Mitochondrial Molecular Adaptations and Life History Strategies Coevolve in Plants

Neeraja M. Krishnan[1#], Basuthkar J. Rao[1]

[1]*Department of Biological Sciences, Tata Institute of Fundamental Research, Homi Bhabha road, Colaba, Mumbai – 400 005, India. Tel: +91 22 22782255 Fax: +91 22 22804610/11*

Corresponding Author
Neeraja M. Krishnan
Department of Biological Sciences,
Tata Institute of Fundamental Research,
Homi Bhabha road, Colaba
Mumbai – 400 005
India.
Tel: +91 22 22782255
Fax: +91 22 22804610/11

Email addresses: neeraja@tifr.res.in, bjrao@tifr.res.in

Keywords: life history strategies; *r-K* axis, Viridiplantae mitochondria, mRNA stability, helix-forming propensity, molecular adaptations, morphology, ecology

Running Title: Molecular and Ecological Adaptations

ABSTRACT

Messenger RNA secondary structure prevents mutations at functionally important sites. Mutations at exposed sites would cause micro-adaptations, niche-specialization, and therefore, can be thought to promote *K*-strategists. Exposing, rather than protecting, conserved sites, is also potentially adaptive because they probably promote macro-adaptive changes. This presumably fits *r*-strategists: their population dynamics tolerate decreased survival. We found that helix-forming tendencies are greater at evolutionary conserved sites of plant mitochondrial mRNAs than at evolutionary variable sites in a majority (73%) of species–gene combinations. *K*-strategists preferentially protect conserved sites in short genes, *r*-strategists protect them most in larger genes. This adaptive scenario resembles our earlier findings in chloroplast genes. Protection levels at various codon positions also display disparity with respect to life history strategies of the plants. Conserved site protection increases overall mRNA folding stabilities for some genes, while decreases it for some others. This contrast exists between homologous genes of *r*- and *K*- strategists. Such compensating interactions between variability, mRNA size, codon position, and secondary structure factors within *r*- and *K*-strategists are most likely, molecular adaptations of plants belonging to the two extreme life history strategies. Our results suggest coevolution between molecular and ecological adaptive strategies.

1. INTRODUCTION

Messenger RNAs seem to have important functional properties beyond mere coding. Their secondary structure seems to play several roles. For example, it regulates transcription and gene expression by affecting mRNA half-lives and turn-over (1-14). Carlini et al (15) suggest that gene expression in two *Drosophila* genes (Adh and Adhr) is dually regulated by synonymous codon bias and mRNA secondary structure. Experiments reveal effects of synonymous codon usage on expression level and mRNA decay in human genes, irrespective of codon-anticodon interaction specificity effects (16). In the extreme case of heat shock proteins, mRNAs are present constantly in the cellular matrix, but are expressed only after high temperatures melt down the secondary structure that prevents initiation of protein synthesis (17-19). Mammalian translation appears to be controlled by mRNA structure near 5' caps, as indicated by fluorescence assays for translation efficiency in single live cells (20)

Another role that has been suggested is that of protection, and we focus here on this aspect of mRNA structure. Site-specific helix-forming propensities in mRNA secondary structure correlate negatively with site-specific substitution rates (21), also after accounting for the replication-caused strand-specific effects on these rates in primate mitochondrial genomes (22-23).

Presumably, helix forming regions in secondary structures are relatively protected from mutations. The tendency for conserved sites to form helices suggests that this is a protection mechanism at functionally important sites, at DNA (22-23) or mRNA levels (21) or both (genetic code creates periodic pattern in mRNA secondary structure; 24) or at protein levels (25). This could bias the mutational spectrum of a gene towards micro-adaptive changes because mutations at the less conserved sites are more likely to produce

functional proteins with slightly altered properties, shifting, narrowing or widening their optima (i.e. for temperature). The opposite, where the more protected sites are more mutation prone, is also a potential adaptation. This would increase adaptability to drastic environmental changes and life history strategies. The former fits better species considered as overall *K*-strategists because it potentially increases micro-adaptation and allows niche specialization, while the latter fits overall *r*-strategists, as its costs by decreasing survival are more likely to be bearable for organisms with many offspring and shorter life-cycles.

Defining *r* and *K* strategies

The sigmoidal growth curve of an ideal population can be approximated by the differential equation: $dN/dt = rN(K - N) / K$, where N is the number of individuals present in a population, t is time, *r* is the intrinsic rate of population growth and *K* is the carrying capacity of the environment. Organisms particularly well adapted to an exponential increase in population size are called *r*-strategists (the *r* coming from the differential equation described above). These *r*-strategists are characterized by great rapidity in their development combined with an ability to produce large numbers of offspring. They tend to inhabit disturbed and transient environments. In contrast to *r*-strategists, *K*-strategists show extreme potential to survive and prosper at or near carrying capacity, though often at the expense of their intrinsic rate of population growth. The variable *K* refers to carrying capacity; they display a bias in their adaptations toward maximizing this parameter.

Plants are one of the best studied subjects as far as life history strategies and environmental habitats are concerned (26-28), and abundant molecular data is available for this group. Therefore we explore these molecular adaptive scenarios of coevolution

between mRNA secondary structure and the conservation of the mRNA's coding properties in plant mitochondria, considering the life history of the plant. In addition, we have already published results from similar analyses of complete chloroplast genomes (29). Comparing them with plant mitochondrial genome co-evolution data would generate further useful insights. The existence of two metabolically important organelles in plant cells, mitochondrion and chloroplast, and their respective genomes, will enable, in the future, independent tests of the hypotheses developed here.

We analyzed the relationships between the site-specific helix-forming tendencies of plant mitochondrial mRNAs and their corresponding site-specific evolutionary variability levels according to four pre-defined rate categories. We analyzed separately the three codon positions, expecting differences due to their different coding properties: the third codon position is relatively freely variable because of the redundancy in the genetic code and is often also used as a neutral control with respect to variation at the other two codon positions for measuring positive selection, adaptation and functional divergence (30-39). Nevertheless, one should remember that even synonymous changes at third codon position have frequently functional consequences, such as, for example, effects on rates and accuracy of protein synthesis (40-41) or costs of ribosomal frameshifts (42).

Our results reveal contrasting relationships between variability, mRNA size, codon position, and secondary structure factors for *r*- and *K*-strategists. They suggest that molecular adaptations divergently manifest themselves as adaptations towards extreme life history strategies, in a gradual way. Our approach towards molecular processes and evolution converges with that described by Skulachev (43).

2 Materials and Methods

2.1 Choosing, Pruning and Aligning Data

We chose sixteen complete mitochondrial genomes belonging to the group *Viridiplantae*, including four chlorophytes (*Nephroselmis olivacea* (NC_008239; (44)), *Prototheca wickerhamii* (NC_001613; (45)), *Pseudendoclonium akinetum* (NC_005926; (46)), *Scenedesmus obliquus* (NC_002254; (47))) and twelve streptophytes (Eudicots: *Arabidopsis thaliana* (NC_001284; (48)), *Beta vulgaris vulgaris* (NC_002511; (49)), *Brassica napus* (NC_008282; (50)), *Nicotiana tabacum* (NC_006581; (51)); Monocots: *Oryza sativa* (NC_007886; (52)), *Triticum aestivum* (NC_007579; (53)), *Zea mays mays* (NC_007982; (54)); Coleochaetales: *Chaetosphaeridium globosum* (NC_004118; (55)); Charales: *Chara vulgaris* (NC_005255; (56)); Marchantiales: *Marchantia polymorpha* (NC_001660; (57)); Mesostigmatales: *Mesostigma viride* (NC_008240; (58)); Bryophyta: *Physcomitrella patens* (NC_007945; (59))). These genomes were selected from twenty complete mitochondrial genomes available in NCBI GenBank (http://www.ncbi.nlm.nih.gov) as of May 2008. Our selection was designed to maximize representation of extreme *r* and *K* life strategists, while simultaneously maximizing the number of homologous protein-coding genes among the species. We observed that discarding some taxa and thereby, decreasing the sampled taxon density yielded more homologous genes. We compiled a dataset of twelve mitochondrial protein-coding genes (atp9, cob, cox1, cox2, cox3, nd1, nd2, nd3, nd4, nd4l, nd5 and nd6) from the above mentioned plant species, that best satisfied both sampling criteria.

