Organellar genes
why do they end up in the nucleus?

Many mitochondrial and plastid proteins are derived from their bacterial endosymbiotic ancestors, but their genes now reside on nuclear chromosomes instead of remaining within the organelle. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the nucleus. To become an active nuclear gene and return to the organelll, the protein must be properly folded and assembled, and in many cases assembled in an orderly manner into a larger protein complex. Finally, the nuclear copy must be properly regulated to achieve a fitness level comparable with the organelar gene. Given the complexity in establishing a nuclear copy, why do organellar genes end up in the nucleus? Recent data suggest that these genes are worse off than their nuclear and free-living counterparts because of a reduction in the efficiency of natural selection, but do these population–genetic processes drive the movement of genes to the nucleus? We are now at a stage where we can begin to discriminate between competing hypotheses using a combination of experimental, natural population, bioinformatic and theoretical approaches.

Jeffrey L. Blanchard
jeffblanchard@oregon.oregonstate.edu
Michael Lynch*
michael.lynch@oregon.oregonstate.edu

*Correspondence
National Center for Genome Resources, 2935 Nokes Park Drive East, Santa Fe, NM 87505, USA
Biology Department, University of Oregon, Eugene, OR 97403, USA

Endosymbiosis, the persistence of one organism within another, is a result of the fusion of two organisms into one. At least two common eukaryotic organelles, the mitochondrion and the plastid, are derived from bacteria whose fate became linked with unicellular eukaryotes some two billion years ago. One of the most obvious features of organelle evolution has been the reduction in genetic autonomy that follows endosymbiosis. Some organellar genes became expendable in the internal environment of the host, and nuclear genes have replaced the function of some organellar genes, but much of the reduction occurred through the transfer of organellar genes to the nucleus.1,2 Many genes might have been transferred to the nucleus early on, but we now realize that there are different rates of gene transfer and loss in eukaryotic lineages.3,4 The most extreme cases lie within the cradle of eukaryotic diversity, the protists. The mitochondrial genome of the retromonad protozoan, Reclinomonas americana, codons for 97 genes5 and the plastid genome of the red alga,
**Steps in time – building a functional organellar gene in the nucleus**

The process of gene transfer begins with the movement of copies of organellar DNA to the nucleus (Fig. 1). DNA can apparently move from the mitochondria into the nucleus at the rate of $5 \times 10^{-6}$ events per cell per generation in yeast. However, most of the mitochondrial DNA remained extrachromosomal and less than 1% of this DNA inserted into the yeast nuclear genome. Because many of the insertions would disrupt nuclear genomes (73% of the genome is protein coding in yeast), the integration rate is likely to be much less than the insertion rate. Surveys of yeast nuclear genomic sequence found few copies of mitochondrial DNA, none comprising a complete mitochondrial genome.

As in yeast, most organellar-derived sections found in nucleus of plants and animals are fragments of genes often less than 130 bp (Refs 12, 14). However, multigene blocks of plastid and/or mitochondrial genomes exist in plant and animal nuclear genomes, and might provide fodder for the transfer process. In felines, a 7.9 kb mitochondrial segment is found in a tandem array of 38–76 repeats, and the sequencing of Arabidopsis chromosome II turned up a 270 kb segment comprising 75% of the large Arabidopsis mitochondrial genome. Multigene chunks of mitochondrial DNA have also been uncovered by the Human Genome Sequencing Project (Ref. 12 and J.L. Blanchard, unpublished). However, because of the extensive editing of plant chloroplast and mitochondrial sequences, and because of changes in the animal mitochondrial genetic code it is not clear whether these large blocks will lead to new genes in the nucleus. Recent gene transfers in plants appear to have been through an RNA intermediate.

Thus, an important parameter for determining the transfer rate is not simply the quantity of organellar-derived DNA in the nucleus, but the potential of the segment to give rise to a functional gene.

We also need to consider that some groups, like vertebrates and plants, have a much larger proportion of non-coding DNA, and that these groups appear to have more organellar-derived DNA in their nuclear genomes than compact genomes like those of Saccharomyces cerevisiae and Caenorhabditis elegans. It is not clear if this is a result of an increase in the likelihood that the insertion would disrupt a functional nuclear gene in compact genomes or is a result of an increased rate of deletion of noncoding DNA as is apparent in Drosophila. Other impediments of gene transfer and a discussion of ‘why organelles have retained genomes’ have been reviewed recently.

