.0070	0.0191	0.0264	0.0026

Note: "n. a." not available

[a]: mean functional genes calculated excluding *His1* and *TAS14*

[b]: mean candidate genes calculated excluding *NtC7*, *His1*, and *TAS14*

[c]: calculated excluding *CT021*

3.4.2 Phylogenetic analyses

In order to understand the relationships between the *S. chilense* populations, phylogenetic trees of the concatenated consensus sequence alignments were constructed (Figure 3.7, B4.1). These trees also include the sequences from the two outgroup species, *S. ochranthum* and *S. lycopersicoides*. Trees were constructed with and without the cultivated tomato *S. lycopersicum* and for all 30 genes, for the 16 candidate genes and for the 14 reference genes. In all six trees the four coastal populations (LA2750, LA2932, LA4107, LA4108) and the four southern high altitude populations (LA2880, LA4118, LA4119, LA4332) form highly supported clades and both clades are nested within the *S. chilense* populations from the central group. Although the coastal populations and the southern high altitude populations are geographically equally distant to the central and northern *S. chilense* populations, these two clades do not form sister clades in any of the three trees. In the coastal clade, the two northern populations, LA2750 and LA2932, and also the two southern populations, LA4107 and LA4108, are sister taxa. Except for LA2750 and LA2932 in the candidate gene tree, this structure within the coastal group is highly supported. Among the populations of the southern high altitude clade, LA4332 is sister taxa to the other three populations. All *S. chilense* populations, except for the two northern populations LA1930 and LA3784, form a highly supported monophyletic clade in all trees. The two northern populations, LA1930 and LA3784, form a sister clade to all other *S. chilense* populations (but with a weaker support in the trees with the *S. lycopersicum* sequence). Furthermore, in the candidate gene tree the cultivated tomato is closer to the other *S. chilense* populations than LA1930 and LA3784 (Figure 3.7C).

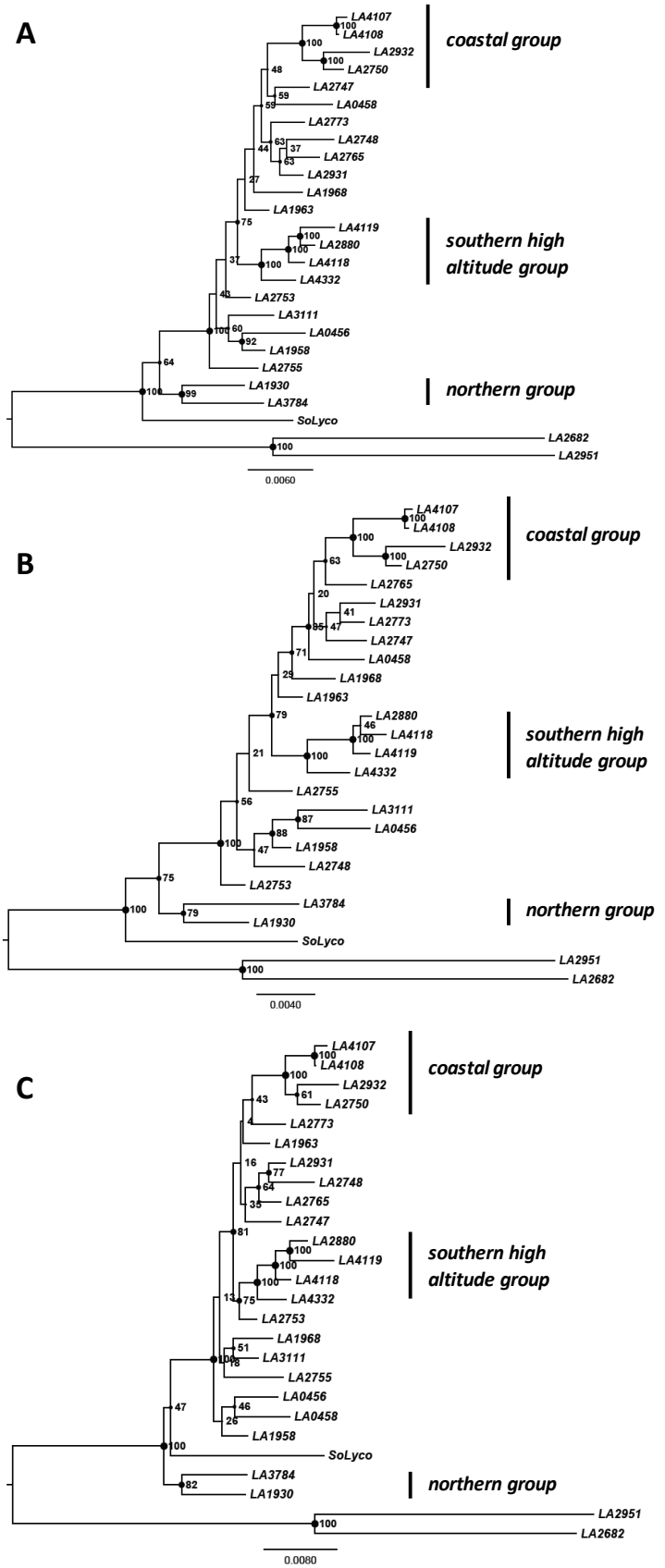

Figure 3.7: Phyolgenetic trees of the consensus sequences with *S. lycopersicum*. SoLyco: sequence of the cultivated tomato *S. lycopersicum*, LA2682: sequence of the outgroup *S. ochranthum*, LA2951: sequence of the outgroup *S. lycopersicoides*. A) phylogenetic tree for all 30 genes, B) phylogenetic tree for the 14 reference genes, C) phylogenetic tree for the 16 candidate genes.

3.4.3 Footprints of selection

The amino acid changes in the consensus sequence data set were summarized (Tables 3.11, B4.1). These amino acid changes represent high frequency variants in the whole data set and can be regarded as candidates for local adaptation. In the whole data set 169 amino acid changes could be detected. This corresponds to 2.25 % of the total number of amino acid positions. More than half of these changes, namely 116, are in candidate genes and correspond to 3.06 % changed amino acid positions. The candidate genes with the highest number of amino acid changes are *TAS14* (8.55 %) and *NtC7* (7.58 %). The functional genes *pLC30-15*, *ER5*, *LTP*, *TSW12*, and *His1* have also a high percentage of amino acid changes (4 – 6 %). The candidate gene with the lowest amount of amino acid changes are *CT208* (0 %), *AREB2* (1.16 %), *DREB3* (1.21 %), and *TPP* (1.34 %). More than half of the amino acid changes (53.25 %) are present in more than one sequence.

Some of the shared amino acid changes are connected to the geographic groups (3.2.2). Five of the amino acid changes are present exclusively in the consensus sequences from the two northern populations, LA1930 and LA3784. Three of them are in the candidate gene *NtC7*. The other two amino acid changes are in *TPP* and *GBSSI*. Seven amino acid changes are only present in consensus sequences from coastal populations and are in the genes *NtC7*, *AREB1*, *JERF1*, *TAS14*, *His1*, *CT182*, and *CT251*. Four amino acid changes are only present in consensus sequences from populations from the high altitudes in the south and are in the genes *NtC7*, *JERF1*, *DREB3*, and *CT192*.

Table 3.11: Summary of the amino acid changes in the consensus sequence data set.

Gene	# a. a.	# a. a. changes	% a. a. changes	Number of amino acid changes			
				unique	northern group	coastal group	southern high altitude group
sensory gene							
NtC7	264	20	7.58 %	7	3	1	1
regulatory genes							
AREB1	430	10	2.33 %	6	0	1	0
AREB2	346	4	1.16 %	3	0	0	0
JERF1	358	10	2.79 %	5	0	1	1
JERF3	304	10	3.29 %	5	0	0	0
DREB3	248	3	1.21 %	1	0	0	1
functional genes							
dhn1	163	4	2.45 %	2	0	0	0
pLC30-15	178	10	5.62 %	5	0	0	0
TAS14	152	13	8.55 %	11	0	0	0
ER5	138	6	4.35 %	5	0	1	0
le25	74	2	2.70 %	0	0	0	0
LTP	102	5	4.90 %	1	0	0	0
TSW12	107	5	4.67 %	0	0	0	0
CT208	366	0	0.00 %	0	0	0	0
His1	185	9	4.86 %	5	0	1	0
TPP	372	5	1.34 %	2	1	0	0
total	3787	116	3.06 %	58	4	5	3
reference genes							
CT021	250	5	2.00 %	3	0	0	0
CT066	448	4	0.89 %	3	0	0	0
CT093	343	3	0.87 %	2	0	0	0
CT114	209	1	0.48 %	0	0	0	0
CT143	138	0	0.00 %	0	0	0	0
CT166	154	0	0.00 %	0	0	0	0
CT179	192	0	0.00 %	0	0	0	0
CT182	149	2	1.34 %	1	0	1	0
CT189	113	0	0.00 %	0	0	0	0
CT192	247	4	1.62 %	2	0	0	1
CT198	108	1	0.93 %	0	0	0	0
CT251	442	15	3.39 %	3	0	1	0
CT268	646	16	2.48 %	6	0	0	0
GBSSI	283	2	0.71 %	1	1	0	0
total	3722	53	1.42 %	21	1	2	1
total	7509	169	2.25 %	79	5	7	4

Note: “a. a.” amino acid

3.5 Differential behaviour after application of salt

86 young *S. chilense* plants were exposed to salinity stress and allowed to recover thereafter. After recovery 39 plants (45.35 %) showed no signs of stress, 32 plants (37.21 %) showed slight symptoms and 15 plants (17.44 %) showed severe symptoms of stress. The populations were different in their response, ranging from 100 % no signs of stress (LA1930, LA2932) to 100 % severe symptoms (LA4117A; Figure B5.1A). A generalized linear model revealed that the differences observed in this data set could be explained by altitude ($p > 0.001$). The altitudinal groups showed significant differences in their response (Figure 3.8A). The group from low altitudes (< 500 m) performed significantly better than the group from intermediate altitudes (501 – 2000 m) and the group from high altitudes (> 2001 m). The group from intermediate altitudes further performed significantly better than the group from high altitudes.

During recovery, several plants started to develop new side branches. The populations were also different in this characteristic, ranging from 100 % with new side branches (LA2750) to 100 % without new side branches (LA1930, LA1938, LA3111; Figure B5.1B). The low altitudinal group differed significantly for this characteristic from the intermediate group (Figure 3.8B). The high altitude group did not differ significantly from the two other groups.

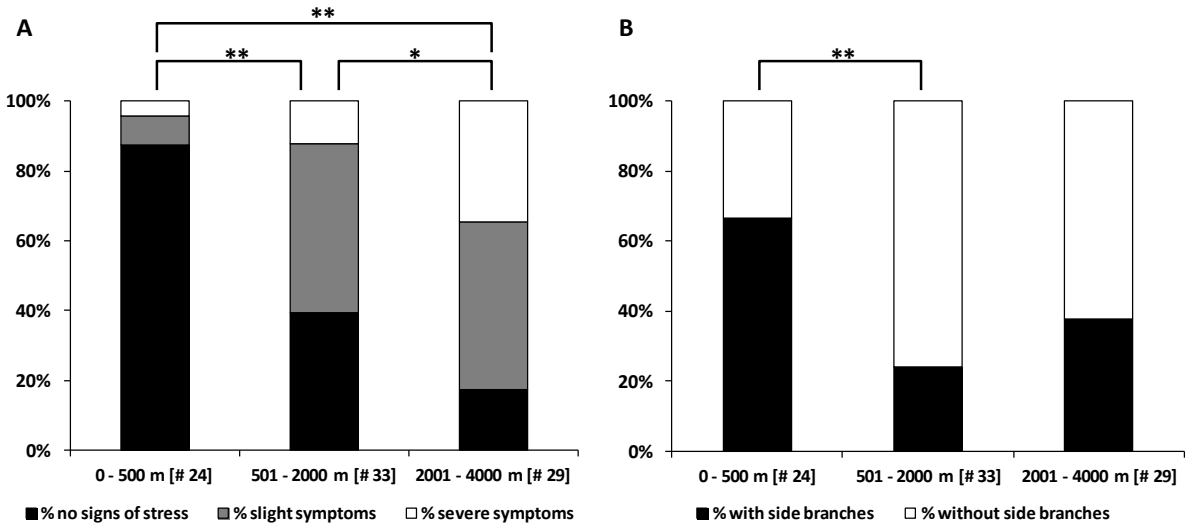

Figure 3.8: Results of group analysis of the young *S. chilense* plants after the salt treatment. Overall conditions (A) and development of side branches (B) after four weeks of recovery. Young *S. chilense* plants are grouped according to altitude. Number in brackets gives number of young *S. chilense* plants per group. Wilcoxon rank sum test was applied to test for differences between the groups (*: p-value < 0.05, **: p-value < 0.01).

CHAPTER 4: DISCUSSION

4.1 Evaluation of the data generation

30 genes were sequenced from the 23 *S. chilense* populations using a pooling approach. This approach was chosen since it allows for sequencing of a large number of individuals per population (25 in this case) in an efficient way. However, pooling approaches have also limitations. Low frequency variants may be lost during pooling. Single SNPs cannot be traced back to the individual level and linkage between SNPs (*i.e.* haplotypes) cannot be analysed. Nevertheless, pooling approaches were successfully applied to different types of studies including studies on selection and local adaptation (*e.g.* Obbard *et al.* 2009; Turner *et al.* 2010; Amaral *et al.* 2011; Zhou *et al.* 2011; Fischer *et al.* 2013b). However, the pooling approaches and procedures differed between the studies ranging from whole genomes (*e.g.* Turner *et al.* 2010) to PCR products (*e.g.* Obbard *et al.* 2009) and from pooling before the DNA extraction (*e.g.* Zhou *et al.* 2011) to pooling of DNA samples (*e.g.* Turner *et al.* 2010). In this study, the pooling started after the DNA extraction. This avoids possible biases caused by different DNA content in different leaves or uneven grinding of plant material. Furthermore, pre-pools were mixed for the PCR amplification and the final pools only afterwards. This way problems occurring during PCR amplification were more likely detected and could be taken into account. *AREB2*, for example, which could be amplified only from one individual of LA2750, was therefore excluded. If all 25 individuals were pooled before the PCR amplification, this would not have been noticed and would have led to a bias in the analyses.

Mapping and analyses of pooled sequence data are challenging, especially for non model organisms like wild tomatoes. The genome of a cultivated tomato, *S. lycopersicum* cv. Heinz, was published in 2012 (The Tomato Genome Sequencing Consortium 2012). Although sequencing of other cultivars and wild tomato genomes and transcriptomes is currently in process (*e.g.* http://solgenomics.net/organism/sol100/view), so far only little data is available. SNP calling is also an important issue. Stringent SNP calling conditions were applied in this study. A SNP was considered to be a true SNP *e.g.* only when it was at a position with a sequencing depth of > 2,000 and present in more than 1 % of the reads. This criterion is very stringent in comparison to other studies that set the thresholds for sequencing depth and minimum number of reads carrying the SNP much lower (*e.g.* Turner

et al. 2010; Mullen *et al.* 2012; Fischer *et al.* 2013b). Therefore it is more likely that the present study underestimates the genetic variation. This study also employs more analytical methods than many previous studies. Usually only SNP analysis, like number of SNPs per sample, and basic summary statistics were calculated (*e.g.* Amaral *et al.* 2011). Some studies extended their analyses to synonymous and nonsynonymous sites (*e.g.* Obbard *et al.* 2009). Basic summary statistics, including θ_W, π, Tajima's *D*, and pairwise F_{ST}, were also calculated in this study. In addition, several other methods were applied including BayeScan and the analyses of synonymous and nonsynonymous polymorphism and divergence.

To evaluate the repeatability of this approach, one population was sequenced twice. Comparisons between the two repetitions showed significant correlations. This indicates that the approach is repeatable.

4.2 Demographic history of *Solanum chilense*

4.2.1 North-south cline indicates migration from north to south

The 23 populations of *S. chilense* vary greatly in their degree of genetic variation (Table 3.1). The difference between the more polymorphic populations and the least polymorphic populations is twofold. Overall the genetic variation decreases from north to south (Figure 3.1). The populations with the highest genetic variation are the two populations from the north, LA1930 and LA3784, two populations from the central region, LA1963 and LA2765, and one population from the coast, LA2750. The populations with the lowest degree of genetic variation are the two most southern coastal populations LA4107 and LA4108. This north-south cline is indicated by all estimators of genetic variation. The data also shows an increase in the π_a/π_s ratio from north to south. This indicates that towards the south the proportion of nonsynonymous nucleotide diversity increases. An increase of nonsynonymous nucleotide diversity could be the result of bottlenecks during colonization events. Populations that undergo a bottleneck are initially small and strongly affected by genetic drift. Genetic drift could increase the frequency of nonsynonymous deleterious alleles. Interestingly, the increase in the π_a/π_s ratio is stronger in the coastal populations than in the high altitude populations in the south. This may suggest that the coastal populations underwent stronger bottlenecks than the high altitude populations in the south. Differences in genetic variation could be also explained by selection. The populations in the south could be more frequently subject to selection than the populations in the north.

This north-south cline is also present for the Tajima's *D* values (Table 3.1, Figure 3.1). Tajima's *D* values decrease from north to south. Overall the Tajima's *D* values are negative indicating that most populations are in the process of population expansion (after a bottleneck). This effect seems to be stronger in the south than in the north. The populations with the lowest Tajima's *D* values are the four coastal populations and LA4119 which is from the high altitude region in the south. Several high altitude populations from the central region have relatively high Tajima's *D* values. This may suggest that some populations in the mountains are less expanding possibly due to the geological and geographical constraints in the mountains. Soil and rocks could physically limit the plant populations in their expansion.

The genetic differentiation between the *S. chilense* populations and three other wild tomato species, *S. peruvianum*, *S. arcanum*, *S. habrochaites*, also follows this north-south cline (Figure 3.5). The *S. chilense* populations in the north are genetically closer to the three

other wild tomato species than the southern *S. chilense* populations. However, this north-south cline is not present for the divergence from the two outgroup species, *S. ochranthum* and *S. lycopersicoides*, and from the cultivated tomato *S. lycopersicum*. The divergence is almost constant for all *S. chilense* populations (Figure 3.5). This may indicate that the north-south cline is a recent within species characteristic and therefore affects only relationships with closely related species. The cultivated tomato falls within the wild tomato clade, but is more distant than *S. peruvianum*, *S. arcanum*, and *S. habrochaites* (Peralta & Spooner 2001; Sarkinen *et al.* 2013). A slight trend for lower divergence from *S. ochranthum*, *S. lycopersicoides*, and *S. lycopersicum* in the south is present. This could be due to the lower genetic variation of the populations in the south. A positive correlation between genetic diversity and divergence has been reported for wild tomatoes (Roselius *et al.* 2005). Overall this north-south cline is in agreement with the hypothesis that *S. chilense* is derived from *S. peruvianum* or a *S. peruvianum*-like ancestor (Baudry *et al.* 2001). *S. chilense* would then have likely originated in the northern range of its current distribution and migrated to the south.

Seed banks were shown to contribute to the observed genetic variation in *S. chilense* (Tellier *et al.* 2011b). Lower genetic variation in the southern range of the species could indicate that seed banks are shorter in the south. One explanation for this is that the populations in the south are younger because either the southern range was colonized later or generation times are longer in the south or both. Longer generation times could be due to the fact that the southern range of the species distribution is characterized by more extreme environmental conditions than the northern range. The coastal region of Chile is greatly affected by the El Niño's southern oscillations, which provide heavy rainfall, and some *Solanum* populations were reported to be present only during El Niño years (Chetelat *et al.* 2009). The El Niño strongly affects plant growth, reproduction, and seed banks in coastal Chile (reviewed in Holmgren *et al.* 2001; Gutierrez & Meserve 2003). If seed germination and/or reproduction was linked to the El Niño, then generation times would be longer for populations affected by the El Niño. However, mainly coastal populations should be affected by the El Niño and not the high altitude populations (Houston 2006; Chetelat *et al.* 2009). The southern range has populations in both habitats, therefore elongated generation times due to the El Niño are unlikely to happen in both groups. Another explanation is adaptation to shorter seed banks. Overall seeds tend to be bigger in the south than in the north (K. Böndel, A. Tellier, personal observation). This may suggest that bigger seeds are advantageous in more extreme environments. This has been shown in several plant species. Seed size has a positive effect on drought tolerance in pearl millet (Manga & Yadav 1995). A

study on related alpine species pairs showed that the species from higher altitudes, and thus more extreme conditions, have larger seeds (Pluess *et al.* 2005). If the seeds are bigger, the number of seeds is lower under the assumption that fruit size does not change. Therefore, adaptation for bigger seeds and/or shorter seed banks could also explain this pattern. However, the populations in the southern range are either from the coast or from high altitudes. These two environments are both extreme, but different. Therefore, it is unlikely that bigger seeds are advantageous in both.

4.2.2 Isolation by distance and genetic differentiation define population groups

In this study a significant pattern of isolation by distance was observed for the 23 *S. chilense* populations (Figure 3.4). A pattern of isolation by distance was also reported for other wild tomato species including *S. peruvianum*, *S. pimpinellifolium* and *S. lycopersicu*m (Nakazato & Housworth 2011; Nakazato *et al.* 2012) as well as in the related Solanaceae species *S. lycopersicoides* and *S. sitiens* (Albrecht *et al.* 2010), which occur in sympatry with *S. chilense* in northern Chile (Peralta *et al.* 2008; Chetelat *et al.* 2009). This indicates that isolation by distance may be the predominant form in wild tomatoes and in related Solanaceae species.

The analyses of genetic differentiation between the *S. chilense* populations further revealed that the populations of this study can be clustered in one central group and three peripheral groups. The central group comprises the populations from the central region of the species distribution (app. 16° 30' S - 21° 30' S; Figure 2.1). Two of the peripheral groups are in the south of the central group. One includes the four populations from the high altitude region around Calama, San Pedro de Atacama and the nearby Salar de Atacama (LA4332, LA4118, LA4119, LA2880). This group is also morphologically distinctive, *e.g.* the leaf segments are broader (Chetelat *et al.* 2009, M. Mboup, K. Böndel, T. Nosenko, personal observation). However, it has to be noted that LA4332 is genetically close to the other three of this group and also to many populations from the central group. This suggests that LA4332 is a 'link' between the southern high altitude group and the central group. The same conclusion was reached by the TGRC collectors (Chetelat *et al.* 2009). The second peripheral group in the south includes the four populations from the coast near Tocopilla (LA2750, LA2932) and Taltal (LA4107, LA4108). The third group comprises only two populations, LA1930 and LA3784, located north of the central group. This group most likely coincides with the origin of the species. LA1930 and LA3784 have high genetic variation, are

genetically close to the other wild tomato species, *S. peruvianum*, *S. arcanum* and *S. habrochaites*, and occur in sympatry with the sister species *S. peruvanum*.

On average the within group F_{ST} values are lower than the between group F_{ST} values (Figure 3.3, Table B2.2). This indicates that gene flow between populations from the same group is higher than between populations from different groups. This is also supported by the finding that isolation by distance is present for the whole sample and absent from the central group (Figure 3.4). However, the four populations of the coastal group most likely form two sup-groups, because F_{ST} values between Tocopilla (LA2750, LA2932) and Taltal (LA4107, LA4108) are in the range of the observed between group F_{ST} values. Also the two populations from the northern group have an F_{ST} value in the range of the between group F_{ST} values. Although this sample is a good representative of the whole species, it does not include enough populations from these peripheral groups to allow further conclusions about their internal structure. The north-south cline suggests that colonization occurred from north to south (4.2.1). Therefore, it is likely that three bottlenecks happened during the colonization: 1) when the central group derived from the northern group, 2) when the coastal group derived from the central group, and 3) when the southern high altitude group derived from the central group. An interesting aspect of these groups is also that the genetic differentiation between the two peripheral groups in the south is similar to the genetic differentiation between these two groups and the northern group. This is surprising since the two peripheral groups are geographically much closer. This finding could be explained by the climatic conditions in Chile. The coastal populations receive most rainfall during the winter season while the populations at the high altitudes receive most rainfall during the summer season (Houston 2006). This could lead to a temporal barrier to gene flow.

The phylogenetic analysis confirms the hypothesis of a north-south colonization (Figure 3.7). This analysis shows that the two peripheral groups in the south are derived from the central group. They form highly supported monophyletic clades nested within the central populations. However, the two peripheral groups in the south are not sister clades. This indicates that they arose independently from the central group. Interestingly, the two populations from the northern group (LA1930, LA3784) form a sister clade to the rest of the *S. chilense* populations. In the candidate gene tree the sequence of the cultivated tomato is even closer to the central *S. chilense* clade than the two northern populations. This could indicate that the northern *S. chilense* populations are either an early diverged group of *S. chilense* and/or that they are partly admixed with other wild tomato species, *i.e.* that interspecies gene flow is occurring. The latter is supported by the interspecies F_{ST} values which showed that the northern populations are the closest to *S. peruvianum*,

S. habrochaites and *S. arcanum* (Figure 3.5). The split between *S. chilense* and *S. peruvianum* is recent (≤ 0.55 mya Stadler *et al.* 2008; 0.73 (or 5.1) mya Naduvilezhath *et al.* 2011; 0.7 (or 4.6) mya Mathew *et al.* 2013; 0.74 mya Sarkinen *et al.* 2013). Therefore, it is also possible that speciation is not complete yet and that gene flow is ongoing in regions where both species co-occur. Speciation with gene flow has actually been reported for the species pair *S. peruvianum* and *S. chilense* (Stadler *et al.* 2005; Stadler *et al.* 2008).

Further support for gene flow between the northern *S. chilense* populations and other wild tomato species, *e.g. S. peruvianum*, comes from previous population genetic studies of *S. chilense* and *S. peruvianum* that included *S. chilense* samples from this northern region. A haplotypic pattern was reported for the *pLC30-15* gene in the Quicacha population, which is from this northern area (Xia *et al.* 2010). It was hypothesized that the intron of one of these haplotypes could have originated from a population of *S. peruvianum*. A trans-species polymorphism at the *CBF2* locus is shared between the Quicacha population and a *S. peruvianum* population (Mboup *et al.* 2012). Adaptive introgression of an allele of the *Pto* gene was found between an individual of LA1930 (*S. chilense*) and *S. peruvianum* (Böndel 2010; Hörger 2011; Hörger *et al.* in preparation). Alternatively, incomplete lineage sorting could also cause this pattern. Ancestral alleles, that were lost during the first bottleneck in the majority of the *S. chilense* populations, could still be present in the northern populations. Seed banks could further explain the observed pattern. If seed banks were larger in the north, for example due to different soil conditions, seeds from pre-speciation times that carry ancestral alleles could be present and germinate. However, this study does not allow deriving any definite conclusions about the northern *S. chilense* populations. This study did only detect that the northern *S. chilense* populations are different and require further investigation. A *S. peruvianum* sample of similar size as the *S. chilense* sample of this study could help to further investigate the status of the northern *S. chilense* populations.

4.3 Local adaptation in *Solanum chilense*

The comparison of genes involved in the abiotic stress response (*i.e.* candidate genes) with reference genes showed differences in their mode of evolution.

4.3.1 Gene evolution on the species level

Overall the candidate genes have more genetic variation than the reference genes (Table 3.2). Especially the nonsynonymous variation is elevated. This indicates that the candidate genes are less conserved than the reference genes and that they possibly maintain more adaptive variation. Furthermore, the more negative Tajima's *D* values observed for the candidate genes may indicate that this excess is mainly at low frequency. This would be in agreement with the observation of an excess of nonsynonymous variation at adaptive genes in *Arabidopsis* (Hancock *et al.* 2011). If the environmental conditions change, the standing genetic variation could become advantageous (reviewed in Barrett & Schluter 2008; Innan & Kim 2008). The candidate genes have also on average higher divergence compared to the two outgroup species, *S. ochranthum* and *S. lycopersicoides*, and to the cultivated tomato, *S. lycopersicum*, than the reference genes (Tables B3.11 – B3.28). This further confirms that the candidate genes have more potential for local adaptation and that they are more likely under positive or balancing selection.

The 16 candidate genes come from different pathways and from different levels of the abiotic stress response (Figure 4.1). Overall the functional genes have higher nonsynonymous genetic variation, higher divergence and more negative Tajima's *D* values than the regulatory genes (Tables 3.3, 3.4) and also more significant McDonald-Kreitman tests. This suggests that standing genetic variation is higher in the functional genes and therefore also the potential for adaptation (reviewed in Barrett & Schluter 2008). The functional genes are located rather downstream in their corresponding abiotic stress response pathways. This is consistent with analyses of the anthocyanin pathway that showed that upstream genes evolve under stronger constraint than downstream genes (Rausher *et al.* 1999; Rausher *et al.* 2008). This pattern can be explained with a higher connectivity of the upstream genes in the gene network (Rausher *et al.* 1999; Rausher *et al.* 2008) and genes with higher connectivity should evolve under stronger constraint (reviewed in Olson-Manning *et al.* 2012).

Despite this overall trend between regulatory and functional genes, the candidate genes show variable evolutionary histories (Tables 3.3, 3.4). Differences are mainly between

genes from the same gene family. The AREB/ABF transcription factor *AREB1* has more genetic variation, higher divergence and a higher Tajima's *D* value than *AREB2*. A similar pattern can be observed between the ERF protein encoding genes *JERF1* and *JERF3*. This indicates that *AREB2* and *JERF3* are under stronger constraint than *AREB1* and *JERF1*. The three dehydrin genes, *dhn1*, *pLC30-15*, and *TAS14*, have comparable synonymous nucleotide diversity, but differ greatly in their nonsynonymous nucleotide diversity (Figure 3.6) and also in their Tajima's *D* values. This suggests that *dhn1* evolves under purifying selection and *TAS14* rather under balancing selection, while *pLC30-15* is intermediate between them. The two lipid transfer protein encoding genes do also show different patterns. Different evolutionary patterns for genes from the same gene family are not unexpected. After gene duplication the copies may acquire different functions and therefore evolve under different selection regimes (reviewed in Flagel & Wendel 2009; Magadum *et al.* 2013). Members of the same gene family are often observed to evolve under different selection types in wild tomatoes (Fischer *et al.* 2011; Mboup *et al.* 2012).

The functional genes *le25* and *ER5* also belong to the *lea* genes, but fall into different functional groups than the dehydrins. Both genes are characterized by high genetic variation and very negative Tajima's *D* values. The gene *le25* is further characterized by a π_a/π_s ratio of above 1 (Figure 3.6). The transcription factor *DREB3* is also characterized by high genetic variation and a low Tajima's *D* value. This may suggest that the standing genetic variation is especially high in these genes. The overall low genetic variation of the alcohol dehydrogenase *CT208* is in agreement with the previously reported selective sweep scenario (Arunyawat *et al.* 2007). The *TPP* gene has low nonsynonymous nucleotide diversity, but high synonymous nucleotide diversity and a relative high Tajima's *D* value. This may imply a trend for balancing selection. *His1* is close to the π_a/π_s ratio of 1, but otherwise average. This may indicate relaxed constraint.

This data set includes also one sensory gene *NtC7*. The high level of polymorphism and the mean positive Tajima's *D* value indicate that *NtC7* evolves under balancing selection. Since *NtC7* is the only sensory gene in this data set, it cannot be concluded whether this pattern is common for sensory genes or not.

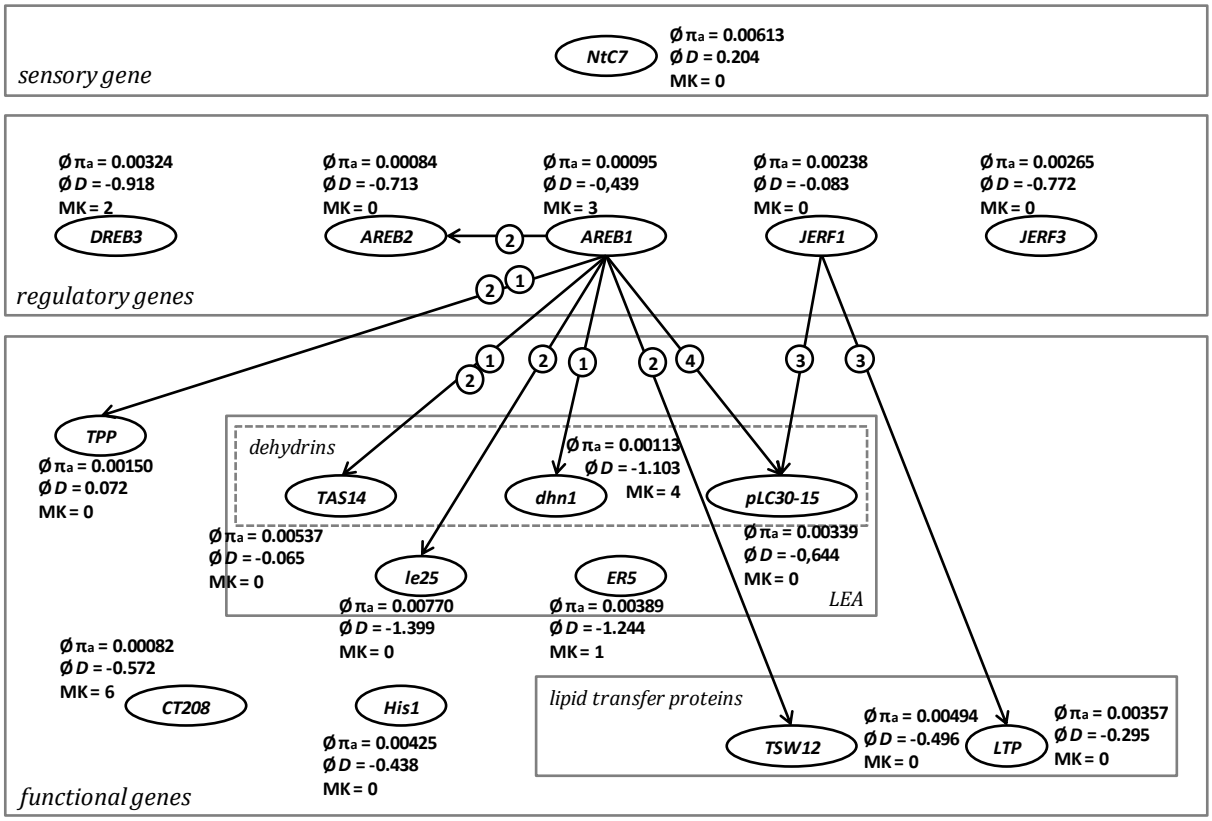

Figure 4.1: Candidate gene network with mean nonsynonymous nucleotide diversity, mean Tajima's *D*, and number of significant McDonald-Kreitman test statistics. Mean nonsynonymous nucleotide diversity (π_a), mean Tajima's *D* (*D*) and number of populations with a significant McDonald-Kreitman test (Fisher's exact test, p-value < 0.05) with both outgroup species, *S. ochranthum* and *S. lycopersicoides* (MK) are given for each gene. For details about the gene classifications and arrows see Figure 2.2.

4.3.2 Different methods to detect outliers in the data set

Several methods were applied to detect outliers in the data set. The McDonald-Kreitman test compares synonymous and nonsynonymous polymorphism within a population to synonymous and nonsynonymous divergence compared to an outgroup species (McDonald & Kreitman 1991). In this data set, 22 genes are significant in at least one population and with at least one outgroup species. Furthermore, every population has at least one gene that is significant with at least one outgroup species (Tables B3.29 - B3.32). The McDonald-Kreitman test is significant due to an excess of nonsynonymous polymorphism for almost all genes. The exception is the reference gene *CT268*. Comparisons of the results for the two outgroup species revealed that only 14 genes are significant in the

same population with both outgroup species (Table 3.5). The majority of these genes are significant in coastal populations.

The proportion of adaptive amino acid changes was estimated based on the π_a/π_s and K_a/K_s ratios (Tables B3.33 – B3.36). Like for the McDonald-Kreitman test, the overlap between the two outgroup species is rather low. On average the outgroup species *S. ochranthum* and *S. lycopersicoides* have similar divergence from the *S. chilense* populations. The observed discrepancies between the two outgroup species suggest that the two outgroup species themselves are influenced by selection or some demographic factor. *S. lycopersicoides* occurs also in extreme environments in northern Chile and is exposed to the same abiotic stresses than *S. chilense* (Peralta *et al.* 2008). Therefore, it is possible that selection acts upon some of the abiotic stress-related genes of this data set. *S. lycopersicoides* also exhibits a patchy distribution and it is hypothesized that the species may be in the process of shrinkage (Chetelat *et al.* 2009). *S. ochranthum* occurs in more mesic environments in Peru (Peralta *et al.* 2008). Therefore, environmental stresses should have less influence. However, neither demography nor selection can be excluded as possible sources of any bias. Overall these observations suggest that several sequences from different outgroup species should be used for population genetic analyses to avoid any outgroup specific effect. A comparison between the McDonald-Kreitman test results and the α values showed almost no overlap between significant McDonald-Kreitman tests and positive α value. This is not unexpected since the McDonald-Kreitman tests were mainly significant due to an excess of nonsynonymous polymorphism and not due to an excess of nonsynonymous divergence. An excess of nonsynonymous polymorphism should lead to negative α values. The majority of genes with a significant McDonald-Kreitman test are found in the coastal populations. This is consistent with the observation that the coastal populations have the lowest number of genes with a positive α value. The only gene with significant McDonald-Kreitman tests due to an excess of nonsynonymous divergence, *CT268*, has also the best overlap with positive α values.

The program BayeScan identifies outlier SNPs based on their frequencies. BayeScan identified 39 SNPs that are most likely under positive selection and 29 outlier SNPs that are most likely under balancing selection (Tables 3.6, B3.37). Overall the outlier SNPs under positive selection are either at high frequency in the northern or coastal group or at high frequency in all populations except for the northern or coastal group. This indicates that although BayeScan is relatively robust for demography, the underlying demographic structure influenced the outcome of this analysis. In fact it has been shown that populations that underwent a strong bottleneck could cause false positives (Foll & Gaggiotti 2008).

Therefore, this analysis provides further evidence for the demographic groups and their possible origin (4.2.2). The outlier SNPs for balancing selection seem to be less affected by demography. They are either present in only a few populations or in several populations that come from different regions of the species range. This indicates that these SNPs might be under balancing selection on the species level. The overlap between a) the BayeScan results and the McDonald-Kreitman tests and b) the BayeScan results and the proportion of adaptive amino acid substitutions is very low. The largest overlap is for the gene *AREB1*. BayeScan identified four outlier SNPs for positive selection, which are at high frequency in coastal populations. Three of the coastal populations have a significant McDonald-Kreitman test with both outgroup species. However, the four outlier SNPs are in intronic regions and the McDonald-Kreitman test is significant due to an excess of nonsynonymous polymorphism. BayeScan further identified an outlier SNP for balancing selection in *AREB1*, which is in intermediate frequency in five populations. These five populations have a positive α value with both outgroup species. Therefore, a clear selection scenario is not observed and demography could be responsible. Further analyses are required to understand if selection or demography is responsible for the observed pattern, especially for the one in the coastal populations. The best overlap is probably for the gene *CT268* in the population LA3784. This gene has a positive α value with both outgroup species, a significant McDonald-Kreitman test with *S. lycopersicoides* and BayeScan identified one outlier SNP for balancing selection. However, also in this case the selection scenario is not clear.

The two peripheral groups in the south show trends for isolation by distance for the average F_{ST} values over all genes and over the reference genes. Several genes, however, have almost no genetic differentiation between the populations (Tables 3.7, 3.8). This suggests that selection is acting upon these genes. The genes *AREB2*, *JERF3*, *DREB3*, *CT208*, *CT093*, and *CT182* have very low genetic differentiation in the southern high altitude group. Many southern high altitude populations have positive α values for these genes. This indicates that these genes are under positive selection in the southern high altitude group. BayeScan also identified two SNPs for positive selection in the gene *JERF3*. The genes *JERF3*, *dhn1*, *le25*, *CT208*, *CT166*, and *CT192* show almost no genetic differentiation in the coastal group. McDonald-Kreitman tests and BayeScan infer that *JERF3*, *dhn1*, *le25*, *CT208*, and *CT192* could evolve under balancing selection in this group. Many coastal populations have significant McDonald-Kreitman test statistics for these genes and BayeScan identified outlier SNPs under balancing selection in *JERF3* (1 SNP: LA2750, LA2932) and *CT192* (6 SNPs: LA4107, LA4108). *CT166* evolves rather under positive selection as suggested by one outlier SNP for

positive selection. However, overall the coastal populations show contrasting patterns that could also be explained by other factors than local adaptation (see 4.3.3). Two genes, *CT208* and *JERF3*, have almost no genetic differentiation in both groups. Between group F_{ST} values are also low for *CT208*. This is in agreement with previous results that inferred that *CT208* is under a species wide selective sweep (Arunyawat *et al.* 2007). Between group F_{ST} values for *JERF3* are high indicating that *JERF3* is involved in two different selection events. These findings imply that *JERF3* could be the most interesting gene of this data set.

