2-26-2008

MicroRNAs and the Advent of Vertebrate Morphological Complexity

Alysha M. Heimberg
Dartmouth College

Lorenzo F. Sempere
Dartmouth College

Vanessa N. Moy
Dartmouth College

Phillip C. J. Donoghue
University of Bristol

Kevin J. Peterson
Dartmouth College

MicroRNAs and the advent of vertebrate morphological complexity

Alysha M. Heimberg*, Lorenzo F. Sempere†, Vanessa N. Moy*, Philip C. J. Donoghue‡§, and Kevin J. Peterson*§

*Department of Biological Sciences, Dartmouth College, Hanover, NH 03755; †Department of Biochemistry, Dartmouth Medical School, Hanover, NH 03755; and ‡Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, United Kingdom

Communicated by Victor R. Ambros, Dartmouth Medical School, Hanover, NH, December 29, 2007 (received for review November 7, 2007)

The causal basis of vertebrate complexity has been sought in genome duplication events (GDEs) that occurred during the emergence of vertebrates, but evidence beyond coincidence is wanting. MicroRNAs (miRNAs) have recently been identified as a viable causal factor in increasing organismal complexity through the action of these ≈22-nt noncoding RNAs in regulating gene expression. Because miRNAs are continuously being added to animalian genomes, and, once integrated into a gene regulatory network, are strongly conserved in primary sequence and rarely secondarily lost, their evolutionary history can be accurately reconstructed. Here, using a combination of Northern analyses and genomic searches, we show that 41 miRNA families evolved at the base of Vertebrata, as they are found and/or detected in lamprey, but not in either ascidians or amphioxus (or any other nonchordate taxon). When placed into temporal context, the rate of miRNA acquisition and the extent of phenotypic evolution are anomalously high early in vertebrate history, far outstripping any other episode in chordate evolution. The genomic position of miRNA paralogues in humans, together with gene trees incorporating lamprey orthologues, indicates that although GDEs can account for an increase in the diversity of miRNA family members, which occurred before the last common ancestor of all living vertebrates, GDEs cannot account for the origin of these novel families themselves. We hypothesize that lying behind the origin of vertebrate complexity is the dramatic expansion of the noncoding RNA inventory including miRNAs, rather than an increase in the protein-encoding inventory caused by GDEs.

genome duplication | lamprey | macroevolution | shark | chordate

Vertebrates are widely perceived to be more complex than their spineless relatives, the ascidian urochordates and the cephalochordate amphioxus (1, 2). Nonetheless, demonstrating this difference in morphological complexity is difficult, and determining its causal basis has proven even less tractable. Typically, causality has been sought in the phenomenon of genome duplication (2–5), thought to have occurred twice during the emergence of vertebrates (6), once before and after the divergence of the lamprey and gnathostome lineages (4, 7). However, the absence of any obvious increase in morphological complexity associated with other known genome duplication events (GDEs) (8), especially within the actinopterygian fishes (9), suggests that the causal link between morphological complexity and GDEs is tenuous at best (10). Further, given the similarity of the developmental tool kit across Metazoa despite the unambiguous differences in organismal complexity between, for example, vertebrates and cnidarians (11), a consequential increase in the protein-coding repertoire cannot provide sufficient explanation for differences in morphological complexity.

An alternative explanation for increasing morphological complexity has been increasing the complexity of gene regulatory networks (12). Although usually considered from the perspective of protein-coding genes (13), vertebrates are also distinguished from invertebrates by the transcribed, noncoding complements of their genome, with mammalian genomes transcribing over an order of magnitude more noncoding RNA as compared with either worm or fly (14). Importantly, it is among this noncoding sequence that a variety of new classes of regulatory factors has been discovered, including microRNAs (miRNAs), which has been postulated as developmental and evolutionary determinants of organismal complexity (15, 16). Indeed, vertebrates possess many more miRNAs than any invertebrate sampled to date (17), and >50 new miRNA families are thought to have evolved in the vertebrate lineage sometime after its split from the invertebrate chordates and before the divergence of osteichthyan fishes (17–19). Nonetheless, how this increase in the miRNA repertoire correlates to the emergence of vertebrate complexity is currently unclear because groups such as lampreys and sharks, from which we may infer the miRNA complement of early vertebrates, have yet to be sampled.

miRNAs are unusual with respect to all other known genetic elements (17, 20, 21) in that they are continuously being added to metazoan genomes, and once integrated into a gene regulatory network, the primary sequence of the mature ≈22-nt sequence comes under intense negative selection, with mutations occurring only very rarely. In addition, the new miRNA is only rarely secondarily lost. These three properties (continuous addition, conservation of primary sequences, and rarity of secondary loss) allow for the accurate reconstruction of the miRNA complement of any last common ancestor, including the last common ancestor of all living vertebrates.

