AIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100

Silver Spring, Maryland 20910

301/587-8202

Centimeter

Inches

MANUFACTURED TO AIM STANDARDS
BY APPLIED IMAGE, INC.

1 of 1

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

JUN 05 1992

GENOMIC DEFINITION OF SPECIES

by

Radomir Crkvenjakov¹ & Radoje Drmanac¹

Biological and Medical Research Division

For submission to
Journal of Theoretical Biology

June 1992

RECEIVED
JUN 10 1992
OSTI

The submitted manuscript has been authored by a contractor of the U. S. Government under contract No. W-31-109-ENG-38. Accordingly, the U. S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U. S. Government purposes.

¹Permanent address of authors:

Institute for Molecular Genetics and Genetic Engineering (R.C. and R.D.) and Dept. of Biology, Faculty of Sciences, Univ. of Belgrade (R.C.), P.O. Box 794, 11000 Belgrade, Yugoslavia

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

GENOMIC DEFINITION OF SPECIES

by

Radomir Crkvenjakov and Radoje Drmanac

INTRODUCTION

A genome is the sum total of the DNA sequences in the cells of an individual organism. The common usage that species possess genomes comes naturally to biochemists, who have shown that all protein and nucleic acid molecules are at the same time species- and individual-specific, with minor individual variations being superimposed on a consensus sequence that is constant for a species. By extension, this property is attributed to the common features of DNA in the chromosomes of members of a given species and is called (species) genome.

The definition of species based on chromosomes, genes, or genome common to its member organisms has been implied or mentioned in passing numerous times. Some population biologists think that members of species have similar "homeostatic genotypes," which are to a degree resistant to mutation or environmental change in the production of a basic phenotype (Mayr 1970; Ehrlich and Raven 1969; Wiley 1981). Molecular biology textbook writers seem to imply the existence of species genomes (Alberts et al. 1989, Singer and Berg 1991). More explicit statements can be found in the recent literature (Depew and Weber 1988; Collier 1988; Wesson 1991). We attempt here to outline the theoretical consequences of the view that genomes are simultaneously organism- and species-specific.

Another important concept is the genome program as a set of instructions for complex functions and development. From the operon concept in 1961, there has been a steady accumulation of evidence for the pattern of time and space control of gene expression, which is not only determined by the various sequence elements but also by their actual position and order in DNA. The alternative programs for the lytic and integrative life cycles of lambda phage, now completely understood on sequence level (Ptashne 1986), provide a paradigm for a single genome encoding a variety of alternative programs needed for the development of various forms in polytypic species (such as queens, drones, and workers in bees). The conservation of homeobox gene batteries in metazoan animals (Murtha et al. 1991) suggests the importance of DNA structures larger than genes in animal development. Now, it is even possible to propose simple models for genetic programs leading from zygote to adult arthropod (Bodnar et al. 1989): Canalization (Waddington 1957) and developmental constraints (Albrecht 1980) are concepts which describe the elasticity of development to perturbation and the restricted range of allowed outcomes of acceptable change. These systemic properties of developing organisms can be explained either on a nongenetic, cellular interaction level or can be directly or indirectly attributed to the species genome program encoded in the DNA sequences. In the definition proposed here, we have tried to clarify the relation between the concepts of genome, DNA sequence, and genome program on a general level.

The philosophy of science has recently contributed to the ongoing debate about the correct definition of species. A definition justifying the expectations of biologists and

philosophers alike would have to satisfy three criteria (slightly modified from Kitcher [1987]):

UNIVERSALITY I. All organisms (evaluated over a full life cycle) must belong to exactly one species.

UNIVERSALITY II (Uniformity). A species category is characterized by principle(s) valid for all species taxa (individual species).

ULTIMACY. The principle(s) by which one explains why a group of organisms constitute a species should also be fundamental to the explanation of biological diversity.

