A BIOCHEMICAL PHYLOGENY OF THE PROTISTS

MARK A. RAGAN
Department of Biology
Dalhousie University
Nova Scotia, Canada

DAVID J. CHAPMAN
Department of Biology
University of California
Los Angeles, California

ACADEMIC PRESS New York San Francisco London 1978
A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Ragan, Mark A
A biochemical phylogeny of the protists.

Bibliography: p.
Includes indexes.
1. Unicellular organisms–Classification.
2. Chemotaxonomy. I. Chapman, D. J., joint author.
II. Title. [DNLM: 1. Biochemistry. 2. Evolution.
3. Microbiology. QW4 R141b]
QR12.R34 576 76-52739
ISBN 0–12–575550–3

PRINTED IN THE UNITED STATES OF AMERICA

Preface

Biochemical characters have been used for more than a century in the reconstruction of phylogenies. It is only in the last two decades, however, that the underlying concepts of biochemistry and molecular biology have begun to exert a significant influence upon phylogenetics. Perhaps the most spectacular example of this progress has been the realization that amino acid sequences of most proteins are a direct reflection of the nucleic acid-based genome, and hence, of the phylogenetic history of organisms.

The proliferation of biochemical data is especially welcome in phylogenetic examination of the protists, where there is a chronic shortage of objective phylogenetic characters. Yet biochemical characters have been utilized by different authors to support quite different and sometimes mutually exclusive phylogenies. Moreover, there is considerable difficulty in approaching much of the relevant primary literature, not only because it is often extremely specialized (if not arcane), but also because it is well dispersed throughout numerous books and scientific journals (we have drawn data from 180 journals and 118 monographs, symposia, and theses). These problems loom especially large for the student and the nonspecialist approaching biochemical phylogenetics for the first time. In this book we seek to examine a broad spectrum of biochemical characters; to point out which ones have been useful in phylogenetics, and the underlying bases of such usefulness; and to illustrate methods of deducing phylogenies from biochemical data. Our efforts will have been justified to the extent that this book serves not as an arbiter of phylogenetic questions, but as a stimulus and guide to further thought and research.

A preliminary draft of some of this work had its origin at the University of Chicago in February 1972; helpful discussions were held at that time with Drs. J. H. Law and T. H. Steck. Dr. G. S. Getz kindly provided then-unpublished data. During preparation of the final draft, valuable advice has been received from Drs. N. J. Antia, T. Christensen, J. S. Craigie, W. F. Doolittle, L. J. Goad, T. W. Goodwin, D. O. Hall, M. V. Laycock, H. Matsubara, K. K. Rao, and J. W. Schopf. Drs. J. S. Craigie, M. V. Laycock, and P. J. McLaughlin have supplied unpublished data. Many

colleagues have provided us with access to manuscripts in advance of publication. To all of the above we express our thanks. We also thank the scientists and publishers who have generously permitted us to use copyrighted materials. These have been acknowledged throughout the text. However, we must assume all responsibility for the interpretation of these data. Our special thanks go to Marjorie McDonald. Her typing skills brought the final manuscript to the light of day.

M. A. R. wishes to thank the Isaac Walton Killam Trust and the National Research Council of Canada for financial support while the final draft was being completed. D. J. C. expresses his thanks to the National Science Foundation and the Regents of the University of California for their support of his research over the years.

MARK A. RAGAN
DAVID J. CHAPMAN

Contents

Preface ix

1 Introduction 1

1.1 What Are the Protists? 1
1.2 Why Are the Protists Interesting? 2
1.3 Systematics, Taxonomy, and Phylogeny 3
1.4 Why a Biochemical Phylogeny? 3

2 Biochemistry and Evolution 6

2.1 The Nature of Biochemical Data and the Central Dogma 6
2.2 Data Weighting and the Central Dogma 7
2.3 Prebiotic Processes and "Biochemical Predestination" 10
2.4 Darwinian and Non-Darwinian Evolution 12
2.5 Biochemistry and Morphology 15
2.6 Ontogeny and Phylogeny 16
2.7 A Note on the Biochemical Method 16

