Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates

Jeffrey D. Palmer*, Keith L. Adams, Yangrae Cho†, Christopher L. Parkinson‡, Yin-Long Qiu§, and Keming Song¶

Department of Biology, Indiana University, Bloomington, IN 47405

We summarize our recent studies showing that angiosperm mitochondrial (mt) genomes have experienced remarkably high rates of gene loss and concomitant transfer to the nucleus and of intron acquisition by horizontal transfer. Moreover, we find substantial lineage-specific variation in rates of these structural mutations and also point mutations. These findings mostly arise from a Southern blot survey of gene and intron distribution in 281 diverse angiosperms. These blots reveal numerous losses of mt ribosomal protein genes but, with one exception, only rare loss of respiratory genes. Some lineages of angiosperms have kept all of their mt ribosomal protein genes whereas others have lost most of them. These many losses appear to reflect remarkably high (and variable) rates of functional transfer of mt ribosomal protein genes to the nucleus in angiosperms. The recent transfer of *cox2* to the nucleus in legumes provides both an example of interorganellar gene transfer in action and a starting point for discussion of the roles of mechanistic and selective forces in determining the distribution of genetic labor between organellar and nuclear genomes. Plant mt genomes also acquire sequences by horizontal transfer. A striking example of this is a homing group I intron in the mt *cox1* gene. This extraordinarily invasive mobile element has probably been acquired over 1,000 times separately during angiosperm evolution via a recent wave of cross-species horizontal transfers. Finally, whereas all previously examined angiosperm mtDNAs have low rates of synonymous substitutions, mtDNAs of two distantly related angiosperms have highly accelerated substitution rates.

The evolutionary dynamics of plant mitochondrial (mt) genomes have long been known to be unusual compared with those of animals and most other eukaryotes at both the sequence level (exceptionally low rate of point mutations) and structural level (high rates of rearrangement, duplication, genome growth and shrinkage, and incorporation of foreign DNA). The rate of synonymous substitutions (a useful approximation of the neutral point mutation rate) was shown in the 1980s to be lower in angiosperm mitochondria than in any other characterized genome, and fully 50–100 times lower than in vertebrate mitochondria (1, 2). This gulf largely persists despite the more recent discovery of modest substitutional rate heterogeneity within angiosperms (3, 4) and vertebrates (5, 6).

Angiosperms have by far the largest mtDNAs, at least 200 kb to over 2,000 kb in size (larger than some bacterial genomes) (7, 8). These genomes grow and shrink relatively rapidly; for example, within the cucumber family, mt genome size varies by more than six-fold (9). Plant mitochondria rival the eukaryotic nucleus (especially the plant nucleus) in terms of the C-value paradox they present: i.e., larger plant mt genomes do not appear to contain more genes than smaller ones, but simply have more spacer DNA (intron content and size also do not vary significantly across angiosperms). This paradox extends to plant/animal comparisons. For example, the one sequenced angiosperm mt genome (from *Arabidopsis*; refs. 10 and 11) is 367 kb in size yet contains only one more RNA gene and twice the number of protein genes (27 vs. 13) than our own mt genome, which is over 20 times smaller (16.6 kb). Angiosperm mtDNAs are large in part because of frequent duplications. These most commonly result in small (2–10 members) repeat families whose elements range up to a few hundred base pairs in size, although large duplications and triplications of up to 20 kb are almost always found at least once and sometimes several times within a genome. Angiosperm mtDNAs also grow promiscuously via the relatively frequent capture of sequences from the chloroplast and nucleus (8, 10, 11). The functional significance of this foreign DNA seems entirely limited to chloroplast-derived tRNA genes, which provide many of the tRNAs used in plant mt protein synthesis (12).

Recombination between the small and large repeats scattered throughout angiosperm mtDNAs creates a very dynamic genome structure, both evolutionarily and in real time. Recombination between repeats of about 2 kb and larger is so frequent as to create a dynamic equilibrium in which an individual plant's mtDNA exists as a nearly equimolar mixture of recombinational isomers differing only in the relative orientation of the single copy sequences flanking the rapidly recombining repeats (7, 13). Plants such as maize, with many different sets of these large, usually direct "recombination repeats" somehow manage to perpetuate their mt genomes despite their dissolution into a bewildering complexity of subgenomic molecules via repeat-mediated deletion events (13, 14). Recombination between smaller repeats appears to occur less frequently, although perhaps frequently enough to help maintain a reservoir of low-level, rearranged forms of the genome (termed "sublimons") that persist together with the main mt genome. On an evolutionary time-scale, recombination between short dispersed inverted repeats generates large inversions frequently enough to scramble gene order almost completely, even among relatively closely members of the same genus (2, 14). The combined forces of

This paper was presented at the National Academy of Sciences colloquium "Variation and Evolution in Plants and Microorganisms: Toward a New Synthesis 50 Years After Stebbins," held January 27–29, 2000, at the Arnold and Mabel Beckman Center in Irvine, CA.

Abbreviation: mt, mitochondrial.

*To whom reprint requests should be addressed at: Department of Biology, 1001 East Third Street, Indiana University, Bloomington, IN 47405. E-mail: jpalmer@bio.indiana.edu.

†Present address: Stanford Genome Center, 855 California Avenue, Palo Alto, CA 94304.

‡Present address: Department of Biology, University of Central Florida, Orlando, FL 32816-2368.

§Present address: Department of Biology, University of Massachusetts, Amherst, MA 01003-5810.

¶Present Address: Sigma Chemical Company, 3300 South Second Street, St. Louis, MO 63118.

frequent duplication and inversion have led to the fairly common creation of novel, chimeric genes in plant mitochondria. A number of these chimeric genes lead to cytoplasmic-nuclear incompatibilities manifest as cytoplasmic male sterility (13, 15).

The above picture, encapsulated as the title of a 1988 paper (2)—"Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence"—with its corollary that animal mtDNA evolves oppositely in all respects, was largely complete by the late 1980s (7). This picture was derived from comparison, at both fine- and broad-scale taxonomic levels, of a relatively small number of angiosperms, belonging primarily to but five economically important families [the crucifers (Brassicaceae), the cucurbits (Cucurbitaceae), the legumes (Fabaceae), the grasses (Poaceae), and the nightshades (Solanaceae)], and with two genera (*Brassica* and *Zea*) serving as exemplars (2, 14). The completion of the *Arabidopsis* mt genome sequence in 1997 (10, 11) gave a more fine-grained and comprehensive picture of the extent to which foreign DNA uptake and internal duplication have influenced the size, structure, and evolutionary potential of a particular mt genome, but without significantly changing prior notions of the structural dynamics of plant mt genomes.

