CrossMark
← click for updates

REVIEW

Towards understanding the evolution and functional diversification of DNA-containing plant organelles [version 1; referees: 3 approved]

Dario Leister[1,2]

[1]Plant Molecular Biology, Department Biology I, Ludwig-Maximilians-Universität, Planegg-Martinsried, 82152, Germany
[2]Copenhagen Plant Science Center (CPSC), University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

Abstract

Plastids and mitochondria derive from prokaryotic symbionts that lost most of their genes after the establishment of endosymbiosis. In consequence, relatively few of the thousands of different proteins in these organelles are actually encoded there. Most are now specified by nuclear genes. The most direct way to reconstruct the evolutionary history of plastids and mitochondria is to sequence and analyze their relatively small genomes. However, understanding the functional diversification of these organelles requires the identification of their complete protein repertoires – which is the ultimate goal of organellar proteomics. In the meantime, judicious combination of proteomics-based data with analyses of nuclear genes that include interspecies comparisons and/or predictions of subcellular location is the method of choice. Such genome-wide approaches can now make use of the entire sequences of plant nuclear genomes that have emerged since 2000. Here I review the results of these attempts to reconstruct the evolution and functions of plant DNA-containing organelles, focusing in particular on data from nuclear genomes. In addition, I discuss proteomic approaches to the direct identification of organellar proteins and briefly refer to ongoing research on non-coding nuclear DNAs of organellar origin (specifically, nuclear mitochondrial DNA and nuclear plastid DNA).

F1000 Faculty Reviews

This article is included in the F1000 Faculty Reviews channel.

Open Peer Review

Referee Status: ☑☑☑

	Invited Referees		
	1	**2**	**3**
version 1 published 11 Mar 2016	☑	☑	☑

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty. In order to make these reviews as comprehensive and accessible as possible, peer review takes place before publication; the referees are listed below, but their reports are not formally published.

1 **Felix Kessler**, University of Neuchâtel Switzerland

2 **William Martin**, University of Düsseldorf Germany

3 **John Allen**, University College London UK

Discuss this article

Comments (0)

Corresponding author: Dario Leister (leister@lrz.uni-muenchen.de)

Competing interests: The author(s) declare that they have no competing interests.

Introduction

The progenitors of the non-nuclear DNA-containing organelles of plants – plastids and mitochondria – were originally acquired as cyanobacterial and proteobacterial endosymbionts, respectively (reviewed in [1–4]). As they co-evolved with their host cells, the original endosymbionts lost most of their genetic repertoires, either definitively or through transfer to the host's nuclear genome. In parallel, having picked up suitable signal sequences, the products of many nuclear genes of endosymbiotic origin were re-routed back to their original compartment, together with new nucleus-encoded proteins, via intracellular trafficking routes[5–10]. As a result, complex organellar proteomes now consist of several thousand different proteins – similar in the total number of different proteins, though less so in composition, to the proteomes of their closest prokaryotic relatives.

To reconstruct the evolutionary history of plastids and mitochondria, analysis of the coding regions of the relatively small residual organellar genomes is the most straightforward approach and has helped us to understand such post-endosymbiotic events as gene loss, nuclear transfer of organellar genes, and organelle evolution in general. Moreover, coding and non-coding organellar DNA can be used as a barcode to elucidate relationships between species[11]. However, to approach the diversification of the functions of organelles in a comprehensive way, ideally their entire proteomes must be identified. Since only partial organellar proteomes can be identified by proteomics, a powerful complementation (or alternative when proteomics is impracticable) is to bioinformatically analyze the corresponding complement of their nuclear genes. This is a formidable challenge and only became feasible when entire nuclear genome sequences of plant species became available. In this review, I summarize genome-wide approaches to the definition of the protein contents of organelles, as well as interspecies comparisons of entire organellar and nuclear genomes (phylogenomics) that have contributed to our understanding of the evolution of organellar proteomes. In addition, I will discuss selected proteomic analyses of organellar proteins and briefly introduce non-coding nuclear DNA sequences of organellar origin as "by-products" of organelle evolution.

