Stable Chloroplast: Myth or Reality?

Shailesh Joshi[1,2] and Dyfed Evans[1,2]

Email: Shailesh.Joshi@sugar.org.za or
Dyfed.Evans@sugar.org.za

[1] South African Sugarcane Research Institute,
Mount Edgecombe, Durban, 4300

[2] School of Life Sciences, University of
KwaZulu-Natal, Durban, 4000

Introduction

Chloroplasts principally encode the photosynthetic machinery in Viridiplantae. It has long been accepted that in photosynthetic plants chloroplast genomic structure is uniquely stable as it is maternally and clonally inherited. The first chloroplast genomes sequenced supported this view, with the only major structural rearrangements being identified between the *Poales* and the *Poaceae*. Due to its purported stability the chloroplast genome is routinely used both for evolutionary analysis and for clock calibration. Recently, additional chloroplast genomes across all plant lineages have been sequenced and they revealed, for the first time, major structural changes between plant lineages. This will have profound implications on any studies performed based on the chloroplast genome. The current study was undertaken to address the potential issue of global chloroplast (in)stability.

Materials and Methods

Chloroplast genomes representing **220** plant species across the angiosperms were downloaded from GenBank. The sequences were colinearized using a combination of **MUMmer**, **BLAST** and **MAUVE**. Rearrangements were recorded. The rearranged sequences were initially aligned with **MAFFT** and a first pass phylogeny was created with **PhyML**. This tree was used as input for **SATe**, with a final phylogeny and alignment generated from **30** iterations of this application. The final alignment was improved using a combination of manual curation along with sub-alignment optimization using **MUSCLE**. Repeat regions (di, tri, tetra, penta, palindromic, inverted, microsatellite, tandem, introgressed) were identified using components of the **EMBOSS** package. Unique regions within the chloroplast genomes were identified using **nucmer** with the closest evolutionary precursor (based on **SATE** alignment) employed as a reference.
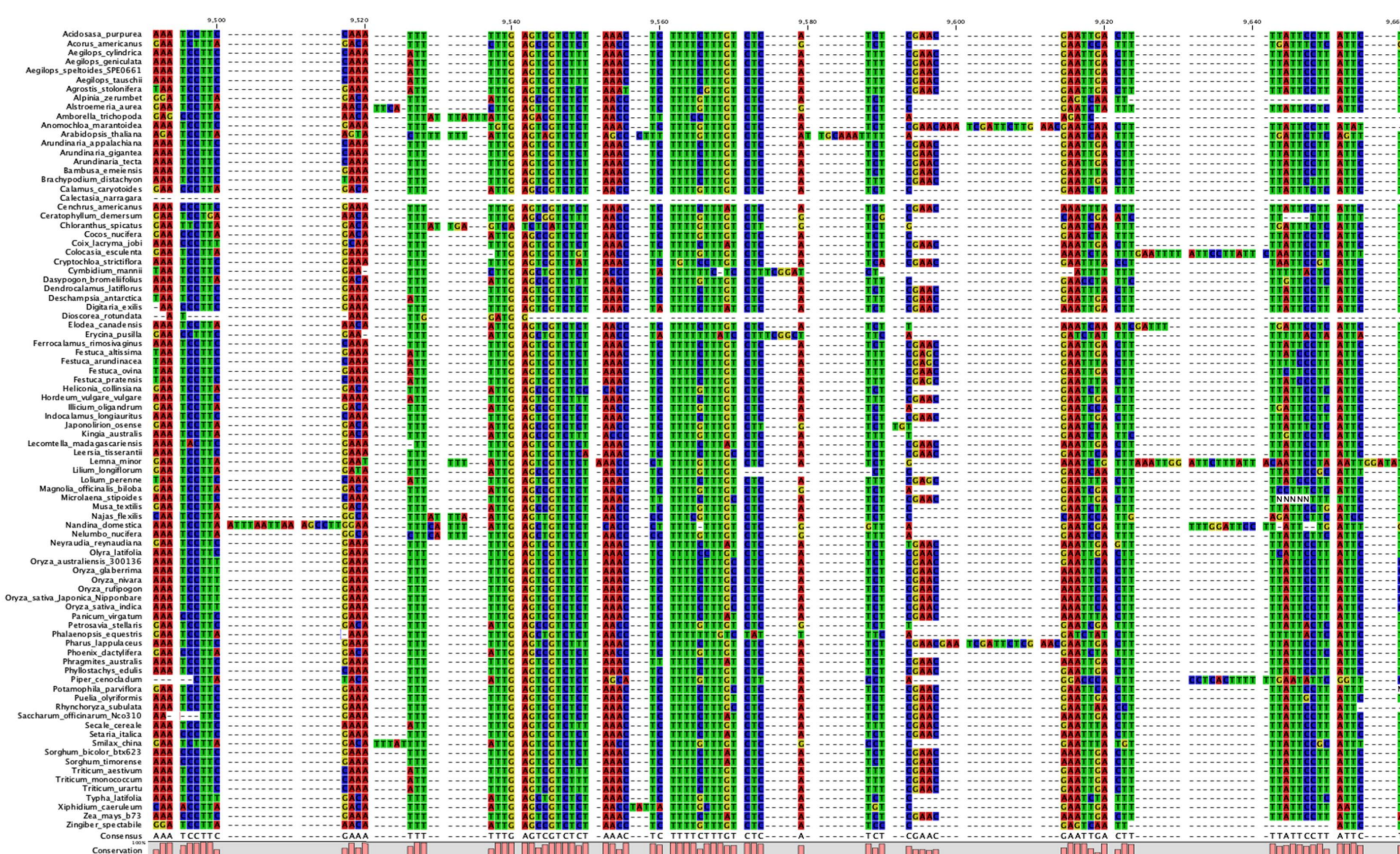

