Wulfenia **10** (2003): 1–14

Mitteilungen des
Kärntner Botanikzentrums
Klagenfurt

The New Phylogenetics: An essay

Igor Y. Pavlinov

Summary: This essay deals with the biological and philosophical roots of the modern approach to phylogenetic reconstructions which might be called the "New Phylogenetics". Its emergence was first stimulated by new premises in ontology and epistemology that made "population thinking" inadequate and led to the rebirth of "phylogenetic thinking". The new phylogenetics is the result of the joining of three independently developed scientific branches – cladistics, numerical phyletics and molecular biology. Some characteristic features of each component of the new phylogenetics discriminating it from both classical and pure phenetic approaches, are considered. The new phylogenetics is historically bound and will be replaced in time by some other phylogenetic approaches that would bear less of reductionist epistemology.

Zusammenfassung: Dieser Essay behandelt die biologischen und philosophischen Wurzeln der modernen, „Neue Phylogenetik" genannten Auffassung phylogenetischer Rekonstruktionen. Ihr Entstehen wurde durch neue Prämissen in Ontologie und Epistemologie angeregt, die das "Populations-Denken" unzulänglich machten und zur Wiedergeburt des "phylogenetischen Denkens" führten. Die neue Phylogenetik ist das Produkt der Vereinigung dreier unabhängig voneinander entwickelter Wissenschaftszweige – Kladistik, numerischer Phyletik und Molekularbiologie. Einige charakteristische Merkmale jeder der Komponenten der neuen Phylogenetik, die sie sowohl von klassischen, als auch von rein phenetischen Auffassungen unterscheidet, werden behandelt. Die neue Phylogenetik ist historisch gebunden und wird zu gegebener Zeit durch andere phylogenetische Auffassungen, die eine weniger reduktionistische Epistemologie enthalten, ersetzt werden.

Keywords: Phylogenetics, cladistics, phenetics, molecular phylogenetics, phylogeny, biophilosophy, epistemology

The second half of the 20th century witnessed the growth of a great interest in phylogenetic reconstructions at macrotaxonomic level which replaced microevolutionary studies of the "new systematics" and phenetics dominating the 40s–70s. However, it was not just a homage to classical approaches to historical reconstructions in biology. Instead, a quite new approach emerged which, by analogy with the "new systematics" mentioned above, could be called the "*New Phylogenetics*".

As a rule, a new scientific branch comes to existence due to divergent evolution of its predecessor. For instance, various schools of phylogenetics (in broad sense) diverged as a result of different definitions of the concept of monophyly. Contrary to this, the new phylogenetics evolved from a kind of "reticulate" evolution: the marriage of the Hennigian cladistics with numerical tree-constructing techniques molecular biology gave birth to it. Thus, it can be defined as a branch of phylogenetics aimed at *elaboration of "parsimonious" phylogenetic hypotheses on the basis of mainly molecular data by means of numerical methods of cladistic analysis.*

The new phylogenetics raised from various sources and therefore is arranged by the block principle: cladistics properly constitutes its biological core as a discipline that defines phylogeny; numerical phyletics provides a specific toolkit for studying phylogeny and a definition *how* it can be studied; and, molecular biology provides phylogenetics with a *new kind of data* never available before.

However, it would be hasty to connect the raise of the new phylogenetics solely to these three particular disciplines. As a matter of fact, there seemed to exist more profound causes of emergence of the modern approach to historical reconstructions in biology.

The new phylogenetics – why?

We have to consider that each scientific paradigm constitutes a kind of unity of interrelated epistemological and ontological premises. The first one defines “allowable” methods of scientific exploration of the objective world; the second one defines the parts of that world which are “allowable” for scientific exploration by those methods.

Taking this preamble into consideration, it should be recalled that the positivistic epistemology was ruling over biology during the first half of the 20th century. According to this, the world explored by “positive science” consists of observable and measurable entities. This world is simple and allows as simple descriptions and explanations which mean reduction of complex phenomena to sets of their elementary constituents. Respectively, unobservable and nonmeasurable entities were declared as “metaphysical” and were set aside from “positive science”. Besides, the latter deals with strictly (or at least numerically) universal statements about equilibrium systems and principally reversible processes. Earlier Popperian epistemology agreed with these basic proclamations, just having changed the way how scientific generalizations could be tested as true or false.

In historical biology, this philosophy was adopted by so called *population thinking* in which the evolutionary idea was reduced to the microevolutionary theory. According to this, any and all entities and processes are to be denied inasmuch as they could not be reduced to populations and population interactions. From this point of view, historical groups of organisms were declared as “non-science fictions”, and the biological diversity was reduced to the diversity of species. It is more than evident that such an evolutionary reductionism did not favour phylogenetic reconstructions. As a result, classical phylogenetics dealing with such reconstructions was ruled out from the mainstream of biology. Of course, it did not mean rejection of phylogeny at all: it was evolutionary causation that attributed to the population level thus having made phylogeny a kind of epiphenomenon of population interactions.