The different approaches to detect selection in this data set showed only little overlap between the results. This could be explained by strong demographic effects like bottlenecks (*e.g.* BayeScan). The fact that also reference genes were detected implies that they are also influenced by selection. This is not surprising since they are functional genes. Some of them were shown to evolve under purifying selection (Tellier *et al.* 2011a). Nevertheless, previous studies showed that they are valuable as reference genes for demographic or selection analyses (*e.g.* Stadler *et al.* 2008; Fischer *et al.* 2011; Tellier *et al.* 2011b; Mboup *et al.* 2012).

4.3.3 Signatures of selection in the coastal populations

The populations from the coastal group show a characteristic pattern that distinguishes them from all other populations.

A trend for isolation by distance is present in the coastal group for the average F_{ST} values over all genes and over the reference genes. However, this pattern is not present for some genes (Table 3.7). The coastal populations show almost no genetic differentiation for the genes *CT208*, *dhn1*, *le25*, *JERF3*, *CT166, and CT192*. A selective sweep scenario was previously reported for the alcohol dehydrogenase *CT208* (Arunyawat *et al.* 2007), which explains the low genetic differentiation. The gene *CT166* is a reference gene that encodes a ferredoxin--NADP reductase. BayeScan identified one outlier SNP for positive selection in *CT166*. This SNP is in high frequency in the four coastal populations and also in the two northern populations. Since this SNP is intronic and also present in low frequency in several other populations, it possibly increased in frequency by chance during the bottleneck that led to the coastal group. *CT192* is also a reference gene and encodes a ribosomal protein S3 kinase. BayeScan identified six outlier SNPs for balancing selection which are all present in the coastal populations LA4107 and LA4108 and also in populations from the other three groups. High Tajima's *D* values also suggest balancing selection in LA4107 and LA4108. *CT192* is linked to a disease resistance protein encoding gene (Solyc04g015210) in the tomato genome (http://solgenomics.net/). Resistance genes evolve frequently under

balancing selection (Delph & Kelly 2014). Therefore, it is possible that this disease resistance gene is the actual target of balancing selection. Remarkably, the other genes were reported to be involved in the salt stress response. *JERF3* is induced by salt stress (Wang *et al.* 2004). The expression of *dhn1* and *le25* is induced by the transcription factor *AREB1* and *AREB1* is induced by salt stress (Yanez *et al.* 2009; Orellana *et al.* 2010). Interestingly, *AREB1* shows a pattern of isolation by distance in the coastal populations. However, BayeScan identified four outlier SNPs for positive selection in *AREB1* and all of them are either in high frequency or fixed in the coastal populations (Table 3.6). A positive selection scenario for *AREB1* in the coastal populations is further supported by very negative Tajima's *D* values in comparison to the reference genes. The Tajima's *D* values of *dhn1* and *le25* are also more negative than in the reference genes. However, for both genes purifying rather than positive selection seems to be responsible for this pattern. Overall, *dhn1* has low variation and divergence and *le25* has an excess of low frequency variants. *JERF3* is also characterized by more negative Tajima's *D* values than the reference genes. In this case however, these values are due to an excess of high frequency variants indicating positive selection. This is also supported by high divergence in comparison to the outgroup species. However, BayeScan identified one outlier SNP for balancing selection which is in intermediate frequency in two coastal populations. A salt stress experiment revealed that plants from altitudes of below 500 m perform better under salt stress than plants from higher altitudes (Figure 3.8). These findings suggest that the coastal populations are adapted to cope with salt stress and that local adaptation to the coastal environment is likely. If the abiotic stress-related genes of this data set are responsible for the local adaptation at the coast or if other genes are involved, remains to be illuminated.

The coastal populations are further characterized by a high number of genes with significant McDonald-Kreitman tests with either both or one of the outgroup species (Figure 4.2). LA2750 has even 18 genes with a significant McDonald-Kreitman test. The McDonald-Kreitman tests are significant due to an excess of nonsynonymous polymorphism, which is classically interpreted as a balancing selection signature. However, the McDonald-Kreitman test does not take frequency into account. Given the overall negative Tajima's *D* values, balancing selection scenarios in these genes in these populations are an unlikely explanation. It rather indicates that the nonsynonymous polymorphisms are in either low or high frequency. The latter can be ruled out for most of the genes, since the divergence in comparison to the other species is not elevated in the coastal populations. The coastal populations have further elevated π_a/π_s ratios (Figure 4.2). These findings suggest that the coastal populations accumulate nonsynonymous polymorphism and that selection is not

eliminating them. In small populations genetic drift can counteract selection and thus, maintain nonsynonymous and/or deleterious variation. Three of the coastal populations have low genetic variation and therefore a rather small effective population size. This favours genetic drift over selection. The fourth coastal population, LA2750, however, is characterized by high genetic variation. Genetic drift is therefore unlikely to explain the excess of nonsynonymous variation in LA2750. An excess of nonsynonymous variation can also indicate relaxed constraint or a high degree of standing genetic variation. Since LA2750 has 18 genes out of 29 with a significant McDonald-Kreitman test, a genome wide effect would be required to explain this pattern.

Another explanation is partial selfing. *S. chilense* is reported to be a self-incompatible wild tomato species (*e.g.* Moyle 2008). However, partial selfing in populations from the marginal ranges of their species distribution is common in self-incompatible wild tomato species including *S. habrochaites* (Rick *et al.* 1979), *S. pennellii* (Rick & Tanksley 1981) and also *S. peruvianum* (Graham *et al.* 2003), the sister species of *S. chilense*. Since the coastal populations are geographically isolated from the central populations of *S. chilense* and probably underwent a bottleneck during the colonization of the coast, the occurrence of partial selfing can be expected. Selfing can ensure reproduction after a bottleneck when other plants and/or pollinators are not available (reviewed in Zuellig *et al.* 2014). This has been shown for the species *Leavenworthia alabamica* (Busch *et al.* 2011). A reduction of genetic variation, especially π, and an excess of nonsynonymous variation is expected for a selfing species/population in respect to an outcrossing species/population (reviewed in Glemin *et al.* 2006; Zuellig *et al.* 2014). This has been reported for sister species of the genus *Capsella* (Slotte *et al.* 2013) and also for sister species of the genus *Collinsia* (Hazzouri *et al.* 2013). If selfing is present in the coastal populations and to what extent, cannot be concluded from the genetic data. Only studying the populations under controlled conditions or analysis of the S-locus will provide evidence. Selfing could have also accelerated local adaptation. After the initial colonization of the coastal regions populations were rather small and genetic drift was strong. Selfing could have counteracted genetic drift and thus, helped to maintain possible advantageous genetic variation.

Figure 4.2: Map with the *S. chilense* populations, mean π_a/π_s ratio and significant McDonald-Kreitman test statistics. The mean π_a/π_s ratio of the 30 genes (first number in brackets) and the number of genes with a significant McDonald-Kreitman test (Fisher's exact test, p-value < 0.05) with both outgroup species (second number in brackets) is given for each population.

4.3.4 Adaptation in high altitude populations

The candidate genes evolve under different selection regimes in the high altitude populations. The genes show signatures of positive, purifying and also balancing selection. The high altitude habitats of the central group are characterized by low temperatures and the populations of the southern high altitude group are confronted with low temperatures and also with drought conditions (Table A1.1). Therefore, local adaptation to these stresses is expected.

Some genes evolve under positive or purifying selection in high altitude populations. The transcription factor *AREB2* seems to evolve under purifying selection in several high altitude populations including LA3111, LA2755, LA4332, LA4118, LA4119, and LA2880. These populations are characterized by very negative Tajima's *D* values and/or by low genetic variation and divergence. Another transcription factor, *JERF3*, evolves under positive selection in the southern high altitude group. This is supported by negative Tajima's *D* values, low genetic differentiation and two outlier SNPs for positive selection identified by BayeScan. *JERF3* is induced by cold stress (Wang *et al.* 2004). This suggests that the positive selection in the high altitude populations could be associated with adaptation to low temperatures.

The dehydrin gene *pLC30-15* evolves under positive selection in the high altitude population LA3111. This is supported by a very negative Tajima's *D* value, by high nonsynonymous divergence compared to the two outgroup species, *S. ochranthum* and *S. lycopersicoides*, and to the cultivated tomato, *S. lycopersicum*. BayeScan further identified two outlier SNPs for positive selection. Previously *pLC30-15* was reported to evolve under diversifying selection in the Quicacha population (Xia *et al.* 2010). Quicacha is close to the population LA3784 of this study. The two outlier SNPs are also present in high frequency in LA3784 indicating positive selection also for this population. This is further supported by the Tajima's *D* value and the divergence in comparison to the three other species. Also two other populations of this region, LA0456 and LA1958, have negative Tajima's *D* values. This suggests that *pLC30-15* is under selection in the northern range of the *S. chilense* species distribution. Since the strongest signature is observed in LA3111, which is occupying one of the coldest habitats, this could imply that *pLC30-15* is involved in adaptation to low temperatures. *pLC30-15* was recently also reported to be upregulated under low temperature stress (Steige 2011; Fischer *et al.* 2013a).

The *lea* gene *le25* shows negative Tajima's *D* values and high divergence in several of the high altitude populations including LA2765, LA3111, and LA4332. Tomato *le25* was

shown to enhance freezing tolerance in yeast (Imai *et al.* 1996). Therefore, a role of *le25* in cold adaptation in the high altitude populations is likely.

Several genes evolve under balancing selection in high altitude populations. The transcription factors *AREB1* and *AREB2* show signatures of balancing selection in some of the high altitude populations. The populations LA2880, LA3111, and LA4118 have elevated Tajima's *D* values for *AREB1*. LA2880 and LA4118 have further elevated π values. LA2773, LA2931, and LA2765 have elevated Tajima's *D* values for *AREB2*. Also other transcription factors with high Tajima's *D* and/or elevated π values are present in the high altitude populations: *JERF1* in LA2773 and LA2880 and *JERF3* in LA2755 and LA3111. LA2755 and LA3111 have also a significant McDonald-Kreitman test for *JERF3* with *S. ochranthum*. Two outlier SNPs for balancing selection were detected in *JERF3*. Both of them are in intermediate frequency in LA2755 and one of them also in LA3111. *JERF1* and *JERF3* have further elevated levels of nonsynonymous nucleotide diversity in these populations.

One of the dehydrins, *TAS14*, has also elevated π, π_a and high Tajima's *D* values in the southern high altitude populations. The McDonald-Kreitman test for LA4332 with *S. lycopersicoides* is also significant. Interestingly, *TAS14* is induced by the transcription factor *AREB1* (Yanez *et al.* 2009; Orellana *et al.* 2010) and both seem to be under balancing selection in LA2880 and LA4118. A similar scenario is observed in the neighbouring populations LA2747 and LA2773. Although LA2773 is from higher altitudes than LA2747, they have a low genetic differentiation. Both populations show signatures of balancing selection for the transcription factor *JERF1* and the functional gene *le25*. This is for both genes due to an excess of π, π_a and Tajima's *D* values relative to the reference genes. Unlike for *AREB1* and *TAS14* a direct interaction between *JERF1* and *le25* has not been reported yet. However, both genes are involved in the stress response of the same abiotic stresses, namely salt and cold (Imai *et al.* 1996; Zhang *et al.* 2004; Wu *et al.* 2007). Therefore, an interaction cannot be excluded.

Since balancing selection in abiotic stress-related genes is uncommon, this finding will be discussed in detail in the following part.

4.3.5 Signatures of balancing selection in abiotic stress genes

Stress responsive genes are expected to evolve under different types of selection depending on the nature of the stress. Balancing selection is primarily expected for genes that are involved in the biotic stress response like resistance genes in plants (reviewed in Delph & Kelly 2014). In wild tomatoes, several genes of the Pto/Prf pathway were shown to

evolve under balancing selection (Rose *et al.* 2007; Rose *et al.* 2011; Hörger *et al.* in preparation). A pattern of balancing selection at the species level is common for abiotic stress genes. Environmental heterogeneity and spatial variation lead to differences in allele frequency over the species range (reviewed in Delph & Kelly 2014). For example, different alleles might be advantageous at the south than at the north and therefore, when the whole species is sampled and analysed, balancing selection signatures are detected. Examples at the population level, however, are scarce (but see Mboup *et al.* 2012). Signatures of balancing selection were detected in some of the abiotic stress genes of this data set.

Two genes, *NtC7* and *TPP*, show balancing selection signatures in multiple populations. The sensory gene *NtC7* has high levels of nonsynonymous nucleotide diversity in all populations and the majority of populations have higher Tajima's *D* values than the mean of the reference genes. This indicates that *NtC7* evolves under balancing selection in *S. chilense*. BayeScan further identified one outlier SNP which is at intermediate frequency in ten populations. All of these populations are characterized by a Tajima's *D* value that exceeds the mean of the reference genes. Another interesting aspect about *NtC7* is that *S. chilense* and also the cultivated tomato are highly diverged from the recovered sequences of *S. ochranthum* and *S. lycopersicoides*. Therefore, this pattern might expand through all wild tomatoes. 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 (Tamura *et al.* 2003). Originally, *NtC7* was identified in a wounding experiment (Hara *et al.* 2000) and was also shown to be induced by osmotic stresses (Tamura *et al.* 2003). These findings could imply that *NtC7* plays a role not only in the abiotic stress response but also in the biotic stress response and therefore evolves under balancing selection. The second gene, the trehalose-6-phosphate phosphatase (*TPP*) has in half of the populations Tajima's *D* values that exceed the mean of the reference genes. However, nonsynonymous nucleotide diversity is relatively low for *TPP*. Since the populations with the balancing selection signatures come from all regions of the species distribution, the balancing selection signature for this gene is hard to connect to geographic or climatic features. An involvement of trehalose in the biotic stress response has been suggested (reviewed in Fernandez *et al.* 2010). Therefore, biotic stress may explain the balancing selection pattern of *TPP*. Another explanation could be that *TPP* is linked to a gene under balancing selection in the genome. Since especially the synonymous nucleotide diversity is high, a recent gene duplication and accidental amplification of both copies could also be responsible for this pattern.

Several genes show balancing selection signature in a few populations. As discussed before, several transcription factors show balancing selection signatures in high altitude

populations. Two *lea* genes are also under balancing selection in high altitude populations. Interestingly, *TAS14* is under balancing selection in the same populations as the transcription factor *AREB1*, which was reported to induce *TAS14* (Yanez *et al.* 2009; Orellana *et al.* 2010). *le25* and the transcription factor *JERF1* are under balancing selection in the neighbouring populations LA2747 and LA2773. In summary, this suggests that balancing selection for abiotic stress-related genes, especially transcription factors, is common in high altitude populations. High altitude regions are characterized by different climatic and therefore stress conditions. The *S. chilense* populations in the southern high altitude region are confronted with low temperatures, drought, and partly also with high salt concentrations in the soil (reviewed in Chetelat *et al.* 2009). The latter is due to their vicinity to the salt lakes in the Atacama Desert. Since most of these genes were reported to be involved in more than one abiotic stress response (Godoy *et al.* 1990; Kahn *et al.* 1993; Imai *et al.* 1996; Wang *et al.* 2004; Zhang *et al.* 2004; Yanez *et al.* 2009; Orellana *et al.* 2010), this could offer an explanation for the balancing selection signatures. Recently one gene of the *CBF* transcription factor family was reported to evolve under balancing selection in wild tomato species (Mboup *et al.* 2012). Expression analysis has revealed that one allele was induced by cold stress and another one by drought stress. Such a scenario is also possible for the transcription factors in this study.

Varying climatic conditions could explain the observed balancing selection signatures. But also an interaction of these genes with the biotic stress response cannot be excluded. Transcription factors could be involved in crosstalk between abiotic and biotic pathways (Fujita *et al.* 2006; Abuqamar *et al.* 2009). Furthermore, it is possible that the genes of this study are not the target of balancing selection, but that they are linked to genes under balancing selection in the tomato genome.

4.4 The utility of consensus sequences in molecular evolution

In this thesis, a new approach to analyse pooled sequence data was tested. This approach is the analyses of the consensus sequences. I anticipated that this approach could be of use for the analyses of pooled sequence data sets in the future. A consensus sequence corresponds to the average allele of the population and can therefore be considered as a good representation of the population.

The consensus sequence data set was analysed like a species wide sample (Stadler *et al.* 2009) to study gene evolution on the species level (Tables 3.9, 3.10). This analysis revealed that the candidate genes have more variation, especially nonsynonymous variation, higher divergence and more negative Tajima's *D* values than the reference genes. It also revealed that the functional genes have more variation, a higher divergence and more negative Tajima's *D* values than the regulatory genes. Both findings correspond to the results for the population means of the genes. Overall the genetic variation estimated for the consensus sequence data set and averaged over the populations correlate quite well (θ_W: $R^2 = 0.609$, $\rho = 0.787$, $p < 0.001$; π: $R^2 = 0.457$, $\rho = 0.752$, $p < 0.001$). However, there is some discrepancy in the Tajima's *D* values for some of the genes (Tajima's *D*: $R^2 = 0.216$, $\rho = 0.497$, $p = 0.006$). *NtC7*, for example, has a negative Tajima's *D* value in the consensus sequence data set while the average over the populations shows a positive Tajima's *D* value. This indicates that in most populations the within populations variation is at intermediate frequency - thus causing high Tajima's *D* values in the populations - but that this frequency is often below 50 % and the polymorphisms do not show up in the consensus sequences. *CT066*, for example, has a slightly negative Tajima's *D* for the average of the population data and a very positive Tajima's *D* for the consensus sequence data set. This could indicate spatial variation on the species level. Within the populations Tajima's *D* values are negative, either due to positive or purifying selection, but overall different mutations are at high frequency. This results in many shared SNPs in the consensus sequence data set and thus to a positive Tajima's *D* value.

The phylogenetic trees, which were constructed using concatenated alignments of the consensus sequences (Figures 3.7, B4.1), confirmed the results of the genetic differentiation analyses of the population data. The population data showed a north-south cline which supports the theory of a north-south colonization. The data further suggested that the 23 *S. chilense* populations fall into four groups: one central group and three peripheral groups. The consensus sequence trees showed that the two peripheral groups that are located south of the central group, form monophyletic clades nested within the clade of the central group.

The northern group, however, was sister clade to this central clade. Overall the consensus sequence trees support the north-south colonization and also the four groups.

Screening the consensus sequence alignments for nonsynonymous SNPs, *i.e.* amino acid changes, revealed that the candidate genes have more amino acid changes than the reference genes (Table 3.11). This is not surprising since nonsynonymous nucleotide diversity is also higher for the candidate genes. However, this screening was aimed to investigate in which sequences these amino acid changes are present and if these sequences represent populations that share any habitat characteristics. More than 50 % of the amino acid changes are present in more than one sequence. However, the majority of these shared amino acid changes are present in few to several populations from different environments or from the different groups. This may reflect standing genetic variation on the species level. Some amino acid changes are exclusive to either the coastal populations or to high altitude populations. These amino acid changes might represent nonsynonymous SNPs that increased in frequency because they were advantageous in the corresponding environment.

Taken together these observations suggest that the analysis of the consensus sequences is a valuable approach to screen large pooled sequence data sets. This way genes or regions that could be interesting in the context of local adaptation could be identified. Furthermore, consensus sequences could be of use to study gene evolution on the species level. This analysis is similar to a species wide sampling approach (Stadler *et al.* 2009) and focuses on the collecting phase of the coalescence (Pannell 2003). Finally, consensus sequences could also be useful for phylogenetic studies. Since consensus sequences represent average alleles, *i.e.* they show at each position what the majority of individuals have, they are fairly good representations of species. Usually only one allele is used. If one allele/individual is chosen randomly, the risk is high to choose one that is an outlier, a migrant or a nonfunctional allele. Several alleles would be better, but this is often not feasible. Therefore, pooled sequencing and using the consensus sequence provides a cost-effective and safe alternative. This strategy could also help to avoid problems with outgroup sequences, *e.g.* ancestral misidentification (Hernandez *et al.* 2007).

4.5 Conclusion and outlook

In this thesis I investigated the evolutionary history of the wild tomato species *S. chilense*.

I showed that demography plays an important role in the evolution of *S. chilense*. The genetic data presented here suggests that *S. chilense* most likely originated in the northern part of its current distribution and migrated towards south. During this north-south colonization, it went through at least three bottlenecks that resulted in four population groups within the sample: one central group and three peripheral groups. Deeper analyses of these groups may reveal further substructures within the groups.

I applied several methods to detect signatures of local adaptation. Although in general the overlap between the results obtained using the different methods is not large, signatures of local adaptation were detected in several populations and genes. Interestingly, most signatures were detected in the two peripheral groups in the south. This supports the theory that populations that recently colonized a new territory should show more likely signatures of local adaptation (Innan & Kim 2008) and contradicts the theory that this is more likely for large populations from the centre of a species distribution (*e.g.* Willi *et al.* 2006).

The populations from the coastal group show a distinct genetic pattern that may be explained by partial selfing. The occurrence of partial selfing is not uncommon in populations from marginal ranges in other wild tomato species (Rick *et al.* 1979; Rick & Tanksley 1981; Graham *et al.* 2003). Additional experiments are required to verify this hypothesis.

I also demonstrated an advantage of a new approach to analyse pooled sequence data. The analysis of the consensus sequence data mainly confirmed the results of the complete data set. Therefore, this approach could be of use for future studies.

This study may provide new avenues for crop production improvement. Local adaptation, possibly associated with environmental conditions, was detected in several candidate genes. These genes could serve as new material for crossing or transgenic experiments with the cultivated tomato.

This study revealed that demography is important in plant evolution and that the knowledge about the demographic history of a plant species is important to understand selection and adaptation events.

APPENDIX A: Material and methods

Table A1.1: Geographic and climatic information of the wild tomato populations.

Population	Province, Country	Latitude and Longitude	Altitude[a]	Temperature[b]	Precipitation[c]
Solanum chilense					
LA0456	Moquegua, Peru	17° 15' 0'' S, 71° 12' 0'' W	200	16.9	8
LA0458	Tacna, Peru	17° 58' 0'' S, 70° 11' 0'' W	800	17.5	18
LA1930	Arequipa, Peru	15° 17' 30'' S, 74° 36' 0'' W	500	19.6	6
LA1958	Moquegua, Peru	17° 15' 0'' S, 71° 15' 0'' W	1250	17.1	6
LA1963	Tacna, Peru	18° 4' 0'' S, 70° 19' 0'' W	200	18.2	11
LA1968	Tacna, Peru	17° 45' 42'' S, 70° 10' 35'' W	1680	15.8	18
LA2747	Arica and Parinacota, Chile	18° 35' 0'' S, 69° 54' 0'' W	800	16.5	9
LA2748	Tarapaca, Chile	21° 12' 48'' S, 69° 32' 52'' W	800	17.7	1
LA2750	Antofagasta, Chile	22° 4' 13'' S, 70° 9' 47'' W	300	18.8	2
LA2753	Tarapaca, Chile	19° 51' 23'' S, 69° 20' 14'' W	1650	14.1	17
LA2755	Tarapaca, Chile	19° 41' 3'' S, 69° 10' 52'' W	3200	7.3	103
LA2765	Arica and Parinacota, Chile	18° 46' 0'' S, 69° 41' 0'' W	2400	12.3	51
LA2773	Arica and Parinacota, Chile	18° 22' 0'' S, 69° 38' 0'' W	3400	10.3	142
LA2880	Antofagasta, Chile	23° 49' 0'' S, 68° 13' 0'' W	2500	13.9	36
LA2931	Tarapaca, Chile	20° 55' 0'' S, 69° 4' 0'' W	2275	12.3	18
LA2932	Antofagasta, Chile	22° 28' 5'' S, 70° 13' 30'' W	300-400	18.5	2
LA3111	Tacna, Peru	17° 28' 0'' S, 70° 2' 0'' W	3070	11.6	121
LA3784	Arequipa, Peru	15° 43' 52'' S, 73° 51' 2'' W	1100	17.3	18
LA4107	Antofagasta, Chile	25° 19' 8'' S, 70° 26' 46'' W	86	18.2	22
LA4108	Antofagasta, Chile	25° 3' 9'' S, 70° 28' 33'' W	80	18.4	18
LA4118	Antofagasta, Chile	23° 9' 27'' S, 68° 2' 6'' W	2440	13.5	48
LA4119	Antofagasta, Chile	23° 33' 14'' S, 67° 56' 2'' W	2980	11.4	40
LA4332	Antofagasta, Chile	22° 36' 32'' S, 68° 31' 19'' W	2968	10.5	29
Solanum ochranthum					
LA2682	Cusco, Peru	13° 38' 2'' S, 72° 14' 11'' W	2500	13.4	740
Solanum lycopersicoides					
LA2951	Tarapaca, Chile	19° 19' 0'' S, 69° 27' 0'' W	2200	13.0	30

Note: geographic data from the Tomato Genetics Resource Center, UC Davis (http://tgrc.usdavis.edu/), climatic data from WorldClim database (http://www.worldclim.org/, Hijmans *et al.* 2005)
[a]: altitude in meter above sea level
[b]: mean annual temperature in °C
[c]: annual precipitation in mm

Table A1.2: List of the 16 candidate genes. Locus name and location in the SL2.40 release of the *S. lycopersicum* genome (http://solgenomics.net/) and references are given.

Gene	SL2.40 Locus	SL2.40 Location	Description	Reference
sensory gene				
NtC7	Solyc03g083480	SL2.40ch03:46907988..46909219	receptor-like membrane protein	Tamura *et al.* 2003
regulatory genes				
AREB1	Solyc04g078840	SL2.40ch04:61097408..61094101	bZIP transcription factor	Yanez *et al.* 2009, Orellana *et al.* 2010
AREB2	Solyc11g044560	SL2.40ch11:35955757..35950455	bZIP transcription factor	Orellana *et al.* 2010
JERF1	Solyc06g063070	SL2.40ch06:36209812..36207061	ethylene responsive factor (ERF) protein	Zhang *et al.* 2004, Wu *et al.* 2007
JERF3	Solyc03g123500	SL2.40ch03:64386170..64388613	ethylene responsive factor (ERF) protein	Wang *et al.* 2004
DREB3	Solyc04g072900	SL2.40ch04:57461458..57462252	AP2/EREP transcription factor family	Islam & Wang 2009
functional genes				
dhn1	Solyc02g084840	SL2.40ch02:42499621..42498653	group 2 LEA (dehydrin)	Baudo *et al.* 1996, Yanez *et al.* 2009
pLC30-15	Solyc04g082200	SL2.40ch04:63552237..63550865	group 2 LEA (dehydrin)	Chen *et al.* 1993, Xia *et al.* 2010 as *NtERD10C*: Wu *et al.* 2007
TAS14	Solyc02g084850	SL2.40ch02:42504804..42503669	group 2 LEA (dehydrin)	Godoy *et al.* 1990, Godoy *et al.* 1994, Del Mar Parra *et al.* 1996, Yanez *et al.* 2009, Orellana *et al.* 2010 as *le4*: Cohen & Bray 1990, Cohen *et al.* 1991, Kahn *et al.* 1993
ER5	Solyc01g095140	SL2.40ch01:78283227..78284039	atypical hydrophobic group of LEA	Zegzouti *et al.* 1997, Yanez *et al.* 2009
le25	Solyc10g078770	SL2.40ch1059807276..59806503	group 4 LEA	Cohen & Bray 1990, Cohen *et al.* 1991, Cohen & Bray 1992, Kahn *et al.* 1993, Imai *et al.* 1996, Orellana *et al.* 2010
LTP	Solyc10g075100	SL2.40ch10:58135114..58135744	lipid transfer protein	Tapia *et al.* direct submission, GenBank as *NtLTP1* Buhot *et al.* 2004, Wu *et al.* 2007
TSW12	Solyc10g075110	SL2.40ch10:58157020..58157628	lipid transfer protein	Torres-Schumann *et al.* 1992, Orellana *et al.* 2010
CT208	Solyc09g064370	SL2.40ch09:57150099..57154745	alcohol dehydrogenase class III	Arunyawat *et al.* 2007
His1	Solyc02g084240	SL2.40ch02:41975280..41976442	H1 histone gene	Wei & O'Connell 1996 as *le20*: Kahn *et al.* 1993
TPP	Solyc03g083960	SL2.40ch03:47442646..47445417	trehalose-6-phosphate phosphatase	Yanez *et al.* 2009, Orellana *et al.* 2010

Table A1.3: List of the 14 reference genes. Locus name and location in the SL2.40 release of the *S. lycopersicum* genome (http://solgenomics.net/) are given.

Gene	SL2.40 Locus	SL2.40 Location	Description
CT021	Solyc06g035580	SL2.40ch06:21293012..21287092	choline dehydrogenase
CT066	Solyc10g054440	SL2.40ch10:50894596..50892473	arginine decarboxylase 1
CT093	Solyc05g010420	SL2.40ch05:4655463..4654381	S-adenosylmethionine decarboxylase
CT114	Solyc07g066600	SL2.40ch07:65208684..65205704	phosphoglycerate kinase
CT143	Solyc09g009040	SL2.40ch09:2385547..2393745	delta(14)-sterol reductase
CT166	Solyc02g083810	SL2.40ch02:41650004..41646719	ferredoxin--NADP reductase
CT179	Solyc03g120470	SL2.40ch03:62913100..62903860	aquaporin
CT182	Solyc11g011960	SL2.40ch11:4919067..4912805	UTP-glucose 1 phosphate uridylyltransferase
CT189	Solyc12g039120	SL2.40ch12:37884761..37883271	40S ribosomal protein S19-like
CT192	Solyc04g015130	SL2.40ch04:5295967..5292846	ribosomal protein S6 kinase alpha-3
CT198	Solyc09g082650	SL2.40ch09:63765012..63769755	acireductone dioxygenase; submergence induced protein 2-like
CT251	Solyc02g036370	SL2.40ch02:21311945..21309276	MYB transcription factor
CT268	Solyc01g007130	SL2.40ch01:1677070..1680707	LRR receptor-like serine/threonine-protein kinase FEI 2
GBSSI	Solyc08g083320	SL2.40ch08:62985075..62981703	granule-bound starch synthase

Table A1.4: PCR primers newly designed for this study.

Gene		Sequence 5' -> 3'	Binding position[a]	T_m[b]	Reference sequences[c]
candidate genes					
dhn1	5'	CATGGCACACTACGAGAACC	-1 - 19	63	contig2434591, SGN-E715270, SGN-E1239943
	3'	CTAGTGGTGTCCAGGGCC	614 - 631		contig2434591, a333613, SGN-E715270, SGN-E1239943
DREB3	5'	GGTTCCACTTCCACAGAGATG	-18 - 3	61	AF506825, contig1994590, SGN-E747086
	3'	CATGAGATGGATACTTCTGCAAC	750 - 772		AF506825, a4484, SGN-E747086
ER5	5'	GCAATGGCAGATTTCATGGAG	-3 - 18	63	U77719, contig77074
	3'	CAGTGTCTTCTTGTTTGTCACCG	539 - 561		
JERF1	5'	GGTGGTGCAATTATCTCCG	7 - 25	65	AY044235, contig4859496
	3'	GCAGCTTAGTAGGCACCTCC	2099 - 2118		AY044235, contig4859496, a11227
JERF3	5'	GTGGTGGTTCTATAATCTCCG	5 - 25	57	AY383630, contig5087938
	3'	CCATCATCAGCTACGGGG	1795 - 1812		AY383630, contig671788
His1	5'	GAGAAGATGACGGCAATCG	-6 - 13	58	Z11842, U01890, contig6478302, a629
	3'	GCCCTTTTAGCTGCTGGAG	718 - 736		Z11842, U01890, contig6478302, a629, AF253416
le25	5'	GCAGACAGGAAAGGACGC	3 - 20	64	M76552, contig6495420, a78300
	3'	TAGAAAGTTGTATGATTGCCAG	752 - 773		
LTP	5'	TGGTTAACAAGATTGCATGC	8 - 27	63	DQ073079, contig6454490
	3'	GATCAGCTTTACTGAACTCTGC	618 - 639		
NtC7	5'	GCCCAGACTACTTCTCAATGTG	67 - 88	69	contig6552527, SGN-E1244697
	3'	GGATCATCGTTCGATGTGTTC	979 - 999		contig6552527, SGN-E1244697, a298013
TAS14	5'	CAAAGATGGCACAATACGGC	-5 - 15	61	U26423, X51904, contig333989
	3'	CAAAGGTGTTCAATGCATCCC	600 - 620		U26423, X51904, contig6679983, a11908
TPP	5'	CTGGGCACAATGGACCTG	-9 - 9	63	contig6570537, SGN-E1244732, a23269
	3'	GGCTACAACTTGACTTCTTCC	2123 - 2143		
TSW12	5'	GAACAATATGGAAATGGTTAGC	-7 - 15	60	X56040, contig303575
	3'	GATCAACTTTACTGAACCCTGC	596 - 617		
reference genes					
CT021	5'	TTTCTCCGTCATCTCCCTCG	690 - 709	67	AK327423, contig 3808320, a56356
	3'	GTGCGTAACCAATCCAACTCC	2383 - 2403		
CT114	5'	GGCGGTGAAGAAGAGTGTCGG	3 - 23	68	contig2771685
	3'	CCTCCACCGAGCAATAACACGTC	1611 - 1633		AY941647, contig2771685
CT182	5'	GGCAGATCAGTTATTGAAGTTCG	2460 - 2482	67	contig2189117, a2009
	3'	CTTCATATGAGATTAAGGTGCCAC	4085 - 4108		
CT192	5'	GCTCCCACTGATATTCAGAATGTG	253 - 276	57	AK247888, contig6684196
	3'	CTTCTCTAACAAGGACATTCTCAGG	2037 - 2061		AK247888, contig6684196, a10549

[a]: binding position relative to the start codon (1st base) of the SL2.40 sequence
[b]: annealing temperature in °C
[c]: reference sequences from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) or the SOL Genomics network (http://solgenomics.net/)

Table A1.5: PCR primers used in previous studies.

Gene		Sequence 5' -> 3'	Binding position[a]	T_m[b]	Reference
candidate genes					
AREB1	5'	ATGGGGAGTAATTATCATTTCAAGAAC	1 - 27	64	Yanez *et al.* 2009
	3'	TTACCATGGACCAGTTTGTGTCCGTCT	2794 - 2820		
AREB2	5'	ATGGGATCTTACCTGAACTTCAAGAACTTTGC	1 - 32	70	Orellana *et al.* 2010
	3'	CCAAGGTCCCGTCACTGTCCTTCG	2370 - 2393		
CT208	5'	GCTACACAAGGTCAAGTCATCA	4 - 25	62	Arunyawat *et al.* 2007
	3'	AAAAGAGTTCATCTGTCATCTTG	4342 - 4364		T. Städler, W. Stephan unpublished work
CT208 LA3784	5'	GCTACACAAGGTCAAGTCATCA	4 - 25	60	Arunyawat *et al.* 2007
	3'	TAATGGACAGCACTAAATCAG	1336 - 1356		
pLC30-15	5'	CGAACAGAACAAGGCATCAG	15 - 34	61	Steige 2011
	3'	GCCTTTGAGTGGTATCCTG	788 - 806		
reference genes					
CT066	5'	CGCTGTCCCTCTTACCACCC	108 - 127	64.5	Arunyawat *et al.* 2007
	3'	ATGATAGGTGCGAACAGGGTC	1474 - 1494		
CT093	5'	CTCCCCTCGGCTACAGCATT	-403 - -384	63.5	Arunyawat *et al.* 2007
	3'	AGCAGCCCTTCAGAACGGACT	1031 - 1051		
CT143	5'	ATGGTTGGGTTCATTGTGG	6302 - 6320	61	Hörger *et al.* in preparation
	3'	CATCTAGTGTACAAGTTGGTTCTG	7952 - 7975		
CT166	5'	TGGAGCAGAGGTCAAGATTACT	1370 - 1391	62	Arunyawat *et al.* 2007
	3'	CATTCCATTGCTCTGCCTTC	2727 - 2746		
CT179	5'	CGAAATCATCTCCACACTCA	72 - 91	62.5	Arunyawat *et al.* 2007
	3'	TAAGACCAGCCAAACTACCAC	1034 - 1054		
CT189	5'	TGGAGGCAGCAAGAAGTGTGA	2 - 22	64	Hörger *et al.* in preparation
	3'	CCCGCTGACCACTGGATGTGAT	1428 - 1449		
CT198	5'	CTACCGAATTACGAGGAG	3711 - 3728	54	Arunyawat *et al.* 2007
	3'	TTAGTGCCACAATACAAC	4489 - 4506		
CT251	5'	TCGGACTCGATACTTCCTTG	360 - 379	58.5	Arunyawat *et al.* 2007
	3'	TCTCTTCATCCAGTTATCCG	2134 - 2153		
CT268	5'	CTATGGAGTTATATTTTCACCACA	23 - 46	58	Arunyawat *et al.* 2007
	3'	ACTTTTGAGAGGACATCAATTT	1988 - 2009		
GBSSI	5'	GATGGGCTCCAATCAAGAACTAAT	142 - 165	68	Peralta & Spooner 2001
	3'	GCCATTCACAATCCCAGTTATGC	1540 - 1562		

[a]: binding position relative to the start codon (1st base) of the SL2.40 sequence
[b]: annealing temperature in °C

Table A1.6: PCR and sequencing primers used for the outgroup sequencing.

Gene		Sequence 5' -> 3'	Gene		Sequence 5' -> 3'
PCR primer					
CT021	5'	CCACCGTGCTCGCGTCTTAG	*CT021*	3'	CCAAAGGCCCACTCGAAG
sequencing primer					
AREB1	5'	TAATACAGGGCTTGCATTTGG	*CT208*	5'	ATCATTTCATGGGAACTTCAAC
AREB1	5'	GCCTTTAATGTCTGCATGC	*CT208*	5'	TCCAAAGCCAATACACTGTG
AREB1	3'	TTAACTATGTCTATGGAAGC	*CT208*	5'	TTGTGTAGAGAACTTCCTTATGG
AREB1	3'	GCTTATCATTCTTCCAACCC	*CT208*	5'	TGAACATTGCCTTACTCCTTGGG
AREB2	5'	GTTGGCAGAGTTTCTCACACCTCC	*CT208*	5'	TTATGTCAGATTGAAGTGGCGG
AREB2	5'	CAAATACTAGTAGTGTCATGCAGGG	*CT208*	5'	AAGCCGATCACAAGTTCCTTC
JERF1	5'	TGAGTTGGTTAGTGTTGTATGG	*CT208*	3'	GCAATCTACTGCTTATCATAC
JERF1	3'	CTTTCTTGCCTCTGATCCTTC	*CT208*	3'	GTAGTGGAAGAGTTGGATAAGT
JERF3	3'	CCAAGGTCTCTGTCTGATCC	*CT208*	3'	TAATGGACAGCACTAAATCAG
TPP	5'	CGTGATAAGGTGCTTATCAATGC	*CT208*	3'	CTCTACACAATCTACCACTATGAC
CT093	5'	GGAAAGCTTGCTGCCGGTAG	*CT208*	3'	ATGGATTAGATTACCTTGCC
CT093	3'	TGAGGCAGAGTAAACATGCC	*CT208*	3'	AAACGGAGAGAGAGATGTATGG
CT114	5'	GATAGCTTCAGAAGATTGAGAAACTAG	*CT208*	3'	AGAGTTGGATAAATGAATAGTGAC
CT143	5'	TTCTTTGTTAGAGCTGGTATGAT	*CT251*	5'	GGAATTGGGCTATGCACCTGA
CT143	3'	ACCACAAGGTAAACTAAA	*CT251*	5'	ATCCATTGATTGTATTAGTTG
CT166	5'	GTGTTCTATCTAATGAGTTC	*CT251*	5'	AACTGCACATGATGATCC
CT166	3'	CCTTCTCCTTCATTTTCTCG	*CT251*	5'	ACTTTTAAGTGGAGTTGC
CT179	5'	GTGAGCATACTAGAAGGA	*CT251*	3'	TAGAGTTGTTTCCGGAAG
CT179	3'	AATAGCTACAAGCCCAGC	*CT251*	3'	GGTGATGCAAGCTCTGTG
CT189	3'	GGTTTTCTCCTGCTTTTTC	*CT268*	5'	CCATCTTCACGACCCAA
CT189	3'	TCCTTGCCATGGAAGGCTG	*CT268*	5'	GGTACTTTCAAGGATGCA
CT192	5'	CCTCTCTCTACAGCCACTTTG	*CT268*	3'	TTGCAGTTTATCATTGTT?
CT208	5'	TGTTCTTGGTCATGAGGCTG	*CT268*	3'	AGATTGTCTAATGGCATATTT
CT208	5'	AATTGAGAGCACTTCTGAGAT	*GBSSI*	3'	TGGCAATGAAGAGAACATCC

Table A2.1: Geographic and climatic information of the *S. chilense* populations in the salt stress experiment. Number of plants used in the experiment given.

