Clusters of microRNAs emerge by new hairpins in existing transcripts

Antonio Marco*, Maria Ninova, Matthew Ronshaugen and Sam Griffiths-Jones

Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK

ABSTRACT

Genetic linkage may result in the expression of multiple products from a polycistronic transcript, under the control of a single promoter. In animals, protein-coding polycistronic transcripts are rare. However, microRNAs are frequently clustered in the genomes of animals, and these clusters are often transcribed as a single unit. The evolution of microRNA clusters has been the subject of much speculation, and a selective advantage of clusters of functionally related microRNAs is often proposed. However, the origin of microRNA clusters has not been so far explored. Here, we study the evolution of microRNA clusters in *Drosophila melanogaster.* We observed that the majority of microRNA clusters arose by the *de novo* formation of new microRNA-like hairpins in existing microRNA transcripts. Some clusters also emerged by tandem duplication of a single microRNA. Comparative genomics show that these clusters are unlikely to split or undergo rearrangements. We did not find any instances of clusters appearing by rearrangement of pre-existing microRNA genes. We propose a model for microRNA cluster evolution in which selection over one of the microRNAs in the cluster interferes with the evolution of the other linked microRNAs. Our analysis suggests that the study of microRNAs and small RNAs must consider linkage associations.

INTRODUCTION

MicroRNAs are small endogenous RNA sequences involved in the regulation of essentially all biological processes in animals and plants (1–3). MicroRNAs are produced from longer transcripts by the RNA interference machinery [reviewed in (4,5)]. A striking feature of these molecules is that their loci are often clustered in the genome (6–8). According to miRBase (9), >30% of animal microRNAs are organized into clusters, some of which have been experimentally shown to produce polycistronic transcripts (10–12). Hence, multiple microRNAs can be produced from the same primary transcript. Further studies including microRNA co-expression and primary transcript identification suggest that the majority of microRNA clusters are transcribed as a single unit (13–16).

The evolutionary importance of microRNA clusters has been the subject of much speculation. Many clusters contain members of the same family, suggesting an important role of gene duplication in their evolution (17,18). However, clusters often contain members of different microRNA families, particularly in animal genomes [reviewed in (1)]. As co-transcription is often used to imply a functional relationship, unrelated microRNAs in the same cluster are often assumed to have similar targeting properties, for example targeting genes in the same pathway (19). However, the origin and evolution of microRNA clusters has not been investigated in detail.

There are a number of known types of polycistronic transcripts, each of which suggests a possible mode of evolution for polycistronic microRNAs. Bacterial operons are formed by multiple protein coding loci under the control of a single promoter. These loci are transcribed as a single transcriptional unit and then the different open reading frames are translated separately by the ribosome (20). The evolutionary origin of bacterial operons has been extensively debated, and several models of evolution have been proposed (21). A common feature of the many models is that genes in the same operon are functionally related, i.e. participate in the same biochemical pathway (21,22). We define this general model as the 'put together' model, which suggests that functionally related products become regulated under a common promoter during evolution (Figure 1A). Under this hypothesis, evolutionarily unrelated microRNAs scattered around the genome may become clustered together during evolution. This mode of

---

*To whom correspondence should be addressed. Tel: +44 1612 751565; Fax: +44 1612 755082; Email: amarco.bio@gmail.com
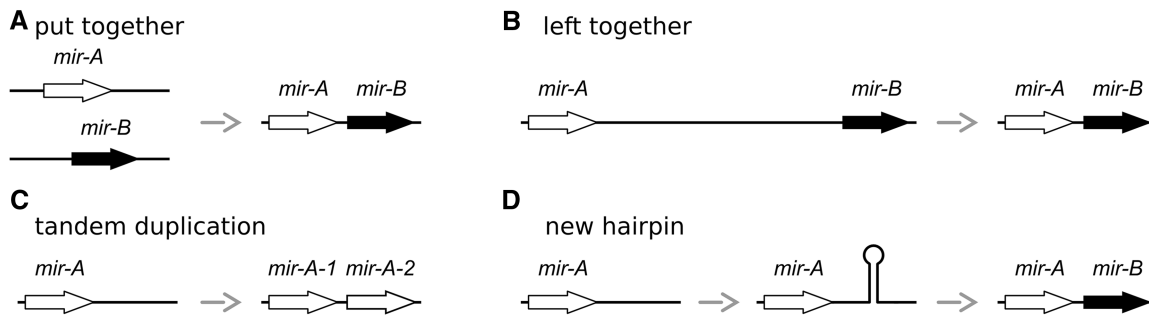
Figure 1. Mechanisms of microRNA cluster emergence. (**A**) Put together: microRNAs in different genomic loci involved in related functional pathways end up being clustered in the genome. (**B**) Left together: microRNAs in different genomic loci become clustered in the genome as a by-product of genome rearrangements. (**C**) Tandem duplication: a microRNA is duplicated in tandem producing a polycistronic transcript. (**D**) New hairpin: a novel microRNA emerges within the primary transcript of an existing microRNA.

