

REVIEW

# Life at the extreme: lessons from the genome

Dong-Ha Oh<sup>1,2†</sup>, Maheshi Dassanayake<sup>1,†</sup>, Hans J Bohnert<sup>1,2,3</sup> and John M Cheeseman<sup>1,\*</sup>

## Abstract

Extremophile plants thrive in places where most plant species cannot survive. Recent developments in high-throughput technologies and comparative genomics are shedding light on the evolutionary mechanisms leading to their adaptation.

**Keywords** extremophile, *Thellungiella*, copy number variation, tandem duplications, orphan genes, taxonomically restricted genes, epigenetics, transposons, microRNA

Vascular plants have adapted to virtually all terrestrial environments, no matter how benign or stressful. Extremophiles are the plants operating in the most challenging environments [1], such as those dominated by the extreme cold in Antarctica [2], wide temperature swings and extreme drought in deserts [3], or salinity in combination with a broad range of other stresses. This last group, the halophytes, are the best documented [4]; the Kew Gardens database [5] recognizes over 1,500 species. Table 1 summarizes some examples of extremophile transcriptomes and genomes that have been published in recent years, at increasing levels of complexity as new sequencing technologies have become available. Six of these plants and their ecological contexts, not all familiar to most plant biologists, are illustrated in Figure 1.

Because of their diverse life forms and life history strategies and in some cases their experimental tractability, halophytes have attracted more attention than the other groups at the molecular level. These include shrubs and forbs (such as *Salicornia* spp. (Table 1, Figure 1d), *Chenopodium* spp., *Atriplex* spp.), grasses (such as *Festuca rubra* (Table 1), *Spartina* spp., *Aeluropus* spp.,

and two adapted to saline sodic deserts, *Leptochloa fusca* and *Leymus chinensis*), trees (several mangroves, especially *Avicennia* and members of the Rhizophoraceae), and desert succulents (especially *Mesembryanthemum crystallinum*, Table 1, Figure 1c). Perhaps most importantly, from the standpoint of comparative genomics, the halophytes also include highly salt-tolerant close relatives of *Arabidopsis thaliana*.

Extremophiles are not simply outliers, plants with little to offer to the mainstream defined by poorly stress-adapted model plants. They occupy one end of a continuum of plant abilities to withstand stress. In all extreme environments, multiple stresses arise concurrently. For example, saline environments are often poor in essential nutrients (especially N and P), but replete to the point of toxicity in others (for example Mg, sulfate or micronutrients). They may experience seasonal swings between flooding and drought-related salt pans (for example, as shown in Figure 1b). Daily and seasonal temperature ranges may be very broad, or, increasingly over the past century, they may be natural or agricultural ecosystems degraded by overgrazing or inappropriate irrigation management. Understanding plants endemic to these environments provides us with the opportunity to understand the successful and unsuccessful adjustments that less tolerant plants make when faced with lesser stresses [1,6].

Plant environmental responses are coordinated through crosstalk among multiple signaling and stress-response networks, and one of the major goals of modern plant biology is to understand these. For example, dehydration response elements, redox controls and the downstream processes they regulate are central to drought and cold responses [7]. In addition, abscisic acid mediates a broad range of environmental responses [8]. But networks are often, if not always, more complicated than can be revealed by analysis of genes 'known to be involved' in particular responses; using Gaussian graphical methods, for example, Ma *et al.* [9] visualized response networks to salt involved in signaling and adaptation - including a large number of unknown and uncharacterized genes. Clearly, 450 million years of land plant evolution has generated biological complexity that cannot be represented by the sequence of a single species, such as *A. thaliana*, or even a single representative of each major

<sup>†</sup>These authors contributed equally.

\*Correspondence: j-cheese@life.illinois.edu

<sup>1</sup>Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Full list of author information is available at the end of the article

**Table 1. Recent studies on extremophile genomes and transcriptomes**

Species	Order	Methods	Stress studied	References
<i>Lobularia maritima</i>	Brassicales	Enriched cDNA library	Salt	[49]
<i>Thellungiella parvula</i>	Brassicales	NGS - genome/transcriptome	Salt	[10]
<i>Thellungiella salsuginea</i>	Brassicales	EST sequencing; enriched cDNA library	Salt, cold	[67]
<i>Amaranthus hypochondriacus</i>	Caryophyllales	NGS - transcriptome	Drought	[68]
<i>Limonium sinense</i>	Caryophyllales	EST sequencing; enriched cDNA library	Salt	[25]
<i>Mesembryanthemum crystallinum</i>	Caryophyllales	EST sequencing	Salt	[69,70]
<i>Salicornia brachiata</i>	Caryophyllales	EST sequencing; enriched cDNA library	Salt	[71]
<i>Avicennia marina</i>	Lamiales	EST sequencing; enriched cDNA library	Salt	[72]
<i>Rhizophora mangle</i>	Malpighiales	NGS - transcriptome	Salt, heat, low nutrient, flooding	[26,73]
<i>Populus euphratica</i>	Malpighiales	Microarray; NGS - transcriptome	Salt, drought	[14,74,75]
<i>Populus x canadensis</i>	Malpighiales	Microarray	Flooding	[76]
<i>Heritiera littoralis</i>	Malvales	NGS - transcriptome	Salt, heat, low nutrient, flooding	[26,73]
<i>Festuca rubra ssp. littoralis</i>	Poales	Enriched cDNA library	Salt	[50]
<i>Spartina alterniflora</i>	Poales	EST sequencing; enriched cDNA library	Salt	[77]
<i>Craterostigma plantagineum</i>	Scrophulariales	NGS - transcriptome	Desiccation	[13]
<i>Solanum commersonii</i>	Solanales	Microarray	Cold	[78]

EST, expressed sequence tag; NGS, next-generation sequencing.

clade. By scrutinizing the few plant genomes that are available, however, the plant biology community is beginning to identify characters of developmental, physiological, and environmental integrative quality that can be deduced and refined into hypotheses for further scrutiny.

Next-generation sequencing (NGS) technologies (especially Roche 454 and Illumina-Solexa) brought with them the promise of high-quality, high-volume, low-cost genomes and transcriptomes. In fact, it is meeting this expectation. Using the resulting datasets, it is now possible to address the evolutionary mechanisms leading to adaptation to extreme environments. The recently sequenced genome of *Thellungiella parvula* [10] exemplifies such efforts, providing resources for high-resolution genome-wide comparison with its non-extremophile relative, *A. thaliana*.