Population	Province, Country	Latitude and Longitude	Altitude[a]	Temperature[b]	Precipitation[c]	No.
low altitude						
LA1930	Arequipa, Peru	15° 17' 30'' S, 74° 36' 0'' W	500	19.6	6	3
LA1963	Tacna, Peru	18° 4' 0'' S, 70° 19' 0'' W	200	18.2	11	6
LA2750	Antofagasta, Chile	22° 4' 13'' S, 70° 9' 47'' W	300	18.8	2	4
LA2932	Antofagasta, Chile	22° 28' 5'' S, 70° 13' 30'' W	300-400	18.5	2	4
LA4107	Antofagasta, Chile	25° 19' 8'' S, 70° 26' 46'' W	86	18.2	22	7
intermediate altitude						
LA1938	Arequipa, Peru	15° 41' 0'' S, 73° 50' 0'' W	1400	16.3	36	9
LA1958	Moquegua, Peru	17° 15' 0'' S, 71° 15' 0'' W	1250	17.1	6	7
LA1968	Tacna, Peru	17° 45' 42'' S, 70° 10' 35'' W	1680	15.8	18	7
LA2747	Arica and Parinacota, Chile	18° 35' 0'' S, 69° 54' 0'' W	800	16.5	9	5
LA2748	Tarapaca, Chile	21° 12' 48'' S, 69° 32' 52'' W	800	17.7	1	5
high altitude						
LA2773	Arica and Parinacota, Chile	18° 22' 0'' S, 69° 38' 0'' W	3400	10.3	142	4
LA2931	Tarapaca, Chile	20° 55' 0'' S, 69° 4' 0'' W	2275	12.3	18	8
LA3111	Tacna, Peru	17° 28' 0'' S, 70° 2' 0'' W	3070	11.6	121	4
LA4117A	Antofagasta, Chile	22° 54' 27'' S, 67° 56' 27'' W	3540	8.5	39	6
LA4332	Antofagasta, Chile	22° 36' 32'' S, 68° 31' 19'' W	2968	10.5	29	7

Note: geographic data from the Tomato Genetics Resource Center, UC Davis (http://tgrc.usdavis.edu/), climatic data from WorldClim database (http://www.worldclim.org/, Hijmans *et al.* 2005)
[a]: altitude in meter above sea level
[b]: mean annual temperature in °C
[c]: annual precipitation in mm

APPENDIX B: Results

Table B1.1: Statistical summary of the sequencing. Number of paired end reads per population returned from the GATC Biotech AG Konstanz, Germany, number of mapped paired end reads, and percentage of mapped reads.

population	# paired end reads	# mapped paired end reads[a]	% mapped paired end reads
LA0456	11,114,074	10,501,130	94.48 %
LA0458	7,943,450	6,790,151	85.48 %
LA1930	13,462,829	12,977,089	96.39 %
LA1958	13,900,378	13,402,041	96.41 %
LA1963	7,656,756	7,391,431	96.53 %
LA1968	12,237,620	10,800,104	88.25 %
LA1968‘ [b]	9,720,649	8,884,645	91.40 %
LA2747	9,192,638	8,680,353	94.43 %
LA2748	9,699,656	9,019,231	92.99 %
LA2750	13,793,177	13,283,244	96.30 %
LA2753	10,871,432	10,225,897	94.06 %
LA2755	9,266,336	8,677,702	93.65 %
LA2765	7,327,729	6,522,545	89.01 %
LA2773	13,835,315	13,476,581	97.41 %
LA2880	14,601,371	14,174,198	97.07 %
LA2931	7,571,610	6,849,597	90.46 %
LA2932	8,653,157	8,039,770	92.91 %
LA3111	11,618,239	11,166,004	96.11 %
LA3784	6,328,429	5,459,486	86.27 %
LA4107	7,750,520	6,543,895	84.43 %
LA4108	8,995,301	8,109,526	90.15 %
LA4118	10,339,623	9,055,910	87.58 %
LA4119	9,228,000	7,511,681	81.40 %
LA4332	6,860,625	6,434,875	93.79 %

[a]: mapped with Stampy v.1.0.20 (Lunter & Goodson 2011)
[b]: repetition of LA1968

Table B2.1: Pairwise population genetic differentiation. Populations are sorted from north to south (left to right and top to bottom). Cells above the diagonal represent mean F_{ST} values of all 30 genes and cells below the diagonal represent mean F_{ST} values of the 14 reference genes.

	LA1930	**LA3784**	**LA0456**	**LA1958**	**LA3111**	**LA1968**	**LA0458**	**LA1963**	**LA2773**	**LA2747**	**LA2765**	**LA2755**	**LA2753**	**LA2931**	**LA2748**	**LA2750**	**LA2932**	**LA4332**	**LA4118**	**LA4119**	**LA2880**	**LA4108**	**LA4107**
LA1930	-	0.207	0.293	0.212	0.291	0.267	0.337	0.236	0.253	0.268	0.235	0.277	0.243	0.221	0.263	0.353	0.382	0.311	0.353	0.359	0.365	0.416	0.436
LA3784	0.168	-	0.318	0.239	0.308	0.295	0.383	0.263	0.284	0.300	0.278	0.298	0.270	0.270	0.314	0.385	0.418	0.347	0.379	0.387	0.401	0.447	0.467
LA0456	0.258	0.280	-	0.143	0.213	0.211	0.299	0.179	0.211	0.227	0.204	0.215	0.183	0.195	0.216	0.344	0.398	0.255	0.306	0.335	0.345	0.415	0.444
LA1958	0.204	0.238	0.159	-	0.123	0.109	0.179	0.094	0.110	0.116	0.106	0.124	0.092	0.088	0.131	0.221	0.266	0.145	0.177	0.199	0.207	0.266	0.290
LA3111	0.290	0.316	0.213	0.124	-	0.136	0.233	0.094	0.147	0.174	0.148	0.179	0.125	0.135	0.180	0.265	0.325	0.176	0.237	0.268	0.277	0.322	0.354
LA1968	0.251	0.280	0.242	0.111	0.138	-	0.198	0.088	0.119	0.125	0.119	0.139	0.108	0.106	0.148	0.223	0.279	0.156	0.211	0.224	0.225	0.264	0.289
LA0458	0.324	0.376	0.317	0.171	0.205	0.158	-	0.169	0.198	0.206	0.209	0.248	0.205	0.193	0.251	0.324	0.397	0.256	0.311	0.328	0.345	0.404	0.432
LA1963	0.228	0.260	0.203	0.091	0.106	0.071	0.133	-	0.076	0.089	0.079	0.116	0.078	0.069	0.104	0.183	0.227	0.103	0.166	0.180	0.177	0.211	0.237
LA2773	0.241	0.271	0.219	0.098	0.159	0.106	0.162	0.067	-	0.056	0.072	0.140	0.088	0.061	0.104	0.190	0.234	0.126	0.186	0.205	0.205	0.232	0.258
LA2747	0.269	0.300	0.256	0.125	0.205	0.120	0.183	0.085	0.050	-	0.076	0.143	0.100	0.076	0.128	0.206	0.248	0.131	0.197	0.212	0.206	0.251	0.277
LA2765	0.241	0.271	0.209	0.112	0.160	0.124	0.190	0.076	0.064	0.079	-	0.123	0.087	0.070	0.105	0.181	0.217	0.128	0.174	0.196	0.188	0.230	0.257
LA2755	0.235	0.267	0.195	0.108	0.169	0.122	0.205	0.097	0.108	0.129	0.105	-	0.102	0.131	0.185	0.255	0.313	0.168	0.226	0.266	0.264	0.292	0.321
LA2753	0.233	0.259	0.187	0.092	0.136	0.105	0.187	0.082	0.086	0.113	0.090	0.089	-	0.089	0.128	0.192	0.239	0.131	0.177	0.198	0.201	0.237	0.266
LA2931	0.221	0.266	0.198	0.076	0.143	0.098	0.141	0.070	0.053	0.072	0.086	0.100	0.092	-	0.087	0.184	0.231	0.123	0.180	0.203	0.204	0.234	0.259
LA2748	0.274	0.294	0.195	0.102	0.158	0.121	0.199	0.085	0.087	0.116	0.099	0.130	0.101	0.086	-	0.236	0.287	0.156	0.219	0.241	0.241	0.276	0.308
LA2750	0.359	0.378	0.384	0.238	0.305	0.226	0.303	0.198	0.199	0.226	0.198	0.262	0.216	0.203	0.232	-	0.070	0.254	0.286	0.321	0.321	0.223	0.262
LA2932	0.429	0.443	0.472	0.328	0.404	0.318	0.432	0.279	0.279	0.307	0.270	0.350	0.309	0.300	0.331	0.082	-	0.305	0.353	0.390	0.396	0.314	0.363
LA4332	0.278	0.336	0.260	0.139	0.163	0.128	0.172	0.078	0.088	0.106	0.114	0.147	0.127	0.091	0.100	0.255	0.342	-	0.133	0.174	0.181	0.300	0.335
LA4118	0.339	0.372	0.348	0.200	0.257	0.220	0.290	0.179	0.181	0.215	0.179	0.225	0.195	0.164	0.180	0.324	0.438	0.158	-	0.111	0.133	0.383	0.426
LA4119	0.350	0.395	0.370	0.209	0.254	0.214	0.287	0.171	0.209	0.234	0.208	0.261	0.201	0.185	0.200	0.343	0.456	0.161	0.093	-	0.101	0.389	0.424
LA2880	0.390	0.432	0.432	0.257	0.326	0.249	0.352	0.201	0.234	0.254	0.223	0.280	0.238	0.232	0.245	0.393	0.511	0.225	0.190	0.118	-	0.410	0.448
LA4108	0.422	0.456	0.462	0.280	0.347	0.242	0.353	0.210	0.245	0.285	0.238	0.274	0.254	0.252	0.246	0.256	0.390	0.284	0.403	0.377	0.451	-	0.065
LA4107	0.433	0.465	0.477	0.297	0.365	0.255	0.365	0.224	0.264	0.298	0.257	0.295	0.270	0.268	0.267	0.271	0.416	0.305	0.429	0.400	0.481	0.067	-

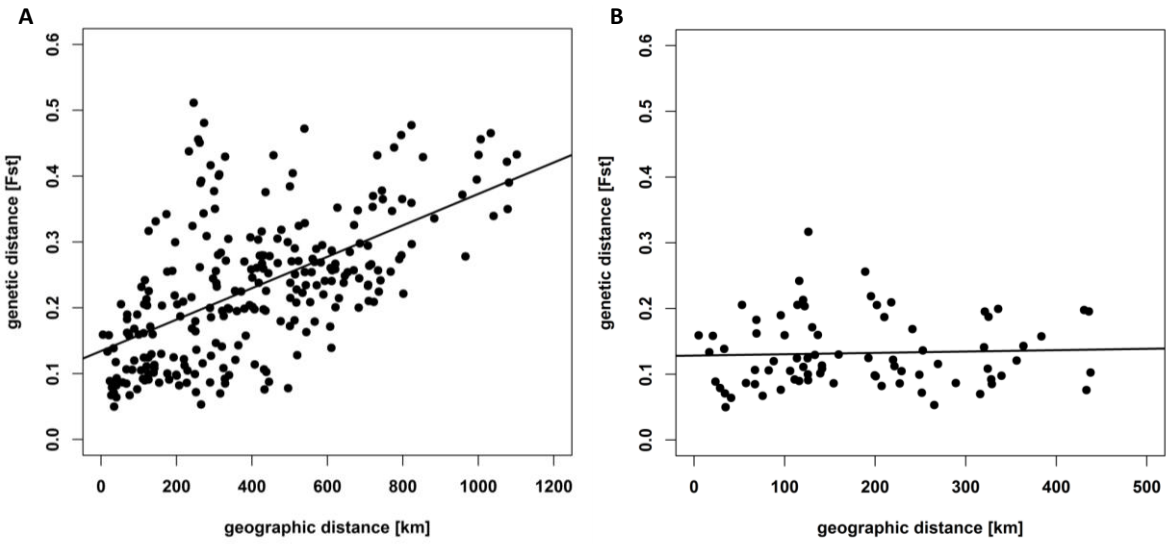

Figure B2.1: Isolation by distance for the average over the 14 reference genes. Genetic distance (F_{ST}) between populations plotted against geographic distance (km). A) A pattern of isolation by distance is observed for all 23 *S. chilense* populations (R^2 = 0.332, Mantel test p-value < 0.001. B) A pattern of isolation by distance is not observed for the 13 *S. chilense* populations from the central group (R^2 = 0.002, Mantel test not significant).

Table B2.2: Summary of within and between group genetic differentiation.

	Mean F_{ST}	Min F_{ST} (pair)	Max F_{ST} (pair)
<u>all 30 genes</u>			
within groups			
northern group (1)	0.207 (± 0)	0.207 (LA1930/LA3784)	0.207 (LA1930/LA3784)
central group (78)	0.140 (± 0.0539)	0.056 (LA2747/LA2773)	0.299 (LA0456/LA0458)
southern high altitude group (6)	0.139 (± 0.0326)	0.101 (LA2880/LA4119)	0.181 (LA2880/LA4332)
coastal group (6)	0.216 (± 0.1246)	0.065 (LA4107/LA4108)	0.363 (LA2932/LA4107)
between groups			
central group – northern group (26)	0.278 (± 0.0376)	0.212 (LA1958/LA1930)	0.383 (LA0458/LA3784)
central group – southern high altitude group (52)	0.211 (± 0.0586)	0.103 (LA1963/LA4332)	0.345 (LA0458/LA2880)
central group – coastal group (52)	0.275 (± 0.0658)	0.181 (LA2765/LA2750)	0.444 (LA0456/LA4107)
northern group – southern high altitude group (8)	0.363 (± 0.0277)	0.311 (LA1930/LA4332)	0.401 (LA3784/LA2880)
northern group – coastal group (8)	0.413 (± 0.0378)	0.353 (LA1930/LA2750)	0.467 (LA3784/LA4107)
southern high altitude group – coastal group (16)	0.359 (± 0.0573)	0.254 (LA4332/LA2750)	0.448 (LA2880/LA4107)
<u>the 14 reference genes</u>			
within groups			
northern group (1)	0.168	0.168 (LA1930/LA3784)	0.168 (LA1930/LA3784)
central group (78)	0.132 (± 0.0533)	0.050 (LA2747/LA2773)	0.317 (LA0456/LA0458)
southern high altitude group (6)	0.157 (± 0.0478)	0.093 (LA4118/LA4119)	0.225 (LA2880/LA4332)
coastal group (6)	0.247 (± 0.1478)	0.067 (LA4107/LA4108)	0.416 (LA2932/LA4107)
between groups			
central group – northern group (26)	0.267 (± 0.0362)	0.204 (LA1958/LA1930)	0.376 (LA0458/LA3784)
central group – southern high altitude group (52)	0.213 (± 0.0736)	0.078 (LA1963/LA4332)	0.432 (LA0456/LA2880)
central group – coastal group (52)	0.292 (± 0.0698)	0.198 (LA1963/LA2750)	0.477 (LA0456/LA4107)
northern group – southern high altitude group (8)	0.362 (± 0.0467)	0.278 (LA1930/LA4332)	0.432 (LA3784/LA2880)
northern group – coastal group (8)	0.423 (± 0.0368)	0.359 (LA1930/LA2750)	0.465 (LA3784/LA4107)
southern high altitude group – coastal group (16)	0.387 (± 0.0734)	0.255 (LA4332/LA2750)	0.511 (LA2880/LA2932)

Table B2.3: Interspecies genetic differentiation between the *S. chilense* populations and three other wild tomato species.

Population	*S. peruvianum*[a] species wide sample	*S. arcanum*[b] pooled population sample	*S. habrochaites*[b] pooled population sample
LA0456	0.384	0.480	0.613
LA0458	0.444	0.570	0.701
LA1930	0.321	0.436	0.537
LA1958	0.328	0.434	0.551
LA1963	0.301	0.415	0.547
LA1968	0.338	0.441	0.579
LA2747	0.350	0.450	0.564
LA2748	0.343	0.461	0.592
LA2750	0.440	0.539	0.637
LA2753	0.298	0.408	0.540
LA2755	0.311	0.429	0.542
LA2765	0.291	0.390	0.513
LA2773	0.319	0.424	0.537
LA2880	0.461	0.584	0.728
LA2931	0.321	0.422	0.550
LA2932	0.489	0.604	0.697
LA3111	0.387	0.465	0.595
LA3784	0.304	0.397	0.479
LA4107	0.524	0.636	0.773
LA4108	0.517	0.629	0.766
LA4118	0.415	0.545	0.690
LA4119	0.454	0.558	0.700
LA4332	0.379	0.500	0.619
mean	0.379	0.488	0.611

[a]: average over nine reference genes, *S. peruvianum* sequence data from Tellier *et al.* (2011b) and Hörger *et al.* (in preparation)
[b]: average over seven reference genes, *S. arcanum* and *S. habrochaites* sequence data from Tellier *et al.* (2011a)

Table B2.4: Divergence from other Solanaceae.

Population	*S. ochranthum*		*S. lycopersicoides*		*S. lycopersicum*	
	K_{all}	K_{ref}	K_{all}	K_{ref}	K_{all}	K_{ref}
LA0456	0.0359	0.0331	0.0341	0.0316	0.0179	0.0165
LA0458	0.0347	0.0328	0.0333	0.0316	0.0182	0.0182
LA1930	0.0359	0.0334	0.0353	0.0321	0.0185	0.0165
LA1958	0.0353	0.0328	0.0336	0.0314	0.0186	0.0177
LA1963	0.0348	0.0323	0.0331	0.0311	0.0181	0.0171
LA1968	0.0350	0.0329	0.0335	0.0320	0.0183	0.0175
LA2747	0.0348	0.0332	0.0332	0.0317	0.0185	0.0181
LA2748	0.0348	0.0330	0.0331	0.0314	0.0181	0.0176
LA2750	0.0356	0.0338	0.0348	0.0328	0.0191	0.0191
LA2753	0.0348	0.0323	0.0337	0.0314	0.0183	0.0172
LA2755	0.0348	0.0324	0.0335	0.0315	0.0180	0.0171
LA2765	0.0347	0.0329	0.0333	0.0314	0.0179	0.0176
LA2773	0.0349	0.0329	0.0332	0.0313	0.0181	0.0175
LA2880	0.0346	0.0327	0.0332	0.0319	0.0181	0.0171
LA2931	0.0349	0.0327	0.0334	0.0316	0.0179	0.0173
LA2932	0.0356	0.0337	0.0342	0.0324	0.0187	0.0190
LA3111	0.0354	0.0329	0.0342	0.0319	0.0186	0.0175
LA3784	0.0367	0.0337	0.0360	0.0329	0.0193	0.0175
LA4107	0.0346	0.0326	0.0334	0.0318	0.0182	0.0180
LA4108	0.0347	0.0326	0.0334	0.0315	0.0181	0.0177
LA4118	0.0349	0.0326	0.0334	0.0320	0.0182	0.0175
LA4119	0.0349	0.0328	0.0333	0.0318	0.0182	0.0174
LA4332	0.0347	0.0321	0.0337	0.0317	0.0183	0.0175
mean	0.0351	0.0329	0.0337	0.0318	0.0183	0.0176

Note: “all” averaged over all genes, “ref” averaged over the reference genes

Table B3.1: Watterson estimator, Θ_W, for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	0.00817	0.00537	0.00323	0.00880	0.00865	0.01253	0.00249	0.00690	0.01338	0.01383	0.00537	0.00782	0.00973	0.00512	0.00376	0.00682	0.00762
LA0458	0.00890	0.00401	0.00146*	0.00716	0.00976	0.00721	0.00250	0.00215	0.00276	0.01351	0.00751	0.00430	0.00440	0.00157*	0.00308	0.00459	0.00531
LA1930	0.01147	0.00969	0.00586	0.01184	0.01118	0.01401	0.01145	0.01200	0.01523	0.01493	0.01181	0.01337	0.01707	0.00877	0.00718	0.00999	0.01162
LA1958	0.01730	0.00781	0.00462	0.01174	0.00930	0.01410	0.00748	0.01198	0.01052	0.00968	0.01395	0.00785	0.01243	0.00956	0.00752	0.00845	0.01027
LA1963	0.01850#	0.00669	0.00647	0.01211	0.00763	0.01280	0.00660	0.01508	0.01572	0.01484	0.01288	0.00638	0.01233	0.00666	0.01101	0.00682	0.01078
LA1968	0.01307	0.00823	0.00586	0.00979	0.01104	0.01156	0.00824	0.00615	0.01148	0.00581	0.00859	0.00701	0.01324	0.00757	0.00790	0.00952	0.00907
LA2747	0.01005	0.00744	0.00438	0.01050	0.00781	0.01111	0.00709	0.01267	0.00906	0.00901	0.00966	0.00687	0.01295	0.00248	0.00584	0.00625	0.00832
LA2748	0.01199	0.00625	0.00386	0.01058	0.00639	0.00923	0.00540	0.00661	0.00971	0.01158	0.02254#	0.00597	0.01449	0.00597	0.00827	0.00746	0.00914
LA2750	0.00691	0.00805	n. a.	0.01016	0.01026	0.02156#	0.01164	0.00768	0.00827	0.01424	0.03327#	0.01485	0.01290	0.00692	0.00481	0.00645	0.01186
LA2753	0.01218	0.00803	0.00532	0.01189	0.00771	0.01508	0.00876	0.01046	0.01324	0.00710	0.02147#	0.00882	0.01566	0.00659	0.00930	0.00858	0.01064
LA2755	0.01251	0.00612	0.00125*	0.01025	0.00825	0.01074	0.00754	0.00658	0.01077	0.00581	0.00859	0.00666	0.01015	0.00293	0.00821	0.00525	0.00760
LA2765	0.01318	0.00990	0.00648	0.01507	0.01266	0.01623	0.01294	0.01093	0.01543	0.00516	0.02653#	0.00938	0.01724	0.01043	0.00859	0.01498	0.01282
LA2773	0.01088	0.00762	0.00396	0.01071	0.00757	0.01191	0.00668	0.01051	0.00783	0.00839	0.00859	0.00920	0.01368	0.00740	0.00756	0.00579	0.00864
LA2880	0.00745	0.00476	0.00042*	0.00619	0.00451	0.01249	0.00208	0.00497	0.00664	0.00516	0.01825#	0.00334	0.01467	0.00084*	0.00547	0.00208	0.00621
LA2931	0.01320	0.00946	0.00376	0.01225	0.01174	0.01238	0.00790	0.00920	0.01059	0.01359	0.00859	0.01068	0.01779	0.00867	0.00694	0.01025	0.01044
LA2932	0.00633	0.00412	0.00344	0.00554	0.00474	0.00880	0.00371	0.00805	0.00459	0.00866	0.01073	0.00868	0.00921	0.00225	0.00240	0.00371	0.00594
LA3111	0.01163	0.00646	0.00366	0.01057	0.00978	0.01271	0.00577	0.00994	0.01021	0.01558	0.02267#	0.00525	0.01107	0.00662	0.00581	0.00788	0.00972
LA3784	0.00920	0.00574	0.00613	0.01238	0.00674	0.01273	0.01028	0.00434	0.01839#	0.01484	0.02683#	0.01567	0.01680	0.01009	0.01162	0.00679	0.01179
LA4107	0.00440	0.00297	0.00261	0.00361	0.00461	0.00521	0.00411	0.00694	0.00419	0.00452	0.01503	0.00441	0.00261	0.00247	0.00657	0.00207	0.00477
LA4108	0.00611	0.00279	0.00492	0.00327	0.00490	0.00826	0.00329	0.00362	0.00326	0.00772	0.02469#	0.00342	0.00432	0.00210	0.00550	0.00461	0.00580
LA4118	0.00889	0.00612	0.00167	0.00480	0.00666	0.00917	0.00208	0.00529	0.00716	0.01613	0.00966	0.00393	0.00827	0.00124*	0.00718	0.00284	0.00632
LA4119	0.00988	0.00758	0.00188	0.00753	0.00711	0.01464	0.01340	0.00810	0.00859	0.01480	0.01742	0.00685	0.01508	0.00187	0.01111	0.00380	0.00935
LA4332	0.01284	0.00564	0.00229	0.01115	0.00958	0.01048	0.00583	0.00776	0.00906	0.01549	0.02039#	0.00789	0.01285	0.00507	0.00482	0.00647	0.00922
Mean	0.01065	0.00656	0.00380	0.00947	0.00820	0.01195	0.00684	0.00817	0.00983	0.01089	0.01587	0.00776	0.01213	0.00536	0.00698	0.00658	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.2: Watterson estimator, Θ_W, for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.01060	0.00070*	0.00418	0.00296	0.00738	0.00716	0.00794	0.00210	0.00739	0.00628	0.01153	0.00756	0.00431	0.00183	0.00585
LA0458	0.00232	0.00350	0.00218	0.00517	0.01176	0.00067*	0.01496	0.00408	0.00135*	0.00586	0.00154*	0.00364	0.00268	0.00250	0.00444
LA1930	0.01569	0.00629	0.00710	0.00605	0.01636	0.00642	0.01120	0.00981	0.01051	0.01006	0.01853#	0.00796	0.00839	0.00682	0.01008
LA1958	0.01058	0.00507	0.00602	0.00518	0.01474	0.00801	0.01166	0.01008	0.00608	0.00874	0.01367	0.00817	0.00677	0.00233	0.00836
LA1963	0.01318	0.00805	0.00620	0.00623	0.02147#	0.01155	0.00963	0.01206	0.00827	0.01228	0.01079	0.00944	0.00804	0.00366	0.01006
LA1968	0.00656	0.00717	0.00763	0.00650	0.01589	0.01104	0.01103	0.01155	0.00355	0.01277	0.00699	0.00815	0.00746	0.00615	0.00875
LA2747	0.00864	0.00524	0.00691	0.00577	0.00939	0.00797	0.01294	0.00829	0.00507	0.01079	0.00764	0.00758	0.00641	0.00266	0.00752
LA2748	0.00868	0.00682	0.00563	0.00500	0.01806#	0.00745	0.00807	0.01170	0.00254	0.00916	0.00699	0.00676	0.00614	0.00266	0.00755
LA2750	0.01508	0.01225	0.01401	0.00677	0.01005	0.00898	0.01523	0.00911	0.00739	0.00610	0.01430	0.00832	0.01002	0.00932	0.01049
LA2753	0.01153	0.00682	0.00673	0.00519	0.01586	0.01014	0.01045	0.01060	0.00355	0.00980	0.01128	0.01029	0.00746	0.00416	0.00885
LA2755	0.00587	0.00612	0.00621	0.00654	0.01316	0.00915	0.01035	0.00650	0.00604	0.00820	0.00182	0.00829	0.00572	0.00316	0.00694
LA2765	0.01504	0.00909	0.00817	0.00588	0.01883#	0.00898	0.01892#	0.01406	0.00843	0.00895	0.00852	0.01114	0.00839	0.00584	0.01073
LA2773	0.00923	0.00524	0.00564	0.00581	0.01248	0.00746	0.00923	0.01189	0.00391	0.00842	0.00912	0.00844	0.00711	0.00216	0.00758
LA2880	0.00718	0.00175	0.00345	0.00059*	0.00435	0.00628	0.00460	0.00552	0.00304	0.00352	0.00335	0.00463	0.00093*	0.00067*	0.00356
LA2931	0.01232	0.00594	0.00618	0.00794	0.01538	0.01019	0.00959	0.00989	0.00152*	0.00953	0.01395	0.00505	0.00736	0.00382	0.00848
LA2932	0.01178	0.00682	0.00345	0.00236	0.00738	0.00321	0.01137	0.00519	0.00254	0.00260	0.00576	0.00479	0.00397	0.00133*	0.00518
LA3111	0.01061	0.00612	0.00618	0.00338	0.01517	0.00680	0.00890	0.00974	0.00236	0.00822	0.00408	0.00885	0.00676	0.00283	0.00714
LA3784	0.01120	0.00613	0.00968	0.00710	0.01920#	0.00857	0.01371	0.01034	0.00763	0.01107	0.01458	0.01184	0.01014	0.00516	0.01045
LA4107	0.00683	0.00454	0.00564	0.00309	0.00602	0.00186	0.00541	0.00509	0.00118*	0.00413	0.00516	0.00461	0.00176	0.00284	0.00415
LA4108	0.00673	0.00402	0.00656	0.00279	0.00645	0.00101*	0.00598	0.00342	0.00102*	0.00377	0.00367	0.00480	0.00187	0.00333	0.00396
LA4118	0.00800	0.00402	0.00454	0.00221	0.01162	0.00541	0.00559	0.00502	0.00254	0.00796	0.00431	0.00462	0.00338	0.00234	0.00511
LA4119	0.00919	0.00875	0.00366	0.00460	0.00904	0.00798	0.01230	0.00688	0.00373	0.00622	0.00564	0.00787	0.00578	0.00420	0.00685
LA4332	0.00853	0.00577	0.00418	0.00368	0.01292	0.00728	n. a.	0.00729	0.00355	0.00743	0.00337	0.00816	0.00699	0.00233	0.00627
Mean	0.00980	0.00592	0.00609	0.00482	0.01274	0.00711	0.01041	0.00827	0.00449	0.00791	0.00811	0.00743	0.00599	0.00357	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.3: Nucleotide diversity, π, for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	0.00466	0.00387	0.00225	0.01243	0.01212	0.00776	0.00207	0.00371	0.01485	0.00787	0.00187	0.01160	0.00937	0.00528	0.00538	0.00974	0.00718
LA0458	0.01355	0.00294	0.00066	0.00884	0.00363	0.00659	0.00292	0.00271	0.00427	0.00797	0.00513	0.00241	0.00294	0.00143	0.00438	0.00702	0.00484
LA1930	0.01129	0.00851	0.00437	0.00871	0.01118	0.01093	0.00883	0.01358	0.01830#	0.01039	0.00621	0.01151	0.01400	0.01093	0.00656	0.01108	0.01040
LA1958	0.01916#	0.00833	0.00642	0.01309	0.01162	0.01252	0.00718	0.00982	0.01222	0.01030	0.00977	0.00765	0.01274	0.01201	0.00784	0.01177	0.01078
LA1963	0.01924#	0.00644	0.00617	0.01219	0.00758	0.01002	0.00323	0.01631	0.01186	0.01127	0.00747	0.00600	0.01116	0.00484	0.00781	0.00859	0.00939
LA1968	0.01656#	0.00876	0.00212	0.01056	0.01105	0.00908	0.00484	0.00442	0.01470	0.00321	0.00593	0.00694	0.01043	0.00550	0.00824	0.00821	0.00816
LA2747	0.01254	0.00793	0.00463	0.01181	0.00781	0.00964	0.00692	0.01002	0.00636	0.00763	0.00969	0.00560	0.00971	0.00185	0.00622	0.00818	0.00791
LA2748	0.01251	0.00508	0.00352	0.00931	0.00834	0.00717	0.00505	0.00290	0.00891	0.00694	0.01192	0.00653	0.01052	0.00890	0.00761	0.00606	0.00758
LA2750	0.00427	0.00237	n. a.	0.00833	0.00370	0.00891	0.00317	0.00770	0.00453	0.00309	0.00911	0.00780	0.00863	0.00276	0.00196	0.00233	0.00524
LA2753	0.01477	0.00876	0.00506	0.01539	0.00832	0.01097	0.00552	0.00928	0.01262	0.00544	0.01051	0.01127	0.01615	0.00723	0.00680	0.00812	0.00976
LA2755	0.01478	0.00558	0.00072	0.00679	0.01174	0.00968	0.00658	0.00799	0.00950	0.00157	0.00430	0.00848	0.01161	0.00267	0.00837	0.00739	0.00736
LA2765	0.01074	0.00871	0.00682	0.01378	0.00976	0.01019	0.00815	0.01010	0.01015	0.00314	0.01161	0.00788	0.01172	0.00602	0.00875	0.00905	0.00916
LA2773	0.01231	0.00838	0.00569	0.01314	0.00599	0.01078	0.00524	0.01024	0.00827	0.00637	0.01141	0.00588	0.01388	0.00514	0.01050	0.00760	0.00880
LA2880	0.00950	0.00572	0.00007*	0.00984	0.00132	0.00861	0.00037*	0.00446	0.01126	0.00250	0.00877	0.00514	0.01279	0.00040*	0.00608	0.00183	0.00554
LA2931	0.01663#	0.00939	0.00475	0.01289	0.01015	0.00928	0.00514	0.00711	0.00869	0.00909	0.00761	0.00797	0.01134	0.00820	0.00539	0.01226	0.00912
LA2932	0.00828	0.00136	0.00519	0.00691	0.00188	0.00593	0.00073	0.00593	0.00487	0.00500	0.00365	0.00722	0.01021	0.00202	0.00280	0.00207	0.00463
LA3111	0.01486	0.00740	0.00164	0.01151	0.01136	0.00842	0.00400	0.00438	0.00958	0.00944	0.00482	0.00459	0.01023	0.00907	0.00373	0.00861	0.00773
LA3784	0.00576	0.00688	0.00597	0.01414	0.00191	0.00794	0.00881	0.00253	0.01791#	0.01324	0.01618	0.01107	0.01272	0.00990	0.00791	0.00888	0.00949
LA4107	0.00247	0.00106	0.00110	0.00210	0.00256	0.00156	0.00122	0.00212	0.00147	0.00208	0.00466	0.00160	0.00137	0.00114	0.00147	0.00037*	0.00177
LA4108	0.00242	0.00150	0.00176	0.00155	0.00218	0.00378	0.00059*	0.00307	0.00058*	0.00377	0.00692	0.00119	0.00253	0.00192	0.00393	0.00591	0.00272
LA4118	0.01200	0.00799	0.00057*	0.00257	0.00287	0.00790	0.00212	0.00388	0.01116	0.00739	0.00466	0.00391	0.00839	0.00044*	0.00383	0.00196	0.00510
LA4119	0.01197	0.00457	0.00035*	0.00462	0.00182	0.00966	0.00269	0.00592	0.00886	0.00320	0.00774	0.00605	0.01544	0.00055*	0.00331	0.00198	0.00555
LA4332	0.01658#	0.00456	0.00138	0.00641	0.00364	0.01090	0.00240	0.00489	0.01009	0.00732	0.00542	0.00955	0.00965	0.00293	0.00345	0.00857	0.00673
Mean	0.01160	0.00592	0.00324	0.00943	0.00663	0.00862	0.00425	0.00665	0.00961	0.00644	0.00762	0.00686	0.01033	0.00483	0.00575	0.00685	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.4: Nucleotide diversity, π, for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.01442	0.00035*	0.00421	0.00211	0.01307	0.00758	0.01002	0.00119	0.00484	0.00414	0.00848	0.01124	0.00578	0.00120	0.00633
LA0458	0.00241	0.00545	0.00132	0.00571	0.01112	0.00012*	0.00860	0.00589	0.00102	0.00980	0.00240	0.00381	0.00409	0.00139	0.00451
LA1930	0.01209	0.00653	0.00682	0.00287	0.01852#	0.00465	0.00899	0.00989	0.00532	0.01339	0.02548#	0.01078	0.00800	0.00381	0.00980
LA1958	0.00844	0.00638	0.00449	0.00615	0.02364#	0.00871	0.01004	0.00949	0.00706	0.00942	0.01388	0.00805	0.00871	0.00183	0.00902
LA1963	0.01022	0.00727	0.00416	0.00484	0.02310#	0.01203	0.01012	0.01039	0.00419	0.01192	0.00568	0.00897	0.00919	0.00337	0.00896
LA1968	0.00759	0.00645	0.00449	0.00469	0.02182#	0.01261	0.00731	0.00900	0.00210	0.01014	0.00551	0.00834	0.00868	0.00408	0.00806
LA2747	0.01271	0.00725	0.00556	0.00585	0.01259	0.01142	0.01001	0.00671	0.00247	0.01236	0.00719	0.00530	0.00704	0.00284	0.00781
LA2748	0.00975	0.00769	0.00334	0.00415	0.02502#	0.01195	0.00907	0.00836	0.00269	0.00688	0.00489	0.00379	0.00848	0.00144	0.00768
LA2750	0.01009	0.00965	0.00452	0.00177	0.00743	0.00270	0.00874	0.00553	0.00322	0.00177	0.01422	0.00269	0.00345	0.00210	0.00556
LA2753	0.01425	0.00736	0.00383	0.00633	0.02282#	0.01034	0.01195	0.00907	0.00329	0.01103	0.00536	0.01132	0.00883	0.00256	0.00917
LA2755	0.00781	0.00915	0.00508	0.00480	0.01602	0.01456	0.00764	0.00648	0.01046	0.01283	0.00062	0.00998	0.00674	0.00164	0.00813
LA2765	0.01121	0.00910	0.00551	0.00401	0.01553	0.01271	0.01098	0.01070	0.00735	0.00944	0.00710	0.01082	0.00850	0.00288	0.00899
LA2773	0.01128	0.00636	0.00493	0.00595	0.01512	0.01184	0.00921	0.01132	0.00325	0.01089	0.00771	0.01013	0.00757	0.00193	0.00839
LA2880	0.00615	0.00031*	0.00280	0.00071	0.00689	0.00438	0.00528	0.00709	0.00203	0.00526	0.00356	0.00397	0.00030*	0.00045*	0.00351
LA2931	0.01080	0.00551	0.00558	0.00690	0.01913#	0.01143	0.01085	0.00943	0.00193	0.01147	0.00981	0.00531	0.00893	0.00159	0.00848
LA2932	0.00412	0.00654	0.00349	0.00090	0.00558	0.00102	0.00537	0.00495	0.00120	0.00104	0.00472	0.00186	0.00151	0.00084	0.00308
LA3111	0.01474	0.00747	0.00513	0.00435	0.01051	0.00488	0.00846	0.00726	0.00179	0.01190	0.00492	0.01218	0.00635	0.00155	0.00725
LA3784	0.01126	0.00558	0.00858	0.00442	0.01591	0.01050	0.00919	0.00665	0.00251	0.01515	0.01789#	0.01150	0.01276	0.00347	0.00967
LA4107	0.00431	0.00114	0.00176	0.00132	0.00671	0.00060	0.00245	0.00227	0.00055*	0.00601	0.00155	0.00245	0.00101	0.00125	0.00238
LA4108	0.00658	0.00164	0.00237	0.00110	0.00631	0.00021*	0.00364	0.00162	0.00064	0.00648	0.00071	0.00238	0.00113	0.00166	0.00260
LA4118	0.00775	0.00299	0.00237	0.00052*	0.01658#	0.00814	0.00301	0.00588	0.00238	0.00757	0.00438	0.00337	0.00471	0.00114	0.00506
LA4119	0.00795	0.00608	0.00066	0.00100	0.00289	0.00918	0.00584	0.00493	0.00415	0.00532	0.00305	0.00480	0.00257	0.00130	0.00426
LA4332	0.01168	0.00680	0.00183	0.00338	0.01952#	0.00916	n. a.	0.00586	0.00245	0.01041	0.00351	0.00861	0.00689	0.00095	0.00700
Mean	0.00946	0.00578	0.00404	0.00364	0.01460	0.00786	0.00803	0.00696	0.00334	0.00890	0.00707	0.00703	0.00614	0.00197	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.5: Synonymous nucleotide diversity, π_s, for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	0.00510	0.00304	0.00089	0.01348	0.00632	0.01654	0.00043	0.00533	0.01193	0.01644	0*	0.01525	0.00518	0.00927	0.00573	0.00875	0.00773
LA0458	0.00903	0.00868	0.00018	0.00689	0.00363	0.01930	0.00566	0.00990	0.01058	0.01975	0*	0.00209	0*	0.00246	0*	0.01281	0.00694
LA1930	0.01474	0.01114	0.01046	0.01426	0.00843	0.03934#	0.01917	0.01399	0.00335	0.02296	0.01450	0.00847	0.01588	0.00732	0.00054	0.01315	0.01361
LA1958	0.02613	0.01530	0.00482	0.01519	0.01595	0.04849#	0.01805	0.00833	0.00685	0.04577#	0.00519	0.01397	0.01332	0.00932	0.00693	0.01352	0.01670
LA1963	0.02080	0.01546	0.00380	0.01134	0.00887	0.03606#	0.00043	0.02650	0.01649	0.03463#	0.00708	0.00439	0.01274	0.00535	0.00808	0.01791	0.01437
LA1968	0.02360	0.01601	0.00328	0.01277	0.00630	0.03071	0.00957	0.00645	0.00554	0.00674	0*	0.00686	0.00848	0.00574	0.00108	0.01234	0.00972
LA2747	0.00928	0.01758	0.00279	0.00725	0.00470	0.03559#	0.01707	0.01075	0.00480	0.02488	0.00884	0.00660	0.01603	0.00489	0.00858	0.01670	0.01227
LA2748	0.01650	0.00753	0.00116	0.00906	0.00812	0.02709	0.01155	0.00249	0.00168	0.02197	0.00785	0.00718	0.01581	0.00872	0.00934	0.01086	0.01043
LA2750	0.00236	0.00347	n. a.	0.00951	0.00135	0.01874	0.00174	0.00780	0.00169	0.00132	0.00978	0.01038	0.01365	0.00422	0.00178	0.00395	0.00612
LA2753	0.01368	0.01509	0.00566	0.01905	0.00317	0.03379#	0.00675	0.00683	0.01156	0.01143	0.00687	0.00896	0.01723	0.00705	0.00243	0.01316	0.01142
LA2755	0.02284	0.01252	0.00214	0.00768	0.00825	0.03927#	0.02101	0.01304	0.00384	0.00432	0.00414	0.00972	0.00981	0.00626	0.01274	0.01014	0.01173
LA2765	0.01110	0.01311	0.00258	0.01548	0.00683	0.03813#	0.02070	0.00932	0.00858	0.00853	0.01283	0.01048	0.00963	0.00663	0.01190	0.01511	0.01256
LA2773	0.01602	0.01664	0.00650	0.01007	0.00429	0.03837#	0.00980	0.00818	0.01155	0.01710	0.01135	0.00916	0.02453	0.00785	0.01693	0.01690	0.01408
LA2880	0.00851	0.00719	0.00018	0.00214	0.00045	0.02682	0*	0.00305	0.01046	0*	0.00091	0.00616	0.01832	0.00017	0.00538	0.00237	0.00576
LA2931	0.01683	0.01896	0.00281	0.01056	0.01161	0.03377#	0.01420	0.01216	0.01254	0.02222	0.00428	0.00661	0.01849	0.00774	0.00568	0.02270	0.01382
LA2932	0.00152	0.00312	0.00400	0.00620	0.00022	0.01840	0.00044	0.00472	0.00084	0.00622	0*	0.00868	0.01382	0.00471	0.00165	0.00412	0.00492
LA3111	0.00722	0.01321	0.00514	0.01260	0.00879	0.03296#	0.00605	0.00480	0.02106	0.02542	0.00288	0.00384	0.01163	0.00893	0.00072	0.01622	0.01134
LA3784	0.00433	0.00550	0.00978	0.01778	0.00200	0.02567	0.02215	0.00604	0.00768	0.02847	0.01373	0.00456	0.02473	0.00683	0.00193	0.01087	0.01200
LA4107	0.00347	0.00047	0.00107	0.00322	0.00632	0.00276	0.00087	0.00255	0.00168	0*	0*	0.00054	0*	0.00051	0.00197	0.00032	0.00161
LA4108	0.00024	0.00016	0.00179	0.00288	0.00045	0.00585	0.00087	0.00292	0.00168	0.00260	0.00270	0*	0.00053	0.00105	0.00602	0.01208	0.00261
LA4118	0.00721	0.01398	0.00086	0.00432	0.00741	0.02660	0.00043	0.00840	0.01131	0.01620	0.00259	0.00525	0.00576	0*	0.00143	0.00185	0.00710
LA4119	0.00789	0.00914	0.00072	0.00200	0.00162	0.03220	0.00225	0.00554	0.00952	0.00227	0.00428	0.00828	0.02727	0.00034	0.00116	0.00254	0.00731
LA4332	0.01447	0.00763	0.00219	0.00425	0.00677	0.03720#	0.00043	0.00871	0.00979	0.02329	0.00504	0.00636	0.01159	0.00593	0.00473	0.01490	0.01021
Mean	0.01143	0.01021	0.00331	0.00948	0.00573	0.02885	0.00824	0.00817	0.00804	0.01576	0.00543	0.00712	0.01280	0.00527	0.00508	0.01101	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.6: Synonymous nucleotide diversity, π_s, for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.01652	0.00114	0.01386	0.00135	0.01534	0.01380	0.02726	0*	0.00944	0.00616	0.03000	0.02191	0.01904	0.00313	0.01278
LA0458	0.01069	0.01864	0.00397	0.00963	0.00975	0*	0.01748	0*	0*	0.01613	0.00592	0.00221	0.01183	0.00042	0.00762
LA1930	0.02099	0.01514	0.01370	0.00531	0.02782	0.00087	0.01908	0.00796	0.01364	0.02183	0.04840#	0.01957	0.02046	0.00845	0.01737
LA1958	0.02300	0.02187	0.01156	0.00732	0.02863	0.01549	0.03018	0.00634	0.01362	0.01362	0.04494#	0.01349	0.02334	0.00316	0.01833
LA1963	0.02453	0.02433	0.00877	0.00583	0.02757	0.01871	0.02494	0.00551	0.00733	0.01830	0.01193	0.01851	0.02346	0.00591	0.01612
LA1968	0.00488	0.02346	0.00851	0.00677	0.02655	0.02004	0.01637	0.00333	0.00055	0.01927	0.01452	0.01879	0.02425	0.00753	0.01392
LA2747	0.02135	0.02336	0.01275	0.00803	0.00902	0.01517	0.01480	0.00354	0*	0.02183	0.01461	0.00731	0.01616	0.00368	0.01226
LA2748	0.02467	0.02369	0.00836	0.00524	0.03292#	0.02146	0.02187	0.00430	0*	0.01067	0.00994	0.00554	0.02627	0.00490	0.01427
LA2750	0.01785	0.02676	0.00589	0.00085	0.01170	0.00259	0.00779	0.00537	0.00215	0.00133	0.02749	0.00156	0.00751	0.00125	0.00858
LA2753	0.02668	0.01985	0.00745	0.00957	0.02253	0.01963	0.03086	0.00728	0*	0.01708	0.00910	0.02651	0.02542	0.00293	0.01606
LA2755	0.02128	0.02820	0.01124	0.00273	0.00873	0.02385	0.01694	0.00578	0.03076	0.02039	0*	0.02325	0.01863	0.00276	0.01532
LA2765	0.01393	0.03029	0.01099	0.00590	0.02081	0.02022	0.02547	0.01222	0.00164	0.01117	0.01491	0.02820	0.02735	0.00603	0.01637
LA2773	0.01626	0.02157	0.01080	0.00703	0.00967	0.01604	0.02749	0.01230	0*	0.02104	0.01450	0.02024	0.02159	0.00157	0.01429
LA2880	0.01191	0.00028	0.00913	0*	0.01034	0.00482	0.01124	0*	0.00109	0.00940	0.00484	0.00170	0.00099	0*	0.00470
LA2931	0.02072	0.01584	0.01540	0.00744	0.02365	0.01885	0.02928	0.00567	0*	0.02450	0.01423	0.00404	0.02386	0.00245	0.01471
LA2932	0.00074	0.01924	0.00687	0.00029	0.00795	0*	0.00735	0.01063	0*	0.00053	0.01169	0.00205	0.00283	0*	0.00501
LA3111	0.01579	0.02391	0.01225	0.00876	0.01160	0.01096	0.01265	0.00637	0.00208	0.02240	0.00867	0.02724	0.01726	0.00416	0.01315
LA3784	0.02494	0.01753	0.01542	0.00372	0.02676	0.01046	0.01688	0.00479	0.01437	0.02027	0.05828#	0.02171	0.03516#	0.00519	0.01968
LA4107	0.00739	0.00214	0.00019	0.00029	0.01042	0*	0.00178	0.00716	0*	0.00781	0.00802	0.00111	0.00192	0.00159	0.00356
LA4108	0.01222	0.00480	0.00256	0.00058	0.00868	0*	0.00794	0.00677	0*	0.00874	0.00055	0.00099	0.00123	0.00063	0.00398
LA4118	0.00304	0.00973	0.00573	0.00028	0.01557	0.00043	0.00491	0*	0*	0.01306	0.00664	0.00243	0.01372	0.00212	0.00555
LA4119	0.00621	0.01823	0.00021	0.00143	0*	0.01136	0.00957	0.00052	0.00626	0.00864	0*	0.00173	0.00828	0.00136	0.00527
LA4332	0.02070	0.02241	0.00323	0.00451	0.02100	0.01588	n. a.	0.00196	0.00111	0.01988	0.00416	0.01656	0.01837	0.00186	0.01166
Mean	0.01593	0.01793	0.00865	0.00447	0.01683	0.01133	0.01737	0.00512	0.00452	0.01452	0.01580	0.01246	0.01691	0.00309	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.7: Nonsynonymous nucleotide diversity, π_a, for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	0.00321	0.00018	0.00083	0.00382	0.00357	0.00481	0.00027	0.00235	0.00615	0.00549	0.00138	0.00928#	0.00518	0.00072	0.00268	0.00315	0.00332
LA0458	0.00742	0.00044	0.00005	0.00138	0.00149	0.00326	0.00013	0.00248	0.00020	0.00500	0.00647	0.00109	0.00018	0.00010	0.00403	0.00024	0.00212
LA1930	0.00609	0.00213	0.00069	0.00290	0.00596	0.00333	0.00289	0.00367	0.00504	0.00601	0.00376	0.00273	0.00708	0.00236	0.00502	0.00236	0.00388
LA1958	0.00922#	0.00098	0.00134	0.00178	0.00393	0.00247	0.00137	0.00300	0.00663	0.00278	0.00858	0.00424	0.01129#	0.00221	0.00744	0.00424	0.00447
LA1963	0.00996#	0.00086	0.00157	0.00194	0.00297	0.00356	0.00093	0.00996#	0.00583	0.00703	0.00792	0.00368	0.00434	0.00059	0.00597	0.00089	0.00425
LA1968	0.00925#	0.00053	0.00090	0.00193	0.00433	0.00208	0.00173	0.00366	0.00728	0.00276	0.00538	0.00181	0.00582	0.00157	0.00573	0.00206	0.00355
LA2747	0.00565	0.00189	0.00087	0.00299	0.00352	0.00302	0.00027	0.00573	0.00627	0.00397	0.01064#	0.00449	0.00340	0*	0.00419	0.00052	0.00359
LA2748	0.00609	0.00114	0.00181	0.00197	0.00311	0.00145	0.00053	0.00104	0.00449	0.00316	0.01264#	0.00257	0.00337	0.00058	0.00485	0.00142	0.00314
LA2750	0.00495	0.00193	n. a.	0.00222	0.00331	0.00622	0.00420	0.00629	0.00440	0.00252	0.00727	0.00364	0.00571	0.00178	0.00176	0.00174	0.00386
LA2753	0.00899	0.00123	0.00120	0.00553	0.00155	0.00559	0.00301	0.00301	0.00857	0.00303	0.01143#	0.00705	0.00804	0.00173	0.00543	0.00226	0.00485
LA2755	0.00766	0.00080	0.00055	0.00333	0.00520	0.00277	0.00027	0.00480	0.00695	0.00071	0.00447	0.00340	0.00798	0.00005	0.00714	0.00244	0.00366
LA2765	0.00585	0.00188	0.00112	0.00395	0.00377	0.00214	0.00146	0.00407	0.00476	0.00234	0.01136#	0.00669	0.00522	0.00088	0.00517	0.00195	0.00391
LA2773	0.00552	0.00068	0.00053	0.00390	0.00264	0.00406	0.00040	0.00383	0.00644	0.00500	0.01214#	0.00385	0.00539	0.00094	0.00528	0.00029	0.00381
LA2880	0.00561	0.00009	0.00005	0.00126	0.00039	0.00390	0.00027	0.00091	0.00498	0.00299	0.01098#	0.00450	0.00343	0*	0.00051	0.00029	0.00251
LA2931	0.00634	0.00114	0.00160	0.00282	0.00273	0.00256	0.00027	0.00362	0.00741	0.00669	0.00729	0.00229	0.00484	0.00121	0.00464	0.00338	0.00368
LA2932	0.00399	0.00057	0.00058	0.00063	0.00192	0.00312	0.00053	0.00578	0.00436	0.00286	0.00426	0.00185	0.00303	0.00039	0.00195	0.00058	0.00228
LA3111	0.00617	0.00095	0.00025	0.00202	0.00609	0.00216	0.00091	0.00217	0.00356	0.00394	0.00333	0.00130	0.00642	0.00138	0.00128	0.00063	0.00266
LA3784	0.00326	0.00082	0.00214	0.00460	0.00123	0.00236	0.00149	0.00283	0.00922#	0.01159#	0.01565#	0.00558	0.00838	0.00037	0.00563	0.00125	0.00478
LA4107	0.00126	0.00076	0.00076	0.00047	0.00357	0.00133	0.00066	0.00179	0.00081	0.00071	0.00523	0*	0.00069	0.00075	0.01250#	0.00010	0.00196
LA4108	0.00110	0.00056	0.00053	0.00050	0.00045	0.00348	0.00053	0.00217	0.00081	0.00089	0.00713	0.00037	0.00224	0.00059	0.00235	0.00057	0.00152
LA4118	0.00696	0.00070	0.00029	0.00184	0.00131	0.00274	0*	0.00084	0.00544	0.00489	0.00548	0.00310	0.00266	0.00010	0.00043	0.00140	0.00239
LA4119	0.00726	0.00060	0.00010	0.00154	0.00113	0.00383	0.00257	0.00213	0.00475	0.00257	0.00958#	0.00509	0.00520	0.00046	0.00216	0.00005	0.00306
LA4332	0.00916	0.00107	0.00061	0.00138	0.00139	0.00419	0.00119	0.00180	0.00926#	0.00256	0.00470	0.00345	0.00378	0.00021	0.00152	0.00259	0.00305
Mean	0.00613	0.00095	0.00084	0.00238	0.00285	0.00324	0.00113	0.00339	0.00537	0.00389	0.00770	0.00357	0.00494	0.00082	0.00425	0.00150	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.8: Nonsynonymous nucleotide diversity, π_a, for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.00233	0.00013	0.00069	0.00050	0*	0*	0.00021	0*	0.00099	0.00093	0.00237	0.00670	0.00206	0.00059	0.00125
LA0458	0.00103	0.00127	0.00012	0.00198	0.00026	0.00011	0.00088	0.00014	0*	0.00323	0*	0.00218	0.00194	0.00185	0.00107
LA1930	0.00447	0.00405	0.00298	0.00091	0.00149	0.00176	0.00041	0.00262	0.00265	0.00282	0.00226	0.00551	0.00456	0.00140	0.00271
LA1958	0.00138	0.00112	0.00197	0.00182	0.00133	0.00011	0.00051	0.00083	0.00016	0.00362	0.00193	0.00433	0.00427	0.00059	0.00171
LA1963	0.00268	0.00109	0.00148	0.00172	0.00200	0.00033	0.00094	0.00085	0*	0.00296	0.00109	0.00512	0.00520	0.00120	0.00190
LA1968	0.00335	0.00048	0.00240	0.00130	0.00234	0.00044	0.00093	0.00029	0.00033	0.00162	0.00111	0.00505	0.00435	0.00189	0.00185
LA2747	0.00178	0.00108	0.00230	0.00165	0.00013	0.00011	0.00051	0.00042	0*	0.00318	0.00145	0.00246	0.00455	0.00062	0.00145
LA2748	0.00191	0.00207	0.00173	0.00173	0.00272	0*	0.00010	0*	0.00016	0.00105	0.00162	0.00183	0.00348	0.00101	0.00139
LA2750	0.00239	0.00294	0.00256	0.00148	0.00213	0.00109	0.00152	0.00257	0.00181	0.00196	0.00403	0.00243	0.00179	0.00294	0.00226
LA2753	0.00426	0.00257	0.00164	0.00273	0.00366	0.00112	0.00071	0.00160	0.00124	0.00245	0.00348	0.00546	0.00402	0.00213	0.00265
LA2755	0.00303	0.00208	0.00339	0.00173	0.00258	0.00061	0.00050	0.00014	0.00033	0.00286	0.00016	0.00566	0.00323	0.00096	0.00195
LA2765	0.00250	0.00172	0.00316	0.00168	0.00078	0.00065	0.00051	0*	0.00033	0.00144	0.00219	0.00583	0.00322	0.00134	0.00181
LA2773	0.00137	0.00110	0.00148	0.00136	0.00013	0*	0.00069	0.00107	0*	0.00217	0.00130	0.00608	0.00367	0.00071	0.00151
LA2880	0.00108	0.00023	0.00073	0.00009	0.00013	0.00011	0.00010	0*	0*	0.00102	0*	0.00328	0.00011	0.00059	0.00053
LA2931	0.00233	0.00166	0.00211	0.00199	0.00177	0.00011	0.00106	0*	0*	0.00249	0.00357	0.00431	0.00481	0.00076	0.00193
LA2932	0.00028	0.00249	0.00070	0.00056	0.00039	0.00022	0.00099	0.00156	0.00118	0.00145	0.00160	0.00096	0.00117	0.00105	0.00104
LA3111	0.00133	0.00159	0.00180	0.00136	0.00072	0*	0.00050	0.00109	0*	0.00442	0.00229	0.00652	0.00334	0.00081	0.00184
LA3784	0.00313	0.00155	0.00634	0.00102	0.00076	0.00022	0.00103	0.00113	0.00319	0.00332	0.00328	0.00664	0.00653	0.00124	0.00281
LA4107	0.00090	0.00072	0.00160	0.00093	0.00013	0.00043	0.00010	0.00206	0.00016	0.00153	0.00016	0.00180	0.00051	0.00085	0.00085
LA4108	0.00228	0.00059	0.00168	0.00065	0.00013	0.00011	0.00010	0.00014	0*	0.00166	0.00016	0.00083	0.00088	0.00116	0.00074
LA4118	0.00147	0.00040	0.00057	0.00074	0.00142	0.00022	0.00021	0*	0.00016	0.00146	0.00227	0.00213	0.00222	0.00085	0.00101
LA4119	0.00151	0.00168	0.00058	0.00123	0.00026	0.00022	0.00093	0.00014	0*	0.00113	0*	0.00426	0.00095	0.00014	0.00093
LA4332	0.00223	0.00103	0.00042	0.00177	0.00074	0.00022	n. a.	0.00120	0*	0.00282	0.00151	0.00383	0.00370	0.00085	0.00156
Mean	0.00213	0.00146	0.00184	0.00134	0.00113	0.00036	0.00061	0.00078	0.00055	0.00224	0.00164	0.00405	0.00307	0.00111	