Population thinking at methodological level was accompanied by what might be called *phenetic thinking*. Its epistemology was (and is) based on the famous declaration that “science begins with measurements”. Another important idea that inspired phenetists was one of *overall similarity* supposing summation of similarities calculated for each of the equally weighted characters. And to make the similarity really “overall”, as many characters as possible were supposed to be included in numerical investigations.

At the beginning of the second half of the 20th century, serious changes occurred both in ontology and in epistemology that highly influenced historical biology in general and phylogenetics in particular.

It was of great importance that the emergence of a new ontological worldview routed in the Prigoginian *non-equilibrium thermodynamics*, or synergetics. It postulated that “the world is not being but becoming” and that the temporal behaviour of non-equilibrium systems is basically irreversible. This worldview gave raise to such significant aphorisms as “each being is developing” and “any developing being is a victim of its own history”. It means that, in order to understand the beings of our time, one has to trace their histories – a possibility which is provided due to particular historical reconstructions. Both the physical world and the biota were supposed to be such non-equilibrium systems. And historical development of the biota, with phylogenesis as an integral part, appeared to be such kind of temporal behaviour.

This ontology made mainly inadequate both classical positivistic and earlier Popperian epistemologies. The metaphysics of history returned to the bosom of biological science which made phylogenetic reconstructions justified by their own and an integral part of causal explanations in biology. The gradual loss of explanatory power of the microevolutionary theory made it actual to put forward a sacramental question: "What is beyond Neo-Darwinism?" The answer appeared to be *phylogenetic thinking*, or "tree thinking" according to O'Hara, which replaced population thinking. It meant that biological diversity was modelled by an inclusive hierarchy of monophyletic groups produced by phylogenetic processes irreducible to population ones. Elucidation of interaction not among populations but rather among phylogenetic lineages turned out to be the primary goal of evolutionary biology.

Comparable with new epistemology, the most influential on the new phylogenetics was the development of the most recent version of so called Popperian, or post-positivistic paradigm of the philosophy of science. That influence was basically three-fold. Firstly, this version rejected the previously declared idea of the "poverty of historicism" and acknowledged usefulness of historical explanations as of at least "metaphysical research program". This allowed phylogenetics to overcome a provincial complex formerly cultivated by the predominance of the physicalist philosophy of science. Secondly, Popperism presumes that any general statement in science has an epistemological status of hypothesis. This makes any phylogenetic reconstruction a kind of hypothesis which has to be elaborated and then tested with more or less formalized rules. At last, Popperism provided a methodology of the new phylogenetics with the principle of parsimony which meant reductionism at epistemological level.

This reductionism, not very different from positivism integrated into population and phenetic thinking, became inherent to the new-born phylogenetic mode of argumentation. It appeared to be very important for putting a demarcation line between the new and classical phylogenetic approaches. And each of the roots of the new phylogenetics mentioned above – cladistics, "numeristics", and molecular phylogenetics – contributed its own portion of reductionism. Cladistics based its keystone monophyly concept on the Darwinian model of evolution at species level. Numerical phyletics was largely phenetic in its understanding of the procedure of phylogenetic reconstructions. And molecular phylogenetics based its basical concept of the "molecular clock" on the theory of neutral mutations developed by population genetics.

It is an interesting fact that all three root disciplines which formed the new phylogenetics emerged – though independently but quite simultaneously of each other. It were the middle sixties when the "American era" of cladistics began with a translation of the fundamental monograph by Willi Hennig and the term "cladistics" was coined. At the same time, the first publication written by pheneticians appeared to show how a dendrogram produced by means of a strictly phenetic numerical method could be interpreted as a phylogram providing certain assumptions. It was during the same years when the first papers were published which contained constructed molecular trees called "phylogenetic" by their authors. At last, a vivid discussion of an application of the Popperian paradigm to systematics and phylogenetics was started at the end of the sixties with a heavy dispute between the proponents of evolutionary taxonomy and cladistics.

The cladistic root

To make impact of cladistics proper to the new phylogenetics more clear-cut, we should consider that it was originally a classificatory doctrine and not a phylogenetic one (although it was infrequently called so). That is, it was aimed at solving certain problems relating rather to the pattern and not to the process per se.

The classic phylogenetic idea was that classification is to be based on phylogeny but it should not strictly reproduce it. This lead to many well known ambiguities that made classifications unstable. And cladistics raised to develop a method which would allow to reflect as unambiguously as possible the results of the phylogenetic process by means of classification.

The now well known cladistic resolution was: to achieve this goal, it is needed to reduce the process in question to a less ambiguous component. It seemed to be *cladistic history* producing a hierarchy of strictly monophyletic (*holophyletic*) groups which is best represented by a branching pattern of a dichotomous cladogram. First of all, this reduction implies elimination of adaptationist treatment of evolutionary process characteristic for classic phylogenetics and requesting for detailed evolutionary scenarios. As a consequence evolution is no longer treated as a regular process; instead, a *principle of minimum evolution* is adopted according to which random changes prevail over regular ones.

