



Plastid endosymbiosis, genome evolution and the origin of green plants

John W. Stiller

Department of Biology, Howell Science Complex, N108, East Carolina University, Greenville, NC 27858, USA

Evolutionary relationships among complex, multicellular eukaryotes are generally interpreted within the framework of molecular sequence-based phylogenies that suggest green plants and animals are only distantly related on the eukaryotic tree. However, important anomalies have been reported in phylogenomic analyses, including several that relate specifically to green plant evolution. In addition, plants and animals share molecular, biochemical and genome-level features that suggest a relatively close relationship between the two groups. This article explores the impacts of plastid endosymbioses on nuclear genomes, how they can explain incongruent phylogenetic signals in molecular data sets and reconcile conflicts among different sources of comparative data. Specifically, I argue that the large influx of plastid DNA into plant and algal nuclear genomes has resulted in tree-building artifacts that obscure a relatively close evolutionary relationship between green plants and animals.

Phylogenies and the origin of plants

Phylogenetic analyses using molecular sequence data have provided influential new hypotheses regarding eukaryotic evolutionary relationships and the origin of green plants (Viridiplantae; see Glossary). Most sequence-based phylogenies strongly support a sister relationship between animals and fungi (opisthokonts; see Glossary), with green plants as a distant outgroup [1]. Recovery of opisthokonts is among the most robust results of broad-scale eukaryotic molecular systematics. The grouping, now, is included in most introductory biology texts, and is used as a benchmark for interpreting the evolution of genome-level characters and complex metabolic processes.

Comparably strong phylogenetic relationships have been harder to pin down for green plants. Molecular phylogenetics that recovered all plastids as monophyletic with respect to cyanobacteria led to the hypothesis of a single endosymbiotic origin in the common ancestor of green plants and red algae, the two major groups containing primary plastids (see Glossary) [2,3]. Whether available data strongly favor a single plastid endosymbiosis has been addressed elsewhere, and the issue remains difficult to resolve [4]; however, even a single plastid

origin does not necessarily imply a close evolutionary relationship between Viridiplantae and Rhodophyta (see Glossary) [5,6].

Many phylogenetic analyses based on nuclear sequence data strongly reject a red–green sister relationship [5,7,8], and inferred differences in RNAP II transcription are difficult to reconcile with their putatively monophyletic relationship [9]. In contrast, recent phylogenomic analyses provided strong support for a hypothesis in which three primary plastid lineages (green plants, red algae and glaucocystophytes) share a common ancestor [10]. This study noted, however, that a relatively large amount of sequence is required to recover this relationship [10], and subsequent analyses of the same data showed that the tree-building signal supporting monophyly of primary plastid eukaryotes was due to more rapidly evolving sequences [8]. This suggests that the result might be a phylogenetic artifact caused by rate variation in sequence evolution [8], although the source of this variation is unclear.

Glossary

Amoebozoans: a group that includes *Dicystelium* and diverse amoeboid protists, and is recovered as the sister group to opisthokonts in sequence-based phylogenies [48].

Glaucocystophyta: a small and relatively rare group of protists whose primary plastids (cyanelles) can be distinguished by the presence of a vestigial bacterial cell wall and cyanobacteria-like pigments.

Long-branch attraction: a catch phrase for tree-building artifacts, named for the tendency of rapidly evolving sequences to attract each other artificially in phylogenetic analyses. The correlated tendency of slowly evolving sequences to cluster artificially (short branch exclusion) can be a major but underappreciated problem in phylogenetic analyses at great evolutionary distances.

Opisthokonts: a group comprising animals, fungi and their respective protistan relatives, proposed originally [49] based on the presence of a posterior flagellum as a synapomorphic character.

Primary plastids: plastids (chloroplasts) believed to be derived directly from a cyanobacterial endosymbiont based on the presence of two surrounding membranes.

Purifying selection: the elimination of deleterious alleles from a population or species by natural selection.

Rhodophyta: commonly called red algae, a diverse and ancient eukaryotic group with primary plastids that contain phycobiliproteins as the major accessory pigments.

Synapomorphy: a derived evolutionary character that is believed to have originated in the common ancestor of two or more taxa.

Viridiplantae: the monophyletic group comprising all green algae and land plants. It is used synonymously with the term green plants in this article.

