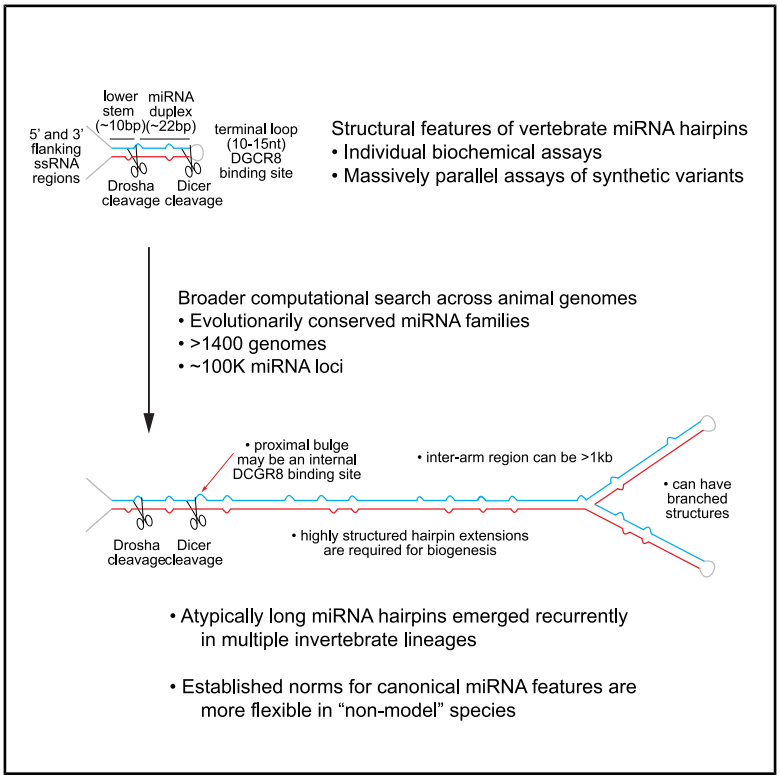
Cell Reports

Article

Repeated emergence of giant microRNA hairpins across invertebrates

Graphical abstract



Highlights

- Canonical miRNA hairpins generally have fixed stem lengths that enable their processing
- Large-scale searches for conserved miRNA loci reveal atypical, highly extended hairpins
- miRNA hairpin extension has occurred recurrently among invertebrate species
- Long miRNA hairpins require extended base-pairing and an internal DGCR8 binding site

Authors

Mir-Mammad Javad-zada, Xiaodong Zou, Katsutomo Okamura, ..., Leemor Joshua-Tor, Bastian Fromm, Eric C. Lai

Correspondence

laie@mskcc.org

In brief

Javad-Zada et al. analyze ~1,500 animal species for conserved miRNA hairpins, and find ones with atypically long stems. These occur recurrently within invertebrates and can be >1 kilobase. Structure-function analyses demonstrate that extensive stem base-pairing and a proximal bulge that mimics an internal DGCR8 binding site are needed for biogenesis.

Javad-zada et al., 2025, Cell Reports *44*, 116243
September 23, 2025
https://doi.org/10.1016/j.celrep.2025.116243

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Repeated emergence of giant microRNA hairpins across invertebrates

Mir-Mammad Javad-zada,[1,2,7] Xiaodong Zou,[1,7] Katsutomo Okamura,[3] Renfu Shang,[1] Ankur Garg,[4,5] Ethan Lee,[1] Leemor Joshua-Tor,[4,5] Bastian Fromm,[6] and Eric C. Lai[1,2,8,9,*]

[1]Developmental Biology Program, Sloan Kettering Institute, New York, NY 10065, USA
[2]Computational Biology Program, Weill Cornell Medicine, New York, NY 10065, USA
[3]Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Nara, Japan
[4]W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, NY 11724, USA
[5]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, NY 11724, USA
[6]The Arctic University Museum of Norway, UiT - The Arctic University of Norway, Lars Thørings veg 10, Tromsø 9006, Norway
[7]These authors contributed equally
[8]Senior author
[9]Lead contact
*Correspondence: laie@mskcc.org
https://doi.org/10.1016/j.celrep.2025.116243

SUMMARY

While canonical microRNA (miRNA) hairpins typically bear 30–35 base pair (bp) stems and <15 nucleotide terminal loops, certain miRNA hairpins are much longer. While vertebrates lack long miRNA hairpins, these emerged multiple times within invertebrate lineages. Systematic assessments across >1,400 genomes provided evolutionary insights into the elongation of well-conserved miRNA precursors, which can harbor >1 kb between conserved miRNA and star species and generally form highly base-paired structures. Experiments in *Drosophila* and humans showed that flies were preferentially capable of maturing certain long precursors, but human cells had a partial capacity. However, neither could handle extreme miRNA hairpins. Finally, analysis of structural variants revealed that extensive stem structure and a local bulge near the dicing site are critical for biogenesis of a lengthened miRNA precursor; the latter likely represents an internal DGCR8 interaction platform. Altogether, we document unanticipated structural complexity in conserved miRNAs and emphasize bioinformatic challenges for their complete annotation.

INTRODUCTION

MicroRNAs (miRNAs) are an extensive class of ∼22-nucleotide (nt) regulatory RNAs that derive from hairpin precursors.[1] In animals, these are initially produced as longer primary miRNA (pri-miRNA) transcripts, which can be tens to >100 kilobases (kb) in length.[2,3] Embedded within these are one or more local hairpin structures that are recognized by the nuclear Microprocessor complex, composed of the RNase III enzyme Drosha and its double-stranded RNA binding domain (dsRBD) partner DGCR8 (Pasha in invertebrates).[4] Microprocessor cleaves ∼10 base pairs (bps) from the hairpin base to liberate a pre-miRNA hairpin, and its activity is assisted by various cofactors, including SRSF3,[5,6] ERH,[7] and SAFB1/2.[8]

After its export to the cytoplasm, the pre-miRNA is cleaved on the hairpin loop side by the RNase III enzyme Dicer and its dsRBD partner TRBP, yielding ∼22-nt duplex species with 2-nt 3′ overhangs.[9] The duplex is asymmetrically loaded into an Argonaute effector protein and resolved into single-stranded form, preferentially retaining the mature (guide) miRNA species, and discarding the passenger (miRNA*) species.[10] The Argonaute-miRNA complex represses targets, typically via 7-nt Watson-Crick complementarity between positions 2–8 of the miRNA (the "seed") and the target 3′ untranslated region (3′ UTR).[11–16]

The stepwise processing of miRNA hairpins by the Microprocessor and Dicer complexes constrains their stem lengths, which usually exhibit ∼3 helical turns (∼33 bps). The connecting region excised upon Dicer cleavage (often referred to as the terminal loop) cannot be too small, as this abrogates effective recognition by DiGeorge Syndrome Critical Region Gene 8 (DGCR8).[7,17–19] However, hairpin loop regions evidently should not be too long either, since animal pre-miRNAs exhibit restricted length ranges. Specifically, most mammalian pre-miRNAs are 55–70 nt, with most terminal loop regions ∼10–15 nt in length.[20,21] However, there are exceptions. For example, certain mammalian let-7 members harbor terminal loops of several dozen nucleotides, which carry binding sites for the RBP Lin28 to inhibit let-7 maturation.[22–25]

Studies from 15 to 20 years ago identified certain miRNA loops that are even longer. For example, *Caenorhabditis elegans mir-229* bears a 54-nt loop[26] while *Drosophila melanogaster mir-989* carries a 99-nt loop.[27] Other, less studied species reveal unexpected miRNA structures. For example, the sponge *Amphimedon queenslandica* has a striking number of long miRNA hairpin loops, with a majority >100 nts in length.[28] Well-curated

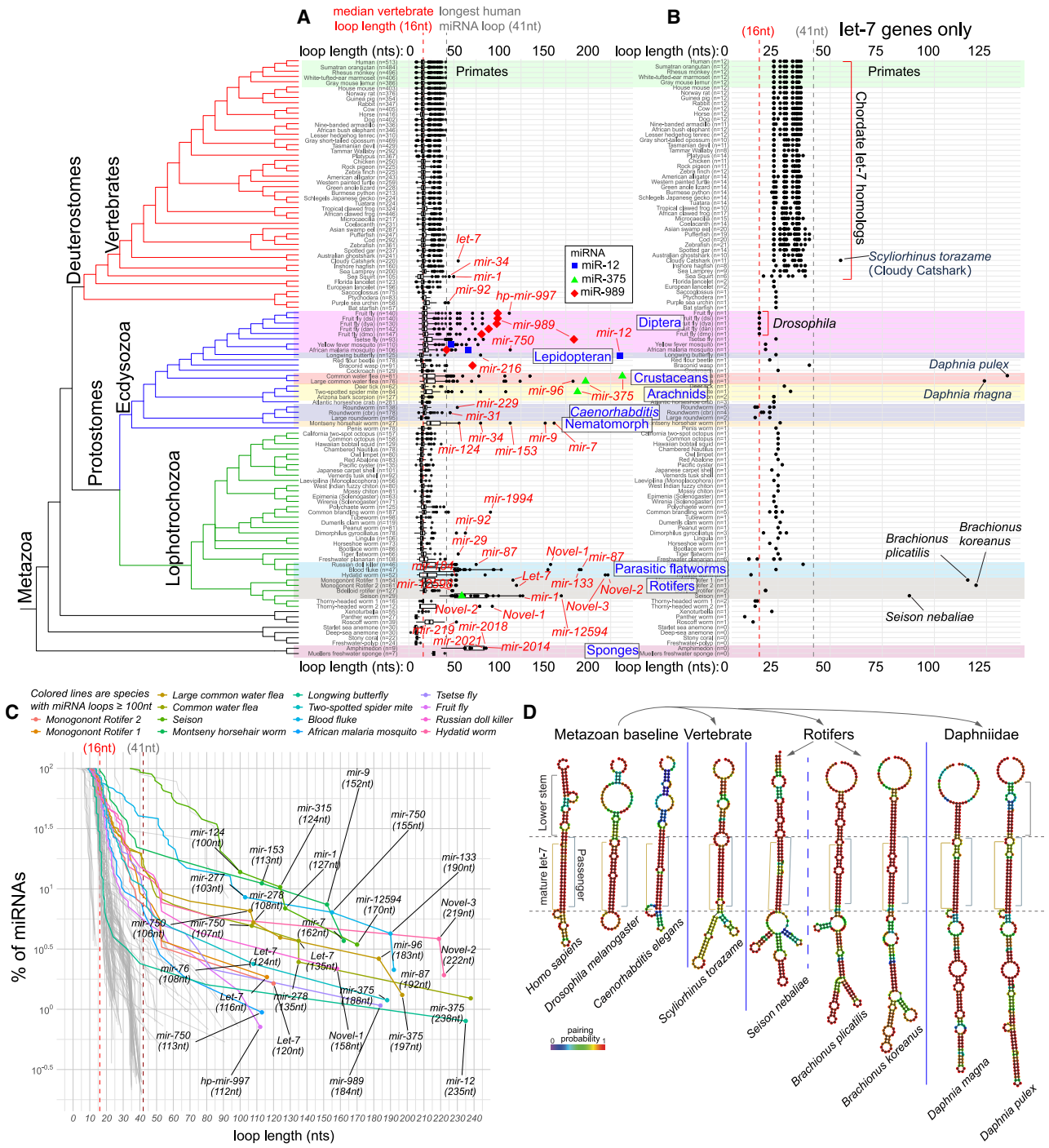


Figure 1. Recurrent emergence of metazoan miRNA hairpins of extended lengths

(A) Lengths of canonical miRNA Dicer-excised regions (designated here as hairpin "loops") across a broad selection of metazoan phylogeny, as curated by MirGeneDB 3.0. The phylogenetic relationships are shown to the left, with selected subclades labeled. Deuterostome miRNA hairpins exhibit uniform loop lengths, with only a single canonical chordate miRNA with a loop >50 nt, out of 13,788 deuterostome miRNA homologs. By contrast, numerous invertebrate canonical miRNAs have >50 nt loops, with some ranging from 100 to 250 nt. The phylogenetic distribution of these makes it evident that atypical miRNA structures emerged multiple times.

(B) The broad conservation of let-7 makes it informative to compare length alterations across animals. Chordate let-7 members are not only abundant in their respective genomes, but they also comprise the longest miRNA hairpins in these species. Protostomes have fewer copies or typically single orthologs. In *Drosophila* genomes, let-7 is shorter than all chordate let-7 members at the average chordate length. However, there are several cases of highly extended let-7 hairpins, including in multiple *Daphnia* species and multiple rotifers.

(legend continued on next page)

loci in MirGeneDB also reveal larger hairpins in the invertebrate groups such as flatworms,[29] crustaceans, and insects.[21] Finally, plant miRNAs, although not the focus of this study, can harbor hairpins ranging from hundreds of nucleotides.[30,31]

We analyzed heterogeneity of miRNA hairpin structures across a broad set of animals, focusing on well-conserved loci that are confidently identified from genome sequences alone. By screening >1,400 genomes, we find that atypically long miRNA precursors emerged recurrently within numerous invertebrate subclades. At the extreme, some conserved miRNA precursors exhibit inter-arm regions extending >1 kb. We delineate specific biogenesis requirements for model extended miRNA hairpins, including the necessity for an extended stem structure across the pre-miRNA and a bulged region adjacent to the prospective small RNA duplex, likely comprising an internal DGCR8 binding site. Altogether, we reveal surprising diversity in animal miRNA structures and unanticipated mechanistic questions on their biogenesis.

RESULTS

Wide variation in miRNA precursor lengths across the breadth of metazoan species

While pre-miRNA hairpins are generally ~55–70 nts in length, we evaluated hairpin lengths more broadly using the MirGeneDB database.[21] This contains a refined set of miRNA loci that were well-vetted to generate small RNA duplexes associated with characteristic hairpin precursors. The current version (MirGeneDB 3.0) houses miRNA annotations from a wide variety of metazoan species, including vertebrates, invertebrates, and a handful of basal metazoans, altogether comprising more than 1700 distinct seed families from 114 species.

As noted,[21] certain species harbor outliers in miRNA precursor lengths. The basal stem region that defines the Drosha cleavage site is generally consistent (~10 bp).[32] Thus, heterogeneity of miRNA precursors is typically due to variability of Dicer-cleaved regions. These are often referred to as miRNA hairpin "terminal loops." While the term loop may imply an unstructured region, in reality, these can contain both duplex and single-stranded regions.[17] To avoid ambiguity, we initially refer to these as "Dicer-excised regions."

