

RESEARCH HIGHLIGHT

# Genomes of extremophile crucifers: new platforms for comparative genomics and beyond

Simon M Dittami<sup>1,2</sup> and Thierry Tonon<sup>1,2\*</sup>

## Abstract

Recent reports describe the genome sequencing of *Thellungiella salsuginea* and *Thellungiella parvula*, two extremophile crucifers closely related to the stress-sensitive model plant *Arabidopsis thaliana*.

**Keywords** Abiotic stress, adaptation, comparative genomics, extremophile plants, genome evolution

Abiotic stresses, such as salinity, are important factors limiting plant productivity. Over millions of years, extremophile plants have evolved efficient ways of sustaining growth in challenging environments, and the study of these plants can provide valuable insights into the underlying mechanisms of their stress tolerance [1].

## Exploring natural genomic variability

Until recently, studies of extremophiles focused on physiological parameters as well as targeted and medium-throughput gene expression analysis. The development of rapid and cost-efficient next-generation sequencing technologies has cleared the way for studying both the molecular bases of these adaptations on a global scale and the evolutionary mechanisms driving them. In this context, comparisons of closely related extremophile and non-extremophile species are particularly valuable, because their genomes can be easily and reliably aligned, allowing genomic changes to be pinpointed at high resolution. The family of Brassicaceae is especially suited for this type of study, because of the development of *Arabidopsis thaliana* as a model organism, and because of the availability of a number of closely related extremophiles in possession of many of the characteristics that make *Arabidopsis* a good model (including a small genome, the possibility of genetic transformation, and a

short life-cycle). The utility of Brassicaceae genomics in the study of stress tolerance was illustrated in 2011 by the publication of *Arabidopsis lyrata* [2] and *Thellungiella parvula* [3] genomes. Very recently, the sequencing of *Thellungiella salsuginea* [4] has complemented this set of resources. These three species have evolved to differ from *A. thaliana* in their stress tolerance within the last 7 to 12 million years (Figure 1). Their study has focused on structural changes as witnesses of forces driving the evolution of these genomes [2-4], as well as relating these structural changes to functions that may be related to stress tolerance [3,4].

## Structural changes and evolutionary forces

Structural genomic comparisons show that, despite significant differences in genome sizes, large regions of collinearity are present between the four species [2,4,5]. Differences in genome size could be explained mainly by differences in the intergenic regions, where repeated sequences and transposable elements are commonly found. While in *A. thaliana* large numbers of micro-deletions in these areas are suspected, resulting in a reduction of genome size [2], recent activity of transposable elements is thought to be one of the main reasons for the expansion of the *T. salsuginea* [4] and *A. lyrata* [2] genomes. The latter elements are also thought to be one of the factors at the origin of so-called taxonomically restricted, or orphan, genes and gene families, which are new genes that have recently arisen in a taxon. Wu *et al.* [4] show the *T. salsuginea* genome to contain 984 families of such genes, the functions of which still remain to be explored. Finally, tandem duplication, segmental duplication and retrotransposon-related gene duplication may act on genome structure, and also may be related to functional adaptation, both via modifying gene expression and by providing an opportunity for functional diversification. Retrotransposition as a source of gene duplication is found to be especially common in the extremophile *T. salsuginea*, relative to *A. thaliana* [4].

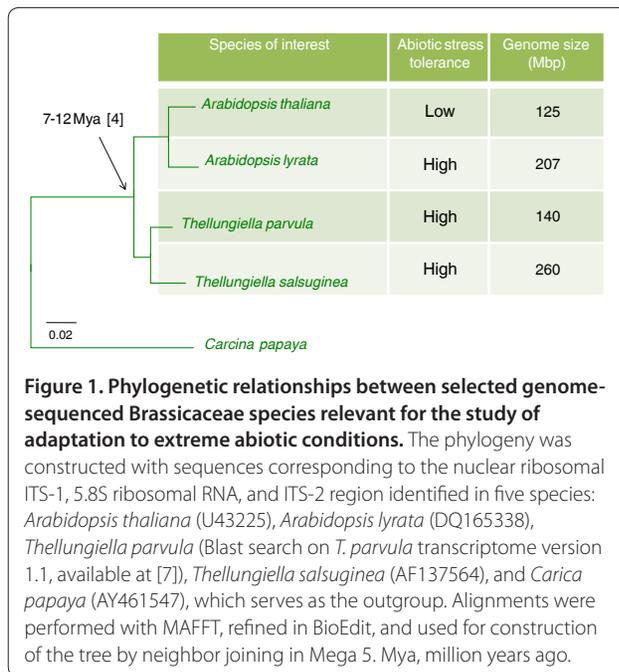
## How genomic changes impact stress tolerance

To elucidate how and which structural genomic changes may influence the stress tolerance of extremophile

