Copyright © 1989 by the Genetics Society of America

Molecular Evolution of the Small Subunit of Ribulose Bisphosphate Carboxylase: Nucleotide Substitution and Gene Conversion

R. B. Meagher,* S. Berry-Lowe† and K. Rice‡

**Department of Genetics and †Department of Botany, University of Georgia, Athens, Georgia 30602, and ‡Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138*

Manuscript received February 6, 1989
Accepted for publication September 1, 1989

ABSTRACT

The nucleotide sequences encoding the mature portion of 31 ribulose 1,5-bisphosphate carboxylase small subunit (SSU) genes from 17 genera of plants, green algae and cyanobacteria were examined. Among the 465 pairwise sequence comparisons, SSU multigene family members within the same species were more similar to each other in nonsynonymous or replacement nucleotide substitutions (RNS) than they were to SSU sequences in any other organism. The concerted evolution of independent SSU gene lineages within closely related plant species suggests that homogenization of RNS positions has occurred at least once in the life of each genus. The rate of expected RNS among mature SSU sequences was calculated to be 1.25×10^{-9}/site/yr for the first 70 million years (MY) of divergence with a significant slowing to 0.13×10^{-9}/site/yr for the next 1,400 MY. The data suggest that mature SSU sequences do not accumulate more than 20% differences in the RNS positions without compensatory changes in other components of this enzyme system. During the first 70 MY of divergence between species, the rate of expected synonymous or silent nucleotide substitutions (SNS) is $\sim 6.6 \times 10^{-9}$/site/yr. This is five times the RNS rate and is similar to the silent rate observed in animals. In striking contrast, SNS and RNS do not show this correlation among SSU gene family members within a species. A mechanism involving gene conversion within the exons followed by selection for biased gene conversion products with conservation of RNS positions and divergence of SNS positions is discussed. A SSU gene tree based on corrected RNS for 31 SSU sequences is presented and agrees well with a species tree based on morphological and cytogenetic traits for the 17 genera examined. SSU gene comparisons may be useful in predicting phylogenetic relationships and in some cases divergence times of various plant, algal and cyanobacterial species.

RIBULOSE 1,5-bisphosphate carboxylase is the most abundant protein in green plants and is composed of eight identical large subunits (LSU) and eight identical small subunits (SSU). Investigation of genetic segregation of isoelectric variants of the large and small subunit polypeptides revealed that the large subunit is maternally inherited and encoded by the chloroplast genome, whereas the SSU is encoded by the nuclear genome (MIZIORKO and LORMER 1983). Genetic analysis of isoelectric focusing of SSU isoforms suggests that between one and five genes encode the small subunit (KAWASHIMIA and WILDMAN 1972; CHEN *et al.* 1975; GATENBY and COCKING 1978; CHEN and WILDMAN 1980). A number of laboratories confirmed that the small subunit is encoded by a small multigene family in higher plants and green algae. Subfamilies of SSU sequence have been identified within several SSU gene families and in some cases there are tandem repeats of one or more of the subfamily representatives (DUNSMUIR, SMITH and BEDBROOK 1983; PICHERSKY *et al.* 1986; DEAN *et al.* 1987). There is at least 70% amino acid similarity among the mature peptide regions of all higher plant SSU sequences, making this a reasonable gene for sequence comparisons between distant plant species. Initial studies on the divergence of SSU genes suggested that a mechanism of concerted evolution has homogenized SSU sequences within some species (PICHERSKY *et al.* 1986; DEAN *et al.* 1987; O'NEAL *et al.* 1987).

The small subunit genes in eukaryotes encode a precursor composed of a transit peptide of approximately 50 amino acids and the mature peptide of approximately 120 amino acids. The transit peptide is removed as the mature SSU peptide is transported across the organellar membrane into the chloroplast (CHUA and SCHMIDT 1978; SMITH and ELLIS 1979). The transit sequences of the small subunit genes are quite divergent, varying greatly in number of amino acids and in amino acid sequence (BERRY-LOWE *et al.* 1982; MISHKIND, WESSLER and SCHMIDT 1985). Thus quantitative analysis of divergence is difficult for this portion of the SSU sequence.

Nucleotide sequences of structural genes have been

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

used to quantify the divergence between genes and proteins (WILSON, CARLSON and WHITE 1977; PERLER *et al.* 1980; MIYATA and HAYASHIDA 1982). Nucleotide changes within a coding region can be separated into silent (synonymous) nucleotide substitutions (SNS) and replacement (nonsynonymous) nucleotide substitutions (RNS). SNS are nucleotide changes which do not change the encoded amino acid, whereas RNS change the encoded amino acid. The SNS rate is fairly constant for different genes in the animal kingdom and SNS increase linearly with time. The rate of SNS has been estimated to be 5.37×10^{-9}/site/yr or 1.0% SNS/1.9 million years (MY) (MIYATA and HAYASHIDA 1982; MIYATA *et al.* 1982). For conserved coding regions the RNS rate is much slower and RNS accumulation is linear for long periods of time for some genes. However, rates of accumulation of RNS vary according to the degree of conservation of the protein sequence. Initial studies examining the divergence of SSU amino acid sequences suggested that the RNS clock may be linear within plant subclasses (MARTIN and DOWD 1988). However, no attempt has been made to account for the problems which arise in comparing the products of this multigene family.

This manuscript analyzes the divergence of the SSU DNA sequences encoding the mature peptide from a variety of plant, algal and cyanobacterial species. The fractions of SNS and RNS are analyzed among 31 small subunit genes isolated from 17 different photosynthetic organisms, utilizing an extension of the method of PERLER *et al.* (1980). Phenograms have been constructed using the method of SAITOU and NEI (1987). Rates of RNS and SNS accumulation for the SSU mature sequence have been examined. The evolutionary relationships among multigene family members are discussed.

MATERIALS AND METHODS

Computational analysis of sequence: The divergence analysis program DIV (HIGHTOWER and MEAGHER 1985; RICE 1987) was written to operate under UNIX, based on the methods of PERLER *et al.* (1980) and SHAH, HIGHTOWER and MEAGHER (1983). Pathways of nucleotide substitution between codons are considered equiprobable by this method. The substitutions are categorized 1, 2 or 3, depending on the number of possible substitutions that will create no change in amino acid (SNS) or will change the amino acid (RNS) and are statistically weighted as such. For example, the change of a proline codon from CCA to CCG is a category 3 SNS since a change from A to three other bases results in a silent change. The change of a lysine codon from AAA to an asparginine codon AAC is a category 2 RNS change because changing the third position to C or T results in an amino acid replacement. The assumption is made in this method that substitution of any nucleotide occurs at any nucleotide site with equal probability. The fraction of SNS or RNS are calculated by comparing the number of changes observed with the potential number of SNS or RNS sites over the two sequences analyzed. Multiple hit kinetics are also taken into consideration for each category of change and result in corrected SNS and corrected RNS values. There are approximately five category 2 SNS sites and five category 1 RNS sites generated by isoleucine codons between any two SSU sequences. In the event that a large fraction of changes occur at these sites, which represent such a small portion of the total sequence, they can drastically bias the multiple hit correction. As recommended by PERLER *et al.* (1980) these two categories have been omitted when calculating the corrected values (HIGHTOWER and MEAGHER 1986).

In calculating both the rates of nucleotide substitution and the ratios of SNS to RNS, a statement of potential error is often presented along with the calculated value in the text. The calculated error represents the standard deviation of all the data considered (MCCLAVE and DIETRICH 1982).

Constructing gene trees: Phenograms showing the relationships of SSU sequences were constructed based on RNS and SNS values using the neighbor joining (NJ) method (SAITOU and NEI 1987) and the unweighted pair group mean averaging method (UPGMA) (SOKAL and SNEATH 1963). The NJ method begins with a star-like topology, and it iteratively constructs a binary tree by connecting sequences or groups of sequences that successively minimize sums of branch lengths of intermediately produced trees. The NJ method is less sensitive than UPGMA to variations in substitution rates across lineages, and it produces an unrooted tree. The UPGMA method was used to allow comparison of NJ trees with those produced by a well studied method that produce a rooted tree.

Small subunit genes examined: The 42 different SSU sequences analyzed are presented in the third column of Table 1 and include 32 genomic sequences and 10 cDNA sequences. These sequences were retrieved from the NIH and EMBL data bases or from published manuscripts and theses and are referenced in column 5 of Table 1. An initial screening of the 861 pairwise comparisons revealed several gene sequences which were either identical in nucleotide sequence, differed by fewer than 2 silent nucleotide changes, or differed by the deletion of a single codon. The similar or duplicate gene/sequences (column 3, Table I) were grouped together and given a single SSU sequence abbreviation (column 1, Table 1), reducing the number of sequences analyzed to 31, or to 465 pairwise comparisons. The genetic mapping (VALLEJOS, TANKSLEY and BERNATZKEY 1986) and recently published physical mapping (SUGITA *et al.* 1987) of the tomato SSU sequences, combined with the comparative sequence analysis presented in this manuscript suggest that the cDNA sequences, Tm4 and Tm5, represent the RNA product of Tm2 and need not have been included in the final set of sequences analyzed. The nucleotide differences between these three sequences probably represent discrepancies in sequencing or recording into the data base and have little impact on the text of the manuscript. Therefore, Tm4 and Tm5 sequences have been omitted from most discussions in the text.

Alignment and some editing of the mature portion of the SSU sequences is required because the PERLER *et al.* (1980) analysis proceeds codon by codon and therefore can only be carried out on sequences of the same length. Each sequence was aligned for maximum homology (SMITH and WATERMAN 1981) with the mature soybean sequence, Soy1 which contains 369 nucleotides (BERRY-LOWE *et al.* 1982). Only alignments which kept codons intact were accepted. The alignment of four example sequences are shown in Figure 1. Based on the numbering of the 123 codons in the mature soybean sequence, Soy1 (BERRY-LOWE *et al.* 1982), the

TABLE 1

Ribulose-1,5-bisphosphate carboxylase small subunit sequences examined

Abbrev.	Genus, species	Gene/Clone	IVS No.	Reference
Anb1	*Anabaena* 7120	An606	0	NIERZWICKI-BAUER, CURTIS and HASELKORN (1984)
Anc1	*Anacystis nidulans* 6301	pANP1155	0	SHINOZAKI and SUGIURA (1983)
Cmy1	*Chlamydomonas reinhardtii*	rbcS1	3	GOLDSCHMIDT-CLERMONT and RAHIRE (1986)
Cmy2		rbcS2	3	GOLDSCHMIDT-CLERMONT and RAHIRE (1986)
Cuc1	*Cucumis sativus*	pCu5/1539	cDNA	GREENLAND, THOMAS and WALDEN (1987)
Cyn1	*Cyanophora paradoxa*	rbcS	0	STARNES *et al.* (1985)
Fla1	*Flaveria trinervia*	None	cDNA	ADAMS *et al.* (1987)
Ice1	*Mesembryanthemum cyrstallinum*	rbcS	cDNA	DEROCHER *et al.* (1987)
Lem1	*Lemna gibba*	5A ~ 5B	1	STIEKEMA, WIMPEE and TOBIN (1983)
Lem2		40A ~ 40 B	1	WIMPEE (1984)
Lem3		pLgSSU	cDNA	STIEKEMA, WIMPEE and TOBIN (1983)
Pea1	*Pisum sativum*	rbcS-3A ~ 3C	2	FLUHR *et al.* (1986)
Pea2		pPS-2.4	2	CORUZZI *et al.* (1984)
Pea3		ss3.6	2	CASHMORE (1983)
Pet1	*Petunia* "Mitchell"	611	2	DEAN *et al.* (1987)
Pet2		ssu11A, 511	2	TUMER *et al.* (1986) and DEAN *et al.* (1987)
Pet3*		301, ssu8	3	DEAN *et al.* (1987) and TUMER *et al.* (1986)
Pot1	*Solanum tuberosum*	rbcS-c	cDNA	WOLTER *et al.* (1988)
Soy1	*Glycine max*	SRS1	2	BERRY-LOWE *et al.* (1982)
Soy2		SRS4	2	GRANDBASTIEN *et al.* (1986)
Sun1	*Helianthus annuus*	rbcS	2	WAKSMAN, LEBRUN and FREYSSINET (1987)
Tm1	*Lycopersicon esculentum*	RbcS-1	2	SUGITA *et al.* (1987) and PICHERSKY *et al.* (1986)
		LESS17	cDNA	MCKNIGHT, ALEXANDER and BABCOCK (1986)
Tm2*		Rbcs-2	3	SUGITA *et al.* (1987)
Tm3		Rbcs-3A ~3B, 3C	2	SUGITA *et al.* (1987) and PICHERSKY *et al.* (1986)
Tm4		Rbcs-2A	cDNA	PICHERSKY *et al.* (1986)
Tm5		LESS5	cDNA	MCKNIGHT, ALEXANDER and BABCOCK (1986)
Tob1	*Nicotiana sylvestris*	pSTV34, pSEM1, 2	cDNA 2	PINCK *et al.* (1984)
Tob2*	*Nicotiana tabacum*	NtSS23	3	MAZUR and CHUI (1985)
Wh1	*Triticum aestivum* var. ERA	pWS4.3	1	BROGLIE *et al.* (1983)
Wh2		pW9	cDNA	BROGLIE *et al.* (1983)
Zea1	*Zea mays*	rbcs ~ ssu	1	LEBRUN, WAKSMAN and FREYSSINET (1987) and MATSUOKA *et al.* (1987)

Abbrev. is the abbreviation used in this manuscript to describe a particular SSU sequence. cDNA sequences are underlined. The genes within *Solanaceae* with three introns which are discussed in detail in the text are marked with an asterisk. Gene/clone is the designation given to this sequence in the original published manuscript. When two genes/clones differ by only one or two nucleotide substitutions they are given the same Abbrev. and shown to be approximately (~) the same sequence. Genes/clones which are identical in sequence are separated by a comma. IVS No. refers to the number of intervening sequences which interrupt the transcript of this SSU gene, unless a cDNA clone was analyzed.

