Available online at www.sciencedirect.com
ScienceDirect
Gene 383 (2006) 33–37
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Alveolate and chlorophycean mitochondrial *cox2* genes split twice independently

Ross F. Waller *, Patrick J. Keeling

Canadian Institute for Advanced Research, Botany Department, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, Canada V6T 1Z4

Received 1 April 2006; received in revised form 20 June 2006; accepted 6 July 2006
Available online 20 July 2006
Received by G. Pesole

Abstract

The mitochondrial gene for COXII is typically encoded in the organelle genome, however in some members of two unrelated groups, Apicomplexa and Chlorophyceae, *cox2* is split into two genes, and both are encoded in the nucleus. Rare genomic changes (RGCs) have acquired popularity as phylogenetic markers, and accordingly this rearrangement of *cox2* has been used to infer a possible source of the apicomplexan plastid, the apicoplast, a topic that continues to attract much debate. Accurate interpretation of RGCs, however, is critically dependent on appropriate sampling of the character state of interest amongst relevant taxa. Dinoflagellates form the sister taxon to Apicomplexa, and therefore the state of their *cox2* is essential to the interpretation of this apparent RGC. Here we present the first complete *cox2* data from dinoflagellates, that suggests despite the remarkable similarity of *cox2* seen in Alveolates and Chlorophyceae, this gene reorganization arose independently in these two groups, not through lateral transfer as previously suggested.

Keywords: Dinoflagellate; Apicomplexa; Rare genomic changes; Apicoplast

1. Introduction

Resolving many major events in eukaryotic evolution, and the relationships among participate lineages, is a difficult challenge to evolutionary biologists due to the great antiquity of these events. In particular, molecular phylogenies are often unable to provide robust resolution to extremely deeply branching lineages. Increasingly, and quite appropriately, many biologists are looking for the inheritance of characters that result from discrete and relatively improbable molecular events – such as gene fusions, replacements and rearrangements – to provide markers for deep relationships and divergences (e.g. Stechmann and Cavalier-Smith, 2002).

One such example is a peculiar rearrangement of the mitochondrial COXII gene, a subunit of cytochrome *c* oxidase (Funes et al., 2002; Waller et al., 2003). This rearrangement has been interpreted to address the vexing question of the origin of the secondary plastid (known as the apicoplast) in apicomplexan parasites such as the malaria pathogen *Plasmodium* (Funes et al., 2002, 2004; Hackett et al., 2004a,b). In almost all eukaryotes COXII is encoded by a single gene (*cox2*) in the mitochondrion. In apicomplexans, however, *cox2* has been split into two genes (*cox2a* and *cox2b*) and both have been relocated to the nucleus (Funes et al., 2002; Gardner et al., 2002). This derivation of *cox2* is only otherwise known in select green algae from the class Chlorophyceae (Funes et al., 2002). The apparent rarity of this state implies that such a rearrangement is unlikely to have occurred twice. Since apicomplexans contain a plastid whose origin from either the green or red algal lineage has attracted much debate (Köhler et al., 1997; Blanchard and Hicks, 1999; Fast et al., 2001; Waller et al., 2003, 2006; Funes et al., 2004; Patron et al., 2004; Waller and McFadden, 2005), this *cox2* arrangement in apicomplexans was used to implicate a green algal origin of the apicoplast (Funes et al., 2002). Apicomplexans were thus

Abbreviations: RGC, rare genomic change; COXII, cytochrome *c* oxidase subunit 2; EST, expressed sequence tag; AU, approximately unbiased; ML, maximum likelihood; NJ, neighbor joining.

* Corresponding author. Present address: School of Botany, University of Melbourne, 3010, Australia. Tel.: +61 3 8344 5042; fax: +61 3 9347 5460.
E-mail address: r.waller@unimelb.edu.au (R.F. Waller).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.07.003

proposed to have inherited the derived *cox2* through endosymbiosis of a chlorophycean alga, and gene transfer from the endosymbiont nucleus to the host nucleus of both *cox2a* and *cox2b* leads to the replacement of the endogenous mitochondrial *cox2* in apicomplexans (Funes et al., 2002). However, a fuller understanding of *cox2* evolution in apicomplexans requires knowledge of the state of *cox2* in the sister lineage to Apicomplexa, the dinoflagellates. In this study we have acquired such data from dinoflagellates that argues for an alternative hypothesis for the origin of the split *cox2* gene in apicomplexans.

