

# Nucleotide Substitution Rates in Legume Chloroplast DNA Depend on the Presence of the Inverted Repeat

Antoinette S. Perry, Kenneth H. Wolfe

Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland

Received: 30 November 2001 / Accepted: 27 March 2002

**Abstract.** The chloroplast genomes of some species of legumes lack the large inverted repeat (IR) that is a trademark of most land-plant chloroplasts. Our analysis of chloroplast genes in legume species that have an IR shows that the synonymous (silent) substitution rate in IR genes is 2.3-fold lower than in single-copy (SC) genes, which is largely in agreement with earlier findings. Given that all genes in species that lack the IR are single-copy, what level of synonymous substitution exists in these genes? We report a uniform substitution rate in IR-less genomes, and moreover, we find this rate to be at the level otherwise reserved for SC genes. In other words, the synonymous substitution rate has accelerated in the remaining copy of the duplicate region. We propose that this acceleration is a direct result of the decrease in the copy number of the sequence, rather than an intrinsic property of the genes normally located in the IR.

**Key words:** Inverted repeat — Legume chloroplast — DNA — Nucleotide substitution rates

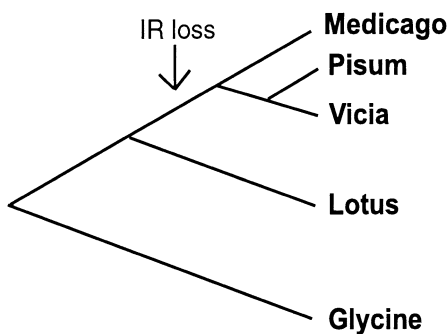
## Introduction

The outstanding structural feature of the chloroplast genome of most land plants is a large inverted repeat (IR). The two copies of the repeat are completely

identical and are separated by one large and one small single-copy region (LSC and SSC, respectively). The IR is typically 25 kb in angiosperms and contains genes coding for four rRNAs, six tRNAs, and seven proteins (Palmer 1991; Sugiura 1992; Gaut 1998).

Many species of legumes (in the subfamily Papilionoideae) such as *Pisum sativum* (pea), *Vicia faba* (broad bean), and *Medicago truncatula* (barrel medic) are exceptions to this pattern and lack the IR, while other legumes such as *Lotus japonicus* and *Glycine max* (soybean) have retained the repeat (Kolodner and Tewari 1979; Koller and Delius 1980; Palmer and Thompson 1981; Palmer et al. 1988; Kato et al. 2000). Clearly, the presence of the IR is not obligatory for chloroplast function. The structurally conserved features of the IR such as its size, position, gene content, gene orientation, and almost-universal presence in land plants allow it to be considered an evolutionary relic which originated early in plastid evolution and has remained widespread not because it confers a selective advantage upon the plant but, instead, because the structure itself is resistant to recombinational loss (Palmer and Thompson 1981; Lavin et al. 1990; Liston 1995). The loss of the IR is hypothesized to have occurred by means of a single deletion of one complete copy of the repeat during the evolution of the subfamily Papilionoideae (Fig. 1) (Wolfe 1988; Lavin et al. 1990; Herdenberger et al. 1990). There are more major rearrangements within the legume family than among the rest of the angiosperms; however, most of the genomes that have lost the IR have undergone more rearrangements than those that have not (Palmer and Thompson 1981, 1982).

Correspondence to: Prof. Kenneth H. Wolfe; email: khwolfe@tcd.ie



**Fig. 1.** Phylogeny of the legume species analyzed in this study (subfamily Papilionoideae), indicating the loss of the inverted repeat (Lavin et al. 1990; Doyle 1995; Doyle et al. 1997; Palmer et al. 1998).

Through restriction analysis and the measurement of numbers of nucleotide substitutions, it has been found that genes in the IR have lower rates of nucleotide substitution than genes in the SC regions of chloroplast genomes (Curtis and Clegg 1984; Wolfe et al. 1987). In studies where the levels of synonymous nucleotide substitutions were calculated in IR- and SC-located genes compared between a pair of species, the extent of substitution was always found to be higher in SC genes, by a factor of between 4- and 17-fold (Wolfe et al. 1987; Wolfe 1991; Gaut 1998). Several of these comparisons involved small numbers of genes, but in the comparison between the complete tobacco and the complete rice chloroplast genomes, the difference was about five-fold (Wolfe 1991; Gaut 1998). Because silent nucleotide substitutions are likely to be neutral with regard to natural selection, Wolfe et al. (1987) proposed that the difference in substitution rates between the IR and the SC was due either to a reduced mutation rate in the IR or to a bias against mutant sequences during DNA repair. Birky and Walsh (1992) pointed out that for chloroplast genes these two possibilities are actually the same thing, because the measured mutation rate is the rate of appearance of homoplasmic cells, in which the mutant site has already become fixed within the multiple copies of the genome per organelle and within the multiple organelles per cell. They showed that if gene conversion has even a very small bias in favor of wild-type sequences, the difference in copy number between IR and SC genes will result in a difference in the rate of appearance of homoplasmic mutant cells ( $v$ ), even if the point mutation rate per genome per generation ( $\mu$ ) is the same in SC and IR genes.

