# De novo assembly and annotation of the grey reindeer lichen (Cladonia rangiferina) transcriptome

**Sini Junttila, Stephen Rudd**✉

Turku Centre for Biotechnology, University of Turku, Turku, Finland

## Motivation and Objectives

Lichens are symbiotic organisms that have a remarkable ability to survive in some of the most extreme terrestrial climates on earth. Lichens can endure frequent desiccation and wetting cycles and are able to survive in a dehydrated state for decades at a time. Genetic resources have been established in lichen species for their taxonomic classification, but no lichen species have been characterised yet using genomics, and the molecular mechanisms underlying the lichen symbiosis and the fundamentals of desiccation tolerance remain undescribed. Research on lichen gene expression is very limited, and as yet there is little in the way of either high-throughput genome sequence or expressed sequence tag (EST) data available for any lichen species. For non-model organisms, de novo genome assembly of short read data is complicated but many transcriptomes have been sequenced from non-model species and published over the last years (Alagna *et al.*, 2009, Novaes *et al.*, 2008, Vera et al. 2008). The annotation process remains challenging, especially for species with no close relatives with a sequenced reference genome.

Our objective was to de novo assemble and annotate a lichen transcriptome using both next-generation sequencing and traditional Sanger sequencing. We produced additional Sanger EST sequences from axenically grown symbiotic partners (*C. rangiferina* and *Asterochloris* sp.) to train classification models for predicting the genome of origin of lichen sequences. We have obtained a basic view of the ongoing molecular processes and have identified the most active biological pathways in *C. rangiferina* (Junttila and Rudd, 2012). Our transcriptome data brings an increase to the amount of publicly available lichen sequences and provides a starting point for further studies into lichen functional transcriptomics.

## Methods

cDNA from lichen tissue was sequenced with Roche GS FLX platform, which was chosen for its longer read length compared to other next-generation sequencing platforms, and Sanger sequencing with ABI PRISM 3130xl Genetic Analyzer capillary DNA sequencer was performed to complement the FLX run data with its long sequence reads. EST sequences from cDNA libraries prepared from the axenically grown symbiotic partners, the fungus and the alga, were obtained using Sanger sequencing.

The lichen sequences were de novo assembled with CLC Genomics Workbench software version 4.9 (CLCBio, Denmark). Prior to the assembly the sequences were trimmed in the CLC Genomics Workbench, and sequences shorter than 15bp were removed from the analysis. Eclat (Friedel *et al.*, 2005) was used to identify the genome of origin for the assembled contigs and singletons derived from lichen tissue. The Sanger sequences obtained from the axenically grown algal and fungal symbiont cDNA libraries were used to train Eclat and build a model file for the classification. The minimum sequence length for classification was set at 100 bp.

Blast2GO (Conesa *et al.,* 2005) tool was used for BLASTX, Gene Ontology (GO) term annotation, Interpro scans and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the contigs and singletons. BLASTX (Altschul *et al.,* 1990) was used to compare the assembled contigs and singletons to a non-redundant (nr) protein sequence database from the NCBI GenBank database (Benson *et al.* 2008). BLASTX matches were filtered using an arbitrary cut-off of 1e-10.

## Results and Discussion

Altogether 243,729 high quality reads were de novo assembled into 16,204 contigs and 49,587 singletons. 62.8% of the sequences were classified as being of fungal origin while the remaining 37.2% were predicted as being of algal origin. In

the annotation 34.4% of the sequences had a BLAST match, 29.3% of the sequences had a GO term match and 27.9% of the sequences had a domain or structural match following an InterPro search. 60 KEGG pathways with more than 10 associated sequences were identified.

Our results present a first transcriptome sequencing and de novo assembly for a lichen species and describe the ongoing molecular processes and the most active pathways in *C. rangiferina.* This brings a meaningful contribution to publicly available lichen sequence information. These data provide a first glimpse into the molecular nature of the lichen symbiosis and characterize the transcriptional space of this remarkable organism. These data will also enable further studies aimed at deciphering the genetic mechanisms behind lichen desiccation tolerance.

## Acknowledgements

## References

Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, et al. (2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* **10**:399. doi: 10.1186/1471-2164-10-399

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *J Mol Biol* **215**(3):403–410.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. *Nucleic Acids Res* **36**:D25–30

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**(18):3674–3676.

Friedel CC, Jahn KH, Sommer S, Rudd S, Mewes HW, et al. (2005) Support vector machines for separation of mixed plant-pathogen EST collections based on codon usage. *Bioinformatics* **21**(8):1383–1388.

Junttila S and Rudd S (2012) Characterization of a transcriptome from a non-model organism, Cladonia rangiferina, the grey reindeer lichen, using high-throughput next generation sequencing and EST sequence data. *BMC Genomics* **13**:575. doi: 10.1186/1471-2164-13-575.

Novaes E, Drost D, Farmerie W, Pappas GJ, Grattapaglia D et al. (2008) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. *BMC Genomics* **9**:312. doi: 10.1186/1471-2164-9-312

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL et al. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* **17**(7):1636–1647. doi: 10.1111/j.1365-294X.2008.03666.x