Here, we look at three notable evolutionary features reflected in the genomes that may contribute to adaptations to abiotic stress. These are gene duplication, lineage-specific, largely functionally uncharacterized genes, and epigenomic modifications effected by abiotic stress.

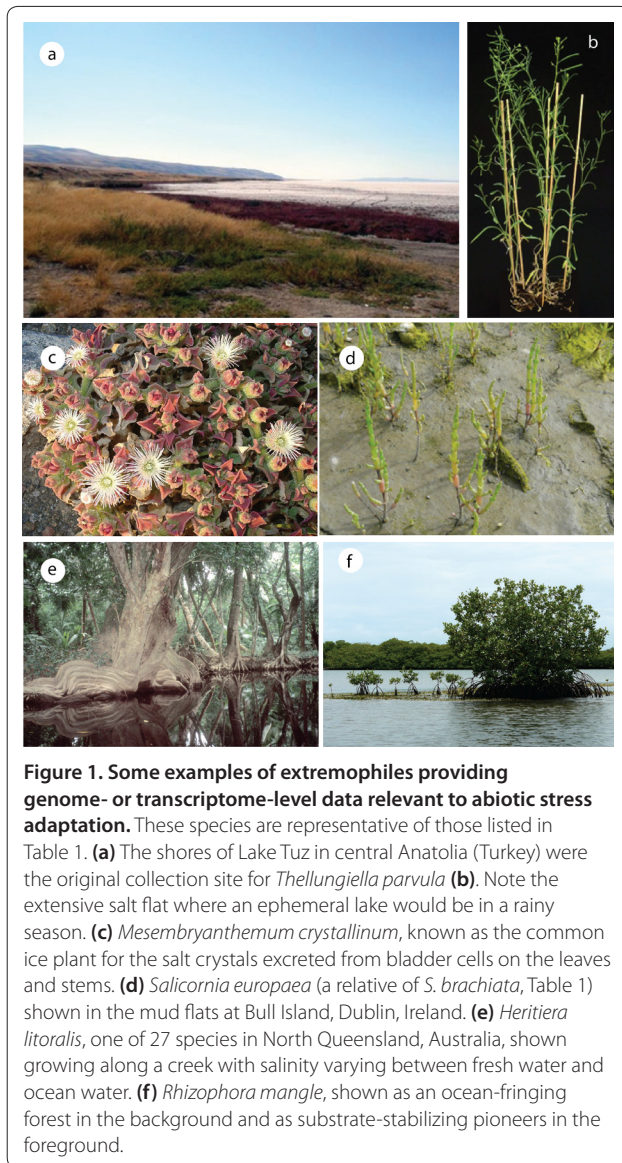
### Genomic resources: the harvest of cheap deep sequencing

Clearly, the search for genetic mechanisms for environmental adaptation was never on hold pending the invention of NGS. Differences in individual genes unquestionably have a big role in adaptation to stress. In some cases, they have been inferred from the primary sequences of well-characterized genes, such as the 37-amino-acid

stretch in L-myo-inositol-1-phosphate synthase, which distinguishes the salt-tolerant wild rice (*Porteresia coarctata*) from domesticated rice (*Oryza sativa*) [11], or the single-amino-acid variation in *AtHKT1;1* (which encodes the high-affinity K<sup>+</sup> transporter 1) that distinguishes coastal from inland clines of *Arabidopsis* [12]. In other cases, they have been implicated by the constitutively higher expression - in the absence of stress - of genes that are induced by stress in *Arabidopsis*, as in the resurrection plant *Craterostigma plantagineum* [13], the salt-tolerant poplar *Populus euphratica* [14], or the *Arabidopsis* relatives *T. parvula* and *Thellungiella salsuginea* (formerly *T. halophila*) [15-17].

But genomes are far more than collections of protein coding sequences. To extend the search for 'genetic mechanisms' beyond this level of primary DNA or cDNA sequences, high-quality genomic resources are a paramount necessity. Especially critical are the genomes of closely related species, or even genotypes, that have adapted to different climates and habitats (that is, that have different lifestyles). Such genomes are beginning to appear, albeit few being proper extremophiles. The strawberry, apple, and peach genomes in the Rosaceae, for example, have begun to reveal how artificial selection for fruit quality has shaped these genomes [18]. Differences reflecting natural selection should also be discernible, for a start, from resources such as those summarized in Table 1.

However, given the long history of *Arabidopsis* as a model system, the new genomes most immediately useful for comparative studies at this point are likely to be those



closely related to it. One of these is the genome of *Arabidopsis lyrata* [19], a potential comparative model for drought tolerance [20], and *T. parvula* (Figure 1a,b) will be perhaps even more useful for elucidating a broad range of environmental adaptations [10]. This species and the congeneric *T. salsuginea* are endemic to regions that experience temperature extremes, poor, degraded, and toxic soils, and especially very high salinities [6,21]. The *T. parvula* genome is of particular interest because chromosomal assemblies that approach the coverage of *A. thaliana* are available. Moreover, because the *Thellungiella* species share many of the characteristics that led to the acceptance of *Arabidopsis* as a model (size, growth habit, seed amount, mutants, and transformation ability), they have been recognized as excellent candidates for comparative genomics studies [15,22].

### Data prospecting and data mining - finding the gems in the genome

Given the evolutionary continuum of genome-level adaptations to abiotic stress, the signatures of the critical adaptive mechanisms must be archived in the genomes of extremophiles. These are the gems in the genome; the challenge is to find and understand them. Comparisons of known genes and transfers between species - the mainstay approach before cheap deep sequencing - can now be supplemented with more extensive genome prospecting, and thereafter with large scale data mining. In this section, we consider three issues as they apply to the problem: what has been explored so far, what has been found, and what is needed to move forward.

First, comparing gene expression at the broad level reflected in Gene Ontology (GO) profiles, stress-tolerant and -sensitive species show different patterns [10]. Salt-tolerant extremophiles, on the one hand, seem to have a bias towards ion transporters in the gene function GO category that is not found in glycophytic species such as *Arabidopsis*. This bias is evident, for example, both in *T. parvula* and *T. salsuginea* [23,24] and in the unrelated salt marsh halophyte *Limonium sinense* [25]. *Arabidopsis*, on the other hand, has invested in an arsenal of pathogen-responsive and developmentally related genes. It is reasonable to suppose - although future research could prove otherwise - that transporters would be critical to salt stress tolerance, and that developmental flexibility and pathogen protection would be important for a winter annual in a high resource environment.