following editorial changes were made in the remaining sequences (for examples see Figure 1). Place marker codons (XXX) were inserted to make up for deletion of codons relative to the soybean sequence. After codon position two, Amb1 and Cyn1 had 2 and 3 marker codons inserted, respectively; after position 33 Zea1 received one marker codon; after position 45 Anb1 and Anc1 received one codon while Cyn1 received 4 codons; after position 47 Lem1, 2, 3, and Zea1 received one codon and Anb1 received 8 codons; after position 48 Anc1 received 8 codons; after position 55 Cyn1 received 7 codons; after position 62 Anc1 and Anb1 each received 3 codons; after position 65 Cyn1 received 1 codon; after position 121 Cyn1 received 2 marker codons and the three Lem sequences received 2 codons. Comparisons between codons where one or both codons were marker codons were not scored at all and had little impact on the calculated nucleotide substitutions of two mature SSU sequences, since the values were based on a fraction of all possible diverging sites (see below) and not the absolute numbers of changes. An extra 1 and 5 codons were deleted from the Cmy1 and Cmy2 sequences after positions 44 and 60, respectively. After codon 47 one codon was deleted from Wh1 and Wh2 sequences. All extra codons after codon 123 were deleted from the Cuc1, Ice1, Cmy1-2 and Wh1-2 sequences. These editorial changes resulted in all 123 codons being compared among dicots sequences and approximately 106 codons being compared between angiosperms and the most distantly related sequence from cyanobacterial species, Cyn1 (Figure 1). In this case, 47–49 of the 106 codons (318 nt) compared encode identical amino acids. For comparison, when the nuclear encoded 18S rRNA sequence from maize is compared to the 16S rRNA sequence from *Anacystis* there are 1090 alignable positions and 529 of these are nucleotide differences (LAKE 1987). Even in these distant comparisons many fewer ambiguities arise in the alignment of SSU sequences than for rRNA small subunit sequences (LAKE 1987). Furthermore, differences in the total numbers of SSU codons compared should

```
                                                  27                                                  54
Soy1  ATG CAG GTG TGG CCA CCA ATT GGC AAG AAG AAG TTC GAG ACT CTT TCC TAC TTG
      MET Gln Val Trp Pro Pro Ile Gly Lys Lys Lys Phe Glu Thr Leu Ser Tyr Leu
Pet1  ATG CAG GTA TGG CCA CCA TAT GGA AAG AAG AAG TAC GAG ACA CTC TCA TAC CTT
       -   -   -   -   -   - Tyr   -   -   -   - Tyr   -   -   -   -   -   -
Zea1  ATG CAG GTG TGG CCG GCC TAC GGC AAC AAG AAG TTC GAG ACG CTG TCG TAC CTG
       -   -   -   -   - Ala Tyr   - Asn   -   -   -   -   -   -   -   -   -
Cyn1  ATG CAA XXX XXX XXX CTT AGA GTA GAA CGT AAG TTC GAA ACT TTT TCT TAT TTA
       -   -               Leu Arg Val Glu Arg   -   -   -   - Phe   -   -   -
                                                   9                                                  18
                                                  81                                                 108
Soy1  CCA GAC CTC GAT GAT GCC CAA TTG GCC AAG GAA GTC GAA TAC CTT CTA AGG AAA
      Pro Asp Leu Asp Asp Ala Gln Leu Ala Lys Glu Val Glu Tyr Leu Leu Arg Lys
Pet1  CCC GAT TTG ACC GAC GAA CAA TTG CTC AAA GAA ATT GAG TAC CTT TTG AAC AAG
       -   -   - Thr   - Glu   -   - Leu   -   - Ile   -   -   -   - Asn   -
Zea1  CCG CCG CTG TCG ACG GAC GAC CTG CTG AAG CAG GTG GAC TAC CTG CTG CGC AAC
       - Pro   - Ser Thr Asp Asp   - Leu   - Gln   - Asp   -   -   -   - Asn
Cyn1  CCA CCA TTA AAC GAC CAA CAG ATT GCG CGT CAA TTA CAA TAC GCA CTT TCC AAT
       - Pro   - Asn   - Gln   - Ile   - Arg Gln Leu Gln   - Ala   - Ser Asn
                                                  27                                                  36
                                                 135                                                 162
Soy1  GGA TGG ATT CCT TGC TTG GAA TTC GAG TTG GAG CAC GGT TTC GTG TAC CGT GAG
      Gly Trp Ile Pro Cys Leu Glu Phe Glu Leu Glu His Gly Phe Val Tyr Arg Glu
Pet1  GGA TGG GTT CCT TGC TTG GAA TTC GAG ACT GAG CAC GGA TTT GTC TAC CGT GAA
       -   - Val   -   -   -   -   -   - Thr   -   -   -   -   -   -   -   -
Zea1  GGC TGG ATA CCC TGC CTC GAG TTC AGC AAG GTC XXX GGC TTC GTG TAC CGC GAG
       -   -   -   -   -   -   -   - Ser Lys Val       -   -   -   -   -   -
Cyn1  GGT TAT AGC CCA GCA ATC GAA TTC AGT XXX XXX XXX XXX TTT ACA GGT AAA GCT
       - Tyr Ser   - Ala Ile   -   - Ser                   -   - Thr Gly Lys Ala
                                                  45                                                  54
                                                 189                                                 216
Soy1  CAC AAC AAG TCA CCT CGA TAC TAT GAT GGA CGC TAC TGG ACC ATG TGG AAG CTG
      His Asn Lys Ser Pro Arg Tyr Tyr Asp Gly Arg Tyr Trp Thr MET Trp Lys Leu
Pet1  TAT CAC GCA TCA CCA GGA TAC TAT GAT GGA AGG TAC TGG ACA ATG TGG AAG TTG
      Tyr His Ala   -   - Gly   -   -   -   -   -   -   -   -   -   -   -   -
Zea1  AAC TCC ACC TCC CCG TGC TAC TAC GAC GGC CGC TAC TGG ACC ATG TGG AAG CTG
      Asn Ser Thr   -   - Cys   -   -   -   -   -   -   -   -   -   -   -   -
Cyn1  GAA XXX XXX XXX XXX XXX XXX XXX GAC TTA GTA XXX TGG ACT TTA TGG AAA TTA
      Glu                               -   Leu Val       -   -   Leu   -   -   -
                                                  63                                                  72
                                                 243                                                 270
Soy1  CCT ATG TTT GGC TGC ACT GAT GCT TCT CAG GTG TTG AAG GAG CTT CAA GAG GCT
      Pro MET Phe Gly Cys Thr Asp Ala Ser Gln Val Leu Lys Glu Leu Gln Glu Ala
Pet1  CCC ATG TTT GGT TGC ACT GAT GCT ACC CAA GTC TTG GGT GAG CTC CAA GAG GCC
       -   -   -   -   -   -   -   - Thr   -   -   - Gly   -   -   -   -   -
Zea1  CCC ATG TTC GGC TGC AAC GAC GCC ACC CAG GTG TAC AAG GAG CTG CAG GAG GCC
       -   -   -   -   - Asn   -   - Thr   -   - Tyr   -   -   -   -   -   -
Cyn1  CCT TTA TTT GGT GCA CAA TCT CCT GAA GAA GTA CTT AGC GAA ATT CAA GCT TGT
       - Leu   -   - Ala Gln Ser Pro Glu Glu   -   - Ser   - Ile   - Ala Cys
                                                  81                                                  90
                                                 297                                                 324
Soy1  AAG ACT GCA TAC CCC AAT GGC TTC ATC CGT ATC ATT GGA TTT GAC AAC GTT CGC
      Lys Thr Ala Tyr Pro Asn Gly Phe Ile Arg Ile Ile Gly Phe Asp Asn Val Arg
Pet1  AAG AAG GCT TAC CCA AAT GCA TGG ATC AGA ATC ATC GGA TTC GAC AAC GTT CGT
       - Lys   -   -   -   - Ala Trp   -   -   -   -   -   -   -   -   -   -
Zea1  ATC AAA TCC TAC CCG GAC GCC TTC CAC CGC GTC ATC GGC TTC GAC AAC ATC AAG
      Ile Lys Ser   -   - Asp Ala   - His   - Val   -   -   -   -   - Ile Lys
Cyn1  AAA CAA CAG TTC CCT AAT GCT TAC ATT CGT GTT GTA GCA TTT GAC TCT ATC AGA
       - Gln Gln Phe   -   - Ala Tyr   -   - Val Val Ala   -   - Ser Ile   -
                                                  99                                                 108
                                                 351
Soy1  CAA GTG CAG TGC ATC AGC TTC ATC GCC TAC AAG CCC CCA GGC TTC
      Gln Val Gln Cys Ile Ser Phe Ile Ala Tyr Lys Pro Pro Gly Phe
Pet1  CAA GTG CAG TGC ATC AGT TTC ATT GCC TAC AAG CCC GAA GGC TAC
       -   -   -   -   -   -   -   -   -   -   - Glu   - Tyr
Zea1  CAG ACG CAG TGC GTC AGC TTC ATC GCC TAC AAG CCC CCG GGC AGC
       - Thr   -   - Val   -   -   -   -   -   -   -   - Ser
Cyn1  CAA GTT CAA ACT TTA ATG TTC TTA GTT TAC AAA CCA TTA XXX XXX
       -   -   - Thr Leu MET   - Leu Val   -   -   - Leu
                                                 117                                 123
```

FIGURE 1.—Alignment of mature SSU sequences. The nucleotide sequence (~369 nt) and encoded mature SSU polypeptide sequence (~123 aa) are presented for four SSU genes with varying degrees of divergence (soybean, Soy1: petunia, Pet1; maize, Zea1; and cyanophora, Cyn1). These example sequences illustrate the editing, alignment, and degree of homology among numerous distantly related SSU sequences compared in this study (for more details see MATERIALS AND METHODS). Insertions of codons to maintain correct codon alignment are marked with XXX. Amino acids which are identical with the soybean Soy1 sequence are marked with a (–). Nucleotide numbers for the mature sequence are listed above the sequences and mature amino acid numbers below and are based on the unedited mature soybean sequence, Soy1 (BERRY-LOWE *et al.* 1982).

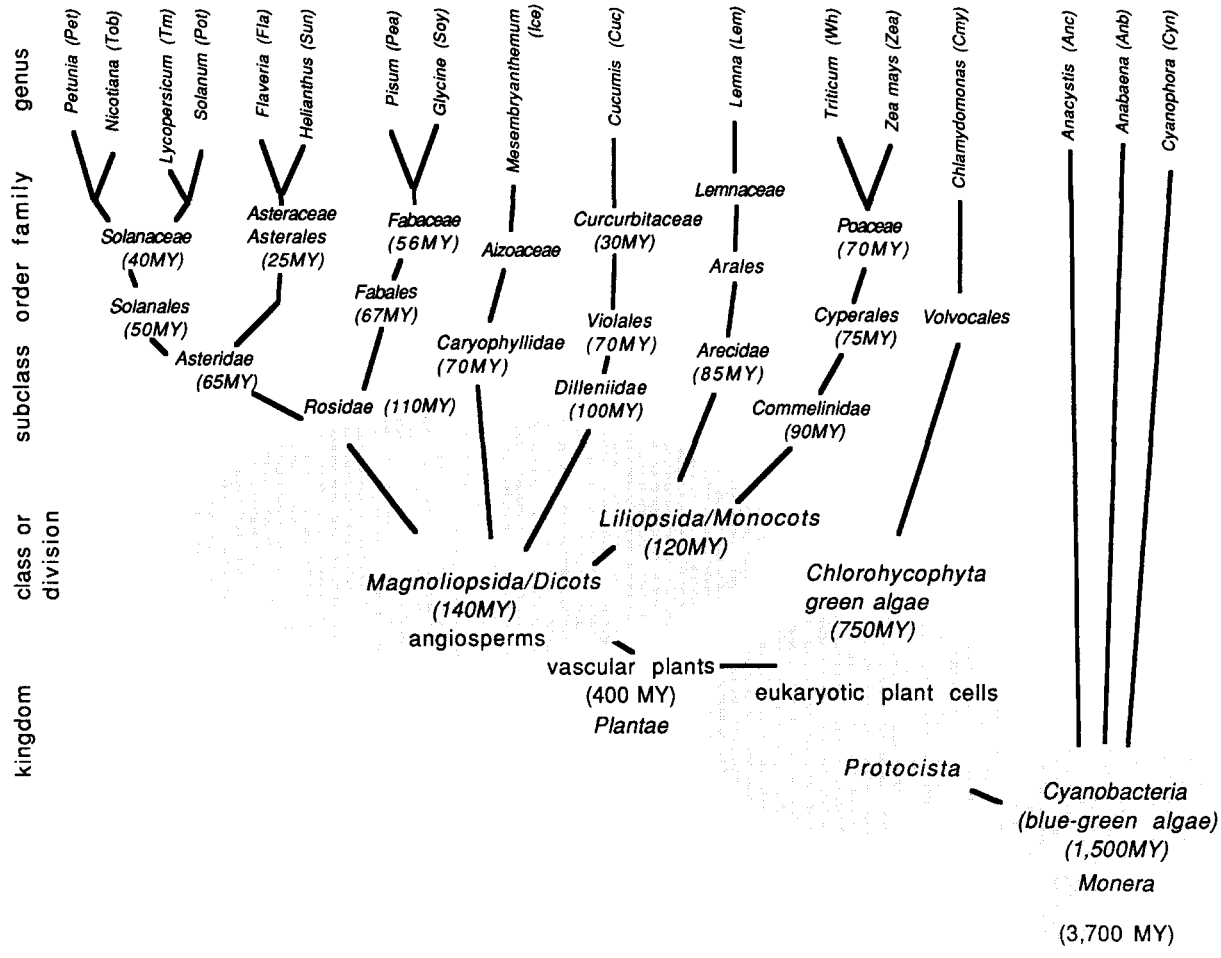

FIGURE 2.—Species tree relating the 17 organisms from which SSU sequences were examined. The phylogenetic relationships indicated are based on mega-fossils, pollen, morphology and chemistry of the various plant groups. When relevant, the taxonomic names of the genus, family, order, subclass, class and kingdom are indicated. The time in millions of years (MY) refer to the estimated time of divergence of the various species from a common ancestor or the age of the group. Additional details are given in MATERIALS AND METHODS.

not bias the calculation of RNS and SNS since the potential number of silent or replacement sites is taken into account in each pairwise comparison (PERLER *et al.* 1980).