If genes in the IR of chloroplast (cp)DNA have a significantly lower synonymous substitution rate ( $K_s$ ) than genes in the SC regions, then what rate exists for such genes in species with only one segment of the repeat? If a biased gene conversion mechanism is in operation, then the model of Birky and Walsh (1992)

predicts that the reduced nucleotide substitution rate would disappear in IR-less genomes. Thus all genes in an IR-less genome are predicted to have the same synonymous substitution rate. Moreover, if the number of chloroplast genomes per cell remains unchanged following deletion of the IR, then the copy number of the formerly IR genes becomes halved, whereas the copy number of the SC genes is unchanged, so the uniform substitution rate of an IR-less genome is predicted to be equal to the rate previously seen in the SC.

We carried out two tests using legume cpDNA sequences. The first investigates whether the synonymous substitution rate is uniform in IR-less genomes. The second uses relative rate tests to examine whether, in a genome that has lost the IR, the formerly IR-located genes accelerate, or the formerly SC-located genes decelerate, with respect to their synonymous substitution rates. Our results are consistent with predictions of the Birky and Walsh (1992) model.

## Data and Methods

The fully sequenced and annotated *Lotus japonicus* chloroplast genome (GenBank accession No. AP002983) was used as a starting point for finding sequences. *Lotus* has a total of 84 chloroplast protein coding genes, including 70 SC-located and 7 IR-located genes (Kato et al. 2000). The protein sequence for each of these genes was used as a query in a TBLASTN similarity search against the nonredundant (nr) protein database at the NCBI web site. The results were filtered to show results only from the family Fabaceae. Sequences from *Pisum* (GenBank accession Nos. M11005, X03853, X03852, AJ004882, AF223227, X05423, X05394, M27309, X03912, X05917, M57711, X56315, X53525, K01516, X15767, X82511, AF153442, X00535, X15645, X05029, X59015, X14804, X13157, X59015, Z80872) and *Glycine* (GenBank accession Nos. X00152, AF142700, Z95552, X64329, AF289093, M18442, U26948, X07676, BG044152, BG839779, AF347627, AF238058, X06429, X07675, X05013) which had a similarity score of 70 or higher were used to obtain the corresponding genes in these species. Sequences were not available for a number of *Pisum* genes and these were taken from the unannotated *Pisum sativum* chloroplast genome (J. Gray, personal communication). Genes in the sequenced but unannotated *Medicago truncatula* chloroplast genome (S. Lin, H. Wu, H. Jia, P. Zhang, R. Dixon, G. May, R. Gonzales, and B.A. Roe, unpublished data; GenBank accession No. AC093544) were located using TBLASTN.

We used EST data to augment the relatively small number of annotated *Glycine* chloroplast sequences that were available in GenBank. *Glycine* ESTs with significant BLASTN matches to the *Lotus* chloroplast genome were downloaded and assembled into contigs using **phrap** (P. Green, unpublished, 1994–1999; <http://www.phrap.org>). GenBank accession numbers for EST representatives of the contigs used are BG839049, AW277587, BE660484, BF425151, BE800200, BF597550, BI674546, AW185635, BG727066, BI426407, AW200884, AW757173, X06429, BF598451, BI095058, AW311191, BG041795, BF597254, and BI972886. Protein sequences from the *Lotus* genome were then searched against this database of contigs using TBLASTN to identify genuine *Glycine* chloroplast protein coding sequences. Only genes for which the number of synonymous sites ( $L_s$ ) was higher than 50 were used for analysis.

Sets of homologous legume gene sequences were translated and protein sequences were aligned using ClustalW (Thompson et al. 1994). The protein sequence alignments were then used as templates to make DNA sequence alignments with gaps at equivalent positions. The method of Li (1993) was used to estimate the number of synonymous substitutions per synonymous site ( $K_s$ ). The standard error of the mean  $K_s$  was calculated as the square root of the mean variance according to the method of Wolfe et al. (1987).