Corresponding author: Stiller, J.W. (stillerj@ecu.edu).
Available online xxxxxx.

A persistent artifact related to green plant origins

The difficulty in recovering a well-supported sister relationship with red algae is not the only phylogenetic curiosity relating to green plants origins. Early attempts at multigene phylogenies of major eukaryotic lineages typically recovered animals and fungi as sister taxa, or placed animals with green plants, but seldom suggested a close relationship between green plants and fungi [11,12]. The presence of a second directional tree-building signal was characterized as a persistent artifact of phylogenetic reconstruction [11]; the fact that animals and fungi were recovered as sister taxa by most sequences led to the presumption that trees supporting a plant–animal relationship reflect this artifact [11]. After this preliminary characterization, most multigene investigations of eukaryotic relationships have not examined the problem further; those that have provide evidence that these two conflicting phylogenetic signals are a consistent feature of molecular sequence data [13,14].

In addition to sequence-based phylogenies, an animal–fungi relationship is supported by a number of putatively shared-derived characters (synapomorphies). These include an insertion in the elongation factor-1 α gene [11], the fusion of the dihydrofolate and thymidylate synthase genes [15], the acquisition of archaeal tyrosyl-tRNA synthetase by horizontal gene transfer [16], and the presence of orthologous genes absent from most other eukaryotes [17]. Interestingly, comparable molecular features that could be considered synapomorphic have been reported for green plants and animals, suggesting that the two competing phylogenetic signals are not limited to tree-building algorithms. The pre-mRNA capping pathway is perhaps the most compelling example. Green plants and animals are unique in their use of a different class of triphosphatase enzyme that also is fused to guanylyl transferase, thereby producing a multifunctional capping enzyme (CE) [18]. Additional evidence of a close animal–green plant relationship includes similarities in regulatory elements for RNA polymerase III transcription [19], homologies in mechanisms for innate immune responses [20] and uniquely shared orthologous gene families [21–23]. In fact, genomic surveys of gene orthologs and protein domains consistently show a stronger similarity between plants and animals, with fungi as the outlier [24–27]. Although this might be attributable to an extensive gene loss from smaller fungal genomes, efforts to sample more broadly and to incorporate weighting based on genome size have not changed the result [26].

The shared presence of complex biochemical, genetic and proteomic networks further support a close relationship between green plants and metazoans. The strongly conserved retinoblastoma protein (Rb)–E2F cell cycle control pathway [28], which regulates G1 to S phase progression, is found only in animals and green plants. Moreover, this is only one of the apparently homologous mechanisms that control cell development in both lineages, including master regulators of stem cell differentiation [29] and control of reproduction by polycomb group complexes [30]. Evidence for these conserved functions has not yet surfaced from the numerous fungal genomes sequenced to

date. The evolution of such complex genomic features and metabolic networks generally are interpreted within the framework of sequence-based phylogenies, which presumes a relatively distant relationship between green plants and animals. This can sometimes lead to more complicated and even strained evolutionary assumptions (Box 1). Thus, even as strong consensus relationships emerge from sequence-based phylogenomics, serious questions remain [1]. This is particularly true for inferences at great genetic distances, such as those relating to green plant origins, where lineage-specific biases can confound tree-building algorithms [1,7,31–33].

The impact of plastid endosymbioses on nuclear genomes

Although the common presence of plastids has influenced hypotheses about eukaryotic phylogenetic relationships, the potential biases these plastids might introduce into nuclear phylogenomic inference require more scrutiny. In particular, it is important to explore whether their presence can explain existing conflicts among molecular data sets, and/or the tree-building anomalies outlined above.

Box 1. Precursors or remnants of complex metabolic pathways?

The evolution of a complex molecular or biochemical feature is generally interpreted in light of sequence-based trees. When any genes involved in a given pathway are found in a member of a lineage, it is assumed that the entire pathway is present, or at least was ancestrally in the group. This can lead to the construction of assumptions and complicated scenarios to explain characters that appear to be synapomorphic for green plants and animals. Such explanations might not be realistic biologically.