Here, we show by using a combination of genomic searches and Northern analyses that the vast majority of miRNAs found previously in osteichthyans [i.e., those shared between teleost fishes and mammals (17, 19)], actually evolved at the base of the Vertebrata, before the divergence between the living jawless (lamprey) and jawed fishes, but after the divergence of vertebrates from their invertebrate chordate relatives. Because the origin of these novel miRNA families cannot be ascribed to the GDEs associated with early vertebrate history, we argue that lying behind the origins of vertebrate complexity might be the evolution of novel miRNA families.

Results and Discussion

The Emergence of Vertebrates Is Characterized By an Unprecedented Increase In the Rate of miRNA Family Innovation. Because miRNAs, once fixed, rarely change the primary sequence of the mature region and are rarely eliminated from the genome, it is possible to determine their phylogenetic origin through analysis and detection in living representatives (17, 22). This approach obviates the need for libraries from every species, as the conserved set of miRNAs between two species can be deduced if the complement of one (e.g., mouse) has been determined from libraries and the second (e.g., lamprey) queried for these miRNAs by other means like Northern analysis. We stress that the miRNAs that have evolved within the lamprey lineage will be

Author contributions: K.J.P. designed research; A.M.H., L.F.S., V.N.M., and P.C.J.D. performed research; A.M.H. and K.J.P. analyzed data; and A.M.H., P.C.J.D., and K.J.P. wrote the paper.

The authors declare no conflict of interest.

§To whom correspondence may be addressed. E-mail: phil.donoghue@bristol.ac.uk or kevin.j.peterson@dartmouth.edu.

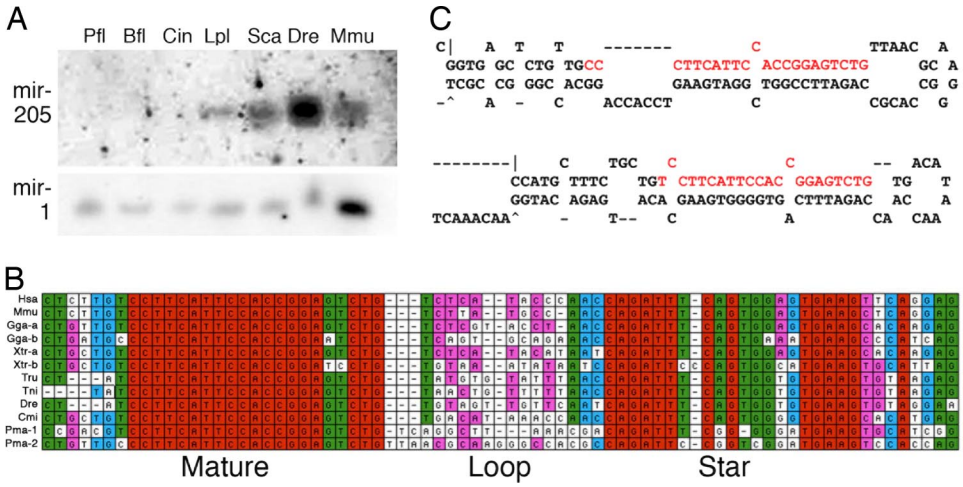

Fig. 1. miRNA discovery in lamprey and shark using mir-205 as an example. (*A*) Northern analysis. miR-205 is clearly detected in all vertebrates examined [brook lamprey, *Lampetra planeri* (Lpl); cat shark, *Scyliorhinus canicula* (Sca); zebrafish, *Danio rerio* (Dre); mouse, *Mus musculus* (Mmu)], but not in any invertebrate including the hemichordate *Ptychodera flava* (Pfl), the cephalochordate *Branchiostoma floridae* (Bfl), and the ascidian *Ciona intestinalis* (Cin). However, more primitive miRNAs, including miR-1, are clearly detected in all samples. (*B*) Alignment of the stem-loop sequence for mir-205 from nine vertebrates. Two copies of mir-205 were found in the sea lamprey *Petromyzon marinus* (Pma), and a single copy was found in the genomic traces of the elephant shark *Callorhinchus milii* (Cmi). All three regions of a miRNA gene (the mature sequences, the star sequences, and the loop region) are clearly discernable from the alignment. (*C*) Predicted secondary structure of the *mir-205* orthologue from Pma-2 (*Upper*) and Cml (*Lower*) as determined by Mfold (42). The initial ΔG values for formation are −44.1 and −36.4 kcal/mol, respectively. Other abbreviations: Hsa, *Homo sapiens*; Gga, *Gallus gallus*; Xtr, *Xenopus tropicalis*; Tru, *Tetrafugu rubipes*; Tni, *Tetraodon nigroviridis*.