A few years ago, we set up a large-scale Southern blot survey, of 281 diverse angiosperms, to better elucidate the evolution of two fundamental classes of mt features—their content of genes and of introns—that were poorly characterized relative to the traits whose evolution is described above. Our hope was that by surveying hundreds of diverse plants, we could discern the basic tempo and pattern of gene and intron loss and gain and, furthermore, identify especially attractive candidates for follow-up study to learn about the processes and mechanisms underlying these kinds of structural changes. An additional, utilitarian goal, whose achievements (see, e.g., ref. 16) will be not be described in this report, was to use these presence/absence characters to help unravel the phylogeny of angiosperms and other land plants. This paper will summarize our results, both recently published and unpublished, on gene content and intron content evolution in angiosperm evolution. We show that rates of gene loss, of accompanying gene transfer to the nucleus, and of intron acquisition by cross-species horizontal transfer can be remarkably high for particular classes of these genetic elements, and that these rates also vary substantially across lineages of flowering plants. As a completely unexpected bonus, these surveys have also led to the discovery of two exceptional groups of plants with vastly elevated rates of synonymous substitutions.

Southern Blot Survey for Changes in mt Gene and Intron Content. We went to some lengths to sample angiosperm diversity, extracting total DNAs from 281 species that represent 278 genera and 191 families of angiosperms (species listed at www.bio.indiana.edu/~palmerlab). Twelve sets of pseudoreplicate filter-blots were made, each set containing one digest (with either *Bam*HI or *Hin*dIII) of each of the 281 DNAs. The digested DNAs were arranged according to the presumptive phylogenetic relationships of their cognate plants as understood about 5 years ago. To date, the sets of blots have been sequentially hybridized with nearly 100 different probes, mostly for segments of various mt genes and introns, but also for several chloroplast genes and introns.

Virtually all mt genes and introns tested hybridized well across the full spectrum of angiosperms examined, and some even hybridized well across additional blots containing the full diversity of land plants, a roughly 450-million-year time span (see, e.g., ref. 16). The success of these hybridizations, carried out at moderate stringencies [washes at 60°C in 2× standard saline citrate (SSC)/0.1% SDS], across such large timespans testifies to the very low substitution rates of the great majority of plant mtDNAs. A probe sequence was inferred to be absent from the mt genome of a particular filter-bound preparation of total DNA if there was no detectable hybridization on an overexposed autoradiograph against two layers of controls: good hybridization to the DNA in question using other mt probes and good hybridization to other DNAs with the probe in question.

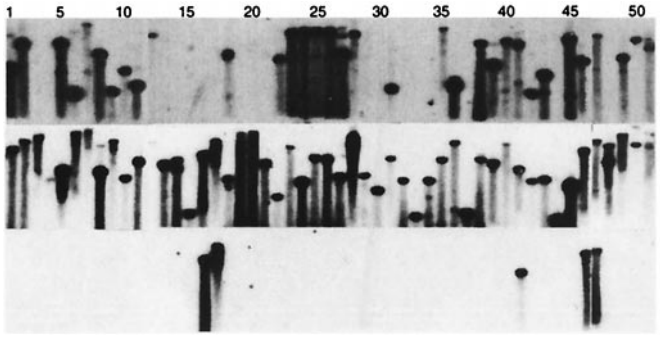

Fig. 1. Southern blot survey illustrating three distinct presence/absence patterns of mitochondrial genes and introns. *Bam*HI-cut total DNAs from 51 of 281 angiosperms surveyed were arranged according to presumptive phylogenetic relationships, were electrophoresed, and were blotted and hybridized with probes internal to the *rps7* gene (*Upper*; gene mostly present, with several losses evident), the *cox1* coding sequence (*Middle*; gene universally present), and the *cox1* intron (*Lower*; intron rarely present, each presence thought to reflect an independent acquisition by horizontal transfer).

Fig. 1 shows examples of the three general categories of results obtained with the various mt probes used. Many probes, such as rRNA probes and the *cox1* exon probe used in the middle panel of Fig. 1, hybridized strongly to essentially all DNAs tested; i.e., the lane-to-lane variations in hybridization intensity were reproducible across all probes in this category. We interpret these variations as primarily reflecting differences in amount of mtDNA loaded per lane, and conclude that each mtDNA probably contains an intact copy of the sequence probed (see the penultimate section for an explanation of the weak *cox1* hybridization in lane 4 of the middle panel). Many other probes, such as the *rps7* gene probe used in the top panel of Fig. 1, while hybridizing strongly to many lanes (normalizing intensities to the category 1 results described above), hybridized at best only very weakly (again normalizing) and often not at all to many other lanes (i.e., lanes 3–4, 13–17, 19–21, 29–30, 32–34, 37, 44, 48). We conclude that most or all of the *rps7* probe region is probably missing from these mt genomes. The third category of results was obtained with but a single probe, the *cox1* intron shown in Fig. 1 *Bottom*. This probe gave a singularly patchy, sporadic hybridization pattern, whose molecular basis will be explained in a later section.

Ribosomal Protein Genes Are Lost Frequently, Respiratory Genes Only Rarely. The survey blots were hybridized with probes for each of the 14 ribosomal protein genes known from angiosperm mt genomes, and with probes for 11 of the 21 known respiratory genes. The cases of inferred gene absences were plotted onto a multigene phylogeny of the surveyed angiosperms (17) to estimate the number of phylogenetically separate gene losses. A total of only two losses were inferred among 10 of the respiratory genes; these include the previously described loss of mt *cox2* in the legume *Vigna* (18–21) and the loss of *nad3* in the Piperaceae. The small respiratory gene *sdh4* was an exception, with about 10 separate losses inferred (K.A., Y.-L.Q., and J.D.P., unpublished data). In striking contrast to the respiratory genes as a group, probes for all 14 ribosomal protein gene probes failed to hybridize to mtDNAs of many, disparately related angiosperms (see, e.g., Fig. 1 *Upper*), suggesting numerous gene losses (at least 10 losses for most genes, over 200 losses in total). The losses vary in phylogenetic depth, with most being limited to one or two related families, whereas several encompass many related families or even orders (K.A., Y.-L.Q., and J.D.P., unpublished data). Probes for *rps2* and *rps11* did not hybridize to the lanes of

most higher eudicots (a group comprising 182 of the 281 angiosperms in our survey), suggesting relatively ancient losses early in the evolution of eudicots. Both *rps2* and *rps11* have been isolated from the nucleus of *Arabidopsis* (ref. 22; *rps11* expressed sequence tags from the GenBank database), suggesting that gene loss followed functional transfer to the nucleus. The relatively few losses of these two genes (4 and 6, respectively) reflect the reduced potential for many ("subsequent") losses when such an ancient loss occurs.