evolution has been suggested to explain the existence of clusters of microRNAs from different families (19). Operons have also been found in some animals, particularly in the nematode *Caenorhabditis elegans* (23) and the ascidian *Ciona intestinalis* (24). Operon formation in nematodes is found to be a one-way phenomenon owing to molecular constraints (23). Comparative genomics analysis of *C. elegans* and related species reveals that their operons appeared as a by-product of genome reduction, leaving unrelated genes under the control of a single promoter (25,26). We define this mechanism as the '*left together*' model (Figure 1B), under which microRNAs would be organized into clusters as a stochastic by-product of genome reorganization. More recently, polycistronic transcripts encoding small peptides have been found in arthropods (27). For example, the gene *mille-pattes* is an essential gene during early development and codes for a number of small peptides (27). As these peptides are similar in sequence, an origin of polycistronic transcription by tandem gene duplication is plausible. MicroRNA cluster formation by gene duplication has been observed in animals (17) and probably dominates the evolution of plant microRNA clusters (18,28). This is the '*tandem duplication*' model (Figure 1C).

However, a fourth mechanism of cluster formation is possible in the case of microRNAs. Any transcript with a hairpin structure is potentially a target of the RNases Drosha and Dicer. The cleavage of a precursor microRNA is largely independent of its specific nucleotide sequence (29,30). Thus, many transcribed hairpins in the genome are potential targets of Drosha and Dicer. Indeed, microRNAs arise *de novo* in the genome at a high rate (31,32). Hence, it is plausible that the emergence of a new hairpin near to an existing microRNA could lead to formation of a microRNA cluster, as has been suggested for the vertebrate mir-17 cluster, for example (33). We call this the '*new hairpin*' model (Figure 1D).

The evolutionary origin of microRNA clusters has not been systematically studied. We explore in this article the source of all *Drosophila melanogaster* clusters by tracing the evolution of their microRNAs and evaluate the relative contribution of the different microRNA cluster formation models.

MATERIALS AND METHODS

MicroRNA sequences, genomic coordinates and expression data sets for *D. melanogaster* were extracted from miRBase version 18 (9). We define a cluster of microRNAs as a group of microRNA precursors with an inter-microRNA distance of <10 kb on the same genomic strand. The degree of co-expression of clustered microRNAs was calculated as the Pearson correlation coefficient of the absolute read counts between all tissues/developmental stages from available RNAseq experiments. We compile homologous microRNAs in animals from miRBase microRNA family annotation, and from BLAST searches (34) with parameters: w = 4, r = 2, q = −3, against multiple genome sequences (Supplementary Table S1). We also included in our analysis the microRNA families described by Sempere, Wheeler and collaborators (35,36). We aligned sequences with Clustal X 2.0 (37) and MAFFT 6.85 (38), manually refined the alignments with RALEE (39) and reconstructed evolutionary trees with standard phylogenetic methods: neighbor-joining (40) and maximum likelihood (41), using MEGA5 (42).

To determine the evolutionary origin of each cluster, we first determined the age of each of the microRNAs in the cluster by analysing sequence alignments and phylogenetic trees of microRNA families (Supplementary Data Set S2). We then identified the two original (oldest) microRNAs and examined the nature of the event that led to these two microRNAs to be clustered together. If the two oldest members of a cluster belong to the same microRNA family, we inferred that the cluster emerged by tandem duplication (Figure 1C). Otherwise, the cluster was formed by one of the other models (Figure 1A, B and D). If the two original microRNAs derive from disparate loci in any other genome, the cluster may have originated by a fusion event. Otherwise, if the two original microRNAs always appear together, we conclude that the cluster was formed by *de novo* emergence of a novel microRNA family. Multiple sequence alignments of related microRNAs are available in the supporting information Supplementary Data Set S2. MicroRNA expression data sets are detailed in Supplementary Table S2.

RESULTS

MicroRNA clusters in *Drosophila melanogaster*

We have studied the genomic distribution and evolutionary origin of 238 *D. melanogaster* microRNAs (see 'Materials and Methods' section). These microRNAs are highly clustered in the genome, with 74 (31%) of the annotated sequences <10 kb away from another microRNA. Analysis of expression data from different tissues/developmental stages shows that, on average, microRNAs separated by <10 kb are highly co-expressed (Figure 2A). The median distance between two clustered microRNAs is only ~130 nt, indicating that clustered microRNAs are, in general, tightly linked in the genome. This observation is in agreement with previous analysis on a more limited data set (43) and supports 10 kb as an appropriate global threshold for defining clusters of microRNAs that are co-expressed. These clusters are most likely produced from single primary transcripts under the control of a single promoter (16). Using this criterion, we defined 21 *Drosophila* microRNA clusters (Table 1).