Whole-transcriptome analyses of two mangrove species, *Heritiera littoralis* (Malvaceae; Figure 1d) and *Rhizophora mangle* (Rhizophoraceae; Figure 1e), showed a similar high representation of transport-related genes. Interestingly, despite these species having different life histories and physiological strategies in their adaptation to tropical intertidal habitats, their transcriptomes showed strikingly similar allocations in GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional categories, suggesting convergent evolution as 'mangroves' [26].

Going beyond transcriptomes, at the genome level, where are the gems, that is, what are targets currently considered most promising as being part of integrative mechanisms that lead to stress adaptation? At this point, there are few genomes complete enough to allow detailed comparisons, essentially only *T. parvula* and *A. thaliana*. In these two, although the gene spaces show extensive overall colinearity, there are also major translocations of gene-rich regions and extensive changes in intergenic sequences [10,15]. Beyond this, there are three promising, potentially adaptive linkages to explore. These involve gene duplication, lineage-specific sequences, and epigenetic regulation. We look at these briefly below, with particular reference to their contributions as reflected in

the newly released genome of *T. parvula* and the testable predictions that follow.

### Stress adaptation by gene duplication

A striking feature of all plant genomes is gene enrichment due to duplication events. Suggested by Haldane in 1932 [27] and later popularized by Ohno [28], gene duplication as an evolutionary mechanism that adds new biological function is a well-established idea. Both the duplication rate and the proportion of retained duplicates seem to be greater in plants than in the other domains of life [29]. With respect to individual genes, the result is termed copy number variation (CNV). From resequencing the genomes of 80 individual *Arabidopsis* ecotypes, it seems that natural selection has led to CNVs covering 2.2 Mb of the reference genome [30]. CNVs can also arise in a short time. For example, they appeared in *Arabidopsis* in several generations under the selection pressure of a continuous stress in the laboratory [31]. These were distributed with a 42%:58% ratio between those initiated by transposable elements (TEs) and those involving tandem duplications.

Practically all angiosperms have polyploidy somewhere in their history, either current or long past. The initially increased gene dosage following duplication is often assumed to be beneficial for survival in new habitats, at least in the short term [32]. But although there are certainly polyploid species known for their extreme adaptations to abiotic stresses, an equal fraction are adapted to less harsh conditions, and there are also diploid extremophiles (including *Thellungiella* spp.). Thus, there is little overall evidence that polyploidy itself is a major evolutionary driving force leading to extremophiles.

In most plants, including *T. parvula*, genomes enriched by polyploidy have subsequently experienced extensive gene losses [33]. Their modern genomes reflect this. On the other hand, the copy numbers of other genes have increased as a result of segmental or tandem duplication events and duplication-translocation events. Individual copies of duplicated genes have, in many cases, also assumed new functionality resulting from mutation (neo-functionalization), or become specialized by acquisition of new promoters or regulatory elements (sub-functionalization). One such example is found in allopolyploid cotton (*Gossypium hirsutum*), in which reciprocal silencing of alcohol dehydrogenase homologs led to their expression in different tissues under distinct abiotic stresses [34].

An example of changes in transcript expression and neo-functionalization is provided by homologs encoding HKT1, a plasma membrane Na<sup>+</sup>/K<sup>+</sup> transporter considered to be a genetic determinant of salt tolerance [12,35]. HKT1 exists as tandem duplicated copies in both *Thellungiella* species [10,17]. One copy encodes new

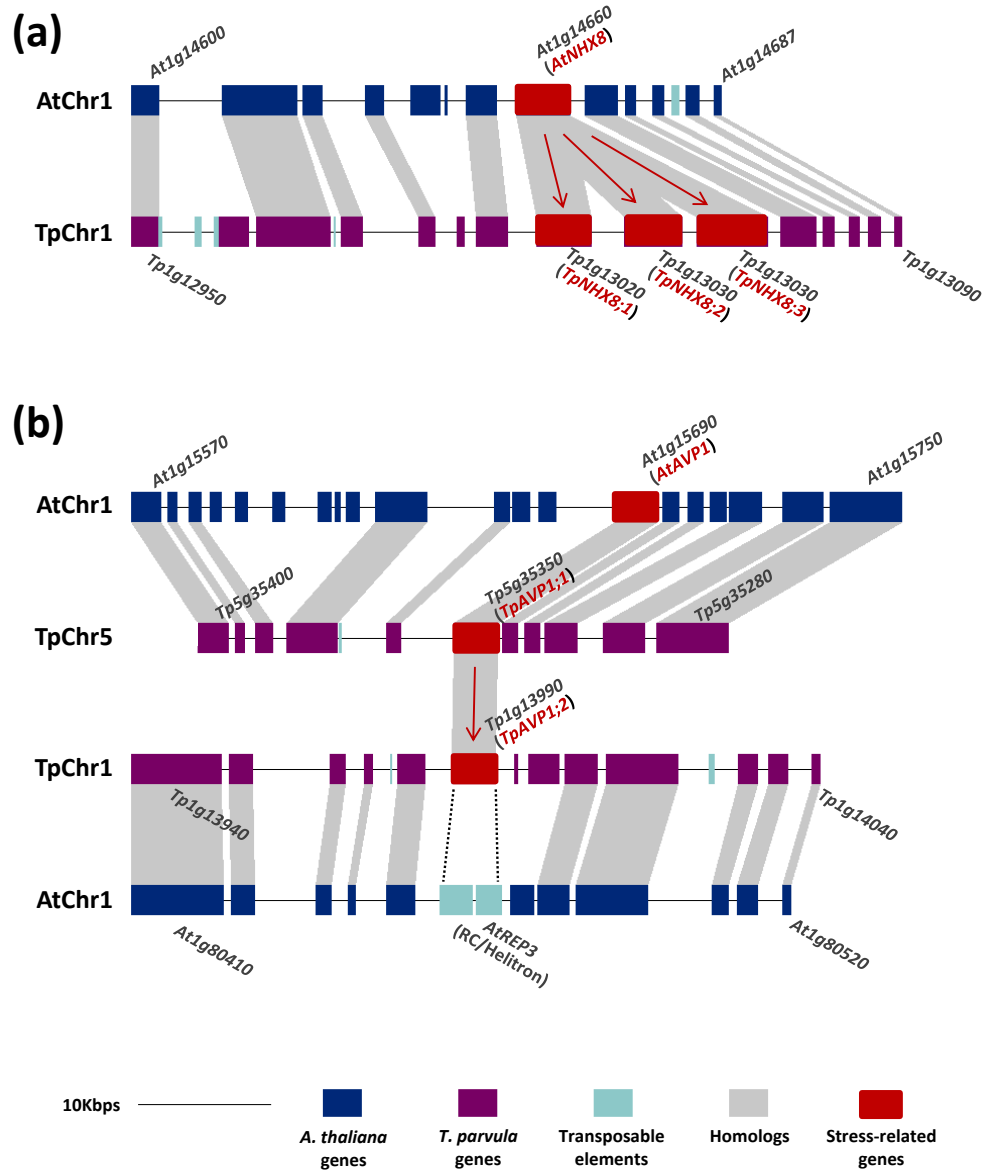
protein functionality and also has an expression pattern different from that of the *Arabidopsis* counterpart [17]. This copy, called *TsHKT1;2* in *T. salsuginea*, is induced under salt stress and leads to continued uptake of potassium ions. By contrast, *TsHKT1;1* in *Thellungiella* behaves like the single-copy *AtHKT1*; because this protein transports sodium ions under salt stress [36], it exacerbates stress unless its expression is downregulated [37].