Phylogenomic approaches employing organellar DNA sequences

Traditionally, plant molecular phylogenetics has involved amplifying, sequencing, and analyzing one or a few genes from many species. Alternatively, entire genomes can be sequenced and analyzed (phylogenomics), providing much larger amounts of data per taxon but often for a smaller number of species[12]. Nowadays, ample sequence information on DNA-containing organelles is available, i.e. the ChloroMitoSSRDB database currently provides access to 2161 organellar genomes (1982 mitochondrial and 179 chloroplast genomes)[13]. Because of their small size, mitochondrial and plastid genomes from different species were the first to be analyzed by phylogenomic approaches. The outcome of such interspecific comparisons turns out to be highly dependent on the sample size. This is illustrated by two pioneering studies performed 4 years apart by the same group with a view to reconstructing plastid evolution[14,15]. In these analyses, 9 and 15 plastid genomes, respectively, were compared, and a total of 210 and 274 different protein-coding plastid genes were identified. Of these, 45 and 44,

respectively, were found in all plastid genomes in the respective set, while 44 and 117 proteins found in at least one plastid genome had nucleus-encoded counterparts in other species[14,15].

Whereas the first complete plastid DNA (ptDNA) sequences were published 30 years ago[16,17], it took a while longer for the first two plant mitochondrial genomes to be sequenced[18,19], primarily because plant mitochondrial DNAs (mtDNAs) are much larger (e.g. ~370 kbps: *Arabidopsis thaliana*) than animal mtDNAs[20,21] or ptDNAs (e.g. ~150 kbps for *A. thaliana*). Because mitochondria are common to all eukaryotes, their phylogenetic and phylogenomic analysis markedly contributed to the elucidation of the deep branching order of all eukaryotes, including protist, fungal, animal, and plant lineages (reviewed by [22]). However, in the mitochondria of land plants, frequent genomic rearrangements, the incorporation of foreign DNA from nuclear and chloroplast genomes, and peculiarities of gene expression – most notably RNA editing and *trans*-splicing – are significantly more prominent than in chloroplasts (reviewed by [23]). Furthermore, the physical organization of plant mtDNAs includes a mixture of linear, circular, and branched structures, resulting from homologous recombination – which appears to be an essential characteristic of plant mitochondrial genetic processes, both in shaping and in maintaining the genome (reviewed by [24]).

Estimating organellar proteomes
Plastids

The first publication predicting the size and evolutionary origin of the chloroplast proteome encoded in the (at that time incompletely sequenced) nuclear genome of the flowering plant *A. thaliana* identified the genes for chloroplast proteins based on the fact that their predicted products bore chloroplast transit peptides (cTPs)[25] (Table 1). The study predicted between 1900 and 2500 nucleus-encoded chloroplast proteins, of which a minimum of 35% derived from the cyanobacterial ancestor. In the entire *A. thaliana* genome sequence, 3574 (14.0%) genes coding for chloroplast proteins were identified by a prediction program[26], but the total number of cTPs obtained was not corrected for the expected numbers of false positives and negatives. Such genome-wide predictions have been repeated several times, employing different versions (with continuously improved annotation) of the Arabidopsis genome and different types or combinations of predictors (see Table 1). Interspecies comparisons of the sets of predicted chloroplast proteins have also been performed. The first such comparison published, between Arabidopsis and rice, conservatively estimated that some 2100 (*A. thaliana*) and 4800 (*Oryza sativa*) proteins carried cTPs, and defined a subset of around 900 tentative chloroplast proteins, predominantly derived from the cyanobacterial endosymbiont and with functions mostly related to metabolism, energy, and transcription, that is shared by both species[27].

As outlined above and shown in Table 1, genome-wide cTP predictions vary markedly in their outcome, depending on the type or combination of predictors used, and their sensitivity and specificity. In fact, a detailed comparative analysis of the performance of five different predictors for subcellular targeting demonstrated a disappointingly small overlap between the outcomes of different predictions. Conversely, when all predicted proteins that had been identified by at least one of the programs were considered, far

Table 1. Overview of organellar proteome size predictions and selected proteomics approaches in Arabidopsis. Note that for the predictor TPpred only the total number of 3194 Arabidopsis proteins with either chloroplast transit peptides (cTP) or mitochondrial transit peptides (mTP) was reported[68].

