Phylogenetic Inference using RevBayes

Chromosome Evolution

William A. Freyman and Sebastian Höhna

Introduction

A central organizing component of the higher-order architecture of the genome is chromosome number, and changes in chromosome number have long been understood to play a fundamental role in evolution. This tutorial will introduce phylogenetic models of chromosome number evolution, and demonstrate how to use RevBayes to estimate the rates of chromosome number change and ancestral chromosome numbers. We will begin by providing an overview of the basic ChromEvol model and an example RevBayes analysis. This is followed by a discussion of a number of model extensions that enable joint inference of phylogeny and chromosome numbers, tests for correlated rates of phenotype and chromosome evolution, and allow for cladogenetic changes in chromosome number. Next, we will introduce the ChromoSSE model which jointly estimates diversification rates and chromosome number evolution. Finally, we will demonstrate how to summarize the results of a RevBayes chromosome number analysis and generate plots of ancestral chromosome numbers.

Contents

The Chromosome Number Evolution tutorial contains several sections:

Recommended tutorials

This tutorial assumes the reader is familiar with the content covered in the following RevBayes tutorials:

- **Rev Basics**
- **Molecular Models of Character Evolution**
- **Discrete Morphology Models of Character Evolution.**
- **Running and Diagnosing an MCMC Analysis**

1 Overview of Chromosome Number Evolution Models

Chromosome changes represent major evolutionary mechanisms that have long been a focal point of study. Changes in chromosome number such as the gain or loss of a single chromosome (dysploidy), or the doubling of the entire genome (polyploidy), can have phenotypic consequences, affect the rates of recombination, and increase reproductive isolation among lineages and thus drive diversification (Stebbins 1971). Recently, evolutionary biologists have increasingly studied the macroevolutionary consequences of chromosome changes within a molecular phylogenetic framework, mostly utilizing the likelihood-based ChromEvol models of chromosome number evolution introduced by Mayrose et al. (2010). The ChromEvol models have permitted phylogenetic studies of ancient whole genome duplication events, rapid "catastrophic" chromosome speciation, major reevaluations of the evolution of angiosperms, and new insights into the fate of polyploid lineages (e.g. Pires and Hertweck 2008; Mayrose et al. 2011; Tank et al. 2015).

Here we describe the ChromEvol model as implemented in `RevBayes`, which except for one detail noted below is the same mathematical model introduced in Mayrose et al. (2010). In further sections, we will explore multiple extensions to this useful but basic model of chromosome number evolution.

1.1 The ChromEvol Model

In ChromEvol the evolution of chromosome number is represented as a continuous-time Markov process, similar to models of molecular evolution and discrete morphological evolution. The continuous-time Markov process is described by an instantaneous rate matrix Q where the value of each element represents the instantaneous rate of change within a lineage from a genome of i chromosomes to a genome of j chromosomes. For all elements of Q in which either $i = 0$ or $j = 0$ we define $Q_{ij} = 0$. For the off-diagonal elements $i \neq j$ with positive values of i and j, Q is determined by:

$$Q_{ij} = \begin{cases} \gamma_a & j = i + 1, \\ \delta_a & j = i - 1, \\ \rho_a & j = 2i, \\ \eta_a & j = 1.5i, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where γ_a, δ_a, ρ_a, and η_a are the rates of chromosome gains, losses, polyploidizations, and demi-polyploidizations.

If we are interested in modeling scenarios in which the probability of fusion or fission events are positively or negatively correlated with the number of chromosomes we can define Q as:

$$Q_{ij} = \begin{cases} \gamma_a e^{\gamma_m (i-1)} & j = i + 1, \\ \delta_a e^{\delta_m (i-1)} & j = i - 1, \\ \rho_a & j = 2i, \\ \eta_a & j = 1.5i, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

where γ_m and δ_m are rate modifiers of chromosome gain and loss, respectively, that allow the rates of chromosome gain and loss to depend on the current number of chromosomes. If the rate modifier $\gamma_m = 0$, then $\gamma_a e^{0(i-1)} = \gamma_a$. If the rate modifier $\gamma_m > 0$, then $\gamma_a e^{\gamma_m (i-1)} \geq \gamma_a$, and if $\gamma_m < 0$ then $\gamma_a e^{\gamma_m (i-1)} \leq \gamma_a$. Note that this parameterization differs slightly from the original ChromEvol model; here we assume the rates of chromosome change can vary exponentially as a function of the current chromosome number,

whereas ChromEvol as originally described by Mayrose et al. (2010) assumes a linear function. The theoretical reasons for this difference are described in Freyman and Höhna (2016), however in practice on most empirical datasets the difference appears insignificant.

Demi-polyploidization is the union of a reduced and an unreduced gametes that produces a cytotype with 1.5 times the number of chromosomes. The number of chromosomes in a genome must of course be an integer, so for odd values of i, Q_{ij} is set to $\eta/2$ for the two integer values of j resulting when $j = 1.5i$ is rounded up and down.

As in all continuous-time Markov models, the diagonal elements $i = j$ of Q are defined as:

$$Q_{ii} = -\sum_{i \neq j}^{C_m} Q_{ij}. \tag{3}$$

The probability of anagenetically evolving from chromosome number i to j along a branch of length t is then calculated by exponentiation of the instantaneous rate matrix:

$$P_{ij}(t) = e^{-Qt}. \tag{4}$$

Given a phylogeny and chromosome counts of the extant lineages, this model can be used in either a maximum likelihood or Bayesian inference framework to estimate the rates of chromosome change and the ancestral chromosome numbers.

1.2 Next Steps

The basic ChromEvol model as described above can be extended in a number of useful ways that will be covered in further sections. In the next section, however, we'll set up and run a simple `RevBayes` analysis using the ChromEvol model before moving on to the more complex models.

References

Freyman, W. A. and S. Höhna. 2016. Cladogenetic and anagenetic models of chromosome number evolution: a bayesian model averaging approach. bioRxiv Page 086629.

Mayrose, I., M. S. Barker, and S. P. Otto. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. Systematic Biology 59:132–144.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. Science 333:1257–1257.

Pires, J. C. and K. L. Hertweck. 2008. A renaissance of cytogenetics: Studies in polyploidy and chromosomal evolution. Annals of the Missouri Botanical Garden 95:275–281.

Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Edward Arnold Ltd., London.

Tank, D. C., J. M. Eastman, M. W. Pennell, P. S. Soltis, D. E. Soltis, C. E. Hinchliff, J. W. Brown, E. B. Sessa, and L. J. Harmon. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytologist 207:454–467.

Version dated: May 1, 2017