Our blot surveys will not detect mt pseudogenes unless much or all of the probe region is missing, and thus our survey probably underestimates the number of gene losses. Several ribosomal protein pseudogenes have been reported in angiosperm mt genomes, for example, of *rps14* and *rps19* in *Arabidopsis* (10, 11, 23). In some cases, weak hybridization by a probe was observed to DNA from a species known to contain only a fragment of a gene in the mitochondrion (e.g., *rps12* in *Oenothera*).

The hybridization results show that ribosomal protein gene content in angiosperm mitochondrial genomes varies considerably, as suggested previously (10, 20), although the magnitude of the variation and frequency of gene loss are unexpectedly high. The high number of ribosomal protein gene losses compared with the low number of respiratory gene losses is also striking. In some angiosperm lineages, the rate of ribosomal protein gene loss appears to be comparable to, or even higher than, the silent substitution rate (K.A., Y.-L.Q., and J.D.P., unpublished work), whereas in many other lineages, including the most ancient ones, there has been no loss at all. These latter lineages have retained all 14 ribosomal proteins that were present in the common ancestor of angiosperms, whereas several fairly recently arisen angiosperm lineages have lost 10 or more of the 14 genes (in one case, apparently all 14). The rate of ribosomal protein gene loss thus varies enormously across angiosperms lineages; some factor(s) must have triggered a rapid rate of loss in certain recent lineages.

The mt gene losses detected by our survey could be explained by functional transfer of the gene to the nucleus, by functional substitution by another protein (see, e.g., ref. 23), or by the protein being dispensable in certain plants. Six ribosomal protein genes (22–28) that are present in the mitochondrion of many flowering plants, along with the respiratory gene *cox2* (18–21), have been reported to have been transferred to the nucleus. Thus, the most likely explanation for loss of a gene from the mitochondrion is transfer to the nucleus.

Do Multiple Gene Losses Reflect Multiple Gene Transfers? Assuming that most of the genes lost from the mitochondrion have been transferred to the nucleus, then the many separate losses of each mt ribosomal protein gene and of *sdh4* could reflect either an equivalent number of separate transfers, each more or less coincident in time with the loss, or a smaller number of earlier transfers (as few as one for each gene), with several or all losses stemming from the same ancestral transfer of a particular gene. The latter, early-transfer/multiple-dependent-loss model predicts a prolonged period of retention of dual intact and expressed genes in both compartments after gene transfer. This model seems inconsistent with both theory (29, 30) and with empirical results (see the next section) indicating fairly rapid loss of one compartment's gene or the other after gene transfer/duplication. On the other hand, the one-loss = one-transfer model and other multiple transfer models seem unlikely considering the complex series of events required for each successful functional transfer (i.e., reverse transcription, movement to the nucleus, chromosomal integration, and functional activation, which in almost all cases requires acquisition of sequences conferring both proper expression and also targeting of the now cytoplasmically synthesized protein to the mitochondrion).

Nuclear sequences for each of three mitochondrially derived ribosomal protein genes have been reported from two separate lineages of mt gene loss as defined by our blot survey, while we have been studying the transferred *rps10* gene in a number of angiosperms. These genes provide a useful starting point for investigating the number and timing of gene transfers during angiosperm evolution. In the *rps14* loss lineage that includes maize and rice, the transferred gene is located within an intron of the *sdh2* gene, and the *sdh2* targeting sequence is alternatively spliced to *rps14* transcripts (27, 28), whereas nuclear *rps14* in *Arabidopsis* shares none of these features (31). Rice *rps11* was duplicated in the nucleus after transfer but before targeting sequence acquisition (26), with the two *rps11* genes having acquired their targeting sequences from two different nuclear genes for mt proteins (*atpB* and *coxVb*), whereas the targeting sequence of pea *rps11* (32) has no similarity to any sequences currently in the databases. Finally, nuclear *rps19* genes of *Arabidopsis* (23) and soybean (expressed sequence tags in the GenBank database), both rosids, also have unrelated targeting sequences and structures. The dissimilar structural features of members of each pair of these three genes strongly suggests that each was derived from a separate activation event. Because activation probably occurs relatively soon after transfer, before the nuclear gene is permanently disabled by mutation (refs. 21 and 33; see next section), we think that these ribosomal protein genes were not only independently activated but also independently transferred to the nucleus. Considering that *rps14*, *rps11*, and *rps19* have been lost from the mt genomes of many different angiosperm lineages, as revealed by our Southern blot survey, it is possible that each gene has been independently transferred many times, not just twice. Indeed, we have recently obtained evidence for many, recent independent transfers of *rps10*, which has been lost from the mt genome over 20 times among the 281 angiosperms surveyed by blots (K.A., D. Daley, Y.-L.Q., J. Whelan, and J.D.P., unpublished work). It is increasingly evident, therefore, that functional transfer of mt genes to the nucleus occurs at a surprisingly high frequency in angiosperm evolution, especially, it would appear, in some groups of flowering plants (see preceding section).

Mitochondrial Gene Transfer in Action: Recent Transfer of *cox2* to the Nucleus in Legumes. The most extensively studied example of recent mt gene transfer in flowering plants (or any group of eukaryotes) is the cytochrome oxidase subunit 2 gene (*cox2*) in legumes (18–21). *Cox2*, present in the mitochondrion of virtually all plants, was transferred to the nucleus during recent legume evolution (21). Examination of nuclear and mt *cox2* presence and expression in over 25 legume genera has revealed a wide range of intermediate stages in the process of mt gene transfer, providing a portrayal of the gene transfer process in action (Fig. 2; 21). *Cox2* was transferred to the nucleus via an edited RNA intermediate (18, 19). Once nuclear *cox2* was activated, a state of dual intact and expressed genes—of transcompartmental functional redundancy—was established; this transition stage persists most fully (i.e., with both compartments' *cox2* genes highly expressed in terms of steady-stated, properly processed RNAs; COX2 protein levels have not been assayed) only in *Dumasia* among the many studied legumes. Four other, phylogenetically disparate legumes also retain intact and expressed copies of *cox2* in both compartments, but with only one of the two genes expressed at a high enough level, in the one tissue type examined thus far, to presumably support respiration (Fig. 2). Silencing of either nuclear or mt *cox2* has occurred multiple times and in a variety of ways, including disruptive insertions or deletions, cessation of transcription or RNA editing, and partial to complete gene loss (21). Based on phylogenetic evidence, we infer that mt *cox2* and nuclear *cox2* have been silenced approximately three to five times each during the evolution of the studied legumes (Fig. 2). Although *cox2* in legumes is the only known example of gene inactivation after recent transfer and activation in the nucleus, a comparative phylogenetic approach