By displaying hairpin length distributions alongside phylogenetic relationships of these species, it becomes clear that vertebrates have highly constrained Dicer-excised regions. Across homologs of all canonical miRNAs across several dozen vertebrates, only a single locus harbored an inter-arm length >50 nts (Figure 1A; Table S1). This sole outlier is Let-7-P2b3 from the cloudy catshark, *Scyliorhinus torazame*, whose Dicer-excised region is 55 nts (Figure 1B). However, across the invertebrates, there are numerous species with Dicer-excised regions that far exceed the maximums found in any vertebrate (Figures 1A, 1B, and S1). The extremes of miRNA Dicer-excised regions are visualized in the ranked cumulative distribution plot (Figure 1C).

let-7 was the first miRNA known to be broadly conserved across the Metazoa,[33] providing an opportunity to examine whether other let-7 copies altered their structures. Curiously, several invertebrate let-7 homologs exhibit extended Dicer-excised regions (Figure 1B). We infer at least three separate lengthening events for let-7, the longest of which reside in the Daphniidae water fleas (Figure 1D). All extended let-7 precursors adopt highly structured states, either as long straight hairpins or with branches (Figure 1D).

Other striking examples of pan-bilaterian miRNAs with anomalous hairpins included *mir-375* and *mir-96*. These harbor very long Dicer-excised regions in *Daphnia* species, which are branchiopod crustaceans (Figure 1A; Table S1). Additional prominent outliers were absent from vertebrates, but conserved among subclades of invertebrates. For example, *mir-12* and *mir-278* are broadly conserved across protostomes and bilateria, respectively, but became exceptionally long in certain lepidopteran species and *Daphnia* species, respectively (Figure 1A).

In most cases, median lengths of Dicer-excised regions in the invertebrate species are comparable to vertebrate medians (15 ± 4 nt loops) (Figure 1A; Table S1). Exceptions include sponges (*Amphimedon queenslandica*, 68 ± 17.3-nt loops),[28] certain parasitic blood flukes (*Schistosoma mansoni*, 34± 42.0-nt loops), and the recently discovered case of unusually long miRNA hairpins in rotifers (*Seison nebaliae*, 62 ± 29.7-nt loops)[34] (Figure 1A). Nevertheless, as there are overall similar constraints on hairpin lengths across metazoans (Table S1), this attribute is not simply relaxed in invertebrates. Instead, we infer that active processes drove selective elongation of miRNA hairpins.

Repeated emergence of unusually long miRNA hairpins among invertebrate subclades

We observe punctuated emergence of extended miRNA hairpins along specific invertebrate branches (Figure 1A), likely reflecting independent evolution. To examine evolutionary trajectories of miRNA hairpin lengthening in greater detail, we analyzed more species. Building on MirGeneDB 3.0,[21] we used MirMachine[35] to collect miRNA orthologs across 1414 genomes comprising representative invertebrate subclades with evidence for miRNA precursor lengthening (Figure 1). Our survey included 319 Drosophilidae (which include the vinegar flies, Drosophilids),[36] 421 other dipteran species (flies in general), 546 lepidopteran species (butterflies and moths), and 128 crustacean species.

We estimated lengths of Dicer-excised regions from MirMachine outputs by subtracting flanking single-stranded regions and approximate mature/passenger sequences. To benchmark this approach, we compared the lengths of Dicer-excised regions of *Drosophila melanogaster* miRNAs inferred from MirMachine, with experimental measurements from small RNA sequencing.[21,27] The vast majority of estimated Dicer-excised regions were within a few nucleotides of ground truth (Figure S2 and Table S2). Given heterogeneity in the Dicer-excised regions of all miRNA loci, due to alternative cleavage

(C) Plot of metazoan miRNA loop lengths, focusing on those with >100 nt Dicer-excised regions. This emphasizes that hairpin lengthening occurred selectively on a few miRNAs.
(D) Secondary structures of selected let-7 members that illustrate multiple lengthening events during metazoan evolution.

yielding isomirs, these minor discrepancies would not affect our conclusions on long miRNA precursors.

Overall, we predicted 118,852 homologs of 81 well-conserved miRNA families across the 1,414 Diptera/Lepidoptera/Crustacea/Drosophilidae genomes (Table S3). MirMachine is recognized to overpredict some loci,[35] including by permitting hits with very low free energy to have relaxed constraints on the miRNA sequence. We decided to filter these by flagging loci with an imperfect match to the seed region of the queried miRNA family, or an unusually long perfect hairpin stem (>32-nt consecutive bp). We also noted some MirMachine hits were shorter than expected, and flagged these (see STAR Methods). While many flagged loci (9,969, 8.38% of total predictions) are likely genuine miRNAs, to be conservative, we only used the 108,883 non-flagged loci for subsequent analyses. Finally, we note that some miRNA families had unusually large numbers of predicted copies. Whether this is due to assembly errors or genuine genomic amplification is not yet clear. We call attention to them by highlighting families with >15 copies in a given genome in Table S4.

Using only the confident, non-flagged loci, we identified 40 miRNA families with multiple members bearing an inferred Dicer-excised region ≥42-nt long, the maximum of known vertebrate miRNAs (excepting Sto-Let-7-P2b3, Figure 1); five additional miRNA families had a single member with ≥42-nt loop. The families of "long" miRNA precursors and their numbers are summarized in Table S5. This table also provides a series of worksheets for each miRNA family with lengthening detected, including long members and all of their conventionally sized copies across species. Overall, these data provide abundant evidence for recurrent extension of miRNA precursors during evolution.

For instance, across 128 crustaceans, several miRNAs are specifically extended in different species (Figure 2A), but the very longest miRNAs occurred among nine species that all turn out to be *Daphniidae* species (water fleas) (Figure S3). The longest *Daphniidae* miRNAs are *mir-375*, *let-7*, *mir-278*, and *mir-750*, whose Dicer-excised regions are often >100 nt across these families (Figure 2B). We also draw attention to *mir-989*, which is broadly conserved across arthropods, but became highly extended across a broad swath of dipterans in the Calyptratae subsection (Figure S4). Many Acalyptratae subsection species, the so-called "true flies," harbor shorter *mir-989* orthologs. However, within this lineage, numerous Ephydroidea superfamily species once again extended *mir-989*; this species branch contains the Drosophilidae (Figure S4). Yet, even within the Drosophilidae, there is again substantial variation in *mir-989* lengths. This highlights the episodic lengthening and shortening of this miRNA hairpin across Dipteran evolution.

Altogether, these large-scale miRNA annotations are a trove of information on the emergence of atypical precursors of well-conserved miRNAs, on both short and long evolutionary timescales, along with their more conventionally sized miRNA copies across species (Tables S3, S4, and S5).

Extraordinary hairpin structures of *mir-12* orthologs in butterflies and moths

mir-12 is broadly conserved across Protostomia, including Ecdysozoa (e.g., arthropods and nematodes), but harbors extremely long miRNA precursors within the Lepidoptera (Figures 2C and S5). It is instructive to describe how we identified these miRNA copies, since MirMachine failed to identify *mir-12* in most lepidopterans. In fact, we recovered clear BLAST hits to mature miR-12 across lepidopteran genomes, but their hairpin structures were not initially evident.

Typically, one inspects small RNA data for a local, sense-strand, read cluster that pinpoints the passenger strand.[37,38] Lacking such data, we resorted to evaluating larger and larger genomic windows for pairing to mature miR-12 sequences. This eventually identified characteristic miR-12 duplexes for these lepidopteran species, which collectively comprise the most strikingly lengthened precursors of any species known. Of 523 species surveyed, 486 have >200 nt Dicer-excised regions, i.e., over twice the maximum length in *D. melanogaster*. The Dicer-excised regions are >500 nt in 188 species, >800 in 35 species, and even >1 kb in seven species of butterflies or moths (Table S6).

The proposal of such extraordinary miRNA precursors demands validation. Is it conceivable that *mir-12* is a pseudogene in Lepidoptera? The full conservation of mature miR-12 and nearly precise matches to its presumed passenger strand, amidst otherwise diverged genomic regions, argues that lepidopteran *mir-12* loci are conserved for the purpose of miRNA biogenesis (Figure 2D). Fortunately, we could test this using the available small RNA data.

The first deep sequencing study that reported lepidopteran miR-12 comes from the silkworm *Bombyx mori*.[39] Although its mature sequence is clearly orthologous to *Drosophila* miR-12, no star species was mentioned, and the proposed hairpin exhibits atypically modest base-pairing. Another study from the tobacco hornworm moth *Manduca sexta* similarly recovered hundreds of miR-12 reads, but no star species.[40] In the postman butterfly *Heliconius melpomene,* miR-12 was annotated by homology, but no precursor structure was proposed.[41] Finally, studies of the fall armyworm *Spodoptera frugiperda*[42] and the cabbage looper moth *Trichoplusia ni*[43] annotated orthologs of miR-10, miR-11, miR-13, and miR-14 that are shared with *Drosophila*, but "skip" over miR-12.[44]

As we found *Heliconius melpomene* small RNA data reveal precise miR-12 and star duplex reads separated by 234 nts,[21] we asked if atypical precursors could be validated from other lepidopteran species. Indeed, we identified specific duplex small RNAs emanating from highly extended precursors for each of the five butterfly and moth species (Figure 2E). This validates unprecedented distances between duplex small RNAs, some over 600 nts. Notably, these loci lack intervening small RNAs that might reflect additional RNase III cleavages, as is the case for hairpin RNA structures that generate endo-siRNAs.[45]

Evolutionary trajectories of hairpin lengthening of broadly conserved miRNAs

We sought insights into the evolutionary changes in sequence and structure associated with extended miRNA structures. *mir-375* was informative, as it is conserved between invertebrates and vertebrates and generally represented by a single ortholog. Some lepidopterans acquired substantially long loops (*Metalampra italica*, 54 nt and *Doxocopa laurentia*, 50 nt), while a Chelicerate (*Tetranychus urticae*, a spider mite) lengthened its loop to 180 nt

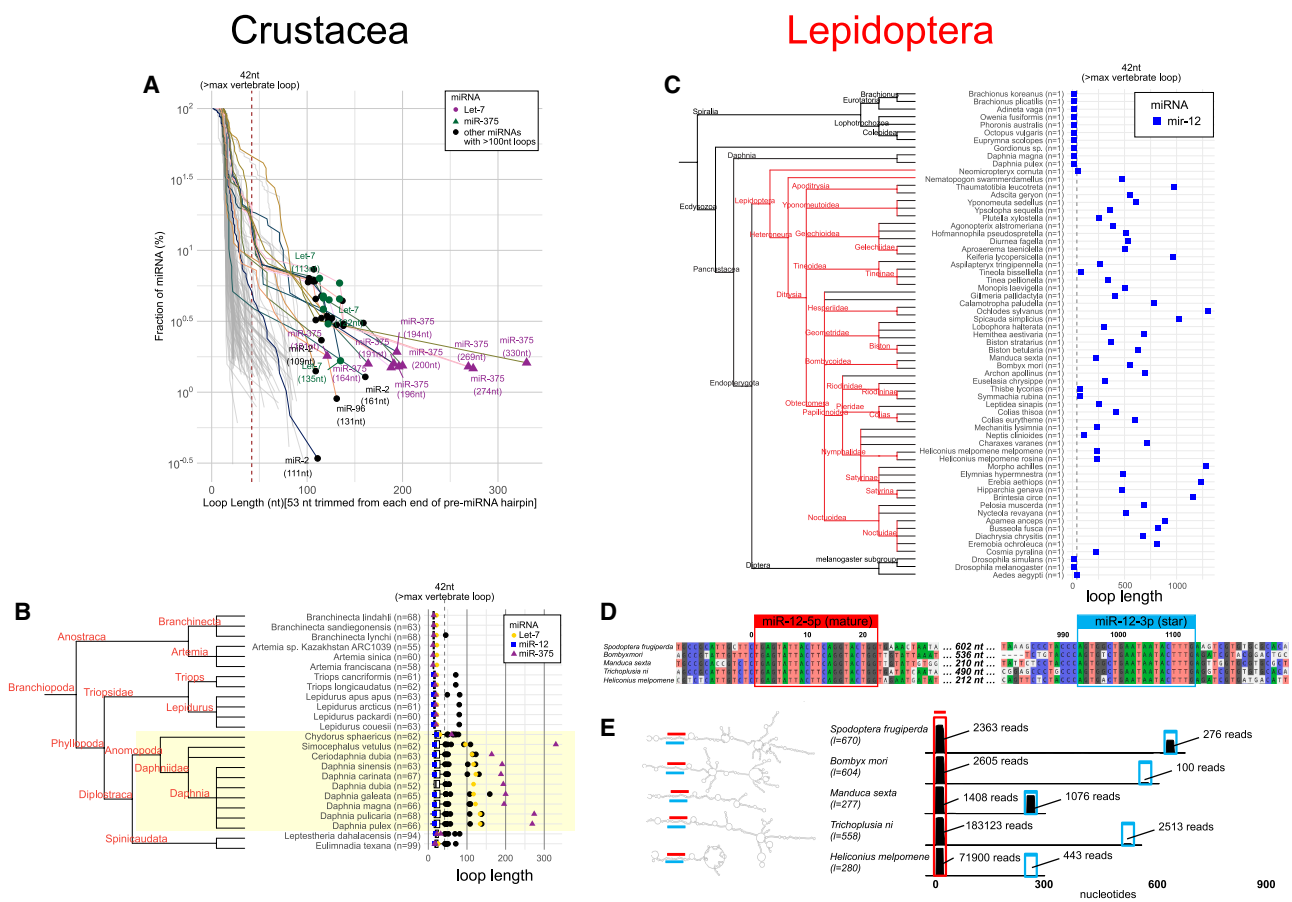


Figure 2. Marked elongation of specific miRNA hairpins in different invertebrate lineages

(A) Ranked cumulative distribution plot of 128 crustacean species revealed that specific miRNAs acquired highly elongated structures. Out of 9,766 total miRNAs analyzed in crustaceans, 42 harbor Dicer-excised regions >100 nt.

(B) A striking example is broadly conserved *mir-375*, which became highly elongated within the Anomopoda order (branchiopod crustaceans). This set includes *Chydorus sphaericus* and nine *Daphniidae* species (*Daphnia* species: *Ceriodaphnia dubia* and *Simocephalus vetulus*). This miRNA loop length plot compares a selection of crustacean species and other outgroup species for comparison.

(C) Loop lengths of a subset of *mir-12* orthologs. *mir-12* bears a typical miRNA hairpin in diverse ecdysozoan and spiralian species, but has become extremely elongated across hundreds of lepidopteran species. Although certain lepidopterans have a relatively normal *mir-12* structure, the majority of species harbor ~500 nt precursors, with most of the remainder being >200 nt. The most marked extensions occur within *Papilionoidea* (butterflies), which can extend >1 kb.

(D) Alignment of the regions surrounding the mature and passenger strand sequences for *mir-12* in selected Lepidoptera.

(E) Structures (left) and small RNA patterns (right) from *mir-12* loci in lepidopteran species with available small RNA data.

(Figures 3A and 3A'). However, inspection of 128 crustaceans revealed that the top 10 lengths of *mir-375* precursors all comprised representatives of the Anomopoda order, which includes branchiopod crustaceans (*Daphnia* species). The longest crustacean miRNA precursor was *mir-375* from *Simocephalus vetulus*, which, despite its name, is within the *Daphniidae* family; its Dicer-excised region is 330 nt. Finally, *Chydorus sphaericus* was the other Anomopodan species analyzed; its Dicer-excised region is 60 nt, shorter than that of the *Daphniidae* but still longer than all vertebrates.