unique to this lineage, and not relevant to our question, which is the original complement of miRNAs in latest common ancestors of living clades, including the last common ancestor of all living vertebrates. Thus, we set out to ascertain the evolutionary history of the 129 chordate-specific miRNA families throughout the phylum Vertebrata by using both genomic queries and Northern analysis (Fig. 1). The data are presented in Fig. 2 and summarized in Fig. 3 and Table 1. We found that the phylum Chordata is characterized by the evolution of two miRNA families, *mir-216* and *mir-217*, and the newly recognized clade Olfactores (23) is characterized by three miRNA families, *mir-126*, *mir-135*, and *mir-155* (see also ref. 24), with only *mir-135* not found in the genome of one of the two ascidians queried (and also the only invertebrate not investigated by Northern analysis). An additional two miRNA families were found in the elephant shark (*Callorhinchus milii*) genome (25), but not in lamprey. Another eight miRNA families were found to be exclusive to the osteichthyans, eight families were restricted to Tetrapoda (frog + bird + mammal), and two families were restricted to Amniota (bird + mammal) (Fig. 2). Sixty-three miRNA families were restricted to the two eutherian mammal taxa considered, mouse and human (data not shown). And finally, of the 56 chordate-specific miRNA families shared between teleosts and mammals, more than two-thirds (41 families) were either found in the genome of the sea lamprey (*Petromyzon marinus*) and/or detected by Northern analysis in total RNA preparations from the brook lamprey (*Lampetra planeri*).

As expected (17, 22) there were few examples of secondary loss; where detectable, either by miRNA or taxon investigated (Fig. 2, bold), most of the expected miRNAs not found in genomic traces are detectable by Northern analysis (Fig. 2, gray circles). Only taxa not yet investigated by experimental means (e.g., chicken) showed an unusually high number of missing miRNAs, keeping in mind that in this case only 12 of 66 total miRNA families could not be found in the genome. Indeed, the phylogenetic relationships of these taxa, as deduced by molecular sequence analyses (23), are entirely in accord with the hierarchical structure of the miRNA data obtained herein, which is only possible if miRNAs, once evolved, are not secondarily lost and conserve the mature miRNA primary sequence, as argued by Sempere *et al.* (17).

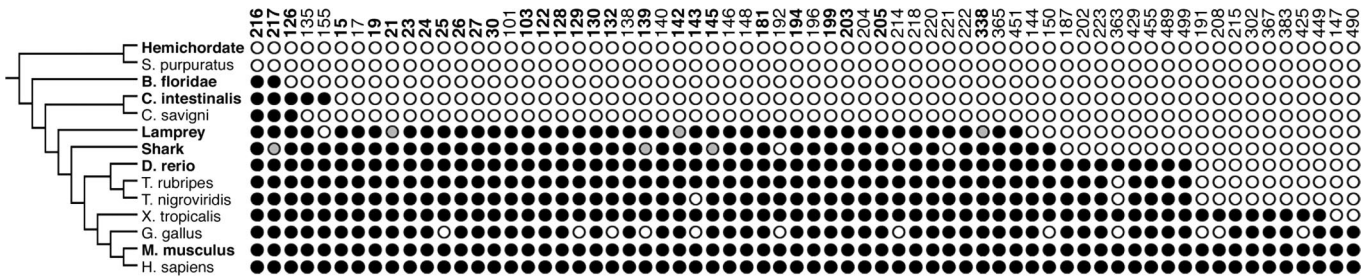

Fig. 2. Distribution of miRNAs across Deuterostomia. miRNAs discovered by genomic searches (and in many cases confirmed by Northern analyses; see Table 1) are indicated by a black dot. Those not found in the genome of the indicated taxon, but detected in a total RNA preparation, are indicated by gray circles. miRNAs not found in the genome and not detected by Northern analysis are indicated by white circles. As expected (17), miRNAs, once evolved, are rarely secondarily lost and the mature sequence rarely changes in primary sequence, allowing for an accurate reconstruction of the miRNA complement of the last common ancestor. The few instances of potential secondary loss often involve miRNAs that were not detectable by Northern analysis (plain text, e.g., mir-135) and/or in taxa not yet explored by experimental means (plain text, e.g., the chicken *Gallus gallus*). miRNAs and taxa in bold were explored by Northern analysis. miRNA families are given by the lowest numbered member; for the full complement of miRNA members for each family see Table 1.

