Is plant mitochondrial RNA editing a source of phylogenetic incongruence? An answer from in silico and in vivo data sets

ID - 247

Picardi Ernesto[1], Quagliariello Carla[1]

[1]Dipartimento di Biologia Cellulare, Università della Calabria, 87036 Arcavacata di Rende, Italy

Motivation

In plant mitochondria a post-transcriptional RNA editing process generates at a specific site C-to-U conversions, usually resulting in the restoration of phylogenetically conserved codons and amino acid residues. The RNA editing process seems, thus, to function as a repair mechanism to correct otherwise deleterious genomic mutations and may have additional significance in regulation of gene and protein function. This process as an additional step in plant mitochondrial RNA processing poses a number of intriguing unsolved evolutionary questions. Sites undergoing RNA editing evolve, indeed, at much higher rate than sites not subjected to such a modification. As a result, the diversity between plant species as to whether a given gene editable site requires editing or whether the edited version, a T, is already genomically-encoded is thought to strongly affect the evolution of plant mitochondrial genomes, representing an important source of sequence variability and potentially informative characters. To date no clear and convincing evidence has established whether or not editing sites of plant mitochondrial genes represent a potential source of phylogenetic incongruence and how much does RNA editing really affects the topology of reconstructed phylogenetic trees. Here by means of computer simulation, the effect of RNA editing on tree reconstruction process has been investigated for the first time on in silico generated sequences. In parallel, 20 different genuine plant mitochondrial gene sequences have been analyzed by the same original approach.

Methods

Genomic and cDNA like sequences have been simulated by the program EdiPy with an editing percentage ranging from 1 to 10 and along three different tree topologies of 12, 18 and 24 taxa with average branch lengths deduced from plant mitochondrial genes via maximum likelihood (ML). Each set of genomic and cDNA sequences has been submitted to PHYML program to estimate the ML phylogenetic tree. Differences between inferred and starting trees have been quantified by the topological distance using the Treedist program of the PHYLIP package. Tree topology accuracy values have been calculated as the proportion of correctly inferred topologies over the total number of detected trees. Moreover, plant mitochondrial genes and edited cDNA have been downloaded from a REDIdb database that we previously developed, taking into account only genes with at least seven sequences. For each gene and cDNA alignment the PHYML program has been used to reconstruct genomic and cDNA trees. The comparison between genomic and cDNA inferred trees has been performed by a new metric, called ratioDt. This ratio, defined here as the ratio between the truly detected topological distance and the maximum value that it could assume, ranges from 0 to 1. The ratio Dt approaches to 0 for identical trees and increases as the match worsens.

Results

Our results from simulated data sets indicate that editing `noise` in tree topology inference is mainly manifested at the cDNA level. In particular, editing sites tends to confound tree topologies when artificial genomic and cDNA sequences are generated shorter than 500 nt and with an editing percentage higher than 5.0%. In case of genuine plant mitochondrial genes, the ratioDt increases when the editing percentage goes up from about 3.0 to 14.0%. Moreover, when the average gene length is higher than 1,000 nt (rps3, matR and atp1) no differences can be detected in the comparison between inferred genomic and cDNA topologies. On the contrary, genes shorter than 1,000 nt shows heterogeneous ratioDt values depending on the editing percentages and the total number of variable characters. On the whole, in silico and in vivo results undoubtedly suggests that editing sites can contribute in generating misleading phylogenetic trees when mitochondrial genes are short, highly edited and the number of analyzed sequences increases. Since there is no evidence up to now that mitochondrial DNA sequences are misleading in phylogenetic analyses, our findings encourage the use of genomic mitochondrial rather than cDNA sequences for reconstructing phylogenetic events in land plants.

Email: e.picardi@unical.it