EEB 122: Principles of Evolution, Ecology, and Behavior
Professor Stephen Stearns
Teaching Fellow: Andrea Hodgins-Davis

By submitting this essay, I attest that it is my own work, completed in accordance with University regulations. —Jared Shenson

"Junk": Breeding Innovation and Complexity

by Jared Shenson

We have long sought an answer to one of the most difficult questions evolution has posed: how did life become so complex? Centered around biology's original findings that the nuclear genome consisted only of protein-coding DNA and "junk" DNA, scientists have long held that major morphological development and complexity was the direct result of increasing numbers of and complexity of protein-coding genes. This was a natural conclusion as the prevailing understanding of the genome and gene regulation did not predict any other possible paths to such dramatic evolutionary changes. As biologists have more closely examined the so-called "junk" DNA in the past decade, it has become clear that most of it is not junk at all, but rather contains important non- coding sequences. With major research into the contents of these sequences in the last five years, a new, potentially causal factor in the evolution of morphological complexity – that is to say the increase of differentiated cell types (Valentine et al., 1993) or the number of different parts of an organism (McShea, 1996) – has been discovered that challenges the old theory of protein-coding gene expansion: the evolution of microRNAs.

MicroRNAs (miRNAs) are small 18-24 nucleotide RNAs that are able to regulate the expression of their target protein-encoding genes posttranscriptionally (Gu et al., 2009). miRNAs exist within the genome as DNA sequences usually located within intronic regions and are transcribed into long pri-miRNAs by RNA polymerase II. These preliminary miRNAs are then further processed by Drosha into pre-miRNAs during which they begin to exhibit characteristic

hairpin structures. These precursor miRNAs are then processed once again by Dicer to form the mature form of the miRNAs. As of 2006, more than 3000 mature miRNAs had been identified across all species (Lee et al., 2007). Once the mature form of the miRNA has been derived, the miRNA can bind to a target messenger RNA (mRNA) transcript, which exhibits a seven nucleotide section of the 3' untranslated region (UTR) complementary to the miRNA. There currently are three competing models to explain how this binding prevents the mRNA from being further processed; however, this paper will not go into examining them. The three models are: RNAi-like degradation of the mRNA; preventing translation by disrupting elongation or 5' capping; and increased destabilization of the mRNA half-life (Lee et al., 2007). Along with the *cis*-regulatory elements, miRNAs represent a large percentage of the highly important gene regulatory network. It is this network that many scientists now argue is responsible for the majority of morphological complexity, and many point to miRNAs as the key player in the process of gene regulation.

If we proceed with these two assumptions – (1) that miRNA evolution is responsible for morphological complexity because (2) miRNAs are crucial components of gene regulation – then we are left with a similarly challenging question to tackle: what is responsible for the evolution of new miRNAs? By answering this question and thereby understanding the manner and timeframe in which novel miRNAs arise, we will be better equipped to understand the evolution of morphological complexity – what it entails for the organism and what drives it – both in the past and for the future. This paper seeks to consider these two questions and concludes that there exists a substantial amount of evidence to support the argument that miRNAs are responsible for driving increased morphological complexity and are themselves evolved by a combination of innovation by de novo and duplication processes.

Currently two outstanding views exist to explain the increase in the miRNA complement throughout evolutionary history: (1) the de novo model whereby new miRNAs are produced from random sequences (Heimberg et al., 2008; Wheeler et al., 2009), and (2) the duplication model in which existing miRNA gene families are expanded by local or genome-wide duplication events (Hertel et al., 2006; Sempere et al., 2006; Gu et al., 2009). While strong arguments exist for both models, evidence has suggested that the correct answer may be a combination of the two approaches.

Nearly 40 years ago, Japanese geneticist Susumu Ohno proposed the theory of a two-fold genome duplication to explain the evolutionary processes that had taken the earliest eukaryotes and transformed them into organisms of increasing morphological complexity (Ohno 1970). It has been argued that genome duplication would be a viable causal factor for evolutionary change because "it creates redundant copies of protein-encoding genes with impunity," allowing for the duplicated copies to accumulate potentially beneficial mutations while the original copies continue to produce the expected protein products (Hunter 2008). This process then leads to new genes that can in turn enable increased organism complexity.