In *T. parvula* and in *A. thaliana*, a major source of CNV has been tandem duplication [10]. The extant populations of unique tandem duplicates reflect the fact that both copies originated since the species diverged about 11 million years ago [38] and that selective gene loss has occurred in each taxon in response to environmental selective pressures. Either through gene duplication or expression strength differences, a large number of other seemingly stress-relevant genes that have not been recognized in *Arabidopsis* show the hallmarks of CNV in *Thellungiella*, including a variety of ion transporters and membrane-located proton ATPases [10]. Such a difference might be expected, as *Thellungiella* shares only 40% of salt-induced regulation of transcript expression with *A. thaliana* [39].

Tandem duplications seem to have a more important role in shaping genomes for stress adaptations than polyploidy, segmental transposition-duplications, or ectopic duplication and translocation [40]; recombination and tandem duplication events may both become accelerated by environmental challenges [29]. As the result of unequal crossing-over during recombination, tandem duplications vary in their 'genetic neighborhoods,' with copies receiving different regulatory motifs that can lead to drastic changes in expression [40]. A comparative study on plant genomes ranging from *Arabidopsis* to *Physcomitrella* showed genes associated with defense, transport functions, or abiotic stress responses enriched in tandem duplicates, whereas duplicates due to other mechanisms included genes enriched in other intracellular regulatory roles [41].

The *A. thaliana* and *T. parvula* genomes have approximately 10% of their total genes in tandem duplicates [10], and they are clearly implicated in the species' dramatically different stress tolerance strategies. This is exemplified by the amplification of *NHX8* homologs (Figure 2a), known to encode a putative Li<sup>+</sup> transporter in *A. thaliana* [42]. The duplication leads to a constitutively higher expression in *T. parvula* than in *A. thaliana*, which might be responsible for the apparently enhanced tolerance of *T. parvula* to high Li<sup>+</sup> in its natural habitat in central Anatolia [43].

Gene duplication may also result from single gene/segmental transposition-duplication or ectopic duplication/translocation [44] in such a way that any syntenic evidence for its ancestral origin is lost. Comparisons of



**Figure 2. Gain of stress-related gene copies through duplication in *Thellungiella parvula*.** Genomic regions containing homologs of *NHX8*, encoding a plasma membrane  $\text{Li}^+$  transporter [42], and *AVP1*, encoding a vacuolar proton transporter [79], were compared between *T. parvula* (Tp) and *Arabidopsis thaliana* (At). Shown are five colinear genes adjacent to *NHX8* and *AVP1* in the two species. Red arrows indicate duplications. **(a)** *NHX8* is duplicated in tandem into three copies in *T. parvula*. **(b)** *AVP1* homologs are duplicated and translocated from *T. parvula* chromosome 5 to chromosome 1. The colinear genomic region in *A. thaliana* chromosome 1 contains rolling-circle (RC)/helitron transposable elements in the place of an *AVP1* homolog (dashed lines), suggesting a possible involvement of transposable elements in the translocation in an ancestor of the two species. The naming of *T. parvula* genes is according to version 2 of the genome [80].

*T. parvula* and *A. thaliana* genomes indicate multiple translocation-duplication events involving stress-related genes, exemplified by the duplications of orthologs of *CBL10*, encoding a calcium sensor [10], and *AVP1*, encoding a vacuolar proton transporter (Figure 2b) in *T. parvula*. The details of the relationship between this mechanism and stress-adaptive evolution deserve further exploration.

From these initial observations, there are a number of important questions for future studies. For example, how do duplications arise and become stabilized in targeted regions of the genome? Can stress increase the rate of their generation? How rapidly can new regulatory sequences evolve to become operational and do they evolve along with duplicated genes or independently? How rapidly can neofunctionalization occur and how is it

balanced by gene loss? And how is tandem duplication called into play to adjust expression levels?

### Stress adaptation through lineage-specific sequences

In any single genome, the suite of genes shaped by stress during adaptation should reflect, above all, the nature of the stresses. In turn, physiological and developmental changes will mirror genomic changes. Thus, both the suite of altered genes and their regulatory sequences can be expected to demonstrate lineage specificity.

Lineage-specific or taxonomically restricted genes (TRGs) are protein-coding genes that do not share sequence similarity outside the lineage. For that reason, they are also sometimes referred to as 'orphan genes' [45], or 'unknown'. Indeed, with each new EST collection or genome, the number of new unknowns (or 'unknown unknowns') proliferates. Regardless of the taxon, and in all the examples included in Table 1, 10 to 20% of the genes in eukaryote genomes or transcriptomes are TRGs [46]. In the Brassicaceae, family-specific TRGs are enriched for genes responsive to abiotic stresses [47]. It should be noted here that 'stress-responsive' or 'stress-related' are not labels indicating that the functions of the genes are then known. They simply mean that expression is induced by stress. In *Arabidopsis*, but not in *T. parvula*, the expansion is pronounced in pathogen-responsive genes; in *T. parvula*, but not in *Arabidopsis*, the expansion is pronounced in abiotic stress-related genes [10]. Across the spectrum of plant stress tolerance, pools of rapidly evolving TRGs may function as a reservoir of adaptive potential to challenging environments.