A prime example is the Rb–E2F pathway, which controls G1–S cell cycle progression and cell differentiation. The pathway is conserved across animals and plants, including the unicellular green alga *Ostreococcus* [50], but has not been reported from any of the numerous fungal genomes sequenced to date. Although also absent from most protistan genomes, possible Rb homologs have been found in a handful of organisms scattered over the eukaryotic tree. Outside plants and animals, the function of a putative Rb homolog has been characterized experimentally only in the slime mold *Dictyostelium*, where it appears to be important for stalk differentiation but not in cell cycle control [51].

Rb-like proteins were certainly present at some earlier point in eukaryotic evolution but, as in *Dictyostelium*, might have carried out related and/or different functions that were less central to the basic cell cycle. Consequently, they would not have come under the same intense purifying selection (see Glossary) that has preserved their core functions in animals and green plants. The most parsimonious interpretation of the full Rb–E2F pathway is that it originated in a common ancestor of metazoans and plants; this clearly conflicts with sequence-based phylogenetic inferences.

It is an inadequate practice to make the *a priori* judgment that such a complex and strongly conserved molecular system, present in both animals and green plants, was simply lost from numerous other eukaryotes. Specifically, reasonable evolutionary hypotheses should include explanations for how core cellular processes could be retained under strong selection in two distant and unrelated taxa, when they have been lost independently and repeatedly across the breadth of eukaryotic diversity. Given the theoretical and empirical problems with sequence-based deep phylogenetic reconstruction, it might be biologically more realistic to assume that the trees are wrong, and that features such as the Rb–E2F pathway and CE are just what they appear to be: shared, derived characters uniting green plants with animals.

It has been recognized for some time that the establishment of fully integrated plastids involved the transfer of much of the original endosymbiont's genome into the host cell nucleus [34]. It was originally thought that the process was largely complete early in the symbiotic association [35]; however, later investigations showed that gene transfer continued [36], and that similarities in the genes retained in plastid genomes were more consistent with a pattern of independent but convergent transfers [37].

The transfer of genes required for plastid function is only part of the overall impact on the nucleus. Approximately 18% of genes encoded in the *Arabidopsis* nucleus have a probable cyanobacterial origin [38]; many of the proteins they encode are not targeted back to the chloroplast. Rather, they have taken on alternative functions or have replaced original eukaryotic orthologs involved in host cell metabolism [38]; an example involving pyrimidine biosynthesis is discussed below. Given that most *Arabidopsis* genes are too divergent for their complete ancestry to be identified using comparative methods [38], additional undetected cyanobacterial replacements are likely. Even this, however, might be the tip of an iceberg.

Fine-scale analyses of the *Arabidopsis* genome led to the discovery that, despite its highly reduced state, the remnant plastid genome continues to pepper the nucleus with sequences of cyanobacterial ancestry, so called NUPTS (nuclear plastid DNA) [39,40]. Similar results have been found for sequences of mitochondrial origin (NUMTS) in both plants and animals [40], indicating that a general phenomenon is at play in the coevolution of organelle and nuclear genomes. These sequences, which appear to be distributed at random in nuclear chromosomes, range in size from entire copies of the organelle genome to tens of base pairs in length [39,40]. Uptake rates of plastid DNA by the nucleus are low in some taxa, presumably because they now contain a single chloroplast per cell [41,42]. Nevertheless, it is likely that ancestral nuclear genomes of green plants, red algae and glaucocystophytes experienced a steady rain of cyanobacterial sequences until the plastid endosymbiosis was fully integrated, with additional and perhaps continuous input in those taxa harboring multiple plastids per cell [42]. Thus, it is important to consider the probable consequences of so much plastid sequence replacement on phylogenomic investigations.

Implications for nuclear gene phylogenies

The most obvious potential complication is an inadvertent inclusion of endosymbiont genes in host cell phylogenies. In principle, this should be avoidable if the transferred genes have an obvious cyanobacterial signature. It is unclear, however, that the problem has received sufficient scrutiny in phylogenomic analyses. There might be any number of plant nuclear genes with cyanobacterial ancestry that are no longer easily recognizable as such and, therefore, are treated as *bona fide*, if divergent, eukaryotic orthologs.