The Chelicerate and Anomopodan *mir-375* extensions are unrelated, implying convergent lengthening (Figure 3A). However, further inspection revealed stepwise extension of Daphniidae orthologs. *Daphnia magna* and *Daphnia carinata mir-375* have overall similar structures, even though their sequences have diverged substantially. *Daphnia pulex* and *Daphnia pulicaria* have stem extensions relative to these, and *Simocephalus vetulus* is the longest of all (Figure 3A).

Lepidopteran *mir-12* was also informative, although difficult to visualize on account of its extraordinary precursors (Figure 3B). Some dipterans moderately extended their copies of *mir-12*, but most crustacean and dipteran copies are normal. In contrast, most lepidopteran *mir-12* orthologs harbor extreme lengths (Table S6 and Figure 3B'). We highlight a few in Figure 3B, along with typical *mir-12* orthologs for comparison. The predicted structures of these giant miRNA hairpins are diverse, and they harbor limited sequence similarity (Figure S6A). We infer that a lepidopteran ancestor had extended its *mir-12* gene, but that ongoing evolution across their phylogeny led to the current variety of *mir-12* structures. Similarly, the extended *mir-989* precursors across

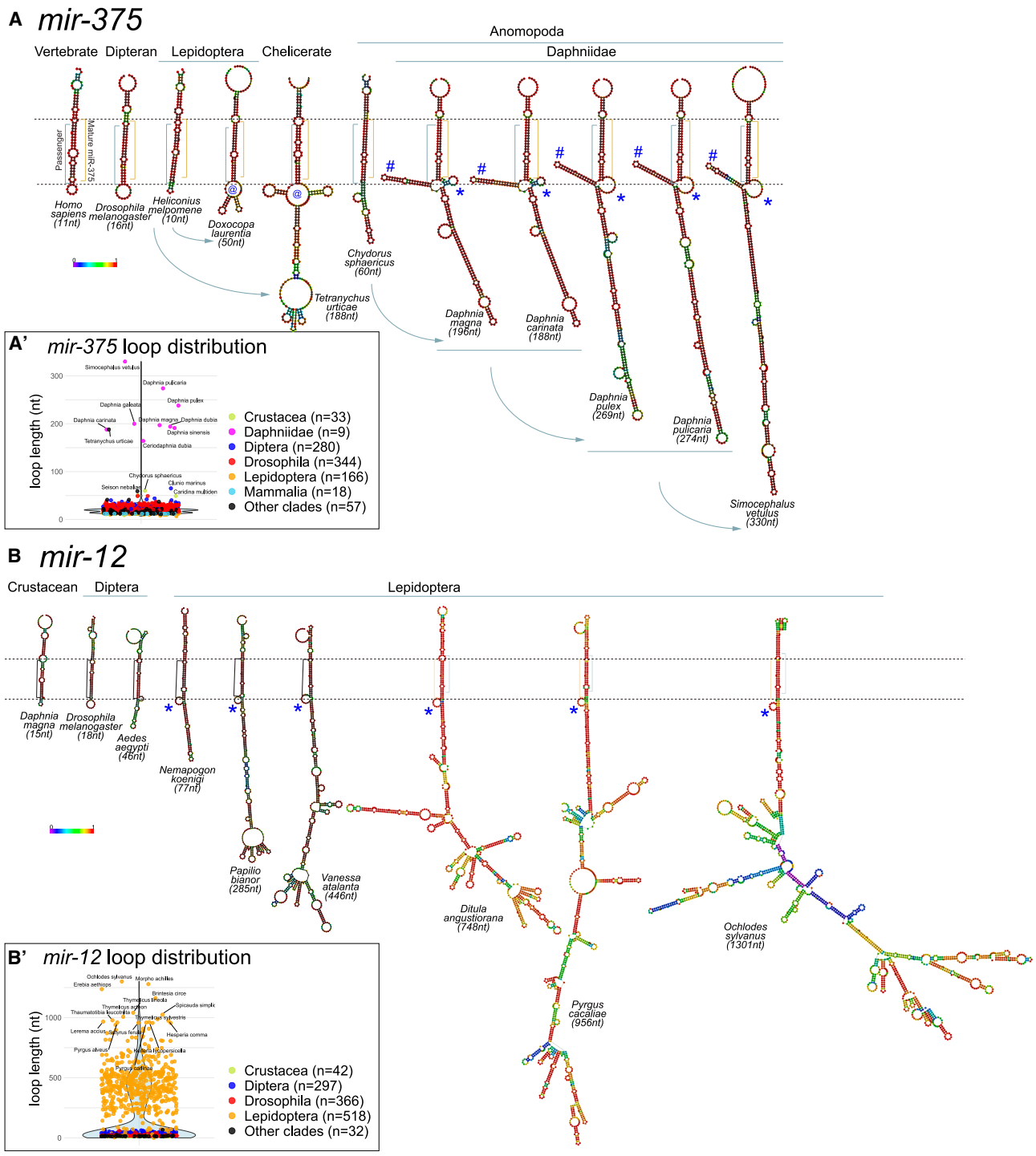


Figure 3. Structural variation and evolutionary dynamics of miRNA precursor lengthening

Shown are miRNA examples that highlight stepwise lengthening of miRNA precursors and/or extreme examples of elongation.

(A) *mir-375* is conserved across vertebrates and invertebrates, but has selectively lengthened in a small subset of species that include a Chelicerate (the spider mite *Tetranychus urticae*) and several Daphniidae species (water fleas). The distribution of *mir-375* loop lengths across all species is shown in (A'). The *Tetranychus mir-375* hairpin is quite distinct from the Daphniidae copies, consistent with independent paths for lengthening. However, the latter clearly share sequence and structural features within the extension proximal to the Dicer cleavage site, although a subset of Daphniidae species exhibit additional lengthening, outlining stepwise extensions. These share a prominent branched stem at the dicing site (designated by #), which is opposed by a bulged region (marked by *).

(legend continued on next page)

Diptera have divergent loop sequences, but maintain long continuous stem structures (Figures S6B and S6C).

In these examples, there are prominent unpaired regions adjacent to the dicing site. All Daphniidae *mir-375* members bear a prominent branched stem at the dicing site that is opposed by a bulged region on the opposite side of the stem (Figure 3A, # and * labels, respectively). All the lepidopteran *mir-12* members bear a similar bulged region in this region (Figure 3B, * labels) as do Dipteran *mir-989* members (Figure S6C). Such features may be relevant to the biogenesis of atypical miRNA hairpins, a hypothesis we address later.

Extensions of highly lengthened miRNA precursors are highly structured

It was visually evident that extended Dicer-excised regions harbor substantial base-pairing. We evaluated this systematically by plotting total hairpin length against minimum free energy (MFE) for each homolog of a given miRNA. With normal miRNAs, there can be substantial variation in the MFE across orthologous hairpins, with some exhibiting less predicted structure than others (Figure S7). Despite such ''cloud-like'' distribution, typical miRNA precursors are, by definition, highly structured.

When extended miRNA hairpins are rare and/or modest in size, their added lengths have negligible influence on the correlation of length and MFE (Figure S7). However, among the collection of highly extended miRNA precursors (i.e., the 10 miRNA families with multiple members with 100-nt Dicer-excised regions), we observe similar correlations of length and MFE as their shorter counterparts. This can be seen with *mir-12* and *mir-375* members that have been mentioned (Figures 4A and 4B), but also with the *mir-216-P2* (Figure 4C) and *mir-989*, *let-7*, *mir-750*, *mir-278*, *mir-2*, *mir-184*, and *mir-282* families (Figure S7). We overlay the correlations for these miRNAs in a single plot (Figure 4D) and provide detailed length-MFE comparisons in Figure S7.

RNA secondary structure programs will predict base-pairing, even with an arbitrary sequence. However, it is evident that extremely long Dicer-excised regions are generally just as base-paired as typical pre-miRNA hairpins; the only outlier was *mir-29*. Since this principle is shared across diverse species, we infer that extensive structure maintains biogenesis of extended miRNA precursors.

Fly and human cells have distinct capacities to mature miRNAs from long hairpins

As mentioned, *Drosophila* harbors some markedly extended miRNA hairpins,[27] but vertebrates exhibit tight restriction of canonical miRNA hairpin lengths (Figure 1A). Accordingly, one might assume flies are more proficient at handling such substrates than humans. However, this notion has not been directly tested. As we shall see, the experimental outcomes were more complex than imagined.

We first tested *Drosophila melanogaster mir-989* (Figure 5A) for biogenesis in fly S2R$^+$ cells and human HEK293T cells. Northern blotting showed *mir-989* was effectively processed into mature miRNAs in S2R+, but not HEK293T (Figure 5B). These data align with expectations. With this in mind, we selected additional long (125–233 nt) miRNA hairpins (*mir-278*, *mir-375*, and/or *mir-12*) from *Daphnia magna* and/or *Heliconius melpomene* (Figure 5A). We compared these to expression constructs of their respective *Drosophila melanogaster* orthologs, as well as *Homo sapiens mir-375*. We also tested their repression activities using luciferase sensors.

In some cases, *Drosophila* cells were better at processing an extended miRNA than human cells, exemplified by *Daphnia magna mir-278* (Figure 5B). However, the biogenesis of this extended *mir-278* homolog was markedly compromised in S2R+ cells, compared to its fly counterpart. This was reflected in the modest repression capacity of *Daphnia mir-278* in S2R+ cells (Figure 5C). Still, human cells were not incapable of handling long miRNA hairpin constructs. Although its maturation and activity were modest, long *Daphnia mir-375* exhibited comparable properties in S2R+ and HEK293T cells (Figures 5B and 5C).

Finally, exceptionally long miRNA loci presented a challenge to both species. Fly and human cells processed *Drosophila melanogaster* and *Daphnia magna mir-12*, which are of comparable size. However, neither could convert *Heliconius melpomene mir-12* into mature miRNAs (Figure 5B), nor could they repress a miR-12 (Figure 5C). With optimal northern blots, we observed not only heterogeneous long species (presumably corresponding to primary transcripts), but also discrete bands that correspond to *Heliconius melpomene pre-mir-12* hairpins (Figure 5B). Therefore, such atypical hairpins can at least be partly recognized and cleaved by the Microprocessor machinery of flies and humans.

Based on this, we directly tested if the longer apical loops of exceptional invertebrate pri-miRNAs impact their processing via recombinant Microprocessor. To do so, we utilized purified human Microprocessor complex,[6] and compared its *in vitro* cleavage activity on long *Daphnia magna pri-let-7* (Figure 1D) and *pri-mir-375* (Figure 5A), compared to their human pri-miRNA counterparts. The control human pri-miRNAs were productively and accurately processed, as shown by the accumulation of specific pre-miRNA species over time. However, both *Daphnia* pri-miRNA substrates were comparatively poorly processed. Neither pre-miRNA nor 3′ ssRNA or 5′ ssRNA species accumulated over the same time course, and there were substantial products of unexpected size, presumably reflecting nicked substrates (Figure 5D). Thus, while human Microprocessor harbors some activity on atypically long pri-miRNA hairpins, it is substantially impaired at cleaving such substrates.

We conclude that *Drosophila* can preferentially handle some long miRNA substrates, but that extreme extension of miRNA hairpins compromises biogenesis in both fly and human cells. This further supports that hairpin lengthening is not a neutral process, but more likely associated with regulatory strategies that enable the biogenesis of selectively lengthened loci in their cognate species.

(B) *mir-12* became extraordinarily long in most lepidopteran species; this can be seen by plotting all *mir-12* loop lengths across all species in B'. Shown are some outgroup orthologs as comparison, along with a selection of lepidopteran *mir-12* orthologs that span 285 nt to 1.3 kb. Collectively, these atypical *mir-12* loop sequences harbor diverse sequences and structures. All of the extended *mir-12* structures share a bulged region adjacent to the dicing site (* symbols).

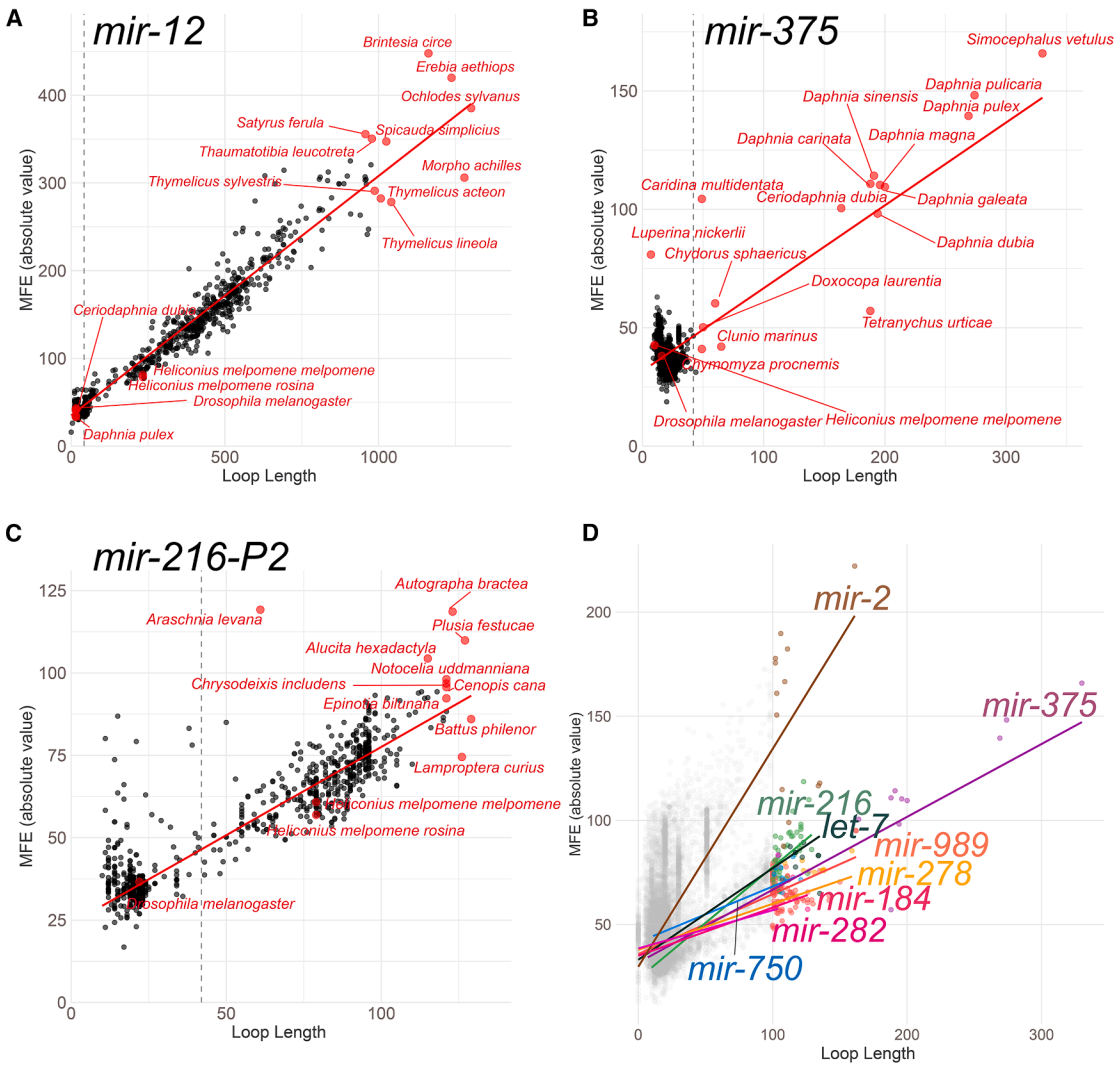


Figure 4. miRNA hairpin lengthening is associated with extensive secondary structure

Correlation of pri-miRNA hairpin lengths and minimum free energy (MFE) for *mir-12* homologs (A), *mir-375* homologs (B), and *mir-216-P2* homologs (C). Note that MFE is usually expressed as a negative value, but we plotted absolute values for convenient visualization of positive correlation instead of negative correlation. Typical canonical miRNAs reside in the lower left of these plots. Since these are all straight hairpins, they are within the range of the lowest free energy that they could adopt for their respective lengths and sequence contents. As miRNA precursors extend into atypical lengths, they tend to maintain similar degrees of correlation of length with absolute MFE. This indicates that, by and large, extended miRNA precursors generally maintain long stems or branched structures. (C) Overlay of the regression values of pri-miRNA hairpin lengths and absolute MFE across the ten families of miRNAs that contain at least two homologs with >100 nt Dicer-excised regions (excluding *mir-12*, as its loop lengths are too long to plot on the same graph as the others, see (A).