**Cases studies of natural and experimental gene transfer**

Once the organellar gene is assimilated into the nuclear genome, there are still a number of steps to becoming an active nuclear gene and returning to the organelle as a functional protein. The gene must be expressed and its protein product must acquire a transit peptide to allow access to the organelle. As the copy number of organellar genomes and organelles can be quite high in some tissues, simply providing enough of the protein is likely to be a chore for the nucleus. Finally, the problem comes of coordinating the expression to different light levels or energy needs. The complexity of this process will be dependent on the individual gene. Although it seems an insurmountable challenge to move an organellar gene to the nucleus, the
challenge has been met both naturally and through human intervention.

Following the origin of land plants, at least eleven mitochondrial genes (coxII, coxIII, rps2, rps7, rps30, rps11, rps12, rps14, rps19, rpl6, adhC) and one plastid gene (rpl22) have made a successful trek to the nucleus\(^1\)\(^2\)\(^3\). Kodawaki and colleagues have recently developed two remarkable gene transfer stories. In both cases the mitochondrial gene acquired the N-terminal region of a pre-existing mitochondrial targeted protein, but the pathways were different. A plant mitochondrial gene became functional in the plant nucleus after integration through a recombination–duplication event, positioning the N-terminal region of a cytochrome oxidase subunit in front of the rpl11 protein-coding region\(^4\). A second copy of the rpl1 gene is also found in the nucleus but with the N-terminal region of a mitochondrial ATPase subunit. The mitochondrial copy of rps14 gene integrated into an intron in the nuclear gene encoding sucinate dehydrogenase and now both proteins are produced in the nucleus through alternative splicing in rice\(^5\). A similar gene organization and alternative splicing pattern was discovered independently in maize\(^6\). In both of these cases the plant organelle gene has become an active nuclear copy by acquiring the N-terminal region of a pre-existing organelle targeted gene, thus providing an instant solution to expression, targeting, and regulatory problems.

Palmer et al. have recently developed a framework for studying the gene transfer process in action\(^7\). In legumes, the integration and activation of the mitochondrial coxII gene appears to be closely linked, as in the two previous examples. Intense sampling of 392 legume genera revealed that in some legume lineages the mitochondrial copy has been lost or become nonfunctional, other lineages retained the mitochondrial copy but the nuclear copy has become nonfunctional, and in some lineages both copies are expressed at varying ratios. Further investigation is under way to test whether there is tissue-specific expression in the plants that express both copies. Differential expression of nuclear and mitochondrial copies of the atp9 gene has been observed in Neurospora (Ref. 30). The possibility also exists of regulation at the translational level (both the nuclear and organelar transcripts might exist but only one copy makes a functional protein) or at the protein level (targeting, modification, proper assembly into the cytochrome oxidase complex).

The above natural transfer pathways are similar to the approaches taken by scientists to experimentally manipulate organelle genes into the nucleus in fungal and plant systems. In fungi, the mitochondrial genetic code is different from the nuclear code and the mitochondrial genes need to be modified substantially to create an active nuclear copy. In spite of these difficulties, a nuclear transgene was constructed with roughly 70% of the wild-type activity\(^7\). In plants, the process is somewhat simpler. For the tobacco rbcL gene, the plastid gene was inserted into an expression cassette containing the upstream targeting sequence and regulatory regions of the nuclear rbcS gene\(^8\). The rbcL nuclear transgene complements the plastid null mutant, but has only 3–10% of the wild-type levels of RUBISCO activity. The rbcL gene is highly expressed in the wild-type plant under photosynthetic conditions, and as plants contain multiple copies of the nuclear rbcS gene (whose product interacts with a one to one stoichiometry with the RBCL protein) a single copy might not be sufficient to restore wild-type activity.

Transfer of genetic information into organelles

Although the transfer of genes is highly polar, organelar genomes are not refractory to exogenous DNA. Plant mitochondrial genomes have long been known to harbor partial and complete copies of plastid and nuclear genes and the recent sequencing of the Arabidopsis mitochondrial genome has revealed a wonderful mixture of exogenous sequences\(^9\). In the genome can be found 16 sections of plastid DNA, fragments of nuclear genes, retrotransposons and sequences similar to plant-pathogen RNA viruses: 49% of the genome is still unaccounted for. The integration of plastid DNA has resulted in six functional tmRNAs. Thus, the Arabidopsis mitochondrial RNA repertoire now includes 12 "native" tmRNAs, six plastid derived tmRNAs (now mitochondrial encoded), and probably another 10 or so that are imported from the nucleus (and derived from eukaryotic type nuclear tmRNAs)\(^10\). There are also well-documented insertions of introns into plastid and mitochondrial genomes (for a recent example see Ref. 34).
Drift, bottlenecks and the ratchet