The number of microRNAs in each cluster is variable, although the majority are small: of size 2–3 (Figure 2B; white boxes). The distribution of the number of different microRNA families in the same cluster shows that only 4 of the 21 clusters are formed by a single family (Figure 2B; black boxes). We plotted the size of each cluster against the number of families and observed that clusters of sizes 2 and 3 (the most abundant; Figure 2B) are more likely to be composed of members of different microRNA families (Figure 2C). This suggests that the initial microRNA
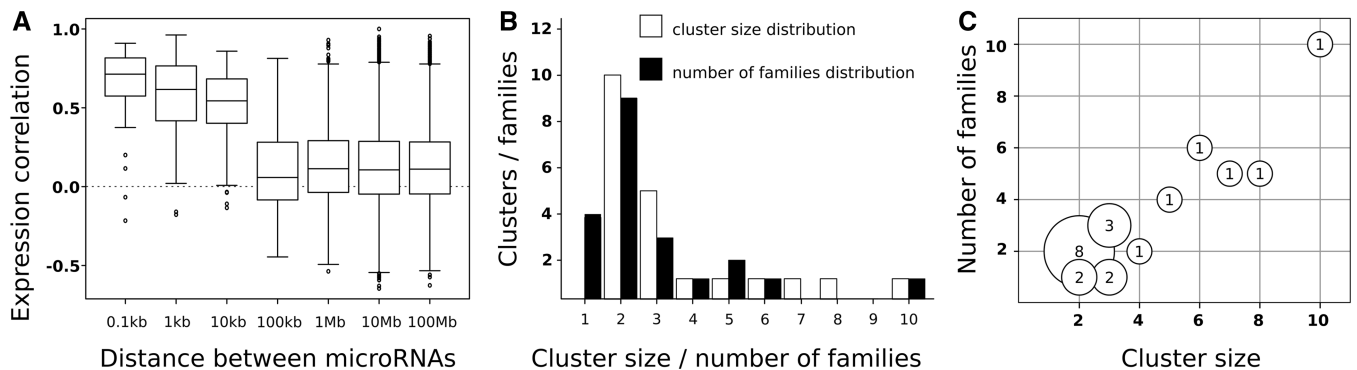

Figure 2. Clusters of microRNAs in the *D. melanogaster* genome. (**A**) Box-plots of expression correlation (Pearson) between pairs of neighboring microRNAs as a function of the genomic distance. (**B**) Frequency distribution of the number of different microRNA families in each cluster (black boxes) and the number of microRNAs per cluster (white boxes). (**C**) Bubble-plot of microRNA cluster sizes against the number of families. The number in each bubble is the number of instances of clusters of a given size (*y*-axis) with a given number of families (*x*-axis).

Table 1. Origin of *D. melanogaster* microRNA clusters

Cluster	Source	Lineage	Notes
999/4969	New hairpin	Melanogaster	Original miRNA: mir-999
982/303/983-1/983-2/984	New hairpin	Melanogaster	Multiple emergence within a conserved gene
969/210	New hairpin	*Drosophila*	Original microRNA: mir-210
124/287	New hairpin	*Drosophila*	Original microRNA: mir-124
972/973/974/2499/4966/975/976/977/978/979	New hairpin	*Drosophila*	
959/960/961/962/963/964	New hairpin	*Drosophila*	
1002/968	New hairpin	*Drosophila*	
281-2/281-1	Duplication	*Drosophila*	
310/311/312/313/2498/991/992	Duplication	*Drosophila*	Probably two clusters: 310/311/312/313 and 2498/991/992
6-3/6-2/6-1/5/4/286/3/309	New hairpin	Insects	Cluster may be older (see main text)
998/11	New hairpin	Insects	
994/318	New hairpin	Insects	
279/996	Duplication	Insects	
9c/306/79/9b	Unknown	Insects	
283/304/12	New hairpin	Protostomes	
275/305	New hairpin	Protostomes	
317/277/34	New hairpin	Protostomes	Original microRNA: mir-34
13b-1/13a/2c	Duplication	Protostomes	The original mir-2 cluster probably emerged by *de novo* acquisition of mir-2 nearby mir-71 (see main text)
2a-2/2a-1/2b-2	Duplication	Protostomes	The original mir-2 cluster probably emerged by *de novo* acquisition of mir-2 nearby mir-71 (see main text)
92a/92b	Duplication	Metazoans	Duplications in insects and chordates may be independent
100/let-7/125	Unknown	Metazoans	mir-100 and mir-125 are paralogs

cluster-forming event is rarely tandem duplication (Figure 1C), and alternative models should be considered (Figure 1).

Evolutionary origin of MicroRNA clusters

We reconstructed the evolutionary origin of all *D. melanogaster* microRNA clusters by phylogenetic analyses of their members and prediction of homologous microRNAs in other animal species (see 'Materials and Methods' section). A summary of the 21 identified clusters is shown in Table 1, and a more detailed analysis in the Supplementary Data Set S1. Seven clusters (33%) are specific to drosophilids (Table 1 and Figure 3). Collectively, 14 clusters (the majority of our data set) emerged within the insects (Figure 3), that is, the Melanogaster, *Drosophila* and insect lineages in Table 1. Two clusters are conserved among all metazoans: the mir-125/let-7/mir-100 and the mir-92a/mir-92b clusters.