\*Correspondence: tonon@sb-roscoff.fr

<sup>1</sup>UPMC Univ Paris 6, UMR 7139 Marine Plants and Biomolecules, Station Biologique, 29680, Roscoff, France

<sup>2</sup>CNRS, UMR 7139 Marine Plants and Biomolecules, Station Biologique, 29680, Roscoff, France



organisms, both Dassanayake *et al.* [3] and Wu *et al.* [4] take advantage of *a priori* knowledge of gene function, as inferred through sequence homology, by applying this functional information to a focus on copy number variations. A simple method for identifying the functional categories whose constituent genes are subject to recent expansion is to perform gene set enrichment analyses. These analyses are frequently based on gene ontology annotations and can be used to identify functional groups of genes (or annotations) that are statistically over- or underrepresented in one genome compared with another. In both *Thellungiella* studies [3,4], gene set enrichment analysis shows that numerous categories of genes already known to be related to abiotic stress, including ‘response to salt stress’, ‘abscisic acid stimulus’, ‘transporter activity’ and ‘development’, are indeed overrepresented in the genomes of these halophytes compared with *A. thaliana*. This finding is of fundamental importance, as it demonstrates that positive selection for duplications of existing stress-response genes is likely to play a part in the adaptation to high abiotic stress environments.

Manual analysis and annotation of gene families that have undergone selective expansion in extremophiles is then used by Wu *et al.* [4] in *T. salsuginea* to highlight specific genes and processes involved in stress tolerance, including genes encoding transcription factors, abscisic acid biosynthetic enzymes, a key enzyme involved in wax production, and the sodium transporter HKT1. The latter gene is of particular interest in *T. salsuginea* as *HKT1* has not only been triplicated with respect to *A. thaliana*, but one of the three copies has also undergone a significant

increase in expression, most likely related to changes in the *cis*-regulatory region. Such changes have also been recorded for another sodium transporter encoded by the *salt overly sensitive 1 (SOS1)* gene. For this gene, both *T. parvula* and *T. salsuginea* copies exhibit homologous promoter regions and high expression, whereas the *A. thaliana* transporter differs in promoter sequence and is expressed at lower levels [5], providing an example of how genomic changes in non-coding sequences may affect physiology.

### Future challenges and directions

The recent work on *Thellungiella* [3,4] shows that, even for closely related species, successful adaptation to abiotic stress is likely to involve the combination of numerous genomic changes related to known but also to novel genes, which are driven by various evolutionary mechanisms. The studies provide a starting point for a fresh approach, namely the genome-wide, and thus holistic, study of the molecular bases of adaptation, and pave the way for the use of systems biology tools to construct and model metabolic and regulatory networks. These models can, in a next step, be expanded to combine other forms of high-throughput data, such as metabolomic results and/or observations on small RNAs, DNA methylation and other forms of epigenetic regulation [6]. At the same time, targeted approaches aiming to identify the function of individual genes are still needed, as even the best models will be incomplete until they are able to incorporate and understand the numerous orphan genes and gene families, which may well be the most innovative features of extremophile genomes. In the same vein, the most complete regulatory networks would remain hypothetical until backed up by experimental evidence.

We anticipate that the discussed studies, exploiting natural genetic variability of land plants to study the evolutionary processes of adaptation to extreme environments, will be inspirational for the development of similar approaches in other organisms. For instance, algae have also colonized environments spanning a wide range of abiotic factors. Meta-comparisons of the evolutionary mechanisms underlying adaptation across lineages and kingdoms will then provide insights into both conserved and specific mechanisms, and increase our understanding of the general principles underlying adaptation.

### Competing interests

The authors declare that they have no competing interests.

Published: 16 August 2012

### References

1. 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.
2. 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.
3. 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.
  4. Wu HJ, Zhang Z, Wang JY, Oh DH, Dassanayake M, Liu B, Huang Q, Sun HX, Xia R, Wu Y, Wang YN, Yang Z, Liu Y, Zhang W, Zhang H, Chu J, Yan C, Fang S, Zhang J, Wang Y, Zhang F, Wang G, Lee SY, Cheeseman JM, Yang B, Li B, Min J, Yang L, Wang J, Chu C, *et al.*: **Insights into salt tolerance from the genome of *Thellungiella salsuginea*.** *Proc Natl Acad Sci U S A* 2012, **109**:12219-12224.
  5. Oh DH, 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 DJ, Zhu JK, 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.
  6. Oh DH, Dassanayake M, Bohnert HJ, Cheeseman JM: **Life at the extreme: lessons from the genome.** *Genome Biol* 2012, **13**:241.
  7. **BLAST against *T. parvula* sequences** [<http://thellungiella.org/blast/thellungiella.html>]

doi:10.1186/gb-2012-13-8-166

**Cite this article as:** Dittami SM, Tonon T: Genomes of extremophile crucifers: new platforms for comparative genomics and beyond. *Genome Biology* 2012, **13**:166.