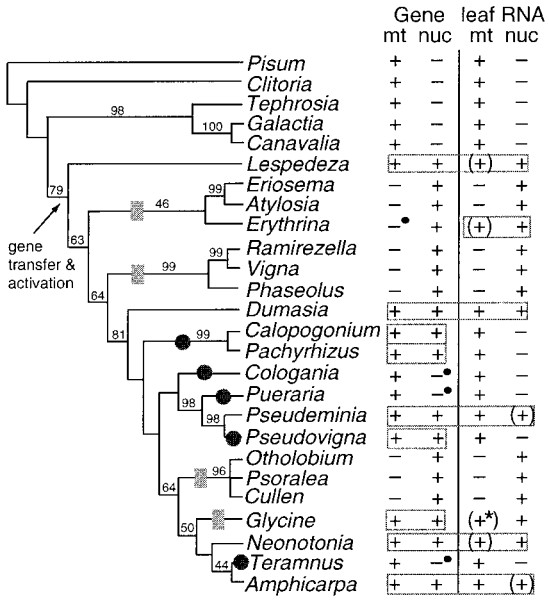

Fig. 2. Summary of legume *cox2* gene distribution and expression data in a phylogenetic context. The left two columns indicate the presence (+) or absence (−) of an intact *cox2* gene in the mitochondrion (mt) or nucleus (nuc) of the indicated species. Bullets indicate genes containing small insertions or deletions that disrupt the reading frame or intron splicing. The right two columns indicate the presence (+) or absence (−) of detectable mitochondrial and nuclear *cox2* transcripts in young leaves. Parentheses indicate transcripts present at low levels; the asterisk indicates transcripts that are not properly edited. Boxing highlights dual intact genes and/or dual transcription and proper processing (of dual *cox2* genes, intact or not) in a given plant. Light rectangles and dark circles indicate loss or silencing of mt and nuclear *cox2*, respectively. The phylogenetic tree is one of three equally parsimonious trees obtained from parsimony analysis of a data set consisting of 2,154 bp of two chloroplast gene sequences (*rbcL* and *ndhF*) and 557 chloroplast restriction sites. Bootstrap values above 40% are shown. The figure is modified from ref. 21.

might reveal that the nuclear copy of other recently transferred organelle genes has become inactivated in one or more species related to the single plant studied so far.

Roles of Selection and Chance in Mitochondrial Gene Transfers During Angiosperm Evolution. All but the last step (gene loss) in the complicated and evolutionarily unidirectional process by which mt genes move to the nucleus and disappear from the mitochondrion may be driven largely by mechanistic forces and chance mutations. These prior steps include reverse transcription (which could also occur after either of the next two steps), exit from the mitochondrion, entry into the nucleus, integration into the nuclear genome, gain of a nuclear promoter and other elements conferring properly regulated expression, and gain of a mt targeting sequence. Nucleic acids could escape from the mitochondrion by several mechanisms, as thoroughly discussed by Thorsness and Weber (33). The rate of mtDNA escape and uptake by the nucleus has been estimated to be relatively high (33, 34). Once inside the nucleus, nucleic acids can integrate into the nuclear genome by double-strand break repair, as shown recently for yeast (35), and perhaps by other mechanisms. Clues have been revealed as to the mechanisms of gene activation and targeting sequence acquisition, including gain of a targeting sequence (and perhaps upstream promoter and other regulatory elements) from a preexisting gene for a mt protein by a shuffling process (ref. 26; our unpublished data) and by integration into a preexisting gene for a mt protein (27, 28).

After the nuclear copy of a transferred gene is activated and gains a mt targeting sequence, both genes must be expressed at least transiently, as described above for *cox2* in legumes (Fig. 2; ref. 21). It is possible that the genes in both genomes could become fixed. Both nuclear and mt *atp9* genes have been retained in *Neurospora crassa* (36) and *Aspergillus nidulans* (37, 38), and both are functional at certain times during the life cycle of *Neurospora* (39). However, the most commonly observed outcome is that one gene (usually the mt gene, although this is influenced by sampling biases) becomes silenced and lost. Inactivation of the nuclear copy of a transferred gene results in a failed transfer, but the opportunity for repeated "attempts" at transfer can create a gene transfer ratchet (40). Both selection and chance factors may play a role in determining which gene is retained and which gene is inactivated. Our finding of approximately equal numbers of *cox2* silencings in legume mt and nuclear genomes raises the possibility that it is largely random as to which gene became silenced in a given species, with disabling mutations inactivating either *cox2* gene at comparable frequencies. Alternatively, if the rate of production of disabling mutations is higher in one genome or the other, then the equal numbers of silencings would reflect selection favoring the gene's retention in the high-rate genome. This is difficult to assess because, although we know that substitution rates are much higher in legume nuclear than mt genes (1), we do not know what the overall rate of disabling mutations is in either genome.

Several hypotheses have been proposed for selection favoring retention of the nuclear copy of a transferred and activated organelle gene and loss of the organellar copy. Presence of a gene in the nucleus allows for crossing over during meiosis, perhaps enabling beneficial mutations to be fixed more rapidly than in asexual organelle genomes (41). Alternatively, the progressive accumulation of detrimental mutations in asexual mt genomes by Muller's ratchet may favor transfer of genes to the nucleus. Evidence for Muller's ratchet has been found in tRNA genes of animal mitochondria (42) and in the genomes of endosymbiotic bacteria (43). However, the rate of nucleotide substitutions is very low in plant mitochondria (about 10-fold lower than in the nucleus), which should counterbalance the effects of Muller's ratchet (44, 45) and negate (for plants) the hypothesis (46) that a nuclear location is favored because it provides relief from the effects of oxygen free radical damage incurred by organellar genes. Selection for a small, compact genome, although perhaps operating in other eukaryotes, is unlikely to be a factor favoring continued gene transfer in plants, because plant mt genomes readily incorporate and retain foreign DNA (8, 10, 11, 47, 48) and are very large and mostly noncoding (7–11). Finally, there is the possibility that genes for some organellar proteins may be better regulated in the nucleus (33). Although this possibility is intriguing, we are unaware of any evidence to support or refute it.