The intellectual underpinnings of our definition can be traced to Goldschmidt (1940) and many others since. The recent contributions to evolutionary theory of Brooks and Wiley (1988), Wicken (1987), Collier (1988), and the program of the Artificial Life Community (see, for instance, Langdon 1989), treat some of the questions raised here.

Our variant of genomic definition is intentionally cast in the most general and informatic terms for the purpose of providing a definition able:

- i.) To encompass sexual and asexual species,
- ii.) To encompass both cellular organisms and viruses,
- iii.) To encompass all nucleic acid-based self-replicating organisms, including protobionts based on RNA (RNA world), and

iv.) To provide a rational link between biology and computer science (artificial life community, computational physics).

Probably most biologists will find at least one (but not the rest) of these goals as worthy of theoretical pursuit if treated separately. It is interesting that a single species concept, if correct, might achieve all four simultaneously.

Any new species definition is useful to the extent that it broadens the area of discussion of this open scientific problem. We believe the majority of other definitions to be correct as well, but less general, since they do not cover some of the goals of our definition. Due to the lack of the complete overlap, their comparison with the definition proposed here will not be attempted. In one of the traditions of biology, our definition is reductionistic. The new terms reflect the basic assumptions of molecular genetics. The proposed definition is based on the view that belonging to a species is an intrinsic property of organisms and that this property is genetically defined.

DEFINITION AND ASSOCIATED PRINCIPLES

Our proposal for the definition of a biological species is as follows:

A species comprises a group of actual and potential biological organisms built according to a historically unique genome program that is recorded, and at least in part expressed, in the structure of their genomic nucleic

acid molecule(s), having intragroup sequence differences which can be fully interconverted in the process of organismal reproduction.

The major premises on which the definition is founded are:

- The unity of living world is based on the tree of reproduction connecting all organisms. Starting with the first organism, reproduction cycles consist in part of imperfect replication of nucleic acids, i.e., an organism is any minimal physical system whose reproduction is based on nucleic acid replication.
- The universal pattern of evolution is the multigenerational existence of groups of phenotypically similar organisms, i.e., species (Fig 1), in which each group has a unique founder group or, more rarely, a unique combination of founder groups (hybridization, symbiosis). This premise implies the impossibility of reversion; the newly formed group cannot produce the members of the founder group as progeny, nor members of any other unrelated group, i.e., there is a discontinuity, an unbridgeable gap, between organisms of different species, despite a common ancestor. It is not required that founder species be extinguished after giving birth to a new species. Also, the premise

allows for a limited possibility that organisms of a given founder group or a specific combination of founder groups can start a given new group in several independent events during evolution.

- Every individual organism belongs to one and only one species (Kitcher 1987), as the first member of the group, as any of the other members, or as a barren mutant (Fig 2).
- Identical sequences of entire genomes in two organisms can not code different instructions within the natural phylogeny based on nucleic acids. However, various genomic sequences can code the identical instructions.
- Starting from the first organism(s) onward in phylogeny, there is a stage in the reproduction process at which parts of the phenotype(s) [i.e., expression of genome(s)] of the parent(s) are used as initial conditions for the existence of the first-generation offspring. Only some of the genomes of the offspring are compatible with "maternal" initial conditions, i.e., organisms encoded by these genomes can reach reproductive maturity and leave their offspring in a given environment. In some of first-generation offspring, in much rarer instances, a genome acquires differences and

produces changed initial conditions for the second-generation offspring organism, which will not be capable of supporting its genome through reproduction if it were a revertant back to the genome(s) of its grandparent(s). In even rarer cases, a filial genome is not compatible with its own initial conditions, i.e., offspring could not have a genome identical to their parent(s).

- Simultaneously living organisms of one species can exist physically isolated from each other, and some generations can be represented by a single living organism (Fig 2).

