

The *rbcL* gene sequence from chestnut indicates a slow rate of evolution in the Fagaceae¹

NATHALIE FRASCARIA, LAURENT MAGGIA, MARTIN MICHAUD, AND JEAN BOUSQUET²

Centre de recherche en biologie forestière, Département des sciences forestières,
Faculté de foresterie et de géomatique, Université Laval, Ste-Foy, QC G1K 7P4, Canada

Corresponding Editor: G. Bellemare

Received November 10, 1992

Accepted April 1, 1993

FRASCARIA, N., MAGGIA, L., MICHAUD, M., and BOUSQUET, J. 1993. The *rbcL* gene sequence from chestnut indicates a slow rate of evolution in the Fagaceae. *Genome*, **36**: 668–671.

The nucleotide sequence was obtained for the chloroplast gene coding for the large subunit of the ribulose 1,5-bisphosphate carboxylase (*rbcL*) of chestnut (*Castanea sativa* Mill.), a member of the woody family Fagaceae. Amplification primers downstream and upstream the *rbcL* open reading frame are also described. By comparing with other angiosperm sequences, we show that the rate of evolution of *rbcL* in the family Fagaceae is much slower than that observed for the families of annuals analyzed.

Key words: angiosperms, *Castanea sativa*, Fagaceae, phylogeny, *rbcL*.

FRASCARIA, N., MAGGIA, L., MICHAUD, M., et BOUSQUET, J. 1993. The *rbcL* gene sequence from chestnut indicates a slow rate of evolution in the Fagaceae. *Genome*, **36** : 668–671.

La séquence nucléotidique du gène chloroplastique codant pour la grosse sous-unité de l'enzyme ribulose 1,5-diphosphate carboxylase (*rbcL*) a été déterminée pour le châtaignier (*Castanea sativa* Mill.), un membre de la famille ligneuse des Fagaceae. Des amorces d'amplification en amont et en aval du cadre de lecture de *rbcL* sont aussi décrites. La comparaison avec d'autres séquences d'angiospermes a permis de démontrer que le taux d'évolution du gène *rbcL* dans la famille des Fagaceae est beaucoup plus lent que celui observé chez les familles d'annuelles analysées.

Mots clés : angiospermes, *Castanea sativa*, Fagaceae, phylogénie, *rbcL*.

DNA sequence comparisons of chloroplast-encoded genes is now routinely used to investigate phylogenetic relationships among angiosperms and other plants (Wolfe et al. 1989). In recent years, there has been a growing interest in using sequences of the gene coding for the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (*rbcL*) to estimate plant phylogenies (Ritland and Clegg 1987). The substitution rate of *rbcL* appears appropriate for studies involving taxa that diverged from 10's of million to 100's of million years (Zurawski et al. 1984). In this paper, we report the complete nucleotide sequence of the *rbcL* open reading frame (ORF) for a woody angiosperm, the chestnut of the family Fagaceae, *Castanea sativa* Mill. We show that the rate of evolution of *rbcL* in this family is much slower than that observed for families of annual angiosperms. Implications for phylogenetic studies using *rbcL* are briefly discussed.

rbcL was amplified symmetrically following previously published procedures (Bousquet et al. 1992a). Seven primers, internal to the ORF, were used for amplification and sequencing and were designed from the alignment of previously published ORF sequences of dicots (Bousquet et al. 1992b): for sequencing the 5'–3' strand, 334⁺: 5'-TC TGT TAC TAA CAT GTT TAC TTC-3'; 691⁺: 5'-GA AAC AGG TGA AAT CAA AGG GCA TTA-3'; 1144⁺: 5'-GG TAT TCA CGT TTG GCA TAT GCC TGC-3'; and for sequencing the opposite strand, 216⁻: 5'-TC GGT CCA CAC AGT TGT CCA TGT-3'; 537⁻: 5'-CC CAA

