

Combinatorial generation of taxonomic diversity: Implication of symbiogenesis for the Proterozoic fossil record

Lynn Margulis and Joel E. Cohen

L.M., *Biology Department, University of Massachusetts, Amherst, MA 01003, USA*; J.E.C., *Rockefeller University, 1230 York Avenue, Box 20, New York, NY 10020-6399, USA*

Symbiogenesis is the emergence of new species with identifiably new physiologies and structures as a consequence of stable integration of symbionts. The development of symbiotic associations may lead to evolutionary innovation. Because of the pervasive influence of symbiogenesis in the origin of eukaryotic organisms, the Latin binomials of taxonomy should be recognized as applying to individuals who are greater than single homologous genetic units. All eukaryotes are composite (more than a single organismal type) and should be named and described accordingly. Symbiogenetic recombination of genomes may generate a striking diversity of both higher taxa and individual "species." A small number of symbionts, such as twenty different bacterial strains, individually and in all possible combinations with a host coleopteran, for example, could potentially generate more than a million distinct new species of beetles. A relatively small number of associates potentially can generate as much biological diversity as has been observed. The upper limit for one host and n symbionts is $2^n + n$ combinations. Rejecting the cladistic restriction of taxon origination by dichotomization of lineages, symbiogenesis requires depiction of evolution by anastomosing branches to form net-shaped phylogenies. We infer a relative paucity of symbiogenetically generated diversity in the Archean Eon. A far more significant amount accompanied the origin of undulipodiated and aerobic prototists in the Proterozoic Eon. Symbiogenesis may be especially significant for the emergence of skeletalized animals in the late Proterozoic and of plants, organisms far more desiccation-resistant than algae, in the Phanerozoic Eon. The polyphyletic acquisition of calcium-precipitating microbial symbionts may underlie the appearance of hard parts at the Proterozoic-Phanerozoic transition.

Symbiosis is the protracted physical association of organisms of different species. Symbiogenesis is the emergence of new species with new structures and physiologies, e.g., mitochondria and oxygen respiration, fish luminous organs, lichens, and oak galls, as a consequence of stable symbiotic associations. The term symbiogenesis was introduced by the Russian biologist Mereschkovsky in 1909 (Khakhina 1979). The importance of symbiogenesis as a mechanism of evolutionary innovation is explored in Margulis & Fester 1991. Symbiogenesis importantly supplements the gradual

accumulation of base-pair mutations, karyotypic rearrangements, and polyploidization. Yet to be determined is the relative importance of these evolutionary mechanisms, which are not mutually exclusive. Nevertheless, it is recognized that, unlike prokaryotes (bacteria, cyanobacteria, and actinobacteria – the last misnamed “actinomycetes,” as if they were fungi), all nucleated organisms (animals, plants, fungi, and prototists) are evolutionarily derived from early events of symbiogenesis that involved nucleocytoplasm and certain classes of crucial cellular organelles, e.g., oxygen-respiring mitochondria and photosynthetic plastids. The question as to whether other eukaryotic organelles such as peroxisomes (de Duve 1991), hydrogenosomes (Müller 1988; Johnson *et al.* 1990; Lahti & Johnson 1991), and kinetosomes (Margulis & McMenamin 1990) are also symbiotically derived from bacteria is unresolved. We argue here that, because of the pervasive influence of symbiogenesis in the origin of eukaryotic organisms, no individual eukaryotes began with fewer than two types of genomic systems. Hence their Latin binomial names should be reinterpreted as applying to ensembles of genomes, bionts, or symbionts that are integrated to form holobionts.

We also draw attention to the power of symbiogenetic recombination of genomes to generate a diversity of higher taxa and individual “species.” We pursue the implications of these ideas for the Archean and Proterozoic fossil record, pointing out that taxonomic practices across the disciplines (bacteriology, mycology, zoology, etc.) are incommensurate.

For quite different perspectives on the forces and patterns of evolution, see, e.g., Nei 1987 (especially Chapter 6 on “Genomic Evolution”), Eldredge 1989, Feldman 1989, and Raup 1991.