Why are a few protein genes preferentially retained by mt genomes across all or most eukaryotes? One view is that the products of these genes, all of which function in respiration, are highly hydrophobic and difficult to both import into mitochondria and properly insert (posttranslationally) into the inner mt membrane (see, e.g., refs. 33, 49, and 50). Evidence for this includes experiments in which cytoplasmically synthesized cytochrome *b*, a highly hydrophobic protein with eight transmembrane helices, could not be imported in its entirety, with successful import limited to regions comprising only three to four transmembrane domains (51). In general, genes whose products have many hydrophobic transmembrane domains are usually located in the mitochondrion whereas genes whose products have few such domains are more often transferred to the nucleus (49, 52). Indeed, the only two protein genes contained in all of the many completely sequenced mt genomes encode what are by some criteria (51) the two most hydrophobic proteins present in the mitochondrion, cytochrome *b* and subunit 1 of cytochrome oxidase (52, 53). Although the hydrophobicity hypothesis cannot account for the distribution of every mt gene in every eukaryote,

it seems likely to be a factor favoring the retention of certain respiratory genes. A second hypothesis for retention of certain genes in organelles is that their products are toxic when present in the cytosol or in some other, inappropriate cellular compartment to which they might be misrouted after cytosolic synthesis (54). Although this toxicity hypothesis is eminently testable, we are unaware of any empirical evidence for it. A third hypothesis for organellar gene retention is to allow their expression to be directly and quickly regulated by the redox state of the organelle (ref. 55; reviewed in ref. 45). Evidence for redox regulation of organellar gene expression has been reported for chloroplasts (56) but not, to our knowledge, for mitochondria.

Although selective factors may be responsible for the transfer of some genes to the nucleus and the retention of others in the mitochondrion, chance factors may also be at work. At the stage of dual expression, whether the nuclear or mt copy of a transferred gene is retained may in some cases depend solely on the roll of the evolutionary dice—on which gene first sustains a gene-inactivating mutation, or a mutation that is either deleterious or beneficial to the gene product's function. In the latter two cases, selection would be involved in the sense that it would act to either fix the gene with the beneficial mutation or eliminate the gene with the deleterious mutation.

We are left with a picture of organelle gene transfer as a complex, historically contingent process whose outcome undoubtedly depends on a combination of mechanistically driven factors and chance mutations, together with selective forces. The process seems to be driven by the high rate of physical duplication of organelle genes into the nucleus (which appears to be true for all eukaryotes, regardless of whether functional gene transfer is still occurring), and proceeds seemingly exclusively in one direction: from organelles to nucleus. Indeed, with one disputed exception, a *mutS* homolog in coral mt DNA (57, 58), there are no examples known of the reverse process, of functional genes moving from the nucleus to the mitochondrion or chloroplast.

Why has gene transfer been so pervasively unidirectional? Flowering plant mitochondria are certainly able to accept foreign sequences: Numerous examples are known of the uptake of chloroplast DNA (8, 10, 11, 47), nuclear DNA (10, 11, 59), and sequences from other organisms (refs. 48 and 60; see below), and a few chloroplast-derived genes are expressed in the mitochondrion (12, 61, 62). Nonetheless, the initial driving force (the rate of physical transfer/duplication of sequences from one genome into the other) may be much stronger toward the nucleus than in the reverse direction; certainly this seems to be the case for yeast by several orders of magnitude (33, 34). Compounding this, each mt gene physically transferred to the nucleus can potentially result in functional transfer whereas only a small fraction of nuclear genes could play a useful role if transferred to the mitochondrion. The pervasively unidirectional flow of mt genes to the nucleus may, therefore, be driven largely, perhaps even entirely, by a huge imbalance in the relative likelihood of gene movement and potential functionality in one direction versus the other.

Explosive Invasion of Plant Mitochondria by a Group I Intron. Thus far, we have discussed intracellular horizontal evolution entirely as a means of relocating plant mt genes to the nucleus. As mentioned in the introduction, plant mt genomes are also well known to acquire foreign sequences by intracellular gene transfer, from both the chloroplast and nucleus. We have recently described (48, 63), and will briefly review here, a case of horizontal evolution that stands out in three respects: (*i*) It is the first case of cross-species acquisition of DNA by plant mt genomes; (*ii*) it is unparalleled with respect to how frequently the same piece of DNA has been acquired, over and over again, during angiosperm evolution; and (*iii*) all of these many invasions have occurred very recently, as an explosive wave within the last 10 million years or so.

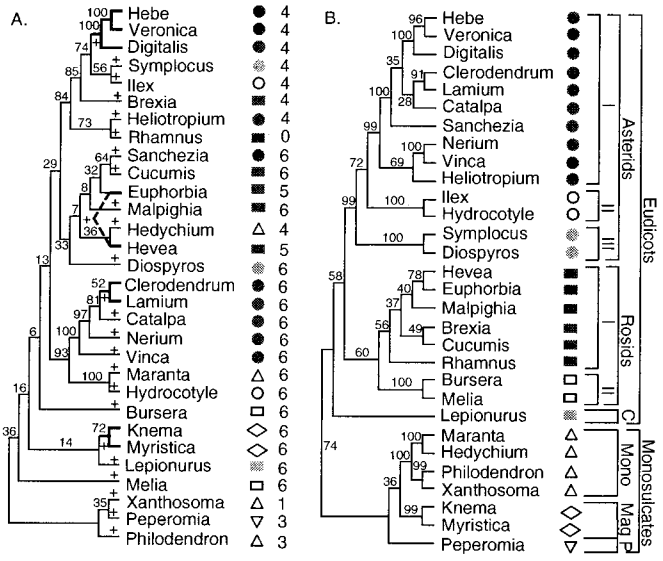

Fig. 3. Phylogenetic evidence for horizontal transfer of *cox1* introns. (*A*) Maximum-likelihood tree of 30 angiosperm *cox1* introns. Numbers on the tree are bootstrap values. Plus signs on the tree mark 25 inferred gains of the intron among these taxa. Symbols to the immediate right of names are as in Fig. 3*B*. Numbers at far right indicate number of 3′-flanking nucleotides changed by co-conversion (see text). Bold branches mark four small clades of introns thought to have originated from the same intron gain event. (*B*) Organismal tree from a maximum-likelihood analysis of a combined data set of chloroplast *rbcL* and mt *cox1* coding sequences. Numbers are bootstrap values. Symbols mark the nine major groups of angiosperms represented in this analysis. The figure is modified from ref. 48.

The piece of DNA in question here is a homing group I intron. These introns encode site-specific endonucleases with relatively long target sites that catalyze their efficient spread from intron-containing to intron-lacking alleles of the same gene in genetic crosses. A few cases of the evolutionary spread of these introns by horizontal homing between species were known when, in 1995, we in collaboration with Jack Vaughn's group reported (60) that the angiosperm *Peperomia* had acquired, quite recently (64), a group I intron in its mt *cox1* gene by long-distance horizontal transfer, most likely from a fungus. We subsequently discovered a closely related form of this intron, located at the same position in *cox1*, in a very distantly related angiosperm, *Veronica*. This stimulated us to use the *Veronica* intron as a probe against our survey blots of 281 angiosperm DNAs. As shown in Fig. 1 *Lower*, the intron probe hybridized strongly to relatively few DNAs, in an unusually patchy manner phylogenetically [and always to the same band as a *cox1* exon probe (Fig. 1 *Middle*), indicating that the hybridizing region is always located in the same gene]. All told, 48 of the 281 angiosperm DNAs on the blots, scattered across most of the major groups represented, hybridized well to the intron (48).

