

Reconsidering the phylogenetic utility of miRNA in animals

Casey W. Dunn¹

Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912

Phylogenetics, the study of evolutionary relationships, is critical to understanding the history of life on Earth, integrating and synthesizing information collected across diverse organisms, and testing hypotheses about evolutionary patterns and processes. Most phylogenetic studies are now based on detailed analyses of gene sequence evolution. Over the last several decades, these studies have provided a much better understanding of the structure of the Tree of Life. This progress has also highlighted and helped define many tough phylogenetic questions that have not been answered yet with available data and analysis methods. This has inspired the search for new categories of data that could succeed where analyses of gene sequence evolution have not yet provided strong, consistent phylogenetic support. In PNAS, Thomson et al. (1) examine one of these new categories of data (2–5): the presence and absence of microRNA (miRNA) families. They find that there are several reasons why they may not be as useful for estimating phylogenies as has been proposed.

miRNAs are small RNA genes that are involved in posttranscriptional gene regulation (6, 7). They originated independently in plants, animals, and several other organisms (8). miRNAs and the enzymes needed to process them are missing entirely from comb jellies (ctenophores), suggesting that they may have arisen within animals rather than before the most recent common ancestor of animals (9). Different miRNA families are found in different animals, providing a potential source of information on animal relationships. The sequence of each miRNA family is highly conserved, which makes them relatively easy to tell apart.

Why was it believed that investigating the presence and absence of miRNAs could succeed where other types of phylogenetic character data have failed? It comes down to homoplasy: any evolutionary pattern of character evolution that deviates from the simplest possible evolutionary history (Fig. 1). Homoplasy can be due to a variety of processes, including convergence and

secondary reversal. Homoplasy is common in molecular sequence data, which is one of the main reasons that answering some phylogenetic questions has been so tough. The proponents of miRNA phylogenetic analyses have argued that homoplasy is exceptionally rare in miRNA presence/absence data and that miRNA therefore provides a clear picture of evolutionary relationships (2–5). Because of the way they originate, the convergent acquisition of miRNA families is extremely unlikely, and each is thought to have a single evolutionary origin (3). This means that if the same miRNA family is present in two different species, it was also present in the most recent common ancestor of these species. In addition, however, it has also been claimed that miRNA is rarely lost once it is gained (2, 3, 5). If this is true, then all species that share a particular miRNA are more closely related to each other than to any species that lacks this miRNA. The lack of homoplasy would lead to a ratchet-like evolution of miRNA content through time. miRNA families would accumulate in different lineages without being lost, and their distribution in living species would provide an unambiguous picture of phylogenetic relationships. This has led some to conclude that miRNA “are potentially the near homoplasy-free data set that systematists have long wished for” (3).

Thomson et al. (1) evaluate three specific issues: the hypothesis that secondary loss of miRNA is rare, the suitability of the analysis methods that have been applied to miRNA evolution, and the ability to correctly detect the presence or absence of miRNA. They find critical problems on each of these fronts. Their analyses indicate that secondary loss of miRNA is widespread. This fundamentally undermines what had been proposed as their unique value, remarkably low homoplasy, for phylogenetic inference. Thomson et al. (1) then apply Bayesian phylogenetic inference methods that can better accommodate secondary loss, revealing considerable phylogenetic uncertainty. Finally, Thomson et al. find that miRNA families often go undetected

