Nucleotide sequence and evolution of the 18S ribosomal RNA gene in maize mitochondria

Shiaoman Chao[1], Ronald Sederoff and Charles S.Levings III

Department of Genetics, North Carolina State University, Raleigh, NC 27695, USA

Received 9 January 1984; Revised and Accepted 20 July 1984

ABSTRACT

The nucleotide sequence of the gene coding for the 18S ribosomal RNA of maize mitochondria has been determined and a model for the secondary structure is proposed. Dot matrix analysis has been used to compare the extent and distribution of sequence similarities of the entire maize mitochondrial 18S rRNA sequence with that of 15 other small subunit rRNA sequences. The mitochondrial gene shows great similarity to the eubacterial sequences and to the maize chloroplast, and less similarity to mitochondrial rRNA genes in animals and fungi. We propose that this similarity is due to a slow rate of nucleotide divergence in plant mtDNA compared to the mtDNA of animals. Sequence comparisons indicate that the evolution of the maize mitochondrial 18S, chloroplast 16S and nuclear 17S ribosomal genes have been essentially independent, in spite of evidence for DNA transfer between organelles and the nucleus.

INTRODUCTION

Ribosomal RNAs (rRNAs) and the DNA sequences that code them, have been studied actively by molecular biologists to learn about the regulation and functions of rRNA genes. These RNAs are essential components of the translational machinery and their functions have been highly conserved over at least a billion years. Comparisons of ribosomal RNA sequences from diverse organisms provide insight into ancient evolutionary relationships. In addition, comparative sequence analysis is a powerful approach toward understanding rRNA structure and mechanisms of the translation process.

The rRNAs of plant mitochondria are of particular interest because they are substantially different from those of animal and fungal mitochondria. The rRNAs of animal mitochondria are 16S and 12S, while those of yeast are 21S and 15S (1-3). In contrast, the corresponding rRNAs of plant mitochondria are larger, 26S and 18S. Plant mitochondrial ribosomes contain a 5S RNA which is not present in any other mitochondria. Moreover, the organization of the ribosomal genes in plant mitochondria is unique. The 5S and 18S genes are closely linked, while the 26S gene is located at a greater distance (4-6).

This organization is also different from that of eubacteria and

chloroplasts which have clusters of ribosomal genes in the order 16S-23S-5S (7,8) or from eukaryotic nuclei where the 5S genes are organized in separate clusters of tandem repeats (9). In this report, we present the sequence of the 18S rRNA gene from maize mitochondria and propose a secondary structure for the RNA molecule. In addition, we have compared the 18S maize sequence with other small subunit rRNA genes to explore evolutionary relationships.

MATERIALS AND METHODS

1. Origin of cloned fragments

The 18S ribosomal RNA gene examined in this study was derived from normal maize cytoplasm (Zea mays L.). The original BamHI clone of the 18S gene (ZmmtN542) has been described previously (10,11). The gene was identified by hybridization with 18S rRNA purified on agarose gels under denaturing conditions. Methods of purification, labelling and hybridization have been described (11).

2. Restriction analysis

The restriction map was constructed by single and double digests. The subclone 542-2 contains a large internal PstI fragment of 542 and covers most of the 18S gene. Length estimation for restriction mapping used the function of Southern (12).

3. Sequencing strategy

The 542 clone or subclone 542-2 was digested with TaqI, Sau3A or HaeIII and cloned into M13 using vectors mp7, mp8 or mp9 (13). DNA sequencing was done using the dideoxy chain termination method (14,15). Methods of electrophoresis and autoradiography were previously described (11). Overlapping clones were detected from nucleotide sequence information. The complete sequence for both strands has been determined.

4. Dot matrix analysis

Sequences of the maize 18S rRNA gene and other small subunit rRNAs were compared on a dot matrix computer program provided by M. Edgell (UNC Chapel Hill). The program allows variation in the length of similarity of sequence needed to register a line, and variation in the percent of similarity acceptable for the defined length. We selected a criterion of 20 for length and a percent match of 75 (15/20).

The diagonal lines from the dot matrix were used to align sequences for detailed comparisons and as an aid in the analysis of secondary structure. To estimate the similarity of two complete sequences, we used the alignment length of the dot matrix by summing the lines along the diagonal, and correcting for overlapping noncontinuous lines. The alignment length was divided by the average length of both sequences being compared. The percent similarity was used to establish a similarity matrix.

5. Secondary structure analysis

The secondary structure model was constructed by comparison of aligned sequences. Positions of alignment between Escherichia coli and maize mitochondria from the dot matrix analysis were used to determine starting positions in comparing potential structural features. The criterion for establishment of base paired regions (helical components) by comparative sequence analysis has been described by Woese et al. (16). Woese et al. (16) consider a double helical region "proven" if it is formed in two different small subunit rRNAs and a covariance can be demonstrated in at least two base pairs in the helix; i.e. the covariance requires that a canonical Watson-Crick pair in one species can be replaced by a different