Phylogenetic relationships among species: The approximate phylogenetic relationships among the 17 species examined were inferred from a literature survey of the fossil record and morphological traits (Figure 2). CRONQUIST's (1981) taxonomic treatment was used for the 13 angiosperms. Estimates of the divergence time of two species from a common ancestor were based on CRONQUIST's (1981) estimates of the time of origin of each class, subclass, order or family on megafossil and/or fossil pollen data. Nomenclature for a few subfamilies or tribes not described in CRONQUIST (1981) were taken from BECK (1976) or HEYWOOD (1978). Putative phylogenetic relationships and divergence times relating the three kingdoms, represented by vascular plants, green algae and blue green algae were obtained from TAYLOR (1981), MARGULIS and SCHWARTZ (1982), DOYLE and HICKEY (1982), JONES and LUCHSINGER (1986), GUNDERSON *et al.* (1987), and OCHMAN and WILSON (1987). CREPET and TAYLOR (1985) have mega-fossil data which extends back 56 MY for *Papilionoideae*, the subfamily to which pea and soybean belong, indicating that their family *Fabaceae* must be at least as ancient. The radiation of the angiosperm subclasses examined (*Commelinidae, Arecidae, Dilleniidae, Caryophyllidae* and *Rosidae*) lies early enough in the evolution of all angiosperms (140 MY) that no attempt was made to distinguish further their possible phylogenetic relationships or their divergence times. The time of origin and the time of divergence of most angiosperm subclasses will be assumed to be 140 MY. However, most taxonomic treatments based on fossil data suggest a recent divergence of the *Asteridae* from the *Rosidae*, about 65 MY before the present. The divergence of aquatic monocots from other monocots occurred some time after the divergence of monocots from dicots and was placed at 120 MY ago. These dates represent a minimum estimate of the origin and divergence of each group since the most ancient members (and in particular the common ancestors of the groups examined) were probably not prevalent enough to have been detected as fossils.

In certain cases the calculated divergence times may represent overestimates of the actual divergence times for several species. The phylogenetic relationship of the two tribes, *Cestreae* and *Solaneae*, within the family *Solanaceae* is not clear and thus the divergence time of all four species from a common ancestor will be assumed to be that of the origin of the family, 40 MY ago. Likewise, the time of family origin will be used to estimate the divergence time between the two *Asteraceae* species (25 MY), two *Fabaceae* species (56 MY) and two *Poaceae* species (70 MY). The oldest fossil record for green algal-like eukaryotic cells are dated at 750

	Pet3	Pea1	Pet1	Pet2	Lem2	Tm1	Tm2	Tm3	Soy2	Cyn1	Zea1	Fla1	Sun1	Ice1	Cmy2	Cmy1	Anc1	Soy1	Anb1	Lem3	Wh1	Pea2	Pea3	Cuc1	Tob1	Tm4	Tm5	Lem1	Wh2	Pot1	Tob2
Pet3	0	.8128	.4020	.6479	1.565	.7367	.3676	.5578	.9276	NaN	1.816	.9087	1.114	.7377	1.959	1.598	4.438	.8642	2.567	1.423	1.786	.8597	.8620	.8953	.4806	.3818	.3796	1.852	1.425	.5317	.4119
Pea1	.2104	0	.7178	.9968	NaN	.9875	.7539	.9400	.6478	1.084	NaN	.8868	1.013	1.163	NaN	NaN	1.656	.7306	1.961	NaN	NaN	.0421	.0807	.8620	.8845	.7539	.7757	NaN	1.759	.9249	.8845
Pet1	.0095	.1949	0	.7982	1.773	.7834	.3886	.5311	.9150	NaN	NaN	1.040	1.093	.8806	NaN	NaN	2.137	.9803	1.934	1.606	NaN	.7541	.7794	1.018	.4397	.3732	.4024	2.260	NaN	.5635	.3769
Pet2	.0071	.2040	.0103	0	1.078	1.115	.8188	.6552	1.024	NaN	1.426	.9985	1.111	.8241	1.614	1.382	1.579	1.034	NaN	1.052	1.511	1.060	1.222	1.065	.7063	.8744	.8315	1.204	1.395	.9543	.6183
Lem2	.1896	.2568	.1966	.1785	0	NaN	NaN	1.717	2.260	NaN	.4903	1.449	1.982	1.221	.4434	.3769	.9247	1.207	2.150	.0991	.5459	NaN	NaN	1.057	NaN	2.010	NaN	.0748	.5945	NaN	2.096
Tm1	.0878	.2092	.0895	.1015	.2077	0	.6699	.6149	.7252	NaN	NaN	1.127	1.120	1.136	1.902	5.016	NaN	.7596	1.749	NaN	1.794	.9434	1.041	.9688	.5664	.6863	.6824	NaN	1.609	.2419	.5664
Tm2	.0816	.1938	.0711	.0827	.1903	.0223	0	.5710	.8302	NaN	NaN	1.088	.9520	.9338	NaN	NaN	NaN	.8467	NaN	NaN	2.185	.7768	.8181	.811	.4427	.0272	.0087	NaN	1.839	.5188	.3698
Tm3	.0826	.1929	.0722	.0838	.1952	.0174	.0071	0	.8488	NaN	1.384	1.037	1.081	.6938	NaN	4.807	1.224	.9466	NaN	1.431	NaN	.9279	1.016	.7835	.4977	.5901	.5861	NaN	4.694	.3945	.4800
Soy2	.1235	.1467	.1318	.1295	.2013	.1619	.1569	.1591	0	2.427	1.297	1.053	1.243	.8594	1.305	1.524	NaN	.2322	1.898	1.256	1.369	.7163	.6728	.5267	.7637	.8879	.8465	NaN	1.368	.6018	.7226
Cyn1	.5664	.6528	.5701	.5680	.6061	.5880	.5987	.6093	.5464	0	NaN	1.416	1.136	2.181	NaN	NaN	NaN	NaN	1.623	NaN	NaN	1.058	1.055	NaN	NaN	NaN	NaN	NaN	3.703	NaN	NaN
Zea1	.2311	.2506	.2391	.2321	.1963	.2648	.2741	.2689	.2410	.5996	0	1.786	NaN	1.408	.4222	.4058	.8329	1.212	NaN	.5617	.4836	NaN	NaN	1.717	1.662	NaN	NaN	.5070	.6324	1.604	1.471
Fla1	.1535	.2119	.1668	.1659	.2126	.1715	.1780	.1795	.1394	.5998	.2473	0	.4634	1.003	NaN	NaN	1.406	.8657	2.667	1.292	1.371	.9311	1.005	.9914	.8913	1.112	1.117	1.460	1.295	.9039	.8913
Sun1	.2062	.1911	.1938	.2050	.2293	.2198	.2057	.2021	.1772	.6017	.3018	.0719	0	.9525	1.584	1.971	NaN	.9618	1.633	1.697	1.345	1.024	1.192	1.157	1.056	1.003	.9822	NaN	1.258	.8907	1.056
Ice1	.1463	.1965	.1451	.1507	.1876	.1557	.1515	.1499	.1539	.5539	.2313	.1762	.2072	0	1.522	1.289	1.408	.9097	1.650	1.294	1.258	1.225	1.206	.7483	.9182	.8578	.9712	1.397	1.196	.7642	.8705
Cmy2	.4289	.4235	.4300	.4260	.4172	.4105	.4188	.4151	.4056	.4947	.4434	.4285	.4383	.4198	0	.0471	1.027	1.151	1.753	.4836	.7601	NaN	NaN	1.408	NaN	NaN	NaN	.4892	.8009	1.850	NaN
Cmy1	.4204	.4401	.4215	.4175	.4124	.4085	.4104	.4105	.4152	.4911	.4420	.435	.4500	.4058	.0139	0	1.006	1.278	2.235	.4675	.7748	NaN	NaN	1.283	NaN	NaN	NaN	.4198	.8712	1.955	4.858
Anc1	.5851	.5986	.5677	.5519	.5923	.5954	.5955	.6031	.5330	.4468	.6184	.6009	.5490	.5706	.5133	.5101	0	1.288	1.596	.9347	1.437	1.799	1.671	1.425	1.581	NaN	NaN	.9049	1.465	NaN	1.305
Soy1	.1226	.1401	.1307	.1286	.214	.1591	.1492	.1485	.0169	.5464	.2303	.1492	.1781	.1630	.427	.4426	.5334	0	1.355	1.102	1.335	.8056	.8094	.6104	.7824	.9068	.8641	1.222	1.114	.6833	.7412
Anb1	.5382	.5325	.5296	.5234	.5293	.5719	.5750	.5746	.5014	.3392	.5505	.5383	.5578	.5527	.5907	.5693	.2880	.4978	0	2.237	2.117	2.008	1.712	1.909	NaN	NaN	NaN	5.650	1.976	1.992	NaN
Lem3	.2030	.2639	.2101	.1919	.0218	.2130	.1956	.2087	.2219	.6152	.2234	.2267	.2448	.2135	.4525	.4475	.6241	.2306	.5518	0	.5462	NaN	NaN	1.091	NaN	1.806	NaN	.0625	.5567	1.738	4.483
Wh1	.2440	.2573	.2557	.2428	.1768	.2388	.2552	.2579	.2378	.5900	.1683	.2288	.2478	.2300	.4371	.4351	.5957	.2477	.6155	.1918	0	NaN	NaN	2.089	1.605	1.990	2.212	.5645	.2261	1.336	1.525
Pea2	.2104	0	.1949	.2040	.2568	.2092	.1938	.1929	.1467	.6528	.2506	.2119	.1911	.1965	.4235	.4401	.5986	.1401	.5325	.2639	.2573	0	.0971	.9471	.8896	.7768	.8035	NaN	2.351	.9179	.8896
Pea3	.2016	.0041	.1860	.1951	.2448	.2004	.1849	.1840	.1574	.6411	.2348	.2215	.2000	.1961	.4446	.4457	.6038	.1409	.5203	.2520	.2437	.0041	0	.9002	.9555	.8580	.8470	NaN	5.080	.9963	.9555
Cuc1	.1946	.2364	.1931	.1962	.1911	.2045	.1997	.2042	.1696	.5469	.2591	.2072	.2165	.2275	.4440	.4559	.5858	.1706	.5749	.2156	.2313	.2364	.2336	0	.8919	.899	.8289	1.210	2.290	.7482	.8492
Tob1	.0949	.1670	.0843	.0962	.2083	.0460	.0311	.0304	.1542	.5999	.2747	.1696	.1938	.1461	.3990	.4078	.5827	.1467	.5711	.2138	.2568	.1670	.1610	.1942	0	.4420	.4524	NaN	1.514	.4316	.0620
Tm4	.0816	.1938	.0711	.0827	.1903	.0223	0	.0071	.1569	.5987	.2741	.1780	.2057	.1515	.4188	.4104	.5955	.1492	.5804	.1956	.2552	.1938	.1849	.1997	.0311	0	.0366	NaN	1.743	.5316	.3705
Tm5	.0851	.1900	.0746	.0862	.1863	.0255	.0031	.0102	.1531	.6037	.2697	.1742	.2017	.1477	.4133	.4049	.6084	.1455	.5677	.1916	.2509	.1900	.1810	.1997	.0279	.0031	0	NaN	1.851	.5290	.3799
Lem1	.1917	.2591	.1987	.1805	0	.2084	.1911	.1960	.2034	.6101	.1923	.2148	.2350	.1897	.4124	.4076	.5962	.2161	.5329	.0218	.1759	.2591	.2471	.1962	.2104	.1911	.1871	0	.5801	NaN	NaN
Wh2	.2520	.2596	.2637	.2508	.1793	.2467	.2611	.2604	.2404	.5778	.1622	.2346	.2481	.2305	.4430	.4412	.5947	.2536	.6088	.1943	.0057	.2596	.2462	.2318	.2626	.2611	.2567	.1784	0	1.371	1.441
Pot1	.0877	.1906	.0824	.0943	.2084	.0319	.0419	.0369	.1417	.6211	.2649	.1643	.2042	.1363	.4302	.4217	.5813	.1389	.5302	.2138	.2345	.1906	.1817	.2121	.0527	.0419	.0386	.2092	.2424	0	.4316
Tob2	.0949	.1670	.0843	.0962	.2083	.0460	.0311	.0304	.1542	.5999	.2747	.1696	.1938	.1461	.3990	.4078	.5827	.1467	.5711	.2138	.2568	.1670	.1610	.1942	0	.0311	.0279	.2104	.2626	.0527	0

FIGURE 3.—Nucleotide substitution among SSU sequences. The corrected RNS values (below the diagonal) and the corrected SNS values (above the diagonal) are shown for 31 pairwise comparisons of the SSU mature coding sequence. The largest correction for multiple hit kinetics in the RNS data occurred for the Cyn1/Pea1 comparison correcting 0.3964 to 0.6528 RNS, shown in the table. NaN refers to the corrected SNS values which approach infinity and thus were not numbers.

MY. Furthermore, there is increasing evidence that divergence of the plant, animal and fungal groups occurred at 1000 MY and that the split between green algae is higher plants occurred 600–800 MY. Therefore, a value of 750 MY is used for the divergence of green algae and higher plants from a common ancestor. Although the blue green algae can be traced back 3700 MY, the period of greatest radiation of cyanobacterial species based on fossil data (MARGULIS and SCHWARTZ 1982), ribosomal RNA sequence data (SOGIN *et al.* 1987) and oxygen accumulation (OCHMANN and WILSON 1987) is estimated to be from 600 to 2500 MY. Thus, the origin and divergence of the two blue green algal species, *Anacystis* and *Anabaena*, and one blue green algal symbiont, *Cyanophora*, is placed approximately at 1500 MY. It has been estimated that the divergence of blue green bacteria from the endosymbiotic precursor of mitochondria and chloroplasts may have occurred 1300–1500 MY ago (OCHMANN and WILSON 1987). Even though eukaryotic plant cells probably arose less than 800 MY, it is not clear that any of the extant blue green bacteria examined are in this direct line to eukaryotes and thus the divergence time of these blue green bacteria from the plant kingdom and from the green alga examined is presumed to be 1500 MY.

