Phytotaxa 1: 3–20 (2009)
www.mapress.com/phytotaxa/

Article

ISSN 1179-3155 (print edition)
PHYTOTAXA
ISSN 1179-3163 (online edition)

Phylogenetic analysis of non-coding plastid DNA in the presence of short inversions

SAMULI LEHTONEN[1*], LEENA MYLLYS[2] & SANNA HUTTUNEN[1]

[1] *Department of Biology, FI-20014 University of Turku, Finland*
[2] *Botanical Museum, Finnish Museum of Natural History, PO Box 7, FI-00014 University of Helsinki, Finland*
**Corresponding author; samile@utu.fi*

Abstract

The presence of short inversions in non-coding plastid DNA is universal and an integral part of its evolution. We studied the effect of these structural changes in phylogenetic inference by measuring phylogenetic accuracy as a topological congruence among topologies inferred from plastid and nuclear sequences in ferns (*Lindsaea*) and mosses (Brachytheciaceae). We randomly replicated ten subsets of both groups and tested the performance of various coding schemes and phylogenetic methods. Our results demonstrate that short inversions do interfere with phylogenetic analysis, leading into incongruent topologies. The phylogenetic accuracy was increased by removing inversions or applying various coding schemes to them. Parsimony analyses resulted in greater accuracy than maximum likelihood or Bayesian analyses, but this was at least partly explained by slightly poorer resolution of parsimony trees. Due to the great importance of non-coding plastid DNA in species-level plant systematics and the frequency of structural rearrangements in it, we recommend these kinds of data be carefully explored before analysis, and the possible rearrangements to be optimised.

Key words: Brachytheciaceae, congruence, ferns, hairpin, ITS, *Lindsaea*, mosses, *rpoC1*, SPR-distance, secondary structure, *trnD-trnT, trnH-psbA, trnL-trnF*

Introduction

Non-coding plastid DNA sequences are widely and increasingly used in lower level plant systematic studies (Shaw *et al.* 2005). It has been noticed that evolutionary processes of these sequences often result in a formation of peculiar structural patterns, which are most parsimoniously explained as small-scale rearrangements (e.g. Kelchner 2002). Most notably, minute inversions (less than 50 bp in length) are common and widespread in the plant plastid genomes (Kim & Lee 2005), and they have been reported from a variety of plants ranging from angiosperms (e.g. Kelchner & Wendel 1996, Sang *et al.* 1997, Graham *et al.* 2000, Kim & Lee 2005, Bain & Jansen 2006, Storchova & Olson 2007, Catalano *et al.* 2008), pteridophytes (Lehtonen & Tuomisto 2007) and bryophytes (Quandt *et al.* 2003, Huttunen & Ignatov 2004, Hernandez-Maqueda *et al.* 2008). In addition to non-coding cpDNA regions, minute inversions occur in mitochondrial genomes of plants (Dumolin-Lapègue *et al.* 1998) and are also widespread in animals (Blakeley *et al.* 2006, Chaisson *et al.* 2006).

This type of short inversions is easily overlooked and treated as multiple independent substitutions in phylogenetic inference (Kelchner & Wendel 1996, Kelchner 2000, Mes *et al.* 2000). Furthermore, heterogeneous mutations leading to a stem-loop secondary structure may create a 'mutational trigger'—a sequence pattern that increases the likelihood of subsequent change (Kelchner 2000). Hence, the apparent character non-independence biases analyses and may become a major problem for the phylogenetic inference

Accepted by M. Christenhusz: 6 Oct. 2009; published: 20 Oct. 2009

al. 2000). Although repeats do not necessarily affect tree topology, they are often more informative and less homoplastic than single base substitutions (Graham *et al.* 2000). Their treatment in phylogenetic analyses may also be important for phylogeny reconstruction, especially if gaps are treated as a fifth character state. Several base pairs long repeats will then lead to unnecessary upweighting of an indel event which often is the result of one event.

Acknowledgements

We thank Maarten Christenhusz, Michael Kessler, Hanna Tuomisto, and the curators of AAU, BM, GOET, L, TUR and U for providing plant material and allowing its use for molecular study. The study was funded by an Academy of Finland grant to Hanna Tuomisto. The Willi Hennig Society is acknowledged for making TNT publicly available.

References

Allard, M. & Miyamoto, M.M. (1992) Testing phylogenetic approaches with empirical data, as illustrated with the parsimony method. *Molecular Biology and Evolution* 9: 778–786.

Bain, J.F. & Jansen, R.K. (2006) A chloroplast DNA hairpin structure provides useful phylogenetic data within tribe Senecioneae (Asteraceae). *Canadian Journal of Botany* 84: 862–868.