3 Phylogenetics 18

3.1 Phylogenetic Trees 18
3.2 Choice of Experimental Organisms 19
3.3 The Endosymbiotic Theory of Organelle Evolution 20

4 Nucleic Acids 28

4.1 Evolution of the Genetic Code 28
4.2 DNA: Structure and Composition 29
4.3 RNA: Structures and Composition 42
4.4 Ribosomes 51

5 Proteins 56

Part I: General Considerations 56

5.1 Polypeptides, Proteins, and Enzymes 56
5.2 Phylogeny and the Structure of Enzymes: General and Theoretical Considerations 57
5.3 Enzyme Aggregates 60
5.4 Evolution of Biosynthetic Pathways 60
5.5 Evolution of Metabolic Energy Production 62
5.6 Biochemical Methods in the Phylogenetic Study of Proteins 64

Part II: Heme Proteins, Metalloproteins, and Histones 65

5.7 Cytochromes 65
5.8 Heme Proteins Other than Cytochromes 69
5.9 Phycobiliproteins 69
5.10 Ferredoxins 73
5.11 Metalloproteins Other than Ferredoxins 76
5.12 Histones 77

6 Proteins: Enzymes 81

6.1 Embden–Meyerhof–Parnas Glycolysis 81
6.2 Pentose Phosphate Pathway 87
6.3 Tricarboxylic Acid Cycle 89
6.4 Hatch–Slack Pathway 93
6.5 Glycolate Oxidation 94
6.6 Calvin Cycle 97
6.7 Polysaccharide Biosynthetic Enzymes 100
6.8 Fatty Acid Synthases 104
6.9 Lysine Biosynthetic Pathways 105
6.10 Ornithine Biosynthesis and the Ornithine–Citrulline Cycle 111
6.11 Isoleucine Biosynthesis Pathways 112
6.12 Tryptophan Biosynthesis and the Polyaromatic Biosynthetic Pathway 115
6.13 Glutamate Dehydrogenases 128
6.14 Lactate and Malate Dehydrogenases 129
6.15 Nitrate Reductases 129

7 Metabolites: I 131

7.1 Monosaccharides and Oligosaccharides 131
7.2 Polysaccharides 136
7.3 Amino Acid Distribution 143

8 Metabolites: II 147

8.1 Acetate-Derived Biosynthetic Pathways 147
8.2 Fatty Acids 148
8.3 Lipids 159
8.4 Acetylenic Compounds 164
8.5 Compounds Derived from Shikimic Acid 166
8.6 Other Routes to Aromatic Compounds 167

9 Metabolites: III 169

9.1 The Mevalonic Acid Pathway: Isoprenoids 169
9.2 Phytol, Geranylgeraniol, and Farnesol 171
9.3 Quinones 172
9.4 Carotenoids 173
9.5 Sterols 185

10 Metabolites: IV 197

10.1 The δ-Aminolevulinic Acid Pathway 197
10.2 Chlorophylls 198
10.3 Photosynthesis 204

11 Miscellaneous Simple Molecules 206

11.1 Sulfate Reduction 206
11.2 Nitrogen Utilization 207
11.3 Carbon Monoxide Production 208
11.4 Mineral Nutrition and Vitamin Requirements 209

12 A Biochemical Phylogeny 211

12.1 Other Phylogenies in the Literature 211
12.2 A Biochemical Phylogeny 211
12.3 Comparison of These Phylogenies 213

13 Conclusions and Speculations 233

13.1 Reflections on the Biochemical Phylogeny 233
13.2 Time Course of Evolution 234
13.3 Important Unresolved Questions 238