RESULTS

The mature portion of SSU is a well conserved polypeptide found in most eukaryotic photosynthetic cells and in the blue green bacteria. The molecular evolution of the DNA sequences encoding the mature SSU polypeptide has been investigated in order to begin to understand the evolutionary constraints and genetic mechanisms which have guided changes in these genes over several thousand million years.

Silent and replacement nucleotide substitution: The fractions of corrected SNS and RNS were examined among the 31 representative mature SSU sequences (Table 1) and the resulting 465 pairwise sequence comparisons are shown in Figure 3. Corrected SNS as a fraction of the total potential SNS sites are shown in the upper half of Figure 3, while the corrected RNS as fraction of all potential RNS sites are shown on the lower half of Figure 3. The values of corrected RNS are summarized in Table 2 for several categories of comparisons of SSU genes. Two or more gene family members with measurable sequence divergence were examined from eight different eukaryotic species. Among all comparisons the lowest RNS values were obtained for the comparisons among the genes within each of these SSU gene families and ranged from 0 to 0.026 (0–2.6%). The fraction of RNS was also small (0.027–0.1) for comparisons among SSU genes from the four different species of family *Solanaceae* (Pet1-3; Tob1, 2; Pot1; Tm1-3) and between the two species of family *Asteraceae* (Sun1; Fla1). Significantly more RNS was obtained for SSU comparisons among three subfamilies of monocots, the five families of dicots examined (group C, 0.12–0.207) and between the SSU sequences from monocots and dicots (group D, 0.207–0.3). A distinct increase in RNS is obtained for comparisons of the SSU sequences from the *Protocista* green algae *Chlamydomonas* (Cmy) with all other eukaryotic sequences (group E, 0.34–0.46). Another significant increase in RNS is observed for all comparisons of the prokaryotic cyanobacterial SSU sequences and eukaryotic sequences (group F, 0.49–0.65).

All 465 pairwise gene comparisons fit within these classes of RNS with the following exceptions involving groups C, D and E (Table 2). All comparisons of soybean with pea SSU sequences give RNS values greater than most comparisons within a family and

TABLE 2

Summary of corrected RNS values for 465 SSU comparisons

Group	Majority of SSU comparisons in category	Corrected RNS	No. comparisons	No. exceptions
A	Within gene families	0–0.026	23	0
B	Among different *Solanaceae* or among *Asteraceae* species	0.027–0.10	42	0
C	Among subfamilies of monocots or among families of dicots	0.12–0.207	151	32[a]
D	Between monocots and dicots	0.207–0.302	111	13[b]
E	Between green algae and angiosperms	0.34–0.46	54	2
F	Between cyanobacteria and eukaryotes	0.49–0.65	84	0

[a] Most of the exceptions which do not fit in this category are comparisons of the SSU sequences of the monocot, *Lemna*, with SSU sequences in dicot species.

[b] Most of the exceptions which do not fit into this class are comparisons of SSU sequences among a few distant subclasses of dicots.

fall in group C. The break between groups C and D (0.207) is the least precisely defined. Eleven comparisons of sequences among dicot subclasses show large RNS values (>0.207) and also fall within group D comparisons, monocots with dicots. Twenty five comparisons of Lemna sequences with dicot sequences give lower RNS values than other monocot *vs.* dicot comparisons and fall in group C. The comparisons among the three cyanobacterial sequences (Cyn1, Anc1 and Anb1) which ranged from 0.288 to 0.41 do not fit within the headings provided in Table 2 and have been listed as exceptions. The taxonomic relationships of the members of this group to one another are obscure.

The numbers of silent nucleotide substitutions were so high in comparison across subclasses, divisions and kingdoms that most silent site changes occurred at sites that had already changed at least once. Thus most of these observed SNS values give corrected SNS values greater than 1 (or >100% SNS). However, most of the sequence comparisons within individual subclasses show low SNS values and will be discussed in greater detail below.

A phenogram based on RNS reveals concerted evolution within all gene families: A tree of SSU sequences (Figure 4) shows the relationship of each of the 31 SSU sequences to each other based on the corrected RNS data (Figure 3). This gene tree was developed using the neighbor joining method of SAITOU and NEI (1987). A similar tree (not shown) with only a few structural differences was obtained using the unweighted pair group mean averaging method, UPGMA (SOKAL and SNEATH 1963).

It is logical to assume that the ancestral plant species at each node in the phenogram contained its own SSU multigene family. The individual members of this gene family would have diverged to some degree from each other. Derivatives of this plant would be expected to maintain these SSU "gene lineages" with some sequence changes. However, enough of the original sequence of each "gene lineage" should be maintained so that each could be identified as a distinct and ancient lineage. Therefore, it is reasonable to attempt to identify ancient lineages of SSU se-

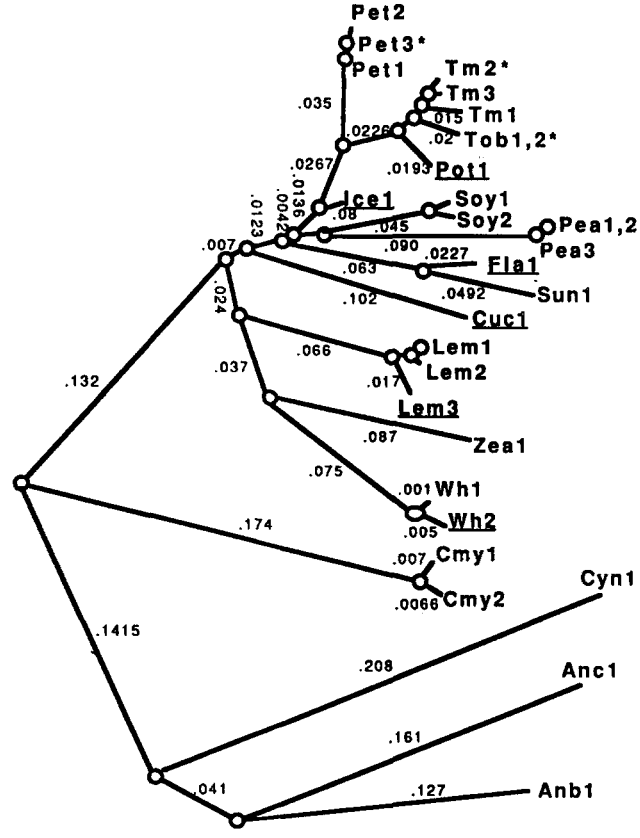

FIGURE 4.—Phenogram relating mature SSU sequences based on corrected RNS. This gene tree was constructed by the neighbor joining method (SAITOU and NEI 1987) using corrected RNS values (Table 2). The RNS distances relating all sequences as calculated by this method are shown for branches of significant length. They are approximately proportional to the fractional RNS values (Table 2) between any two particular species shown in the tree. Although this is an unrooted tree, it has been displayed in this orientation to show its similarity to the species tree in Figure 2. SSU sequence designations may be referred to Table 1. The Pet3*, Tm2* and Tob2* are highlighted by an asterisk to distinguish their three-intron structure from the remaining dicot sequences, which have only two introns. Mature SSU sequences derived from cDNA clones are underlined. The portion of the RNS tree relating SSU sequences from family *Solanaceae* is expanded in Figure 5A.

quences within SSU gene families which predate the divergence of the species. Although SSU sequences are moderately well conserved, these lineages should show some differences in RNS positions relative to each other since significant RNS has accumulated in the SSU genes among all divergent species (BERRY-LOWE *et al.* 1982; PICHERSKY *et al.* 1986). However, the most striking feature of the neighbor joining tree (Figure 4), which is confirmed in the UPGMA tree, is that all the mature SSU sequences within any single species are more similar to each other than to the SSU genes in even the most closely related species. This result holds for all of the SSU gene family members examined within eight different multigene families representing members of the monocots, dicots, and one green algal species.

If ancient lineages of SSU sequences exist, which predate species divergence, they might be most obvious in comparisons of SSU sequences among the relatively closely related species of *Solanaceae*. The majority of SSU sequences and gene subfamilies have been characterized in petunia and tomato. The three different SSU gene subfamily representatives examined from tomato (Tm1-Tm3) are distributed on three different chromosomes (VALLEJOS, TANKSLEY and BERNATZKEY 1986) and the three gene subfamily representatives examined from Petunia (Pet1-3) are not closely linked (DEAN *et al.* 1987). Pet1, Pet2 and Pet3, representing the three divergent SSU gene subfamilies in petunia, have very few RNS between them. (Also see group A in Table 2.) However, when petunia SSU sequences are compared with SSU sequences from tomato, tobacco and potato, or with any other higher plant species, significantly more RNS are observed across species. Based on the conservation in RNS positions for individual members of these SSU gene families, no obvious lineages are conserved across these species within *Solanaceae*.

Concerted evolution among SSU genes evidently occurs within genes with different numbers of introns, located on nonhomologous chromosomes. All of the SSU genes examined in division *Dicotyledona* have only two introns with the exception of the family *Solanaceae* where a single gene family member with three introns has been characterized in several species. These three intron genes have been found in petunia (Pet3*), tomato (Tm2*) and tobacco (Tob2*) and are marked with an asterisk in Table 1, Figure 4, and Figure 5A, where the expanded examination of RNS between *Solanaceae* species is shown. The position of the third intron is identical in Pet3*, Tm2* and Tob2*. All other genes examined from the gene families within *Solanaceae* have the two introns typical of SSU sequences from dicots. It is unlikely that a third intron arose independently in precisely the same location in each of these *Solanaceae* species. Therefore Pet3*,

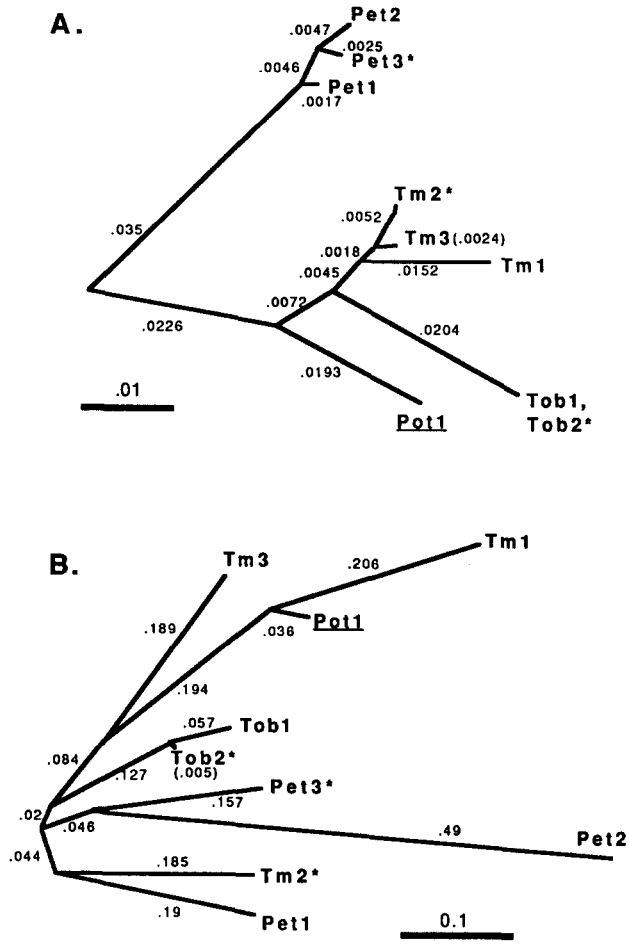

FIGURE 5.—Phenograms relating mature SSU sequences from family *Solanaceae*. Bars indicate relative scales of the two trees based on fractional corrected RNS (A) or fractional corrected SNS (B). For more details see Figure 3.

Tm2* and Tob2* would be expected to share a single common ancestral gene, belonging to an ancient gene lineage, which predates the divergence of these genera of *Solanaceae* and which had diverged at that time from the gene lineage containing two introns. Yet, the mature portions of each of the SSU genes containing three introns have fewer RNS when compared to members of their own gene family than when compared to their three intron counterparts in the other species of *Solanaceae*. Thus, concerted evolution has obscured variation among ancient lineages predicted based on intron structure and on divergent chromosomal locations. A similar relationship has been observed among two classes of 7S seed storage protein genes in *Fabales*, where the genes with and without an insertion in soybean show more homology than gene comparisons of either class of 7S gene between soybean and common bean (DOYLE *et al.* 1986).

Ancient lineages of SSU sequences which predate species divergence also are not clearly revealed by SNS comparisons. Figure 4B shows the NJ tree for the corrected SNS between the *Solanaceae* SSU genes examined. As expected, SNS accumulate much more

rapidly than RNS between closely related species (Figure 3) and this is revealed in the longer branch lengths between most sequences. The SNS tree is distinct from the RNS tree in that it does not reveal sequence similarity between all the SSU genes within a gene family. However, there is not enough similarity in SNS among members of the three-intron SSU gene lineage across species which would completely distinguish this lineage from the other SSU gene family members within a species. For example, Tob2*, a three intron gene, is closer to Tob1, a gene containing two introns, than to Tm2* or Pet3*. Likewise, Pet3* is as close to Pet1 as it is to Tm2*. The variability in SNS accumulation and branch length is due in part to the fact that many of the SNS positions have been changed more than once and will be addressed in the discussion.

Since ancient lineages of SSU genes which predate species divergence cannot be detected within the eight multigene families examined a mechanism of concerted evolution must regularly homogenize the SSU gene sequences. The fact that two-intron and three-intron containing genes can share the same nucleotide sequences suggests that gene conversion plays a significant role in the concerted evolution of this gene family (NAGYLAKI and PETES 1982; NAGYLAKI 1988). SSU multigene families have been found in all eukaryotes in which the SSU genes have been examined in detail. It is likely, therefore, that a mechanism of concerted evolution has acted on the SSU multigene families throughout the history of most photosynthetic eukaryotes. Furthermore, based on the data in the family *Solanaceae* and the limited sequence difference within each of the gene families examined, it is likely that concerted evolution has acted on the SSU sequences at least once in the life of each genus.

Correlation between the gene tree and the species tree: As with ribosomal gene families, the inability to observe ancient lineages of SSU genes within gene families based on DNA sequence allows mature SSU sequences to be compared across plant species. In other words, the divergence time of the individual multigene family members need not be considered as significant when comparing SSU sequences between different species. Figure 2 presents the proposed phylogenetic relationships among the 17 genera examined from organismal and geologic considerations as described in MATERIALS AND METHODS. There is an obvious similarity in the neighbor joining SSU gene tree (Figure 4) and the species tree (Figure 2).