In *Arabidopsis*, 3.4% of all genes share sequence similarity only within the Brassicaceae, and another 5% lack similarity with any sequences deposited in public databases [48]. Because the *Arabidopsis* genome is the most fully annotated, it can be expected that the more evolutionarily distant from *Arabidopsis* a species is, the larger will be the number of TRGs, especially if the species is highly adapted to an environment in which *Arabidopsis* cannot survive. In the *T. parvula* genome, 11% of the annotated non-transposon putative protein-coding genes show no sequence similarity with *A. thaliana* genes. About two-thirds of those also lack similarity with any known plant sequence [10]. In *Lobularia maritima* (sweet alyssum), a salt-tolerant coastal relative of *Arabidopsis* [49], 35% of the salt-induced transcriptome is 'unknown', as are half of the salt-stress-induced transcripts from a facultative halophyte, *Festuca rubra* ssp. *litoralis* [50] and nearly 55% of the contigs in two mangrove transcriptomes (*R. mangle* and *H. littoralis*) [26].

Regulatory elements in the untranslated regions and promoters also show lineage specificity. For example, a detailed comparison of the upstream regulatory region of

*SOS1*, a gene critical for salt tolerance in both *Arabidopsis* and *Thellungiella* [51], showed conserved repeat sequences and secondary structures in *Thellungiella* spp. and other halophytes that are absent in *Arabidopsis*. These differences in regions that are not transcribed are correlated with differences in expression observed for *SOS1* in *Thellungiella* [15,16].

TEs seem to have a key role in generating TRGs [31], because novel chimeric genes originate when active retrotransposons recruit new exons from flanking sequences [52]. About 10% of the *Arabidopsis* TRGs showed degenerate sequence conservation with transposable elements, a proportion double that among non-TRGs [47]. In the *T. parvula* genome, TRGs are enriched in pericentromeric TE-rich regions, suggesting roles of transposons in their evolution [10].

Without sequence similarities on which to base annotation, 'orphan genes' usually lack assignable functions [10,26]. Clearly, this is a major obstacle to elucidating the genetic basis for any characteristic, not just for understanding stress tolerance, and overcoming this is an important target. Again, there are associated questions to be addressed. For example, why do duplications, especially those associated with TEs, seem to be clustered in centromeric regions? And how do lineage-specific, taxonomically restricted, or 'orphan' genes fit in the overall picture of functioning organisms? With regard to this last question, network analysis has already proved to be a good starting place. As has already been demonstrated in *Arabidopsis* transcriptional network models, the correlated expression of TRGs and genes with assigned functions in response to stresses provides, even without definitive annotations, useful linkages for visualizing co-expression patterns and identifying 'hub' genes that have core roles in regulating pathways [53,54]. Although still limited for extremophiles, RNA-sequencing experiments performed under both transient and chronic stress conditions should, before long, contribute the expression data needed for extending similar networks to non-model - or new model - species.

### Epigenetic modifications and non-coding RNAs

Beyond adaptations embedded in the basic nucleotide sequence of a genome, epigenetic controls have key roles in ensuring plant survival and reproduction under suboptimal growth conditions [55,56]. Selective hypermethylation on salt stress adaptation in the extremophile Crassulacean acid metabolism (CAM) plant *Mesembryanthemum crystallinum*, for example, indicates both specific and global epigenetic restructuring in plant abiotic stress response regulation [57].

Methylation, alone or in combination with small interfering RNA degradation pathways, can also regulate transposon activity [58]. Although most TEs are inactive

at any time, the proportion that is active is highly dynamic and stress responsive [59,60]. TE copies can vary significantly within single species (for example, maize haplotypes [58]), or between closely related species; in *T. parvula* and *T. salsuginea*, TEs make up about 7.4% [10] and up to 50% (Q Xie, personal communication) of the genome, respectively.

The potential influence of retrotransposon-rich gene neighborhoods undoubtedly varies in ways yet to be fully appreciated. It may, for example, be represented in the *HKT1* locus in *T. parvula* [10], as it is for *Arabidopsis* *TIP1;2*, the aquaporin whose high basal expression has been caused by TEs in the promoter region [61].

Plant microRNAs (miRNAs) also act epigenetically, through target mRNA cleavage or translational inhibition, and their effects are further compounded by feedback regulation. The majority are lineage specific or species specific. Even conserved miRNAs, however, have species-specific functions, as demonstrated by comparisons of *Arabidopsis* and poplar [62]. Only 80% of known miRNAs identified in the *T. parvula* genome share sequence similarity with *A. thaliana* miRNAs. Another 10% are found in Brassicaceae species, but not in *A. thaliana* [10].

An *in silico* comparison of the target sequences of miRNAs in the mRNAs of mangroves and *Arabidopsis* showed that both the conservation of miRNA targets in stress-responsive genes and their placements within those genes are lineage specific. They may also be similarly represented in unrelated species showing similar ecological affinities [23].

Both methylation and miRNA-based epigenetic regulation are fields of intense activity at present and, from the standpoint of stress adaptation, how miRNA targeting comes about and varies between species is an important question. Another is how the functions of miRNAs and protein-coding genes are regulated and coordinated. Can epigenetic signatures due to stress adaptation be trans-generational, and if so, for how many generations? The concept of trans-generational epigenetic stress signatures has support from some studies. For example, when *Arabidopsis* parent populations were exposed to abiotic stresses that increased global methylation, their progeny were more stress tolerant [63]. Similarly, in rice, parents with hypermethylation of particular loci in response to low-nutrient stress produced progeny with increased tolerance [64]. In dandelion (*Taraxacum officinale*), exposure to stress resulted in heritable markers, again implying epigenetic heritability for stress adaptation [65]. In *Arabidopsis* mutants impaired for small interfering RNA biogenesis, increased copy numbers of the *ONSEN* retrotransposon element were induced by heat stress. *ONSEN* insertion, in turn, rendered adjacent genes heat inducible. Unlike in wild-type plants, these numbers

failed to decay over a period of 20 to 30 days. Because transposition was particularly active during flower development and before gametogenesis, the effect was trans-generational [60].

### Concluding remarks

To know that the phenomena we have presented here operate is not sufficient. By themselves, sequences provide only the raw materials for addressing more important questions. On the one hand, they set the stage for exploring how genomes have evolved in plants with different adaptations to environmental conditions. On the other, and more fundamentally, expanding genomic resources bring the opportunity to explore mechanisms of genome evolution themselves.

The recently completed genome sequences of *T. parvula* [10] and the soon to be available genome of *T. salsuginea* [66] are critical resources, enabling high-resolution genome-wide comparisons between extremophiles and their non-extremophile crucifer relatives. Along with a dozen other transcriptomes of extremophile plants and numerous genomes from non-extremophiles, they have supported the ideas, first, that there is a basal set of genes shared between all plants, and second, that a subset of these has experienced selective modification and amplification of a sort required for adaptation to and success in changing or stressful environments. With sequencing technologies evolving rapidly, a 'third generation' of instruments will undoubtedly have an even greater transforming effect.