Studies of the human genome subsequent to the first complete sequencing of the genome have shown that at least one complete genome duplication event occurred, while the second event may or may not have. McLysaght et al. (2002) argue that at least one genome duplication event must have taken place to explain the extensive evidence of genomic duplication they found among early chordate evolution. Additionally, research by Gu et al. (2002) reveals that small-scale local duplications are occurring constantly within the genome. They postulate that given

the age distribution of those human gene families studied, a two-round genome duplication model cannot explain the observed data; rather they suggest that the data supports a hypothesis of one complete genome duplication event followed by continued local duplication. These results, however, do not conclusively prove that local or genome-wide duplication events will cause an increase in morphological complexity (Spring, 2002). However, many still attributed the duplication model to increased complexity, causing some scientists to more recently call for more rigorous experimentation and proof, beyond mere coincidence, that genome duplication can actually explain morphological complexity (Heimberg et al., 2008).

The first to propose that the theory of genome duplication additionally holds in the case of miRNA evolution and provide substantiated evidence to support the theory was Hertel et al. (2006). Hertel and colleagues performed BLAST searches of the MicroRNA Registry looking for matches in sequences to four available metazoan genomes (*Homo sapiens, Mus musculus, Drosophila melanogaster, Caenorhabditis elegans*). A BLAST (Basic Local Alignment Search Tool) search is a programming algorithm implemented within sequence databases to allow for the comparison of a queried nucleotide string to the known sequences in the database, and, subsequently, the identification of exact or near matches between the queried string and the known sequences, defined to a certain threshold. The results gathered by Hertel et al. were then placed in context with the gene and species phylogenies to identify the relationships between the evolution of species and miRNAs. The team's searches and subsequent analysis revealed the innovation of miRNAs across three specific episodes of time that also corresponded to major developmental innovations. First, a group of 20 miRNAs was discovered that can be found in both protostomes and deuterostomes – the two taxa that make up the Bilateria: those organisms with bilateral symmetry and three germ layers. Second, 56 new miRNAs were found at the

branch of the phylogenetic tree leading to vertebrates (members of the phylum Chordata, a descendent of the deuterostomes). And third, a smaller group of miRNAs were found to correlate with the development of eutherian mammals (the branch of placental mammals opposing the marsupials of the metatheria) (Hertel et al., 2006). In addition to these findings of novel sequences, Hertel et al. found many paralogous miRNA sequences across species throughout the metazoa. They classified these paralogs – homologous sequences that were separated by a genome duplication event – into two types of duplication events: local (tandem) duplications and non-local duplications. Among the non-local duplications, they found that nearly all of the identified sequences can be traced back to only two points in the metazoan (all multicellular animals) phylogeny. Given this strong evidence for miRNA radiation from the teleost (bony fishes) ancestor and later the vertebrate ancestor, Hertel and colleagues concluded that miRNAs evolved in a manner consistent with the 2R model of genome-wide duplications proposed by Ohno. Additionally, they argued that not all miRNA innovation is derived from genome-wide duplication events (GDE) but rather that miRNA innovation is an ongoing process resulting from local duplications and occasional de novo gene appearance.

The work of Hertel et al., although leaning heavily towards the view that duplication events are responsible for miRNA evolution, makes an interesting case for both sides of the argument. They acknowledge that some de novo innovation clearly occurred – citing some 200 miRNA genes found in vertebrates but not in bilaterians – but do not make a claim as to how important this novel development of miRNAs was. This makes it difficult to assess to what degree duplication and de novo innovation each play in the evolution of miRNAs in the metazoa. Additionally, their results are based on a relatively small dataset of only four species' genomes which weakens the authors' conclusions (Heimberg et al., 2008).

Interestingly, a similar dual-explanation theory was proposed two years prior when Tanzer and Stadler (2004) examined the history of the mir17 miRNA cluster. They concluded that the paralogs of the cluster's miRNA genes could likely be explained by an initial phase of local (tandem) duplications, followed by duplications of the entire cluster, and then loss of individual miRNAs from the copied clusters. However, they acknowledged that, at least in the case of the *mir-92* transcript – one of the miRNAs in the cluster – some of the observed variation could have been due to the de novo creation of a second hairpin structure and therefore not the result of gene duplication. While their study was of a very small representation of the entire miRNA complement, their conclusions were mostly in agreement with those of Hertel et al. in that genome duplication events were largely responsible for the expansion of the miRNA repertoire, but that some de novo innovation likely occurred as well.

