Plastids, Genomes, and the Probability of Gene Transfer

Nick Lane*

Department of Genetics, Evolution and Environment, University College London

***Corresponding author:** E-mail: nick.lane@ucl.ac.uk.

Accepted: 24 January 2011

Key words: gene transfer, plastids, genomes.

In a snowball fight, the amount of snow that sticks to your coat depends on the number and size of snowballs that hit you and the stickiness of your coat. Much the same goes for the bombardment of nuclear genomes by organellar genes, according to genome sequence data published in *GBE* this week (Smith et al. 2011). The new findings suggest that organisms with more plastids per cell have a higher probability of undergoing plastid-to-nucleus DNA transfer than organisms with only one or a few plastids per cell. The report is consistent with the "limited transfer window" hypothesis for organelle-to-nucleus gene transfer, but the ramifications extend more generally to the processes that fashion eukaryotic chromosomes.

The limited transfer window hypothesis was proposed by Barbrook et al. (2006) to explain why plastids in nonphotosynthetic organisms almost always retain a small genome. Both mitochondria and chloroplasts have lost the vast majority of their genes, through gene transfer to the nucleus, and by simple loss, retaining only those needed for the local control of chemiosmotic electron and proton transfer, according to the Colocation for Redox Regulation (CoRR) hypothesis (Allen 2003; Puthiyaveetil et al. 2008), recently backed by compelling evidence in chloroplasts (Shimizu et al. 2010). But although redox regulation is both necessary and sufficient to account for genes in chloroplasts and mitochondria, it cannot explain why plastids in nonphotosynthetic organisms retain genes. Specific biochemical reasons, such as heme synthesis and even protein synthesis for nearby mitochondria, may explain the retention of plastid DNA in particular cases (Barbrook et al. 2006), but do not offer a general explanation. The limited transfer window hypothesis does.

Many protists, including the apicomplexan parasites such as *Plasmodium* (the malarial parasite) and algae such as *Chlamydomonas*, retain a single plastid. This makes gene loss much more difficult: lysis of the single plastid is likely to be lethal to the host cell as well as the plastid. The retention of plastid genes might therefore not reflect a need so much as "an inability to get them out," as Barbrook et al. (2006) put it. This inability should be reflected not only in the retention of genes in plastids but also in a low rate of transfer of plastid genes to the nucleus—the fate of at least some DNA relinquished from lysed organelles. Although limited genomic evidence in 2006 was consistent with this prediction, Smith et al. (2011) report on 30 newly available genome sequences in diverse monoplastidic and polyplastidic species. These genome sequences unequivocally support the limited transfer window hypothesis.

The findings are not inherently surprising, but the scale of the differences is striking and fits into a larger picture of genome bombardment. Species with multiple plastids have an average of 80 times more plastid sequences incorporated into the nuclear genome (*nupts* or nuclear plastid sequences) than monoplastidic species. Not only the number but the mean length of nuclear inserts is greater in polyplastidic species. The same goes for nuclear mitochondrial sequences (*numts*), as reported for some species by Smith et al. (2011) and in a larger study of *numts* by Hazkani-Covo et al. (2010). The content of *nupts* and *numts* therefore depends in part on what amounts to the number of snowballs thrown at the target.

Two recent studies show how high this rate can be. *numts*, for example, accumulate within a single lifetime. In rats, real-time polymerase chain reaction quantification and fluorescence in situ hybridization demonstrate up to four times as many nuclear chromosomal insertions of two mitochondrial genes (COX III and 16S rRNA) in old versus young rats (Caro et al. 2010). This bombardment of mitochondrial genes may play a role in ageing, as seems to be the case in yeast. In *Saccharomyces cerevisiae*, the migration

frequency of mitochondrial DNA (mtDNA) fragments to the nucleus rises about 30-fold during the yeast chronological lifespan, which apparently contributes to ageing by promoting genomic instability (Cheng and Ivessa, 2010).

