Journal of Statistical Physics, Vol. 110, Nos. 3–6, March 2003 (© 2003)

Renormalization and Taxonomy

Y. Oono[1]

Received January 29, 2002; accepted September 25, 2002

Taxonomists now seem to embrace the conclusion that natural classification not based on phylogeny is impossible in principle. Classification in physics suggests, however, that a natural classification of objects is possible when there are two disparate levels of their description. This is the essence of classification based on the renormalization-group (RG) philosophy, and the resultant classification scheme is regarded sufficiently objective. Thus, we can argue that it is still premature to conclude that biological classification is objective only when it is based on molecular phylogeny. The analogy to RG suggests the developmental taxonomy dependent *solely* on ontogeny.

KEY WORDS: Renormalization group; taxonomy; developmental process.

1. INTRODUCTION

Taxonomy has often been (especially by physicists) compared with stamp collecting, and regarded as a secondary subject of science. However, it is not only practically important but is one of the fundamental fields of biology.

There is an urgent need of recording biodiversity that is rapidly being lost. To describe and understand the vast diversity of the biological world, the attitude of the ordinary taxonomists (the so-called α-taxonomists) is: we have only to catalogue all the species in the world. Whatever the concept of "species" may be, some people estimate that there are no less than 100 million species of plants and animals (further worse, each species seems to host at least one special parasitic bacterium species). Therefore, cataloging all the "species" is simply impossible; naturally, we need description and understanding at coarser levels than species, at the so-called higher order taxonomic levels.

---

[1] Department of Physics, 1110 W. Green Street, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801-3080; e-mail: y-oono@uiuc.edu

Many people tend to believe biological systems are classifiable into a hierarchical scheme, and this hierarchical classifiability may be one of the major characteristics of the so-called complex systems.(1) However, there seems no agreement on its objective basis. Higher order taxonomic units are often claimed to be arbitrary.(2) Hey(3) asserts that when we devise taxa, we are not objective, and that we must keep in mind that different human observers will find different taxa. Therefore, the current orthodox opinion asserts that molecule-based phylogenetic taxonomy is *the* natural taxonomy of organisms. Logically speaking, however, genealogy and classification are distinct. Then, we should first reflect on what classification, especially, objective classification is.

Classification of objects and phenomena appears in physics and mathematics as well. For example, we have universality classes for various phase transition phenomena. Physicists never believe these classifications unnatural and man-made. Therefore, although biologists seem to give up the possibility of natural classification that does not rely on genealogy, at least logically, there seems to be a possibility of natural (and objective) classification without referring to phylogeny.(4)

2. CLASSIFICATION AND DEVELOPMENTAL TAXONOMY

Classifying objects requires identification of (not quite the same) objects, so classification and ignoring certain features = coarse-graining are inseparable. If there is no *a priori* (un)importance criterion that allows us to ignore features, there cannot be any natural or objective classification.

Consider two levels of describing an object, the level where we observe it at our scale and the microscopic level supposedly governing (or generating) the object. First, we look at the object with sufficient resolving power to observe (microscopic) details. Then, we reduce the resolving power of our observation (or increase the cutoff length scale). Along this procedure, the quantities that fade out of sight are irrelevant, and those that persist or grow more important are relevant. This is the standard RG procedure.(5) A collection of phenomena (or objects) sharing the same relevant quantities (features) may be regarded as a group (universality class). The classification according to this grouping is a logical consequence of the two levels we have chosen. If these two levels are "natural" (e.g., without any more reasonable choice), the classification is natural. That is, if there are two disparate levels of description of objects dictated by the objects, and if the map from the detailed description level to the less detailed level (observation level) is dictated by these levels, a natural classification scheme is induced. In practice, we may study how the images of the map change when the preimages are perturbed. Preimages sharing the same image are in the same group. Notice that larger classification units have larger basins of attraction, if the map is interpreted as a dynamical system.