Approach	(Estimated) number	Reference
chloroplast		
cTP prediction (ChloroP) and correction for false positives/negatives	1900–2500	25
cTP prediction (TargetP)	3574	26
cTP prediction (TargetP)	3646	69
cTP prediction (TargetP) and correction for false positives/negatives	3130	31
cTP prediction (combination of predictors) and correction for false positives/negatives	2090	27
cTP prediction (Predotar)	1591	70
cTP prediction (TargetP)	4255	71
Mass spectrometry	690	72
Mass spectrometry + literature search	916	30
mitochondrion		
mTP prediction (TargetP)	2897	26
mTP prediction (TargetP) and correction for false positives/negatives	3135	31
mTP predictions (combinations of predictors) and correction for false positives/negatives	2957	73
mass spectrometry	416	28
mTP predictions (combination of predictors)	2955–4514	28
mTP prediction (Predotar)	1105	70

too many proteins were found to have been assigned to a specific compartment[28]. This clearly shows that predictive models inevitably involve a trade-off. Tightly constrained models which pinpoint *only* proteins that are truly located in the respective compartment (i.e. with high specificity) will fail to detect all of the proteins actually localized there (many false negatives), whereas saturated predictions that identify *most* of the truly located proteins (i.e. with high sensitivity) will also turn up many proteins that are actually destined for other compartments (many false positives). Moreover, a subset of chloroplast proteins does not contain cTPs, either because these proteins are inserted in the outer membrane or because they employ another ER-dependent pathway for targeting and import into chloroplasts (reviewed by 9,29) – although the latter fraction may well be quite small[30].

Instead of first predicting the entire set of chloroplast proteins and then analyzing their homology with proteins from other species (in particular cyanobacteria, to identify proteins derived from the original endosymbiont), one can do the reverse. In fact, a comparison of all *A. thaliana* proteins with those encoded in cyanobacterial genomes, other prokaryotic reference genomes, and yeast allowed its authors to extrapolate that ~4500 *A. thaliana* protein-coding genes had been acquired from the cyanobacterial ancestor of plastids[15] and the products of some 1300 should belong to the predicted chloroplast proteome of 3100 proteins[31]. Since then, the identity of the ancient cyanobacterial endosymbiont that gave rise to all contemporary plastids was narrowed down to the progenitors of diazotrophic cyanobacterial lineages because the gene set possessed by their modern-day representatives shows the greatest similarity to that predicted for the plastid ancestor[32].

Interspecies comparisons of nuclear genomes that do not also consider the predicted subcellular location of their products do not in themselves permit reliable conclusions regarding plastid or mitochondrial functions. However, if the species to be compared are appropriately selected, indirect but important conclusions can be drawn with respect to the protein repertoires of organelles and their evolutionary diversification. An early phylogenomic study

compared all protein-coding genes from only one plant species (*A. thaliana*) with the genes from several animals, yeasts, and combined sets of bacteria and Archaea[33] and identified 3848 plant-specific proteins, of which about 27% were predicted to localize to chloroplasts or mitochondria. In 2007, the phylogenomic comparison of several photosynthetic eukaryotes with non-photosynthetic eukaryotes, cyanobacteria, non-photosynthetic eubacteria, and Archaea enabled researchers to define sets of plant proteins with plastid-associated functions without having to depend primarily on cTP predictions[34]. The original set, the so-called *GreenCut*, comprised proteins that were conserved in the green algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri*, the moss *Physcomitrella patens*, and the flowering plant *A. thaliana*, but were absent from non-photosynthetic organisms, and consisted of 349 proteins in *C. reinhardtii*. The more restrictive *PlastidCut* (with 90 proteins in *C. reinhardtii*) was made up of *GreenCut* proteins which were also conserved in one diatom and one red alga species. In 2011, a revised version of this analysis (with *GreenCut2* and *PlastidCut2*) became available, which was based on the analysis of a larger set of sequenced genomes[35]. To qualify for *GreenCut2*, a protein must (i) have orthologs in *A. thaliana*, *P. patens*, *O. sativa*, *Populus trichocarpa*, *C. reinhardtii*, and one of the three *Ostreococcus* species with fully sequenced genomes and (ii) not have orthologs in a number of bacterial, fungal, and animal species. *GreenCut2* contained 597 *Chlamydomonas* (and 710 *Arabidopsis* orthologs due to gene duplications) and *PlastidCut2* covers 124 proteins in *C. reinhardtii*. A subset (84%) of the *PlastidCut2* proteins were experimentally localized to, or are predicted to be targeted to, the plastid and 52% of all *GreenCut2* proteins were experimentally localized to the chloroplast, implying that the majority of *GreenCut2* proteins are involved in plastid-specific functions. In line with this tentative assignment of plastid-related functions of *GreenCut* proteins, mutations in *GreenCut2* genes were sixfold over-represented in a screen for photosynthetic mutants in *C. reinhardtii* which used large-scale random insertional mutagenesis[36]. However, it is intriguing that 6% (11%) of all *PlastidCut2* (*GreenCut2*) proteins have been experimentally located in non-plastid sites.

