Varying constraints impact the miR-6/5/4/286/3/309 microRNA cluster in *Drosophila* leg development.

Zhe Qu[1,^], Wing Chung Yiu[1,^], Ho Yin Yip[1,^], Wenyan Nong[1,^], Clare W.C. Yu[1], Ivy H.T. Lee[1], Annette Y.P. Wong[1], Nicola W.Y. Wong[1], Fiona K.M. Cheung[1], Ting Fung Chan[2], Kwok Fai Lau[3], Silin Zhong[2], Ka Hou Chu[4], Stephen S. Tobe[5], David E.K. Ferrier[6], William G. Bendena[8], Jerome H.L. Hui[1,*]

[1] School of Life Sciences, Simon F.S. Li Marine Science Laboratory, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong
[2] School of Life Sciences, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong
[3] School of Life Sciences, The Chinese University of Hong Kong
[4] School of Life Sciences, Simon F.S. Li Marine Science Laboratory, The Chinese University of Hong Kong
[5] Department of Cell and Systems Biology, University of Toronto, Canada
[6] School of Biology, University of St. Andrews, United Kingdom
[7] Department of Biology, Queen's University, Canada

^co-first authors, Correspondence: jeromehui@cuhk.edu.hk

Abstract

A striking feature of microRNAs is that they are often clustered in the genomes of animals. The functional and evolutionary consequences of this clustering remain obscure. Here, we investigated a microRNA cluster miR-6/5/4/286/3/309 that arose and conserved in the drosophilid lineages. Small RNA sequencing revealed expression of this microRNA cluster in *Drosophila melanogaster* leg discs, and overexpression of the whole cluster resulted in leg appendage shortening. Mutants are fertile and can mate, but males cannot efficiently grasp the female abdomen for courtship. Further mutants expressing different combinations of members in the microRNA cluster were constructed. Expression of individual ancient microRNA in the cluster resulted in normal phenotype, but either the expression of several ancient microRNAs together (miR-5/4/286/3/309) or recently evolved clustered microRNAs (miR-6-1/2/3) can recapitulate the phenotypes generated by the whole-cluster overexpression mutants. Transcriptomic analyses revealed the differentially expressed genes between mutants and wild-type, and screening of overexpression and loss-of-function mutants revealed down-regulation of leg patterning gene cassettes in generation of the leg-shortening phenotype. Further cell transfection with different combinations of members of this microRNA cluster revealed a suite of downstream genes targeted by all cluster members, as well as complements of targets that are unique for distinct cluster members. This study demonstrates the importance of microRNA clustering where new members can reinforce the selection forces on both the flanking sequences (i.e. regulatory polycistronic promoter in controlling expression level) and the gene regulatory network of existing microRNA.

Introduction

Operation of genes within multigenic clusters is widespread, but the functional and evolutionary implications of this are often poorly understood. In microbes, the operon concept and recent anti-phage defensive system are well known for their importance (Doron et al 2018). In animals there are various examples of protein-coding genes operating within clusters that exhibit multigenic regulatory elements, such as the homeobox genes in the Hox cluster (Deschamps and Duboule 2017; Hui et al 2012). In addition to these cases of clustered protein-coding genes, non-protein encoding genes such as those producing microRNAs are also often found to be in co-regulated clusters (e.g. Altuvia et al 2005; Fromm et al 2015; Lagos-Quintana et al 2001; Mohammed et al 2014; Bartel 2018). Since these microRNA clusters are relatively recent discoveries compared to protein-coding gene clusters, much less is known about the range of functional consequences of this clustering and the resultant evolutionary impacts.

There is an important difference between protein-coding versus microRNA gene clusters in animals. The individual genes in protein-coding gene clusters tend to have their own promoters, whereas the microRNA clusters are often comprised of members transcribed as a single unit or polycistronic transcript regulated by a single promoter (Kozomara and Griffiths-Jones 2014). In addition, microRNA genes in a cluster are sometimes found to be conserved in sequence and orientation (e.g. the miR-17 cluster in mammals, Tanzer and Stadler 2004). *De novo* formation of new hairpins in existing microRNA transcripts has been proposed as the major mechanism giving rise to members in the cluster (Marco et al 2013; Wang et al 2016). MicroRNAs in the same cluster are proposed to possess similar targeting properties or regulate genes in the same pathway (e.g. Kim and Nam 2006; Yuan et al 2009; Wang et al 2011; Hausser and Zavolan 2014; Wang et al 2016). For instance, synchronised expression of cluster microRNAs found in normal human cells is mis-regulated during disease development (Dambal et al 2015; Nojima et al 2016); and has been implicated in cancer formation (Ventura et al 2008; Kim et al 2009). The range of functional and evolutionary implications of polycistronic microRNAs in general remains controversial with regards to whether they are non-adaptive, the by-product of a tight genomic linkage, or simply expressed together due to unknown functional constraints (e.g. Marco et al 2013). A fundamental issue in these controversies is that most of these studies rely on correlating expression of the cluster with *in silico* prediction of their target genes. Systematic dissection for the functions of individual microRNAs in the cluster versus the function of all of them as one whole cluster remains to be tested.

