151. A first step towards an evolutionary model of gene expression

Gunter Weiß [1] [2]

Keywords: gene expression, evolution, probabilistic model

1 Introduction

Differences in expression levels between two species have arisen by some evolutionary process. Mutations in the upstream part of a gene are one possible mechanism that leads to changes of expression levels over time. The reasoning here might be that mutations at promoter, repressor or enhancer binding sites lead to an alteration of the efficiency of the transcription process. Here, I propose a simple probabilistic model that already catches some features we see in comparative data sets. I describe the mutation process as a Poisson process and model the effect of a mutation on the expression level multiplicatively. This is reasonable, since it is more intuitive that mutations that hit sites relevant for transcription do not add or subtract a value from the expression level, but rather increase or decrease the rate of transcription, i.e. a multiplicative effect on expression. Using the resulting compound Poisson process model I am able to reproduce a couple of interesting features one observes while comparing real data expression profiles.

2 Probabilistic model and its features

Mutational model. The mutational model is a Poisson process with rate λt, where t describes a time period of evolution. Let $N(t)$ be the number of mutations within time period t, that hit the (relevant) upstream region of a gene.

Modeling the effect of a mutation.

The effect of a mutation is assumed to be multiplicative, i.e. a mutation changes the expression level by a factor. Thus, I change to the logarithm of variables, such that the model becomes additive. Then, the effect X of a single mutation is defined as the logarithm of the ratio of expression level after and before the mutation. A mutation could cause an increase or decrease in expression level. However, I assume that the net effect of mutations in terms of expression changes equals zero. This means that over evolutionary time the amount of mRNA molecules in a cell remains stable. I also assume that a single random mutation causes frequently a decrease and only rarely an increase of expression. ¿From an evolutionary point of view this seems reasonable, since the already highly specialized transcription machinery will only rarely benefit from a random alteration of the binding sites.

Therefore, the distribution that models the effect of a mutation should have two features: its mean should be zero and its skewness should be positive.

Compound Poisson process.

Putting the two parts together we get the following model for the evolution of gene expression. Let $Y(t)$ be the logarithm of the ratio of expression levels at time t and time 0. We have $Y(t) = \sum_{i=1}^{N(t)} X_i$, with the X_i independent and identically distributed as X. This defines a compound Poisson process with independent increments. Obviously, I assume effects of several mutations as mutually independent.

[1] WE Informatik, Heinrich-Heine Universität Düsseldorf. E-mail: weiss@cs.uni-duesseldorf.de

[2] Max-Planck-Institut für evolutionäre Anthropologie, Leipzig. E-mail: weiss@eva.mpg.de

Differences in expression levels between two samples.
For analyses of available comparative data [1, 2] I investigate the distribution of differences between expression levels of two samples (e.g. human / chimp).

Assume that the two samples involved (or their expression levels) diverged at time 0. Let $Y_j(t_j)$ be the ratio of expression levels at time t_j and time 0 in sample $j = 1, 2$. The actual expression difference between two samples is the random variable $Z = Y_1(t_1) - Y_2(t_2)$.

3 Predictions of the model, qualitative results

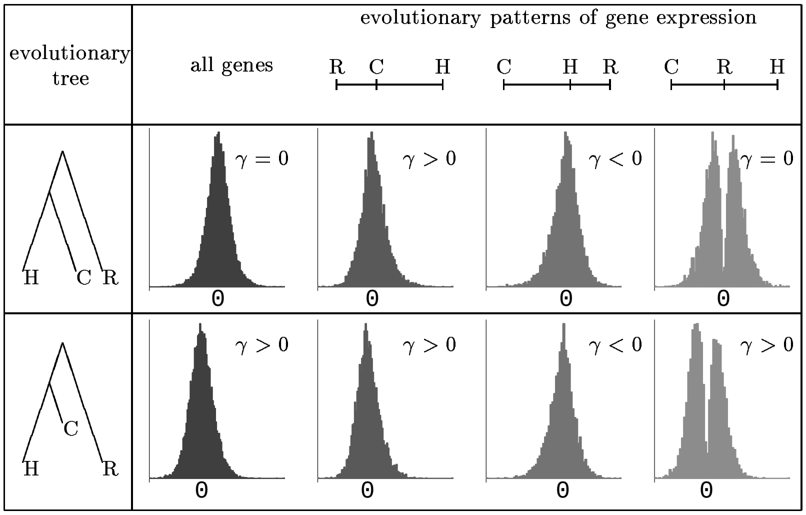

Figure 1: Simulated distributions ($\log_2\left(\frac{H}{C}\right)$) of differences in expression level between genes from two samples H and C; simulated distributions ($\log_2\left(\frac{H}{C}\right)$) of differences in expression level between two samples H and C, when genes are grouped according to their "evolutionary" pattern using an outgroup sample R. γ denotes the skewness of the distribution.

- Model parameters can be estimated via a system of equations containing moments of Z.
- A symmetric distribution of Z is the result of an equal amount of elapsed evolutionary time in both lineages. The skewness of Z is a measure for the difference in evolutionary rate.
- Given data from an outgroup species, grouping genes according to their "evolutionary" pattern disclose the assumed skewness in the distribution of X even when Z is symmetric.
- These predictions fit very well with data characteristics we find when comparing primate transcriptomes [1, 2].

References

[1] Enard, W.*et al.* 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–3.

[2] Khaitovich, P., Weiss, G. and Pääbo, S. 2003. Evolution of the Primate Transcriptome. *in preparation*