Of the 597 *GreenCut2* proteins in *C. reinhardtii*, 105 were missing in at least one of the other green algae analyzed, and diatoms too display a reduced number of *GreenCut2* proteins. These findings suggest that (i) adaptation of green algae to specific environmental niches leads to genome specialization and/or reduction and (ii) several core plastid functions in the green lineage are either not essential or are performed by different pathways/processes in diatoms[35]. In contrast, almost all *GreenCut2* proteins are conserved in the other plant genomes analyzed, suggesting that the *GreenCut2* proteins are especially relevant to, and representative of, all land plants of the green lineage[35]. The suggestion that the extent of conservation of the *GreenCut2* inventory in a plant could serve as an indicator of a particular genome's degree of specialization might be an oversimplification[35] – at least when applied to plastid proteome complexity – because one must take account of the fact that plants contain multiple types of plastids, such that each variant might be of similar complexity to those from green algae. Indeed, analysis of chloroplast differentiation in maize, rice, and tomato reveals remarkably dynamic changes in plastid proteomes during plant development. For instance, to accommodate C_4 photosynthesis,

maize chloroplasts differentiate along the developmental axis of the leaf blade, leading from an undifferentiated leaf base into highly specialized bundle sheath (BS) and mesophyll (M) types. Hundreds of proteins detected by proteomics show differential BS/M accumulation[37], displaying five developmental transitions[38]. Analysis of etioplast-to-chloroplast differentiation in rice by proteomics has shown that etioplast metabolism is already primed to accommodate the metabolic changes that occur during the onset of photosynthesis, such that only minor metabolic network reconstruction and modification of enzyme levels occurs during the first phase of etioplast-to-chloroplast differentiation[39]. During the chloroplast-to-chromoplast transition in tomato, proteomic analyses detected a strong decrease in the abundance of proteins required for the light reactions and carbohydrate metabolism, and an increase in terpenoid biosynthesis and stress-response proteins was noted[40].

Mitochondria

The first phylogenomic approach that indirectly addressed the evolution of nuclear genes for mitochondrial proteins compared the nuclear protein-coding genes from *Saccharomyces cerevisiae* to the ones encoded by Bacteria and Archaea and found that about 75% of all yeast nuclear genes of tentatively prokaryotic origin are more similar to eubacterial than to archaebacterial homologs[41]. This suggested that the common ancestor of eukaryotes may also have possessed a majority of eubacterial genes, though it is still unclear how many of these ultimately come from the ancestral mitochondrial genome. Subsequent analysis of a sample of 27 sequenced eukaryotic and 994 sequenced prokaryotic genomes identified a set of 571 genes that was presumed to be present in the common ancestor of eukaryotes, underscoring the archaebacterial (host) nature of the eukaryotic informational genes and the eubacterial (mitochondrial) nature of eukaryotic energy metabolism[42]. A similar type of analysis indicated that gene transfer from bacteria to eukaryotes is episodic and coincides with major evolutionary transitions at the origin of chloroplasts and mitochondria[43].

Plant proteomics has also contributed to our understanding of the evolution of the mitochondrial proteome. For instance, a comparison of more than 347 mitochondrial proteins identified by proteomics in Chlamydomonas, with their homologs predicted from 354 sequenced genomes, indicated that Arabidopsis is the non-algal eukaryote most closely related to *C. reinhardtii* and that free-living α-proteobacteria belonging to the orders Rhizobiales and Rhodobacterales better reflect the gene content of the ancestor of the chlorophyte mitochondrion than parasitic α-proteobacteria do[44].

Non-coding nuclear sequences of chloroplast or mitochondrial origin

The continuous transfer of genetic material from organelles to the nucleus can result in various outcomes with respect to the functionality of the resulting nuclear sequences (reviewed in 3,45–47): (i) rarely, but with high impact on gene evolution, functional genes are generated when the transferred open reading frame recruits appropriate elements for its expression. The product of the relocated gene can then be retargeted to its original compartment or acquire new subcellular locations and functions[31]; (ii) Parts of the transferred organellar DNA can remain/become functional as material for new exons in other genes[48]; (iii) In the vast majority of cases, the

transferred organellar DNA becomes non-functional and accumulates mutations, resulting in the so-called nuclear mtDNA (NUMT) sequences (see e.g. 49–55) and nuclear ptDNA (NUPT) sequences (see e.g. 56–62). In plants, NUPTs and NUMTs can account for several hundred kbps of nuclear genomes, ranging from very small insertions to larger segments of mtDNA and/or ptDNA >100 kbps in length[63], which further facilitates study of the fate of alien DNA in the nuclear genome.