(D) Overlay of the MFE/loop length comparisons for all miRNA families bearing long hairpin homologs.

Functional evidence that long hairpin *mir-989* is a canonical miRNA substrate

We sought mechanistic insights into the maturation of an atypically long miRNA precursor. Since all of the highly extended miRNA precursors tested were modestly processed (Figures 5B and 5D), we selected *Drosophila melanogaster mir-989* as our subject. It is the longest endogenous miRNA hairpin in this species,[27] and is efficiently processed into active small RNAs (Figures 5B and 5C). Nevertheless, caution is needed when assigning atypical loci to a particular biogenesis pathway. This is due to the fact that, unlike mammalian cells, *Drosophila* harbors a distinct endogenous RNAi pathway that can efficiently process long hairpins with stems of several hundred base pairs into siRNAs.[45,46] Indeed, *mir-989* is comparable in length to *hp-mir-997-1* (Figure 6A), which we previously clarified is not an miRNA, but in fact an RNAi-dependent hairpin RNA (hpRNA) substrate that yields endo-siRNAs.[46] Thus, functional tests are warranted to assess if a given long hairpin is processed using miRNA or siRNA factors.

We generated a panel of dsRNA-mediated knockdown conditions in S2 cells, and transfected them with expression constructs for *mir-1* (canonical miRNA), *hp-mir-997-1* (hpRNA), and *mir-989* (Figure 6B). As expected, the biogenesis of mature miR-1 was

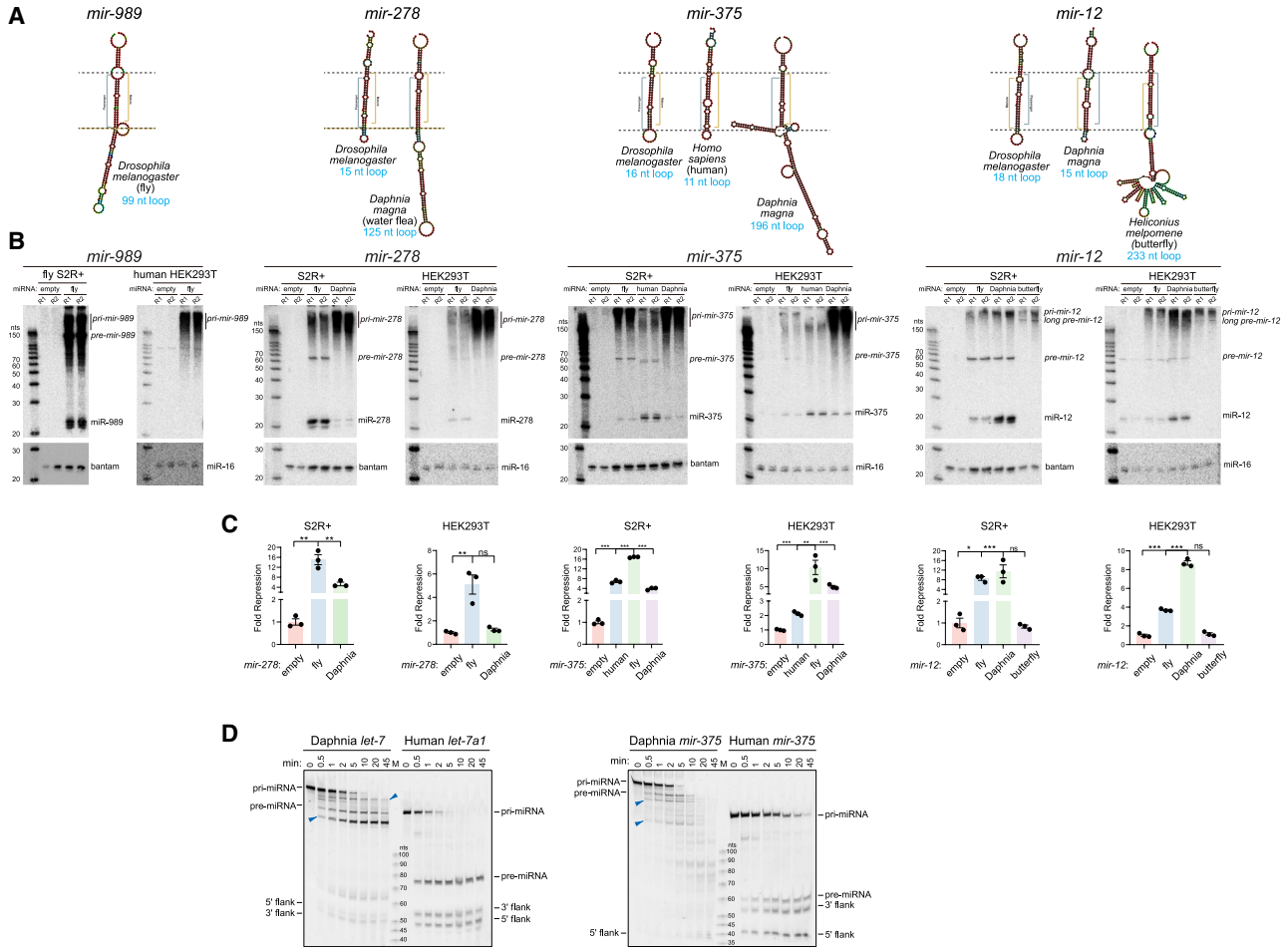


Figure 5. Biogenesis and activity of extended miRNA hairpins in fly and human cells

Functional studies of miRNA biogenesis and function in *Drosophila* S2R+ cells and human HEK293T cells.

(A) Structures of hairpin precursors that show broad length variation across the indicated species.

(B) Northern blots of total RNAs from S2R+ and HEK293T cells transfected with the indicated miRNA expression constructs. Bantam and miR-16 were used as loading controls in fly and human cells, respectively. Some long miRNA precursors are better processed in fly cells compared to human cells (e.g., *mir-989* and *mir-278*), but long *mir-375* was equivalently processed in both species. The long miRNA precursors of *Daphnia magna* (*mir-278* and *mir-375*) generally exhibited poorer biogenesis compared to their fly and/or human orthologs. The long *Heliconius melpomene* mir-12 was not detectably processed in either species.

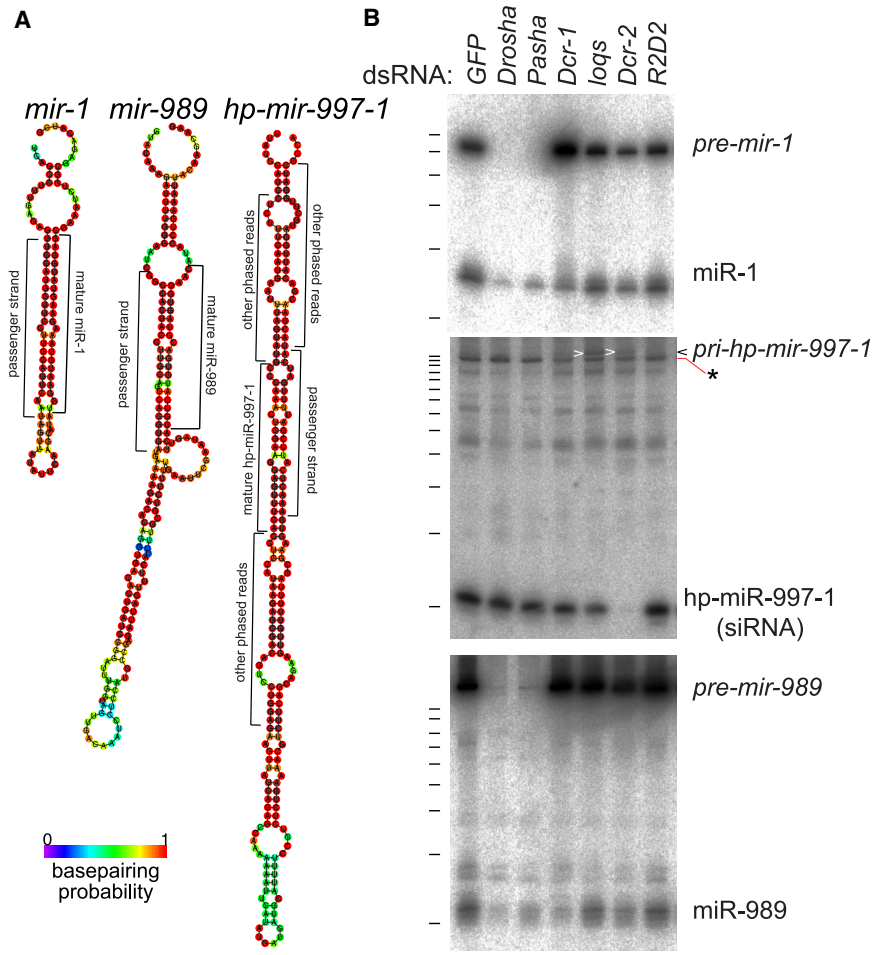
(C) Functional repression assays using luciferase sensors bearing two antisense matches to the cognate miRNA. The capacity of these miRNAs to repress their sensors was well correlated with the relative accumulation of mature miRNAs.

(D) *In vitro* cleavage assays of different pri-miRNAs using recombinant human Microprocessor. Human *pri-let-7* and *pri-mir-375* are productively cleaved, with time-dependent accumulation of pre-miRNA hairpins and 5′/3′ ssRNA flanks of expected sizes. By contrast, long *Daphnia pri-let-7* and *pri-mir-375* hairpins are modestly cleaved, but their pre-miRNAs do not accumulate over time. The RNA species marked by blue arrowheads likely represent nicked product species.

highly dependent on Drosha, Pasha, and Dicer-1 (Figure 6B). Moreover, the *pre-mir-1* hairpin disappeared upon Drosha and Pasha depletion, while it increased upon Dicer-1 depletion. In contrast, the accumulation of mature hp-miR-997-1 siRNA was strongly dependent on Dicer-2 and partly dependent on Loquacious (Loqs, and presumably the Loqs-PD isoform) (Figure 6B), consistent with other hpRNAs we studied.[46,47] Among these perturbations, depletion of Dcr-2 and Loqs both induced the accumulation of a new species that was larger than a presumed background hybridizing band present in all conditions (Figure 6B). These criteria indicate that it is likely a primary *hp-mir-997-1* species. These data confirm that small RNA expression constructs can be used to assess selective biogenesis requirements for miRNAs and siRNAs.

With these knockdown validations in hand, we assessed the maturation of the transfected *mir-989* construct. Its requirements mirrored those of miR-1, and not of hp-miR-997-1, thus establishing it as a canonical miRNA locus. In particular, mature miR-989 products were strongly reduced upon depletion of miRNA biogenesis factors Drosha, Pasha, and Dicer-1 (Figure 6B). More tellingly, the normally strong *pre-mir-989* band was nearly absent upon loss of either Microprocessor factor.

Next, we performed *in vitro* assays to test if cleavage of *pre-mir-989* was selective for the miRNA Dicer. We used *in vitro*



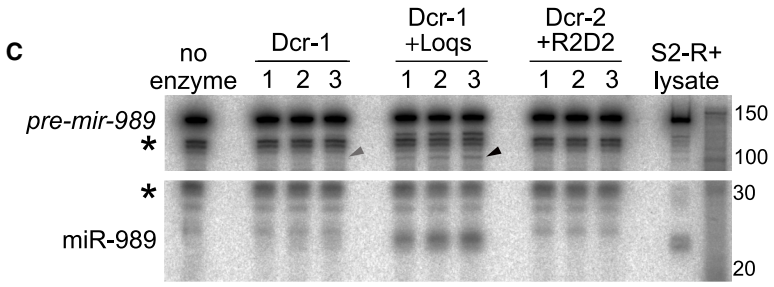


Figure 6. Mechanistic studies of long miRNA hairpin processing in cells and *in vitro*

(A) Precursor structures of canonical miRNA *mir-1*, hairpin RNA *hp-mir-997-1*, and *mir-989*.

(B) Assays of small RNA biogenesis upon dsRNA-mediated depletion of a panel of core miRNA and RNAi factors. Cells were then transfected with the indicated upstream activating sequence (UAS)-DsRed-small RNA expression constructs and activated using Ub-Gal4, and RNAs were isolated for northern blotting. The maturation of miR-989 shows similar dependencies on core miRNA factors as the control miR-1, while the maturation of hp-miR-997-1 requires siRNA factors.

(C) *In vitro* processing of *pre-mir-989* using recombinant Dicer complexes. *In vitro* transcribed, internally radiolabeled *pre-mir-989* was incubated with purified Dicer-1, Dicer1 + Loqs, and Dicer-2 + R2R2 was resolved on a urea acrylamide gel. The S2R+ cytoplasmic extract was used as a positive control. The numbers above the gel panel indicate the time of incubation in hours. The asterisk indicates a band that represents a contaminating shorter *pre-mir-989* fragment that was present in the substrate preparation.

transcription to prepare internally radiolabeled *pre-mir-989*, and incubated it with recombinant Dcr-1 ± its cofactor Loqs-PB or with Dcr-2/R2D2 complex. Although we purified *pre-mir-989*, it appeared prone to hydrolysis, as evidenced by heterogeneous species shorter than the full hairpin. However, we could establish the migration of *in vitro* processed miR-989 by incubating *pre-mir-989* with S2R+ cell lysate, which generated a characteristic miRNA-sized species (Figure 6C). With this reference, we found that incubation of *pre-mir-989* with Dcr-1 yielded a modest amount of miR-989. However, Dcr-1 cleavage of *pre-mir-989* was stimulated by Loqs-PB protein, as expected from their known structural collaboration.[48] In contrast, recombinant Dcr-2/R2D2 did not generate mature miR-989, indicating specificity to this reaction (Figure 6C). Overall, these functional tests in cells and *in vitro* demonstrate that *mir-989* is indeed a canonical miRNA, despite its highly extended structure that is comparable in length to some hpRNAs.

