A Study of the 5S Ribosomal RNAs of the Vibrionaceae

by

Michael Terrell MacDonell

University of Maryland, Department of
Microbiology, College Park, MD
Contract N00014-81-K-0638

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1984

APPROVAL SHEET

Title of dissertation: A Study of the 5S Ribosomal RNAs of the Vibrionaceae

Name of Candidate: Michael Terrell MacDonell
Doctor of Philosophy, 1984

Thesis and Abstract Approved: Rita R. Colwell
Rita R. Colwell
Professor of Microbiology
Department of Microbiology

Date Approved: November 23, 1984

CURRICULUM VITAE

Name: Michael Terrell MacDonell.

PII Redacted

Degree and date to be conferred: Ph.D., 1984.

Secondary education: Choctawhatchee High School, Shalimar, FL, 1966.

Collegiate institutions attended:	Dates	Degree	Date of Degree
Okaloosa-Walton Junior College	1972-1973		
University of Florida	1973-1974	B.S.	1975
University of Poznan, Poland	1975-1976		
University of West Florida	1981-1982	M.S.	1982
University of Maryland	1982-1984	Ph.D.	1984

Major: Microbiology
Minor: Biochemistry

Professional publications:

MacDonell, M.T., D.B. Roszak, and H.E. Gómez. 1980. Hematocrito por aproximacion. Proc 1er Nat. Health Workshop, Managua.

MacDonell, M.T. and M.A. Hood. 1982. Isolation and characterization of ultramicrobacteria from a Gulf Coast estuary. Appl. Environ. Microbiol. 43:556-571.

Baker, R.M., M.T. MacDonell, F.L. Singleton, and M.A. Hood. 1982. Effects of surface association and osmolarity on *Vibrio cholerae*. Proc. 7th Annu. Trop. Subtrop. Seafood Technol. Conf. of the Americas, pp. 224-238.

MacDonell, M.T., F.L. Singleton, and M.A. Hood. 1982. Diluent composition for use of API-20E in characterizing marine and estuarine bacteria. Appl. Environ. Microbiol. 44:423-427.

MacDonell, M.T. 1982. Isolation, characterization and distribution of ultramicrobacteria from a Gulf Coast Estuary. M.S. Thesis, The University of West Florida, Pensacola.

MacDonell, M.T., F.L. Singleton, D.B. Roszak, M.A. Hood, D.L. Tison, and R.J. Seidler. 1983. Rapid GC mol% screening of primary culture lysates using horizontal slab gel electrophoresis. J. Microbiol. Meth. 1:81-88.

MacDonell, M.T. 1983. Rapid estimation of Most Probable Number using a hand-held calculator. Can. J. Microbiol. 29:621-623.

MacDonell, M.T., R.M. Baker, F.L. Singleton, and M.A. Hood. 1984. Effects of surface association and osmolarity on seawater microcosm populations of an environmental isolate of *Vibrio cholerae*. *In* R.R. Colwell (ed.), Vibrios in the Environment, Wiley Interscience, New York. pp. 535-549.

MacDonell, M.T. and M.A. Hood. 1984. Ultramicrovibrios in Gulf Coast estuarine waters: isolation, characterization and incidence. *In* R.R. Colwell (ed.), Vibrios in the Environment, Wiley Interscience, New York. pp. 551-562.

MacDonell, M.T., E. Russek, and R.R. Colwell. 1983. An interactive microcomputer program for the computation of Most Probable Number. J. Microbiological Meth. 2:1-7.

MacDonell, M.T. and R.R. Colwell. 1984. Identical 5S rRNA nucleotide sequence of *Vibrio cholerae* strains representing temporal, geographical and ecological diversity. Appl. Environ. Microbiol. 48:119-121.

MacDonell, M.T. and R.R. Colwell. 1984. The nucleotide base sequence of *Vibrionaceae* 5S rRNA. FEBS Lett. 175:183-188.

MacDonell, M.T. and R.R. Colwell. 1984. The contribution of numerical taxonomy to the systematics of Gram-negative bacteria. *In* M. Goodfellow and D. Jones (eds.), Current Topics in Computer Assisted Bacterial Systematics. Academic Press, London.

MacDonell, M.T. and R.R. Colwell. 1984. Nucleotide sequence of 5S ribosomal RNA from *Vibrio marinus*. Micorbiological Sciences. (In press).

MacDonell, M.T. and R.R. Colwell. 1984. Microcomputer program for the computation of free energy of the secondary structure of RNA using Ninio's rules. J. Microbiol. Meth. (In press).

MacDonell, M.T. and R.R. Colwell. 1984. Nucleotide sequences of 5S rRNA from two *Vibrio* species, *V. fluvialis* and *V. parahaemolyticus*. MIRCEN J. Appl. Microbiol. Biotechnol. (In press).

Colwell, R.R., H.-S. Xu, N. Roberts, R. Siebeling, A. Huq, P. West, I. Huq, K.M.S. Aziz, D. Maneval, P. Brayton, D. Roszak, and M.T. MacDonell. 1984. Ecology and systematics of *Vibrio cholerae* in endemic and non-endemic areas. Proc. 19th Joint Conf. on Cholera, United States - Japan Cooperative Med. Science Program, 1983. (In press).

MacDonell, M.T. 1984. A study of the 5S ribosomal RNAs of the *Vibrionaceae*. Ph.D. dissertation, The University of Maryland, College Park.

MacDonell, M.T. and J.N. Hansen. 1985. Isolation and purification of RNAs suuitable for sequence analysis. *In* R.R. Colwell and R. Grigorova, (eds.), Methods in Microbiology 18. (In press).

Professional positions held:

December 1984 - Present	Post-doctoral Research Fellow, Department of Microbiology, University of Maryland, College Park.
January 1983 - December, 1984	Graduate Research Assistant, Department of Microbiology, University of Maryland, College Park.
July 1984 - August 1984	Lecturer, Department of Microbiology University of Maryland, College Park.
January 1984 - May 1984	Lecturer, University of Maryland University College Statewide Undergraduate Programs, College Park
January 1983 - December 1983	Graduate Teaching Assistant, University of Maryland University College, Statewide Undergraduate Programs, College Park.
August 1982 - January 1983	Graduate Teaching Assistant, Department of Microbiology, University of Maryland, College Park.
March 1981 - August 1982	Graduate Teaching Assistant, Department of Biology University of West Florida.
June 1979 - September 1980	Laboratory Supervisor, Thaeler Memorial Hospital; Laboratory Advisor: Atlantic Sector and Medical Technologist, Nicaraguan Ministry of Health.
February 1978 - June 1979	Medical Technologist, Thaeler Memorial Hospital, Bilwaskarma, Zelaya Province, Nicaragua.
September 1975 - August 1976	Research Assistant, Department of Biochemistry, University of Poznan, Poland.

ABSTRACT

Title of dissertation: A Study of the 5S Ribosomal RNAs of the Vibrionaceae

Michael Terrell MacDonell, Doctor of Philosophy, 1984

Dissertation directed by: Rita R. Colwell, Professor of Microbiology
The University of Maryland, College Park

The sequence of the 120 nucleotide bases of 5S ribosomal RNA (5S rRNA) has been determined for each of 23 eubacterial strains classified as species of the family *Vibrionaceae*. The methods employed included the resolution of limited enzymatic digests of end-labeled RNAs by high voltage polyacrylamide gel electrophoresis, as well as thin-layer chromatography. Several of the 5S rRNAs, for which the primary structures were determined, were also subjected to secondary structure analysis, employing nuclease S1, which hydrolyzes regions of nucleic acids not participating in the formation of helices, followed by resolution of the partial digests on thin sequencing gels.

Sequence data obtained from this study were compiled and analyzed, using statistical methods and group-specific signature analyses, for the purpose of constructing a phylogenetic taxonomy of the Vibrionaceae. Dot matrix maps were generated for 5S rRNA sequences determined in this study, in order to analyze the occurrence, extent, and complexity of palindromic and repeated base sequences. Data from nuclease S1 analyses, *i.e.*, observations of interactions among nucleotide bases,

were used in the evaluation of several secondary structure models derived from most probable base-pairing schema.

Results of sequence determinations of 5S rRNAs indicate that the family *Vibrionaceae* is heterogenous, as presently defined. The present genus *Vibrio* may comprise as many as three genera, one of which is composed of the majority of named *Vibrio* species. Results of dot matrix analyses indicate that the 5S rRNA molecule is composed of several inverted repeats (palindromes). Analyses of the extent of degeneracy of palindromic sequences suggests that palindrome analysis may be useful in determining the extent of evolution of a given species, with respect to others in the *Vibrionaceae*. Furthermore, composites of palindromic sequences shared by species of the *Vibrionaceae* suggest a partial base sequence of an ancestral 5S rRNA.

Dedicated to peace.

ACKNOWLEDGEMENTS

I wish to express my gratitude to Professor Rita R. Colwell for inviting me to her laboratory, and for providing not only direction and guidance but also an atmosphere of excitement which seems a necessary component to the discovery process. Aside from whatever modest skills I may have acquired, I place the highest value on the patience, judgement, and perspective I have gained through her example.

I wish to express my appreciation of each and every member of Professor Colwell's research group, past and present, since each, in his own way, has contributed to my success. I am grateful to Dr. D. J. Grimes for the advice and guidance he so willingly provided, and to the members of my dissertation committee, Professor Cyril Ponnamperuma, Dr. J. Norman Hansen, Dr. Ron Weiner, and Dr. William Brusilow, for their advice, interest, and friendship. I am especially grateful to Dr. Sam W. Joseph, for his enthusiasm, and for having faith in me, and to Dr. Emilio Weiss for having provided a role model.

I gratefully acknowledge the generous support of Sea Grant College, The Society of the Sigma Xi, The Interscience Conference on Antimicrobial Agents and Chemotherapy (American Society for Microbiology), The National Science Foundation (grant DEB-82-08418), The Office of Naval Research (grant N00014-81-K-0638), and The Graduate School of the University of Maryland.

Most of all I wish to express my sincerest gratitude to my partner in life Darlene Barbara Roszak, for the seemingly endless supply of love and support.

TABLE OF CONTENTS

LIST OF TABLES

Table 1. Bacterial Strains Employed for Sequence Analysis.

Nomenclature	Strain	Obtained from
Aeromonas hydrophila	9071	ATCC[1]
Aeromonas media	33907	ATCC
Alteromonas putrifaciens	8071	ATCC
Escherichia coli	MRE 600	PL Biochemicals
Photobacterium angustum	25915	ATCC
Photobacterium leiognathi	25521	ATCC
Photobacterium logei	15382	ATCC
Plesiomonas shigelloides	14029	ATCC
Vibrio alginolyticus	17749	ATCC
Vibrio anguillarum	19264	ATCC
Vibrio carchariae	35084	ATCC
Vibrio cholerae	14033	ATCC
	E 8498	J.P. Craig
Vibrio 'cincinnati'	vp cin	R.B. Bode
Vibrio damsela	33539	ATCC
Vibrio diazotrophicus	33466	ATCC
Vibrio fischeri	7744	ATCC
Vibrio fluvialis	NCTC[2] 11328	J.V. Lee
Vibrio gazogenes	19988	ATCC
Vibrio harinus (MP-1)	15381	ATCC
Vibrio metschnikovii	7708	ATCC
Vibrio mimicus	33655	ATCC
Vibrio natriegens	14048	ATCC
Vibrio parahaemolyticus	17802	ATCC
Vibrio psychroerythrus	27364	ATCC
Vibrio vulnificus	27562	ATCC

Key: [1] American Type Culture Collection, Rockville, MD
[2] National Collection of Type Cultures, London, UK

MATERIALS AND METHODS

A. BACTERIAL STRAINS

All bacterial strains (see Table 1) employed in the study were purchased from The American Type Culture Collection (ATCC, Rockville, MD), with the following exceptions: *V. cholerae* strain E-8498 was provided by J.P. Craig, *Vibrio* "cincinnatii" was provided by P.B. Bode, and *V. fluvialis* strain NCTC 11328 (ATCC 33812) was provided by J.V. Lee.

B. BACTERIAL CULTURE MEDIA

A broth medium consisting of equal parts of Marine Broth 2216 (Difco Laboratories, Detroit, MI) and Tryptic Soy Broth (Difco) was used for the batch cultivation of marine and estuarine strains. Tryptic Soy Broth (Difco) was used for the cultivation of non-salt-requiring bacterial strains. Psychrophilic marine strain *V. psychroerythrus* ATCC 27364) was grown in Marine Broth 2216 (Difco).

C. RNA EXTRACTION

Washed, pelleted bacterial cells from approximately 300 ml of an exponential phase broth culture were found to be adequate for recovery of adequate quantities of 5S rRNA species. When necessary, Gram-negative bacterial cells were lysed using the freeze-thaw technique (Zablen *et al.*, 1975). Resuspension of the washed cell pellet in Tris-borate-EDTA (TBE) buffer usually resulted in immediate cell lysis, eliminating the need to continue with freezing and thawing

consequence of a polythetic approach that reliance on *a priori* assumptions in the development of a taxonomy of the *Vibrionaceae* was minimized. The power of polyphasic taxonomy, in the case of the *Vibrionaceae*, in providing a natural classification is that it allows incorporation of new dimensions as the science of systematics advances. Thus, the ability to elucidate phylogenetic characteristics, *i.e.*, characteristics whose possession or absence constitutes a phylogenetic marker, can now be incorporated into an expanded polyphasic taxonomy for the *Vibrionaceae*. Even more attractive is that having a phylogenetic basis for the family structure, it is possible to select, *a posteriori*, those phenotypic characters embedded in the evolutionary history of the organisms. Selection of such characters is vital because workers in the field must rely on a determinative scheme for identification of bacterial taxa, especially when confronting new, previously unidentified species.

The objective of this study, therefore, was to construct a classification of the *Vibrionaceae*, using comparative sequencing of the 5S rRNAs as a means of elucidating the phylogeny of the family. In addition, the 5S rRNA molecules were analyzed for: relationships between primary (and secondary) structures and environmental parameters; evolutionary markers in the primary structures; and information of evolutionary significance at the level of the genetic transcript, *viz.*, conservation of reading frames, repeated sequences, inverted repeats, and conservation of hairpin loops and helices.

fixation and ribosomal RNA sequence determinations have shown that the Vibrionaceae and Enterobacteriaceae share a relatively common evolution (Baumann *et al.*, 1984; Baumann and Schubert, 1984). The proximity in phylogeny of the two families is reflected in the fact that, taken together, they contribute more than ninety percent of the species comprising RNA superfamily I, one of at least five superfamilies of the eubacterial kingdom (De Vos and De Ley, 1983).

In contrast to their relative similarity, the identification schemes for the *Vibrionaceae* and *Enterobacteriaceae* is strikingly different, due, in part, to the influence of clinical microbiology on the taxonomy of the Enterobacteriaceae, resulting in "overclassification" of species and genera. In fact, the phylogenetic depth of the entire family Enterobacteriaceae is less than that of the genus *Vibrio*, suggesting that the genera of the Enterobacteriaceae, as presently defined, more closely approximate species, on a basis of phylogenetic relationships (MacDonell and Colwell, 1984b).

The Vibrionaceae, as presently defined, consists of the genera *Vibrio* (28 species), *Aeromonas* (4 species), *Photobacterium* (3 species) and *Plesiomonas* (1 species) (Baumann and Schubert, 1984), all of which are associated with either aquatic or marine environments. Whereas, the taxonomy of the *Enterobacteriaceae* was heavily influenced by clinical criteria, the taxonomy of the Vibrionaceae, as presently defined, derives largely from a polyphasic approach (Citerella and Colwell, 1970; Colwell, 1970; West and Colwell, 1984). Using this approach, dependence upon key characteristics was abandoned, and instead, a very wide range of equally weighted characters, both phenetic and genetic, were employed in order to generate clusters (species) of related strains (Colwell, 1968). It was a

helical regions, there is a tendency for the base sequence to be much less highly conserved. In fact, the rate of mutation in these hypervariable regions is approximately twice that of conserved regions (MacDonell and Colwell, 1984a). That two significantly different rates of mutation exist in 5S rRNA sequences provides a coarse and fine focus for interpretation of sequence comparisons. Compilations of 5S rRNA sequences indicate that sequence differences occur in "hypervariable" regions down to, and possibly including, species, indicating applicability to phylogenetic inferences within given genera. Sequence differences in the more highly conserved regions, however, appear to occur only at the genus and family level. Thus, the different mutation rates allow an extension of the range of taxonomic levels for which sequences of 5S rRNA can be compared. At present, the total published library of bacterial 5S rRNA sequences comprises ca. 50 in number.

C. THE FAMILY VIBRIONACEAE

In 1965, Veron proposed the family Vibrionaceae for non-enteric, Gram-negative rods, and suggested two major criteria for differentiation of these strains from those of the family Enterobacteriaceae: (1) possession of a cytochrome oxidase; and (2) motility by means of a single polar flagellum (Veron, 1975). The classification scheme of Veron (1965) was not designed to reflect phylogenetic relationships, rather it was proposed for convenience in separating the two groups. It is interesting to note that, although the species assigned to both families have undergone revision, it has been extensive in the case of the Vibrionaceae (see Baumann *et al.*, 1980). Comparative studies focussing on bacterial evolution, employing new methods, such as quantitative microcomplement

(1580 bases compared with 120 in 5S rRNAs) is lost because of the need to resort to oligomer catalogs. At this writing, only seventeen complete 16S rRNA sequences are known to exist, although more than 4000 16S rRNA oligomer catalogs have been constructed (A. Böck, personal communication).

Although *E. coli* 5S rRNA was one of the first nucleic acid molecules for which the complete nucleotide base sequence was determined (Brownlee *et al.*, 1967), its secondary structure is still a subject of controversy. Application of rules derived from more than a decade of research on the primary and secondary structure of tRNA (Tinoco *et al.*, 1971; Ninio, 1979) has provided several structures, all of which are consistent with the known physical data, but none has gained acceptance as the representation of spatial conformation. The concept suggests that the role of 5S rRNA in the 50S ribosomal subunit may be modulatory, thus, it may well involve switching between two (or more) conformations.

It is now reasonably well established that comparisons among 5S rRNAs provide a rather firm basis for evaluating evolutionary relatedness among bacterial species (Luehrsen and Fox, 1981; Dekio *et al.*, 1984; MacDonell and Colwell, 1984b; Sogin *et al.*, 1972). In fact, 5S rRNAs represent ideal material for sequence determination since they are easily isolated and purified to homogeneity (MacDonell and Hansen, 1985), are small enough to be sequenced in their entirety, and appear to lack post-transcriptionally modified bases (Luehrsen and Fox, 1981; MacDonell and Colwell, 1984d). With regard to comparative sequencing, 5S rRNAs consist of two qualitatively different regions. In one, that generally associated with single-stranded portions of the molecule, the nucleotide base sequence is highly conserved. In the other, usually associated with

which consists of 23S, 16S and 5S rDNA, along with sequences coding for one or more tRNAs (Lund *et al.*, 1979). Pioneering work in the field of rRNA sequence analysis involved comparisons among catalogs of oligomers prepared from ribonuclease T1 digests of ribosomal RNAs (Sogin *et al.*, 1972). This method gave rise to the employment of two-enzyme comparative catalogs, from which ribonucleotide base sequences could be inferred with a relatively high degree of accuracy (Uchida *et al.*, 1974). Chemical sequencing methods for RNA were developed and improved to the point where the primary structures of rRNA molecules could be determined both routinely and unambiguously (Peattie, 1979). Enzymatic methods (Donis-Keller *et al.*, 1979; Donis-Keller, 1980) were slower in development and were more restricted in range, since they could not be employed with either tRNAs or eukaryotic RNAs in which substantial post-transcriptional modification of nucleotide bases occur. Prokaryotic 5S rRNAs, however, provide ideal substrates for enzymatic methods, and can be sequenced unambiguously using established techniques (MacDonell and Colwell, 1984d). It is still considered impractical, in terms of both time and material, to sequence large ribosomal RNAs. Although oligomer catalogs from digests of the relatively small 5S rRNAs are virtually useless, the sequence of its 120 nucleotide bases can be determined readily. 5S rRNA studies, therefore, generally employ sequence determinations, whereas 16S rRNA studies still depend on oligomer cataloging. Comparisons among populations of oligomers, however, do not allow the same resolution as comparisons among sequences, since similar oligomer populations do not necessarily reflect unique sequences. It is frustrating, therefore, that in the construction of evolutionary trees, much of the potential resolution offered by 16S rRNAs

phosphatase, superoxide dismutase, and glutamine synthetase (Bang *et al.*, 1981; Baumann *et al.*, 1980; Baumann *et al.*, 1983; Woolkalis and Baumann, 1981). Comparative sequencing of ribosomal RNAs, however, remains unrivaled as a method for evaluating phylogenetic relationships.

B. RIBOSOMAL RNA AND EVOLUTION

The ribosome is believed to have evolved in three stages: (1) establishment of a relatively simple archetypal mechanism, along with a primitive set of codon assignments; (2) increase in complexity of the translation mechanism, in which the codon assignments assumed their present form; and (3) increase in the efficiency, *i.e.*, rapidity and precision of the process of translation (Woese, 1970). Allowing that, in the process of evolution, central features of primitive processes remain the central features of their evolved counterparts, it is reasonable to expect the modern bacterial translation apparatus to reflect characteristics of its evolutionary predecessor (Fox and Woese, 1975). Furthermore, since it can be shown that ribosomal RNAs are very highly conserved, they should reflect the base composition, and to some extent, the base sequence of the primitive rDNA cistron, and, therefore, the ancestral genome. Similarly, primary structures of ribosomal components, particularly the ribosomal RNAs, provide unique insights phylogenetic relationships among bacterial taxa, as well as into the nature of biochemical evolution.

Ribosomal RNAs comprise approximately 80% of the total bacterial RNA, and are processed post-transcriptionally from a single large (30S) RNA polymer (Dunn and Studier, 1973; Nikolaev *et al.*, 1973). This single transcript is coded for by the "ribosomal RNA transcription unit" (rRTU).

compositions provided an interesting contrast to classically derived taxonomic schema (Colwell and Mandel, 1964; Thornley, 1967), and the impact of DNA base composition on bacterial systematics has been such that it is now a requirement for the minimum description of a new bacterial species by the International Committee of Systematic Bacteriology.

Within a decade after the advent of DNA base composition analysis, DNA/DNA hybridization methods significantly extended the sensitivity with which the bacterial genome could be probed, and provided a means by which the primary genetic structures of two distinct genomes could be compared directly, permitting inference of taxonomic relationships based on phylogeny, and in which phylogenetic distances between closely related strains could be quantitated by extent of DNA homology shared by the strains.

More recently, the ability to routinely determine sequences of nucleotide bases in genetic material (Sanger *et al.*, 1977; Donis-Keller, 1979; Peattie, 1979; Maxam and Gilbert, 1980; MacDonell and Colwell, 1984d) has elevated the sophistication of nucleic acid methods. Large quantities of phylogenetic information can be obtained, stored on the computer, and retrieved permitting sequence data comparisons as the sequences are determined, *i.e.*, on-line, if desired.

The introduction of comparative sequencing, and to a lesser extent, immunologic comparisons, of key biological molecules has allowed construction of geneologies of species. Biological molecules used for tracing of natural relationships include quinones (Collins and Jones, 1979), cytochromes (Schleifer *et al.*, 1982), ferridoxins (Schwartz and Dayhoff, 1978) and microcomplement fixation techniques employing alkaline

Woese *et al.*, 1984). The authority reflected in the existing system of nomenclature, even though not warrented, suggests the existence of natural relationships, and, by implication, a phylogeny, however speculative.

We might consider the period between the biochemical taxonomy of Orla-Jensen (1909), and the publication by Fox *et al.* of *The Phylogeny of the Prokaryotes* (1981), as the determinative era in microbial systematics. It was during the last two decades of this era that the enormous advances in molecular biology occurred which brought the construction of a natural taxonomy of bacterial species closer to reality. Methods for nucleic acid analysis, including DNA/DNA and DNA/RNA hybridization, oligomer cataloging, and nucleic acid sequencing, have attained some prominence in microbial taxonomy in recent years, although the significance of some results obtained using these methods are open to interpretation. Nevertheless, the newer methods permit analysis, in great detail, of the molecular genetic structure, from which direct evidence of natural relationships among bacterial species can be deduced.

The first of the new generation of methods to emerge was the determination of bacterial DNA base composition (Lee *et al*, 1956; Belozersky and Spirin, 1960), which provided a direct, although crude, probe into the bacterial genome. Results of comparisons of "base ratio" determinations clearly demonstrated the significance of G+C molar ratios in bacterial taxonomy (Colwell and Mandel, 1964; Hill, 1966; MacDonell and Colwell, 1984f), and even now provide a powerful and routine means of confirming relatedness (or the lack of relatedness) among phenetically similar strains. The earliest compilations of bacterial base

waning need to establish a bacterial phylogeny was all the more insidious since several generations of microbiologists have had to satisfy themselves with a taxonomy in a constant state of flux. For these microbiologists, also, the elucidation of natural relationships among bacterial species appeared to be an intractable problem, if not a myth. A result is that many microbiologists have regarded the construction of a phylogenetic taxonomy of bacterial species as a game of speculation and an activity irrelevant to the practical aspects of microbiology.

Interestingly, *a priori* determinative approaches to bacterial taxonomy, a school of thought for which the later editions of Bergey's manual might be regarded as a cornerstone, appeared to be reasonably successful. Determinative approaches, indeed, provided a practical basis for allocation of bacterial strains to taxa, although at the expense of regarding phylogenetic relationships as secondary in importance. It is ironic that, in an atmosphere of skepticism and suspicion of putative phylogenetic treatments, a level of skill and technology has been acheived that is sufficient to determine natural, *i.e.*, phylogenetic relationships with relative ease.

Borne largely out of clinical considerations and reinforced by successive editions of Bergey's manual, an exaggerated emphasis on the identification of bacteria by means of a minimum number of key characteristics, *i.e.*, a determinative approach, has promoted a false confidence that certain "key" phenetic characters are *a priori* sufficient for the definition of natural taxa. Unfortunately, such practice has been reinforced by the use of Latin and Greek nomenclature, with associated rules, reserved, by convention, for classifications which reflect true phylogenetic relationships (Stackebrandt and Woese, 1984;

INTRODUCTION

A. HISTORIC APPROACHES TO BACTERIAL SYSTEMATICS

The need to define natural relationships among bacterial species has dominated classical microbial systematics since its beginnings more than a century ago. Indeed, numerous systems of taxonomy purporting to reflect a natural order have been proposed (Cohn, 1872; Bergonzini, 1879; Orla-Jensen, 1909), in general based on either morphological or physiological criteria. Despite transient popularity, none of these systems, founded necessarily on purely phenetic traits, can be said to represent the true natural order. Bisset (1950, 1952, 1957) attempted an ambitious bacterial taxonomy based on what he speculated to be phylogenetic criteria, but was unsuccessful since there existed at that time no basis for evaluating whether or not a feature was a phylogenetic characteristic. Thus, the "bacterial phylogeny" of Bisset (1950, 1952, 1957) was, itself, speculative (Sneath, 1962). The failure to achieve a phylogenetic taxonomy is not surprising since methods involving direct genetic comparison necessary for (i) detection of true evolutionary relationships; and (ii) identification of valid phylogenetic characteristics, have only recently become available.