We can find no cases where clustered microRNAs in *D. melanogaster* have homologs that derive from disparate loci in any other genome. We therefore conclude that none of the *D. melanogaster* clusters emerged by the union of pre-existing single microRNAs. This rules out two of our evolutionary models of cluster origin: 'put together' and 'left together'. The initial cluster-forming events for all extant microRNA clusters are predicted to be tandem duplication and hairpin formation (Figure 3), with the latter being the most common (13 of the 21 cases). The seven new clusters that emerged in the last common ancestor of drosophilids are conserved in all extant (studied) species, supporting the notion that these clusters are evolutionarily constrained after their emergence (Figure 3). Around 15% (14/99) of the microRNAs that emerged *de novo* in the Melanogaster lineage are clustered with another microRNA. However, >50% (35/66) of the microRNAs that emerged *de novo* before the split of the *Drosophila* lineage are clustered. As we look at sets of microRNAs of increasing age, the proportion that have arisen by *de novo* hairpin formation quickly approaches the 30% of observed clustered microRNAs in most species. This indicates that microRNAs in clusters are less likely to be lost after

they emerge than non-clustered microRNAs. We conclude that microRNA clusters in *D. melanogaster* primarily originated by *de novo* hairpin formation.

MicroRNA clusters are evolutionarily stable to genomic reorganizations

A fraction of the microRNAs that emerged within the dipteran lineage are <10 kb apart from another microRNA (62 of 178). We therefore speculate that clusters are important generators of microRNAs that may later become independent transcripts by translocation or duplication out of the original cluster. Thus, we explored whether extant non-clustered *D. melanogaster* microRNAs are clustered in any other animal genome, by systematic search for potential microRNA homologs of *Drosophila* microRNAs in other species (see 'Materials and Methods' section). On first inspection, it does appear that *Drosophila* non-clustered microRNAs have clustered homologs in other species (Table 2). However, close examination of this data set reveals that the majority of these clusters were the product of independent local tandem duplication or new hairpin formation. For instance, in mammalian genomes mir-7 is clustered with mir-1179, a mammal-specific microRNA, showing that the creation of new clusters by new hairpin formation also happens in other clades (Table 2). Similarly, mir-285 has been tandemly duplicated in the vertebrate lineage (Table 2).

We have found two instances of microRNA clusters in animals whose individual microRNAs are apparently not clustered in *Drosophila* (mir-1/mir-133 and mir-276a/mir-276b; Table 2). However, both pairs of microRNAs are also linked in the *Drosophila* genome, although with an inter-microRNA distance of >10 kb [see also (44)], thereby escaping our conservative cluster definition. There are two further cases of *Drosophila* non-clustered microRNAs that are clustered in another organism. First, mir-87 forms a cluster of two duplicates in most studied animals, yet *Drosophila* conserves only a single copy. This may be a rare case of 'acquired individuality' by loss of one of the microRNAs in a cluster. The other case is mir-276a/b. These two microRNAs are not clustered in any species except in the crustacean *Daphnia pulex*. The most
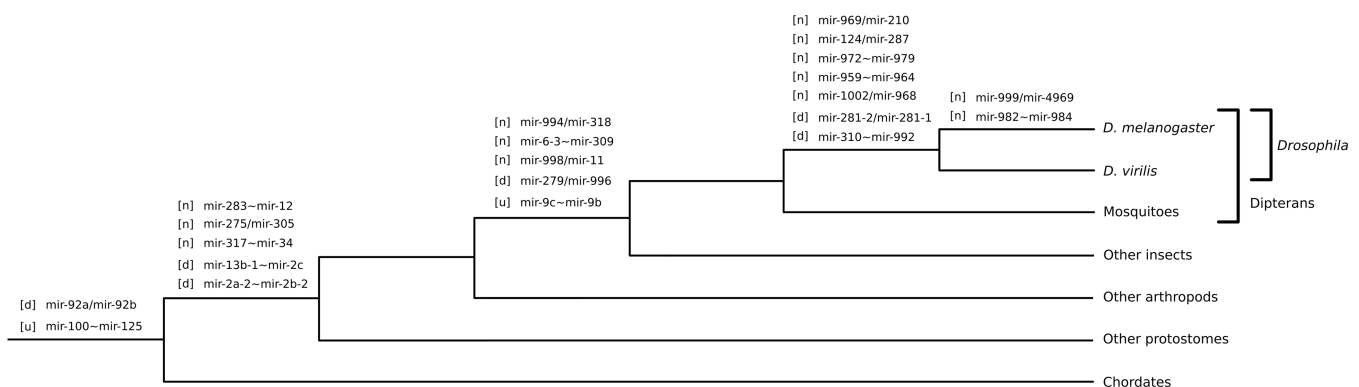

Figure 3. Origin of *D. melanogaster* microRNA clusters. Clusters emerging in a given lineage are listed on the corresponding branch of the evolutionary tree. Clusters that formed by the emergence of new hairpins in existing transcripts are labeled with a [n], and clusters formed by tandem duplication with a [d]. The label [u] indicates that we cannot infer whether the cluster originally came from a tandem duplication or a new hairpin formation. For clusters with more than two members, only the first and last microRNA are shown separated by a tilde.