The basic assertion of the proposed definition is that the genome is the organizational level which accounts for the species characteristics given as premises. Any phenotype as derived from a genome has a lesser chance to give a correct picture than the genome itself. In many cases, organisms of a single or closely related species are more different than organisms of evolutionarily more distant species when only separate characteristics are compared. Because genomes require wider molecular systems of which they are part, one might claim that organisms themselves are the only level on which species can be described. However, if one holds that a component of an organism called the *genome* determines species characteristics, this does not mean the other components of the system are not necessary, that the wholeness of organism is denied, or that interactions of organisms do not play a significant

role. It only means that the genome is the most proximal cause for such behavior of the system.

In our definition, the phrase "genomic nucleic acid molecule(s), having intragroup sequence differences which can be fully interconverted" is the basis for recognizing one species from another. Reproductive isolation is a specific phenotypical consequence of this principle and is valid for most cases concerning sexual organisms. Any vertical (mutation, deletion, duplication, translocation, transposition) or horizontal (sexual recombination, transfection, symbiotic merge) changes of genomic sequences in reproduction contribute to the generation of sequence differences. Some of these will lead to phenotypic organismal diversity within species or, very rarely, to the creation of a new species. However, the individual characteristics of genomic DNA within species can be repetitively produced starting from different parents, from clones in asexual species, etc.

Within a species, the assumption that second-generation offspring can be revertants to the grandparental genomic sequence means that any organism within a species has a capacity to start a line, eventually giving an organism possessing a genomic sequence identical to that of any other member of the species. At least one pathway of interconversion is equivalent to reversion to the genome of a common ancestor and then proceeding with changes of this genome equivalent to those in the line leading from the ancestor to the reference organism (Fig 3). One must keep in mind that a potential for reversion to the grandparental genome is realized very rarely (especially in the case of complex genomes). Other pathways are possible, including organisms with new genomic

sequence variants. As the broader term *genome convertibility* includes reversibility, since it does not require the same steps in both directions.

For the simplest organisms, protobionts and viruses, the totality of possible genomic sequences which are interconvertible could be defined and thus the consequences of the genomic species definition shown in practice. The potential convertibility of total genomic sequences, which is very rarely realized in species with longer genomes, can be of a high practical value if correctly estimated from the limited knowledge of genomic sequences and their meanings (instructions). The rates of actual convertibility of defined parts of genomic sequence, and the extent of allowable differences and probability of their conversion could be measured and estimated, respectively. The elaborated principle of genome sequence convertibility outlined above can serve as a nearly universal basis for species recognition (Drmanac and Crkvenjakov, in preparation). It can be a useful parameter for systematics as the natural and general level for identification and definition of species.

The organisms possessing genome sequences that are potentially interconvertible (even if reproductively isolated) or that are nonconvertible with genome sequences of organisms of other species (even when they are phenotypically very similar) would constitute a species. Thus, a reversible mutation responsible for reproductive isolation does not constitute the splitting of one species into two.

There are two aspects of nonconvertibility: probabilistic impossibility and systemic impossibility. For any complex material

system, one can define physically very improbable changes, or combinations of changes, and even more importantly, changes which are incompatible with the system's existence as a whole. The evolution of a genome is inseparable from the evolution of a wider organized molecular system (cell, organism) of which it is a component. This fact prohibits the intentional interspecies conversion of genomic sequence only or the transplantation of a genome in a cell of another species, causing in both cases nonviability of the organisms thus obtained. The multitude of simultaneous changes needed to overcome this barrier is an example of probabilistic impossibility, and lethal mutations are an example of systemic impossibility.

For purposes of the analysis, it is necessary to consider the genome as consisting of a genomic sequence and genome program. This is seen from problems arising by considering genomic sequences only phenotypically (morphologically), since individual genomic sequences are nonidentical but similar, just as organisms are. The premise that similar but nonidentical sequences can encode the same instructions allows the possibility that a group of organisms can share common instructions. Thus, the question is transformed into defining a common denominator of evolutionary links and the discreteness of species on the level of genome instructions.