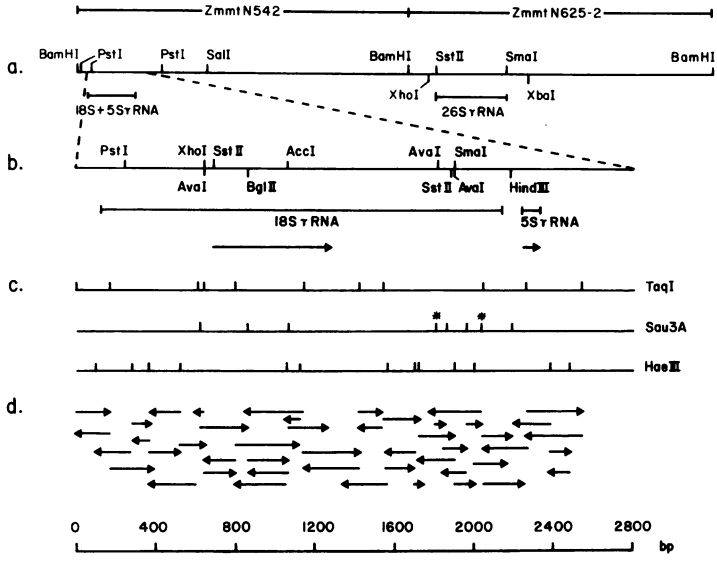

Fig. 1. Restriction maps and sequencing strategy
(a) Restriction maps of clones 542 (16.6 kb) and 625-2 (14.7 kb). Restriction sites SstII and SmaI on clone 625-2 are based on the results of Stern et al. (6). The two BamHI clones are adjacent on the map according to the results of Iams and Sinclair (5). The regions coding for rRNA genes are indicated.
(b) Restriction map of the segment containing the complete 18S rRNA gene, 5S rRNA gene and intergenic region. Arrows indicate the transcription directions of the 18S and 5S rRNA genes.
(c) Detailed restriction sites on the fragment described in (b). The sites where Sau 3A has cut at GATT or GGTC sequences is denoted by *.
(d) Sequencing strategy of the fragment described in (b). Sequences were read from the cloning site to the positions indicated by arrows. The complete sequence was determined for both strands. The position of adjacent clones has been determined from overlapping sequences obtained from clones cut with a different enzyme.

```
MAIZE MT   ATCATAGTC. AAAATCTGAG TTTGATCCTG GCTCAGAAGG AACGCTAGCT ATATGCTTAA CACATGCAAG  69
E. COLI    .........A AATTGAAGAG TTTGATCATG GCTCAGATTG AACGCTGGCG GCAGGCCTAA CACATGCAAG  61

MAIZE MT   TCGAACGTTG TTTTCGGGGA GCTGGGCAGA AGGAAAAGAG GCTCCTAGCT AAAGTTGTCT CGCCCTGCCT 139
E. COLI    TCGAACGGTA ACAGGAAGAA GCT..... .......... .......... .......... .......... 84

MAIZE MT   GCAGGGCGCG CGCAACGGCT TTGACCTAAC GGCCTCCGTT TGCTGGAATC GGAATAGTTG AGAACAAAGT 209
E. COLI    .......... .......... .......... .......... .......... ..TGCTCTTT GCTGACGAGT 102

MAIZE MT   GGCGAACGGG TGCGTAACGC GTGGGAATCT GCCGAACAGT TCGGGCCAA. .......... .........AT 260
E. COLI    GGCGGACGGG TGAGTAATGT CTGGGAAACT GCCTGATGGA GGGGGATAAC TACTGGAAAC GGTAGCTAAT 172

MAIZE MT   CCTGAAGAAA GCTCAAAAGC .......... .......... ......GCTG TTTGATGAGC CTGCGTAGTA 304
E. COLI    ACCGCATAAC GTCGCAAGAC CAAAGAGGGG GACCTTCGGG CCTCTTGCCA TCGGATGTGC CCAGATGGGA 242

MAIZE MT   TTAGGTAGTT GGTCAGGTAA AGGCTGACCA AGCCAATGAT GCTTAGCTGG TCTTTTCGGA TGATCAGCCA 374
E. COLI    TTAGCTAGTA GGTGGGGTAA CGGCTCACCT AGGCGACGAT CCCTAGCTGG TCTGAGAGGA TGACCAGCCA 312

MAIZE MT   CACTGGGACT GAGACACGGC CCGGACTCCC ACGGGGGGCA GCAGTGGGGA ATCTTGGACA ATGGGCGAAA 444
E. COLI    CACTGGAACT GAGACACGGT CCAGACTCCT ACGGGAGGCA GCAGTGGGGA ATATTGCACA ATGGGCGCAA 382

MAIZE MT   GCCCGATCCA GCAATATCGC GTGAGTGAAG AAGGGCAATG CCGCTTGTAA AGCTCTTTCG TCGAGTGCGC 514
E. COLI    GCCTGATGCA GCCATGCCGC GTGTATGAAG AAGGCCT..T CGGGTTGTAA AGTACTTTCA GCGGGGAGGA 450

MAIZE MT   GATCA..... .......... ........TG ACAGGACTCG AGGAAGAAGC CCCGGCTAAC TCCGTGCCAG 561
E. COLI    AGGGAGTAAA GTTAATACCT TTGCTCATTG ACGTTACCCG CAGAAGAAGC ACCGGCTAAC TCCGTGCCAG 520

MAIZE MT   CAGCCGCGGT AAGACGGGGG GGGCAAGTGT TCTTCGGAAT GACTGGGCGT AAAGGGCACG TAGGCGGTGA 631
E. COLI    CAGCCGCGGT AATACGGAGG GTGCAAGCGT TAATCGGAAT TACTGGGCGT AAAGCGCACG CAGGCGGTTT 590

MAIZE MT   ATCGGGTTGA AAGTGAAAGT CGCCAA..AA AGTGGCGGAA TGCTCTCGAA ACCAATTCAC TTGAGTGAGA 699
E. COLI    GTTAAGTCAG ATGTGAAATC CCCGGGCTCA ACCTGGGAAC TGCATCTGAT ACTGGCAAGC TTGAGTCTCG 660

MAIZE MT   CAGAGGAGAG TGGAATTTCG TGTGTAGGGG TGAAATCCGT AGATCTACGA AGGAACGCCA AAAGCGAAGG 769
E. COLI    TAGAGGGGGG TAGAATTCCA GGTGTAGCGG TGAAATGCGT AGAGATCTGG AGGAATACCG GTGGCGAAGG 730

MAIZE MT   CAGCTCTCTG GGTCCCTACC GACGCTGGGG TGCGAAAGCA TGGGGAGCGA ACAGGATTAG ATACCCTGGT 839
E. COLI    CGGCCCCCTG GACGAAGACT GACGCTCAGG TGCGAAAGCG TGGGGAGCAA ACAGGATTAG ATACCCTGGT 800

MAIZE MT   AGTCCATGCC GTAAACGATG AGTGTTCGCC CTTGGTCTGT CTACGCTACG CTACGCGGAT CAGGGGCCCA 909
E. COLI    AGTCCACGCC GTAAACGATG TCGACTTGGA GGTTGTGCCC T......... ..TGAGGCGT GGCTTCCGGA 859

MAIZE MT   GCTAACGCGT GAAACACTCC GCCTGGGGAG TACGGTCGCA AGACCGAAAC TCAAAGGAAT TGACGGGGGC 979
E. COLI    GCTAACGCGT TAAGTCGACC GCCTGGGGAG TACGGCCGCA AGGTTAAAAC TCAAATGAAT TGACGGGGGC 929

MAIZE MT   CTGCACAAGC GGTGGAGCAT GTGGTTTAAT TCGATACAAC GCGCAAAACC TTACCAGCCC TTGACATATG 1049
E. COLI    CCGCACAAGC GGTGGAGCAT GTGGTTTAAT TCGATGCAAC GCGAAGAACC TTACCTGGTC TTGACATCCA 999

MAIZE MT   AACAACAAAA CCTGTCTTT. AACGGGATGG TACTTACTTT CATACAGGTG CTGCATGGCT GTCGTCAGCT 1118
E. COLI    CGGAAGTTTT CAGAGATGAG AATGTGCCTT CGGGAACCGT GAGACAGGTG CTGCATGGCT GTCGTCAGCT 1069

MAIZE MT   CGTGTCGTGA GATGTTTGGT CAAGTCCTAT AACGAGCGAA ACCCTCGTTT TGTGTTGCTG AGACATGCGC 1188
E. COLI    CGTGTTGTGA AATGTTGGGT TAAGTCCCGC AACGAGCGCA ACCCTTATCC TTTGTTGCCA GC........ 1131

MAIZE MT   CTAAGGAGAA ATAGCCGAGG AGCCGAGTGA CGTGCCAGCG CTACTACTTG ATTGAGTGCC AGCACGTAGC 1258
E. COLI    .......... .......... .......... .......... .......... .......... .......... 1131

MAIZE MT   TGTGCTTTCC GCAAGAATTT CACCATTGGG AGCCTCGAAG CACTTTCACG TGTGAACCGA AGTCGTCTTG 1328
E. COLI    .......... .......... .......... .......... .......... .......... .......... 1131

MAIZE MT   CCGAACTCAA GACCCACGGA GACCTACCTA TAGTGACGTC AAAGTACCAG TGAGCATGGA GGTTTGGTTA 1398
E. COLI    .......... .......... .......... .......... .......... .......... .......... 1131

MAIZE MT   GGCTTGGTTA CGACGACGTC GAGTTGGCGG CGGAGGAAGA CTCGGCATGA AGGCCAGCCG CCCGGTGGTG 1468
E. COLI    .......... .......... .......... .......... .......... .......... .......... 1131

MAIZE MT   TGGTACGTAG TGGTAATAGT ACGCGCCCCG CTCCGAAACA AAGAAAAAGG TGCGTGCCGC ..ACTCACGA 1536
E. COLI    .......... .......... .......... .......... ........GG TCCG.GCCGG GAACTCAAAG 1152

MAIZE MT   GGGACTGCCA GTGAGATACT GGAGGAAGGT GGGGATGACG TCAAGTCCGC ATGGCCCTTA TGGGCTGGGC 1606
E. COLI    GAGACTGCCA GTGATAAACT GGAGGAAGGT GGGGATGACG TCAAGTCATC ATGGCCCTTA CGACCAGGGC 1222
```

```
MAIZE MT  CACACACGTG CTACAATGGC AATGACAATG GGAAGCAAGG CTGTAAGGCG GAGCG.AATC CGGAAAGATT  1675
E. COLI   TACACACGTG CTACAATGGC GCATACAAAG AGAAGCGACC TCGCGAGAGC AAGCGGACCT CATAAAGTGC  1292

MAIZE MT  GCCTCAGTTC GGATTGTTCT CTGCAACTCG GGAACATGAA GTAGAAATCG CTAGTAATCG CGGATCAGCA  1745
E. COLI   GTCGTAGTCC GGATTGGAGT CTGCAACTCG ACTCCATGAA GTCGGAATCG CTAGTAATCG TGGATCAGAA  1362

MAIZE MT  TGCCGCGGTG AATATGTACC CGGGCCCTGT ACACACCGCC CGTCACACCC TGGGAATTGG TTTCGCCCGA  1815
E. COLI   TGCCACGGTG AATACGTTCC CGGGCCTTGT ACACACCGCC CGTCACACCA TGGGAGTGGG TTGCAAAAGA  1432

MAIZE MT  AGCATCGGAC CAATGATCAC CCATGACTTC TGTGTACCAC TAGTGCCACA AAGGCCTTTG GTGGTCTTAT  1885
E. COLI   AGTAGGTAGC TTAA..... ..........  .......... .......... ....CCTTCG GGAG.....  1456

MAIZE MT  TGGCGCATAC CACGGTGGGG TCTTCGACTG GGGTGAAGTC GTAACAAGGT AGCCGTAGGG GAACCTGTGG  1955
E. COLI   .GGCGCTTAC CACTTTGTGA TTCATGACTG GGGTGAAGTC GTAACAAGGT AACCGTAGGG GAACCTGCGG  1525

MAIZE MT  CTGGATTGAA TCC...                                                           1968
E. COLI   TTGGATCACC TCCTTA                                                           1541
```