The image of asexual organellar genomes rising up to many hundreds of copies per cell then crashing down to a few copies during germ-line divisions invokes analogies with Muller’s ratchet: the progressive loss of fitness in finite populations because of genetic drift. If organellar genomes are effectively haploid and inherited unparentally, then they will be more subject to random genetic drift than nuclear genes, because of a fourfold decrease (twofold because it is haploid and twofold because of uniparental inheritance) in their effective population size. Furthermore, although there is recent evidence for recombination between mitochondrial and bacterial genomes in natural populations, organellar genes still appear to be more vulnerable to genetic drift through linkage disequilibrium effects. Because organelles are a linked set of genes, fixation of a beneficial mutation might result in a deleterious mutation hitchhiking along at another site, or the rate of fixation of a beneficial mutation might be slowed by deleterious mutations at other sites. The important consequence of these population-genetic processes is that a mutation will be less visible to natural selection (more subject to genetic drift) in an organellar gene than in an identical nuclear gene.

One ‘click’ of Muller’s ratchet corresponds to the loss of the original genotype and the fixation of a less fit genotype. The click rate of the ratchet depends on the effective population size, mutation rate, recombination rate and the distribution of mutational effects. One way in which this can be visualized experimentally is by picking a bacterial colony at random and streaking it onto a new plate. With each transfer, the population goes through a bottleneck of one and the result over time is a steady accumulation of deleterious mutations.

Muller’s ratchet has been cited as playing a role in the transfer of organellar genes to the nucleus, reduction in organellar and bacterial genome sizes, changes in animal mitochondrial genetic code, occurrence of mRNA editing and a role for organellar genomes in extinction processes. An important assumption in invoking a ratchet-like process is that the population cannot regain its original fitness because compensatory mutations are rare and have small positive mutational effects.

Direct comparison of nuclear, organellar and bacterial tRNA structures suggests that organellar tRNAs are less thermodynamically stable than their nuclear counterparts and eubacterial ancestors, and a greater variance in the organellar tRNA loop and stem regions indicates that more sites are subject to drift. Thus, the relatively greater width of the selective sieve in organelles (Fig. 3) is not simply a balance of deleterious and compensatory mutations that results in no net change in fitness or a faster recycling of the same sites. The reduction in fitness from Muller’s ratchet is a very long-term process and is not likely to endanger species on timescales of less than a million years.

How does the organelle gene become superfluous?

Does a successful replacement event depend on the fitness of the nuclear copy starting near the fitness of the organellar gene? It is likely that, as in the experimental systems, the nuclear copy would initially be less fit than an established organellar gene. One solution requires that both the organellar and nuclear copy become essential until the nuclear copy becomes as fit as the organellar gene. This could happen through one of three different routes. (1) The establishment of the nuclear copy could be a beneficial event, making up for a suboptimal organellar gene. This would increase the integration rate as the gene would be fixed by positive selection instead of through drift. (2) Both genes might become essential, owing to the accumulation of deleterious mutations in one or both copies to the extent that the loss of one copy is not tolerated (Fig. 2). A similar model of deleterious mutation accumulation has been proposed to explain the persistence of genes following genome duplications. (3) A dosage-compensation mechanism triggered by the import of the nuclear protein could lead to the downregulation of the organellar gene and/or protein. Once both copies become essential (at least temporarily), the population-genetic processes, such as the efficiency of selection on beneficial or deleterious mutations, will determine the winner.
Other organelles, including bacteriocytes (non-cell autonomous organelles found in many insects), show signs of decreased rRNA stability and high rates of substitution relative to their free-living and nuclear counterparts. Selection might be relaxed against some organelar genes, as in nonphotosynthetic plastids, but some simple calculations suggest that the reduced fitness is primarily a result of the smaller effective organelar population sizes. However, theoretical models suggest that the germ-line bottleneck, instead of facilitating Muller’s ratchet, can act to slow any ratchet-like process by increasing genetic variability in the population thereby making selection more effective.

Some organelar genes do not show signs of the accumulation of deleterious mutations, but are subject to greater genetic drift than their nuclear counterparts. Plant chloroplast tRNAs are indistinguishable in their stem stability from their nuclear counterparts, but have wider sieve widths (Fig. 3). Therefore, in plant chloroplast tRNAs, deleterious mutations might be fixed by drift followed by compensatory mutations. The result is that relatively more organelar sites are subject to drift, but with no net change in fitness. Thus, although genetic drift is a component of Muller’s ratchet, increased genetic drift does not imply a role for Muller’s ratchet.