Considering the inevitability of individual differences, what is the theoretical framework for discreteness and uniqueness of species? This framework is provided by the genome program as an integral set of instructions encoded in the sequence. It encompasses all meanings of the term "genetic program." The analysis of its usage in molecular genetics provides a more general

picture of a genome program as a formal, logical entity. For brevity, only some conclusions based on assumed pattern of evolution and species characteristics will be listed. One generalization is that a given, historically unique genome program produces each and every organism of one species, with its particular genomic sequences and phenotypic traits. The program is understood as a specific pathway of self-reproduction of a particular material system, i.e., a specimen biological organism of a species. A species genome program is the product of ancestral programs and environmental constraints. It is a framework limiting the range of individual differences, adaptational possibilities, and capacities for starting a new species. The program provides those constraints and also provides for the stability and long life of the species. The program and species are historically unique individuals (*sensu* Ghiselin 1987). In spite of random changes in nucleic acids as the material record of the program, the uniqueness of the program survives as such. The death of the last organism carrying a given program can happen. The species with its program might be extinguished before or after giving birth to a species that has a new program. The program includes constraints which limit changes of both the record and the whole system. The genome program consists at least of the developmental subprogram (ontogenic program), reproduction subprogram, and the inscription rule, which is a specific way of making the program record for development and reproduction. By the inscription rule we mean the class of phenomena of which an example is that one protein with the same time-space expression pattern can be coded in two species by

genes with different numbers of introns and exons or even by genes without introns.

In our definition, species are considered as inevitable products of the process of self-reproduction of biological organisms. The formation of discrete and historical entities by self-reproducing systems in general and biological organisms in particular is the consequence of several irreducible properties of these systems. These properties are realized in the first self-reproducing system(s) (biological organism[s]) which was formed from nonreproducing material systems. Other, derived properties can be added on in organisms formed in an evolutionary chain, but the basic ones cannot be lost. The basic properties of self-reproducing systems are as follows:

- i) The wholeness of the organism, which makes it distinguishable from the environment and cannot be reduced to any one part of the system; and
- ii) The existence of a system part that has both the property of a structural element and the property of material record of a formal entity: a program for self-reproduction. The system retains the capacity for self-reproduction within the range of changes of this part, and its copying of this part by reproduction is generally imperfect.

The whole of the system had to emerge once at the very beginning of the process, since its basic properties cannot be attributed to the system's components *per se* (i.e., when they are outside the system). The first system's component, which can not be regenerated outside the system after emergence (only within it

with the help of other components), is the material record of program. The system had to contain various accessory components in addition to the record component. These accessories are helpers which need not necessarily have been formed within a system (but could have been) and which are internalized within a system in the course of evolution.

The wholeness of the first biological system can be described in light of the dual properties of the nucleic acid molecule. This molecule provides the basis for replicable program inscription through chemical complementarity. Simultaneously, as a material component, it constitutes part of the system and has a structural role, which in interaction with internalized components from the environment (accessory components) forms the whole of the reproducing organism. The structural role of nucleic acids has been retained in present-day organisms, as has the duality of structure and inscription.

Nucleic acids also satisfy the necessary property of self-reproducing systems of generally not requiring perfect copying of its program record. Replication based on complementary pairing is not entirely sequence-specific in the sense that the fidelity of any base pair replication is mostly independent of neighboring sequences, but is dependent on some structure of either the replicating enzyme or the nucleic acid itself. Not only the sequence-derived structure can alter fidelity of replication and even prevent it, but also several different sequences can give the same derived structural element, as evident in modern organisms. Complementary pairing opens the additional possibility of large and immediate sequence changes via recombination.

