

APBC020

Sat DNA Analyzer, The First Computing Solution For Satellite-DNA Evolutionary Analysis

Navajas-Pérez R¹, Ruiz Rejón M¹, Garridos Ramos MA¹, Aznarte JL², Rubio Escudero, C²

¹Departamento de Genética, and ² Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Campus de Fuentenueva s/n, 18071, Granada, SPAIN

Repetitive DNA families are influenced by several molecular mechanisms of non-reciprocal exchanges that can gradually spread a variant sequence throughout a family within a sexual population in a process called molecular drive. It should explain the evolutionary pattern of repetitive sequences known as concerted evolution, which leads to high levels of family homogeneity for species-diagnostic mutations.

Assuming that molecular drive is a time-dependent process, then the expected stages of transition during the spread of a variant repeat unit toward its fixation can be defined according to the model of Strachan et al. (1985). This method analyses the patterns of variation at each nucleotide position considered independently, comparing repeats of a repetitive family belonging to a pair of species. The classification includes six stages (Classes 1-6) in the spread of variant repeats through the family and the species. In brief, the Class 1 site represents complete homogeneity across all repeat units sampled from a pair of species, whereas Classes 2, 3 and 4 represent intermediate stages in which one of the species shows a polymorphism. The frequency of the new nucleotide variant at the site considered is low in Class 2 and intermediate in Class 3, while Class 4 represents sites in which a mutation has replaced the progenitor base in most members of the repetitive family in the other species. Class 5 represents diagnostic sites in which a new variant is fully homogenized and fixed in all the members of one of the species while the other species retains the progenitor nucleotide. A Class 6 site represents an additional step over stage 5 (new variants appear in some of the members of the repetitive family at a site fully divergent between the two species).

Up to date, non-specific software has been developed in order to analyze this type of sequences and the data processing becomes a time-demanding and tedious process. We here try to fill this gap. SatDNA Analyzer is a program, implemented in C++, devoted to the classification of Strachan Transition Stages in samples of satellite DNA amongst all units of a given satellite-DNA family when comparing it between a pair of species. Furthermore, this program implements several other utilities for satellite-DNA analysis evolution such as the design of the average consensus sequences, the average base pair contents, the distribution of variant sites, the transition to transversion ratio and different estimates of intra-specific variation and inter-specific variation. Aprioristic hypotheses on factors influencing the molecular drive process and the rates and biases of concerted evolution can be tested with this program. SatDNA Analyzer has been designed to operate on Windows and Linux, and is freely available at <http://satdna.sourceforge.net/>