Note: "n. a." not available

*: in lower 2.5 % of the density distribution

Table B3.9: Tajima's *D* values for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	-1.465	-0.974	-1.010	1.456	1.398	-1.292	-0.434	-1.488	0.364	-1.382	-1.580	1.499	-0.119	0.113	1.251	1.494	-0.136
LA0458	1.794#	-0.921	-1.668	0.817	-2.202	-0.280	0.424	0.679	1.396	-1.315	-0.841	-1.234	-0.949	-0.297	1.186	1.803#	-0.100
LA1930	-0.057	-0.436	-0.882	-0.940	-0.003	-0.753	-0.756	0.446	0.673	-0.984	-1.381	-0.457	-0.612	0.888	-0.277	0.389	-0.321
LA1958	0.382	0.235	1.335	0.408	0.876	-0.382	-0.129	-0.613	0.521	0.198	-0.898	-0.081	0.084	0.923	0.138	1.388	0.274
LA1963	0.140	-0.131	-0.161	0.023	-0.021	-0.739	-1.581	0.278	-0.825	-0.780	-1.243	-0.177	-0.312	-0.978	-0.973	0.910	-0.411
LA1968	0.933	0.229	-2.217	0.277	0.004	-0.722	-1.316	-0.898	0.916	-1.254	-0.845	-0.029	-0.705	-0.979	0.140	-0.488	-0.435
LA2747	0.858	0.234	0.193	0.442	0.004	-0.446	-0.075	-0.711	-0.946	-0.464	0.008	-0.561	-0.830	-0.863	0.202	1.071	-0.118
LA2748	0.152	-0.661	-0.296	-0.424	1.048	-0.738	-0.198	-1.800	-0.264	-1.261	-1.512	0.278	-0.917	1.746#	-0.261	-0.656	-0.360
LA2750	-1.288	-2.502*	n. a.	-0.636	-2.248	-2.053	-2.406	0.009	-1.423	-2.524*	-2.424*	-1.585	-1.097	-2.150	-1.796	-2.226	-1.757
LA2753	0.741	0.322	-0.171	1.047	0.277	-0.937	-1.186	-0.380	-0.156	-0.682	-1.628	0.879	0.106	0.342	-0.887	-0.187	-0.156
LA2755	0.633	-0.311	-1.260	-1.196	1.474	-0.330	-0.403	0.689	-0.380	-2.044	-1.362	0.828	0.468	-0.302	0.064	1.405	-0.127
LA2765	-0.649	-0.429	0.182	-0.306	-0.809	-1.283	-1.235	-0.255	-1.145	-1.070	-1.830	-0.507	-1.087	-1.525	0.063	-1.418	-0.832
LA2773	0.457	0.356	1.478	0.807	-0.723	-0.321	-0.668	-0.088	0.176	-0.721	0.898	-1.142	0.049	-1.093	1.257	1.084	0.113
LA2880	0.930	0.703	-1.862	2.050#	-2.367	-1.054	-1.989	-0.318	2.113#	-1.405	-1.622	1.424	-0.428	-1.624	0.345	-0.386	-0.343
LA2931	0.907	-0.028	0.888	0.185	-0.478	-0.851	-1.106	-0.752	-0.580	-1.061	-0.310	-0.821	-1.234	-0.192	-0.713	0.693	-0.341
LA2932	1.027	-2.301	1.700	0.852	-2.027	-1.074	-2.244	-0.866	0.171	-1.267	-1.888	-0.529	0.347	-0.350	0.433	-1.485	-0.594
LA3111	0.969	0.508	-1.863	0.318	0.567	-1.147	-0.930	-1.862	-0.199	-1.282	-2.510*	-0.362	-0.250	1.321	-1.117	0.324	-0.470
LA3784	-1.285	0.699	-0.091	0.506	-2.466*	-1.279	-0.466	-1.265	-0.088	-0.350	-1.298	-0.990	-0.822	-0.066	-1.074	1.080	-0.578
LA4107	-1.417	-2.152	-1.898	-1.398	-1.495	-2.168	-2.013	-2.249	-1.818	-1.425	-2.093	-1.786	-1.210	-1.842	-2.459*	-2.601*	-1.876
LA4108	-2.016	-1.542	-2.212	-1.752	-1.869	-1.773	-2.240	-0.448	-2.172	-1.514	-2.332	-1.728	-1.187	-0.285	-0.884	0.960	-1.437
LA4118	1.202	1.076	-2.035	-1.589	-1.957	-0.455	0.044	-0.832	1.717#	-1.772	-1.449	-0.009	0.045	-2.076	-1.496	-1.021	-0.663
LA4119	0.727	-1.407	-2.553*	-1.352	-2.571*	-1.164	-2.674*	-0.880	0.101	-2.540*	-1.721	-0.357	0.080	-2.373	-2.341	-1.606	-1.415
LA4332	1.022	-0.669	-1.279	-1.511	-2.173	0.137	-1.783	-1.212	0.358	-1.717	-2.327	0.653	-0.826	-1.492	-0.862	1.131	-0.784
Mean	0.204	-0.439	-0.713	-0.083	-0.772	-0.918	-1.103	-0.644	-0.065	-1.244	-1.399	-0.295	-0.496	-0.572	-0.438	0.072	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.10: Tajima's *D* values for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	1.261	-1.145	0.025	-0.911	2.660#	0.202	0.865	-1.298	-1.182	-1.173	-0.896	1.685	1.151	-1.010	0.017
LA0458	0.128	1.783#	-1.163	0.352	-0.191	-1.862	-1.474	1.450	-0.662	2.308#	1.359	0.153	1.705	-1.365	0.180
LA1930	-0.813	0.128	-0.136	-1.794	0.471	-0.934	-0.674	0.030	-1.722	1.166	1.309	1.233	-0.163	-1.506	-0.243
LA1958	-0.707	0.857	-0.857	0.632	2.144#	0.299	-0.477	-0.206	0.542	0.274	0.052	-0.049	0.998	-0.645	0.204
LA1963	-0.794	-0.334	-1.105	-0.759	0.272	0.145	0.170	-0.486	-1.695	-0.103	-1.596	-0.177	0.502	-0.253	-0.444
LA1968	0.533	-0.343	-1.405	-0.957	1.328	0.498	-1.149	-0.777	-1.313	-0.731	-0.687	0.078	0.573	-1.141	-0.392
LA2747	1.633	1.272	-0.664	0.048	1.189	1.492	-0.781	-0.656	-1.708	0.513	-0.190	-1.043	0.341	0.210	0.118
LA2748	0.428	0.434	-1.356	-0.577	1.377	2.072#	0.411	-1.002	0.181	-0.874	-0.973	-1.514	1.321	-1.423	-0.107
LA2750	-1.175	-0.744	-2.387	-2.536*	-0.914	-2.423*	-1.480	-1.364	-1.932	-2.443*	-0.017	-2.356	-2.318	-2.692*	-1.770
LA2753	0.828	0.268	-1.458	0.740	1.562	0.068	0.487	-0.503	-0.235	0.440	-1.776	0.351	0.644	-1.258	0.011
LA2755	1.122	1.672	-0.611	-0.910	0.770	2.051#	-0.888	-0.011	2.469#	1.972#	-1.683	0.712	0.618	-1.522	0.411
LA2765	-0.905	0.005	-1.120	-1.081	-0.627	1.441	-1.472	-0.843	-0.443	0.192	-0.548	-0.100	0.044	-1.710	-0.512
LA2773	0.774	0.709	-0.415	0.083	0.750	2.010#	-0.010	-0.167	-0.546	1.024	-0.515	0.700	0.225	-0.320	0.307
LA2880	-0.493	-2.347	-0.600	0.475	1.949#	-1.025	0.464	0.956	-1.049	1.636	0.187	-0.477	-1.841	-0.738	-0.207
LA2931	-0.435	-0.245	-0.325	-0.454	0.868	0.426	0.443	-0.164	0.752	0.718	-1.020	0.175	0.745	-1.896	-0.029
LA2932	-2.287	-0.137	0.032	-1.919	-0.842	-2.166	-1.801	-0.155	-1.623	-1.915	-0.569	-2.054	-2.086	-1.009	-1.324
LA3111	1.364	0.742	-0.570	0.927	-1.094	-0.961	-0.164	-0.888	-0.736	1.566	0.617	1.312	-0.210	-1.414	0.035
LA3784	0.019	-0.304	-0.393	-1.298	-0.612	0.778	-1.139	-1.246	-2.301	1.303	0.782	-0.100	0.915	-1.090	-0.335
LA4107	-1.262	-2.462*	-2.296	-1.840	0.389	-1.968	-1.743	-1.845	-1.425	1.524	-2.186	-1.571	-1.300	-1.745	-1.409
LA4108	-0.075	-1.923	-2.158	-1.921	-0.076	-2.017	-1.262	-1.673	-0.948	2.376#	-2.385	-1.698	-1.223	-1.599	-1.184
LA4118	-0.107	-0.829	-1.564	-2.351	1.503	1.692	-1.481	0.573	-0.192	-0.171	0.049	-0.908	1.305	-1.560	-0.289
LA4119	-0.471	-1.055	-2.583*	-2.607*	-2.376	0.516	-1.803	-0.967	0.361	-0.496	-1.446	-1.355	-1.913	-2.248	-1.317
LA4332	1.282	0.599	-1.823	-0.269	1.808#	0.884	n. a.	-0.673	-0.997	1.392	0.120	0.194	-0.051	-1.798	0.051
Mean	-0.007	-0.148	-1.084	-0.823	0.535	0.053	-0.680	-0.518	-0.713	0.456	-0.522	-0.296	-0.001	-1.293	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.11: Divergence from *S. ochranthum* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean[a]
LA0456	0.14532[#]	0.03792	0.03237	0.05683	0.03349	0.02748	0.02508	0.02619	n. a.	0.04036	0.04495	0.05957	0.05332	0.03541	0.02943	0.02896	0.03861
LA0458	0.14436[#]	0.03867	0.03206	0.05355	0.03206	0.02486	0.02697	0.02439	n. a.	0.03878	0.03861	0.05746	0.04809	0.03550	0.02879	0.02613	0.03670
LA1930	0.15184[#]	0.04049	0.03161	0.05720	0.03225	0.02733	0.03060	0.02812	n. a.	0.03853	0.03822	0.05746	0.05606	0.03521	0.02784	0.02720	0.03848
LA1958	0.14677[#]	0.03880	0.03251	0.05564	0.03201	0.02776	0.02683	0.02598	n. a.	0.03517	0.04257	0.05850	0.05008	0.03569	0.03057	0.02873	0.03771
LA1963	0.14694[#]	0.03852	0.03173	0.05586	0.03320	0.02533	0.02517	0.02945	n. a.	0.03669	0.03911	0.05603	0.05176	0.03496	0.02872	0.02665	0.03727
LA1968	0.14617[#]	0.03809	0.03242	0.05555	0.03311	0.02457	0.02610	0.02714	n. a.	0.03099	0.04079	0.05881	0.05335	0.03477	0.02892	0.02777	0.03719
LA2747	0.14605[#]	0.03853	0.03158	0.05408	0.03353	0.02581	0.02720	0.02414	n. a.	0.03541	0.03861	0.05458	0.04998	0.03453	0.02862	0.02577	0.03644
LA2748	0.15013[#]	0.03790	0.03069	0.05525	0.03341	0.02365	0.02638	0.02198	n. a.	0.03593	0.04238	0.05543	0.05383	0.03470	0.02889	0.02467	0.03663
LA2750	0.14762[#]	0.03795	n. a.	0.05571	0.03662	0.02489	0.02450	0.02675	n. a.	0.03378	0.03931	0.05755	0.04985	0.03507	0.02600	0.02772	0.03748
LA2753	0.14732[#]	0.03854	0.03214	0.05673	0.03472	0.02637	0.02603	0.02691	n. a.	0.03233	0.04208	0.05513	0.05294	0.03478	0.02963	0.02698	0.03736
LA2755	0.15211[#]	0.03766	0.03206	0.05594	0.03390	0.02720	0.02866	0.02591	n. a.	0.02983	0.04020	0.05536	0.05371	0.03484	0.02784	0.02785	0.03716
LA2765	0.14496[#]	0.03897	0.03013	0.05460	0.03210	0.02336	0.02866	0.02827	n. a.	0.03116	0.04184	0.05463	0.05002	0.03488	0.02763	0.02706	0.03659
LA2773	0.14592[#]	0.03812	0.03197	0.05724	0.03387	0.02509	0.02628	0.02801	n. a.	0.03407	0.03901	0.05432	0.05278	0.03459	0.02932	0.02502	0.03695
LA2880	0.14920[#]	0.03902	0.03265	0.05447	0.03382	0.02248	0.02518	0.02516	n. a.	0.03099	0.03851	0.05629	0.05387	0.03460	0.02827	0.02644	0.03642
LA2931	0.14470[#]	0.03908	0.03067	0.05561	0.03408	0.02411	0.02605	0.02417	n. a.	0.03458	0.03891	0.05795	0.05490	0.03450	0.02644	0.02628	0.03699
LA2932	0.14903[#]	0.03755	0.03096	0.05682	0.03569	0.02312	0.02522	0.02617	n. a.	0.03520	0.03663	0.05667	0.05230	0.04293	0.02628	0.02756	0.03745
LA3111	0.14805[#]	0.03905	0.03219	0.05422	0.03292	0.02453	0.02566	0.03058	n. a.	0.03548	0.04366	0.05610	0.05579	0.03550	0.02980	0.02569	0.03780
LA3784	0.15446[#]	0.03990	0.03340	0.05321	0.03096	0.02899	0.03455	0.03189	n. a.	0.04536	0.04208	0.05604	0.05494	0.03030	0.02870	0.02751	0.03916
LA4107	0.14934[#]	0.04040	0.03211	0.05561	0.03527	0.02330	0.02554	0.02680	n. a.	0.03233	0.03723	0.05596	0.04780	0.03498	0.02836	0.02844	0.03660
LA4108	0.14878[#]	0.03978	0.03193	0.05513	0.03471	0.02189	0.02512	0.02712	n. a.	0.03238	0.03842	0.05760	0.05047	0.03472	0.02689	0.02854	0.03675
LA4118	0.15158[#]	0.03941	0.03271	0.05366	0.03471	0.02618	0.02514	0.02485	n. a.	0.03395	0.04010	0.05816	0.05374	0.03410	0.02845	0.02633	0.03716
LA4119	0.14800[#]	0.03904	0.03270	0.05421	0.03488	0.02463	0.02618	0.02641	n. a.	0.03052	0.03849	0.05787	0.05476	0.03501	0.02655	0.02536	0.03693
LA4332	0.14939[#]	0.04045	0.03267	0.05337	0.03368	0.02567	0.02451	0.02624	n. a.	0.03388	0.04248	0.05747	0.05085	0.03454	0.02801	0.02631	0.03709
Mean	0.14818	0.03886	0.03197	0.05524	0.03370	0.02516	0.02659	0.02664	n. a.	0.03468	0.04018	0.05674	0.05240	0.03505	0.02826	0.02691	

Note: "n. a." not available

[#]: in upper 2.5 % of the density distribution

[a]: population means calculated excluding *NtC7*, *TAS14* and *His1*

Table B3.12: Divergence from *S. ochranthum* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean[a]
LA0456	0.05988	0.02630	0.01543*	0.02140	0.03204	0.03266	0.03985	0.04213	0.03889	0.05487	0.03698	0.04397	0.02540	0.02098	0.03315
LA0458	0.06391#	0.02797	0.01495*	0.02123	0.03435	0.03234	0.03706	0.04497	0.03445	0.05619	0.03497	0.04317	0.02288	0.02145	0.03277
LA1930	0.06084	0.02964	0.01570*	0.02303	0.03317	0.03196	0.03645	0.04217	0.03774	0.05599	0.04031	0.04106	0.02470	0.02184	0.03337
LA1958	0.05886	0.02788	0.01496*	0.02137	0.03296	0.03223	0.03849	0.04300	0.03573	0.05395	0.03858	0.04257	0.02431	0.02095	0.03284
LA1963	0.06122	0.02953	0.01505*	0.02053	0.03087	0.03130	0.03715	0.04181	0.03535	0.05575	0.03560	0.04217	0.02411	0.02067	0.03230
LA1968	0.06259#	0.02968	0.01638	0.01952	0.03216	0.03118	0.03862	0.04486	0.03498	0.05404	0.03708	0.04338	0.02427	0.02117	0.03287
LA2747	0.06347#	0.02994	0.01494*	0.02079	0.03478	0.03208	0.03807	0.04347	0.03345	0.05571	0.03759	0.04343	0.02509	0.02203	0.03318
LA2748	0.06139#	0.02828	0.01581*	0.02038	0.03466	0.03079	0.03840	0.04322	0.03485	0.05531	0.03681	0.04397	0.02538	0.02063	0.03296
LA2750	0.06142#	0.03042	0.01616	0.02073	0.03357	0.03082	0.03759	0.04562	0.03600	0.05489	0.04186	0.04543	0.02460	0.02159	0.03379
LA2753	0.06146#	0.02826	0.01546*	0.02070	0.03115	0.03231	0.03746	0.04327	0.03557	0.05374	0.03479	0.04117	0.02435	0.02128	0.03227
LA2755	0.06248#	0.02941	0.01509*	0.02043	0.02886	0.03277	0.03776	0.04310	0.03540	0.05539	0.03542	0.04174	0.02375	0.02149	0.03235
LA2765	0.06316#	0.02905	0.01535*	0.02014	0.03465	0.03183	0.03800	0.04354	0.03317	0.05589	0.03738	0.04286	0.02448	0.02147	0.03291
LA2773	0.06335#	0.02835	0.01532*	0.02096	0.03404	0.03093	0.03948	0.04236	0.03365	0.05513	0.03762	0.04354	0.02446	0.02151	0.03287
LA2880	0.06394#	0.03100	0.01529*	0.01974	0.03282	0.02898	0.04002	0.04339	0.03444	0.05524	0.03563	0.04378	0.02437	0.02092	0.03274
LA2931	0.06298#	0.02761	0.01517*	0.02150	0.03275	0.03224	0.03738	0.04265	0.03484	0.05374	0.03894	0.04395	0.02397	0.02091	0.03274
LA2932	0.06623#	0.03318	0.01459*	0.01962	0.03457	0.03043	0.03884	0.04540	0.03731	0.05440	0.04003	0.04450	0.02436	0.02095	0.03371
LA3111	0.06226#	0.02831	0.01615	0.02060	0.03457	0.03157	0.03956	0.04612	0.03507	0.05461	0.03355	0.04314	0.02378	0.02072	0.03290
LA3784	0.06239#	0.02778	0.01692	0.02364	0.03202	0.03144	0.03789	0.04212	0.03945	0.05867	0.03992	0.04159	0.02605	0.02123	0.03375
LA4107	0.05971	0.03287	0.01577*	0.01958	0.03203	0.03023	0.03379	0.04210	0.03458	0.05509	0.03621	0.04669	0.02314	0.02191	0.03261
LA4108	0.06042	0.03267	0.01589	0.01924	0.03255	0.02999	0.03612	0.04187	0.03490	0.05522	0.03330	0.04645	0.02347	0.02165	0.03256
LA4118	0.06234#	0.02698	0.01538*	0.01999	0.03292	0.03115	0.03734	0.04356	0.03490	0.05558	0.03552	0.04458	0.02526	0.02113	0.03264
LA4119	0.06314#	0.02884	0.01668	0.01966	0.03514	0.02896	0.03750	0.03988	0.03524	0.05727	0.03622	0.04504	0.02507	0.02138	0.03284
LA4332	0.06277#	0.02913	0.01548*	0.01992	0.03365	0.03224	n. a.	0.04325	0.03423	0.05323	0.03554	0.04335	0.02454	0.02098	0.03213
Mean	0.06218#	0.02926	0.01556*	0.02064	0.03306	0.03132	0.03786	0.04321	0.03540	0.05521	0.03695	0.04354	0.02443	0.02125	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution
[a]: population means calculated excluding *CT021*