This idea has been utilized to study the long time asymptotic behavior of differential equations.(6, 7) The two levels mentioned above adapted to this case are the level of the equation dictating the system and the level of its long time asymptotic behavior. The map is the time-evolution itself. We perturb the original system and watch whether it affects long time behavior or not, studying the secular behavior arising from it. The resultant RG equation characterizes the universality class, allowing us to classify the differential equations.

The two representative description levels of organisms are the molecular level and the phenotype level. The former is the level that generates the system we are interested in. The latter is also almost the unique level maximally disparate from the former. It is the asymptotic outcome of the molecular specification for each organism. A natural classification scheme of organisms must then be the classification of molecular level specifications according to the natural relation between the two levels. This relation is ontogeny = time evolution. Therefore, *developmental taxonomy* is the most straightforward application of the idea explained above as the RG theoretical classification; the analogy with the RG theory for differential equations is straightforward.

Developmental taxonomy is not new at all, although the extreme assertion that classification must use *only* the developmental process might be new. The question is whether we can make higher order taxonomic levels objective from this point of view. If we pursue the analogy with differential equations, we should modify the molecular level and watch whether marked changes can result or not. If the answer is affirmative, the modification is regarded relevant to the higher order taxonomic level. However, since we do not know the logic of genetic control of developmental processes, artificial modification of genetic systems is not yet practical for developmental taxonomy. Therefore, for the time being, the means we can employ is mostly comparative. This is of course the traditional and still the main-stream approach of the so-called evo-devo program.

3. POSSIBILITY AND OUTLOOK

Notable trends in the evo-devo program are the efforts to describe comparative results as explicitly as possible in terms of genes. A typical proposal is as follows.(8) The phylum level classification of animals is characterized by the set of *Hox* genes. The distinction in the broad pattern enhancers (= sequences that are targets for regulation of transcription) characterizes the class level. Changes within *Hox* domains, and changes in the downstream genes (directly) controlled by *Hox* genes correspond to the order and lower taxonomic levels. In other words, this proposal asserts that the genetic control hierarchy corresponds to the taxonomic hierarchy.

The problem is, however, that such straightforward "genocentrism" does not simply work. This is partly because it is likely that the more important is the structure, the more ways there are to make it. This is exactly the situation we encounter in physics. The so-called phylotypic stage (= the developmental stage where the phylum specific body plan is blocked out) is a typical example. The expression "developmental hourglass" is used to illustrate the diversity of its formation modes(9) (the constricted part corresponds to the phylotypic stage).

Developmental process of a structure is governed by a gene network, but the relation between the structure (phenotype) and the gene network governing it is not as simple as the latter determining the former; for example, a genetic system can adapt to an already existing phenotype. A recent work by Salazar-Ciudad *et al.*(10) may illustrate this point of view. Even after consolidation, itineration among equivalent gene networks (that is, networks giving the same phenotypic effects) could happen just as there are many different primary structures giving the same tertiary structure of a protein family.(11) Phenotypes are generally more stable and long-lasting than genotypes just as thermodynamics transcends mechanics. Therefore, to understand classification of organisms, we must somehow go beyond genocentristic understanding of developmental processes.

Then, a logical conclusion is that developmental taxonomy is impossible (premature) because of our poor understanding of ontogeny. This is certainly a sound conclusion at present, and this corroborates the old saying that to accomplish the natural classification we must have complete understanding of biological systems. However, still there are two things we may be able to do: to explore the limitation of genocentrism, and to look for at least tentatively some characteristics of higher level taxonomic units in terms of the gene control hierarchy.(12)

Developmental taxonomy suggests at least two interesting (and perhaps deep) problems. One is related to the observation that there seems to be a selector gene for each network, making the latter into a semi-independent module. Clarke and Mittenthal wrote a seminal paper(13) on the theoretical understanding of this modularization with a switch. The possibility of hierarchical classification could be a consequence of modularization of the gene network structure. How general is this structure, and what is its general consequence? This modular structure seems to allow replacement of selector genes without affecting very much the network under their control. This replacement may correspond to some orders.

