Molecular Phylogenetics and Evolution 48 (2008) 313–325

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Reticulate or tree-like chloroplast DNA evolution in *Sileneae* (Caryophyllaceae)?

Per Erixon *, Bengt Oxelman

Department of Systematic Botany Evolutionary Biology Centre, Uppsala University, Norbyv. 18D, SE-75236 Uppsala, Sweden
Department of Plant and Environmental Sciences, Göteborg University, Box 461, SE-40530 Göteborg, Sweden

ARTICLE INFO

Article history:
Received 7 December 2007
Revised 4 April 2008
Accepted 7 April 2008
Available online 14 April 2008

Keywords:
Sileneae
cpDNA
Chloroplast recombination
Incongruence
Indel characters

ABSTRACT

Despite sampling of up to 25 kb of chloroplast DNA sequence from 24 species in *Sileneae* a number of nodes in the phylogeny remain poorly supported and it is not expected that additional sequence sampling will converge to a reliable phylogenetic hypothesis in these parts of the tree. The main reason for this is probably a combination of rapid radiation and substitution rate heterogeneity. Poor resolution among closely related species are often explained by low levels of variation in chloroplast data, but the problem with our data appear to be high levels of homoplasy. Tree-like cpDNA evolution cannot be rejected, but apparent incongruent patterns between different regions are evaluated with the possibility of ancient interspecific chloroplast recombination as explanatory model. However, several major phylogenetic relationships, previously not recognized, are confidently resolved, e.g. the grouping of the two SW Anatolian taxa *S. cryptoneura* and *S. sordida* strongly disagrees with previous studies on nuclear DNA sequence data, and indicate a possible case of homoploid hybrid origin. The closely related *S. atocioides* and *S. aegyptiaca* form a sister group to *Lychnis* and the rest of *Silene*, thus suggesting that *Silene* may be paraphyletic, despite recent revisions based on molecular data.

1. Introduction

The concept of the chloroplast genome as a single evolving unit is fundamental to plant systematics. This assumption rests on the fact that there are very few documented cases of recombination among chloroplast genomes of flowering plants (Medgyesy et al., 1985; Houliston and Olson, 2006), no published evidence for lateral transfer (LGT) of chloroplast genes in land plants (see Archibald et al., 2003 for an example in chlorarachniophytes, and Rice and Palmer, 2006 for a cryptophyte/haptophyte example), and that chloroplast DNA (cpDNA) is predominantly uniparentally inherited (Wolfe and Randle, 2004). Therefore, DNA sequences from different parts of the chloroplast genome are assumed to have evolved on a single tree topology, even if the species containing those sequences have a history of hybridization. Sequence data from the nuclear genome are very different in this respect, because hybridization can cause nuclear genomes to merge and potentially recombine.

Early studies on plant phylogeny utilizing chloroplast DNA sequences were typically based on a single gene (e.g. *rbcL*, Chase et al., 1993). Biotechnological development has facilitated inclusion of more genes to increase resolution and support. For example, the early evolutionary diversification of flowering plants has been addressed by concatenation of a large number of chloroplast genes. Graham and Olmstead (2000) based their phylogenetic study on basal angiosperm relationships on 17 genes yielding 13,806 DNA characters and Goremykin et al. (2003) made a study on less taxa from the same group on sequences from 13 completely sequenced chloroplast genomes, based on 30,017 DNA characters from protein-coding genes. However, if taxon sampling is sparse, and the branches long, phylogenetic methods can be inconsistent (Felsenstein, 1978), leading to erroneous conclusions (Soltis et al., 2004; Qiu et al., 2005; Rydin and Källersjö, 2002; Stefanovic et al., 2004).

Closely related taxa should be less problematic, because the branches of the phylogenetic tree are shorter and therefore less sensitive to inconsistencies of the phylogenetic model (Felsenstein, 1978; Philippe and Laurent, 1998; Bergsten, 2005). Thus, it might be expected that inferring phylogenies of closely related taxa should be easy, given that enough information is at hand. Ideally we would like to have branches with so few substitutions per site that the probability of multiple substitutions is close to zero, and at the same time sequences that are long enough to contain information for every branch.