With the basic phylogenetic analysis of miRNA evolution with the work of Hertel et al. (2006) as a starting point, Sempere et al. (2006) performed their own phylogenetic analysis of miRNAs through original total RNA analysis as well as BLAST searches in the microRNA Registry. By expanding their dataset, especially with the use of direct analysis of total RNA preparations from many more organisms than were available in the microRNA Registry to Hertel et al., Sempere et al.'s work represents a greater study of the phylogenetic distribution of miRNAs throughout evolutionary history. The conclusions concerning the distribution patterns of miRNA drawn by Sempere et al. based on the results of their two methods of inquiry are in line with those of Hertel et al., demonstrating that a large number of miRNA genes are shared across early eutherian ancestors with new groups of miRNA genes evolving with each new branch on the phylogenetic tree. From this data, Sempere et al. make a few important conclusions that have ramifications for the future study of miRNA evolution: once formed, a

miRNA gene typically experiences little if any change to its primary sequence and also remains within the genome throughout long stretches of evolutionary time as a result of intense negative selection against mutations. Given that fixation typically occurs and thus the rate of mutation is very low, the continued expansion of the miRNA repertoire observed by Sempere et al. after the initial intense miRNA radiation suggests that novel miRNA acquisition is continuous. This conclusion further adds to the argument made by Hertel et al. concerning the same subject. However, Sempere et al. do not conclude whether this continual acquisition is the result of de novo innovation or duplication events.

The three properties of miRNA described by Sempere et al. are very important in that they enable evolutionary biologists to accomplish two tasks: phylogeny construction and miRNA complement determination. Many methods have been employed over the years to construct phylogenetic trees, including phenotypic profiling and study of mitochondrial DNA. With the evidence that miRNA acquisition is closely correlated with taxonomic development, evolutionary biologists have a new method by which to construct a phylogenetic tree (Lee et al., 2007). Additionally, some scientists have proposed a biological clock based on miRNA acquisition, which can then be used to approximately date the divergence of different branches of the phylogenetic tree derived from other sources such as mtDNA or fossils (Gu et al., 2009). This knowledge becomes particularly salient in attempting to unravel the phylogenetic history of the relationship between sponges, cnidarians and triploblasts, as well as the placement of acoelomates in relation to platyhelminthes (Wheeler et al., 2009). Concerning the latter task of miRNA complement determination, once an accurate phylogenetic tree has been constructed, scientists can look for expected miRNA presence in specific species and, more importantly,

construct the miRNA complement of the last common ancestor of a certain clade, including the vertebrates (Heimberg et al., 2008).

Additional evidence and strong support for the theory of genome duplication events comes from the work of Gu et al. (2009). In work in 2007, Huang and Gu laid the foundation, developing a novel computational method for classifying miRNA families by locating homologous miRNA called the Phylogeny-Bootstrap-Cluster pipeline, which they claim "considerably improves the reliability of miRNA classification" (Gu et al., 2009). In the paper outlining the formation of this method, Huang and Gu (2007) argued for the existence of both local and genome-wide duplication events to explain the development of the specific miRNA families. The evidence for this claim, however, was largely left out as it was not the goal of their paper to discuss this theory. In the work published by Gu et al. in 2009, the team first employed the Phylogeny-Bootstrap-Cluster technique to analyze miRNA families followed by the nearest-neighbor clock technique (Gu et al., 2002) to perform molecular dating of the suspected miRNA duplication events.

Gu and his colleagues analyzed the genomes of six species – human, mouse, chicken, zebrafish, fruit fly, and worm – and created a map of all miRNA families found within these species. Using the well-known and widely-accepted neighbor-joining method with the simple Jukes-Cantor distance model (the best phylogeny model based on miRNAs according to Nei and Kumar, 2000), they created an inferred phylogenetic tree for the miRNA families observed and used their bootstrapping method to evaluate the statistical reliability of the constructed tree. As the large majority of the reliability values were above 80%, Gu et al. felt fairly confident with their inferred tree. With this tree, the duplicated miRNAs were located and the age of the duplication event that created the copies estimated by the nearest-neighbor clock method. In this