Specific structural features are required for the biogenesis of long miRNA hairpins

The entry of canonical miRNAs is gated by the heterotrimeric Microprocessor complex, composed of one Drosha and two Pasha (known as DGCR8 in vertebrates) proteins. Pasha binds within the terminal loop, while Drosha recognizes the junction between the basal stem and the single-stranded flanking regions.[6,32,49,50] Since the terminal loop of these extended miRNA precursors does not seem capable of positioning Microprocessor appropriately, we were curious to gain further mechanistic insights into the processing of these unusual miRNA substrates.

We designed mutant *mir-989* hairpins that perturb various aspects of its extended structure. These include manipulations that move the internal bulge near the dicing site from the 5p arm to the 3p arm (M1, "opposite bulge"), that remove this internal bulge (M2, "no bulge"), that abrogate secondary structure within the inter-arm region (M3, "unpaired loop"), truncate the extended stem to a normal miRNA size (M4, "normal stem"), or that further lengthen its stem by grafting in the region from the *Daphnia magna mir-278* (M5, "extended stem") (Figure 7A).

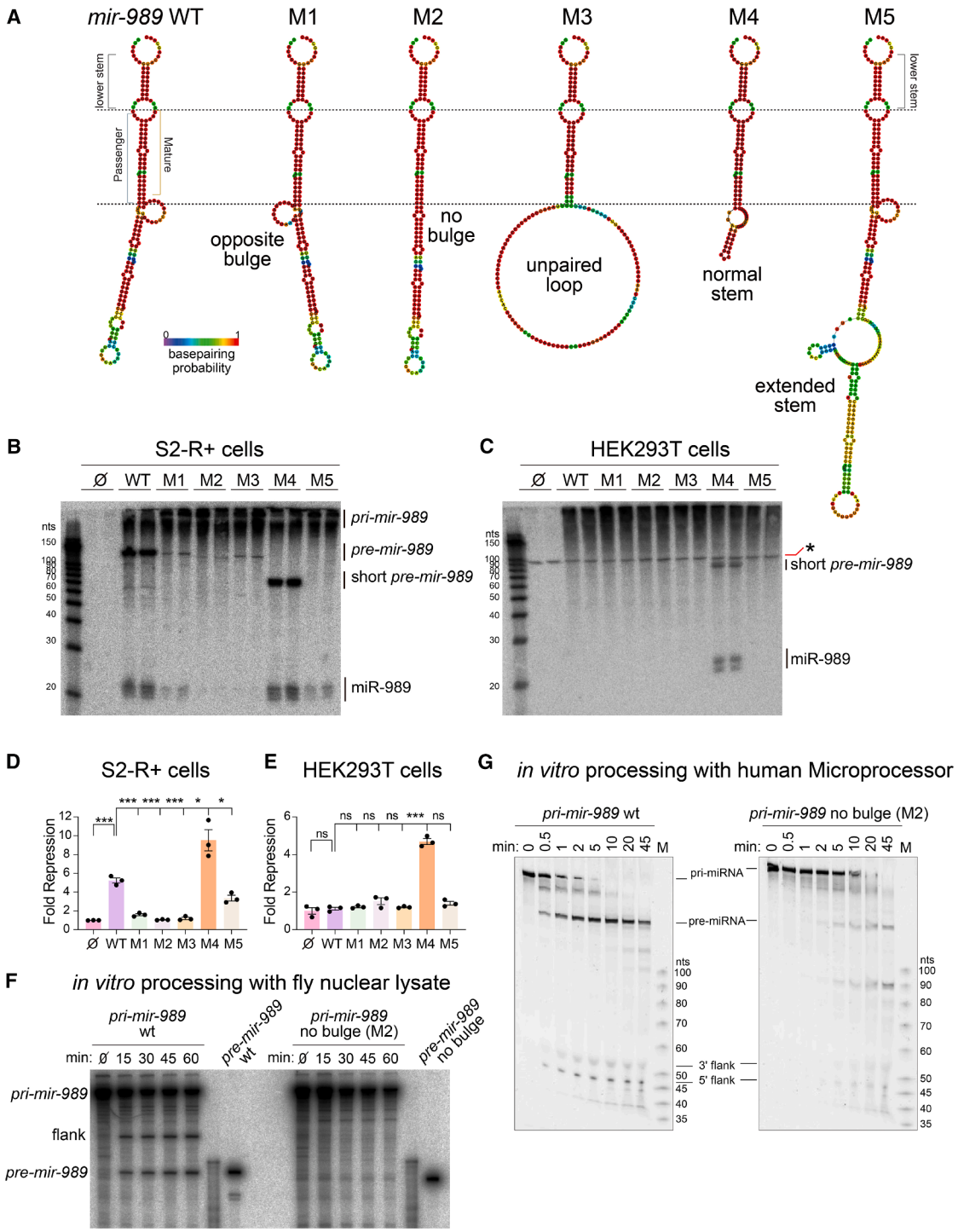


Figure 7. Structural features required for the biogenesis and function of miR-989

(A) Structures of wild type (WT) and variant *mir-989* (M1-M5) hairpins used for functional tests.

(B) Maturation of the panel of *mir-989* constructs in *Drosophila melanogaster* S2R+ cells, assessed by northern blotting. Alteration or removal of the internal bulge strongly compromised miR-989 biogenesis, as did further stem lengthening.

(C) Maturation of the panel of *mir-989* constructs in *Homo sapiens* HEK293T cells. Only the shortened *mir-989* construct was matured.

(D and E) Luciferase sensor assays of *mir-989* constructs in fly (D) and human (E) cells. Their repression activities parallel their maturation capacities.

(legend continued on next page)

As before, we cloned these variants into *Drosophila* and mammalian expression plasmids and compared their biogenesis and activity in S2R+ and HEK293T cells. In fly cells, the wild-type (WT) construct yielded substantial amounts of *pre-mir-989* and mature miR-989. However, except for M4, which converts *mir-989* into a typical miRNA structure, all other variants prevented the accumulation of both pre-miRNA and mature miRNA (Figure 7B). In particular, removal of the internal bulge near the dicing site (M2) and the abrogation of stem structure beyond the miRNA/passenger duplex (M3) had the most severe effects on biogenesis, with M2 suppressing detectable *pre-mir-989* (Figure 7B).

In contrast, *mir-989* WT and its mutants were not processed in HEK293T cells (Figure 7C). This was consistent with other data that human cells are generally poorly able to handle such extended miRNA substrates (Figure 5). The exception was M4, which rescued miR-989 processing upon shortening to the typical vertebrate length (Figure 7C). We also tested all of these constructs against miR-989 luciferase sensors in fly and human cells. All of their repression capacities correlated with their relative biogenesis of mature miR-989 (Figures 7D and 7E).

The fly data suggested that the M2 variant, which removes the internal bulge near the dicing site, may directly impede Microprocessor recognition. To test this directly, we used *in vitro* processing assays using internally radiolabeled *pri-mir-989* substrates and S2 cell nuclear extracts. Tests using WT *pri-mir-989* showed time-dependent accumulation of *pre-mir-989* hairpin, whose size matched synthetic *pre-mir-989* (Figure 7F). We next tested the M2 variant of *pri-mir-989* (Figure 7F), which more closely resembles a hpRNA substrate (Figure 6A). In fact, this mutant substrate was not substantially cropped *in vitro*. These data support the notion that this single-stranded region may serve as an internal site for Pasha recognition, which would be compatible with the positioning of Drosha to cleave the pre-miRNA hairpin, despite its extended structure. As noted, other extended miRNA hairpins exhibit analogous unpaired regions near their dicing site (Figures 3 and S6).

Finally, we tested *pri-mir-989* for cleavage by recombinant human Microprocessor. Curiously, even though miR-989 was not substantially matured in HEK293T cells (Figure 5B), and its length far exceeds the longest canonical miRNA hairpin in vertebrates (Figure 1A), it was reasonably processed by human Microprocessor *in vitro* (Figure 7G compared to Figure 5D for other substrates). This provided us the opportunity to test whether an internal bulge near the dicing site could facilitate its cleavage. Tests of *pri-mir-989-M2*, which lacks the bulge, indeed showed that it generated *pre-mir-989* far less efficiently than in WT (Figure 7G).

These structure-function data demonstrate how atypical miRNA precursor lengthening in invertebrates is facilitated by maintenance of extended duplex structure and presence of a likely internal Pasha/DGCR8 docking site. These are general features across a range of extended miRNA loci, and an internal hairpin bulge enables cleavage of an extended pri-miRNA hairpin by human Microprocessor (Figure 7G). In summary, specific features enable biogenesis of atypically long miRNA precursors, and suggest that as yet unappreciated small RNA loci may await discovery even in vertebrates.

DISCUSSION

Unanticipated features of small RNA biogenesis revealed through broad evolutionary analyses

While the overall structure of the miRNA biogenesis pathway is conserved across animals, some species handle the details differently. For example, *C. elegans* pri-miRNAs are generally deficient in features that enable efficient biogenesis in human cells.[5] This led to the identification of sequence motifs that enhance miRNA biogenesis and are more prevalent in mammalian miRNA loci. Conversely, the action of *C. elegans* Microprocessor has mechanistic differences from the human Microprocessor.[51,52]

Here, we document that while vertebrate miRNA hairpins are homogenous in length, there was repeated emergence of atypically long miRNA precursors across multiple invertebrate clades. This prompts a more nuanced view on what constitutes an ideal canonical miRNA substrate, which can differ across species. Direct experimental comparisons indicate that some long miRNA precursors are better processed in fly cells than in human cells. Consistent with this, we reported that a short artificial hairpin extension into the terminal loop of human *mir-144* strongly abrogated its biogenesis.[53] Thus, it is not happenstance that long miRNA precursors do not exist in mammals; there is functional restriction of their pre-miRNA lengths.

We showed that an internal bulge near the dicing site is likely a surrogate platform for DGCR8, which normally associates with the terminal hairpin loop. This explains how certain miRNA precursors can extend beyond typical lengths. In fact, early biochemical studies of human Microprocessor showed that it does not require a terminal loop *per se* for substrate recognition and cleavage.[32] That is, two annealed RNA strands that reconstitute pri-miRNA-like structure can be efficiently cleaved by Microprocessor, at least *in vitro*. Nevertheless, our tests using recombinant human Microprocessor directly show that elongated miRNA hairpins are suboptimal substrates. The absence of long canonical miRNA precursors in vertebrates may be rationalized by the propensity of long inverted repeat RNA to activate the interferon response,[54] owing to its similarity to viral dsRNA.[55] Thus, such substrates might be evolutionarily purged in mammals. We note that invertebrates such as planarians, cnidarians, and molluscs harbor an apoptotic response to dsRNA,[56]

(F) *In vitro* cleavage of *pri-mir-989* using S2R+ nuclear lysates. *pre-mir-989* was prepared by *in vitro* transcription in the presence of radioactive nucleotides and used as a size marker. Wild-type *pri-mir-989* is efficiently cropped to yield *pre-mir-989* hairpin, but removal of the internal bulge (M2) abrogated its processing. (G) *In vitro* cleavage of *pri-mir-989* using recombinant human Microprocessor. Wild-type *pri-mir-989* is specifically and effectively cleaved to yield *pre-mir-989*, but the M2 variant lacking the internal bulge is highly compromised for processing.

and these species also appear to lack highly extended miRNA hairpins (Figure 1).

Challenges for complete miRNA annotations from small RNA data

miRNAs are usually annotated from homogenous ~22 nt small RNA duplexes that form 2-nt 3′ overhangs at either end and map within a short inverted repeat.[37,57] These features lie at the heart of computational strategies to identify miRNAs from deep sequencing data,[58] which must accurately distinguish these from heterogeneous small RNAs that fortuitously map to hairpins.[38] A key consideration to complete annotations is the fact that many functional miRNAs derive from alternative biogenesis strategies, which necessitate alternative annotation criteria.[59] We now appreciate many valid reasons why small RNA loci of interest may evade canonical miRNA annotation. These include inappropriate hairpin sizes due to bypass of Microprocessor[60–63] or bypass of Dicer,[64–66] failure of quality control pathways,[67] or unusually high numbers of genomic mappings.[68] In fact, numerous hairpin RNAs in cells are poorly processed into small RNAs, but may still associate with miRNA factors.[69,70] Each of these exceptions requires special procedures for full annotation.

This study concerns the existence of atypically long miRNA precursors, which, although poorly appreciated, were recognized even from early miRNA annotation efforts across diverse species.[26,27,30] Importantly, atypical distances between the miRNA and passenger strands can mask genuine miRNA hairpins from annotation, if inappropriately short genomic regions are submitted for RNA folding. Even current programs designed for the detection of miRNA orthologs[35] do not properly retrieve obvious miRNA orthologs with unusual precursors (e.g., ~1.5 kb between mature and passenger sequences, as with certain lepidopteran *mir-12* orthologs). We recovered these atypical structures in the pursuit of well-conserved miRNAs, but such atypical miRNAs are prone to being overlooked during *de novo* annotation of miRNAs. Although miRNAs in some well-studied organisms were proposed to be complete, e.g., in humans,[71] it is warranted to revisit troves of small RNA data with this broader perspective, especially in invertebrates.

Limitations of the study

This study is largely based on miRNA discovery using MirMachine. Because there are currently no small RNA data available for the vast majority of species analyzed, we decided to filter loci lacking perfect matches to seed regions. Some of these might be genuine, but there could be an error in the genome assembly, an error in the alignment, or it might be an miRNA paralog bearing seed divergence. Our current analysis did not utilize these loci, but future usage of these catalogs of predicted miRNAs would require further investigation of the "filtered" loci. We provide these for further inspection (Tables S3 and S5).

Another limitation of this algorithm is that it does not assign potential miRNA orthologs. Thus, in all of our outputs, we list all members of each family assigned by MirMachine (Tables S3 and S5), but have not attempted to infer specific orthologies among paralog members within each species. While orthologs could be identifiable among closely related species, as we analyzed a broad swath of invertebrate species, 1-to-1 orthologs likely cannot be determined for many expanded miRNA families. While it is beyond the scope of this study to undertake systematic ortholog annotation, it is likely that different family members may exhibit distinct evolutionary length dynamics.