We defined all species belonging to Spermatophyta as relatively fitting the *K*-strategists life history syndrome, due to their relative large sizes, longevities, and low offspring

numbers. This group also includes the large fern species of our sample. Other plants were considered from an ecological point of view as relatively *r*-strategists (see for review of the concepts (60)). This way of grouping largely parallels that of micro-organisms (*r*-strategists) and macro-organisms (*K*-strategists). It is possible that the differences we detect below might be as well due to the relative extent of pluri-cellularity (or complexity at the level of cellular differentiation of the organism), rather than the ecological interpretation we propose. However, diversity of cellular differentiation in itself might well be part of the *K* strategy syndrome.

We aligned the amino acid and nucleotide sequences of protein coding genes from each of these datasets using ClustalW (61). DNA alignments were further corrected using the amino acid alignment as reference by introducing three gaps for each gap found in the amino acid alignment. This was done in order to preserve the correct coding frame. We considered only the coding sequences of the genes for our analyses, irrespective of the strand on which they are coded, because this is the sequence expressed as mRNA.

2.2 Secondary Structure Calculations

We used the RNA folding version of the mFold software (62) available online (http://www.bioinfo.rpi.edu/applications/mfold/rna/) and obtained secondary structures for all the twelve genes, for each of the sixteen genomes. We considered all alternative structures that were within 50% sub-optimality of the most optimal secondary structure. We calculated the average stability across the 50% sub-optimality range of secondary structures. We also averaged over the same range, the propensity of each site in the alignment to be part of a helix. The helix-forming propensity is indicated by a binary value (0 for loop and 1 for stem). We thus obtained the average site-specific helix-forming propensity of each nucleotide in the gene sequence by dividing the number of

secondary structures in which that site participates in helix formation by the total number of secondary structures.

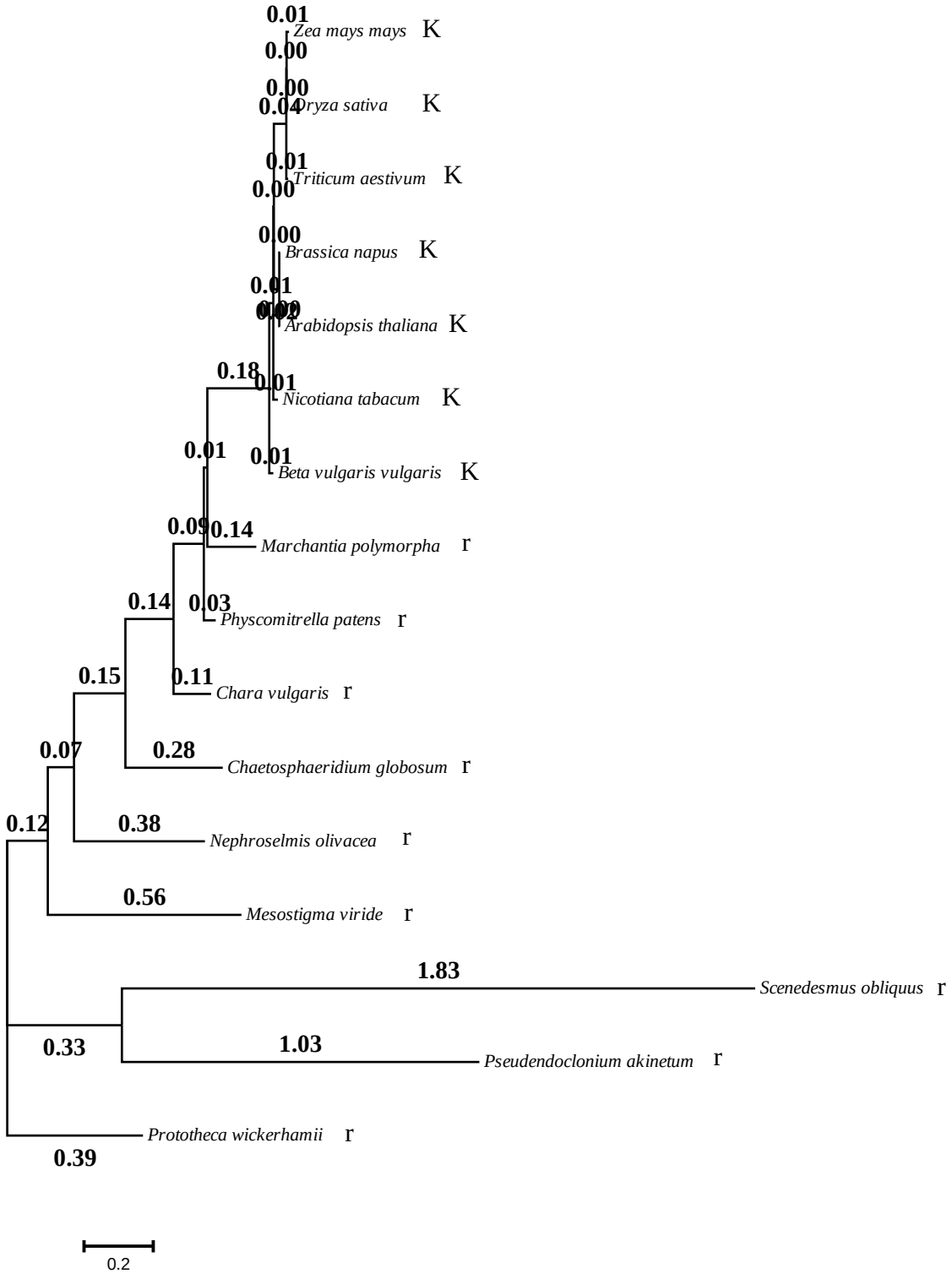

Figure 1 ***Maximum Likelihood Phylogeny of*** **Viridiplantae** ***Mitochondrial Genomes*** **The mitochondrial phylogeny for sixteen Viridiplantae genomes is inferred using PHYLIP v 3.67 (63) under *dnaml* criterion which uses maximum likelihood. Phylogeny was calculated using the corrected nucleotide sequence alignments. Branch lengths were also optimized by maximum likelihood and are proportional to the number of nucleotide substitutions per site as per the scale bar. The life history strategy adopted by the plant species is indicated as *K* or *r*, next to the species name on the phylogeny.**

2.3 Calculations of Variability Categories for Each Site

We inferred the mitochondrial phylogeny for this dataset using PHYLIP v 3.67. (63) under *dnaml* criterion, which uses maximum likelihood. Phylogenies were calculated using the corrected nucleotide sequence alignments. Branch lengths were also optimized by maximum likelihood and are proportional to the number of nucleotide substitutions per site. Figure 1 depicts the phylogeny we used, which also indicates the life history strategies of the plant species. The *r*-strategists and *K*-strategists form two distinct phylogenetic clusters, which stresses the need to test the working hypotheses proposed here at ulterior stages with larger numbers of species while taking into account phylogenetic relationships. Rate variation was allowed using a gamma distribution (64) with shape parameter an= 1.75. The inferred phylogeny was further used to infer in which predefined average rate categories each site fits, such as to maximize the overall likelihood of the data, given the phylogeny. The number of site-specific evolutionary variability categories depicting different average rate parameters and the corresponding rates were decided based on probabilistic cut-offs evaluated by hidden Markov Models by PHYLIP while inferring phylogenies. These cut-offs indicated a high probability of sites fitting an average rate parameter =·0.5. Based on this information, we arbitrarily chose four rate categories: M1, M2, M3 and M4, which represented average rate parameters of 0.1, 0.2, 0.3 and 0.4. Maintaining a constant average rate across the genome reduces the sensitivity in detecting heterogeneity at synonymous sites. Our

analysis incorporates average rate variation and is therefore sensitive to heterogeneity even at the synonymous sites.

We further performed similar analyses on individual codon positions for this dataset and its corresponding phylogeny. Qualitatively, classification of sites under these four categories helps to assess the evolutionary variability composition of gene sequences based on their relative fits to the four evolutionary rates. Determining the relative frequency of sites fitting each of the four categories, we could comparatively assess differences in evolutionary variability compositions if any, across different gene sequences.