TTT AGG TTT AAT AGT ACA TCC-3'; 979⁻: 5'-AAT ATG ATC TCC ACC AGA CAA ACG TAA-3'; 1303⁻: 5'-TCC CTC ATT ACG AGC TTG TAC ACA-3'. Primers upstream and downstream the coding region were designed from comparing Hamamelideae sequences (Bousquet et al. 1992a): upstream the primer -34⁺ was complementary to the 3'–5' strand, starting at the position -34: 5'-AAG TTC TTA ATT CAT GAG TCG TAG-3'; downstream the primer 1517⁻ was complementary to the 5'–3' strand, starting at the position 1517: 5'-TTA GTA AAA GAT TGG GCC GAG TTT-3'. Direct sequencing of the two DNA strands was performed with the dideoxynucleotide chain-termination procedure using the Sequenase version 2.0 kit with dITP nucleotide mixes (USB) as recommended by the manufacturer, except that times and temperatures were modified (Zimmerman and Fuscoe 1991).

The coding region of the chestnut *rbcL* was 1428 bp long. The chestnut sequence was very similar to red oak (*Quercus rubra*) (Bousquet et al. 1992a) (Fig. 1), another member of the family Fagaceae, with only 12 nucleotide differences for a DNA homology of 99.1%. Otherwise, the chestnut sequence was 90.1% identical to tobacco (*Nicotiana tabacum*) (Shinozaki and Sugiura 1982), 86.5% to maize (*Zea mays*) (Gaut et al. 1992), and 83.5% to Douglas-fir (*Pseudotsuga menziesii*) (Hipkins et al. 1990).

Synonymous and nonsynonymous rates of evolution and their standard errors in different annual and woody families of dicots and monocots were calculated (Li et al. 1985). Overall rates of substitution were estimated using a weighted average of K_s and K_a , the two-parameter method of Kimura (1980), and the one-parameter method

¹The nucleotide sequence reported appears in the GenBank and EMBL under the accession number M94936.

²Author to whom correspondence should be addressed.