In chloroplast phylogenetic studies, protein-coding genes have often been dismissed for low-level relationships, presumably because of conservative negative selection that is expected to be associated with these. Instead, systematists have often used non-coding regions such as intergenic spacers and introns for shallow phylogenetic problems (e.g. Shaw et al., 2005, 2007).

* Corresponding author. Fax: +46 18 4716457.
E-mail address: per.erixon@ebc.uu.se (P. Erixon).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.04.015

With decreased sequencing costs and increased amounts of available cpDNA sequence data, lack of information because of low sequence variability becomes less of a problem. Low-level systematics and relatively shallow phylogenetic problems could potentially benefit from the simplicity of cpDNA data, because biparentally inherited sequences and recombination are more problematic for closely related taxa (Álvarez and Wendel, 2003; Poke et al., 2006). Analyses based on cpDNA sequences do indeed often exhibit little homoplasy (e.g. Archibald et al., 2005; Popp and Oxelman 2001, 2004; Popp et al., 2005; Frajman and Oxelman, 2007) relative to nrDNA.

Because of the above-mentioned properties of the chloroplast genome, different DNA sequence regions are often concatenated without evaluation of possible incongruence (e.g. Bremer et al., 2002; Nishiyama et al., 2004). If incongruences among different regions from the chloroplast are found, they will be interpreted as artifacts, rather than the result of conflicting evolutionary histories (e.g. Graham and Olmstead, 2000). Hard incongruences (mutually well supported conflicting phylogenies) between chloroplast regions from closely related taxa are not known to us.

Given the assumption that the chloroplast genome has evolved according to a single tree topology, poorly resolved phylogenetic trees should be the result of either poor fit of the model used for phylogenetic inference, or lack of information due to limited number of characters. If, on the other hand, the tree model is invalid, for example due to recombination, this could result in poor phylogenetic resolution.

The tribe *Sileneae* (Caryophyllaceae), a phylogenetically relatively shallow group in the angiosperm tree, includes several taxa that have properties of interest to, e.g. breeding system (e.g. Desfeux et al., 1996; Andersson-Ceplitis, 2002; McCauley et al., 2005), evolution of sex chromosomes (e.g. Vyskot and Hobza, 2004; Filatov, 2005), pollination (e.g. Kephart et al., 2006), host–parasite interactions (e.g. Hood and Antonovics, 2000), aberrant evolution of mitochondrial genes (e.g. Städler and Delph, 2002; Mower et al., 2007; Sloan et al., 2008), and strong positive selection in a chloroplast gene (Erixon and Oxelman, 2008). The group has been studied phylogenetically based on sequences from the nuclear ribosomal internal transcribed spacers (ITS) region (Oxelman and Lidén, 1995; Desfeux et al., 1996; Oxelman et al., 2001; Burleigh and Holtsford, 2003; Eggens et al., 2007; Frajman and Oxelman, 2007) and parts of four nuclear low-copy number RNA polymerase genes (Popp and Oxelman, 2004, 2007; Popp et al., 2005; Eggens et al., 2007). Hitherto, the only chloroplast regions used for inferring relationships in *Sileneae* have been the *rps16* intron (Oxelman et al., 1997; Popp and Oxelman, 2004), the *psbE-petL* spacer (Popp et al., 2005) and the *trnL/F* intron/spacer region (Burleigh and Holtsford, 2003). The data from the previous chloroplast DNA studies are not possible to combine because of different taxon sampling, but Popp and Oxelman (2004) presented a combined analysis of data from the *rps16* intron, ITS and RNA polymerase genes, resulting in a highly resolved and supported tree. However, a few nodes were not well supported, and some recent data suggest strong disagreement between nuclear and plastid regions for particular groups (e.g. Frajman and Oxelman, 2007), and also between different nuclear regions (Frajman et al., unpublished) for some taxa. The sampling of Popp and Oxelman (2004) included only 29 of the c. 700 species in the tribe, and it is possible that there actually are more hard incongruences among the individual genes, that are undetected because of lack of information (i.e. the null hypothesis of a single tree topology could not be rejected). Many groups in *Sileneae* have a complex evolutionary history with several cases of allopolyploidization and hybridization (Oxelman, 1996; Popp and Oxelman, 2001; Popp et al., 2005; Frajman and Oxelman, 2007). In light of the more prevalent degree of recombination in the nuclear genome, we anticipate that a "backbone" of well-supported chloroplast DNA phylogenetic relationships could potentially provide a useful framework onto which nuclear phylogenies could be anchored.