Beyond the influence of whole-gene replacements, a more subtle and insidious problem would be the insertion or recombination of smaller cyanobacterial sequences into nuclear genes. Given the number of whole-gene replacements that have been tolerated, it is reasonable to speculate

that recombination between homologous or even similar but non-homologous cyanobacterial and nuclear genes was common early in the endosymbiotic relationship, before most genes were lost from the plastid. The presence of short, plastid-derived sequences would be hard to detect, particularly with ancient transfers that do not retain a clear cyanobacterial signature. Nevertheless, anecdotal cases of horizontal transfer of gene segments have been reported in eukaryotes, even in the absence of a clear mechanism to account for their movement [43]. At least one possible partial-gene recombination has strongly influenced the idea of a rhodophyte–green plant sister relationship [44], although no obvious cyanobacterial connection was identified in this case. Given established rates of organelle sequence influx into nuclear genomes and the length of time that nuclei of primary plastid eukaryotes have been exposed to DNA transfer from plastids, the cumulative impact is potentially enormous.

A plastid-impact hypothesis

The pervasive impact of plastid sequences on nuclear genomes suggests a unifying hypothesis for reconciling conflicts in phylogenomic and comparative investigations. It begins with the assumption that the similarities between plants and animals, as well as the directional tree-building signal in many nuclear genes favoring their close relationship, are present because the two groups share a relatively close ancestor. To provide a straightforward interpretation of features such as CE, as well as a clear demonstration of the potential impact of cryptic plastid sequences, it is assumed that plants and metazoans are immediate sister taxa. This need not be the case, however; the hypothesis requires only that animals and plants share a common ancestor to the exclusion of most other eukaryotic lineages, including red algae.

Under this assumption, how might the historical influx of plastid DNA into the plant nucleus impact phylogenetic reconstruction? If a given nuclear gene has no cyanobacterial imprint, it should indicate (ideally) a close relationship between plants and metazoans. A tree-building signal from these genes [Figure 1, gene (a)] can account for the so-called 'persistent artifact' that attracts these two groups in phylogenetic reconstruction. Given the potential magnitude of plastid impacts, combined with an acknowledged loss of historical information at deep levels [1,31,32], a relatively small fraction of plant nuclear genes might retain an accurate signal that reflects the evolutionary history of the host cell.

By contrast, any cyanobacterial imprint on a given nuclear gene will increase its genetic distance from eukaryotic orthologs. If included in phylogenetic reconstruction, such cryptic sequences will draw plant sequences away from their host cell relatives and toward longer branch outgroups [Figure 1, gene (b)]. This should result in a strong clustering of animals and fungi, because their sequences have changed less from those present in the common ancestor of all three groups. This tree-building artifact, directly related to the well-known problem of long-branch attraction (see Glossary), might be a far more important feature of deep phylogenies than has generally been recognized [7].