Figure 2: Portion of the whole chloroplast alignment.

Figure 1: Large scale chloroplast rearrangements.

LSC Region
IR_B Region
SSC Region
IR_A Region
Inversion
Single Gene Inversion

Results and Discussion

Our novel multi-genome based bioinformatics approach (colinearization of chloroplast genomes, iterative alignment with **SATE**, global repeat discovery and novel sequence quantification) allows for accurate whole chloroplast alignments revealing changes at the macro, gene and micro scales. The current study includes 220 chloroplast genomes covering 340MY of plant evolution and includes both photosynthetic and non-photosynthetic seed plants.

At macro levels, when plants become non-photosynthetic, we observed massive gene loss and trancuation in the chloroplast genome (eg. *Petrosavia, Rhizanthella* and *Epifagus*) **(Fig 1)**.

Results and Discussion   Contd….

We further described the complete loss of the IRa region from chloroplasts in the lentil lineage (*Lens*). Two recently sequenced eudicot chloroplasts (*Pelargonium* and *Trifolium*) also revealed large scale genomic rearrangements (Fig 1).

At the gene level there have been specific gene losses that characterise new plant lineages as noted in the five events (two gene losses and three rearrangements) separating early dicots from the *Poaceae*. This phenomenon was also noted in orchid (*Calectasia*) and *Dioscoreal* lineages **(Fig 2)**.

Our results at the micro level demonstrate that short nucleotide repeats (10-20nt) are considerably less common than the large scale repeats (50-300nt), which are distributed throughout the genome of all lineages. However, the duplication in each event is lineage specific and does not reoccur in other genera (Fig 2).