The other problem is about the nature of genetic information coded in the base sequence. The diversity of the enhancer structure even among closely related species implies the flexibility of the enhancer stressed by Dover.(14) We tend to believe that vitally important base sequences (in contrast to junks) evolve very slowly or are well preserved (e.g., histone 3). As to structural genes this is understandable. However, as to control genes or coding of developmental information it is conceivable that important information is coded in fast changing sequences lest parasites and pathogens exploit the organism. It is dangerous to conclude simply that the highly non-preserved sequences are unimportant. The vital message must be preserved, but encoding schemes may evolve fast (and perhaps must change fast to outwit potential exploiters). The buffering mechanism(16) in genetic systems could enable (or may even enhance) this cryptic encoding.

ACKNOWLEDGMENTS

I gratefully acknowledge constructive criticisms of Jay Mittenthal. I wish to dedicate this essay to Mike Fisher on the occasion of his 70th birthday. I am indebted to him for his constant encouragement and support throughout my career in the US. Mike's recent biophysics research encourages my own perhaps wilder projects.

REFERENCES

1. Y. Oono, *Int. J. Mod. Phys. B* **12**:245 (1998).
2. There is a serious problem even at the species level. According to cladistics, if a different taxonomic group sprouts out from a group, the latter is no more a natural taxonomic group and is called a paraphyletic group. If we accept approximate self-similar nature of the phylogenetic tree, a conclusion is unavoidable that most species would be paraphyletic and so not good taxonomic units. Developmental taxonomy could be a natural way to avoid this problem
3. J. Hey, *Trend. Ecol. Evol.* **16**:326 (2001).
4. A more biological version appeared as an invited paper in *Bull. Biogeographical Soc. Japan* **56**:45 (2001) [Japanese].
5. M. E. Fisher, *Scaling, Universality and Renormalization Group Theory*, in Lecture Notes in Physics, Vol. 186, No. 1, F. J. W. Hahne, ed. (Springer, Berlin, 1983); *Rev. Mod. Phys.* **70**:653 (1998).
6. An elementary introduction may be found in Y. Oono, *Internat. J. Mod. Phys. B* **14**:1327 (2000). The relation between reduction and RG was first recognized in L.-Y. Chen, N. Goldenfeld, Y. Oono, and G. Paquette, *Physica A* **204**:111 (1993).
7. K. Nozaki and Y. Oono, *Phys. Rev. E* **63**:046101-1 (2001).
8. G. Gellon and W. McGinnis, *BioEssays* **20**:116 (1998).
9. R. A. Raff, *The Shape of Life* (Chicago University Press, Chicago, 1995).
10. I. Salazar-Ciudad, S. A. Newman, and R. V. Sole, *Evolution & Development* **3**:84 (2001).

11. M. O. Dayhoff, R. M. Schwartz, and B. C. Orecutt, *Atlas of protein Sequence and Structure*, Vol. 5, Suppl. 3, M. O. Dayhoff, ed. (National Biomedical Research Foundation, Washington, D.C. 1978), pp. 345–352.
12. However, still within the genocentrism framework, some suggestions are possible: (1) Large scale gene network organization characterizes superphylum classification units such as Ecdysozoa; (2) It is likely that the genetic system evolves to place crucial gene networks under the control of a small number of key genes.(13) Orders may be characterized by these selectors; (3) Difference in enhancer structures among closely related species(14, 15)) may imply that genera are characterized by essentially the same phenotypes with different enhancer structures.
13. B. S. Clarke and J. E. Mittenthal, *Bull. Math. Biol.* **54**:1 (1992).
14. G. Dover, *BioEssays* **22**:1153 (2000).
15. D. L. Stern, *Evolution* **54**:1079 (2000).
16. E.g., C. Queitsch, T. A. Sangster, and S. Lindquist, *Nature* **417**:618 (2002).