Fig. 2. Complete nucleotide sequence of maize mt 18S rRNA gene. Both ends were located by using the sequence determined for wheat mt 18S rRNA gene (20,21). The alignment presented was based on both dot matrix analysis (Fig. 4) and comparison of secondary structure (Fig. 3).

pair in the other species. Where the maize mitochondrial sequence was similar in sequence and/or formed the same structure in the correct position as that of a "proven" helix according to Woese's criteria, we considered that component of the maize sequence established for our model. To designate a specific base paired region, the number of the first and last nucleotides are given for both strands, e.g. 25-28/964-967 describes the helical region formed by bases 25-28 and 964-967 in the maize mt 18S sequence.

RESULTS

1. Restriction map of the region coding mt rRNAs.

In maize mtDNA, restriction mapping and hybridization analysis of cloned fragments have shown that the rRNA genes are located 16 kb apart (5,6) (Fig. 1a). The 18S and 26S rRNA genes of the maize mitochondrial genome are present only once in contrast with wheat which has several copies (17). The 5S mt rRNA gene is closely linked to the 18S rRNA gene in several higher plants (18). In maize mtDNA, the distance between the 18S and 5S rRNA genes is about 108 bp and both genes are transcribed in the same direction (11). We have carried out a detailed restriction analysis of the 18S and 5S region in connection with nucleotide sequencing (Fig. 1b). Two BamHI clones containing the rRNA genes have been isolated: the 26S rRNA gene is contained in a 14.7 kb fragment (N625-2), and the 18S and 5S genes are contained in a 16.6 kb fragment (N542) (Fig. 1a).