method, the branch lengths of the tree are calibrated based on the nearest established split-times of major species splits, including primate-rodent, mammal-bird, tetrapod-teleost, and vertebrate-*Drosophila*. A simple formula is then used to calculate the offset of the duplication event from the nearest split-time based on the ratio of branch lengths. This method is considerably more accurate than using a global molecular clock as the basis for the calculation (Gu et al., 2009). However, the analysis performed by Gu and his colleagues relies on an assumption about miRNAs that not all researchers agree upon: that miRNAs have undergone rapid gain and loss throughout evolution and it is not uncommon for an miRNA gene to be lost (argued against by Hertel et al., 2006; Sempere et al., 2006; Heimberg et al., 2008). It is unclear at this time which theory is correct.

The conclusions drawn by Gu et al. from their dating analysis reveal an interesting connection: the age distribution of the miRNA duplication events seems to align very nearly with the distribution of protein-encoding gene duplication events throughout evolutionary history. The one major exception noted by Gu and his colleagues was that following the mammalian radiation approximately 100 million years ago (mya), there was a large increase in protein-encoding genes but no similar expansion of miRNA genes. They attribute this to the large proportion of local or tandem duplications that occurred during the mammalian radiation, arguing that these kinds of duplications are much more common among protein-encoding genes than miRNA. Instead they suggest that the new miRNAs present in this most recent time period are the result of de novo innovation. The notion that miRNAs are not likely to undergo local duplication was not supported by earlier work by Hertel et al. (2006) who suggested that local duplication events are very common and explain much of the continued miRNA expansion. Additionally, it is important to consider that the age distributions determined by Gu et al. must be interpreted

within context and that they may not accurately follow the fossil record. Gu et al. claim in the paper that they have addressed this concern and validated that this is a non-issue, but the data shown to prove this are not clearly explained. That said the ultimate conclusion reached by their analysis supports the hypothesis that genome duplication events during the early stages of vertebrate development are particularly important in the evolution of miRNAs.

One more minor but additionally relevant theory to explain some of the genome duplication that appears to be present among miRNA paralogs was proposed by Prochnik et al. (2007). In their work where they performed BLAST searches for known miRNA sequences from *Homo sapiens*, *C. elegans*, and *Drosophila melanogaster* in the genomes of a number of organisms, they discovered a number of miRNA genes that were clustered together. They proposed that due to the close proximity of these genes, they may have evolved from a duplication event. They further added, however, that this duplication may not be for the purposes of increasing diversity (although they acknowledge it could be), but rather for the simple purpose of increasing the level of expressed miRNA product in the cell.

The work of Hertel et al. (2006) and others who supported similar duplication models went largely unchallenged for two years until the first solid evidence against their theory emerged in a paper published in 2008 by Heimberg et al. The team sought to evaluate the previously reported evidence (Hertel et al., 2006) that a vast majority of known miRNAs in vertebrate species appeared to have arisen in osteichtyans – the clade of teleost fishes and mammals. This evidence presented by Hertel et al. and further supported by Gu et al. (2009) gave support to the theory of genome duplication breeding miRNA innovation because the osteichtyans represent the period of vertebrate history during which, most scientists agree, at least one genome duplication occurred. But Heimberg and her colleagues concluded, based on

research of the 129 chordate-specific miRNA families that these miRNA families evolved at the base of the Vertebrata, not during the osteichtyans.

Similar to the work of other groups, Heimberg et al. employed a dual analysis approach, performing both genomic searches by the BLAST method as well as Northern blot analyses of total RNA preparations. The phylogenetic relationships derived from these analyses were consistent with those produced by other groups; however, the molecular dating of the innovation of the novel miRNA families revealed a shift in the major radiation of miRNA families to an older time, specifically the Vertebrate branch, the branch immediately following the split of invertebrates from vertebrates. This initial data additionally supports the three conclusions of Sempere et al. (2006) concerning the conservation of mature miRNA sequences. Heimberg et al. had a brilliant idea, however, whose results appear to be inconsistent with the theory of genome duplication to explain miRNA innovation. Instead of considering purely the raw number of newly evolved miRNA families within a branch of the phylogenetic tree, as Hertel et al. had done, Heimberg et al. calculated the rate of miRNA acquisition measured in families per million years. When these new data were considered, the rate of miRNA acquisition in the eutherian branch was “not significantly higher” than during other times of rapid vertebrate evolution. Thus Heimberg et al. concluded that there is no clear indication of a spike in miRNA innovation that could be attributed to any large-scale genome duplication event, only a consistent, steady rate of innovation. The relatively small (approximately 3-10 fold) increase in miRNA acquisition rate for the early vertebrates indicates, according to Heimberg and colleagues, that there was one genome duplication event that occurred, but this event can only account for increased diversity among miRNA families that already existed by increasing the number of miRNA paralogs.

