Different evolutionary histories of two cation/proton exchanger gene families in plants

Inês S Pires1,2, Sónia Negrão1, Melissa M Pentony2, Isabel A Abreu1, Margarida M Oliveira1 and Michael D Purugganan2*

Abstract

Background: Gene duplication events have been proposed to be involved in the adaptation of plants to stress conditions; precisely how is unclear. To address this question, we studied the evolution of two families of antiporters. Cation/proton exchangers are important for normal cell function and in plants, Na⁺,K⁺/H⁺ antiporters have also been implicated in salt tolerance. Two well-known plant cation/proton antiporters are NHX1 and SOS1, which perform Na⁺ and K⁺ compartmentalization into the vacuole and Na⁺ efflux from the cell, respectively. However, our knowledge about the evolution of NHX and SOS1 stress responsive gene families is still limited.

Results: In this study we performed a comprehensive molecular evolutionary analysis of the NHX and SOS1 families. Using available sequences from a total of 33 plant species, we estimated gene family phylogenies and gene duplication histories, as well as examined heterogeneous selection pressure on amino acid sites. Our results show that, while the NHX family expanded and specialized, the SOS1 family remained a low copy gene family that appears to have undergone neofunctionalization during its evolutionary history. Additionally, we found that both families are under purifying selection although SOS1 is less constrained.

Conclusions: We propose that the different evolution histories are related with the proteins’ function and localization, and that the NHX and SOS1 families are examples of two different evolutionary paths through which duplication events may result in adaptive evolution of stress tolerance.

Background

It has been argued that gene duplications underlie mechanisms to achieve stress adaptation [1,2]. There is little evidence, however, to support this. Additionally, it is not known if there is one or multiple ways of achieving stress adaptation through gene duplication. To address this question, we focused our attention on salt tolerance, and the evolutionary histories of two gene families involved in this trait.

Salt tolerance is a complex trait that is thought to have originated multiple times in plants [3], and it is important to understand the nature and molecular evolution of salt tolerance mechanisms throughout the history of land plants. This will be key not only to our understanding of how plants adapt against the disruptive effects of high salt concentrations in the soil, but also to point out directions for possible crop improvement. Indeed farmers are increasingly facing loss of crop production due to this abiotic stress, since currently > 20% of the world’s arable land is affected by high salinity [4].

Cation/proton exchangers are essential to the normal function of the cell. Besides helping regulating internal pH, cell volume, and cytoplasmic ion homeostasis [5-7], these transporters have also been shown to be involved in vesicular trafficking and protein targeting [8,9]. In plants, Na⁺,K⁺/H⁺ antiporters are also associated with salt tolerance [10,11], and a series of studies have targeted these cation exchangers to improve agronomically important crops [12,13].

All Na⁺,K⁺/H⁺ exchangers belong to a superfamily of monovalent cation/proton antiporters (CPA) and are divided in two families, CPA1 and CPA2 [14]. The CPA1 family is composed of two main groups, one containing plasma membrane-bound proteins and the other intracellular proteins [14]. Within these two groups are well-characterized antiporters associated with salt tolerance.
in plants, the salt overly sensitive 1 (SOS1) and sodium hydrogen exchanger (NHX) proteins. SOS1 proteins are localized in the plasma membrane [15,16], and are responsible for Na⁺ efflux from the cell. NHX proteins, on the other hand, are intracellular proteins that compartmentalize Na⁺ and K⁺ in the vacuole [17-19].

In Arabidopsis thaliana, the NHX protein family has six members that are classified into two classes [17]. The class I proteins, AtNHX1-4, are localized in the tonoplast [17] and have equal affinity for Na⁺ and K⁺ [7,20]. Class II proteins include AtNHX5 and 6, which are localized in the endosomal compartment of the Golgi and trans-Golgi network [18] and have a higher affinity for K⁺ compared to Na⁺ [6].

Previous reports have shown that the Arabidopsis NHX genes which encode these proteins have different expression patterns and responsiveness to abiotic stresses [21,22]. AtNHX1 and 2 were shown to be expressed at high levels in all plant tissues, while AtNHX3 and 4 were mainly expressed in root and flower tissues, respectively [21,22]. AtNHX5 was also expressed in all tissues but at lower levels [21,22], while AtNHX6 expression was detected only in shoots and roots by RT-PCR [22]. In addition, AtNHX1-3 were shown to be responsive to both salt stress and abscisic acid (ABA) [22,23]. AtNHX5 was only responsive to salt stress, suggesting that its response is ABA-independent [22]. Moreover, all NHX genes promoted the recovery of a salt sensitive yeast mutant [21,22] and numerous studies have shown that overexpression of Arabidopsis NHX1 and 5, or rice NHX1 (among others), resulted in increased salt tolerance in transgenic plants [10,24-27].

The other protein that characterizes this cation/proton exchanger family is SOS1, which belongs to the well-known salt tolerance Salt-Overly-Sensitive (SOS) pathway [28-30]. Salt stress elicits a cytosolic calcium signal, which functions as a major secondary-messenger signalling molecule. A myristoylated calcineurin B-like protein (SOS3) senses the salt-elicited calcium signal, and upon Ca²⁺ binding SOS3 undergoes dimerization and enhances the serine/threonine protein kinase activity of SOS2. The SOS3/SOS2 complex is targeted to the plasma membrane and activates SOS1 [28-30].