2. Materials and methods

2.1. Characterization of dinoflagellate cox2 sequences

Homologues of COXII coding sequence were identified by searching ongoing or complete dinoflagellate expressed sequence tag (EST) projects on *Karlodinium micrum* CCMP 415 (Patron et al., 2006), and *Oxyrrhis marina* CCMP 1788 (EST sequencing in progress). All COXII-encoding ESTs were completely sequenced from both strands, and these have been deposited in GenBank (accession numbers DQ462645–DQ462648). The presence of predicted N-terminal extensions, that could serve as mitochondrial transit peptides, was assessed by alignment to COXII sequences from other organisms, and using the online transit peptide predictor MitoProt II (Claros and Vincens, 1996). Protein mesohydrophobicity (average regional hydrophobicity over a 60- to 80-residue window) and maximal local hydrophobicity (over 17 residues) were calculated with MitoProt II (Claros and Vincens, 1996) using the GES hydrophobicity scale.

2.2. Phylogenetic analyses

COXII amino acid sequence alignments were constructed using Clustal X (Thompson et al., 1997) with dinoflagellate sequences and homologues available in public databases. Existing genomic sequence data from *Perkinsus marinus* (www.tigr.org/tdb/e2k1/pmg), the sister group to dinoflagellates, was also searched for *cox2* sequences. Alignments were manually adjusted resulting in a dataset of 140 aligned homologous residues. Due to relatively divergent ciliate COXII sequences, the dataset including ciliate taxa was reduced to 120 residues. Phylogenetic analyses were carried out using maximum likelihood and distance. Maximum likelihood trees were inferred using PhyML 2.4.4 (Guindon and Gascuel, 2003). Site-to-site rate variation was modelled on a gamma distribution with 8 rate categories and invariable sites. The shape parameter alpha and proportion of invariable sites were estimated by PhyML using the WAG substitution matrix. Distances were calculated by TREE-PUZZLE 5.2 (Strimmer and von Haeseler, 1996) with the WAG substitution matrix, with 8 rate categories and invariable sites all estimated from the data. Bootstraps were calculated using puzzleboot (shell script by A. Roger and M. Holder, http://www.tree-puzzle.de) and the same conditions. Trees were inferred using WEIGHBOR 1.0.1a (Bruno et al., 2000). Alpha and Γ parameters were 1.42 and 0.16, respectively, for the dataset without ciliates, and 1.60 and 0.14, respectively including ciliates.

3. Results and discussion

3.1. cox2 in dinoflagellates

Dinoflagellate algae are the closest relatives to apicomplexans, and together with ciliates these three groups comprise the Alveolata. The state of dinoflagellate *cox2* is therefore critical to interpreting the timing and origin of the *cox2* split in apicomplexans. Expressed sequence tag (EST) data from two dinoflagellates, *Alexandrium tamarense* and *Karenia brevis*, reported sequence corresponding to the *cox2b* domain apparently encoded in the nucleus (Hackett et al., 2004a,b). This data indicated that, as with apicomplexans, *cox2* is likely split in dinoflagellates, however the location of a coding sequence corresponding to *cox2a* remained unknown. Of the three chlorophycean algae with split *cox2* genes, *Scenedesmus* represents an intermediate state of *cox2* evolution, with the *cox2a* gene still encoded in the mitochondrion (i.e., *Scenedesmus* diverged from other green algae prior to the migration of *cox2a* to the nucleus) (Funes et al., 2002). It remained possible, therefore, that dinoflagellate *cox2a* might also still remain in the mitochondrion — an evolutionary intermediate analogous to *Scenedesmus*. We have surveyed ESTs from the dinoflagellate *K. micrum* (Patron et al., 2006) and the non-photosynthetic sister lineage to dinoflagellates, *O. marina*, and recovered discrete, full-length cDNAs for both *cox2a* and *cox2b* genes from both lineages (GenBank accession numbers DQ462645–DQ462648). Congruent with a nucleus location, both genes from both lineages encode an N-terminal extension with the