Appendix 241

Bibliography 245

Taxonomic Index 285

Subject Index 296

1

Introduction

1.1 WHAT ARE THE PROTISTS?

The word "protist" was coined by Haeckel (1866) to describe the morphologically simple forms of life, including bacteria, fungi, many algae, protozoa, and sponges. Seventy years later Chatton (1937) emphasized the two basic types of cellular organization, which he designated as "procaryotic" and "eucaryotic." This basic division is now a well-established tenet of taxonomic and phylogenetic thinking. Copeland (1938) and Stanier and van Niel (1941) subsequently reclassified the prokaryotic protists (bacteria and blue-green algae) as "monera," and retained the term "protista" for eukaryotic forms. Later Dougherty and Allen (1953) recognized the lower protists (prokaryotes), "mesoprotists" (red algae), and "metaprotists" (remaining eukaryotic algae, fungi, and protozoa). In more recent years the Dinophyceae (dinoflagellates) have also been considered "mesoprotists," particularly in the Russian literature and by Dodge (1965), but in a different context from that used by Dougherty and Allen (1953). These latter workers considered the Rhodophyceae (red algae) to be "intermediate" between the prokaryotic Cyanophyceae (blue-green algae) and other eukaryotic algae. The distinction was made primarily on the basis of biochemistry (especially pigments) and the lack of flagellate structures in the Rhodophyceae. Dodge (1965), on the other hand, established his "mesoprotists" or "mesocaryotes" by the sole criteria of nuclear structure and nuclear behavior. We do not intend to argue the merits of the distinction into "mesoprotists" or "mesocaryotes," except to mention that the terms have not received the universal acceptance accorded the prokaryote and eukaryote. The word "protist" will be used here to encompass the bacteria, blue-

green algae, actinomycetes, eukaryotic algae (including red algae and dinoflagellates), fungi, water molds, euglenoids, and protozoa. This approach retains the original sense of the term and avoids the tendency to create new and unnecessary terminology. It is, in effect, a convenience term, and does not imply or suggest a taxonomic or systematic entity. Other simple forms of life, including viruses and pleuropneumonia-like organisms, will not be discussed in this treatment. Viruses have been considered by Dougherty (1955), Evans (1960), and Joklik (1974).

We have chosen as convenient reference points, three taxonomic schemes for the eukaryotic organisms under discussion (Appendix). Original taxonomic designations have been retained, in preference to more recent name changes or combinations. This approach may not represent adherence to systematic protocol or rules of nomenclature. However, we believe that less confusion will result if the old name, under which the biochemical or chemical investigations were carried out, is retained. For example, we will retain *Anacystis nidulans* (not *Lauterbornia*) and *Porphyridium cruentum* (not *P. purpureum*).

Many protists have been very incompletely studied by biochemists, e.g., *Kakabekia*-like organisms (Siegel and Giumarro, 1966; Siegel *et al.*, 1967), the Chloromonadophyceae, and cyanellae symbionts. The emphasis on certain algal groups, photosynthetic prokaryotes, and fungi, is symptomatic of our state of knowledge and the emphasis placed upon these organisms as experimental material in biochemical studies. Wherever possible, however, the lesser known groups have been considered, and projections, ideas, or suggestions with regard to these organisms have been put forward.

1.2 WHY ARE THE PROTISTS INTERESTING?

If one considers organisms around us that are readily visible, it becomes clear that for the most part they fall into two major categories: vascular plants and higher animals. Although a closer investigation would probably reveal mosses, ferns, earthworms, and insects, many morphological similarities within groups are nonetheless apparent among all these organisms. The biochemical processes characteristic of these organisms are even more uniform: all animals use very similar respiratory cytochromes, and all higher plants utilize identical chlorophylls in photosynthesis, and possess a very similar energy conversion or photosynthetic apparatus. It is in the protists that these relatively narrow ranges of body form, physiology, and especially biochemistry are found to vary most widely. The observed variations in morphology, physiology, and biochemistry often provide insights into the involvements of each in life processes.

Protists are very intimately involved in almost every aspect of most important ecological processes including oxygen production, disease, decay of organic matter, and nutrient cycling. The increasing incursion of industrialized man into previously undisturbed and perhaps finely balanced communities is likely to bring about significant changes in the lives of protists and consequently in human life.