Over the past several decades bacterial taxonomy has undergone frequent and extensive rearrangement. Unfortunately, there was also a slow but significant loss of interest in the historic search for a natural classification of the bacteria. Indeed, some microbiologists lost confidence in the belief that a natural classification of bacterial species could be constructed (Shimwell and Carr, 1960; Cowan, 1962). The

ABBREVIATIONS

BAC	*bis*-acrylylcystamine
biovar	biological variant
DNA	deoxyribonucleic acid
C	centigrade
g	gram
G	gravity
K(nuc)	"nucleotide constant" (see Kimura, 1980), a measure of evolutionary distance
M	molar
NA	nucleic acid
n-mer	oligomer of subunit length n
pmol	picomole, *i.e.*, 10^{-12} mole
R	purine
RNA	ribonucleic acid
RNY	"RNY" rule codon, *i.e.*, purine, any base, pyrimidine (see Shepherd, 1981; 1982)
S_{AB}	Evolutionary distance coefficient
S-value	Similarity value, coefficient
STOP	termination codon (UAA, UAG, or UGA)
TBE	Tris-borate-EDTA buffer
ug	microgram, *i.e.*, 10^{-6} gram
ul	microliter, *i.e.*, 10^{-6} liter
UPG	unweighted pair-group
UPGMA	unweighted pair-group mathematical average
WPG	weighted pair-group
Y	pyrimidine

LIST OF FIGURES

the cells, and significantly reducing exposure of RNAs to endogenous RNAses and to lysozyme, a possible source of foreign RNases. Cell lysis was evidenced by a rapid and marked increase in viscosity (as well as a noticable clearing) of the cell suspension.

The method of Kirby (1956), with the modifications discussed below, was used for extracting 5S rRNA. In the first step, *i.e.*, crude nucleic acid extraction, the crude cell lysate was brought to 1M NaCl and vortexed with an equal volume of chloroform for 10 minutes. The aqueous and organic phases were separated by centrifugation at 12,000 x G for 15 minutes at 4 C. Substitution of chloroform for phenol in the first extraction step increased yields of the total nucleic acid fraction as much as 50% (MacDonell and Hansen, 1985). Since chloroform and phenol remove proteins from aqueous solutions by different mechanisms, virtually all chloroform-precipitable proteinaceous material was removed with a single chloroform extraction. The aqueous (upper) phase, containing the nucleic acid fraction, was removed from underneath with a hook-shaped pasteur pipette. The aqueous phase was precipitated with 2 volumes of cold (-80 C) absolute ethanol and mixed several times by inversion, after which it was placed on dry ice for 10 minutes. After chilling, it was collected by centrifugation (12,000 x G for 10 minutes at 4 C), and resuspended in sterile 50 mM TBE buffer. The nucleic acid pellet was re-extracted, using a phenol solution of the following composition: 79% (w/v) phenol (redistilled), 11% (w/v) m-cresol (redistilled), 10% (v/v) TBE buffer, and 0.05% (w/v) 8-hydroxyquinoline. After collection, the aqueous phase was precipitated with 2 volumes of cold (-80 C) absolute ethanol, mixed several times by inversion and chilled on dry ice for 10 minutes. The

total nucleic acid (NA) fraction was collected by centrifugation as described above. After the alcoholic supernatant had been decanted, the nucleic acid pellet was dried partially under vacuum for several minutes in order to remove excess ethanol. The nucleic acid was resuspended in as small a volume of sterile distilled water as would permit complete dissolution, generally 10 to 20 ml.

D. ISOLATION OF RNA

To the aqueous nucleic acid solution was added an equal volume of redistilled 2-methoxyethanol, 2.5 M K_2HPO_4; and 0.05 volume of 3% (v/v) H_3PO_4. This was mixed vigorously for 10 minutes and separated by centrifugation for 10 minutes at 12,000 x G at 4 C. The aqueous (upper) phase, containing the total RNA fraction, was collected, and precipitated by addition of 2 volumes of cold ethanol followed by chilling on dry ice, as described above. Prior to collection by centrifugation, a substantial quantity of fluffy white RNA, occupying as much as 1/3 of the volume of the centrifuge tube, was usually evident. The RNA pellet was dissolved in a minimum volume (a few milliliters) of autoclave-sterilized TBE buffer.

E. DEAE-CELLULOSE CHROMATOGRAPHY

The small oligomer fraction, representing a significant proportion of the crude RNA extract, was removed before fractionation and purification of 5S rRNA by polyacrylamide gel electrophoresis. The fine cellulose particles were removed (Peterson, 1980) from approximately 1 gram of washed and TBE-equilibrated DEAE-cellulose (Cellex-D, Bio Rad, Richmond, CA). The prepared DEAE-cellulose was

packed in a 5 ml sterile plastic syringe barrel, to a depth of 5 to 6 cm. The crude RNA solution was then applied to the column and was rinsed with ten column volumes of TBE buffer. A rinse with 0.2 M NaCl in TBE buffer was sufficient to remove small oligomers (less than about 50 bases). The fraction containing 5S rRNA was eluted with 1 M NaCl, 7 M urea in 50 mM TBE buffer (Hansen, 1981), and precipitated by addition of 2 volumes of cold absolute ethanol and chilling on dry ice for 10 minutes. The precipitated RNA was collected by centrifugation, as described above, and suspended in a tracking dye-buffer of the following composition: 8 M urea, 0.05% xylene cyanol and 0.05% brom-phenol blue in TBE buffer.

F. PURIFICATION OF 5S rRNA

5S rRNA was isolated and purified using 5% (w/v) acrylamide preparative stacking gels, in which the upper- and lowermost stacks (approximately 25% of the length) consisted of acrylamide/ bisacrylamide, and the center stack (middle 50%) consisted of acrylamide/bis- acrylylcystamine (BAC, Bio Rad, Richmond, CA). Substitution of bis-acrylylcystamine, a thiol-soluble cross-linking agent, for N,N'-methylene-bis-acrylamide produces a soluble acrylamide gel (Hansen, 1976; Hansen *et al.*, 1980; Hansen, 1981). The use of conventional (bis) acrylamide in the upper stack facilitates the formation of clean, square-bottomed slots (since 5% BAC is somewhat sticky and tends to adhere to slot formers). After casting the first (lowest) stack, the gel form (along with the degassed, uncatalyzed BAC-acrylamide) was placed in an incubator set at 42-45 C and left to equilibrate for approximately an hour. It was necessary to warm the

gel form thoroughly, prior to pouring the gel, since failure to do so was found to favor the formation of insoluble C-N bonds, a consequence of reduced temperature during polymerization. After completion of polymerization, the BAC-acrylamide was poured to within 3 or 4 cm from the top of the gel form and polymerization allowed to proceed for 60 minutes at 42 C to 45 C. After removal of the gel form from the incubator, the remaining bis-acrylamide solution was degassed, initiated and poured. A large 5-tooth preparative slot former was inserted. Gels (150 mm x 175 mm x 4 mm) were pre-electrophoresed for at least 30 minutes at 4 to 5 W (constant power), *i.e., ca.* 20 to 25 mA. After pre-electrophoresis, the sample wells were loaded with 50 to 100 microliters of RNA/tracking dye solution and electrophoresed at 7 W (constant power), *i.e., ca.* 40 mA. The current was adjusted so that the glass plates or the gel form reached 55 C to 65 C, in order to inhibit formation of secondary structure. Electrophoresis of the RNA solution was continued until the brom-phenol blue band began to exit the BAC-acrylamide gel stack, after which the gel was removed, stained for 30 minutes with a 1 ug/ml aqueous solution of ethidium bromide in distilled water, and viewed on an ultraviolet (short wavelength) transilluminator, or alternatively, imaged by UV-shadowing (Hassur and Whitlock, 1974).

The 5S rRNA band, located about 85% of the distance from the brom-phenol blue band to the xylene cyanol band, was excised with a sterile blade (see Carmichael, 1980; Hecht and Woese, 1968; Loening, 1967). The excised gel plugs were placed into sterile siliconized glass test tubes, and solubilized with 5 ml of sterile TBE buffer and 75 ul of 2-mercaptoethanol (Hansen, 1981).

G. RECOVERY OF RNA FROM LIQUIFIED GEL

DEAE-cellulose columns were prepared as follows. A small plug of glass wool was aseptically placed in the bottom of a long tip pasteur pipette. Both the glass wool and the pasteur pipettes were silanized and autoclave-sterilized. To this was added sufficient DEAE-cellulose slurry to form a bed of about 1 cm in height. This was pre-rinsed with 1 to 2 ml of the final elution buffer consisting of 1 M NaCl, 7 M urea in 50 mM TBE buffer followed by equilibration with 5 ml of fresh TBE buffer. The solution containing RNA and solubilized gel was added to the column and rinsed with 2 ml of fresh sterile TBE buffer followed by 2 ml of 0.2 M NaCl in TBE buffer. The 5S rRNA fraction was eluted with 500 ul of 1 M NaCl in TBE, added in aliquots of 100 ul (Hansen, 1981). 5S rRNA thus recovered was found to be of sufficient purity for sequence analysis, *i.e.*, >99% pure.

H. SILANIZING GLASS AND PLASTIC WARE

All glassware used in the isolation and purification of RNAs was silanized (Schlief and Wensink, 1981), and rinsed in reagent grade water (Milli-Q reagent water system, Millipore Inc., Bedford, MA). Glassware was baked at 300 C for 4 hours. Plastic micropipette tips and microfuge tubes were silanized as follows. Approximately 2 ml of a 1:1 ratio of chloroform and dimethyl, dichlorosilane were pipetted onto a watch glass placed in the bottom of a glass vacuum desiccator from which the desiccant had been removed. Tubes and pipette tips to be silanized were distributed inside, and vacuum was applied for about 10 seconds before being released abruptly. This process was repeated 10

to 15 times, after which the plastic ware was removed, rinsed thoroughly with distilled water, and sterilized for 5 minutes in an autoclave followed by a drying cycle.

I. END-LABELING RNA WITH ^{32}P

Two methods of end-labeling RNA species with ^{32}P were employed. In the first, [5' ^{32}P] cytidine-bis-phosphate was covalently attached to the 3' (OH) terminal of the RNA. The second involved transfer of the (gamma) phosphate of [gamma ^{32}P] ATP to a dephosphorylated 5' terminal nucleotide. End-labeling the 5' terminus of nucleic acids required a dephosphorylation step, since forward reaction in the case of phosphate transfer on the 5' terminal is strongly favored by dephosphorylation of the 5' terminal nucleotide. Although labeling of the 3' terminal base (using RNA ligase) generally requires a dephosphorylation step, in order to avoid formation of concatemers (chains of oligomers attached 5' to 3'). 5S rRNAs are folded in such a way as to produce a 3' overhang, which obscures the 5' terminal base. Therefore, dephosphorylation of the 5' terminus prior to 3' end-labeling was omitted. Since the unequivocal determination of the complete sequence requires construction of sequence ladders from both ends, separate aliquots of the 5S rRNA species were labeled on both termini, in turn.

J. DEPHOSPHORYLATION OF THE 5' TERMINUS

High yields of end-labeled RNA by the forward phosphorylation of 5' hydroxyls required dephosphorylation of the native 5' terminus (Richardson, 1965; Lillehaug *et al.*, 1976). Calf intestinal (alkaline)

phosphatase (CIP) was found to be superior to bacterial alkaline phosphatase (BAP) for the dephosphorylation of bacterial RNAs. Furthermore, highly purified (molecular biology grade) CIP is commercially available (Boehringer, Indianapolis, IN). Approximately 25 pmol (1 ug) of 5S rRNA is sufficient for a complete sequence analysis. With 250 uCi of fresh (*i.e.*, less than a week old) [gamma-^{32}P] ATP, end-label yields on the order of 10 million cpm/ug of RNA were commonplace.

To a silanized autoclave-sterilized microfuge tube was added 1 ug 5S rRNA in 10 ul sterile water; 0.1 U of CIP, 1 mM $MgCl_2$, 90 ul CIP buffer, 50 mM Tris-HCl, pH 9.0, 1 mM spermidine. This was incubated for 30 minutes at 37 C. The reaction was terminated by the addition of 100 ul of phenol/m-cresol solution (described above), and briefly vortexed. Organic and aqueous phases were separated by centrifugation in a microfuge for 5 minutes at 4 C. The aqueous (upper) phase was collected, using a sterile micropipette, placed in a sterile microfuge tube and chilled in an ice bath. To the organic phase was added 100 ul of TBE. This was vortexed briefly, and phases were separated by centrifugation, as described above. The aqueous phase was collected and pooled. 100 ul of phenol/m-cresol was added, vortexed briefly, and centrifuged, as described above. The aqueous phase was collected and to it was added 500 ul diethyl ether (stored over water). The mixture was vortexed briefly, centrifuged for 10 seconds and the upper (ether) phase was discarded. Ether extraction was repeated once. The ether-extracted RNA solution was degassed on a vacuum line for several minutes to remove traces of dissolved ether, and placed in an ice bath. 50 ul of 2 M sodium acetate (pH 5.5) was added, vortexed briefly, and

returned to the ice bath. The RNA was precipitated with 400 ul cold (-80 C) absolute ethanol, mixed several times by inversion, and chilled on dry ice for 10 minutes. The RNA was collected by centrifugation for 5 minutes in a microfuge. Ethanol was gently removed using a drawn capillary, taking care to avoid disturbing the RNA pellet (usually not visible). 400 ul of cold absolute ethanol was carefully layered over the RNA, chilled on dry ice for 2 minutes, and centrifuged for 1 minute in a microfuge. The ethanol was discarded, as described above. The dephosphorylated RNA was placed in a vacuum desiccator for 15 minutes to remove all remaining traces of ethanol.

K. LABELING THE 5' TERMINUS

The method employed to end-label the 5' terminal base of 5S rRNA was a modification of the method described by D'Alessio (1983). The dephosphorylated 5S rRNA was resuspended in 10 ul of water. To this was added 5 ul of 10 uM ATP, and 2 ul of 10X kinase buffer of the following composition: 0.5 M Tris-HCl, pH 9.0; 10 mM $MgCl_2$; and 10 mM spermidine. It was then placed in an ice bath. 250 uCi of [gamma-^{32}P] ATP (Amersham, Springfield, IL), packed in a 1:1 ratio of ethanol and water, was evaporated to dryness under a nitrogen jet, after which the reaction mixture was added and vortexed briefly. Five U of T_4 polynucleotide kinase (PNK) was added. The mixture was vortexed briefly and incubated for 15 minutes at 37 C. After incubation, the reaction was terminated by addition of 30 ul of ammonium acetate and 25 ug of phenol-extracted tRNA (carrier). The ^{32}P-RNA was precipitated by the addition of 200 ul of cold (-80 C) absolute ethanol, and mixed by inversion. The mixture was chilled on dry ice for 10 minutes and collected by centrifugation in a microfuge for 5 minutes. The alcoholic (radioactive) supernatant was collected with a drawn capillary and discarded as radioactive waste. The ^{32}P-RNA was resuspended in 100 ul of cold 0.5 M sodium acetate by swirling, precipitated in 300 ul of cold ethanol, and collected as described above, and overlayed with 500 ul of cold ethanol, chilled on dry ice, centrifuged, and collected as described above. The ^{32}P-RNA was placed in a vacuum desiccator for 10 to 15 minutes to remove remaining traces of alcohol.

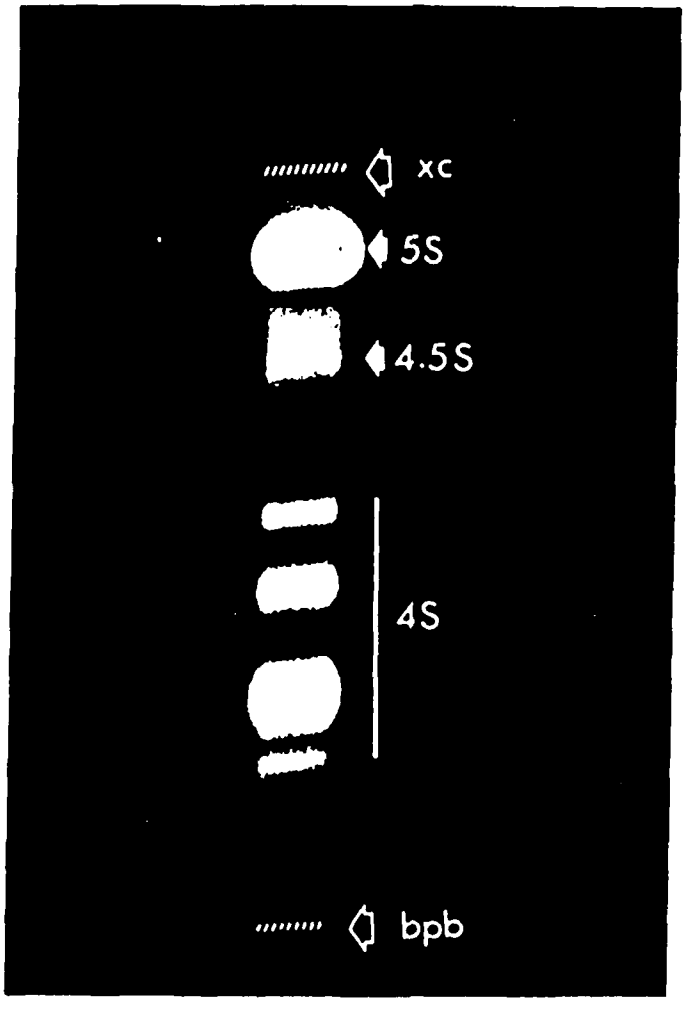
L. LABELING THE 3' TERMINUS

To approximately 1 ug of RNA, resuspended in 10 ul of water, was added 2 ul dimethyl sulphoxide (DMSO); 2 ul of 0.2 M ATP; and 2 ul of 10 X ligase buffer of the following composition: 0.5 M HEPES, pH 7.5; 0.1 M $MgCl_2$; and 33 mM dithiothreitol (DTT). 250 uCi of [^{32}P] pCp (ICN Radiochemicals, Palo Alto, CA), packed in water, was evaporated to dryness under a nitrogen jet, after which the reaction mixture was added and vortexed briefly. The reaction mixture was placed in an ice bath, and 10 U of RNA ligase was added. The mixture was vortexed briefly, and incubated for 4 hours at 4 C. Termination of the reaction, precipitation, and collection of ^{32}P-RNA was exactly as described above for 5' end-labeling.

M. ELECTROPHORETIC PURIFICATION OF ^{32}P-RNA

The end-labeled RNA was resuspended in a sterile siliconized microfuge tube with 20 ul of tracking dye/buffer (see Isolation and Purification section) and placed in an ice bath. A 5% polyacrylamide/8 M urea denaturing gel, 40 cm in length, was cast and pre-electrophoresed for 30 minutes. The ^{32}P-RNA/tracking dye mixture was heated briefly to 90 C, chilled, loaded on the gel, and electrophoresed until the xylene cyanol (upper) band travelled 3/4 of the length of the gel. The glass plates were removed, and the gel covered with plastic film and autoradiographed using X-ray film (XAR-5, Kodak Corp., Rochester, NY). The image of the 5S rRNA band was identified (Figure 1) and cut from the film using a razor blade, and the film used as a template to locate the position of the 5S rRNA in the gel. After positioning the film template, the 5S rRNA band was excised and placed in a sterile

Figure 1. Relative locations of RNA bands in a 5% acrylamide denaturing gel. Bands were generated by the fractionation of ^{32}P-RNA on a 5% acrylamide gel containing 8 M urea. Locations of xylene cyanol FF (xc) and brom-phenol blue (bpb) are indicated by arrows.


xc
5S
4.5S
4S
bpb

siliconized 1 ml (blue) Eppendorf pipette tip which had been heat sealed and plugged with a small quantity of sterile siliconized glass wool.

N. ELUTION OF ^{32}P-RNA FROM bis-ACRYLAMIDE GELS

Elution of ^{32}P-RNA from conventional bis-acrylamide gels was accomplished as follows. The gel slice was crushed to a paste against the walls of the pipette tip using a sterile siliconized glass rod. The glass rod was rinsed into the gel paste with 300 ul of 0.5 M ammonium acetate, 1 mM EDTA in 100 ul aliquots. Phenol-extracted carrier tRNA was added to a ratio of approximately 50:1 (in the majority of cases, about 50 ul). The top of the pipette tip was sealed with parafilm and incubated overnight at 37 C. The ^{32}P-RNA was recovered by carefully cutting the heat-sealed tip with a sterile blade, then removing the parafilm from the top of the pipette tip, and allowing the RNA solution to collect in a sterile 1.5 ml microfuge tube. The pipette tip was rinsed with an additional 100 ul of ammonium acetate, precipitated with 2 volumes of cold (-20 C) absolute ethanol, mixed several times by inversion, and chilled on dry ice, for 10 minutes. The ^{32}P-RNA was collected by centrifugation in a microfuge for 5 minutes at 4 C, and stored in sterile EDTA (0.01 M, pH 6.5) at -20 C.

O. TERMINAL ANALYSIS

1. IDENTIFICATION OF THE 5' TERMINAL NUCLEOTIDE

5,000 to 20,000 cpm of 5' end-labeled RNA was exhaustively digested with nuclease P1 and chromatographed on a PEI-cellulose thin layer plate (with fluorescent indicator). Standards composed of the 4 common nucleotide base monophosphates were spotted on the TLC plate and

chromatographed alongside the labeled digest, using 0.25 M LiCl as the mobile phase (Randerath and Randerath, 1967). After completion of the chromatography, the reference monophosphates were located with a short wavelength black light, and their positions marked on the TLC plate. Next, the TLC plate was autoradiographed overnight in order to determine location of the unique radioactive terminal base. R_f values of nucleotide monophosphates on PEI-cellulose, developed with 0.25 M LiCl were taken from Randerath and Randerath (1967).

2. IDENTIFICATION OF THE 3' TERMINAL NUCLEOTIDE

Analysis of the 3' terminal base follows the same rationale as 5' terminal analysis, except that RNase T2 was substituted for nuclease P1 in the exhaustive digestion of the ^{32}P-RNA.

B. ENZYMATIC SEQUENCE ANALYSIS OF RNA

Sequences of RNAs not containing modified bases (bacterial 5S rRNAs) were determined using enzymatic methods and chromatography on ultrathin polyacrylamide gels using an enzymatic approach (Donis-Keller, 1980; MacDonell and Colwell, 1984d). A number of specific endoribonucleases have been characterized and are commercially available. These include T1, U2, Phy M, B.c., CL3, and M1 (PL Biochemicals, Milwaukee, WI). Endoribonucleases T1 (Sato and Egami, 1957) and U2 (Uchida *et al.*, 1974), exhibit a high degree of specificity. RNases T1 and U2 hydrolyze the phosphate backbone, producing 5'-phosphates at guanines and adenines respectively. Enzymes highly specific for cytidine and uridine have yet to be characterized. Therefore, it is necessary to employ several RNases in concert in order

sequence variations, however, were observed in virtually every region of the molecule. Taken as classes, several sequence variations in the "hypervariable" B/B' helix were found to be characteristic of 5S rRNAs from clusters of similar strains. The existence of these characteristic sequences, known as "group-specific signatures" appears to be common to all ribosomal RNAs (Delihas and Andersen, 1982; Kuntzel *et al.*, 1983). Based on the possession of common group specific signatures, the 5S rRNA sequences determined in this study could be grouped into five sets. These groups, as well as the characteristic signature sequences, are listed in figure 4.

B. CLUSTER ANALYSIS

Comparison of the 5S rRNA sequences was accomplished by cluster analysis, using both unweighted pair-group (UPG), and weighted pair-group mathematical average (WPGMA) methods. In the case of cluster analysis by UPG, three different dendrograms were generated from the 5S rRNA sequence data. These correspond to single linkage, average linkage, and complete linkage clustering (figure 5a-c). In the case of cluster analysis by single linkage, strains (and clusters of strains) were linked at the level of the highest degree of relatedness of any two of their component sequences. In the case of complete linkage, strains (and clusters) were linked at the highest level of similarity in the sequence of one cluster compared with every sequence for strains of another cluster. Average linkage portrays the mathematical average similarity (S-value) between clusters. For WPGMA cluster analysis, the algorithm of Kimura (1980), which estimates evolutionary distances from pairwise comparisons of nucleic acid sequences, was used. This

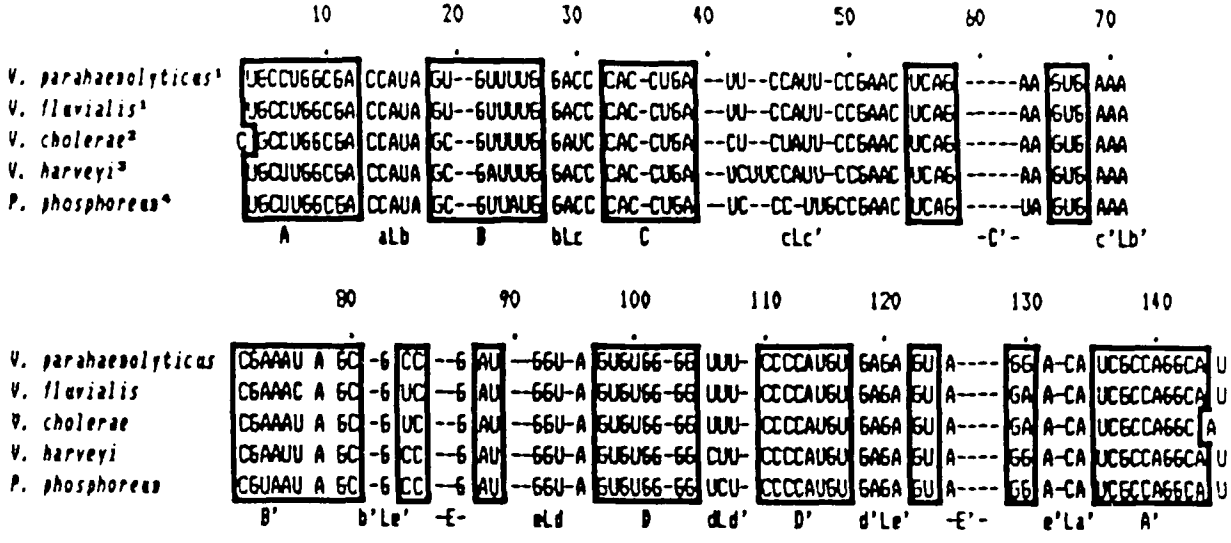
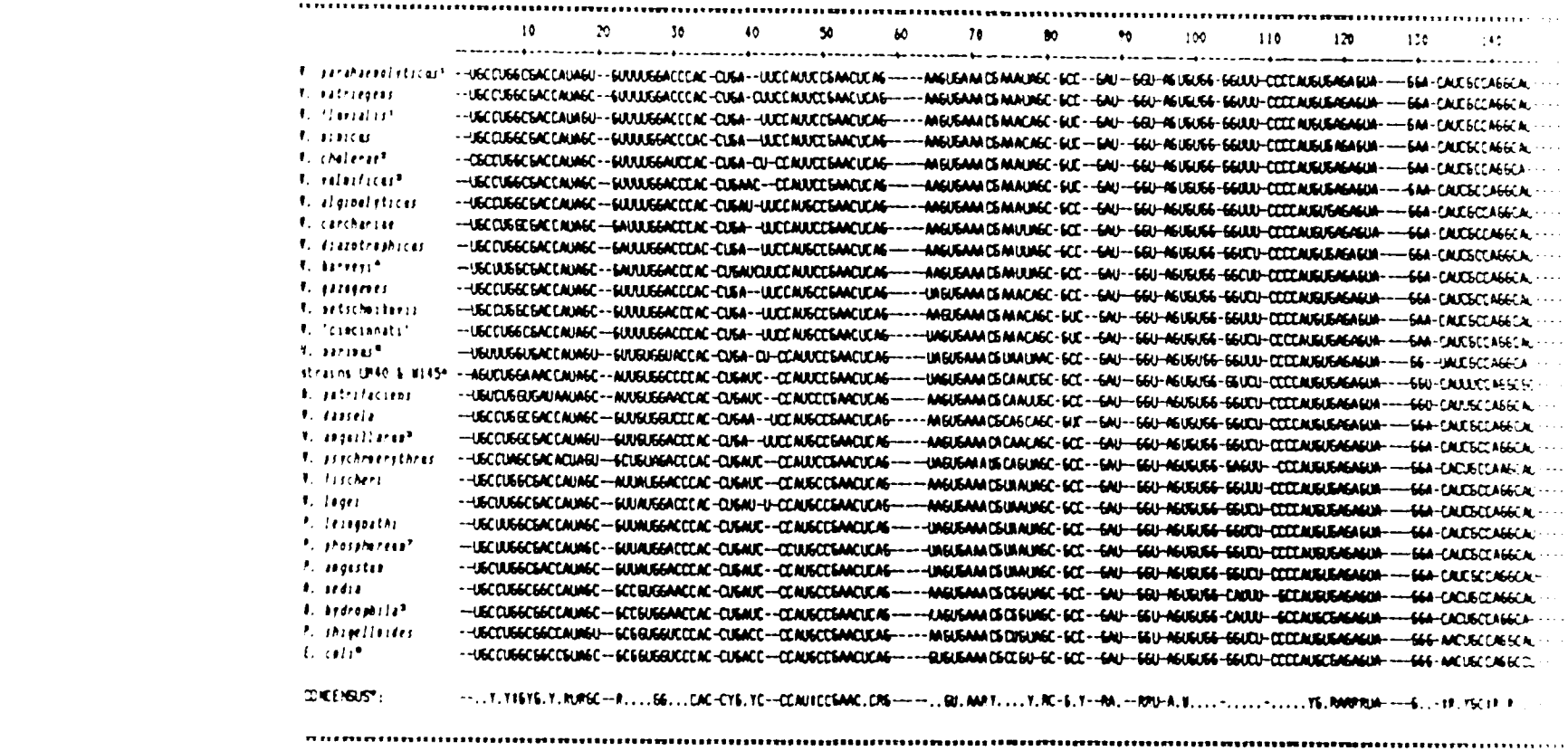

10 20 30 40 50 60 70
V. parahaemolyticus[1]
V. fluvialis[1]
V. cholerae[2]
V. harveyi[3]
P. phosphoreus[4]
A aLb B bLc C cLc' -C'- c'Lb'
80 90 100 110 120 130 140
V. parahaemolyticus
V. fluvialis
V. cholerae
V. harveyi
P. phosphoreus
B' b'Le' -E- eLd D dLd' D' d'Le' -E'- e'La' A'

Figure 3. Alignment of 5S rRNA sequences. Nucleotide base sequences of five 5S rRNAs are depicted in the alignment scheme of Erdmann *et al.* (1983). Boxed areas indicate regions assumed to participate in helices. Hypothetical loops and helices are indicated using the lettering scheme of Hori and Osawa (1979), wherein helix B/B', for example, would derive from the base-pairing of regions B and B'. Loops (L) are flanked by the adjacent helix designations, indicated in lower case.