Table B3.13: Divergence from *S. lycopersicoides* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean[a]
LA0456	0.14979#	0.03915	0.01928	0.06991#	0.03252	0.02395	0.03427	0.03065	0.06565	0.03410	0.04386	0.04717	0.03390	0.03369	n. a.	0.03285	0.03656
LA0458	0.15034#	0.03980	0.01886	0.06788	0.03241	0.02068	0.03593	0.02977	0.06755	0.03268	0.03768	0.04736	0.02821	0.03367	n. a.	0.03112	0.03508
LA1930	0.15845#	0.04149	0.01721*	0.07101#	0.03312	0.02468	0.04433	0.03341	0.07147#	0.04070	0.03729	0.04803	0.04364	0.03274	n. a.	0.03201	0.03844
LA1958	0.15205#	0.03902	0.01906	0.06941#	0.03218	0.02451	0.03586	0.03192	0.06506	0.02688	0.04193	0.04620	0.03111	0.03426	n. a.	0.03335	0.03582
LA1963	0.15263#	0.04084	0.01826	0.06879#	0.03372	0.02303	0.03424	0.03147	0.06209	0.02780	0.03816	0.04328	0.03084	0.03395	n. a.	0.03156	0.03507
LA1968	0.15280#	0.03881	0.01912	0.06717	0.03334	0.02101	0.03417	0.03183	0.06485	0.02214	0.03981	0.04645	0.03439	0.03352	n. a.	0.03242	0.03494
LA2747	0.15144#	0.03907	0.01817*	0.06863#	0.03267	0.02197	0.03589	0.03251	0.05938	0.02934	0.03768	0.04105	0.03017	0.03334	n. a.	0.03063	0.03470
LA2748	0.15612#	0.03874	0.01643*	0.06687	0.03407	0.02107	0.03598	0.03129	0.06334	0.02986	0.04135	0.04259	0.03084	0.03339	n. a.	0.02914	0.03474
LA2750	0.15414#	0.04026	n. a.	0.06961#	0.03672	0.01820	0.03381	0.03364	0.05741	0.02765	0.03865	0.04702	0.03307	0.03395	n. a.	0.03196	0.03705
LA2753	0.15304#	0.03850	0.01876	0.06976#	0.03476	0.02356	0.03661	0.03371	0.06058	0.02347	0.04106	0.04490	0.03839	0.03297	n. a.	0.03114	0.03597
LA2755	0.15759#	0.03911	0.01886	0.06917#	0.03475	0.02350	0.03702	0.03278	0.06343	0.02104	0.03923	0.04521	0.03546	0.03367	n. a.	0.03182	0.03551
LA2765	0.15237#	0.03908	0.01741*	0.06584	0.04025	0.02122	0.03733	0.03337	0.0628	0.02231	0.04119	0.04183	0.03232	0.03335	n. a.	0.03162	0.03516
LA2773	0.15146#	0.03892	0.01837	0.07042#	0.03388	0.02083	0.03558	0.03251	0.05831	0.02520	0.03807	0.04466	0.03421	0.03302	n. a.	0.02989	0.03504
LA2880	0.15326#	0.03946	0.01944	0.06868#	0.03380	0.02107	0.03433	0.02984	0.06226	0.02214	0.03758	0.04553	0.03362	0.03354	n. a.	0.03071	0.03460
LA2931	0.15062#	0.03963	0.01655*	0.06971#	0.03384	0.02265	0.03527	0.03307	0.06000	0.02568	0.03797	0.04387	0.03374	0.03339	n. a.	0.03108	0.03511
LA2932	0.15499#	0.03983	0.01694*	0.07052#	0.03567	0.01985	0.03981	0.03283	0.05854	0.03158	0.03575	0.04610	0.03569	0.03319	n. a.	0.03180	0.03612
LA3111	0.15401#	0.04006	0.01897	0.06802#	0.03349	0.02292	0.03482	0.03590	0.06103	0.03087	0.04235	0.04619	0.03715	0.03389	n. a.	0.02986	0.03650
LA3784	0.15873#	0.04086	0.02028	0.06789	0.03095	0.02584	0.04679	0.03863	0.06300	0.04046	0.04126	0.04681	0.03921	0.02315	n. a.	0.03244	0.03804
LA4107	0.15594#	0.04269	0.01888	0.06935#	0.03531	0.01994	0.04004	0.03205	0.06023	0.02168	0.03594	0.04159	0.02955	0.03368	n. a.	0.03511	0.03506
LA4108	0.15561#	0.04198	0.01861	0.06883#	0.03473	0.02006	0.03962	0.03344	0.06011	0.02363	0.03633	0.04313	0.03039	0.03364	n. a.	0.03325	0.03520
LA4118	0.15795#	0.04035	0.01950	0.06799	0.03456	0.02056	0.03426	0.02921	0.06500	0.02509	0.03894	0.04538	0.03323	0.03350	n. a.	0.03059	0.03486
LA4119	0.15229#	0.03882	0.01949	0.06786	0.03485	0.02138	0.03549	0.03038	0.06290	0.02438	0.03755	0.04445	0.03473	0.03395	n. a.	0.02921	0.03481
LA4332	0.15617#	0.04248	0.01947	0.06706	0.03371	0.02112	0.03373	0.03248	0.06324	0.02780	0.04145	0.04430	0.03527	0.03355	n. a.	0.03036	0.03560
Mean	0.15399#	0.03995	0.01854	0.06871#	0.03414	0.02190	0.03675	0.03246	0.06253	0.02767	0.03918	0.04492	0.03388	0.03309	n. a.	0.03147	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution
[a]: population means calculated excluding *NtC7*, *TAS14* and *His1*

Table B3.14: Divergence from *S. lycopersicoides* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean[a]
LA0456	n. a.	0.02838	0.01832	0.03041	0.03110	0.03592	0.02851	0.03245	0.04353	0.04940	0.03414	0.03534	0.01931	0.02420	0.03162
LA0458	n. a.	0.02935	0.01740*	0.03080	0.03162	0.03560	0.02824	0.03519	0.04062	0.05077	0.03358	0.03589	0.01703*	0.02478	0.03161
LA1930	n. a.	0.03181	0.01900	0.03213	0.03317	0.03501	0.02708	0.03211	0.04216	0.04951	0.03742	0.03433	0.01881	0.02495	0.03211
LA1958	n. a.	0.02894	0.01756*	0.03081	0.03131	0.03518	0.02846	0.03262	0.04150	0.04900	0.03514	0.03424	0.01929	0.02433	0.03141
LA1963	n. a.	0.03005	0.01856	0.03031	0.02856	0.03453	0.02733	0.03251	0.04057	0.05045	0.03418	0.03390	0.01892	0.02379	0.03105
LA1968	n. a.	0.02995	0.01951	0.02903	0.03172	0.03544	0.03073	0.03440	0.04103	0.04895	0.03566	0.03582	0.01906	0.02426	0.03197
LA2747	n. a.	0.03020	0.01908	0.03034	0.02853	0.03651	0.02965	0.03304	0.03968	0.04967	0.03509	0.03556	0.01943	0.02486	0.03166
LA2748	n. a.	0.02904	0.01799*	0.02982	0.03026	0.03336	0.02990	0.03270	0.04123	0.04994	0.03419	0.03633	0.02004	0.02382	0.03143
LA2750	n. a.	0.03108	0.02090	0.03037	0.03036	0.03413	0.03021	0.03485	0.04180	0.04893	0.04142	0.03788	0.01963	0.02482	0.03280
LA2753	n. a.	0.02971	0.01822	0.03005	0.03053	0.03616	0.03080	0.03258	0.04221	0.04857	0.03340	0.03214	0.01884	0.02452	0.03136
LA2755	n. a.	0.02988	0.01774*	0.03008	0.03190	0.03696	0.02872	0.03286	0.04130	0.05022	0.03270	0.03397	0.01819	0.02476	0.03148
LA2765	n. a.	0.02951	0.01905	0.02981	0.02575	0.03514	0.03088	0.03359	0.04055	0.05029	0.03486	0.03499	0.01910	0.02477	0.03141
LA2773	n. a.	0.02888	0.01856	0.03060	0.02901	0.03441	0.02993	0.03250	0.03956	0.04912	0.03514	0.03553	0.01904	0.02477	0.03131
LA2880	n. a.	0.02990	0.01729*	0.02938	0.03753	0.03095	0.02983	0.03293	0.04199	0.05044	0.03295	0.03598	0.02060	0.02495	0.03190
LA2931	n. a.	0.02910	0.01859	0.03093	0.03027	0.03537	0.02851	0.03238	0.04107	0.04905	0.03632	0.03560	0.01944	0.02416	0.03160
LA2932	n. a.	0.03258	0.01938	0.02928	0.03132	0.03369	0.02989	0.03439	0.04265	0.04844	0.03850	0.03693	0.01952	0.02420	0.03237
LA3111	n. a.	0.02966	0.01938	0.03041	0.03142	0.03534	0.03111	0.03698	0.04090	0.04967	0.03248	0.03514	0.01795*	0.02399	0.03188
LA3784	n. a.	0.02962	0.02108	0.03298	0.03107	0.03489	0.02863	0.03301	0.04469	0.05304	0.03847	0.03482	0.02099	0.02437	0.03290
LA4107	n. a.	0.03178	0.01820	0.02921	0.03120	0.03349	0.02932	0.03425	0.04062	0.04807	0.03351	0.03911	0.01882	0.02523	0.03175
LA4108	n. a.	0.03157	0.01841	0.02886	0.03148	0.03325	0.03057	0.03321	0.04019	0.04960	0.03058	0.03878	0.01862	0.02495	0.03154
LA4118	n. a.	0.02789	0.01750*	0.02957	0.03574	0.03389	0.02945	0.03313	0.04126	0.05073	0.03416	0.03689	0.02117	0.02444	0.03199
LA4119	n. a.	0.02904	0.01487*	0.02934	0.04052	0.03129	0.03081	0.02908	0.04173	0.05161	0.03202	0.03687	0.02110	0.02488	0.03178
LA4332	n. a.	0.02905	0.01773*	0.02966	0.03282	0.03612	n. a.	0.03283	0.04035	0.04908	0.03416	0.03567	0.01895	0.02419	0.03172
Mean	n. a.	0.02987	0.01845	0.03018	0.03162	0.03464	0.02948	0.03320	0.04136	0.04976	0.03479	0.03573	0.01930	0.02452	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
[a]: population means calculated excluding *CT021*

Table B3.15: Divergence from *S. lycopersicum* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	0.02253	0.01000	0.01361	0.02306	0.02125	0.01196	0.00942	0.01556	0.02672	0.01935	0.01962	0.03825#	0.01826	0.01612	0.02484	0.01656	0.01919
LA0458	0.02463	0.00973	0.01326	0.02201	0.01838	0.01251	0.00873	0.01474	0.02204	0.02254	0.01346	0.03752#	0.01787	0.01540	0.02420	0.01532	0.01827
LA1930	0.02200	0.01206	0.01340	0.02194	0.01891	0.01431	0.01432	0.02006	0.03016	0.02657	0.01250	0.03902#	0.02341	0.01557	0.02325	0.01663	0.02026
LA1958	0.02619	0.01089	0.01300	0.02080	0.01987	0.01369	0.01181	0.01876	0.02180	0.02087	0.01769	0.03837#	0.01730	0.01716	0.02536	0.01665	0.01939
LA1963	0.02553	0.00924	0.01294	0.02230	0.01987	0.01135	0.00961	0.02165	0.02700	0.02121	0.01394	0.03582	0.01894	0.01655	0.02410	0.01358	0.01898
LA1968	0.02327	0.01085	0.01310	0.02420	0.02022	0.01204	0.01081	0.01628	0.02670	0.01763	0.01558	0.03732#	0.01885	0.01553	0.02440	0.01736	0.01901
LA2747	0.02507	0.01293	0.01326	0.02188	0.01929	0.01332	0.01219	0.01787	0.02419	0.02058	0.01346	0.03609	0.01612	0.01493	0.02437	0.01464	0.01876
LA2748	0.02323	0.01122	0.01293	0.02154	0.02028	0.01072	0.01114	0.01577	0.02328	0.01902	0.01654	0.03581	0.02126	0.01566	0.02272	0.01478	0.01849
LA2750	0.02631	0.00903	n. a.	0.02409	0.02257	0.01275	0.00916	0.01736	0.02770	0.01606	0.01442	0.03794#	0.01689	0.01620	0.02138	0.01476	0.01911
LA2753	0.02494	0.01172	0.01310	0.02259	0.02058	0.01201	0.01178	0.02176	0.02388	0.01931	0.01683	0.03805#	0.01756	0.01582	0.02346	0.01391	0.01921
LA2755	0.02551	0.00875	0.01346	0.01940	0.02095	0.01289	0.01347	0.01700	0.02702	0.01803	0.01500	0.03608	0.01917	0.01533	0.02325	0.01617	0.01884
LA2765	0.02212	0.01218	0.01370	0.02102	0.01857	0.01214	0.01342	0.01813	0.02590	0.01699	0.01129	0.03465	0.01785	0.01586	0.02302	0.01498	0.01824
LA2773	0.02312	0.01087	0.01332	0.02297	0.01978	0.01147	0.01121	0.02042	0.02418	0.01873	0.01385	0.03354	0.01925	0.01550	0.02471	0.01546	0.01865
LA2880	0.02579	0.01143	0.01402	0.01977	0.01956	0.01381	0.00944	0.01555	0.02658	0.01913	0.01337	0.03714#	0.02464	0.01553	0.02392	0.01398	0.01898
LA2931	0.02269	0.01201	0.01229	0.02272	0.02026	0.01074	0.01060	0.01855	0.02392	0.02058	0.01375	0.03455	0.01757	0.01604	0.02177	0.01571	0.01836
LA2932	0.02484	0.00867	0.01315	0.02512	0.02224	0.00999	0.00959	0.01612	0.02757	0.01779	0.01154	0.03688#	0.01835	0.01523	0.02166	0.01514	0.01837
LA3111	0.02484	0.00996	0.01284	0.02221	0.02075	0.00999	0.01006	0.02736	0.02593	0.02076	0.01706	0.03615	0.01845	0.01598	0.02524	0.01452	0.01951
LA3784	0.02442	0.01253	0.01455	0.02188	0.01739	0.01442	0.01433	0.02577	0.02630	0.02694	0.01702	0.03520	0.02194	0.01102	0.02413	0.01702	0.02030
LA4107	0.02571	0.01103	0.01418	0.02320	0.02181	0.01098	0.00987	0.01718	0.02700	0.01659	0.01212	0.03364	0.01529	0.01612	0.02368	0.01644	0.01843
LA4108	0.02593	0.01078	0.01356	0.02274	0.02119	0.01246	0.00950	0.01770	0.02442	0.01775	0.01327	0.03505	0.01779	0.01578	0.02228	0.01541	0.01848
LA4118	0.02758	0.01015	0.01405	0.01728	0.01972	0.01522	0.00940	0.01476	0.02730	0.02006	0.01490	0.03763#	0.01840	0.01573	0.02392	0.01422	0.01877
LA4119	0.02595	0.01338	0.01410	0.01744	0.02047	0.01487	0.01051	0.01678	0.02725	0.01867	0.01298	0.03811#	0.02057	0.01578	0.02334	0.01307	0.01895
LA4332	0.02587	0.01202	0.01397	0.01810	0.01951	0.01375	0.00892	0.01683	0.02599	0.01965	0.01721	0.03929#	0.01929	0.01578	0.02339	0.01515	0.01905
Mean	0.02470	0.01093	0.01345	0.02166	0.02015	0.01250	0.01084	0.01835	0.02578	0.01977	0.01467	0.03661#	0.01891	0.01559	0.02358	0.01528	

Note: "n. a." not available

#: in upper 2.5 % of the density distribution

Table B3.16: Divergence from *S. lycopersicum* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.02971	0.01037	0.00960	0.00814	0.02135	0.01661	0.01666	0.01748	0.01409	0.02103	0.02515	0.01698	0.01305	0.01025	0.01646
LA0458	0.03767#	0.01280	0.00920	0.00819	0.02117	0.01746	0.02102	0.01845	0.02233	0.02091	0.02368	0.01888	0.01190	0.01074	0.01817
LA1930	0.02343	0.01399	0.01028	0.01012	0.01769	0.01645	0.01934	0.02027	0.01231	0.01809	0.02921	0.01723	0.01348	0.00956	0.01653
LA1958	0.03590	0.01193	0.00840	0.00859	0.02249	0.01685	0.01852	0.01598	0.02150	0.01950	0.02715	0.01708	0.01313	0.01037	0.01767
LA1963	0.03615	0.01382	0.00927	0.00731*	0.02078	0.01532	0.01640	0.01620	0.02058	0.01940	0.02427	0.01604	0.01413	0.01022	0.01714
LA1968	0.03839#	0.01359	0.01037	0.00633*	0.01965	0.01621	0.01535	0.01821	0.02303	0.01591	0.02666	0.01693	0.01414	0.01060	0.01753
LA2747	0.03699#	0.01435	0.00955	0.00742*	0.02364	0.01725	0.02017	0.01652	0.02146	0.01817	0.02313	0.01889	0.01432	0.01151	0.01810
LA2748	0.03386	0.01296	0.01008	0.00722*	0.02269	0.01410	0.01875	0.01622	0.02328	0.02083	0.02453	0.01830	0.01436	0.00985	0.01765
LA2750	0.03532	0.01382	0.01223	0.00762*	0.02330	0.01584	0.02078	0.01969	0.02379	0.01957	0.03254	0.01886	0.01275	0.01081	0.01907
LA2753	0.03595	0.01250	0.00971	0.00790*	0.02049	0.01645	0.01557	0.01743	0.02407	0.01785	0.02296	0.01617	0.01288	0.01059	0.01718
LA2755	0.03715#	0.01341	0.00844	0.00710*	0.01794	0.01681	0.01800	0.01782	0.01893	0.02038	0.02311	0.01644	0.01351	0.01073	0.01713
LA2765	0.03648#	0.01350	0.00963	0.00702*	0.02497	0.01517	0.01758	0.01721	0.01820	0.02101	0.02470	0.01691	0.01302	0.01058	0.01757
LA2773	0.03441	0.01319	0.00971	0.00754*	0.02075	0.01481	0.01946	0.01679	0.02131	0.01817	0.02530	0.01788	0.01438	0.01093	0.01747
LA2880	0.03539	0.01502	0.00950	0.00663*	0.01136	0.01076	0.01741	0.01734	0.02372	0.02070	0.02448	0.02035	0.01529	0.01081	0.01705
LA2931	0.03461	0.01175	0.00968	0.00727*	0.01944	0.01559	0.01683	0.01617	0.02279	0.01762	0.02681	0.01856	0.01530	0.01032	0.01734
LA2932	0.04020#	0.01616	0.01004	0.00646*	0.02410	0.01556	0.02070	0.01929	0.02484	0.01903	0.02893	0.01838	0.01220	0.01019	0.01901
LA3111	0.03159	0.01247	0.01020	0.00764*	0.02468	0.01651	0.01649	0.02048	0.02283	0.01866	0.02476	0.01701	0.01153	0.01009	0.01750
LA3784	0.03574	0.01219	0.01148	0.01086	0.01859	0.01465	0.01904	0.01626	0.01188	0.02050	0.03282	0.01736	0.01443	0.00922	0.01750
LA4107	0.03536	0.01538	0.00997	0.00646*	0.02274	0.01540	0.01612	0.01932	0.02235	0.02013	0.02385	0.02117	0.01300	0.01117	0.01803
LA4108	0.03539	0.01513	0.01016	0.00611*	0.02301	0.01517	0.01717	0.01818	0.02256	0.01990	0.02091	0.01997	0.01284	0.01099	0.01768
LA4118	0.03534	0.01204	0.00961	0.00685*	0.01692	0.01408	0.01735	0.01700	0.02467	0.02034	0.02454	0.02008	0.01569	0.01043	0.01750
LA4119	0.03578	0.01312	0.00940	0.00710*	0.01424	0.01373	0.01789	0.01459	0.02415	0.02238	0.02502	0.02018	0.01573	0.01016	0.01739
LA4332	0.03495	0.01389	0.00980	0.00688*	0.02078	0.01725	n. a.	0.01717	0.02277	0.01824	0.02364	0.01735	0.01403	0.01021	0.01746
Mean	0.03503	0.01336	0.00984	0.00751*	0.02056	0.01557	0.01803	0.01757	0.02119	0.01949	0.02557	0.01813	0.01370	0.01045	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.17: Synonymous divergence from *S. ochranthum* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean[a]
LA0456	0.27336#	0.04845	0.04640	0.06947	0.06466	0.06852	0.04898	0.04328	n. a.	0.12400#	0.07078	0.11020	0.09572	0.06464	0.03968	0.03712	0.06863
LA0458	0.27233#	0.05685	0.04640	0.05906	0.07821	0.06754	0.05682	0.03939	n. a.	0.11639	0.07068	0.10706	0.09211	0.06656	0.03582	0.04779	0.06960
LA1930	0.26937#	0.05299	0.05072	0.07814	0.05670	0.08140	0.05954	0.02543*	n. a.	0.10675	0.07966	0.10026	0.09490	0.06609	0.03598	0.03918	0.06860
LA1958	0.27914#	0.05677	0.04686	0.06600	0.06431	0.08140	0.05902	0.03455	n. a.	0.10932	0.07366	0.10693	0.09363	0.06418	0.03997	0.03928	0.06892
LA1963	0.28556#	0.06395	0.04619	0.06475	0.06261	0.07078	0.04923	0.04299	n. a.	0.10691	0.07499	0.10489	0.09632	0.06540	0.03756	0.04203	0.06854
LA1968	0.27940#	0.05523	0.04839	0.06291	0.06116	0.06486	0.05490	0.02751	n. a.	0.08644	0.07076	0.10088	0.09888	0.06175	0.03646	0.04295	0.06436
LA2747	0.27010#	0.05512	0.04630	0.05890	0.06027	0.07428	0.06000	0.02962	n. a.	0.10475	0.07688	0.09835	0.09788	0.06167	0.03844	0.04070	0.06652
LA2748	0.27359#	0.04960	0.04239	0.06141	0.06309	0.06258	0.05388	0.02398*	n. a.	0.11041	0.07422	0.09709	0.09938	0.06524	0.03900	0.04754	0.06545
LA2750	0.27497#	0.04686	n. a.	0.06129	0.06292	0.06717	0.05013	0.02770	n. a.	0.09954	0.07581	0.11473	0.10346	0.06484	0.03683	0.04249	0.06808
LA2753	0.27736#	0.05142	0.04864	0.0685	0.05818	0.07277	0.05332	0.02647*	n. a.	0.08855	0.07437	0.10874	0.08507	0.06425	0.03717	0.03948	0.06460
LA2755	0.29527#	0.05549	0.04992	0.07115	0.06019	0.08402	0.06981	0.03287	n. a.	0.08394	0.07314	0.10731	0.09972	0.06167	0.03782	0.04078	0.06846
LA2765	0.26898#	0.05025	0.04371	0.06411	0.05950	0.06046	0.06576	0.03055	n. a.	0.08839	0.08162	0.10136	0.09567	0.06482	0.04061	0.04107	0.06517
LA2773	0.27385#	0.05455	0.04825	0.06068	0.06200	0.06878	0.05192	0.03319	n. a.	0.09294	0.07753	0.10453	0.09789	0.06396	0.04310	0.04528	0.06627
LA2880	0.27828#	0.05139	0.05103	0.05947	0.07421	0.05180	0.04900	0.03121	n. a.	0.08132	0.07111	0.10302	0.10263	0.06433	0.03725	0.03769	0.06371
LA2931	0.28516#	0.05526	0.04389	0.06252	0.06536	0.06528	0.05681	0.03169	n. a.	0.09517	0.07312	0.09843	0.10083	0.06342	0.03790	0.04483	0.06589
LA2932	0.27428#	0.04658	0.04464	0.06163	0.06273	0.06147	0.04984	0.02473*	n. a.	0.09720	0.07069	0.10631	0.10357	0.06426	0.03682	0.04397	0.06443
LA3111	0.27166#	0.05318	0.05011	0.06379	0.06232	0.06992	0.05315	0.03380	n. a.	0.10649	0.07658	0.10598	0.10545	0.06321	0.03617	0.04252	0.06819
LA3784	0.28974#	0.04959	0.04989	0.06853	0.06797	0.08678	0.08110	0.01501*	n. a.	0.13328#	0.07831	0.10951	0.09967	0.08129	0.03659	0.04116	0.07401
LA4107	0.27666#	0.05337	0.04723	0.05629	0.06585	0.06649	0.04946	0.02282*	n. a.	0.09830	0.07060	0.09441	0.09333	0.06432	0.04583	0.05303	0.06427
LA4108	0.27378#	0.05345	0.04697	0.05523	0.06283	0.05424	0.04946	0.02329*	n. a.	0.09830	0.07198	0.09415	0.09362	0.06487	0.04212	0.04787	0.06279
LA4118	0.28285#	0.05440	0.05064	0.05958	0.07196	0.07597	0.04923	0.03602	n. a.	0.09437	0.07223	0.10464	0.09697	0.06424	0.03617	0.03640	0.06667
LA4119	0.27955#	0.04779	0.05092	0.06202	0.07070	0.06453	0.05187	0.03311	n. a.	0.08275	0.07294	0.08358	0.10617	0.06424	0.02962	0.03815	0.06375
LA4332	0.29136#	0.05339	0.04950	0.05874	0.07119	0.06976	0.04922	0.02753	n. a.	0.09948	0.07363	0.09818	0.10247	0.06338	0.03950	0.04162	0.06601
Mean	0.27811#	0.05287	0.04768	0.06322	0.06474	0.06917	0.05532	0.03029	n. a.	0.10022	0.07414	0.10263	0.09806	0.06490	0.03810	0.04230	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution
[a]: population means calculated excluding *NtC7*, *TAS14* and *His1*

Table B3.18: Synonymous divergence from *S. ochranthum* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean[a]
LA0456	0.11770	0.08982	0.04270	0.02773	0.08137	0.06987	0.09735	0.08633	0.05379	0.07255	0.05842	0.07450	0.06364	0.03725	0.06579
LA0458	0.11872#	0.09776	0.04114	0.03035	0.07837	0.07568	0.08156	0.08634	0.02772	0.09473	0.05945	0.07972	0.05569	0.03678	0.06502
LA1930	0.12344#	0.09557	0.03552	0.02363*	0.08322	0.05364	0.08633	0.09203	0.05512	0.08745	0.06111	0.06726	0.05988	0.03252	0.06410
LA1958	0.11457	0.09542	0.03552	0.02603*	0.08277	0.06454	0.08776	0.09003	0.03160	0.07498	0.06441	0.07288	0.05868	0.03725	0.06322
LA1963	0.12027#	0.10269	0.03952	0.03004	0.07952	0.05999	0.08709	0.08950	0.03049	0.08918	0.05778	0.07442	0.05622	0.03558	0.06400
LA1968	0.11203	0.10553	0.04102	0.02743*	0.09147	0.06265	0.09013	0.08835	0.02799	0.08129	0.06442	0.07874	0.05646	0.03725	0.06559
LA2747	0.12239#	0.10456	0.03649	0.02888	0.08122	0.05507	0.07897	0.08835	0.02772	0.08285	0.05833	0.08071	0.05765	0.03735	0.06293
LA2748	0.11903#	0.09481	0.04295	0.03097	0.08113	0.05745	0.08794	0.08892	0.02772	0.07271	0.05555	0.08266	0.06189	0.03431	0.06300
LA2750	0.11677	0.10062	0.03706	0.02728*	0.07750	0.04429	0.07992	0.09759	0.02883	0.09231	0.06439	0.08522	0.06033	0.03704	0.06403
LA2753	0.12157#	0.09407	0.04016	0.02560*	0.07817	0.05966	0.08281	0.09180	0.02772	0.07825	0.05721	0.07830	0.05961	0.03558	0.06223
LA2755	0.11631	0.10057	0.03754	0.02728*	0.06785	0.06073	0.08452	0.09092	0.04102	0.08718	0.05529	0.07490	0.05663	0.03861	0.06331
LA2765	0.12920#	0.09914	0.03692	0.03123	0.08165	0.05674	0.09200	0.09462	0.02744	0.08690	0.05693	0.08150	0.06085	0.03870	0.06497
LA2773	0.12138#	0.09604	0.03854	0.02804	0.08103	0.05633	0.08572	0.09492	0.02772	0.08370	0.05834	0.07799	0.05732	0.03745	0.06332
LA2880	0.12898#	0.10707	0.04162	0.03466	0.07593	0.04598	0.09191	0.08633	0.02827	0.09474	0.05834	0.08090	0.05867	0.04163	0.06508
LA2931	0.12357#	0.09243	0.04037	0.02728*	0.07924	0.06194	0.08813	0.08979	0.02772	0.08499	0.06118	0.07791	0.05479	0.03611	0.06322
LA2932	0.12472#	0.11235	0.03328	0.02732*	0.07835	0.04340	0.08320	0.10188	0.02772	0.09205	0.06747	0.08529	0.06107	0.03663	0.06539
LA3111	0.11782	0.09773	0.04168	0.03235	0.08143	0.07527	0.09220	0.10014	0.02883	0.08211	0.05497	0.07642	0.05690	0.03568	0.06582
LA3784	0.12998#	0.09384	0.03394	0.02158*	0.08483	0.05116	0.08876	0.08931	0.06948	0.08047	0.06691	0.07037	0.06293	0.02891	0.06481
LA4107	0.10642	0.11408	0.04287	0.02732*	0.07552	0.04342	0.07466	0.10822	0.02772	0.09401	0.05975	0.08773	0.05709	0.03745	0.06537
LA4108	0.10807	0.11363	0.04335	0.02748	0.07430	0.04340	0.08573	0.10506	0.02772	0.09442	0.05558	0.08561	0.05809	0.03699	0.06549
LA4118	0.12377#	0.09017	0.04104	0.03481	0.06670	0.04362	0.08824	0.08633	0.02772	0.09386	0.06047	0.07738	0.06192	0.03767	0.06230
LA4119	0.13292#	0.09705	0.04717	0.03557	0.07107	0.05159	0.08929	0.08660	0.03188	0.09315	0.05556	0.08108	0.06063	0.03480	0.06426
LA4332	0.12280#	0.10045	0.04094	0.02949	0.07493	0.06683	n. a.	0.08750	0.02827	0.08779	0.05774	0.07835	0.05597	0.03694	0.06210
Mean	0.12054#	0.09980	0.03962	0.02880	0.07859	0.05666	0.08656	0.09221	0.03305	0.08616	0.05955	0.07869	0.05882	0.03646	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution
[a]: population means calculated excluding *CT021*

Table B3.19: Synonymous divergence from *S. lycopersicoides* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His*	*TPP*	Mean[a]
LA0456	0.28546#	0.06338	0.02275*	0.07751	0.04165	0.07562	0.08716	0.03415	0.22913	0.07265	0.02252*	0.06998	0.02980	0.05659	n. a.	0.03540	0.05301
LA0458	0.28449#	0.07052	0.02275*	0.06798	0.05521	0.07050	0.09370	0.03179	0.23737#	0.06764	0.02249*	0.06671	0.02632	0.05370	n. a.	0.05450	0.05414
LA1930	0.29092#	0.06577	0.02699	0.08711	0.04437	0.09063	0.09028	0.04689	0.23224#	0.05673	0.03105	0.05983	0.04587	0.05342	n. a.	0.03605	0.05654
LA1958	0.29348#	0.06960	0.02320	0.07301	0.04135	0.08974	0.09325	0.02750	0.22861	0.06338	0.02524	0.06663	0.02751	0.05444	n. a.	0.03922	0.05339
LA1963	0.29556#	0.07772	0.02255*	0.06812	0.03966	0.08353	0.08720	0.04690	0.23727#	0.05819	0.02657	0.06455	0.03053	0.06009	n. a.	0.04448	0.05462
LA1968	0.29043#	0.06827	0.02463	0.07021	0.03820	0.07154	0.09356	0.03287	0.22740	0.03734	0.02252*	0.06053	0.03366	0.05070	n. a.	0.04089	0.04961
LA2747	0.28215#	0.06948	0.02266*	0.06623	0.03734	0.07923	0.09500	0.03744	0.22672	0.05567	0.02837	0.05804	0.03342	0.05053	n. a.	0.04168	0.05193
LA2748	0.28370#	0.06549	0.01879*	0.06532	0.04017	0.07383	0.09478	0.03254	0.22578	0.06143	0.02654	0.05809	0.03444	0.05461	n. a.	0.04474	0.05160
LA2750	0.28732#	0.06793	n. a.	0.07014	0.03978	0.06125	0.08788	0.02859	0.24856#	0.05012	0.02787	0.07442	0.03842	0.05848	n. a.	0.04039	0.05377
LA2753	0.28709#	0.06699	0.02493	0.07492	0.03512	0.08566	0.08001	0.03499	0.23322#	0.03937	0.02607	0.06847	0.03123	0.05775	n. a.	0.03831	0.05106
LA2755	0.30773#	0.07187	0.02619	0.08006	0.03711	0.08817	0.10241	0.04042	0.22528	0.03510	0.02478	0.06697	0.03447	0.05190	n. a.	0.04114	0.05389
LA2765	0.28099#	0.06564	0.02012*	0.07052	0.03704	0.07435	0.10137	0.03009	0.24809#	0.03930	0.03135	0.06103	0.03189	0.05668	n. a.	0.04003	0.05072
LA2773	0.28585#	0.06855	0.02456	0.06729	0.03909	0.07125	0.09087	0.03269	0.23031#	0.04387	0.03520	0.06422	0.04053	0.05488	n. a.	0.04605	0.05223
LA2880	0.29063#	0.06617	0.02728	0.06836	0.05116	0.06907	0.08698	0.02307*	0.23423#	0.03254	0.02293*	0.06267	0.03684	0.06004	n. a.	0.03587	0.04946
LA2931	0.29674#	0.06933	0.02000*	0.06234	0.04222	0.08196	0.09508	0.03880	0.23241#	0.04629	0.02477	0.05809	0.03702	0.05383	n. a.	0.04473	0.05188
LA2932	0.28614#	0.06750	0.02103*	0.07056	0.03976	0.06889	0.08815	0.02437	0.24645#	0.04720	0.02250*	0.06604	0.03854	0.05586	n. a.	0.04185	0.05017
LA3111	0.28352#	0.06803	0.02638	0.07147	0.03928	0.08590	0.09182	0.04301	0.24277#	0.05799	0.02541	0.06563	0.03987	0.05345	n. a.	0.03831	0.05435
LA3784	0.30142#	0.06281	0.02613	0.07806	0.04527	0.09545	0.10750	0.04843	0.22962	0.08454	0.03076	0.06934	0.04852	0.07162	n. a.	0.04037	0.06222
LA4107	0.28867#	0.07427	0.02338	0.06483	0.04284	0.07338	0.08716	0.03268	0.24719#	0.04917	0.02247*	0.05407	0.02667	0.06003	n. a.	0.06311	0.05185
LA4108	0.28588#	0.07433	0.02316*	0.06395	0.03984	0.06915	0.08716	0.03359	0.24741#	0.04917	0.02381	0.05380	0.02694	0.06058	n. a.	0.05325	0.05067
LA4118	0.29493#	0.07044	0.02691	0.06852	0.04780	0.07149	0.08696	0.02610	0.23834#	0.04549	0.02389	0.06429	0.03041	0.05995	n. a.	0.03457	0.05052
LA4119	0.28499#	0.06444	0.02714	0.05736	0.04699	0.07281	0.09103	0.02567	0.24232#	0.03360	0.02471	0.03960	0.03724	0.05995	n. a.	0.03627	0.04745
LA4332	0.30212#	0.07294	0.02581	0.06659	0.04815	0.07085	0.08695	0.03519	0.23022#	0.05017	0.02523	0.05783	0.03629	0.05669	n. a.	0.03950	0.05171
Mean	0.29001#	0.06876	0.02397	0.07002	0.04215	0.07714	0.09158	0.03425	0.23569#	0.05117	0.02596	0.06221	0.03463	0.05677	n. a.	0.04220	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution
[a]: population means calculated excluding *NtC7*, *TAS14* and *His1*

Table B3.20: Synonymous divergence from *S. lycopersicoides* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean[a]
LA0456	n. a.	0.09912	0.05697	0.04314	0.03499	0.05906	0.08705	0.04317	0.08150	0.06932	0.05842	0.07820	0.04112	0.03683	0.06068
LA0458	n. a.	0.10600	0.05549	0.04576	0.05035	0.06487	0.07829	0.04318	0.05543	0.08913	0.05945	0.07999	0.03430	0.03678	0.06146
LA1930	n. a.	0.10538	0.05380	0.03917	0.04831	0.04326	0.07486	0.04890	0.08322	0.08286	0.06111	0.07429	0.03829	0.03251	0.06046
LA1958	n. a.	0.10190	0.05380	0.04144	0.04373	0.05361	0.08452	0.04690	0.05931	0.07261	0.06441	0.07623	0.03974	0.03725	0.05965
LA1963	n. a.	0.10829	0.05412	0.04607	0.04414	0.04916	0.08238	0.04634	0.05820	0.08541	0.05778	0.07897	0.03781	0.03464	0.06025
LA1968	n. a.	0.10982	0.05536	0.04284	0.05367	0.05159	0.09013	0.04519	0.05570	0.08098	0.06442	0.08240	0.03793	0.03621	0.06202
LA2747	n. a.	0.10921	0.05197	0.04441	0.04426	0.04423	0.07838	0.04519	0.05543	0.08021	0.05833	0.08128	0.03713	0.03683	0.05899
LA2748	n. a.	0.09947	0.05729	0.04638	0.04860	0.04661	0.08676	0.04576	0.05542	0.07038	0.05555	0.08294	0.04278	0.03410	0.05939
LA2750	n. a.	0.10521	0.06047	0.04286	0.05190	0.03387	0.07984	0.05442	0.05655	0.08658	0.06439	0.08554	0.04285	0.03704	0.06166
LA2753	n. a.	0.10162	0.05499	0.04627	0.04214	0.04882	0.08725	0.04864	0.05543	0.07570	0.05721	0.08169	0.03972	0.03558	0.05962
LA2755	n. a.	0.10472	0.05176	0.04269	0.04164	0.04988	0.08363	0.04777	0.06872	0.08302	0.05529	0.08248	0.03649	0.03861	0.06052
LA2765	n. a.	0.10338	0.05127	0.04708	0.03819	0.04622	0.08621	0.05178	0.05515	0.08238	0.05693	0.08504	0.04153	0.03849	0.06028
LA2773	n. a.	0.09962	0.05347	0.04376	0.04610	0.04550	0.08366	0.05165	0.05543	0.08121	0.05834	0.08106	0.03786	0.03745	0.05962
LA2880	n. a.	0.10910	0.05587	0.05007	0.05075	0.03514	0.08455	0.04317	0.05598	0.08850	0.05834	0.08067	0.04656	0.04162	0.06156
LA2931	n. a.	0.09986	0.05462	0.04269	0.04897	0.05112	0.08343	0.04634	0.05543	0.08390	0.06118	0.07808	0.03932	0.03590	0.06006
LA2932	n. a.	0.11258	0.05476	0.04285	0.05034	0.03255	0.08319	0.05872	0.05542	0.08629	0.06747	0.08558	0.04414	0.03663	0.06235
LA3111	n. a.	0.10522	0.05605	0.04961	0.03135	0.05840	0.09073	0.05699	0.05654	0.08222	0.05497	0.08385	0.03562	0.03568	0.06133
LA3784	n. a.	0.10256	0.05527	0.03711	0.05216	0.04033	0.08522	0.04610	0.09830	0.07616	0.06691	0.08242	0.04505	0.02891	0.06281
LA4107	n. a.	0.11255	0.05713	0.04285	0.05116	0.03257	0.08207	0.06506	0.05542	0.08786	0.05975	0.08792	0.04252	0.03745	0.06264
LA4108	n. a.	0.11209	0.05808	0.04301	0.05116	0.03255	0.08602	0.06189	0.05543	0.08821	0.05558	0.08590	0.04114	0.03699	0.06216
LA4118	n. a.	0.09468	0.05538	0.05022	0.05251	0.03277	0.08824	0.04317	0.05542	0.08800	0.06047	0.07742	0.04835	0.03767	0.06033
LA4119	n. a.	0.10265	0.05243	0.05118	0.05077	0.04075	0.08958	0.04345	0.05958	0.08695	0.05556	0.08058	0.04769	0.03480	0.06123
LA4332	n. a.	0.10383	0.05567	0.04626	0.04325	0.05602	n. a.	0.04433	0.05598	0.08567	0.05774	0.08089	0.03574	0.03673	0.05851
Mean	n. a.	0.10473	0.05504	0.04468	0.04654	0.04560	0.08436	0.04905	0.06083	0.08233	0.05955	0.08145	0.04059	0.03629	

