Zootaxa 2946: 60–64 (2011)
www.mapress.com/zootaxa/

Editorial

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Opening Pandora's Molecular Box*

MALTE C. EBACH[1], MARCELO R. DE CARVALHO[2] & DAVID M. WILLIAMS[3]

[1]*Evolution & Ecology Research Centre, School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia E-mail: mcebach@gmail.com (corresponding author)*

[2]*Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, Trav. 14, no. 101, São Paulo, 05508-900, SP, Brazil E-mail: mrcarvalho@ib.usp.br*

[3]*David M. Williams, Botany Department, The Natural History Museum, Cromwell Road, London SW7 5BD, UK E-mail: dmw@nhm.ac.uk*

**In*: Carvalho, M.R. de & Craig, M.T. (Eds) (2011) Morphological and Molecular Approaches to the Phylogeny of Fishes: Integration or Conflict?. *Zootaxa*, 2946, 1–142.

Tell you what they're gonna do
Started doing it already
Got to find something new
Looking for it in genetix

Found a new game to play
Think it's impossible to lose...
(The Stranglers - Genetix)

Introduction

Mooi & Gill (2010) have prised open the cap of the molecular systematics vial and caused a debate to take-off in the ichthyological community. Molecular trees and their supporting evidence are the first two items to leave this Pandora's box, closely followed by DNA barcoding and DNA taxonomy. In short, the debate is fuelled by the nature of molecular data: can nucleotide sequences provide the necessary evidence for relationship? The majority (Wiley et al., 2011) believe that DNA contains informative data; however, in our view, they have failed to ascertain the truth of their claim. Not all data are informative. Data may provide supporting evidence, conflicting evidence, or no evidence at all. Assuming that all data are informative *apriori to analysis* is a theoretical position, not an empirical one. We claim that systematics is, quite the contrary, empirical, and relies on evidence rather than on implicit measurements of data. Consequently, this assertion leads back to the original question of evidence in molecular systematics, namely molecular homology.

Comparatively few authors deal with the comparison of molecular homology and morphological homology. A lack of theory on part of molecular systematists has led to a rather basic understanding of molecular relationship (i.e. similarity between aligned sequences). Similarity as relationship, whether it be 'special similarity' (Farris, 1977) or 'overall similarity' (Sneath & Sokal, 1973), is nothing more than two objects compared in some way. Homology, however, is a three-item relationship in which two homologs are more closely related to each other than they are to a third. This means homology can be defined as 'affinity' or 'sameness', that homologous relationships can be observed and quantified. Similarity is just one increasingly superficial aspect of homology and not, as some claim, part of a 'test' (contra Patterson, 1982 and de Pinna, 1991; but see Rieppel & Kearney, 2002). Afterall, we do not "test", even with congruence, our initial similarity assessments among different taxa once the characters are deemed to be true homologs. This misunderstanding of the difference between molecular similarity and molecular homology lies at the heart of Mooi & Gill's argument. Without addressing homology in molecular systematists we

 Accepted by M. Craig: 2 May 2011; published: 8 Jul. 2011

create an unsustainable science---one that has the tendency to make unsupportable claims of relationship. Unless addressed, this approach may lead to the undoing of molecular systematics.

A Worst Case Scenario

Consider the following (fictional) account:

> "Darwin Year 2059 marks the 200th anniversary of Origin of Species and the 250th birthday of its author. Celebrations worldwide are overshadowed by an evolutionary break-through—Phylogeology. Scientists at an undisclosed institution have successfully analyzed 100 species through mass spectrometry—a technique commonly used in geochemistry. The procedure is quick and simple. Samples of whole organisms are broken down into their atomic components, analyzed for percentages of 50 common elements present in living organisms. The discovery that each species has its own chemical signature heralds a new age in taxon identification. A small tissue sample, like skin, blood, scales, feathers or leaves alone are required to accurately identify a known species. Further research suggests that these signatures share similarities with closely related species, thus eliminating the need for costly and problematic molecular data. Gone are datasets plagued by xenology, gene duplication, bad alignments, and paralogy. Phylogeology only needs the chemical signature in order to identify taxa and determine their phylogenetic relationships. Phylogeology will also revolutionize taxonomy. Para-taxonomists will be able to discover new species and assign object identifiers for each new name. Specimens can be kept online and in museum collections on glass slides or small vials, therefore reducing the cost of storing specimens.
>
> Already institutions are investing heavily in mass spectrometers and cheap and effective taxon identification. The amount of time and money saved in training alone would be phenomenal. A technician can be trained within a week to use the mass spectrometer, identify taxa and determine phylogenetic relationships. Scientists estimate that all known species and their phylogenetic relationships can be sequenced and determined within two years. Research funding bodies have praised this as an end to taxonomic and phylogenetic "dark ages"—no more taxonomic impediment! Now all organisms can be catalogued with handheld mass spectrometers. A whole new generation of "para-taxonomists" can discover new species without the rigmarole of costly taxonomic monography and molecular phylogenetic analysis".

The assumption that closely related organisms share similarities, for example, similar geographical distributions, behavioral patterns and chemical make-up, is not a new idea. For instance, DNA fingerprinting relies on similarities in DNA to associate criminals with crime scenes and match the bones of victims to their families based on statistical probabilities. On no account does DNA definitively disclose who begot whom within a family genealogy. Molecular systematists, however, argue that they do provide such proof, and they use the same 'evidence' paleontologists did in the early to mid 20th century, namely similarity and taxonomic authority to link one specimen or taxon to another in a series of ancestor-descendant ghost lineages. The basis of their claim lies in the misinterpretation of molecular data as the 'units of hereditary'. A similar catechism 'units of evolution' is employed in population dynamics, evolutionary taxonomy and paleontology to substantiate their claims of inter-breeding populations, ancestral taxa or missing links. These 'units' seem to be getting continually smaller, from Haeckel's stammgruppen (i.e. paraphyletic taxa), to Mayr's populations (i.e. individuals), to alleles, to chromosomes, to DNA, to RNA. The above fictional example of phylogeology takes this apparent "reductionism" to a final and logical conclusion—if there ever was a true unit of hereditary, surely it must be the atom? If we equate phylogeology as an analogy to molecular systematics, we would reach the conclusion that any two similar quantifiable objects can be used to relate taxa. Why then treat DNA base-pairs as the 'magic molecule' when atoms are far more accurate and free of any error (e.g. xenology, mutation, paralogy etc.)? One may dismiss this as an irrelevant assumption, but yet it is the same claim used to favor molecules over morphology.

The 'worst case scenario' above is nothing more than an analogy to the rise of molecules over morphology, or of information over knowledge. At first molecular data was justified based on the remarkable overlap between molecular and morphological trees. Following this, and based on the premise of the molecular and morphological

overlap, morphological rather than molecular characters were mapped onto molecular trees in order to show support for the use of molecular data as 'evidence'. Now, when molecular and morphological trees conflict it is the morphological trees that are questioned for validity (Scotland et al. 2003). Yet one may ask where the evidence is for rejecting morphological over molecular trees? Ironically, that evidence lies in the original overlap between both types of data. If we now trace this contradiction via our analogy, we may state that phylogeological analysis between taxa, using only their chemical compositions, uncovers a tree that is identical to a molecular tree---evidence that atomic trees can uncover genealogies. We can now map the molecular characters onto the atomic tree. Additionally we find that mass spectroscopy is a very quick, cheap and viable alternative that takes only minutes to learn. Further analysis shows that atomic and molecular trees conflict and the former are used as 'evidence' because molecular data is known to be riddled with errors, duplications, dodgy primers and so on. By 2059 systematics will have moved into the atomic era.

Citing evidence

The phylogeology analogy has helped to highlight two important points:

1. Reducing the size of data increases its number (herein the Law of Large Numbers) and;
2. Similarity is secondary evidence.