Such a simplified view of evolution deprives cladistics of the lack of an a priori weighting of characters according to their "adaptive significance" which was propagated by classic approaches and was strongly criticized by positivistic methodology. According to cladistics, each character is to be evaluated as just a *signifier of cladistic event(s)*, and the more characters are shown to indicate the event the more reliably supported it is. Such character treatment leads directly to involving numerical techniques into cladistics reconstruction (see the next chapter).

One of the most profound input of cladistics into the new phylogenetics at operational level (adopted by proponents of the classic approach as well) was the *synapomorphy principle* deduced from the postulated relation between character transformations and the nested hierarchy of holophyletic groups. This principle asserts it that only a derived character state inherited from the nearest common ancestor (called *apomorphies*) defines a holophyletic group while an initial character state (*plesiomorphy*) cannot do so. Thus, nested similarity of taxa by derived states of all characters (called *synapormophy*) defines the nested hierarchy of holophyletic groups, that is the phylogenetic pattern. Contrary to this, *symplesiomorphy* relates to more distant origin and does not bear on uncovering of this hierarchy. In turn, the latter corresponds to (allows to hypothesize) the consequence of *cladistic events* constituting the contents of cladistically interpreted phylogeny.

As far as phylogeny (process) and not the intergroup relations (pattern) is concerned, the basic task of cladistic analysis can be formulated as *sequencing* of the cladistic events. At more technical level of consideration, both hierarchy of holophyletic groups and the corresponding sequence of cladistic events are usually represented graphically by a stylized phylogenetic tree known as the *cladogram*. As a consequence of propositions of the graph theory the desirable sequence is defined ultimately by the position of the *root* of the cladogram. And the principle in question serves as an instrument allowing to accomplish the task.

This is provided by a *hierarchy of synapomorphies assumption* according to which the tree root corresponds to that level of hierarchy at which the number of symplesiomorphies is maximal while the number of synapomorphies respectively is minimal. Rooting of the tree by reference to the synapomorphy principle can be achieved by two methods, a direct and an indirect one, corresponding to a *strong* respectively a *weak* formulation of this principle.

At earlier stages of developments of cladistic methodology, the synapomorphy principle was adopted in the strong form which meant that apomorphic and plesiomorphic states for a given character (known as *character polarity*) were to be recognized prior to cladogram constructing. According to the assumption given above, polarities of the entire set of the characters could serve as the basis for defining root of the tree: it is the one which corresponds to the maximal amount of symplesiomorphies over the set of all characters.

Subsequent development of cladistic methodology employing a more rigid perusal of the parsimony principle rejected the idea of prior determination of character polarities as burden with surplus of evolutionary scenario. The problem of rooting the cladogram was suggested to be resolved by means of the *outgroup concept* which corresponds to the weak formulation of the synapomorphy principle. Though pretty technical, this method deserves a little bit more close consideration as it highly affects the procedure and results of phylogenetic reconstructions.

For the sake of the parsimony principle, evolutionary scenario underlying the reconstruction is reduced to an assumption that the study group (or *ingroup*) is monophyletic in relation to the outgroup and that they both are *sister groups*. Correspondingly, the root of the tree for the ingroup is axiomatically defined by the node at which the outgroup joins the tree. It is evident that this node is the one at which the number of synapomorphies for the ingroup is minimal, as the ingroup is just at the starting point of its history. Hence, from the hierarchy of synapomorphies presumption mentioned above it follows that this node axiomatically defines the entire hierarchy of synapomorphies for the ingroup which, in turn, is interpreted as reflecting sequence of all cladistic events in the history of this group.

The result of the procedure just shortly outlined is that the synapomorphies for the holophyletic groups are defined a posteriori to reconstruct the cladogram. Thus, paradoxically enough, the weak formulation of the synapomorphy principle makes it possible to recognize cladistic relationships on the basis of similarities that are not cladistically interpreted initially: they appear to be pure phenetic. The latter allows to employ numerical estimates of similarity without prior estimation of character polarity as a part of procedure of phylogenetic reconstruction. It might be advantageous under some circumstances, as it provides certain problems in demarcation between correct phylogenetic and phenetic interpretations of results of data analysis (see the next chapter).

From the classical viewpoint, cladistics is a formalized approach allowing to recognize monophyletic groups and to give their synapomorphy based diagnoses. Parsimony cladistics stops with this aim; evolutionary cladistics goes farther and “grafts” a “flesh” of evolutionary scenarios onto a cladogram “skeleton”. Application of some formal rules based on the same synapomorphy principle makes it possible: (a) to estimate the degree of divergence of members of the groups by a number of respective apomorphies, (b) to recognize occasionally "cladistic ancestors" of these groups as those belonging to their basic radiation having no autapomorphies, (c) to reveal parallels and reversions by analysis of apomorphy distributions over the cladogram and to make some conclusions about evolutionary trends of the entire study group.