The exceptionally patchy phylogenetic distribution of the intron (see Fig. 2 of ref. 48) caused us to sequence the intron from 30 diverse intron-containing taxa. We then compared the congruence of intron (Fig. 3*A*) and "organismal" (Fig. 3*B*) phylogenies to assess the relative contributions of vertical and horizontal transmission to the intron's evolutionary history in angiosperms. These phylogenies are highly incongruent. From this, we concluded that the intron had been independently acquired, by cross-species horizontal transfer, many times separately among the examined plants. For example, consider the closely related rosids *Bursera* and *Melia*, whose intron-hybridizing DNAs are in adjacent lanes in Fig. 1 (lanes 46 and 47; recall that DNAs are arranged according to presumptive phylogenetic order in these blots) and which group with 100% bootstrap support in the organismal tree of Fig. 3*B*. Their *cox1* intron

sequences do not, however, group together (Fig. 3*A*), suggesting that *Bursera* and *Melia* acquired their introns independently of one another. Three, more convincing pairs of examples of phylogenetic evidence for independent acquisition consist of *Ilex*/*Hydrocotyle*, *Symplocus*/*Diospyros*, and *Maranta*/*Hedychium*. Each pair again receives 100% bootstrap support in Fig. 3*B*, and in each case the two members of the pair are now separated by multiple, well supported nodes in the intron tree (Fig. 3*A*).

All told, we inferred at least 32 separate cases of intron gain to account for the intron's presence in the 48 angiosperms revealed to contain the intron by the 281-taxa Southern blot survey (48). Some 25 of these cases are marked on Fig. 3*A* by plus signs, while 7 additional gains were inferred by criterion *ii*, as we shall now describe. Overall, the inferences of independent intron gain were based on four criteria: (*i*) the many incongruencies, some strongly supported, some less so, between intron and organismal phylogenies (Fig. 3); (*ii*) the highly disjunct phylogenetic distribution of intron-containing plants; (*iii*) different lengths of co-conversion tracts among otherwise related introns (Fig. 3); and (*iv*) the existence of ancestrally intron-lacking taxa within families containing the intron. This last form of evidence also relates to co-conversion, the process by which donor exonic sequences flanking the intron replace recipient exonic sequences when the intron is inserted into the cox1 gene. Space limitations preclude any meaningful discussion of the complicated logic behind the two criteria that are largely or entirely based on co-conversion tract evidence; the interested reader is instead referred to refs. 48 and 63.

More extensive sampling within the monocot family Araceae showed that 6 of the 14 Araceae sampled contain the intron and that these 6 taxa probably acquired their introns by at least 3 and quite possibly 5 separate horizontal transfers (48, 63). In addition, unpublished studies from our lab and that of Claude dePamphilis reveal many more cases of independent gain of this promiscuous group I intron. Given that we have still sampled only a tiny fraction of the >300,000 species of angiosperms, we are confident that the intron has been horizontally acquired at least hundreds of times during angiosperm evolution and probably over 1,000 times. Equally remarkably, all of these transfers seem to have occurred very recently, in the last 10 million years or so of angiosperm evolution.

Many fascinating questions can be asked about the evolution of this wildly invasive group I intron. What does the inevitably complex historical network of horizontal transfers look like; i.e., who is the donor and who is the recipient in each specific instance of intron transfer? Phylogenetic evidence suggests at least one, perhaps initiating long-distance transfer of the intron from a fungus to a flowering plant (48). Have many or most of the transfers occurred via this long-distance route, in which case all of the fungi must themselves be closely related [because all of the plant introns are (48)]? Or have most transfers, perhaps all but the first, occurred via "short"-distance transfer, i.e., from angiosperm to angiosperm? These two alternative models make contrasting phylogenetic predictions as elaborated elsewhere (48). Has transfer, especially if largely plant-to-plant, been mediated by vectoring agents, and if so which ones (e.g., viruses, bacteria, aphids, mycorrhizal fungi, etc.)? Or has it occurred by transformation-like uptake of DNA from the environment or by the occasional direct fusion (perhaps pollen-mediated) of two unrelated plants? To some extent, the answers are probably yes, yes, and yes; considering the large number of independent transfers, each a unique and rare (except on the evolutionary timescale) historical event, almost any imaginable kind of vector and method of intron transfer could have been used at least once. Why has the intron burst on the angiosperm scene in such a rampant manner only so recently? Has this recent wave of lateral transfers been triggered by some key shift in the intron's invasiveness within angiosperms, and if so, what has changed?

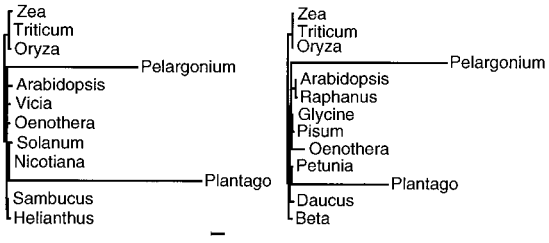

Fig. 4. Accelerated evolution of mitochondrial *cob* (left tree) and *cox2* (right tree) genes in *P. hortorum* and *P. rugelii*. Trees are from a maximum likelihood analysis of third codon positions only (344 positions for *cob* and 116 for *cox2*). Tree topologies were constrained to match current views of phylogenetic relationships for the organisms whose genes are analyzed. Trees are shown at the same scale; scale bar indicates 0.1 substitution/nucleotide. *Pelargonium* and *Plantago* sequences are from our unpublished data, and the rest are from the GenBank database.

We hope to provide at least partial answers to some of these fascinating but challenging questions over the coming years.

In passing, we note that the overwhelmingly horizontal evolution of this remarkable group I intron is in striking contrast to the vertical pattern of evolution of the 23 other introns in angiosperm mt genomes, all of which are group II introns (10). We have used probes for 11 of these introns in our Southern blots surveys (ref. 16; Y.-L.Q. and J.D.P., unpublished data). All 11 are present in most or all major groups of angiosperms, and in many other groups of vascular plants, and thus were clearly present in the common ancestor of all angiosperms. These group II introns appear to have been transmitted in a strictly vertical manner, including occasional to frequent losses.