Table 1 Average site-specific stem-forming tendencies were obtained across the initial 50% sub-optimal structures obtained from the RNA folding version of mFold (62) for twelve mitochondrial mRNAs of sixteen plant species and the coefficients of correlations between these site-specific stemminess measures and the variability category of that site are represented below as 'protectedness' of variable sites. A positive coefficient indicates that variable sites are protected and/or conserved sites are exposed. A negative coefficient indicates that variable sites are exposed and/or conserved sites are protected. Variability categories of each site were obtained using the 'dnaml' command available in PHYLIP v3.67 package. The life-strategies, i.e. '*r*' or '*K*', adopted by the plant species are indicated within brackets, next to the scientific names. Numbers in bold indicate significant correlations according to 2-tailed t-tests. The last three rows, indicated by N, N_K, and N_R indicate the total number of negative correlation coefficients over all the sixteen species, all the *K*-strategists and all the *r*-strategists, respectively, for each gene. The last column indicated by N is the total number of negative correlation coefficients over all genes for each species.

	Exposedness of conserved Sites (r)												
	atp9	cob	cox1	cox2	cox3	nd1	nd2	nd3	nd4	nd4l	nd5	nd6	N
Arabidopsis thaliana (K)	-0.07	-0.02	-0.02	-0.02	**-0.07**	-0.03	0.05	0.03	0.02	**-0.17**	0.00	**0.08**	7
Brassica napus (K)	-0.04	-0.03	-0.02	0.00	-0.03	-0.04	0.05	0.05	-0.02	**-0.20**	-0.04	**0.08**	8
Nicotiana tabacum (K)	-0.12	-0.01	0.00	-0.02	-0.04	-0.02	0.01	-0.04	0.00	**-0.21**	0.00	-0.04	10
Oryza sativa (indica cultivar-K)	-0.16	-0.02	-0.01	-0.01	-0.02	0.02	0.00	0.01	-0.04	**-0.13**	**-0.05**	0.01	8
Triticum aestivum (K)	-0.11	-0.01	-0.01	**-0.08**	-0.03	-0.01	0.01	0.00	-0.02	**-0.13**	-0.03	0.00	11
Zea mays subsp. mays (K)	-0.21	-0.03	0.00	0.01	-0.03	0.00	0.03	0.01	-0.04	**-0.12**	-0.04	-0.03	7
Beta vulgaris subsp. vulgaris (K)	-0.12	-0.03	-0.03	-0.03	**-0.06**	-0.03	0.00	-0.03	-0.03	**-0.18**	-0.02	0.02	10
Chaetosphaeridium globosum (r)	-0.07	-0.02	-0.04	-0.06	-0.03	**-0.06**	-0.02	-0.02	-0.03	-0.08	-0.01	0.02	11
Chara vulgaris (r)	0.00	-0.03	-0.03	-0.06	**-0.08**	-0.04	**-0.06**	-0.07	-0.01	**-0.11**	**-0.07**	0.00	12
Marchantia polymorpha (r)	-0.02	0.00	-0.05	0.04	**-0.08**	**-0.10**	-0.05	0.01	-0.02	-0.10	-0.02	-0.02	10
Mesostigma viride (r)	-0.12	-0.01	**-0.05**	-0.02	**-0.09**	**-0.08**	**-0.06**	-0.01	-0.04	-0.10	**-0.10**	-0.02	12
Nephroselmis olivacea (r)	-0.12	-0.02	**-0.07**	-0.06	-0.04	-0.03	-0.03	0.08	-0.01	-0.02	-0.04	-0.03	11
Physcomitrella patens (r)	-0.03	-0.02	**-0.06**	**-0.09**	-0.05	-0.04	-0.03	-0.04	-0.03	-0.04	-0.04	-0.05	12
Prototheca wickerhamii (r)	-0.09	-0.07	**-0.05**	0.00	**-0.11**	-0.03	-0.03	**0.09**	0.02	0.01	0.00	-0.05	7
Pseudendoclonium akinetum (r)	-0.12	-0.02	-0.02	-0.05	-0.01	**-0.07**	0.00	0.00	**-0.05**	**-0.18**	**-0.09**	**0.07**	9
Scenedesmus obliquus (r)	0.03	0.03	0.00	0.03	-0.01	0.04	0.01	0.00	-0.01	0.01	0.01	0.02	2
N	15	15	14	11	16	13	7	7	14	14	12	9	
N_K	6	6	5	4	6	4	0	2	5	6	4	3	
N_R	9	9	9	7	10	9	7	5	9	8	8	6	

3 Results

3.1 Protection by secondary structure of functionally important sites

The coefficients of correlation (*r*s) between the site-specific variability levels and site-specific stemminess in the secondary structure of mRNAs are represented for all genes and species in Table 1. Positive *r*s indicate that highly conserved sites in the gene tend to be part of loops in the secondary structure and that the variable sites form helices. Negative *r*s indicate that highly conserved sites form helices, and the more variable sites are part of loops. Negative *r*s also fit general expectations that the functionally important (conserved) sites are those most protected by the secondary structure. This expected scenario is found in 73% of all combinations of genes and species, and is slightly weaker in *K*-strategists (68%) than in *r*-strategists (77%). Each % value is significantly greater than 50% by a sign test, but the difference between them was not statistically significant by chi square test.

Some genes have greater proportions of significant correlation coefficients in the expected negative direction for one strategist type than the other. This difference is strong for cox3 (4 of 9 negative *r*s in *r*-strategists and 2 of 7 in *K*-strategists), nd2 (7 of 9 negative *r*s (significant at $P < 0.05$ for 2 cases) in *r*-strategists and none in *K*-strategists), nd1 and cox1 (4 of 9 negative *r*s in *r*-strategists and none in *K*-strategists) and nd4L (2 of 9 negative *r*s in *r*-strategists and 7 of 7 in *K*-strategists) genes. The differences between the mean *r*s in the two life history types are significant (2-tailed t-test for unequal sample sizes) for cox1, nd1, nd2, and nd4L genes (Figure 2). These differences between them are affected by gene size as a confounding factor, according to which the trends switch directions. A greater percentage of negative correlation coefficients (69%) was found for

smaller genes (with average size lesser than 950 nucleotides) than larger genes with average size greater than 950 and lesser than 2200 nucleotides, (61%), in *K*-strategists, though this difference was not statistically significant.

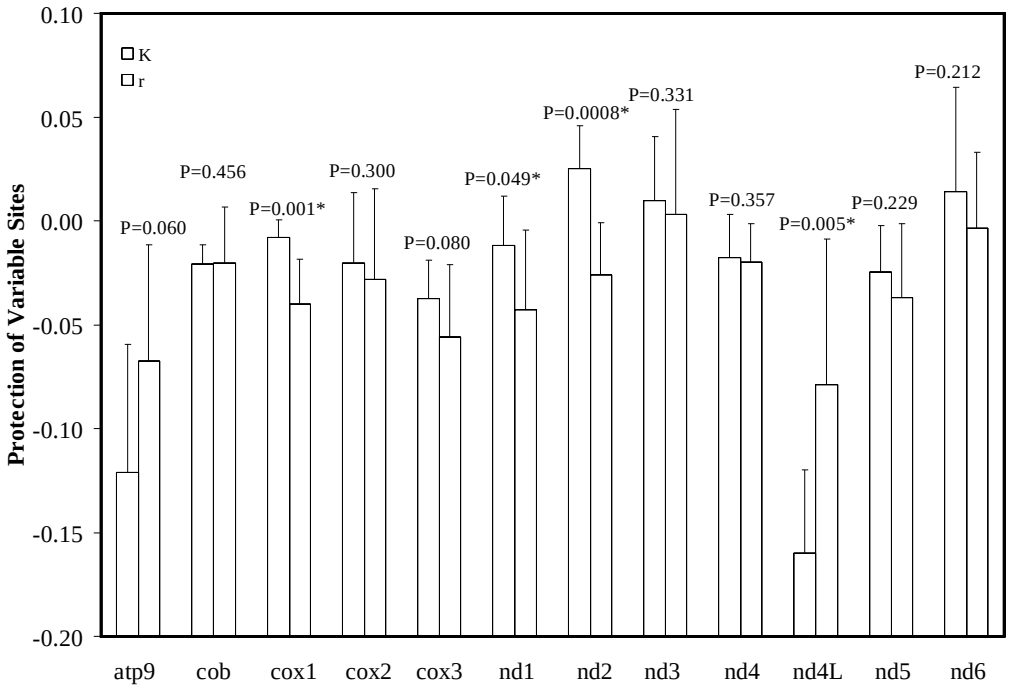

Figure 2 ***Level of 'protectedness' of variable sites in mitochondrial genes vis-à-vis life history strategies of plant species***
The Y-axes are the correlation coefficients between the variability category of each site in a gene and the tendency of that site to be part of helices in the RNA secondary structure. These measures are averaged across all 7 *K*-strategists and 9 *r*-strategists. The variation in variability-protection correlation measures are indicated by the standard deviation bars. The double (**) and single asterisks (*) indicate significant differences at the 5% and 10% levels of significance, according to a 2-tailed t-test.