Table 2. Non-clustered *Drosophila* microRNAs that are clustered in other species

microRNA	Clustered homolog[a]	Cluster source
mir-1/mir-133	Clustered together in animals;>10 kb in *D. melanogaster*	New hairpin
mir-7	Clustered with mir-1179 in mammals	New hairpin
mir-7	Clustered with mir-3529 in *Gallus gallus*	New hairpin
mir-7	Clustered with mir-1720 in *G. gallus*	New hairpin
mir-8	Tandem copies in chordates (mir-200)	Duplication
mir-10	Clustered with mir-2886 in *Bos taurus*	New hairpin
mir-10	Clustered with mir-1713 in *G. gallus*	New hairpin
mir-31a	Tandem duplication in *Rattus norvegicus*	Duplication
mir-31a	Tandem duplication in *Schmidtea mediterranea*	Duplication
mir-33	Tandem duplication in *Branchiostoma floridae*	Duplication
mir-87	Tandem duplication in insects. One copy lost in *Drosophila*	Duplication
mir-137	Clustered with mir-2682 in *Homo sapiens*	New hairpin
mir-184	Tandem duplication in *Capitella teleta*	Duplication
mir-193	Clustered with mir-365 in vertebrates	New hairpin
mir-219	Clustered with mir-2964 in vertebrates	New hairpin
mir-252	Tandem duplication in *Acyrthosiphon pisum*	Duplication
mir-252	Tandem duplication and novel mir-2001 in *Lottia gigantea* and *C. teleta*	Duplication/new hairpin
mir-263a/b	Clustered together in *Daphnia pulex*. Not clustered in other insects	Duplication
mir-276a/b	Clustered together in *Drosophila* lineage >10 kb in *D. melanogaster*	Duplication
mir-285	Tandem duplication in vertebrates	Duplication
mir-285	Clustered with mir-3556 and mir-3587 in *R. norvegicus*	New hairpin

[a]As annotated in miRBase (http://mirbase.org).

likely explanation is that mir-276a/b in *Daphnia* resulted from an independent, lineage-specific, gene duplication. We also observed that mir-9 and mir-279 microRNAs appear clustered in some insects (*Apis mellifera* and *Tribolium castaneum* according to miRBase), suggesting that an original cluster may have split in *Drosophila*. However, the evolution of the mir-9 family is particularly complex and will be better understood as new genome sequences become available. In summary, clusters of microRNAs are evolutionary units that are rarely the source of singleton microRNAs. In most cases, after a cluster is formed in the genome, it either stays together or it is lost as a whole.

DISCUSSION

In this work, we have investigated the evolutionary origin of microRNA clusters studying the model organism *D. melanogaster*. Contrary to observations in other types of polycistronic transcripts, microRNA clusters mostly emerged by tandem duplication and *de novo* hairpin formation in existing microRNA transcripts, with the latter being the dominant mechanism. Only two clusters are conserved in all metazoans, mir-92a/mir-92b and mir-125/let-7/mir-100. However, mir-92a/mir-92b may be the product of independent duplications in different animal lineages, i.e., mir-92a/mir-92b of protostomes and deuterostomes may not be orthologous clusters (Supplementary Figure S1). Although the statistical support of our phylogenetic analysis is weak (low bootstrap values), the fact that there is only one copy in *Daphnia pulex* also supports an insect specific duplication of mir-92. Moreover, mir-92a in *Drosophila* is hosted inside an intron, whereas mir-92b is not, suggesting that the two microRNAs may not be part of the same transcript. The other cluster, mir-125/let-7/mir-100, is probably the only

conserved cluster in most metazoans. Indeed, mir-100 is the evolutionarily most ancient microRNA, and it is conserved in metazoans and cnidarians (45,46).