Individuals formed by genome integration

Paleontologists face the necessity of devising a useful taxonomy for the geological remains of formerly living communities, such as bioturbated sediments (*Paleodictyon*), stromatolites (such as *Conophyton*), horizontally aligned biogenetic gas holes or burrows (*Skolithos*), and fossil coral reefs (*Axixtes*). They recognize that such structures were most likely generated by communities composed of a great and unknowable diversity of organisms. Each member of the structure may have a distinct genome only remotely related to the others. So the concept of a *form-taxon* is used to describe body and trace fossils, including burrows, tracks, stromatolites, coral reefs, etc., each with a particular set of morphologically distinguishable characteristics. Each form-taxon, with its distinctive characteristics, labels a recognizable, repeatable morphological unit, often called a *morphotype*. The morphotype may even be an entire community. We argue that, notwithstanding the Linnaean claim that Latin binomials refer to individual members of a single species, in many and perhaps the overwhelming majority of cases (e.g., all eukaryotes), species names in contemporary taxonomy also refer to aggregates of individuals with diverse genomes, i.e. communities. For example, the cephalopod mollusk *Euprymna scolopes* forms a light organ with ciliated, microvillous appendages bearing pores that lead to empty spaces. The cilia sweep in bacterial symbionts that will develop into luminous colonies characteristic of this species of squid. When the light organ, which is embedded in the ink sac, has become replete with a dense, single-type

luminous bacterial population, the cilia are no longer needed and are absorbed. This sequence of events, repeated each generation, describes the cyclical symbiont integration in the ontogeny of normal *Euprymna* squid (McFall-Ngai & Ruby 1991). The origins of permanently integrated microbial symbionts, like the twenty or so species of *Caedibacter* known in the ciliate genus *Paramecium*, are more difficult to discern. The relation between speciation and genome acquisition in these and other genera is insufficiently investigated.

Combinatorics of symbiotic genomes

The combination of genomes in symbiosis has a power to generate diversity in form-taxa that may not be generally appreciated. With a single host that has no symbionts, only one genomic combination is possible. This is the case with nearly all the bacteria. Exceptions would be *Pelochromatium roseum* or other consortia bacteria (regular associations of a single flagellated heterotroph with clustered anoxygenic photoautotrophs) or encysted *Bdellovibrio* (bdelloplast) that contains, at some points in its development, the genomes of both *Bdellovibrio* and its *Chromatium* host (Tudor & Conti 1977; Tudor & Bende 1986). With a host and one symbiont, three genomic combinations are possible: the host alone, the symbiont alone, and the host and symbiont together. Probable examples are *Giardia* (a diplomonad), *Neocallimastix* (a chytrid), *Retortomonas* (a mastigote), *Vairimorpha* (a microsporidian), calonymphids, and other anaerobic mastigotes that lack mitochondria but display two- or three-componented reproducing karyomastigotes (Kirby 1952; classes Retortomonadida, Diplomonadida, Parabasalia, etc., in Margulis *et al.* 1990). With a host and two symbionts, six combinations are possible: the host alone, each symbiont alone, the host with symbiont 1, the host with symbiont 2, and the host with symbionts 1 and 2 together. In general, the number of genomic combinations that can be generated in this way by a host with n symbionts, assuming that each symbiont in addition can survive by itself, is $2^n + n$. With a host and ten symbionts, the number of potential taxa formed by recombination is 1,034. With a host and 20 symbionts, the number of possible genomic combinations is 1,048,596. With a host and 25 symbionts, the number of possible combinations is 33,554,457. This number approximates the minimal number of species on Earth estimated by some authors (e.g., May 1990; T. Erwin, oral presentation, 1994).

While genomic symbiosis has enormous power to generate diversity, that power may not always be used. For example, the platymonad marine worm *Convoluta* occurs without any photosynthetic symbionts as *Convoluta convoluta*. It is also found in regular and predictable combination with at least two kinds of photosynthetic symbionts, one at a time. With diatoms the yellowish worm is called *Convoluta paradoxa*, and when the symbionts are the green alga *Tetraselmis* (which is the same as *Prasinomonas*), all worms are not only bright green but they are functionally photosynthetic. The green form is called *Convoluta roscoffensis* (Smith & Douglas 1989). Some argue that *C. roscoffensis* should be removed from *Convoluta* to another genus, implying a still more profound effect of the cyclical symbiont integration that is characteristic of these marine worms. It is unlikely that those *Convoluta* occur with more than a single type of photosynthetic symbiont at the same time.