According to these size categories, the large genes were nd1, nd2, nd4, nd5, cox1, and cob with average sizes 991, 1751, 1705, 2200, 1633 and 1318 nucleotides, and the small genes were atp9, nd3, nd4L, nd6, cox2 and cox3 with average sizes 276, 473, 324, 797, 861 and 946 nucleotides, respectively. In *r*-strategists, the relationship with gene size is reversed: there are fewer negative rs (70%) in shorter genes than in larger genes (83%). This means that the large genes in *r*-strategists have significantly more times negative rs

than any size in *K*- and small size in *r*-strategists ($P < 0.05$, 1 tailed chi2 test (a 1 tailed test is arguably justified because one can expect that cost minimization is more needed for large (hence more costly) genes, especially in *r*-strategists disposing of less little reserves)). Hence, minimization of mutation costs is greatest for large genes of *r*-strategists.

3.2 Protection by Secondary Structure at Different Codon Positions

The overall trend at second codon positions fits best expectations that conserved sites tend to form helices, as indicated by the relatively larger percentage (73%, which is significantly more than 50% by sign test) of negative correlation coefficients for all combinations of genes and species (Table 2B). The trend at first and third codon positions is similar and weak (56-57%, not significantly different from 50% by sign test; Table 2A and 2C). The average variability levels across genes are not very different for the first, second and third codon positions (1.68, 1.46 and 1.57 respectively).

The proportion of negative *r*s for each gene correlates negatively with the average variability levels of those genes at the first and second codon position, but positively at the third codon positions. The percentage of negative *r*s is significantly lower (by 2 tailed chisquare tests, $P < 0.05$) in *K*-strategists than *r*-strategists at first and second codon positions (61:50 and 80:64, respectively), as reported above for the complete sequences. However, at third codon positions, *K*-strategists yield more negative *r*s (62%) than *r*-strategists (54%), although this difference is not statistically significant by chi-square test. Different selection pressures at the synonymous and non-synonymous sites are therefore, an additional factor regulating the coevolution of molecular and ecological life history switches.

A

	Exposedness of conserved Sites (r) Codon Position I												
	atp9	cob	cox1	cox2	cox3	nd1	nd2	nd3	nd4	nd4l	nd5	nd6	N
Arabidopsis thaliana (K)	-0.12	-0.02	-0.02	0.07	-0.07	0.03	**0.08**	0.02	0.02	**-0.24**	0.04	0.08	5
Brassica napus (K)	-0.09	-0.01	0.00	0.09	-0.02	0.04	0.08	0.09	-0.01	**-0.20**	0.04	0.08	5
Nicotiana tabacum (K)	-0.10	-0.02	0.07	-0.10	-0.03	0.09	-0.01	-0.10	-0.01	**-0.29**	0.07	-0.04	9
Oryza sativa (indica cultivar-K)	-0.14	0.04	0.04	0.04	-0.06	**0.11**	0.04	0.07	-0.04	-0.19	-0.04	0.01	5
Triticum aestivum (K)	-0.08	0.03	0.07	-0.08	-0.05	0.08	**0.09**	0.05	-0.01	**-0.19**	-0.03	0.00	7
Zea mays subsp. mays (K)	-0.15	0.03	0.05	0.04	-0.07	0.07	**0.11**	0.06	-0.02	-0.19	-0.04	-0.03	6
Beta vulgaris subsp. vulgaris (r)	-0.17	-0.05	0.02	-0.09	-0.03	0.05	0.04	0.00	-0.02	**-0.29**	0.03	0.02	6
Chaetosphaeridium globosum (r)	-0.18	**-0.13**	-0.01	**-0.13**	-0.07	-0.08	0.01	-0.07	-0.04	**-0.26**	0.00	0.02	10
Chara vulgaris (r)	-0.14	-0.07	-0.03	-0.02	-0.05	0.04	-0.03	-0.02	0.01	**-0.21**	0.00	0.00	10
Marchantia polymorpha (r)	-0.16	-0.06	-0.05	0.04	-0.05	0.03	0.04	0.00	-0.01	**-0.20**	-0.01	-0.02	9
Mesostigma viride (r)	-0.05	**-0.12**	-0.04	0.06	-0.08	0.00	-0.07	0.13	-0.02	-0.18	**-0.10**	0.06	9
Nephroselmis olivacea (r)	-0.11	0.02	-0.03	-0.08	0.01	0.05	-0.02	0.02	0.01	0.01	-0.03	-0.03	6
Physcomitrella patens (r)	**-0.24**	**-0.12**	-0.06	0.02	0.01	0.01	-0.04	0.01	0.01	-0.07	-0.01	-0.05	7
Prototheca wickerhamii (r)	-0.04	-0.08	**-0.12**	-0.07	-0.08	0.01	-0.01	-0.11	0.02	-0.01	0.02	-0.05	9
Pseudendoclonium akinetum (r)	-0.14	-0.05	0.01	-0.02	-0.07	-0.04	0.02	0.08	-0.06	**-0.21**	**-0.09**	0.07	8
Scenedesmus obliquus (r)	0.00	0.06	0.04	0.09	-0.01	**0.11**	0.01	0.05	-0.06	-0.03	0.05	0.02	3
N	15	11	8	8	14	3	6	5	11	15	10	8	
N_K	6	3	1	2	6	0	1	1	5	6	3	3	
N_R	9	8	7	6	8	3	5	4	6	9	7	5	

B

	Exposedness of conserved Sites (r) Codon Position II												
	atp9	cob	cox1	cox2	cox3	nd1	nd2	nd3	nd4	nd4l	nd5	nd6	N
Arabidopsis thaliana (K)	-0.06	-0.08	0.00	0.01	**-0.13**	-0.10	0.04	0.08	0.06	**-0.20**	0.04	0.08	5
Brassica napus (K)	0.05	-0.07	-0.01	0.04	-0.04	-0.09	0.05	0.09	-0.05	-0.17	**-0.12**	0.08	7
Nicotiana tabacum (K)	**-0.26**	-0.04	-0.06	0.10	-0.05	-0.11	0.05	**0.19**	-0.05	-0.08	-0.06	0.01	8
Oryza sativa (indica cultivar-K)	**-0.21**	-0.08	-0.03	0.05	0.02	-0.09	0.01	0.09	-0.05	-0.13	**-0.10**	0.02	7
Triticum aestivum (K)	-0.13	-0.09	-0.02	-0.07	-0.02	**-0.15**	-0.02	0.09	-0.01	-0.14	**-0.08**	0.01	10
Zea mays subsp. mays (K)	-0.07	**-0.17**	-0.01	0.03	-0.04	-0.09	0.00	0.10	**-0.08**	-0.14	**-0.09**	-0.02	9
Beta vulgaris subsp. vulgaris (r)	-0.16	**-0.14**	-0.05	0.08	-0.10	**-0.12**	0.00	0.08	-0.06	-0.16	**-0.09**	0.03	8
Chaetosphaeridium globosum (r)	-0.06	-0.04	-0.05	-0.03	-0.02	-0.05	-0.06	0.13	-0.01	-0.03	-0.03	0.04	10
Chara vulgaris (r)	-0.01	**-0.10**	-0.01	-0.04	-0.11	-0.09	**-0.14**	-0.07	-0.06	-0.14	**-0.16**	0.01	11
Marchantia polymorpha (r)	-0.14	-0.03	-0.05	0.08	-0.08	**-0.22**	**-0.08**	-0.07	-0.04	-0.13	-0.02	-0.01	11
Mesostigma viride (r)	-0.08	0.01	-0.06	-0.01	**-0.15**	**-0.19**	-0.01	-0.15	**-0.08**	-0.09	**-0.12**	-0.07	11
Nephroselmis olivacea (r)	-0.10	-0.10	-0.05	0.03	-0.10	**-0.17**	-0.02	**0.23**	0.01	-0.09	-0.04	-0.02	9
Physcomitrella patens (r)	-0.10	**-0.10**	**-0.09**	-0.10	**-0.14**	-0.10	0.00	0.02	-0.08	-0.18	-0.06	-0.06	10
Prototheca wickerhamii (r)	-0.05	**-0.14**	-0.03	**0.15**	**-0.16**	-0.09	-0.07	0.07	0.06	-0.11	-0.02	-0.02	9
Pseudendoclonium akinetum (r)	-0.15	-0.06	-0.05	0.01	0.05	**-0.12**	-0.01	-0.14	-0.07	**-0.22**	**-0.12**	0.07	9
Scenedesmus obliquus (r)	0.03	0.06	-0.03	0.05	-0.01	-0.03	-0.02	0.07	-0.01	-0.03	0.00	0.03	7
N	14	14	15	5	14	16	9	4	13	16	15	6	
N_K	5	6	5	1	5	6	1	0	5	6	5	1	
N_R	9	8	10	4	9	10	8	4	8	10	10	5	