SOS1 has been called NHX7 by several authors, and was thus thought to be part of the NHX gene family [12,16]. Previous studies, however, regarding the phylogeny of cation/hydrogen transporters [14,31] suggest that it is distinct from the other NHX proteins and should more appropriately retain its designation as SOS1. For this reason, we refer to the Arabidopsis proteins SOS1/ NHX7 and NHX8 (highly similar to, but shorter than SOS1) as SOS1A and SOS1B, respectively. SOS1B, like SOS1A, is localized to the plasma membrane [15,16]. While SOS1A is a Na⁺/H⁺ antiporter [32], however, SOS1B seems to only transport Li⁺ [16]. In addition, the tonoplast exchangers in Arabidopsis are also regulated by the SOS pathway, specifically by SOS2 [33], thus suggesting coordination between tonoplast and plasma membrane antiporters.

The importance of these two families of proteins in plant salt tolerance is well established. By promoting the efflux of Na⁺, SOS1A helps protect the cell from the deleterious effects of this ion. Additionally, SOS1A seems to have an important role in long-distance Na⁺ transport, thus helping to regulate the Na⁺/K⁺ ratio in roots and shoots [17]. On the other hand, NHX family members have been described as essential to Na⁺ compartmentalization in the vacuole, protecting the cell from the deleterious effects of this ion and maintaining cytoplasmic ion homeostasis [10,34]. Recent studies, however, suggest instead that increased salt tolerance of NHX over-expressing plants result from an improved ability to retain K⁺ after stress induction [35-37].

Despite the key roles these genes play in salt tolerance in plants, little is known about the evolutionary histories of both gene families. Understanding the evolution of Na⁺,K⁺/H⁺ antiporters can help clarify the mechanisms leading to plant stress adaptation associated with gene duplication events [1,2,38]. In this study, we report a comprehensive molecular evolutionary analysis of both the NHX and SOS1 plant protein families. We reconstructed the phylogeny and the history of duplication events for each family, as well as determined the selection pressure on amino acid sites within these proteins. Our purpose here is not to identify novel NHX and SOS1 plant genes, or to do a phylogeny analysis of the entire monovalent CPA gene family, as in others [39]. Instead, we show that the NHX and SOS1 gene families have very different evolution trajectories and suggest that these divergent evolutionary histories are related to the evolution of their function and cellular localization. Finally, we suggest that the NHX and SOS1 families represent examples of two different paths in the molecular evolution of stress tolerance.

Results and discussion
Phylogeny of the NHX family exhibits multiple independent duplication events
We selected 121 genes in 33 taxa that appear, or are described, to belong to the NHX family. The distribution of NHX genes among the various species is shown in Table 1. Among these genes, 108 are found in 28 angiosperm species, 2 in 1 gymnosperm taxa, and 9 in 2 non-seed plant taxa. We also identified 1 sequence from the alga C. reinhardtii, and we used 1 sequence from the budding yeast S. cerevisiae as an outgroup. The NHX phylogeny agrees with the classification of these proteins in two distinct evolutionary groups. We obtained two main clades, with ~99% bootstrap support, which
showed evolutionary divergence of NHX protein groups according to their protein localization, as previously reported [22,40] (Figure 1 and Additional file 1: Figure S1 for more detail and unrooted tree). Proteins localized in the endosomal compartment of the Golgi and trans-Golgi network grouped in clade 1 (closer to the yeast NHX1), while proteins localized in the tonoplast grouped in clade 2.

We observe that multiple independent duplication events have occurred throughout the evolutionary history of the NHX family. Based on the reconciled phylogeny (Figure 2), we estimate 27 independent gene duplication and 40 gene loss events during the diversification of this gene family. Gene loss events, however, are not displayed on the phylogeny since they might not represent true losses but may be due to partial gene inventories due to incomplete whole genome sequencing data (e.g., *Thellungiella halophila*) and to some genes still remaining to be identified. In this reconciled tree, all the branches with bootstrap support inferior to a set value (here 75%) are rearranged to achieve the most parsimonious duplication and loss history for the gene family. Nevertheless our results, together with the fact that non-plant species also have multiple NHX genes [9,39], suggest that multiple copies of NHX proteins were already present in a common ancestor of modern land plants (even though AtNHX1-4 and their orthologs are plant specific) [39].

Interestingly, orthologs of Arabidopsis NHX1 and 2 might be specific to (at least) seed plants. NHX1 and 2 proteins consistently grouped separately from proteins from basal plant lineages such as spikemoss (*Selaginella moellendorffii*), moss (*Physcomitrella patens*), and the green photosynthetic alga (*Chlamydomonas reinhardtii*) (Figure 1 and Additional file 1: Figure S1). Due to the limited number of gymnosperm sequences (only two NHX proteins from *Picea sitchensis*) we could not assess divergence from the more closely related species to the angiosperms. Sequences from more gymnosperm species are available as EST sequences in PlantGDB database, which coupled with greater availability of genomics resource data from other non-flowering plant taxa species, could potentially be used in the future to examine more closely the precise origins of specific gene family clades. Nevertheless, bootstrap supports obtained for these branches were not very high (Figure 1 and Additional file 1: Figure S1) and more sequences will be necessary to be analyzed to confirm this.

