

# SatDNA Analyzer, the first computing solution for satellite-DNA evolutionary analysis

Rafael Navajas-Pérez<sup>1</sup>, Cristina Rubio-Escudero<sup>2</sup>, José Luis Aznarte<sup>2</sup>, Manuel Ruiz Rejón<sup>1</sup>  
and Manuel A. Garrido-Ramos<sup>1</sup>

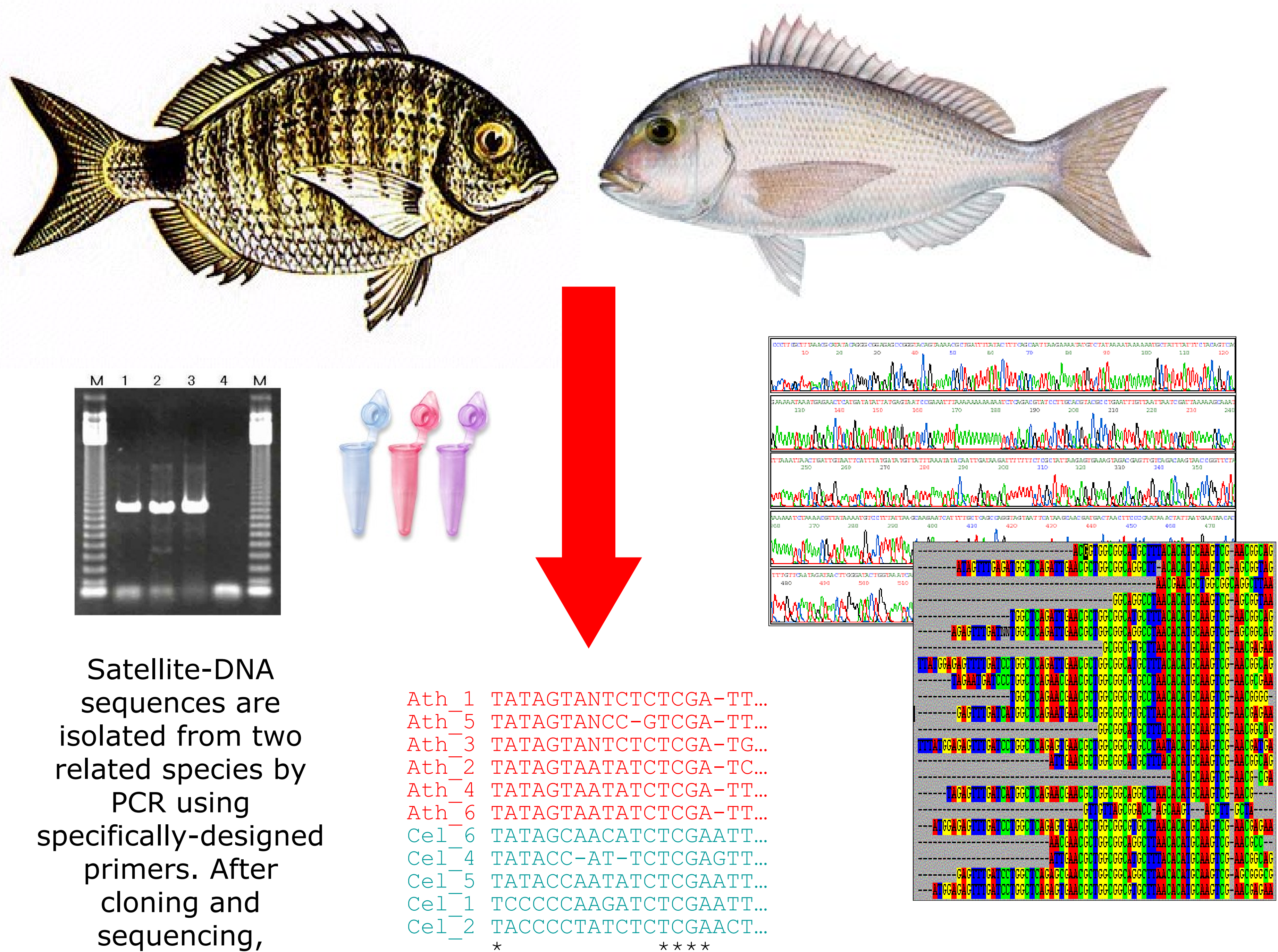
<sup>1</sup>Departamento de Genética, Facultad de Ciencias, Universidad de Granada, 18071, Granada, SPAIN

<sup>2</sup>Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada, SPAIN



satDNA Analyzer (satellite-DNA Analyzer), is a software package for the analysis of satellite-DNA sequences from aligned DNA sequence data. It allows for the analysis of the patterns of variation at each nucleotide position considered independently amongst all units of a given satellite-DNA family when comparing two species. The program classifies each site accordingly as monomorphic or polymorphic, discriminates shared from non-shared polymorphisms and classifies each non-shared polymorphism according to the model proposed by Strachan et al. (1985) in six different stages of transition during the spread of a variant repeat unit toward its fixation. Unfortunately, by this date, such calculations are manually performed in an arduous and time-consuming manner. **Trying to fill this gap, we have designed satDNA Analyzer 1.2.** This software supposes a time saving feature since every utility is automatized and collected in a single software package, so the user does not need to use different programs. Additionally, it significantly avoids data miscalculations due to human errors, very prone to occur specially in large files.

Make it easy, use satDNA Analyzer!!



Satellite DNAs are tandemly arrayed highly repetitive DNA sequences of the eukaryotic genomes located in the constitutive heterochromatin. The repeats comprising a satellite-DNA family do not evolve independently of one another but rather follow concerted evolution. Basically, it states that once a mutation arises it spreads through the rest of repeats of one species toward the fixation of the new variant, being the rate of fixation higher than the rate of mutation. This leads to high levels of intraspecific homogeneity and interspecific divergence. Assuming molecular clock, Strachan et al. (1985) designed a method for concerted evolution quantification. This method compares position to position two sets of aligned sequences belonging to two different species. It allows classifying any position in terms of six stages. 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).

satDNA Analyzer v1.2 - Universidad de Granada

Summary for species S and species J

Strachan Type I	490
Strachan Type II (from 0% to 25%)	210
Strachan Type III (from 25% to 50%)	22
Strachan Type IV (from 50% to 75%)	28
Strachan Type V	48
Strachan Type VI	20
Shared Polymorphism	91
Shared Polymorphisms due to indels	2
Irrelevant Positions	36
Irrelevant Positions due to indels	134
Non-shared Polymorphism	498
TOTAL	1011

Legend: Strachan Type I, Type II, Type III, Type IV, Type V, Type VI;  
Shared polymorphism, Shared polymorphism due to indels, Irrelevant position;

SatDNA Analyzer reads most common DNA sequence alignments formats

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 and inter-specific variation that make it a powerful tool for satellite-DNA evolutionary analysis.

Screenshots of satDNA Analyzer html output (left) Alignment showing in colors the different types of positions the software recognizes. (right) Example of a table display for statistical calculations.

## References:

Strachan, T., Webb, D. and Dover, GA. (1985). Transition stages of molecular drive in multiple-copy DNA families in *Drosophila* EMBO J. 1985 July; 4(7): 1701-1708.

Navajas-Pérez, R., Rubio-Escudero, C., Aznarte, J.L., Ruiz Rejón, M. & Garrido-Ramos, MA. (in press) SatDNA Analyzer: a computing tool for satellite-DNA evolutionary analysis. Bioinformatics.

APBC 2007

The Fifth Asia Pacific Bioinformatics Conference  
Hong Kong, 15 - 17 January 2007