EVOLUTION

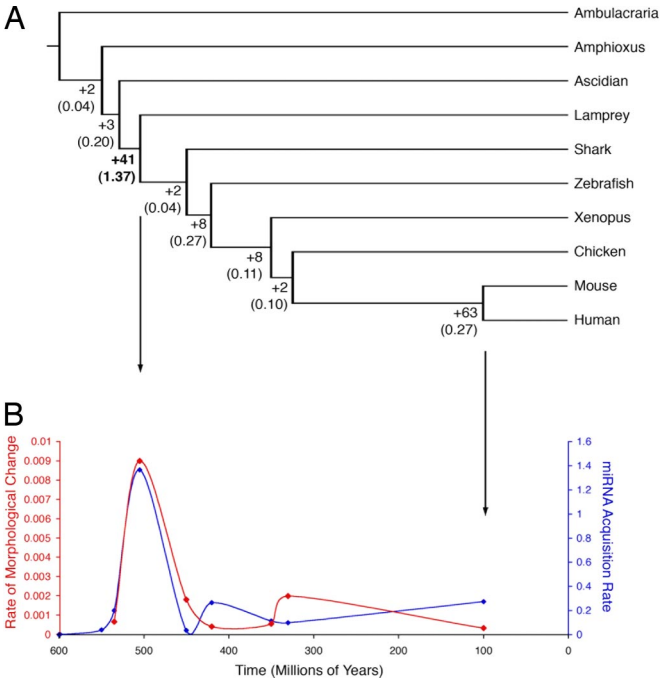

Fig. 3. Evolutionary history of the 129 chordate-specific families of miRNAs found in eutherian mammals. (*A*) Cladogram derived from the history of miRNA family acquisition, with the number of new families (Table 1) indicated at the node and the rate of acquisition (number of new families per million years) shown parenthetically. Divergence times taken from estimates were derived from a molecular clock analysis (26) and the fossil record (44). (*B*) miRNA family acquisition rate (blue) plotted with rate of morphological change (2) (red) against absolute time. The spike for both miRNA acquisition and MCI are both outliers as compared with any other time in vertebrate history, as determined by a Dixon's *D* test. Points along the curves are tied to the nodes in *A* (two of which are indicated by arrows).

Table 1. Evolutionary acquisition of miRNA families

Taxon	No. of families	miRNA family
Eumetazoa	1	(10^{2d}, 99^{2d}, 100)
Nephrozoa	29	(let7$^{11d/s}$, 98), (1^{2d}, 206), 7^{3d}, (8, 141, 200$^{3d/s}$), 9^{3d}, 22, (29$^{4d/s}$, 285), 31, 33^{2d}, 34$^{3d/s}$, 71^{lost}, 79^{lost}, 92^{3d}, (96, 182, 183, 263^{2d}), 124$^{3d/s}$, 125^{3d}, 133^{3d}, 137, 153^{2d}, 184, 190^{2d}, 193^{2d}, 210, 219^{2d}, 252^{lost}, 278^{lost}, 281^{lost}, 315^{lost}, 375
Chordata	2	*216*^{2s}, **217**
Olfactores	3	**126**, 135^{3d}, *155*
Vertebrata	41	(15^{2d}, 16^{2d}, *195*), (17, 18^{2d}, 20^{2d}, *93*, *106*^{2d}), **19**$^{3d/s}$, **21**, **23**^{2d}, **24**^{2d}, **25**, **26**^{3d}, **27**^{2d}, **30**$^{6d/s}$, 101^{2d}, (**103**^{2d}, **107**), ***122***, **128**^{2d}, ***129***^{2d}, (*130*^{2d}, **301**^{2d}), (**132**, *212*), 138^{2d}, 139, 140, ***142***, **143**, **145**, 146^{2d}, (148^{2d}, *152*), **181**$^{6d/s}$, 192, ***194***^{2d}, *196*^{3d}, **199**^{3d}, **203**, (204, *211*), **205**, 214, *218*^{2d}, *220*$^{3d/s}$, *221*, *222*, ***338***, 365^{2d}, *451*
Gnathostomata	2	144, 150
Osteichthyes	8	187, 202, 223, 363, *429*, *455*, 489, 499
Tetrapoda	8	(191, 637), 208^{2s}, 215, 302^{4s}, 367, 383, 425, 449^{2s}
Amniota	2	147^{2d}, 490
Mammalia	63	(28, 151, 708), 127, (134, 412), 136, 149, (154, 323, 329^{2s}, 369, 377, 381, 382, 410, 453, 485, 487^{2s}, 494, 495, 496, 539, 655, 656), 185, 186, (188, 362, 500, 501, 502, 532, 660), 224, 296, 297, (299, 579), 320, (324, 544), (325, 493), 326, (328, 483), (330, 560), 331, 335, 339, 340, (342, 610), (345, 378), 346, 361, 376^{4s}, 370, (374^{2s}, 542), (379, 380, 411, 758), 384, 409, (422, 423), 431, 433, 448, 450^{3s}, 484, (486, 612), 488, 491, (497, 600), 503, 505, (506, 507, 508, 509^{3s}, 510, 513^{2s}, 514^{3s}, 652), (511^{2s}, 802), 551^{2d}, (568, 620), 592, 615, 668, 671, 675, 770, 801, 871, 872, 873, 874, 875, 876, 877

Bold type indicates detected by both Northern analysis and genomic searches in this study at the taxonomic level indicated. Bold italics indicates detected by Northern analysis in this study, not found by genomic searches, at the taxonomic level indicated. Underline indicates not detected by Northern analysis in this study, but found by genomic searches, at the taxonomic level indicated. Italics indicates not assayed by Northern analysis in this study, but found by genomic searches, at the taxonomic level indicated. Superscript indicates the number of known paralogues on different (d) or same (s) chromosome(s). Parentheses indicate microRNA family members (18) with a few minor modifications. Lost indicates the family is secondarily lost in the eutherian lineage. Note that *mir-278* is found in the genome of the hemichordate *Saccoglossus kowalevskii*; *mir-252*, and *mir-315* are found in the genomes of both the hemichordate and the lamprey; and *mir-71* is found in the genomes of sea urchin and amphioxus, but not ascidian or vertebrate; *mir-281* is found in nonvertebrate deuterostomes. Plain text indicates phylogenetic position determined by Sempere *et al.* (17), Prochnik *et al.* (24), and Huang and Gu (18) in conjunction with miRBase.