### Purifying selection on NHX proteins

We examined the sequences of the plant NHX gene family to determine whether there was significant heterogeneity in selective pressure among amino acid sites. Selective pressure at the protein level can be measured as $dN/dS$, where $dN$ and $dS$ are the number of non-synonymous and synonymous substitutions per non-synonymous and synonymous site, respectively [41]. If non-synonymous mutations are favored by positive selection, non-synonymous mutations are fixed at a faster

---

**Table 1 Species and number of NHX and SOS1 sequences used in this study**

| Species          | Number of putative NHX sequences | Number of putative SOS1 sequences |
|------------------|----------------------------------|---------------------------------|
| Yeast            | Sacharomyces cerevisae           | 1                               | -                               |
| Alga             | Chlamydomonas reinhardtii        | 1                               | -                               |
| Moss             | Physcomitrella patens            | 7                               | 2                               |
| Spikemoss        | Selaginella moellendorffii       | 2                               | 2                               |
| Gymnosperm       | Picea stichensis                 | 2                               | -                               |
| Angiosperms      | Brachypodium distachyon          | 5                               | 2                               |
| Monocots         | Oryza sativa                     | 5                               | 1                               |
|                  | Setaria italica                  | 5                               | 1                               |
|                  | Zea mays                         | 5                               | 1                               |
|                  | Sorghum bicolor                  | 4                               | 1                               |
| Angiosperms      | Aquilegia coerulea               | 3                               | 2                               |
| Dicots           | Mesembryanthemum crystallinum    | 3                               | 1                               |
|                  | Atriplex gmelinii                | 1                               | -                               |
|                  | Atriplex dimorphostegia          | 1                               | -                               |
|                  | Suaeda salsa                     | 1                               | -                               |
|                  | Salicornia europaea              | 1                               | -                               |
|                  | Chenopodium glaucum              | 1                               | -                               |
|                  | Kalidium foliatum                | 1                               | -                               |
|                  | Mimus guttatus                   | 6                               | 1                               |
|                  | Vitis vinifera                   | 5                               | 1                               |
|                  | Eucalyptus grandis               | 4                               | 2                               |
|                  | Citrus clementina                | 5                               | 1                               |
|                  | Citrus sinensis                  | 3                               | 1                               |
|                  | Carica papaya                    | 4                               | 1                               |
|                  | Thellungiella halophila          | 1                               | 2                               |
|                  | Arabidopsis thaliana             | 6                               | 2                               |
|                  | Prunus persica                   | 6                               | 1                               |
|                  | Cucumis sativus                  | 3                               | 1                               |
|                  | Glycine max                      | 8                               | 1                               |
|                  | Medicago truncatula              | 3                               | -                               |
|                  | Populus trichocarpa              | 6                               | 2                               |
|                  | Ricinus communis                 | 5                               | 1                               |
|                  | Manihot esculenta                | 7                               | 2                               |
| **TOTAL**        | **33**                           | **121**                         | **32**                          |
rate and dN/dS > 1 [41]. In the case of purifying selection dN/dS < 1. However, if selection has no effect on fitness due to neutral evolution, both mutations are fixed at the same rate and dN/dS = 1 [41].

For NHX coding region sequences from both clade 1 and 2, the average dN/dS was estimated at ~0.08 and no sites were predicted to be under positive selection. Additionally, no difference in dN/dS between NHX genes was identified. We believe that these results might reflect the key function of NHX proteins, which besides salt tolerance [10,11] include normal cell functions in vesicular trafficking and protein targeting [8,9]. Additionally, NHX proteins are predicted to have ~12 transmembrane domains [42] and are probably constrained at the structural level. Our results are consistent with the purifying selection expected for these proteins [43] and agree with results recently obtained by Hudson et al. [44].

**Phylogeny of the SOS1 family exhibits few and recent duplication events**

Unlike the NHX family, plant species appear to have lower numbers of SOS1 genes in their genomes. We have selected 32 genes in 22 taxa that appear, or are described, to belong to the SOS1 family. The distribution of SOS1 genes among the various species is shown in Table 1. Among these genes, 22 are found in 16 angiosperm species, and 4 in 2 non-seed plant taxa. In this case we used 2 sequences from the moss *P. patens* as outgroup sequences.

We found that eight plant species, including *A. thaliana*, seem to have two SOS1-like genes (Table 2). At the time of submission of this manuscript, however, a new BLASTp search using Phytozome v8.0 (http://www.phytozome.net/, verified in June 2012) suggested that more species might possess more than one Putative SOS1 protein. The
Figure 2 Phenylogeny of the NHX family that indicates multiple independent duplication events throughout its evolutionary history.
The reconciled NHX tree obtained using NOTUNG v2.6, has a duplication/loss score of 60.5 and shows 27 independent gene duplication and 40 gene loss (not shown) events.
existence of two SOS1 genes has been described previously for Arabidopsis thaliana [16] and two other plant species not included in this study, namely Neptune grass (Cymodocea nodosa) [45] and quinoa (Chenopodium quinoa) [46]. Nevertheless, to confirm that these seven species have two SOS1-like proteins like in Arabidopsis, it is necessary to perform further functional studies.

As in Arabidopsis, the species with two SOS1-like proteins usually displayed differences in protein length (Table 2). However, two classes of length differences were identified - one in which both duplicates had similar lengths (Δ < 110 amino acid residues), and another in which the difference in duplicates length was substantially larger (Δ > 400 amino acid residues). Nevertheless, as in Arabidopsis, the differences in length were mainly due to shortening in the C-terminus instead of deletions within the protein primary sequence. This is important to notice, because it was shown by Quintero et al. [47] that it is within the C-terminus of SOS1 that lays the activation site and the auto-inhibitory domain of the protein. Additionally, SOS1-like proteins from A. coerulea, E. grandis, and M. esculenta have deletions in the N-terminus of the protein when compared to the Arabidopsis SOS1 proteins. This suggests that SOS1-like proteins from these three species might have a smaller transmembrane region, but further biochemical analyses need to be performed to confirm this.