1.3 SYSTEMATICS, TAXONOMY, AND PHYLOGENY

Since the days of Aristotle, scientists have engaged in the classification of organisms into hierarchical systems. Such activity not only helps to organize our knowledge of different forms of life, but should also allow people to make certain deductions about these organisms.

Taxonomy is the study of the bases, principles, procedures, and rules of classification (Heywood, 1973), or is the classificatory process itself (systematics). Apart from the nomenclatural rules, much of taxonomy is a matter of opinion. The desire for a "perfect" taxonomy must be tempered by pragmatism and by the need for convenience and ready applicability.

Phylogeny can be considered as a taxonomy in which the resulting system is thought to be representative of the historical evolution of the organisms considered. Descent from common ancestors with evolutionary modifications is studied in its various manifestations: morphological, biochemical, or otherwise. Consequently a phylogeny, adequately constructed, is a more powerful conceptual framework than is a taxonomy alone. The problem comes, however, in the modification "adequately constructed." What is an adequate construction, given that the fossil record is, and probably forever will be, incomplete or unintelligible concerning the details of protistan phylogeny? Is it intellectually valid to utilize biochemical characters in the reconstruction of phylogeny? What is to be done with data that do not appear to be in agreement with most other data? Can some criteria be considered more significant than others, and if so, on what basis? These questions will be considered in the following pages.

1.4 WHY A BIOCHEMICAL PHYLOGENY?

Why is it useful to construct a phylogeny of the protists using biochemical data? There are several answers to this question, none of which is complete in itself.

1. Biochemical data are *genetic*, being directly coded in the DNA ("primary semantides" or "primary semantophores" of Zuckerkandl and

Pauling, 1965a,b). Ribonucleic acids ("secondary semantides") and proteins ("tertiary semantides") are produced sequentially from the primary semantides, and consequently provide insight, although less directly, into the primary genetic makeup of the organism. There is of course environmental input into biochemical and physiological processes of living organisms, but where desired this can be disregarded or minimized by examination at the proper biochemical level.

2. In recent years there has been an enormous increase in the number of biochemical data available from protists. Many of these data have come from biochemists who utilize certain protists as favorable experimental systems, while other data have come from scientists interested in protists themselves. Construction of a biochemical phylogeny might help to organize some of these data, and suggest fruitful areas of further research.

3. Biochemical data have already been used to support a wide range of mutually conflicting and mutually exclusive phylogenies. It is important to examine these data to determine if they are internally consistent, and if so, how to use them properly in constructing phylogenies.

4. Biochemical methods may in certain circumstances be easier to apply than are more traditional examinations of morphology, ultrastructure, or life history.

Needless to say, there are also difficulties inherent in biochemical techniques and in their application to phylogenetics. It is not always easy to collect the necessary biochemical data, whether reaction pathways, molecular structures, or chemical compositions. There may be problems of the absence of a character (is it due to the repression of a gene, or to a critical mutational step that has occurred recently, or to insensitivity of the analytical method?). The culture conditions or an abnormal environment for the protist may cause some subtle, "unnatural" change in its biochemistry. Finally, only a percent or two of all known protists have even been studied at all by biochemists.

This raises the problem of the representative taxon: What species is representative of the genus, what genus of the family, and so on up the taxonomic ladder? Indeed, the question "Is there a representative taxon?" is rarely asked, and it is very easy to end up answering the question with circular reasoning. The concept of the "type" (e.g., type species of the genus; type genus of the family) in taxonomy is well understood. This, however, almost invariably applies only to the morphological realm. In view of the increasing use of chemical and biochemical data in taxonomy, it is appropriate to raise the question (Chapman and Ragan, 1977) of whether or not there should be a "chemical" type taxon, and if so, should this be the same as the morphological type. We believe these are basic questions, and are

even more important when it is realized that taxonomy and phylogeny are closely interwoven, and that biochemistry and morphology receive different emphasis in the two disciplines. Although conceptually convenient, typological systematics is poorly suited to deal with evolutionary events, as the confusion in primate paleontology has been vividly demonstrating in recent years (Paleontology Correspondent, 1974).