These data indicate that there were two periods in vertebrate evolutionary history when a seemingly inordinate number of miRNAs were acquired, once at the base of vertebrates, and once along the stem-lineage leading to eutherian mammals, specifically the branch intermediate of the last common ancestors of Amniota and Archontoglires (see also ref. 19). However, when the rate of miRNA acquisition was considered (the number of miRNA families acquired per million years of evolutionary history), as opposed to the raw number of newly evolved miRNA families, the eutherian rate of miRNA acquisition is not significantly higher than other episodes in vertebrate evolutionary history (Fig. 3). To ask about the rate of miRNA acquistion in early vertebrate history, the split between urochordates and vertebrates needs to be estimated. Peterson *et al.* (26) dated the origin of crown-group Chordata to ≈550 Ma, setting the maximum for this split, and the fossil record suggests that crown-group Olfactores had arisen by the end of the Early Cambrian ≈520 Ma (27); the age of crown-group Vertebrata, as estimated from the fossil record, is ≈505 Ma (28, 29). Thus, the rate of acquisition varies from 0.91 to 2.73, depending on when the speciation event occurred, and is 3–10 times higher than the rate at which they were acquired within the eutherian stem lineage (Fig. 3*B*).

The Rate of miRNA Family Acquisition Correlates to the Increase in Vertebrate Morphological Complexity. Using a midpoint estimate for the origin of Olfactores (i.e., 535 Ma), we then asked how the rate of miRNA acquisition along the vertebrate stem compares with changes in morphological complexity, as calculated by Aburomia *et al.* (2). Our data suggest that the vertebrate stem-lineage is similarly anomalous with respect to both miRNA acquisition rate and both the relative (Fig. 3*B*) and absolute (data not shown) amount of morphological change when compared with any other point in chordate evolution, as assessed by a Dixon's *D* test ($P < 0.01$; ref. 30) for both miRNAs and the morphological complexity index (MCI) (Fig. 3*B*).

miRNA Family Innovation Is Not the Result of Genome Duplication. Hertel *et al.* (19) have cogently argued that nonlocal duplication of miRNAs, resulting in paralogues located on different chromosomes, occurs exclusively in association with whole GDE. Supporting their argument, if the evolutionary history of all 153 miRNA families conserved in eutherian mammals is traced

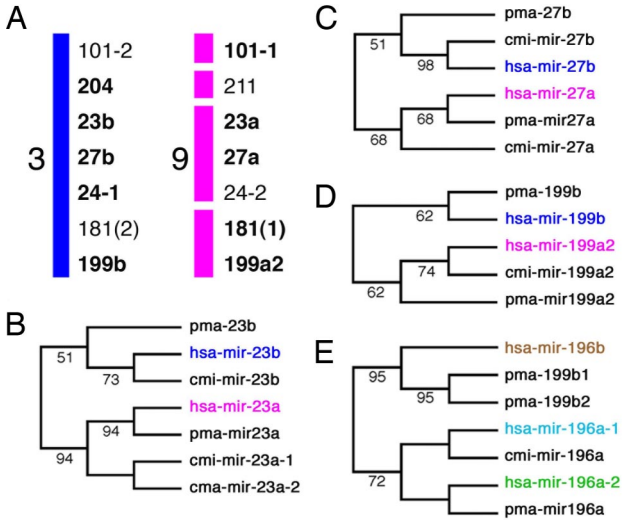

Fig. 4. Fixation of vertebrate-specific miRNA families preceded the GDEs, which preceded the divergence between lamprey and gnathostomes. (*A*) Two of the reconstructed paleochromosomes of the ancestral osteichthyan, as determined by Kohn *et al.* (32), with their paralogous miRNA sets. miRNAs indicated in bold were found in the lamprey genome and phylogentically group with the indicated miRNA (*B–D*). (*B–D*) Midpoint-rooted distance trees (see *Materials and Methods*) of the indicated miRNA with the human, shark, and lamprey paralogues. Bootstrap values (1,000 replications) are indicated at the nodes. Note that lamprey has both paralogues of each of these three families. (*E*) Midpoint rooted distance phylogram of *mir-196*. Two copies were found in the lamprey genome, and both group with *mir-196b*, which is located in the HoxA cluster and reconstructed as part of paleochromosome 7 (brown; ref. 23), suggesting a tandem duplication of the miRNA, if not the entire cluster, in the lamprey lineage. No paralogues were found that group with the HoxB-associated (light blue) or HoxC-associated (light green) *mir-196s*. Both of these results are consistent with what is known about the evolution of the Hox clusters themselves in the lamprey lineage (34).

through phylogeny (including the families that evolved before the last common ancestor of all living chordates), 70 are present in the last common ancestor of lampreys and mammals, and of these 70, 47 miRNAs (67.1%) have at least two paralogues on separate chromosomes in both human and mouse (Table 1). In contrast, of the 83 eutherian families that evolved after the split from lampreys, only two miRNAs (2.4%) have at least two paralogues on separate chromosomes.