Highly Accelerated Substitution Rates in Two Lineages of Plants. As already emphasized, our wide-scale Southern blot survey for presence or absence of mt genes and introns is predicated entirely on the uncommonly low rates of nucleotide substitutions observed to date in plant mitochondria. If a lineage of plants were to sustain for very long a radically higher substitution rate (say at the 50- to 100-fold higher level characteristic of mammalian mitochondria), then all of its mt genes might hybridize poorly or not at all, as if the genome no longer existed. Poor hybridization with all mt probes tested was observed for two of the 281 angiosperms on our blots, *Pelargonium hortorum* (the common garden geranium) (Fig. 1, lane 4) and *Plantago rugelii* (plantain, a common lawn weed); this was despite normal loadings of total DNA in these two lanes and strong hybridization with all chloroplast probes used. To explore this, portions of several mt protein and rRNA genes were PCR-amplified from both taxa and sequenced (Y.C., C.L.P., Y.-L.Q., and J.D.P., unpublished work). In all cases, the genes are exceptionally divergent. Most critically, in the case of the protein genes, most of the enhanced divergence is confined to synonymous sites. This indicates that the neutral point mutation rate, the rate of occurrence of nucleotide substitutions irrespective of selection, is markedly enhanced in both genera. To illustrate this effect for *cob* and *cox2*, Fig. 4 shows phylogenetic trees constructed with third codon positions only. Most third position changes are silent, and therefore the extremely long branch lengths leading to the *Pelargonium* and *Plantago* sequences in each tree graphically illustrate the high point mutation rate in these two distantly unrelated angiosperms.

Analysis of chloroplast and nuclear gene sequences from both plants indicates that these two genomes are not undergoing accelerated evolution (consistent with the strong hybridization of chloroplast probes mentioned above). Thus, the mutation rate increases in these two plants are restricted to their mitochondrial genomes. This distinguishes these cases of rate variation from those reported by Eyre-Walker and Gaut (3), in which all three genomes of grasses were shown to exhibit higher rates of synonymous substitution than in palms. Another distinction is

the magnitude of the rate variation: Grasses and palms differ only several-fold in their (plant-wide) substitution rates, whereas *Pelargonium* and *Plantago* show elevated (mt-specific) rates some 50–100 times higher than normal, putting them on a par with the very rapidly evolving mt genomes of mammals. Furthermore, analysis of several other species from the *Pelargonium* (Geraniaceae) and *Plantago* (Plantaginaceae) families shows a range of enhanced mt divergences in both families, as if sequential increases in the mt mutation rate had occurred during their evolution (Y.C., C.L.P., and J.D.P., unpublished work).

The magnitude and recency of these mutation rate shifts appear to be unprecedented for evolutionary lineages of species (as opposed to the well known but ephemeral mutator strains of laboratory mutant cultures, of wild strains of bacteria, or of human colon cancers). It remains to be seen whether the same sorts of underlying mechanisms are involved, such as changes in the fidelity and efficacy of DNA replication and mismatch repair (65).

Postscript. Plant mt genomes continue to spring marvelous evolutionary surprises. The discovery that certain angiosperm groups are rapidly moving a large set of mt ribosomal proteins to the nucleus seems remarkable in two contexts: first, that they still have these so "easily transferred" genes left to transfer after a roughly 2-billion-year period of mt existence; second, that animals lost all of these ribosomal protein genes at least 0.6 billion years ago (i.e., before they became animals) and that there has been absolutely no functional gene transfer within the long period of metazoan evolution (plants still do it, animals don't!). For reasons as yet unfathomable, rates of functional gene transfer appear to vary hugely across lineages and over time. The very recent and explosive burst of *cox1* intron invasions into angiosperm mt genomes and the discovery of unprecedentedly large increases in the mt point mutation rate in two groups of angiosperms also speak to the surprising fluidity of the forces that control the rates of all manner of classes of mutations. These discoveries pave the way for more reductionist studies aimed at elucidating molecular mechanisms underlying these striking evolutionary patterns and rate changes. They also point to the opportunity afforded by microarray technology to mine new veins of molecular evolutionary gold by scaling up by orders of magnitude the Southern blot approach so successfully used thus far.

We thank Jeff Doyle, Jane Doyle, Peter Kuhlman, Jackie Nugent, Phil Roessler, and Andy Shirk for various contributions and Jeff Blanchard, Dan Daley, Will Fischer, Patrick Keeling, and Jim Whelan for helpful discussions. This study was supported by National Institutes of Health Research Grant GM-35087 to J.D.P., U.S. Department of Agriculture Plant Biotechnology Fellowship 95-38420-2214 to K.L.A., and National Institutes of Health Postdoctoral Fellowships GM-17923 and GM-19225 to Y.L.Q. and C.L.P., respectively.