Conclusions

As yet, no single prediction program and no single proteomics experiment can accurately identify the full complement of proteins located in plastids or mitochondria. At least for model plants like *C. reinhardtii* and *A. thaliana*, a combination of predictions, large-scale fluorescence tagging, epitope tagging, proteomics of multiple subfractions of organelles, and studies of individual genes/proteins will remain the method of choice for identifying entire organelle proteomes. To this end, public and searchable databases with a web-accessible interface like SUBA3 (http://suba3.plantenergy. uwa.edu.au/)[64] and PPDB (http://ppdb.tc.cornell.edu/)[65] are now available, which integrate the results of various prediction programs of subcellular targeting proteins with large-scale proteomic datasets from cellular compartments. It needs to be remembered, however, that in the case of plants with distinct plastid variants, prediction programs will have their limitations. Here, only proteomics can reliably discriminate the diverse proteomes in the several differentiation types of plastids.

Evolutionary trees obtained by phylogenomic analyses have changed our perspective on the origin of eukaryotes by supporting hypotheses which postulate that the mitochondrial endosymbiont was acquired by an archaeon, thus placing eukaryotes within the Archaea. Therefore, phylogenomic analyses provided support for only two primary domains of life – Archaea and Bacteria – and eukaryotes arose through partnership between them (reviewed by 66). Moreover, the outcomes of phylogenomic analyses also strikingly illustrate the concept of "evolutionary tinkering"[67]. The nucleus can recruit novel exons even from "junk DNA" derived from plastids and mitochondria, and genes from cyanobacteria or proteobacteria now code in plants for many proteins that are not in their original compartment but have ended up elsewhere in the cell.

References   F | F1000 recommended

1. Zimorski V, Ku C, Martin WF, *et al.*: **Endosymbiotic theory for organelle origins.** *Curr Opin Microbiol.* 2014; **22**: 38–48.
    PubMed Abstract | **Publisher Full Text**

2. Martin W: **Evolutionary origins of metabolic compartmentalization in eukaryotes.** *Philos Trans R Soc Lond B Biol Sci.* 2010; **365**(1541): 847–55.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Kleine T, Maier UG, Leister D: **DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis.** *Annu Rev Plant Biol.* 2009; **60**: 115–38.
    PubMed Abstract | **Publisher Full Text**

4. Keeling PJ: **The number, speed, and impact of plastid endosymbioses in eukaryotic evolution.** *Annu Rev Plant Biol.* 2013; **64**: 583–607.
    PubMed Abstract | **Publisher Full Text**

5. Shi LX, Theg SM: **The chloroplast protein import system: from algae to trees.** *Biochim Biophys Acta.* 2013; **1833**(2): 314–31.
    PubMed Abstract | **Publisher Full Text**

6. Schleiff E, Becker T: **Common ground for protein translocation: access control for mitochondria and chloroplasts.** *Nat Rev Mol Cell Biol.* 2011; **12**(1): 48–59.
    PubMed Abstract | **Publisher Full Text**

7. Strittmatter P, Soll J, Bölter B: **The chloroplast protein import machinery: a review.** *Methods Mol Biol.* 2010; **619**: 307–21.
    PubMed Abstract | **Publisher Full Text**

8. Millar AH, Whelan J, Small I: **Recent surprises in protein targeting to mitochondria and plastids.** *Curr Opin Plant Biol.* 2006; **9**(6): 610–5.
    PubMed Abstract | **Publisher Full Text**

9. Jarvis P: **Targeting of nucleus-encoded proteins to chloroplasts in plants.** *New Phytol.* 2008; **179**(2): 257–85.
    PubMed Abstract | **Publisher Full Text**

10. Murcha MW, Kmiec B, Kubiszewski-Jakubiak S, *et al.*: **Protein import into plant mitochondria: signals, machinery, processing, and regulation.** *J Exp Bot.* 2014; **65**(22): 6301–35.
    PubMed Abstract | **Publisher Full Text**

11. F | Costa FO, Carvalho GR: **New insights into molecular evolution: prospects from the Barcode of Life Initiative (BOLI).** *Theory Biosci.* 2010; **129**(2–3): 149–57.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