Tandem duplication has been described as an important source of polycistronic microRNAs in plants (18,47) and in animals (17). Our analysis supports the view that this mechanism is more important in the formation of clusters in plants (3,47), as we find only six cases in which a tandem duplication is the original microRNA cluster-forming event (Table 1). The remaining clustered duplicates arose after the cluster-forming event. Two of the five clusters, mir-13b-1/mir-13a/mir-2c and mir-2a-2/mir-2a-1/mir-2b-2, are derived from a single ancestral mir-2/mir-13 cluster (48,49). All members of the mir-2/mir-13 ancestral cluster belong to the same family (the mir-2 family), suggesting that the ancestral cluster originated by tandem duplication. However, we have previously shown that the mir-2 cluster originally appeared by the *de novo* birth of the first mir-2 family member within the mir-71 transcript (48,49). Later, the mir-2 family expanded by duplication and mir-71 was lost in several lineages, including the *Drosophila* genus (49). This example shows that cluster formation by novel acquisition of a hairpin may be masked by subsequent microRNA gene loss. Hence, our approach is likely to underestimate the number of clusters formed by novel hairpin formation. Another caveat is that the actual age of some clusters may be greater than we detect with our conservative methodology. Ongoing work in our laboratory suggests, for instance, that the mir-6-3∼mir-309 cluster may be conserved beyond dipterans (Ninova, Ronshaugen and Griffiths-Jones; in preparation).

Tandem duplication is important in the evolution of already existing clusters and may generate novel functions of existing microRNAs (43). With the available data, we can only speculate why duplication is much less frequent

in cluster formation in animals than in plants. Plant microRNAs frequently target gene transcripts with high complementarity, whereas animal microRNAs bind their targets with more mismatches (50). Two tandemly duplicated microRNAs could therefore quickly diversify in their targeting properties in plants, whereas it may take longer to accumulate sufficient changes in animals to modify their targets. Tandemly duplicated microRNAs in animals are therefore more likely to be functionally redundant in the long term. For instance, members of the mir-2 family have, in general, the same targets (49,51,52). In addition, an animal microRNA duplicated in tandem may produce a gene dosage imbalance. However, the emergence of a new microRNA in an existing microRNA transcript will not affect the existing regulatory network. Protein-coding genes tend to diversify their expression pattern after duplication (53). However, duplicated microRNAs encoded in the same transcript may not be able to diversify unless they break the linkage. Some authors have suggested that, as plant microRNAs have high complementarity to their targets, it is less likely that novel microRNAs acquire functional targets in plants, explaining why *de novo* emergence is less important than duplication important in these species [see discussion in (47)]. However, this explanation assumes that a new microRNA will have functional targets as soon as it emerges in the genome. Our analyses indicate that that may not be always true, as linkage associations could play an important role in the fixation of new microRNAs. Further analyses of the increasing amount of plant microRNA data sets will clarify the evolutionary fate of novel microRNAs in plants.

Our data show that clusters of microRNAs generally evolve as single units and are lost as a whole, probably because of the tight linkage of the microRNAs. This cluster stability is known for nematode gene clusters as well (25,54), where cluster (operon) formation is described as a 'one-way' evolutionary process (23). Our comparative genomics exploration of animal microRNAs also indicates that microRNA clusters often gain new microRNAs (either by tandem duplication of further new hairpin acquisitions); yet, they rarely split or suffer rearrangements. In principle, microRNA hairpins can arise randomly in any genomic position. However, new hairpins within microRNA encoding transcripts may be more likely to become functional microRNAs, as these transcripts are already interacting with the small RNA processing machinery. Indeed, it has been found recently that primary microRNA transcripts include various sequence motifs that are required for the proper processing of precursor microRNAs (55). Clustered microRNAs are actually close to each other (median distance of 130 nt in our study), suggesting that any regulatory motif in the primary transcript may affect all the microRNAs in the cluster. MicroRNAs can also be lost from existing clusters, although this is relatively infrequent. A notable case is the mir-125/let-7/mir-100 cluster, which is highly conserved across the animal kingdom, although in both Nematodes (56) and in Platyhelminthes (57), mir-125 and let-7 are not clustered, and mir-100 is lost. This exceptional case shows that highly conserved linkage associations between microRNAs can be lost during evolution without major consequences.

Recombination between two closely linked loci by crossing-over is unlikely. Consequently, selection operating on one microRNA in a cluster results in greatly reduced selection efficiency in the neighboring microRNAs owing to a phenomena called the Hill–Robertson interference (HRI) (58,59). Both positive and purifying selection results in HRI, the former by selective sweeps (60), and the latter by background selection (61). This type of interference between linked loci has been used to explain the quantitatively reduced impact of selection compared with non-adaptive forces across whole genomes (62), and it is likely to account for the evolutionary pattern of tightly linked sequences such as clustered microRNAs.

We propose an evolutionary model for the origin and evolution of microRNA clusters, which we call the 'drift-draft' model. New microRNA hairpins often emerge *de novo* in an existing transcript (44,63). Under our model of microRNA evolution, we envision two scenarios. First, the new microRNA appears within a primary microRNA transcript; therefore, both microRNAs will be tightly linked in the genome. The older microRNA is subject to strong purifying selection so that the new microRNA is (almost) invisible to natural selection owing to HRI as recombination between the two microRNAs is virtually absent. In a second scenario, a novel microRNA may appear and provide selective advantage to the host genome. Owing to HRI, positive selection will drive the evolution of the novel microRNA, whereas, again, non-adaptive forces would dominate the evolutionary fate of the other microRNAs in the cluster. Our drift-draft model is consistent with the observations that most clusters contain members of only a few families, that clusters are relatively young and that they evolve as a single unit. It also explains why tandem duplication may happen within pre-formed clusters: changes in the number of microRNAs linked to a selectively constrained neighbor will have a minor impact on the function of the cluster. Future development of theoretical models and analysis of population polymorphism data will explore the validity of this model.