C

	Exposedness of conserved Sites (r) Codon Position III												
	atp9	cob	cox1	cox2	cox3	nd1	nd2	nd3	nd4	nd4l	nd5	nd6	N
Arabidopsis thaliana (K)	0.05	0.02	-0.02	**-0.13**	-0.04	-0.05	0.03	-0.13	0.00	-0.14	0.04	0.10	6
Brassica napus (K)	0.01	-0.02	0.01	-0.09	-0.02	-0.10	0.03	-0.05	0.02	**-0.24**	-0.04	0.10	7
Nicotiana tabacum (K)	0.00	0.02	-0.02	-0.06	-0.07	-0.06	-0.01	**-0.22**	0.00	**-0.23**	-0.01	-0.05	9
Oryza sativa (indica cultivar-K)	-0.05	-0.02	0.00	-0.09	0.00	-0.02	-0.05	-0.12	-0.01	-0.05	-0.02	0.03	9
Triticum aestivum (K)	-0.07	0.00	-0.02	-0.10	0.00	-0.02	-0.04	-0.13	-0.04	-0.04	0.00	0.03	8
Zea mays subsp. mays (K)	**-0.26**	0.01	0.02	-0.03	0.01	-0.03	-0.04	-0.14	-0.03	-0.02	-0.01	-0.01	9
Beta vulgaris subsp. vulgaris (K)	-0.03	0.08	-0.04	-0.07	-0.08	-0.04	-0.03	-0.18	-0.01	-0.09	-0.01	-0.04	11
Chaetosphaeridium globosum (r)	0.08	**0.11**	-0.01	-0.02	0.05	-0.03	0.01	-0.10	-0.02	0.05	0.00	0.06	5
Chara vulgaris (r)	0.09	0.05	-0.02	-0.11	-0.10	-0.04	-0.01	-0.13	0.01	-0.02	-0.05	0.02	8
Marchantia polymorpha (r)	0.19	0.08	-0.06	0.02	-0.07	**-0.12**	**-0.09**	**-0.26**	-0.01	0.07	-0.01	-0.05	8
Mesostigma viride (r)	-0.16	**0.10**	-0.08	-0.09	-0.04	-0.07	**-0.10**	0.02	-0.01	-0.02	-0.05	-0.05	10
Nephroselmis olivacea (r)	0.02	0.01	**-0.11**	**-0.13**	0.02	0.03	-0.03	-0.01	0.00	0.02	-0.04	0.06	5
Physcomitrella patens (r)	**0.26**	**0.15**	-0.01	**-0.16**	-0.04	-0.05	-0.06	-0.16	-0.05	0.05	-0.03	-0.05	9
Prototheca wickerhamii (r)	-0.12	-0.02	-0.02	-0.07	-0.10	-0.04	-0.01	0.14	0.00	0.13	0.01	0.07	7
Pseudendoclonium akinetum (r)	0.01	0.04	-0.07	**-0.15**	0.03	-0.10	0.00	0.05	-0.03	-0.16	-0.04	0.02	6
Scenedesmus obliquus (r)	0.15	-0.01	-0.04	-0.03	0.03	0.01	0.05	-0.11	0.05	0.12	-0.03	-0.04	6
N	6	4	13	15	9	14	11	13	9	10	12	7	
N_K	4	2	4	7	4	7	5	7	4	7	5	3	
N_R	2	2	9	8	5	7	6	6	5	3	7	4	

Table 2 Average site-specific stem-forming tendencies were obtained across the initial 50% sub-optimal structures obtained from the RNA folding version of mFold (62) for twelve mitochondrial mRNAs of sixteen plant species and the coefficients of correlations between these site-specific stemminess measures and the variability category of that site are represented below as 'protectedness of variable sites for individual codon positions (Table 2a, 2b and 2c for the first, second and third codon positions). The other details are as in Table 1.

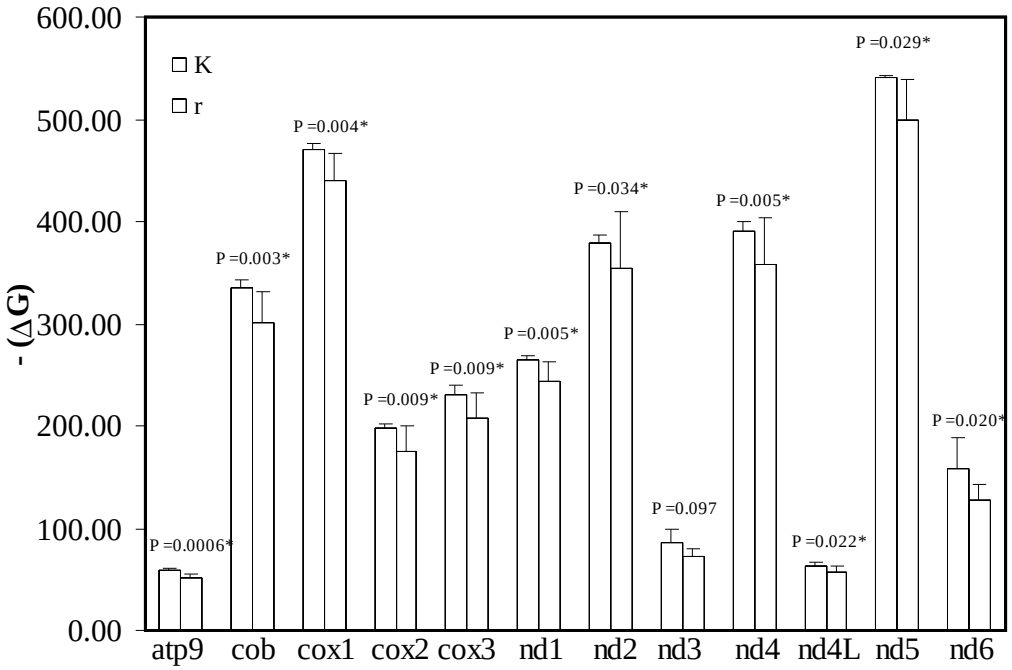

Figure 3 ***RNA stability in mitochondrial genes vis-à-vis life history strategies of plant species*** **The Y-axes are the correlation coefficients between the variability category of each site in a gene and the secondary structure stability (negative of ΔG) of mitochondrial mRNAs. These measures are averaged across all 7 *K*-strategists and 9 *r*-strategists. The variation in the stability correlation measures are indicated by the standard deviation bars. The double (**) and single asterisks (*) indicate significant differences at the 5% and 10% levels of significance, according to a 2-tailed t-test.**

3.3 mRNA secondary structure adaptations and species life history strategies

The overall stability of genes is significantly greater in *K*-strategists than in *r*-strategists, except for nd3, for which the P-value of the differences in stabilities between the two life history types is 0.097 (by 2-tailed t-test for unequal sample sizes; Figure 3). We also explored whether the variability or conservation levels of helix-forming sites enhances or decreases overall mRNA stability. We used regression analyses to estimate their contribution to stability in *r*- versus *K*-strategists. The t-statistics of these regression analyses are plotted for all genes in Figure 4. The t-statistics are in opposite directions for *K*- and *r*-strategists, for the majority of genes (8 among 12). atp9, cob, cox3, nd2 are genes for which the t-statistics are in the same direction. For most genes, rs between variability and stemminess correlate positively with stability, irrespective of their sign. This is significant for cob, cox1, cox3 and nd2. Significant negative correlations with stability occur for nd4L and nd6 genes, for *K*-strategists. Variability correlates negatively with stemminess for nd4L of *K*-strategists, and predominantly positively for nd6 gene in *K*-strategists. This is also true for the *r*-strategists to a lesser extent or with lower significance, than for *K*-strategists. However, these rs increase (though not significantly) with stability in both genes for *r*-strategists. This suggests that different genes adapted their nucleotide compositions and secondary structure to either up-regulate stability as in *r*-strategists, or down-regulate stability as in *K*-strategists.