Hertel *et al.* (19) further argued that a significant complement of the vertebrate miRNA families can be explained by the whole GDEs. Thus, GDEs have been hypothesized to explain both the dramatic increases in morphological complexity (2–5) and anomalously high rates of miRNA family innovation (19, 31). However, because lampreys often have one, if not two, copies of each of these miRNAs (Fig. 4 *A–D*), and many of the paralogy groups that we detected in lampreys can be located to different chromosomes on the reconstructed protokaryotype of osteichthyans (32) (Fig. 4*A*), the end product of the GDEs, the establishment of these paralogy groups must have occurred before the divergence of lampreys and gnathostomes. Further, because we detect paralogues of these families in the lamprey, the family itself must have evolved before any GDE. Hence, although genome duplications events can dramatically increase the diversity within existing miRNA families (see above), they cannot account for the origin of these novel families in the first instance (what we term here miRNA disparity).

The results of our phylogenetic analyses require that the dramatic increase in miRNA disparity preceded any GDE, which itself preceded the speciation event between lamprey and human. To compare our results to those that have used protein-encoding genes, we searched the sea lamprey genome for *mir-10* and *mir-196*, two miRNAs that are embedded within the Hox clusters: one copy of *mir-10* is found in each of the HoxB and HoxD clusters, whereas one copy of *mir-196* is found in each of the HoxA, HoxB, and HoxC clusters (33). We were unable to find any copies of *mir-10* in the genome of the sea lamprey, but we did find two paralogues of *mir-196*. Both of these genes group together phylogenetically (Fig. 4*E*) and are collectively the sister genes of *mir-196b*, the miRNA embedded in the HoxA cluster. Hence, we find evidence for a HoxA cluster or clusters (which contains *mir-196b* but not *mir-10*), but the potential absence of the HoxB, HoxC, and HoxD clusters, consistent with the data obtained from direct examination of the Hox genes themselves (34), attesting, if nothing more, to the accuracy of our phylogenetic analyses. Therefore, when considered in the context of genome evolution, it seems clear that at least one round of genome duplication occurred after the divergence between urochordates and vertebrates and before the divergence between lamprey and human, and that this GDE dramatically increased the diversity, but not the disparity, of miRNAs by increasing the number of miRNA paralogues.

Emergence of miRNA Disparity and Vertebrate Morphological Complexity: Coincidence or Causality? At an acquisition rate between 0.91 and 2.73 families per million years, miRNA families were established at a faster rate during the emergence of vertebrates than during any other episode in vertebrate history, far outstripping previous reports of anomalously high acquisition rates in stem-eutherians (19). Of course, it could be argued that the correlation between miRNA acquisition and morphological complexity is exactly that, a simple correlation. However, we suggest otherwise given that many of these 41 miRNA families are expressed in vertebrate-specific cell types or tissues, as determined in the zebrafish (35), such as the pronephros (miR-30), the liver and pancreas (miR-122), the thymus (miR-142), and melanocytes (miR-204), or in organs that are clearly more complex than their homologues in other chordates, for example, the brain (e.g., miR-128a, miR-129, miR-132, miR-218) and the pharynx (e.g., miR-23a, miR-27a, miR-140, miR-214). Given the role miRNAs play in the specification of cell and tissue types (36, 37), we suggest that the origin of these cellular novelties was predicated on the origin and fixation of these novel miRNAs (17). Our analyses indicate that the better part of osteichthyan miRNA diversity is generic to vertebrates as a whole, entirely coincident with the emergence of vertebrate complexity, and implicated in the development of vertebrate innovations and elaborations. Of course, the next step is to demonstrate functional causality between the novel miRNA families and vertebrate phenotype.

Insight Into the Mechanistic Basis of miRNA Innovation. Such rapid rates of miRNA innovation, and the correlation with the dramatic increase in morphological complexity, beg the question of how and why this burst of miRNA family innovation occurred. The basis of miRNA innovation in animals is controversial as there are a number of competing, but weakly supported, hypotheses, ranging from local to large-scale duplication and divergence, to serendipitous emergence from random hairpins in the genome (19, 31). Although much of the diversity of vertebrate-specific miRNAs can be explained by the GDEs, it cannot explain the establishment of most of the 41 vertebrate-specific miRNA families themselves, which must have occurred still earlier in the lineage leading to vertebrates after its separation from the tunicate lineage. Given that miRNA families preserve no evolutionary footprint of innovation through duplication and divergence, such as in their coding sequence or their syntenic arrangement within and between chromosomes, it may be inferred that their innovation occurred via a different mechanism. This notion suggests that, although whole genome duplication can increase the diversity of miRNAs within a

family, it does not lead to fundamental innovation or disparity, such as in the establishment of new miRNA families. In fact, where there is an unusually high rate of miRNA family acquisition, for example at the base of Nephrozoa (Table 1), there is no evidence for a GDE (11), and where there is a GDE, for example at the base of the teleost lineage (38), there is no evidence, despite extensive library searches (39), for an increase in the number of teleost-specific miRNA families. Indeed, in contrast to the rhetoric, no good evidence has been marshaled in support of the much-vaunted hypothesis that GDEs can confer increasing organismal complexity (10). Instead, we suggest that changes in the global transcriptional status of the vertebrate genome (14, 15, 40), which may have lead to the creation of more hairpins, and hence potentially more miRNAs, led to the dramatic increase in organismal complexity in this one metazoan lineage.