References:

[1]MacDonell and Colwell (1984e)
[2]MacDonell and Colwell (1984a)
[3]Luehrsen and Fox (1981)
[4]Woese *et al.* (1975)

Table 2. List of 5S rRNA Sequences



Keys: U = uracil residue; C = cytosine residue; - = no base; Y = pyrimidine residue; B = unique to prokaryotes; G = guanine residue; A = adenine residue; . = any base may occur; R = purine residue

References: [1] MacDonell and Colwell (1984c); [2] MacDonell and Colwell (1984d); [3] MacDonell and Colwell (1984b); [4] Moese et. al. (1975); [5] Erdmann et. al. (1982); [6] MacDonell and Colwell (1984a); [7] Luehrsen and Fox (1981); [8] Dessing et. al. (1984); [9] Brownlee et. al. (1967)

-
T1
Phy M
U2
OH⁻
Bc
M1
++
T1
Phy M
U2
OH⁻
Bc
M1
++

2b. Sequence ladders generated from limited enzymatic digest of 5S rRNAs purified from V.vulnificus *(left) and* A. hydrophila *(right). Identities of the endoribonucleases are listed at the tops of sequence lanes. A portion of the nucleotide base sequence of V.* vulnificus *is listed to the left of the enzyme T1 lane.*

G U/A A [N] C/U -C
UACCGCGGAACCACUCCAUCCGGACUCAGAAGUGAAACGCAAUUG

Figure 2. Sequence ladders of 5S rRNAs generated using an enzymatic approach.

2a. Sequence ladders generated from the limited enzymatic digest of 5S rRNA purified from *Alteromonas putrifaciens*. Lanes are identified as to the substrate nucleotide (see text for explanation). Identities of nucleotide bases are listed to the left of the "5" lane.

RESULTS

A. 5S rRNA SEQUENCES

The nucleotide base sequences of 5S rRNAs from 26 bacterial strains were determined by the enzymatic method (Donis-Keller, 1980; MacDonell and Colwell, 1984d). Sequences were deduced from "sequence ladders" generated by electrophoresis of limited digests of uniquely end-labeled 5S rRNAs on ultrathin polyacrylamide sequencing gels (Sanger and Coulsen, 1978) and imaged by autoradiography. A typical autoradiogram, annotated to indicate the character of digest lanes and identity of base sequence, is shown in figure 2. The generation of several (typically 3 to 10) such sets of sequence ladders was necessary in order to deduce each composite, *i.e.*, complete, nucleotide sequence. The sequences of 5S rRNAs from strains representing 25 named, or suspected *Vibrio* species (as well as *E. coli*, included in the study as a control) are listed in Table 2. Sequence alignments are based on the recommendation of Erdmann *et al.* (1983). Helical regions (De Wachter *et al.*, 1982; Erdmann *et al.*, 1984; MacDonell and Colwell, 1984d) are indicated as boxed-in areas, and loops and helices are designated by the lettering scheme of Hori and Osawa (1979). See figure 3. Although a length of 120 nucleotide bases was typical, a range of variation in length, from 119 to 122 nucleotides, was observed. With only a single exception in 26 sequences, *i.e.*, the sequence of 5S rRNA from *V. marinus*, the site of length variation was restricted to the base 40 - 44 region of loop cLc'. Sequence variation was found to be restricted, to a large degree, to two regions, referred to as "hypervariable" by Fox *et al.*, (1977). These regions correspond to helix B/B' and loop cLc' (Figure 3). Minor

T. NUCLEASE S1 LIMITED DIGESTS

End-labeled 5S rRNAs from some species were subjected to limited, i.e., "single-hit" digests using nuclease S1 (PL Biochemicals, Milwaukee, WI). Sequence ladders from these digests, and from conventional sequencing digests, were generated simultaneously in order to determine the locations of single-stranded and helical (double-stranded) regions in the 5S rRNA molecule. S1 digests were prepared as described by Maniatis *et al.* (1982), except that the concentration of Zn^{++} ions was reduced to 5% of that recommended for hydrolysis of DNA. This was done to reduce exposure of RNA to Zn^{++} ions which have been shown to efficiently introduce lesions in the phosphate backbones of RNAs (Butzow and Eichhorn, 1975).

U. DATA MANAGEMENT

Storage of sequence data, sequence comparisons, free energy determinations, and generation of dot matrices, evolutionary trees and dendrograms, as well as generation of graphics, was done using either of 2 computer systems. These were (1) TRS-80 model 100 computer, outfitted with 64 kilobytes of random access memory (RAM) (Tandy Corp., Ft. Worth, TX) or (2) Commodore VIC-20 (Commodore Business Machines, West Chester, PA), outfitted with 32 kilobytes of RAM, and an 80-column board (Protecto Enterprises, Barrington, IL). All computer programs employed in this study are listed in appendix A.

xylene cyanol (i.e., upper, band. The wrapped gels were placed on a sheet of cardboard and, in a darkroom, the radioactive bands were imaged by exposing the gel to a sheet of X-ray film for approximately 10 seconds, the exact exposure time being determined empirically. Using a sterile knife blade, the photographic image of the 5S rRNA band was carefully cut from the film. The film was aligned, and used as a template for excising the ^{32}P-5S rRNA from the gel.

2. AUTORADIOGRAPHY OF SEQUENCE LADDERS

In the case of the imaging of sequence ladders produced by the limited enzymatic digests, very small quantities of radiation often were involved. Whether or not to use intensification screens for the autoradiography was dictated by the yield of labeled termini. Autoradiography of sequence ladders required at least 50,000 cpm of ^{32}P-RNA per lane, but this could be reduced to 10,000 by using intensification screens. In general, however, it was possible to read screened autoradiograms only to about 2/3 the distance as those exposed without screens.

Thin sequencing gels were removed from the gel form and wrapped in plastic wrap. In the darkroom, the gels were covered with X-ray film, placed in a film casette, and allowed to expose overnight. When intensification screens were used, the cassette was placed in a -70 C freezer. Otherwise they were placed in a -20 C freezer.

2. REACTION BUFFERS

RNases T1, Phy M, and M1: 0.025 mM sodium citrate, pH 5.0; 7 M urea; 1 mM EDTA; and 0.05% (w/v) each of xylene cyanol FF and brom-phenol blue.

RNase U2: 0.025 mM sodium citrate, pH 3.5; 7 M urea; 1 mM EDTA; and 0.05% (w/v) each of xylene cyanol FF and brom-phenol blue.

RNase B.c.: 0.025 mM sodium citrate, pH 5.0.

S. AUTORADIOGRAPHY OF ^{32}P-RNA

The generation of autoradiograms was necessary at two stages. The first was in locating the ^{32}P-RNA after purification by gel electrophoresis. The second applied to the imaging of end-labeled fragments resulting from limited enzymatic digests, *i.e.*, the "sequence ladder". Kodak XAR-5 film (Kodak Corp., Rochester, NY) was used for autoradiography. Films were developed with either D-11 or with Kodak GBX developer.

1. LOCATING THE PURIFIED ^{32}P-RNA BAND

After the RNA had been end-labeled and subsequently purified by gel electrophoresis, the RNA band was located and excised so that the purified ^{32}P-RNA could be recovered. Relatively large amounts (>10^7 dpm) of radiation were involved; therefore, typical exposure times ranged from 10 to 30 seconds. After completion of purification of ^{32}P-RNA electrophoresis, the gel was removed from the gel form and covered with plastic wrap. When necessary, brittle denaturing gels were trimmed with a pizza cutter in order to produce smooth edges. In 5% denaturing acrylamide gels, 5S rRNA bands were located just below the

10% (w/v), depending on the oligomer size range under investigation. A lane containing a limited alkaline hydrolysis of an aliquot of labeled RNA was run alongside the sequence lanes in order to mark the position of each length of n-mer, from n=1 to n=122, in the autoradiogram of the sequencing gel.

Sequencing gels were cast and run on a 40 cm model SD sequencing system (BRL, Gaithersburg, MD), to which power was supplied by a Bio Rad (Richmond, CA) model 3000 high voltage transformer. Simultaneous electrophoresis of the limited RNA digests in adjacent lanes of a thin denaturing polyacrylamide gel consistently resulted in a unique set of well-resolved bands from which the RNA sequence was read directly.

Adjustment of the enzyme-to-substrate ratio so as to achieve "single-hit" conditions was approached two ways: (1) serial 10-fold dilutions of the endoribonucleases to achieve the proper titre; and (2) adjustment of substrate concentration by addition of carrier tRNA. Adjustment of substrate concentration was generally more useful in controlling enzyme/substrate ratios.

R. BUFFERS

1. ENZYME DILUTION BUFFERS:

RNases T1 and Phy M: 25 mM sodium citrate, pH 5.0; 7 M urea; 1 mM EDTA; 0.05% (w/v) xylene cyanol FF; and 0.05% (w/v) brom-phenol blue.

RNase U2: 25 mM sodium citrate, pH 3.5; 1 mM EDTA; 0.05% (w/v) xylene cyanol FF; and 0.05% (w/v) brom-phenol blue.

RNases B.c. and M1: 25 mM sodium citrate, pH 5.0.

to unequivocally identify the pyrimidine bases. Three of these: Phy M, specific for adenine and uridine (Donis-Keller, 1980); B.c., specific for cytidine and uridine (Lockard *et al.*, 1978); and M1, which hydrolyzes at adenine, guanine and uridine, but not cytidine residues (PL Biochemicals, unpublished data), were of particular value. Endoribonucleases useful for sequence analysis, except for RNase M1, hydrolyze the phosphate backbone, producing 5' phosphates at the site of specificity. RNase M1, however, hydrolyzes the phosphate backbone to produce a 5' nucleotide phosphate. This results in the shifting of the RNase M1 lane one position out of frame, and must be taken into consideration in analyzing the subsequent sequence ladders. RNase CL3 (Bogusky *et al.*, 1980; Levy and Karpetsky, 1980), with a putative specificity for cytidine residues was not employed because of inadequate specificity.

Q. LIMITED ENZYMATIC DIGESTS

"Sequence ladders" from which the nucleotide base sequence of RNA molecules was read directly, were generated by adjusting the ratio of enzyme (endoribonuclease) to substrate (RNA) so that each end-labeled RNA molecule was clipped, on average, at exactly one site. This resulted in a nested set of fragments in which every possible end-labeled sequence was represented. Limited digests of separate aliquots of ^{32}P-RNA, using the various endoribonucleases, were used to generate sets of fragments which were terminated by the enzyme-specific base on one end and the ^{32}P label on the other. These fragments were separated on thin 8 M urea polyacrylamide denaturing gels (Sanger and Coulsen, 1978) which ranged in total acrylamide concentration from 7% to

Figure 4. Group-Specific Signatures in the *Vibrionaceae*. The hypervariable B/B' helix was found to constitute the smallest sequence segment by which the major clusters of strains could be distinguished. The five consensus sequences, therefore, are considered to represent group-specific signatures (see Delihas and Andersen, 1982).

5'...GUUUUG$_{27}$...
3'...YAAAGC$_{72}$...

V. parahaemolyticus
V. natriegens
V. fluvialis
V. cholerae
V. vulnificus
V. mimicus
V. alginolyticus
V. gazogenes
V. metschnikovii
V. "cincinnatii"

5'...RUUAUG$_{27}$...
3'...UAAUGC$_{72}$...

P. phosphoreum
P. logei
P. angustum
P. leiognathi
V. fischeri

5'...GAUUUG$_{27}$...
3'...YUANGC$_{72}$...

V. carchariae
V. harveyi
V. diazotrophicus

5'...RUUGUG$_{27}$...
3'...YAAYRC$_{72}$...

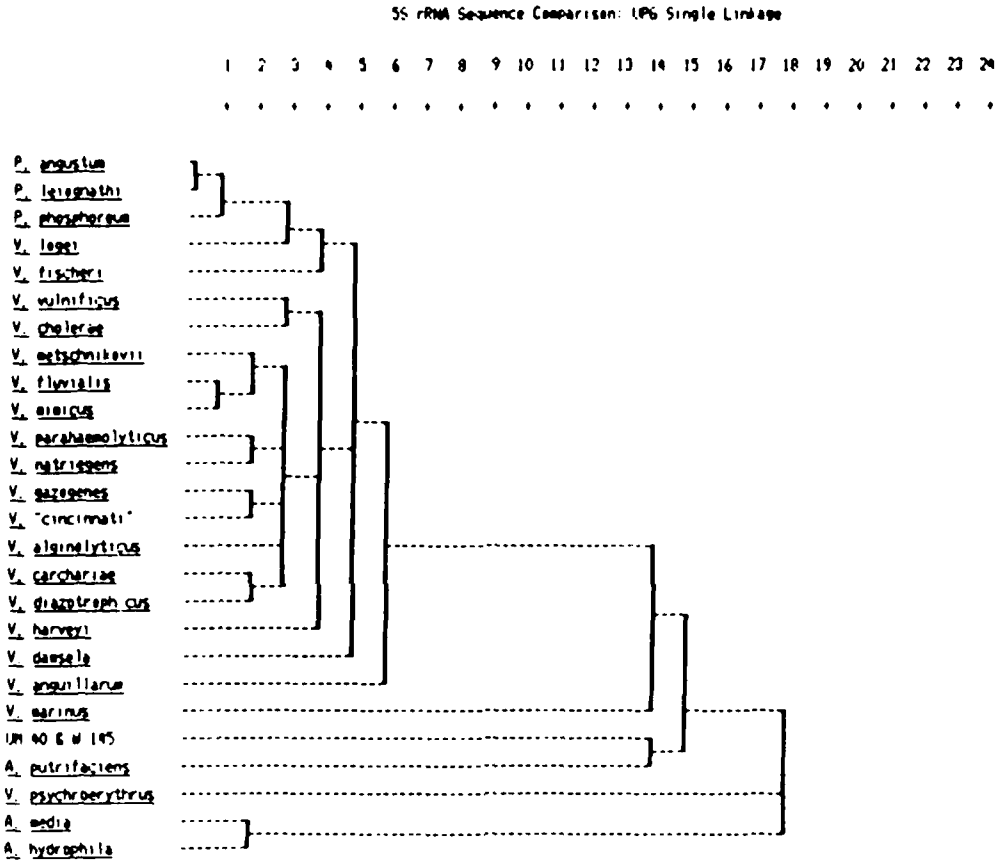
V. marinus
V. anguillarum
V. damsela
A. putrifaciens
strains UM40 and W145

5'...GCnGUG$_{27}$...
3'...UGnYGC$_{72}$...

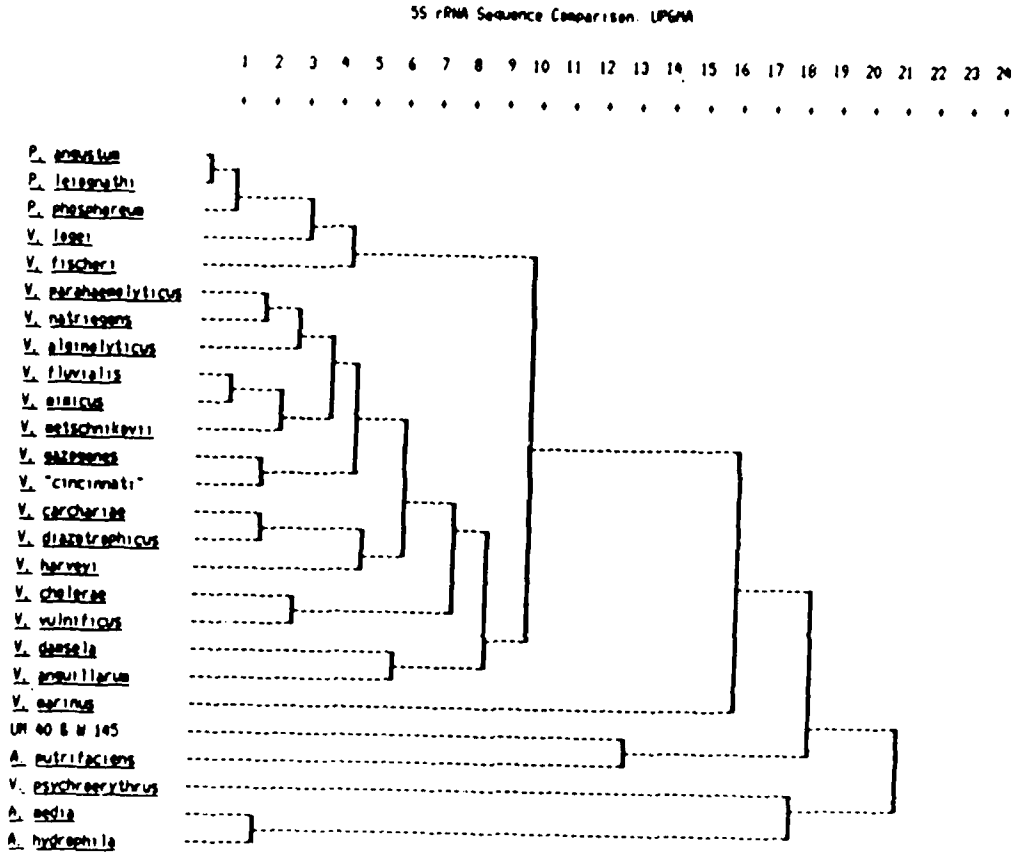
P. shigelloides
E. coli
A. hydrophila
A. media

Figure 5. Unweighted pair group (UPG) analysis dendrograms.

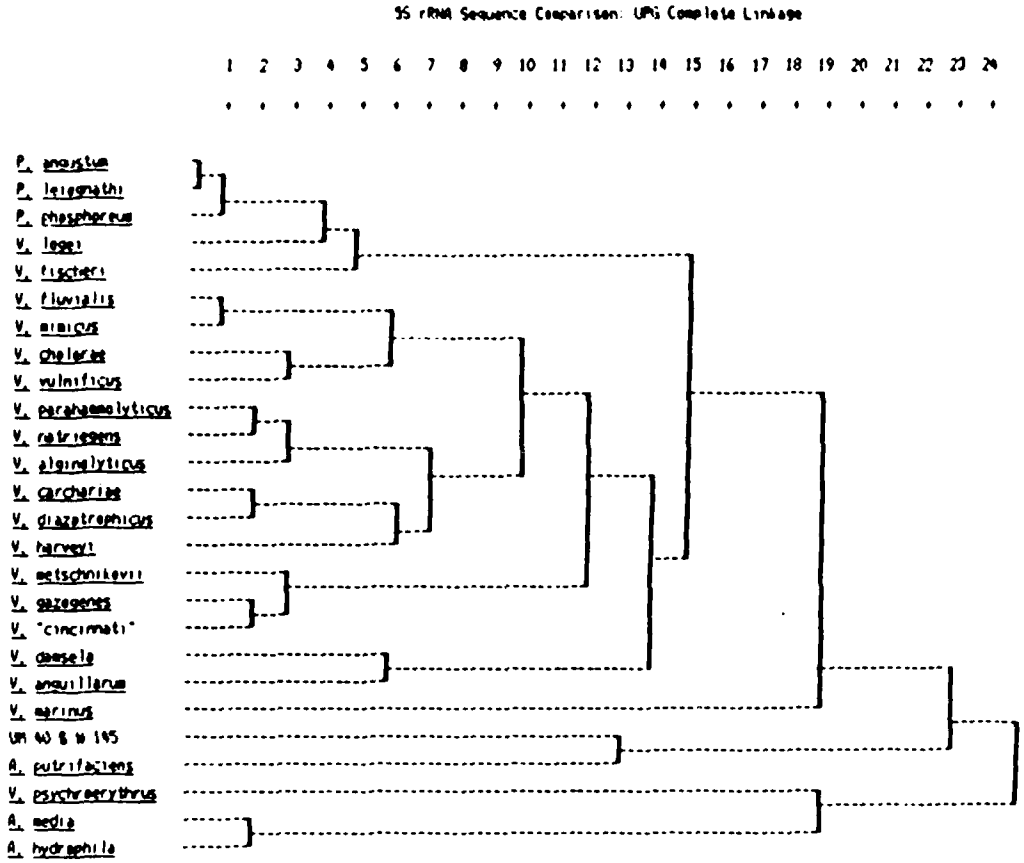
5a. The UPG single linkage dendrogram (*i.e.*, evolutionary tree) resulting from the clustering of species on a basis of overall 5S rRNA sequence homology. Numbers denote difference matrix elements, indicating the total number of base differences out of 120 (the nucleotide base length of 5S rRNA).


5S rRNA Sequence Comparison: UPG Single Linkage
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
P. angustum
P. leiognathi
P. phosphoreum
V. logei
V. fischeri
V. vulnificus
V. cholerae
V. metschnikovii
V. fluvialis
V. mimicus
V. parahaemolyticus
V. natriegens
V. gazogenes
V. "cincinnati"
V. alginolyticus
V. carchariae
V. diazotrophicus
V. harveyi
V. damsela
V. anguillarum
V. marinus
UH 40 & W 145
A. putrefaciens
V. psychroerythrus
A. media
A. hydrophila

5b. The UPGMA, or UPG average linkage, dendrogram resulting from the clustering of species on a basis of overall 5S rRNA sequence homology.


5S rRNA Sequence Comparison. UPGMA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
P. angustum
P. leiognathi
P. phosphoreum
V. logei
V. fischeri
V. parahaemolyticus
V. natriegens
V. alginolyticus
V. fluvialis
V. sinaicus
V. metschnikovii
V. gazogenes
V. "cincinnati"
V. carchariae
V. diazotrophicus
V. harveyi
V. cholerae
V. vulnificus
V. damsela
V. anguillarum
V. ordalii
UM 40 & M 145
A. nutrifaciens
V. psychroerythrus
A. media
A. hydrophila

5c. The UPG complete linkage dendrogram resulting from the clustering of species on a basis of overall 5S rRNA sequence homology.


5S rRNA Sequence Comparison: UPG Complete Linkage
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
P. angustum
P. leiognathi
P. phosphoreum
V. logei
V. fischeri
V. fluvialis
V. mimicus
V. cholerae
V. vulnificus
V. parahaemolyticus
V. natriegens
V. alginolyticus
V. carchariae
V. diazotrophicus
V. harveyi
V. metschnikovii
V. gazogenes
V. "cincinnati"
V. damsela
V. anguillarum
V. marinus
UH 40 B w 195
A. putrefaciens
V. psychroerythrus
A. media
A. hydrophila

algorithm, in effect, weights transversions against transitions in nucleotide base mutations in such a way that transversions are penalized. A mutation from purine to pyrimidine, or from pyrimidine to purine constitutes a transversion, and a mutation from one pyrimidine to another, or from one purine to another constitutes a transition. Deletions and insertions are assessed as transversions. The algorithm assigns a coefficient of evolutionary "distance", K(nuc), to each pairwise comparison, and is not dependent on there being an identical, *i.e.*, common ancestral sequence.

Computer programs, written in BASIC language, were developed to compute S-values for sequence comparisons, as well as K(nuc) values (see Appendix A). The UPG difference matrix, and the evolutionary distance, *i.e.*, K(nuc), matrix are shown in figure 6. Results of cluster analyses, were used to construct dendrograms and to detect evolutionary relationships.

1. UPG Analysis

Dendrograms generated from UPG cluster analyses using single-, average-, and complete linkage groups are shown in figure 5a-c, respectively. Dendrograms generated from single linkage clustering (Figure 5a) were helpful only in indicating maximum relatedness. For example, it is apparent that *Vibrio* 5S rRNA sequences are significantly homologous, with very few differences, thus, the majority of the species examined appear to belong to a single large cluster. Both average- and complete linkage clustering, on the other hand, provided additional information concerning relationships among 5S rRNAs of individual species. In both cases, two distinct similarity groups were observed.

Figure 6. UPG and K(nuc) difference matrices.

6a. UPG difference matrix. Numbers indicate the base differences (of 120 total) between the 5S rRNAs from pairs of species. The difference matrix element for each pair of species is located at the intersection of that pair. Asterisks indicate identities.

	PH	LE	LO	FI	PA	NA	AL	FL	MI	ME	GA	CI	CA	DI	HA	CH	VU	DA	AN	MA	UM	PU	PS
PH	0	1	4	5	12	10	9	15	14	12	9	11	13	11	11	12	10	11	12	15	17	15	19
LE		0	3	4	11	9	8	14	13	11	8	10	12	10	10	11	9	10	11	14	16	14	18
LO			0	5	8	8	5	11	10	8	7	9	9	7	9	10	10	9	8	15	19	15	21
FI				0	9	7	6	12	11	9	10	12	10	10	10	9	7	10	11	16	16	12	20
PA					0	2	3	3	4	4	5	6	3	5	7	6	7	10	6	14	21	18	21
NA						0	3	5	4	4	5	7	3	5	5	4	7	10	8	14	19	16	21
AL							0	6	5	3	4	6	4	4	6	7	7	8	7	17	20	16	22
FL								0	1	3	6	4	6	8	10	5	6	11	7	17	24	21	24
MI									0	2	5	3	5	7	9	4	5	10	8	18	23	20	25
ME										0	3	3	5	5	9	6	7	8	6	18	23	19	25
GA											0	2	6	4	10	9	10	7	5	17	20	18	24
CI												0	8	6	12	7	8	9	7	17	22	20	26
CA													0	2	4	7	8	10	9	17	22	19	22
DI														0	6	9	10	8	7	19	22	18	24
HA															0	9	9	13	13	17	22	19	19
CH																0	3	14	12	14	19	17	23
VU																	0	11	13	17	19	16	21
DA																		0	6	19	19	16	22
AN																			0	17	21	17	23
MA																				0	20	18	24
UM																					0	13	26
PU																						0	23

KEY: PH = *P. phosphoreum*, LE = *P. leiognathi (P. angustum)*, LO = *P. logei*, FI = *V. fischeri*, PA = *V. parahaemolyticus*, NA = *V. natriegens*, AL = *V. alginolyticus*, FL = *V. flavialis*, MI = *V. nimicus*, ME = *V. metschnikovii*, GA = *V. gazogenes*, CI = *V.* 'cincinnati', CA = *V. carchariae*, DI = *V. diazotrophicus*, HA = *V. harveyi*, CH = *V. cholerae*, VU = *V. vulnificus*, DA = *V. damsela*, AN = *V. anguillarum*, MA = *V. marinus*, UM = strains UM 40 and M 145, PU = *A. putrefaciens*, PS = *P. shigelloides*

6b. K(nuc), or evolutionary distance, difference matrix. Values at the intersections of pairs of species indicate the "evolutionary distance" between those species (Kimura, 1980). K(nuc) values are listed as factors of 10^{-4} to facilitate data handling. The K(nuc) value corresponding to the evolutionary distance between *P. leiognathi* and *P. phosphoreum*, therefore, should be read as .0031.