The numerical phyletic root

Originally, cladistics did not imply use of numerical techniques for producing phylogenetic trees. Neither Hennig himself nor his earlier adherents in Europe and in the United States dared to use them. However, as it was indicated above, their approach was explicitly based on counting synapomorphies to reveal nested branching pattern of relations among monophyletic groups. And this alone served as a precondition of introducing strictly numerical approaches into the new phylogenetics which are known as *numerical cladistic analysis*, or numerical phyletics.

Such a direction of development of the new phylogenetics towards its "numerization" appeared to be demanded by the phylogenetic community of the second half of the 20th century. On the one hand, it corresponded to a commonly acknowledged positivist formula which is connecting science with measurements. On the other hand, certain parts of the phylogenetic community was prepared to adopt the numerical approaches by preceding development of numerical phenetics. At last, computer technologies made numerical calculations of phylogenetic trees using standard computer programs, quite an easy and hence attractive job.

As indicated above, it was phenetists and not phylogenetists that first did exercises of such a kind. Several years elapsed before earlier research in quantitative cladistics properly raised. At present, numerical approaches in phylogenetic reconstructions absolutely predominate, especially when concerning with molecular data, and numerical phenetics is largely replaced by numerical phyletics. Again, this discriminates strongly the new phylogenetics from the classical one.

Numerical phyletics is an analytical approach which calculates both the similarities among organisms and the trees to reveal their phylogenetic relationships. It is based on the following basic premises differing drastically from those of the classical approaches. First of all, the organism is reduced to a set of non-correlated characters. Then, as a consequence of the previous fact, goes the reduction of the evolutionary process, according to the concept of minimum evolution, to a stochastic one in which characters and their states are allowed to be combined freely. Using a set of formally uncorrelated characters allows to assess kinship relationships by employing numerical similarity coefficient(s) in which inputs of particular characters are "dissolved". The first two statements relate to ontology of the analyzed entities, the last one concerns epistemology of phylogenetic reconstructions.

All these essential features of the new phylogenetics are considered the severe shortages of this approach by the traditional morphological school which takes a) the organism as a whole in its evolutionary developments, and b) the evolution as an orderly process in which properties of the organisms don't change randomly. Contrary to this, numerical cladistic reductionism is highly acknowledged in molecular phylogenetics. Indeed, it opens a wide possibility to compare organisms by a very large set of traits for which any prior polarities are more than uncertain; a situation which is lightening up when molecular data are coming to mind.

Use of basic phenetic ideas in estimating phylogenetic trees made phenetic thinking implicitly penetrating the new phylogenetics. One of the consequences appeared to be turning the latter from methodology (in a wider, philosophical sense) to technology. Such a trend quite fits predictions of the Kuhnean model of paradigmatic development of any scientific discipline: discussions of theoretical problems are unavoidably being replaced by discussions of technical puzzles. The phylogenetic methods are now substantiated en mass not by their biological consistency but by their mathematic correctness (which is not bad itself but quite insufficient), how fast they process the large data matrices and the like.

Like the phenetics, recent numerical phyletics provides a great variety of particular algorithms of tree constructing based on different theories. Parsimony analysis, maximum likelihood analysis, compatibility analysis, etc … And, discouraging enough, such things as neighbourhood joining, UPGMA and the like could also be met in some papers. The first list contains basically cladistic algorithms, the second list contains phenetic ones. How it might happened?

It happened because of the lack of clear understanding of *the* biological ideas underlying technical devices of cladistic analysis. It makes sense among different procedures conventional and arbitrary. This concerns, first of all, the methods of tree rooting: some of them make phylogenetic trees while some turn them into phenograms.

In this connection, we should underline that the use of certain phenetic approximations in numerical phyletics is warranted by a weak formulation of the synapomorphy principle. A key part of this formulation is the outgroup concept which makes tree rooting based on phenetic similarity estimates phylogenetically consistent (see preceding chapter).

But if a new phylogenetist is not aware of this circumstance why, indeed, is she / he obliged to apply outgroup rooting and not midpoint rooting? – especially if the latter one is also available in some computer programs designed for phylogenetic reconstructions as it is proclaimed by their names: popular PAUP is just one of the examples. Now suppose a molecular tree-maker working with an equipment which facilities include a tree-producing program based on, say, UPGMA algorithm. She / he inputs DNA samples, obtains a tree, calls it “phylogenetic”, and publishes a paper entitled “Phylogenetic analysis of …”. But, methodologically, it is *the* phenetic and not phylogenetic analysis: it produces *the* phenogram and not phylogram, even though it is based on genetic and not phenetic data.

One of the merits of numerical cladistic analysis, from an evolutionary point of view, is a possibility to manipulate with datasets: to determine or not prior character polarity, to study different subsets of characters and / or taxa separately. By virtue of these manipulations one can explore various evolutionary scenarios for the study group. For instance, changing character polarity allows to find out which particular scenario assumption about ordered or random character transformations, produces a phylogenetic hypothesis most corroborated or least falsified by other kinds of data (geographic distribution etc). Another possibility is to study separately various biological subgroups, such as larvae and imagoes in insects, to see how concordantly they were evolving. At last, obtaining trees for different character sets which can not be combined in one datamatrix and then calculating a probabilistic supertree for the set of initial trees allows to reveal which monophyletic groups are strongly and which are weakly supported by the entire data pool.