In the absence of knowledge of divergence times between species, rates of evolution can be compared between lineages by the "relative rate" method (Wu and Li, 1985; Gaut 1998). This method requires three homologous nucleotide sequences for each gene studied: two ingroup sequences (*Lotus* and *Medicago*) and an outgroup sequence (*Glycine*). Synonymous nucleotide substitution rates were calculated between an ingroup and an outgroup sequence and overall  $K_s$  values for IR and SC regions were obtained by pooling gene values and weighting each according to its number of sites compared ( $L_s$ ).

## Results and Discussion

We first calculated synonymous substitution rates between genes for which sequence data were available in both the IR-containing genomes, *Lotus* and *Glycine*, and the IR-less genomes, *Medicago* and *Pisum*. We compared genes between *Lotus* and *Glycine* and found that SC genes have a 2.3-fold higher nucleotide substitution rate than genes present twice, in the IR. The mean  $K_s$  value between *Lotus* and *Glycine* for six IR-located genes (*rpl2*, *rpl23*, *rps7*, *ndhB*, and the IR-located portions of *ycf1* and *rps12*) is  $0.120 \pm 0.018$ , compared with a mean of  $0.282 \pm 0.008$  for 30 SC genes (Fig. 2). Therefore, our  $K_s$  data for these legumes confirm the earlier findings by Wolfe et al. (1987), that chloroplast genomes with an IR display two levels of synonymous substitution and that the rate depends on the location and hence the copy number of the gene. When the same set of 30 SC and 6 IR genes is compared between *Medicago* and *Pisum*, the 2.3-fold difference in synonymous substitution rates is not maintained in these genomes where the repeat has been lost. The six formerly IR-located genes have a mean  $K_s$  of  $0.173 \pm 0.021$ , while the rate calculated for the SC genes is, remarkably, lower, at  $0.136 \pm 0.006$  (Fig. 2).

We next used all available gene sequences from the fully sequenced *Medicago* and *Pisum* genomes to produce a more comprehensive calculation of  $K_s$  between these two IR-less legumes. We were able to compare all seven protein coding IR-located genes and 49 of 70 SC genes (all genes with  $\geq 50$  synonymous sites were compared). The mean  $K_s$  value ( $0.219 \pm 0.015$ ) for the genes that are normally located in the IR was again higher than the mean  $K_s$  for the SC genes ( $0.135 \pm 0.005$ ) (Table 1).

However, it is clear from Table 1 that the average results are being influenced by a small number of very fast-evolving genes. The level of nonsynonymous substitution ( $K_a$ ) reflects the rate of evolution of an

amino acid sequence. Table 1 shows that two SC genes, *ycf4* and *accD*, have significantly higher  $K_a$  values than other chloroplast SC genes studied and hence, are evolving at much faster rates. Similarly, two IR-located genes, *ycf1* and *ycf2* (sequence data were not available for *Lotus* versus *Glycine*) have much higher nonsynonymous rates than their neighboring IR genes. There is a correlation between  $K_s$  and  $K_a$  in almost all genomes (Li 1997), so the high  $K_a$  values may be artificially inflating the  $K_s$  values for these genes. Typically, in any given gene  $K_a < K_s$ , but each of these genes has a  $K_a/K_s$  ratio approaching 1, which suggests that these genes are under very little evolutionary constraint in legumes and are accumulating amino acid changes almost as quickly as silent ones. In fact, *ycf4* is not shown in Fig. 2 because its  $K_s$  value between *Medicago* and *Pisum* is outside the boundaries of the plot. Although *ycf4* was reported by Southern blot analyses to be missing from the chloroplast genomes of several legumes (Doyle et al. 1995), homologous but very diverged ORFs are present in all four legume species analyzed here. We suggest that the substitution rate in another fast-evolving gene, *ycf1*, might be affected by its location on the border of the IR and the SC, especially as shifts of these border positions have been reported (Goulding et al. 1996; Kato et al. 2000).

Elimination of the fastest-evolving genes (those with a  $K_a \geq 0.15$ , i.e., *ycf4*, *accD*, *clpP*, *ycf1*, and *ycf2*) reveals a three-fold difference in synonymous substitution rates between SC and IR genes in the IR-containing genomes, *Lotus*, and *Glycine* (Table 1). In sharp contrast, when the high- $K_a$  genes are omitted, comparisons involving IR-less genomes (*Medicago* versus *Pisum*) show a uniform synonymous substitution rate (Table 1). The five remaining genes that are in the IR (*rpl2*, *rpl23*, *ndhB*, *rps7*, and *rps12*) have an average  $K_s$  of  $0.114 \pm 0.017$ , which is not significantly different from the rate calculated for 46 SC genes,  $0.119 \pm 0.004$ .