The SOS1 phylogeny (Figure 3 and Additional file 1: Figure S2) exhibits three distinct clades: (1) one that contains the moss and spikemoss proteins (~94% bootstrap support), (2) another that has monocot proteins (~91% bootstrap support), and (3) finally, one clade with dicots SOS1 sequences. The SOS1 protein phylogeny is almost identical to the accepted species phylogeny of the study species, although there appears to be some long-branch attraction that alters the position within the dicot clade of one Eucalyptus grandis and one A. thaliana protein.

Unlike the NHX family, the SOS1 family has undergone fewer gene duplication events. Reconciliation of the SOS1 tree with the species tree (Figure 3 and Additional file 1: Figure S3) estimates that 8 independent gene duplication events and one loss have occurred within the land plants (Figure 4). Once again, gene losses may simply represent unidentified genes; in fact Oh et al. [2] recently reported that Thellungiella parvula has three copies of the A. thaliana SOS1B and that these result from recent tandem duplications.

Interestingly, gene duplication events in the SOS1 family appear toward the terminal nodes of the reconciled tree, indicating that these events were recent in evolutionary time. The analyses of some species that have a second copy of SOS1 further support this observation. For example, Populus trichocarpa is thought to have undergone a recent whole genome duplication event [48] and monocots such as Brachypodium distachyon are thought to be more prone to tandem duplications which suggests that any gene duplication in one monocot species is probably recent and not shared with the common ancestor of this group [49].

### Purifying selection on SOS1 proteins with few amino acids under positive selection

Like the NHX proteins, the SOS1 proteins are generally subject to purifying selection, but our analysis indicates that some amino acid sites appear to be under positive selection. Among the land plants the SOS1 proteins had an average dN/dS of ~0.16. Using alignments that largely represented orthologous sequences, the glutamine at position 958 (Q958) in A. thaliana is evolving neutrally (dN/dS ~ 1.00) and two amino acid residues are under positive selection (dN/dS ~ 1.23) - a valine at position 366 (V366) and a serine at position 738 (S738) (Figure 5). Other residues (L843, C902, and P915 in A. thaliana) were identified in our analysis as strong candidate residues to be under neutral or positive selection, but these were not as well supported as the other three sites. Nevertheless, these other residues might be indeed under neutral

| Species                                   | Putative SOS1 – longer gene | Putative SOS1 – shorter gene | Δ Length (AA) |
|-------------------------------------------|-----------------------------|-----------------------------|--------------|
| Physcomitrella patens                    | Pp1s12_8V6.1                | Pp1s15_101V6.1              | 93           |
| Selaginella moellendorffii                | 74518                       | 75049                       | 84           |
| Brachypodium distachyon                   | Bradi4g00290.1              | Bradi4g17380.1              | 419          |
| Aquilegia coerulea                       | Aca_000299m                 | Aca_002550m                 | 562          |
| Eucalyptus grandis                       | Egrandis_v1_0.001457m       | Egrandis_v1_0.008144m       | 454          |
| Arabidopsis thaliana                     | At2g01980                   | At1g14660                   | 413          |
| Populus trichocarpa                      | POPTR_0010s11130.1          | POPTR_0008s14030.1          | 2            |
| Manihot esculenta                        | cassava41_025141m           | cassava41_002710m           | 107          |

Δ represents the difference in protein length.
or positive selection since the tests designed to detect positive selection are very stringent.

In general, structurally constrained sites should be under stronger purifying selection, while unstructured sites have higher levels of amino acid replacements [43]. According to different prediction tools available at ExPASy (http://expasy.org/tools/), however, the positively selected residue V366 is in a predicted transmembrane helix. Amino acid site Q958 is predicted to be in a transition zone between coiled and beta-sheet regions, and residue S738 is predicted to be in a random coil.

Two of these residues, V366 and S738, are in the cytoplasmic domain of AtSOS1A (Figure 5) but do not fall in the auto-inhibitory domain (amino acids 1013–1047) or near sites known to be relevant for protein activation by phosphorylation (S1138 and S1136) [47]. Additionally, mutations of these three residues under neutral or positive selection are not predicted to result in changes that affect secondary or tertiary structure, or create new sites for putative posttranslational modifications (data not shown). Nevertheless, residue S738 is in close proximity to residues S742 and V743, which were shown to affect the activity of the *A. thaliana* protein when mutated [47]. Like S742 and V743, S738 is not well conserved between species. Mutations within the transmembrane pore were also shown to affect AtSOS1A activity [47]. Thus, it may be that the roles of residues S738 and V366 in the activity of the SOS1A protein are worth study.

Relaxation of selection on SOS1 gene duplicates

The SOS1 family apparently exhibits a relaxation of purifying selective pressure upon duplication, as previously reported in other gene families [50] and, specifically for ion transporters [44]. This relaxation of selection, however, seems to be present in only one duplicated gene copy, and only in species that have two SOS1-like proteins differing in length. Although the free-ratio model [51] is not generally considered to be a good method to predict dN/dS because it has too many free parameters, the comparison...
between values of dN/dS is suggestive (Additional file 1: Figure S4). Species which have two Putative proteins similar in length (< 110 amino acid residues) showed a difference in dN/dS values less than 0.1, while species that retain two putative proteins with very different length (> 400 amino acid residues) displayed a difference in dN/dS values always greater than 0.2. Furthermore, in the cases where the two putative SOS1-like proteins substantially differed in length, the shortest appeared to be evolving faster than the longer protein from the same species (Additional file 1: Figure S4).