As output increases in amount and quality and cost comes down, it seems clear that the genome sequence of any plant species deemed important, and eventually multiple ecotypes of each, can, as needed, become available. The value and importance of this cannot be overstated in a world where the population is rising much faster than total agricultural production and land degradation is rapidly reducing the area useable for crops. Extremophiles provide not only a model for what is possible, but for the traits that may be necessary for crops in the future.

### Competing interests

The authors declare that they have no competing interests.

### Author contributions

D-HO and MD were responsible for assembling the literature, the figures and all comparisons between *Thellungiella* spp. and *Arabidopsis*. All authors contributed equally to the final writing.

### Acknowledgements

D-HO and HJB are supported by World Class University Program (R32-10148) at Gyeongsang National University, Republic of Korea and the Next-generation BioGreen21 Program (SSAC, PJ008025), Rural Development Administration, Republic of Korea. We thank Mike Barker (University of Arizona) for his insights and discussions concerning polyploidy and stress adaptation. Sunhee Jeon and Hyewon Hong are gratefully acknowledged for permission to use the photos in Figures 1a,b.

#### Author details

<sup>1</sup>Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>2</sup>Division of Applied Life Science (BK21 program), Gyeongsang National University, Jinju 660-701, Korea. <sup>3</sup>College of Science, King Abdulaziz University, Jeddah, Saudi Arabia.

Published: 6 March 2012

#### References

- Amtmann A, Bohnert HJ, Bressan RA: **Abiotic stress and plant genome evolution. Search for new models.** *Plant Physiol* 2005, **138**:127-130.
- Komarkova V, Poncet S, Poncet J: **Two native antarctic vascular plants, *Deschampsia antarctica* and *Colobanthus quitensis*: a new southernmost locality in the Antarctic peninsula area.** *Arctic Alpine Res* 1985, **17**:401-416.
- Willert DJ, Eller BM, Werger MJA, Brinckmann E: **Desert succulents and their life strategies.** *Vegetatio* 1990, **90**:133-143.
- Flowers TJ, Troke PF, Yeo a R: **The mechanism of salt tolerance in halophytes.** *Annu Rev Plant Physiol* 1977, **28**:89-121.
- Royal Botanic Gardens, Kew: **Salt Tolerance (eHALOPH)** [http://data.kew.org/sid/halophyte.html]
- Amtmann A: **Learning from evolution: *Thellungiella* generates new knowledge on essential and critical components of abiotic stress tolerance in plants.** *Mol Plant* 2009, **2**:3-12.
- Pastori GM, Foyer CH: **Update on stress tolerance common components, networks, and pathways of cross-tolerance to stress. The central role of "redox" and abscisic acid-mediated controls.** *Plant Physiol* 2002, **129**:460-468.
- Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K: **Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport.** *Plant Cell Physiol* 2010, **51**:1821-1839.
- Ma S, Gong Q, Bohnert HJ: **Dissecting salt stress pathways.** *J Exp Bot* 2006, **57**:1097-1107.
- Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ, Cheeseman JM: **The genome of the extremophile crucifer *Thellungiella parvula*.** *Nat Genet* 2011, **43**:913-918.
- Majee M, Maitra S, Dastidar KG, Pattnaik S, Chatterjee A, Hait NC, Das KP, Majumder AL: **A novel salt-tolerant L-myo-inositol-1-phosphate synthase from *Porteresia coarctata* (Roxb.) Tateoka, a halophytic wild rice.** *J Biol Chem* 2004, **279**:28539-28552.
- Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, Vitek O, Salt DE: **A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *ATHKT1;1*.** *PLoS Genet* 2010, **6**:e1001193.
- Rodríguez MCS, Edsgård D, Hussain SS, Alquezar D, Rasmussen M, Gilbert T, Nielsen BH, Bartels D, Mundy J: **Transcriptomes of the desiccation-tolerant resurrection plant *Craterostigma plantagineum*.** *Plant J* 2010, **63**:212-228.
- Dennis J, Katja B, Basem K: **Pathway analysis of the transcriptome and metabolome of salt sensitive and tolerant poplar species reveals evolutionary adaptation of stress tolerance mechanisms.** *BMC Plant Biol* 2010, **10**:150.
- Oh D-H, Dassanayake M, Haas JS, Kropornika A, Wright C, D'Urzo MP, Hong H, Ali S, Hernandez A, Lambert GM, Inan G, Galbraith DW, Bressan RA, Yun D-J, Zhu J-K, Cheeseman JM, Bohnert HJ: **Genome structures and halophyte-specific gene expression of the extremophile *Thellungiella parvula* in comparison with *Thellungiella salsuginea* (*Thellungiella halophila*) and *Arabidopsis*.** *Plant Physiol* 2010, **154**:1040-1052.
- Dassanayake M, Oh D-H, Hong H, Bohnert HJ, Cheeseman JM: **Transcription strength and halophytic lifestyle.** *Trends Plant Sci* 2011, **16**:1-3.