Materials and Methods

Northern analysis and genomic queries were done as described (17). Total RNA was extracted from the hemichordate *Ptychodera flava*, the cephalochordate *Branchiostoma floridae* (amphioxus), the urochordate *Ciona intestinalis*, and the vertebrates *Lampetra planeri* (brook lamprey), *Scyliorhinus canicula* (cat shark), *Danio rerio* (zebrafish), and *Mus musculus* (mouse). Genomic queries of taxa not represented in miRBase (release 10.0, Sanger Institute, Cambridge, United Kingdom; ref. 41) used the full stem-loop (pre-miRNA) sequence of the human miRNA to Blast against genomic traces and, if possible, the unassembled genome. Blast parameters used the default settings for blastn. Genomes searched included the sea urchin (*Strongylocentrotus purpuratus*), hemichordate (*Saccoglossus kowalevskii*), amphioxus (*Branchiostoma floridiae*), ascidians (*Ciona intestinalis* and *Ciona savigni*), sea lamprey (*Petromyzon marinus*) (all deposited at the National Center for Biotechnology Information, version 2.2.14, April 2007), and the elephant shark (*Callorhinchus milii*) (25), which was available at http://blast.fugu-sg.org. For the lamprey miRNAs we also queried against the unassembled sea lamprey genome available by Pre-Ensembl (version 43, November 2006, Sanger Institute and European Bioinformatics Institute, Cambridge, United Kingdom). On occasion, we also searched the genomes of the teleost fishes *Takifugu rubripes* and *Tetraodon nigroviridis*, and the chicken *Gallus gallus*, for miRNAs not deposited at miRBase. Putative orthologues were determined by selecting all subject sequences that showed at least 75% similarity in the mature sequence (with 100% similarity across the seed region, nucleotides 2–7) with the query pre-miRNA sequence. To confirm orthology we first aligned the sequence with known pre-miRNAs, and then folded these putative orthologous sequences by using the web-based program Mfold (42) with standard minimum free energy values (43) to confirm a stable secondary structure. Alignments of the stem-loop sequences and the phylogenetic analyses used MacVector (version 7.2.3–2004; Accelrys). Distance analyses used the neighbor-joining algorithm with the Tamua-Nei correction; bootstrap values were derived from 1,000 replications. Family assignment of particular miRNAs followed Huang and Gu (18) with a few minor modifications after miRBase. To calculate the rate of miRNA, acquisition was calculated by dividing the number of miRNAs acquired at each node by the time elapsed between nodes in millions of years, for example, 41 miRNAs/15 million years = 2.73 miRNAs/million years. The following divergence times were used: Chordata, 550 Ma; Olfactores, 550–520 Ma; Vertebrata, 505 Ma; Gnathostomata, 450 Ma; Osteichthyes, 420 Ma; Tetrapoda, 350 Ma; Amniota, 330 Ma; Eutheria, 100 Ma. The 550-Ma divergence estimate is based on a detailed molecular clock analysis (26); the rest were taken directly from the fossil record (27, 44).

ACKNOWLEDGMENTS. We thank V. Ambros and C. Cole (Dartmouth College) for the gift of some of the StarFire probes and other Northern reagents used in this study; C. Cole for hosting L.F.S. in his laboratory; V. Ambros for discussion; E. Bryleva (Dartmouth College) for the gift of murine tissue; J. Garcia (University of Barcelona, Barcelona) for amphioxus material; M. McPeek for running the statistical analyses; and anonymous referees for helpful suggestions and criticisms. L.F.S. is a postdoctoral fellow supported by Susan G. Komen Breast Cancer Foundation Grant PDF0503563. P.C.J.D. was funded by the National Endowment for Science, Technology and the Arts and The Leverhulme Trust. K.J.P. is supported by the National Science Foundation.