Evolution is a historical process that can profitably be examined from a number of viewpoints. The aim of this biochemical phylogeny is to introduce one possible way of viewing evolution. Other approaches may be more powerful in describing Darwinian selection and evolution at the organismal and population levels. It is not intended that our phylogeny, based upon one approach, should be used to the exclusion of other phylogenies or methods of construction. They should not be mutually exclusive, but rather they should act as a check and balance upon each other, since the greatest rewards lie in the final synthesis of all possible approaches to evolution and construction of a phylogeny.

2

Biochemistry and Evolution

2.1 THE NATURE OF BIOCHEMICAL DATA AND THE CENTRAL DOGMA

Data are statements of information derived from observation and consideration of characters.* Biochemical data are not necessarily true, due to the possibility of experimental error in the observation of organisms and in the execution of experiments. Some biochemical data incorporate other data, as will be seen, and hierarchies of biochemical data exist. In these regards such data do not differ from others except in their subject matter.

Data are rarely used alone; they are interpreted into facts, and are then used in hypothesis building and hypothesis testing. Are biochemical facts in any sense different from nonbiochemical ones? The answer is a qualified affirmative. In phylogenetics, for example, it can be argued that there is less subjectivity in interpreting biochemical data than in the interpretation of morphological or other types of data. This may be the "statistical objectivity" of Turner (1967). This reduced degree of subjectivity, the argument goes, increases the chances that the resulting facts will be of lasting value, and will place biochemical phylogenies on more solid grounds than is the case with other phylogenies. The opposing view is that this reduced

* The term "character" is often, but incorrectly, used interchangeably with the word "datum." Hennig (1966) has defined the former term in the context of the "character-bearing semaphoront," the unit of biological systematics. Semaphoronts are considered to be "individuals in given short periods of their lifetime" (Hennig, 1965).

degree of subjectivity is in reality a reduced level of interpretation, and as a consequence the resulting biochemical phylogeny is based upon more limited facts than are other phylogenies.

If there is indeed a difference between biochemical data and other types of data, it is that a *natural framework of interpretation* presents itself with biochemical data. The natural framework is the *biosynthetic history* of the observed molecule.

Retracing the biosynthesis of a metabolite (back through the appropriate biosynthetic enzymes, through the mRNA, to the DNA sequence(s) responsible for its biosynthesis) solves several problems endemic to phylogenetics: First, it demonstrates homology, if any, between two molecules. Second, differences in biosynthetic pathways may readily be observed. Third, problems arising from gene repression are, at least in theory, avoided. Finally, this approach provides natural groupings, aiding in the logistics of phylogeny building and data processing. This process is a natural one in that it follows, in retrograde, the actual history of the biochemical compound through real time.* The validity and usefulness of this method are suggested here without proof, but evidence for its applicability will be examined in later pages.

The transfer of information from DNA through RNA to the cellular proteins, some of which act as enzymes in the biosynthesis of various metabolites, is certainly one of the most fundamental life processes. In appreciation of the basic importance of these biological molecules and of this information transfer, the above process is often termed the "Central Dogma" of molecular biology. Although apparent evolutionary modifications of this process (reverse transcriptases, proviruses) have been discovered more recently, the importance of the Central Dogma processes becomes increasingly obvious with further biophysical and biochemical research.

2.2 DATA WEIGHTING AND THE CENTRAL DOGMA

Experience has shown—and the following chapters will document—that certain characters are of considerable phylogenetic usefulness, and that others are relatively uninteresting phylogenetically. From this observation, by no means limited to biochemical data (Mayr, 1969), there arises the perennial question: Is one type of datum *intrinsically more likely* to be of use in phylogenies than is a second type? Is it possible to predict phylogenetic usefulness *a priori*? If so, it would be routinely possible to

* Other meanings of the term "natural" have been discussed by Sneath and Sokal (1973).

weight heavily the inherently more useful data and to ignore the lesser, especially if the latter appeared to contradict the more useful ones.