We use roughly 25 kb of DNA sequence from the large single copy region (LSC) in the chloroplast of 24 species in *Sileneae*. As previous cpDNA phylogenies based on short regions (e.g. *rps16*, Oxelman et al., 1997) indicate that the lack of resolution in some parts of the tree is due to lack of information rather than homoplasy, we wanted to collect large amounts of data for relatively few taxa to test this. Our approach is to sequence long contiguous regions, including both coding and non-coding regions, rather than to try to find many "super-informative" short regions. Shaw et al. (2005) evaluated 21 non-coding cpDNA regions and concluded that regions differ in variability. They did not find any extremely variable regions and their results also showed large differences between different taxonomic groups.

In this paper we will (1) explore if the major phylogenetic relationships in *Sileneae* become well resolved in a tree-like fashion by 25 kb of cpDNA sequence, (2) investigate the effect of incrementally adding data, and (3) explore potential causes of weak tree resolution.

2. Materials and methods

2.1. Plant material

An attempt was made to represent as much of the phylogenetic range of variation as possible within the tribe *Sileneae*. All genera (as recognized by Oxelman et al., 2001) in *Sileneae* (*Agrostemma* L., *Atocion* Adans., *Eudianthe* (Rchb.) Rchb., *Heliosperma* (=*Ixoca*) (Rchb.) Rchb., *Lychnis* L., *Petrocoptis* A. Br. ex Endl., *Silene* L., and *Viscaria* Bernh.) including several representatives of the two approximately equally sized subgenera in *Silene* (*Behen* (Moench) Bunge and *Silene*). The three species in *Silene* subgenus *Silene* was chosen to represent the three major clades in this group based on previous studies (Oxelman and Lidén, 1995; Oxelman et al., 1997; Eggens et al., 2007). The resolution of subgenus *Behen* on the basis of previous molecular studies is poor (Oxelman et al., 1997; Popp and Oxelman, 2004). Consequently, we sampled more taxa from this group (eight taxa), mainly based on morphological variation and previous taxonomic work (Chowdhuri, 1957). *Silene sordida* and *S. cryptoneura* were included because they do not fall in either of the two subgenera and there are incongruent patterns between nuclear ribosomal ITS data (Oxelman and Lidén, 1995) and chloroplast data (Oxelman et al., 1997). *Silene aegyptiaca* and *S. atocioides* are morphologically very similar, but preliminary unpublished results showed surprisingly high sequence divergence and the two taxa did not group with *S. cryptoneura*, which could have been expected from morphological and phytogeographical patterns, as well as previous classifications (Coode and Cullen, 1967; Chowdhuri, 1957). Plant materials used in this study are presented with voucher data and GenBank/EMBL accession numbers in Table 1.

2.2. PCR and sequencing

The choice of the three largest sequence regions used in this study was guided by the identification of large contiguous regions with high proportions of non-coding sequences in the spinach genome (GenBank Accession No. NC_002202). The largest region (>18 kb; size measures in this section refer to the spinach genome) between *rbcL* and *petB* (Fig. 1, region 4) was initially divided into four subregions (4.3–4.9 kb), by constructing four primer pairs (*rbcL*-F2/*cemA*-R, *cemA*-F/*petG*-R, *petL*-F2/*clpP*-R, *clpP*-F/