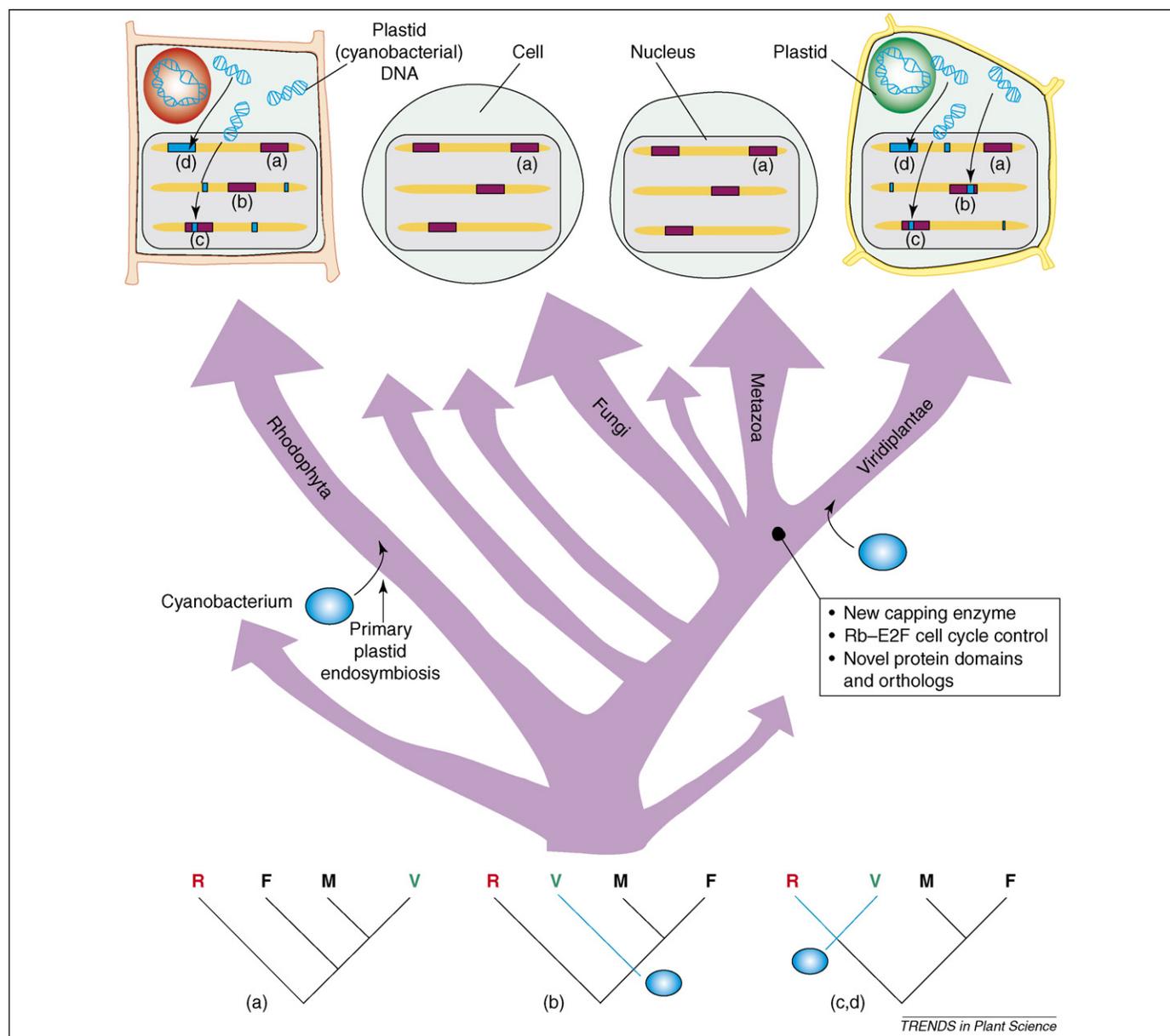


Figure 1. Hypothetical phylogeny of green plants, red algae, animals and fungi based on the potential impacts of plastid DNA on the nuclei of photosynthetic organisms. Unlabelled branches indicate the possibility of intervening eukaryotic lineages not addressed in this article. Examples of potential synapomorphies supporting a sister relationship between green plants and animals are shown in the white box. Cells with nuclei from each of the four groups are depicted at the top of their respective tree branches; plastids are present in the green plant and red algal cells as sources of DNA that can be integrated into the nucleus. The nuclei contain three chromosomes (in orange) with four orthologous eukaryotic loci (in purple); for simplicity, genome rearrangements are not depicted. Blue-green regions in green plant and red algal chromosomes indicate either insertions or recombinations of plastid-derived cyanobacterial sequences. The predicted impacts of these sequences on host cell phylogenetic inference are shown in the three tree-building scenarios at the bottom of the figure. Sequences from gene (a) produce a phylogeny reflecting historical relationships because there have been no impacts from plastid DNA; this explains the so-called persistent artifact supporting a green plant–animal sister relationship found in various sequences. Gene (b) contains a region of plastid DNA that has impacted only the green plant nuclear gene, thereby increasing its distance from its eukaryotic orthologs and leading to a tree supporting opisthokonts but not clustering green plants with red algae. Genes (c) and (d) contain parallel plastid impacts in green plants and red algae. Regardless of whether they have a common endosymbiotic origin, plastids all have a cyanobacterial ancestry, which draws red and green sequences together in phylogenetic reconstruction. To illustrate their potential impacts on nucleus-based phylogenomics most emphatically, plastids are depicted as having multiple origins; however, independent ancestors for red and green host cells are also fully compatible with a single plastid origin [5,6]. Thus, the plastid impact hypothesis is consistent with either scenario of plastid origins. Abbreviations: F, Fungi; M, Metazoa; R, Rhodophyta; V, Viridiplantae.

