

ORAL PRESENTATION

Open Access

Haploid transcriptome analysis reveals allelic gene expression variants, co-expressed gene groups, and linkages between expression and copy number variation

Jukka-Pekka Verta^{1*}, Sebastien Caron¹, Isabelle Giguère¹, Brian Boyle¹, Christian R Landry², John MacKay¹

From IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery
Arraial d Ajuda, Bahia, Brazil. 26 June - 2 July 2011

Background

Genetic variation can cause changes in gene expression (mRNA abundance) among individuals. This so-called heritable variation in gene expression is affected by genetic variants that are co-segregating with the gene locus (local/*cis* effects) and/or segregating independently from it (distant/*trans* effects). Genetic variation in gene expression can be measured to estimate the extent of variation in gene expression within a population, and to determine to what degree expression alleles of different genes are connected within regulatory networks. Furthermore, determining whether variation in the expression of a gene is linked to local or distant effects allows us to make inferences about how heritable variation may change depending on gene function, the number of interacting partners, genetic architecture and evolutionary history [1,2].

Detecting heritable variation in gene expression can be a challenging task in diploid organisms, mainly because of tissue-specificity and dominance effects of allelic expression. For example, up to 70% of gene expression alleles in *Drosophila* may be masked by dominance [3]. We developed an experimental system to overcome these obstacles by utilizing the conifer seed's maternally derived haploid tissue, the megagametophyte. Analyzing a set of sibling megagametophytes allows us to first, measure separately the expression each of the two alleles in the maternal genome in the absence of dominance and second, identify genes whose expression levels are co-segregating. In addition, the megagametophyte allows

us to categorize the underlying genetic variants into local or distant with a simple co-segregation assay.

Methods

We set out to characterize segregating variation in gene expression in white spruce (*Picea glauca* [Moench] Voss). We analyzed the transcriptomes of germinating sibling megagametophytes from two controlled-crossed families (C9412516: male 2388 x female 77111 & C9612856: male 80109 x female 80112) with a custom microarray comprised of 32,000 spotted oligonucleotides, which represent over 25,000 unique white spruce genes. Each megagametophyte was split into two halves to provide technical replicates that were analyzed separately. A separate comparison of microarray results and RNA-Seq data has been carried out to validate the quality of the microarray.

The single-color microarray data was background-corrected and quantile-normalized with the R package Limma [4]. We used the R package Mclust [5] to test for unimodal vs. bimodal expression distributions of each gene across sibling megagametophytes. Genes that exhibited at least two expression alleles, which segregated within the 95% IC of their expected proportions, were selected. The segregation had to be repeatable with a replicate sample set to be considered valid.

Results and discussion

Analysis of two families of sibling megagametophytes (n=18) identified close to a thousand genes with segregating gene expression patterns in both of the two families. Approximately 10% of these genes were shared. Zero replicates are expected to have the same clustering

* Correspondence: jukka-pekka.verta.1@ulaval.ca

¹Centre d'étude de la forêt, Université Laval, Québec QC, Canada
Full list of author information is available at the end of the article

pattern by chance alone (binomial $p \approx 3.810 \cdot 10^{-6}$ | $n=18$). The number of variable genes is comparable to that found between 50 nearly isogenic *Drosophila* lines [6].

We have discovered a large number of genes with gene expression patterns segregating in a Mendelian way in white spruce. We are presently analyzing the relationships between gene function and paralog number with the variation in its expression in order to determine whether heritable variation in gene expression is associated with same genetic attributes in white spruce as what has been reported in model organisms. We have also begun investigating the contribution of local vs. distant effects on the expression alleles identified in the megagametophytes, and their nature, by studying the co-segregation of gene expression, SNPs and copy number variations (CNVs). Preliminary comparative genomic hybridization data suggests a significant portion of genes which show segregating gene expression alleles also exhibit CNVs. In follow-up experiments, we will address the amount of dominance between the gene expression alleles by comparing gene expression in megagametophytes versus self-fertilized embryos.

Acknowledgments

The seed material was provided by J. Beaulieu (Canadian wood fibre centre).

Author details

¹Centre d'étude de la forêt, Université Laval, Québec QC, Canada.

²Department of Biology, Université Laval, Québec QC, Canada.

Published: 13 September 2011

References

1. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL: **Genetic properties influencing the evolvability of gene expression.** *Science* 2007, **317**:118-121.
2. Kliebenstein D: **A role for gene duplication and natural variation of gene expression in the evolution of metabolism.** *PLoS One* 2008, **3**:e1838.
3. Lemos B, Araripe LO, Fontanillas P, Hartl DL: **Dominance and the evolutionary accumulation of cis-and trans-effects on gene expression.** *PNAS* 2008, **105**:14471-14476.
4. Smyth GK, Speed T: **Normalization of cDNA microarray data.** *Methods* 2003, **31**:265-273.
5. Fraley C, Raftery AE: **MCLUST version 3 for R: Normal mixture modeling and model-based clustering.** *University of Washington Tech Report* 2009, **504**.
6. Hsieh W, Passador-Gurgel G, Stone EA, Gibson G: **Mixture modeling of transcript abundance classes in natural populations.** *Genome Biology* 2007, **8**:R98.

doi:10.1186/1753-6561-5-S7-O47

Cite this article as: Verta et al.: Haploid transcriptome analysis reveals allelic gene expression variants, co-expressed gene groups, and linkages between expression and copy number variation. *BMC Proceedings* 2011 **5**(Suppl 7):O47.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