Cs 1	ATG	TCA	CCA	CAA	ACA	GAG	ACT	AAA	GCA	AGT	GTT	GGA	TTC	AAA	GCT	GGT	GTT	51
Qr 1	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	51
52	AAA	GAT	TAT	AAA	TTG	ACT	TAT	tAT	ACT	CCT	GAC	TAT	CAA	ACC	AAA	GAT	ACT	102
52	---	---	---	---	---	---	---	c-	---	---	---	---	---	---	---	---	---	102
103	GAT	ATC	TTG	GCA	GCC	TTC	CGA	GTA	ACT	CCT	CAA	CCT	GGA	GTT	CCG	CCg	GAG	153
103	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--c	---	153
154	GAA	GCA	GGG	GCC	GCG	GTA	GCT	GCT	GAA	TCT	TCC	ACT	GGG	ACA	TGG	AgA	ACT	204
154	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	-c-	---	204
205	GTG	TGG	ACT	GAC	GGG	CTT	ACC	AGT	CTT	GAT	CGT	TAC	AAA	GGA	CGA	TGC	TAC	255
205	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	255
256	CAC	ATC	GAG	CCg	GTT	GCT	GGA	GAA	GAA	AAT	CAA	TTT	ATT	GCT	TAT	GTA	GCT	306
256	---	---	---	-a	---	---	---	---	---	---	---	---	---	---	---	---	---	306
307	TAC	CCC	TTA	GAC	CTC	TTT	GAA	GAA	GGT	TCT	GTT	ACT	AAC	ATG	TTT	ACT	TCC	357
307	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	357
358	ATT	GTG	GGT	AAT	GTA	TTT	GGA	TTC	AAG	GCC	CTG	CGC	GCT	CTA	CGT	CTG	GAG	408
358	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	408
409	GAT	TTG	CGA	ATC	CCT	ACT	TCT	TAT	TCT	AAA	ACT	TTC	CAA	GGT	CCG	CCT	CAT	459
409	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	459
460	GGC	ATC	CAA	GTT	GAG	AGG	GAT	AAA	TTA	AAC	AAG	TAT	GGC	CGC	CCC	CTA	TTA	510
460	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	510
511	GGA	TGT	ACT	ATT	AAA	CCT	AAA	TTG	GGA	TTA	TCC	GCT	AAG	AAT	TAC	GGT	AGA	561
511	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	561
562	GCA	GTT	TAT	GAA	TGT	CTC	CGC	GGT	GGG	CTT	GAT	TTT	ACC	AAA	GAT	GAT	GAG	612
562	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	612
613	AAC	GTT	AAT	TCC	CAA	CCA	TTT	ATG	CGT	TGG	AGA	GAC	CGT	TTC	cTA	TTT	TGT	663
613	---	---	---	---	---	---	---	---	---	---	---	---	---	---	g--	---	---	663
664	GCC	GAA	GCA	ATT	TAT	AAA	GCG	CAG	GCT	GAA	ACA	GGT	GAA	ATC	AAg	GGG	CAT	714
664	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--a	---	---	714
715	TAC	TTG	AAT	GCT	ACT	GCA	GGT	ACA	TGC	GAA	GAA	ATG	ATC	AAA	AGG	GCT	GTA	765
715	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	765
766	TTT	GCC	AGA	GAA	CTG	GGA	GTT	CCT	ATC	GTA	ATG	CAT	GAC	TAC	TTA	ACC	GGG	816
766	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	816
817	GGG	TTC	ACT	GCA	AAT	ACT	AGC	TTG	GCT	CAT	TAT	TGC	CGA	GAT	AAT	GGT	CTA	867
817	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	867
868	CTT	CTT	CAC	ATC	CAT	CGT	GCa	ATG	CAT	GCA	GTT	ATT	GAT	CGA	CAG	AAG	AAT	918
868	---	---	---	---	---	---	-c	---	---	---	---	---	---	---	---	---	---	918
919	CAT	GGT	ATA	CAC	TTT	CGT	GTA	CTA	GCT	AAA	GCA	TTA	CGT	ATG	TCT	GGT	GGA	969
919	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	969
970	GAT	CAT	ATT	CAT	GCC	GGT	ACC	GTA	GTA	GGT	AAA	CTT	GAA	GGG	GAA	AGA	GAA	1020
970	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1020
1021	ATC	ACT	TTA	GGC	TTT	GTT	GAT	TTA	CTA	CGT	GAT	GAT	TAT	ATT	GAA	AAA	GAT	1071
1021	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1071
1072	CGA	AGC	CGC	GGT	ATT	TAT	TTC	ACT	CAA	GAT	TGG	GTC	TCT	TTA	CCA	GGT	GTT	1122
1072	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1122
1123	CTG	CCC	GTG	GCT	TCA	GGG	GGT	ATT	CAC	GTT	TGG	CAT	ATG	CCT	GCT	TTG	ACC	1173
1123	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1173
1174	GAA	ATC	TTT	GGA	GAT	GAT	TCC	GTA	CTA	CAA	TTt	GGC	GGA	GGA	ACT	TTg	GGG	1224
1174	---	---	---	---	---	---	---	---	---	---	--c	---	---	---	---	--a	---	1224
1225	CAC	CCT	TGG	GGA	AAT	GCA	Cct	GGT	GCc	GTA	GCT	AAT	CGA	GTA	GCT	CTA	GAA	1275
1225	---	---	---	---	---	---	-c	---	-t	---	---	---	---	---	---	---	---	1275
1276	GCA	TGT	GTA	CAA	GCT	CGT	AAT	GAG	GGA	CGT	GAT	CTT	GCT	CGT	GAG	GGT	AAT	1326
1276	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1326
1327	GAA	ATT	ATt	CGT	GAG	GCT	GCT	AAA	TGG	AGT	CCT	GAG	CTA	GCT	GCT	GCT	TGT	1377
1327	---	---	-c	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1377
1378	GAA	GTA	TGG	AAA	GAG	ATC	AAA	TTT	GAA	TTC	CCA	GCA	ATG	GAT	ACT	TTG	TAA	1428
1378	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1428