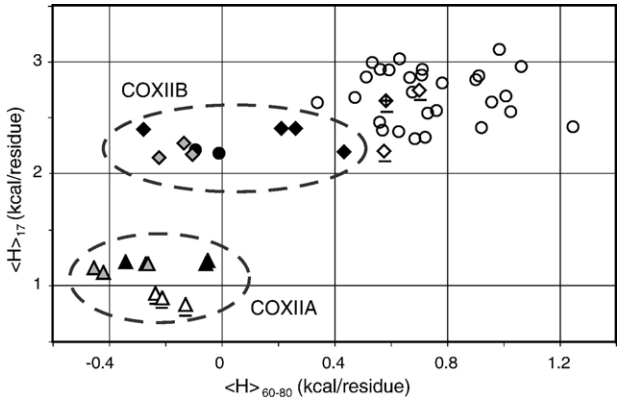

Fig. 1. Mesohydrophobicity ($<H>_{60–80}$) versus maximal local hydrophobicity ($<H>_{17}$) plot of COXII proteins represented in the phylogenies (Fig. 2). Nucleus-encoded split COXII components that are imported into the mitochondrion show reduced hydrophobicity compared to intact mitochondrion-encoded COXII. Circles indicate intact COXII proteins, triangles and diamonds represent split COXIIA and COXIIB respectively. Grey symbols indicate proteins from dinoflagellates, black from apicomplexans (triangles and diamonds) or ciliates (circles), and underscored symbols from chlorophycean aglae (note, mitochondrion-encoded COXIIA from *Scenedesmus* is indicated by a crossed-diamond). Ciliate COXIIs display similar reduced hydrophobicity to other alveolate homologues, despite being encoded in the mitochondrion. The GES hydrophobicity scale was used.

features of mitochondria-targeting transit peptides (predicted according to MitoProt II (Claros and Vincens, 1996)). Moreover, both predicted proteins obey a defined trend towards reduced hydrophobicity compared to the single COXII protein that occurs for all nucleus-encoded versions of COXII (Fig. 1). This modification of protein hydrophobicity is believed to be necessary for targeting this otherwise hydrophobic protein back across the mitochondrial membranes (Claros and Vincens, 1996; Daley et al., 2002; Waller et al., 2003). Together these data demonstrate that dinoflagellates share the derived state of a split, nucleus-encoded *cox2* as found in apicomplexans.

3.2. COXII phylogeny

Early phylogenetic analyses of COXII (Funes et al., 2002) indicated a sister relationship between the chlorophycean and apicomplexan split proteins, and it was argued that this indicated a single common ancestry of split COXII, and therefore lateral transfer of *cox2a* and *cox2b* to apicomplexans from a chlorophycean alga (Funes et al., 2002). However several factors compromised the utility of this initial analysis. Data from the apicomplexans' closest relatives was either unavailable (dinoflagellates) or neglected (ciliates), and hence vertical inheritance of COXII in apicomplexans, compared to lateral inheritance, was not rigorously tested. Further, in this and all subsequently published COXII phylogenies (Waller et al., 2003; Hackett et al., 2004b) the clade comprising split chlorophycean *cox2* genes does not branch with other green alga or plants, and there is very little robust resolution among other major clades, suggesting the phylogenetic information in this molecule is limited. Nevertheless, to test alveolate COXII relationships, and if better resolution of split COXIIs can be achieved with more complete representation of taxa, we have included the *K. micrum* and *O. marina* COXIIA and COXIIB sequences in COXII phylogenies. We have also included existing genomic data from the deep branching dinoflagellate *Perkinsus marina* (from the *P. marina* genome at TIGR: www.tigr.org/tdb/e2k1/pmg). While this *P. marina* genomic data is still incomplete, *cox2a* and *cox2b* sequences are represented on distinct scaffolds, thus implying a split *cox2*. Protein phylogenies were constructed using maximum likelihood and distance methods, both including and excluding the