It is noteworthy that the above mentioned evolutionary scenarios supplementary to proper cladistic analysis (see the last pragraph of the previous chapter) are easy to “write” by applying quite formal numerical techniques. It requests that the tree resulted from the analysis is not a standard ultrametric cladogram, but a metric *phylogram* with branch lengths directly proportional to patristic distances. In case of molecular data, the phylogram is calculated directly from the datamatrix, for morphological data this distance is calculated as the number of apomorphies between two nodes. Based on such a tree, divergence is estimated numerically as respective patristic distance; if this distance between a terminal group and preceding (counting from the tree root) node is zero, then this group coincides with that node and is considered as a cladistic ancestor.

As hierarchy of monophyletic groups is deduced from hierarchy of synapomorphies, there are evolutionary scenarios under which cladistic analysis fails to reveal phylogenetic pattern. I.e.: it is unable to produce a strictly dichotomous cladogram; instead, a polychotomous tree appears which corresponds to incompletely resolved cladistic relationships. Most obviously it happens because of mosaic (irregular) evolution of the study group with discordant character transformations causing their inability to testify consistently the sequence of cladistic events in the history of the group.

The molecular phylogenetic root

Molecular phylogenetics introduced a fundamentally new kind of data into phylogenetic reconstructions unavailable in times of classical approaches. On the one hand, they fit well to the emergence of a new image of biology as a basically “molecular” science. On the other hand, one of the undoubted merits of these data became a possibility of direct comparison of organisms taking quite distant position on the tree of life – say, mammals, flower plants and cyanobacteria.

First experiences in molecular phylogenetics took place in the middle of the 60s of the 20th century, simultaneously with the first publications in proper cladistics and numerical phyletics. Molecular phylogenetics appeared to be numerical from the very beginning of its development. At that time it was basically phenetic by its methodology, because standard phenetic clusterization algorithms were employed for graphical representation of distance matrices calculated from molecular data. Serious debates with “morphological” phylogenetists became necessary to make it evident that, for the molecular trees to be phylogenetic, their constructing had to follow certain methodological principles elaborated by cladistics. Among these principles, one based on the outgroup concept appeared to be of most fundamental significance: as it was explained above, its neglecting makes analysis yet phenetic rather than phylogenetic, a point not accepted and traced so by some practising molecular tree-makers.

The principal idea of molecular phylogenetics borrowed directly from Sir Charles Darwin is more than clear-cut: phylogenetic reconstructions are to be based on structures which changes during the course of evolution and are basically out of control by natural selection, that is non-adaptive. The theory of neutral mutations, developed within the population genetics, has led to the theory of the so called *neutral evolution* which was accompanied by the principle of minimum evolution mentioned above. The essential of living beings was assumed to be their genetic information written in the form of DNA chains; accordingly, the essential of evolutionary process was assumed to be changes of this information – hence, of those chains. All these premises constituted an evolutionary model of clearly reductionist kind.

Methodology of molecular phylogenetics is based on this model and hence is also a reductionist one. At first we should assume that in order to reconstruct *the* true phylogeny, it is requested to compare organisms by their DNAs, to obtain some numerical estimate of similarity by these DNAs which give, by definition, estimate of similarity of their genetic programs, and then to convert the similarity matrix in a tree which is presumed to be phylogenetic just because it is based on DNA data.

As evolution is a temporal process, the *molecular clock concept* was elaborated by molecular phylogenetics to tie the DNA similarity among organisms with real time of their divergence. At first, this concept was considered as the most promising; however, the subsequent investigations showed that tempo and modes of evolutionary changes of various fragments of DNA (and, later, of RNA) are quite different, so the concept appeared to be generally untrue. Besides, calibration of this clock was based initially on comparison of molecular distances with known paleontological records which makes the entire approach strongly dependent on not very reliable data.

History of molecular phylogenetics repeats, maybe with some variation, the pathway of any newly born approach. As always, it promised to resolve all phylogenetic problems within an observable period. Similarity was first taken for homology and each tree obtained for a particular DNA a RNA fragment first declared reflecting the phylogeny. However, it became clear pretty soon that different fragments may provide nonconcordant trees which obliged to treat *gene trees* (semophyleses of morphologists) and *species trees* (phylogenies proper) separately. Similarity was then acknowledged to be different from homology at molecular level just as it is at morphological level, and various kinds of molecular homology are now acknowledged: *orthology* is identical with homogeny, while *paralogy* corresponds with serial homology recognized by morphologists long ago. Thus, at present, proponents of molecular phylogenetics become aware of the methodological problems they met which are mostly the same morphologists have been trying to solve for more than a hundred years.