	PH	LE	LO	FI	PA	NA	AL	FL	MI	ME	GA	CI	CA	DI	HA	CH	VU	DA	AN	MA	UM	PU	PS	MD	HY
PH	▮	31	128	129	364	305	270	436	411	351	270	316	108	353	340	343	287	308	329	428	484	421	519	491	557
LE		▮	95	96	328	270	236	400	375	316	236	282	376	318	305	308	252	273	294	392	447	385	482	455	520
LO			▮	128	227	236	138	297	273	216	204	250	273	216	270	273	284	239	196	428	555	417	590	487	552
FI				▮	270	214	181	340	316	289	293	340	318	305	293	250	195	263	284	454	458	324	548	391	455
PA					▮	52	84	64	86	96	129	174	85	140	192	150	193	287	162	415	639	523	615	588	654
NA						▮	94	117	95	106	138	183	95	149	137	95	138	296	216	403	573	461	604	523	588
AL							▮	15	128	74	106	150	128	116	170	192	203	227	183	510	612	472	643	486	550
FL								▮	21	74	151	106	151	207	261	128	171	311	184	491	722	604	700	671	740
MI									▮	53	129	85	129	184	237	106	149	287	207	517	694	577	730	643	711
ME										▮	74	74	139	129	248	160	204	229	151	528	706	561	740	577	643
GA											▮	42	172	117	282	239	284	218	139	491	612	548	700	615	682
CI												▮	218	162	329	193	238	263	184	544	666	600	761	671	740
CA													▮	53	105	193	238	311	252	514	692	573	666	586	652
DI														▮	159	250	296	253	195	579	706	561	734	573	639
HA															▮	237	248	387	364	525	666	550	561	561	627
CH																▮	85	403	321	421	621	480	658	544	610
VU																	▮	308	634	491	560	447	595	511	577
DA																		▮	160	539	555	465	638	480	544
AN																			▮	480	627	486	661	556	622
MA																				▮	571	506	704	735	733
UM																					▮	371	766	755	797
PU																						▮	646	529	595
PS																							▮	489	554
MD																								▮	53

KEY: PH = *P. phosphoreum*, LE = *P. leiognathi (P. angustum)*, LO = *P. logei*, FI = *V. fischeri*,
PA = *V. parahaemolyticus*, NA = *V. natriegens*, AL = *V. alginolyticus*, FL = *V. fluvialis*,
MI = *V. mimicus*, ME = *V. metschnikovii*, GA = *V. gazogenes*, CI= *V. 'cincinnati'*
CA = *V. carchariae*, DI = *V. diazotrophicus*, HA = *V. harveyi*, CH = *V. cholerae*,
VU = *V. vulnificus*, DA = *V. damsela*, AN = *V. anguillarum*, MA = *V. marinus*,
UM = strains UM 40 and W 145, PU = *A. putrefaciens*, PS = *P. shigelloides*, MD = *A. media*
HY = *B. hydrophila*

The first comprised *V. parahaemolyticus*, *V. natriegens*, *V. alginolyticus*, *V. fluvialis*, *V. metschnikovii*, *V. mimicus*, *V. gazogenes*, *V. vulnificus*, *V. cholerae*, *V. carchariae*, *V. diazotrophicus*, *V. harveyi*, and a new *Vibrio* species tentatively designated *V. cincinnatii*. The second cluster comprised the *Photobacterium* species: *P. phosphoreum*, *P. leiognathi*, *P. angustum*, and *V. logei*, and "*V.*" *fischeri* (Table 3). Although the generic composition of the clusters was identical, intra-phenetic differences were observed. For example, the *V. fluvialis* - *V. mimicus* and *V. cholerae* - *V. vulnificus* doublets clustered by complete linkage, whereas by average linkage, they appeared to be more distantly related, not only to each other, but also to the central *Vibrio* cluster.

The second major cluster, containing the *Photobacterium* species, did not change with linkage method employed. Although *V. damsela* and *V. anguillarum* formed a doublet related to the main *Vibrio* cluster by average linkage clustering (Figure 5b), by both single- and complete linkage, they represented a separate cluster. *A. media* and *A. hydrophila* were only distantly related to the other species of the *Vibrionaceae* by cluster analysis, regardless of linkage method. By average- and complete linkage, however, *V. psychroerythrus* clustered with *A. media* and *A. hydrophila*, but at a relatively low similarity value.

2. Evolutionary Distance

The dendrogram generated from the cluster analysis using the Kimura algorithm for estimation of evolutionary distance (Figure 7), bears a striking resemblance to that generated by UPGMA (average linkage, figure

Table 5. Palindrome Relatedness Groups

Palindrome:	GCGAUUUAG------CUGACU-C-A-GCCG-ACUCAGUC-------GAUUUAGCG
V. gazogenes	GCG UUU G CUGA U C A GCCG ACUCAGU GA AGCG
V. cincinnati	GCG UUU G CUGA U C A GCCG ACUCAGU GA AGCG
V. sinicus	GCG UUU G CUGA U C A CCG ACUCAG GA AGCG
V. fluvialis	G G UUU G CUGA U C A CCG ACUCAG GA AGCG
V. diazotrophicus	GCGAUUU G CUGA U C A GCCG ACUCAG GA UUAGCG
V. carchariae	GCGAUUU G CUGA U C A CCG ACUCAG GA UUAGCG
V. sinicus	GCG UUU G CUGA U C A CCG ACUCAG GA AGCG
V. fluvialis	G G UUU G CUGA U C A CCG ACUCAG GA AGCG
V. diazotrophicus	GCGAUUU G CUGA U C A GCCG ACUCAG GA UUAGCG
V. carchariae	GCGAUUU G CUGA U C A CCG ACUCAG GA UUAGCG
V. parahaemolyticus	G G UUU G CUGA U C A CCG ACUCAG GA UAGCG
V. fluvialis	G G UUU G CUGA U C A CCG ACUCAG GA AGCG
V. diazotrophicus	GCGAUUU G CUGA U C A GCCG ACUCAG GA UUAGCG
V. alginolyticus	GCG UUU G CUGA U C A GCCG ACUCAG GA UAGCG
P. logei	GCG U U G CUGA U C A GCCG ACUCAG G UAGCG
A. hydrophila	GCG U G CUGA U C A GCCG ACUCAG G UAGCG
V. diazotrophicus	GCGAUUU G CUGA U C A GCCG ACUCAG GA UUAGCG
V. alginolyticus	GCG UUU G CUGA U C A GCCG ACUCAG GA UAGCG
P. logei	GCG U U G CUGA U C A GCCG ACUCAG G UAGCG
V. damsela	GCG U U G CUGA U C A GCCG ACUCAG G AGCG
V. anguillarum	G G U U G CUGA U C A GCCG ACUCAG GCG
V. diazotrophicus	GCGAUUU G CUGA U C A GCCG ACUCAG GA UUAGCG
V. alginolyticus	GCG UUU G CUGA U C A GCCG ACUCAG GA UAGCG
P. logei	GCG U U G CUGA U C A GCCG ACUCAG G UAGCG
V. fischeri	GC U U G CUGA U C A GCCG ACUCAG G UAGCG
A. putrefaciens	GC U U G CUGA C A CCG ACUCAG G U GCG
V. diazotrophicus	GCGAUUU G CUGA U C A GCCG ACUCAG GA UUAGCG
V. alginolyticus	GCG UUU G CUGA U C A GCCG ACUCAG GA UAGCG
V. vulnificus	GCG UUU G CUGA C A CCG ACUCAG GA UAGCG
A. putrefaciens	GC U U G CUGA C A CCG ACUCAG G U GCG
V. diazotrophicus	GCGAUUU G CUGA U C A GCCG ACUCAG GA UUAGCG
V. alginolyticus	GCG UUU G CUGA U C A GCCG ACUCAG GA UAGCG
V. sinicus	GCG UUU G CUGA U C A CCG ACUCAG GA AGCG
V. fluvialis	G G UUU G CUGA U C A CCG ACUCAG GA AGCG

V. logei GCG-U-U-G------CUGA-U-C-A-GCCG-ACUCAG------G---UAGCG

V. alginolyticus GCG-UUU-G------CUGA-U-C-A-GCCG-ACUCAG------GA--UAGCG

V. anguillarum G-G-U-U-G------CUGA-U-C-A-GCCG-ACUCAG----------GCG

V. damsela GCG-U-U-G------CUGA-U-C-A-GCCG-ACUCAG------G----AGCG

V. diazotrophicus GCGAUUU-G------CUGA-U-C-A-GCCG-ACUCAG------GA-UUAGCG

Visual inspection of these sequences suggested the possibility that base sequences from certain species of the *Vibrionaceae*, representing this palindromic region, might be arranged in such a way that a sequential drift in base sequence could be detected. The five species of the above example are arranged thus:

V. diazotrophicus GCGAUUU-G------CUGA-U-C-A-GCCG-ACUCAG------GA-UUAGCG

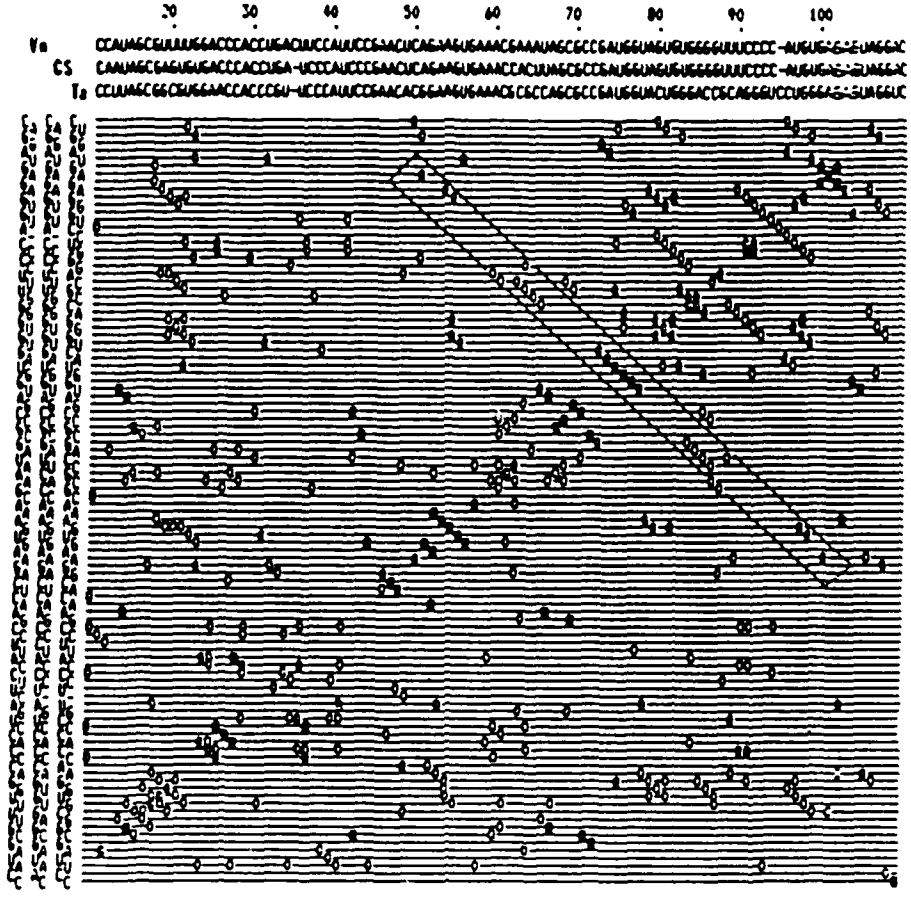
V. alginolyticus GCG-UUU-G------CUGA-U-C-A-GCCG-ACUCAG------GA--UAGCG

V. logei GCG-U-U-G------CUGA-U-C-A-GCCG-ACUCAG------G---UAGCG

V. damsela GCG-U-U-G------CUGA-U-C-A-GCCG-ACUCAG------G----AGCG

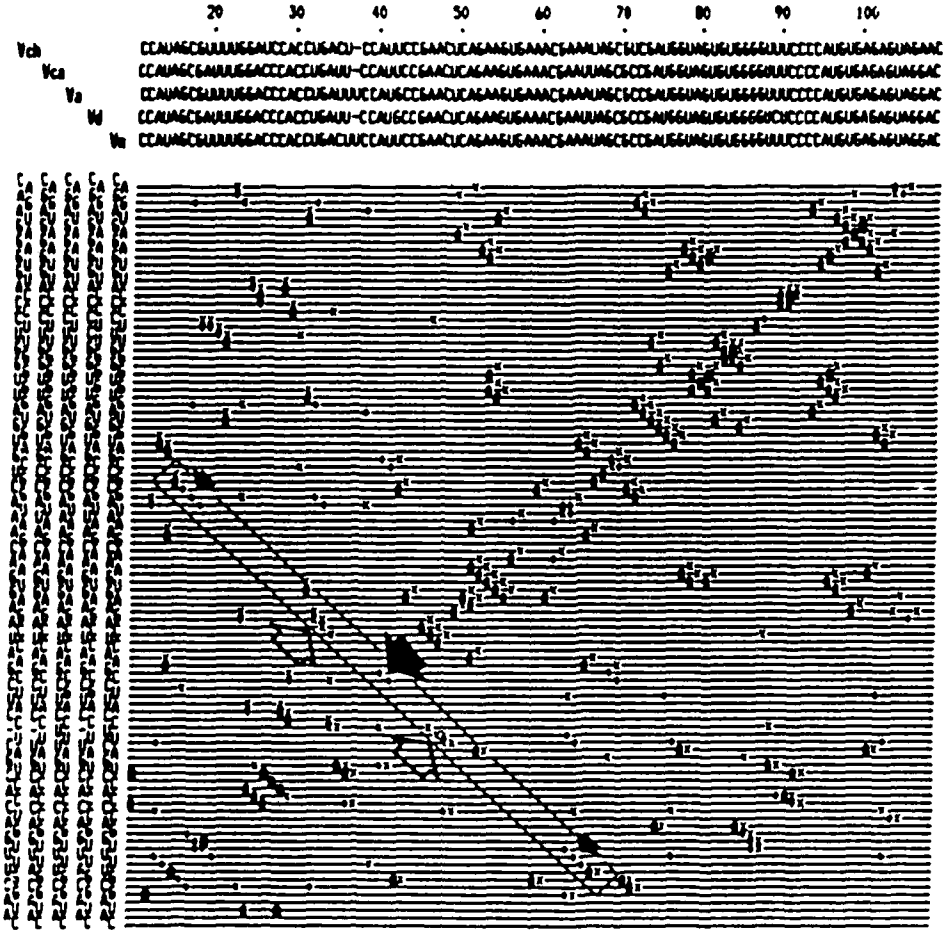
V. anguillarum G-G-U-U-G------CUGA-U-C-A-GCCG-ACUCAG----------GCG

When the same approach to arrangement of sequences of the putative 54-base palindrome was applied to all the *Vibrionaceae* 5S rRNA sequences, eight groupings, all of which appear to suggest evolutionary relationships among component species were obtained (Table 5). The relationships implied in the groupings suggest that a single, common, simultaneous solution to the phylogeny of the eight species can be portrayed as an evolutionary tree. See figure 9.


20
30
40
50
60
70
80
90
100

8c. Composite dot matrix map. The boxed-in region indicates the position of a degenerated 52-base palindrome, portions or which are contributed by species from several RNA superfamilies.

Key: Vn = *V. natriegens*, CS = *Calyptogena* symbiont (Stahl *et al.*, 1984), Ta = *Thermus aquaticus* (Dams *et al.*, 1983). Frequency of occurrence: (open square) = 1, (circle) = 2, (solid square) = 3.


20
30
40
50
60
70
80
90
100
Vch
Vca
Va
Vd
Vu

8b. This is a composite dot matrix in which the palindromic sequences from 5 *Vibrio* species have been "overlayed" by use of a computer, and indicates the existence of a degenerated 54-base palindrome (boxed-in region). Although the palidrome is not evident in the sequence of any single *Vibrio* species, composites of numerous combinations of sequences from *Vibrio* species yield the same result. The small solid arrows indicate sequences common to *V. carchariae* and *V. diazotrophicus*. The large solid arrow indicates sequences common to *V. diazotrophicus* and *V. alginolyticus*. Open arrows indicate the portion of the palindrome contributed by *V. cholerae*.

Key: Vch = *V. cholerae*, Vca = *V. carchariae*, Va = *V. alginolyticus*, Vd = *V. diazotrophicus*, Vn = *V. natriegens*. Frequency of occurrence: x = 1, (open diamond) = 2, (solid diamond) = 3, (circle) = 4, (solid square) = 5.

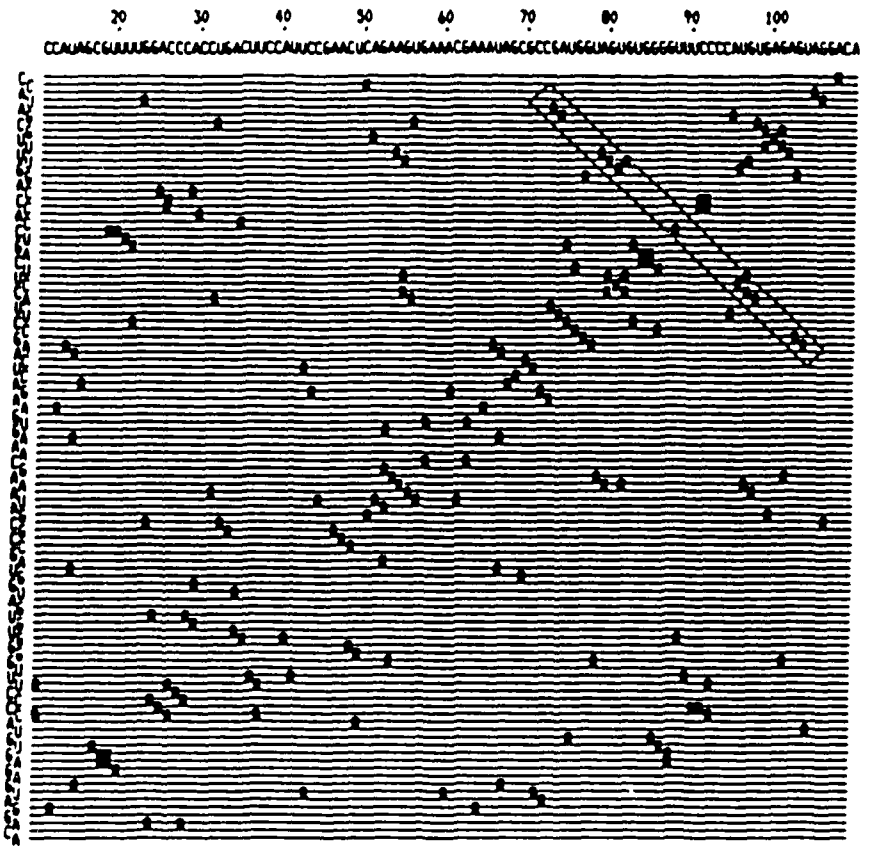

20
30
40
50
60
70
80
90
100

Figure 8. Dot matrix analysis of inverted repeats (palindromes). Dot matrices were generated by comparing truncated sequences, consisting of bases 11 to 110 (of 120) of 5S rRNAs arranged 5' to 3' versus 3' to 5'. The search algorithm was set to ignore matches less than 3 consecutive bases in length.

8a. This pattern (boxed-in region), is typical of all of the 5S rRNAs of the genus *Vibrio sensu stricto*, and indicates the existence of a degenerated 33-base inverted repeat (palindrome). The sequence shown here is from *V. vulnificus*.

D. REPEATED AND PALINDROMIC SEQUENCES

A computer program was developed (MacDonell and Colwell, 1984, manuscript in review) which searches nucleic acid sequences for repeated and palindromic sequences, in order to measure the frequency of occurrence and extent of conservation. Data for 5S rRNA sequences determined in this study and for those reported in the literature (Erdmann *et al.*, 1984) were analyzed by the program. Results of the analyses revealed several palindromes (inverted repeats). Repeated sequences, however, appear to occur infrequently, and are limited to regions of <10 bases. Dot matrices with palindromic or repeated sequences are shown in figure 8. Two regions in the base sequences of 5S rRNAs from species of the Vibrionaceae appear to have been derived from palindromes. These are (1) a 54-base region extending from G_{14} through G_{70}:

GCGAUU--G------CUGACU-C-A-GCCG-ACUCAGUC------G--UUAGCG

and (2) a 33-base region extending from G_{73} through G_{104}:

GAUG--AGUGU----UUU----UGUGA--GUAG

which appears to have derived from two smaller palindromes:

GAUGGUAG------------------------

and

--------------------AUGUGAGAGUA-

Although the sequence of the 33-base palindrome is stable in all *Vibrio* 5S rRNAs examined to date, slight differences in the extent of conservation of the sequence of the 54-base palindrome were detected by comparison of the 5S rRNA sequences. Five species thusly compared are listed below.

Table 4. Relative Stability of Helical and Single-stranded Regions

	Frequency of mutation per 100 bases
Helix[1]	
A	5.2
B	12
C	0
C'	2.4
B'	11.2
E	4.3
D	2.7
D'	3.3
E'	4.3
A'	4
All helices	5.3
Loop[1]	
aLb	2.6
bLc	10.9
cLc'	15.6
c'Lb'	0
b'Le	0
eLd	0
dLd'	14.5
d'Le'	0
e'La'	7.2
All Loops	10.9

Key: [1] Refer to Figure 3

5b). The single difference in the clustering of component phena is the location of the *V. cholerae - V. vulnificus* cluster. The location of the "*V. cholerae - V. vulnificus*" doublet in the Kimura dendrogram more closely resembles UPG single linkage clustering (Figure 5a). The composition and relative locations of the *Photobacterium* cluster, the *A. media - A. hydrophila - V. psychroerythrus* cluster, and the *V. damsela - V. anguillarum* doublet are identical to that predicted by UPGMA analysis.

C. CONSERVED AND HYPERVARIABLE REGIONS

As suggested by Fox *et al.* (1977), certain regions of the 5S rRNA molecule are characterized by a high degree of sequence variation (hypervariable regions), whereas other regions were apparently resistant to mutation. Locations of the hypervariable regions, detected by analysis of the 26 5S rRNA sequences determined in this study are as follows (listed in order of increasing stability): helix B/B', the base 40 - 48 region of loop cLc', loop e'La', base pair 83/130, and to much lesser degrees, helix D/D' and loop dLd' (Figure 3). Highly conserved regions in the sequences of 5S rRNAs were detected at (in order of increasing variability) region C of helix C/C', base 49 - 58 region overlapping loop cLc' and helix C/C', base 87 - 100 region overlapping helices D/D' and E/E', and terminal helix A/A' (Figure 3). The relative stabilities of base sequences associated with these regions, expressed as percent stability, are listed in Table 4.

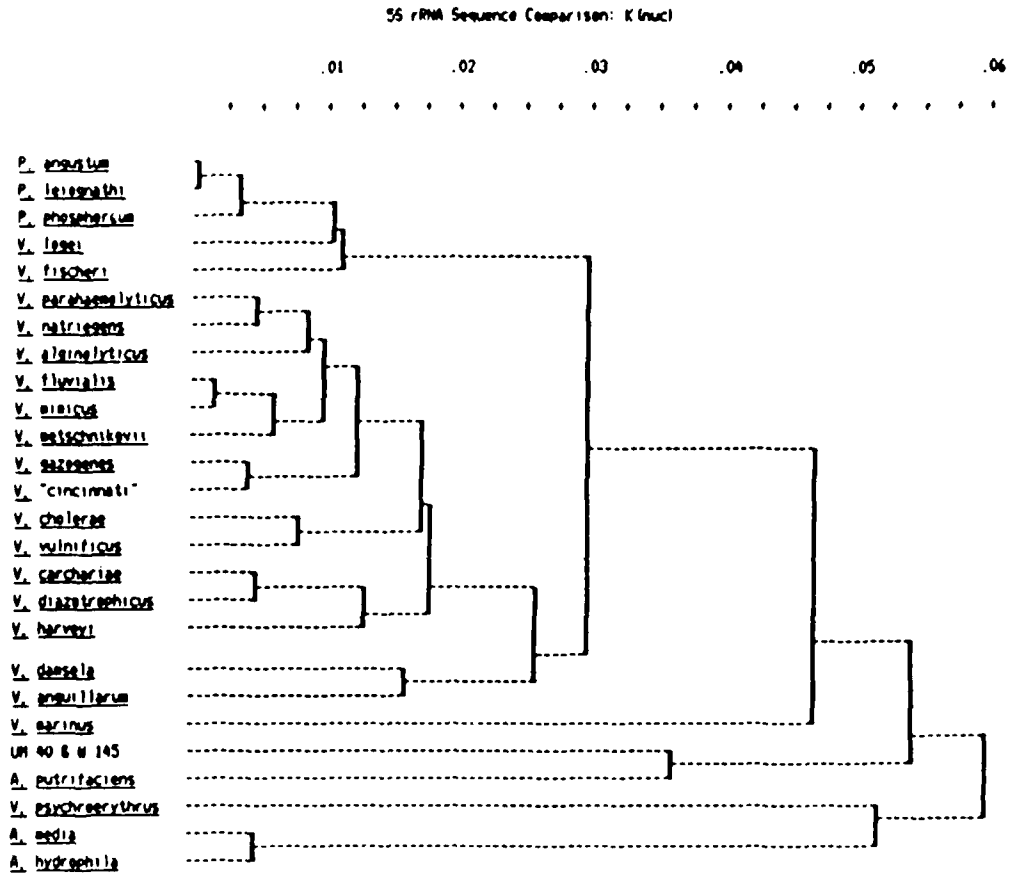

5S rRNA Sequence Comparison: K (nuc)
.01
.02
.03
.04
.05
.06
P. angustum
P. leiognathi
P. phosphoreum
V. logei
V. fischeri
V. parahaemolyticus
V. natriegens
V. alginolyticus
V. fluvialis
V. mimicus
V. metschnikovii
V. gazogenes
V. "cincinnati"
V. cholerae
V. vulnificus
V. carchariae
V. diazotrophicus
V. harveyi
V. damsela
V. anguillarum
V. salmonicida
UM 40 & M 145
A. putrefaciens
V. ichthyoenteri
A. media
A. hydrophila

Figure 7. **Evolutionary distance (Knuc) dendrogram.** This dendrogram (*i.e.*, evolutionary tree) was derived by the WPG method from K(nuc) values computed using the algorithm of Kimura (1980), listed in figure 6b. Refer to the text for an explanation of the K(nuc) coefficient.