To determine whether this homogenous synonymous substitution rate seen in IR-less genomes is due to an acceleration in the rate of the formerly IR genes or a retardation in the rate of SC genes, we compared the relative rates of substitution in the two sets of genes in the outgroup *Glycine* (with an IR) versus *Lotus* (with) and versus *Medicago* (without) (see Fig. 1). The average  $K_s$  value was calculated for each pair for 27 SC genes and 6 IR genes. We found the  $K_s$  values in SC genes for the *Glycine/Lotus* and *Glycine/Medicago* comparisons to be very similar,  $0.280 \pm 0.008$  and  $0.284 \pm 0.008$ , respectively. A plot of  $K_s$  values for the 27 SC genes shows that the rates of synonymous substitution of these genes in the *Medicago* and *Lotus* lineages have been equal since their divergence from the *Glycine* lineage (Fig. 3). This result suggests that SC genes are subject to the same

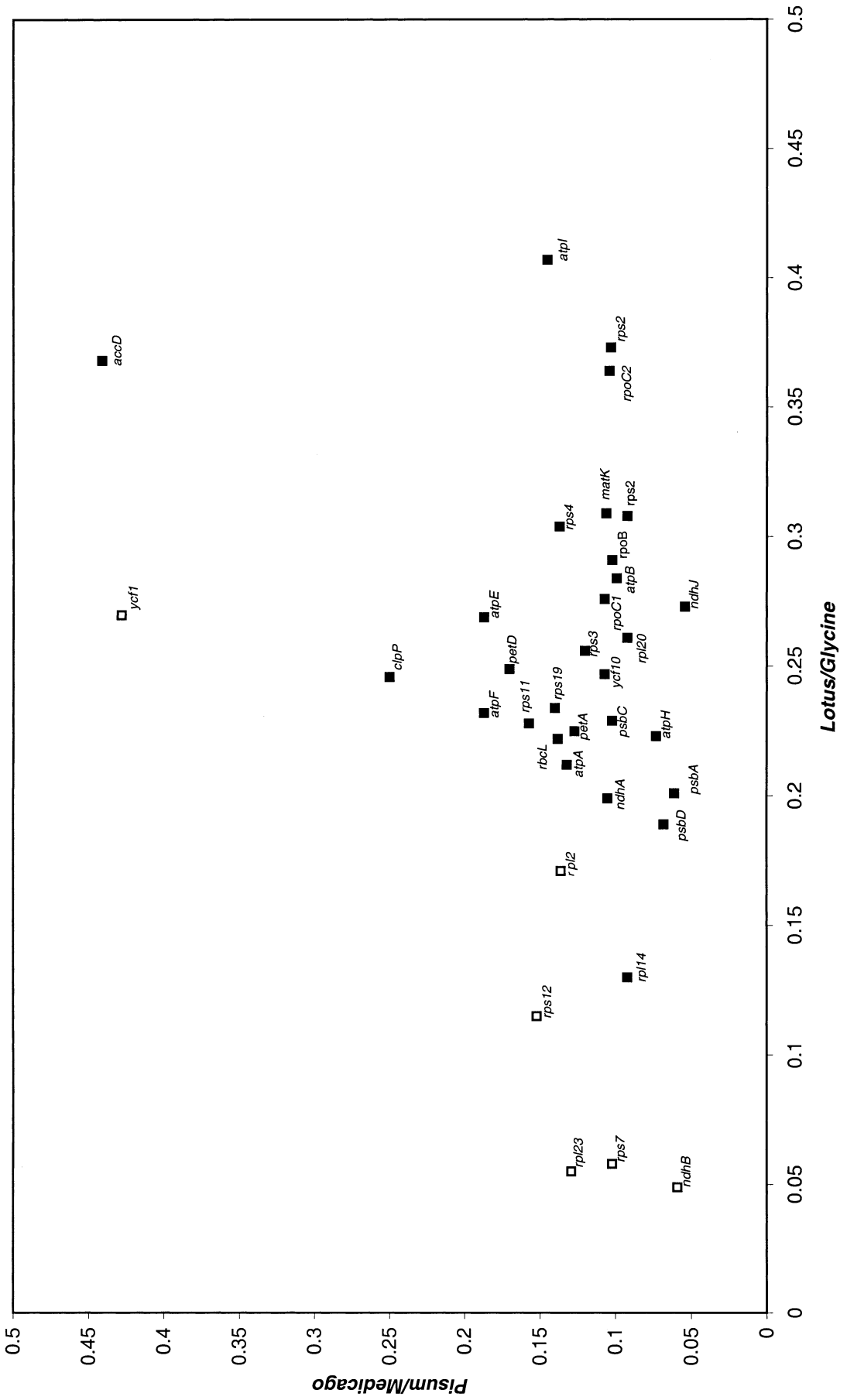
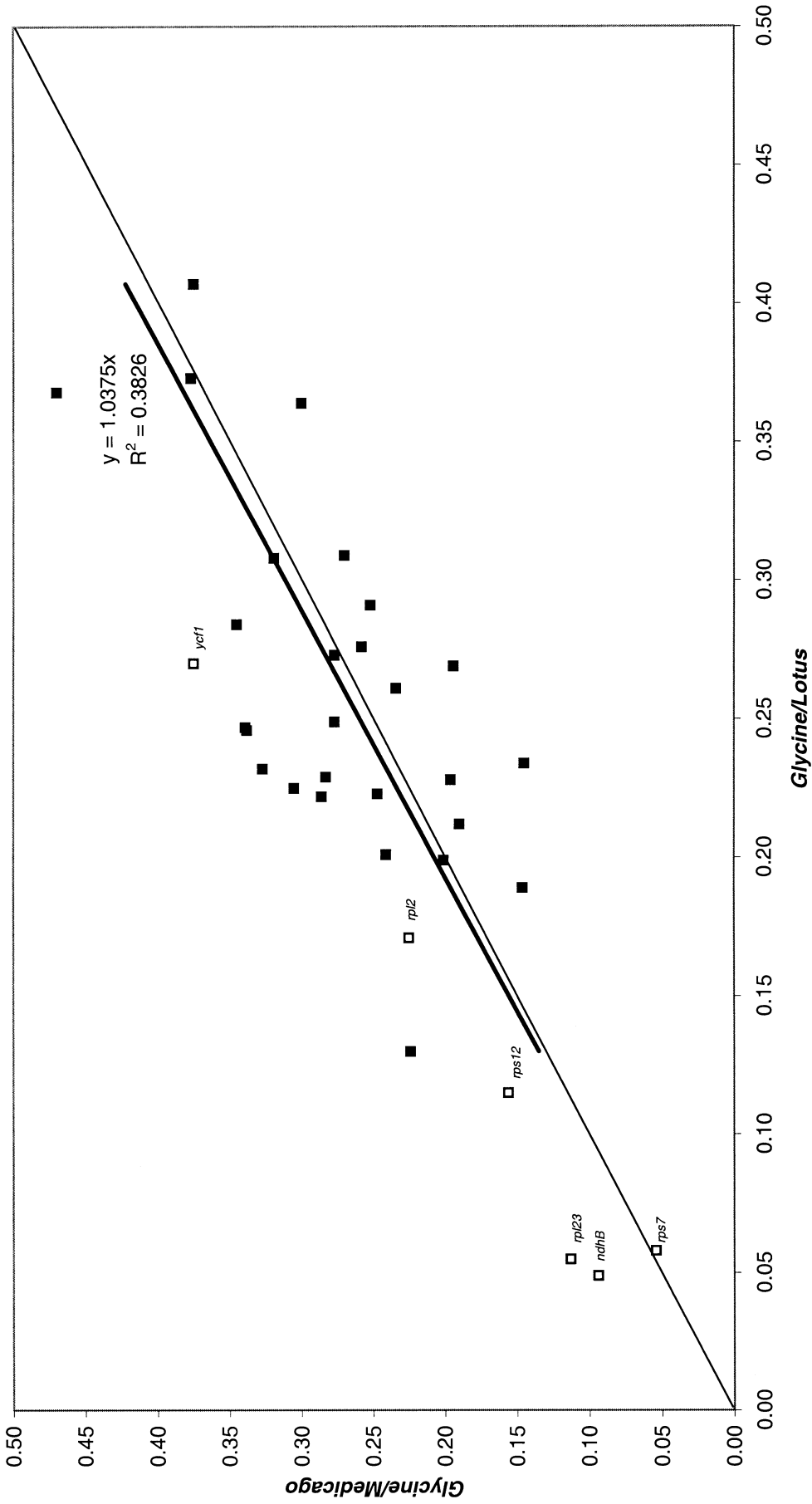


Fig. 2. Comparison of  $K_s$  values between *Lotus* and *Glycine* (both containing an IR) and between *Medicago* and *Pisum* (both lacking an IR), for 30 SC-located and 6 IR-located genes. SC and IR genes are represented by filled and open boxes, respectively.