The Law of Large Numbers is used as a way to support data as 'evidence'. For instance, take three taxa: A, B and C (Nelson & Platnick, 1981). Taxa A, B and C share 99 data points, whereas taxa B and C share one additional unique data point. The resulting relationship is: A(BC). In this case the one unique data point is more meaningful in terms of evidence because it identifies the closer relationship between B and C in comparison to A. The number of data points is not sufficient to make that data 'evidence' because evidence is dependent upon the meaning it provides. Mammals, for instance, can be related by the presence of hair and lactating glands. These two characteristics relate a large body of organisms. Hair and lactating glands are both data and evidence. Stating that mammals are related based on a seemingly endless list of similarities is secondary. Naturally, a taxon like Mammalia will also share many similarities that cannot be used as primary evidence, such as 'walking on all four limbs or 'presence of an endoskeleton', because these are also characters that appear in other groups. Secondary evidence is dependent on primary or independent evidence. Molecular data, we argue, is secondary evidence, because it relies on primary or independent evidence currently found in morphological data.

A sample taken from a DNA sequence contains a string of base-pairs that are similar within a group of organisms. In order to find primary evidence, molecular systematists would need to understand what that sequence does, how it expresses itself phenotypically, and how that sequence relates to the organism. In other words evidence contains meaning, that is, a greater significance in context to the whole organism.

One way to find independent evidence in context to the organism is to reopen the debate concerning molecular homologies. If data do point to a series of similarities that match morphological trees, then surely there could be molecular homologs that are related. The molecular homolog would have structure (i.e. A[TT] on position, say, 616), and ontogeny (i.e. the sequence produces a certain hormone in the hippocampus), and a 'geography' and 'topographical relationship' (i.e. it affects development of the thyroid). In other words molecular homologies are supported by a context (i.e. the threefold parallelism of Form, Time and Space) and a form of independence (i.e. A[TT] relates certain taxa but not others). Using the threefold parallelism of Agassiz (1859; Williams and Ebach, 2004) may be a step back into the ancient literature, but it is gigantic leap forward for comparative biology, an area that molecular systematics has avoided addressing. We believe that by introducing fundamental concepts like the threefold parallelism and morphological homology, molecular systematics can build up a theory that would serve as a foundation to finding molecular homologies.

Taxonomies

A viable molecular taxonomy would treat molecules as morphology. If we can apply the threefold parallelism to molecules, then there is no need to treat molecules any differently to morphology. First, molecular homologs need

to be given in context of the organism. A molecular homolog is a part of an organism that manifests itself in other taxa. Moreover the molecular homolog is part of the threefold parallelism. Molecular homologies, then, are defined as the relationship between two or more molecular homologs (i.e. the smallest being a three-item relationship), a notion that may open up methodological innovations.

Once a molecular homology is established there is no reason to dismiss molecular characters from traditional taxonomy. If molecular characters can be used as evidence for homology, then they would serve as valuable characters for taxonomy, as we will show. However, before we do, it is important to demonstrate the difference between molecular characters in taxonomy and the recent calls for DNA Taxonomy and Barcoding, which lies in the difference between using meaningful molecular homologs and homologies and using arbitrary, meaningless sequences. On one hand, the threefold parallelism demonstrates that molecular data do not have to be fundamentally different from morphological data. Morphological and molecular data have form, they occur in space and they certainly have a developmental aspect, whereas both DNA Taxonomy and Barcoding assume that quantifying molecular data at some level is sufficient and proponents ignore any theoretical aspect of molecular features. As in the phylogeology example above, both DNA Taxonomy and Barcoding are used simply as a means to an end—to catalog and recover data respectively for identification purposes (see Ebach & de Carvalho, 2010; Will et al., 2005).

And then there was Hope...