The examples of *Convoluta* and others (Table 1) show that the process of symbiogenesis is currently active at the level of individual species as labeled by conventional Latin binomials. Genomic symbiosis – i.e. acquisition and integration of microbial symbionts – may have played a powerful role in the origin of higher taxa, such as the 33 formally recognized phyla of animals (Margulis & Schwartz 1988). Conventional gradual accumulation of mutations, probably crucial for maintenance of symbionts and emergence of new holobiont properties, may then have differentiated these groups further at the species level (Margulis 1976; Margulis 1993).

TABLE 1 Taxa-specific symbioses: very few examples.

Host	Symbiont 1	Symbiont 2	Symbiosis name	New features, comments
Protists:				
<i>Devescovina</i> ¹	unidentif. fusi-form bacterium	peritrichous bacterium	"Rubberneckia"	gliding and swimming motility
<i>Mesodinium</i>	none	none	<i>Mesodinium album</i>	heterotrophic mesodinium
<i>Mesodinium</i>	partial cryptomonad	none	<i>Mesodinium rubrum</i>	photosynthetic, fast-swimming ciliate
<i>Metopus</i>	methanogen	none	<i>Metopus contortus</i>	life in anoxic environments
		?	<i>Metopus paleoformis</i>	life in anoxic environments
<i>Paramecium</i>	none	none	<i>Paramecium aurelia</i>	ciliate
<i>Paramecium</i>	<i>Caedibacter</i>	none	<i>Paramecium aurelia</i>	killer-strain ciliate
<i>Paramecium</i>	none	<i>Chlorella vulgaris</i>	<i>Paramecium bursaria</i>	photosynthetic ciliate
<i>Paramecium</i>	<i>Caedibacter</i>	<i>Chlorella</i>	no such organism	
<i>Plagiopyla</i>	methanogen	"hydrogenosome"	<i>Plagiopyla</i> sp.	life in anoxic environments
Animals:				
<i>Convoluta</i>	none	none	<i>Convoluta convoluta</i>	heterotrophic worm
<i>Convoluta</i>	<i>Tetraselmis</i>	none	<i>Convoluta roscoffensis</i>	photosynthetic worm
<i>Convoluta</i>	none	diatom	<i>Convoluta paradoxa</i>	photosynthetic worm
<i>Convoluta</i>	<i>Tetraselmis</i>	diatom	no such organism	
<i>Gazza leiognathid</i> fish	vibrio gram-negative bacterium	none	<i>Gazza minuta</i> (ponyfish)	gas-bladder light organ, luminous fish
<i>Hydra</i>	none	none	<i>Hydra</i> sp.	brown hydra
<i>Hydra</i>	<i>Chlorella</i>	none	<i>Hydra viridis</i>	photosynthetic hydra
<i>Hydra</i>	<i>Chlorella</i>	<i>Aeromonas</i>	<i>Hydra viridis</i>	photosynthetic hydra
<i>Monastraea</i> ²	<i>Symbiodinium</i> ?	none	<i>M. annularis</i> , morphotypes I, II & III	carbonate reef formation

¹ Tamm in Margulis 1993

² Knowlton *et al.* 1992

The role of an additional symbiont may depend on the number and physiological features of other symbionts, if any, already associated with a given host. For example, when one additional symbiont joins *Convoluta*, the species name changes. By contrast, a domestic cow may have a large number of stably associated rumen ciliates and cellulolytic bacteria and an even larger number of transient rumen ciliates and spore-forming bacteria. When one or another of the transient rumen microorganisms arrives or departs, even in huge numbers, it is more customary to change the description of the "health" of the cow than its species classification.

TABLE 2 Inconsistent names of taxa.