In the light of our observations, the emergence of polycistronic microRNAs is largely non-adaptive, and the maintenance of the clusters is most likely a byproduct of tight genomic linkage. However, a potential role of natural selection in functional diversification of clusters is yet to be elucidated. The linkage of microRNAs to other loci (microRNAs or other genes) has been so far ignored in microRNA evolutionary studies. The impact of genomic linkage has been shown to be a crucial factor in the evolution of protein coding genes but may be even more important in the evolution of microRNAs and other small RNA coding loci.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–2, Supplementary Figure 1 and Supplementary Data Sets 1–2.

REFERENCES

1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
2. Kloosterman,W.P. and Plasterk,R.H.A. (2006) The diverse functions of microRNAs in animal development and disease. *Dev. Cell*, **11**, 441–450.
3. Axtell,M.J., Westholm,J.O. and Lai,E.C. (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.*, **12**, 221.
4. Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
5. Marco,A. (2012) Regulatory RNAs in the light of Drosophila genomics. *Brief. Funct. Genomics*, **11**, 356–365.
6. Lagos-Quintana,M., Rauhut,R., Lendeckel,W. and Tuschl,T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
7. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, **294**, 858–862.
8. Altuvia,Y., Landgraf,P., Lithwick,G., Elefant,N., Pfeffer,S., Aravin,A., Brownstein,M.J., Tuschl,T. and Margalit,H. (2005) Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.*, **33**, 2697–2706.
9. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
10. Lee,Y., Jeon,K., Lee,J.-T., Kim,S. and Kim,V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**, 4663–4670.
11. Lee,Y., Kim,M., Han,J., Yeom,K.-H., Lee,S., Baek,S.H. and Kim,V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*, **23**, 4051–4060.
12. Aravin,A.A., Lagos-Quintana,M., Yalcin,A., Zavolan,M., Marks,D., Snyder,B., Gaasterland,T., Meyer,J. and Tuschl,T. (2003) The small RNA profile during Drosophila melanogaster development. *Dev. Cell*, **5**, 337–350.
13. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
14. Saini,H.K., Griffiths-Jones,S. and Enright,A.J. (2007) Genomic analysis of human microRNA transcripts. *Proc. Natl Acad. Sci. USA*, **104**, 17719–17724.
15. Saini,H.K., Enright,A.J. and Griffiths-Jones,S. (2008) Annotation of mammalian primary microRNAs. *BMC Genomics*, **9**, 564.
16. Ryazansky,S.S., Gvozdev,V.A. and Berezikov,E. (2011) Evidence for post-transcriptional regulation of clustered microRNAs in Drosophila. *BMC Genomics*, **12**, 371–371.
17. Hertel,J., Lindemeyer,M., Missal,K., Fried,C., Tanzer,A., Flamm,C., Hofacker,I.L. and Stadler,P.F. (2006) The expansion of the metazoan microRNA repertoire. *BMC Genomics*, **7**, 25.
18. Maher,C., Stein,L. and Ware,D. (2006) Evolution of Arabidopsis microRNA families through duplication events. *Genome Res.*, **16**, 510–519.
19. Kim,V.N. and Nam,J.-W. (2006) Genomics of microRNA. *Trends Genet.*, **22**, 165–173.
20. Alberts,B., Johnson,A., Lewis,J., Raff,M., Roberts,K. and Walter,P. (2008) *Molecular Biology of the Cell*, 5th edn. Garland Science, New York.
21. Lawrence,J. (1999) Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr. Opin. Genet. Dev.*, **9**, 642–648.
22. Salgado,H., Moreno-Hagelsieb,G., Smith,T.F. and Collado-Vides,J. (2000) Operons in Escherichia coli: genomic analyses and predictions. *Proc. Natl Acad. Sci. USA*, **97**, 6652–6657.
23. Blumenthal,T. (2004) Operons in eukaryotes. *Brief. Funct. Genomic Proteomic*, **3**, 199–211.
24. Satou,Y., Mineta,K., Ogasawara,M., Sasakura,Y., Shoguchi,E., Ueno,K., Yamada,L., Matsumoto,J., Wasserscheid,J., Dewar,K. *et al.* (2008) Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. *Genome Biol.*, **9**, R152.
25. Qian,W. and Zhang,J. (2008) Evolutionary dynamics of nematode operons: easy come, slow go. *Genome Res.*, **18**, 412–421.
26. Blumenthal,T. and Gleason,K.S. (2003) Caenorhabditis elegans operons: form and function. *Nat. Rev. Genet.*, **4**, 112–120.
27. Savard,J., Marques-Souza,H., Aranda,M. and Tautz,D. (2006) A segmentation gene in tribolium produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell*, **126**, 559–569.
28. Li,A. and Mao,L. (2007) Evolution of plant microRNA gene families. *Cell Res.*, **17**, 212–218.
29. 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.
30. Han,J., Lee,Y., Yeom,K., Nam,J., Heo,I., Rhee,J., Sohn,S., Cho,Y., Zhang,B. and Kim,V. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.
31. Nozawa,M., Miura,S. and Nei,M. (2010) Origins and evolution of microRNA genes in Drosophila species. *Genome Biol. Evol.*, **2**, 180–189.
32. Roux,J., Gonzàlez-Porta,M. and Robinson-Rechavi,M. (2012) Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. *Nucl. Acids Res.*, **40**, 5890–5900.
33. Tanzer,A. and Stadler,P.F. (2004) Molecular evolution of a microRNA cluster. *J. Mol. Biol.*, **339**, 327–335.
34. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
35. Sempere,L.F., Cole,C.N., McPeek,M.A. and Peterson,K.J. (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool. B Mol. Dev. Evol.*, **306**, 575–588.
36. Wheeler,B., Heimberg,A., Moy,V., Sperling,E., Holstein,T., Heber,S. and Peterson,K. (2009) The deep evolution of metazoan microRNAs. *Evol. Dev.*, **11**, 68, 50.
37. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
38. Katoh,K. and Toh,H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinformatics*, **9**, 286–298.
39. Griffiths-Jones,S. (2005) RALEE–RNA ALignment editor in Emacs. *Bioinformatics*, **21**, 257–259.
40. Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
41. Felsenstein,J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**, 368–376.
42. Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M. and Kumar,S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.
43. 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.