1. Valentine JW, Collins AG, Meyer CP (1994) Morphological complexity increase in metazoans. *Paleobiology* 20:131–142.
2. Aburomia R, Khaner O, Sidow A (2003) Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. *J Struct Funct Genomics* 3:45–52.
3. Ohno S (1970) *Evolution by Gene Duplication* (Springer, Heidelberg).
4. Holland PWH, Garcia-Fernàndez J, Williams, Nic A, Sidow A (1994) in *The Evolution of Developmental Regulatory Mechanisms*, eds Akam M, Holland P, Ingham P, Wray G (The Company of Biologists Limited, Cambridge, UK), pp 125–133.
5. Sidow A (1996) Genome duplications in the evolution of early vertebrates. *Curr Opin Genet Dev* 6:715–722.
6. Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *Plos Biol* 3:e314.
7. Stadler PF, *et al.* (2004) Evidence for independent *Hox* gene duplications in the hagfish lineage: A PCR-based gene inventory of *Eptatretus stoutii*. *Mol Phylogenet Evol* 32:686–694.
8. Spring J (2002) Genome duplications strikes back. *Nat Genet* 31:128–129.
9. Stellwag EJ (2004) Are genome evolution, organism complexity, and species diversity linked? *Integrat Comp Biol* 44:358–365.
10. Donoghue PCJ, Purnell MA (2005) Genome duplication, extinction, and vertebrate evolution. *Trends Ecol Evol* 20:312–319.
11. Putnam NH, *et al.* (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317:86–94.
12. Valentine JW (2000) Two genomic paths to the evolution of complexity in bodyplans. *Paleobiology* 26:513–519.
13. Vogel C, Chothia C (2006) Protein family expansions and biological complexity. *PLoS Comput Biol* 2:e48.
14. Frith MC, Pheasant M, Mattick JS (2005) The amazing complexity of the human transcriptome. *Eur J Hum Genet* 13:894–897.
15. Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA, eukaryotic complexity. *BioEssays* 29:288–299.
16. Lee C-T, Risom T, Strauss WM (2007) Evolutionary conservation of microRNA regulatory circuits: An examination of microRNA gene complexity and conserved microRNA-target interactions *DNA Cell Biol* 26:209–218.
17. Sempere LF, Cole CN, McPeek MA, Peterson KJ (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zool* 306:575–588.
18. Huang Y, Gu X (2007) A bootstrap based analysis pipeline for efficient classification of phylogenetically related animal miRNAs. *BMC Genomics* 8:66.
19. Hertel J, *et al.* (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7:25.
20. Ruby JG, *et al.* (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* 17:1850–1864.
21. Stark A, *et al.* (2007) Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res* 17:1865–1879.
22. Sempere LF, Martinez P, Cole C, Baguñà J, Peterson KJ (2007) Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes. *Evol Dev* 9:409–415.
23. Delsuc F, Brinkman H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965–968.
24. Prochnik SE, Rokhsar D, Aboobaker AA (2007) Evidence for a microRNA expansion in the bilaterian ancestor. *Dev Genes Evol* 217:73–77.
25. Venkatesh B, *et al.* (2007) Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome. *PLoS Biol* 5:e101.
26. Peterson KJ, Cotton JA, Gehling JG, Pisani D (2007) The Ediacaran emergence of bilaterians: Congruence between the genetic and geologic fossil records. *Philos Trans R Soc London Ser B*, in press.
27. Chen J-Y, Huang D-Y, Peng Q-Q, Chi H-M, Feng M (2003) The first tunicate from the Early Cambrian of South China. *Proc Natl Acad Sci USA* 100:8314–8318.
28. Donoghue PCJ, Forey PL, Aldridge RJ (2000) Conodont affinity and chordate phylogeny. *Biol Rev* 75:191–251.
29. Müller KJ, Hinz-Schallreuter I (1998) Internal structure of Cambrian conodonts. *J Paleontol* 72:91–112.
30. Sokal RR, Rohlf FJ (1994) *Biometry* (Freeman, New York).
31. Chen K, Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8:93–103.
32. Kohn M, *et al.* (2006) Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends Genet* 22:203–210.
33. Tanzer A, Amemiya CT, Kim C-B, Stadler PF (2005) Evolution of microRNAs located within *Hox* gene clusters. *J Exp Zool* 304:1–10.
34. Fried C, Prohaska SJ, Stadler PF (2003) Independent Hox-cluster duplications in lamprey. *J Exp Zool* 299:18–25.
35. Wienholds E, *et al.* (2005) MicroRNA expression in zebrafish embryonic development. *Science* 309:310–311.
36. Plasterk RHA (2006) MicroRNAs in animal development. *Cell* 124:877–881.
37. Wienholds E, Plasterk RHA (2005) MicroRNA function in animal development. *FEBS Lett* 579:5911–5922.
38. Meyer A, Van de Peer Y (2005) From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *BioEssays* 27:937–945.
39. Kloosterman WP, *et al.* (2006) Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res* 34:2558–2569.
40. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer, Sunderland, MA).
41. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: MicroRNA sequences, targets, and gene nomenclature. *Nucleic Acids Res* 34:D140–D144.
42. Zuker M (1999) Mfold web server for nucleic acid folding and hybridization predictions. *Nucleic Acids Res* 31:3406–3415.
43. Ambros V, *et al.* (2003) A uniform system for microRNA annotation. *RNA* 9:277–279.
44. Benton MJ, Donoghue PCJ (2007) Paleontological evidence to date the tree of life. *Mol Biol Evol* 24:26–53.