Generally, imperfect reproduction leads to organisms that differ from each other, but their genomic inscriptions are interconvertible, i.e., can be repeatedly generated from previous organisms. One can ask what is required, after sufficient time has elapsed, for the inevitable occurrence of organisms so changed that they, although being able to reproduce, are unable to revert to the range of parental types no matter how much time is allowed for it. Or, more directly put, what is the cause of existence of the species category. There are two aspects of this question. One is, why there is discreteness, i.e., gaps between species? The other is, why is the birth of an organism of a new species from an old species a unique and irreversible event? One can expect that certain levels of complexity of both system and environment are necessary to provide the sufficient conditions for organismal and hence species occurrences. This granted, the inevitability of species occurrences as entities separated by gaps is a consequence of the discrete nature of physical laws whose resultant derivative is a genome program.

The irreversibility and uniqueness are the result of a probability so low as to represent a virtual impossibility of reproduction in certain connecting stages through which systems would have to pass to both convert back and remain self-reproducing. An alternative concept sometimes mentioned is that speciation is a historical event and hence is irreversible.

DISCUSSION

The common denominator of all organisms is the existence and inseparability of a program, its inscription in a nucleic acid sequence and a derived molecular system, and the common descent of these inseparable entities via reproduction. We believe that a more general statement about biology is impossible at this time. Other assertions that have been advanced exclude one or another form of extinct or extant life. The species is equally as basic to biology as is the organism. Therefore, it is reasonable to see species explained in terms of program, inscription, and system.

The program is the primitive term (i.e., the meaning is assumed) of our definition, but we view it as a closed and finite information system or a heuristic algorithm. A system's program is a formal, logical entity and can be recorded by any material means to exist as a copy outside of the system simultaneously, before, or after it (Langdon 1989). Biological organisms are the only natural systems built according to a program. It is not known if other kinds of program-based systems exhibiting life-like properties are physically impossible. At least the possibility of artificial life based on such systems should be taken into account (Langdon 1989).

The fact that the program has evolved in some complex species to reside in more than one physically separable material system, for example, a breeding pair in sexual species, is not actually a major divide in evolution separating apparent and real species. As evident in many independent evolutionary lines, sexual species coexist with asexual ones. This indicates that, at least in some

cases, relatively less profound genomic change than implied by the divide hypothesis is required to transform one into another.

The view of species through nonconvertibility of genomes is broader than the conventional one, since it includes transient species, if they existed, connecting a long-living ancestor/descendant pair of species. The transient species can have a small number of generations. For any given pair of closely related but morphologically well-separated species, the nature of genomic changes required to transform one into another determines if the transient species is required or not. Our definition is neutral to this central question of speciation mechanisms except to predict that the question is not resolvable on the phenotypic level alone.

Our assertion is that the species category is a class of individuals, each individual bearing a historically unique genome program. Two different inscriptions can generally represent either a single or two different genome programs. The inscriptions representing one program (i.e., within species) are more similar among themselves than they are to any given inscription of another program. Due to the inseparability of the program and its inscription, there must be a rule that transforms one into the other. However, there is no reason *a priori* to expect the rule to be the same for all species taxa; rather, the rule is a historically frozen event shared in part by smaller or larger subgroups of species (higher taxons). Therefore, the species definition based on DNA sequences or on any other derived phenotypic property can never satisfy the criterion of absolute operationality, which requires a transpecific law that errorlessly places an individual organism into its species. However, the

genomic DNA sequence as an organizational level of the system most directly connected with the genome program must represent the basis of a definition much closer to universal operability than any based on the other phenotypic levels of organization, which are connected to a genome program much more indirectly and without exception via a DNA sequence.

Due to its generality, the present definition is vague on specifics of species determination and cannot be falsified on the basis of assertion that it specifically predicts the genome organization of any one species. However, it can be falsified by providing a specific biological example which lies outside its realm of validity. The new science of genomics is in its infancy. Once sufficient data about genomes are collected and analyzed in an empirical framework, the systematics theory most consistent with the reality will be discerned. However, the example of viruses already illustrates the natural displacement of systematics based on more derived phenotypic levels by systematics based on sequences. Furthermore, the potential of sequence-based systematics to achieve the higher degree of operability with the availability of the genomic sequence data is apparent as well to practitioners at the cutting edge of systematics of more complex organisms.