Based on fossil evidence, the lineages of primary plastid eukaryotes diverged from each other over a billion years ago [45]. Thus, even under the assumption of a single plastid origin, most genetic incursions into green plant genomes are unlikely to be shared with red algae or glaucocystophytes. Consequently, they are expected to produce an artifact favoring the opisthokonts, but not necessarily support a relationship among primary plastid eukaryotes. By contrast, any parallel replacement or

recombination that did occur would tend to draw the respective photosynthetic taxa together in phylogenetic analyses [Figure 1, genes (c,d)]. Because parallel events are expected to be rarer, it should take far more sequence data to build strong support for a relationship among primary plastid eukaryotes than for a grouping of animals with fungi. Moreover, these genes should appear to evolve relatively rapidly owing to the cryptic presence of divergent cyanobacterial sequences. These are

precisely the patterns reported in phylogenomic investigations [8,10].

The potential effects described above are consistent with (i) the dual tree-building signals that group animals either with fungi or with green plants, (ii) the presence of unique genomic and proteomic features shared by green plants and animals, (iii) the weak phylogenetic attraction among eukaryotes with primary plastids, and (iv) the finding that this attraction is produced by sequences that evolve more rapidly.

Although the specific mechanistic effects of cryptic plastid sequences on nuclear phylogenomics have yet to be demonstrated, a prominent example shows how plastid gene replacement can mislead interpretations of plant and animal origins: in the preliminary investigations of the *de novo* pyrimidine biosynthesis pathway, it was shown that animals, fungi and amoebozoans (see Glossary) share the fusion of three genes (*CPS II*, *ACT* and *DHO*) that encode a single CAD complex, which was proposed as a synapomorphic character uniting the three groups [15]. In green plants, however, several of the genes in question were replaced by plastid homologs, making the ancestral state of CAD-encoding genes impossible to determine. Subsequent sequencing of the red alga *Cyanidioschyzon* genome demonstrated the presence of both the fused CAD complex and additional plastid-derived homologs [46]. Thus, the CAD fusion probably occurred before the origin of the common ancestor of all these diverse groups and, subsequently, was lost from green plants through plastid gene replacement [46].

The original and evidently incorrect phylogenetic inference regarding the CAD fusion is precisely the artifact predicted for sequence-based trees [Figure 1, gene (b)]. It is worth asking, hypothetically, what the phylogenetic conclusion would have been, had the first red alga investigated also lost the ancestral CAD complex and retained only cyanobacterial-derived homologs? It seems likely that parallel replacements would have been considered a second synapomorphy (see Glossary) uniting red algae and green plants as sister groups [the artifact shown in Figure 1, genes (c,d)].

Future perspectives

The large impact of cyanobacterial sequences on plant nuclear genomes appears indisputable, but their influence on phylogenomic analyses remains unclear. Nevertheless, the potential that plastid endosymbioses affect nuclear phylogenetic inference so strongly suggests some general recommendations for the future investigations.

- (i) Major cellular changes, such as the incorporation of plastids, can lead to lineage specific effects with strong potential impacts on phylogenetic inference. This is true even in the absence of plastid–nuclear gene recombination; for example, ecophysiological selection plays an important role in shaping divergent patterns of amino acid content between animals and green plants, due to differences in the mechanisms of acquisition and allocation of nutritional resources [47]. These and other results from comparative genomics should be incorporated into an appreciation of how sequence data produce trees of relatedness.

- (ii) Individual sequences should be examined carefully for evidence of biases that might introduce tree-building artifacts into larger data sets; consistent but conflicting patterns should be investigated thoroughly.
- (iii) There should be no *a priori* assumption that the strongest tree-building signal in a given data set reflects evolutionary history rather than bias in the data.
- (iv) The evolution of complex genomic and proteomic characters should not be interpreted automatically through the prism of sequence-based phylogenies, or any other single methodological approach. As more eukaryotic genomes become available, different kinds of comparative data can be treated cladistically, with the cumulative signal from sequence-based trees representing just one of the characters used to polarize relationships.