Over evolutionary time, the accumulation of *nupts* and *numts* depends on the rates of fixation in the germ line and their subsequent loss: the stickiness of the coat. Large genomes in general accumulate more noncoding DNA, and *numts* and *nupts* are no exception to this rule. According to Smith et al. (2011), there is a reasonably strong relationship between genome size and NUPT content—the forces that govern the expansion and contraction of noncoding DNA impact the accumulation of *nupts* in nuclear genomes. Likewise, Hazkani-Covo et al. (2010) reported a similar relationship for *numts*. Thus, large genomes retain more *numts* and *nupts*, whereas smaller genomes are more likely to lose them, along with other noncoding DNA.

This begs a question of wider significance to eukaryotic chromosomal dynamics. If the retention of DNA depends on the general rules governing genome size (whatever they may be), then regardless of the total amount, the proportional representation of noncoding DNA should reflect the rates of bombardment from different sources, whether organelles, endosymbionts, free-living bacteria, or other organisms. It should reflect opportunity; and here the evolutionary patterns are interesting.

First, there is no obvious relationship between organellar genome size and either *nupt* or *numt* content—the size of the snowball does not seem to matter. Presumably, this reflects the relatively small variation in organellar genome size relative to bacterial and host cell genomes generally—the rate of hits (number of organelles), and the rate of loss (host genome size), matter much more than organellar genome size (size of snowball). The one exception to this rule, as noted by Smith et al. (2011) is *Volvox carteri*, which has a "prodigious" genome of 525 kb, the largest plastid genome sequenced to date (and more than 300 kb larger than any other plastid genome in their data set). Despite being monoplastidic, *V. carteri* has accumulated more *nupts* than some polyplastidic species.

This at least hints at the possibility that snowball size matters, so long as the genome is big enough. That would almost certainly be the case in the early days of the eukaryotic cell, when the bacterial endosymbionts that became mitochondria still had genomes measured in megabases. A bombardment of giant organellar snowballs may have helped fashion eukaryotic chromosomes. The fact that 75% of eukaryotic genes that have prokaryotic sequence similarities are related to bacterial genes rather than archaeal genes (i.e., the putative host cell) is consistent with this view (Esser et al. 2004).

Second, some endosymbionts should have equal, if not more opportunity, being equally plentiful and genomically larger than most plastids and mitochondria. Until relatively recently, this did not seem to be the case, if only because bacterial genes have often been annotated out of complete genome sequences as presumed contaminants. This changed when Hotopp et al. (2007) reported widespread lateral gene transfer from *Wolbachia* to insects and nematodes, with repeated transfers ranging from nearly the entire *Wolbachia* genome (>1 Mb) to short insertions. Similar findings have been reported more widely since (Saridaki and Bourtzis, 2010). In line with transfers from organelles, the actual number of transfers, as well as the severity of symptoms, depends on the number of endosymbionts per oocyte, which can approach 500,000 in some insects (Jeong and Stouthamer 2009)—comparable with the number of mitochondria. All of this is to be expected if endosymbiotic gene transfer depends on lysis of multiple endosymbionts. However, a systematic study of *Wolbachia* inserts comparable with the *nupt* and *numt* studies discussed here has yet to be reported.

Although organellar and endosymbiotic gene transfers appear to be common and important, it is less certain how far eukaryotic genomes have been sculpted by lateral transfer from free-living bacteria and other organisms. In specific cases such as Bdelloid rotifers, bacterial genes seem to be common, especially in telomeric regions (Gladyshev et al. 2008); but in general, relatively few transfers to multicellular eukaryotes have been verified.

Even in protists, notably parasites, the role of lateral gene transfer across domains is problematic. To take one example at face value, there seem to have been many transfers of genes encoding proteins involved in anaerobic metabolism and fermentation from bacteria to anaerobic eukaryotic microbes (Hug et al. 2010). Although these cases look like lateral gene transfer there are two serious reservations. The first concerns the range of metabolisms encoded—a tiny subset of the bacterial complement, even in such restricted environments (Ginger et al. 2010). If these are genuinely lateral acquisitions, why is the same small group of genes acquired repeatedly, and independently, to the exclusion of all others, such that the entire eukaryotic domain has the metabolic capability of a single bacterium?