The turning of molecular biology towards the genome can be seen from the interest in various genome projects. The genome is the next level up from the gene in complexity and is becoming accessible through comparative analysis of genomic sequences. It is a relatively contained and information-rich level that gives the hope of extracting general rules of behavior from a complete material description, which is still unthinkable for other even

more complex levels of biological organization. Due to the experimental developments, theory has to find a place for the genome that is different from that of a mere sum of genes. In our opinion, both Rosenberg (1985) and Wicken (1987) correctly point out that the most profitable approach for formulating adequate evolutionary theory is to consider the period of emergence of life in the primordial broth. From our point of view, this is also true for the definition of species: it must be able to encompass both the acellular and cellular life.

The empirical fact that RNA molecules can catalyze RNA phosphodiester bond formation led to the proposal of a protobiotic "RNA world" (Sharp 1985), now widely accepted among molecular biologists. Thus there is a reason now to consider only the option of RNA-based life, the first living system which had a genome consisting of RNA and had no energy-transforming machinery. At the time of emergence, the first biological organism had not yet evolved the complex paraphernalia necessary for today's survival. It was directly subject to natural selection. Curiously, the simultaneous theoretical treatment of genome and the emergence of first organisms argued for here seem not to have been attempted previously.

Regarding our four goals, we have provided criteria of convertibility of genomic inscriptions which bonds together the clones of asexual species as firmly as mating bonds the organisms of sexual species. By uniting the acellular and cellular species by the requirement that they both reproduce on the basis of a nucleic-acid-inscribed genome, the viruses, despite their lack of metabolic machinery, are shown to be living. There is no real

difference in the reproduction of protobiotic RNA organism in primordial soup and of a lambda phage infecting bacteria. Both genomes utilize the material and energy resources available in their respective environments. If one was alive by virtue of being ancestral to cellular life, the other must be too. The placement of genome program in the center of biological species phenomena provides a theoretical basis for the already perceived need to expand the interactions between biology and computer science. The genome program as an algorithm lies in the realm of information science (already called computational genetics), and, by extension, in part in the realms of theoretical physics and mathematics. It is hoped that this bridge, regardless of the present work, will be shown as real and productive by the future empirical work.

Summarizing, this is the first biological species definition which satisfies the logical requirements of philosophers by conforming to the three principles of Kitcher (1987), as follows. Each organism belongs to exactly one species, since it is a system with one and only one genome program. All species taxa belong to the species category since each has a unique genome program. Finally, the ultimacy principle is satisfied by the assertion that the genome level is the fundamental one for explanation of biological diversity. This is based on the analysis of the case of the first organism/species proposed to have consisted of a replicating RNA molecule and accessory components. The laws valid for it are the only common laws for all living beings descended from it.

ACKNOWLEDGMENTS

We thank Ivan Labat for helpful comments on the manuscript, David E. Nadziejka for editorial assistance, and Nikola Tucic for use of his library on evolutionary biology. An earlier version of this manuscript was completed at the Institute of Molecular Genetics and Genetical Engineering, Belgrade, Yugoslavia. The work was supported in part by the Science Fund of Serbia, Yugoslavia; U.S. Department of Energy grant DE-FG02-88ER60699; U.S.-Yugoslav Joint Board grant JF-820; and by the U.S. Department of Energy, Office of Health and Environmental Research, under contract No. W-31-109-Eng-38.