We also confirmed that these results were not affected by a saturation of nucleotide substitutions on the shorter sequences (Additional file 1: Figure S5), except possibly for AtSOS1B.

More data is needed to verify if this trend is statistically significant. Nevertheless, we can speculate that the possible relationship between the differences in protein length and dN/dS might be due to the relative recent occurrence of the duplications in SOS1, which could indicate that not enough time has elapsed for differentiation between duplicated genes to have occurred (e.g., woody plants, such as Populus, have a slower evolution rate than Arabidopsis) [48]. Another hypothesis to explain the different dN/dS values observed between SOS1 proteins that differ in length is that the duplication of some SOS1 genes did not result in functionally equivalent gene copies, and hence functional divergence between copies is immediately observed.

Conclusions

Contrasting evolutionary histories between the NHX and SOS1 gene families

NHX and SOS1 plant gene families exhibit markedly different evolutionary histories - while the NHX family expanded and developed functionally specialized members throughout the history of the land plants (Figures 1 and 2), the SOS1 family remained a low copy gene family (Figures 3 and 4). It is clear that after gene duplication events, the NHX family members have undergone a series of protein subcellular relocalization and spatial subfunctionalization events. This can be observed by phylogenetically mapping functional and expression information collected from studies on Arabidopsis proteins (Figure 6) [6,17-23]. Rice (Oryza sativa) proteins also exhibit functional differentiation, although they are not always functionally equivalent to their Arabidopsis homologues. For example, both AtNHX5 and OsNHX5, which are both part of clade 1 (Figure 1), have higher affinity to K+ compared to Na+ [6,40], although AtNHX5 appears to be unresponsive to ABA [22] while OsNHX5 responds to this hormone [40].

In contrast, the SOS1 genes are found only as single- or low-copy genes in most plant species, suggesting either low duplication rates or a higher rate of gene deletion after gene duplication events for these loci. Although it is unclear why SOS1 is constrained to be a single- or low-copy gene family, the few cases where a second SOS1-like gene has remained in the genome, it appears to have undergone neofunctionalization. For example, Arabidopsis SOS1B is only able to transport Li+ [16], and the C. nodosa SOS1B is described as having a different function from CnSOS1A [45]. Moreover, we found a difference in dN/dS between proteins that are diverging in protein length, thus supporting the observed neofunctionalization or possible loss of function of a second copy of SSO1. This difference
might occur because intracellular membrane proteins can be duplicated and subfunctionalized, in contrast to cytoplasmatic membrane proteins such as the NHXs. In a simplistic view, this process arises from fewer possibilities for neofunctionalization, in contrast to cytoplasmatic membrane proteins. In our work, we do not detect a relaxation of selective constraint after duplication in the NHX gene family, in agreement with a previous study [44]. Furthermore, the difference in average dN/dS between NHX (~0.08) and SOS1 proteins (~0.16) supports previous results [50,52] indicating that in eukaryotes, old duplicated genes evolve under stronger purifying selection. In contrast, SOS1 proteins were targeted to the tonoplast. Consequently, an organism with an improved ability to tolerate salt stress arose, since it was able to better retain K+ after stress induction [36,37].

On the other hand, the duplication followed by neofunctionalization of SOS1 has generated an increased capacity to tolerate soils with high Li+ content in the case of Arabidopsis and related species. This is especially relevant for the Arabidopsis wild relative, Thellungiella spp., that lives in habitats with naturally high Li+ concentrations [2]; indeed, Thellungiella parvula appears to have three copies of SOS1, probably due to environmental adaptation [2]. Altogether, our results represent two examples of different molecular evolutionary trajectories in land plant genomes that result in organismal stress adaptation through gene duplication.

Methods

NHX and SOS1 protein sequences

We used rice protein sequences OsNHX1 (Os07g47100) and OsSOS1A (Os12g44360), in Node Consensus BLASTp of Phytozome v7.0 [53] to obtain the homologous sequences from several plant species (from mosses to eudicots). Sequences belonging to families with an e-value lower than 10^-100 were used in the analysis. Independent BLASTp searches in Phytodrome v7.0 and in NCBI (http://www.ncbi.nlm.nih.gov/) allowed retrieval of homologous sequences from OsNHX1 from the algae Chlamydomonas reinhardtii and the gymnosperm Picea sitchensis. In these particular
cases, sequences with an e-value lower than $10^{-10}$ were used. Additionally, we also retrieved from NCBI sequences identified as being NHX- or SOS1-like from the yeast *Saccharomyces cerevisiae* (ScNHX1 – YDR456W) and from salt tolerant plant species (in grey background in Additional file 2: Table S1).

In order to improve the quality of the alignments obtained, sequences shorter than 500 or longer than 940 or 1200 amino acids from the NHX and SOS1 families, respectively, were removed from the study (38 out of 158 NHX like sequences, and 6 out of 38 SOS1 like sequences, were excluded). This exclusion was performed because the alignments these sequences produced had a large number of discontinuous aligned stretches and we had low confidence on the homology assignments from these alignments. The identifiers of all the sequences used, and their respective database origin, are listed in the Additional file 2: Table S1.