Note: "n. a." not available

[a]: population means calculated excluding *CT021*

Table B3.21: Synonymous divergence from *S. lycopersicum* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	0.03220	0.00559	0.00462	0.02566	0.03011	0.03963	0.01109	0.04262	0.00738	0.07264#	0.02253	0.06678#	0.04298	0.01378	0.03074	0.02256	0.02943
LA0458	0.03580	0.01413	0.00462	0.02355	0.04355	0.04017	0.01772	0.03882	0.01258	0.07544#	0.02250	0.06673#	0.03947	0.01088	0.02686	0.04087	0.03211
LA1930	0.03647	0.01586	0.00888	0.02099	0.02238	0.04162	0.02524	0.03550	0.00185*	0.07341#	0.02836	0.05985	0.05351	0.01062	0.02705	0.02454	0.03038
LA1958	0.04523	0.01314	0.00508	0.02718	0.03017	0.04445	0.01956	0.03402	0.00448	0.07848#	0.02525	0.06612	0.04073	0.01171	0.03105	0.02597	0.03141
LA1963	0.04397	0.01413	0.00444	0.02194	0.02815	0.02887	0.01098	0.04130	0.01347	0.06988#	0.02657	0.06457	0.04368	0.01729	0.02858	0.03204	0.03062
LA1968	0.03942	0.01398	0.00645	0.02566	0.02670	0.03170	0.01730	0.02665	0.00314	0.06091	0.02252	0.06055	0.04655	0.00788	0.02748	0.03351	0.02815
LA2747	0.03852	0.01896	0.00453	0.01858	0.02579	0.04065	0.02250	0.02457	0.00269	0.07203#	0.02838	0.05806	0.04394	0.00771	0.03161	0.02872	0.02920
LA2748	0.03865	0.01554	0.00064*	0.02024	0.02867	0.03354	0.01859	0.02362	0.00090*	0.06534	0.02385	0.05811	0.04601	0.01181	0.03001	0.03838	0.02837
LA2750	0.04159	0.00964	n. a.	0.02559	0.02857	0.03333	0.01178	0.02730	0.02346	0.05011	0.02788	0.07445#	0.05127	0.01561	0.02785	0.02711	0.03170
LA2753	0.03562	0.01710	0.00698	0.02694	0.02358	0.02788	0.01381	0.02428	0.00900	0.06495	0.02607	0.06849#	0.03123	0.01497	0.02819	0.02601	0.02782
LA2755	0.05339	0.01095	0.00807	0.02740	0.02572	0.03811	0.03174	0.03237	0.00223*	0.06723#	0.02478	0.06699#	0.04657	0.00908	0.02886	0.02779	0.03133
LA2765	0.03451	0.01429	0.00199*	0.02493	0.02511	0.03533	0.02648	0.02825	0.00453	0.05960	0.00724	0.06267	0.04093	0.01387	0.03163	0.02967	0.02756
LA2773	0.04106	0.01782	0.00644	0.02113	0.02754	0.02747	0.01620	0.01645	0.00717	0.05630	0.03521	0.06424	0.04211	0.01207	0.03771	0.03788	0.02918
LA2880	0.04475	0.01631	0.00915	0.02054	0.03954	0.03668	0.01087	0.03075	0.00944	0.06507	0.02294	0.06270	0.05000	0.01722	0.02973	0.02267	0.03052
LA2931	0.03164	0.01842	0.00165*	0.01622	0.03071	0.02838	0.01919	0.02051	0.00854	0.06650#	0.02478	0.05798	0.04700	0.01104	0.02883	0.03449	0.02787
LA2932	0.04116	0.00940	0.00290	0.02613	0.02818	0.02219	0.01121	0.02436	0.02198	0.05353	0.02250	0.06607	0.05139	0.01302	0.02784	0.02850	0.02815
LA3111	0.03881	0.01289	0.00825	0.02394	0.02790	0.02389	0.01457	0.03330	0.01797	0.06645#	0.02542	0.06566	0.05299	0.01069	0.02722	0.03001	0.03000
LA3784	0.02755	0.01209	0.00799	0.02235	0.03346	0.04043	0.02858	0.03705	0.00450	0.09169#	0.03077	0.06937#	0.05300	0.00387	0.02769	0.02838	0.03242
LA4107	0.04447	0.01582	0.00964	0.02090	0.03123	0.03137	0.01119	0.02249	0.02246	0.04916	0.02248	0.05409	0.04000	0.01721	0.03661	0.04928	0.02990
LA4108	0.04173	0.01571	0.00930	0.01984	0.02831	0.03401	0.01119	0.02295	0.02248	0.05047	0.02381	0.05382	0.04027	0.01774	0.03314	0.03961	0.02902
LA4118	0.04792	0.01437	0.00879	0.02403	0.03728	0.05060	0.01109	0.03293	0.01348	0.06667#	0.02389	0.06458	0.04375	0.01713	0.02758	0.02152	0.03160
LA4119	0.04432	0.01918	0.00908	0.01874	0.03502	0.04298	0.01225	0.03174	0.01797	0.06539	0.02472	0.05473	0.05086	0.01713	0.02961	0.02276	0.03103
LA4332	0.04085	0.01565	0.00770	0.02220	0.03651	0.03699	0.01098	0.02577	0.00713	0.07089#	0.02524	0.05772	0.04937	0.01387	0.03056	0.02935	0.03005
Mean	0.03998	0.01439	0.00624	0.02281	0.03018	0.03523	0.01670	0.02946	0.01038	0.06575	0.02468	0.06280	0.04555	0.01288	0.02984	0.03051	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.22: Synonymous divergence from *S. lycopersicum* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.03664	0.03580	0.02844	0.00071*	0.04516	0.02661	0.04029	0.01302	0.03049	0.02805	0.03406	0.03448	0.03790	0.01644	0.02915
LA0458	0.03701	0.04942	0.02699	0.00966	0.00650	0.03243	0.04541	0.01302	0.05543	0.03609	0.03180	0.04473	0.03352	0.01588	0.03128
LA1930	0.04018	0.04274	0.02524	0.01074	0.02964	0.01081	0.04839	0.01822	0.03261	0.02833	0.04452	0.03609	0.03553	0.01187	0.02964
LA1958	0.03703	0.04309	0.02380	0.00582	0.02664	0.02078	0.04337	0.01640	0.05155	0.02920	0.04783	0.04157	0.03567	0.01644	0.03137
LA1963	0.03543	0.05330	0.02548	0.00454	0.02461	0.01667	0.04473	0.01588	0.05432	0.03113	0.03124	0.03583	0.03975	0.01477	0.03055
LA1968	0.03201	0.05399	0.02688	0.00398	0.01870	0.01929	0.03397	0.01484	0.05570	0.02753	0.03677	0.03910	0.04113	0.01644	0.03002
LA2747	0.03332	0.05460	0.02347	0.00472	0.00650	0.01171	0.04309	0.01484	0.05543	0.03090	0.03069	0.04552	0.03876	0.01654	0.02929
LA2748	0.03664	0.04484	0.02874	0.00398	0.02664	0.01409	0.04353	0.01536	0.05542	0.02752	0.02791	0.04232	0.04105	0.01352	0.03011
LA2750	0.03286	0.04519	0.03188	0.00043*	0.01006	0.00130*	0.05430	0.02318	0.05641	0.03385	0.04145	0.04289	0.03663	0.01620	0.03047
LA2753	0.03607	0.04191	0.02643	0.00668	0.01934	0.01627	0.04611	0.01796	0.05543	0.02912	0.03068	0.03696	0.03663	0.01488	0.02961
LA2755	0.03888	0.04737	0.02332	0.00142*	0.01077	0.01735	0.04513	0.01718	0.05099	0.03381	0.02764	0.03685	0.03777	0.01780	0.02902
LA2765	0.03235	0.04940	0.02276	0.00401	0.01787	0.01367	0.04415	0.02108	0.05515	0.03259	0.02930	0.04164	0.03867	0.01710	0.02998
LA2773	0.03198	0.04826	0.02493	0.00426	0.00589	0.01300	0.04485	0.02061	0.05543	0.03114	0.03069	0.03906	0.04115	0.01653	0.02913
LA2880	0.03276	0.05986	0.02732	0.00710	0.01056	0.00260*	0.04779	0.01302	0.05598	0.03559	0.03069	0.04532	0.04766	0.02081	0.03122
LA2931	0.03245	0.03942	0.02612	0.00455	0.01524	0.01863	0.04700	0.01614	0.05543	0.03256	0.03460	0.04217	0.04344	0.01530	0.03022
LA2932	0.03102	0.05653	0.02624	0.00014*	0.00528	0*	0.05389	0.02708	0.05348	0.03320	0.03982	0.04296	0.03618	0.01570	0.03011
LA3111	0.03395	0.04569	0.02750	0.00696	0.03868	0.03201	0.03485	0.02551	0.05654	0.03377	0.02735	0.03744	0.03396	0.01540	0.03212
LA3784	0.04436	0.04166	0.02666	0.00801	0.02415	0.00780	0.04689	0.01564	0.04064	0.02751	0.06415	0.03566	0.04088	0.00811	0.03087
LA4107	0.03603	0.05686	0.02862	0.00014*	0.01218	0*	0.05244	0.03281	0.05542	0.03482	0.03209	0.04556	0.03665	0.01653	0.03144
LA4108	0.03440	0.05607	0.02952	0.00029*	0.01421	0*	0.05662	0.02995	0.05543	0.03516	0.02793	0.04291	0.03532	0.01603	0.03099
LA4118	0.03557	0.04108	0.02689	0.00724	0.01660	0.00022*	0.04412	0.01302	0.05542	0.03587	0.03286	0.04214	0.04869	0.01686	0.02976
LA4119	0.03710	0.04862	0.03141	0.00761	0.02031	0.00824	0.04606	0.01328	0.05958	0.03926	0.02778	0.04393	0.04878	0.01206	0.03172
LA4332	0.03394	0.05342	0.02712	0.00329	0.02092	0.02357	n. a.	0.01406	0.05598	0.03280	0.03011	0.03904	0.03735	0.01613	0.02983
Mean	0.03530	0.04822	0.02677	0.00462	0.01854	0.01335	0.04577	0.01835	0.05253	0.03217	0.03443	0.04062	0.03926	0.01554	

Note: "n. a." not available

*: in lower 2.5 % of the density distribution

Table B3.23: Nonsynonymous divergence from *S. ochranthum* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean[a]
LA0456	0.10512[#]	0.00721	0.01270	0.02421	0.01310	0.01087	0.00374	0.01716	n. a.	0.01437	0.02979	0.04724[#]	0.01239	0.00168	0.01636	0.00955	0.01569
LA0458	0.10680[#]	0.00818	0.01184	0.02103	0.01228	0.01000	0.00366	0.01775	n. a.	0.01325	0.02704	0.04719[#]	0.00914	0.00134	0.01453	0.00816	0.01468
LA1930	0.10942[#]	0.00853	0.01333	0.02387	0.01576	0.00919	0.00417	0.01819	n. a.	0.01251	0.02384	0.04383[#]	0.02198	0.00319	0.01441	0.00832	0.01590
LA1958	0.10784[#]	0.00737	0.01268	0.02107	0.01479	0.00912	0.00366	0.01860	n. a.	0.01053	0.02848	0.04626[#]	0.01648	0.00302	0.01883	0.01020	0.01556
LA1963	0.10836[#]	0.00699	0.01265	0.02185	0.01499	0.00990	0.00402	0.01852	n. a.	0.01316	0.02660	0.04453[#]	0.01195	0.00160	0.01614	0.00869	0.01503
LA1968	0.10665[#]	0.00730	0.01235	0.02131	0.01562	0.00899	0.00431	0.01848	n. a.	0.01045	0.02863	0.04747[#]	0.01466	0.00222	0.01733	0.00965	0.01550
LA2747	0.10446[#]	0.00780	0.01234	0.02173	0.01425	0.00941	0.00374	0.01786	n. a.	0.01188	0.02529	0.04284[#]	0.01086	0.00145	0.01593	0.00819	0.01443
LA2748	0.10701[#]	0.00803	0.01420	0.02105	0.01493	0.00866	0.00388	0.01613	n. a.	0.01081	0.03075	0.04301[#]	0.01077	0.00160	0.01657	0.00883	0.01482
LA2750	0.10707[#]	0.00810	n. a.	0.02081	0.01539	0.00984	0.00596	0.02068	n. a.	0.01021	0.02545	0.04274	0.01228	0.00240	0.01770	0.00904	0.01524
LA2753	0.10750[#]	0.00770	0.01226	0.02374	0.01416	0.01090	0.00022	0.01753	n. a.	0.01080	0.03054	0.04332[#]	0.01504	0.00230	0.01733	0.00921	0.01521
LA2755	0.10847[#]	0.00754	0.01218	0.02441	0.01637	0.00889	0.00374	0.02006	n. a.	0.00929	0.02892	0.04358[#]	0.01557	0.00132	0.01661	0.00966	0.01550
LA2765	0.10461[#]	0.00792	0.01367	0.02197	0.01463	0.00907	0.00431	0.01918	n. a.	0.01027	0.02919	0.04473[#]	0.01198	0.00176	0.01496	0.00883	0.01519
LA2773	0.10312[#]	0.00734	0.01218	0.02212	0.01577	0.01004	0.00381	0.01809	n. a.	0.01349	0.02556	0.04284[#]	0.01231	0.00178	0.01593	0.00816	0.01488
LA2880	0.10679[#]	0.00716	0.01184	0.02055	0.01337	0.01101	0.00374	0.01744	n. a.	0.01145	0.02647	0.04586[#]	0.00923	0.00129	0.01587	0.00818	0.01443
LA2931	0.10188[#]	0.00773	0.01392	0.02032	0.01398	0.00912	0.00377	0.01800	n. a.	0.01288	0.02587	0.04358[#]	0.01168	0.00199	0.01580	0.00936	0.01478
LA2932	0.10910[#]	0.00742	0.01278	0.01981	0.01479	0.00950	0.00384	0.02065	n. a.	0.01079	0.02413	0.04180	0.01178	0.00150	0.01717	0.00840	0.01440
LA3111	0.10716[#]	0.00718	0.01192	0.02201	0.01552	0.00884	0.00413	0.02505	n. a.	0.01163	0.03272	0.04671[#]	0.01567	0.00214	0.01651	0.00832	0.01630
LA3784	0.10850[#]	0.00796	0.01351	0.02249	0.01505	0.00897	0.00449	0.02176	n. a.	0.01897	0.02800	0.04032	0.02261	0.00018	0.01619	0.00762	0.01630
LA4107	0.11015[#]	0.00957	0.01227	0.02105	0.01455	0.00840	0.00395	0.02095	n. a.	0.00929	0.02473	0.04273	0.00950	0.00170	0.01427	0.00808	0.01437
LA4108	0.10972[#]	0.00865	0.01217	0.02099	0.01347	0.00985	0.00388	0.02090	n. a.	0.00938	0.02575	0.04273	0.01096	0.00160	0.01509	0.00840	0.01452
LA4118	0.10696[#]	0.00697	0.01197	0.02078	0.01388	0.00929	0.00359	0.01809	n. a.	0.01185	0.02906	0.04614[#]	0.01155	0.00134	0.01614	0.00913	0.01490
LA4119	0.1062[#]	0.00764	0.01190	0.02116	0.01402	0.01067	0.00477	0.01844	n. a.	0.01022	0.02575	0.04025	0.01057	0.00152	0.01517	0.00822	0.01424
LA4332	0.10977[#]	0.00803	0.01221	0.02018	0.01388	0.01047	0.00410	0.01860	n. a.	0.01035	0.03153	0.04330[#]	0.01105	0.00139	0.01722	0.00950	0.01497
Mean	0.10707[#]	0.00775	0.01259	0.02167	0.01455	0.00961	0.00389	0.01905	n. a.	0.01164	0.02757	0.04404[#]	0.01304	0.00175	0.01618	0.00877	

Note: "n. a." not available

[#]: in upper 2.5 % of the density distribution

[a]: population means calculated excluding *NtC7*, *TAS14* and *His1*

Table B3.24: Nonsynonymous divergence from *S. ochranthum* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean[a]
LA0456	0.02375	0.00695	0.00640	0.00653	0*	0*	0.00267	0.00393	0.00461	0.00865	0.00824	0.02759	0.01493	0.00824	0.00760
LA0458	0.02472	0.00671	0.00599	0.00808	0.00013	0.00006*	0.00302	0.00401	0.00412	0.00744	0.00824	0.02726	0.01390	0.00939	0.00757
LA1930	0.02603	0.00970	0.00631	0.00676	0.00078	0.00127	0.00278	0.00566	0.00548	0.00686	0.00618	0.02590	0.01506	0.01019	0.00792
LA1958	0.02344	0.00708	0.00631	0.00753	0.00090	0.00006*	0.00283	0.00440	0.00420	0.00855	0.00758	0.02772	0.01464	0.00824	0.00770
LA1963	0.02324	0.00649	0.00678	0.00783	0.00116	0.00016	0.00308	0.00440	0.00412	0.00773	0.00865	0.02695	0.01533	0.00862	0.00779
LA1968	0.02430	0.00601	0.00738	0.00703	0.00149	0.00022	0.00308	0.00409	0.00428	0.00761	0.00882	0.02734	0.01553	0.00897	0.00783
LA2747	0.02305	0.00646	0.00691	0.00782	0.00006*	0.00005*	0.00283	0.00417	0.00412	0.00836	0.00775	0.02738	0.01624	0.00827	0.00772
LA2748	0.02206	0.00732	0.00638	0.00768	0.00168	0*	0.00262	0.00393	0.00420	0.00815	0.00915	0.02808	0.01534	0.00853	0.00793
LA2750	0.02430	0.00807	0.00731	0.00942	0.00107	0.00055	0.00338	0.00739	0.00502	0.00739	0.00651	0.02854	0.01437	0.00942	0.00834
LA2753	0.02400	0.00769	0.00680	0.00827	0.00213	0.00066	0.00297	0.00519	0.00486	0.00799	0.00989	0.02445	0.01458	0.00912	0.00805
LA2755	0.02419	0.00705	0.00724	0.00773	0.00310	0.00033	0.00283	0.00401	0.00428	0.00692	0.00832	0.02620	0.01470	0.00846	0.00778
LA2765	0.02368	0.00725	0.00763	0.00832	0.00039	0.00041	0.00284	0.00393	0.00428	0.00673	0.00833	0.02692	0.01454	0.00837	0.00769
LA2773	0.02371	0.00725	0.00655	0.00763	0.00006*	0*	0.00293	0.00455	0.00412	0.00773	0.00767	0.02798	0.01552	0.00836	0.00772
LA2880	0.02274	0.00700	0.00566	0.00872	0.00006*	0.00005*	0.00262	0.00393	0.00412	0.00502	0.00824	0.02756	0.01504	0.00824	0.00740
LA2931	0.02297	0.00746	0.00623	0.00728	0.00097	0.00005*	0.00324	0.00393	0.00412	0.00753	0.01013	0.02899	0.01560	0.00827	0.00798
LA2932	0.02478	0.00866	0.00634	0.00907	0.00019	0.00011	0.00310	0.00692	0.00478	0.00739	0.00503	0.02739	0.01430	0.00855	0.00783
LA3111	0.02328	0.00672	0.00702	0.00763	0.00039	0*	0.00283	0.00715	0.00412	0.00828	0.00956	0.02657	0.01474	0.00837	0.00795
LA3784	0.02216	0.00759	0.00976	0.00681	0.00039	0.00011	0.00308	0.00464	0.00583	0.00899	0.00643	0.02678	0.01597	0.00986	0.00817
LA4107	0.02329	0.00608	0.00703	0.00927	0.00006*	0.00022	0.00264	0.00519	0.00420	0.00629	0.00832	0.02903	0.01333	0.00849	0.00770
LA4108	0.02380	0.00605	0.00688	0.00912	0.00006*	0.00005*	0.00262	0.00401	0.00412	0.00631	0.00832	0.02773	0.01343	0.00864	0.00749
LA4118	0.02271	0.00702	0.00593	0.00912	0.00077	0.00011	0.00267	0.00393	0.00420	0.00531	0.00989	0.02853	0.01529	0.00846	0.00779
LA4119	0.02122	0.00747	0.00675	0.00701	0.00013	0.00011	0.00305	0.00401	0.00412	0.00619	0.00858	0.02958	0.01535	0.00855	0.00776
LA4332	0.02309	0.00677	0.00614	0.00757	0.00039	0.00011	n. a.	0.00464	0.00412	0.00640	0.00915	0.02699	0.01601	0.00846	0.00806
Mean	0.02350	0.00717	0.00677	0.00792	0.00071	0.00020	0.00290	0.00470	0.00441	0.00730	0.00822	0.02745	0.01495	0.00870	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
[a]: population means calculated excluding *CT021*

Table B3.25: Nonsynonymous divergence from *S. lycopersicoides* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean[a]
LA0456	0.09951#	0.01123	0.00727	0.02160	0.02301	0.00620	0.00013	0.01440	0.01425	0.00964	0.03602	0.04610#	0.00335	0.00297	n. a.	0.01015	0.01477
LA0458	0.10114#	0.01214	0.00643	0.01841	0.02210	0.00543	0.00007	0.01670	0.01083	0.00877	0.03335	0.04690#	0.00009	0.00263	n. a.	0.00922	0.01402
LA1930	0.10367#	0.01248	0.00789	0.02125	0.02539	0.00512	0.00186	0.01538	0.01528	0.01671	0.03024	0.04820#	0.01293	0.00449	n. a.	0.01089	0.01637
LA1958	0.10202#	0.01138	0.00724	0.01847	0.02460	0.00450	0.00086	0.01578	0.01481	0.00607	0.03530	0.04673#	0.00733	0.00432	n. a.	0.01174	0.01495
LA1963	0.10273#	0.01103	0.00730	0.01925	0.02480	0.00528	0.00047	0.01571	0.01255	0.00868	0.03291	0.04387#	0.00290	0.00289	n. a.	0.00972	0.01422
LA1968	0.10148#	0.01132	0.00691	0.01871	0.02542	0.00437	0.00086	0.01569	0.01395	0.00598	0.03489	0.04662#	0.00561	0.00351	n. a.	0.01064	0.01466
LA2747	0.09896#	0.01216	0.00691	0.01911	0.02406	0.00486	0.00013	0.01502	0.01711	0.00741	0.03164	0.03970	0.00181	0.00274	n. a.	0.00938	0.01346
LA2748	0.10176#	0.01286	0.00871	0.01845	0.02474	0.00405	0.00027	0.01338	0.01179	0.00635	0.03674	0.04197	0.00181	0.00289	n. a.	0.00987	0.01401
LA2750	0.10149#	0.01212	n. a.	0.01823	0.02525	0.00407	0.00232	0.02034	0.01222	0.00578	0.03208	0.04265#	0.00322	0.00369	n. a.	0.01012	0.01499
LA2753	0.10314#	0.01170	0.00712	0.02113	0.02393	0.00549	0.00298	0.01476	0.01737	0.00634	0.03675	0.04237	0.00581	0.00359	n. a.	0.00992	0.01476
LA2755	0.10331#	0.01154	0.00676	0.02181	0.02574	0.00471	0.00013	0.01726	0.01816	0.00482	0.03517	0.04254#	0.00652	0.00261	n. a.	0.01015	0.01460
LA2765	0.09907#	0.01194	0.00822	0.01937	0.02444	0.00441	0.00073	0.01633	0.01507	0.00581	0.03571	0.04397#	0.00283	0.00305	n. a.	0.01016	0.01438
LA2773	0.09756#	0.01162	0.00676	0.01951	0.02558	0.00554	0.00020	0.01530	0.01486	0.00902	0.02993	0.04179	0.00326	0.00307	n. a.	0.00925	0.01391
LA2880	0.09967#	0.01119	0.00642	0.01794	0.02317	0.00636	0.00013	0.01724	0.01265	0.00698	0.03279	0.04396#	0.00199	0.00258	n. a.	0.00928	0.01385
LA2931	0.09650#	0.01172	0.00841	0.01900	0.02381	0.00461	0.00013	0.01522	0.01517	0.00841	0.03221	0.04330#	0.00263	0.00328	n. a.	0.01060	0.01410
LA2932	0.10368#	0.01143	0.00735	0.01720	0.02469	0.00499	0.00027	0.02005	0.01373	0.00616	0.03052	0.04180	0.00272	0.00279	n. a.	0.00946	0.01380
LA3111	0.10144#	0.01121	0.00653	0.01939	0.02533	0.00429	0.00047	0.02471	0.01147	0.00716	0.03929	0.04671#	0.00661	0.00344	n. a.	0.00933	0.01573
LA3784	0.10280#	0.01195	0.00804	0.01988	0.02486	0.00445	0.00080	0.01891	0.02058	0.01647	0.03428	0.04261#	0.01357	0.00018	n. a.	0.01040	0.01588
LA4107	0.10437#	0.01345	0.00685	0.01844	0.02439	0.00391	0.00033	0.02071	0.01641	0.00482	0.03053	0.04273#	0.00037	0.00300	n. a.	0.00915	0.01374
LA4108	0.10396#	0.01261	0.00672	0.01838	0.02332	0.00530	0.00027	0.02065	0.01651	0.00491	0.03039	0.04254#	0.00183	0.00289	n. a.	0.00943	0.01379
LA4118	0.10022#	0.01101	0.00655	0.01818	0.02369	0.00474	0*	0.01758	0.01319	0.00738	0.03502	0.04567#	0.00274	0.00263	n. a.	0.01019	0.01426
LA4119	0.09913#	0.01176	0.00647	0.01976	0.02403	0.00594	0.00129	0.01863	0.01244	0.00576	0.03209	0.03817	0.00390	0.00281	n. a.	0.00798	0.01374
LA4332	0.10414#	0.01202	0.00678	0.01757	0.02368	0.00586	0.00060	0.01581	0.01773	0.00583	0.03771	0.04121	0.00210	0.00269	n. a.	0.01021	0.01401
Mean	0.10138#	0.01182	0.00717	0.01918	0.02435	0.00498	0.00067	0.01720	0.01470	0.00762	0.03372	0.04357#	0.00417	0.00299	n. a.	0.00988	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution
[a]: population means calculated excluding *NtC7*, *TAS14* and *His1*

Table B3.26: Nonsynonymous divergence from *S. lycopersicoides* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean[a]
LA0456	n. a.	0.00573	0.00492	0.01401	0*	0*	0.00267	0*	0.00049	0.00768	0.00824	0.01791	0.01350	0.01296	0.00678
LA0458	n. a.	0.00549	0.00451	0.01555	0.00013	0.00006	0.00302	0.00008	0*	0.00660	0.00824	0.01802	0.01247	0.01411	0.00679
LA1930	n. a.	0.00844	0.00482	0.01427	0.00078	0.00127	0.00278	0.00173	0.00133	0.00609	0.00618	0.01650	0.01363	0.01492	0.00713
LA1958	n. a.	0.00565	0.00482	0.01501	0.00090	0.00006	0.00283	0.00047	0.00008	0.00803	0.00758	0.01778	0.01358	0.01296	0.00690
LA1963	n. a.	0.00518	0.00529	0.01530	0.00116	0.00016	0.00308	0.00047	0*	0.00687	0.00865	0.01632	0.01390	0.01334	0.00690
LA1968	n. a.	0.00480	0.00590	0.01451	0.00149	0.00022	0.00309	0.00016	0.00016	0.00676	0.00882	0.01739	0.01410	0.01368	0.00701
LA2747	n. a.	0.00524	0.00543	0.01533	0.00006	0.00005	0.00283	0.00024	0*	0.00742	0.00775	0.01759	0.01481	0.01299	0.00690
LA2748	n. a.	0.00609	0.00489	0.01515	0.00168	0*	0.00262	0*	0.00008	0.00724	0.00915	0.01833	0.01392	0.01325	0.00711
LA2750	n. a.	0.00681	0.00582	0.01690	0.00107	0.00055	0.00341	0.00346	0.00091	0.00662	0.00651	0.01887	0.01294	0.01413	0.00754
LA2753	n. a.	0.00648	0.00532	0.01575	0.00213	0.00066	0.00297	0.00126	0.00074	0.00709	0.00989	0.01409	0.01315	0.01384	0.00718
LA2755	n. a.	0.00603	0.00576	0.01520	0.00310	0.00033	0.00283	0.00008	0.00016	0.00614	0.00832	0.01567	0.01327	0.01318	0.00693
LA2765	n. a.	0.00578	0.00615	0.01583	0.00039	0.00033	0.00284	0*	0.00016	0.00597	0.00833	0.01719	0.01311	0.01315	0.00686
LA2773	n. a.	0.00603	0.00507	0.01510	0.00006	0*	0.00293	0.00063	0*	0.00686	0.00767	0.01771	0.01409	0.01310	0.00687
LA2880	n. a.	0.00464	0.00418	0.01620	0.00006	0.00005	0.00262	0*	0*	0.00445	0.00824	0.01763	0.01361	0.01296	0.00651
LA2931	n. a.	0.00623	0.00474	0.01475	0.00097	0.00005	0.00324	0.00008	0*	0.00668	0.01013	0.01895	0.01416	0.01299	0.00715
LA2932	n. a.	0.00741	0.00486	0.01658	0.00019	0.00011	0.00310	0.00299	0.00066	0.00656	0.00503	0.01760	0.01287	0.01328	0.00702
LA3111	n. a.	0.00562	0.00554	0.01511	0.00039	0*	0.00283	0.00322	0*	0.00742	0.00956	0.01607	0.01330	0.01309	0.00709
LA3784	n. a.	0.00636	0.00828	0.01433	0.00039	0.00011	0.00308	0.00071	0.00161	0.00798	0.00643	0.01712	0.01454	0.01457	0.00735
LA4107	n. a.	0.00491	0.00554	0.01678	0.00006	0.00022	0.00264	0.00126	0.00008	0.00558	0.00832	0.01926	0.01189	0.01322	0.00690
LA4108	n. a.	0.00485	0.00540	0.01663	0.00006	0.00005	0.00262	0.00008	0*	0.00560	0.00832	0.01786	0.01199	0.01337	0.00668
LA4118	n. a.	0.00573	0.00445	0.01660	0.00077	0.00011	0.00267	0*	0.00008	0.00471	0.00989	0.01873	0.01387	0.01318	0.00698
LA4119	n. a.	0.00551	0.00352	0.01469	0.00013	0.00011	0.00305	0.00008	0*	0.00547	0.00858	0.01927	0.01390	0.01365	0.00677
LA4332	n. a.	0.00515	0.00466	0.01508	0.00039	0.00011	n. a.	0.00071	0*	0.00568	0.00915	0.01719	0.01458	0.01318	0.00716
Mean	n. a.	0.00583	0.00521	0.01542	0.00071	0.00020	0.00290	0.00077	0.00028435	0.00650	0.00822	0.01752	0.01353	0.01344	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
[a]: population means calculated excluding *CT021*

Table B3.27: Nonsynonymous divergence from *S. lycopersicum* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	0.01748	0.00343	0.00589	0.01222	0.00638	0.00508	0.00013	0.00671	0.00491	0.00964	0.01483	0.03887#	0.00335	0.00039	0.01636	0.00190	0.00922
LA0458	0.02096	0.00435	0.00515	0.01060	0.00564	0.00430	0.00010	0.00758	0.00010	0.01324	0.01215	0.03797#	0.00009	0.00005	0.01453	0.00013	0.00856
LA1930	0.01603	0.00468	0.00661	0.01319	0.00907	0.00400	0.00186	0.00761	0.00134	0.01726	0.00904	0.03453#	0.01293	0.00189	0.01441	0.00175	0.00976
LA1958	0.01959	0.00389	0.00545	0.01076	0.00812	0.00344	0.00086	0.00820	0.00572	0.01053	0.01412	0.03827#	0.00733	0.00173	0.01776	0.00324	0.00994
LA1963	0.01949	0.00386	0.00581	0.01143	0.00844	0.00413	0.00050	0.01025	0.00499	0.01316	0.01172	0.03570#	0.00290	0.00031	0.01614	0.00058	0.00934
LA1968	0.01732	0.00361	0.00531	0.01090	0.00893	0.00326	0.00090	0.00800	0.00405	0.01045	0.01370	0.03883#	0.00561	0.00093	0.01733	0.00161	0.00942
LA2747	0.01932	0.00455	0.00563	0.01130	0.00759	0.00376	0.00017	0.00861	0.00717	0.01170	0.01045	0.03675#	0.00181	0.00015	0.01593	0.00028	0.00907
LA2748	0.01714	0.00506	0.00640	0.01077	0.00825	0.00292	0.00030	0.00569	0.00343	0.01081	0.01554	0.03456#	0.00181	0.00031	0.01657	0.00084	0.00878
LA2750	0.02038	0.00432	n. a.	0.01046	0.00865	0.00407	0.00232	0.01047	0.00789	0.01030	0.01088	0.03334	0.00322	0.00111	0.01770	0.00096	0.00974
LA2753	0.02087	0.00399	0.00568	0.01272	0.00740	0.00513	0.00298	0.00707	0.01039	0.01080	0.01555	0.03477#	0.00581	0.00101	0.01733	0.00144	0.01018
LA2755	0.01845	0.00379	0.00548	0.01288	0.00968	0.00360	0.00013	0.00958	0.00874	0.00928	0.01398	0.03513#	0.00652	0.00003	0.01661	0.00255	0.00978
LA2765	0.01788	0.00414	0.00694	0.01119	0.00795	0.00328	0.00073	0.00861	0.00464	0.00991	0.01357	0.03542#	0.00283	0.00046	0.01496	0.00107	0.00897
LA2773	0.01555	0.00375	0.00548	0.01173	0.00910	0.00441	0.00023	0.00829	0.00561	0.01349	0.00875	0.03438#	0.00326	0.00049	0.01485	0.00015	0.00872
LA2880	0.02008	0.00339	0.00514	0.01013	0.00671	0.00520	0.00013	0.00728	0.00395	0.01127	0.01159	0.03826#	0.00199	0*	0.01587	0.00015	0.00882
LA2931	0.01864	0.00392	0.00668	0.01087	0.00727	0.00349	0.00013	0.00753	0.00592	0.01288	0.01102	0.03442#	0.00263	0.00070	0.01580	0.00205	0.00900
LA2932	0.01903	0.00363	0.00607	0.00942	0.00819	0.00387	0.00027	0.01025	0.00800	0.01079	0.00932	0.03230	0.00272	0.00021	0.01717	0.00033	0.00885
LA3111	0.02091	0.00390	0.00520	0.01153	0.00883	0.00322	0.00047	0.01466	0.00468	0.01163	0.01624	0.03721#	0.00661	0.00085	0.01651	0.00033	0.01017
LA3784	0.02392	0.00413	0.00676	0.01272	0.00838	0.00336	0.00080	0.01114	0.00696	0.01593	0.01308	0.03081	0.01357	0.00018	0.01603	0.00124	0.01056
LA4107	0.02004	0.00565	0.00557	0.01065	0.00786	0.00281	0.00037	0.01073	0.01060	0.00929	0.00989	0.03323	0.00037	0.00041	0.01428	0.00005	0.00886
LA4108	0.02001	0.00477	0.00528	0.01059	0.00678	0.00415	0.00030	0.01068	0.01080	0.00938	0.01088	0.03342	0.00183	0.00031	0.01510	0.00030	0.00904
LA4118	0.02064	0.00370	0.00527	0.01040	0.00723	0.00361	0*	0.00792	0.00530	0.01167	0.01412	0.03702#	0.00274	0.00005	0.01614	0.00104	0.00918
LA4119	0.01777	0.00374	0.00519	0.01147	0.00722	0.00476	0.00132	0.00822	0.00540	0.01022	0.01088	0.04226#	0.00390	0.00023	0.01517	0.00003	0.00924
LA4332	0.02073	0.00420	0.00550	0.00981	0.00723	0.00471	0.00063	0.00814	0.00833	0.00999	0.01652	0.03584#	0.00210	0.00010	0.01722	0.00164	0.00954
Mean	0.01923	0.00411	0.00575	0.01121	0.00787	0.00394	0.00068	0.00884	0.00604	0.01146	0.01251	0.03580#	0.00417	0.00052	0.01608	0.00103	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution
#: in upper 2.5 % of the density distribution

Table B3.28: Nonsynonymous divergence from *S. lycopersicum* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.00520	0.00233	0.00403	0.00028	0*	0*	0.00010	0*	0.00049	0.00212	0.00412	0.00865	0.00616	0.00352	0.00264
LA0458	0.00541	0.00209	0.00451	0.00176	0.00013	0.00006	0.00044	0.00007	0*	0.00274	0.00412	0.00949	0.00591	0.00468	0.00296
LA1930	0.00636	0.00505	0.00429	0.00046	0.00078	0.00127	0.00021	0.00159	0.00133	0.00208	0.00206	0.00787	0.00739	0.00546	0.00330
LA1958	0.00427	0.00227	0.00478	0.00120	0.00090	0.00006	0.00026	0.00043	0.00008	0.00306	0.00346	0.00839	0.00664	0.00352	0.00281
LA1963	0.00412	0.00171	0.00381	0.00153	0.00116	0.00016	0.00051	0.00043	0*	0.00204	0.00453	0.00833	0.00699	0.00390	0.00280
LA1968	0.00502	0.00139	0.00519	0.00074	0.00149	0.00022	0.00051	0.00014	0.00016	0.00148	0.00470	0.00859	0.00664	0.00425	0.00289
LA2747	0.00387	0.00181	0.00448	0.00144	0.00006	0.00005	0.00026	0.00022	0*	0.00208	0.00363	0.00909	0.00756	0.00355	0.00272
LA2748	0.00319	0.00268	0.00491	0.00134	0.00168	0*	0.00005	0*	0.00008	0.00167	0.00503	0.00874	0.00691	0.00381	0.00286
LA2750	0.00493	0.00338	0.00584	0.00305	0.00107	0.00055	0.00076	0.00325	0.00095	0.00104	0.00239	0.00927	0.00567	0.00473	0.00335
LA2753	0.00566	0.00308	0.00486	0.00190	0.00213	0.00066	0.00039	0.00115	0.00074	0.00174	0.00577	0.00887	0.00618	0.00440	0.00340
LA2755	0.00556	0.00263	0.00534	0.00143	0.00310	0.00033	0.00026	0.00007	0.00016	0.00191	0.00420	0.00894	0.00670	0.00374	0.00317
LA2765	0.00444	0.00229	0.00517	0.00191	0.00039	0.00033	0.00026	0*	0.00016	0.00085	0.00420	0.00947	0.00590	0.00360	0.00278
LA2773	0.00448	0.00215	0.00448	0.00130	0.00006	0*	0.00036	0.00058	0*	0.00137	0.00354	0.01048	0.00695	0.00363	0.00281
LA2880	0.00359	0.00125	0.00420	0.00236	0.00006	0.00005	0.00005	0*	0*	0.00260	0.00412	0.00951	0.00629	0.00352	0.00269
LA2931	0.00380	0.00283	0.00457	0.00120	0.00097	0.00005	0.00067	0*	0*	0.00163	0.00601	0.00999	0.00748	0.00355	0.00305
LA2932	0.00545	0.00399	0.00421	0.00260	0.00019	0.00011	0.00052	0.00274	0.00066	0.00100	0.00091	0.00871	0.00554	0.00382	0.00289
LA3111	0.00409	0.00222	0.00459	0.00134	0.00039	0*	0.00026	0.00296	0*	0.00356	0.00544	0.00947	0.00529	0.00365	0.00309
LA3784	0.00479	0.00295	0.00682	0.00051	0.00039	0.00011	0.00051	0.00065	0.00161	0.00271	0.00231	0.00964	0.00708	0.00512	0.00323
LA4107	0.00409	0.00149	0.00554	0.00279	0.00006	0.00022	0.00005	0.00115	0.00008	0.00098	0.00420	0.01081	0.00593	0.00375	0.00294
LA4108	0.00498	0.00143	0.00541	0.00265	0.00006	0.00005	0.00005	0.00007	0*	0.00115	0.00420	0.00815	0.00608	0.00391	0.00273
LA4118	0.00349	0.00233	0.00445	0.00268	0.00077	0.00011	0.00010	0*	0.00008	0.00256	0.00577	0.00930	0.00652	0.00374	0.00299
LA4119	0.00363	0.00209	0.00352	0.00299	0.00013	0.00011	0.00047	0.00007	0*	0.00331	0.00442	0.00993	0.00650	0.00346	0.00290
LA4332	0.00391	0.00175	0.00467	0.00121	0.00039	0.00011	n. a.	0.00065	0*	0.00219	0.00503	0.00787	0.00759	0.00374	0.00301
Mean	0.00454	0.00240	0.00477	0.00168	0.00071	0.00020	0.00032	0.00071	0.00029	0.00199	0.00409	0.00911	0.00652	0.00396	

Note: "n. a." not available
*: in lower 2.5 % of the density distribution

Table B3.29: Candidate genes with significant McDonald-Kreitman test statistic with *S. ochranthum*.