Certain problems are exposed by the character concept in molecular phylogenetics. Earlier attempts to connect a character to some biologically meaningful part of the genome, to a gene or to a codon, failed for various reasons. At the moment, a character of molecular phylogenetics is understood as the site, that is, as a certain position in the DNA/RNA sequence. Respectively, various nucleotide bases which can take this position are treated as character state. From this evidently reductionist character definition a problem results that doesn’t trouble molecular phylogenetists but morphologists. It concerns incomparability of character amounts used by these two clusters of investigators: it appeared that the former are able to manipulate with many thousands of separate traits (the usual length of sequenced chains) in contrast to several dozens (or hundreds, at most) of morphological traits. Because of this, combining these two kinds of data into one datamatrix results in the molecular traits just “devour” morphological ones during parsimony analysis. Therefore, a more reasonable practice seems to be combining the resultant trees to find the nodes of their agreement.

Results of molecular phylogenetics actually repeat those obtained by morphological data in some instances, thus corroborating them, while contradicting them to more or less degree in others. Frequently such contradictions appear to be so just at first glance, when predominating concepts are considered. As each of the approaches doesn’t produce a single phylogenetic scheme, the conflicts sometimes happen to find pretty simple solutions: whatever paradox a particular molecular phylogeny might look, a similar morphological one appeares to have been forwarded some time ago and has just been rejected and forgotten by recent morphologists. Such a situation, for instance, occurred with so called Ecdysozoa, a concept connecting arthropods with nematomorphs rather than with annelids. However, in several instances molecular phylogeny may contradict all the morphological reconstructions known to have been suggested before and the groups assumed monophyletic in the former are not (by the moment?) corroborated by any of morphological synapomorphies. An example is the split of eutherian mammals into afrotherian and laurasiatherian lineages by a set of molecular data: the problem is, how could it be that evolution, provided it has actually produced such a dichotomy, did not leave any trace in morphological features?

A huge amount of data gathered today in the Internet genebanks whips up numerical phyletics to elaborate new tree producing algorithms that allow to make phylogenetic reconstructions for thousands of organisms at once within a reasonable time. However, there is a kind of pitfall in such reconstructions resulted from methodology of these algorithms. They all are based on minimum tree length optimality criteria: this means, for instance, that a structure of relationships among mammal taxa may depend on optimization of the tree fragment with bacteria, and vice versa. And one may wonder if there is a biological sound in such an interdependence.

The hypothetico-deductive background

As it was stated at the very beginning of this essay, the new phylogenetics has been developing under the influence of the hypothetic-deductive argumentation scheme. According to it, the body of scientific knowledge consists of hypotheses which are to be formulated and tested following certain principles. Thus, for the science of the new phylogenetics to meet this condition, its generalizations are to be considered as hypotheses, referred to as *phylogenetic hypotheses*. Also a notion of *cladistic hypotheses* is spread in the literature, as it is usually elaborated by means of cladistic analysis.

One of those principles, maybe of universal meaning, states that any hypothesis can be formulated within a framework of certain theory only, the latter defining certain general properties of the object being hypothesized. In respect to phylogenetics it means that, in order to put forward a hypothesis about a particular phylogenesis, one has to know what is *the* phylogenesis in general. This knowledge is provided by a kind of theory or, rather, of *model of phylogenesis* which presets an ontological background for the phylogenetic reconstruction.

Very important for understanding this point is Goedel's *incompleteness theorem* which was originally proved in the formal arithmetic and now is usually taken in a more wide, epistemological meaning. It serves as an argument in favour of the idea that, at a given level of generality, any proposition can be exhaustively explained only in terms of another proposition belonging to the higher level of generality. This means that the above mentioned model of phylogenesis does not appear as a "deus ex machina"; instead, it can be explicitly formulated only in terms of a more general theory or model, which is an evolutionary theory, whatever might be its particular version.

Taking all this into account, it seems that a popular slogan of the orthodox proponents of parsimony cladistics: no theory is better than a bad theory – is essentially wrong. It belongs to positivist epistemology rather than to post-positivist one, that is, it is largely outdated (or it does mean that parsimony cladistics' reconstructions are not hypotheses about phylogeny).

For this general idea to be operational, it is to "rule" some way over methods of phylogenetic analyses "turning" comparative data into phylogenetic hypotheses. This is provided by the *correspondence principle* which establishes the following causal relation between the background theory and a particular phylogenetic reconstruction.

Provided the general evolutionary theory or model (such as "descent with modification") is valid, some properties of historical development are asumed which are most relevant to the phylogenetic tasks. One of them is a certain relation between the history and the kinship formalized by the principle of *irreversible kinship decrease*. Another is a certain relation between the kinship and the similarity: in the new phylogenetics, this is the relation between holophyly and synapomorphy. Based on the latter relation, algorithms are elaborated which allow to express kinship in terms of similarity in phylogenetically correct manner: the parsimony analysis based on outgroup concept lined out above is an example. At last, this or any other algorithm, in form of a particular method, is applied to a dataset and a partial phylogenetic hypothesis is formulated which is consistent with underlying evolutionary model due to the scheme just outlined.