Instead, argues Heimberg et al., the acquisition of novel miRNA families as seen at the base of the Vertebrata (what they term "miRNA disparity") can be attributed only to de novo innovation.

Further evidence in support of the de novo innovation theory was marshaled by Wheeler et al. (2009) when the team performed a large-scale study of the miRNA complements of 18 animal species. It was the largest, most comprehensive study of miRNAs to date, with the motivation for this research given as a desire to confirm that the trends described in the model systems that had been studied previously – including *Drosophila melanogaster*, *C. elegans*, and *Mus musculus* – held true across a broader range of species. Using a combination of 454 sequencing – a form of high-throughput DNA and RNA sequencing developed by 454 Life Sciences in Branford, CT – and genomic database searches, Wheeler et al. found that the evolutionary trends developed from data gathered during the study of the model systems by many previous researchers generally holds true for all miRNA families and taxa explored. This finding further confirms the already agreed upon conclusions developed by Sempere et al. (2006) that miRNA families are continuously added to the genome, experience only rare substitutions to the mature sequence, and are rarely lost secondarily. To additionally confirm these results, they calculated a substitution rate for miRNA mutations across all 14 taxa analyzed and found it to be only 3.5%. For comparison, they also calculated the rate for mutation for 18S rDNA, a gene integral to the functioning of ribosomes in eukaryotic cells, and therefore known to be one of the least mutation-prone genes across evolutionary history. For 18S rDNA, Wheeler et al. concluded that there was a 7.3% substitution rate, therefore arguing that miRNAs evolve more than twice as slowly as one of the most conserved genes in the entire genome.

Wheeler et al.'s research, however, reveals a number of new methods by which de novo innovation of miRNAs occurs, in turn adding further support to the initial work by Heimberg et

al. (2008). The first method, as originally identified and reported by Okamura et al. (2008), is the star sequence – a rarer partner miRNA of the mature sequence. During the processing the of the pre-miRNA by Dicer, each half of the hairpin structure is processed separately, and at equilibrium, the proportions of each strand are not equal. The strand in greater number is the regular miRNA and the rarer partner strand the miRNA*, also known as the star sequence of the miRNA. These miRNA* are not perfect complements to the miRNA and thus can sometimes bind to the miRNA in regulatory complexes. Star sequences had been identified prior to the work by Okamura et al.; however, it was previously believed that they represented on average 15% or less of the total miRNA / miRNA* products and therefore were not relevant to regulatory processes. Okamura et al. showed that in many cases the proportion of star sequences could reach as high as 40% and described the importance of the star sequence in increasing the complexity of gene regulatory networks – by adding an additional level of control of miRNA sequences and hence the target mRNA sequences as well – and in controlling the evolution of miRNA and mRNA.

Here, Wheeler et al. confirms these observations and notes that the star sequences are expressed highly for many more miRNAs than originally identified. Secondly, the team identifies a new process by which the first few nucleotides at the 5' end of the mature miRNA product can be shifted so as to produce a new seed sequence (positions 2-8) of the miRNA transcript. They coined this "seed shifting" to describe this process that they conclude is "phylogenetically conserved and evolutionarily derived." The miRNA produced following a seed shift presumably targets new and different mRNA transcripts than before, although the majority of the sequence is conserved. The final new process discovered by Wheeler and colleagues was 5' end editing, occurring posttranscriptionally. Under this process, the mature miRNA product is

edited after being transcribed, typically by the insertion of a single nucleotide in between the first and second positions in the seed sequence. This gives rise to at minimum two potential mature miRNA products from a single locus (the original product and the edited version). When these three potential innovation methods are considered together, a single miRNA gene can produce multiple different transcripts, as seen in one example observed by Wheeler et al. where species from two echinoderm taxa each expressed a star sequence in conjunction with a 5' end editing process, thereby expressing four different transcripts from one miR-200 gene, each with the potential to regulate a different mRNA transcript. This evidence lends very strong support to the de novo innovation hypothesis by suggesting a number of new ways for de novo innovation to occur and thus can explain observed variances among paralogs that otherwise might be attributed to genome duplication events followed by mutations.

