Molecular evolution analysis of genetic network related to plant trichome development

A. V. Doroshkov[1], D.K. Konstantinov[1,2]D. A. Afonnikov[1] and K. V. Gunbin[1,2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*
ad@bionet.nsc.ru

Trichomes are involved in many significant functions such as the transpiration, thermoregulation and protection from insect attacks. On the other hand specialized cell formation is an fruitful model system for analyzing the molecular mechanisms of plant cell differentiation, including cell fate choices, cell cycle control, and cell morphogenesis. In plants, epidermal cells are easily accessible and allow in vivo study. Unicellular trichome formation is a classical experimental model for identification of the activator–inhibitor and the activator–depletion pattern formation models, studying the interplay between cell cycle and cell differentiation and numerous of genes involved in these processes were found (Wang et. al., 2014). However, the evolution of specialized epidermal cell formation genetic network remains unclear. In this study, we analyze the phylogenetic relationships of genes associated with the formation of trichomes and root hairs from various species of plants.

Using the text mining technology we reconstructed the network of interactions between known leaf pubescence genes using *A. thaliana* as a model organism.We doing so using step-by-step procedure: (1), the additional gene pasted to the initial experimentally proved core gene network if this gene co-expressed (and protein co-localized) with more than the half of genes from core gene network or gene subnetwork, co-expression and co-localization data should be provided by several independenthigh-throughput experiments; (2), repeating step (1) until the opportunity for genes addition will not be exhausted.Genetic networks containing target genes were reconstructed using Cytoscape and Pathway Studio software. For each network node (gene) the extraction of sets of homological protein sequences was carried out using the reciprocal DELTA-BLAST search. Multiple sequence alignment was conducted with MAFFT algorithm. The PhyML maximum likelihood algorithm was used to reconstruct the phylogenetic trees and bootstrap resampling technique was used for testing the topology. The following comparative phylogenetic procedures intended to differentiation between functionally conservative and functionally variable gene network components were implemented: (1) the statistical comparison of mean branch lengths from the last common ancestors (LCAs) of land and seed plants to the extant tree tips (for identification of slowly and rapidly evolved protein families); (2) searching for inner branch length outliers and relating these outliers to the absolute evolutionary time (for identification of evolutionary stages characterized by slow or rapid changes); (3) comparing the stemness of phylogenetic trees (in order to identification of protein families under relaxed negative selection in the present time). The

statistical significance of comparative phylogenetic analyses were obtained based on trees from the bootstrap resampling procedure. Additionally, to test the influence of positive selection on the gene network components evolution we applied: (1) several nonsynonymous to synonymous substitution rate tests implemented in PAML 4 to cases without synonymous substitutions saturation (superficial tree branches); (2) several radical to conservative amino acid substitution rate testsimplemented in SAMEM (Gunbin et. al., 2012) to cases with synonymous substitutions saturation (deep inner tree branches).

The reconstructed gene network of *A. thaliana* contained a highly connected part and several satellite fragments. Core fragment characterized by high connectivity with protein-protein interactions. This fragment contain proteins of the known "trichome initiatoion complex": WD40 (TTG1), HLH (GL3, EGL3, TT8, MYC1), and 13 R2R3-MYB factors, 7 complex inhibitors (R3-MYB factors). In addition, it was necessary to mention a gene subnetwork consisting of 8 proteins, 7 of which are zinc finger proteins of C2H2 type that are responsible for modifying the "trichome initiatoion complex" functioning in response to hormonal stimulis. Analysis of nonsynonymous to synonymous substitution rates particular tree length differences were conducted for all these proteins. Note, the "trichome initiatoion complex" according to the protein annotations, also participates in the formation of unicellular outgrowths of the root surface, the root hairs.

The main stages of gene network evolution has been traced down to the evolutionary time of appearance of its components. It was found that the LCA of dicotyledonus plants in a comparison with *A. thaliana* have a reduced set of R2R3-MYB factors (7 versus 13), and at the LCA of all flowering plants also have a reduced set of R3-MYB inhibitors (1 versus 7). It is interesting to note that the duplication events in "trichome initiatoion complex" genes were not found in the lineage of monocotyledonous plants, compared with dicotelidonus plants lineage. For instance, rice corn and barley have a reduced set of genes resembling the ancestor set of all flowering plants. This is consistent with the fact that monocotyledonous plants form monocellular simply shaped hairs whereas dicotyledonus plants form branching (Arabidopsis) or even multicellular (potatoes, a tomato) trichomes with secreting function. Our results argue that there is a fraction of genes involved in the formation of trichomes and root hairs passed several relatively young duplication events and do not reveal a direct correspondence between monocotyledonous and dicotyledonous plants. Thiat means that a part of the cellular morphogenesis mechanisms evolved independently in dicots and monocots which reflects morphology differences. At the same time, we observe a good correspondence between studied genes of cell morphogenesis inside dicotyledonous as well as monocotyledonous clades.

The first appearance of the components of the core gene network was dated to more than 500 million years ago (MYA), even before the divergence of mosses from flowering plants. The

ancestral gene network for LCA of this evolutionary old period was predicted to have WD40 protein (TTG1), 2 HLH proteins (GL3/EGL3 & TT8), 4 R2R3-MYB factors, R3-MYB factor, 2 zinc finger protein of C2H2 type, and also a number of additional components. According to the *A. thaliana* experimental data it is well known that function of "trichome initiatoion complex" components is considerably redundant. Thus,the predicted minimumal ancestral gene network still should perform the cell fate specification via pattern formation. At the same time, we did not found any genes encoding proteins having the domain structure corresponding to trichome-related genetic network in the *Osterococcus* and *Chlamidomonas* genomes (monocelluar algae). Therefore, it is most likely that the origination of this gene network should be dated to the primitive multicellular plants. After that, there was a series of gene network complication events on various lineages leading to higher plants, which are related to trichome cellular structure complication.

This study was funded by Russian Science Foundation grant №14-14-00734.

1. Gunbin, K. V., Suslov, V. V., Genaev, M. A., & Afonnikov, D. A. (2012). Computer system for analysis of molecular evolution modes (SAMEM): analysis of molecular evolution modes at deep inner branches of the phylogenetic tree. In silico biology, **11(3, 4)**: 109-123.

2. Wang, S., Normal, N., Zeng, Q., Wang, Y., Pattanaik, S., & Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Regulation of Cell Fate Determination in Plants, 12.