**Table 1.**  $K_s$  and  $K_a$  values between *Medicago* and *Pisum* and between *Lotus* and *Glycine*

| Gene                       | <i>Medicago/Pisum</i> |               | <i>Lotus/Glycine</i> <sup>a</sup> |               |
|----------------------------|-----------------------|---------------|-----------------------------------|---------------|
|                            | $K_s$                 | $K_a$         | $K_s$                             | $K_a$         |
| <i>SC genes</i>            |                       |               |                                   |               |
| <i>ycf4</i>                | 0.646                 | 0.537         | 0.459                             | 0.458         |
| <i>accD</i>                | 0.441                 | 0.303         | 0.368                             | 0.182         |
| <i>clpP</i>                | 0.249                 | 0.150         | 0.246                             | 0.031         |
| <i>rps3</i>                | 0.120                 | 0.142         | 0.256                             | 0.056         |
| <i>rp120</i>               | 0.092                 | 0.120         | 0.261                             | 0.093         |
| <i>rps14</i>               | 0.271                 | 0.110         | —                                 | —             |
| <i>atpE</i>                | 0.187                 | 0.084         | 0.269                             | 0.068         |
| <i>ycf10</i>               | 0.107                 | 0.073         | 0.247                             | 0.083         |
| <i>rps19</i>               | 0.140                 | 0.053         | 0.234                             | 0.041         |
| <i>matK</i>                | 0.106                 | 0.051         | 0.309                             | 0.148         |
| <i>rps15</i>               | 0.271                 | 0.049         | —                                 | —             |
| <i>atpF</i>                | 0.187                 | 0.045         | 0.232                             | 0.050         |
| <i>rps11</i>               | 0.157                 | 0.041         | 0.228                             | 0.045         |
| <i>rpoC2</i>               | 0.104                 | 0.038         | 0.364                             | 0.097         |
| <i>rps8</i>                | 0.143                 | 0.036         | —                                 | —             |
| <i>rps4</i>                | 0.131                 | 0.033         | 0.304                             | 0.169         |
| <i>psbH</i>                | 0.164                 | 0.032         | —                                 | —             |
| <i>rpoA</i>                | 0.125                 | 0.029         | —                                 | —             |
| <i>rps2</i>                | 0.103                 | 0.029         | 0.373                             | 0.135         |
| <i>petB</i>                | 0.190                 | 0.028         | —                                 | —             |
| <i>ycf5</i>                | 0.140                 | 0.028         | —                                 | —             |
| <i>ndhD</i>                | 0.154                 | 0.026         | —                                 | —             |
| <i>rbcL</i>                | 0.138                 | 0.026         | 0.222                             | 0.018         |
| <i>psaB</i>                | 0.180                 | 0.025         | —                                 | —             |
| <i>atpA</i>                | 0.132                 | 0.024         | 0.212                             | 0.020         |
| <i>ndhJ</i>                | 0.054                 | 0.022         | 0.273                             | 0.069         |
| <i>petA</i>                | 0.127                 | 0.021         | 0.225                             | 0.040         |
| <i>rpoC1</i>               | 0.107                 | 0.021         | 0.276                             | 0.044         |
| <i>atpB</i>                | 0.099                 | 0.020         | 0.284                             | 0.035         |
| <i>psbG</i>                | 0.092                 | 0.020         | 0.308                             | 0.099         |
| <i>ndhG</i>                | 0.107                 | 0.016         | —                                 | —             |
| <i>psaC</i>                | 0.037                 | 0.016         | —                                 | —             |
| <i>psbB</i>                | 0.122                 | 0.014         | —                                 | —             |
| <i>ndhH</i>                | 0.107                 | 0.013         | —                                 | —             |
| <i>ndhA</i>                | 0.105                 | 0.013         | 0.199                             | 0.073         |
| <i>rpoB</i>                | 0.102                 | 0.012         | 0.291                             | 0.051         |
| <i>atpH</i>                | 0.073                 | 0.012         | 0.223                             | 0.000         |
| <i>ndhF</i>                | 0.120                 | 0.011         | —                                 | —             |
| <i>psaA</i>                | 0.086                 | 0.011         | —                                 | —             |
| <i>psbC</i>                | 0.102                 | 0.010         | 0.229                             | 0.049         |
| <i>ndhI</i>                | 0.087                 | 0.010         | —                                 | —             |
| <i>psbD</i>                | 0.068                 | 0.007         | 0.189                             | 0.000         |
| <i>ndhC</i>                | 0.068                 | 0.007         | —                                 | —             |
| <i>rpl14</i>               | 0.092                 | 0.006         | 0.130                             | 0.025         |
| <i>psbE</i>                | 0.041                 | 0.005         | —                                 | —             |
| <i>petD</i>                | 0.170                 | 0.004         | 0.249                             | 0.032         |
| <i>psbA</i>                | 0.061                 | 0.003         | 0.201                             | 0.004         |
| <i>atpI</i>                | 0.145                 | 0.002         | 0.407                             | 0.121         |
| <i>ycf3</i>                | 0.103                 | 0.000         | —                                 | —             |
| Mean                       | 0.135 ± 0.005         | 0.041 ± 0.001 | 0.282 ± 0.008                     | 0.076 ± 0.002 |
| Adjusted mean <sup>b</sup> | 0.119 ± 0.00          | 0.026 ± 0.001 | 0.275 ± 0.008                     | 0.063 ± 0.002 |
| <i>IR genes</i>            |                       |               |                                   |               |
| <i>ycf1</i> <sup>c</sup>   | 0.428                 | 0.178         | 0.270                             | 0.137         |
| <i>ycf2</i>                | 0.247                 | 0.159         | —                                 | —             |
| <i>rpl23</i>               | 0.129                 | 0.109         | 0.055                             | 0.039         |
| <i>rps7</i>                | 0.102                 | 0.058         | 0.058                             | 0.007         |
| <i>rps12</i> <sup>c</sup>  | 0.152                 | 0.023         | 0.115                             | 0.005         |
| <i>rpl2</i>                | 0.136                 | 0.020         | 0.171                             | 0.020         |
| <i>ndhB</i>                | 0.059                 | 0.017         | 0.049                             | 0.025         |
| Mean                       | 0.219 ± 0.015         | 0.125 ± 0.005 | 0.120 ± 0.018                     | 0.040 ± 0.005 |
| Adjusted mean <sup>b</sup> | 0.114 ± 0.017         | 0.039 ± 0.005 | 0.092 ± 0.015                     | 0.018 ± 0.004 |