2. Sequence analysis of the 18S rRNA gene

Using the plasmid clone N542 and subcloned fragments in M13, we have determined the nucleotide sequence of the 18S rRNA gene. The

dideoxynucleotide sequencing procedures of Sanger (14,15) were used with overlapping Taq 1, Sau 3A and Hae III cloned fragments (Fig. 1c and 1d). The entire sequence is 1968bp (Fig. 2). The sequence was obtained for both strands, and the strand equivalent to the RNA transcript is presented. The G + C content of the rRNA gene is 54%, while the average G + C content of the mitochondrial genome is 47% (19). The location of the ends of the gene have been presented based on comparison with the rRNA sequences determined for the ends of wheat mitochondrial 18S rRNA (20,21). Consequently, the location of the ends of the maize mitochondrial 18S RNA sequence is provisional. The sequence of the mitochondrial 5S rRNA gene and its flanking regions has been published separately (11).

3. Secondary structure of 18s rRNA from maize mitochondria.

Comparison of the secondary structure of ribosomal RNAs is a useful method for investigation of functional sites in ribosomes and the evolution of the translational machinery. Comparative analysis of primary structure is a reliable and established approach for determination of RNA secondary structure (16,22). Such methods have been used to propose and refine the secondary structure of many ribosomal genes, particularly for the 5S and the 16S rRNAs.

Three different secondary structures have been proposed for the 16S rRNA of E. coli, and all agree to a first approximation. We have used the model of Woese et al. (16) with minor modifications to establish a secondary structure model for the maize mitochondrial 18S sequence (Fig. 3). We have chosen E. coli as a model because of the high degree of sequence similarity between the sequences, and because the E. coli secondary structure was established by comparative sequence data and by a variety of protection or accessibility experiments (16,23).

The secondary structure, which we propose, has a great deal of similarity with that of E. coli due to apparent conservation of sequence, and compensatory changes in regions of sequence variation (Fig. 4). For example, region 628-689 in maize mitochondrial 18S has the same structure as E. coli 588-651, but is highly divergent in sequence; 37 out of 62 nucleotides of the comparable segment of the maize sequence are different from E. coli. Similarly, region 821-879 in E. coli corresponds to 859-928 in maize and has the same structure. The loop at the end of the stem is 5 nucleotides in E. coli and 15 nucleotides in maize mitochondria, but both have similar stems, 20 (E. coli) and 21 (maize), with a bulge loop of 14 nucleotides in both cases. Only 12 of 40 nucleotides (30%) in the stem are identical. Most nucleotides have compensating changes which preserve the structure but not the sequence. The bulge loop is more conserved with 11 of 14 nucleotides being identical in maize and E. coli.

The secondary structure we propose contains most of the major features

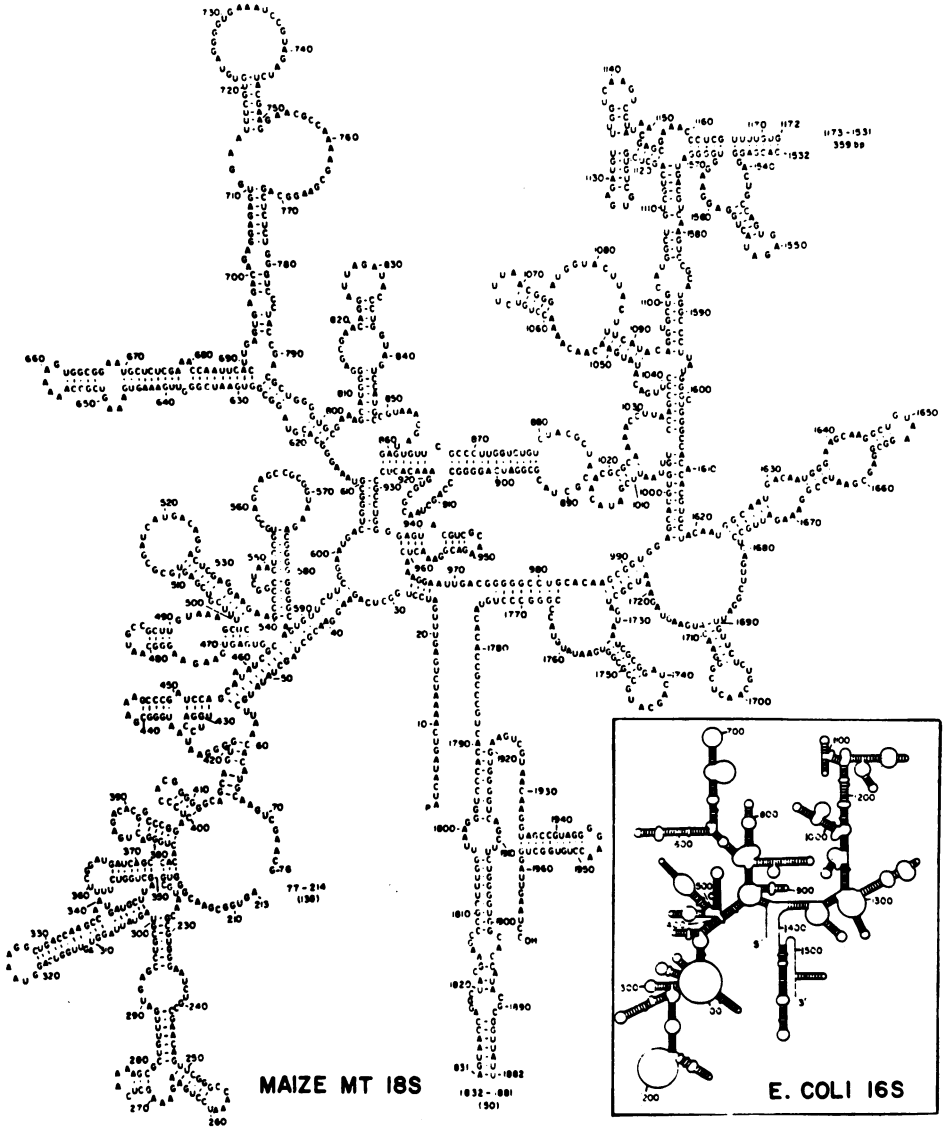

Fig. 3. A secondary structure model for maize mt 18S rRNA. The schematic structure for E. coli 16S rRNA is taken from Woese et al. (16).

of the E. coli model, with a major extension for a 359 bp insert. Both models have a major 5' domain, a central domain and a 3' domain containing a major and minor element. The 359 bp insert occurs in the 3' major domain. Since insufficient sequence information is available for