From the facts given above it follows that there cannot be a phylogenetic method "neutral" in respect to some evolutionary theory. Returning to the very beginning of this essay, the following paradigmic example can be given to show dependence of selection of methods, with which biological diversity is investigated, upon properties of this diversity defined by a biologically sound theory. Diminishing of population thinking and rise of phylogenetics has led to replacing standard statistical methods by various clusterization (broadly speaking) techniques. The cause is that the statistics deals with so called *statistical ensembles* with random distribution of variables. Biological population, with certain reservation, can be taken for such an ensemble but not a monophyletic group. Its membership and character state distribution are both bounded by kinship best represented by nested hierarchy that cannot be analyzed adequately by routine statistical approaches.

It is needless to say that there is a noticeable discordance amongst different schools of the new phylogenetics concerning contents of the background evolutionary theory: they are considered in short in the next chapter. Here it is proper to consider just few points relevant to the subject under consideration.

First, the requirement is correct formulation of operational conditions under which a phylogenetic hypothesis is elaborated. Among other things, it involves the number of characters upon which such a hypothesis is to be based. Parsimony cladistics asserts that the phylogenetic hypothesis is to be deduced from the entire set of characters under investigation: the respective tree summarizes, in a minimally controversial manner, all available information. However, this is a phenetic position: the more characters are used the better is the estimate of overall similarity. And parsimony principle is appealed to just to formulate an optimization criterion, which is minimum tree length defined by overall sum of patristic distances.

However, initial hypothetic-deductive understanding of the parsimony principles is significantly different: it requests minimization of available data as a source of a hypothesis for the latter to be as much falsifiable as possible. That is why the Popperian paradigm considers, for instance, a conjecture as a valid source of scientific hypothesis. At the operational level, it means that not all but some of the characters are better used in formulating a hypothesis, others being used in its testing. This means a kind of differential character weighting, an approach advocated by both classical phylogenetics and evolutionary cladistics. In particular, an approach based on compatibility analysis seems to be most warranted by hypothetic-deductive methodology.

Another topic is the correct formulation of the epistemological status of the cladistic hypothesis. The problem here is that the ontological background of the new phylogenetics expressed by minimum evolution principle is of such type that it contradicts some fundamental properties of the development processes, especially their certain integrity. In particular, this background presumes that random changes prevail over regular ones during evolution. As it follows from concluding comments in the chapter above on numerical phyletics, such an assumption, be it true, would lead to completely unresolved cladograms. To the extent that in practice the trees are obtained with pretty high resolution, it indicates that evolution actually is not random.

Thus, taking into account that cladistic reconstruction of the history of monophyletic groups is fulfilled under premises as weak as possible about properties of evolutionary development, it seems reasonable to treat it as a specific kind of *null hypothesis*. Such a treatment seems to be in accordance with the cladistic analysis being aimed at searching for more and more “short” trees, pending on data available, which means that results are less and less biologically sound.

Schools of the new phylogenetics

Analogy between biological evolution and development of a scientific paradigm means that the latter usually evolves in divergent fashion. For instance, classical phylogenetics during the 50s–60s of the 20th century diverged into Simpsonian evolutionary taxonomy and Hennigian cladistics. The same occurred with the new phylogenetics: it diverged in several schools and branches. Moreover, it was initially doomed to be differentiated by several developmental trajectories, because it was born from various sources and is arranged by the block principle.

According to this, the new phylogenetics is divided, first of all and quite apparently, by a “substrate” criterion into morphological and molecular schools. Their discrepancy is evident both in their biological backgrounds and in the particular results. Another cause of differentiation is the methodical one: disputes between advocates of parsimony and maximum likelihood methods examplify this. However, it is more interesting to explore another, methodological aspect of the emergence of various schools of the new phylogenetics.

Optimally principle asserts it that any developing complex system is able to be optimized by only one crucial parameter at a time. And observing development of a system from this point of view allows to reveal such a parameter.

As far as phylogenetics is concerned, the divergence of classical domain into above mentioned evolutionary taxonomy and cladistics in the middle of the 20th century is usually attributed to different definitions of monophyly adopted by them, “wide” and “narrow” ones, respectively. True as it may look, cladistic refinement of that concept did not appear from the blue, it was actually caused by certain methodological premises summarized by the parsimony principle. So, it was the adherence to the latter principle by which the new phylogenetics was being optimized from the very beginning of its history. And it is not surprising that trajectories of its further development seemed to be determined also by the parsimony principle. Namely, its schools appeared and channeled because of different interpretations of two subjects: which and to what extent the prior judgments about evolution are to be minimized in accordance to requirements of that principle.