To this day, having considered the most current and relevant published research, it remains uncertain which of the two models of miRNA innovation fits best. There is strong evidence to support both sides of the debate as illustrated throughout this section. However, many of the researchers point to the strong potential for a dual-process evolutionary model, combining one major genome duplication event with de novo innovation to develop the remaining majority of miRNA genes. While we await further research to either confirm or deny this hypothesis, it appears that perhaps this is the most accurate representation of miRNA innovation throughout evolutionary history.

The Evolution of Morphological Complexity

While the debate over the de novo model versus duplication model for the evolution of novel miRNAs remains heated and as yet unresolved, the discussion concerning the relation of miRNA to the evolution of morphological complexity has been decidedly more agreed upon.

Nearly every paper this author has read concerning this topic, including every paper mentioned within the section above, has concluded that miRNAs represent a logical and highly likely explanation for the continued innovation of morphological complexity throughout evolutionary history. However, most researchers acknowledge that the evidence given to support this conclusion is largely circumstantial, and that, while there is a large quantity of evidence spanning a number of different explanations to support it, we cannot yet conclude definitively that miRNA innovation is responsible for increased organismal complexity.

Perhaps the greatest evidence in support of this theory comes from the discovered importance of miRNAs in key roles in animal development. Many of the miRNAs identified so far are implicated in specific temporal regulation of gene products during the early animal development cycle, "particularly in the establishment, temporal control, and maintenance of cell, tissue, and organ-specific identity" (Prochnik et al., 2007). Others are involved in the regulation of basic cellular processes and gene products (Lee et al., 2007). Additionally, the loss of some miRNA function has been implicated in pancreatic cancer development (Morris, personal communication) as well as other cancers. It is important to note that there are a large number of discovered roles for miRNAs and that these roles can be both specialized or generalized. It has been estimated that as many as a third or more of the genes in a cell may be regulated by miRNAs (Lai, 2005). Together, these discoveries suggest an incredibly important role for miRNAs in the overall regulation of cells and therefore the organism. If we are to measure organism complexity by the number of different cell types (Valentine et al., 1993), it is evident that increased gene regulation by miRNAs increases complexity by enabling new cell types and therefore tissue and organ types to develop and be regulated.

As illustrated by many of the papers described in the previous section about miRNA innovation (Hertel et al., 2006; Heimberg et al., 2008; Wheeler et al., 2009), studies of the phylogenetic distribution of miRNA innovation showed that the novel miRNAs evolved in conjunction with the development of new taxa. Each subsequent taxa possessed nearly all, if not all, of the miRNAs present in the complement of its ancestor along with at least one new miRNA (see *Figure 1*). This idea was further supported by the original conclusions proposed by Sempere et al. (2006) concerning the high level of conservation of mature miRNA sequences and continued acquisition through time. If miRNAs did not play such integral roles in the regulation and proper functioning of cells this high level of conservation would not be expected. Furthermore, the high correlation between miRNA innovation and species evolution provides strong evidence for a causative relationship.

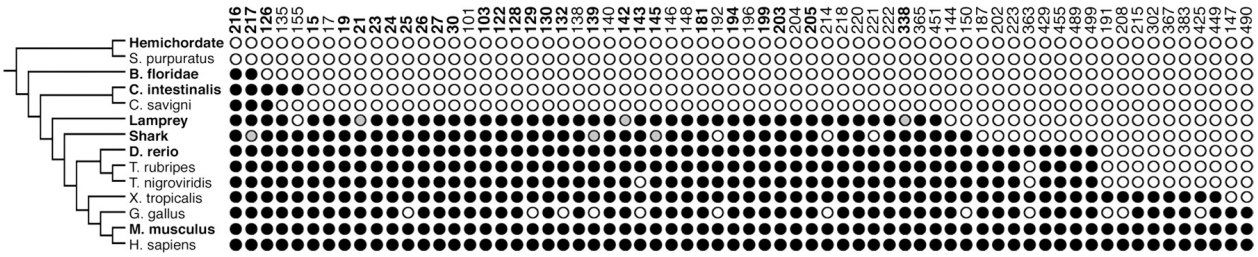

Figure 1. "Distribution of miRNAs across Deuterostomia. miRNAs discovered by genomic searches are indicated by a black dot. Those not found in the genome of the indicated taxon, but detected in a total RNA preparation, are indicated by gray circles. miRNAs not found in the genome and not detected by Northern analysis are indicated by white circles. miRNAs and taxa in bold were explored by Northern analysis." (Heimberg et al., 2008)