Table 3. Genera *sensu strictu*: *Vibrio* and *Photobacterium*

Vibrio sensu strictu[1]	*Photobacterium sensu strictu*[1]
V. alginolyticus	*P. angustum*[2]
V. carchariae	*P. fischeri*
V. cholerae	*P. leiognathi*
V. "cincinnatii"	*P. logei*
V. diazotrophicus	*P. phosphoreum*
V. fluvialis	
V. gazogenes	
V. harveyi	
V. metschnikovii	
V. mimicus	
V. natriegens	
V. parahaemolyticus	
V. vulnificus	

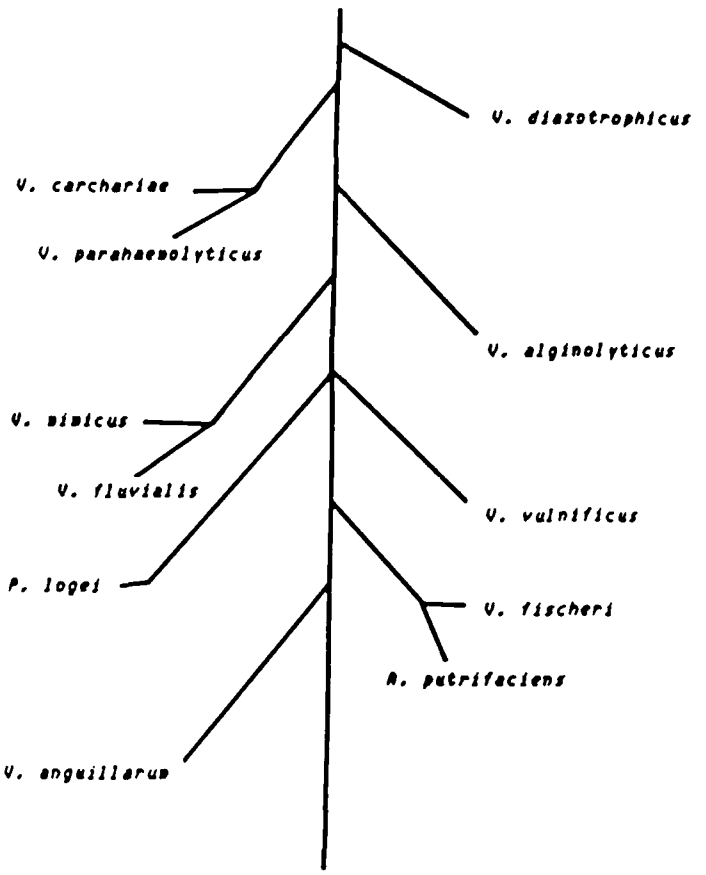
Vibrio species not fitting into genus *Vibrio*[3]

V. anguillarum
V. damsela
V. fischeri (see *Photobacterium*)
V. logei (see *Photobacterium*)
V. marinus
V. psychroerythrus (see Results)

Key:

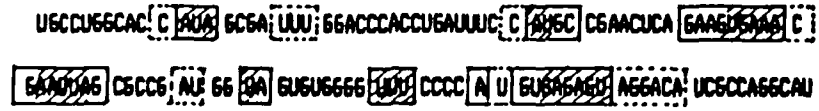
[1] On a basis of 5S rRNA phylogeny
[2] Possibly a synonym or biovar of *P. leiognathi* (see Results)
[3] The following named *Vibrio* species have not yet been evaluated:
V. campbelli, *V. costicola*, *V. furnisii*, *V. hollisae*, *V. nereis*, *V. nigrapulchritudo*, *V. ordalii*, *V. orientalis*, *V. pelagius*, *V. proteolyticus*, *V. splendidus*

Figure 9. Evolutionary tree based on palindromic sequence analysis. This dendrogram is the graphic representation of one simultaneous solution for the eight groups of "sequence drift" observed for the 54-base palindrome (see Table 5). Since the 54-base palindrome is most conserved in the 5S rRNA from *V. diazotrophicus* and most randomized in that from *A.hydrophila*, it is suggested that *V. diazotrophicus* may represent the least highly evolved of the *Vibrio* species. Three bifurcations are evident, indicating the existence of "intermediate" common ancestors, suggesting a high level of relatedness between *V. carchariae* and *V. parahaemolyticus*; *V. fischeri* and *A. putrifaciens*; and *V. nigicus* and *V. fluvialis*.


V. diazotrophicus
V. carchariae
V. parahaemolyticus
V. alginolyticus
V. nimicus
V. fluvialis
V. vulnificus
P. logei
V. fischeri
A. putrifaciens
V. anguillarum
A. hydrophila

E. NUCLEASE S1 LIMITED DIGESTS

Nuclease S1 (from *Aspergillus oryzae*) preferentially hydrolyzes single-stranded (ss) regions in nucleic acids, with relative affinities of approximately 2000:1 (ss:ds) compared with double-stranded regions (Vogt, 1980). Limited, *i.e.*, "single-hit" digests of ^{32}P-5S rRNA, using nuclease S1, were separated on thin sequencing gels alongside conventional sequencing ladders in order to map the locations of nucleotide bases participating in the formation of single-stranded regions (Figure 10). Using this method, composite maps were generated for 5S rRNAs from *P. shigelloides*, *V. alginolyticus*, *V. diazotrophicus*, *V. natriegens*, and *V. psychroerythrus*. Based on these data, the following consensus map was constructed (using the *V. diazotrophicus* sequence):



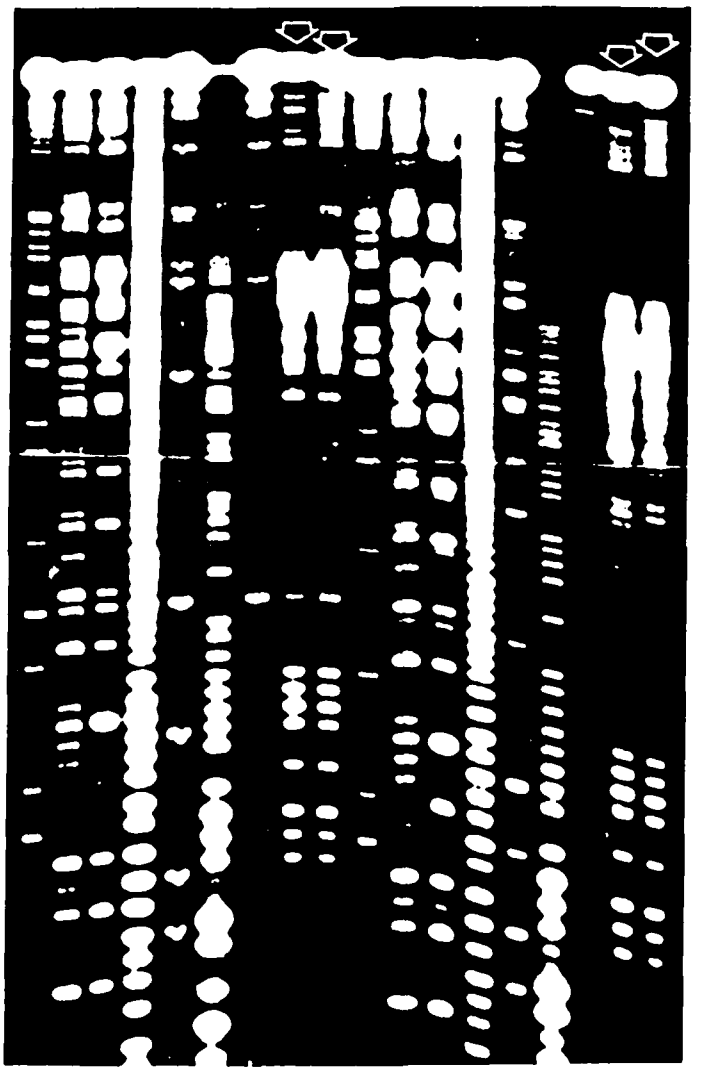
Where broken boxes identify regions which were occasionally clipped (frequency < 50%); solid boxes identify regions which were frequently clipped (frequency > 50%); and hatched areas identify regions which were invariably clipped (frequency = 100%).

No significant differences in hydrolysis patterns were detected over a wide range of enzyme-to-substrate ratios (10^2-fold range); incubation times (30 seconds to 6 minutes); incubation temperatures (10 to 35 C); or 5S rRNA species employed as a substrate.

F. FREQUENCY OF OCCURRENCE OF "STOP" AND "RNY" CODONS

Since the statistical distribution of "stop" (UAA, UAG, UGA), and 'RNY" (purine, any base, pyrimidine) codons in a reading frame is

Figure 10. **Nuclease S1 hydrolysis ladders.** Limited nuclease S1 digests were run alongside sequence ladders in order to indicate the locations of single-stranded regions. Nuclease S1 and enzymatic sequence ladders were generated from limited digests of 5S rRNA from *V. logei* and *P. leiognathi*. Identity of the endonuclease lanes are as indicated in figure 2. Arrows mark the positions of lanes resulting from limited S1 digestion. See text for composite S1 digest map.



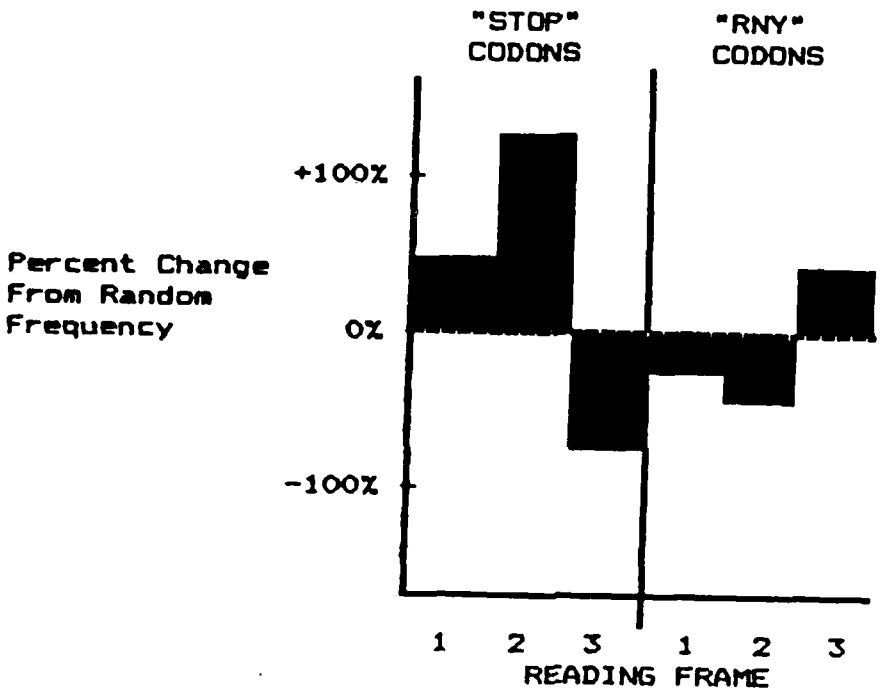
diagnostic for the existence of a direct coding function (Shepherd, 1981; 1982), a computer program was written (appendix A) to search each reading frame, in turn, and compile the frequency of occurrence of "STOP" and "RNY" for each. Results obtained from application of the search algorithm to 5S rRNA sequences determined in this study are presented in Table 6. Of the 28 species of the *Vibrionaceae* for which 5S rRNA sequences have been determined, only *A. hydrophila*, *A. media*, and *V. psychroerythrus* had at least one stop codon in each reading frame. The average frequency of occurrence of "stop" codons in each of the three reading frames was 6.8% (frame 1), 10.8% (frame 2), and 1.45% (frame 3). The expected frequency, based on random occurrence, is $3/4^3$ or 4.7%. The average frequency of occurrence of "RNY" codons in each reading frame was 22.6% (frame 1), 18.7% (frame 2), and 34.9% (frame 3). The probability of occurrence of "RNY" codons in a random sequence is $1/2^2$ or 25%. By comparing each observed frequency with that expected from a randon sequence, percent change (Δ%) values for each reading frame were determined. These are presented in the form of a histogram in Figure 11.

Table 6. Occurrence of "RNY" and "STOP" Codons in 5S rRNA Reading Frames

	"STOP" CODONS ----FRAME----			"STOP" FREQ. ----FRAME----			"RNY" CODONS ----FRAME----			"RNY" FREQ. ----FRAME----		
	1	2	3	1	2	3	1	2	3	1	2	3
A. hydrophila	1	4	1	2.6	10.3	2.6	7	7	16	18	18	41
A. media	1	4	2	2.6	10.3	5.1	7	7	16	18	18	41
A. putrefaciens	3	5	0	7.7	12.8	0	9	7	15	23	18	38.5
ANT 300	1	3	3	2.6	7.7	7.7	11	8	12	28	20.5	31
P. leiognathi	4	5	0	10.3	12.8	0	9	8	13	23	20.5	33
P. angustum	4	5	0	10.3	12.8	0	9	8	13	23	20.5	33
P. shigelloides	3	4	0	7.7	10.3	0	8	9	12	20.5	23	31
V. alginolyticus	0	5	2	0	12.8	5.1	10	8	11	25.6	20.5	28
V. anguillarum	3	3	0	7.7	7.7	0	9	6	13	23	15.4	33
V. carchariae	3	4	0	7.7	10.3	0	9	7	15	23	18	38.5
V. cholerae	3	4	0	7.7	10.3	0	9	6	15	23	15.4	38.5
V. "cincinnati"	4	3	0	10.3	7.7	0	9	8	13	23	20.5	33
V. damsela												
V. diazotrophicus	3	4	0	7.7	10.3	0	10	8	13	25.6	20.5	33
V. fischeri	3	5	0	7.7	12.8	0	9	7	13	23	18	33
V. fluvialis	3	3	0	7.7	7.7	0	9	6	14	23	15.4	36
V. gazogenes	4	3	0	10.3	7.7	0	9	8	13	23	20.5	33
V. logei	3	5	0	7.7	12.8	0	9	7	13	23	18	33
V. marinus	4	6	0	10.3	15.4	0	6	8	16	15.4	20.5	41
V. metschnikovii												
V. nimicus	3	3	0	7.7	7.7	0	9	6	14	23	15.4	36
V. natriegens	0	5	2	0	12.8	5.1	10	8	11	25.6	20.5	28
V. parahaemolyticus	0	5	2	0	12.8	5.1	9	8	14	23	20.5	36
V. psychroerythrus	5	5	1	12.8	12.8	2.6	8	6	16	20.5	15.4	41
V. vulnificus	3	4	0	7.7	10.3	0	9	7	13	23	18	33

Values expected for random sequences: "STOP", 4.6% (1.8/molecule); "RNY", 25% (9.75/molecule).

Figure 11. Histogram of "RNY" and "STOP" codon frequency. The data for occurrence of "RNY" and "STOP" codons (Table 6) are presented as a percent of expected random frequency. Reading frames 1 (starting with the 5' terminus) and 2 both exhibit (1) an increased frequency of "STOP" codons and (2) a decreased frequency of "RNY" codons. Reading frame 3, however, contains almost no "STOP" codons, and a significantly increased number of "RNY" codons, suggesting a conserved direct coding function.


"STOP"
CODONS
"RNY"
CODONS
+100%
0%
-100%
Percent Change
From Random
Frequency
1 2 3 1 2 3
READING FRAME

DISCUSSION

A. MOLECULAR BASIS OF PHYLOGENETIC INFERENCE

A logical interpretation of the phylogenetic data presented here requires the following assumptions: that prokaryotic species are monophyletic, and that eubacterial species share a common ancestor. Both are reasonable assumptions since there are no data available suggesting otherwise. The first assumption, that prokaryotes are monophyletic, suggests that prokaryotes, unlike eukaryotes, evolved as cells, and not assemblages of organelles or sub-cellular components with independent phylogenetic histories. Therefore, phylogeny of a part, the ribosome for example, should reflect the phylogeny of the whole, or that the phylogeny of bacterial ribosomal RNA is equivalent to the phylogeny of the bacterial cell. Although true for prokaryotes, this is not believed to be the case for eukaryotes, since there is ample evidence suggesting that chloroplasts, mitochondria, and nuclei all are of distinct and independent prokaryotic phylogenies.

The second assumption, in part related to the first, suggests that all prokaryotic species share a common origin. Although difficult to prove, there is evidence to support this view. For example, only a single kind of protein synthesis apparatus occurs in all known forms of life (Vogel, *et al.*, 1984). Furthermore, tRNAs and ribosomal RNAs from the broadest possible range of species share striking similarities, strongly suggesting a common origin. Since the sequences can be considered highly conserved over a time frame of more than a billion

years, nucleotide base sequences of ribosomal RNAs provide the information from which evolutionary relationships can be deduced. Comparisons among sequences of ribosomal RNAs appear to be ideal for inference of phylogenetic relationships (see Kuntzel *et al.*, 1981; Kuntzel, 1982; Kuntzel *et al.*, 1983). Unfortunately, the number of known 5S rRNA sequences is small, with only 43 eubacterial sequences published to date (Erdmann *et al.*, 1984; Dekio *et al.*, 1984; MacDonell and Colwell, 1984a,b,d,e), including 21 Gram-negative species. This is too few to establish, conclusively, a phylogeny for eubacteria.

A major reason for the small number of eubacterial 5S rRNA sequences presently available for analysis is that more emphasis has been placed on investigation of deep phylogenetic branching than of phylogenetic relationships at the genus and species level. The focal point has been, primarily, the saltation of archaebacteria and eubacteria, and deep branches within each group (Stackebrandt and Woese, 1984). In addition, several laboratories undertaking pioneering work on comparative sequencing of ribosomal RNA as a phylogenetic probe have focussed on characterization of archaebacterial species (see Fox *et al.*, 1982; Woese, 1982). Clearly, methods for comparative sequencing of ribosomal RNAs are just now being developed. It is too soon to expect large compilations of rRNA sequences, but the situation will alter in the immediate future.

The main objective of this study was to analyze relationships among species of a single taxon of the eubacteria, the family *Vibrionaceae*, and, thereby, provide information needed for resolution of several issues. The first, and most immediate, was the clarification of

taxonomic relationships among the 36 species of the *Vibrionaceae* and assessment of the utility of the polyphasic approach to bacterial systematics, upon whose foundations the taxonomy of the *Vibrionaceae* (Citarella and Colwell, 1970; Colwell, 1970; Colwell, 1971) was constructed. The second was to evaluate the ability of comparative sequence analysis to clarify phylogenetic relationships at the species level. Lastly, and most important, considering the impact on bacterial systematics, was to provide a basis for selection of phenotypic characters correlated with phylogeny for the *Vibrionaceae*, the purpose being replacement of tables of key characteristics constructed from *a priori* assumptions of phylogenetic relationships.

Several methods exist for estimation of evolutionary relationships among nucleic acid sequences, based on sequence similarities (Klotz *et al.*, 1979; Kimura, 1980; Li, 1981). These derive, for the most part, from modifications of clustering of difference matrix data by unweighted pair-group (UPG) algorithms. The modifications are, in general, based on empirical or statistical models which correct for the tendency for transitions (purine to purine, or pyrimidine to pyrimidine) to occur with significantly greater frequency than transversions (purine to pyrimidine, or pyrimidine to purine) (Kimura, 1980); different rates of mutation (Klotz *et al.*, 1979; Li, 1981); or compensate for lack of reference ancestral or prototype nucleotide sequences (Kimura, 1980; Li, 1981). Application of the UPG method, despite its inherent inability to compensate for the interactions mentioned above, produced virtually the same clustering (Figure 5b) as the Kimura algorithm (Figure 7), the latter being considered by many investigators as the most appropriate

for nucleic acid sequence comparisons. Either, however, appears to be more favorable for estimation of phylogenic relationships than S_{AB} coefficients (Woese, 1982; Woese *et al.*, 1984; Stackebrandt and Woese, 1984) commonly employed in 16S rRNA studies. A correlation between S_{AB} values and sequence homology has not been shown, since S_{AB} values greater than 0.4 tend to overestimate homologies significantly, whereas values below 0.4 tend to underestimate to the same degree (Hori and Osawa, 1982). The result is an unacceptable skewing of evolutionary distances that are estimated using the S_{AB} coefficient, not altogether surprising, since the S_{AB} was developed to estimate evolutionary distances from oligomer catalogs. Were methods for direct determination of nucleotide base sequences available, a more robust coefficient might have resulted.

B. PHYLOGENY OF THE VIBRIONACEAE

Unweighted pair-group (UPG) analysis, using single-, average-, and complete linkage (Figure 5a-c), as well as estimation of evolutionary distance, employing the Kimura (1980) algorithm and comparative sequence data (Figure 7), were applied to construct evolutionary trees. As expected, the UPG single-, and UPG complete linkage analyses were of limited utility in defining phylogenetic relationships, although helpful, in some cases, such as analysis of relationships between *V. anguillarum* and the major *Vibrio* cluster. Dendrograms generated using 5S rRNA sequence comparisons and UPGMA (average linkage) and using estimates of evolutionary distance (Kimura, 1980) were in remarkably

good agreement, suggesting that most of the named *Vibrio* species share a recent, common ancestor. These species are concluded to comprise the genus *Vibrio sensu strictu* (MacDonell and Colwell, 1984b). *V. anguillarum* and *V. damsela*, and, presumably, *V. ordalii* and *V. tubiashii*, which were considered previously to be biovars of *V. anguillarum* are concluded to comprise a separate genus.

V. marinus, based on the criterion of comparative analysis of 5S rRNA sequences, shares only a remote common ancestor with any of the named *Vibrio* species (MacDonell and Colwell, 1984b), and, as a consequence, the 5S rRNA sequence is concluded not to permit clustering of this species with those of either of the two groups described above. Whether *V. marinus* constitutes a species of the family *Vibrionaceae sensu strictu* requires further study.

Motivated by the observation of Van Landschoot and DeLey (1983), based on rRNA/DNA hybridizations, that *Alteromonas putrifaciens* may share a moderately recent common ancestor with the genus *Vibrio*, the sequence of the 5S rRNA from that species was determined. Although it failed to cluster with species of either the genus *Vibrio sensu strictu* or *V. anguillarum* and *V. damsela*, it shares, with both groups, a level of relatedness sufficiently high to suggest that it probably is a member of the *Vibrionaceae*. Interestingly, *A. putrifaciens* shares a common ancestor with two abyssal marine isolates, UM40 and W145 (Deming *et al.*, 1984) at an phylogenetic depth suggesting a mutual relatedness at approximately the family level.

The genus *Photobacterium* comprises at least four species, *P. phosphoreum*, *P. leiognathi*, *V. fischeri*, and *V. logei*. *P. angustum*

possesses a 5S rRNA the sequence of which is identical to that of *P. leiognathi*, suggesting, but not proving, they may be biovars of the same species. Although there is no reason why two species cannot share an identical 5S rRNA sequence, it appears likely that the time scale in which a common ancestor gives rise to two distinct species is sufficiently large that one (or several) base differences out of the 120 comprising the 5S rRNA molecule would result. The rate of spontaneous mutation in 5S rRNA sequences is in the order of 10^{-9} years (Stahl *et al.*, 1984).

V. psychroerythrus was originally described by D'Aoust and Kushner (1972) with the suggestion that it be allocated to the genus *Vibrio*, although it was never properly validated as a *Vibrio* species. From comparison of its 5S rRNA sequence with those for the other species of the *Vibrionaceae* included in this study, it is concluded to be misnamed. The only species with which it shares other than the most remote common ancestor are *A. hydrophila* and *A. media*, which, based on 5S rRNA sequences, as well as rRNA/DNA hybridizations (J. DeLey, personal communication) cannot be considered members of the *Vibrionaceae*. A proposal to create a new family composed of the species *V. psychroerythrus*, *A. hydrophila*, *A. media* and related species is being prepared (MacDonell and Colwell, manuscript in preparation).

The correct taxonomic position of *P. shigelloides* has been a topic of controversy since the original description was proposed by Fergusen and Henderson (1947), and has been shifted back and forth between the families *Enterobacteriaceae* and *Vibrionaceae* (Hendrie *et al.*, 1971). The reason for the apparent difficulty in resolving the taxonomic

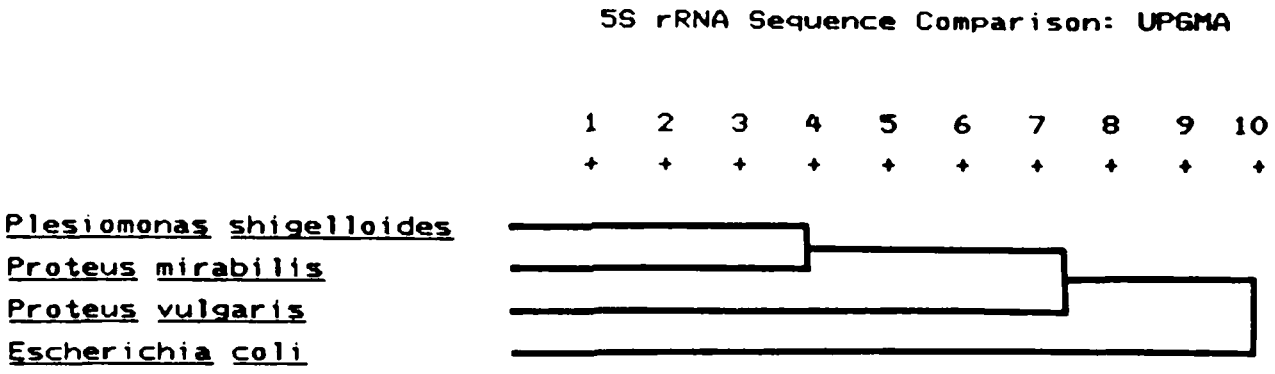
position of this species is that it possesses phenotypic characteristics considered, *a priori*, to be characteristic of each family. The 5S rRNA sequence, however, appears to be useful in resolving the controversy since an extensive homology with the 5S rRNA from *Proteus mirabilis* (*Enterobacteriaceae*) was observed. In fact, it is evident that *Proteus mirabilis* shares with *Plesiomonas shigelloides* a much more recent common ancestor than with *Proteus vulgaris* (figure 12). It is concluded, therefore, that *P. shigelloides* should be reassigned to the *Proteae*.

Compared with the classical, *i.e.*, alpha, taxonomy for other bacterial groups, *viz.*, *Pseudomonadaceae* and *Bacillaceae*, the taxonomy of the *Vibrionaceae*, derived extensively from application of a polyphasic approach, which now includes 5S rRNA, is remarkably consonant in all phases. In fact, had it been known that the DNA base composition, in the range of 45 to 51 mol% GC, is a phylogenetic characteristic of the genus *Vibrio sensu strictu*, the polyphasic approach, without the 5S rRNA data, would have delineated accurately the genera of the *Vibrionaceae*. Even without prior knowledge of the utility of DNA base composition for construction of a taxonomy of *Vibrio spp.*, the polyphasic approach has clearly provided a means of estimating a phylogenetic taxonomy.

C. PALINDROMIC SEQUENCES

Since the limit of resolution of comparative sequence analyses of bacterial 5S rRNAs occurs at the species-biovar boundary, it was speculated that, if repeated or palindromic sequences exist in the 5S

Figure 12. Evolutionary relationships between *P. shigelloides* and *Proteus spp.*. Sequence homologies among 5S rRNAs from the species *Plesiomonas shigelloides*, *Proteus mirabilis*, and *Proteus vulgaris*, analyzed by UPGMA, indicate that among the three species, the most recent common ancestor is shared by *P. shigelloides* and *P. mirabilis*. From this, as well as overall homology between the sequence of the 5S rRNA from *P. shigelloides* and those from *Proteus* species, it is concluded that *P. shigelloides* is mis-named, and should be allocated to the *Proteae*.


5S rRNA Sequence Comparison: UPGMA
1 2 3 4 5 6 7 8 9 10
Plesiomonas shigelloides
Proteus mirabilis
Proteus vulgaris
Escherichia coli

rRNA primary structure, it might be possible to map mutations in a sequential fashion and in such a way that insight into evolutionary relationships at the level of biovar could be gleaned. It was gratifying, therefore, to discover that several degenerated palindromic (repeated invert) sequences do exist in *Vibrionaceae* 5S rRNA. While the simultaneous solution of the course of sequential mutations in the least highly conserved palindrome (figure 9) appears to lack the sensitivity necessary to define relationships at the biovar level, it did corroborate evolutionary relationships suggested by UPG analysis using difference matrix and K(nuc) values. It is probable that after a sufficient number of 5S rRNA sequences from a wide range of bacterial species have been determined, palindromic sequences may yield a unique method for analyzing evolutionary trees based on nucleic acid data. Even though there is no fossil record of ancestral nucleic acid sequences, perfect palindromes presumably may be considered ancestors of degenerated sequences.