- Ali Z, Ali A, Park HC, Aman R, Kropornika A, Bressan RA, Bohnert HJ, Kim W-Y, Lee SY, Oh D-H, Yun D-J: **TSHKT1;2, a HKT1 homolog from the extremophile *Arabidopsis*-relative *Thellungiella salsuginea*, shows K<sup>+</sup>-specificity in the presence of NaCl.** *Plant Physiol* 2012. doi:10.1104/pp.111.193110.
- Genome Database for Rosaceae [http://www.rosaceae.org]
- Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottillar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, Bergelson J, Carrington JC, Gaut BS, Schmutz J, Mayer KF, Van de Peer Y, Grigoriev IV, Nordborg M, Weigel D, Guo YL: **The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change.** *Nat Genet* 2011, **43**:476-481.
- Sletvold N: **Variation in tolerance to drought among Scandinavian populations of *Arabidopsis lyrata*.** *Evol Ecol* 2011. doi:10.1007/s10682-011-9502-x.
- Bressan RA, Zhang C, Zhang H, Hasegawa PM, Bohnert HJ, Zhu JK: **Learning from the *Arabidopsis* experience. The next gene search paradigm.** *Plant Physiol* 2001, **127**:1354-1360.
- Inan G, Zhang Q, Li P, Wang Z, Cao Z, Zhang H, Zhang C, Quist TM, Goodwin SM, Zhu J, Shi H, Damsz B, Charbaji T, Gong Q, Ma S, Fredricksen M, Galbraith DW, Jenks MA, Rhodes D, Hasegawa PM, Bohnert HJ, Joly RJ, Bressan RA, Zhu J-K: **Salt cress. A halophyte and cryophyte *Arabidopsis* relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles.** *Plant Physiol* 2004, **135**:1718-1737.
- Wong CE, Li Y, Whitty BR, Diaz-Camino C, Akhter SR, Brandle JE, Golding GB, Weretilnyk EA, Moffatt BA, Griffith M: **Expressed sequence tags from the Yukon ecotype of *Thellungiella* reveal that gene expression in response to cold, drought and salinity shows little overlap.** *Plant Mol Biol* 2005, **58**:561-574.
- Taji T, Sakurai T, Mochida K, Ishiwata A, Kurotani A, Totoki Y, Toyoda A, Sakaki Y, Seki M, Ono H, Sakata Y, Tanaka S, Shinozaki K: **Large-scale collection and annotation of full-length enriched cDNAs from a model halophyte, *Thellungiella halophila*.** *BMC Plant Biol* 2008, **8**:115.
- Chen S-H, Guo SL, Wang ZL, Zhao JQ, Zhao YX, Zhang H: **Expressed sequence tags from the halophyte *Limonium sinense*.** *DNA Sequence* 2007, **18**:61-67.
- Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM: **Shedding light on an extremophile lifestyle through transcriptomics.** *New Phytol* 2009, **183**:764-775.
- Haldane JBS: *The Causes of Evolution*. Ithaca, NY: Cornell University Press; 1932.
- Ohno S: *Evolution by Gene Duplication*. London: Springer Verlag, New York-Heidelberg-Berlin; 1970.
- Zhang J: **Evolution by gene duplication: an update.** *Trends Ecol Evol* 2003, **18**:292-298.
- Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D: **Whole-genome sequencing of multiple *Arabidopsis thaliana* populations.** *Nat Genet* 2011, **43**:956-963.
- DeBolt S: **Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales.** *Genome Biol Evol* 2010, **2**:441-453.
- Soltis DE, Buggs RJA, Doyle JJ, Soltis PS: **What we still don't know about polyploidy.** *Taxon* 2010, **59**:1387-1403.
- Jiao Y, Wickert NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, DePamphilis CW: **Ancestral polyploidy in seed plants and angiosperms.** *Nature* 2011, **473**:97-100.
- Adams KL: **Insights into the evolution of duplicate gene expression in polyploids from *Gossypium*.** *Botany* 2008, **86**:827-834.
- Kronzucker HJ, Britto DT: **Sodium transport in plants: a critical review.** *New Phytol* 2011, **189**:54-81.
- Rus A, Lee B-ha, Muñoz-Mayor A, Sharkhuu A, Miura K, Zhu JK, Bressan RA, Hasegawa PM: **ATHKT1 facilitates Na<sup>+</sup> homeostasis and K<sup>+</sup> nutrition.** *Plant Physiol* 2004, **136**:2500-2511.
- Oh D-H, Lee SY, Bressan RA, Yun D-J, Bohnert HJ: **Intracellular consequences of SOS1 deficiency during salt stress.** *J Exp Bot* 2010, **61**:1205-1213.
- Franzke A, German D, Al-Shehbaz IA, Mummenhoff K: ***Arabidopsis* family ties: molecular phylogeny and age estimates in Brassicaceae.** *Taxon* 2009, **58**:425-437.
- Gong Q, Li P, Ma S, Indu Rupassara S, Bohnert HJ: **Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*.** *Plant J* 2005, **44**:826-839.
- Zou C, Lehti-Shiu MD, Thomashow M, Shiu S-H: **Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*.** *PLoS Genet* 2009, **5**:e1000581.
- Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H: **Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli.** *Plant Physiol* 2008, **148**:993-1003.
- An R, Chen Q-J, Chai M-F, Lu P-L, Su Z, Qin Z-X, Chen J, Wang X-C: **AtNHX8, a member of the monovalent cation: proton antiporter-1 family in *Arabidopsis thaliana*, encodes a putative Li<sup>+</sup>/H<sup>+</sup> antiporter.** *Plant J* 2007, **49**:718-728.
- Hamzaoui E, Aksoy A: **Phytosociological studies on the halophytic**