Unfortunately no generalized classification of characters has been discovered. Suggestions that character X or character Y is intrinsically likely to be of great value in phylogenetics usually acquire a host of modifying conditions. This is frequently seen whenever an overly zealous researcher maps out the phylogeny of all organisms from the distribution of a single character; in essence this constitutes an extreme form of data weighting, in which all other data are assigned zero weight. It has been recognized since the work of Adanson (1763) that all possible useful data should be included in systematics.

At another extreme are the numerical or phenetic taxonomists who claim that all data must be lumped together without weighting into a computer, which then prints out the best available scheme. Many in this school do not claim that the resulting systems are even phylogenies, but can justly point out that their approach is relatively (although not completely!) free from personal biases in the interpretational stages. Indeed some data, such as isoenzyme patterns, are suited to this treatment. But it has been pointed out that numerical taxonomy is "getting the least out of the most" (Turner, 1967).

Between the two extremes lies current opinion. "To be sure," the current wisdom goes, "some characters have proved to be more useful than others in the construction of phylogenies. To be sure, there is some data weighting, if only subjectively, in most phylogenies, and this is a valid if necessarily *a posteriori* phenomenon. But it would be foolhardy to attempt to *predict* which characters will be of greater use, and which will be relatively useless." Experience has borne out this point of view quite well over the years, although the question has not been adequately discussed for biochemical characters. It is possible that justifications for data weighting could be found in biochemistry and molecular biology even if none was forthcoming for the traditional morphological characters.

There have been suggestions in recent years that phylogenetically interesting biochemical characters possess certain attributes (Erdtman, 1968):

1. Widely distributed identical characters are of little phylogenetic interest.
2. Characters unique to individual species are of little use in phylogenies, due to the impossibility of relating them to similar features in other organisms, and their very limited distribution.
3. Structurally complex molecules can often be weighted more heavily than can structurally simple molecules.

There is a unifying relationship underlying these observations: the concept of biosynthesis (Birch, 1973a,b). Compounds may be perceived as the

products of biosynthetic pathways. It may then be seen, in parallel with the above statements, that

1. If a compound is found in two organisms, but is biosynthesized by different pathways, its biosynthesis takes on considerable phylogenetic significance.

2. If a unique compound found in a given organism can be related biosynthetically (as a further elaboration, as a precursor, or as arising by the action of related enzymes) to a compound in another organism, there is the possibility that the two organisms are related phylogenetically. This approach is particularly useful when the biochemical pathway involves a very significant and unusual chemical modification of a molecule, such that one may assume with some justification that a very specific and discrete enzyme is involved in the establishment of the pathway. To a certain extent one is using an enzyme, albeit hypothesized, as the character. These tertiary semantides are nearer the genome than the episemantic molecules or metabolites. This "biosynthetic approach" has an advantage in that it may reduce the problems posed by the limitations of analytical techniques (the presence-or-absence question) and the inevitable question of possible environmental control or determination of the presence of a given compound. One is no longer using a single compound as the character (with the inherent problems), but rather a character represented by a series of compounds. Nevertheless this approach does not eliminate the possibility of gene repression.

3. Structurally complex compounds tend to be more significant because they presumably require more biosynthetic steps, hence more (specialized) enzymes (but cf. Herout, 1973). An extension of the concept of biosynthesis to include all related events from the level of DNA to metabolites could provide an even more powerful method for examining the relative usefulness of biochemical characters.

It might be argued that this approach does not come to grips with the problem of weighting individual biochemical characters, but instead merely increases the number of biochemical data by retracing the biosynthetic history of the compound [in the terminology of Sneath and Sokal (1973), amassing "logically correlated character complexes"]. Instead, by considering *in toto* a biosynthetic pathway or sequence containing the distinctive feature, rather than individual molecules, one is in fact reducing the number of characters and thus the number of biochemical data. Moreover, DNA is more than just another macromolecule; it is the primary genetic substance of the organism. All information necessary for cell development and activity is contained in the DNA, and if phylogeneticists could "read" DNA as the living cell does, they would be in a position to predict the struc-