Anyone but a jaded systematist would call the above arguments 'anti-molecular'. We feel that this misunderstanding arises from the notion that a 'war' is raging between morphologists and molecular systematists, where one side is made up of technophobes hurling fossils at the clean technology of genetics. It is erroneous to suggest that all molecular systematists are geneticists, just as it is erroneous to say that that all morphologists are palaeontologists (see Ebach and Williams 2005). A more accurate description would be that the field of systematics is divided between those who classify based on homology and those that group based on similarity. Moreover, the latter group is the one that has adopted the technology and ignored the theory. We make this claim because molecular systematics needs to show us the molecular homologs and homologies as defined above to make their data meaningful. These are the foundations of molecular systematics, not computer algorithms or nifty new ways to sequence data. Molecular systematists need to return to theory, rediscover the foundations of systematics and develop the necessary methodologies and numerical implementations. Once discovered, molecular homology has great potential. For instance, the Linnaean system of classification, as it stands today, can accommodate molecular homologs and homologies to classify taxa.

Molecular systematics needs redoing, not undoing. Its undoing lies in the push for DNA Taxonomy and Barcoding and the 'numericalization' (i.e. lack of separate, homological identity) of molecular data without call to supporting theory. It rests in the notion of large numbers, and the faulty belief that homology is similarity. Molecular systematics is selling itself short and leaving itself open to the worst case scenario of phylogeology.

> "It would be too much to hope, however, that we would be done, once and for all, with the wizard and his advices. He is too cunning a tempter to be permanently banished from reasoned discourse, and he will not be long in making his reappearance, cloaked in the attire of the next fashionable movement in systematics. I hope we will be able to recognize him when he comes along" (Nelson, 1978: 111–112).

References

Agassiz, L. (1859) An essay on classification. London: Longman, Brown, Green, Longmans Trübner & Co.

Ebach, M.C. & Williams, D.M. (2005) Molecular systematics is not genetics. *Rivista di Biologia*, 98, 373–376.

Ebach, M.C. & de Carvalho, M.R. (2010) Anti-intellectualism in the DNA barcoding enterprise. *Zoologia*, 27, 165–178.

Farris, J.S. (1979) The information content of the phylogenetic system. *Systematic Biology*, 28, 483–519.

Mooi, R.D. & Gill, A.C. (2010) Phylogenies without synapomorphies---a crisis in fish systematics: time to show some character. *Zootaxa*, 2450, 26–40.

Nelson, G.J. (1978) Professor Michener on phenetics--old and new. *Systematic Zoology*, 27, 104–112.

Nelson, G. & Platnick, N.I. (1981) *Systematics and Biogeography: Cladistics and Vicariance*. New York: Columbia University Press.

Patterson, C. (1982) Morphological characters and homology. *In*: Joysey, K.A. & Friday, A.E. (Eds), *Problems of Phylogenetic Reconstruction*. Academic Press, London, pp. 21–74.
de Pinna, M.C.C. (1991) Concepts and tests of homology in the cladistic paradigm. *Cladistics*, 7, 367–394.
Rieppel, O. & Kearney, M. (2002) Similarity. *Biological Journal of the Linnean Society*, 75, 59–82.
Scotland, R.W., Olmstead, R.G. & Bennett, J.R. (2003) Phylogeny reconstruction: the role of morphology. *Systematic Biology*, 52, 539–548.
Sneath, P.H.A. & Sokal, R.R. (1973) *Numerical taxonomy: the Principles and Practice of Numerical Classification*. San Francisco: W.H. Freeman.
Wiley, E.O., Chakrabarty, P., Craig, M.T., Davis, M.P., Holcroft, N.I., Mayden, R.L. & Smith, W.L. (2011) Will the real phylogeneticists please stand up? *Zootaxa*, 2946, 7–16.
Will, K.W., Mishler, B.D. & Wheeler, Q.D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, 54, 844–851.
Williams, D.M. & Ebach, M.C. (2004) The reform of palaeontology and the rise of biogeography—25 years after 'Ontogeny, phylogeny, paleontology and the biogenetic law' (Nelson, 1978). *Journal of Biogeography*, 31, 685–712.