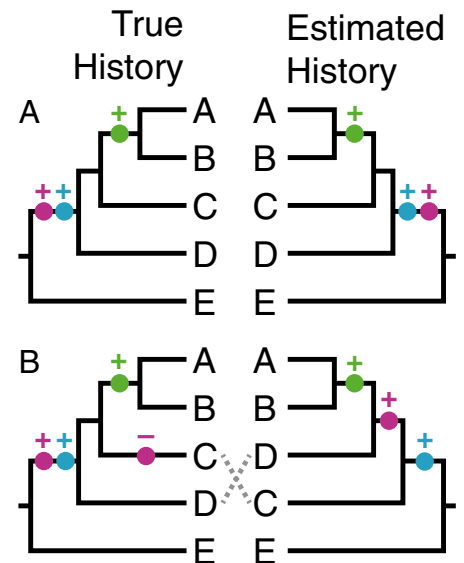


Fig. 1. A simplified example of the impact of unrecognized homoplasy on phylogenetic analysis. Evolutionary changes in each miRNA family are indicated by a different color. (A) If miRNA evolution proceeds without homoplasy, there would be a single origin (indicated by a plus) and no losses for each miRNA family. A simple model can recover true relationships. (B) If there is homoplasy, in this case a single loss (indicated by a minus), and it is not recognized with additional data and robust analysis tools, phylogenetic estimation can provide the wrong answer (C and D are switched).

using standard observation methods. This suggests that, in addition to often being lost, miRNAs are sometimes not detected. Thomson et al. (1) reexamine five previously published phylogenies based on miRNA and find that when these problems are taken into account, the results of all these studies are fundamentally altered. The support for stated claims is found to be very weak in two of the studies, and in three of the studies, a strongly supported contradictory conclusion is reached.

Why were these strong claims about low homoplasy in miRNA evolution made in the first place, and why did it take so long to systematically examine them? The amount of

Author contributions: C.W.D. wrote the paper.

The author declares no conflict of interest.

See companion article on page E3659.

¹Email: casey_dunn@brown.edu.

homoplasy that can be observed depends on how well a character is sampled. It is only in the last decade that a concerted effort has been made to collect miRNA data across phylogenetically diverse animals (2). Initial glimpses of miRNA diversity were consistent with low homoplasy (2), but the sampling was not yet sufficient to rigorously test the hypothesis that they had low rates of loss. The original claims that there are no losses have been tempered through the years as more data became available (5), but it was still claimed that they were remarkably low and little was done analytically to accommodate these losses.

There are multiple historical precedents for the problems identified by Thomson et al. (1). Considerations of each type of new phylogenetic character data often initially assert that homoplasy is low, and these claims are later refuted when sufficient data become available to actually test them (10). Early animal phylogenies were based on the assumption that complex morphological traits such as body cavities and segmentation were difficult to gain and lose and therefore exhibited little homoplasy. It is only when other categories of data, particularly molecular sequence data, were obtained that these old assumptions became testable hypotheses. Many have been soundly rejected (11, 12). Segmented animals such as arthropods and annelids, for example, do not form an evolutionary clade to the exclusion of unsegmented animals. Arthropods and annelids are distantly related (13), and segmentation has been repeatedly lost within annelids. Another historical precedent for underestimated homoplasy is the phylogenetic analysis of short and long interspersed elements, mobile genetic elements that replicate in the genome. The insertion of these elements was proposed to be nearly homoplasy free (14), making them an ideal phylogenetic marker that did not even require statistical analysis (15). In a striking parallel to miRNA, later analyses made it clear that there was homoplasy and that observation problems compromised their phylogenetic utility (10). Rather than continue to repeat this history of underestimating homoplasy, future considerations of new categories of phylogenetic data should start with the null hypothesis that homoplasy is higher. Low homoplasy should be invoked only if sufficient data later reject this null hypothesis.

What are the implications of the findings of Thomson et al. (1)? There is bad news and good news. The bad news is that miRNAs are less useful than had been proposed for

resolving difficult phylogenetic problems. This does not necessarily mean that miRNAs cannot make useful contributions to estimating phylogenies; they just have the same limitations as other data. They may still prove to be useful when analyzed in combination with other data. Scoring the presence and absence of miRNAs based on full genome sequences rather than tissue-specific transcriptomes will help resolve some of the

Thomson et al. find that miRNA families often go undetected using standard observation methods.

sampling issues described by Thomson et al. (1). The major challenge, however, will be to develop phylogenetic analysis methods that can simultaneously evaluate many different types of data, including sequence evolution, gene gain and loss, and genomic rearrangements, in a single unified framework. This echoes past attempts to integrate morphological and molecular sequence data. As molecular datasets grew much larger than morphological datasets, the question became how to weight the datasets. Developing methods that can simultaneously assess multiple categories of data will be one of the major methodological challenges for the field of phylogenetics in coming years.