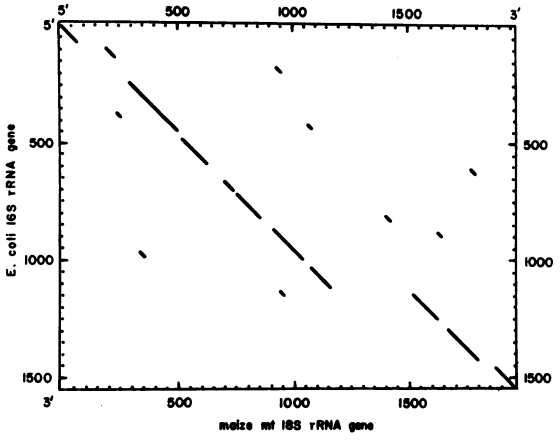

Fig. 4. Dot matrix analysis for sequence comparison between maize mt 18S rRNA gene and E. coli 16S rRNA gene. Diagonal lines indicate regions 20 bp or longer, having at least 75% sequence similarity.

comparison, no structure is proposed for the large insert 1173-1531. Similarly, insufficient comparative data are available to propose a structure for 77-214 and 1832-1881 in the maize mitochondrial 18S rRNA sequence.

The secondary structure model of E. coli 16S (16) has 74 helical regions of which 62 are "proven" by comparative analyses. All but 3 of these helical regions are common to the maize structure. In the E. coli model these regions are located at 153-158/163-168, 455-462/470-477, and 1132-1135/1139-1142.

4. Sequence comparison of mitochondrial 18S rRNA with E. coli 16S rRNA

Previous studies have shown that the T_1 oligonucleotides of wheat mitochondrial 18S rRNA and the 3' terminal sequence have strong similarities with E. coli 16S rRNA (20,24). The sequence of the 18S rRNA gene of maize mitochondria (1968 nucleotides) is significantly larger than that of E. coli (1541 nucleotides). Comparison of these sequences is difficult because of the interrupted distribution of similar regions. In order to align these sequences we have used a dot matrix method that displays regions of similarity where 15 of 20 nucleotides are matched (Fig. 5). A large portion of the molecule shows similarity. However, long regions of similarity are interrupted by regions with no detectable similarity.

A large segment of the maize sequence near the 3' end of 359 nucleotides is not present in the E. coli sequence, and largely accounts for the difference in size between the two sequences. The extent of

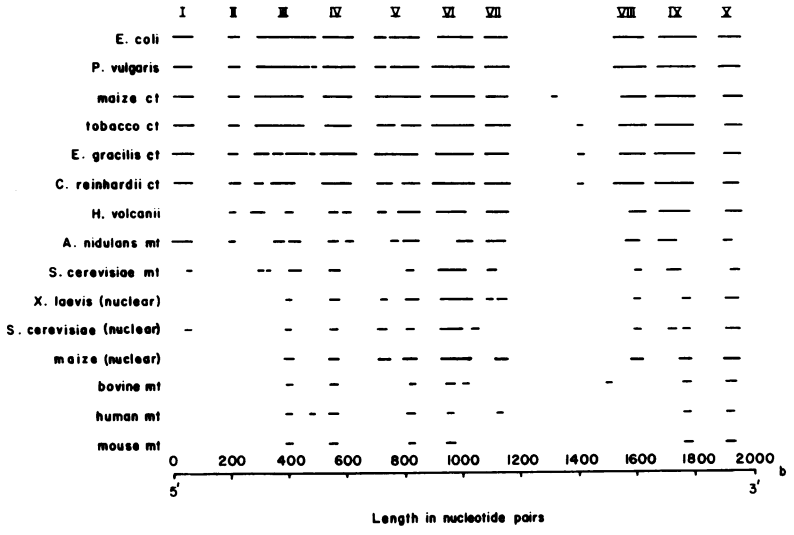

Fig. 5. Combined dot matrix analysis. Sequence comparison of maize mt 18S rRNA gene with 15 known small subunit rRNA gene sequences. The lines presented for each species are the diagonals of the dot matrix plots from comparisons with the maize mitochondrial sequence. The length (in bp) represents the maize 18S sequence, and the positions of similar sequences are shown for each species by the horizontal lines. The ten regions indicated by roman numerals are regions of sequence similarity common to many species.

sequence similarity was estimated from the length of aligned sequence in the dot matrix analysis. We have summed the length of non-overlapping aligned sequences and divided by the average length of the sequences being compared. When calculated by this method, the overall similarity of these sequences is 64%. If the 359 bp gap is excluded, the similarity is 70%. Excluding all gaps, and calculating homology on aligned regions alone gives a value of 81%. Because homology is not uniform within an aligned sequence and often exceeds 75%, this method of calculation might tend to underestimate similarity. Because the dot matrix includes regions with 75% similarity, the percent of alignment length can be higher than the exact nucleotide similarity. However, we have compared this result with an alternative method of calculation which depends on an alignment derived from the secondary structure. When this alignment is used to calculate a more precise estimate of similarity based on the total sequence, the value obtained is 61%.

5. Comparison of maize mt 18S rRNA sequence with 15 other small subunit rRNA sequences.

We have used the dot matrix technique to contrast the sequence of the

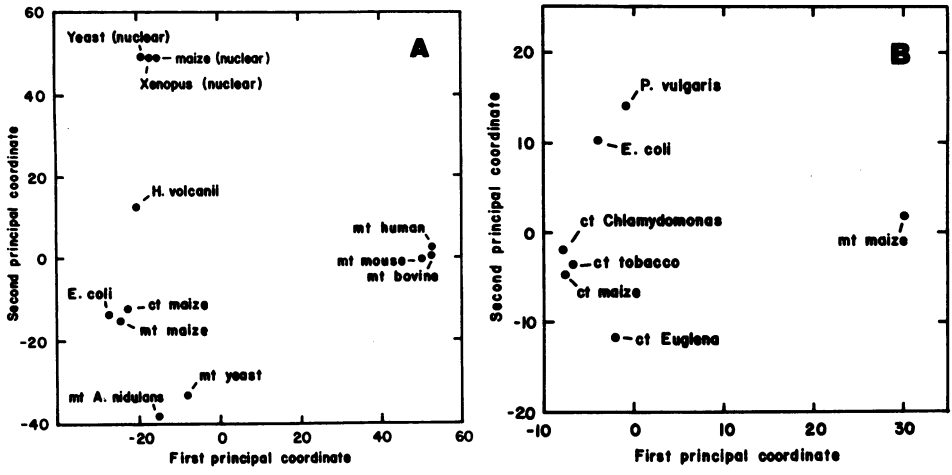

Fig. 6A. Principal coordinates analysis of some small subunit rRNAs. The genetic distance of maize mt 18S rRNA gene with small subunit rRNA gene of animal mt, fungal mt, prokaryotes and nuclear is estimated by principal coordinates analysis. The first principal coordinate accounts for 38% of the total variation and the second principal coordinate accounts for 31%.