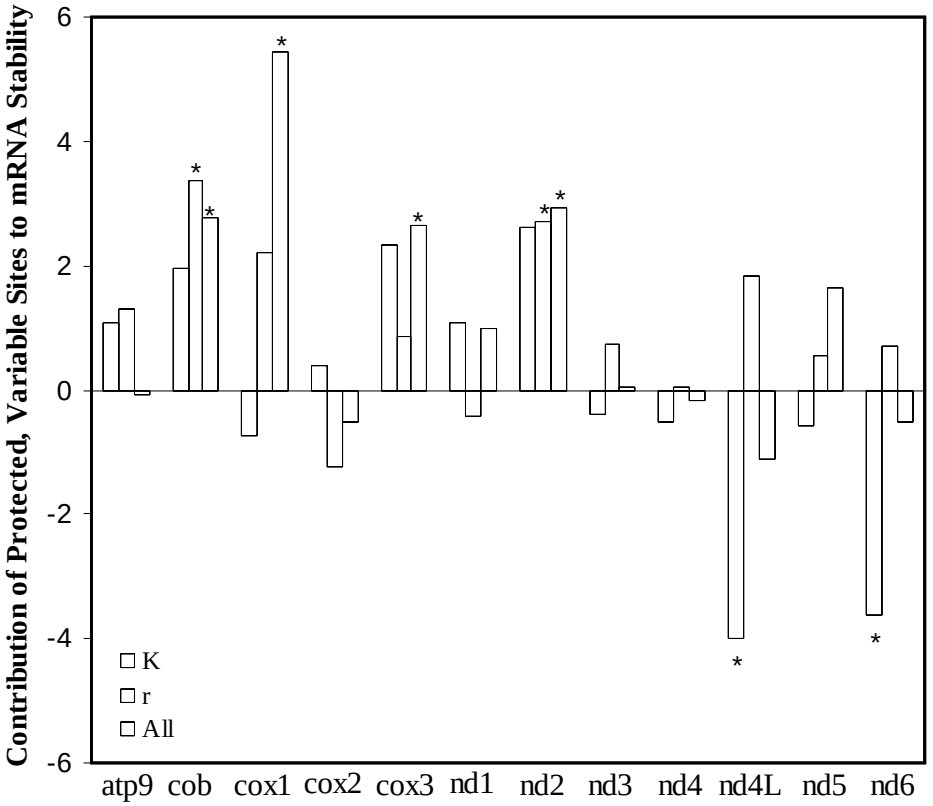

Figure 4 ***Contribution of protected, variable sites (and/or exposed, conserved sites) to RNA stability vis-à-vis life history strategies of plant species***

The coefficient of correlation between the RNA stability, - (ΔG), and exposed extents of conserved sites in that mitochondrial mRNA depicts the Y-axis. The protection of variable sites and exposure of conserved sites is estimated as in Figure 1, by the correlation coefficient between the variability categories of sites and the tendencies of these sites to be part of helices. The difference between this contribution for *K*- and *r*-strategist species is highly significant (P=0.004, 2-tailed t-test). Asterisks denote gene-strategy combinations for which the contribution to mRNA stability of exposed, conserved sites is significant (2-tailed t-test).

4 Discussion

4.1 Regulating mutability

Our major result is that genome regions that are conserved, and hence are probably functionally important, tend to be coded by sequences that participate in the formation of helices at the mRNA level. This probably results in fewer mutations at the mRNA level, increasing the accuracy of protein synthesis for a majority of sites. This suggests that some highly conserved sites are conserved not because of their coding properties and their effects on protein function, but because they protect by complementing other sites in helices, the latter being functionally important beyond mRNA secondary structure. This means that part of the design of mRNAs includes packing and protection constraints. As a result, mutations, also at the DNA level (DNA, during single stranded periods (22-23) also self hybridizes in a manner similar to RNA), are biased as a function of their potential adaptive effects on protein function. This adaptation, as that described for off frame stops that stop protein synthesis after ribosomal shift (42), decreases the costs of inaccurate translation due to mutations in the mRNA.

4.2 Adaptive strategies for mutability at molecular level

Results indicate the existence of three adaptive strategies, for various combinations of plants and genes: 1. the most frequent strategy protects functionally important sites (for example in the *r*- strategist *Mesostigma viride*, and the short gene nd4L, see Table 1) and is most frequent at the second codon position (31 significant cases: 17.7% of all cases; Table 2B), as compared to 11.5 and 10.9% at first and third codon positions, respectively); 2. a rarer strategy exposes functionally important sites (sign tests show a significant excess of positive correlations in the *r*-strategist *Scenedesmus obliquus* (Table 1), at first codon positions in ND1 (Table 2A) and at third codon positions in ND3 (Table 2C)) and is most common at 3rd codon position (10

significant cases: 5.7% of all cases), as compared to 2.6 and 1.5% for first and second codon positions, respectively (2.5% of tests are expected to be positive and significant at $P < 0.05$ according to two tailed tests due to chance); and 3. a mixed strategy, where sequences probably include sites or regions following each of the previous principles, resulting in fewer than expected significant correlations combined with a lack of overall direction in the signs of the correlations when looking at all the cases in these genes or species. These combined criteria suggest that this mixed strategy exists for the *r*-strategists *Nephroselmis olivacea* (first codon position), *Scenedesmus obliquus* (second codon position) and *Chara vulgaris* (third codon position), and for ND3 (first codon position) and ND6 (first and third codon positions). These considerations suggest that in any gene, some regions might tend to follow the protective strategy and others, the exposing strategy, and that our analyses detect consistent tendencies only in the relatively fewer cases where the majority of sites in that sequence follow the same strategy.

4.3 Life history strategies, mRNA secondary structure stability and gene size

There were some differences in the tendencies of *r*- and *K*-strategists with respect to the protection of conserved sites. Overall, and for the same gene length, mRNA secondary structure stability was greater in *K*- than *r*-strategists (Figure 3). The stability increases more with gene length in the former (Figure 5). It can be suspected that higher stabilities reflect slower turn-over rates in *K*-strategists, which could also fit with slower growth. If mRNA half lives are longer, the need for protection against mutations is also greater. This is in line with the observation that the tendency to protect more conserved sites was also stronger in *K*-strategists than in *r*-strategists, and this especially at the first and second codon positions. The opposite was true for the third codon position, which suggests that *K*-strategists increase the

adaptive flexibility of their molecular machinery mainly by synonymous or semi-conservative mutations, which is less true for *r*-strategists.

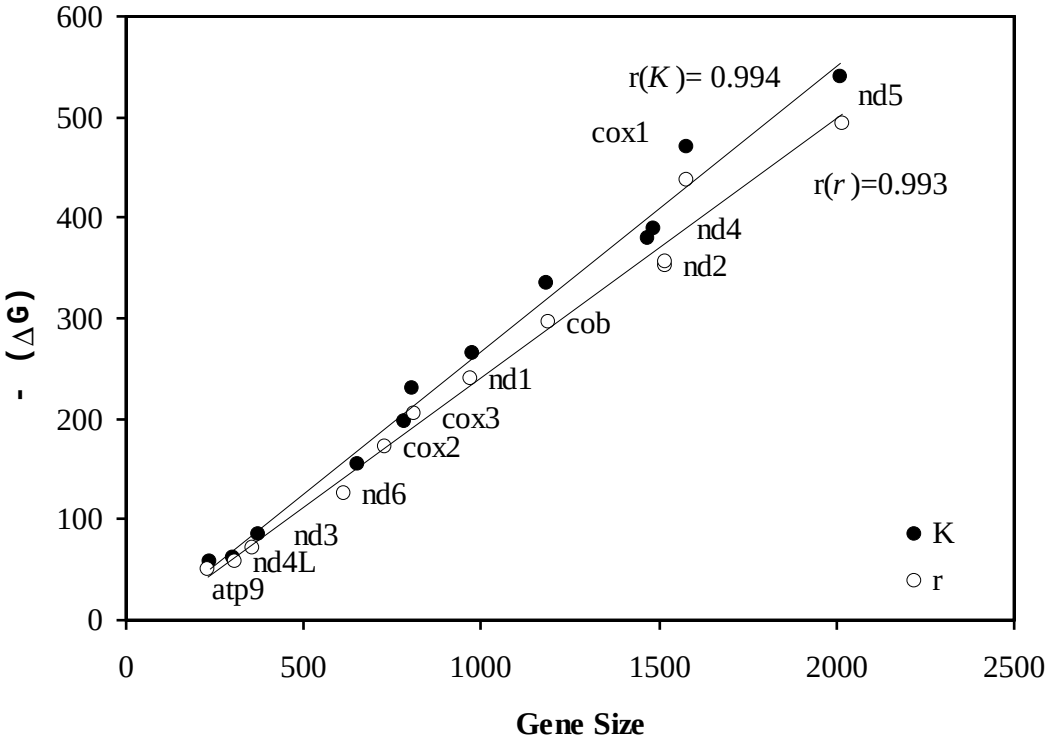

Figure 5 ***Mean stability of mRNA secondary structure as a function of the mRNA's mean length, for r- and K*-strategists**
The average stabilities were calculated over the folding stability of the mRNAs, separately for *K*- and *r*-strategists. The *K*-strategists are indicated by filled circles and the *r*-strategists are indicated by hollow circles.