Quantification of the unique sequences demonstrates the presence of 2.5-7.5% novel sequence in every chloroplast genome examined **(Fig 3)**. The presence of a basal level of novel and repeat sequences points towards a constant chromosomal churn. In conjunction with the large scale rearrangements observed these data contradict the existing perception of a stable chloroplast.
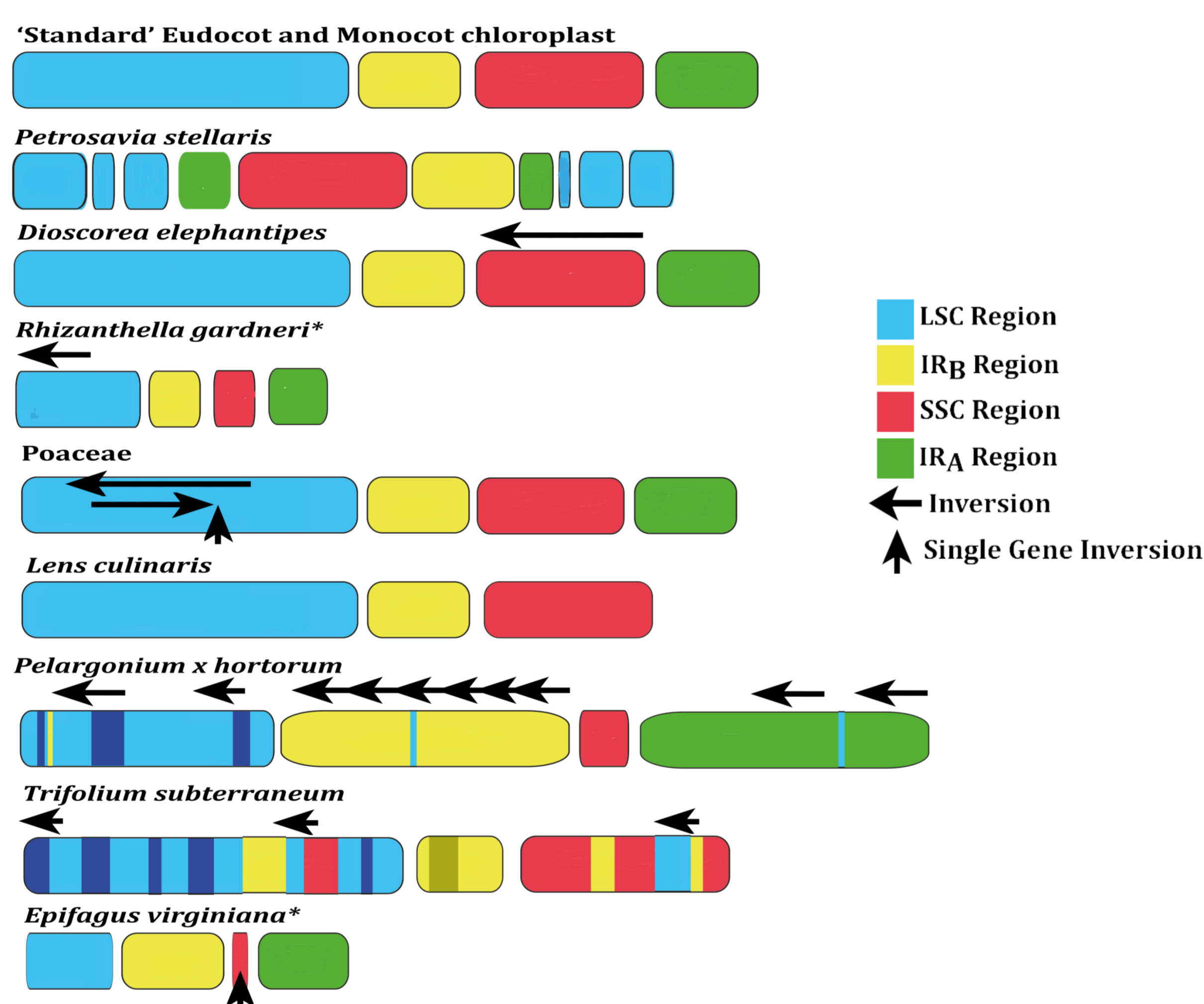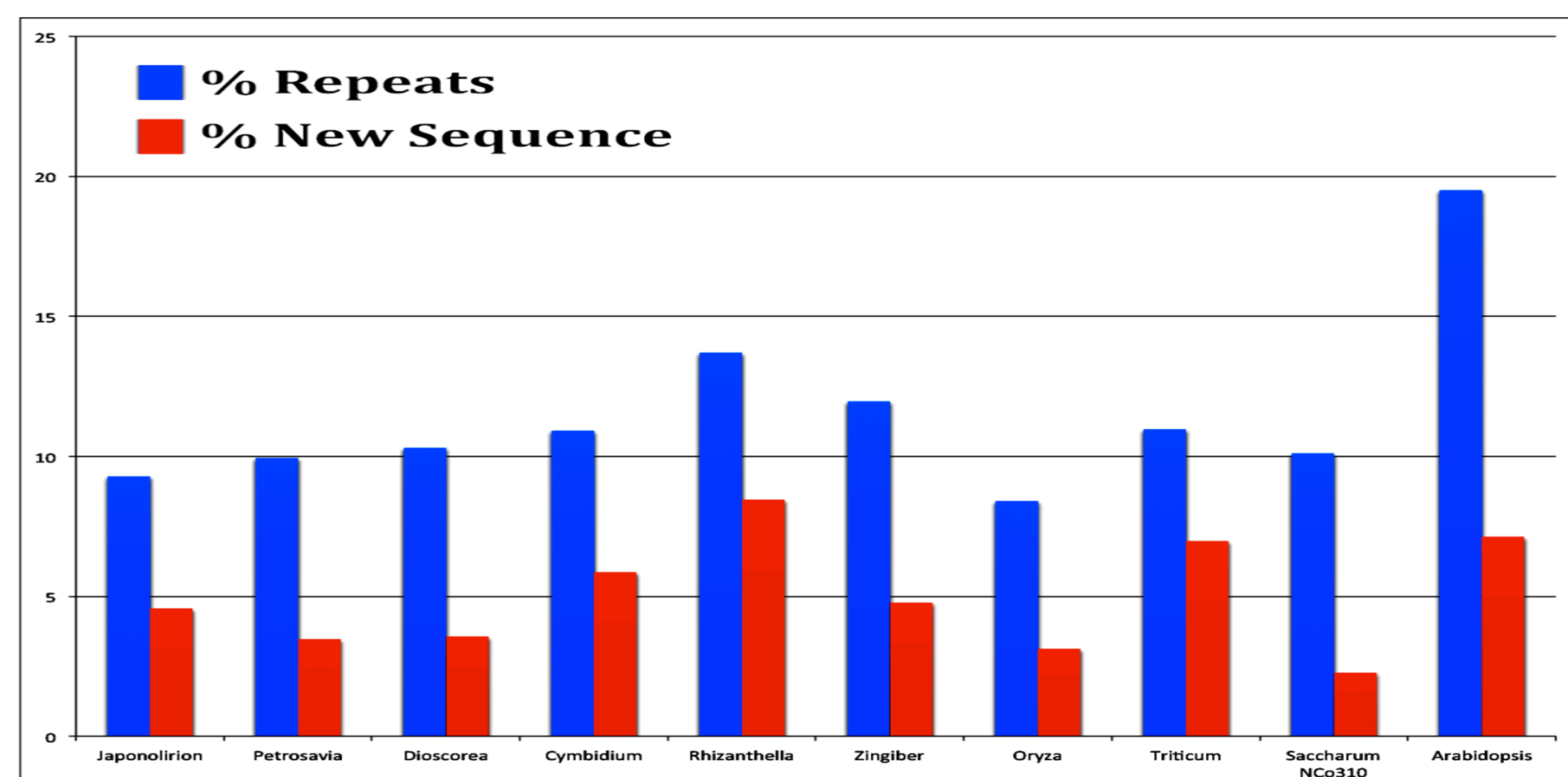

Figure 3: Repeats and novel sequence as percent of total chloroplast genome.

Conclusion

This large scale, multigenome, study allows us to demonstrate that the chloroplast is a dynamic genome, rather than being static and stable as perceived in the traditional, erroneous view. The chloroplast is derived from an ancestral cyanobacterial genome and still retains it plasticity. This means that lineages can have profoundly different clock rates of plastid evolution. Here, we demonstrate alterations in every examined chloroplast genome, at all scales, proving that the stable chloroplast is a myth.