**Phylogenetic analysis**

The phylogeny of NHX proteins was obtained from 121 sequences in 33 different species (Table 1), with the NHX1 protein from the yeast *Saccharomyces cerevisiae* as the outgroup sequence. The SOS1 phylogeny was estimated from 32 sequences in 23 plant species (Table 1), with the sequences from the moss *P. patens* used as outgroups. These outgroups were chosen because the S. cerevisiae sequence (NHA1) most similar to the plant SOS1 genes proved difficult to align with the rest of the plant genes.

Alignments of protein sequences were obtained using MUSCLE [54] and the alignments were refined using Gblocks (default settings) [55,56] in order to minimize the number of positions with missing information (gaps). For each cleaned or non-cleaned alignment the best-fit amino acid substitution model was selected using ProtTest from MEGA5 [57]. Since the Jones-Taylor-Thornton model [58] together with a discrete approximation of the gamma distribution ($JT + \Gamma$) was always within the top five best-fit models, we selected this model to obtain the phylogenies using the neighbor-joining (NJ) method [59]. We ran the NJ analysis with 1000 bootstrap replicates in MEGA5.

Majority-rule consensus trees of 100 bootstrap replicates were also obtained from maximum likelihood (ML) phylogenies obtained using the JTT model plus a gamma rate distribution approximated using 5 categories. For this we used the SEQBOOT, PROML and CONSENSE programs from PHYLIP 3.69 (http://evolution.genetics.washington.edu/phylip.html). Branches with a good support (>75% bootstrap) were generally the same in the consensus trees obtained using either NJ (data not shown) or ML, although generally ML resulted in higher bootstrap support values.

### Estimation of duplication history

Using the information on species relationships available at the Angiosperm Phylogeny Site (www.mobot.org/mobot/research/apweb/) we constructed a cladogram representing the species tree (Additional file 1: Figure S3). Estimation of the gene duplication and loss history of both protein families was performed, using NOTUNG v2.6 [60,61], through reconciliation of the species tree with the NHX and SOS1 ML gene trees (Figures 1 and 2 and Additional file 1: Figure S1 and S2). In order to obtain the most parsimonious estimation of duplication and loss events, the reconciliation of the species and gene trees was followed by rooting, and rearranging branches with less than 75% bootstrap support.

### Determination of selective pressure on amino acid sites

NHX protein sequences were separated into two groups according to the Arabidopsis proteins with which they grouped in the estimated phylogeny. Sequences grouping with Arabidopsis NHX1-4 formed group I, and sequences grouping with Arabidopsis NHX5 and 6 formed group II. Each group was used to obtain a new alignment and a new NJ tree as previously explained. Additionally, for SOS1 sequences three other alignments were obtained in which only orthologs were included or sequences deleted at the N- or C-terminus were excluded. All these alignments were cleaned and analyzed in order to maximize the alignment positions tested. Afterwards, in order to obtain codon-cleaned alignments, we mapped the coding sequences (CDS) of NHX and SOS1 genes to protein alignments.

The codon-cleaned alignments and the NJ phylogenetic trees (both rooted and unrooted), were further used in Codeml from the PAML package [62] to determine heterogeneous selective pressure on amino acid sites. Codon substitution models [41] M0, M1a, M2a, M3, M7, and M8 were applied to the alignments. Each model builds on the preceding one by adding new dN/dS classes. M0 assumes that all sites in an alignment are under purifying selection ($dN/dS < 1$), M1a allows for some sites to be under neutral selection ($dN/dS = 1$), and M2a allows for some sites to be under positive selection ($dN/dS > 1$). M3, M7, and M8 have increased classes of sites that are allowed to be under different selection pressures, being M8 the most complex model with 13 classes of sites. All models were tested at least twice to check for convergence problems and all alignments analyzed had an average synonymous substitution (dS) rate less than 14, which is near the lower limit of the dS range estimated to be disruptive for this type of analyses [63,64]. Afterwards, Likelihood Ratio Tests (LRT) [65] and Akaike Information Criteria (AIC) [66] methods were used to determine the model that best described the data. Neutral or positive selected sites were accepted when model M8 was the
best-fit model to the alignment and when sites came up both in model M2 and M8, with a probability ≥ 95% in model M8.

Comparison of non-synonymous/synonymous substitution rate ratio between clades
Branch-site model [67] from Codeml (PAML) was used to compare the non-synonymous/synonymous substitution rate ratio (dN/dS) between clades or sequences. Bonferroni correction [68] was used to examine significance under these multiple tests. For the Sos1 alignment, the free-ratio model [51] was also used to estimate an independent dN/dS per branch. In both cases, models were tested at least twice to check for convergence problems. Convergence might be an issue, especially for parameter rich models, when there is lack of information in the data, normally resulting from highly similar or divergent sequences (see PAML manual).

Additional files

Additional file 1: Figures S1-S5. Figures depicting NHX and SOS1 phylogenies; the species tree used for reconciliation of gene trees; SOS1 phylogenetic tree where branch lengths represent non-synonymous/synonymous substitution rate ratio (dN/dS); and SOS1 phylogenetic tree where branch lengths represent the rate of synonymous substitutions.