Higher taxa ^a	Name (minimal number of genomes per individual)	Partner (number of genomes per partner)	Basis for name ^b
L	<i>Heterorhabditis bacteriophora</i> (3)	<i>Xenorhabdus</i> 1 (1)	complex complex
L	<i>Heterorhabditis luminescens</i> (3)	<i>Xenorhabdus</i> 2 (1)	
SYM ¹	<i>Chlorochromatium aggregatum</i> (2)	<i>Chlorobium chlorochromatii</i> (1)	complex complex
SYM ¹	<i>Pelochromatium roseum</i> (2)	brownish chromatium (1)	
L	<i>Paramecium aurelia</i> (2 + 1 = 3)	<i>Caedibacter</i> (1)	L, ST
L	<i>Paramecium bursaria</i> (2 + 3 = 5)	<i>Chlorella</i> (1)	L, SP
SYM ²	<i>Cyanophora paradoxa</i> (2 + 1 = 3)	cyanobacterium ^c (1)	small small
SYM ²	<i>Cyanidium caldarium</i> (2 + 1 = 3)	cyanobacterium ^{c, d} (1)	
SYM ³	<i>Cladonia cristatella</i> (2 + 1 = 3)	<i>Nostoc</i> (1)	large large
SYM ³	<i>Cladonia cristatella</i> (2 + 3 = 5)	<i>Trebouxia</i> (3) ^e	
L	<i>Glycine max</i> (3)	<i>Rhizobium</i> (1)	large
L	<i>Glycine max</i> (3 + 1 = 4)		
SYM ⁴	<i>Microcycas</i> (3)	<i>Nostoc</i> (1)	large
L	<i>Convoluta convoluta</i> (2)	<i>Bacillaria</i> (3), diatom ^e <i>Tetraselmis</i> (3), green alga ^e	large, SP large, SP
L	<i>Convoluta paradoxa</i> (2 + 3 = 5)		
L	<i>Convoluta roscoffensis</i> (2 + 3 = 5)		
L	<i>Homo sapiens</i> (2)	unknown	large large
L	<i>Homo sapiens</i> (syphilitic) (2 + 1 = 3)	<i>Treponema</i> (1) plus unknown	

^a L=large; SYM=smaller symbiont. Higher taxa (families, orders, classes, phyla) based on *large* partner when the presence of the small one is irrelevant to taxonomy and on symbiotic complex (SYM) when the entire higher taxon is defined by traits characteristic of the complex and not of its components. SYM¹ = consortia bacteria; SYM² = glaucocystophytes; SYM³ = lichens; SYM⁴ = cycads.

^b Name of genus based on *large* partner when name is independent of presence or absence of smaller partner; on *small* partner when presence of endosymbiont determines genus name; on *complex* when genus is defined by traits of the partnership. SP = specific name determined by presence of endosymbiont; ST = strain name determined by presence of symbiont.

^c Sometimes called a cyanelle.

^d Sometimes called a chloroplast or rhodoplast.

^e Plastid, mitochondrion, nucleocytoplasm.

In the prototists – a huge taxon (Kingdom Protista or Protoctista) estimated to encompass 250,000 species – the relative sizes of the symbionts (bionts) that form the “individuals” (holobionts) are far more equal than those of plants, animals, and fungi. Therefore, both the clearly symbiogenetic provenance and the nomenclatorial confusion are far more evident in these eukaryotic microorganisms than in other large taxa (Corliss 1992). Given new results of molecular biology, the taxonomy and practical systematics of the group of former animals (province of zoology), former plants (province of botany), and former fungi (province of mycology) have reached nearly crisis levels (Margulis 1992a). Recognition of the “legitimacy of having distinct high-level ranks for protist species that seem to be widely separated phylogenetically from fellow protists or from eukaryotic assemblages” is fervently pleaded by Corliss (1992).

Not all possible symbiotic combinations are likely to be realized in practice. For example, the presence of one type of photosynthesizer, one hydrogen-sulfide generator, or one dinitrogen fixer probably precludes any selection pressure for a second of the same type. Furthermore, not all of the combinations realized in nature may be distinguishable (Table 2). An open empirical question is to determine the actual relation between the number of possible symbionts associated with a host and the number of “species” conventionally distinguished for the corresponding group of genomic combinations.

The relative poverty of species in the Archean fossil record and their prokaryotic level of organization are well established (Schopf 1983b). This suggests that the major integration of microbial symbionts to form individuals of higher levels of complexity did not occur until the beginning of the Proterozoic Eon associated with the appearance of *Grypania* (Han & Runnegar 1992; Runnegar, this volume) and the later Ediacaran prototists and animals (McMenamin 1993). The remarkably sudden appearance of large marine animals at the end of the Proterozoic and through the lower Phanerozoic may be related to symbiont acquisition, especially of calcium-precipitating bacteria by soft-bodied animals (Lowenstam & Weiner 1989). This well-known discontinuity in the fossil record may correlate with symbiotic consortia having 7–9 different components and having the capacity to generate hundreds of distinct morphotypes (species). The techniques of molecular biology permit analysis of complex genomes of eukaryotes and recognition of their elemental composition by identification of the original metabolism, morphology, and genomes of microbes that comprise them.