12. F | Martin W, Deusch O, Stawski N, *et al.*: **Chloroplast genome phylogenetics:** why we need independent approaches to plant molecular evolution. *Trends Plant Sci.* 2005; **10**(5): 203–9.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

13. F | Sablok G, Padma Raju GV, Mudunuri SB, *et al.*: **ChloroMitoSSRDB 2.00: more genomes, more repeats, unifying SSRs search patterns and on-the-fly repeat detection.** *Database (Oxford).* 2015; **2015**: pii: bav084.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

14. F | Martin W, Stoebe B, Goremykin V, *et al.*: **Gene transfer to the nucleus and the evolution of chloroplasts.** *Nature.* 1998; **393**(6681): 162–5.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

15. F | Martin W, Rujan T, Richly E, *et al.*: **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.* 2002; **99**(19): 12246–51.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

16. Shinozaki K, Ohme M, Tanaka M, *et al.*: **The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression.** *EMBO J.* 1986; **5**(9): 2043–9.
    PubMed Abstract | **Free Full Text**

17. Ohyama K, Fukuzawa H, Kohchi T, *et al.*: **Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA.** *Nature.* 1986; **322**: 572–4.
    Publisher Full Text

18. Unseld M, Marienfeld JR, Brandt P, *et al.*: **The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides.** *Nat Genet.* 1997; **15**(1): 57–61.
    PubMed Abstract | **Publisher Full Text**

19. Oda K, Yamato K, Ohta E, *et al.*: **Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome.** *J Mol Biol.* 1992; **223**(1): 1–7.
    PubMed Abstract | **Publisher Full Text**

20. Gray MW: **Evolution of organellar genomes.** *Curr Opin Genet Dev.* 1999; **9**(6): 678–87.
    PubMed Abstract | **Publisher Full Text**

21. Burger G, Gray MW, Lang BF: **Mitochondrial genomes: anything goes.**

Trends Genet. 2003; **19**(12): 709–16.
PubMed Abstract | **Publisher Full Text**

22. Bullerwell CE, Gray MW: **Evolution of the mitochondrial genome: protist connections to animals, fungi and plants.** *Curr Opin Microbiol.* 2004; **7**(5): 528–34.
PubMed Abstract | **Publisher Full Text**

23. Knoop V: **The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective.** *Curr Genet.* 2004; **46**(3): 123–39.
PubMed Abstract | **Publisher Full Text**

24. Gualberto JM, Mileshina D, Wallet C, *et al.*: **The plant mitochondrial genome: dynamics and maintenance.** *Biochimie.* 2014; **100**: 107–20.
PubMed Abstract | **Publisher Full Text**

25. Abdallah F, Salamini F, Leister D: **A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*.** *Trends Plant Sci.* 2000; **5**(4): 141–2.
PubMed Abstract | **Publisher Full Text**

26. Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature.* 2000; **408**(6814): 796–815.
PubMed Abstract | **Publisher Full Text**

27. Richly E, Leister D: **An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice.** *Gene.* 2004; **329**: 11–6.
PubMed Abstract | **Publisher Full Text**

28. [F] Heazlewood JL, Tonti-Filippini JS, Gout AM, *et al.*: **Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins.** *Plant Cell.* 2004; **16**(1): 241–56.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

29. Radhamony RN, Theg SM: **Evidence for an ER to Golgi to chloroplast protein transport pathway.** *Trends Cell Biol.* 2006; **16**(8): 385–7.
PubMed Abstract | **Publisher Full Text**

30. Zybailov B, Rutschow H, Friso G, *et al.*: **Sorting signals, N-terminal modifications and abundance of the chloroplast proteome.** *PLoS One.* 2008; **3**(4): e1994.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

31. Leister D: **Chloroplast research in the genomic age.** *Trends Genet.* 2003; **19**(1): 47–56.
PubMed Abstract | **Publisher Full Text**

32. [F] Dagan T, Roettger M, Stucken K, *et al.*: **Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids.** *Genome Biol Evol.* 2013; **5**(1): 31–44.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

33. [F] Gutiérrez RA, Green PJ, Keegstra K, *et al.*: **Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants from other organisms?** *Genome Biol.* 2004; **5**(8): R53.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

34. [F] Merchant SS, Prochnik SE, Vallon O, *et al.*: **The *Chlamydomonas* genome reveals the evolution of key animal and plant functions.** *Science.* 2007; **318**(5848): 245–50.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

35. [F] Karpowicz SJ, Prochnik SE, Grossman AR, *et al.*: **The GreenCut2 resource, a phylogenomically derived inventory of proteins specific to the plant lineage.** *J Biol Chem.* 2011; **286**(24): 21427–39.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