These two topics were deeply explored in research by Lee et al. (2007) who concluded that there is indeed a strong conservation of miRNA genes throughout evolutionary history and a strong correlation between miRNA innovation and evolutionary development. Using the example of the well-known *let-7* miRNA family, Lee et al. point out that the family has been highly conserved and implicated during crucial events in the development process for species

across the entire metazoan kingdom. This, they argue, suggests that when new miRNA families arise and become so highly conserved, they drive species evolution. This point was also argued by Prochnik et al. (2007) who suggest that *let*-7 "provides a single data point linking miRNAs to the transition from early animals to the more complex triploblastic bilaterians." By this definition, they conclude, *let-7* indicates a clear example of miRNAs increasing in regulatory complexity and therefore leading to increasing morphological complexity; and other miRNA families behave the same way. Additionally, Lee and colleagues draw upon the idea first discussed by Valentine et al. (1993) which says that structural complexity can be measured in terms of the number of differentiated cell types. Under this definition they argue that miRNAs have enabled new cell types (as evidenced by their regulation of cell state and type) and therefore are responsible for increased complexity. Although fairly widely accepted already, Lee et al. also concludes that most of the genome expansion during evolutionary history was the result of the addition of new non-coding DNA rather than protein-coding DNA. Evidence for this is provided by the "C-value enigma" (first explained by Gregory, 2005) which examines the genome size variability among species and organisms within a specific species. Given this fact, it is logical to conclude that whatever genomic elements have caused species evolution must be among non-coding DNA, and this therefore fits with the theory that it is specifically miRNAs.

While these observations have all been discussed within the context of animals, much of the same has been observed in plants where miRNAs, although slightly different in terms of exact functioning from and much less numerous than animal miRNAs, still play key roles in regulating plant development. Some of the observed increased complexities that are controlled by miRNA regulation include leaf growth, flowering and the use of roots for liquid exchange. These events are all temporally controlled just like development within animals (Hunter 2008).

Currently scientists are still unsure as to whether the miRNAs seen in algae, plants and animals derived from a common ancestral unicellular organism or rather evolved independently; however, the presence of miRNAs across all eukaryotic organisms – whether uni- or multi-cellular – further suggests that the development of novel miRNAs was particularly instrumental in the development of morphological complexity.

With all of this strong evidence in favor of miRNA induced organism evolution, there is bound to be at least some argument against this theory. While few (Gu et al., 2009) still argue that increases in the protein-coding gene repertoire could explain the observed increases in morphological complexity, some point to a different component of the gene regulatory network – the cis-regulatory genes – as a more likely causative factor in the evolution of complexity. As described by Hunter (2008), the issue with miRNAs as the chief driving factor of increased complexity is the paradox that most miRNAs are pleiotropic in nature and pleiotropic genes are inherently highly conserved, meaning that miRNAs could not undergo evolution easily without breaking their functions and therefore being selected against. This paradox could be explained, however, by methods described by Wheeler et al. (2009) such as star sequences and 5' end editing that enable the same genomic sequence to not be modified but produce multiple different products. Hunter argues that the cis-regulatory genes do not face this problem and could be a better fit for rapid evolution because there is one set of the genes for each organ or tissue type they regulate, allowing for experimentation with one set while not affecting the others. While the reasoning given by Hunter is plausible, each cis-regulatory element can only regulate the transcription of one gene, unlike miRNAs which can regulate up to 100 (Lee et al., 2007). This means that beneficial mutations to miRNAs or the de novo innovation of novel miRNA genes can potentially regulate many more gene products, and therefore much more easily effect large-

scale changes. Ultimately, more researchers currently support miRNAs as the larger player in evolutionary history, but further research is necessary to show to what extent cis-regulatory elements influence the advent of morphological complexity.