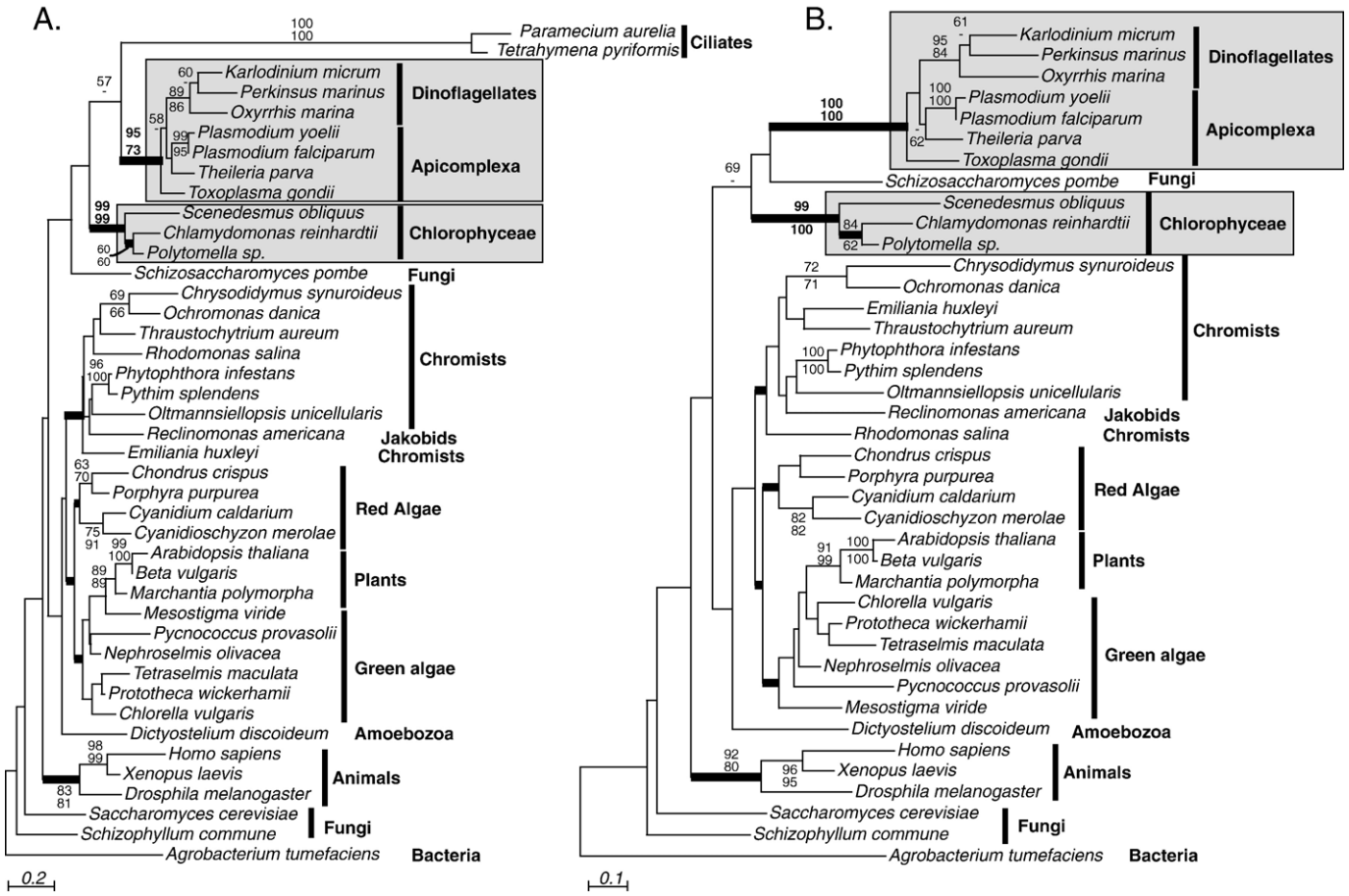

Fig. 2. Protein maximum likelihood phylogeny of COXII with (A) and without (B) ciliate sequences. Taxa with split COXII proteins are shown in grey boxes, numbers at nodes correspond to bootstrap support for nodes over 50% (ML above, NJ below) and major groups are named to the right. Dinoflagellate and apicomplexan proteins form a clade with strong support, however the position of this clade within the COXII phylogeny is not resolved either with or without the more divergent ciliate sequences. AU tests positioning the dinoflagellate/apicomplexan clade at the other major nodes in this phylogeny (indicated by thick nodes) did not reject any of these alternative topologies ($p > 0.05$).

relatively divergent ciliate sequences to control for possible long-branches topology artifacts. In all phylogenies a close relationship between dinoflagellate and apicomplexan sequences is strongly supported by bootstrap analysis (95/73, 100/100; Fig. 2A and B, respectively) as was seen when COXIIB was analyzed alone (Hackett et al., 2004a,b). Similarly, the chlorophycean clade is also strongly supported (99/99, 99/100; Fig. 2). While the best topology grouped the ciliates as sisters to the dinoflagellates and apicomplexans, uniting all alveolates (Fig. 2A), and the alveolates as sisters to chlorophycean algae, neither of these relationships was supported by bootstrapping. The removal of the divergent ciliate sequences weakens the relationship of alveolates and chlorophycean algae further with the fungal sequence from *S. pombe* branching amongst the split COXIIs (Fig. 2B). Furthermore, AU testing of alternative topologies, shifting the dinoflagellate/apicomplexan clade to all other major nodes (Fig. 2) rejected none of these alternate topologies, suggesting that the phylogeny alone cannot be used to support either common or independent origins of split *cox2*.

The two strongest results from the phylogeny do nevertheless provide firm support for 1) dinoflagellate and apicomplexan split *cox2* genes sharing a common origin and, separately, 2) that split *cox2* genes of the chlorophycean algae also share a common origin to the exclusion of the other taxa. This second relationship is significant, given that *Scenedesmus cox2a* is still encoded in the mitochondrion, representing an intermediate