36. Dent RM, Sharifi MN, Malnoë A, *et al.*: **Large-scale insertional mutagenesis of *Chlamydomonas* supports phylogenomic functional prediction of photosynthetic genes and analysis of classical acetate-requiring mutants.** *Plant J.* 2015; **82**(2): 337–51.
PubMed Abstract | **Publisher Full Text**

37. Majeran W, Zybailov B, Ytterberg AJ, *et al.*: **Consequences of C$_4$ differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells.** *Mol Cell Proteomics.* 2008; **7**(9): 1609–38.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

38. Majeran W, Friso G, Ponnala L, *et al.*: **Structural and metabolic transitions of C$_4$ leaf development and differentiation defined by microscopy and quantitative proteomics in maize.** *Plant Cell.* 2010; **22**(11): 3509–42.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

39. [F] Reiland S, Grossmann J, Baerenfaller K, *et al.*: **Integrated proteome and metabolite analysis of the de-etiolation process in plastids from rice (*Oryza sativa* L.).** *Proteomics.* 2011; **11**(9): 1751–63.
PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

40. [F] Barsan C, Zouine M, Maza E, *et al.*: **Proteomic analysis of chloroplast-to-chromoplast transition in tomato reveals metabolic shifts coupled with disrupted thylakoid biogenesis machinery and elevated energy-production components.** *Plant Physiol.* 2012; **160**(2): 708–25.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

41. Esser C, Ahmadinejad N, Wiegand C, *et al.*: **A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes.** *Mol Biol Evol.* 2004; **21**(9): 1643–60.
PubMed Abstract | **Publisher Full Text**

42. [F] Thiergart T, Landan G, Schenk M, *et al.*: **An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin.** *Genome Biol Evol.* 2012; **4**(4): 466–85.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

43. [F] Ku C, Nelson-Sathi S, Roettger M, *et al.*: **Endosymbiotic origin and differential loss of eukaryotic genes.** *Nature.* 2015; **524**(7566): 427–32.
PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

44. [F] Atteia A, Adrait A, Brugière S, *et al.*: **A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor.** *Mol Biol Evol.* 2009; **26**(7): 1533–48.
PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

45. Timmis JN, Ayliffe MA, Huang CY, *et al.*: **Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes.** *Nat Rev Genet.* 2004; **5**(2): 123–35.
PubMed Abstract | **Publisher Full Text**

46. Leister D, Kleine T: **Role of intercompartmental DNA transfer in producing genetic diversity.** *Int Rev Cell Mol Biol.* 2011; **291**: 73–114.
PubMed Abstract | **Publisher Full Text**

47. Bock R, Timmis JN: **Reconstructing evolution: gene transfer from plastids to the nucleus.** *Bioessays.* 2008; **30**(6): 556–66.
PubMed Abstract | **Publisher Full Text**

48. Noutsos C, Kleine T, Armbruster U, *et al.*: **Nuclear insertions of organellar DNA can create novel patches of functional exon sequences.** *Trends Genet.* 2007; **23**(12): 597–601.
PubMed Abstract | **Publisher Full Text**

49. Richly E, Leister D: **NUMTs in sequenced eukaryotic genomes.** *Mol Biol Evol.* 2004; **21**(6): 1081–4.
PubMed Abstract | **Publisher Full Text**

50. Bensasson D, Zhang D, Hartl DL, *et al.*: **Mitochondrial pseudogenes: evolution's misplaced witnesses.** *Trends Ecol Evol.* 2001; **16**(6): 314–21.
PubMed Abstract | **Publisher Full Text**

51. Tsuji J, Frith MC, Tomii K, *et al.*: **Mammalian NUMT insertion is non-random.** *Nucleic Acids Res.* 2012; **40**(18): 9073–88.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

52. Jensen-Seaman MI, Wildschutte JH, Soto-Calderón ID, *et al.*: **A comparative approach shows differences in patterns of numt insertion during hominoid evolution.** *J Mol Evol.* 2009; **68**(6): 688–99.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

53. Hazkani-Covo E, Covo S: ***Numt*-mediated double-strand break repair mitigates deletions during primate genome evolution.** *PLoS Genet.* 2008; **4**(10): e1000237.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

54. Hazkani-Covo E, Graur D: **A comparative analysis of *numt* evolution in human and chimpanzee.** *Mol Biol Evol.* 2007; **24**(1): 13–8.
PubMed Abstract | **Publisher Full Text**