Hennigian approach was first to minimize evolutionary scenario to several simple statements outlined above in short. However, it allowed at least some prior judgment about pathways of character evolution on which prior character polarities were based. In subsequent times, this approach, in its original Hennigian formulation took shape of the so called *evolutionary cladistics* now pretty popular among many morphologists. As for the molecular phylogenetists, at least those implying maximum likelihood or Bayesian probability approaches also belong to this school, with some reservations.

From the Hennigian starting point, two argumentation and development lines in cladistics emerged, each with its own vision of what was the "right" treatment of the parsimony principle.

The so called *parsimony cladistics* reduces evolutionary scenario to exclude possibility of prior character polarity. However, important is that at least reference to evolution remained as a background of cladistic reconstructions. It is this branch of the new phylogenetics that dominates today and deals mostly with elaborating new numerical techniques. Molecular phylogenetics based on parsimony algorithms and correctly using outgroup concept is methodologically a part of this branch of cladistics.

The *pattern cladistics* goes much farther and declares that its aim is just to uncover some hierarchical pattern of monophyletic groups. Thus, a presumption of this pattern is taken as a sole background for cladistic explorations of this kind. It is the reference to monophyly that is reserved only as a reminiscence of evolution in the ontological background of pattern cladistics. This concept is perhaps a cladistic one but it is certainly not a phylogenetic one: it means homage to typology and today it is abandoned.

It is very normal for a scientific discipline to develop such a divergent way and to produce several different approaches to resolve the same problem. However, there is certain jeopardy hidden in multiplicity of particular approaches. On the one hand, it means impossibility to suggest a unified "true" method allowing to uncover the "true phylogeny" which is the only one by definition. It's needless to say that such an inability causes nothing but disappointment in the entire discipline: such was one of the causes of the fall of phenetics during the 70's. On the other hand, it provokes a further evolution of the discipline, now controlled not by search for other technical tools most appropriate from the parsimony point of view, but by trying to formulate another, biologically more sound task. The latter means, evidently, another change of the phylogenetic paradigm.

In place of a conclusion: what is beyond?

The above mentioned analogy between biological evolution and development of a scientific discipline means that there are two processes that come together in the "scientific evolution": falling down of old approaches and emergence of new ones. Taking this into account, it is very easy to foresee that the new phylogenetics, as a historical unit, is doomed to lose its position dominating nowadays and to give place to some other approach to phylogenetic reconstructions. But what in particular would be that *other* approach?

Being a non-equilibrium system, any scientific discipline develops an unpredictable way. However, there are some general development principles (or laws) which constrain the possible pathways of the history, so unpredictability of the latter is not absolute. One of these is the *novelty principle* which postulates that a new approach would emerge with a new feature that negates some essential features of the preceding one (Darwinian “descent with modification”). Another one is the *succession principle* which ensures that any new approach inherits at least some of the features of its predecessor(s). One more is a *dialectical spiral pathway principle* according to which the historical development tends to repeat at least some essential features of more ancient concepts. At last, there is the so called *swing principle* postulating that, for a theoretical concept, there exist at least two opposites and various treatments of this concept are “floating” between them.

Considering emergence and development of the new phylogenetics from this point of view, it is easy to notice that (a) its ideas appeared to negate non-historicism of population thinking which (b) meant certain homage to phylogenetic thinking; however, (c) it inherited some elements of reductionism developed by both positivism and earlier post-positivism.

As for the swing principle, for the phylogenetics, as a part of biology, the core concept among which opposite treatments of approaches use to “swing” is the *integrity*. At one extreme, minimal integrity of historical developments (phylogeny), of its subjects (taxa) and of its results (phylogenetic pattern) is supposed to be. It presumes random evolution of populations due to uncorrelated changes of their traits, a version of population thinking once popular. At another extreme, maximal integrity of the evolution is assumed which means strong succession (= minimal randomness) of changes and certain integrity of both, monophyletic groups and entire phylogenetic pattern. This idea was probably most expressed by the Leibnitzian idea of emergence of the *scala naturae* which had exerted much influence on earlier phylogenetic schools of the 19th century.

As it was underlined above, reductionism is one of the essential features of the new phylogenetics delimiting it from the “old” one. So it won’t be surprising if the next step in the history of phylogenetics would imply the rebirth of some “forgotten” ideas of classical approaches. It would imply paying more attention to integrity of evolving organisms and to elaboration of the methods which would incorporate this integrity into the techniques of phylogenetic tree constructing.

Acknowledgements

The author is indebted to Dr. Roland Eberwein for his suggestion to submit the article, and to Mag. Lucie Weingartner and Ruth Smech-Filley for their efforts to make the English text more readable.

Address of the author:

Dr. Igor Y. Pavlinov
Zoological Museum
Moscow Lomonosov State University
ul. Bol’shaya Nikitskaya 6
Moscow, 125009
Russia
E-mail: igor_pavlinov@zmmu.msu.ru

Zeitschrift/Journal: Wulfenia

Jahr/Year: 2003

Band/Volume: 10

Autor(en)/Author(s): Pavlinov Igor Y.

Artikel/Article: The new Phytologenetics: An essay 1-14