state of *cox2* evolution (Funes et al., 2002). If apicomplexan and dinoflagellate split *cox2* genes were derived from nucleus–nucleus transfer from a chlorophycean endosymbiont, this endosymbiont would have diverged after the divergence of *Scenedesmus* in chlorophycean radiation, and this should be reflected by dinoflagellate and apicomplexan *cox2* sequences branching within the clade of split green algal genes and specifically above *Scenedesmus* — a result that is not observed in any phylogeny. It is formally possible that a putative green algal endosymbiont may have possessed mitochondrion-encoded split *cox2* genes, and that multiple independent mitochondrion–nucleus relocations have occurred in green algae and alveolates, however again, the phylogenies do not provide support for this scenario.

Plastids of dinoflagellates are similar to those of apicomplexans in that they represent secondary endosymbionts (except in unusual cases where subsequent plastid replacement has occurred) (Hackett et al., 2004a). Molecular and pigment data strongly support the origin of the dinoflagellate plastid from within the red algal lineage (Fast et al., 2001; Archibald and Keeling, 2002; Yoon et al., 2005; Waller et al., 2006). Therefore the introduction of a green algal endosymbiont to the alveolate lineage before the dinoflagellate–apicomplexan divergence seems unlikely. Further, as previously shown, ciliate *cox2* sequence contains a large insertion at the same location as the *cox2* split (Waller et al., 2003), as well as modified hydrophobicity (remarkably reduced compared to other COXIIs, Fig. 1), indicating that alveolate *cox2* was already deviating from its canonical form, in some ways preconditioned to making a split and transfer to the nucleus more likely. On balance, there is no direct evidence that alveolate and green algal split *cox2* genes share a common origin. Instead, *cox2* has most likely split and relocated to the nucleus independently in both alveolate and chlorophycean lineages. Accordingly, no inference of the source of the apicomplexan plastid can be drawn from this event. This also serves as a warning that seemingly rare events can occur in parallel in unrelated lineages and need to be thoroughly scrutinized before far-reaching conclusions are drawn from the distribution of such characters.