Fig. 6B. Principal coordinates analysis of the maize 18S rRNA with chloroplast 16S sequences. The genetic distances of maize mt 18S and chloroplast 16S sequences from four species are compared with the 16S sequences of E. coli and P. vulgaris. The first principal coordinate accounts for 63% of the total variation and the second principal coordinate accounts for 28%.

18S rRNA of maize mitochondria with that of mitochondrial small subunit rRNAs of bovine (25), human (1), mouse (2), yeast (26) and Aspergillus nidulans (27). Three nuclear 18S rRNA sequences have been obtained from yeast (28) Xenopus laevis (29) and maize (52). In addition we have compared the 16S rRNA sequences of the eubacterium Proteus vulgaris and the archebacterium Halobacterium volcanii (30,31). Lastly, we have compared the sequence of maize, tobacco, Chlamydomonas reinhardtii and Euglena gracilis chloroplast 16S rRNAs (32-35) with that of maize mt 18S rRNA. To compare these sequences with the maize 18S sequence we have drawn the diagonals from dot matrix plots aligned against the maize mt sequence (Fig. 6). When all of these sequences are compared with the maize sequence, regions of similarity common to many species are observed (Fig. 6). The lengths of similar sequence and the number of similar regions varies. We have divided the molecule into 10 regions (Fig. 6). Some sequences within regions III, IV, V and X are similar among all sequences analyzed. These regions are located in the central and 3' end of the gene. Region VIII is present in most of the genes except those of animal mitochondria. The

segment between regions VII and VIII has little sequence similarity in all sequences examined. Small stretches of similarity 20 nucleotides long are found in this region in the comparison of maize mt with all the chloroplast genes examined, as well as the bovine mt rRNA gene. These 10 regions of similar sequence probably reflect functional conservation of specific sequences in the evolution of the ribosomal genes (16,23,43).

An unusual segment in the 3' domain

The greater length of the maize 18S gene compared to prokaryotic, chloroplast or other mitochondrial sequences is largely explained by a 359bp segment in the 3' domain. This segment is not due to the duplication of an adjacent segment because dot matrix analysis of the 18S gene against itself shows no major duplication in this region. It is unlikely that the 359bp segment is an intron because the size of the purified RNA coincides with the size of the gene, and a cloned fragment completely internal to the 359bp segment hybridizes with the mature RNA.

Estimation of genetic distance between small subunit rRNAs.

In order to explore the evolutionary position of plant mitochondrial genomes, we have compared our sequence with 25 published sequences of other small subunit ribosomal RNAs. We have used the alignment length from our dot matrix analysis to evaluate the sequence similarity in all combinations. The resulting similarity matrix (Table 1) can then be used to estimate the sequence identity (Fig. 6) or genetic distance.

We have estimated the relationships of all 16 sequences using principal coordinates analysis (36)(37). The similarity between different sequences is presented in a two-dimensional plot using the first two principal coordinates as the axes. The clustering of sequences indicates greater similarity. The relationships of the chloroplast, bacterial and plant mitochondrial sequences are better visualized when analysed separately (Fig. 6B).

The results indicate a high degree of similarity between sequences of the eubacteria, chloroplasts, and plant mitochondria. The striking result, supported by previous data, is the distance between the fungal mitochondria, animal mitochondria and that of the 18S rRNA maize mitochondrial sequence. The extent of divergence within mitochondria is far greater than the divergence observed between bacterial genes, or between chloroplast genes or between nuclear genes. The extent of divergence among the mammalian mitochondrial genes, which have diverged within about 80 million years (44), is very high and is comparable to that between nuclear 18S genes of yeast, maize and Xenopus, or between *E. coli* and chloroplasts, all genomes that are thought to have diverged in the Precambrian (> 600 million years).

DISCUSSION

The most intriguing aspect of the comparison of the maize 18S mitochondrial sequence with other small subunit sequences is the high degree of similarity with the eubacterial sequences and the relatively low degree of similarity with other mitochondrial small subunit genes. Undoubtedly, much of the similarity between the most diverse genes can be accounted for by conservation of essential sequences. However, functional conservation alone cannot account for all of the similarities because of the great diversity of small subunit sequences. Similarity could reflect ancestry; therefore the similarity of plant mitochondria with eubacteria could provide support for the origin of mitochondria as eubacterial symbionts. Alternatively, the interaction of chloroplast and mitochondrial genes, by recombination or gene conversion could be responsible for a high level of similar sequences. In addition, the rate of nucleotide divergence may vary in different organelles in different organisms. Other factors may involve selection for sequence specific functions (convergence). The possibility of polyphyletic origin of mitochondria must be considered, as well as the possibility of an exogenous independent origin of the ribosomal genes in mitochondria.

Whatever the origin of mitochondrial genomes, it is clear that they are unusual and exceedingly diverse with respect to mechanisms of transcription, the genetic code, genome size and genome organization. It is not possible to evaluate the possibility of a multiple origin of mitochondria, since the origin of any single mitochondrial system has not been established. While a multiple origin for mitochondrial genomes has been suggested, based on sequence comparisons, by Kuntzel and Kochel (43), we prefer the alternative that the rates of nucleotide divergence are particularly variable for mitochondrial DNA. This view is supported by evidence of a high rate of nucleotide divergence in mammalian mtDNA. Nucleotide divergence in mitochondrial genes is 6 to 10 fold higher than nuclear genes in mammalian systems (44,45). This effect is also apparent in comparison of sequences of the small subunit ribosomal genes in animal mitochondria (Table 1) which vary by the same degree as the yeast and Xenopus nuclear genes (Table 1).