Finally, we do not fully understand how these unusually long miRNAs are processed. For example, it is possible that Microprocessor and/or Dicer have unusual substrate preferences in some species. Alternatively, there might be additional factors and/or features of these long structures that aid biogenesis. Our data strongly support the notion of DGCR8 binding to internal unstructured regions near dicing sites, a scenario that needs direct evaluation.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Eric C. Lai (laie@mskcc.org).

Materials availability

All plasmids generated in this study are available from the lead contact without restriction.

Data and code availability

- This paper analyzes existing, publicly available genomes from GenBank (https://www.ncbi.nlm.nih.gov/genbank/).
- This paper analyzes existing, publicly available small RNA data obtained via miSRA (https://arn.ugr.es/misra/).
- All of the microRNA annotations and their assessed properties are provided within the supplemental information.
- MirMachine code is available at https://github.com/sinanugur/MirMachine and https://mirmachine.org.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We are grateful to Marcin Dziuba, Meghan Duffy, and Lev Yampolsky for samples of *Daphnia magna* genomic DNA and Arnaud Martin for sharing *Heliconius melpomene* tissues used for DNA isolation and information on lepidopteran evolution. We thank Seungjae Lee and David Jee for technical advice; Dieter Ebert for discussion about *Daphnia* species; Michael Hackenberg for help with miSRA; and Marcel Tarbier for technical advice. Work in the K.O. lab was supported by the JSPS Fund for the Promotion of Joint International Research (Returning Researcher Development Research, 17K20145) and the Ohsumi Frontier Science Foundation. Work in the B.F. lab was supported by the Tromsø Forskningsstiftelse starting grant (20_SG_BF "MIRevolution"). L.J.-T. is an Investigator of the Howard Hughes Medical Institute. Work in the E.C.L. lab was supported by the NIH (R01-GM083300 and R01-HD108914) and MSK Core grant P30-CA008748.

AUTHOR CONTRIBUTIONS

Conceptualization, M.-M.J.-z., X.Z., and E.C.L.; methodology, M.-M.J.-z., E.L., and B.F.; investigation, M.-M.J.-z., X.Z., K.O., R.S., A.G., E.L., and B.F.; writing – original draft, M.-M.J.-z. and E.C.L.; writing – review & editing, M.-M.J.-z., X.Z., K.O., R.S., A.G., E.L., L.J.-T., B.F., and E.C.L.; funding acquisition, L.J.-T., B.F., and E.C.L.; supervision, L.J.-T., B.F., and E.C.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

None declared.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Cell lines
  - Genome data
  - Detection of miRNA homologs and sequence processing
  - Small RNA processing and mapping
  - Statistical analysis and visualization
  - Sequence alignment and visualization
  - Phylogenetic relationships
  - Construction of miRNA expression plasmids and miRNA sensor plasmids
  - Northern blotting
  - Luciferase reporter assays
  - miRNA/RNAi factor knockdown tests in S2R + cells
  - *In vitro* cropping assays using *Drosophila* lysates
  - In *vitro* dicing assays using recombinant Dicer-1/Loqs-PB complex
  - Expression and purification of human Microprocessor complex
  - In-vitro transcription (IVT) of pri-miRNAs
  - In-vitro pri-miRNA processing assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116243.

Received: February 12, 2025
Revised: June 16, 2025
Accepted: August 12, 2025

REFERENCES

1. Shang, R., Lee, S., Senavirathne, G., and Lai, E.C. (2023). microRNAs in action: biogenesis, function and regulation. Nat. Rev. Genet. *24*, 816–833. https://doi.org/10.1038/s41576-023-00611-y.
2. Chang, T.C., Pertea, M., Lee, S., Salzberg, S.L., and Mendell, J.T. (2015). Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. Genome Res. *25*, 1401–1409. https://doi.org/10.1101/gr.193607.115.
3. de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Åström, G., Babina, M., Bertin, N., Burroughs, A.M., et al. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat. Biotechnol. *35*, 872–878. https://doi.org/10.1038/nbt.3947.
4. Ruiz-Arroyo, V.M., and Nam, Y. (2022). Dynamic Protein-RNA recognition in primary MicroRNA processing. Curr. Opin. Struct. Biol. *76*, 102442. https://doi.org/10.1016/j.sbi.2022.102442.
5. Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell *152*, 844–858. https://doi.org/10.1016/j.cell.2013.01.031.
6. Garg, A., Shang, R., Cvetanovic, T., Lai, E.C., and Joshua-Tor, L. (2024). The structural landscape of Microprocessor-mediated processing of pri-let-7 miRNAs. Mol. Cell *84*, 4175–4190.e6. https://doi.org/10.1016/j.molcel.2024.09.008.
7. Fang, W., and Bartel, D.P. (2020). MicroRNA Clustering Assists Processing of Suboptimal MicroRNA Hairpins through the Action of the ERH Protein. Mol. Cell *78*, 289–302.e6. https://doi.org/10.1016/j.molcel.2020.01.026.
8. Hutter, K., Lohmuller, M., Jukic, A., Eichin, F., Avci, S., Labi, V., Szabo, T. G., Hoser, S.M., Huttenhofer, A., Villunger, A., and Herzog, S. (2020). SAFB2 Enables the Processing of Suboptimal Stem-Loop Structures in Clustered Primary miRNA Transcripts. Mol. Cell *78*, 876–889.e876. https://doi.org/10.1016/j.molcel.2020.05.011.
9. Zapletal, D., Kubicek, K., Svoboda, P., and Stefl, R. (2023). Dicer structure and function: conserved and evolving features. EMBO Rep. *24*, e57215. https://doi.org/10.15252/embr.202357215.
10. Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat. Rev. Genet. *12*, 19–31. https://doi.org/10.1038/nrg2916.
11. Lai, E.C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. *30*, 363–364. https://doi.org/10.1038/ng865.
12. Lai, E.C., Burks, C., and Posakony, J.W. (1998). The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts. Development *125*, 4077–4088. https://doi.org/10.1242/dev.125.20.4077.
13. Lai, E.C., and Posakony, J.W. (1997). The Bearded box, a novel 3' UTR sequence motif, mediates negative post-transcriptional regulation of *Bearded* and *Enhancer of split* Complex gene expression. Development *124*, 4847–4856. https://doi.org/10.1242/dev.124.23.4847.
14. Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. Cell *75*, 855–862. https://doi.org/10.1016/0092-8674(93)90530-4.
15. Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife *4*, e05005. https://doi.org/10.7554/eLife.05005.
16. McGeary, S.E., Lin, K.S., Shi, C.Y., Pham, T.M., Bisaria, N., Kelley, G.M., and Bartel, D.P. (2019). The biochemical basis of microRNA targeting efficacy. Science *366*, eaav1741. https://doi.org/10.1126/science.aav1741.
17. Shang, R., Baek, S.C., Kim, K., Kim, B., Kim, V.N., and Lai, E.C. (2020). Genomic Clustering Facilitates Nuclear Processing of Suboptimal Pri-miRNA Loci. Mol. Cell *78*, 303–316.e4. https://doi.org/10.1016/j.molcel.2020.02.009.
18. Shang, R., and Lai, E.C. (2023). Parameters of clustered suboptimal miRNA biogenesis. Proc. Natl. Acad. Sci. USA *120*, e2306727120. https://doi.org/10.1073/pnas.2306727120.
19. Zeng, Y., Yi, R., and Cullen, B.R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. EMBO J. *24*, 138–148. https://doi.org/10.1038/sj.emboj.7600491.
20. Baek, S.C., Kim, B., Jang, H., Kim, K., Park, I.S., Min, D.H., and Kim, V.N. (2024). Structural atlas of human primary microRNAs generated by SHAPE-MaP. Mol. Cell *84*, 1158–1172.e6. https://doi.org/10.1016/j.molcel.2024.02.005.
21. Clarke, A.W., Høye, E., Hembrom, A.A., Paynter, V.M., Vinther, J., Wyrożemski, Ł., Biryukova, I., Formaggioni, A., Ovchinnikov, V., Herlyn, H., et al. (2025). MirGeneDB 3.0: improved taxonomic sampling, uniform nomenclature of novel conserved microRNA families and updated covariance models. Nucleic Acids Res. *53*, D116–D128. https://doi.org/10.1093/nar/gkae1094.
22. Newman, M.A., Thomson, J.M., and Hammond, S.M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. RNA *14*, 1539–1549. https://doi.org/10.1261/rna.1155108.
23. Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol. Cell *32*, 276–284. https://doi.org/10.1016/j.molcel.2008.09.014.

24. Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. Science *320*, 97–100. https://doi.org/10.1126/science.1154040.

25. Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E.E., Nitsch, R., and Wulczyn, F.G. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat. Cell Biol. *10*, 987–993. https://doi.org/10.1038/ncb1759.

26. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of Caenorhabditis elegans. Genes Dev. *17*, 991–1008. https://doi.org/10.1101/gad.1074403.

27. Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., and Lai, E.C. (2007). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. *17*, 1850–1864. https://doi.org/10.1101/gr.6597907.

28. Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193–1197. https://doi.org/10.1038/nature07415.

29. Fromm, B., Worren, M.M., Hahn, C., Hovig, E., and Bachmann, L. (2013). Substantial loss of conserved and gain of novel MicroRNA families in flatworms. Mol. Biol. Evol. *30*, 2619–2628. https://doi.org/10.1093/molbev/mst155.

30. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev. *16*, 1616–1626. https://doi.org/10.1101/gad.1004402.

31. Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605–1619. https://doi.org/10.1105/tpc.003210.

32. Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887–901. https://doi.org/10.1016/j.cell.2006.03.043.

33. Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86–89. https://doi.org/10.1038/35040556.

34. Herlyn, H., Hembrom, A.A., Tosar, J.P., Mauer, K.M., Schmidt, H., Dezfuli, B.S., Hankeln, T., Bachmann, L., Sarkies, P., Peterson, K.J., and Fromm, B. (2025). Substantial Hierarchical Reductions of Genetic and Morphological Traits in the Evolution of Rotiferan Parasites. Genome Biol. Evol. *17*, evaf124. https://doi.org/10.1093/gbe/evaf124.

35. Umu, S.U., Paynter, V.M., Trondsen, H., Buschmann, T., Rounge, T.B., Peterson, K.J., and Fromm, B. (2023). Accurate microRNA annotation of animal genomes using trained covariance models of curated microRNA complements in MirMachine. Cell Genom. *3*, 100348. https://doi.org/10.1016/j.xgen.2023.100348.

36. Kim, B.Y., Gellert, H.R., Church, S.H., Suvorov, A., Anderson, S.S., Barmina, O., Beskid, S.G., Comeault, A.A., Crown, K.N., Diamond, S.E., et al. (2024). Single-fly genome assemblies fill major phylogenomic gaps across the Drosophilidae Tree of Life. PLoS Biol. *22*, e3002697. https://doi.org/10.1371/journal.pbio.3002697.

37. Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. (2003). A uniform system for microRNA annotation. RNA *9*, 277–279. https://doi.org/10.1261/rna.2183803.

38. Berezikov, E., Liu, N., Flynt, A.S., Hodges, E., Rooks, M., Hannon, G.J., and Lai, E.C. (2010). Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. Nat. Genet. *42*, 6–10. https://doi.org/10.1038/ng0110-6.

39. Liu, S., Li, D., Li, Q., Zhao, P., Xiang, Z., and Xia, Q. (2010). MicroRNAs of Bombyx mori identified by Solexa sequencing. BMC Genom. *11*, 148. https://doi.org/10.1186/1471-2164-11-148.

40. Zhang, X., Zheng, Y., Jagadeeswaran, G., Ren, R., Sunkar, R., and Jiang, H. (2012). Identification and developmental profiling of conserved and novel microRNAs in Manduca sexta. Insect Biochem. Mol. Biol. *42*, 381–395. https://doi.org/10.1016/j.ibmb.2012.01.006.

41. Surridge, A.K., Lopez-Gomollon, S., Moxon, S., Maroja, L.S., Rathjen, T., Nadeau, N.J., Dalmay, T., and Jiggins, C.D. (2011). Characterisation and expression of microRNAs in developing wings of the neotropical butterfly Heliconius melpomene. BMC Genom. *12*, 62. https://doi.org/10.1186/1471-2164-12-62.

42. Yang, Y., Zhang, Y., Wang, A., Duan, A., Xue, C., Wang, K., Zhao, M., and Zhang, J. (2021). Four MicroRNAs, miR-13b-3p, miR-278-5p, miR-10483-5p, and miR-10485-5p, Mediate Insecticide Tolerance in Spodoptera frugiperda. Front. Genet. *12*, 820778. https://doi.org/10.3389/fgene.2021.820778.

43. Fu, Y., Yang, Y., Zhang, H., Farley, G., Wang, J., Quarles, K.A., Weng, Z., and Zamore, P.D. (2018). The genome of the Hi5 germ cell line from Trichoplusia ni, an agricultural pest and novel model for small RNA biology. eLife *7*, e31628. https://doi.org/10.7554/eLife.31628.

44. Mohammed, J., Flynt, A.S., Panzarino, A.M., Mondal, M.M.H., DeCruz, M., Siepel, A., and Lai, E.C. (2018). Deep experimental profiling of microRNA diversity, deployment, and evolution across the Drosophila genus. Genome Res. *28*, 52–65. https://doi.org/10.1101/gr.226068.117.

45. Vedanayagam, J., Lin, C.J., Papareddy, R., Nodine, M., Flynt, A.S., Wen, J., and Lai, E.C. (2023). Regulatory logic of endogenous RNAi in silencing de novo genomic conflicts. PLoS Genet. *19*, e1010787. https://doi.org/10.1371/journal.pgen.1010787.

46. Wen, J., Duan, H., Bejarano, F., Okamura, K., Fabian, L., Brill, J.A., Bortolamiol-Becet, D., Martin, R., Ruby, J.G., and Lai, E.C. (2015). Adaptive regulation of testis gene expression and control of male fertility by the Drosophila hairpin RNA pathway. Mol. Cell *57*, 165–178. https://doi.org/10.1016/j.molcel.2014.11.025.

47. Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803–806. https://doi.org/10.1038/nature07015.

48. Jouravleva, K., Golovenko, D., Demo, G., Dutcher, R.C., Hall, T.M.T., Zamore, P.D., and Korostelev, A.A. (2022). Structural basis of microRNA biogenesis by Dicer-1 and its partner protein Loqs-PB. Mol. Cell *82*, 4049–4063.e6. https://doi.org/10.1016/j.molcel.2022.09.002.

49. Partin, A.C., Zhang, K., Jeong, B.C., Herrell, E., Li, S., Chiu, W., and Nam, Y. (2020). Cryo-EM Structures of Human Drosha and DGCR8 in Complex with Primary MicroRNA. Mol. Cell *78*, 411–422.e4. https://doi.org/10.1016/j.molcel.2020.02.016.

50. Jin, W., Wang, J., Liu, C.P., Wang, H.W., and Xu, R.M. (2020). Structural Basis for pri-miRNA Recognition by Drosha. Mol. Cell *78*, 423–433.e5. https://doi.org/10.1016/j.molcel.2020.02.024.