44. Campo-Paysaa,F., Sémon,M., Cameron,R.A., Peterson,K.J. and Schubert,M. (2011) microRNA complements in deuterostomes: origin and evolution of microRNAs. *Evol. Dev.*, **13**, 15–27.

45. 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.

46. Wheeler,B.M., Heimberg,A.M., Moy,V.N., Sperling,E.A., Holstein,T.W., Heber,S. and Peterson,K.J. (2009) The deep evolution of metazoan microRNAs. *Evol. Dev.*, **11**, 50–68.

47. Nozawa,M., Miura,S. and Nei,M. (2012) Origins and evolution of microRNA genes in plant species. *Genome Biol. Evol.*, **4**, 230–239.

48. Marco,A., Hui,J.H.L., Ronshaugen,M. and Griffiths-Jones,S. (2010) Functional shifts in insect microRNA evolution. *Genome Biol. Evol.*, **2**, 686–696.

49. Marco,A., Hooks,K. and Griffiths-Jones,S. (2012) Evolution and function of the extended miR-2 microRNA family. *RNA Biol.*, **9**, 242–248.

50. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

51. Boutla,A., Delidakis,C. and Tabler,M. (2003) Developmental defects by antisense mediated inactivation of micro RNAs 2 and 13 in Drosophila and the identification of putative target genes. *Nucleic Acids Res.*, **31**, 4973–4980.

52. Leaman,D., Chen,P.Y., Fak,J., Yalcin,A., Pearce,M., Unnerstall,U., Marks,D.S., Sander,C., Tuschl,T. and Gaul,U. (2005) Antisense-mediated depletion reveals essential and specific functions of microRNAs in Drosophila development. *Cell*, **121**, 1097–1108.

53. Lynch,M. and Force,A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, **154**, 459–473.

54. Cutter,A.D. and Agrawal,A.F. (2010) The evolutionary dynamics of operon distributions in eukaryote genomes. *Genetics*, **185**, 685–693.

55. 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.

56. Griffiths-Jones,S., Hui,J.H.L., Marco,A. and Ronshaugen,M. (2011) MicroRNA evolution by arm switching. *EMBO Rep.*, **12**, 172–177.

57. Marco,A., Kozomara,A., Hui,J., Emery,A., Rollinson,D., Griffiths-Jones,S. and Ronshaugen,M. (2012) Sex-biased expression of microRNAs in Schistosoma mansoni (in review).

58. Hill,W.G. and Robertson,A. (1966) The effect of linkage on limits to artificial selection. *Genet. Res.*, **8**, 269–294.

59. Barton,N.H. (2010) Genetic linkage and natural selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **365**, 2559–2569.

60. Maynard Smith,J. and Haigh,J. (1974) The hitch-hiking effect of a favourable gene. *Genet. Res.*, **23**, 23–35.

61. Charlesworth,B., Morgan,M.T. and Charlesworth,D. (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

62. Lynch,M. (2007) *The Origins of Genome Architecture*, 1st edn. Sinauer Associates Inc, Sunderland.

63. Marco,A., Ninova,M. and Griffiths-Jones,S. (2013) Multiple products from microRNA transcripts. *Biochem. Soc. Trans.*, **41**, 850–854.