To develop a complete, reliable and biologically useful model of plant origins and diversification, it is important that the evolutionary implications of comparative genomic and proteomic data be investigated thoroughly, both for what they imply directly about phylogenetic relationships and for their potential effects on computational algorithms. The plastid impact hypothesis provides one viable explanation for the results of phylogenomic and other comparative research on green plant origins, including existing conflicts and anomalous results, and deserves further investigation.

Acknowledgements

I thank Jinling Huang, Christopher Howell, Trip Lamb and Jason Stajich for their insights and helpful comments, and the efforts made by three anonymous reviewers to improve this article.

References

- 1 Roger, A.J. and Hug, L.A. (2006) The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. *Philos. Trans. Royal Soc. B.* 361, 1039–1054
- 2 Bhattacharya, D. and Medlin, L. (1995) The phylogeny of plastids – a review based on comparisons of small-subunit ribosomal-RNA coding regions. *J. Phycol.* 31, 489–498
- 3 Ragan, M.A. and Gutell, R.R. (1995) Are red algae plants? *Bot. J. Linn. Soc.* 118, 81–105
- 4 Larkum, A.W. *et al.* (2007) Shopping for plastids. *Trends Plant Sci.* 12, 189–195
- 5 Nozaki, H. *et al.* (2003) The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J. Mol. Evol.* 56, 485–497
- 6 Stiller, J.W. and Hall, B.D. (1997) The origin of red algae: implications for plastid evolution. *Proc. Natl. Acad. Sci. U. S. A.* 94, 4520–4525
- 7 Stiller, J.W. and Harrell, L. (2005) The largest subunit of RNA polymerase II from the Glaucocystophyta: functional constraint and short-branch exclusion in deep eukaryotic phylogeny. *BMC Evol. Biol.*, DOI: 10.1186/1471-2148-5-71 (www.biomedcentral.com)
- 8 Nozaki, H. *et al.* (2007) Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Mol. Biol. Evol.*, DOI: 10.1093/molbev/msm091 (mbe.oxfordjournals.org)
- 9 Guo, Z.H. and Stiller, J.W. (2005) Comparative genomics and evolution of proteins associated with RNA polymerase IIC-terminal domain. *Mol. Biol. Evol.* 22, 2166–2178
- 10 Rodriguez-Ezpeleta, N. *et al.* (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* 15, 1325–1330
- 11 Baldauf, S.L. and Palmer, J.D. (1993) Animals and fungi are each others closest relatives – congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. U. S. A.* 90, 11558–11562