Gene	Population	Fisher's exact test, p-value	Fixed differences		Polymorphism	
			S	NS	S	NS
AREB1	LA2750	0.00137	11	6	8	35
AREB1	LA4107	0.00511	13	7	2	12
AREB1	LA4108	0.00670	13	7	1	9
JERF3	LA0458	0.02580	10	7	6	20
JERF3	LA1963	0.04975	10	5	7	15
JERF3	LA2750	0.04846	9	8	6	23
JERF3	LA2755	0.01453	9	6	4	18
JERF3	LA2932	0.01823	11	8	1	10
JERF3	LA3111	0.04861	9	6	6	17
JERF3	LA4107	0.01823	11	8	1	10
DREB3	LA2750	0.01943	7	2	14	33
DREB3	LA4107	0.04718	8	3	4	10
dhn1	LA1968	0.04739	4	1	3	10
dhn1	LA2750	0.02345	4	1	4	16
dhn1	LA4119	0.02767	4	1	4	15
dhn1	LA4332	0.03186	4	1	1	7
ER5	LA0456	0.01937	6	2	2	10
ER5	LA2750	0.00832	6	2	2	12
le25	LA2880	0.03741	3	2	1	13
le25	LA2932	0.04396	3	3	0	9
le25	LA4107	0.02941	3	3	0	11
TSW12	LA2750	0.02563	6	2	4	14
CT208	LA1930	0.01689	13	1	8	8
CT208	LA1958	0.03285	13	1	7	6
CT208	LA1968	0.00112	13	1	7	13
CT208	LA2748	0.02609	13	1	6	6
CT208	LA2750	0.00002	14	1	8	23
CT208	LA2753	0.02128	13	1	5	6
CT208	LA2765	0.01689	13	1	8	8
CT208	LA2773	0.03518	13	1	8	7
CT208	LA2931	0.03285	13	1	7	6
CT208	LA2932	0.03934	13	1	4	4
CT208	LA4107	0.00081	14	1	3	8
CT208	LA4108	0.00130	15	1	1	5
CT208	LA4118	0.01961	15	1	0	2
CT208	LA4119	0.00016	14	1	2	9
CT208	LA4332	0.03934	13	1	4	4

Note: “S” synonymous, “NS” nonsynonymous

Table B3.30: Reference genes with significant McDonald-Kreitman test statistic with *S. ochranthum*.

Gene	Population	Fisher's exact test, p-value	Fixed differences		Polymorphism	
			S	NS	S	NS
CT066	LA1930	0.03000	22	6	15	16
CT066	LA2750	0.00716	21	5	25	27
CT066	LA2880	0.00435	31	5	2	5
CT066	LA4107	0.00012	29	5	7	15
CT066	LA4108	0.00016	31	5	6	12
CT066	LA4119	0.00046	22	5	16	27
CT066	LA4332	0.04540	23	5	16	13
CT093	LA2750	0.03699	7	4	11	29
CT093	LA4107	0.02374	9	4	1	7
CT093	LA4119	0.01150	8	4	1	9
CT114	LA4107	0.04739	4	3	1	10
CT114	LA4118	0.04977	5	3	1	8
CT143	LA2750	0.00110	7	0	6	16
CT143	LA2755	0.04274	6	0	5	6
CT143	LA2932	0.04546	7	0	2	3
CT143	LA4119	0.02778	7	0	0	2
CT179	LA2750	0.04429	8	1	10	13
CT182	LA2750	0.01522	6	1	2	8
CT192	LA2750	0.00131	11	3	4	16
CT251	LA2750	0.01747	24	23	11	32
CT268	LA0456	0.03229	15	19	26	11
GBSSI	LA2750	0.00368	7	5	6	37

Note: “S” synonymous, “NS” nonsynonymous

Table B3.31: Candidate genes with significant McDonald-Kreitman test statistic with *S. lycopersicoides*.

Gene	Population	Fisher's exact test, p-value	Fixed differences		Polymorphism	
			S	NS	S	NS
NtC7	LA4108	0.03361	44	43	1	8
AREB1	LA2750	0.00024	17	10	8	37
AREB1	LA4107	0.01057	19	11	3	12
AREB1	LA4108	0.00836	19	11	1	9
JERF1	LA2750	0.03211	14	13	8	26
DREB3	LA0458	0.04921	8	1	10	11
DREB3	LA2750	0.00540	5	0	15	33
DREB3	LA2932	0.04658	8	1	11	13
DREB3	LA3784	0.03301	10	1	17	15
DREB3	LA4107	0.00371	10	1	4	10
DREB3	LA4108	0.00170	9	1	5	14
dhn1	LA1930	0.01028	7	0	8	11
dhn1	LA1963	0.00140	8	0	1	7
dhn1	LA1968	0.00022	8	0	3	13
dhn1	LA2750	0.00009	8	0	4	18
dhn1	LA2765	0.03017	7	0	11	10
dhn1	LA2880	0.02222	8	0	0	2
dhn1	LA2932	0.01818	8	0	1	3
dhn1	LA3111	0.01499	8	0	2	4
dhn1	LA4107	0.00699	8	0	2	5
dhn1	LA4108	0.01499	8	0	2	4
dhn1	LA4119	0.00012	8	0	5	19
dhn1	LA4332	0.00041	8	0	1	8
TAS14	LA1930	0.04911	9	2	2	5
TAS14	LA2747	0.02742	10	2	3	7
TAS14	LA2748	0.04491	10	2	2	5
TAS14	LA2750	0.01757	11	2	2	6
TAS14	LA2753	0.00333	10	2	2	10
TAS14	LA2755	0.02763	10	2	1	4
TAS14	LA2931	0.04491	10	2	2	5
TAS14	LA3784	0.01228	10	2	3	8
TAS14	LA4332	0.01937	10	2	2	6
ER5	LA2750	0.04412	3	1	2	12
CT208	LA1968	0.01099	11	2	7	13
CT208	LA2750	0.00027	12	2	8	23
CT208	LA4107	0.00398	13	2	3	8
CT208	LA4108	0.00432	14	2	1	5
CT208	LA4118	0.03922	14	2	0	2
CT208	LA4119	0.00095	13	2	2	9

Note: “S” synonymous, “NS” nonsynonymous

Table B3.32: Reference genes with significant McDonald-Kreitman test statistic with *S. lycopersicoides*.

Gene	Population	Fisher's exact test, p-value	Fixed differences		Polymorphism	
			S	NS	S	NS
CT066	LA1930	0.01418	24	5	16	16
CT066	LA2750	0.00299	22	4	26	27
CT066	LA2880	0.00251	31	4	2	5
CT066	LA2932	0.04626	22	4	20	14
CT066	LA3784	0.04036	24	4	18	12
CT066	LA4107	0.00001	29	4	7	16
CT066	LA4108	0.00007	31	4	6	12
CT066	LA4119	0.00017	23	4	17	27
CT066	LA4332	0.02125	24	4	17	13
CT093	LA1930	0.02551	9	2	8	14
CT093	LA2750	0.00133	11	3	11	29
CT093	LA2753	0.02688	9	3	6	14
CT093	LA3784	0.01493	9	2	13	22
CT093	LA4107	0.00590	12	3	1	7
CT093	LA4119	0.00191	9	2	1	9
CT143	LA2750	0.01405	4	0	6	16
CT143	LA4119	0.04762	5	0	0	2
CT179	LA2750	0.04429	8	1	10	13
CT182	LA2750	0.03497	3	0	2	8
CT189	LA2750	0.00630	4	0	2	11
CT189	LA3784	0.00506	6	0	8	16
CT192	LA2750	0.00088	12	3	5	17
CT251	LA2750	0.00081	24	14	11	32
CT251	LA2880	0.01670	21	13	4	13
CT251	LA4118	0.01265	21	14	5	16
CT251	LA4119	0.04647	20	14	10	21
CT268	LA0456	0.00339	7	17	26	11
CT268	LA1958	0.03970	6	14	33	24
CT268	LA2748	0.00999	7	14	35	17
CT268	LA2931	0.03949	6	13	38	26
CT268	LA3784	0.04715	6	13	51	38
CT268	LA4119	0.03231	6	14	30	19
GBSSI	LA2750	0.02619	7	8	6	37

Note: “S” synonymous, “NS” nonsynonymous

Table B3.33: Proportion of adaptive amino acid substitutions from *S. ochranthum* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	-0.637	0.602	-2.407	0.187	-1.788	-0.833	-7.223	-0.112	n. a.	-1.882	n. a.	-0.420	-6.726	-1.988	-0.134	-0.399	-1.697
LA0458	-1.095	0.648	-0.089	0.438	-1.614	-0.141	0.643	0.444	n. a.	-1.224	n. a.	-0.183	n. a.	-1.019	n. a.	0.890	-0.192
LA1930	-0.017	-0.188	0.749	0.334	-1.544	0.250	-1.153	0.633	n. a.	-1.234	0.134	0.263	-0.925	-5.680	-22.212	0.155	-2.029
LA1958	0.087	0.507	-0.027	0.633	-0.071	0.545	-0.224	0.331	n. a.	0.369	-3.276	0.298	-3.816	-4.039	-1.279	-0.208	-0.678
LA1963	-0.262	0.491	-0.509	0.493	-0.399	0.294	-25.486	0.128	n. a.	-0.649	-2.154	-0.975	-1.746	-3.508	-0.719	0.760	-2.283
LA1968	-0.027	0.750	-0.075	0.554	-1.691	0.511	-1.303	0.155	n. a.	-2.387	n. a.	0.439	-3.629	-6.608	-10.162	0.257	-1.658
LA2747	-0.574	0.240	-0.170	-0.118	-2.168	0.330	0.746	0.116	n. a.	-0.407	-2.659	-0.562	-0.912	1.000	-0.178	0.845	-0.298
LA2748	0.056	0.065	-3.658	0.366	-0.618	0.613	0.363	0.379	n. a.	-0.469	-2.886	0.192	-0.967	-1.712	-0.222	0.296	-0.547
LA2750	-4.387	-2.218	n. a.	0.312	-9.024	-1.266	-19.303	-0.080	n. a.	-17.612	-1.214	0.059	-2.524	-10.396	-1.057	-1.070	-4.984
LA2753	-0.696	0.456	0.159	0.162	-1.009	-0.104	-107.076	0.335	n. a.	-1.174	-3.052	-0.975	-1.639	-5.855	-3.793	0.264	-8.266
LA2755	0.087	0.530	-0.053	-0.264	-1.318	0.333	0.760	0.397	n. a.	-0.485	-1.731	0.139	-4.210	0.627	-0.276	-0.016	-0.365
LA2765	-0.355	0.090	-0.388	0.255	-1.245	0.626	-0.076	0.304	n. a.	-1.361	-1.476	-0.447	-3.329	-3.888	-0.179	0.400	-0.738
LA2773	0.085	0.696	0.677	-0.062	-1.419	0.275	0.444	0.141	n. a.	-1.014	-2.244	-0.026	-0.747	-3.303	0.156	0.905	-0.362
LA2880	-0.718	0.910	-0.197	-0.704	-3.810	0.316	n. a.	0.466	n. a.	n. a.	-31.414	-0.641	-1.082	1.000	0.777	0.436	-2.666
LA2931	-0.054	0.570	-0.795	0.178	-0.099	0.457	0.713	0.476	n. a.	-1.225	-3.814	0.218	-1.260	-3.982	-0.960	0.287	-0.619
LA2932	-5.599	-0.147	0.494	0.684	-36.016	-0.097	-14.634	-0.467	n. a.	-3.142	n. a.	0.458	-0.928	-2.547	-1.534	0.263	-4.515
LA3111	-1.166	0.467	0.796	0.535	-1.782	0.482	-0.936	0.390	n. a.	-0.419	-1.706	0.232	-2.715	-3.565	-2.895	0.802	-0.765
LA3784	-1.011	0.071	0.192	0.212	-1.778	0.111	-0.215	0.677	n. a.	-1.860	-2.188	-2.324	-0.494	-23.465	-5.593	0.379	-2.486
LA4107	0.088	-8.018	-1.734	0.610	-1.556	-2.814	-8.499	0.235	n. a.	n. a.	n. a.	1.000	n. a.	-54.640	-19.378	-1.051	-7.980
LA4108	-10.437	-20.627	-0.143	0.543	-3.664	-2.276	-6.766	0.172	n. a.	-2.587	-6.382	n. a.	-35.102	-21.782	-0.090	0.731	-7.743
LA4118	-1.553	0.609	-0.427	-0.221	0.083	0.158	1.000	0.801	n. a.	-1.404	-4.259	-0.339	-2.877	n. a.	0.326	-2.017	-0.723
LA4119	-1.422	0.589	0.406	-1.257	-2.518	0.281	-11.421	0.310	n. a.	-8.167	-5.340	-0.277	-0.915	-56.180	-2.636	0.909	-5.843
LA4332	-0.680	0.068	-0.129	0.055	-0.053	0.250	-32.223	0.694	n. a.	-0.056	-1.178	-0.230	-2.024	-0.615	0.263	0.238	-2.375
Mean	-1.317	-0.993	-0.333	0.171	-3.265	-0.074	-10.539	0.301	n. a.	-2.304	-4.269	-0.186	-3.741	-9.643	-3.263	0.176	

Note: “n. a.” not available

Table B3.34: Proportion of adaptive amino acid substitutions from *S. ochranthum* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	0.301	-0.474	0.668	-0.573	n. a.	n. a.	0.719	n. a.	-0.224	-0.266	0.440	0.174	0.539	0.148	0.132
LA0458	0.537	0.007	0.792	0.228	-15.076	n. a.	-0.360	n. a.	n. a.	-1.550	1.000	-1.885	0.343	-16.253	-2.929
LA1930	-0.010	-1.636	-0.224	0.401	-4.714	-84.443	0.333	-4.352	-0.954	-0.647	0.538	0.269	0.114	0.471	-6.775
LA1958	0.707	0.310	0.041	0.141	-3.272	-6.639	0.476	-1.679	0.912	-1.331	0.635	0.156	0.267	0.156	-0.652
LA1963	0.435	0.291	0.016	-0.132	-3.973	-5.613	-0.066	-2.138	1.000	-0.866	0.390	0.236	0.187	0.162	-0.719
LA1968	-2.165	0.641	-0.568	0.251	-4.411	-5.252	-0.662	-0.881	-2.924	0.102	0.442	0.226	0.348	-0.042	-1.064
LA2747	0.557	0.252	0.047	0.241	-18.510	-6.986	0.038	-1.514	n. a.	-0.444	0.253	0.008	0.000	0.239	-1.986
LA2748	0.582	-0.132	-0.393	-0.331	-2.990	n. a.	0.847	1.000	n. a.	0.122	0.011	0.028	0.466	0.171	-0.052
LA2750	0.357	-0.370	-1.203	-4.042	-12.186	-32.890	-3.614	-5.320	-3.835	-17.408	-0.450	-3.651	-0.001	-8.248	-6.633
LA2753	0.191	-0.584	-0.300	0.117	-4.962	-4.157	0.359	-2.887	n. a.	-0.405	-1.212	0.340	0.353	-1.836	-1.153
LA2755	0.315	-0.052	-0.564	-1.236	-5.468	-3.707	0.118	0.451	0.897	-0.767	n. a.	0.304	0.332	-0.587	-0.766
LA2765	0.021	0.224	-0.391	-0.069	-6.847	-3.449	0.351	1.000	-0.290	-0.665	-0.004	0.374	0.507	-0.027	-0.662
LA2773	0.569	0.324	0.194	0.289	-17.156	n. a.	0.266	-0.815	n. a.	-0.117	0.318	0.163	0.372	-1.026	-1.385
LA2880	0.486	-11.564	0.412	n. a.	-14.911	-19.987	0.688	n. a.	1.000	-1.048	1.000	-4.664	0.567	n. a.	-4.366
LA2931	0.395	-0.298	0.112	-0.002	-5.114	-6.229	0.015	1.000	n. a.	-0.147	-0.515	-1.867	0.292	-0.354	-0.978
LA2932	-0.904	-0.679	0.465	-4.817	-19.229	n. a.	-2.615	-1.161	n. a.	-33.078	-0.836	-0.458	-0.766	n. a.	-5.825
LA3111	0.574	0.033	0.128	0.342	-11.960	n. a.	-0.288	-1.397	1.000	-0.957	-0.519	0.312	0.253	0.170	-0.947
LA3784	0.264	-0.093	-0.430	0.131	-5.177	-8.782	-0.758	-3.541	-1.646	-0.466	0.414	0.196	0.268	0.299	-1.380
LA4107	0.444	-5.313	-50.353	-8.451	-14.703	n. a.	-0.589	-4.999	n. a.	-1.928	0.857	-3.901	-0.138	-1.358	-7.536
LA4108	0.153	-1.309	-3.135	-2.377	-17.546	n. a.	0.588	0.458	n. a.	-1.842	-0.943	-1.588	-2.095	-6.883	-3.043
LA4118	-1.635	0.472	0.312	-9.087	-6.900	-201.884	-0.413	n. a.	n. a.	-0.976	-1.090	-1.377	0.345	-0.785	-18.585
LA4119	-0.523	-0.197	-18.301	-3.365	n. a.	-8.083	-1.845	-4.814	1.000	-0.968	n. a.	-5.750	0.547	0.581	-3.476
LA4332	0.427	0.318	0.133	-0.529	-5.770	-7.417	n. a.	-10.546	1.000	-0.946	-1.291	0.329	0.296	-0.995	-1.922
Mean	0.090	-0.862	-3.154	-1.494	-9.566	-27.035	-0.291	-2.218	-0.236	-2.895	-0.027	-0.958	0.148	-1.714	

Note: “n. a.” not available

Table B3.35: Proportion of adaptive amino acid substitutions from *S. lycopersicoides* for the candidate genes.

Population	*NtC7*	*AREB1*	*AREB2*	*JERF1*	*JERF3*	*DREB3*	*dhn1*	*pLC30-15*	*TAS14*	*ER5*	*le25*	*LTP*	*TSW12*	*CT208*	*His1*	*TPP*	Mean
LA0456	-0.806	0.666	-1.918	-0.017	-0.022	-2.547	-419.987	-0.046	-7.289	-1.517	n. a.	0.076	-7.896	-0.480	n. a.	-0.256	-31.574
LA0458	-1.311	0.706	0.017	0.260	-0.025	-1.193	-29.745	0.523	0.586	-0.953	n. a.	0.258	n. a.	0.170	n. a.	0.889	-2.294
LA1930	-0.159	-0.008	0.774	0.166	-0.236	-0.498	-6.317	0.200	-21.866	0.111	0.734	0.600	-0.582	-2.836	n. a.	0.406	-1.967
LA1958	-0.015	0.608	0.109	0.537	0.586	-0.016	-7.230	0.372	-13.940	0.366	-0.182	0.567	-2.181	-1.988	n. a.	-0.048	-1.497
LA1963	-0.378	0.608	-0.276	0.395	0.465	-0.562	-400.267	-0.122	-5.684	-0.361	0.097	-0.233	-2.586	-1.293	n. a.	0.773	-27.295
LA1968	-0.122	0.800	0.022	0.433	-0.033	-0.109	-18.666	-0.189	-20.421	-1.557	n. a.	0.657	-3.118	-2.951	n. a.	0.358	-3.207
LA2747	-0.736	0.386	-0.023	-0.429	-0.162	-0.383	-10.559	-0.329	-16.309	-0.199	-0.079	0.005	-2.916	1.000	n. a.	0.862	-1.991
LA2748	-0.029	0.229	-2.366	0.230	0.378	0.024	-15.108	-0.016	-50.181	-0.391	-0.163	0.505	-3.056	-0.257	n. a.	0.407	-4.653
LA2750	-4.938	-2.117	n. a.	0.102	-2.863	-3.995	-90.433	-0.133	-51.957	-15.554	0.354	0.388	-3.991	-5.685	n. a.	-0.758	-12.970
LA2753	-0.829	0.533	0.258	-0.029	0.282	-1.581	-10.973	-0.045	-8.954	-0.646	-0.180	-0.272	-1.508	-2.947	n. a.	0.337	-1.770
LA2755	0.001	0.602	0.004	-0.592	0.091	-0.320	-9.124	0.138	-21.452	-0.197	0.239	0.449	-3.301	0.841	n. a.	0.025	-2.173
LA2765	-0.495	0.212	-0.063	0.071	0.163	0.054	-8.794	0.195	-8.133	-0.856	0.223	0.114	-5.108	-1.467	n. a.	0.492	-1.559
LA2773	-0.010	0.759	0.704	-0.336	0.060	-0.361	-17.545	0.000	-7.642	-0.422	-0.258	0.354	-1.732	-1.141	n. a.	0.915	-1.777
LA2880	-0.922	0.926	-0.180	-1.244	-0.914	-0.579	n. a.	0.601	-7.816	n. a.	-7.438	-0.041	-2.466	1.000	n. a.	0.527	-1.427
LA2931	-0.158	0.644	-0.354	0.124	0.583	-0.348	-12.907	0.241	-8.053	-0.657	-0.310	0.535	-2.685	-1.566	n. a.	0.372	-1.636
LA2932	-6.245	-0.079	0.585	0.583	-13.054	-1.341	-392.262	-0.488	-92.168	-2.523	n. a.	0.663	-2.107	-0.658	n. a.	0.377	-36.337
LA3111	-1.388	0.564	0.804	0.409	-0.074	-0.312	-28.385	0.213	-2.578	-0.255	0.252	0.524	-2.330	-1.401	n. a.	0.841	-2.208
LA3784	-1.208	0.216	0.289	-0.016	-0.120	-0.972	-8.039	-0.200	-12.395	-1.090	-0.023	-0.991	-0.212	-20.555	n. a.	0.554	-2.984
LA4107	-0.004	-7.929	-1.424	0.487	0.008	-8.044	-199.368	-0.108	-6.263	n. a.	n. a.	1.000	n. a.	-28.426	n. a.	-1.155	-20.936
LA4108	-11.604	-19.631	-0.020	0.396	-0.708	-6.761	-195.657	-0.209	-6.225	-2.428	-1.069	n. a.	-61.218	-10.779	n. a.	0.734	-22.513
LA4118	-1.841	0.680	-0.385	-0.605	0.643	-0.554	n. a.	0.852	-7.691	-0.861	-0.443	0.169	-4.125	n. a.	n. a.	-1.567	-1.210
LA4119	-1.645	0.640	0.417	-1.235	-0.364	-0.458	-79.602	0.470	-8.719	-5.604	-0.724	0.362	-0.821	-27.864	n. a.	0.911	-8.282
LA4332	-0.836	0.149	-0.060	-0.231	0.583	-0.362	-400.048	0.540	-11.282	0.054	0.376	0.239	-4.636	0.254	n. a.	0.328	-27.662
Mean	-1.551	-0.862	-0.140	-0.024	-0.641	-1.357	-112.429	0.107	-17.236	-1.692	-0.477	0.270	-5.646	-4.956	n. a.	0.275	

Note: “n. a.” not available

Table B3.36: Proportion of adaptive amino acid substitutions from *S. lycopersicoides* for the reference genes.

Population	*CT021*	*CT066*	*CT093*	*CT114*	*CT143*	*CT166*	*CT179*	*CT182*	*CT189*	*CT192*	*CT198*	*CT251*	*CT268*	*GBSSI*	Mean
LA0456	n. a.	-0.973	0.424	-0.140	n. a.	n. a.	0.749	n. a.	-16.443	-0.363	0.440	-0.335	0.670	0.464	-1.551
LA0458	n. a.	-0.316	0.628	0.395	-9.328	n. a.	-0.305	n. a.	n. a.	-1.704	1.000	-3.379	0.549	-10.482	-2.294
LA1930	n. a.	-2.340	-1.428	0.530	-2.317	-67.909	0.421	-8.304	-11.156	-0.758	0.538	-0.268	0.374	0.639	-7.075
LA1958	n. a.	0.076	-0.902	0.314	-1.257	-5.345	0.495	-12.064	-7.709	-1.403	0.635	-0.376	0.465	0.463	-2.047
LA1963	n. a.	0.063	-0.726	0.112	-1.760	-4.419	-0.008	-14.210	n. a.	-1.011	0.390	-0.338	0.397	0.473	-1.753
LA1968	n. a.	0.532	-1.646	0.433	-2.175	-4.149	-0.657	-23.597	-207.875	-0.007	0.442	-0.273	0.517	0.336	-18.317
LA2747	n. a.	0.036	-0.727	0.405	-9.632	-5.414	0.046	-21.340	n. a.	-0.575	0.253	-0.555	0.294	0.522	-3.057
LA2748	n. a.	-0.427	-1.424	-0.011	-1.390	n. a.	0.849	n. a.	n. a.	0.043	0.011	-0.495	0.593	0.470	-0.178
LA2750	n. a.	-0.697	-3.516	-3.416	-7.830	-24.917	-3.568	-6.527	-51.316	-18.274	-0.450	-6.061	0.211	-5.165	-10.117
LA2753	n. a.	-1.030	-1.275	0.162	-2.214	-3.220	0.324	-7.484	n. a.	-0.532	-1.212	-0.194	0.522	-0.869	-1.419
LA2755	n. a.	-0.281	-1.710	-0.780	-2.970	-2.866	0.128	-13.463	-3.608	-0.897	n. a.	-0.281	0.523	-0.019	-2.185
LA2765	n. a.	-0.016	-1.397	0.153	-2.670	-3.502	0.392	n. a.	-68.358	-0.779	-0.004	-0.023	0.627	0.350	-6.269
LA2773	n. a.	0.157	-0.445	0.439	-9.329	n. a.	0.283	-6.132	n. a.	-0.221	0.318	-0.375	0.543	-0.293	-1.369
LA2880	n. a.	-18.314	-0.069	n. a.	-9.634	-15.039	0.713	n. a.	n. a.	-1.158	1.000	-7.828	0.620	n. a.	-5.523
LA2931	n. a.	-0.680	-0.579	0.226	-2.778	-4.966	0.068	1.000	n. a.	-0.276	-0.515	-3.396	0.440	0.143	-0.943
LA2932	n. a.	-0.966	-0.148	-3.991	-11.997	n. a.	-2.615	-1.882	n. a.	-34.987	-0.836	-1.277	-0.418	n. a.	-5.912
LA3111	n. a.	-0.245	-0.487	0.490	-3.989	n. a.	-0.267	-2.029	n. a.	-1.186	-0.519	-0.249	0.482	0.469	-0.685
LA3784	n. a.	-0.426	-1.745	0.290	-2.798	-6.711	-0.688	-14.317	-12.554	-0.563	0.414	-0.472	0.425	0.526	-2.971
LA4107	n. a.	-6.712	-85.840	-7.189	-9.638	n. a.	-0.746	-13.856	n. a.	-2.085	0.857	-6.403	0.050	-0.514	-12.007
LA4108	n. a.	-1.841	-6.058	-1.898	-11.770	n. a.	0.586	-14.998	n. a.	-1.992	-0.943	-3.032	-1.455	-4.094	-4.318
LA4118	n. a.	0.321	-0.238	-6.995	-5.219	-151.419	-0.413	n. a.	n. a.	-1.089	-1.090	-2.623	0.436	-0.146	-15.316
LA4119	n. a.	-0.717	-40.138	-1.997	n. a.	-6.174	-1.854	-145.226	n. a.	-1.079	n. a.	-9.297	0.606	0.738	-20.514
LA4332	n. a.	0.073	-0.553	-0.204	-2.908	-6.055	n. a.	-37.227	n. a.	-1.140	-1.291	-0.088	0.506	-0.274	-4.469
Mean	n. a.	-1.510	-6.522	-1.031	-5.410	-20.807	-0.276	-20.097	-47.377	-3.132	-0.027	-2.070	0.347	-0.774	

Note: "n. a." not available

Table B3.37: Outlier SNPs for balancing selection identified with BayeScan. List of the outlier SNPs (FDR 5 %, q value < 0.05) with negative α value and intermediate frequency (0.2 – 0.8) in at least one population from the run with multiple hits.

SNP	Gene	Type	α value	q value	Frequency 0.2 - 0.8
candidate genes					
90	*AREB1*	NS	-1.309	0.0001	LA2747, LA2748, LA2753, LA2773, LA4332
655	*dhn1*	S	-1.328	0.0213	LA2748
956[a]	*ER5*	intron	-1.360	0.0008	LA1930, LA2932
962[a]	*ER5*	NS	-1.555	0.0148	LA2753
1467[a]	*JERF3*	NS	-0.777	0.0423	LA1963, LA1968, LA2753, LA2755, LA2765, LA3111
1496	*JERF3*	intron	-4.064	0	LA2750, LA2755, LA2932
1865	*le25*	NS	-3.555	0	LA4119
1904[a]	*LTP*	NS	-1.023	0.0031	LA0456, LA2747, LA2748, LA2753, LA2755, LA2765, LA2773, LA2880, LA4119, LA4332
2079[a]	*NtC7*	intron	-1.348	0	LA0458, LA1958, LA1963, LA2747, LA2750, LA2753, LA2773, LA2931, LA2932, LA4332
2800	*CT208*	intron	-2.052	0	LA1930, LA1958, LA1963, LA2753, LA2765, LA2931, LA3111, LA3784
reference genes					
2940	*CT021*	NS	-0.998	0.0003	LA0456, LA1930, LA1958, LA1963, LA2747, LA2748, LA2750, LA2765, LA2773, LA2931, LA3111, LA4108, LA4118, LA4332
2976	*CT021*	intron	-1.327	0.0028	LA2747, LA2773, LA2932, LA4332
2977	*CT021*	intron	-0.688	0.0295	LA1963, LA1968, LA2747, LA2750, LA2765, LA2773, LA2931, LA3111, LA4118, LA4119, LA4332
3380[a]	*CT093*	NS	-1.284	0.0185	LA1958
4541[a]	*CT182*	intron	-1.234	0.0004	LA1958
4746	*CT192*	NS	-0.821	0.0284	LA0458, LA1930, LA2880, LA3111, LA4107, LA4108, LA4118, LA4119
4767	*CT192*	S	-0.995	0.0052	LA0458, LA1930, LA1963, LA2755, LA2880, LA3111, LA4107, LA4108, LA4118, LA4119
4851	*CT192*	intron	-0.840	0.0208	LA0458, LA1930, LA1963, LA2765, LA280, LA3111, LA4107, LA4108, LA4118, LA4119
4909	*CT192*	intron	-0.895	0.0156	LA0458, LA1930, LA1963, LA2880, LA3111, LA4107, LA4108, LA4118, LA4119
4946	*CT192*	intron	-0.734	0.0456	LA0458, LA1930, LA1963, LA2880, LA3111, LA4107, LA4108, LA4118, LA4119
4974	*CT192*	intron	-0.809	0.0300	LA0458, LA1930, LA1963, LA2880, LA3111, LA4107, LA4108, LA4118, LA4119
5130	*CT198*	intron	-1.718	0	LA1930, LA1958, LA1963, LA2747, LA2765, LA4118, LA4332
5166	*CT198*	3' UTR	-0.933	0.0180	LA1963, LA2750, LA2880, LA31111, LA3784, LA4118, LA4119
5181	*CT251*	NS	-0.653	0.0334	LA0458, LA1930, LA1963, LA1968, LA2747, LA2748, LA2753, LA2755, LA2765, LA2773, LA2880, LA2931, LA3111, LA3784, LA4119, LA4332
5340	*CT251*	S	-0.996	0.0006	LA0456, LA1930, LA1963, LA2753, LA2755, LA2765, LA2773, LA3111, LA4332
5388	*CT251*	S	-2.299	0	all except LA3784
5506[a]	*CT268*	S	-1.874	0.0051	LA3784
5642[a]	*CT268*	NS	-1.658	0.0071	LA2932
5735	*GBSSI*	NS	-3.931	0	all except LA2931

Note: "S" synonymous SNP, "NS" nonsynonymous SNP, "UTR" untranslated region
[a]: SNPs that were not identified in the run without multiple hits

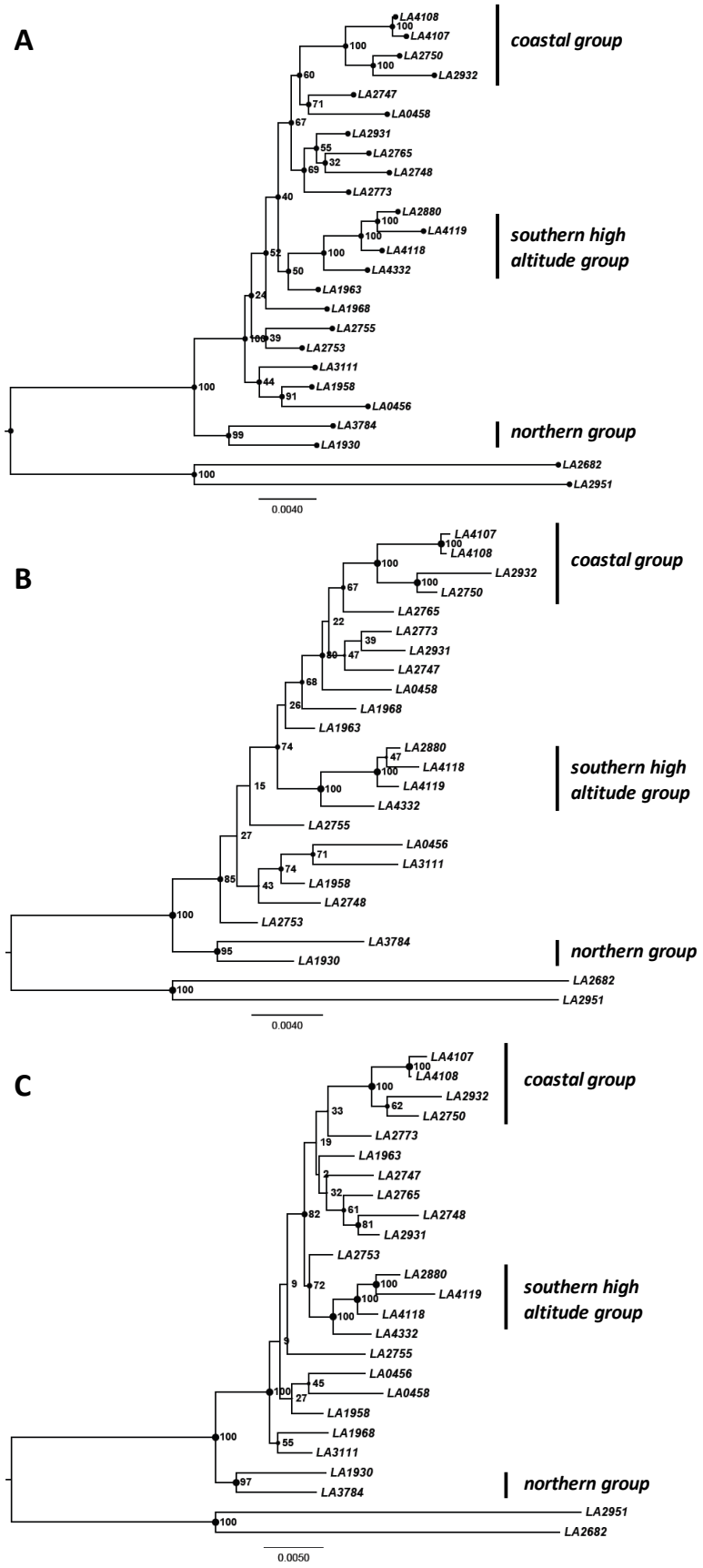

Figure B4.1: Phyolgenetic trees of the consensus sequences without *S. lycopersicum*. LA2682: sequence of the outgroup *S. ochranthum*, LA2951: sequence of the outgroup *S. lycopersicoides*. A) phylogenetic tree for all 30 genes, B) phylogenetic tree for the 14 reference genes, C) phylogenetic tree for the 16 candidate genes.

Table B4.1: Amino acid changes in the consensus sequence data set of *S. chilense*. *S. chilense* populations from north to south. Amino acid information of *S. lycopersicum*, *S. ochranthum* and *S. lycopersicoides* given. Amino acid changes shaded in gray.

