

# Comparative evolutionary analysis of chloroplast genomes

Emiliani G<sup>(1)</sup>, Fondi M<sup>(2)</sup>, Papaleo MC<sup>(2)</sup>, Mengoni A<sup>(2)</sup>, Fani R<sup>(2)</sup>

<sup>(1)</sup> Department of Environmental and Forestry Sciences, University of Florence, Florence

<sup>(2)</sup> Department of Evolutionary Biology, University of Florence, Florence

## Motivation

The chloroplast is the essential organelle in plants that, performing photosynthesis, is required for the photoautotrophic plant growth. It is generally accepted that chloroplasts have arisen from an internalized cyanobacterial ancestor through at least two (primary and secondary endosymbioses) events. Chloroplasts have maintained an independent genome that codes for an important fraction of the proteins required for the photosynthetic activity. The chloroplast genome (cpDNA) consists of circular double stranded DNA molecules of 110–200 kb size, containing RNA genes and a number of protein coding genes, which ranges from about 100 in plants and green algae to 150–200 in non-green algae. The protein-coding genes can be classified into two groups:

- i) genes involved in the chloroplast expression and translation machinery, and
- ii) genes related to photosynthesis.

The chloroplast genome of most plants harbors two large inverted repeats (IRs) of 6–76 kb that divide the cpDNA in one large and one small single-copy region (LSC and SSC, respectively). The non-recombinant, uniparentally inherited nature of organelle genomes makes them potentially useful tools for evolutionary studies. However, detecting useful polymorphism is often difficult due to the slow mutation rates of chloroplast genes and the attempts to reconstruct plastid evolution using traditional molecular approaches (sequence-based analyses of genes) have proven particularly difficult. Furthermore, it is now clear that many of the proteins needed for plastid functions, including those involved in photosynthesis, are now encoded by genes located in the nucleus where they very likely arrived during evolution by Endosymbiotic Gene Transfer (EGT) from the cyanobacterial genome. The availability of fully sequenced chloroplast DNA and complete genomes opens up the possibility to perform a phylogenetic reconstruction and to shed some light on the genome evolutionary patterns, an issue that requires the developing for innovative bioinformatic tools for visualization and comparative analysis. Therefore, the aim of this work was to analyze the available completely sequenced chloroplast genomes in order to:

- i) carry out a comparative evolutionary analysis of chloroplasts,
- ii) a comparative analysis of chloroplast vs a cyanobacterial genome, and
- iii) an analysis of imprints of chloroplast-nuclear genome trade-off.

## Methods

All the 136 available chloroplast genomes from land plants, green and non-green algae and protists (i. e. the chromatophore form *Paulinella chromatophora*), the genome of *Nostoc punctiforme* and the genome of the basal land plants *Physcomyrella patens* were retrieved from GB. A recently developed bioinformatic tool, Blast2Network, that permits:

- 1) immediate visualization/estimation of evolutionary relationships in a large dataset of amino acid and nucleotide sequences, and
- 2) the construction of phylogenetic profiles resembling genomes similarity (in terms of shared genes) and proteins co-occurrence, was used to perform the analysis.

## Results

Despite the high synteny exhibited by the different chloroplast genomes, the comparative analysis revealed interesting discontinuities of similarity along the generally accepted Viridiplantae phylogeny. Whilst the chloroplasts of some basal land plants (i. e. *Psilotum nudum*, *Anthoceros formosae*, *Amborella thricopoda*) confirm their ancestry to the Angiosperm clade, the chloroplast of *P. patens* did not show a high degree of sequence similarity (at least at high sequence similarity threshold value) either with green algae or with angiosperm chloroplasts. The discrepancies between species phylogeny and chloroplasts genomes similarity was also confirmed by phylogenetic profiling of protein sharing. Particularly interesting is the absence of high degree of similarity between the available gymnosperm chloroplast and the other tracheophyte, marking a putative incongruence in the phylogenetic position of this group. The visualisation power of Blast2Network enabled also to immediately compare the distribution and the dimen-

sion of paralogous gene regions within the same chloroplast genome; large IRs, containing protein coding genes, are present in some green and red algae and diatoms (possibly for recent recombination events), while are largely absent in land plants. The construction of high similarity protein cluster enables also the identification of chloroplast proteins conserved “cores” as well as the annotation of previously undescribed proteins. Preliminary comparative analysis of chloroplast and *N. punctiforme* chromosome suggests a unequal contribution of the bacterial genome to different plastids, showing possibly the imprints of primary and secondary endosymbioses. The proposed bioinformatic approach allowed also to individuate potential traces of EGT events between *P. patens* nuclear and chloroplast genome.

**Contact :** giovanni.emiliani@unifi.it