<sup>a</sup> Dashes indicate genes for which no *Glycine* sequence was available.<sup>b</sup> Mean excluding genes with a  $K_a \geq 0.15$ .<sup>c</sup> Only the IR part of the gene was analyzed.



**Fig. 3.** Relative rate comparison of  $K_s$  values for 27 SC genes (filled boxes) and 6 IR genes (open boxes; label) in *Glycine* (the outgroup) versus *Lotus* (with an IR) and versus *Medicago* (without an IR). The best-fit regression for the SC genes (thick line; constrained to pass through the origin) and the  $y = x$  line (thin line) are shown.

synonymous substitution rates regardless of whether an IR is present in the genome. However, the comparison of six (normally) IR-located genes revealed a 1.4-fold higher mean  $K_s$  for *Glycine/Medicago* ( $0.167 \pm 0.022$ ) than for *Glycine/Lotus* ( $0.119 \pm 0.018$ ). Of the six normally IR-located genes studied, five have accumulated more substitutions between *Glycine* and *Medicago* than between *Glycine* and *Lotus* (Fig. 3), suggesting that these genes accelerated in the lineage leading to *Medicago* once the IR was lost.

We conclude that the lower nucleotide substitution rate normally evident for genes located in the IR region is dependent on the presence of both copies of the repeat within the chloroplast genome. In legume species that lack the repeat, the former IR genes have a higher nucleotide substitution rate, equivalent to the rate for the SC regions, and hence the chloroplast genome has a uniform synonymous substitution rate. Our results show that the reduced substitution rate in genes such as *rpl2* is a direct consequence of their location in the IR and not an intrinsic property of the gene. This supports the hypothesis that the rate difference between IR and SC genes is the result of having a slightly biased gene conversion system operating on sequences with different copy numbers (Birky and Walsh 1992).