Acknowledgments

This work was supported by a grant from the Canadian Institutes for Health Research (CIHR) to PJK (MOP-42517). EST sequencing was supported by the Protist EST Program funded by Genome Canada/Genome Atlantic. RFW was supported by a fellowship from the CIHR and Michael Smith Foundation for Health Research (MSFHR). PJK is a Fellow of the Canadian Institute for Advanced Research and a New Investigator of CIHR and MSFHR.

References

Archibald, J.M., Keeling, P.J., 2002. Recycled plastids: a green movement in eukaryotic evolution. Trends Genet. 18, 577–584.

Blanchard, J.L., Hicks, J.S., 1999. The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. J. Eukaryot. Microbiol. 46, 367–375.

Bruno, W.J., Socci, N.D., Halpern, A.L., 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol. Biol. Evol. 17, 189–197.

Claros, M.G., Vincens, P., 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur. J. Biochem. 24, 79–86.

Daley, D.O., Clifton, R., Whelan, J., 2002. Intracellular gene transfer: reduced hydrophobicity facilitates gene transfer for subunit 2 of cytochrome *c* oxidase. Proc. Natl. Acad. Sci. U. S. A. 99, 10510–10515.

Fast, N.M., Kissinger, J.C., Roos, D.S., Keeling, P.J., 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol. Biol. Evol. 18, 418–426.

Funes, S., et al., 2002. A green algal apicoplast ancestor. Science 298, 2155.

Funes, S., Reyes-Prieto, A., Perez-Martinez, X., Gonzalez-Halphen, D., 2004. On the evolutionary origins of apicoplasts: revisiting the rhodophyte vs. chlorophyte controversy. Microbes Infect. 6, 305–311.

Gardner, M.J., et al., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419, 498–511.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hackett, J.D., Anderson, D.M., Erdner, D.L., Bhattacharya, D., 2004a. Dinoflagellates: a remarkable evolutionary experiment. Am. J. Bot. 91, 1523–1534.

Hackett, J.D., et al., 2004b. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr. Biol. 14, 213–218.

Köhler, S., et al., 1997. A plastid of probable green algal origin in apicomplexan parasites. Science 275, 1485–1489.

Patron, N.J., Rogers, M.B., Keeling, P.J., 2004. Gene replacement of fructose-1,6-bisphosphate aldolase (FBA) supports a single photosynthetic ancestor of chromalveolates. Eukaryot. Cell 3, 1169–1175.

Patron, N.J., Waller, R.F., Keeling, P.J., 2006. A tertiary plastid uses genes from two endosymbionts. J. Mol. Biol. 357, 1373–1382.

Stechmann, A., Cavalier-Smith, T., 2002. Rooting the eukaryote tree by using a derived gene fusion. Science 297, 89–91.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Waller, R.F., McFadden, G.I., 2005. The apicoplast: a review of the derived plastid of apicomplexan parasites. Curr. Issues Mol. Biol. 7, 57–79.

Waller, R.F., Keeling, P.J., van Dooren, G.G., McFadden, G.I., 2003. Comment on "A green algal apicoplast ancestor". Science 301, 49.

Waller, R.F., Patron, N.J., Keeling, P.J., 2006. Phylogenetic history of plastid-targeted proteins in the peridinin-containing dinoflagellate *Heterocapsa triquetra*. Int. J. Syst. Evol. Microbiol. 56, 1439–1447.

Yoon, H.S., Hackett, J.D., van Dolah, F.M., Nosenko, T., Lidie, K.L., Bhattacharya, D., 2005. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. Mol. Biol. Evol. 22, 1299–1308.