51. Nguyen, T.L., Nguyen, T.D., Ngo, M.K., and Nguyen, T.A. (2023). Dissection of the Caenorhabditis elegans Microprocessor. Nucleic Acids Res. *51*, 1512–1527. https://doi.org/10.1093/nar/gkac1170.

52. Nguyen, T.L., Nguyen, T.D., Ngo, M.K., Le, T.N.Y., and Nguyen, T.A. (2023). Noncanonical processing by animal Microprocessor. Mol. Cell *83*, 1810–1826.e8. https://doi.org/10.1016/j.molcel.2023.05.004.

53. Shang, R., Kretov, D.A., Adamson, S.I., Treiber, T., Treiber, N., Vedanayagam, J., Chuang, J.H., Meister, G., Cifuentes, D., and Lai, E.C. (2022). Regulated dicing of pre-mir-144 via reshaping of its terminal loop. Nucleic Acids Res. *50*, 7637–7654. https://doi.org/10.1093/nar/gkac568.

54. Gantier, M.P., Baugh, J.A., and Donnelly, S.C. (2007). Nuclear transcription of long hairpin RNA triggers innate immune responses. J. Interferon Cytokine Res. *27*, 789–797. https://doi.org/10.1089/jir.2006.0152.

55. Hur, S. (2019). Double-Stranded RNA Sensors and Modulators in Innate Immunity. Annu. Rev. Immunol. *37*, 349–375. https://doi.org/10.1146/annurev-immunol-042718-041356.

56. Kozlovski, I., Jaimes-Becerra, A., Sharoni, T., Lewandowska, M., Karmi, O., and Moran, Y. (2024). Induction of apoptosis by double-stranded RNA was present in the last common ancestor of cnidarian and bilaterian animals. PLoS Pathog. *20*, e1012320. https://doi.org/10.1371/journal.ppat.1012320.

57. Fromm, B., Billipp, T., Peck, L.E., Johansen, M., Tarver, J.E., King, B.L., Newcomb, J.M., Sempere, L.F., Flatmark, K., Hovig, E., and Peterson, K.J. (2015). A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu. Rev. Genet. *49*, 213–242. https://doi.org/10.1146/annurev-genet-120213-092023.

58. Friedlander, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. *40*, 37–52. https://doi.org/10.1093/nar/gkr688.

59. Yang, J.S., and Lai, E.C. (2011). Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol. Cell *43*, 892–903. https://doi.org/10.1016/j.molcel.2011.07.024.

60. Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83–86. https://doi.org/10.1038/nature05983.

61. Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89–100. https://doi.org/10.1016/j.cell.2007.06.028.

62. Xie, M., Li, M., Vilborg, A., Lee, N., Shu, M.D., Yartseva, V., Šestan, N., and Steitz, J.A. (2013). Mammalian 5′-capped microRNA precursors that generate a single microRNA. Cell *155*, 1568–1580. https://doi.org/10.1016/j.cell.2013.11.027.

63. Zamudio, J.R., Kelly, T.J., and Sharp, P.A. (2014). Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. Cell *156*, 920–934. https://doi.org/10.1016/j.cell.2014.01.041.

64. Cheloufi, S., Dos Santos, C.O., Chong, M.M.W., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584–589. https://doi.org/10.1038/nature09092.

65. Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D., et al. (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694–1698. https://doi.org/10.1126/science.1190809.

66. Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc. Natl. Acad. Sci. USA *107*, 15163–15168. https://doi.org/10.1073/pnas.1006432107.

67. Hasler, D., Lehmann, G., Murakawa, Y., Klironomos, F., Jakob, L., Grässer, F.A., Rajewsky, N., Landthaler, M., and Meister, G. (2016). The Lupus Autoantigen La Prevents Mis-channeling of tRNA Fragments into the Human MicroRNA Pathway. Mol. Cell *63*, 110–124. https://doi.org/10.1016/j.molcel.2016.05.026.

68. Chak, L.L., Mohammed, J., Lai, E.C., Tucker-Kellogg, G., and Okamura, K. (2015). A deeply conserved, noncanonical miRNA hosted by ribosomal DNA. RNA *21*, 375–384. https://doi.org/10.1261/rna.049098.114.

69. Lee, S., Jee, D., Srivastava, S., Yang, A., Ramidi, A., Shang, R., Bortolamiol-Becet, D., Pfeffer, S., Gu, S., Wen, J., and Lai, E.C. (2023). Promiscuous splicing-derived hairpins are dominant substrates of tailing-mediated defense of miRNA biogenesis in mammals. Cell Rep. *42*, 112111. https://doi.org/10.1016/j.celrep.2023.112111.

70. Yao, J., Xu, H., Ferrick-Kiddie, E.A., Nottingham, R.M., Wu, D.C., Ares, M., Jr., and Lambowitz, A.M. (2024). Human cells contain myriad excised linear intron RNAs with links to gene regulation and potential utility as biomarkers. PLoS Genet. *20*, e1011416. https://doi.org/10.1371/journal.pgen.1011416.

71. Fromm, B., Zhong, X., Tarbier, M., Friedländer, M.R., and Hackenberg, M. (2022). The limits of human microRNA annotation have been met. RNA *28*, 781–785. https://doi.org/10.1261/rna.079098.122.

72. Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., and Karsch-Mizrachi, I. (2019). GenBank. Nucleic Acids Res. *47*, D94–D99. https://doi.org/10.1093/nar/gky989.

73. Gruber, A.R., Bernhart, S.H., and Lorenz, R. (2015). The ViennaRNA web services. Methods Mol. Biol. *1269*, 307–326. https://doi.org/10.1007/978-1-4939-2291-8_19.

74. Kodama, Y., Shumway, M., and Leinonen, R.; International Nucleotide Sequence Database Collaboration (2012). The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res. *40*, D54–D56. https://doi.org/10.1093/nar/gkr854.

75. Kang, W., Eldfjell, Y., Fromm, B., Estivill, X., Biryukova, I., and Friedländer, M.R. (2018). miRTrace reveals the organismal origins of microRNA sequencing data. Genome Biol. *19*, 213. https://doi.org/10.1186/s13059-018-1588-9.

76. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

77. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

78. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

79. Ito, K., and Murphy, D. (2013). Application of ggplot2 to Pharmacometric Graphics. CPT Pharmacometrics Syst. Pharmacol. *2*, e79. https://doi.org/10.1038/psp.2013.56.

80. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

81. Procter, J.B., Carstairs, G.M., Soares, B., Mourão, K., Ofoegbu, T.C., Barton, D., Lui, L., Menard, A., Sherstnev, N., Roldan-Martinez, D., et al. (2021). Alignment of Biological Sequences with Jalview. Methods Mol. Biol. *2231*, 203–224. https://doi.org/10.1007/978-1-0716-1036-7_13.

82. Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics *30*, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531.

83. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. *33*, 1635–1638. https://doi.org/10.1093/molbev/msw046.

84. Federhen, S. (2012). The NCBI Taxonomy database. Nucleic Acids Res. *40*, D136–D143. https://doi.org/10.1093/nar/gkr1178.

85. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

86. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

87. Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287–297. https://doi.org/10.1016/j.cell.2007.05.056.

88. Bejarano, F., Bortolamiol-Becet, D., Dai, Q., Sun, K., Saj, A., Chou, Y.T., Raleigh, D.R., Kim, K., Ni, J.Q., Duan, H., et al. (2012). A

genome-wide transgenic resource for conditional expression of Drosophila microRNAs. Development *139*, 2821–2831. https://doi.org/10.1242/dev.079939.

89. Ishizuka, A., Saito, K., Siomi, M.C., and Siomi, H. (2006). In vitro precursor microRNA processing assays using Drosophila Schneider-2 cell lysates. Methods Mol. Biol. *342*, 277–286. https://doi.org/10.1385/1-59745-123-1:277.

90. Bortolamiol-Becet, D., Hu, F., Jee, D., Wen, J., Okamura, K., Lin, C.J., Ameres, S.L., and Lai, E.C. (2015). Selective Suppression of the Splicing-Mediated MicroRNA Pathway by the Terminal Uridyltransferase Tailor. Mol. Cell *59*, 217–228. https://doi.org/10.1016/j.molcel.2015.05.034.

91. Berger, I., Fitzgerald, D.J., and Richmond, T.J. (2004). Baculovirus expression system for heterologous multiprotein complexes. Nat. Biotechnol. *22*, 1583–1587. https://doi.org/10.1038/nbt1036.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
TOP10 Chemically Competent E. coli	Our lab	N/A
E.coli DH5α cells	ThermoFisher Scientific	Cat#18258-012
E.coli DH10Bac cells	ThermoFisher Scientific	Cat#12033-015
Biological samples		
Heliconius melpomene pupal tissue	Arnaud Martin	N/A
Daphnia magna genomic DNA	Marcin Dziuba and Meghan Duffy; Lev Yampolsky	N/A
Chemicals, peptides, and recombinant proteins		
ATP, [γ-P32]	PerkinElmer	Cat#BLU502Z250UC
UTP, [α-P32]	PerkinElmer	Cat#BLU007H250UC
Penicillin-Streptomycin	Life Technologies	Cat#15070063
Trizol reagent	Life Technologies	Cat#15596018
Lipofectamine 2000	Thermo Fisher	Cat#11668030
T4 Polynucleotide Kinase	NEB	Cat#M0201L
T4 DNA Ligase	NEB	Cat#M0202L
Gel Loading Buffer II	Invitrogen	Cat#AM8547
Strep-Tactin Superflow high-capacity resin	IBA Lifesciences	Cat# 2-1208-025
5-aminolevulinic acid hydrochloride	Sigma	Cat#A378505G
d-Desthiobiotin	Millipore-Sigma	Cat#D1411-1G
RNase-free DNase I	NEB	Cat#M0303S
RNase-free water	Sigma	Cat#W3513
Amicon Ultra centrifugal filter (30kDa MWCO)	Millipore-Sigma	Cat#903096
Proteinase K	NEB	Cat#p8107S
SYBR Gold Nucleic acid gel stain	ThermoFisher Scientific	Cat#S11494
HiTrap SP-HP Column	Cytiva	Cat#17115201
Superose 6 increase 10/300 column	Cytiva	Cat#29091596
Opti-MEM Reduced Serum Medium	Thermo Fisher	Cat#51985034
HyClone CCM3 media	Cytiva	Cat#SH30065.02
ESF921 cell culture media	Expression Systems	Cat#96-001-01
Critical commercial assays		
Dual Glo luciferase assay system	Promega	Cat#E2940
MEGAscript™ T7 Transcription Kit	Thermo Fisher	Cat# AM1334
Decade marker system	Thermo Fisher	Cat#AM7778
Phusion™ High-Fidelity DNA Polymerase	NEB	Cat# M0530
SequaGel UreaGel System	National Diagnostics	Cat#EC-833
Experimental models: Cell lines		
HEK293T cells	Our lab	N/A
S2R + cells	Our lab	N/A
Sf9 insect cells	ThermoFisher Scientific	Cat#11496015
High Five insect cells	ThermoFisher Scientific	Cat#B855502
Oligonucleotides		
Oligonucleotides for plasmid construction, see Table S7	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides for *in vitro* transcription, see Table S7	This study	N/A
Oligonucleotides for Northern blotting probes, see Table S7	This study	N/A
Pri-miRNA sequences used in *in-vitro* processing assays, see Table S7	This study	N/A
Recombinant DNA		
All the miRNA expression plasmids, see Table S7	This study	N/A
All the miRNA sensor plasmids, see Table S7	This study	N/A
pFL-Twin-Strep-Drosha$^{317\text{-}1337}$	Garg et al. (Mol Cell, 2024)	N/A
pSPL-6xHis-DGCR8$^{175-751}$	Garg et al. (Mol Cell, 2024)	N/A
pRSF-T7-pri-let-7a1 (human)	This study	N/A
pRSF-T7-pri-miR-375 (human)	This study	N/A
pRSF-T7-pri-let-7 (*Daphnia magna*)	This study	N/A
pRSF-T7-pri-miR-375 (*Daphnia magna*)	This study	N/A
pRSF-T7-pri-miR-989 WT (*Drosophila melanogaster*)	This study	N/A
pRSF-T7-pri-miR-989 M2 (*Drosophila melanogaster*)	This study	N/A
Other		
Illustra MicroSpin G-25 columns	GE Healthcare	Cat#27532501
GeneScreen Plus hybridization transfer membrane	PerkinElmer	Cat#NEF1017001PK
Stericup-GP Sterile Vacuum Filtration System	Millipore	Cat#SCGPU05RE

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines: male *Drosophila* S2R + cells and female human HEK293T cells. We do not anticipate sex-specific differences since the miRNAs studied are equally expressed in both sexes in their respective cognate species. Human cells were checked for mycoplasma biannually.

METHOD DETAILS

Cell lines

HEK293T cells were grown in DME-high glucose media containing 10% FBS, 1% non-essential amino acids, 1% sodium pyruvate, penicillin/streptomycin and 0.1% 2-mercaptoethanol. S2R + cells were grown in Schneider's Insect medium with 10% FBS at 25°C. Cells were regularly tested for potential mycoplasma contamination.

Genome data

We obtained genomes for 421 Diptera, 546 Lepidoptera, and 128 Crustacea, from the NCBI GenBank database.[72] We used all available crustacean genomes, and selected representatives from across the genus classifications of Diptera and Lepidoptera. We also utilized a collection of 319 nanopore genomes of the Drosophilidae (fruitfly) family.[36] These genomes formed the basis for subsequent analyses.

Detection of miRNA homologs and sequence processing

For initial summaries of miRNA properties, we used well-curated loci included in the MirGeneDB 3.0 database, which is based on manual annotation of small RNA data.[21] Since most available genomes lack corresponding small RNA data, we took advantage of the MirMachine (v0.2.13) pipeline, which uses covariance models of all known conserved miRNAs in MirGeneDB to identify their putative homologs directly from genomes.[35] The complete set of sequences is available in Table S3. Bash scripts were employed to compile these sequences into a comprehensive master FASTA file, which included all miRNA hairpins with their respective flanking regions and other pertinent sequence properties.

Although only high confidence predictions were utilized, inspection showed that certain improbable hits were returned. Accordingly, we filtered MirMachine outputs using these criteria: (1) the sequence hit bears a perfect match to the miRNA family seed queried, (2) a minimum length sequence hit returned of 106 nt (which includes 30 nt upstream and downstream flanking sequence), (3) the sequence hit bears no more than 32 nt perfect duplex. The flagged loci are output on separate tables, along with information on relevant criteria (Table S3). Note that many of these flagged loci might be genuine miRNAs, but to be conservative, these loci were not used in further analyses of long miRNA structures.

Some miRNAs are members of large paralogous families (e.g., let-7, Figure 1B). In fact, certain conserved miRNAs can exist in >100 copies, but such unusual loci require functional validation.[68] We noticed that some species harbor a miRNA family with dozens to hundreds of putative copies. Since these might potentially be artifacts in genome assembly, we summarized them in Table S4, so that appropriate caution can be taken to those who might study these further.