Blakeley, E., Rennie, K., Jones, L., Elstener, M., Chrzanowska-Lightowlers, Z., White, C., Shield, J., ilz, D., Turnbull, D., Poulton, J. & Taylor, R. (2006) Sporadic intragenic inversion of the mitochondrial DNA MTND1 gene causing fatal infantile lactic acidosis. *Pediatric Research* 59: 440–444.

Catalano, S.A., Saidman, B.O. & Vilardi, J.C. (2009) Evolution of small inversions in chloroplast genome: a case study from a recurrent inversion in angiosperms. *Cladistics* 25: 93–104.

Catalano, S.A., Vilardi, J.C., Tosto, D. & Saidman, B.O. (2008) Molecular phylogeny and diversification history of *Prosopis* (Fabaceae: Mimosoideae). *Biological Journal of the Linnean Society (London)* 93: 621–640.

Chaisson, M.J., Raphael, B.J. & Pevzner, P.A. (2006) Microinversions in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America* 103: 19824–19829.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madriñan, S., Petersen, G., Seberg, O., Jørgensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L. & Wilkinson, M. (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295–299.

Crayn, D.M. & Quinn, C.J. (2000) The evolution of the *atp*β-*rbc*L intergenic spacer in the Epacrids (Ericales) and its systematic and evolutionary implications. *Molecular Phylogenetetics and Evolution* 16: 238–252.

Cunningham, C.W. (1997) Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Systematic Biology* 46: 464–478.

Demesure, B.N., Sodzi, N. & Petit R.J. (1995) A set of universal primers or amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129–131.

De Rijk, P., Wuyts, J. & De Wachter, R. (2003) RnaViz2: an improved representation of RNA secondary structure. *Bioinformatics* 19: 299–300.

Dixon, M.T., Hillis, D.M. (1993) Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Molecular Biology and Evolution* 10: 256–267.

Drábková, L. Kirschner, J. & Vlcek, C. (2002) Comparison of seven DNA extraction and amplification protocols in historical herbarium specimens of Juncaceae. *Plant Molecular Biology Reporter* 20: 161–175.

Dryander, J. (1797) *Lindsaea* a new genus of ferns. *Transactions of the Linnean Society of London* 3: 39–43.

Dumolin-Lapègue, S., Pemonge, M.-H. & Petit, R.J. (1998) Association between chloroplast and mitochondrial lineages in oaks. *Molecular Biology and Evolution* 15: 1321–1331.

Farris, J.S. 2000 Corroboration versus “strongest evidence”. *Cladistics* 16: 385–393.

Farris, J.S., Källersjö, M. & Kluge, A.G. (1995) Testing significance of incongruence. *Cladistics* 10: 315–319.

Fitzhugh, K. (2006) The “requirement of total evidence” and its role in phylogenetic systematics. *Biology & Philosophy* 21: 309–351.

Giribet, G. & Wheeler, W.C. (1999) On gaps. *Molecular Phylogenetics and Evolution* 13: 132–143.