Gene	Position[a]	*S. lycopersicum*	*S. ochranthum*	*S. lycopersicoides*	LA1930	LA3784	LA0456	LA1958	LA3111	LA1968	LA0458	LA1963	LA2773	LA2747	LA2765	LA2755	LA2753	LA2931	LA2748	LA2750	LA2932	LA4332	LA4118	LA4119	LA2880	LA4108	LA4107
NtC7	39	E	D	H	D	D	D	D	D	D	D	D	D	D	D	D	D	D	G	D	D	D	D	D	D	D	D
NtC7	40	L	I	I	V	V	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
NtC7	45	R	R	R	K	K	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
NtC7	69	F	Y	Y	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	L	L
NtC7	89	R	R	R	R	R	R	R	R	R	M	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
NtC7	97	N	D	D	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	N	N
NtC7	99	T	N	N	T	N	T	T	N	T	N	N	N	N	T	N	N	N	T	N	T	N	N	N	T	N	N
NtC7	100	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	K	E	E	E	E
NtC7	116	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	S	T	T	T	T	T	T	T	T
NtC7	124	R	H	Q	R	R	R	R	R	R	R	R	R	R	R	C	R	R	R	R	R	R	R	R	R	R	R
NtC7	136	G	T	T	G	G	G	G	D	G	D	D	G	D	D	D	D	D	G	D	D	D	D	G	G	D	D
NtC7	152	A	N	N	A	A	N	N	K	N	K	K	N	K	K	K	K	N	N	K	K	N	K	K	N	K	K
NtC7	162	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	C	F	F	F	F	F	F
NtC7	170	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	R	Q	Q	Q	Q	Q	R	Q	Q	Q	Q	Q
NtC7	171	A	A	A	A	S	A	S	S	A	A	S	A	S	S	A	A	S	A	S	S	A	A	S	S	S	S
NtC7	176	V	N	N	V	V	V	G	G	V	V	G	V	G	G	G	G	V	V	G	G	G	V	V	V	G	G
NtC7	225	G	A	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	R	R	G	G
NtC7	255	S	S	S	T	T	T	T	T	T	T	T	T	T	T	S	T	T	T	T	T	T	T	T	T	T	T
NtC7	269	F	A	A	F	A	A	A	A	A	A	F	F	A	F	F	F	A	A	A	F	F	A	A	A	F	F
NtC7	285	I	I	I	I	M	M	M	I	I	M	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
AREB1	77	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	I	I
AREB1	126	G	A	G	G	R	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
AREB1	131	S	S	S	S	C	S	S	S	S	S	C	S	S	S	S	S	S	S	S	S	S	S	S	S	C	C
AREB1	170	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
AREB1	249	S	P	P	S	S	S	S	S	S	S	S	S	S	S	S	S	S	P	S	S	S	S	S	S	S	S
AREB1	251	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	S	P	P	P	P	P	P	P	P
AREB1	293	A	A	A	T	T	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A
AREB1	309	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	T	M	M	M	M	M
AREB1	342	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	H
AREB1	438	C	*na*	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	S	C	C	C	C	C	C	C	C
AREB2	68	A	A	A	V	A	A	A	A	A	A	A	A	A	V	A	A	V	V	A	V	A	A	A	A	A	A
AREB2	177	T	T	T	T	S	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
AREB2	235	T	T	T	T	T	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
AREB2	238	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	A	V	V	V	V	V	V	V	V	V
JERF1	29	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	F	F	S	S	S
JERF1	73	D	D	D	E	D	E	D	D	E	D	E	D	D	D	E	E	D	D	D	D	D	D	D	D	D	D
JERF1	85	R	R	R	K	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
JERF1	91	C	C	C	C	C	S	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
JERF1	196	I	I	I	T	T	T	T	T	T	T	T	T	T	I	T	T	T	T	I	I	I	I	I	I	I	I
JERF1	226	F	F	F	F	F	I	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
JERF1	237	S	P	P	Y	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
JERF1	240	A	A	A	D	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
JERF1	287	F	F	V	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	S	S
JERF1	366	V	V	V	V	V	I	V	V	V	V	V	V	V	V	I	V	V	V	I	I	V	V	V	V	I	I
JERF3	30	D	N	D	D	D	D	D	N	N	N	D	D	D	D	N	D	D	D	D	D	D	D	D	D	D	D
JERF3	47	H	H	H	H	H	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
JERF3	52	V	V	V	V	V	V	I	I	I	V	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
JERF3	88	T	T	T	T	T	T	T	T	T	T	T	T	T	T	S	T	T	T	T	T	T	T	T	T	T	T
JERF3	93	K	K	K	R	S	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K
JERF3	172	D	D	D	D	D	V	D	D	D	D	D	V	D	D	D	V	D	D	V	V	D	D	D	D	V	V
JERF3	174	A	A	A	A	A	A	A	A	A	A	A	A	T	T	A	A	A	A	A	A	T	T	T	T	A	A
JERF3	188	Q	Q	Q	K	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
JERF3	207	N	N	N	N	N	R	N	N	N	N	N	R	N	N	N	R	R	K	R	R	N	N	N	N	R	R
JERF3	208	S	S	S	S	G	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
DREB3	17	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	P	P	Q	Q
DREB3	32	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	L	Q	Q	Q	Q	Q	Q	Q
DREB3	192	S	S	S	S	S	A	S	S	S	A	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
dhn1	56	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	Q	H	H	H
dhn1	57	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	A	A	T	T
dhn1	58	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	G	T	T	G	G	T	T
dhn1	72	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A
pLC30-15	14	E	E	E	E	G	E	E	G	E	E	E	E	G	G	G	G	G	G	E	E	G	E	E	E	E	E
pLC30-15	40	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S
pLC30-15	42	A	A	A	A	A	A	A	V	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
pLC30-15	57	N	N	N	N	K	N	N	K	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
pLC30-15	75	L	L	L	L	L	L	H	L	L	H	L	L	L	H	L	L	L	L	L	H	H	H	H	H	L	L
pLC30-15	120	K	K	K	K	K	K	K	K	K	K	K	K	K	K	N	K	K	K	K	K	K	K	K	K	K	K
pLC30-15	138	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	T	A	T	T	T	T	T	A	A

Table B4.1: continued.

pLC30-15	171	P	P	P	P	P	P	P	T	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
pLC30-15	172	P	P	P	P	L	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
pLC30-15	173	A	A	A	A	A	A	A	A	A	A	A	A	V	A	A	A	A	A	A	A	A	A	A	A	A	A
TAS14	27	Q	*na*	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	P	Q	Q	Q	Q	Q	Q	Q
TAS14	28	G	*na*	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	S	G	G	G	G	G	G	G
TAS14	30	G	*na*	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	D	G	G	G	G	G	G	G
TAS14	61	M	*na*	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	V
TAS14	62	M	*na*	-	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	V	M	M	M
TAS14	63	G	*na*	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	R	G	G	G	G
TAS14	115	H	*na*	H	H	H	H	H	H	H	H	H	H	H	H	H	Y	H	H	H	H	H	H	H	H	H	H
TAS14	150	M	*na*	M	M	M	V	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
TAS14	156	Q	*na*	-	Q	Q	Q	Q	E	Q	Q	E	Q	Q	Q	Q	E	Q	Q	E	E	Q	E	E	Q	E	E
TAS14	157	H	*na*	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	R	H	H	H	H	H
TAS14	159	G	*na*	G	G	V	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
TAS14	169	-	*na*	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
TAS14	182	K	*na*	K	K	K	K	N	K	K	K	K	N	N	N	N	N	N	K	K	K	N	K	K	K	N	N
ER5	35	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	K	S	S	S
ER5	73	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	T	T	P	P	P	P	T	T
ER5	76	G	G	G	G	G	G	G	G	G	R	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
ER5	90	V	V	V	V	V	V	V	V	V	V	V	L	V	V	V	V	V	V	V	V	V	V	V	V	V	V
ER5	113	K	K	K	T	R	T	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K
ER5	127	D	N	D	N	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
le25	21	N	N	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	N	S	S	S
le25	50	A	A	A	A	A	G	G	G	G	A	G	A	A	A	G	G	G	A	A	A	G	G	A	A	A	A
LTP	33	T	S	V	S	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
LTP	35	A	A	A	A	A	S	S	S	S	S	A	A	A	A	A	A	A	A	A	A	A	S	S	S	A	A
LTP	49	N	G	G	D	G	D	D	D	D	D	G	G	G	G	G	G	G	G	G	G	G	G	G	G	D	D
LTP	72	T	T	T	N	T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
LTP	79	S	T	A	S	S	S	S	S	S	S	S	S	A	S	S	S	S	S	S	S	S	S	S	S	S	S
TSW12	42	Q	Q	Q	K	K	Q	Q	K	K	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
TSW12	44	R	R	R	S	S	R	R	S	S	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
TSW12	52	G	G	G	G	G	N	N	G	G	N	N	N	N	N	G	N	N	N	N	N	N	N	N	N	N	N
TSW12	56	N	G	G	S	S	G	G	S	S	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
TSW12	85	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	N	D	N	D	D	D
His1	12	V	V	*na*	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	T	V	V	V
His1	53	V	V	*na*	A	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
His1	63	K	K	*na*	K	T	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K
His1	70	N	N	*na*	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	H	H	N	N	N	N	N	N
His1	128	K	K	*na*	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	E
His1	131	T	T	*na*	T	T	T	T	T	T	T	T	T	T	T	A	T	T	T	T	T	T	T	T	T	T	T
His1	140	P	P	*na*	P	P	L	L	L	L	P	L	L	L	L	P	L	L	L	L	L	L	L	P	L	P	P
His1	144	S	S	*na*	S	S	S	S	S	S	P	S	S	S	S	S	S	S	S	S	S	S	S	P	S	S	S
His1	148	P	P	*na*	P	P	T	T	T	T	P	T	P	T	T	P	T	T	T	T	T	T	T	P	T	T	T
TPP	17	L	*na*	L	L	L	L	L	L	L	L	L	L	L	L	M	L	L	L	L	L	L	L	L	L	L	L
TPP	81	S	S	S	S	S	S	S	S	P	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
TPP	95	R	R	C	R	R	R	R	R	R	R	R	R	R	R	C	R	R	R	R	R	R	R	R	R	R	R
TPP	196	I	V	I	V	V	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
TPP	229	V	V	V	V	V	V	I	V	V	V	V	V	V	V	I	V	V	V	V	V	V	V	V	V	V	V
CT021	223	T	T	*na*	N	T	T	T	T	T	T	T	N	N	N	T	T	T	T	T	N	N	T	T	T	T	T
CT021	252	R	R	*na*	R	R	R	R	R	R	S	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
CT021	262	L	L	*na*	L	W	L	L	L	W	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
CT021	301	G	G	*na*	R	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
CT021	303	S	S	*na*	T	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
CT066	131	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	F	Y	Y	Y	Y	Y
CT066	217	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	F	L	L	L	L	L
CT066	240	A	A	A	E	E	E	E	A	A	A	A	E	A	E	A	E	E	E	E	A	A	E	A	A	A	A
CT066	443	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	L	R	R	R	R	R	R	R
CT093	4	D	D	D	D	D	D	Y	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
CT093	192	N	N	N	N	D	D	N	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
CT093	304	H	R	R	R	H	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
CT114	131	D	D	D	D	D	D	D	E	D	E	E	E	E	E	D	D	D	D	E	E	D	E	E	E	E	E
CT182	133	P	P	P	P	P	P	P	S	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
CT182	184	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S	N	N	N	N	N	N
CT192	111	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	I	T	T	T
CT192	138	S	S	S	S	N	N	N	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S
CT192	148	I	T	T	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	T	T	T	T	I	I
CT192	152	A	A	A	A	A	A	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CT198	122	S	S	S	S	S	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S	N	N	N	N	N	N
CT251	37	V	V	V	V	V	V	V	V	V	L	V	L	V	L	V	V	L	L	V	V	V	L	L	L	V	V
CT251	86	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
CT251	201	V	V	V	V	V	V	V	V	V	V	V	I	V	V	V	V	V	V	V	V	V	V	V	V	V	V
CT251	213	P	P	P	A	P	P	A	P	A	A	A	A	A	A	A	P	A	A	A	A	A	A	A	A	A	A
CT251	217	V	V	V	A	V	A	A	A	A	A	A	A	A	A	A	V	A	A	A	A	A	A	A	A	A	A
CT251	249	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	D	D	N	N	N	N	N	N
CT251	252	V	V	V	V	V	V	A	V	A	A	A	A	A	A	V	A	A	A	A	A	A	A	A	A	A	A
CT251	314	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	Q	Q	Q	P	P	P	P	Q	Q
CT251	320	T	T	T	T	S	T	T	S	T	T	T	T	T	T	S	T	T	T	T	T	T	T	T	T	T	T

Table B4.1: continued.

CT251	332	V	A	A	V	A	V	V	A	V	V	V	V	V	V	A	V	V	V	V	V	V	V	V	V	V	V
CT251	351	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	R
CT251	371	E	E	E	G	E	G	G	E	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
CT251	410	S	S	S	L	S	S	L	S	L	L	L	L	L	L	S	S	L	L	L	L	L	L	L	L	L	L
CT251	415	V	V	V	I	V	V	I	V	I	I	I	I	I	I	I	V	I	I	I	I	I	I	I	I	I	I
CT251	444	V	M	M	M	M	M	M	V	V	V	M	V	V	V	M	M	M	V	V	V	V	V	V	V	V	V
CT268	113	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	K	Q	Q	Q	Q	Q
CT268	158	E	Q	E	E	E	E	G	G	G	G	G	G	E	G	E	E	G	G	E	E	G	E	E	E	E	E
CT268	214	S	S	S	S	S	A	S	S	S	S	S	A	A	S	S	S	S	A	S	S	A	A	A	A	S	S
CT268	366	P	P	P	P	P	P	P	P	P	P	P	P	P	P	H	P	P	P	P	P	P	P	P	P	P	P
CT268	377	D	D	D	D	G	D	D	D	D	D	D	D	G	D	D	D	D	D	D	D	G	D	D	D	G	G
CT268	398	P	P	P	P	P	P	P	P	P	S	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
CT268	405	T	S	S	S	T	T	T	T	T	S	T	T	T	T	S	T	T	T	T	T	T	T	T	T	S	S
CT268	448	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	N	T	T	T	T	T	T	T	T	T
CT268	472	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	S	P	P	P	P	P	P	P	P	P
CT268	476	T	T	T	A	T	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
CT268	516	A	A	A	A	A	A	A	A	S	S	S	S	S	A	S	A	S	A	A	A	S	A	A	A	A	A
CT268	525	F	F	F	S	S	C	C	F	S	S	S	S	S	C	S	F	F	C	F	F	S	C	C	C	F	F
CT268	530	C	F	F	F	F	F	F	C	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
CT268	549	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G
CT268	560	P	P	P	L	P	L	L	L	L	P	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
CT268	595	S	S	S	S	S	P	P	S	S	S	S	S	S	P	S	P	S	P	S	S	S	P	P	P	S	S
GBSSI	57	K	K	K	T	T	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K	K
GBSSI	185	L	L	L	L	L	L	L	L	L	M	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L

Note: "na" not available

[a]: amino acid position relative to the first amino acid of the *S. lycopersicum* (SL2.40) sequence

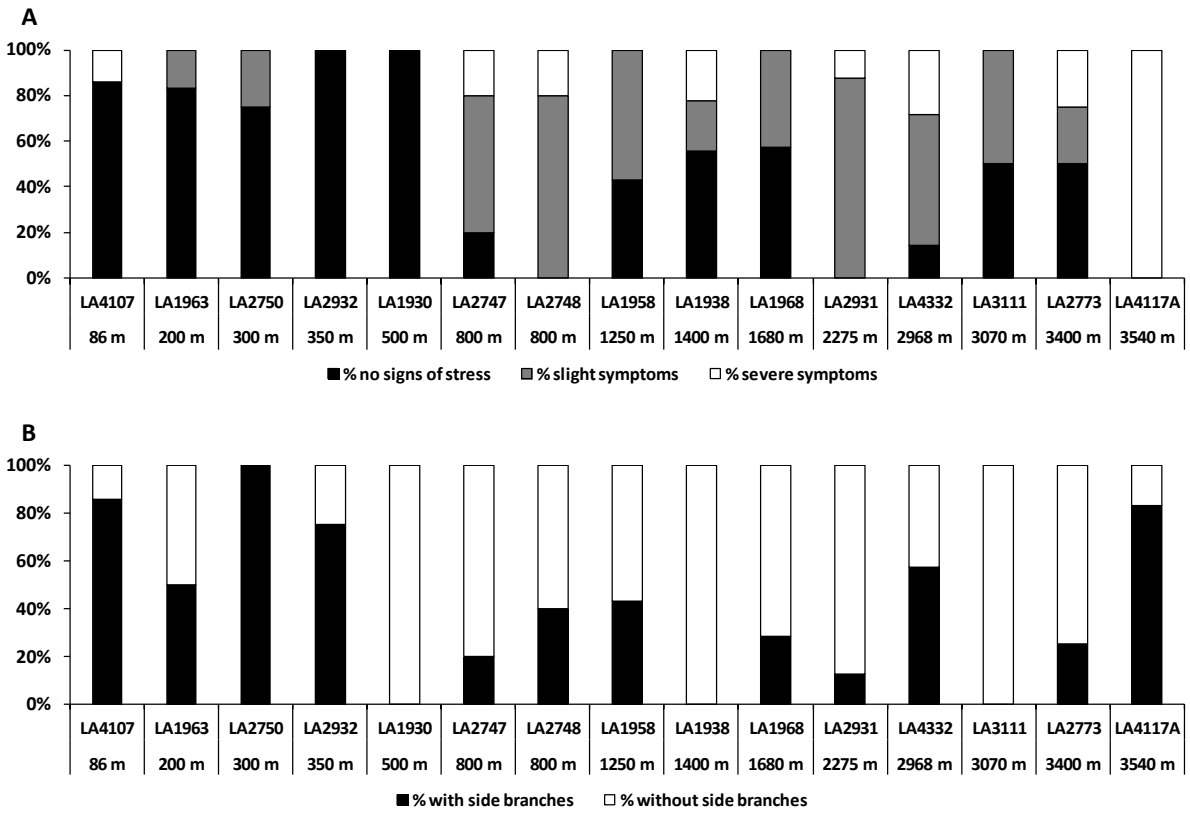

Figure B5.1: Results of population analysis of the young *S. chilense* plants after the salt treatment. Overall conditions (A) and development of side branches (B) after four weeks of recovery. Altitude of each population is given.

Bibliography

Abuqamar S., Luo H., Laluk K., Mickelbart M.V. & Mengiste T. (2009) Crosstalk between biotic and abiotic stress responses in tomato is mediated by the AIM1 transcription factor. *Plant J* 58, 347-60.

Abzhanov A. (2010) Darwin's Galapagos finches in modern biology. *Philos Trans R Soc Lond B Biol Sci* 365, 1001-7.

Albrecht E., Escobar M. & Chetelat R.T. (2010) Genetic diversity and population structure in the tomato-like nightshades Solanum lycopersicoides and S. sitiens. *Ann Bot* 105, 535-54.

Allagulova Ch R., Gimalov F.R., Shakirova F.M. & Vakhitov V.A. (2003) The plant dehydrins: structure and putative functions. *Biochemistry (Mosc)* 68, 945-51.

Alvarez-Ponce D. & Fares M.A. (2012) Evolutionary rate and duplicability in the Arabidopsis thaliana protein-protein interaction network. *Genome Biol Evol* 4, 1263-74.

Amaral A.J., Ferretti L., Megens H.J., Crooijmans R.P., Nie H., Ramos-Onsins S.E., Perez-Enciso M., Schook L.B. & Groenen M.A. (2011) Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. *PLoS One* 6, e14782.

Arunyawat U., Stephan W. & Stadler T. (2007) Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. *Mol Biol Evol* 24, 2310-22.

Barrett R.D. & Schluter D. (2008) Adaptation from standing genetic variation. *Trends Ecol Evol* 23, 38-44.

Battaglia M., Olvera-Carrillo Y., Garciarrubio A., Campos F. & Covarrubias A.A. (2008) The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* 148, 6-24.

Baudo M.M., Meza-Zepeda L.A., Palva E.T. & 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, 331-6.

Baudry E., Kerdelhue C., Innan H. & Stephan W. (2001) Species and recombination effects on DNA variability in the tomato genus. *Genetics* 158, 1725-35.

Baxter I., Brazelton J.N., Yu D., Huang Y.S., Lahner B., Yakubova E., Li Y., Bergelson J., Borevitz J.O., Nordborg M., Vitek O. & Salt D.E. (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genet* 6, e1001193.

Böndel K. (2010) Sequence evolution of resistance genes across species boundaries in wild tomatoes. *Diplom Thesis, LMU Munich*.

Borsani O., Cuartero J., Fernandez J.A., Valpuesta V. & Botella M.A. (2001) Identification of two loci in tomato reveals distinct mechanisms for salt tolerance. *Plant Cell* 13, 873-87.

Borsani O., Cuartero J., Valpuesta V. & Botella M.A. (2002) Tomato tos1 mutation identifies a gene essential for osmotic tolerance and abscisic acid sensitivity. *Plant J* 32, 905-14.

Boyer J.S. (1982) Plant productivity and environment. *Science* 218, 443-8.

Brücher P. (2013) Evolutionsgenetische Analyse von Toleranzgenen in Solanum chilense. *Bachelor Thesis, LMU Munich*.

Buhot N., Gomes E., Milat M.L., Ponchet M., Marion D., Lequeu J., Delrot S., Coutos-Thevenot P. & Blein J.P. (2004) Modulation of the biological activity of a tobacco LTP1 by lipid complexation. *Mol Biol Cell* 15, 5047-52.

Burbidge A., Grieve T.M., Jackson A., Thompson A., McCarty D.R. & Taylor I.B. (1999) Characterization of the ABA-deficient tomato mutant notabilis and its relationship with maize Vp14. *Plant J* 17, 427-31.

Busch J.W., Joly S. & Schoen D.J. (2011) Demographic signatures accompanying the evolution of selfing in Leavenworthia alabamica. *Mol Biol Evol* 28, 1717-29.

Catalan A., Hutter S. & Parsch J. (2012) Population and sex differences in Drosophila melanogaster brain gene expression. *BMC Genomics* 13, 654.

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

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

Chetelat R.T., Pertuze R.A., Faundez L., Graham E.B. & Jones C.M. (2009) Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. *Euphytica* 167, 77-93.

Cohen A. & Bray E.A. (1990) Characterization of three mRNAs that accumulate in wilted tomato leaves in response to elevated levels of endogenous abscisic acid. *Planta* 182, 27-33.

Cohen A. & Bray E.A. (1992) Nucleotide sequence of an ABA-induced tomato gene that is expressed in wilted vegetative organs and developing seeds. *Plant Mol Biol* 18, 411-3.

Cohen A., Plant A.L., Moses M.S. & Bray E.A. (1991) Organ-Specific and Environmentally Regulated Expression of Two Abscisic Acid-Induced Genes of Tomato : Nucleotide Sequence and Analysis of the Corresponding cDNAs. *Plant Physiol* 97, 1367-74.

Colosimo P.F., Hosemann K.E., Balabhadra S., Villarreal G., Jr., Dickson M., Grimwood J., Schmutz J., Myers R.M., Schluter D. & Kingsley D.M. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307, 1928-33.

Cook L.M. (2003) The rise and fall of the Carbonaria form of the peppered moth. *Q Rev Biol* 78, 399-417.

Cutler S.R., Rodriguez P.L., Finkelstein R.R. & Abrams S.R. (2010) Abscisic acid: emergence of a core signaling network. *Annu Rev Plant Biol* 61, 651-79.

Darwin C. (1859) On the Origin of Species by Means of Natural Selection. *Murray, London*.

Davies J. & Davies D. (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* 74, 417-33.

Delph L.F. & Kelly J.K. (2014) On the importance of balancing selection in plants. *New Phytol* 201, 45-56.

Eckert C.G., Samis K.E. & Lougheed S.C. (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Mol Ecol* 17, 1170-88.

Fay J.C., Wyckoff G.J. & Wu C.I. (2001) Positive and negative selection on the human genome. *Genetics* 158, 1227-34.

Fernandez O., Bethencourt L., Quero A., Sangwan R.S. & Clement C. (2010) Trehalose and plant stress responses: friend or foe? *Trends Plant Sci* 15, 409-17.

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, 1032-44.

Fischer I., Steige K.A., Stephan W. & Mboup M. (2013a) Sequence evolution and expression regulation of stress-responsive genes in natural populations of wild tomato. *PLoS One* 8, e78182.

Fischer M.C., Rellstab C., Tedder A., Zoller S., Gugerli F., Shimizu K.K., Holderegger R. & Widmer A. (2013b) Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. *Mol Ecol* 22, 5594-607.

Flagel L.E. & Wendel J.F. (2009) Gene duplication and evolutionary novelty in plants. *New Phytol* 183, 557-64.

Flowers T.J. (2004) Improving crop salt tolerance. *J Exp Bot* 55, 307-19.

Foll M. & Gaggiotti O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180, 977-93.

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, 436-42.

Glemin S., Bazin E. & Charlesworth D. (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc Biol Sci* 273, 3011-9.

Godfray H.C., Beddington J.R., Crute I.R., Haddad L., Lawrence D., Muir J.F., Pretty J., Robinson S., Thomas S.M. & Toulmin C. (2010) Food security: the challenge of feeding 9 billion people. *Science* 327, 812-8.

Godoy J.A., Lunar R., Torres-Schumann S., Moreno J., Rodrigo R.M. & Pintor-Toro J.A. (1994) Expression, tissue distribution and subcellular localization of dehydrin TAS14 in salt-stressed tomato plants. *Plant Mol Biol* 26, 1921-34.

Godoy J.A., Pardo J.M. & Pintor-Toro J.A. (1990) A tomato cDNA inducible by salt stress and abscisic acid: nucleotide sequence and expression pattern. *Plant Mol Biol* 15, 695-705.

Golldack D., Luking I. & Yang O. (2011) Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Rep* 30, 1383-91.

Gong P., Zhang J., Li H., Yang C., Zhang C., Zhang X., Khurram Z., Zhang Y., Wang T., Fei Z. & Ye Z. (2010) Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *J Exp Bot* 61, 3563-75.

Graham E.B., Shannon S.M., Petersen J.P. & Chetelat R.T. (2003) A self-compatible population of *Lycopersicon peruvianum* collected from N. Chile. *TGC* 53, 22-4.

Grant B.R. (2003) Evolution in Darwin's finches: a review of a study on Isla Daphne Major in the Galapagos archipelago. *Zoology (Jena)* 106, 255-9.

Grivet D., Sebastiani F., Alia R., Bataillon T., Torre S., Zabal-Aguirre M., Vendramin G.G. & Gonzalez-Martinez S.C. (2011) Molecular footprints of local adaptation in two Mediterranean conifers. *Mol Biol Evol* 28, 101-16.

Gur A. & Zamir D. (2004) Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol* 2, e245.

Gutierrez J.R. & Meserve P.L. (2003) El Nino effects on soil seed bank dynamics in north-central Chile. *Oecologia* 134, 511-7.

Hancock A.M., Brachi B., Faure N., Horton M.W., Jarymowycz L.B., Sperone F.G., Toomajian C., Roux F. & Bergelson J. (2011) Adaptation to climate across the Arabidopsis thaliana genome. *Science* 334, 83-6.

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, 684-91.

Hazzouri K.M., Escobar J.S., Ness R.W., Killian Newman L., Randle A.M., Kalisz S. & Wright S.I. (2013) Comparative population genomics in Collinsia sister species reveals evidence for

reduced effective population size, relaxed selection, and evolution of biased gene conversion with an ongoing mating system shift. *Evolution* 67, 1263-78.

Hernandez R.D., Williamson S.H. & Bustamante C.D. (2007) Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol Biol Evol* 24, 1792-800.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25, 1965-78.

Hirayama T. & Shinozaki K. (2010) Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J* 61, 1041-52.

Holmgren M., Scheffer M., Ezcurra E., Gutierrez J.R. & Mohren G.M. (2001) El Nino effects on the dynamics of terrestrial ecosystems. *Trends Ecol Evol* 16, 89-94.

Hörger A.C. (2011) Evolution of disease resistance genes in wild tomato species. *Dissertation, LMU Munich*.

Hörger A.C., Böndel K., Laurent S., Rose L.E., Stephan W. & Tellier A. (in preparation) Balancing selection at a pathogen resistance gene favours introgression between two closely related tomato species.

Houston J. (2006) Variability of precipitation in the Atacama Desert: Its causes and hydrological impact. *Int. J. Climatol.* 26, 2181-98.

Hudson R.R., Slatkin M. & Maddison W.P. (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* 132, 583-9.

Imai R., Chang L., Ohta A., Bray E.A. & Takagi M. (1996) A lea-class gene of tomato confers salt and freezing tolerance when expressed in Saccharomyces cerevisiae. *Gene* 170, 243-8.

Innan H. & Kim Y. (2008) Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* 179, 1713-20.

Islam M.S. & Wang M.H. (2009) Expression of dehydration responsive element-binding protein-3 (DREB3) under different abiotic stresses in tomato. *BMB Rep* 42, 611-6.

Kahn T.L., Fender S.E., Bray E.A. & O'Connell M.A. (1993) Characterization of Expression of Drought- and Abscisic Acid-Regulated Tomato Genes in the Drought-Resistant Species Lycopersicon pennellii. *Plant Physiol* 103, 597-605.

Kodama Y., Tamura T., Hirasawa W., Nakamura K. & Sano H. (2009) A novel protein phosphorylation pathway involved in osmotic-stress response in tobacco plants. *Biochimie*.

Kovacs D., Agoston B. & Tompa P. (2008a) Disordered plant LEA proteins as molecular chaperones. *Plant Signal Behav* 3, 710-3.

Kovacs D., Kalmar E., Torok Z. & Tompa P. (2008b) Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiol* 147, 381-90.

Kuck P. & Meusemann K. (2010) FASconCAT: Convenient handling of data matrices. *Mol Phylogenet Evol* 56, 1115-8.

Leimu R. & Fischer M. (2008) A meta-analysis of local adaptation in plants. *PLoS One* 3, e4010.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. & Durbin R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-9.

Librado P. & Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451-2.

Liu H., Ouyang B., Zhang J., Wang T., Li H., Zhang Y., Yu C. & Ye Z. (2012) Differential modulation of photosynthesis, signaling, and transcriptional regulation between tolerant and sensitive tomato genotypes under cold stress. *PLoS One* 7, e50785.

Lunter G. & Goodson M. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* 21, 936-9.

Maddison W.P. & Maddison D.R. (2010) Mesquite: a modular system for evolutionary analysis. Version 2.74 http://mesquiteproject.org.

Magadum S., Banerjee U., Murugan P., Gangapur D. & Ravikesavan R. (2013) Gene duplication as a major force in evolution. *J Genet* 92, 155-61.

Manga V.K. & Yadav O.P. (1995) Effect of seed size on developmental traits and ability to tolerate drought in pearl millet. *Journal of Arid Environments* 29, 169-72.

Mantel N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27, 209-20.

Mardis E.R. (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9, 387-402.

Mathew L.A., Staab P.R., Rose L.E. & Metzler D. (2013) Why to account for finite sites in population genetic studies and how to do this with Jaatha 2.0. *Ecol Evol* 3, 3647-62.

Mazel D. & Davies J. (1999) Antibiotic resistance in microbes. *Cell Mol Life Sci* 56, 742-54.

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, 3641-52.

McDonald J.H. & Kreitman M. (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351, 652-4.

Metzker M.L. (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11, 31-46.

Mittler R. (2006) Abiotic stress, the field environment and stress combination. *Trends Plant Sci* 11, 15-9.

Miyazono K., Miyakawa T., Sawano Y., Kubota K., Kang H.J., Asano A., Miyauchi Y., Takahashi M., Zhi Y., Fujita Y., Yoshida T., Kodaira K.S., Yamaguchi-Shinozaki K. & Tanokura M. (2009) Structural basis of abscisic acid signalling. *Nature* 462, 609-14.

Moyle L.C. (2008) Ecological and evolutionary genomics in the wild tomatoes (Solanum sect. Lycopersicon). *Evolution* 62, 2995-3013.

Moyle L.C. & Muir C.D. (2010) Reciprocal insights into adaptation from agricultural and evolutionary studies in tomato. *Evolutionary Applications* 3, 409-21.

Mullen M.P., Creevey C.J., Berry D.P., McCabe M.S., Magee D.A., Howard D.J., Killeen A.P., Park S.D., McGettigan P.A., Lucy M.C., Machugh D.E. & Waters S.M. (2012) Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples. *BMC Genomics* 13, 16.

Naduvilezhath L., Rose L.E. & Metzler D. (2011) Jaatha: a fast composite-likelihood approach to estimate demographic parameters. *Mol Ecol* 20, 2709-23.

Nakazato T., Franklin R.A., Kirk B.C. & Housworth E.A. (2012) Population structure, demographic history, and evolutionary patterns of a green-fruited tomato, Solanum peruvianum (Solanaceae), revealed by spatial genetics analyses. *Am J Bot* 99, 1207-16.

Nakazato T. & Housworth E.A. (2011) Spatial genetics of wild tomato species reveals roles of the Andean geography on demographic history. *Am J Bot* 98, 88-98.

Nakazato T., Warren D.L. & Moyle L.C. (2010) Ecological and geographic modes of species divergence in wild tomatoes. *Am J Bot* 97, 680-93.

Nei M. (1987) Molecular evolutionary genetics. *Columbia University Press, New York*.

Nookaew I., Papini M., Pornputtapong N., Scalcinati G., Fagerberg L., Uhlen 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, 10084-97.

Obbard D.J., Welch J.J., Kim K.W. & Jiggins F.M. (2009) Quantifying adaptive evolution in the Drosophila immune system. *PLoS Genet* 5, e1000698.

Olson-Manning C.F., Wagner M.R. & Mitchell-Olds T. (2012) Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat Rev Genet* 13, 867-77.

Orellana S., Yanez M., Espinoza A., Verdugo I., Gonzalez E., Ruiz-Lara S. & Casaretto J.A. (2010) The transcription factor SlAREB1 confers drought, salt stress tolerance and regulates biotic and abiotic stress-related genes in tomato. *Plant Cell Environ* 33, 2191-208.

Pannell J.R. (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution* 57, 949-61.

Park S.J., Jiang K., Schatz M.C. & Lippman Z.B. (2012) Rate of meristem maturation determines inflorescence architecture in tomato. *Proc Natl Acad Sci U S A* 109, 639-44.

Parra M.M., del Pozo O., Luna R., Godoy J.A. & Pintor-Toro J.A. (1996) Structure of the dehydrin tas14 gene of tomato and its developmental and environmental regulation in transgenic tobacco. *Plant Mol Biol* 32, 453-60.

Peralta I.E. & Spooner D.M. (2001) Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (Solanum L. section Lycopersicon [Mill.] Wettst. subsection Lycopersicon). *Am J Bot* 88, 1888-902.

Peralta I.E., Spooner D.M. & Knapp S. (2008) Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicon*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Syst. Bot. Monogr.* 84.

Pluess A.R., Schutz W. & Stocklin J. (2005) Seed weight increases with altitude in the Swiss Alps between related species but not among populations of individual species. *Oecologia* 144, 55-61.

R Development Core Team (2005) R: a Language and Environment for Statistical Computing. *R Foundation for statistical Computing (ed.): Vienna, Austria*.

Rand D.M. & Kann L.M. (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. *Mol Biol Evol* 13, 735-48.

Ranjan A., Ichihashi Y. & Sinha N.R. (2012) The tomato genome: implications for plant breeding, genomics and evolution. *Genome Biol* 13, 167.

Rausher M.D., 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, 137-44.

Rausher M.D., Miller R.E. & Tiffin P. (1999) Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol* 16, 266-74.

Rick C.M., Fobes J.F. & Tanksley S.D. (1979) Evolution of mating systems in Lycopersicon hirsutum as deduced from genetic variation in electrophoretic and morphological characters. *Pl. Syst. Evol.* 132, 279-98.

Rick C.M. & Tanksley S.D. (1981) Genetic variation in Solanum pennellii: Comparisons with two other sympatric tomato species. *Pl. Syst. Evol.* 139, 11-45.

Rose L.E., Grzeskowiak L., Horger A.C., Groth M. & Stephan W. (2011) Targets of selection in a disease resistance network in wild tomatoes. *Mol Plant Pathol* 12, 921-7.

Rose L.E., Michelmore R.W. & Langley C.H. (2007) Natural variation in the Pto disease resistance gene within species of wild tomato (Lycopersicon). II. Population genetics of Pto. *Genetics* 175, 1307-19.

Roselius K., Stephan W. & Stadler T. (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* 171, 753-63.

Sarkinen T., Bohs L., Olmstead R.G. & Knapp S. (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol Biol* 13, 214.

Scott S.J. & Jones R.A. (1982) Low temperature seed germination of Lycopersicon species evaluated by survival analysis. *Euphytica* 31, 869-83.

Scott S.J. & Jones R.A. (1985) Quantifying seed germination responses to low temperatures: Variation among *Lycopersicon* spp. *Environmental and Experimental Botany* 25, 129-37.

Shendure J. & Ji H. (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26, 1135-45.

Shinozaki K. & Yamaguchi-Shinozaki K. (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Curr Opin Plant Biol* 3, 217-23.

Shinozaki K. & Yamaguchi-Shinozaki K. (2007) Gene networks involved in drought stress response and tolerance. *J Exp Bot* 58, 221-7.

Shinozaki K., Yamaguchi-Shinozaki K. & Seki M. (2003) Regulatory network of gene expression in the drought and cold stress responses. *Curr Opin Plant Biol* 6, 410-7.

Slotte T., Hazzouri K.M., Agren J.A., Koenig D., Maumus F., Guo Y.L., Steige K., Platts A.E., Escobar J.S., Newman L.K., Wang W., Mandakova T., Vello E., Smith L.M., Henz S.R., Steffen J., Takuno S., Brandvain Y., Coop G., Andolfatto P., Hu T.T., Blanchette M., Clark R.M., Quesneville H., Nordborg M., Gaut B.S., Lysak M.A., Jenkins J., Grimwood J., Chapman J., Prochnik S., Shu S., Rokhsar D., Schmutz J., Weigel D. & Wright S.I. (2013) The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nat Genet* 45, 831-5.

Spooner D.M., Peralta I.E. & Knapp S. (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [Solanum L. section Lycopersicon (Mill.) Wettst.]. *Taxon* 51, 43-61.

Stadler T., Arunyawat U. & Stephan W. (2008) Population genetics of speciation in two closely related wild tomatoes (Solanum section Lycopersicon). *Genetics* 178, 339-50.

Stadler T., Haubold B., Merino C., Stephan W. & Pfaffelhuber P. (2009) The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* 182, 205-16.

Stadler T., Roselius K. & Stephan W. (2005) Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. *Evolution* 59, 1268-79.

Stamatakis A., Hoover P. & Rougemont J. (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 57, 758-71.

Stamatakis A., Ludwig T. & Meier H. (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21, 456-63.

Steige K.A. (2011) Population Genetic Analysis and Expression Patterns of the dehydrin *pLC30-15* in wild tomato species. *Diplom Thesis, LMU Munich*.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585-95.

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, 454-62.

Tanksley S.D., Ganal M.W., Prince J.P., de Vicente M.C., Bonierbale M.W., Broun P., Fulton T.M., Giovannoni J.J., Grandillo S., Martin G.B. & et al. (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132, 1141-60.

Tavaré S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in Life Sciences (American Mathematical Society)* 17, 57-86.

Tellier A., Fischer I., Merino C., Xia H., Camus-Kulandaivelu L., Stadler T. & Stephan W. (2011a) Fitness effects of derived deleterious mutations in four closely related wild tomato species with spatial structure. *Heredity (Edinb)* 107, 189-99.

Tellier A., Laurent S.J., Lainer H., Pavlidis P. & Stephan W. (2011b) Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proc Natl Acad Sci U S A* 108, 17052-7.

Tester M. & Langridge P. (2010) Breeding technologies to increase crop production in a changing world. *Science* 327, 818-22.

The Tomato Genome Sequencing Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635-41.

Thomashow M.F. (2010) Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. *Plant Physiol* 154, 571-7.

Thompson A.J., Jackson A.C., Symonds R.C., Mulholland B.J., Dadswell A.R., Blake P.S., Burbidge A. & Taylor I.B. (2000) Ectopic expression of a tomato 9-cis-epoxycarotenoid dioxygenase gene causes over-production of abscisic acid. *Plant J* 23, 363-74.

Toivainen T., Pyhajarvi T., Niittyvuopio A. & Savolainen O. (2014) A recent local sweep at the PHYA locus in the Northern European Spiterstulen population of Arabidopsis lyrata. *Mol Ecol* 23, 1040-52.

Torres-Schumann S., Godoy J.A. & Pintor-Toro J.A. (1992) A probable lipid transfer protein gene is induced by NaCl in stems of tomato plants. *Plant Mol Biol* 18, 749-57.

Turner T.L., Bourne E.C., Von Wettberg E.J., Hu T.T. & Nuzhdin S.V. (2010) Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nat Genet* 42, 260-3.

Umezawa T., Nakashima K., Miyakawa T., Kuromori T., Tanokura M., Shinozaki K. & Yamaguchi-Shinozaki K. (2010) Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. *Plant Cell Physiol* 51, 1821-39.

van Hoek A.H., Mevius D., Guerra B., Mullany P., Roberts A.P. & Aarts H.J. (2011) Acquired antibiotic resistance genes: an overview. *Front Microbiol* 2, 203.

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, 183-92.

Wasilewska A., Vlad F., Sirichandra C., Redko Y., Jammes F., Valon C., Frei dit Frey N. & Leung J. (2008) An update on abscisic acid signaling in plants and more. *Mol Plant* 1, 198-217.

Watterson G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7, 256-76.

Wei T. & O'Connell M.A. (1996) Structure and characterization of a putative drought-inducible H1 histone gene. *Plant Mol Biol* 30, 255-68.

Willi Y., Van Buskirk J. & Hoffmann A.A. (2006) Limits to the Adaptive Potential of Small Populations. *Annu Rev Ecol Evol Syst* 37, 433-58.

Wu L., Chen X., Ren H., Zhang Z., Zhang H., Wang J., Wang X.C. & 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, 815-25.

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*.

Xiong L., Schumaker K.S. & Zhu J.K. (2002) Cell signaling during cold, drought, and salt stress. *Plant Cell* 14 Suppl, S165-83.

Yamaguchi-Shinozaki K. & Shinozaki K. (2005) Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci* 10, 88-94.

Yanez M., Caceres S., Orellana S., Bastias A., Verdugo I., Ruiz-Lara S. & Casaretto J.A. (2009) An abiotic stress-responsive bZIP transcription factor from wild and cultivated tomatoes regulates stress-related genes. *Plant Cell Rep* 28, 1497-507.

Zanne A.E., Tank D.C., Cornwell W.K., Eastman J.M., Smith S.A., FitzJohn R.G., McGlinn D.J., O'Meara B.C., Moles A.T., Reich P.B., Royer D.L., Soltis D.E., Stevens P.F., Westoby M., Wright I.J., Aarssen L., Bertin R.I., Calaminus A., Govaerts R., Hemmings F., Leishman M.R., Oleksyn J., Soltis P.S., Swenson N.G., Warman L. & Beaulieu J.M. (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature* 506, 89-92.

Zegzouti H., Jones B., Marty C., Lelievre J.M., Latche A., Pech J.C. & Bouzayen M. (1997) ER5, a tomato cDNA encoding an ethylene-responsive LEA-like protein: characterization and expression in response to drought, ABA and wounding. *Plant Mol Biol* 35, 847-54.

Zhang H., Huang Z., Xie B., Chen Q., Tian X., Zhang X., Lu X., Huang D. & Huang R. (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, 262-70.

Zhou R., Ling S., Zhao W., Osada N., Chen S., Zhang M., He Z., Bao H., Zhong C., Zhang B., Lu X., Turissini D., Duke N.C., Lu J., Shi S. & Wu C.I. (2011) Population genetics in nonmodel organisms: II. natural selection in marginal habitats revealed by deep sequencing on dual platforms. *Mol Biol Evol* 28, 2833-42.

Zuellig M.P., Kenney A.M. & Sweigart A.L. (2014) Evolutionary genetics of plant adaptation: insights from new model systems. *Curr Opin Plant Biol* 18C, 44-50.

Acknowledgements

First, I would like to thank my supervisor Wolfgang Stephan for giving me the opportunity to stay in the Tomato Group after my Diploma studies and to do this exciting project as my PhD thesis.

A great thank you to my co-supervisors Mamadou Mboup and Aurélien Tellier for starting this project together and for all the discussions, their help, advice and support throughout the years. Many thanks also to Tetyana Nosenko for joining this project, for her help, advice and for the many interesting discussions during the last months.

Further I would like to acknowledge all people from the evolutionary biology group for the good and friendly working atmosphere during all the years that I was part of this group.

A great thank you to Hilde Lainer for the excellent assistance in preparing the samples for the *S. chilense* sequencing, to Gaby Kumpfmüller, who joined later on and helped with the outgroup sequencing, and to Gisela Brinkmann and Simone Lange for their help and advice since the day that I did my first project in this lab. I am very grateful to Anja Hörger and Iris Fischer who introduced me to the topics and work routines in the Munich Tomato Group during my diploma studies – I do not know if I had ended up working with tomatoes without them – and for their help and advice ever since. Special thanks also to Kim Steige for the good time in the lab and all the helpful discussions, and to Tina Jancke, Katharina Irl, and Susi Voigt. Many thanks to Pablo Duchén, Stefan Laurent, and Stephan Hutter for their help and advice with all computer, program, and programming related questions and topics, and to Ricardo Wilches for his help with BayeScan. Many thanks also to Paula Brücher and Armin Scheben who helped with the analyses of this data set during their Bachelor and Master studies.

I would also like to thank all people from the evolutionary biology group who are or were involved in teaching during my undergraduate years. I would not have been able to complete this thesis without what I learned about evolutionary biology and population genetics during these lectures and courses.

Finally, I would like to thank my parents and grandparents for their support.