55. Lopez JV, Yuhki N, Masuda R, *et al.*: ***Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat.** *J Mol Evol.* 1994; **39**(2): 174–90.
PubMed Abstract

56. Rousseau-Gueutin M, Ayliffe MA, Timmis JN: **Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution.** *Plant Physiol.* 2011; **157**(4): 2181–93.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

57. Huang CY, Grünheit N, Ahmadinejad N, *et al.*: **Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes.** *Plant Physiol.* 2005; **138**(3): 1723–33.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

58. Ayliffe MA, Scott NS, Timmis JN: **Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants.** *Mol Biol Evol.* 1998; **15**(6): 738–45.
PubMed Abstract | **Publisher Full Text**

59. Huang CY, Ayliffe MA, Timmis JN: **Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco.** *Proc Natl Acad Sci U S A.* 2004; **101**(26): 9710–5.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

60. Fuentes I, Karcher D, Bock R: **Experimental reconstruction of the functional transfer of intron-containing plastid genes to the nucleus.** *Curr Biol.* 2012; **22**(9): 763–71.
PubMed Abstract | **Publisher Full Text**

61. [F] Stegemann S, Hartmann S, Ruf S, *et al.*: **High-frequency gene transfer from the chloroplast genome to the nucleus.** *Proc Natl Acad Sci U S A.* 2003; **100**(15): 8828–33.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

62. Richly E, Leister D: **NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs.** *Mol Biol Evol.* 2004; **21**(10): 1972–80.
PubMed Abstract | **Publisher Full Text**

63. Noutsos C, Richly E, Leister D: **Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants.** *Genome Res.* 2005; **15**(5): 616–28.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

64. Tanz SK, Castleden I, Hooper CM, *et al.*: **SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis.** *Nucleic Acids Res.* 2013; **41**(Database issue): D1185–91.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

65. Sun Q, Zybailov B, Majeran W, *et al.*: **PPDB, the Plant Proteomics Database at Cornell.** *Nucleic Acids Res.* 2009; **37**(Database issue): D969–74.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

66. **F** Williams TA, Foster PG, Cox CJ, *et al.*: **An archaeal origin of eukaryotes supports only two primary domains of life.** *Nature.* 2013; **504**(7479): 231–6.
   PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

67. Jacob F: **Evolution and tinkering.** *Science.* 1977; **196**(4295): 1161–6.
   PubMed Abstract | **Publisher Full Text**

68. Indio V, Martelli PL, Savojardo C, *et al.*: **The prediction of organelle-targeting peptides in eukaryotic proteins with Grammatical-Restrained Hidden Conditional Random Fields.** *Bioinformatics.* 2013; **29**(8): 981–8.
   PubMed Abstract | **Publisher Full Text**

69. Peltier J, Emanuelsson O, Kalume DE, *et al.*: **Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction.** *Plant Cell.* 2002; **14**(1): 211–36.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

70. Small I, Peeters N, Legeai F, *et al.*: **Predotar: A tool for rapidly screening proteomes for *N*-terminal targeting sequences.** *Proteomics.* 2004; **4**(6): 1581–90.
   PubMed Abstract | **Publisher Full Text**

71. Sun Q, Emanuelsson O, van Wijk KJ: **Analysis of curated and predicted plastid subproteomes of Arabidopsis. Subcellular compartmentalization leads to distinctive proteome properties.** *Plant Physiol.* 2004; **135**(2): 723–34.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

72. Kleffmann T, Russenberger D, von Zychlinski A: **The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions.** *Curr Biol.* 2004; **14**(5): 354–62.
   PubMed Abstract | **Publisher Full Text**

73. **F** Richly E, Chinnery PF, Leister D: **Evolutionary diversification of mitochondrial proteomes: implications for human disease.** *Trends Genet.* 2003; **19**(7): 356–62.
   PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

Open Peer Review

Current Referee Status: ☑ ☑ ☑

Editorial Note on the Review Process

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

The referees who approved this article are:

Version 1

1   **John Allen**, Research Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT, UK
    Competing Interests: No competing interests were disclosed.

2   **William Martin**, Institute of Molecular Evolution, University of Düsseldorf, Düsseldorf, 40225, Germany
    Competing Interests: No competing interests were disclosed.

3   **Felix Kessler**, Laboratory of Plant Physiology, University of Neuchâtel, Neuchâtel, Switzerland
    Competing Interests: No competing interests were disclosed.