The sequence of the wheat mitochondrial 18S rRNA gene has recently been determined (21) and is very similar (97%) to that of the same gene in maize. Our alignment length criterion would give a similarity of 99.8% In contrast, rRNA genes of bovine, mouse and human have 81 to 89% similarity by alignment length (Table 1). A reasonable estimate of the time since divergence of the major mammalian taxa is about 80 million years (44). In the evolution of the subfamilies of grasses, the time of divergence is not known but it seems likely that the subfamilies that gave rise to wheat

Table 1
Percent of alignment length

	abbrev.	length	eco	pvu	hvo	mct	tct	ect	cct	hmt	bmt	mmt	amt	ymt	xnu	mnu	ynu
maize mt		1968	64	65	41	62	63	65	62	16	17	14	36	24	21	21	21
E. coli	eco	1541	-	99	50	84	85	77	88	16	16	18	37	29	26	24	23
P. vulgaris	pvu	1544		-	50	79	81	73	81	18	18	17	37	26	21	21	21
H. volcanii	hvo	1472			-	53	53	49	53	16	15	18	14	11	30	32	30
maize ct	mct	1490				-	100	85	96	19	18	22	37	21	21	24	25
tobacco ct	tct	1486					-	86	96	18	18	20	38	23	20	25	25
Euglena ct	cct	1491						-	86	18	17	20	38	21	22	24	25
Chlamydomonas ct	cct	1475							-	20	18	20	34	21	20	23	22
human mt	hmt	954								-	89	81	11	7	10	9	11
bovine mt	bmt	955									-	86	11	12	11	10	10
mouse mt	mmt	955										-	15	17	11	11	14
A. nidulans mt	amt	1437											-	41	8	6	7
yeast mt	ymt	1654												-	7	9	9
Xenopus nuc	xnu	1825													-	84	84
maize nuc	mnu	1805														-	89
yeast nuc	ynu	1789															-

Table 1. Similarity matrix of small subunit rRNA sequences. Values presented are obtained from dot matrix analysis as shown in Fig. 4. The length of aligned regions is summed, after correction for overlaps and calculated as a percent of the length of the entire molecule. The values presented are calculated using the average of the lengths of the two sequences being compared, because the molecules are different in length.

(Pooideae) and maize (Andropogoneae) had diverged at least by the middle of the Tertiary, about 30 million years ago (53)(54). One tribe of the family, the Oryzeae, was already distinct in the Upper Eocene (57), indicating that the divergence of wheat and maize could have taken place in the early Tertiary, about 60 million years ago. Stebbins recommends an estimated date of 50-70 million years (58).

We have calculated the extent of base mutation (substitutions) between wheat and maize, correcting for multiple hits and transitional bias (21)(44), to be 0.010 substitutions per nucleotide. The comparable calculation for human-bovine is 0.206, for mouse-bovine 0.210 and for mouse-human 0.232. The divergence between the sequence of the rat (55) and the mouse is of interest because these groups have separated more recently. The fossil record of murid rodents is poorly known and the time since divergence of rats and mice could be between 5 and 35 million years (56). Wilson suggests 25-30 million years as the best estimate (59). The nucleotide substitution calculation for the rat-mouse comparison is 0.078 substitutions per nucleotide. Therefore the low amount of divergence between wheat and maize supports the view of a low rate of nucleotide substitution in plant mtDNA compared to animal mtDNA.

Since evidence for a high rate of divergence for mammalian mitochondria is strong, it becomes more reasonable to consider that the similarity

between the plant mitochondrial sequences and eubacteria reflects a slower rate of divergence compared to the divergence between eubacteria and animal mitochondria. If one compares the rRNA sequences within the fungi, the extent of similarity between the mitochondrial sequences of yeast and Aspergillus is less (41%) than between E. coli and maize mitochondria (64%) (Table 1). We propose that the rate of divergence for the 18S gene is slow in plant mitochondria.

In plants, animals and fungi, DNA transfer between nucleus and organelles or between organelles, has occurred repeatedly over long periods of evolutionary time (46-51). In maize, a segment of the chloroplast ribosomal gene cluster has been found in the mitochondria (50). In spite of these indications, it appears from the comparison of sequences of small subunit rRNA genes, that the functionally equivalent genes have been independent in sequence evolution. In animals and fungi, there is no indication of exchange or DNA transfer between nuclear and mitochondrial rRNA genes, since the nuclear genes from diverse sources are much closer to each other than to their animal or fungal mitochondrial genes. If recombination or gene conversion were to occur between chloroplasts and mitochondrial genes in plant cells, the ribosomal genes of maize mitochondria would be more similar to that of maize chloroplast than to other chloroplasts. However, the sequence of the chloroplast 16S rRNA genes from maize (32), tobacco (33), Euglena (34), and Chlamydomonas (35) are equally similar to the maize mitochondrial sequence (Table 1)(Fig. 8). Similarly, the maize nuclear rRNA gene is close to the nuclear rRNA genes of yeast and Xenopus and equivalent in similarity to the maize mitochondrial and maize chloroplast rRNA genes.

ACKNOWLEDGMENTS

We wish to thank Jane Suddith for technical assistance, and Ted Emigh for advice and assistance in the use of the dot matrix analysis for similarity calculation. Harry Noller and Robin Gutell provided advice and unpublished data for the analysis of secondary structure.

In addition we wish to thank M. Gray for making available data on the mitochondrial 18S rRNA sequences from wheat, J. Messing for data on the maize nuclear ribosomal RNA sequences. John Doebley provided assistance with the method of principal coordinates analysis.

This work was supported by the United States Department of Agriculture/Science and Education Administration competitive grants program and the National Science Foundation PCM-8010933. Paper No. 9080 of the Journal series of the North Carolina Agricultural Research Service, Raleigh, NC 27695-7614.