Additional file 2: Table S1. List of all sequences used in this study with reference to the database from which they were retrieved. Known salt tolerant plant species are highlighted in grey background. A simplified version (without special characters and limited to 10 characters) of the original sequences’ identifiers was used in this study, since the majority of theophylogenetic software have limitations for sequence names.

Abbreviations
NHX: Sodium hydrogen exchanger; SOS: Salt overly sensitive; dN: Number of non-synonymous substitutions per non-synonymous site; dS: Number of synonymous substitutions per synonymous site; dN/dS: Non-synonymous/synonymous substitution rate ratio.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
ISP participated in the design of the study, carried out the computational analysis, and drafted the manuscript. MMP helped carrying out the ISP participated in the design of the study, carried out the computational analysis. IAA performed the analysis of the analysis, and drafted the manuscript. MMP helped carrying out the ISP participated in the design of the study, carried out the computational analysis. IAA performed the analysis of the analysis, and drafted the manuscript. We also thank funding by Fundação para a Ciência e a Tecnologia (FCT) – Portugal through National Funds (# PEstOE/EEQB/A0004/ 2011), and in part by a grant from the NSF Plant Genome Research Program to MDP. ISP has a PhD fellowship grant (SFRH/BD/58835/2010) from FCT – Portugal. SN and IAA have postdoctoral fellowship grants (SFRH/BPD/34559/2007 and SFRH/BPD/78314/2011, respectively) also from FCT – Portugal.

Acknowledgements
We are grateful to Ulises Rosas (Punuggan Lab, NYU) for critical reading of the manuscript. We also thank funding by Fundação para a Ciência e a Tecnologia (FCT) – Portugal through National Funds (# PEst-OE/EEQB/A0004/2011), and in part by a grant from the NSF Plant Genome Research Program to MDP. ISP has a PhD fellowship grant (SFRH/BD/58835/2010) from FCT – Portugal. SN and IAA have postdoctoral fellowship grants (SFRH/BPD/34559/2007 and SFRH/BPD/78314/2011, respectively) also from FCT – Portugal.

Author details
1Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal and IBET, Apartado 12, 2781-901 Oeiras, Portugal. 2Department of Biology and Center for Genomics and Systems Biology, New York University, New York, US.

Received: 24 May 2013 Accepted: 27 June 2013
Published: 4 July 2013

References
1. Dassanayake M, Oh D-H, Hong H, Bohnert HJ, Cheeseman JM: Transcription strength and halophytic lifestyle. Trends Plant Sci 2011, 16(1):1–3.
2. Oh D-H, Dassanayake M, Bohnert HJ, Cheeseman JM, Life at the extreme: lessons from the genome. Genome Biol 2012, 13:241.
3. Flowers TJ, Galal HK, Bromham L: Evolution of halophytes: multiple origins of salt tolerance in land plants. Funct Plant Biol 2010, 37:604–612.
4. Zhu J-K: Plant salt tolerance. Trends Plant Sci 2001, 6(2):66–71.
5. Apse MP, Blumwald E: Na+ transport in plants. FEBS Lett 2007, 581(12):2247–2254.
6. Venema K, Belver A, Marín-Manzano MC, Rodríguez-Rosales MP, Donaire JP: A Novel Intracellular K+/H+ Antipporter Related to Na+/H+ Antipporters Is Important for K+ Ion Homeostasis in Plants. J Biol Chem 2003, 278(25):22453–22459.
7. Apse MP, Sotosantos JB, Blumwald E: Vaccum cation/H+ exchange, ion homeostasis, and leaf development are altered in a T-DNA insertional mutant of ANNH1X1, the Arabidopsis vacuolar Na+/H+ antipporter. Plant J 2003, 36:229–239.
8. Bowers K, Levi BP, Patel FL, Stevens TH: The Sodium/Proton Exchanger Nhx1p Is Required for Endosomal Protein Trafficking in the Yeast Saccharomycnes cerevisiae. Mol Biol Cell 2000, 11:4277–4294.
9. Brett CL, Tukaye DN, Mukherjee S, Rao R: The yeast endosomal Na+/H+ exchanger Nhxl regulates cellular pH to control vesicle trafficking. Mol Biol Cell 2005, 16(3):1396–1405.
10. Apse MP: Salt Tolerance Conferred by Overexpression of a Vacular Na+/H+ Antipport in Arabidopsis. Science 1999, 285(5431):1256–1258.
11. Hernandez A, Jiang X, Cubero B, Nieto PM, Bressan RA, Haegewa PM, Pardo JM: Mutants of the Arabidopsis thalana cation/H+ antipporter ANNH1X1 conferring increased salt tolerance in yeast: the endosome/prevacuolar compartment is a target for salt toxicity. J Biol Chem 2009, 284(11):7126–7135.
12. Li T, Zhang Y, Liu H, Wu Y, Li W, Zhang H: Stable expression of Arabidopsis vacuolar Na+/H+ antipporter gene ANNHX1, and salt tolerance in transgenic soybean for over six generations. Chin Sci Bull 2010, 55(12):1127–1134.
13. Cao D, Hou W, Liu W, Yao W, Xu C, Liu X, Han T: Overexpression of ToNHX2 enhances salt tolerance of 'compositive' and whole transgenic soybean plants. Plant Cell Tissue Organ Cult 2011, 107(3):541–552.
14. Brett CL, Donowitz M, Rao R: Evolutionary origins of eukaryotic sodium/ proton exchangers. Am J Physiol Cell Physiol 2005, 288:C223–C239.
15. Shi H, Ishitani M, Kim C, Zhu JK: The Arabidopsis thalana salt tolerance gene SOS1 encodes a putative Na+/H+ antipporter. Plant J 2003, 37(12):6896–6901.
16. An R, Chen Q, Chai MF, Lu PL, Su Z, Qin ZX, Chen J, Wang XC: ANNHX8, a member of the monovalent cation: proton antiporter-1 family in Arabidopsis thalana, encodes a putative Li+ antipporter. Plant J 2007, 49(4):718–728.
17. Pardo JM, Cubero B, Leidi EO, Quintero FJ, Alkaline cation exchangers: roles in cellular homeostasis and stress tolerance. J Exp Bot 2006, 57(S):1181–1199.
18. Basé E, Ohto MA, Esumi T, Tajima H, Zhu Z, Cagnac O, Belmonte M, Peleg Z, Yanozuchi T, Blumwald E: The Arabidopsis intracellular Na+/H+ antiporters NHX5 and NHX6 are endosomically associated and necessary for plant growth and development. Plant J 2011, 63(2):224–239.
19. Basé E, Tajima H, Liang Y-C, Ohto MA, Ishijima K, Nakano R, Esumi T, Itoh R, Belmonte M, Blumwald E: The Arabidopsis Na+/H+ Antipporters NHX1 and NHX2 Control Vacuolar pH and K+ Homeostasis to Regulate Growth, Flower Development, and Reproduction. Plant Cell 2011, 23(9):3482–3497.
20. Venema K, Quintero FJ, Pardo JM, Donaire JP: The arabidopsis Na+/H+ exchanger ANNH1X1 catalyzes low affinity Na+ and K+ transport in reconstituted liposomes. J Biol Chem 2002, 277(4):2413–2418.
21. Aharon GS, Apse MP, Duan S, Hua X, Blumwald E: Characterization of a family of vacuolar Na+/H+ antipporters in Arabidopsis thalana. Plant Sci 2003, 182:245–256.
22. Yokoi S, Quintero FJ, Cubero B, Ruiz MT, Bressan RA, Hasewaga PM, Pardo JM: Differential expression and function of Arabidopsis thaliana NHX Na+/H+ antiporters in the salt stress response. Plant J 2002, 30(5):529–539.