If we had detected a deceleration of SC genes instead of an acceleration of IR genes, we would have had either to reject the Birky/Walsh model in favor of a model such as error-prone recombination repair (Sharp et al. 1989; Birky and Walsh 1992) or to conclude that the number of chloroplast genomes per cell had been increased as a result of deleting the IR. Instead, our finding that, after one copy of the IR was deleted, the other copy accelerated in rate suggests that the number of chloroplast DNA molecules per cell remained unchanged. This in turn suggests that the copy-number control system in chloroplasts detects and "counts" a sequence located in the SC, even though the origins of replication of cpDNA are in the IR (Kunnimalaiyaan and Nielsen, 1997). We would predict that similarly homogeneous synonymous substitution rates will also be found in the cpDNAs of other plant lineages where the IR has been lost independently of legumes (for example, conifers), but there is not sufficient sequence data to test this at present.

**Acknowledgments.** We wish to thank Simon Wong for help with the phrap assembly and John Gray for the *Pisum sativum* chloroplast genome sequence. This study was supported by Science Foundation Ireland.

## References

Birky CW, Walsh JB (1992) Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. *Genetics* 130:677–683

- Curtis SE, Clegg MT (1984) Molecular evolution of chloroplast DNA sequences. *Mol Biol Evol* 1:291–301
- Doyle JJ (1995) DNA data and legume phylogeny: A progress report. In: Crisp M, Doyle J (eds) *Advances in legume systematics 7: Phylogeny*. Royal Botanic Gardens, Kew, pp 11–30
- Doyle JJ, Doyle JL, Palmer JD (1995) Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Syst Bot* 20:272–294
- Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T, Ohashi H (1997) A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: Taxonomic correlations and insights into the evolution of nodulation. *Am J Bot* 84:541–554
- Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht MK, et al. (eds) *Evolutionary biology*. Vol 30, Plenum Press, New York, pp 93–120
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet* 252:195–206
- Herdenberger F, Pillay D, Steinmetz A (1990) Sequence of the *trnH* gene and the inverted repeat structure deletion site of the broad bean chloroplast genome. *Nucleic Acids Res* 18:1297
- Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S (2000) Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Res* 7:323–330
- Kolodner R, Tewari KK (1979) Inverted repeats in chloroplast DNA from higher plants. *Proc Natl Acad Sci USA* 76:41–45
- Koller B, Delius H (1980) *Vicia faba* chloroplast DNA has only one set of ribosomal RNA genes as shown by partial denaturation mapping and R-Loop analysis. *Mol Gen Genet* 178:261–269
- Kunnimalaiyaan M, Nielsen BL (1997) Fine mapping of replication origins (*oriA* and *oriB*) in *Nicotiana tabacum* chloroplast DNA. *Nucleic Acids Res* 25:3681–3686
- Lavin M, Doyle JJ, Palmer JD (1990) Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* 44:390–402
- Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 36:96–99
- Li WH (1997) *Molecular evolution*. Sinauer Associates, Sunderland, MA
- Liston A (1995) Use of the polymerase chain reaction to survey for the loss of the inverted repeat in the legume chloroplast genome. In: Crisp M, Doyle J (eds) *Advances in legume systematics 7: Phylogeny*. Royal Botanic Gardens, Kew, pp 31–40
- Palmer JD (1991) Plastid chromosomes: Structure and evolution. In: Bogorad L, Vasil IK (eds) *Cell culture and somatic cell genetics of plants*. Vol 7A, Academic Press, New York, pp 5–53
- Palmer JD, Thompson WF (1981) Rearrangements in the chloroplast genomes of mung bean and pea. *Proc Natl Acad Sci USA* 78:5533–5537
- Palmer JD, Thompson WF (1982) Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 29:537–550
- Palmer JD, Osorio B, Thompson WF (1988) Evolutionary significance of inversions in legume chloroplast DNAs. *Curr Genet* 14:65–74
- Sharp PM, Shields DC, Wolfe KH, Li WH (1989) Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* 246:808–810
- Sugiura M (1992) The chloroplast genome. *Plant Mol Biol* 19:149–168
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: 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.
- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054–9058
- Wolfe KH (1988) The site of deletion of the inverted repeat in pea chloroplast DNA contains duplicated gene fragments. *Curr Genet* 13:97–99
- Wolfe KH (1991) Protein-coding genes in chloroplast DNA: compilation of nucleotide sequences, data base entries, and rates of molecular evolution. In: Bogorad L, Vasil IK (eds) *Cell culture and somatic cell genetics of plants*. Vol 7B, Academic Press, New York, pp 467–482
- Wu CI, Li WH (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci USA* 82:1741–1745