PRINCIPLES OF NUMERICAL TAXONOMY

ROBERT R. SOKAL

The University of Kansas

PETER H. A. SNEATH

National Institute for Medical Research, London

THE UNIVERSITY OF MANITOBA ELIZABETH DAFOE LIBRARY

W. H. FREEMAN AND COMPANY

SAN FRANCISCO AND LONDON

A SERIES OF

BOOKS IN BIOLOGY

Editors: DOUGLAS M. WHITAKER
RALPH EMERSON
DONALD KENNEDY
GEORGE W. BEADLE (1946–1961)

Principles of Human Genetics (Second Edition)
Curt Stern

Experiments in General Biology
Graham DuShane and David Regnery

Principles of Plant Physiology
James Bonner and Arthur W. Galston

General Genetics
Adrian M. Srb and Ray D. Owen

An Introduction to Bacterial Physiology (Second Edition)
Evelyn L. Oginsky and Wayne W. Umbreit

Laboratory Studies in Biology: Observations and their Implications
Chester A. Lawson, Ralph W. Lewis, Mary Alice Burmester, and Garrett Hardin

Plants in Action: A Laboratory Manual of Plant Physiology
Leonard Machlis and John G. Torrey

Comparative Morphology of Vascular Plants
Adriance S. Foster and Ernest M. Gifford, Jr.

Taxonomy of Flowering Plants
C. L. Porter

Growth, Development, and Pattern
N. J. Berrill

Biology: Its Principles and Implications
Garrett Hardin

Animal Tissue Techniques
Gretchen L. Humason

Microbes in Action: A Laboratory Manual of Microbiology
Harry W. Seeley, Jr., and Paul J. VanDemark

Botanical Histochemistry: Principles and Practice
William A. Jensen

Modern Microbiology
Wayne W. Umbreit

Laboratory Outlines in Biology
Peter Abramoff and Robert G. Thomson

Molecular Biology of Bacterial Viruses
Gunther S. Stent

Principles of Numerical Taxonomy
Robert R. Sokal and Peter H. A. Sneath

Structure and Function in the Nervous Systems of Invertebrates
Theodore Holmes Bullock and G. Adrian Horridge

© Copyright 1963 by W. H. Freeman and Company

The publisher reserves all rights to reproduce this book, in whole or in part, with the exception of the right to use short quotations for review of the book.

Printed in the United States of America.

Library of Congress Catalog Card Number: 63-13914.

TO

JULIE AND JOAN

without whose patience
and understanding this book
could not have been written

We would stress the fact that, from the time of Linnaeus to our own, a weak point in biological science has been the absence of any quantitative meaning in our classificatory terms. What is a Class, and does Class A differ from Class B as much as Class C differs from Class D? The question can be put for the other classificatory grades, such as Order, Family, Genus, and Species. In no case can it be answered fully, and in most cases it cannot be answered at all. . . . Until some adequate reply can be given to such questions as these, our classificatory schemes can never be satisfactory or "natural." They can be little better than mnemonics—mere skeletons or frames on which we hang somewhat disconnected fragments of knowledge. Evolutionary doctrine, which has been at the back of all classificatory systems of the last century, has provided no real answer to these difficulties. Geology has given a fragmentary answer here and there. But to sketch the manner in which the various groups of living things arose is a very different thing from ascribing any quantitative value to those groups. [*Singer, 1959, p. 200.*]

PREFACE

It is widely acknowledged that the science of taxonomy is one of the most neglected disciplines in biology. Although new developments are continually being made in techniques for studying living creatures, in finding new characters, in describing new organisms, and in revising the systematics of previously known organisms, little work has been directed toward the conceptual basis of classification—that is, taxonomy in the restricted sense of the theory of classification. Indeed, the taxonomy of today is but little advanced from that of a hundred, or even two hundred, years ago. Biologists have amassed a wealth of material, both of museum specimens and of new taxonomic characters, but they have had little success in improving their power of digesting this material. The practice of taxonomy has remained intuitive and commonly inarticulate, an art rather than a science. And an uncritical attitude toward some aspects of evolutionary theory has compounded the confusion.

In the last few years there has been increased awareness of the problems in the aims and practices of taxonomy. In particular, there has been interest in the development of numerical methods in taxonomy as

an aid to making systematics a quantitative science, a step which comes in time to every scientific discipline.

Numerical taxonomy is the evaluation by numerical methods of the affinity or similarity between taxonomic units and the employment of these affinities in erecting a hierarchic order of taxa. The ideas on which numerical taxonomy rests go back to Adanson, a contemporary of Linnaeus, and have been repeatedly voiced. The present rapid development of these ideas is presumably a result of the development of computer techniques. Numerical taxonomy aims to develop methods which are objective and repeatable, both in evaluation of taxonomic affinity and in the erection of taxa. In addition, we believe that numerical methods may open up a wide field in the exact measurement of evolutionary rates and may provide a more critical approach to phylogenetic problems. It is worth noting that the success of the intuitive approach of the past lies in the ability of the mind to appreciate swiftly, though inexactly, overall similarity in morphological detail. This is not easy with data in tabular form, as with microbiological, chemical, or physiological data, which are now becoming so abundant; numerical methods are in these cases doubly necessary. The time now seems ripe to attempt a comprehensive treatment of these new advances.