Estimating rates of accumulation of SNS and RNS in SSU sequences: The overall similarity of the RNS-based gene tree and morphologically based species tree allows a calculation of the rates at which SNS and RNS have accumulated among SSU sequences in different organisms. Because all the sequences within a gene family have been homogenized, the divergence of the individual SSU sequences within a multigene family and their gene lineages need not be considered. The fraction of SNS and RNS for SSU genes in different species is compared with the approximate time of divergence of the species from a common ancestor. The divergence times, in millions of years (MY), are shown in Figure 2 and represent only an estimate for the firm establishment of that plant group based on fossil plant and pollen data (see MATERIALS AND METHODS).

Figure 6, A and C, examines RNS and SNS, respectively, as a function of 1500 MY of divergence for 442 pairwise comparisons of SSU sequences among all 17 species examined. Figure 6, B and D, expands data in A and C, respectively, examining only the first 70 MY of divergence. The 23 comparisons made among multigene family members within their parent species were omitted, because no estimate of their divergence time can be made from the fossil record (group A, Table 2).

The fraction of corrected RNS between any two SSU sequences increases relatively rapidly for the first 70–100 MY until it reaches a level of 0.2 (20%; Figure 6A). From this point the corrected RNS increases very slowly for the next 1400 MY of divergence, suggesting that SSU sequences are restricted by some mechanism against further rapid change in RNS. Although the rate of accumulation of corrected RNS is not linear over the entire time period examined, it is instructive to calculate the expected rates of divergence predicted from the initial and late divergence times in Figure 6, B and A, respectively. In calculating expected nucleotide substitution rates, the combined divergence time of both species from a common ancestor must be considered (*i.e.*, twice the time shown in Figures 1 and 6). The expected rate of RNS is nearly linear during the initial 70 MY of divergence, ~1.25 ($\pm$0.32) $\times$ 10^{-9}/site/yr (1% RNA/8 MY), with a rapid decline in the rate beginning during the next 70 MY of divergence. The expected rate of RNS for the divergence of monocots and dicots from green algal and cyanobacterial sequences, covering the next 1400 MY is ~0.13 ($\pm$0.06) $\times$ 10^{-9}/site/yr (1% RNS/75 MY). This rate is 10 times slower than the rate at which RNS accumulate during the first 70 MY. Because this change in the rate of RNS accumulation occurs at such low RNS values it is not due to problems of multiple hit kinetics (see discussion, PERLER *et al.* 1980).

The rate of accumulation of SNS is significantly greater than for RNS and there is a much greater spread in the SNS data, but the same overall shape is observed in these curves (compare Figure 6, A to C). The accumulation of corrected SNS is nearly linear during the first 70 MY of divergence, when it reaches

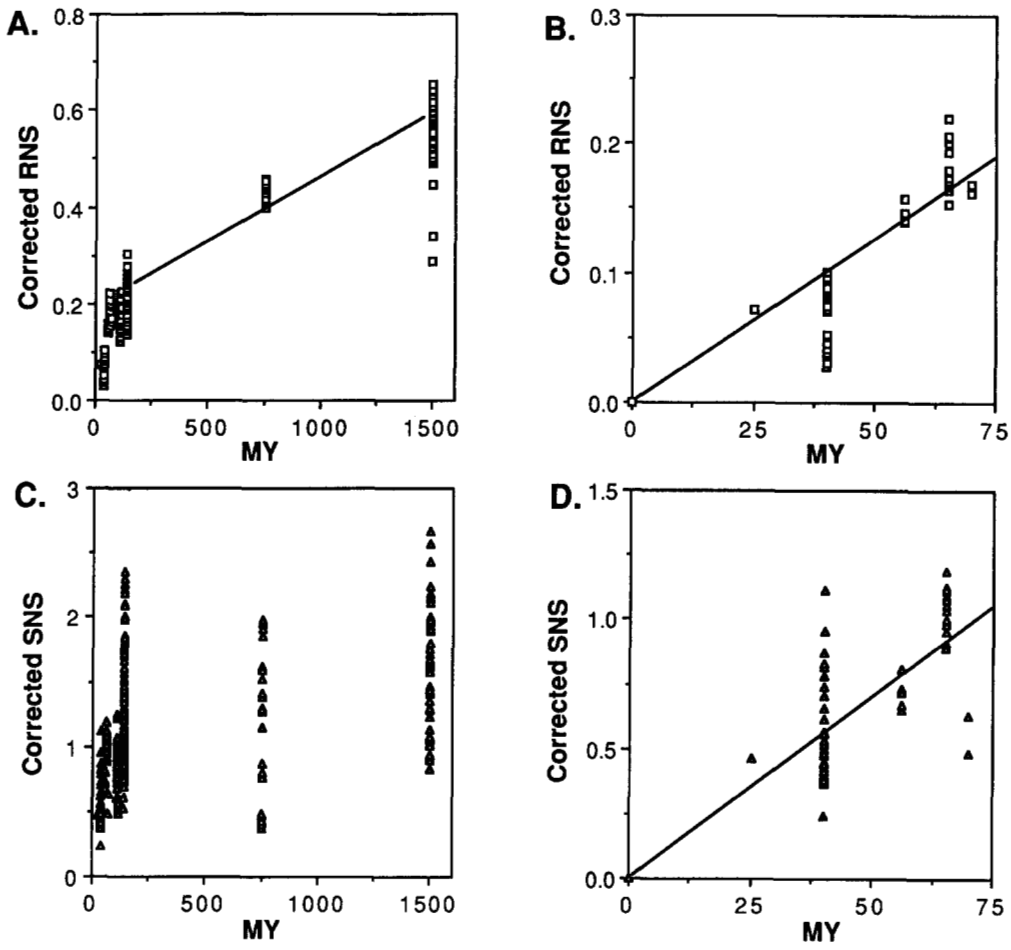

FIGURE 6.—Corrected RNS and corrected SNS as a function of time. In A and C, corrected RNS and corrected SNS, respectively, are examined as a function of 1500 MY of divergence for SSU sequences in different species. In A and C, 442 pairwise comparisons and 330 pairwise, comparisons are presented, respectively. In C, 112 of the original 442 SNS values are not shown because they corrected to extremely high values (see text and Table 2). In B and D, the corrected RNS and corrected SNS, respectively, are examined for the 73 pairs of sequences examined in different species over the initial 70 MY of divergence. The scales for fraction of nucleotide substitutions, corrected RNS and corrected SNS, and time scale in millions of years (MY) are indicated. Lines approximating the initial rates of RNS (B), rates of SNS (D), and long term rate of corrected RNS (A) for the divergence of two sequences from each other are shown. To obtain the expected nucleotide substitution rates for the divergence of two sequences from a common ancestral sequence as discussed in the text these rates must be divided by 2 to adjust for the time of divergence of both sequences from their common ancestral sequence.

a value of 1 (100%; Figure 6D). This slowdown in the accumulation of SNS may be due to a saturation of SNS sites with multiple changes and is discussed below (PERLER *et al.* 1980). For another viewpoint see GOJOBORI (1983). An expected SNS rate of 6.6 ($\pm$1.5) $\times$ 10^{-9}/site/yr (1% SNS/1.5 MY) is obtained for the first 70 MY, five times the initial RNS rate for SSU sequences. This is very close to the expected SNS rate (5.37 $\times$ 10^{-9}; 1% SNS/1.86 MY) derived for functional animal genes (MIYATA and HAYASHIDA 1982; MIYATA *et al.* 1982). The initial estimate of the rates of SNS from several plant nuclear genes is 6 $\times$ 10^{-9} [1% SNS/1.6 MY; WOLFE, LI and SHARP (1987) and WOLF, SHARP and LI (1989)] very similar to these extensive estimates for SSU sequences.

The estimates of initial rates of both RNS and SNS shown in Figure 6, B and D, respectively, have been based on zero divergence at zero time. Without this assumption the SNS and RNS data for the first 70 MY of divergence would intersect the abscissa at 5–10 MY suggesting that on the average the estimates of divergence times between species may be too great (see MATERIALS AND METHODS). Reducing the divergence times for these species by 10 MY would have the effect of increasing the expected rates of RNS and SNS by about 15%.

Transitions *vs.* transversions: Half as many transition (ts) nucleotide substitutions will have occurred relative to the number of transversion (tv) nucleotide substitutions (ts/tv = ½) if transitions and transversions occur randomly (NEI 1987). Estimates of ts/tv ratios taken from closely related plant ribosomal sequences give an observed ts/tv ratio of 2:1 (HAMBY and ZIMMER 1988; ZIMMER *et al.* 1989). An estimate of the ts/tv ratio for any gene comparison requires that the observed nucleotide substitution rate approx-

imate the mutation rate at those sites examined. Therefore, the positions compared should not be under strong selection and should not have accumulated multiple changes which could bias the estimates (BROWN *et al.* 1982). Of the 465 pairwise SSU sequence comparisons, only 10 had a fractional uncorrected SNS lower than 0.06 where the number of multiple hits would be expected to be negligible. These low SNS values are only obtained for comparisons between the most closely related SSU multigene family members within a single species, for example between Pet2 and Pet3*. The SNS ts/tv ratio for these 10 comparisons ranged from 0 to 10 with an average of 2.4. This value can serve as an initial estimate of ts/tv ratios in SSU sequences with the following caveats. (1) The SNS positions in SSU sequences have not been shown to accumulate mutations randomly and even these low SNS values showed some correction for multiple hit kinetics (PERLER *et al.* 1980). (2) It is not possible to determine whether these small differences truly represent nucleotide substitution or the effect of the last event of concerted evolution.

DISCUSSION

Gene conversion contributes to the concerted evolution of all SSU gene family members

Analysis of both replacement and silent site changes does not reveal distinct ancient SSU gene lineages which predate divergence of the species examined from their common ancestors. This is quite different from observations on the evolution of many family gene sequences in animals and plants. For example, ancient lineages of actin genes, representing hundreds of millions of years of divergence are easily identifiable from conserved RNS positions and in the corresponding amino acid sequences (HIGHTOWER and MEAGHER 1985, 1986). The dispersal of SSU genes among different chromosomes, and the existence of a uniquely positioned third intron present in only one gene family member in species of *Solanaceae*, imply that ancient gene lineages which predate species divergence do exist within SSU gene families. Sufficient RNS occur in SSU sequences between all species examined for these lineages to be revealed if they had evolved independently of one another.

The similarity in RNS positions among all SSU gene family members demonstrates that homogenization of SSU sequences has occurred within each species. A mechanism of gene amplification and loss, similar to that which may produce the concerted evolution of ribosomal RNA genes alone cannot account for the homogenization of SSU sequences. Amplification and loss of sequences could not homogenize sequences among two and three intron genes and maintain the three intron structure of one gene. The DNA turnover mechanism involved in this case must also account for the movement of homologous sequences among chromosomes. The SSU sequences therefore, must be homogenized by a mechanism that includes gene conversion. Previous studies have suggested gene conversion as one possible mechanism for the concerted evolution of SSU sequences in *Solanales* (PICHERSKY *et al.* 1986; SUGITA *et al.* 1987; DEAN *et al.* 1987). This gene conversion process must act on SSU sequences in different contexts (*i.e.*, in different regions of the genome). The process must homogenize relatively short stretches of sequence, altering the coding regions of these genes but not the adjacent introns. Strong evidence has been presented for gene conversion between linked globin genes (SLIGHTOM, BLECHL and SMITHIES 1980; SCOTT *et al.* 1984), linked and unlinked rDNA genes (ARNHEIM *et al.* 1980), unlinked catachol oxygenase genes (DOTEN, GREGG and ORNSTON 1987), and between plasmid and chromosomal copies of the *spoC* locus, suggesting that the process can occur between nonallelic regions of a genome. In yeast meiotic gene conversion events have been shown to occur frequently between nonhomologous chromosomes (JINKS-ROBERTSON and PETES 1986). Evidence for gene conversion of small sequence tracts also has been presented (SYMINGTON and PETES 1988; JUDD and PETES 1988; ORNSTON, NEIDLE and HOUGHTON 1988). The evolutionary mechanism acting on SSU sequences must have functioned frequently enough to homogenize RNS mutations which occur among all the multigene family members. Therefore, homogenization must have occurred at least once for each gene in the lifetime of each genus examined. Because the gene conversion process could have occurred in any order between the SSU sequences, the variation in branch lengths in the RNS and SNS trees between different SSU sequences within a species could, among other possibilities, reflect the order and the time of gene conversion events.

One trivial explanation suggested for the homogeneous conservation of RNS positions in all SSU sequences in a gene family is that selection acts independently on all SSU coding sequences to accumulate similar mutations in each gene family member, leading to the production of a homogeneous set of mature SSU polypeptides (PICHERSKY *et al.* 1986; TANKSLEY and PICHERSKY 1988). Such a selective force may exist due to the intimate association of the SSU and LSU polypeptides in the Rubisco holoenzyme (see below) and the fact that the LSU is even more highly conserved than the SSU (YOSHINAGA *et al.* 1988). However, considering that the expected mutation rate at any one nucleotide site is on the order of 6.6×10^{-9}/site/yr, and the rate of changes accumulating in replacement positions is even smaller (1.25×10^{-9}), the chance of accumulating the same mutational changes

in several RNS positions within two or three gene family members without gene conversion is quite small. For example, assuming the average RNS rate for one nucleotide substitution the probability of obtaining the same mutation independently in three genes is 1.9×10^{-27}. Thus, the conclusion remains that gene conversion plays an active role in the concerted evolution of SSU sequences.