- communities of central Anatolia. *Ekoloji* 2009, **14**:1-14.
44. Freeling M: **Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition.** *Annu Rev Plant Biol* 2009, **60**:433-453.
  45. Tautz D, Domazet-Lošo T: **The evolutionary origin of orphan genes.** *Nat Rev Genet* 2011, **12**:692-702.
  46. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG: **More than just orphans: are taxonomically-restricted genes important in evolution?** *Trends Genet* 2009, **25**:404-413.
  47. Donoghue MT, Keshavaiah C, Swamidatta SH, Spillane C: **Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*.** *BMC Evol Biol* 2011, **11**:47.
  48. Haining L, Gaurav M, Shu O, Amy I, Shin-Han S, Xun G, Robin BC: **Comparative analyses reveal distinct sets of lineage-specific genes within *Arabidopsis thaliana*.** *BMC Evol Biol* 2010, **10**:41.
  49. Popova OV, Yang O, Dietz K-J, Gollmack D: **Differential transcript regulation in *Arabidopsis thaliana* and the halotolerant *Lobularia maritima* indicates genes with potential function in plant salt adaptation.** *Gene* 2008, **423**:142-148.
  50. Diédhiou CJ, Popova OV, Gollmack D: **Transcript profiling of the salt-tolerant *Festuca rubra* ssp. *litoralis* reveals a regulatory network controlling salt acclimatization.** *J Plant Physiol* 2009, **166**:697-711.
  51. Oh D-H, Leidi E, Zhang Q, Hwang S-M, Li Y, Quintero FJ, Jiang X, D'Urzo MP, Lee SY, Zhao Y, Bahk JD, Bressan RA, Yun D-J, Pardo JM, Bohnert HJ: **Loss of halophytism by interference with *SOS1* expression.** *Plant Physiol* 2009, **151**:210-222.
  52. Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J, Vang S, Lu Z, Wong GK, Long M, Wang J: **High rate of chimeric gene origination by retroposition in plant genomes.** *Plant Cell* 2006, **18**:1791-1802.
  53. Ma S, Gong Q, Bohnert HJ: **An *Arabidopsis* gene network based on the graphical Gaussian model.** *Genome Res* 2007, **17**:1614-1625.
  54. Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY: **Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*.** *Nat Biotechnol* 2010, **28**:149-156.
  55. Boyko A, Kovalchuk I: **Genome instability and epigenetic modification-heritable responses to environmental stress?** *Curr Opin Plant Biol* 2011, **14**:260-266.
  56. Mirouze M, Paszkowski J: **Epigenetic contribution to stress adaptation in plants.** *Curr Opin Plant Biol* 2011, **14**:267-274.
  57. Dyachenko OV, Zakharchenko NS, Shevchuk TV, Bohnert HJ, Cushman JC, Buryanov YI: **Effect of hypermethylation of CCWGG sequences in DNA of *Mesembryanthemum crystallinum* plants on their adaptation to salt stress.** *Biochemistry (Moscow)* 2006, **71**:461-465.
  58. Wang Q, Dooner HK: **Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus.** *Proc Natl Acad Sci U S A* 2006, **103**:17644-17649.
  59. Grandbastien MA, Audeon C, Bonnard E, Casacuberta JM, Chalhoub B, Costa APP, Le QH, Melayah D, Petit M, Poncet C, Tam SM, Van Sluys MA, Mhiri C: **Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae.** *Cytogenet Genome Res* 2005, **110**:229-241.
  60. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J: **An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress.** *Nature* 2011, **472**:115-119.
  61. Quigley F, Rosenberg JM, Shachar-Hill Y, Bohnert HJ: **From genome to function: the *Arabidopsis* aquaporins.** *Genome Biol* 2002, **3**:RESEARCH0001
  62. Cuperus JT, Fahlgren N, Carrington JC: **Evolution and functional diversification of *MIRNA* genes.** *Plant Cell* 2011, **23**:431-442.
  63. Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, Hollander J, Meins Jr F, Kovalchuk I: **Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of Dicer-like proteins.** *PLoS One* 2010, **5**:e9514.
  64. Kou HP, Li Y, Song XX, Ou XF, Xing SC, Ma J, Von Wettstein D, Liu B: **Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (*Oryza sativa* L.).** *J Plant Physiol* 2011, **168**:1685-1693.
  65. Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A: **Stress-induced DNA methylation changes and their heritability in asexual dandelions.** *New Phytol* 2010, **185**:1108-1118.
  66. **Phytozome: *Thellungiella halophila* (Salt cress)** [http://www.phytozome.net/thellungiella.php]
  67. Taji T, Komatsu K, Katori T, Kawasaki Y, Sakata Y, Tanaka S, Kobayashi M, Toyoda A, Seki M, Shinozaki K: **Comparative genomic analysis of 1047 completely sequenced cDNAs from an *Arabidopsis*-related model halophyte, *Thellungiella halophila*.** *BMC Plant Biol* 2010, **10**:261.
  68. Delano-Frier JP, Aviles-Arnaut H, Casarrubias-Castillo K, Casique-Arroyo G, Castrillon-Arbelaez PA, Herrera-Estrella L, Massange-Sanchez J, Martinez-Gallardo NA, Parra-Cota FI, Vargas-Ortiz E, Estrada-Hernandez MG: **Transcriptomic analysis of grain amaranth (*Amaranthus hypochondriacus*) using 454 pyrosequencing: comparison with *A. tuberculatus*, expression profiling in stems and in response to biotic and abiotic stress.** *BMC Genomics* 2011, **12**:363.
  69. Kore-eda S, Cushman MA, Akselrod I, Bufford D, Fredrickson M, Clark E, Cushman JC: **Transcript profiling of salinity stress responses by large-scale expressed sequence tag analysis in *Mesembryanthemum crystallinum*.** *Gene* 2004, **341**:83-92.
  70. Cushman JC, Tillett RL, Wood JA, Branco JM, Schlauch KA: **Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C3 photosynthesis and Crassulacean acid metabolism (CAM).** *J Exp Bot* 2008, **59**:1875-1894.
  71. Jha B, Agarwal PK, Reddy PS, Lal S, Sopory SK, Reddy MK: **Identification of salt-induced genes from *Salicornia brachiata*, an extreme halophyte through expressed sequence tags analysis.** *Genes Genet Syst* 2009, **84**:111-20.
  72. Mehta PA, Sivaprakash K, Parani M, Venkataraman G, Parida AK: **Generation and analysis of expressed sequence tags from the salt-tolerant mangrove species *Avicennia marina* (Forsk) Vierh.** *Theor Appl Genet* 2005, **110**:416-424.
  73. Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM: **Comparative transcriptomics for mangrove species: an expanding resource.** *Funct Integr Genomics* 2010, **10**:523-532.
  74. Brinker M, Brosché M, Vinocur B, Abo-Ogiala A, Fayyaz P, Janz D, Ottow EA, Cullmann AD, Saborowski J, Kangasjärvi J, Altman A, Polle A: **Linking the salt transcriptome with physiological responses of a salt-resistant *Populus* species as a strategy to identify genes important for stress acclimation.** *Plant Physiol* 2010, **154**:1697-1709.
  75. Qiu Q, Ma T, Hu Q, Liu B, Wu Y, Zhou H, Wang Q, Wang J, Liu J: **Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*.** *Tree Physiol* 2011, **31**:452-461.
  76. Kreuzwieser J, Hauberg J, Howell KA, Carroll A, Rennenberg H, Millar AH, Whelan J: **Differential response of gray poplar leaves and roots underpins stress adaptation during hypoxia.** *Plant Physiol* 2009, **149**:461-473.
  77. Baisakh N, Subudhi PK, Varadwaj P: **Primary responses to salt stress in a halophyte, smooth cordgrass (*Spartina alterniflora* Loisel.).** *Funct Integr Genomic* 2008, **8**:287-300.
  78. Carvallo MA, Pino M-T, Jeknic Z, Zou C, Doherty CJ, Shiu S-H, Chen THH, Thomashow MF: **A comparison of the low temperature transcriptomes and CBF regulons of three plant species that differ in freezing tolerance: *Solanum commersonii*, *Solanum tuberosum*, and *Arabidopsis thaliana*.** *J Exp Bot* 2011, **62**:3807-3819.
  79. Gao F, Gao Q, Duan X, Yue G, Yang A, Zhang J: **Cloning of an H<sup>+</sup>-PPase gene from *Thellungiella halophila* and its heterologous expression to improve tobacco salt tolerance.** *J Exp Bot* 2006, **57**:3259-3270.
  80. ***Thellungiella* - an *Arabidopsis*-like extremophile** [http://thellungiella.org/]

doi:10.1186/gb-2012-13-3-241

Cite this article as: Oh DH, et al.: Life at the extreme: lessons from the genome. *Genome Biology* 2012, **13**:241.