Gene lengths were not systematically different between *K*- and *r*-strategists. This result indicates that cost minimization does not include gene length, when comparing mitochondrial genes of *r* and *K* plants, unlike examples reported for other taxa and genes (65-67).

4.4 Gene size as cofactor of molecular adaptive strategies

On average, the tendencies of *K*-strategists to protect conserved sites are not overwhelmingly lower than those of *r*-strategists. However, the difference becomes striking when considering

gene length. Indeed, in *K*-strategists, the tendency to protect conserved sites was strongest in short genes, and decreased with gene length. The opposite was true in *r*-strategists. Hence, for example, in the relatively large gene ND2, all *K*-strategists tend to expose conserved sites, while a significant majority of *r*-strategists protect them (results significant by sign tests, Table 1).

One can presume that small genes have more regulatory functions, while larger ones tend to have more catalytic, basic maintenance functions. Following this rationale, *r*-strategists would maximize the potential for functional evolution of regulatory functions while being relatively conservative in the evolution of house keeping functions. *K*-strategists would maximize the evolutionary potential of large genes, and be conservative for regulatory genes. The larger reserves and bodies of *K*-strategists enable them to "experiment" at the level of house keeping functions in some body parts without being lethal for the individual plant, but their long life cycles and low reproduction rates make such experimenting at the level of regulatory functions unpractical. The opposite rationale holds with *r*-strategists, who possess relatively little reserves and bodies, but large offspring numbers and short life cycles. The encouraging similarity of these results with results obtained from analyzing chloroplast genomes (29; 8 species common between these two analyses) suggests this to be a general adaptive strategy for plants rather than an organelle-specific strategy.

4.5 Gradual and saltatory evolution

We suggest that protection of conserved sites tends towards gradual, micro-adaptive evolution, while exposing these sites leads to saltatory evolution. The results indicate that the gradual mode is the most frequent, but less so in *r*-strategists. They also suggest that the latter would tend to evolutionary macro-adaptations for regulatory functions, while *K*-strategists would tend to macro-adaptive changes for basic maintenance functions. These rationales are again

compatible with the general concepts of *r*- and *K*- strategies. It is likely that the association between site variability and secondary structure varies not only among plants and protein species, as well as regions of a sequence, but also among the 20 amino acids. Further studies, based on more ample species sampling, and on comparisons taking into account the phylogenetic relatedness among the various plant species, as well as less dichotomic definitions of *r*- and *K*- strategies, are likely to elucidate the links indicated by this study between molecular and whole organism levels of adaptation.

5. Conclusions

1. Sites coding at mRNA level for functionally important sites in the protein are mostly protected by secondary structure. However, in a quarter of the possible combinations of mitochondrial genes and species, the opposite relationship exists: variable sites are more protected by forming parts of helices and conserved sites being in the loop regions.

2. Second codon positions showed greater protection of conserved sites than first and third codon positions, where trends were weaker.

3. In some cases, the third codon position showed greater than expected exposure of conserved sites.

4. Overall, *K*-strategists protect conserved sites lesser than *r*-strategists, despite greater overall stability of secondary structures.

5. This was true for the first and second codon positions, but at third codon positions, *K*-strategists protected conserved sites more than *r*-strategists.

6. Conserved sites were more protected in short than large genes in *K*-strategists. This relationship with gene size is reversed in *r*-strategists.

7. Overall secondary structure stability of genes increases with gene length and is greater for *K*-strategists than *r*-strategists, independent of gene size.

8. In some genes for most plants, we observe a tendency to protect conserved sites, in fewer species-gene combinations, a tendency to expose them. We did not detect any tendency in a larger than expected proportion, suggesting that these sequences consist of a mix of sites that follow each of the 2 strategies, defining a third category of molecular adaptation.

9. The effects of variability or conservation levels of helix forming sites on stability, for the same genes are often opposite for *r*- and *K*-strategists, suggesting that interactions between gene composition, gene size and codon position factors ultimately affect gene stability oppositely in the two extreme life history types.

Literature Cited

1. Gagliardi, D., Stepien, P. P., Temperley, R. J., Lightowlers, R. N., Chrzanowska-Lightowlers, Z. M. 2004. Messenger RNA stability in mitochondria: different means to an end. Trends Genet. 20, 260-267.

2. Gingerich, T. J., Feige, J. J., LaMarre, J. 2004. AU-rich elements and the control of gene expression through regulated mRNA stability. Anim. Health Res. Rev. 5, 49-63.

3. Gutiérrez, R. A., MacIntosh, G. C., Green, P. J. 1999. Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control. Trends Plant Sci. 4, 1360-1385.

4. Hollams, E. M., Giles, K. M., Thomson, A. M., Leedman, P. J. 2002. mRNA stability and the control of gene expression: implications for human disease. Neurochem. Res. 27, 957-980.

5. Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S. C., Gram, H., Han, J. 2005. Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell 120, 623-634.

6. Marashi, S. A., Eslahchi, C., Pezeshk, H., Sadeghi, M. 2006. Impact of RNA structure on the prediction of donor and acceptor splice sites. BMC Bioinformatics 7, 297.

7. McHardy, A. C., Puhler, A., Kalinowski, J. Meyer, F. 2004. Comparing expression level-dependent features in codon usage with protein abundance: an analysis of predictive proteomics. Proteomics 4, 46-58.

8. Monde, R. A., Schuster, G., Stern, D. B. 2000. Processing and degradation of chloroplast mRNA. Biochimie 82, 573-582.

9. Ohme-Takagi, M., Taylor, C. B., Newman, T. C., Green, P. J. 1993. The effect of sequences with high AU content on mRNA stability in tobacco. Proc. Natl. Acad. Sci. U. S. A. 90, 11811-11815.

10. Qi, C., Pekala, P. H. 1999. The influence of mRNA stability on glucose transporter (GLUT1) gene expression. Biochem. Biophys. Res. Commun. 263, 265-269.

11. Storz, G., Opdyke, J. A., Zhang, A. 2004. Controlling mRNA stability and translation with small, noncoding RNAs. Curr. Opin. Microbiol. 7, 140-144.

12. Tourrière, H., Chebli, K., Tazi, J. 2002. mRNA degradation machines in eukaryotic cells. Biochimie 84, 821-837.

13. Wang, L., Wessler, S. R. 2001. Role of mRNA Secondary Structure in Translational Repression of the Maize Transcriptional Activator Lc1,2. Plant Physiol. 125, 1380–1387.

14. Zhang, W., Xiao, W., Wei, H., Zhang, J., Tian, Z. 2006. mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in Escherichia coli. Biochem. Biophys. Res. Commun. 349, 69-78.

15. Carlini, D. B., Chen, Y., Stephan, W. 2001. The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. Genetics 159, 623-633.

16. Duan, J., Antezana, M. A. 2003. Mammalian mutation pressure, synonymous codon choice, and mRNA degradation. J. Mol. Evol. 57, 694-701.

17. Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., Cossart, P. 2002. An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. Cell 110, 551-561.

18. Morita, M. T., Kanemori, M., Yanagi, H., Yura, T. 2000. Dynamic interplay between antagonistic pathways controlling the sigma(32) level in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 97, 5860-5865.

19. Yuzawa, H., Nagai, H., Mori, H., Yura, T. 1993. Heat induction of sigma 32 synthesis mediated by mRNA secondary structure: a primary step of the heat shock response in Escherichia coli. Nucleic Acids Res. 21, 5449-5455.