Concluding Remarks

The research that has been published within the last five years, and particularly within the last three, regarding miRNA innovation and its potential correlation to the evolutionary history of organisms and their increasing complexity has revealed an extraordinarily large set of new data and theories. The most salient, and arguably important, conclusion from the heated debate states that miRNAs are indeed responsible for causing much of morphological complexity. But what of the more contested issue, that of what causes miRNA innovation? It seems that the best conclusion that can be drawn is that a single genome duplication event did occur and is responsible for some miRNA diversity, but that throughout evolutionary history novel miRNAs (the so-called miRNA disparity) have appeared as a result of de novo innovation, not genome or even local level duplication.

While further research is needed to confirm the hypothesis of miRNAs breeding morphological complexity, despite numerous solid examples of circumstantial evidence, if the theory holds, scientists will have a new way of looking at the tree of life and may find that taxa associations previously believed to be correct may no longer hold – as was already demonstrated with the placement of the acoelomates (Wheeler et al., 2009). Furthermore, it gives cause to a reevaluation of species relationships according to their miRNA complement instead of the protein-coding portion of the genome. In short, it challenges all that science had previously embraced regarding the importance of protein-coding genes in evolutionary history, dating back nearly 40 years to the work of Ohno.

As biologists press forward with research on miRNAs, they are bound to discover new and potentially exciting functions of these tiny gene regulators. And with evidence of de novo innovation in conjunction with the connection to the evolution of complexity, this research is poised to make significant breakthroughs in understanding the past, present, and potential future of life on planet earth.

Literature Cited

Gregory, T. 2005. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann Bot (Lond)* 95, 133–146.

Gu X, Su Z, Huang Y. 2009. Simultaneous expansions of microRNAs and protein-coding genes by gene/genome duplications in early vertebrates. *J. Exp. Zool. (Mol. Dev. Evol.)* 312B.

Gu X, Wang Y, Gu J. *Nature Genet.* 31, 205-209.

Heimberg A, Sempere L, Moy V, Donoghue P, Peterson K. 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl. Acad. Sci. USA* 105, 2946-2950.

Hertel J et al. 2006. The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7:25.

Huang Y, Gu X. 2007. A bootstrap based analysis pipeline for efficient classification of phylogenetically related animal miRNAs. *BMC Genomics* 8:66.

Hunter P. 2008. The great leap forward. *Euro. Mol. Bio. Org.* 9:7, 608-611.

Lai E. 2005. miRNAs: Whys and Wherefores of miRNA-Mediated Regulation. *Curr. Bio.* 15:12, 458-460.

Lee C, Risom T, Strauss W. 2007. Evolutionary Conservation of MicroRNA Regulatory Circuits: An Examination of MicroRNA Gene Complexity and Conserved MicroRNA-Target Interactions through Metazoan Phylogeny. *DNA & Cell Bio.* 26:4, 209-218.

McLysaght A, Hokamp K, Wolfe KH. *Nature Genet.* 31, 200-204.

McShea, D. *Evolution* 50:2, 477-492.

Ohno S. 1970. *Evolution by Gene Duplication.* Heidelberg, Germany: Springer.

Okamura K, Phillips M, Tyler D, Duan H, Chou Y, and Lai E. 2008. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat. Struct. Mol. Biol.* 15, 354–363.

Prochnik S, Rokhsar D, Aboobaker A. 2007. Evidence for a microRNA expansion in the bilaterian ancestor. *Dev. Genes Evol.* 217, 73-77.

Sempere LF, Cole CN, Mcpeek MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool. (Mol Dev Evol)* 306B.

Spring J. 2002. Genome duplication strikes back. *Nature Genet.* 31, 128-29.

Tanzer A, Stadler P. 2004. Molecular Evolution of a MicroRNA Cluster. *J. Mol. Bio.* 339, 327-335.

Valentine J, Collins A, Meyer C. 1993. Morphological complexity increases in metazoans. *Paleobiology* 20, 131–142.

Wheeler B, Heimberg M, Moy V, Sperling E, Holstein T, Heber S, Peterson K. 2009. The deep evolution of metazoan microRNAs. *Evol. Dev.* 11:1, 50-68.