SSU gene trees and species trees from nucleotide divergence

Measurement of nucleotide divergence: An attempt has been made to relate RNS and SNS accumulation between two species to the divergence times of those species from a common ancestor (Figure 6). The PERLER *et al.* (1980) method of calculating RNS and SNS is complex and it has been suggested that it may underestimate these values (NEI 1987). For these reasons NEI and GOJOBORI (1986) developed an alternate, simplified unweighted pathway method. FELIX JIN and MASATOSHI NEI generously calculated the divergence among the 31 SSU sequences described herein by the method of NEI and GOJOBORI (1986). Using this calculation the values Dn and Ds refer to the fraction of corrected *n*onsynonymous site substitutions and to the fraction of corrected *s*ynonymous site substitutions, respectively, and may be compared with corrected RNS and corrected SNS. All corrected RNS values (PERLER *et al.* 1980) less than 0.3 were within 1–2% of the corresponding Dn values obtained by the NEI and GOJOBORI method and were sometimes greater than the Dn values. This represents 300 of the 465 corrected RNS values and all comparisons among angiosperms. For comparisons among different kingdoms where the corrected RNS values reached 0.65, the Dn values were as much as 10% larger. Corrected SNS and the corresponding values calculated by the NEI and GOJOBORI method, Ds, were within approximately 2% of each other for divergence times less than 75 MY (corrected SNS < 0.8). For divergence of 140 MY or greater the Ds values were on the average about 25% greater than the corrected SNS values. However, plots of Dn or Ds with respect to time of divergence over the first 70 MY and the plots of Dn *vs.* time over the entire time period also give rates of nucleotide substitution within 2% of those presented based on RNS and SNS, respectively (PERLER *et al.* 1980).

The correction for multiple hit kinetics (JUKES and CANTOR 1969) produces a great deal of scatter in calculating the corrected SNS values for the SSU sequences after 100 MY of divergence (Figure 6C), making SNS estimates for these greater divergence times impossible. Apparently, when the mathematical methods of PERLER *et al.* (1980), and the JUKES and CANTOR (1969) correction for multiple hits are applied to these sequences the formula over- or under-corrects for some classes of changes which occur in the SNS positions of the SSU sequences. The NEI and GOJOBORI (1986) method of calculating the fraction of corrected synonymous nucleotide substitution, Ds, shows the same degree of scatter in the data (not shown). This approach also relies on the JUKES-CANTOR correction. In the examination of the divergence of SNS sites in animal globin sequences the mathematical adjustment for multiple hit kinetics is more consistent in its correction after 80 MY of divergence and at similar levels of observed SNS (PERLER *et al.* 1980). The JUKES-CANTOR correction formula assumes for each nucleotide position within a codon that any of the three remaining bases may be substituted. As the observed fraction of SNS approaches the expected limit of ¾ this equation calculates the log of a number approaching zero and gives these extreme values. As a specific example of this problem, 107 of the comparisons from species more than 100 MY apart and most with observed SNS values >0.5, the corrected SNS values were calculated to be extremely large numbers (values termed NaN in Figure 3). Evidently, the actual level of SNS after the first 100 MY of SSU sequence divergence is so large or the classes of substitutions are so skewed from what is expected that it is unlikely that any method presently available can accurately correct for the level of multiple changes.

The compositions of G plus C nucleotides in the organism (%GC) may also contribute significantly to the apparent scatter in the SNS data. It is known that although the %GC compositions vary widely at SNS sites within different genes that these values drift toward the average GC composition of the organismal genome (SUEOKA 1961; BRINKMAN *et al.* 1987). The %GC in angiosperms alone varies widely from values as high as 80% in the monocots to as low as 30% in the dicots. The contribution of GC composition to the divergence of SNS sites within SSU sequences has not been investigated.

Topology and comparison of the SSU gene trees with species trees: The overall correlation between the topology of the RNS gene tree and the species tree suggests that SSU gene trees might be useful in constructing species trees which show the relative relationship of different species to common ancestors. The utility of this approach in recovering the correct topology will undoubtedly be dependent upon having sufficient SSU sequence data and is subject to the inherent limitations in constructing a gene tree (NEI 1987). The relatively linear rate of corrected RNS and corrected SNS accumulation over short time periods (*i.e.* the first 100 MY of divergence) suggests that the SSU sequence data can probably be used to calculate times of species divergence within an order and perhaps between orders of seed plants. These

data may act to support proposed phylogenetic relationships among all plant species. The tight clustering of all corrected RNS data over the entire 1500 MY of divergence examined may allow even greater divergence times to be approximated.

Discrepancies between the topology of the neighbor joining (NJ) gene tree (Figure 4) and of the species tree (Figure 2) may be due in part to inaccuracies in estimates of divergence relationships among angiosperm orders and subclasses, where the fossil record and determination of the homology and polarity of morphological character changes are weak. Although this manuscript has followed the phylogenetic relationships of angiosperms suggested by CRONQUIST (1981) extremely different phylogenetic schemes are under active consideration by plant systematists (JONES and LUCHSINGER 1986).

For the species examined herein the data may also suggest new relationships. Based on macroscopic morphological characteristics, tobacco has long been grouped in a tribe with petunia and separate from tomato or potato (HEYWOOD 1978). The topology of SSU sequence divergence suggests a different grouping. For example, Pet1 and Tm3 have the minimum corrected RNS between any petunia and any tomato SSU gene (0.074 RNS). Based on an overall RNS rate of 1.25×10^{-9}/site/yr obtained from the first 70 MY years of SSU sequence divergence it can be estimated that the divergence time between these two species is 60 MY and that they have not had a common ancestor for approximately 30 MY. Tobacco and tomato would diverge from a common ancestor 12 MY ago by a similar calculation. Similar short divergence times are calculated for tobacco *vs.* potato or tomato *vs.* potato. PICHERSKY *et al.* (1986) and MCKNIGHT *et al.* (1986) also noted that the tomato SSU genes were far more similar to tobacco than to petunia SSU genes. Recent data on chloroplast genome evolution also suggests that tomato, potato and tobacco are more closely related to each other than they are to petunia (PALMER *et al.* 1988).

Another discrepancy between the branching in the gene tree and the accepted species tree is the apparent high degree of divergence of the two genes of order *Asterales*, Fla1 and Sun1, from four species within order *Solanales*, petunia, tobacco, tomato and potato within *Asteridae*. Analysis of the evolution of chloroplast DNA sequence also suggests a distinct and possibly ancient split between these orders (JANSEN and PALMER 1987; PALMER *et al.* 1988) in contrast to the relatively recent divergence indicated in Figure 2. However, these data still strongly support the simpler phylogenetic relationships proposed by CRONQUIST (1981) over several alternative phylogenetic models (JONES and LUCHSINGER 1986). Perhaps when even more SSU sequence data or data on other plant gene families is available these discrepancies can be resolved.

In the SSU gene tree the six subclasses of angiosperms all show deep branches going close to the root of the gene tree which are not well resolved from each other. This is consistent with the early diversification of angiosperms 110–140 MY ago. However, in one case in particular the gene tree branching order for subclasses is distinct from the branching in the species tree. The branch for Ice1 is found between Soy/Pea (*Rosidae*) and Pet/Tob/Tm and Pot1 (*Asteridae*). It is well accepted from morphological data that *Asteridae* arose from *Rosidae* (CRONQUIST 1981). *Mesembryanthemum* (Ice1), a member of the family *Aizoaceae*, belongs to the subclass *Caryophyllidae*, and is thought to have arisen more recently than *Rosidae*, however, there is no morphological evidence that *Caryophyllidae* arose from *Rosidae* as suggested from the gene tree. Analysis of a large amount of chloroplast DNA sequence data predicts a species tree which agrees with the SSU gene tree in placing *Rosidae* and *Caroyphyllidae* relatively close together (WOLFE, LI and SHARP 1987).

A similar gene tree was obtained by the unweighted pair group mean averaging (UPGMA) analysis method of SOKAL and SNEATH (1963) (data not shown) which had only a few topographical differences from the species tree in Figure 2 or the NJ gene tree presented in Figure 4. These differences centered on the order of the deep branches among dicot orders and subfamilies. The NJ method of gene tree construction gives the central branches the respective order Ice1, Soy/Pea and Fla1/Sun1. Their respective order in the UPGMA tree is Soy, Ice1, Fla1/Sun1, and Pea, showing greater difference from the presumptive species tree. Although the remaining branching in the UPGMA tree is very similar to the NJ tree, the topology of the NJ tree more closely approximates the species tree than does the UPGMA tree. This is consistent with the observations of SAITOU and NEI (1987) that the UPGMA method is not as effective as the NJ method at recovering the correct topology of a tree when rates of evolution are not uniform across lineages.

The relative similarity in RNS among the three cyanobacterial sequences (Anb1, Anc1 and Cyn1) is interesting considering that these organisms are morphologically quite distinct and that the major diversification of cyanobacteria occurred 1500 MY ago (see MATERIALS AND METHODS). The RNS values and the topological relationships shown in Figure 4 suggest that Anb1 and Anc1 may be only 300–500 MY diverged from a common ancestor (relatively late in the diversification of blue green algae), and that Anb1 and Anc1 may be no more than 600–1200 MY diverged from Cyn1. It should be mentioned that al-

though recent data suggests that bacteria have a similar rate of SNS to animals, it is not clear that the bacterial RNS rate for a particular sequence will be similar to other organisms (OCHMAN and WILSON 1987).

Concerted evolution of SSU sequences within a gene family may result from complex selective forces

Discrepancy between RNS and SNS levels within a gene family: It may be possible to elucidate the processes which lead to homogenization of SSU sequences within a multigene family by examining patterns of nucleotide change for the *Solanaceae* and for sets of other closely related species. Assuming that all SSU gene family members within a single organism have interacted via gene conversion at some time in their past, differences in RNS and SNS and in corresponding branch lengths in the RNS and SNS trees (Figure 5) between different SSU sequences within a species can be viewed in two extreme ways. First, the sequence differences between two SSU genes in a gene family might represent the amount of sequence divergence and hence the approximate time since the last round of gene conversion events which converted the two sequences to homogeneity. Second, the level of sequence divergence within a species could measure the amount of homogenization and sequence divergence remaining from a recent gene conversion event without any subsequent nucleotide substitution. The level of homogenization produced immediately following a gene conversion event is dependent upon what proportion of the genes sequence is converted and upon a relatively random repair process which is not expected to produce two identical sequences. In either case, and assuming that the gene conversion events are random, the degree of sequence divergence of RNS and SNS positions within species should follow the initial rates of RNS and SNS established for divergence of SSU sequences between different species. In other words the fivefold difference in initial SNS and RNS rates established among different species should result in there being a fivefold difference in RNS and SNS among gene family members. In fact, however, there is a discrepancy between the RNS and SNS values for comparisons within the SSU gene families of petunia, tomato, soybean, pea, *Lemna* and wheat. For example the 15 comparisons of tomato sequences to petunia sequences the SNS/RNS ratio is 7.4 (±2.5) close to the expected ratio of 5, whereas the ratio of SNS to RNS between two gene family members, Pet1 and Pet2, is 77. The ratios are 42 and 91 for the Pet1/Pet3* and Pet2/Pet3* comparisons, respectively. In other words, there are approximately 70 times more SNS than RNS within the petunia SSU gene family, whereas the expected ratio is about 5.

A comparison of SNS as a function of increasing

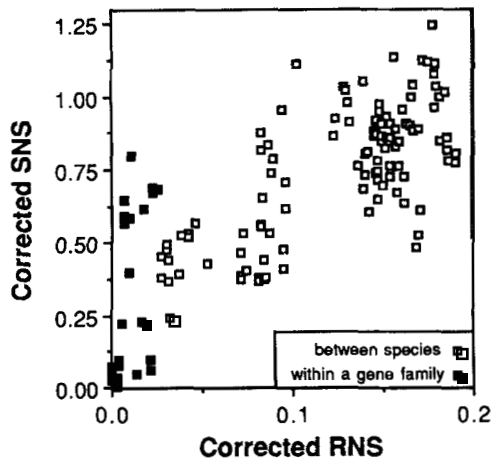

FIGURE 7.—SNS as a function of RNS for sequences with RNS values less than 0.19 (data from Table 2). The open squares show 117 SSU sequence comparisons among different species within the two divisions of angiosperms (coefficient of determination = 0.55). The closed squares show 23 comparisons of SSU sequences within gene families (coefficient of determination = 0.22). Comparisons of SSU sequences within a gene family and known to be unlinked in a genome all fall above 0.4 SNS. Comparison of SSU sequences in a gene family which are known to be closely linked all fall below 0.25 SNS.

RNS for sequences with RNS values less than 0.19 is presented in Figure 7. Although there is some variation in the data for SSU sequence comparisons among different species within the monocots and dicots, the positive correlation of SNS as a function of RNS is apparent and the amount of variation decreases rapidly as RNS approaches zero. The coefficient of determination for these data is 0.55 confirming the linear relationship between SNS and RNS (MCCLAVE and DIETRICH 1982). However, for the 23 comparisons of SSU sequences within gene families the extreme variation in SNS values as a function of RNS is obvious. Within SSU gene families, SNS values as a function of RNS fall far outside the variation observed across species. The coefficient of determination is 0.22 demonstrating that the correlation between SNS and RNS accounts for only a small part of the variation within SSU sequences within a gene family of a single species. Comparisons of the three petunia SSU sequences show SNS values of 0.4, 0.65, and 0.8 which all fall above the bounds set by comparisons between species. It should be noted that, based on physical/genetic mapping studies, these three sequences are not closely linked. Two additional petunia sequences, SSU231 and SSU112, which are physically linked to Pet2 (DEAN *et al.* 1987) were not included in this study. When compared to Pet2, SSU231 and 112 have zero and one replacement nucleotide substitution, respectively, and low SNS values which fall below the bounds set by comparisons between species. This suggests that closely linked SSU genes follow different rules from unlinked SSU genes.

Comparison of the tomato SSU gene family members also suggest a discrepancy between SNS and RNS values. As determined by analysis of RNS all five tomato sequences are more related to each other than they are to petunia sequences having RNS values <0.025. The three genomic sequences analyzed, Tm1, Tm2* and Tm3, are not linked in the tomato genome (VALLEJOIS, TANKSLEY and BERNATZKEY 1986). The three pairwise comparisons show SNS values of 0.57–0.67, which all lie well above the bounds set by comparisons among different species (Figure 7). The only tomato comparisons giving SNS values below these bounds are comparisons of two additional SSU sequences, Rbcs3B and 3C (Table 1; SUGITA *et al.* 1987), which are closely linked to Tm3 in the genome. These sequences have no RNS when compared to Tm3. Again these data indicate that within the tomato gene family that unlinked SSU sequences show far more SNS than expected relative to the level of RNS, whereas, linked loci have very low SNS values. Although the data presented herein do not elucidate the mechanisms by which linked SSU sequences undergo concerted evolution these data suggest the possibility that intrachromosomal gene conversion events homogenizing SSU sequences occur at much higher levels than interchromosomal gene conversion events, as has been shown to be the case in yeast (JINKS-ROBERTSON and PETES 1986).