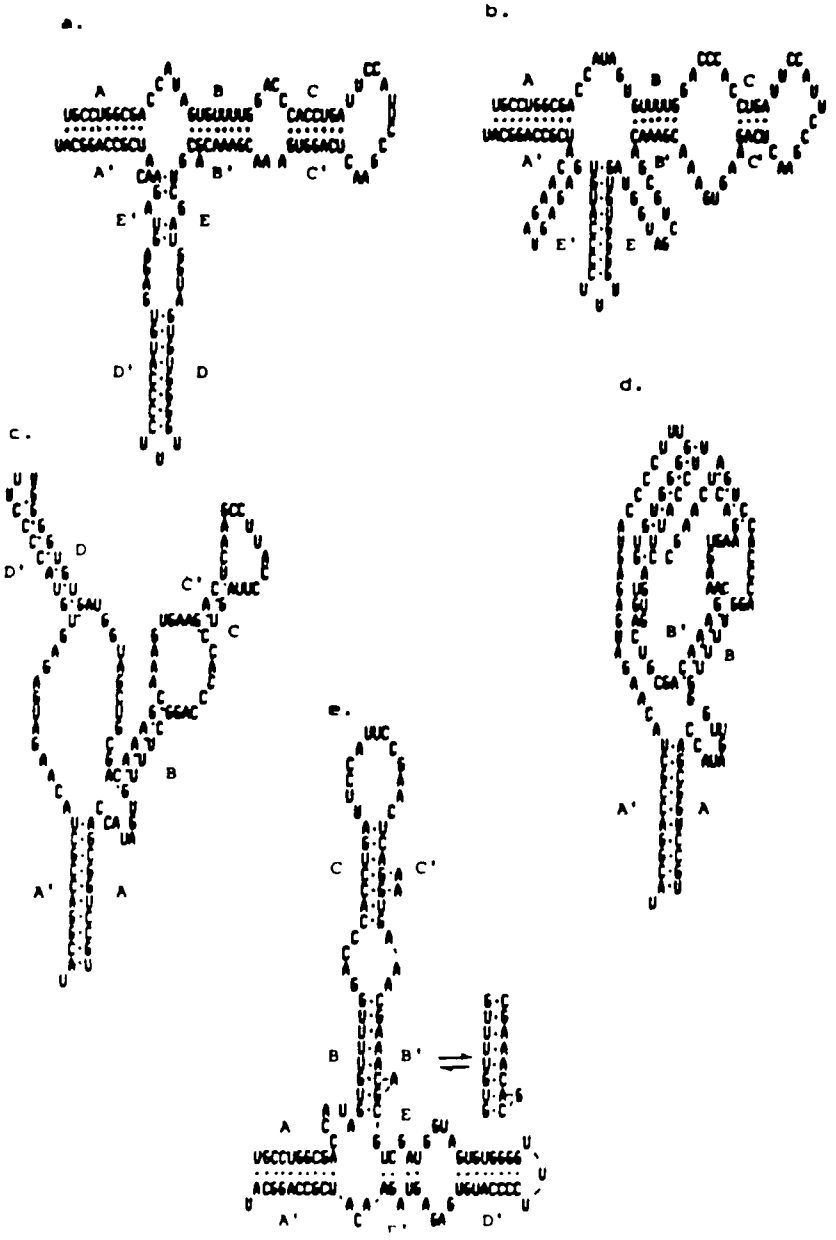
D. NUCLEASE S1 MAPS

Nearly two decades ago, the first 5S rRNA was sequenced (Brownlee *et al.*, 1967). Since that time dozens more sequences have been determined. Yet, the secondary structure of 5S rRNA remains to be resolved fully (Noller, 1984). Most attempts to reveal the secondary structure of 5S rRNAs have relied on base pairing schema for insight into "unique" structures (Tinoco *et al.*, 1971; Dams *et al.*, 1982; Dams *et al.*, 1983; DeWachter *et al.*, 1982; Pieler and Erdmann, 1982). See

figure 13. Unfortunately, no single "permissible" base-pairing scheme exists. For example, Trifonov and Bolshoi (1983), with the aid of a computer to overlay dozens of 5S rRNA sequences, demonstrated that two relatively unrelated secondary structures, termed "Y-form" and "P-form", appear equally probable. Computer analysis of free energy of the secondary structure of each of these, using Ninio's rules (Ninio, 1979; MacDonell and Colwell, 1984c), indicates that both represent thermodynamically stable structures. There have been several chemical approaches taken to resolve the 5S rRNA secondary structure, including NMR (Kime and Moore, 1982; 1983b), nuclear Overhauser effect (Kime and Moore, 1983a), and X-ray scattering (Leontis and Moore, 1984). The concensus of all is that the A/A', D/D' and E/E' regions (Figure 13) form a single cylindrical rod (helix), which, although informative, does not contribute a major insight into the secondary structure problem.

It can be speculated that since 5S rRNA is a relatively small molecule, with at least two thermodynamically stable secondary structures, it may possess a modulatory function. If this is true, it would not be expected to possess a unique secondary structure, but switch between two (or more) structures. In light of this speculation, results of the nuclease S1 limited hydrolysis studies (Figure 10) are encouraging, since they suggest the existence of three types of regions associated with base-pairing: helical, single-stranded, and "intermediate". Helix regions (never hydrolyzed by S1), and loop regions (always hydrolyzed by S1) are general features of several secondary structure models (DeWachter *et al.*, 1982; Pieler *et al.*, 1982; 1984; Trifonov and Bolshoi, 1983), whereas "intermediate" regions

Figure 13. Representation of several 5S rRNA secondary structure models. The proposed secondary structure models of (a) DeWachter *et al.* (1982), (b) Fox and Woese (1975), (c) Trifonov and Bolshoi (1983) (Y-form model), (d) Trifonov and Bolshoi (1983) (P-form model), and (e) MacDonell and Colwell (1984d) are presented for comparison. Although each of these models is supported by experimental evidence, no single model is adequate to explain the accumulated observations from secondary structure analyses. It is probable that the 5S rRNA molecule does not possess a single secondary structure, and that its role may be modulatory, shifting between 2 (or more) conformations. Helices are identified as in Hori and Osawa (1979). The 5S rRNA sequence is that of *V. fluvialis* (MacDonell and Colwell, 1984e)


a.
b.
c.
d.
e.

(occassionally hydrolyzed by S1) are represented as helices in some models and loops in others, best seen by comparing the "P-form" model (Trifonov and Bolshoi, 1983) with the "5-helix" model (DeWachter *et al.*, 1982). Therefore, rather than confirm a particular secondary structure for the 5S rRNA molecule, it is concluded that the nuclease S1 maps reported here provide significant indirect evidence for a modulatory role.

E. A CONSERVED READING FRAME IN 5S rRNA

Shepherd (1981; 1982) showed that a relatively high frequency of "RNY" codons occurs in reading frames having a direct coding function, and that these may represent a vestige of an ancestral comma-less code. Sixteen such codons specify 8 amino acids. Although none of the 3 reading frames of 16S or 23S rRNAs contain other than the statistical distribution of either "RNY" or "STOP" (UAA, UAG, UGA) codons, one frame (third) in 5S rRNAs was found to contain a significantly large number of "RNY" and virtually no "STOP" codons (Erdmann *et al.*, 1983; Colin Clarke, personal communication). Since the number of 5S rRNA sequences determined in this study is approximately equal to the number of published 5S rRNA sequences from which such findings were compiled, the sequences reported herein were screened for frequencies of "RNY"s and "STOP"s (Table 6). The results of the frequencies of "STOP" codons, in each of the 3 frames, in particular, is striking. Furthermore, in each of the 5S rRNAs prepared from species of the genus *Vibrio sensu strictu*, a frame with no "STOP" codons occurs, despite the presence of a

significant number of point mutations in the population as a whole. Such a strongly conservated stop-less reading frame justifies the speculation that frame #3 of 5S rRNA may retain, or participate in, a direct coding function. The hypothetical 39 amino acid polypeptide coded for by the 5S rRNA from *V. fluvialis*, for example, would be:

5'-Pro-Gly-Asp-His-Ser-Val-Leu-Asp-Pro-Pro-Asp-Ser-Ile-Pro-Asn-Ser-Glu-Val-Lys-]
3'-Ala-Gln-Arg-His-Glu-Val-Arg-Val-His-Pro-Phe-Gly-Val-Ser-Gly-Asp-Val-Ser-Asn-Arg-]

Because of the overall conservation of the nucleotide base sequence of 5S rRNAs, only a small number of amino acid residues are susceptible to effects associated with point mutations. Taking this information into account, a concensus polypeptide sequence for the *Vibrionaceae* can be written, as follows:

5'-Pro-Gly-Asp-His-Ser-X-Y-Asp-Pro-Pro-Asp-Ser-(Ile/Met)-Pro-Asn-Ser-Glu-Val-Lys-Arg-]
3'-Ala-Gln-Arg-His-Glu-Val-Arg-Val-His-Pro-Phe-Gly-Val-Ser-Gly-Asp-(Val/Ala)-Ser-(Asn/Ile)-]

Where X may be Val, Asp, Cys, or Ile, and Y may be Leu, Val, or Phe.

Whether or not this polypeptide species (or an analog) occurs in the bacterial cell remains to be determined, and if extant, a function(s) will have to be identified, no doubt contributing even greater understanding of the evolution of the prokaryotes.

APPENDICES

APPENDIX A.

The following are original BASIC language computer programs employed in the management and analysis of nucleic acid data generated in the course of this study.

```
1310 IFT=7THENA=A+2.7
1315 IFT>7THENA=A+(2.9+((T-8)*0.2))
1320 RETURN
1400 CLS
1410 PRINT" ************************************"
1420 PRINT"    ADJUSTMENT FOR NUMBER OF HELICES"
1430 PRINT" ************************************"
1440 PRINT"                ΔG="A
1450 PRINT:INPUT"NUMBER OF HELICES";X
1460 A=A+X
1470 CLS
1480 PRINT" ************************************"
1490 PRINT"  ADJUSTMENT FOR NUMBER OF G:C TERMINI"
1500 PRINT" ************************************"
1510 PRINT"                ΔG="A
1520 PRINT:INPUT"NUMBER OF G:C TERMINI";Z
1530 A=A-(0.7*Z)
1540 CLS
1545 PRINT" ************************************"
1546 PRINT"                ΔG = "A
1550 PRINT" ************************************"
1560 INPUT"         DUMP DATA TO PRINTER?";J$
1570 IFJ$<>"Y"THENCLEAR:GOTO25
1580 INPUT"STRAIN";S$
1590 INPUT"MODEL";M$
1600 LPRINT"STRAIN ="S$
1610 LPRINT"MODEL ="M$
1620 LPRINT"FREE ENERGY (NINIO'S RULES) ="A
1630 CLEAR:GOTO25
```

```
920 IFI=5THENA=A+2.9
930 IFI=6THENA=A+3.2
940 IFI=7THENA=A+3.5
950 IFI>7THENA=A+(3.7+((I-8)*0.2))
960 RETURN
1000 CLS:N=1
1010 PRINT" ========================================"
1020 PRINT"  ADJ. FOR MULTI-BRANCH LOOPS      #"N
1030 PRINT" ========================================"
1040 PRINT"<999 = ENTER>  ΔG="A
1050 PRINT:INPUT"NUMBER OF BASES";M
1055 IFM=999THEN1200
1060 INPUT"NUMBER OF 'U' RESIDUES";UM
1070 GOSUB1080
1075 N=N+1:CLS:GOTO1010
1080 IFUM=0THEN1130
1085 IFM=1THENA=A+0.7
1090 IFM=2THENA=A+1.0
1095 IFM=3THENA=A+1.3
1100 IFM=4THENA=A+1.8
1105 IFM=5THENA=A+2.3
1110 IFM=6THENA=A+2.7
1115 IFM=7THENA=A+2.9
1120 IFM>7THENA=A+(3.1+((M-8)*0.2))
1125 RETURN
1130 IFM=1THENA=A+1.2
1135 IFM=2THENA=A+1.4
1140 IFM=3THENA=A+1.6
1145 IFM=4THENA=A+2.0
1150 IFM=5THENA=A+2.4
1155 IFM=6THENA=A+2.7
1160 IFM=7THENA=A+2.9
1165 IFM=0THENA=A+1.2
1170 IFM>7THENA=A+(3.1+((M-8)*0.2))
1175 RETURN
1200 CLS:N=1
1210 PRINT" ========================================"
1220 PRINT"  ADJ. FOR UNPAIRED TERMINUS       #"N
1230 PRINT" ========================================"
1240 PRINT"<999 = ESCAPE>  ΔG="A
1250 PRINT:INPUT"NUMBER OF BASES";T
1255 IFT=999THEN1400
1260 GOSUB1280
1265 N=N+1
1270 IFN=3THEN1400
1275 CLS:GOTO1210
1280 IFT=1THENA=A+1.5
1285 IFT=2THENA=A+1.7
1290 IFT=3THENA=A+1.9
1295 IFT=4THENA=A+2.1
1300 IFT=5THENA=A+2.3
1305 IFT=6THENA=A+2.5
```

```
340 IFH>7THENA=A+(4.7+((H-8)*0.2)):RETURN
350 RETURN
360 IFH=3THENA=A+6.9
370 IFH=4THENA=A+4.7
380 IFH=5THENA=A+4.0
390 IFH=6THENA=A+4.1
400 IFH=7THENA=A+4.4
405 IFH>7THENA=A+(4.7+((H-8)*0.2))
420 PRINTA
430 RETURN
450 CLS:N=1:B$=""
460 PRINT" ************************************"
470 PRINT"  ADJUSTMENT FOR BULGE LOOPS  LOOP #"N
480 PRINT" ************************************"
490 PRINT"(999 = ESCAPE)  AG="A
500 PRINT:INPUT"NUMBER OF BASES";B
505 IFB=999THEN800
510 INPUT"NUMBER OF 'U' RESIDUES";UB
520 GOSUB540
530 N=N+1:CLS:GOTO460
540 IFUB=0THEN640
550 IFB=1THENA=A+2.2
560 IFB=2THENA=A+3.5
570 IFB=3THENA=A+4.3
580 IFB=4THENA=A+4.8
590 IFB=5THENA=A+5.2
600 IFB=6THENA=A+5.5
610 IFB=7THENA=A+5.7
620 IFB>7THENA=A+(5.9+((B-8)*0.2))RETURN
630 RETURN
640 IFB=1THENA=A+2.8
650 IFB=2THENA=A+3.9
660 IFB=3THENA=A+4.5
670 IFB=4THENA=A+5.8
680 IFB=5THENA=A+5.3
690 IFB=6THENA=A+5.5
700 IFB=7THENA=A+5.7
710 IFB>7THENA=A+(5.9+((B-8)*0.2))
720 RETURN
800 CLS:N=1:B$=""
810 PRINT" ************************************"
820 PRINT"  ADJUSTMENT FOR INTERNAL LOOPS    #"N
830 PRINT" ************************************"
840 PRINT"(999 = ESCAPE)  AG="A
850 PRINT:INPUT"NUMBER OF BASES";I
855 IFI=999THEN1000
860 GOSUB880
870 N=N+1:CLS:GOTO810
880 IFI=1THENA=A+1.5
890 IFI=2THENA=A+1.9
900 IFI=3THENA=A+2.3
910 IFI=4THENA=A+2.6
```

```
110 IFA$="UGCA"THEN14
111 IFA$="UGAA"THEN14
112 IFA$="UGUC"THEN14
113 IFA$="GUAC"THEN14
114 IFA$="GUCA"THEN14
115 IFA$="GUCC"THEN14
116 IFA$="GUAA"THEN14
117 IFA$="GUUC"THEN14
118 IFA$="GUCU"THEN14
119 IFA$="GUUU"THEN14
120 IFA$="GUGA"THEN14
121 IFA$="GUAG"THEN14
122 IFA$="GUGG"THEN14
123 IFA$="UGGUT"THEN14
124 IFA$="UGTGU"THEN14
125 IFA$="GCCG"THEN15
126 IFA$="UAGA"THEN16
127 IFA$="UAAA"THEN16
128 IFA$="UAAG"THEN16
129 IFA$="AUAA"THEN16
130 IFA$="AUGA"THEN16
131 IFA$="AUAG"THEN16
132 IFA$="CGGUT"THEN17
133 IFA$="GCUGT"THEN18
134 IFA$="UGUA"THEN18
135 IFA$="UGCG"THEN2
137 IFA$=""THEN150
140 CLS:GOTO25
150 N=1:B$=""
151 CLS
155 PRINT" ==========================================="
160 PRINT"  ADJUSTMENT FOR HAIRPINS  Hairpin #"N
165 PRINT" ==========================================="
170 PRINT"(999 = ESCAPE)  ▲G="A:PRINT:INPUT"NUMBER OF BASES";H
171 IFH=999THEN450
175 INPUT"NUMBER OF 'U' RESIDUES";UH
180 GOSUB200
190 N=N+1:GOTO151
200 IFUH=0THEN290
210 IFUH=1THEN360
220 IFH=3THENA=A+6.4
230 IFH=4THENA=A+4.3
240 IFH=5THENA=A+3.7
250 IFH=6THENA=A+3.9
260 IFH=7THENA=A+4.3
270 IFH>7THENA=A+(4.7+((H-8)*0.2)):RETURN
280 RETURN
290 IFH=3THENA=A+8.4
300 IFH=4THENA=A+5.9
310 IFH=5THENA=A+4.9
320 IFH=6THENA=A+4.7
330 IFH=7THENA=A+4.7
```

```
 58 IFA$="CGAG"THEN6
 59 IFA$="AUCC"THEN6
 60 IFA$="AUGG"THEN6
 61 IFA$="GCUC"THEN7
 62 IFA$="GCCU"THEN7
 63 IFA$="UAAC"THEN7
 64 IFA$="CGCU"THEN7
 65 IFA$="UACA"THEN7
 66 IFA$="CGUC"THEN7
 67 IFA$="AUAC"THEN7
 68 IFA$="AUCA"THEN7
 69 IFA$="GCUU"THEN8
 70 IFA$="CGGU"THEN8
 71 IFA$="AUGU"THEN8
 72 IFA$="UGGC"THEN8
 73 IFA$="CGUU"THEN8
 74 IFA$="AUGUT"THEN8
 75 IFA$="CGCG"THEN9
 76 IFA$="GCGC"THEN9
 77 IFA$="CGUG"THEN10
 78 IFA$="AUUG"THEN10
 79 IFA$="GUUA"THEN10
 80 IFA$="GUGC"THEN10
 81 IFA$="AUUGT"THEN10
 82 IFA$="GCGUT"THEN10
 83 IFA$="CGUGT"THEN10
 84 IFA$="UAUA"THEN11
 85 IFA$="UAAU"THEN11
 86 IFA$="AUUA"THEN11
 87 IFA$="AUAU"THEN11
 88 IFA$="UGUG"THEN12
 89 IFA$="GUGU"THEN12
 90 IFA$="UACU"THEN12
 91 IFA$="UAUC"THEN12
 92 IFA$="AUUC"THEN12
 93 IFA$="AUCU"THEN12
 94 IFA$="GUGUT"THEN12
 95 IFA$="UGUGT"THEN12
 96 IFA$="UGTUG"THEN12
 97 IFA$="GUTGU"THEN12
 98 IFA$="GUUG"THEN13
 99 IFA$="UAUU"THEN13
100 IFA$="AUUU"THEN13
101 IFA$="GUUGT"THEN13
102 IFA$="UGGU"THEN14
103 IFA$="UGCU"THEN14
104 IFA$="UGGA"THEN14
105 IFA$="UGAG"THEN14
106 IFA$="UGAC"THEN14
107 IFA$="UGCC"THEN14
108 IFA$="UGGG"THEN14
109 IFA$="UGUU"THEN14
```

```
0 REM DETERMINATION OF FREE ENERGY OF SECONDARY STRUCTURE BY NINIO'S RULES
1 CLS:A=0:GOTO25
2 A=A-1.5:GOTO25
3 A=A-3.2:GOTO25
4 A=A-.2:GOTO25
5 A=A-.8:GOTO25
6 A=A+.2:GOTO25
7 A=A+.8:GOTO25
8 A=A-.5:GOTO25
9 A=A-3.8:GOTO25
10 A=A-1:GOTO25
11 A=A-1.7:GOTO25
12 A=A+1:GOTO25
13 A=A+.5:GOTO25
14 A=A+1.5:GOTO25
15 A=A-3.3:GOTO25
16 A=A-1.2:GOTO25
17 A=A-.3:GOTO25
18 A=A-2:GOTO25
21 PRINT" **********************************"
25 CLS:PRINT" COMPUTATION OF ▲G BY NINIO'S RULES"
26 PRINT" **********************************"
27 PRINT" Enter Non-Canonical Bases as 2d Pair"
29 PRINT" ----------------------------------"
30 A$="":PRINT"                ▲G="A:PRIN":PRINT@200,"";:INPUTA$
31 IFA$="C6U6"THEN2
32 IFA$="6CGU"THEN2
33 IFA$="C66C"THEN2
34 IFA$="6UCG"THEN2
35 IFA$="C5AU"THEN3
36 IFA$="6CAU"THEN3
37 IFA$="UAC6"THEN3
38 IFA$="AUC6"THEN3
39 IFA$="C6UA"THEN3
40 IFA$="6CUA"THEN3
41 IFA$="UA6C"THEN3
42 IFA$="AU6C"THEN3
43 IFA$="6CAC"THEN4
44 IFA$="6CCA"THEN4
45 IFA$="C6CA"THEN4
46 IFA$="C6AC"THEN4
47 IFA$="6CCC"THEN5
48 IFA$="6C66"THEN5
49 IFA$="C6CC"THEN5
50 IFA$="C666"THEN5
51 IFA$="6CAA"THEN6
52 IFA$="6C6A"THEN6
53 IFA$="6CA6"THEN6
54 IFA$="C66A"THEN6
55 IFA$="UAGG"THEN6
56 IFA$="UACC"THEN6
57 IFA$="C6AA"THEN6
```

```
10030 IFQ$="V PSYC"THENA$=PP$
10031 IFQ$="P FLUO"THENA$=FF$
10032 IFQ$="VP1"THENA$=HH$
10033 IFQ$="R RUBR"THENA$=RR$
10034 IFQ$="A NIDU"THENA$=ND$
10035 IFQ$="S LIVI"THENA$=LL$
10036 IFQ$="PROC"THENA$=PR$
10037 IFQ$="T AQUA"THENA$=AA$
10038 IFQ$="T THER"THENA$=TT$
10039 IFA$=""THENPRINT"STRAIN NOT LOCATED IN DATABANK":FORI=1TO500:NEXTI:GOTO130
11000 IFR$="V PARA"THENB$=PA$
11001 IFR$="V FLUV"THENB$=FL$
11002 IFR$="V CHOL"THENB$=CH$
11003 IFR$="V HARV"THENB$=HA$
11004 IFR$="V FISC"THENB$=FI$
11005 IFR$="V MARI"THENB$=MA$
11006 IFR$="V VULN"THENB$=VU$
11007 IFR$="HALVIS"THENB$=HA$
11008 IFR$="V ANGU"THENB$=AN$
11009 IFR$="P PHOS"THENB$=PH$
11010 IFR$="A HYDR"THENB$=HY$
11011 IFR$="RIFTIA"THENB$=RF$
11012 IFR$="CALYPT"THENB$=CA$
11013 IFR$="SOL"THENB$=SO$
11014 IFR$="P SHIG"THENB$=SH$
11015 IFR$="P MIRA"THENB$=MI$
11016 IFR$="P VULG"THENB$=VL$
11017 IFR$="Y PEST"THENB$=PE$
11018 IFR$="A VINE"THENB$=VI$
11019 IFR$="P AERU"THENB$=AE$
11020 IFR$="E ADES"THENB$=AD$
11021 IFR$="E ARGS"THENB$=AR$
11022 IFR$="S MARC"THENB$=MR$
11023 IFR$="S TYPH"THENB$=TY$
11024 IFR$="E COLI"THENB$=CO$
11025 IFR$="V ALGI"THENB$=AL$
11026 IFR$="V NATR"THENB$=NN$
11027 IFR$="V MIMI"THENB$=MM$
11028 IFR$="V CARC"THENB$=CC$
11029 IFR$="V DIAZ"THENB$=DD$
11030 IFR$="V PSYC"THENB$=PP$
11031 IFR$="P FLUO"THENB$=FF$
11032 IFR$="VP1"THENB$=HH$
11033 IFR$="R RUBR"THENB$=RR$
11034 IFR$="A NIDU"THENB$=ND$
11035 IFR$="S LIVI"THENB$=LL$
11036 IFR$="PROC"THENB$=PR$
11037 IFR$="T AQUA"THENB$=AA$
11038 IFR$="T THER"THENB$=TT$
11039 IFR$=""THENPRINT"STRAIN NOT LOCATE IN DATABANK":FORI=1TO500:NEXTI:GOTO1311
11040 RETURN
```

```
4025 SI=1/KB
4026 SB=(SA+SI)/2
4028 SJ=SA^2
4029 SK=SB^2
4030 SL=(SJ$P)+(SK$Q)
4031 SM=((SA$P)+(SB$Q))^2
4032 SS=(SL-SM)/N
5145 PRINT"1/2 K(nuc) = "KN/2
5146 PRINT"SE(k) = "SS
5147 PRINT"DUMP TO PRINTER?"
5148 Y$=INKEY$:IFY$=""THEN5148
5149 IFY$="Y"THENGOSUB6000
5150 PRINT"CONTINUE?"
5151 R$=INKEY$:IFR$=""THEN5151
5152 IFR$="Y"THEN5155
5153 MENU
5155 CLEAR:GOTO6
6000 LPRINT"REF STRAIN = "Q$
6001 LPRINT"TEST STRAIN = "R$
6002 LPRINT"1/2 K(nuc) = "KN/2
6003 LPRINT"SE(k) = "SS
6004 LPRINT
6005 RETURN
10000 A$="":B$="":IFQ$="V PARA"THENA$=PA$
10001 IFQ$="V FLUV"THENA$=FL$
10002 IFQ$="V CHOL"THENA$=CH$
10003 IFQ$="V HARV"THENA$=HA$
10004 IFQ$="V FISC"THENA$=FI$
10005 IFQ$="V MARI"THENA$=MA$
10006 IFQ$="V VULN"THENA$=VU$
10007 IFQ$="HALVIS"THENA$=HA$
10008 IFQ$="V ANGU"THENA$=AN$
10009 IFQ$="P PHOS"THENA$=PH$
10010 IFQ$="A HYDR"THENA$=HY$
10011 IFQ$="RIFTIA"THENA$=RF$
10012 IFQ$="CALYPT"THENA$=CA$
10013 IFQ$="SOL"THENA$=SO$
10014 IFQ$="P SHIG"THENA$=SH$
10015 IFQ$="P MIRA"THENA$=MI$
10016 IFQ$="P VULG"THENA$=VL$
10017 IFQ$="Y PEST"THENA$=PE$
10018 IFQ$="A VINE"THENA$=VI$
10019 IFQ$="P AERU"THENA$=AE$
10020 IFQ$="E ADES"THENA$=AD$
10021 IFQ$="E ARGS"THENA$=AR$
10022 IFQ$="S MARC"THENA$=MR$
10023 IFQ$="S TYPH"THENA$=TY$
10024 IFQ$="E COLI"THENA$=CO$
10025 IFQ$="V ALGI"THENA$=AL$
10026 IFQ$="V NATR"THENA$=NN$
10027 IFQ$="V MIMI"THENA$=MM$
10028 IFQ$="V CARC"THENA$=CC$
10029 IFQ$="V DIAZ"THENA$=DD$
```

```
131 INPUT"Test Taxon";R$
132 PRINT:PRINT"Working...":GOSUB10000
133 GOTO3000
200 IF MID$(B$,I,1)="A"THENN=N+1
201 IF MID$(B$,I,1)="G"THENS=S+1:N=N+1
202 IF MID$(B$,I,1)="U"THENV=V+1:N=N+1
203 IF MID$(B$,I,1)="C"THENV=V+1:N=N+1
204 IF MID$(B$,I,1)="-"THENV=V+1:N=N+1
205 GOTO4008
210 IF MID$(B$,I,1)="A"THENS=S+1:N=N+1
211 IF MID$(B$,I,1)="G"THENN=N+1
212 IF MID$(B$,I,1)="U"THENV=V+1:N=N+1
213 IF MID$(B$,I,1)="C"THENV=V+1:N=N+1
214 IF MID$(B$,I,1)="-"THENV=V+1:N=N+1
215 GOTO4008
220 IF MID$(B$,I,1)="A"THENV=V+1:N=N+1
221 IF MID$(B$,I,1)="G"THENV=V+1:N=N+1
222 IF MID$(B$,I,1)="U"THENN=N+1
223 IF MID$(B$,I,1)="C"THENS=S+1:N=N+1
224 IF MID$(B$,I,1)="-"THENV=V+1:N=N+1
225 GOTO4008
230 IF MID$(B$,I,1)="A"THENV=V+1:N=N+1
231 IF MID$(B$,I,1)="G"THENV=V+1:N=N+1
232 IF MID$(B$,I,1)="U"THENS=S+1:N=N+1
233 IF MID$(B$,I,1)="C"THENN=N+1
234 IF MID$(B$,I,1)="-"THENV=V+1:N=N+1
235 GOTO4008
240 IF MID$(B$,I,1)="A"THENV=V+1:N=N+1
241 IF MID$(B$,I,1)="G"THENV=V+1:N=N+1
242 IF MID$(B$,I,1)="U"THENV=V+1:N=N+1
243 IF MID$(B$,I,1)="C"THENV=V+1:N=N+1
244 IF MID$(B$,I,1)="-"THENN=N
245 GOTO4008
275 LPRINT"-------------------------------------------------------------------------------"
276 LPRINT"        DETERMINATION OF EVOLUTIONARY DISTANCE BY KIMURA'S ALGORITHM"
277 LPRINT"-------------------------------------------------------------------------------"
278 LPRINT:LPRINT:LPRINT:LPRINT:LPRINT
279 GOTO6
3000 FORI=BTOE
3001 IF MID$(A$,I,1)="A"THEN 200
3002 IF MID$(A$,I,1)="G"THEN 210
3003 IF MID$(A$,I,1)="U"THEN 220
3004 IF MID$(A$,I,1)="C"THEN 230
3005 IF MID$(A$,I,1)="-"THEN 240
4008 NEXTI
4017 P=S/N:Q=V/N
4018 KA=(1-(2*P)-Q)
4019 KB=(1-(2*Q))
4020 KC=KA*KB
4021 KD=SQR(KC)
4022 KE=LOG(KD)
4023 KN=-KE/2
4024 SA=1/KA
```

```
 1 REM COMPUTATION OF K(nuc) BY THE ALGORITHM OF KIMURA JME 16:111 (80)
 2 REM PROGRAM WRITTEN FOR TRS 80 MODEL 100: M.T. MACDONELL, DEPT MICROBIOLOGY, UNIV. MARYLAND
 3 CLS:PRINT"DUMP HEADING TO PRINTER?
 4 Y$=INKEY$:IFY$=""THEN4
 5 IFY$="Y"THEN275
 6 DATA""
 10 DATA""
 15 DATA""
 20 DATA""
 25 DATA""
 30 DATA""
 35 DATA""
 40 DATA""
 45 DATA""
 50 DATA""
 51 DATA""
 52 DATA""
 53 DATA""
 54 DATA""
 55 DATA""
 56 DATA""
 57 DATA""
 58 DATA""
 59 DATA""
 60 DATA""
 61 DATA""
 62 DATA""
 63 DATA""
 64 DATA""
 65 DATA""
 66 DATA""
 67 DATA""
 68 DATA""
 69 DATA""
 70 DATA""
 71 DATA""
 72 DATA""
 73 DATA""
 74 DATA""
 75 DATA""
 76 DATA""
 77 DATA""
 78 DATA""
 79 DATA""
 80 READAL$,PA$,FL$,CH$,HA$,FI$,MA$,VU$,WA$,AN$,PH$,HY$,RF$,CA$,SO$,SH$,MI$,VL$,PE$,VI$,AE$,AD$,AR$,MR-
$,TY$,CO$,NN$,MM$,CC$,DD$,PP$,FF$,HH$,RR$,ND$,LL$,PP$,AA$,TT$
105 CLS:PRINT:PRINT:PRINT:PRINT:PRINT"          ' CAPS LOCK ON ":FORJ=1TO70:NEXT
110 CLS:LINE(5,5)-(227,18),1,B
111 T=0:N=0
120 PRINT:PRINT" == DETERMINATION OF K(nuc): KIMURA =="
125 PRINT
128 B=1:E=149
130 INPUT"Reference Taxon";D$
```

```
5 CLS:LINE(60,14)-(196,24),1,B:PRINT@92,"SEQUENCE COMPARISON"
10 GOSUB1000
15 INPUT"Test strain taxon";T$
16 PRINT"Input "T$" sequence";:INPUTB$
20 N=LEN(A$)
25 M=LEN(B$)
29 C=0
30 IFM>NTHENN=M
35 FORI=1TON
40 IFMID$(A$,I,1)=MID$(B$,I,1)THEN45
41 C=C+1
45 NEXTI
50 CLS:PRINT:PRINT"Difference matrix element= ":LINE(161,6)-(186,16),1,B:PRINT@67,C
55 X=INT(1000*((N-C)/N))
60 PRINT:PRINT"% Homology=":LINE(71,22)-(102,32),1,B:PRINT@132,X/10
63 PRINT
65 PRINT"Continue?"
66 R$=INKEY$
67 IFR$=""THEN66
70 IFR$="y"THENCLS:GOTO15
71 IFR$="Y"THENCLS:GOTO15
75 MENU
1000 PRINT:INPUT"Reference strain taxon";R$
1001 PRINT"Input "R$" sequence";:INPUTA$
1002 RETURN
```

```
1 REM PROGRAM WRITTEN BY M.T. MACDONELL, DEPT. OF MICROBIOLOGY UNIV OF MARYLAND: SEARCHES FOR 'RNY'
AND 'STOP' CODONS IN EACH OF THREE READING FRAMES OF NUCLEIC ACID SEQUENCES
2 REM WRITTEN IN MICROSOFT BASIC FOR TRS 80-MODEL 100
5 GOSUB205
10 PRINT:PRINT:PRINT:PRINT:PRINT:INPUT"SEQUENCE";A$
11 B$=LEFT$(A$,3)
12 C=0
15 A=0
20 I=1
25 I=LEN(A$)
30 IFMID$(A$,X,1)="A"THEN100
31 IF MID$(A$,X,1)="G"THEN100
40 X=X+3:IFX)ZTHEN120
45 GOTO30
100 IFMID$(A$,X+2,1)="C"THENA=A+1
101 IF MID$(A$,X+2,1)="U"THENA=A+1
105 X=X+3
106 IFX)ZTHEN120
110 GOTO30
120 PRINTA"'RNY' sequences in frame "C+1
125 A$=MID$(A$,2,Z-1)
127 C=C+1
128 IFC=3THEN150
130 GOTO15
150 GOTO1000
205 SOUNDON
209 CLS
210 LINE(20,22)-(210,32),1,B
220 PRINT @ 125," 'RNY' & 'STOP' CODON SEARCH"
224 SOUND1500,1:SOUND2510,1:SOUND1500,1:SOUND2510,1:SOUND1500,1:SOUND2510,1
225 FORI=1TO900:NEXT:CLS
230 RETURN
1000 A=0:X=1:C=0:A$=B$+A$
1001 Z=LEN(A$)
1005 IFMID$(A$,X,3)="UAA"THEN1100
1031 IF MID$(A$,X,3)="UAG"THEN1100
1035 IF MID$(A$,X,3)="UGA"THEN1100
1040 X=X+3:IFX)ZTHEN1120
1045 GOTO1005
1100 A=A+1
1105 X=X+3
1106 IFX)ITHEN1120
1110 GOTO1005
1120 PRINTA"'STOP' codons in frame "C+1
1125 A$=MID$(A$,2,Z-1)
1127 C=C+1
1128 IFC=3THEN1150
1130 A=0:X=1:GOTO1005
1150 PRINT"CONTINUE?"
1151 K$=INKEY$
1152 IFK$=""THEN1151
1153 IFK$="Y"THENCLS:GOTO10
1154 MENU
```

```
4018 ZJ$=CHR$(203):GOTO4020
4019 ZJ$=CHR$(239):GOTO4020
4020 LPRINTZJ$;:
4054 NEXTJ
4055 LPRINT
4056 NEXTI
4057 INPUT"CONTINUE";ZZ$
4058 IFZZ$="N"THENEND
4059 CLEAR:GOTO10
5149 PRINT
5150 PRINT"CONTINUE?"
5151 R$=INKEY$:IFR$=""THEN5151
5152 IFR$="N"THEN4000
5153 GOTO190
10000 A$="":B$=""
10001 IFQ$=""THENA$=$
10039 IFA$=""THENPRINT"STRAIN NOT LOCATED IN DATABANK":FORI=1TO500:NEXTI:GOTO130
11000 IFR$=""THENB$=$
11039 IFR$=""THENPRINT"STRAIN NOT LOCATED IN DATABANK":FORI=1TO500:NEXTI:GOTO1311
11040 RETURN
20000 FORI=1TO0
20001 FORJ=1TO0
20003 LPRINTGZ(J,I);:
20004 NEXTJ
20005 NEXTI
20006 END
```