- 12 Nikoh, N. *et al.* (1994) Phylogenetic relationship of the Kingdoms Animalia, Plantae, and Fungi, inferred from 23 different protein species. *Mol. Biol. Evol.* 11, 762–768
- 13 Philip, G.K. *et al.* (2005) The Opisthokonta and the Ecdysozoa may not be clades: stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. *Mol. Biol. Evol.* 22, 1175–1184
- 14 Stiller, J.W. (2004) Emerging genomic and proteomic evidence on relationships among the animal, plant and fungal kingdoms. *Genom. Proteom. Bioinform.* 2, 70–76
- 15 Stechmann, A. and Cavalier-Smith, T. (2003) The root of the eukaryote tree pinpointed. *Curr. Biol.* 13, R665–R666
- 16 Huang, J.L. *et al.* (2005) The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. *Mol. Biol. Evol.* 22, 2142–2146
- 17 Richards, T.A. and Cavalier-Smith, T. (2005) Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436, 1113–1118
- 18 Hausmann, S. *et al.* (2005) Yeast-like mRNA capping apparatus in *Giardia lamblia*. *J. Biol. Chem.* 280, 12077–12086
- 19 Matsutani, S. (2004) Similarities in transcription factor III C subunits that bind to the posterior regions of internal promoters for RNA polymerase III. *BMC Evol. Biol.*, DOI: 10.1186/1471-2148-4-26 (www.biomedcentral.com)
- 20 Jordan, T. *et al.* (2002) Alternative splicing of transcripts encoding Toll-like plant resistance proteins – what's the functional relevance to innate immunity? *Trends Plant Sci.* 7, 392–398
- 21 Linder, T. *et al.* (2005) A family of putative transcription termination factors shared amongst metazoans and plants. *Curr. Genet.* 48, 265–269
- 22 Pandey, R. *et al.* (2002) Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucl. Acids Res.* 30, 5036–5055
- 23 Verelst, W. and Asard, H. (2003) A phylogenetic study of cytochrome b561 proteins. *Genome Biol.*, DOI: 10.1186/gb-2003-4-6-r38 (genomebiology.com)
- 24 Anantharaman, V. *et al.* (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucl. Acids Res.* 30, 1427–1464
- 25 Koonin, E.V. *et al.* (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5, R7
- 26 Yang, S. *et al.* (2005) Phylogeny determined by protein domain content. *Proc. Natl. Acad. Sci. USA* 102, 373–378
- 27 Wolf, Y.I. *et al.* (2004) Coelomata and not ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res.* 14, 29–36
- 28 Ramirez-Parra, E. *et al.* (1999) The cloning of plant E2F, a retinoblastoma-binding protein, reveals unique and conserved features with animal G(1)/S regulators. *Nucl. Acids Res.* 27, 3527–3533
- 29 Benfey, P.N. (1999) Stem cells: A tale of two kingdoms. *Curr. Biol.* 9, R171–R172
- 30 Guitton, A.E. and Berger, F. (2005) Control of reproduction by polycomb group complexes in animals and plants. *Int. J. Dev. Biol.* 49, 707–716
- 31 Penny, D. *et al.* (2001) Mathematical elegance with biochemical realism: the covarian model of molecular evolution. *J. Mol. Evol.* 53, 711–723
- 32 Lockhart, P. *et al.* (2006) Heterotachy and tree building: a case study with plastids and eubacteria. *Mol. Biol. Evol.* 23, 40–45
- 33 Lockhart, P. and Steel, M. (2005) A tale of two processes. *Syst. Biol.* 54, 948–951
- 34 Martin, W. and Herrmann, R.G. (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* 118, 9–17
- 35 Kowallik, K.V. (1994) From endosymbionts to chloroplasts - evidence for a single prokaryotic eukaryotic endocytobiosis. *Endocyt. Cell Res.* 10, 137–149
- 36 Martin, W. *et al.* (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393, 162–165
- 37 Stiller, J.W. *et al.* (2003) A single origin of plastids revisited: convergent evolution in organellar genome content. *J. Phycol.* 39, 95–105
- 38 Martin, W. *et al.* (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12246–12251
- 39 Noutsos, C. *et al.* (2005) Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res.* 15, 616–628
- 40 Timmis, J.N. *et al.* (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5, 123–135
- 41 Lister, D.L. *et al.* (2003) DNA transfer from chloroplast to nucleus is much rarer in *Chlamydomonas* than in tobacco. *Gene* 316, 33–38
- 42 Barbrook, A.C. *et al.* (2006) Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci.* 11, 101–108
- 43 Keeling, P.J. and Palmer, J.D. (2001) Lateral transfer at the gene and subgenomic levels in the evolution of eukaryotic enolase. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10745–10750
- 44 Stiller, J.W. *et al.* (2001) Are red algae plants? a critical evaluation of three key molecular data sets. *J. Mol. Evol.* 52, 527–539
- 45 Butterfield, N.J. (2000) *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26, 386–404
- 46 Nozaki, H. *et al.* (2005) Phylogenetic implications of the CAD complex from the primitive red alga *Cyanidioschyzon merolae* (Cyanidiales, Rhodophyta). *J. Phycol.* 41, 652–657
- 47 Elser, J.J. *et al.* (2006) Signatures of ecological resource availability in the animal and plant proteomes. *Mol. Biol. Evol.* 23, 1946–1951
- 48 Kuspa, A. and Loomis, W.F. (2006) The genome of *Dictyostelium discoideum*. *Methods Mol. Biol.* 346, 15–30
- 49 Cavalier-Smith, T. (1987) The origin of fungi and pseudofungi. In *Evolutionary Biology of Fungi* (Rayner, A.D.M. *et al.*, eds), pp. 339–353, Cambridge University Press
- 50 Robbens, S. *et al.* (2005) Genome-wide analysis of core cell cycle genes in the unicellular green alga *Ostreococcus tauri*. *Mol. Biol. Evol.* 22, 589–597
- 51 MacWilliams, H. *et al.* (2006) A retinoblastoma ortholog controls stalk/spore preference in *Dictyostelium*. *Development* 133, 1287–1297