1. Wolfe, K. H., Li, W.-H. & Sharp, P. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 9054–9058.
2. Palmer, J. D. & Herbon, L. A. (1988) *J. Mol. Evol.* **28,** 87–97.
3. Eyre-Walker, A. & Gaut, B. S. (1997) *Mol. Biol. Evol.* **14,** 455–460.
4. Laroche, J., Li, P., Maggia, L. & Bousquet, J. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5722–5727.
5. Martin, A. P., Naylor, G. J. P. & Palumbi, S. R. (1992) *Nature (London)* **357,** 153–155.
6. Waddell, P. J., Cao, Y., Hasegawa, M. & Mindell, D. P. (1999) *Syst. Biol.* **48** 119–137.
7. Palmer, J. D. (1990) *Trends. Genet.* **6** 115–120.
8. Palmer, J. D. (1992) in *Plant Gene Research: Cell Organelles*, ed. Herrmann, R. G. (Springer, Vienna), pp. 99–133.
9. Ward, B. L., Anderson, R. S. & Bendich, A. J. (1981) *Cell* **25,** 793–803.
10. Unseld, M., Marienfeld, J. R., Brandt, P. & Brennicke, A. (1997) *Nat. Genet.* **15,** 57–61.
11. Marienfeld, J., Unseld, M. & Brennicke, A. (1999) *Trends Plant Sci.* **4,** 495–502.
12. Miyata, S., Nakazono, M. & Hirai, A. (1998) *Curr. Genet.* **34,** 216–220.
13. Mackenzie, S., He, S. & Lyznik, A. (1994) *Plant Physiol.* **105,** 775–780.
14. Fauron, C. M.-R., Moore, B. & Casper, M. (1995) *Plant Sci.* **112,** 11–32.
15. Hanson, M. R. (1991) *Annu. Rev. Genet.* **25,** 461–486.
16. Qiu, Y.-L., Cho, Y., Cox, J. C. & Palmer, J. D. (1998) *Nature (London)* **394,** 671–674.
17. Soltis, P. S., Soltis, D. E. & Chase, M. W. (1999) *Nature (London)* **402** 402–404.
18. Nugent, J. M. & Palmer, J. D. (1991) *Cell* **66,** 473–481.
19. Covello, P. S. & Gray, M. W. (1992) *EMBO J.* **22,** 3815–3820.
20. Nugent, J. M. &Palmer, J. D. (1993) in *Plant Mitochondria*, eds. Brennicke, A. & Kuck, U. (VCH, New York), pp. 163–170.
21. Adams, K. L., Song, K., Roessler, P. G., Nugent, J. M., Doyle, J. L., Doyle, J. J. & Palmer, J. D. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 13863–13868.
22. Perrotta, G., Grienenberger, J. M. & Gualberto, J. M. (1998) in *Plant Mitochondria: From Gene to Function*, eds. Moller, I. M., Gardestrom, P., Glimelius, K. & Glaser, E. (Backhuys, Leiden, the Netherlands), pp. 37–41.
23. Sanchez, H., Fester, T., Kloska, S., Schroder, W. & Schuster, W. (1996) *EMBO J.* **15,** 2138–2149.
24. Grohmann, L., Brennicke, A. & Schuster, W. (1992) *Nucleic Acids Res.* **20,** 5641–5646.
25. Wischmann, C. & Schuster, W. (1995) *FEBS Lett.* **375,** 152–156.
26. Kadowaki, K., Kubo, N., Ozawa, K. & Hirai, A. (1996) *EMBO J.* **15,** 6652–6661.
27. Figueroa, P., Gómez, I., Holuigue, L., Araya, A. & Jordana, X. (1999) *Plant J.* **18,** 601–609.
28. Kubo, N., Harada, K., Hirai, A. & Kadowaki, K. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9207–9211.
29. Marshall, C. R., Raff, E. C. & Raff, R. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12283–12287.
30. Herrmann, R. G. (1997) in *Eukaryotism and Symbiosis*, eds. Schenk, H. E. A., Herrmann, R. G., Jeon, K. W., Müller, N. E. & Schwemmler, W. (Springer, Vienna), pp. 73–118.
31. Figueroa, P., Gomez, I., Carmona, R., Holuigue, L., Araya, A. & Jordana, X. (1999) *Mol. Gen. Genet.* **262,** 139–144.
32. Kubo, N., Harada, K. & Kadowaki, K. (1998) in *Plant Mitochondria: From Gene to Function*, eds. Moller, I. M., Gardestrom, P., Glimelius, K. & Glaser, E. (Backhuys, Leiden, the Netherlands), pp. 25–27.
33. Thorsness, P. E. & Weber, E. R. (1996) *Int. Rev. Cytol.* **165,** 207–233.
34. Thorsness, P. E. & Fox, T. D. (1990) *Nature (London)* **346,** 376–379.
35. Ricchetti, M., Fairhead, C. & Dujon, B. (1999) *Nature (London)* **402,** 96–100.
36. van den Boogaart, P., Samallo, J. & Agsteribbe, E. (1982) *Nature (London)* **298,** 187–189.
37. Brown, T. A., Ray, J. A., Waring, R. B., Scazzocchio, C. & Davies, R. W. (1984) *Curr. Genet.* **8,** 489–492.
38. Ward, M. & Turner, G. (1986) *Mol. Gen. Genet.* **205,** 331–338.
39. Bittner-Eddy, P., Monroy, A. F. & Brambl, R. (1994) *J. Mol. Biol.* **235,** 881–897.
40. Doolittle, W. F. (1998) *Trends Genet.* **14,** 307–311.
41. Blanchard & Lynch, M. (2000) *Trends Genet.*, in press.
42. Lynch, M. (1996) *Mol. Biol. Evol.* **13,** 209–220.
43. Moran, N. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2873–2878.
44. Martin, W. & Herrmann, R. G. (1998) *Plant Physiol.* **118,** 9–17.
45. Race, H. L., Herrmann, R. G. & Martin, W. (1999) *Trends Genet.* **15,** 364–370.
46. Allen, J. F. & Raven, J. A. (1996) *J. Mol. Evol.* **42,** 482–492.
47. Nugent, J. M. & Palmer, J. D. (1988) *Curr. Genet.* **14,** 501–509.
48. Cho, Y., Qiu, Y.-L., Kuhlman, P. & Palmer, J. D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14244–14249.
49. Popot, J.-L. & de Vitry, C. (1990) *Annu. Rev. Biophys. Biophys. Chem.* **19,** 369–403.
50. Palmer, J. D. (1997) *Science* **275,** 790–791.
51. Claros, M. G., Perea, J., Shu, Y., Samatey, F. A., Popot, J.-L. & Jacq, C. (1995) *Eur. J. Biochem.* **228,** 762–771.
52. Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T. G., *et al.* (1998) *Nucleic Acids Res.* **26,** 865–878.
53. Gray, M. W. (1999) *Curr. Opin. Genet. Dev.* **9,** 678–687.
54. Martin, W. & Schnarrenberger, C. (1997) *Curr. Genet.* **32,** 1–18.
55. Allen, J. F. (1993) *J. Theor. Biol.* **165,** 609–631.
56. Pfannschmidt, T., Nilsson, A. & Allen, J. F. (1999) *Nature (London)* **397,** 625–628.
57. Pont-Kingdon, G. A., Okada, N. A., Macfarlane, J. L., Beagley, C. T., Wolstenholme, D. R., Cavalier-Smith, T. & Clark-Walker, G. D. (1995) *Nature (London)* **375,** 109–111.
58. Pont-Kingdon, G., Okada, N. A., Macfarlane, J. L., Beagley, C. T., Watkins-Sims, C. D., Cavalier-Smith, T., Clark-Walker, G. D. & Wolstenholme, D. R. (1998) *J. Mol. Evol.* **46,** 419–431.
59. Knoop, V., Unseld, M., Marienfeld, J., Brandt, P., Sunkel, S., Ullrich, H. & Brennicke, A. (1996) *Genetics* **142,** 579–585.
60. Vaughn, J. C., Mason, M. T., Sper-Whitis, G. L., Kuhlman, P. & Palmer, J. D. (1995) *J. Mol. Evol.* **41,** 563–572.
61. Joyce, P. B. & Gray, M. W. (1989) *Nucleic Acids Res.* **17,** 5461–5476.
62. Kanno, A., Nakazono, M., Hirai, A. & Kameya, T. (1997) *Plant Mol. Biol.* **34,** 353–356.
63. Cho, Y. & Palmer, J. D. (1999) *Mol. Biol. Evol.* **16,** 1155–1165.
64. Adams, K. L., Clements, M. J. & Vaughn, J. C. (1998) *J. Mol. Evol.* **46,** 689–696.
65. Modrich, P. & Lahue, R. (1996) *Annu. Rev. Biochem.* **65,** 101–133.