The good news is that we are now in a better position to learn about miRNA function. Homoplasy creates challenges for estimating phylogenies, but it is tremendously helpful for testing mechanistic hypotheses

about character evolution. This is because homoplasy provides replication. If a trait has a single evolutionary origin and no losses, it is hard to know if that one change is associated with evolutionary changes in other characters. If the trait has multiple changes, then it is possible to test if other organism characters repeatedly change each time the trait changes. This is particularly helpful in the case of understanding miRNA evolution, as some very bold hypotheses have been proposed for their functional evolution. It has been claimed that increased miRNA diversity has enabled lineages to become more complex than other lineages with fewer miRNA families (2, 16, 17). In this scenario, a ratchet-like accumulation of miRNA families is seen as the cause of a ratchet-like increase in complexity (18). In this way, miRNA has been invoked as an explanation for the “taxonomic hierarchy of animal relationships” (2). However, the relationships of living animals are not hierarchical but are tree-like. Organisms cannot be arrayed on a single spectrum from simple to complex, akin to Aristotle’s *scala naturae*, with a parallel spectrum of miRNA diversity. Organisms have many traits that vary independently, and each living species, which is at the tip of a branch of the phylogenetic tree, can have a mix of complex and simple traits that are gained and lost independently. With greater homoplasy in miRNA evolution, it will be easier to learn if any of these changes in specific complex traits are associated with gains and losses of particular miRNA families. Homoplasy is not always the enemy; it can sometimes be your friend.

- 1 Thomson RC, Plachetzki DC, Mahler DL, Moore BR (2014) A critical appraisal of the use of microRNA data in phylogenetics. *Proc Natl Acad Sci USA* 111:E3659–E3668.
- 2 Sempere LF, Cole CN, McPeck MA, Peterson KJ (2006) The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. *J Exp Zool B Mol Dev Evol* 306(6):575–588.
- 3 Sperling EA, Peterson KJ (2009) *Animal Evolution: Genomes, Fossils, and Trees*, eds Telford MJ, Littlewood DTJ (Oxford Univ Press, Oxford, UK), pp 157–170.
- 4 Dolgin E (2012) Phylogeny: Rewriting evolution. *Nature* 486(7404):460–462.
- 5 Tarver JE, et al. (2013) miRNAs: Small genes with big potential in metazoan phylogenetics. *Mol Biol Evol* 30(11):2369–2382.
- 6 Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-4*. *Cell* 75(5):843–854.
- 7 Ambros V (2004) The functions of animal microRNAs. *Nature* 431(7006):350–355.
- 8 Tarver JE, Donoghue PCJ, Peterson KJ (2012) Do miRNAs have a deep evolutionary history? *BioEssays* 34(10):857–866.
- 9 Maxwell EK, Ryan JF, Schnitzler CE, Browne WE, Baxeavanis AD (2012) MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore *Mnemiopsis leidyi*. *BMC Genomics* 13:714.
- 10 Hillis DM (1999) SINEs of the perfect character. *Proc Natl Acad Sci USA* 96(18):9979–9981.
- 11 Edgecombe GD, et al. (2011) Higher-level metazoan relationships: Recent progress and remaining questions. *Org Divers Evol* 11(2):151–172.
- 12 Wagele JW (2014) *Deep Metazoan Phylogeny: The Backbone of the Tree of Life*, ed Bartholomaeus T (Walter De Gruyter Inc, Berlin).
- 13 Aguinaldo AM, et al. (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387(6632):489–493.
- 14 Nikaido M, Rooney AP, Okada N (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci USA* 96(18):10261–10266.
- 15 Murata S, Takasaki N, Saitoh M, Tachida H, Okada N (1996) Details of retropositional genome dynamics that provide a rationale for a generic division: The distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). *Genetics* 142(3):915–926.
- 16 Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ, Peterson KJ (2008) MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci USA* 105(8):2946–2950.
- 17 Berezikov E (2011) Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* 12(12):846–860.
- 18 Wheeler BM, et al. (2009) The deep evolution of metazoan microRNAs. *Evol Dev* 11(1):50–68.