It is the purpose of this book to present a firm theoretical basis for numerical taxonomy, to show why we believe numerical taxonomy has advantages over conventionally practiced taxonomy, to report on the various advances made in the field so far, and to furnish newcomers in the field with a detailed step-by-step description of the procedures employed in numerical taxonomy.

It is not our intention to treat at length all forms of numerical analysis which have been used in taxonomy for many special problems. There are numerous texts on the use of statistical and mathematical methods in biology; some are listed in the Appendix. We have therefore restricted the scope of this book to methods which are intended to demonstrate taxonomic relationships and to create taxonomic groupings, although some other techniques have been briefly treated for completeness. We have, however, attempted to make the treatment as broad as possible so as to be applicable to zoology, botany, microbiology, and paleontology, and other related sciences. This book is intended to serve practitioners in these sciences and will, we hope, invite them to look at their material in a new way. We hope also that students currently contemplating systematics as a career will wish to acquaint themselves with these methods and to evaluate their applicability to their own research.

In developing a theoretical foundation for numerical taxonomy we found it necessary to re-examine the theory of taxonomy as a whole, and we found much of the currently professed theory of phylogenetic systematics to be unsound and in need of critical re-evaluation. Of necessity, therefore, the book contains a critique of the methods and principles of phylogenetic systematics, in addition to a statement of our own views on numerical taxonomy.

Details on the use of this book are given in Chapter 1, "Introduction." The chapters dealing with the numerical methods will require a knowledge of statistics through correlation and regression. The rest of the book demands little or no mathematical background of the reader.

In preparing this book we have had the active assistance and encouragement of a number of colleagues. Professor Charles D. Michener of the University of Kansas has read and criticized the entire manuscript. He has been instrumental in much of the development of numerical taxonomy. Time and again he has pointed out weaknesses in our reasoning and has raised the problems of the practicing taxonomist relating to numerical taxonomy. Professor George W. Byers (University of Kansas) read the entire manuscript also and we are greatly indebted to him for his meticulous editorial care as well as his constructive criticism on many points. Portions of the book were read by Dr. W. T. Stearn (British Museum, Natural History), Dr. F. J. Rohlf (University of California, Santa Barbara), Professor Charles E. Leone (University of Kansas), and Professor H. Grüneberg, F.R.S., (University College, London). We are grateful for their valuable comments on their specialities.

An informal luncheon group of faculty and graduate students at the Entomology Department of the University of Kansas, "the Biosystematists," has had much to do with the origin of numerical taxonomy in America. The first ideas relating to this field were discussed by its members, and the manuscript for this book was read before this group and criticized by them. We are indebted to our colleagues and students in this group for many valuable suggestions and for helping us express our ideas more clearly than we might otherwise have done.

Preparation of this manuscript has involved the labors of many persons over a considerable period of time. Mr. Richard C. Rinkel and Mr. Anthony J. Boyce have helped in the preparation of the Appendix. A number of secretaries both in England and the United States have helped prepare the manuscript—our thanks go to Mrs. Elizabeth Ferrell, Mrs. Lois Harmon, Mrs. Maxine L. Howe, Miss Helen Lang-

Brown, Miss Barbara Mueller, Mrs. Dolores Vandermeer, and especially to Mrs. Betty Lou Deffley, who did the greatest part of the work. Mr. A. J. Lee of St. Albans, England, skillfully produced the figures from sketchy directions. Mr. Ivan Huber proofread the manuscript.

This book had its inception during the tenure of a Senior Postdoctoral Fellowship of the National Science Foundation (U.S.A.), which RRS spent at the Galton Laboratory, University College, London. Appreciation is expressed to Professor L. S. Penrose, F.R.S., head of the laboratory, for his cordial hospitality and encouragement.

The research leading to this book was financed by grants from the National Science Foundation (U.S.A.), the General Research Fund of the University of Kansas, and the Medical Research Council of Great Britain. We acknowledge gladly the encouragement of our work by our respective institutions, culminating in the preparation of this book. The authors, ordinarily domiciled on opposite sides of the Atlantic, were enabled to come together for completion of this work through a generous grant of the National Science Foundation.

We are indebted to Messrs. Abelard-Schuman Ltd., Publishers, for permission to cite a passage from C. Singer, *A History of Biology*. Other sources are acknowledged when used.

Lawrence, Kansas
March 1963

Robert R. Sokal
Peter H. A. Sneath