To analyze miRNA loops, the lengths of Dicer-excised regions were approximated by removing 53 nucleotides from each end of the MirMachine outputs. Note that the Dicer-excised regions are inferred, and not based on small RNA data, but validation tests indicate that most lengths were within a few nucleotides (Table S2). We classified long loops as ones that exceed all annotated canonical vertebrate miRNAs (excepting cloudy catshark Sto-Let-7-P2b3, i.e., loops $\geq$42 nt). We also predicted the secondary structures of miRNA hairpins and minimum free energy (MFE) values using RNAfold (v2.6.4).[73] The filtered dataset is available as Table S5, and includes all possible homologs for each family (long and short), as well as their identified hairpin precursors and their inferred Dicer-excised regions.

We observed that MirMachine did not recover *mir-12* homologs in most species queried from the Lepidoptera clade. Since *Heliconius melpomene* (butterfly) *mir-12* had one of the longest precursors across animals,[21] we conducted a BLAST-short search of all Lepidopterans using mature *Heliconius melpomene* miR-12 (TGAGTATTACTTCAGGTACTGG), which exhibits a single nucleotide divergence from *Drosophila melanogaster* miR-12. We retrieved BLAST hits with 30 nt upstream and successively larger downstream flanks (500, 1kb, and then 5kb).These hits underwent two filtering steps: first, we retained only sequences containing the *Heliconius melpomene* miR-12 seed sequence (GAGTATT), and second, we filtered for sequences that matched the *Heliconius melpomene* miR-12 star sequence (CAGTGACTGAATAATACTTTG) allowing up to 8 mismatches. The filtered sequences were then submitted to RNAfold to identify duplex pairing that would indicate mature/passenger strand pairing and adjacent lower stem pairing.

Small RNA processing and mapping

To identify relevant Lepidopteran small RNA data, we used the miSRA tool (https://zenodo.org/records/13925083) to screen the SRA repository for existing small RNA sequencing data. Reads were either downloaded directly from miSRA, or as raw reads from SRA[74] and processed using miRTrace.[75] small RNA reads were then mapped to extended precursor sequences using bowtie[76] and visualized using samtools[77] and IGV.[78]

Statistical analysis and visualization

Statistical analyses and visualizations were performed using the R programming language, with the ggplot2 library[79] used to generate scatterplots and boxplots. These visualizations facilitated the exploration of miRNA sequence length distributions and structural characteristics.

Sequence alignment and visualization

miRNA sequence alignments were performed using MAFFT (v7.526)[80] with the –auto parameter to enable automated model selection. The resulting alignments were visualized using JALview[81] and AliView.[82]

Phylogenetic relationships

We used taxonomy IDs from NCBI and retrieved taxonomic information using the NCBITaxa module of ETE3,[83] which interfaces with the NCBI Taxonomy Database.[84] Phylogenetic relationships were inferred using IQ-TREE2 (v2.3.5)[85] with the -s parameter, based on the alignment results of selected miRNA. Phylogenetic trees were constructed and visualized using the Interactive Tree of Life (iTOL) platform[86] for comprehensive and interactive exploration.

Construction of miRNA expression plasmids and miRNA sensor plasmids

Drosophila plasmids. We used PCR to amplify miRNA hairpins with ~300 flanking nts on either side from genomic DNA of the appropriate species (*Drosophila melanogaster*, *Heliconius melpomene*, *Daphnia magna* or *Homo sapiens*), and cloned these into the 3′ UTR of a pUAST-DsRed vector using the NotI and XbaI sites. The series of *D. melanogaster mir-989* variants was constructed using overlapping PCR and appropriate oligonucleotides, using the wild-type *mir-989* construct as template. Luciferase sensors containing two perfect antisense matches to a cognate miRNA were constructed by inserting annealed DNA oligonucleotides into the 3′ UTR of an actin-firefly luciferase plasmid, between the NotI and XhoI sites. All details and oligonucleotide sequences used to clone these constructs are listed in Table S7.

Mammalian plasmids. Plasmids for miRNA expression in mammalian cells were constructed by cloning miRNA fragments amplified as above, into Bgl II and Xho I sites downstream of the CMV promoter. Luciferase reporters were constructed by inserting

annealed DNA oligonucleotides bearing two antisense miRNA matches into the 3′ UTR of the firefly luciferase gene, between NotI and XhoI sites. All the details and oligonucleotide sequences used to clone these constructs are listed in Table S7.

Northern blotting

S2R + cell samples. Co-transfection of miRNA plasmids (2 μg/well for 6-well plate or 1 μg/well for 12-well plate), Ub-GAL4 plasmids (400 ng/well for 6-well plate or 200 ng/well for 12-well plate) and control *mir-279* plasmid (200 ng/well for 6-well plate or 100 ng/well for 12-well plate) were performed in S2R + cells using Effectene Transfection reagent (Qiagen). Three days post-transfection, total RNA was prepared using Trizol reagent (Invitrogen).

HEK293T cell samples. Co-transfection of test miRNA plasmids (2 μg/well for 6-well plate or 1 μg/well for 12-well plate) with control *mir-144* plasmid[53] (200 ng/well for 6-well plate or 100 ng/well for 12-well plate) were performed in HEK293T cells using Lipofectamine 2000. Two or three days post-transfection, total RNA was prepared using Trizol reagent (Invitrogen).

Equal amounts of total RNAs (20 μg) were mixed with 2x RNA loading dye, denatured at 95°C for 5 min, and then fractionated on a 20% urea polyacrylamide gel in 0.5x TBE buffer, until the bromophenol blue dye migrated out of the gel. Then, the gel was transferred to GeneScreen Plus nylon membrane (PerkinElmer) at 300 mA for 1.5 h, UV-crosslinked with 120,000 μJ of energy, and baked at 80°C for 30 min. The blot was hybridized with γ-^{32}P-labeled DNA probes against mature miRNA sequences in hybridization buffer (5x SSC, 7% SDS, 2x Denhardt's solution) at 42°C overnight. The membrane was washed with Non-Stringent Wash Solution (3x SSC, 5% SDS, 10x Denhardt's solution) followed by two rounds with Stringent Wash Solution (1x SSC, 1% SDS). Each wash step is conducted at 42°C for 30 min. The membrane was then sealed in plastic wrap and exposed to a phosphorimager cassette for 1–3 days. For re-probing, the blot was stripped in 1% SDS at 80°C for 30 min before hybridization with the next probe. All the probe sequences are listed in Table S7.

Luciferase reporter assays

Transient co-transfections of S2R + cells were performed in 24-well cell culture plates using Effectene Transfection reagent (Qiagen) according to the manufacturer's protocol. We typically used miRNA plasmids (150–200 ng/well), Ub-Gal4 plasmids (40 ng/well), firefly luciferase plasmids containing miRNA sensors (15 ng/well) and control renilla luciferase plasmids.

Transient co-transfections of HEK293T cells with miRNA expressing plasmids (150–200 ng/well), firefly luciferase plasmids containing miRNA sensors (15 ng/well) and control renilla luciferase plasmids were performed in 24-well cell culture plates using Lipofectamine2000 (Thermo Fisher) according to the manufacturer's protocol. Cells were lysed 24 h post-transfection using 70 μL/well lysis buffer (PBS with 0.2% Triton X-100), and then 10 μL of the cell lysates were used to measure firefly and renilla luciferase activities using the Dual-Glo luciferase assay system (Promega) and a Cytation 5 luminometer plate reader.

miRNA/RNAi factor knockdown tests in S2R + cells

We treated S2R + cells expressing GFP[87] with dsRNAs against cognate miRNA and RNAi factors, as described.[47] We soaked 2-3x10^6 S2R + cells with 15 μg dsRNA in 1 mL serum-free medium for 30 min, and then 1 mL of 20% FBS containing medium was added. After incubating cells for 4 days, cells were transfected with 500 ng pUAST-DsRed-miRNA and 250ng Ub-Gal4 plasmids using Effectene (Qiagen) according to the manufacturer's instructions on a 6-well plate. 15 μg of the same dsRNA as the initial RNAi soaking was added to the culture 8 h after the plasmid transfection. Total RNA was extracted from cells 4 days later by Trizol (Thermo) and used for Northern blotting as described.[47]

In vitro cropping assays using *Drosophila* lysates

To prepare the plasmid encoding *Dme-pri-mir-989* for *in vitro* transcription, the NotI-XhoI fragment of *pDsRed-mir-989*[88] was inserted in the NotI-XhoI site of pBluescript II (pBS). Site-directed mutagenesis was used to remove the internal bulge of the *mir-989* hairpin by using the miR989_noBulge_sense and miR989_NoBulge_3 primers (Table S7). The PCR products using the M13 Forward and Reverse primers with the *pBS-mir-989* (wild-type or no bulge mutant) template were directly used for *in vitro* transcription using the Megascript T7 kit in the presence of 2 μL α-32P-UTP (3000 μCi/mmol, 10mCi/ml EasyTide, PerkinElmer) and 1 μL of 6mM cold UTP (Thermo). ATP, GTP, CTP, T7 polymerase and 10x buffer were added according to the manufacturer's instructions. The resulting labeled transcripts were acrylamide gel-purified following DNase treatment, and 1x10^4 cpm was used for each 10 μL *in vitro* cropping reaction containing 5 μL nuclear extract, 1 μL 3.2mM $MgCl_2$, 1 μL 5mM ATP and 0.5 μL RNase Out (Promega). Nuclear extracts were isolated from S2R + cells as described.[89]

In *vitro* dicing assays using recombinant Dicer-1/Loqs-PB complex

The PCR templates for pre-miRNA production were prepared using the primer pairs of T7-pre-miR-14/pre-miR-14_B, T7_miR989_WT5/miR989_WT3 and T7_miR989_WT5/miR989_NoBulge_3, using no template (for *pre-mir-14*) or pBS-mir-989 plasmids as PCR templates (Table S7). The template was used for *in vitro* transcription using the Megascript T7 kit, and then the 5′-triphosphate group was removed by Calf Intestinal Phosphatase (NEB). The resulting RNA was purified by Phenol/Chloroform/Isoamyl alcohol extraction and Ethanol precipitation, and then 1pmol of pre-miRNAs were 5′ labeled using Polynucleotide kinase (NEB) with 1 μL γ-^{32}P-ATP (3000 μCi/mmol, 10 mCi/mL EasyTide, PerkinElmer) in 10 μL reactions. The labeled pre-miRNAs were gel-purified and

dicing was carried out by incubating 1 nmol of purified pre-miRNA with 0.4 pmol recombinant Dicer-1/Loqs complex in a 20 μL reaction for 1 h at room temperature, using conditions described previously.[90]

Expression and purification of human Microprocessor complex

Human Microprocessor complex was expressed and purified using the MultiBac baculovirus expression system in insect cells (Sf9 or HighFive) as described.[6] Baculovirus generated in Sf9 cells was maintained in HyClone CCM3 Cell Culture Media (Cytiva), while HighFive cells were cultured in ESF921 media (Expression systems). In brief, Drosha isoform 4 (Drosha$^{317\text{-}1337}$) with an N-terminal Twin-strep tag was cloned into a pFL vector while DGCR8 (DGCR8$^{175-751}$) with an N-terminal 6xHis tag was cloned into a pSPL plasmid. Drosha and DGCR8 clones were Cre-fused and expressed in either Sf9 or HighFive cells using the MultiBac baculovirus expression system.[91] The insect cells were infected with baculovirus at 27°C for 60 h, and supplemented with 0.75 mM 5-aminolevulinic acid (5-ALA) during protein expression. Insect cells were harvested and lysed in lysis buffer (50 mM Tris pH 8.0, 750 mM NaCl, 5 mM DTT, and 10% glycerol) supplemented with a protease inhibitor (PI) mix. The Microprocessor was affinity purified using Strep-Tactin Superflow beads (IBA Lifesciences) and eluted with 7 mM desthiobiotin in 25 mM HEPES pH 7.5, 200 mM NaCl, 5 mM DTT, 10% glycerol. The Microprocessor was further purified using a HiTrap SP HP column (Cytiva) (in 25 mM Tris pH 6.8, 75–1000 mM NaCl, 5 mM DTT and 10% glycerol) and over a Superose 6 increase 10/300 size-exclusion column (Cytiva) (in 25 mM HEPES pH 7.5, 400 mM NaCl, 5 mM DTT, and 10% glycerol). Purified protein was flash frozen in liquid N_2 and stored at −80°C.

In-vitro transcription (IVT) of pri-miRNAs

The cDNAs encoding different pri-miRNAs were subcloned into a pRSF plasmid, downstream to the T7 promoter (Table S7). *In vitro* transcription reactions (1.5 mL volume) were set up using in-house purified T7 RNA polymerase. After transcription, DNA templates were digested using RNase-free DNase (NEB), and RNA was purified using phenol-chloroform extraction and isopropanol precipitation. The RNA pellet was dissolved in TE buffer, concentrated, and buffer exchanged into RNase-free water using Amicon Ultra-14 (30 kDa MWCO) and stored at −80°C.

In-vitro pri-miRNA processing assays

The pri-miRNA processing assays were performed as described.[6] In brief, ∼50 nM of pri-miRNA substrate was incubated with 500 nM MP in processing assay buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 2.5 mM DTT, 5 mM $MgCl_2$, 1 U/ml RNasin and 10% glycerol) at 30°C. 15 μL aliquots were removed at 0, 0.5′, 1′, 2′, 5′, 10′, 20′, 45′ time points and quenched with 3 μL of 5X stop buffer (9% SDS and 50 mM EDTA). The samples were treated with 2.5U Proteinase K (NEB) before analysis onto a 10% denaturing urea-PAGE (20 cm × 20 cm) (National Diagnostics). Gels were stained with SybrGold and visualized on a BioRad ChemiDoc imaging system. For all assays $n = 3$.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of the luciferase sensor tests were performed using GraphPad Prism version 9.0.0, and results are presented as the mean ± SEM. Statistical variances were assessed using unpaired Student's t-tests for pairwise group comparisons and one-way analysis of variance (ANOVA) with Bonferroni's post hoc tests for multiple group comparisons. Figure 5B: $N = 2$ biological replicates. Figure 5C: $N = 3$ biological replicates; ns, not significant; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Figures 7B and 7C: $N = 2$, biological replicates. Figures 7D and 7E: $N = 3$ biological replicates; ns, not significant; $^{*}p < 0.05$; $^{***}p < 0.001$.

Boxplots were generated using the ggplot2 package (v3.5.2) in R (v4.5.1). The box indicates the interquartile range (Q1 and Q3) and the box line indicates the median, the whiskers indicate values to 1.5x the IQR, and outliers are shown as discrete points. Boxplots are shown in Figures 1A, 2B, and S3–S5. For linear regression analysis, the lm() function in R was used to fit a least-squares regression model, and the coefficient of determination (R2) was calculated. Linear regression was applied to Figures 4 and S7.