20. Babendure, J. R., Babendure, J. L., Ding, J. H., Tsien, R. Y. 2006. Control of mammalian translation by mRNA structure near caps. RNA 12, 851-861.

21. Seligmann, H., Pollock, D. D. 2003. Function and evolution of secondary structure in human mitochondrial mRNAs. Midsouth Computational and Bioinformatics Society, Abstract 26.

22. Krishnan, N. M., Seligmann, H., Raina, S. Z., Pollock, D. D. 2004a. Detecting Gradients of Asymmetry in Site-Specific Substitutions in Mitochondrial Genomes. DNA Cell Biol. 23, 707-714.

23. Krishnan, N. M., Seligmann, H., Raina, S. Z., Pollock, D. D. 2004b. Phylogenetic analysis of site-specific perturbations in asymmetric mutation gradients; in Currents in Computational Molecular Biology (eds) A Gramada and P E Bourne pp 266-267.

24. Shabalina, S. A., Ogurtsov, A. Y., Spiridonov, N. A. 2006. A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res. 34, 2428-2437.

25. Jia, M., Luo, L. 2006. The relation between mRNA folding and protein structure. Biochem. Biophys. Res. Commun. 343, 177-182.

26. Morin, X., Chuine, I. 2006. Niche breadth, competitive strength and range size of tree species: a trade-off based framework to understand species distribution. Ecol. Lett. 9, 185-195.

27. Gilbert, B., Wright, S. J., Muller-Landau, H. C., Kitajima, K., Hernandez, A. 2006. Life history trade-offs in tropical trees and lianas. Ecology 87, 1281-1288.

28. von Arx, G., Edwards, P. J., Dietz, H. 2006. Evidence for life history changes in high-altitude populations of three perennial forbs. Ecology 87, 665-674.

29. Krishnan, N.M., Seligmann, H., Rao, B.J. (2008). Relationship between mRNA secondary structure and sequence variability in Chloroplast genes: possible life history implications. BMC Genomics 9:48.

30. Aguadé, M. 1999. Positive Selection Drives the Evolution of the Acp29AB Accessory Gland Protein in Drosophila. Genetics 152, 543-551.

31. Barbash, D. A., Awadalla, P., Tarone, A. M. 2004. Functional Divergence Caused by Ancient Positive Selection of a Drosophila Hybrid Incompatibility Locus. PLoS Biol. 2, 839-848.

32. Bielawski, J. P., Yang, Z. 2004. A Maximum Likelihood Method for Detecting Functional Divergence at Individual Codon Sites, with Application to Gene Family Evolution. J. Mol. Evol. 59, 121–132.

33. Biermann, C. H. 1998. The molecular evolution of sperm binding in six species of sea urchin (Echinoida: Strongylocentrotidae). Mol. Biol. Evol. 15, 1761-1771.

34. Castillo-Davis, C. I., Kondrashov, F. A., Hartl, D. L., Kulathinal, R. J. 2004. The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. Genome Res. 14, 802–811.

35 Comeron, J.M., Kreitman, M. 1998. The correlation between synonymous and nonsynonymous substitutions in Drosophila: mutation, selection or relaxed constraints? Genetics 150:767-775.

36. Metz, E. C., Palumbi, S. R. 1996. Positive selection and sequence rearrangements generate extensive polymorphisms in the gamete recognition protein binding. Mol. Biol. Evol. 13, 397-406.

37. Mindell, D. P. 1996. Positive selection and rates of evolution in immunodeficiency viruses from humans and chimpanzees. Proc. Natl. Acad. Sci. U. S. A. 93, 3284–3288.

38. Riginos, C., McDonald, J. H. 2003. Positive selection on an acrosomal sperm protein, M7 lysin, in three species of the Mussel genus Mytilus. Mol. Biol. Evol. 20, 200-207.

39. Schully, S. D., Hellberg, M. E. 2006. Positive selection on nucleotide substitutions and indels in accessory gland proteins of the Drosophila pseudoobscura subgroup. J. Mol. Evol. 62, 793-802.

40.Bennetzen, J. L., Hall, B. D. 1982. Codon selection in yeast. J. Biol. Chem. 257, 3026-3031.

41. Ikemura, T. 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. 158, 573-597.

42. Seligmann, H., Pollock, D. D. 2004. The ambush hypothesis: off frame stop codons arrest early accidental frameshited transcription. DNA Cell Biol. 23, 701-705.

43 Skulachev, V. P. 1998. Biochemical mechanisms of evolution and the role of oxygen Biochem-MOSCOW 63, 1335-1343.

44. Turmel, M., Lemieux, C., Burger, G., Lang, B. F., Otis, C., Plante, I., Gray, M. W.1999. The complete mitochondrial DNA sequences of Nephroselmis olivacea and Pedinomonas minor: Two radically different evolutionary patterns within green algae. Plant Cell 11, 1717-1730.

45. Wolff, G., Plante, I., Lang, B. F., Kuck, U., Burger, G. 1994. Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii: Gene content and genome organization. J. Mol. Biol. 237, 75-86.

46. Pombert, J. F., Otis, C., Lemieux, C., Turmel, M. 2004. The complete mitochondrial DNA sequence of the green alga Pseudendoclonium akinetum (Ulvophyceae) highlights distinctive

evolutionary trends in the Chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae. Mol. Biol. Evol. 21, 922-935.

47. Winkler, M., Kuck, U. 1991. The group IIB intron from the green alga Scenedesmus obliquus mitochondrion: molecular characterization of the in vitro splicing products. Curr. Genet. 20, 495-502.

48. Giege, P., Brennicke, A. 1999. RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. Proc. Natl. Acad. Sci. U. S. A. 96, 15324-15329.

49. Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., Mikami, T. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res. 28, 2571-2576.

50. Handa, H. 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica napus L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. Nucleic Acids Res. 31, 5907-5916.

51. Sugiyama, Y., Watase, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A., Sugiura, M. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol. Genet. Genomics 272, 603-615.

52. Tian, X., Zheng, J., Hu, S., Yu, J. 2006. The Rice Mitochondrial Genomes and Their Variations. Plant Physiol. 142, 401-410.

53. Ogihara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., Shiina, T., Miyashita, N., Nasuda, S., Nakamura, C., Mori, N., Takumi, S., Murata, M., Futo, S., Tsunewaki, K. 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 33, 6235-6250.

54. Clifton, S. W., Minx, P., Fauron, C. M., Gibson, M., Allen, J. O., Sun, H., Thompson, M., Barbazuk, W. B., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, R. K., Newton, K. J. 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol. 136, 3486-3503.

55. Turmel, M., Otis, C., Lemieux, C. 2002a. The chloroplast and mitochondrial genome sequences of the charophyte Chaetosphaeridium globosum: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc. Natl. Acad. Sci. U. S. A. 99, 11275-11280.

56. Turmel, M., Otis, C., Lemieux, C. 2003. The Mitochondrial Genome of Chara vulgaris: Insights into the Mitochondrial DNA Architecture of the Last Common Ancestor of Green Algae and Land Plants. Plant Cell 15, 1888-1903.

57. Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohyama, K. 1992. Transfer RNA genes in the mitochondrial genome from a liverwort, Marchantia polymorpha: the absence of chloroplast-like tRNAs. Nucleic Acids Res. 20, 3773-3777.

58. Turmel, M., Otis, C., Lemieux, C. 2002b. The complete mitochondrial DNA sequence of Mesostigma viride identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. Mol. Biol. Evol. 19, 24-38.

59. Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M., Sato, N. 2007. The mitochondrial genome of the moss Physcomitrella patens sheds new light on mitochondrial evolution in land plants. Mol Biol Evol 24, 699–709.

60. Brookfield, J. F. Y. 1986. The evolution of r-strategies and K-strategies. Biol. J. Linn. Soc. 27, 165-178.

61. Thompson, J. D., Higgins, D. G., Gibson, T. J. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

62. Zuker, M. 2003. mFold web server for nucleic acids folding and hybridization prediction. Nucleic Acids Res. 31, 3406-3415.

63. Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) version 3. 67. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle

64. Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Tree 11, 367-371.

65. Seligmann, H. 2003. Cost minimization of amino acid usage. J. Mol. Evol. 56, 151-161.

66. Brocchieri, L., Karlin, S. 2005. Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Res. 33, 3390-3400.

67. Warringer, J., Blomberg, A. 2006. Evolutionary constraints on yeast protein size. Biomed. Cent. Evol. Biol. 6, 61-71.