Another way to look at the discrepancy between the RNS and SNS values within a gene family is to calculate divergence time from a common ancestor using the nucleotide substitution data. The petunia and tomato sequences are significantly diverged from each other in RNS, so that there is little doubt that novel RNS has accumulated in each gene family since the divergence of each species (Figures 4 and 5A), even while homogenization has occurred within each gene family. The corrected RNS between the most distant of the three petunia SSU subfamily representatives examined in this study, Pet2 *vs.* Pet1 for example, is 0.0103 (1.03% RNS) and well within the limits for gene family members (Table 1, Figure 7). Assuming for the moment that gene conversion produced identity in all nucleotide positions, then based on an initial expected RNS rate of 1.25×10^{-9}/site/yr these data suggest that the last gene conversion event between these sequences occurred approximately 4 MY ago. However, using corrected SNS data for the divergence of these two sequences, 0.798 (80% SNS), and an initial expected SNS rate of 6.6×10^{-9}/site/yr for SSU sequences, it might be concluded that the last homogenization of these sequences occurred 110 MY ago. The discrepancy in RNS and SNS values among a gene family members is striking.

Mechanisms and selective forces: The similarity among SSU gene family members in RNS positions can best be explained by gene conversion events followed by purifying selection. Based on the overall conservation of replacement nucleotide positions in the mature SSU sequence across kingdoms, selection at the amino acid level is a strong force in SSU gene evolution. However, this does not explain how unlinked SSU gene family members obtain extreme divergence in the SNS positions while their RNS positions are homogeneous. One trivial explanation would be that the initial rate calculated to SNS accumulation is underestimated by nearly 15-fold. Although this is possible, it is unlikely since the rates of initial RNS and SNS accumulation were calculated from a large number of comparisons of divergence between different species (73 comparisons; Figures 3 and 6, B and D). It seems unlikely that the estimated rate of SNS accumulation could be in error by more than a factor of two.

There are several mechanisms which have been proposed to alter the mutation rates within sequences. For example, gene conversion between small regions of DNA related only in sequence and taken from nonrelated contexts or non-homologous regions of the genome (unlinked loci) could increase the rate at which mutation occurs (ORNSTON 1982; SHIMKETS, GILL and KAISER 1983; DOTEN, GREGG and ORNSTON 1987). It is likely that gene conversion can occur at reasonable frequency between homologous sequences without recombination (SYMINGTON and PETES 1988). In addition, many of the mechanisms of DNA turnover operating within the proposed process of molecular drive (DOVER 1986) are capable of either accelerating or retarding the rate at which mutations accumulate in multigene families and undoubtedly contribute to the evolution of SSU sequences. However, without strong selection neither gene conversion alone nor any of the mechanisms of DNA turnover can account for both an increase in the similarity in RNS positions and an increase in divergence in SNS positions among multigene family members.

Gene conversion followed by drift and selection of the organisms containing biased gene conversion products offers an alternative explanation. Within this set of mechanisms natural selection will act to conserve RNS positions due to the high degree of conservation of SSU polypeptide sequence. It must also allow homogenization and a high degree of divergence of SNS positions. ANDERSON and ROTH (1977) and ORNSTON and YEH (1981) suggest that when duplicated genes were required there would be a strong selection for sequence divergence. Loss of the duplicate copy occurs at much higher frequency than the original duplication event. However, sequence divergence between duplicated genes should slow homologous recombination and loss of the duplicate copy. Therefore, in the case of the three unlinked petunia

sequences and three unlinked tomato sequences an additional force should act to select for divergence in SNS positions. Recombination between homogeneous but unlinked SSU sequences could lead to chromosomal translocation and in most cases it should be lethal. Therefore, out of numerous SSU gene conversion products in a population of organisms, drift and selection could act differentially on RNS and SNS positions to select only those organisms with the appropriate gene conversion products, maintaining a conserved SSU encoding sequence and at the same time keeping substantial sequence divergence in the SNS positions among family members. The selective force at the replacement positions would maintain normal SSU protein function and a homogeneous set of polypeptides (see below). The selective force at the silent nucleotide positions would maintain sequence divergence in order to stabilize the gene family members at unlinked genetic loci against homologous recombination. Applying these selective mechanisms, gene conversion offers more than sequence homogenization to the concerted evolution of a family of SSU sequences (see below).

These data on the SSU multigene families suggest that the process leading to the concerted evolution of SSU sequences may be distinct from molecular drive as it is now modeled in a few important respects. At any one point in time a number of SSU sequences with varying degrees of divergence in the RNS and SNS positions are observed. Within the proposed process of molecular drive, most of the sequences in a gene family should be nearly identical within an individual organism. Furthermore, by the above model proposed for SSU sequence evolution all the observations associated with the concerted evolution of SSU sequences can be accounted for by the processes of purifying selection and neutral drift, whereas the proposed process of molecular drive operates nearly independently from selection (DOVER 1986).

The concerted evolution of several eukaryotic gene families are affected by gene conversion, including mammalian globins (SLIGHTOM, BLECHL and SMITHIES 1980; SCOTT *et al.* 1984), silk moth chorions (GOLDSMITH and KAFATOS 1984) and probably the major histocompatibility antigens (MELLOR *et al.* 1983; HUGHES and NEI 1988). Ubiquitin genes may undergo gene conversion but primarily within tandem sequence repeats and not between unlinked loci (SHARP and LI 1987). Although gene conversion events probably result from accidents of the DNA repair system, it seems reasonable to ask if gene conversion offers any benefit to the organism. Evidence suggests that gene conversion provides some of the polypeptide sequence diversity required of the MHC antigens. But more specifically, does the gene conversion and the concerted evolution of SSU sequences offer any benefit (PICHERSKY *et al.* 1986)? If so, then there may be a selective reason to maintain or even enhance gene conversion. The SSU polypeptide must be transported into the chloroplast, associate with the large subunit, become integrated into the ribulose-1,5-bisphosphate carboxylase holoenzyme and possibly interact with other machinery involved in carbon fixation and photorespiration. While the SSU polypeptide is always encoded by a multi-gene family in eukaryotes the large subunit is always encoded by a single gene sequence in the chloroplast genome. Because the LSU and SSU polypeptides are in intimate association (CHAPMAN *et al.* 1988), it is likely that they do not evolve independently. Changes in SSU amino acid sequence and structure must in some cases be countered by selection and fixation of complementary changes in the conserved LSU and vice versa. This is a possible explanation for the restriction beyond 20% divergence in RNS for SSU sequences. In other words, the slow long-term rate of RNS occurring during SSU evolution (Figure 6A) could be due to the requirement for compensating mutations in LSU and other proteins with which it interacts. Furthermore, if a heterogeneous population of SSU polypeptides were produced from divergent gene family members it would ultimately become impossible to maintain the integrity of the holoenzyme. Is it possible that specific sequences linked to the SSU gene family members enhance and ensure frequent sequence homogenization?

The majority of genes within most gene families do not appear to interact via gene conversion suggesting that only genes meeting unique requirements will undergo frequent gene conversion. For example, there has been an undetectable level of homogenization between divergent cytoplasmic and muscle actin genes or between the α-globins and β-globins for the last 300–500 MY, whereas within the β-globin cluster the γ_1- and γ_2-globin genes have undergone frequent gene conversion in primates (SCOTT *et al.* 1984) as have the loci for the histocompatability antigens (HUGHES and NEI 1988). Specific sequences flanking gene conversion tracts in primate γ-globin genes have been implicated in enhancing the homogenization of globin sequences (SLIGHTOM *et al.* 1985). Based on the data presented herein, homogenization of the SSU sequences seems to be an integral part of SSU evolution. Therefore, it seems likely that signals within or flanking SSU gene sequences are required to stimulate and ensure gene conversion between unlinked SSU sequences.

We wish to thank WYATT ANDERSON, KEN WOLFE, JONATHAN ARNOLD, NICK ORNSTON, LARRY SHIMKETS, JEFF PALMER and BRENDA SHIRLEY for helpful discussions about this work and the manuscript. MASATOSHI NEI and FELIX JIN generously analyzed the SSU sequences analyzed herein by an alternative unweighted pathway method. ALIX WEAVER kindly determined the UPGMA tree from our data. JIM LAKE provided the values for the divergence or

rRNA sequences between maize and anacystis. We thank these individuals for their significant and generous effort. This project was supported by U.S. Department of Energy grant DE-FG09-86ER13543.

LITERATURE CITED

ADAMS, C. A., M. BABCOCK, F. LEUNG and S. M. SUN, 1987 Sequence of a ribulose 1,5-bisphosphate carboxylase/oxygenase cDNA from the C_4 dicot *Flaveria trinervia*. Nucleic Acids Res. **15:** 1875.

ANDERSON, R. P., and J. J. ROTH, 1977 Tandem genetic duplications in phage and bacteria. Annu. Rev. Microbiol. **31:** 473–505.

ARNHEIM, N., M. KRYSTAL, R. SCHMICKEL, G. WILSON, D. RYDER and E. ZIMMER, 1980 Molecular evidence for genetics exchanges among ribosomal genes on non-homologous chromosomes in man and apes. Proc. Natl. Acad. Sci. USA **77:** 7323–7327.

BECK, C. B., 1976 *Origin and Early Evolution of Angiosperms*. New York: Columbia University Press.

BERRY-LOWE, S. L., T. D. MCKNIGHT, D. M. SHAH and R. B. MEAGHER, 1982 The nucleotide sequence, expression and evolution of one member of a multigene family encoding the small subunit in ribulose 1,5-bisphosphate carboxylase in soybean. J. Mol. Appl. Gen. **1:** 483–498.

BRINKMAN, H., P. MARTINEZ, F. QUIGLEY, W. MARTIN and R. CERFF, 1987 Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. J. Mol. Evol. **26:** 320–328.

BROGLIE, R., G. CORUZZI, G. LAMPPA, B. KEITH and N.-M. CHA, 1983 Structural analysis of nuclear genes coding for the precursor to the small subunit of wheat ribulose-1,5-bisphosphate carboxylase. Bio/Technology **1:** 55–60.

BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–239.

CASHMORE, A. R., 1983 Nuclear genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase, pp. 29–38 in *Genetic Engineering in Plants*, edited by T. KOSUGE, C. MEREDITH and A. HOLLAENDER. Plenum Press, New York.

CHAPMAN, M. S., S. WON-SUH, P. M. G. CURMI, D. CASICO, W. W. SMITH and D. S. EISENBERG, 1988 Tertiary structure of plant RuBisCO: Domains and their contacts. Science **241:** 71–74.

CHEN, K., and S. G. WILDMAN, 1980 Inheritance behavior of information coding for small subunit polypeptides of fraction 1 protein. Biochem. Genet. **18:** 1175–1184.

CHEN, K., S. D. KUNG, J. C. GRAY and S. G. WILDMAN, 1975 Polypeptide composition of fraction 1 protein from *Nicotiana glauca* and from cultivars of *Nicotiana tabacum* including a male sterile line. Biochem. Genetics **13:** 771–778.

CHUA, N., and G. W. SCHMIDT, 1978 Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose-1,5-bisphosphate carboxylase. Proc. Natl. Acad. Sci. USA **75:** 6110–6114.

CORUZZI, G., R. BROGLIE, C. EDWARDS and N.-H. CHA, 1984 Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. EMBO J. **3:** 1671–1679.

CREPET, W. L., and D. W. TAYLOR, 1985 The diversification of the *Leguminosae*: first fossil evidence of the *Mimosoideae* and *Papilionoideae*. Science **222:** 1087–1088.

CRONQUIST, A., 1981 *An Integrated System of Classification of Flowering Plants*. Columbia University Press, New York.

DEAN, C., P. VAN DEN ELZEN, S. TAMAKI, M. BLACK, P. DUNSMUIR and J. BEDBROOK, 1987 Molecular characterization of the rbcS multigene family of Petunia (Mitchell). Mol. Gen. Genet. **206:** 465–474.

DEROCHER, E. J., R. T. RAMAGE, C. B. MICHALOWSKI and H. J. BOHNERT, 1987 Nucleotide sequence of a cDNA encoding *rbcS* from the desert plant *Mesembryanthemum crystallinum*. Nucleic Acids Res. **15:** 6301.

DOTEN, R. C., L. A. GREGG and L. N. ORNSTON, 1987 Influence of *catBCE* sequence on the phenotypic reversion of a pcaE mutation in *Acinetobacter* classifications. J. Bacteriol. **169:** 3175–3180.

DOVER, C. A., 1986 Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Theor. Appl. Genet. **2:** 159–165.

DOYLE, J. A., and L. J. HICKEY, 1982 Pollen and leaves from the mid-Cretaceous Potomac group and their bearing on early angiosperm evolution, pp. 139–206 in *Origin and Early Evolution of Angiosperms*, edited by C. B. BECK. Columbia University Press, New York.

DOYLE, J. J., M. A. SCHULER, W. D. GODETTE, V. ZENGER, R. N. BEACHY and J. L. SLIGHTOM, 1986 The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. J. Biol. Chem. **261:** 9228–9238.

DUNSMUIR, P., S. SMITH and J. BEDBROOK, 1983 A number of different nuclear genes for the small subunit of RuBPCase are transcribed in petunia. Nucleic Acids Res. **11:** 4177–4183.

FLUHR, R., P. MOSES, G. MORELLI, G. CORUZZI and N.-H. CHUA, 1986 Expression dynamics of the pea rbcS multigene family and organ distribution of the transcripts. EMBO J. **5:** 2063–2071.

GATENBY, A. A., and E. C. COCKING, 1978 Polypeptide composition of the subunits of fraction 1 protein in the genus *Lycopersicon*. Plant Sci. Lett. **13:** 171–176.

GOJOBORI, T., 1983 Codon substitution in evolution and the "saturation" of synonymous changes. Genetics **105:** 1011–1027.

GOLDSCHMIDT-CLERMONT, M., and M. RAHIRE, 1986 Sequence, evolution and differential expression of the two genes encoding variant small subunits of ribulose bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. J. Mol. Biol. **191:** 421–432.