REFERENCES

- Albercht, P., 1980, Ontogenesis and Morphological Diversification, *Am. Zool.* 20: 653-667
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D., 1989, *Molecular Biology of the Cell*, 2nd ed., Garland, New York
- Bodnar, J.W., Jones, G.S., and Elis, C.H., Jr., 1989, The Domain Model for Eukaryotic DNA Organization: A Molecular Basis for Constraints on Development and Evolution, *J. Theor. Biol.*, 137, 281-320
- Brooks, D.R., and Wiley, E.O., 1988, *Evolution as Entropy: Toward the Unified Theory of Biology*, 2nd ed., Univ. of Chicago Press, Chicago
- Colier, J., 1988, The Dynamics of Biological Order in *Entropy, Information and Evolution: New Perspective on Physical and Biological Evolution*, B.H. Weber, D.J. Depew, and J.D. Smith, eds., MIT Press, Cambridge, 227-242
- Depew, D.J., and Weber, B.H., 1988, Consequences of Nonequilibrium Thermodynamics for the Darwinian Tradition, in *Entropy, Information and Evolution*, Weber, B.H., Depew, D.J., and Smith, J.D., eds. MIT Press, Cambridge, Mass. pp 317-354
- Ehrlich, P., and Raven, P., 1969, Differentiation of Populations, *Science*, 165: 1228-1232
- Ghiselin, M.T., 1987, Species concepts Individuality and Objectivity, *Biology & Philosophy*, 2: 127-144
- Goldschmidt, R., 1940, *The Material Basis of Evolution*, Yale University Press, New Haven, Conn.

- Kitcher, P., 1987, Ghostly whispers: Mayr, Ghiselin and the "Philosophers" on the Ontological Status of Species, *Biology & Philosophy* 2: 184-191
- Langdon, C.A., 1989, Artificial Life, in *Artificial Life*, Langdon, C. A., ed. Addison-Wesley, Reading, Mass., pp. 1-48
- Martha, M.T., Leckman, J.F., and Ruddie, F.H., 1991, Detection of Homeobox genes in Development and Evolution, *Proc. Natl. Acad. Sci. USA*, 88: 11711-10715
- Mayr, E., 1970, *Populations, Species and Evolution*, Harvard University Press, Cambridge, Mass.
- Ptashne, M., 1986, *A Genetic Switch*, Blackwell Scientific Publications, Palo Alto, Calif.
- Rosenberg, A., 1985, *The Structure of Biological Science*, Cambridge University Press, Cambridge, England
- Sharp, P., 1985, On the Origin of RNA Splicing and Introns, *Cell* 42, 397-400
- Singer, M., and Berg, P., 1991, *Genes & Genomes*, University Science Books, Mill Valley, Calif.
- Waddington, C.H., 1957, *The Strategy of Genes*, Allen and Unwin Ltd., London
- Wicken, J.S., 1987, *Evolution, Thermodynamics, and Information: Extending the Darwinian Program*, Oxford University Press, Oxford
- Wiley, E., 1981, *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*, Wiley & Sons, New York
- Wesson, R., 1991, *Beyond Natural Selection*, MIT Press, Cambridge, Mass.

FIGURE LEGENDS

Fig 1. Irreversibility of speciation in time. Each time horizon contains a certain number of phylogenetically related species (each represented by a different symbol). The irreversible founder/descendent relation is indicated by the same founder(s) in occurrence(s) in which a species appears as a progeny. Also present are different time-cohorts of the same species (identical symbols in two adjacent time-horizons) shown as either giving rise to different new species, or to the same one. Dashed arrows and symbols denote relations excluded by the irreversibility premise (see text).

Fig 2. Single species membership of organisms and origin of new species. Organisms belong either to species A or a new species B. In this example progeny of members of sexually compatible species A and B is always B, as required by the irreversibility premise. Any two A organisms are capable of giving A progeny. Only a limited set of A $\times$ A pairs gives rise to B progeny. B $\times$ B crosses give B only, as well as A $\times$ B crosses when viable.

Fig 3. Genome interconversions within species phylogeny. Member organisms (O) have different subscript numbers to denote their nonidentity as physical objects. However, their genomes (G) can be identical in sequence to ancestral genomes. O₁, O₆ and O₁₃ contain G₁, O₂, O₉ and O₁₀ contain G₂, O₄ and O₁₄ contain G₄, and O₁₁ and O₁₅ contain G₁₁, respectively.

Time

DATE

FILMED

7/13/94

END

1