FIG. 1. Open reading frame sequence (1428 bp) of the *rbcl* gene from *Castanea sativa* Mill. Differences are indicated by lowercase letters between *C. sativa* (Cs) and *Quercus rubra* (Qr) (Bousquet et al. 1992a), another member of the family Fagaceae.

of Jukes and Cantor (1969). For each family analyzed, a rate per year was obtained by dividing the pairwise rate between two taxa by $2T$, where T was the divergence time derived from the fossil record.

Substitution rates per year varied in a range of near one order of magnitude between the two woody and the

two annual angiosperm families compared (Table 1). This difference was apparent at both synonymous and non-synonymous sites (Table 1). The slower rate of evolution observed in the two woody families analyzed confirms a rate heterogeneity trend recently reported among seed plants for *rbcl*, although no rates per year were esti-

TABLE 1. Substitution rates per site per year (and substitution rates per site in parentheses) of *rbcL* within four angiosperm families

Families	Taxa compared	Divergence time (million years)*	Number of sites compared	Substitution rate \pm SE [†]			
				K_s	K_a	K_o	Kimura J.C.
Fagaceae	<i>Castanea-Quercus</i>	60	1428	$2.36 \pm 0.79 \times 10^{-10}$ (0.0283 \pm 0.0095)	$0.23 \pm 0.13 \times 10^{-10}$ (0.0027 \pm 0.0016)	$0.71 \pm 0.21 \times 10^{-10}$ (0.0085 \pm 0.0025)	$0.71 \pm 0.21 \times 10^{-10}$ (0.0085 \pm 0.0025)
Solanaceae	<i>Nicotiana-Petunia</i>	50	1434	$7.99 \pm 1.64 \times 10^{-10}$ (0.0799 \pm 0.0164)	$1.21 \pm 0.33 \times 10^{-10}$ (0.0121 \pm 0.0033)	$2.75 \pm 0.45 \times 10^{-10}$ (0.0275 \pm 0.0045)	$2.77 \pm 0.45 \times 10^{-10}$ (0.0277 \pm 0.0045)
Poaceae	<i>Zea-Triticum</i>	60	1428	$1.34 \pm 2.06 \times 10^{-10}$ (0.1606 \pm 0.0247)	$1.90 \pm 0.38 \times 10^{-10}$ (0.0228 \pm 0.0046)	$4.45 \pm 0.54 \times 10^{-10}$ (0.0534 \pm 0.0065)	$4.37 \pm 0.53 \times 10^{-10}$ (0.0524 \pm 0.0063)
Arecaceae	<i>Serenoa-Calamus</i>	60	985	$1.98 \pm 0.89 \times 10^{-10}$ (0.0237 \pm 0.0107)	$0.54 \pm 0.24 \times 10^{-10}$ (0.0065 \pm 0.0029)	$0.86 \pm 0.28 \times 10^{-10}$ (0.0103 \pm 0.0033)	$0.85 \pm 0.27 \times 10^{-10}$ (0.0102 \pm 0.0032)

*References for divergence times: Fagaceae, Crepet (1989); Solanaceae, Albert et al. (1992); Poaceae, Wolfe et al. (1989); Arecaceae, Wilson et al. (1990).

[†]Abbreviations used: K_s , K_a , and K_o , synonymous, nonsynonymous, and overall numbers of substitutions, respectively (Li et al. 1985); Kimura, Kimura's (1980) two-parameter numbers of substitutions; J.C., Jukes and Cantor's (1969) one-parameter numbers of substitutions.

mated (Bousquet et al. 1992b). This diversity of rates implies that careful analysis of rate heterogeneity among the sequences analyzed must be conducted before using this gene as a molecular clock, which assumes constancy of evolutionary rate over time (Zuckerlandl and Pauling 1965). Procedures for this purpose have recently become available (Wu and Li 1985; Li and Bousquet 1992). Methods of phylogenetic tree construction, which are sensitive to rate heterogeneity (see Sourdis and Nei 1988), must also be used cautiously with this gene.