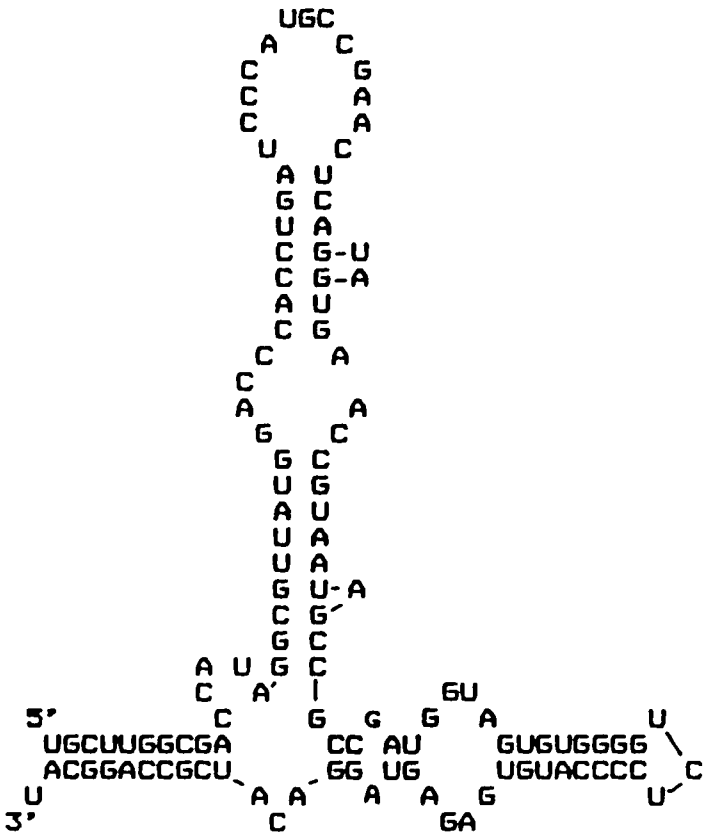
```
1 REM BASIC PROGRAM FOR DOT MATRIX ANALYSIS OF PALINDROMIC AND REPEATED SEQUENCES
3 CLEAR:DIMGZ(100,100):LPRINTCHR$(27)CHR$(77)CHR$(14):LPRINTCHR$(15):LPRINTCHR$(27)CHR$(65)CHR$(4)
5 REM M.T. MACDONELL & R.R. COLWELL, DEPT OF MICROBIOLOGY, UNIV OF MARYLAND
6 DATA"
180 READ
181 CLS:PRINT"DTMX DENSITY MAP PROGRAM":PRINT:PRINT"(1) REPEATED SEQUENCES":PRINT"(2) PALINDROMIC
SEQUENCES":PRINT"(3) COMPARE TWO SEQUENCES":INPUT"CHOICE";K
190 INPUT"Identify Reference Sequence";Q$
191 IFK=1THENM$="REPEATED SEQUENCE SEARCH":R$=Q$:GOTO195
192 IFK=2THENM$="PALINDROME SEARCH":R$=Q$:GOTO196
193 IFK=3THENM$=""
194 INPUT"Identify Test Sequence";R$
195 INPUT"OLIGOMER DEFAULT LENGTH";L
196 PRINT:PRINT"Working...":GOSUB10000
197 IFK=2THENGOTO3000
198 PRINT"DOT MATRIX:  "Q$" x "R$"    "
199 PRINT
200 PRINT
201 PRINT
205 IFLEN(A$)>LEN(B$)THENQ=LEN(A$):GOTO210
206 Q=LEN(B$)
210 FORI=1TOQ:FORJ=1TOQ
220 IFMID$(A$,J,L)=MID$(B$,I,L)THENGZ(J,I)=1+GZ(J,I)
230 NEXTJ
240 PRINTCHR$(13);:PRINT(I$100)"ARRAY ELEMENTS COMPLETED";:NEXTI
250 GOTO5149
3000 Q=LEN(A$):INPUT"OLIGOMER DEFAULT LENGTH";L
3001 B$="":FORI=0TO1STEP-1
3002 Y$=MID$(A$,I,1)
3003 B$=B$+Y$
3004 NEXTI
3005 GOTO198
4000 INPUT"DUMP TO PRINTER";ZZ$:IFZZ$="N"THENEND
4001 IFK=1THENLPRINT"REPEATED SEQUENCE DENSITY MAP":LPRINT:LPRINT:LPRINT
4002 IFK=2THENLPRINT"PALINDROMIC SEQUENCE DENSITY MAP":LPRINT:LPRINT:LPRINT:
4003 LPRINT"REPRESENTATIVE SEQUENCE "Q$:LPRINT:LPRINT:LPRINT:LPRINT"        20        30        40
50        60        70        80        90        100"
4004 LPRINT"         .         .         .         .         .         .         .         .
.":LPRINT:LPRINT
4005 LPRINTA$:LPRINT:LPRINT:LPRINT
4006 FORI=1TOQ:FORJ=1TOQ
4007 IFGZ(J,I)>5THENGZ(J,I)=5
4008 IFGZ(J,I)=0THEN4014
4009 IFGZ(J,I)=1THEN4015
4010 IFGZ(J,I)=2THEN4016
4011 IFGZ(J,I)=3THEN4017
4012 IFGZ(J,I)=4THEN4018
4013 IFGZ(J,I)=5THEN4019
4014 ZJ$=CHR$(241):GOTO4020
4015 ZJ$=CHR$(43):GOTO4020
4016 ZJ$=CHR$(173):GOTO4020
4017 ZJ$=CHR$(174):GOTO4020
```

```
1 REM Program "DATA MANAGER" by M.T. MacDonell August, 1984.  Written for TRS 80 M100 system.
2 MAXFILES=2
5 SOUNDON
9 CLS
10 LINE(20,20)-(210,35),1,B
20 PRINT @ 133,"DATA MANAGER"
24 SOUND1500,1:SOUND2510,1:SOUND1500,1:SOUND2510,1:SOUND1500,1:SOUND2510,1
25 FORI=1TO900:NEXT:CLS
26 PRINT:LINE(88,6)-(103,17),1,B:INPUT"READ OR WRITE";L$
27 IFL$=""THEN26
28 IFL$="R"THEN2000
29 IFL$<>"W"THEN26
30 CLS:PRINT:GOSUB1000:LINE(124,6)-(139,17),1,B:INPUT"INPUT FILE REQUIRED";A$
35 IFA$<>"Y"THEN50
40 CLS:GOSUB1000:PRINT:LINE(118,6)-(163,17),1,B:INPUT"NAME OF INPUT FILE";FI$
41 OPENFI$FOR INPUT AS 1
42 TT=1
50 CLS:GOSUB1000:PRINT:LINE(124,6)-(169,17),1,B:INPUT"NAME OF OUTPUT FILE";FO$
55 OPENFO$FOR OUTPUT AS 2
60 PRINT:GOSUB1000:LINE(112,22)-(133,33),1,B:INPUT"NUMBER OF ENTRIES";N
65 PRINT:LINE(160,38)-(175,49),1,B:INPUT"IS DATA NUMERIC OR STRING";M$
70 IFTT=1THEN100
71 IFLEFT$(M$,1)=""THEN71
72 IFLEFT$(M$,1)="S"THEN900
73 IFLEFT$(M$,1)<>"N"THEN71
74 FORI=1TON
75 GOSUB1000:INPUT"ITEM,DATA";IT$,DA
80 PRINT#2,IT$","DA
90 NEXT
95 GOSUB1000:INPUT"END";A$
96 IFA$="Y"THENCLOSE#1:CLOSE#2:END
97 GOTO9
100 IFLEFT$(M$,1)=""THEN65
101 IFLEFT$(M$,1)="S"THEN170
102 IFLEFT$(M$,1)<>"N"THEN65
105 FORI=1TON
110 INPUT#1,IT$
115 PRINTIT$
120 GOSUB1000:INPUT"DATA";DA
130 PRINT#2,IT$;DA
140 NEXT
150 GOSUB1000:INPUT"END";A$
155 IFA$="Y"THENCLOSE#1:CLOSE#2:END
160 GOTO9
170 FORI=1TON
175 INPUT#1,IT$,DA
180 PRINTIT$,DA
185 GOSUB1000:INPUT"DATA";DA$
190 PRINT#2,IT$,DA$
195 NEXT
197 GOSUB1000:INPUT"END";A$
198 IFA$="Y"THENCLOSE#1:CLOSE#2:END
```

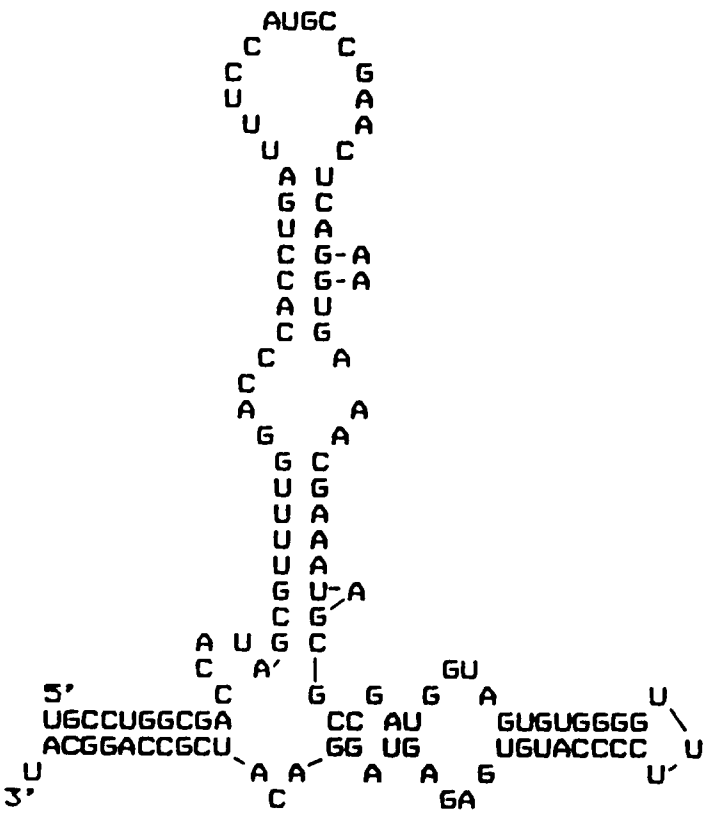
```
900 FORI=1TON
905 GOSUB1000:INPUT"ITEM,DATA";IT$,DA$
910 PRINT#2,IT$","DA$
920 NEXT
950 GOSUB1000:INPUT"END";A$
960 IFA$="Y"THENCLOSE#1:CLOSE#2:END
970 GOTO9
1000 SOUND3000,3
1001 RETURN
2000 CLS:GOSUB1000:PRINT:LINE(124,6)-(169,17),1,B:INPUT"NAME OF OUTPUT FILE";FO$
2005 OPENFO$FOR INPUT AS 2
2010 FORI=1TO100
2020 INPUT#2,IT$,DA$
2025 PRINTIT$,DA$
2030 INPUT"(Key to continue (E to end)";Z$
2035 IFZ$="E"THENCLOSE#1:CLOSE#2:END
2040 NEXT
```

APPENDIX B.

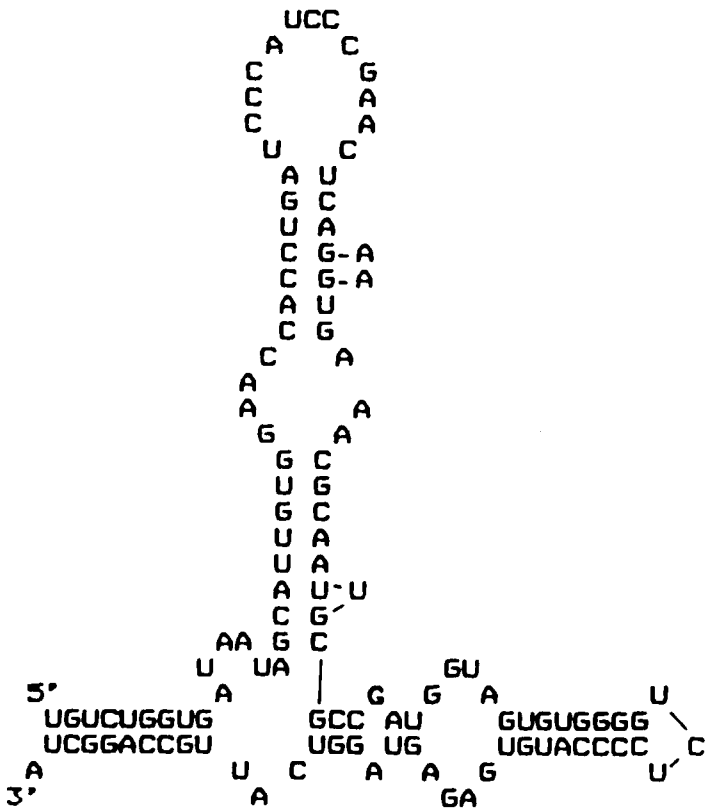
The following are sequences of 5S rRNAs representative of each major cluster (figures 5 and 7). Secondary structure model employed is from MacDonell and Colwell (1984d).



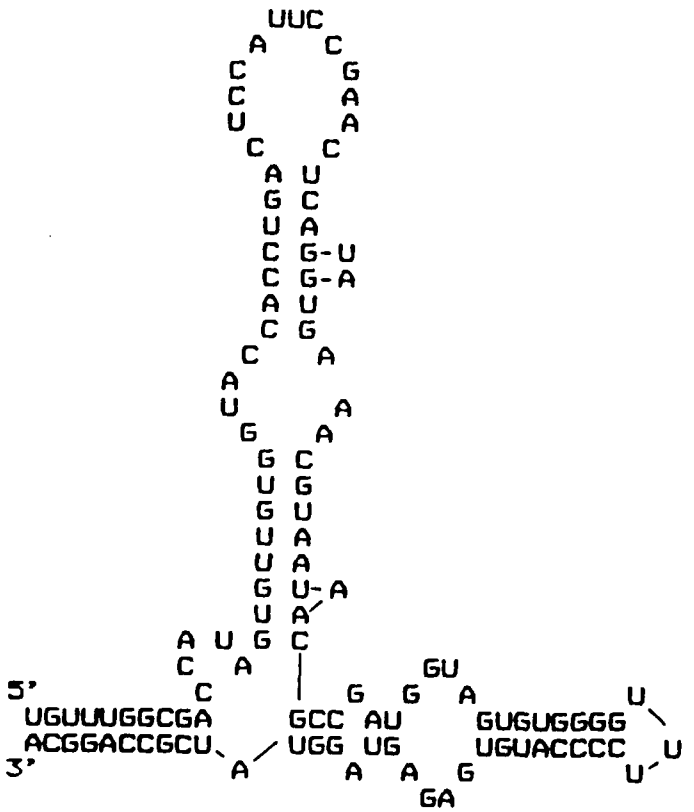
Photobacterium leoignathi



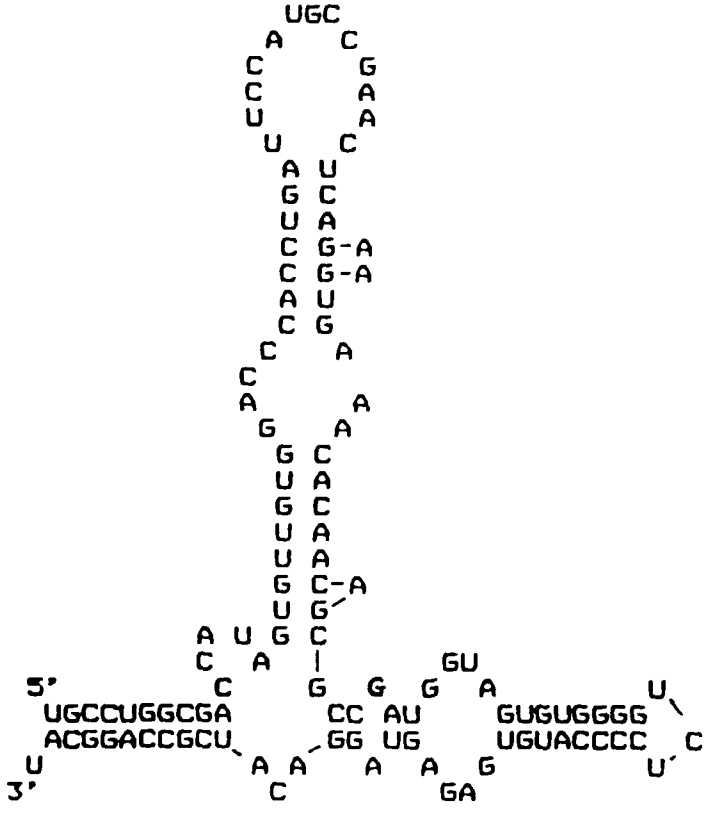
Vibrio alginolyticus



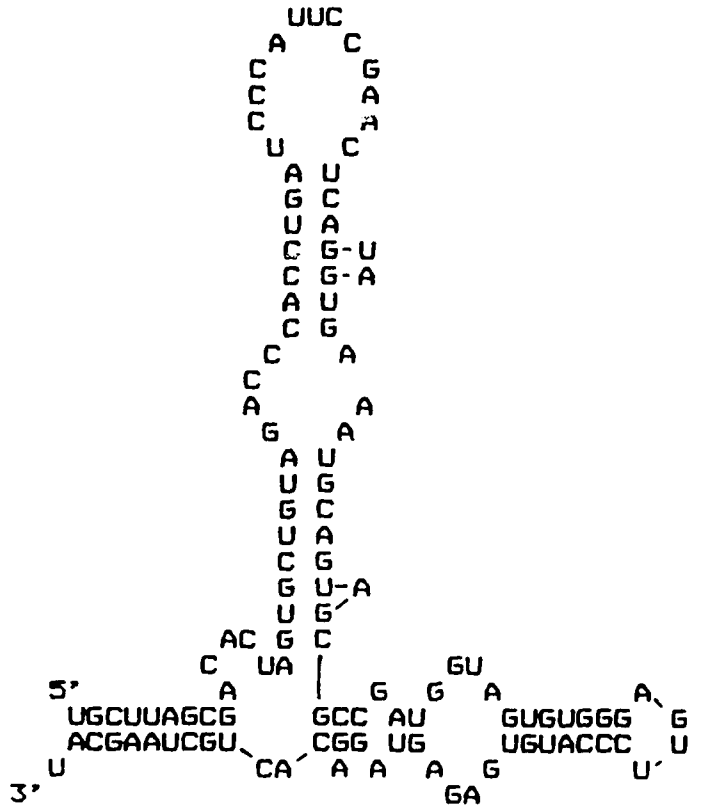
Alteromonas putrifaciens



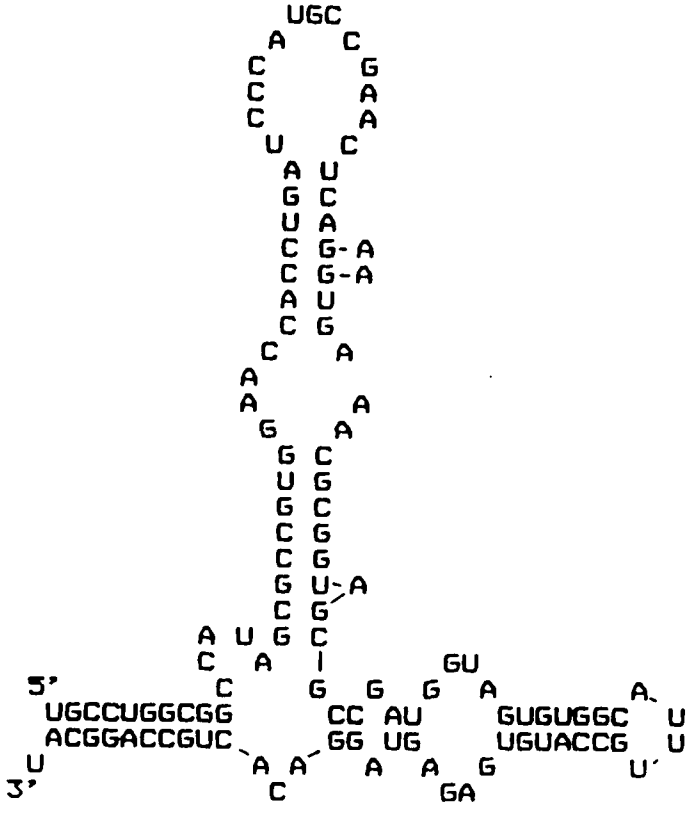
Vibrio marinus (MP-1)