How quickly is the race run?

Several groups have compared mutation and substitution rates of organelar genes with either functional or non-functional copies in the nucleus. In each study, the organelle sequence in the nucleus now conform to the nuclear mutation rate and pattern. The rate at which genes become silenced is dependent on the mutation rate. If all else is equal, the copy in the genome with the lower mutation rate will persist longer. In plants the nuclear mutation rate is orders of magnitude higher than the mitochondrial mutation rate, thus the nuclear copy should be inactivated much more frequently than the mitochondrial copy if the process is driven solely by the mutation rate. Therefore, we would expect to see few, if any, plant mitochondrial gene transfer events, which is in contradiction to the many events previously referenced. Furthermore, Adams et al. have reported a similar number of separate mitochondrial and nuclear coxII gene inactivations within legumes. Thus, the mitochondrial copy seems to be losing more often than the mutation rates would dictate. Further investigation of the legume coxII system should provide valuable information on rates of synonymous versus non-synonymous substitutions, and should provide a test of whether the mitochondrial and/or nuclear copy is subjected to elevated drift.

Selection for smaller organelar genomes either at the organismal or intraorganelar level could tilt the scales towards the nuclear copy by speeding up the rate at which an organelar deletion is fixed. Large mitochondrial deletions can confer a replicative advantage to the smaller mitochondrial genome, in spite of the negative consequences for the host. Case studies of gene transfer in plants suggest that a mitochondrial gene can become non-functional due to a frameshift mutation and/or the formation of stop codons. Some of these nonfunctional organelar genes are still expressed and edited, and some have no obvious defect in the protein-coding sequence. Thus, the first step in nonfunctionalization of an organelar gene often does not involve a large deletion. However, point mutations, like a single substitution in a tRNA gene, can put the mutant mitochondrial genome at a replicative advantage.

In plants, because of the high nuclear mutation rate, a frameshift mutation or other types of gene-inactivating mutations would probably occur within a few million years. Thus, both the ‘integration happens’ and the small-genome-size models predict that the race would be run in or under a few million years. However, the process of gene transfer for coxII and the plastid rpl22 gene has been ongoing for at least 60 million years (but these estimates might need downwards revision; see Ref. 29). Why is the process taking so long? It is likely that, as in the experimental systems, the nuclear copy would initially be less active than an established organelar gene and the loss of the organelar copy would be selected against. Thus, the models mentioned above that require the retention of both copies would need to be invoked.

FIGURE 3. Accelerated substitution rates in organelar genomes

The relative width of the selective sieve is the ratio of organelar or nuclear substitutions standardized by their respective mutation rates. The ratio of the organelle to nuclear selective sieve widths is shown. In all categories, the ratio is greater than one, indicating the mutations in organelar genes are relatively less visible to selection. This is in spite of the tremendous variation in substitution rates that run from an order of magnitude higher in vertebrate mitochondria to an order of magnitude lower in plant mitochondria relative to their respective nuclear genes. Thus, mutation rates are not the determinant of whether a gene is relatively more prone to genetic drift. Data are taken from Refs 45-47. The black, white and gray bars are rRNA, tRNA and protein-coding genes, respectively.
Organellar genes

During the course of time when both copies are present, an organellar gene might lose fitness because of drift, setting the stage for compensatory mutations to come in through the nuclear gene as in Fig. 2. Then beneficial mutations could accrue in the nuclear copy to a level where the organellar gene can be lost with little or no consequence to the host. Thus, a high mutation rate in the nuclear copy (as in plants) might no longer be a handicap, but might hasten the generation of beneficial mutations.

**Time to test hypotheses**

Currently, we cannot provide a satisfactory answer to why the nucleus is the preferred location for organellar genes or why there are different rates of gene transfer in different taxon groups. In the next few years, we can expect progress in defining and estimating the parameters needed to explain why organellar genes end up in the nucleus. Genome projects should allow us to estimate the number of organellar DNA insertions, and allow us to determine whether flanking sequence motifs of DNA insertions are derived primarily from pre-existing genes. As there are many recent examples of gene transfer in plants, surveys of plant species should reveal the time periods over which the organellar gene and the nuclear copy are expressed.

Experimental approaches can be used to test directly for dosage compensation. There is also a need to develop theoretical models, and to test the role of deleterious and beneficial mutations in organellar evolution. Together these efforts will help provide answers to a question that is often too speculatively.

---

**References**