[1]Present address: Department of Biology, University of North Carolina, Chapel Hill, NC 27514, USA

REFERENCES

1. Anderson, S., Bankier, A. T., Barrell, B. G., deBruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. M. H., Smith, A. J. H., Staden, R. and Young, I. G. (1981) Nature (London) 290, 457-464.
2. Bibb, M. J., VanEtten, R. A., Wright, C. T., Walberg, M. W. and Clayton, D. A. (1981) Cell 26, 167-180.
3. Borst, P. and Grivell, L. A. (1978) Cell 15, 705-723.
4. Bonen, L. and Gray, M. W. (1980) Nucleic Acids Res. 8, 319-335.
5. Iams, K. P. and Sinclair, J. H. (1982) Proc. Natl. Acad. Sci. USA 79, 5926-5929.
6. Stern, D. B., Dyer, T. A. and Lonsdale, D. M. (1982) Nucleic Acids Res. 10, 3333-3340.
7. Lindahl, L. and Zengel, J. (1982) Adv. Genetics 21, 53-121.
8. Bedbrook, J. R., Kolodner, R. and Bogorad, L. (1977) Cell 11, 739-749.
9. Long, E. O. and Dawid, I. B. (1980) Ann. Rev. Biochem. 49, 727-764.
10. Spruill, W. M., Levings III, C. S. and Sederoff, R. R. (1980) Dev. Genet. 1, 363-378.
11. Chao, S., Sederoff, R. R. and Levings III, C. S. (1983) Plant Physiol. 71, 190-193.
12. Schaffer, H. E. and Sederoff, R. R. (1981) Anal. Biochem. 115, 113-122.
13. Messing, J., Crea, R. and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309-321
14. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
15. Sanger, F., Coulson, R., Hong, G. F., Hill, D. F. and Petersen, G. B. (1982) J. Mol. Biol. 162, 729-773.
16. Woese, C. R., Gutell, R., Gupta, R. and Noller, H. F. (1983) Microbiol. Rev. 47, 621-669.
17. Gray, M. W., Bonen, L., Falconet, D., Huh, T. Y., Schnare, M. N. and Spencer, D. F. (1982) in Mitochondrial Genes, pp. 483-488. Cold Spring Harbor.
18. Huh, T. Y. and Gray, M. W. (1982) Plant Mol. Biol. 1, 245-249.
19. Shah, D. M. and Levings III, C. S. (1974) Crop Sci. 14, 852-853.
20. Schnare, M. N. and Gray, M. W. (1982) Nucleic Acids Res. 10, 3921-3932.
21. Spencer, D. F., Schnare, M. N. and Gray, M. W. Proc. Natl. Acad. Sci. USA 81, 493-497.
22. Fox, G. E. and Woese, C. R. (1975) Nature 256, 505-507.
23. Noller, H. F. and Woese, C. R. (1981) Science 212, 403-411.
24. Bonen, L., Cunningham, R. L., Gray, M. W. and Doolittle, W. F. (1977) Nucleic Acids Res. 4, 663-671.
25. Anderson, S., deBruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F. and Young, I. G. (1982) J. Mol. Biol. 157, 683-717.
26. Sor, F. and Fukuhara, H. (1980) C. R. Acad. Sci. Paris, t. 291 D. 933-936.
27. Kochel, H. G. and Kuntzel, H. (1981) Nucleic Acids Res. 9, 5689-5696.
28. Rubtsov, P. M., Musakhanov, M. M., Zakharyev, V. M., Krayev, A. S., Skryanbin, K. G. and Bayeo, A. A. (1980) Nucleic Acids Res. 8, 5779-5794.

29. Salim, M. and Maden, E. H. (1981) Nature, 205-208.
30. Carbon, P., Ebel, J. P. and Ehresmann, C. (1981) Nucleic Acids Res. 9, 2325-2333.
31. Gupta, R., Lanter, J. M. and Woese, C. R. (1983) Science 221, 656-659.
32. Schwarz, Zs, and Kossel, H. (1980) Nature 283, 739-742.
33. Tohdoh, N. and Sugiura, M. (1982) Gene 17, 213-218.
34. Graf, L., Roux, E., Stutz, E. and Kossel, H. (1982) Nucleic Acids Res. 10, 6369-6381.
35. Dron, M., Rahire, M. and Rochaix, J. -D. (1982) Nucleic Acids Res. 10, 7609-7620.
36. Gower, J. C. (1966) Biometrika 53, 325-338.
37. Sneath, P. H. A. and Sokol, R. R. (1973) in Numerical Taxonomy - The Principles and Practice of Numerical Classification.
38. Margulis, L. (1970) in Origin of Eukaryotic Cells. Yale University Press, New Haven.
39. Gray, M. W. (1982) Can. J. Biochem. 60, 157-171.
40. Gray, M. W. and Doolittle, W. F. (1982) Microbiol. Rev. 46, 1-42.
41. Wallace, D. C. (1982) Microbiol. Rev. 46, 208-240.
42. Mahler, H. R. (1981) Ann. N. Y. Acad. Sci. 361, 53-75.
43. Kuntzel, H. and Kochel, H. G. (1981) Nature 293, 751-755.
44. Brown, W. M., Prager, E. M., Wang, A. and Wilson, A. C. (1982) J. Mol. Evol. 18, 225-239.
45. Miyata, T., Hayashida, H., Kikuno, R., Hasagawa, M., Kobayashi, M. and Koike, K. (1982) J. Mol. Evol. 19, 28-35.
46. Gellissen, G., Bradfield, J. Y., White, B. N. and Wyatt, G. R. (1983) Nature 301, 631-634.
47. Jacobs, H. T., Posakony, J. W., Grula, J. W., Roberts, J. W., Xin, J., Britten, R. J. and Davidson, E. H. (1983) J. Mol. Biol. 165, 609-632.
48. Timmis, J. N. and Scott, N. S. (1983) Nature 305, 65-67.
49. Farrelly, F. and Butow, R. A. (1983) Nature 301, 296-301.
50. Stern, D. B. and Lonsdale, D. M. (1982) Nature 299, 698-702.
51. Kemble, R. J., Mans, R. J., Gabay-Laughnan, S. and Laughnan, J. R. (1983) Nature 304, 744-747.
52. Messing, J., Carlson, J., Hagen, G., Rubenstein, I. and Oleson, A. DNA 3, 31-40.
53. Clayton, W. D. (1981) Ann. Missouri Bot. Gard. 68, 5-14.
54. Stebbins, G. L. (1981) Ann. Missouri Bot. Gard. 68, 75-86.
55. Kobayashi, M., Seki, T., Yaginuma, K., and Koike, K. (1981) Gene 16, 297-307.
56. Wilson, A., Carlson, S. S., and White, T. J. (1977) Ann. Rev. Biochem. 46, 573-639.
57. Litke, R. (1968) Monatsber. Deutsch. Akad. Wiss. Berlin 10, 462-471.
58. Stebbins, G. L., personal communication.
59. Wilson, A., personal communication.