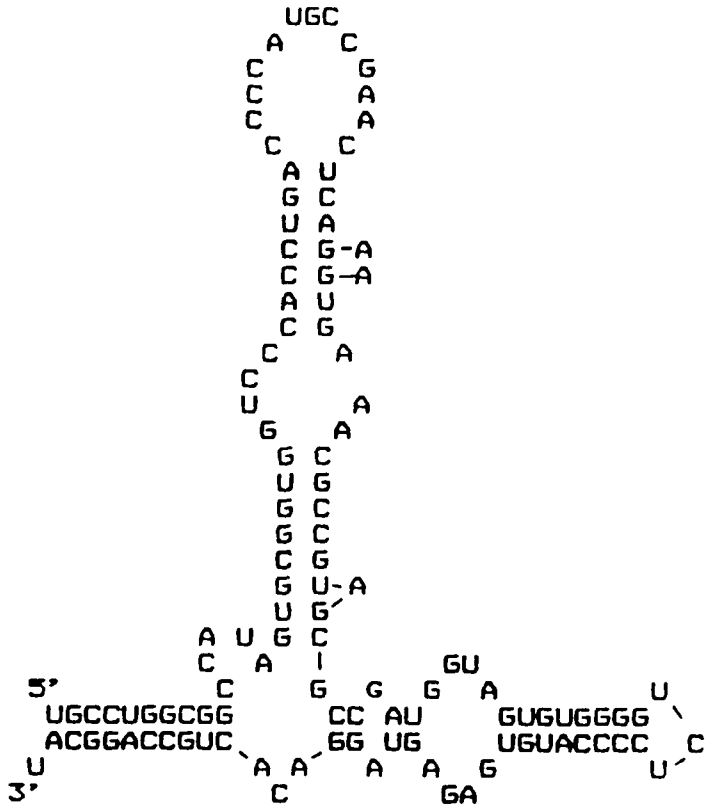
Vibio anguillarum



Vibrio psychroerythrus



Aeromonas hydrophila



Plesiomonas shigelloides

REFERENCES

1. Bang, S.S., L. Baumann, M.J. Woolkalis, and P. Baumann. 1981. Evolutionary relationships in *Vibrio* and *Photobacterium* as determined by immunological studies of superoxide dismutase. Arch. Microbiol 130:111-120.

2. Baumann, L., S.S. Bang, and P. Baumann. 1980. Study of relationship among species of *Vibrio*, *Photobacterium*, and terrestrial enterobacteria by an immunological comparison of glutamine synthetase and superoxide dismutase. Curr. Microbiol. 4:133-138.

3. Baumann, P., L. Baumann, S.S. Bang, and M.L. Woolkalis. 1980. Reevaluation of the taxonomy of *Vibrio*, *Beneckea*, and *Photobacterium*: abolition of the genus *Beneckea*. Curr. Microbiol. 4:127-132.

4. Baumann, P., L. Baumann, M.J. Woolkalis, and S.S. Bang. 1983. Evolutionary relationships in *Vibrio* and *Photobacterium*: a basis for a natural classification. Annu. Rev. Microbiol. 37:369-398.

5. Baumann, P., A.L. Furniss, and J.V. Lee. 1984. Vibrio. *In* N.R. Krieg and J.G. Holt (eds.), Bergey's Manual of Systematic Bacteriology, Williams and Wilkins, Baltimore.

6. Baumann, P. and R.H.W. Schubert. 1984. Vibrionaceae. *In* N.R. Krieg and J.G. Holt (eds.), Bergey's Manual of Systematic Bacteriology, Williams and Wilkins, Baltimore.

7. Belozersky, A.N. and A.S. Spirin. 1960. Chemistry of the nucleic acids of microorganisms. *In* E. Chargaff and J.N. Davidson (eds.), The Nucleic Acids, Vol. 3, Academic Press, New York.

8. Bergonzini, C. 1879. I bacteri. Annuario della Societa dei Naturalisti in Modena, ser. 2, Anno 13:19.

9. Bisset, K.A. 1950. Evolution in bacteria and the significance of the bacterial spore. Nature 166:431.

10. Bisset, K.A. 1952. Bacteria. Livingstone, Edinburgh.

11. Bisset, K.A. 1957. Some observations on the mode of sporulation and relationships of monosporous actinomycetes. J. Gen. Microbiol. 17:562-566.

12. Bogusky, M.S., P.A. Hieter, and C.C. Levy. 1980. Identification of a cytidine-specific ribonuclease from chicken liver. J. Biol. Chem. 255:2160-2163.

13. Brownlee, G.G., F. Sanger, and B.G. Barrell. 1967. Nucleotide sequence of 5S ribosomal RNA from *Escherichia coli*. Nature 215:735-736.

14. Butzow, J.J. and G.L. Eichhorn. 1975. Different susceptibility of DNA and RNA to cleavage by metal ions. Nature 254:358-359.

15. Carmichael, G.G. 1980. Molecular weight determinations of RNA by gel electrophoresis 1:78-82.

16. Citeralla, R.N. and R.R. Colwell. 1970. Polyphasic taxonomy of the genus *Vibrio*: polynucleotide sequence relationships among selected *Vibrio* species. J. Bacteriol. 104:434-442.

17. Cohn, F. 1872. Bacteria: the smallest of living organisms. *Translated by* C.S. Dolley. *Reprinted from* Bull. Hist. Med. 7:49-92.

18. Collins, M.D. and D. Jones. 1979. Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implications. Microbiological Rev. 45:316-354.

19. Colwell, R.R. 1968. Polyphasic taxonomy of bacteria. *In* H. Iizuka and R. Hasegawa (eds.), Culture Collections of Microorganisms. University of Tokyo Press, Tokyo. pp. 421-436.

20. Colwell, R.R. 1970. Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae, Vibrio parahaemolyticus*, and related *Vibrio* species. J. Bacteriol. 104:410-433.

21. Colwell, R.R. 1971. The value of numerical taxonomy in bacterial systematics. *In* A. Perez-Miravete and D. Pelaez (eds.), Recent Advances in Microbiology, Xth Intl. Cong. Microbiol., Mexico City.

22. Colwell, R.R. and M. Mandel. 1964. Adansonian analysis and deoxyribonucleic acid base composition of some Gram-negative bacteria. J. Bacteriol. 87:1412-1422.

23. Cowan, S.T. 1962. The microbial species -- a macromyth? *In* G.C. Ainsworth and P.H.A. Sneath (eds.), Microbial Classification. Society for General Microbiology, Cambridge. pp. 433-455.

24. D'Alessio, J.M. 1983. RNA sequencing. *In* D. Rickwood and B.D. Hames (eds.), Gel Electrophoresis of Nucleic Acids, IRL Press, London.

25. D'Aoust, J.Y. and D.J. Kushner. 1972. *Vibrio psychroerythrus* sp. n.: classification of the psychrophilic marine bacterium NRC 1004. J. Bacteriol. 111:340-342.

26. Dams, E., P. Londei, P. Cammarano, A. Vandenberghe, and R. DeWachter. 1983. Sequences of the 5S rRNAs of the thermoacidophilic archaebacterium *Sulfolobus solfataricus (Caldariella acidophila)* and the thermophilic eubacteria *Bacillus acidocaldarius* and *Thermus aquaticus*. Nucl. Acids Res. 11:4667-4676.

27. Dams, E., A. Vandenberghe, and R. DeWachter. 1982. Sequences of the 5S rRNAs of *Azotobacter vinelandii, Pseudomonas aeruginosa* and *Pseudomonas fluorescens* with some notes on 5S RNA secondary structure. Nucl. Acids Res. 11:1245-1252.

28. Dekio, S., R. Yamasaki, J. Jidoe, H. Hori, and S. Osawa. 1984. Secondary structure and phylogeny of *Staphylococcus* and *Micrococcus* 5S rRNAs. J. Bacteriol. 159:233-237.

29. Delihas, N. and J. Andersen. 1982. Generalized structures of the 5S ribosomal RNAs. Nucl. Acids Res. 10:7323-7344

30. Deming, J.W., H. Hada, R.R. Colwell, K.R. Luehrsen, and G.E. Fox. 1984. The ribonucleotide sequence of 5S rRNA from two strains of deep-sea barophilic bacteria. J. Gen. Microbiol. 130: 1911-1920.

31. De Vos, P. and J. De Ley. 1983. Intra- and intergeneric similarities of *Pseudomonas* and *Xanthomonas* ribosomal ribonucleic acid cistrons. Int. J. Syst. Bacteriol. 33:487-509.

32. DeWachter, R., M.-W. Chen, and A. Vandenberghe. 1982. conservation of secondary structure in 5S ribosomal RNA: a uniform model for eukaryotic, eubacterial, archaebacterial and organelle sequences is energetically favourable. Biochimie 64:311-329.

33. Donis-Keller, H. 1979. Site specific enzymatic cleavage of RNA. Nucl. Acids Res. 7:179-192.

34. Donis-Keller, H. 1980. Phy M: an RNase activity specific for U and A residues useful in RNA sequence analysis. Nucl. Acid Res. 3133-3142.

35. Donis-Keller, H., A.M. Maxam, and W. Gilbert. 1977. Mapping adenines, guanines, and pyrimidines in RNA. Nucl. Acids Res. 4:2527-2538.

36. Douthwaite, S., A. Christensen, and R.A. Garrett. 1983. Higher order structure in the 3'-minor domain of small subunit ribosomal RNAs from a Gram negative bacterium, a Gram positive bacterium, and a eukaryote. J. Mol. Biol 169:249-279.

37. Dunn, J.J. and F.W. Studier. 1973. T7 early RNAs and *E. coli* ribosomal RNAs are cut from large precursor RNAs *in vivo* by ribonuclease III. Proc. Natl. Acad. Sci. (USA) 70:3296-3300.

38. Erdmann, V.A., E. Huysmans, A. Vandenberghe, and R. De Wachter. 1983. Collection of published 5S and 5.8S ribosomal RNA sequences. Nucl. Acids Res. 11:r105-r133.

39. Erdmann, V.A., J. Wolters, E. Huysmans, A. Vandenberghe, and R. De Wachter. 1984. Collection of published 5S and 5.8S ribosomal RNA sequences. Nucl. Acids Res. 12:r133-r166.

40. Ferguson, W.W. and N.D. Henderson. 1947. Description of strain C27: a motile organism with the major antigen of *Shigella sonnei* phase I. J. Bacteriol. 54:179-181.

41. Fox, G.E., K.R. Luehrsen, and C.R. Woese. 1982. Archaebacterial 5S ribosomal RNA. Zbl. Bakt. Hyg., I. Abt. Orig. C3:330-345.

42. Fox, G.E., K.E. Pechman, and C.R. Woese. 1977. Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to prokaryotic systematics. Int. J. Syst. Bacteriol. 27:44-57.

43. Fox, G.E., E. Stackebrandt, R.B. Hespell, J. Gibson, J. Maniloff, T.A. Dyer, R.S. Wolfe, W.E. Balch, R.S. Tanner, L.J. Magrum, L.B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B.J. Lewis, D.A. Stahl, K.R. Luehrsen, K.N. Chen, and C.R. Woese. 1980. The phylogeny of the prokaryotes. Science 209:457-463.

44. Fox, G.E. and C.R. Woese. 1975. 5S RNA secondary structure. Nature 256:505-507.

45. Fox, G.E. and C.R. Woese. 1975. The architechture of 5S rRNA and its relation to function. J. Mol. Evol. 6:61-76.

46. Hansen, J.N. 1976. Electrophoresis of ribonucleic acid on a poly-acrylamide gel which contains disu[illegible]e cross-linkages. Anal. Biochem 76: 37-44.

47. Hansen, J.N., B.H. Pheiffer, and J.A. Boehnert. 1980. Chemical and electrophoretic properties of solubilizable disulfide gels. Anal. Biochem. 105:192-201.

48. Hansen, J.N. 1981. Use of solubilizable acrylamide disulfide gels for isolation of DNA fragments suitable for sequence analysis. Anal. Biochem. 116:146-151.

49. Hassur, S.M. and W. Whitlock. 1974. UV shadowing -- a new and convenient method for the location of ultraviolet-absorbing species in polyacrylamide gels. Anal. Biochem. 59:162-164.

50. Hendrie, M.S., J.M. Shewan, and M. Veron. 1971. *Aeromonas shigelloides* (Bader) Ewing *et. al.*: a proposal that it be transferred to the genus *Vibrio*. Int. J. Syst. Bacteriol. 21:25-27.

51. Hecht, N.B. and C.R. Woese. 1968. Separation of bacterial ribosomal ribonucleic acid from its macromolecular precursors by poly-acrylamide gel electrophoresis. J. Bacteriol. 95:986-990.

52. Hill, L.R. 1966. An index to deoxyribonucleic acid base composition of bacterial species. J. General Microbiol. 44:419-437.

53. Hori, H. and S. Osawa. 1979. Evolutionary change in 5S RNA secondary structure and a phylogenetic tree of 54 5S RNA species. Proc. Natl. Acad. Sci. (USA) 76:381-385.

54. Ikemura, T. and J.E. Dahlberg. 1973. Small ribonucleic acids of *Escherichia coli*. J. Biol. Chem. 248:5024-5032.

55. Kime, M.J. and P.B. Moore. 1982. NMR evidence for the existence of two native conformations of 5S RNA. Nucl. Acids. Res. 10:4973-4983.

56. Kime, M.J. and P.B. Moore. 1983. Nuclear Overhauser experiments at 500 MHz on the downfield proton spectrum of a ribonuclease-resistant fragment of 5S ribonucleic acid. Biochem. 22:2615-2622.

57. Kime, M.J. and P.B. Moore. 1983. Physical evidence for a domain structure in *Escherichia coli* 5S RNA. FEBS Lett. 153:199-203.

58. Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111-120.

59. Kirby, K.S. 1956. A new method for the isolation of ribonucleic acids from mammalian tissues. Biochem J. 64:405-408.

60. Klotz, L.C., N. Komar, R.L. Blanken, and R.M. Mitchell. 1979. Calculation of evolutionary trees from sequence data. Proc. Natl. Acad. Sci. (USA) 76:4516-4520.

61. Kuntzel, H., M. Heidrich, and B. Piechulla. 1981. Phylogenetic tree derived from bacterial, cytosol and organelle 5S rRNA sequences. Nucl. Acids. Res. 9:1451-1461.

62. Kuntzel, H. 1982. Phylogenetic trees derived from mitochondrial, nuclear, eubacterial and archaebacterial rRNA sequences: implications on the origin of eukaryotes. Zbl. Bakt. Hyg., I Abt. Orig. C3:31-39.

63. Kuntzel, H., B. Piechulla, and U. Hahn. 1983. Concensus structure and evolution of 5S rRNA. Nucl. Acids Res. 11: 893-900.

64. Lee, K.Y., R. Wahl, and E. Barbu. 1956. Contenu en bases puriques et pyrimidiques des acides desoxyribonucleiques des bacteries. Ann. Inst. Pasteur 91:212-224.

65. Leontis, N.B. and P.B. Moore. 1984. A small angle X-ray scattering study of a fragment derived from *E. coli* 5S RNA. Nucl. Acids Res. 12:2193-2203.

66. Levy, C.C. and T.P. Karpetsky. 1980. The purification and properties of chicken liver RNase. J. Biol. Chem. 255:2153-2159.

67. Li, W.-H. 1981. Simple method for constructing phylogenetic trees from distance matrices. Proc. Natl. Acad. Sci. (USA) 78:1085-1089.

68. Lillehaug, J.R., R.K. Kleppe, and K. Kleppe. 1976. Phosphorylation of double-stranded DNAs by T_4 polynucleotide kinase. Biochem. 15:1858-1865.

69. Lockard, R.E., B. Alzner-Deweerd, J.E. Heckman, J. MacGee, M.W. Tabor, and U.L. RajBhandary. 1978. Sequence analysis of 5'[^{32}P] labeled mRNA and tRNA using polyacrylamide gel electrophoresis. Nucl. Acids Res. 5:37-56.

70. Loening, U.E. 1967. The fractionation of high-molecular weight ribonucleic acid by polyacrylamide gel electrophoresis. Biochem. J. 102:251-257.

71. Luehrsen, K.R. and G.E. Fox. 1981. The nucleotide sequence of *Beneckea harveyi* 5S rRNA. J. Mol. Evol. 17:52-55.

72. Lund, E., J.E. Dahlberg, and C. Guthrie. 1979. Processing of spacer tRNAs from rRNA transcripts of *E. coli*. *In* J. Abelson, D. Soll, and P. Schimmel (eds.), Transfer RNA. Cold Spring Harbor Monograph Series.

73. MacDonell, M.T. and R.R. Colwell. 1984a. Identical 5S rRNA nucleotide sequence of *Vibrio cholerae* strains representing temporal, geographical, and ecological diversity. Appl. Environ. Microbiol. 48:119-121.

74. MacDonell, M.T. and R.R. Colwell. 1984b. Nucleotide sequence of 5S ribosomal RNA from *Vibrio marinus*. Microbiological Sci. 1:(in press).

75. MacDonell, M.T. and R.R. Colwell. 1984c. Microcomputer program for the computation of free energy of the secondary structure of RNA. J. Microbiological Meth. 3:(in press).

76. MacDonell, M.T. and R.R. Colwell. 1984d. The nucleotide base sequence of *Vibrionaceae* 5S rRNA. FEBS Lett. 175:183-188.

77. MacDonell, M.T. and R.R. Colwell. 1984e. Nucleotide sequences of 5S rRNA from two *Vibrio* species, *V. fluvialis* and *V. parahaemolyticus*. MIRCEN J. Appl. Microbiol. Biotechnol. 1:(in press).

78. MacDonell, M.T. and R.R. Colwell. 1984f. The contribution of numerical taxonomy to the systematics of Gram-negative bacteria. *In* M. Goodfellow and D. Jones (eds.), Current Topics in Computer Assisted Bacterial Systematics, Academic Press, London.

79. MacDonell, M.T. and J.N. Hansen. 1985. Isolation and purification of RNAs suitable for sequence analysis. *In* R.R. Colwell and R. Grigorova (eds.), Methods in Microbiology 18:(in press).

80. MacKay, R.M., D.F. Spencer, M.N. Schnare, W.F. Doolittle, and M.W. Gray. 1982. Comparative sequence analysis as an approach to evaluating structure, function, and evolution of 5S and 5.8S ribosomal RNAs. Can. J. Biochem. 60:480-489.

81. Maniatis, T., E.R. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor.

82. Maxam, A.M. and Gilbert, W. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzmol. 65: 499-560.

83. Nikolaev, N., L. Silengo, and D. Schlessinger. 1963. Synthesis of a large precursor to ribosomal RNA in a mutant of *E. coli*. Proc. Natl. Acad. Sci. (USA) 70:3361-3370.

84. Ninio, J. 1979. Prediction of pairing schemes in RNA molecules -- loop contributions and energy of wobble and non-wobble pairs. Biochimie 61:1133-1150.

85. Noller, H.F. 1984. Structure of ribosomal RNA. Annu. Rev. Biochem. 53:119-162.

86. Orla-Jensen, S. 1909. Die Hauptlinien des naturlichen Bakteriensystems. Zbl. Bakt. 22:305.

87. Peattie, D.A. 1979. Direct chemical method for sequencing RNA. Proc. Natl. Acad. Sci. (USA) 76:1760-1764.

88. Peterson, E.A. 1980. Cellulosic Ion Exchangers. *In* T.S. Work and E. Work (eds.), Laboratory Techniques in Biochemistry and Molecular Biology. Elsevier North Holland, New York.

89. Pieler, T. and V.A. Erdmann. 1982. Three-dimensional structural model of eubacterial 5S RNA that has functional implications. Proc. Natl. Acad. Sci. (USA) 79:4599-4603.

90. Pieler, T., I. Kumagai, and V.A. Erdmann. 1982. Phylogenetic diversity reflected in the three dimensional structure of ribosomal 5S RNA. Zbl. Bakt. Hyg., I. Abt. Orig. C3: 69-78.

91. Pieler, T., A. Schreiber, and V.A. Erdmann. 1984. Comparative structural analysis of eubacterial 5S rRNA by oxidation of adenines in the N-1 position. Nucl. Acids Res. 12:3115-3126.

92. Randerath, K. and E. Randerath. 1967. Thin-layer separation methods for nucleic acid derivatives. Meth. Enzymol. 12A:323-347.

93. Richardson, C.C. 1965. Phosphorylation of nucleic by an enzyme from T4 bacteriophage-infected *E. coli*. Proc. Natl. Acad. Sci. (USA) 54:158-165.

94. Sanger, F., G.G. Brownlee, and B.G. Barrell. 1965. A two-dimensional fractionation procedure for radiolabeled nucleotides. J. Mol. Biol. 13:373-398.

95. Sanger, F. and A.R. Coulsen. 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. 87:107-110.

96. Sato, K. and Egami, F. 1957. Studies on ribonucleases in takadiastase. J. Biochem. 44:753-767.

97. Schleifer, K.H., R. Kilpper-Balz, U. Fischer, A. Faller, and J. Endl. 1982. Identification of *"Micrococcus candidus"* ATCC 14852 as a strain of *Staphylococcus epidermidis* and of *"Micrococcus caseolyticus"* ATCC 13548 and *Micrococcus varians* ATCC 29750 as members of a new species, *Staphylococcus caseolyticus*. Int. J. Syst. Bacteriol. 32:15-20.

98. Schleif, R.F. and P.C. Wensink. 1981. Practical methods in molecular biology. Springer-Verlag, New York.

99. Schwartz, R.M. and M.O. Dayhoff. 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. Science 199:395-403.

100. Shepherd, J.C.W. 1981. Periodic correlations in DNA sequences and evidence suggesting their evolutionary origin in a comma-less genetic code. J. Mol. Evol. 17:94-102.

101. Shepherd, J.C.W. 1982. From primeval message to present-day gene. Cold Spring Harbor Symp. Quant. Biol. 47:1099-1108.

102. Shimwell, J.L. and J.G. Carr. 1961. The type culture concept in *Acetobacter* and other genera. Leeuwenhoek ned. Tijdschr. 27:65-73.

103. Sneath, P.H.A. 1962. The construction of taxonomic groups. *In* G.C. Ainsworth and P.H.A. Sneath (eds.), Microbial Classification. Society for General Microbiology, Cambridge. pp. 289-332.

104. Sogin, S.J., M.L. Sogin and C.R. Woese. 1972. Phylogenetic measurement in procaryotes by primary structural characterization. J. Mol. Evol. 1:173-184.

105. Stackebrandt, E. and C.R. Woese. 1984. The phylogeny of prokaryotes. Microbiological Sci. 1:117-122.

106. Stahl, D.A., D.J. Lane, G.J. Olsen and N.R. Pace. 1984. Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. Science 224:409-411.

107. Thornley, M.J. 1967. A taxonomic study of *Acinetobacter* and related genera. J. Gen. Microbiol. 49:211-257.

108. Tinoco, I., O.C. Uhlenbeck, and M.D. Levine. 1971. Estimation of secondary structure in ribonucleic acids. Nature 230:362-367.

109. Trifonov, E.N. and G. Bolshoi. 1983. Open and closed 5S ribosomal RNA, the only two universal structures encoded in the nucleotide sequences. J. Mol. Biol. 169:1-13.

110. Uchida, T., L. Bonen, H.W. Schaup, B.J. Lewis, L. Zablen, and C.R. Woese. 1974. The use of ribonuclease U_2 in RNA sequence determinations. J. Mol. Evol. 3:66-77.

111. Van Landschoot, A. and J. De Ley. 1983. Intra- and intergeneric similarities of the rRNA cistrons of *Alteromonas*, *Marinomonas* (gen. nov.) and some other Gram-negative bacteria. J. Gen. Microbiol. 129:3057-3074.

112. Veron, M.M. 1965. La position taxonomique des *Vibrio* et de certaines bacteries comparables. C.R. Acad Sci. Paris 261:5243-5246.

113. Vogel, D.W., R.K. Hartmann, M. Bartsch, A.R. Subramanian, W. Kleinow, T.W. O'Brien, T. Pieler, and V.A. Erdmann. 1984. Reconstitution of 50S ribosomal subunits from *Bacillus stearothermophilus* with 5S RNA from spinach chloroplasts and low-M$_r$ RNA from mitochondria of *Locusta migratoria* and bovine liver. FEBS Lett. 169:67-72.

114. Vogt, V.M. 1980. Purification and properties of S1 nuclease from *Aspergillus*. Meth. Enzymol. 65:248-264.

115. West, P.A. and R.R. Colwell. 1984. Identification of *Vibrionaceae* -- an overview. *In* R.R. Colwell (ed.), Vibrios in the Environment. Wiley Interscience, New York.

116. Woese, C.R. 1970. The genetic code in prokaryotes and eukaryotes. *In* H.P. Charles and B.C. Knight (eds.), Organization and Control in Prokaryotic and Eukaryotic Cells: 20th Symp., Soc. Gen. Microbiol. pp 39-54.

116. Woese, C.R. 1982. Archaebacteria and cellular origins: an overview. Zbl. Bakt. Hyg., I Abt. Orig. C3:1-17.

117. Woese, C.R., P. Blanz, and C.M. Hahn. 1984. What isn't a pseudomonad: the importance of nomenclature in bacterial classification. Syst. Appl. Microbiol. 5:179-195.

118. Woese, C.R., C.D. Pribula, G.E. Fox, and L.B. Zablen. 1975. The nucleotide sequence of the ribosomal RNA from a photobacterium. J. Mol. Evol. 5:35-46.

119. Woolkalis, M.J. and P. Baumann. 1981. Evolution of alkaline phosphatase in marine species of *Vibrio*. J. Bacteriol. 147: 36-45.

120. Zablen, L., L. Bonen, R. Meyer, and C.R. Woese. 1975. The phylogenetic status of *Pasteurella pestis*. J. Mol. Evol. 4:347-358.