Acknowledgements

We thank M. Lefranc (ESV, Orsay, France) for providing plant materials. This work was supported by AUPELF-UREF association fellowships to N.F. and L.M., by the Natural Sciences Engineering Research Council of Canada (NSERC) fellowship to L.M. and NSERC operating grant (OGP0046273) to J.B., and by Québec FCAR grants (ER-0693 and NC-0642) to J.B.

- Albert, V.A., Mishler, B.D., and Chase, M.W. 1992. Character-state weighting for restriction site data in phylogenetic reconstruction with an example from chloroplast DNA. *In* Molecular systematics of plants. *Edited by* P.S. Soltis, D.E. Soltis, and J.J. Doyle. Chapman and Hall, New York. pp. 369-401.
- Bousquet, J., Strauss, S.H., and Li, P. 1992a. Complete congruence between morphological and *rbcL*-based molecular phylogenies in birches and related species (Betulaceae). *Mol. Biol. Evol.* **9**: 1076-1088.
- Bousquet, J., Strauss, S.H., Doerksen, A.H., and Price, R.A. 1992b. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 7844-7848.
- Crepet, W.L. 1989. History and implications of the early North American fossil record of Fagaceae. *In* Evolution, systematics, and fossil history of the Hamamelidae. *Edited by* P.R. Crane and S. Blackmore. Clarendon Press, Oxford. pp. 45-66.
- Gaut, B.S., Muse, S.V., Clark, W.D., and Clegg, M.T. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. *J. Mol. Evol.* **35**: 292-303.
- Hipkins, V.D., Tsai, C.H., and Strauss, S.H. 1990. Sequence of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from a gymnosperm, Douglas fir. *Plant Mol. Biol.* **15**: 505-507.
- Jukes, T.H., and Cantor, C.R. 1969. Evolution of protein molecules. *In* Mammalian protein metabolism. *Edited by* H.N. Munro. Academic Press, New York. pp. 21-132.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111-120.
- Li, P., and Bousquet, J. 1992. Relative rate test for nucleotide substitutions between two lineages. *Mol. Biol. Evol.* **9**: 1185-1189.
- Li, W.-H., Wu, C.-I., and Kuo, C.-C. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**: 150-174.
- Ritland, K., and Clegg, M.T. 1987. Evolutionary analysis of plant DNA sequences. *Am. Nat.* **130**: 74-100.
- Shinozaki, K., and Sugiura, M. 1982. The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose 1,5-bisphosphate carboxylase-oxygenase. *Gene*, **20**: 91-102.
- Sourdis, J., and Nei, M. 1988. Relative efficiencies of the maximum parsimony and distance-matrix methods in

- obtaining the correct phylogenetic tree. *Mol. Biol. Evol.* **5**: 298–311.
- Wilson, M.A., Gaut, B., and Clegg, M.T. 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Mol. Biol. Evol.* **7**: 303–314.
- Wolfe, K.H., Gouy, M., Yang, Y.-W., Sharp, P.M., Li, W.-H. 1989. Date of the monocot–dicot divergence from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 6201–6205.
- Wu, C.-I., and Li, W.-H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 1741–1745.
- Zimmerman, L.J., and Fuscoe, J.C. 1991. Direct DNA sequencing of PCR products. *Environ. Mol. Mutagen.* **18**: 274–276.
- Zuckerandl, E., and Pauling, L. 1965. Evolutionary divergence and convergence in proteins. *In* *Evolving genes and proteins*. Edited by V. Bryson and H.G. Vogel. Academic Press, New York. pp. 97–166.
- Zurawski, G., Clegg, M.T., and Brown, A.H.D. 1984. The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. *Genetics*, **106**: 735–749.