Conclusions

What are the implications of this analysis? First, biologists should recognize explicitly that most of their so-called individuals, including all eukaryotes, are in fact genomic combinations;¹ they should consider the possibility of adopting a consistent large-host nomenclature that appropriately recognizes the integrated genomes. Taxonomic nomenclature should be more consistent across fields; Table 3 illustrates the problem.

Second, the role of symbiogenesis as a driving factor in the diversification of life should be investigated empirically in many more groups than it has been so far. A start

¹ E.g., four genomes of algal cells: Nucleocytoplasm, undulipodia, mitochondria, and plastids.

TABLE 3 Phototrophic marine protoctists^a: Identical organisms^b (individuals^c) described by different higher-taxa names.

Taxa assigned	People who use this terminology
phytoplankton, photoplankton, nanoplankton	oceanographers, limnologists
algae, microphytes, phytomonads	phycologists, ecologists, zoologists
aquatic plants	ecologists
green scum	public-at-large
phototrophic protists, photosynthetic eukaryotes	bacteriologists
plants, lower plants, algae, photosynthetic protoctists, chrysophytes, prymnesiophytes, haptomonads, thallophytes	botanists
coccolithophorids, prymnesiophytes, lower plants	paleontologists, geologists
eukaryotic microbes, algae, protists	cell biologists

^a For detailed classification of these organisms see the *Handbook of Protoctista* (Margulis *et al.* 1990).

^b Examples: *Chrysochromulina* (dasmotrophic coccolithophorid), *Dunaliella* (motile green alga), *Emiliana* (coccolithophorid), *Mychonastes* (nonmotile encysting green alga of the chlorella type).

^c If we were to recognize microbiological standards and require growth in pure culture of all the organisms involved, we would not be allowed to name many protoctists, animals, plants, or fungi.

in this direction has been made by McFall-Ngai & Ruby (1991) in their analysis of luminescent squid, by Nealson (1991) in his analysis of "glowworms" (lepidopteran larvae inhabited by nematodes and luminous bacteria), by Schwemmler (1991) in his studies of homopterans such as *Eucelis* with its integrated bacterial symbionts, by Nardon & Grenier (1990) in weevil-bacterial associations, and by Vetter (1991) in his analysis of thiotrophic animals. We predict that between 20 and 22 physiologically distinctive microorganisms (primarily bacteria and fungi) are regularly associated with coleopterans. Genomic combinatorics may explain why, as J.B.S. Haldane observed, God has expressed such an inordinate fondness for His most flamboyant morphotypes: His millions of species of beetles.

Acknowledgments. – L.M. is grateful to the NASA Life Sciences Office (NGR-025-004) and to the Dean of the College of Arts and Sciences, University of Massachusetts, Amherst. J.E.C. acknowledges the support of U.S. National Science Foundation grants BSR87-05047 and BSR92-07293 and the hospitality of Mr. and Mrs. William T. Golden.

This chapter is dedicated to the memory of Heinz A. Lowenstam, pioneer integrator of biological and geological knowledge, a founder of the field of biomineralization.

Early Life on Earth

Nobel Symposium No. 84

Stefan Bengtson, editor



Columbia University Press New York

Columbia University Press
New York Chichester, West Sussex

Copyright © 1994 Columbia University Press
All rights reserved

Library of Congress Cataloging-in-Publication Data

Nobel Symposium (84th : 1992 : Karlskoga, Sweden)

Early Life on Earth / Nobel Symposium No. 84 ; Stefan Bengtson, editor.

p. cm.

Includes bibliographical references and index.

ISBN 0-231-08088-3

1. Evolutionary paleobiology — Congresses. I. Bengtson, Stefan. II. Title.

QE721.2.E85N63 1994

560—dc20

94-3822

CIP



Casebound editions of Columbia University Press books are printed on permanent and durable acid-free paper

Printed in the United States of America

c 10 9 8 7 6 5 4 3 2 1

p 10 9 8 7 6 5 4 3 2 1