23. Li HT, Lu H, Gao XS, Zhang H: Knock-out of Arabidopsis ANTHX gene enhances tolerance to salt stress. Biochim Biophys Acta 2009, 1788(3):637–641.

24. Fukuda A, Nakamura A, Taga G, Tanaka H, Miyao A, Hirooka H, Tanaka Y: Function, intracellular localization and the Importance in Salt Tolerance of a Vacular Na+/H+ Antipporter from Rice. Plant Cell Physiol 2004, 45(1):164–159.

25. Fukuda A, Nakamura A, Tanaka Y: Molecular cloning and expression of the Na+/H+ exchanger gene in Oryza sativa. Biochimica et Biophysica Acta - Gene Structure and Expression 1999, 1446(1–2):149–155.

26. Li M, Lin X, Li H, Pan X, Wu G: Overexpression of AtNHX1 improves tolerance to both salt and water stress in rice (Oryza sativa L.). Plant Cell Tissue Organ Cult 2011, 107(2):283–293.

27. Xue ZY, Zhi DY, Xue GP, Zhang H, Zhao YX, Xia GM: Enhanced salt tolerance of transgenic wheat (Triticum aestivum L.) expressing a vacuolar Na+/H+ antiporter gene with improved grain yields in saline soils in the field and a reduced level of leaf Na+. Plant Sci 2004, 167(4):849–859.

28. Mahajan S, Pandey GK, Tuteja N: Identification of two alternative versions of AtNHX1 gene copy number. Tissue Organ Cult 2011, 98(1):28–33.

29. Quintero FJ, Martinez-Alia V, Cubero B, Ruiz MT, Bressan RA, Hasegawa PM, Pardo JM: Topogenic properties of transmembrane segments of Arabidopsis thaliana proton transporters, AtNhx1 and Avp1, can function in cation uptake into vacuoles to regulate cell turgor and stomatal function in Arabidopsis. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 2014, 42(D1):D1178–D1186.

30. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004, 32(5):1792–1797.

31. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(1):2731–2739.

32. Tautz G, Pires JC: Planta – The neighbor-joining method: a new method for phylogenetic analysis. Mol Biol Evol 1989, 6(5):440–454.

33. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

34. Tautz G, Pires JC: Planta – The neighbor-joining method: a new method for phylogenetic analysis. Mol Biol Evol 1989, 6(5):440–454.

35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

36. Tautz G, Pires JC: Planta – The neighbor-joining method: a new method for phylogenetic analysis. Mol Biol Evol 1989, 6(5):440–454.

37. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

38. Tautz G, Pires JC: Planta – The neighbor-joining method: a new method for phylogenetic analysis. Mol Biol Evol 1989, 6(5):440–454.

39. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

40. Tautz G, Pires JC: Planta – The neighbor-joining method: a new method for phylogenetic analysis. Mol Biol Evol 1989, 6(5):440–454.

41. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

42. Tautz G, Pires JC: Planta – The neighbor-joining method: a new method for phylogenetic analysis. Mol Biol Evol 1989, 6(5):440–454.

43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

44. Tautz G, Pires JC: Planta – The neighbor-joining method: a new method for phylogenetic analysis. Mol Biol Evol 1989, 6(5):440–454.

45. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.