The simplest answer is perhaps that eukaryotes acquired all their metabolic genes from a single facultatively anaerobic bacterium, the ancestor of the mitochondria. If so, why do genes in anaerobic eukaryotes bear sequence similarities to bacterial genes in the same environment? One possible answer brings us to the second major reservation: the role of selection, specifically convergent evolution at the level of genes. The reality of convergence is attested to by the pervasiveness of epistatic (nonadditive) interactions in molecular evolution. For example, of 168 separate site-directed mutations in the *Escherichia coli* gene for isopropylmalate dehydrogenase, to match the sequence at the equivalent site in *Pseudomonas aeruginosa*, nearly 40% impaired enzyme function, "challenging a basic assumption of

molecular phylogenetics that sites in sequences evolve independently of each other" (Lunzer et al. 2010). Often they do not: selection constrains sequences at multiple sites (covarions), which gives rise to sequence similarities that are hard to distinguish from common ancestry, whether vertical or lateral.

One powerful example is the appearance of similar mutation patterns (recurrent combinations) in tumors and over human evolution (giving rise to mtDNA haplogroups), which certainly suggests selective constraints (Zhidkov et al. 2009). Plainly, genes that have a common function in a shared environment should not be taken as evidence for lateral transfer without first ruling out functional interactions.

All in all, there is little evidence that lateral gene transfer from free-living bacteria has played a major role in fashioning eukaryotic chromosomes, but DNA acquired from endosymbionts and organelles seems to be a different matter. The bombardment of genes and DNA from mitochondria and plastids probably shaped early eukaryotic evolution, through processes that can still be quantified and studied today. Opportunity is key: if fortune favors the prepared mind, endosymbiosis favors the prepared chromosome.

Literature Cited

Allen JF. 2003. The function of genomes in bioenergetic organelles. Philos Trans R Soc Lond B Biol Sci. 358:19–38.

Barbrook AC, Howe CJ, Purton S. 2006. Why are plastid genes retained in non-photosynthetic organisms? Trends Plant Sci. 11:101–108.

Caro P, et al. 2010. Mitochondrial DNA sequences are present inside nuclear DNA in rat tissues and increase with age. Mitochondrion. 10:479–486.

Cheng X, Ivessa AS. 2010. The migration of mitochondrial DNA fragments to the nucleus affects the chronological aging process of Saccharomyces cerevisiae. Aging Cell. 9:919–923.

Esser C, et al. 2004. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol Biol Evol. 21:1643–1660.

Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6(2):e1000834.

Hotopp JCD, et al. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science. 317:1753–1756.

Hug LA, Stechmann A, Roger AJ. 2010. Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. Mol Biol Evol. 27:311–324.

Ginger ML, et al. 2010. Intermediary metabolism in protists: a sequence-based view of facultative anaerobic metabolism in evolutionarily diverse eukaryotes. Protist. 161:642–671.

Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in Bdelloid rotifers. Science. 320:1210–1213.

Jeong G, Stouthamer R. 2009. Quantification of Wolbachia copy number in Trichogramma eggs (Hymenoptera: trichogrammatidae): lysozyme treatment significantly improves total gene yield from the Gram-negative bacterium. Entomol Res. 39:66–69.

Lunzer M, Golding GB, Dean AM. 2010. Pervasive cryptic epistasis in molecular evolution. PLoS Genet. 6(10):e1001162.

Puthiyaveetil S, et al. 2008. The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. Proc Natl Acad Sci U S A. 105:10061–10066.

Saridaki A, Bourtzis K. 2010. Wolbachia: more than just a bug in insects genitals. Curr Opin Microbiol. 13:67–72.

Shimizu M, et al. 2010. Sigma factor phosphorylation in the photosynthetic control of photosystem stoichiometry. Proc Natl Acad Sci U S A. 107:10760–10764.

Smith DR, Crosby K, Lee RW. 2011. Correlation between nuclear plasmid abundance and plastid number supports the limited transfer window hypothesis. Genome Biol Evol. doi: 10.1093/gbe/evr001.

Zhidkov I, Livneh EA, Rubin E, Mishmar D. 2009. mtDNA mutation pattern in tumors and human evolution are shaped by similar selective constraints. Genome Res. 19:576–580.