GOLDSMITH, M. R., and F. C. KAFATOS, 1984 Developmentally regulated genes in silkmoths. Annu. Rev. Genet. **18:** 443–487.

GRANDBASTIEN, M. A., S. BERRY-LOWE, B. W. SHIRLEY and R. B. MEAGHER, 1986 Two soybean ribulose-1,5-bisphosphate carboxylase small subunit genes share extensive homology even in distant flanking sequences. Plant Mol. Biol. **7:** 451–465.

GREENLAND, A. J., M. V. THOMAS and R. M. WALDEN, 1987 Expression of two nuclear genes encoding chloroplast proteins during early development of cucumber seedlings. Planta **170:** 99–110.

GUNDERSON, J. H., H. ELWOOD, A. INGOLD, K. KINDLE and M. L. SOGIN, 1987 Phylogenetic relationships between chlorophytes, chrysophytes and oomycetes. Proc. Natl. Acad. Sci. USA **84:** 5023–5827.

HAMBY, R. K., and E. A. ZIMMER, 1988 Ribosomal RNA sequences for infering phylogeny within the grass family *Porceae*. Plant Syst. Evol. **160:** 29–37.

HEYWOOD, V. H., 1978 *Flowering Plants of the World*. Mayflower Books Inc., New York.

HIGHTOWER, R. C., and R. B. MEAGHER, 1985 Divergence and differential expression of soybean actin genes. EMBO J. **4:** 1–8.

HIGHTOWER, R. C., and R. B. MEAGHER, 1986 The molecular evolution of actin. Genetics **114:** 315–332.

HUGHES, A. L., and M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

JANSEN, R. K., and J. D. PALMER, 1987 A chloroplast DNA inversion masks an ancient evolutionary split in the sunflower

family (Asteridaceae). Proc. Natl. Acad. Sci. USA **84:** 5818-5822.

JINKS-ROBERTSON, S., and T. D. PETES, 1986 Chromosomal translocations generated by high-frequency meiotic recombination between repeated yeast genes. Genetics **114:** 731–752.

JONES, S. B., JR., and A. E. LUCHSINGER, 1986 *Plant Systematics.* McGraw-Hill, New York.

JUDD, S. R., and T. D. PETES, 1988 Physical lengths of meiotic and mitotic gene conversion tracts in *Saccharomyces cerevisiae.* Genetics **118:** 401–410.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism,* edited by H. N. MONRO. Academic Press, New York.

KAWASHIMIA, N., and S. G. WILDMAN, 1972 Studies on fraction 1 protein. IV. Mode of inheritance of primary structure in relation to whether chloroplast or nuclear DNA contains the code for a chloroplast protein. Biochim. Biophys. Acta **262:** 42–49.

LAKE, J. A., 1987 A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. Mol. Biol. Evol. **4:** 167–191.

LEBRUN, M., G. WAKSMAN and G. FREYSSINET, 1987 Nucleotide sequence of a gene encoding corn ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcS). Nucleic Acids Res. **15:** 4360.

MARGULIS, L., and V. SCHWARTZ, 1982 *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth.* W. H. Freeman, San Francisco.

MARTIN, P. G., and J. M. DOWD, 1988 A molecular evolutionary clock for angiosperms. Taxon **37:** 364–377.

MATSOUKA, M., Y. KANO-MURAKAMI, Y. TANAKA, Y. OZEKI and N. YAMAMOTO, 1987 Nucleotide sequence of cDNA encoding the small subunit of ribulose-1,5-bisphosphate carboxylase from maize. J. Biochem. **102:** 673–676.

MAZUR, B. J., and C.-F. CHUI, 1985 Sequence of a genomic DNA clone for the small subunit of ribulose biphosphate carboxylase oxygenase from tobacco. Nucleic Acids Res., **13:** 2373–2386.

MCCLAVE, J. T., and I. H. DIETRICH III, 1982 *Statistics.* Dellen Publishing Company, San Francisco.

MCKNIGHT, T. D., D. C. ALEXANDER, M. S. BABCOCK and R. B. SIMPSON, 1986 Nucleotide sequence and molecular evolution of two tomato genes encoding the small sununit of ribulose-1,5-bisphosphate carboxylase. Gene **48:** 23–32.

MISHKIND, M. L., S. R. WESSLER and G. W. SCHMIDT, 1985 Functional determinants in transit sequences import and partial maturation by vascular plant chloroplasts of the ribulose-1,5-bisphosphate carboxylase small subunit of Chlamydomonas. J. Cell Biol. **100:** 226–234.

MELLOR, A. L., E. H. WEISS, RAMACHANDRANK and R. A. FLAVELL, 1983 A potential donor gene for the bond gene conversion event in the C57BL mouse. Nature **306:** 792–795.

MIYATA, T., and H. HAYASHIDA, 1982 Recent divergence from a common ancestor of human IFN-alpha gene. Nature **295:** 165–168.

MIYATA, T., H. HAYASHIDA, R. KIKUNO, M. HASEGAWA, M. KOBAYSHI and K. KOIKE, 1982 Molecular clock of silent substitution: At least six fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. **19:** 28–35.

MIZIORKO, H. M., and G. H. LORIMER, 1983 Ribulose-1,5-bisphosphate carboxylase-oxygenase. Annu. Rev. Biochem. **52:** 507–535.

NAGYLAKI, T., and T. D. PETES, 1982 Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics **100:** 315–337.

NAGYLAKI, T., 1988 Gene conversion, linkage, and the evolution of multigene families. Genetics **120:** 291–301.

NEI, M., 1987 *Molecular Evolutionary Genetics,* pp. 287–326. Columbia University Press, New York.

NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

NIERZWICKI-BAUER, S. A., S. E. CURTIS and R. HASELKORN, 1984 Cotranscription of genes encoding the small and large subunits of ribulose-1,5-bisphosphate carboxylase in the cyanobacterium *Anabaena* 7120. Proc. Natl. Acad. Sci. USA **81:** 5961–5965.

OCHMAN, H., and A. C. WILSON, 1987 Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26:** 74–86.

O'NEAL, J. K., A. R. POKALSKY, K. L. KIEHNE and C. K. SHEWMAKER, 1987 Isolation of tobacco SSU genes: characterization of a transcriptionally active pseudogene. Nucleic Acids. Res. **15:** 8661–8677.

ORNSTON, L. N., 1982 Communication among coevolving genes, pp. 121–127 in *Experiences in Biochemical Perception.* Academic Press, New York.

ORNSTON, L. N., E. L. NEIDLE and J. E. HOUGHTON, 1988 Gene rearrangements, a force for evolutionary change; DNA sequence arrangements, a source of genetic consistency, in *The Bacterial Chromosome: Structure and Functional Organization,* edited by M. RILEY and K. ORLICA. Justin W. Wiley and Sons (in press).

ORNSTON, L. N., and W.-K. YEH, 1981 Toward molecular natural history, pp. 140–145 in *Microbiology.* American Society for Microbiology, Washington, D.C.

PALMER, J. D., R. K. JANSEN, H. J. MICHAELS, M. W. CHASE and J. R. MANHART, 1988 Chloroplast DNA variation and plant phylogeny. Annu. Mo. Bot. Gard. **75:** 1180–1206.

PERLER, F., A. EFSTRATIADIS, P. LOMEDICO, W. GILBERT, R. KOLODNER and J. DODGSON, 1980 The evolution of genes: The chicken preproinsulin gene. Cell **20:** 555–566.

PICHERSKY, E., R. BERNATZKY, S. D. TANKSLEY and A. R. CASHMORE, 1986 Evidence for selection as a mechanism in the concerted evolution of *Lycopersicon esculentum* (tomato) genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc. Natl. Acad. Sci. USA **83:** 3880–3884.

PINCK, M., E. GUILLEY, A. DURR, M. HOFF, L. PINCK and J. FLECK, 1984 Complete sequence of one of the mRNAs coding for the small subunit of ribulose bisphosphate carboxylase of *Nicotiana sylvestris.* Biochemie **66:** 539–545.

RICE, K. A., 1987 Estimating nucleotide sequence divergence from polymorphism in nucleotide sequences and restriction maps. Technical report for Master of Applied Mathematical Sciences, Department of Computer Science, University of Georgia.

SAITOU, N., and M. NEI, 1987 The Neighbor Joining Method: A new method of reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SCOTT, A. F., P. HEATH, S. TRUSKO and S. H. BOYER, 1984 The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution. Mol. Biol. Evol. **1:** 371–389.

SHAH, D. M., R. C. HIGHTOWER and R. B. MEAGHER, 1983 Gene encoding actin in higher plants: intron positions are highly conserved but the coding sequences are not. J. Mol. Appl. Genet. **2:** 111–126.

SHARP, P. M., and W.-H. LI, 1987 Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J. Mol. Evol. **25:** 58–64.

SHIMKETS, L. J., R. E. GILL and D. KAISER, 1983 Developmental cell interactions in *Myxococcus xanthus* and the spoC locus. Proc. Natl. Acad. Sci. USA **80:** 1400–1410.

SHINOZAKI, K., and M. SUGIURA, 1983 The gene for the small

subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase is located close to the gene for the large subunit in the cyanobacterium *Anacystis nidulans* 6301. Nucleic Acids Res. **11:** 6957–6964.

SLIGHTOM, J. L., A. E. BLECHL and O. SMITHIES, 1980 Human fetal G*- and A*-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21:** 627–638.

SLIGHTOM, J. L., L.-Y. E. CHANG, B. F. KOOP and M. GOODMAN, 1985 Chimpanzee fetal $^{G}\gamma$- and $^{A}\gamma$-globin gene nucleotide sequences provide further evidence of gene conversions in hominine evolution. Mol. Biol. Evol. **2:** 370–389.

SMITH, S. M., and R. J. ELLIS, 1979 Processing of small subunit precursor of ribulose bisphosphate carboxylase and its assembly into whole enzyme are stromal events. Nature **278:** 662–664.

SMITH, T. F., and WATERMAN, 1981 Identification of common molecular subsequences. J. Mol. Biol. **147:** 195–197.

SOGIN, M. L., J. H. GUNDERSON, H. J. ELWOOD, R. A. ALONSO and D. A. PEATTIE, 1987 Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. Science **243:** 75–77.

SOKAL, R. R., and P. H. A. SNEATH, 1963 *Principles of Numerical Taxonomy*. W. H. Freeman, San Francisco.

STARNES, S. M., D. H. LAMBERT, E. S. MAXWELL, S. E. STEPHENS, JR., R. D. PORTER and J. M. SHIVELY, 1985 Cotranscription of the large and small subunit genes of ribulose-1,5-bisphosphate carboxylase/oxegenase in *Gyanophora paradoxa*. FEMS Microbiol. Lett. **28:** 165–169.

STIEKEMA, W. J., C. F. WIMPEE and E. M. TOBIN, 1983 Nucleotide sequence encoding the precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase from *Lemna gibba* L.G-3. Nucleic Acids Res. **11:** 8051–8061.

SUEOKA, N., 1961 Variation and heterogeneity of base composition of DNA: A compilation of old and new data. J. Mol. Biol. **3:** 31–40.

SUGITA, M., T. MANZARA, E. PICHERSKY, A. CASHMORE and W. GRUISSEM, 1987 Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. Mol. Gen. Genet. **209:** 247–256.

SYMINGTON, L. S., and T. D. PETES, 1988 Expansions and contractions of the genetic map relative to the physical map of the yeast chromosome III. Mol. Cell. Biol. **8:** 595–604.

TANKSLEY, S. D., and E. PICHERSKY, 1988 Organization and evolution of sequences in the plant nuclear genome, pp. 55–84 in *Plant Evolutionary Biology*, edited by L. D. GOTTLIEB and S. K. JAIN. Chapman & Hall, New York.

TAYLOR, T. N., 1981 *Paleobotany: An Introduction to Fossil Plant Biology*. McGraw-Hill, New York.

TUMER, N. E., W. G. CLARK, G. J. TABOR, M. HIRONAKA, R. T. FRALEY and D. M. SHA, 1986 The genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase are expressed differentially in petunia leaves. Nucleic Acids Res. **14:** 3325–3342.

VALLEJOS, C. E., S. D. TANKSLEY and R. BERNATZKEY, 1986 Localization in the tomato genome of DNA restriction fragments containing the sequences homologous to the rRNA (42S), the major chlorophyll A/13 binding polypeptide and the ribulose bisphosphate carboxylase genes. Genetics **112:** 93–105.

WAKSMAN, G., M. LEBRUN and G. FREYSSINET, 1987 Nucleotide sequence of a gene encoding sunflower ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (rbcS). Nucleic Acids Res. **15:** 7181.

WILSON, A. C., S. S. CARLSON and T. J. WHITE, 1977 Biochemical evolution. Annu. Rev. Biochem. **46:** 573–639.

WIMPEE, C. F., 1984 Ph.D. thesis, Organization and expression of light-regulated genes in *Lemna gibba* L. G-3 (phytochrome, RuBP, SSU). University of California, Los Angeles.

WOLFE, K. H., W.-H. LI and P. M. SHARP, 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84:** 9054–9058.

WOLFE, K. H., P. M. SHARP and W.-H. LI, 1989 Rates of synonymous substitution in plant nuclear genes. J. Mol. Evol. **29:** 208–211.

WOLTER, F. P., C. C. FRITZ, L. WILLMITZER, J. SCHELL and P. H. SCHREIER, 1988 *rbcS* genes in *Solanum tuberosum*: conservation of transit peptide and exon shuffling during evolution. Proc. Natl. Acad. Sci. USA **85:** 846–850.

YOSHINAGA, K., T. OHTA, Y. SUZUKI and M. SUGIURA, 1988 Chlorella chloroplast DNA sequence containing a gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and a part of a possible gene for β' subunit of RNA polymerase. Plant Mol. Biol. **10:** 245–250.

ZIMMER, E. A., R. K. HAMBY, M. L. ARNOLD, D. A. LEBLANC and E. C. THERIOT, 1989 Ribosomal RNA phylogenics and flowering plant evolution, pp. 205–214 in *The Hierarchy of Life: Proceedings of the 70th Nobel Symposium*, edited by B. FERNHOLM, K. BREMMER and H. JORNVALL.

COMMUNICATING EDITOR: M. R. HANSON