Goloboff, P.A. (2008) Calculating SPR distances between trees. *Cladistics* 24: 591–597.
Goloboff, P.A., Farris, J.S., Nixon, K.C. (2008) TNT, a free program for phylogenetic analysis. *Cladistics* 24: 774–786.
Goloboff, P.A., Mattoni, C.I. & Quinteros, A.S. (2006) Continuous characters analyzed as such. *Cladistics* 22: 589–601.
Graham, S.W., Reeves, P.A., Burns, A.C.E. & Olmstead, R.G. (2000) Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161: S83–S96.
Hernandez-Maqueda, R., Quandt, D., Werner, O. & Munoz, J. (2008) Phylogeny and classification of the Grimmiaceae/Ptychomitriaceae complex (Bryophyta) inferred from cpDNA. *Molecular Phylogenetics and Evolution* 46: 863–877.
Hillis, D.M. (1995) Approaches for assessing phylogenetic accuracy. *Systematic Biology* 44: 3–16.
Huelsenbeck, J.P. & Ronquist, F. (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
Huttunen, S., Gardiner, A.A. & Ignatov, M.S. (2007) Advances in knowledge of the Brachytheciaceae (Bryophyta). In: Newton, A.E. & Tangney, R.S. (eds.). Pleurocarpous mosses: systematics and evolution. *Systematics Association Special Volume* 71: 117–143. Kelchner, S.A. (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.
Huttunen, S. & Ignatov, M.S. (2004) Phylogeny of the Brachytheciaceae (Bryophyta) based on morphology and sequence level data. *Cladistics* 20: 151–183.
Kelchner, S.A. (2002) Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *American Journal of Botany* 89: 1651–1669.
Kelchner, S.A. & Clark, L.G. (1997) Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution* 8: 385–397.
Kelchner, S.A. & Wendel, J.F. (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30: 259–262.
Kim, K.-J. & Lee, H.-L. (2005) Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Molecules and Cells* 19: 104–113.
Kluge, A.G. (1989) A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* 38: 7–25.
Källersjö, M, Albert, V.A. & Farris, J.S. (1999) Homoplasy increases phylogenetic structure. *Cladistics* 15: 91–93.
Lehtonen, S. & Myllys, L. (2008) Cladistic analysis of *Echinodorus* (Alismataceae): simultaneous analysis of molecular and morphological data. *Cladistics* 24: 218–239.
Lehtonen, S. & Tuomisto, H. (2007) Ecological and genetic notes on *Lindsaea digitata* (Lindsaeceae), a new fern species from western Amazonia. *Nordic Journal of Botany* 25: 145–151.
Levison, G. & Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 9: 856–871.
Lewis, P.O. (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50: 913–925.
Mathews, D.H., Zuker, M. & Turner, D. H. (2008) RNAstructure 4.6. (computer program) http://rna.urmc.rochester.edu/rnastructure.html (Date of access Aug. 5th, 2009).
Mes, T.H.M., Kuperus, P., Kirschner, J., Stêpánek, J., Oosterveld, P., Storchová, H. & den Nijs, J.C.M. (2000) Hairpins involving both inverted and direct repeats are associated with homoplasious indels in non-coding chloroplast DNA of *Taraxacum* (Lactuceae: Asteraceae). *Genome* 43: 634–641.
Mickevich, M.F. & Farris, J.S. (1981) The implications of congruence in *Menidia*. *Systematic Zoology* 30: 351–370.
Müller, K., Quandt, D., Müller, J. & Neinhuis, C. (2005) PhyDE ® 0.983: Phylogenetic Data Editor. Available at www.phyde.de (Date of access Aug. 5th, 2009).
Ogden, T.H. & Rosenberg, M.S. (2007) How should gaps be treated in parsimony? A comparison of approaches using simulation. *Molecular Phylogenetetics and Evolution* 42: 817–826.
Planet, P.J. (2006) Tree disagreement: measuring and testing incongruence in phylogenies. *Journal of Biomedical Informatics* 39: 86–102.
Posada, D. & Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
Quandt, D., Müller, K. & Huttunen, S. (2003) Characterisation of the chloroplast DNA psbT-H region and the influence of dyad symmetrical elements on phylogenetic reconstructions. *Plant Biology* 5: 400–410.
Quandt, D. & Stech, M. (2004) Molecular evolution of the $trnT_{UGU}$-$trnF_{GAA}$ region in bryophytes. *Plant Biology* 6: 545–554.
Rohlf, F.J. (1982) Consensus indices for comparing classifications. *Mathematical Biosciences* 59: 131–144.
Sang, T, Crawford, D.J. & Stuessy, T.F. (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.
Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E. & Small, R.L. (2005) The tortoise and the hare II: relatively utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.

Simmons, M.P. & Ochoterena, H. (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.
Stech, M., Frahm & J.-P. (1999) The status of *Platyhypnidium mutatum* Ochyra & Vanderpoorten and the systematic value of the Donrichardsiaceae based on molecular data. *Journal of Bryology* 21: 191–195.
Storchová, H. & Olson, M.S. (2007) The architecture of the chloroplast *trn*H-*psb*A non-coding region in angiosperms. *Plant Systematics and Evolution* 268: 235–256.
Swartz, O. (1801) Genera et species filicum. *Schrader's Journal für die Botanik* 1800(2): 1–120.
Swofford, D.L. (1998) PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA.
Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.
Tate, J.A. & Simpson, B.B. (2003) Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28: 723–737.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876–4882.
Varón, A., Vinh, L.S., Bomash, I. & Wheeler, W.C. (in press) POY version 4: phylogenetic analysis using dynamic homologies. *Cladistics*.
Vijverberg, K. & Bachmann, K. (1999) Molecular evolution of a tandemly repeated *trnF* (GAA) gene in the chloroplast genome of Microseris (Asteraceae) and the use of structural mutations in phylogenetic analyses. *Molecular Biology and Evolution* 16: 1329–1340.
Wheeler, W. (2001) Homology and the optimization of DNA sequence data. *Cladistics* 17: S3–S11.
Wheeler, W.C. (2007) Chromosomal character optimization. *Molecular Phylogenetics and Evolution* 44: 1130–1140.
Zwickl, D.J. & Hillis, D.M. (2002) Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51: 588–598.