Results and Discussion

Here, we investigated a microRNA cluster miR-6/5/4/286/3/309 (or named as miR-309-6 cluster in previous studies) that arose and conserved in the drosophilid lineages (Fig. 1A, Supplementary data S1). The miR-309-6 cluster is located between genes CG15125 and CG11018, and is processed from a single ~1.5kb transcript in *Drosophila melanogaster* (Biemar et al 2005). miR-309-6 has high expression level in early embryos (Ninova et al 2014) and has roles in the larval stage as deletion of the cluster resulted in larval lethality (Bushati et al 2008; Chen et al 2014). To identify the targets that are being regulated by miR-309-6 clusters in different drosophilids, *in silico* prediction of the microRNA targets of individual candidates within cluster were carried out using miRanda and also Targetscan for *D. melanogaster* (Supplementary data S2). In *D. melanogaster*, 33.53% of the total targets were shared between at least 2 microRNAs in the cluster (Fig. 1B). In other drosophilid species, ignoring the numbers predicted in *D. willistoni* which is based on small number of available transcriptomes, 12.96%-25.68% of targets were shared between at least 2 microRNAs (Supplementary data S2). This is in partial agreement with previous data that suggested that microRNAs within a cluster may share common target genes (e.g. Kim and Nam 2006; Yuan et al 2009; Wang et al 2011; Hausser and Zavolan 2014; Wang et al 2016).

As bioinformatics prediction of microRNA targets are prone to inclusion of false positives (e.g. Pinzon et al 2017), functional investigation was undertaken. Given that loss-of-function of this whole cluster resulted in larval lethality (Bushati et al 2008; Chen et al 2014), and our investigation of the available line with the whole cluster deletion also resulted in lethality before adult stage. So the strategy of comparing individually deleted members of the cluster would not be ideal due to the subtle phenotypic score. Therefore, two homozygous UAS-miR-309-6 cluster lines were first generated for this study, with the whole cluster independently inserted at 3L:3714826 (5'end of CG32264, named UAS-miR-309-6-I) and 3R:25235447 (5'end of CG10420, named UAS-miR-309-6-II) respectively as confirmed by Splinkerette PCR.

Previous studies mainly focused on the functional importance of this microRNA cluster in embryonic stages (i.e. during maternal-to-zygotic transition), and there is limited information on its function in other developmental stages or tissues (e.g. misorientation of adulty sensory bristles will be caused on the adult notum with miR-3/-309 overexpression, Zhou et al 2018). We first carried out a screening of phenotypes by crossing the UAS-miR-309-6-I and UAS-miR-309-6-II with different GAL4 lines, and found the expression of the whole cluster in the leg discs and wing discs will result in phenotypic changes (to be detailed below). Subsequently, we sequenced the small RNA

contents of leg discs in *D. melanogaster* L3 larvae, and revealed the expression of mature microRNAs contained in this cluster (Fig. 2). This finding is further validated by Taqman microRNA assays (Supplementary data S3), suggesting their potential functional roles during *D. melanogaster* leg development. The expression of miR-309 and miR-3 are lower than other members contained in the cluster, which is consistent with a recent study showing that faster degradation rate of miR-309/-3 (Zhou et al 2018). Subsequently, UAS-miR-309-6-I or UAS-miR-309-6-II were crossed with either GAL4-*Dll* or GAL4-*ptc* respectively to create overexpression of the miR-309-6 cluster in *Distal-less* or *patched* expressing cells. Under the direction of these two promoters, a previously undescribed shortening of the leg appendages resulted in (Fig. 3A-C, Supplementary information S4). In both the parents of GAL4-*Dll* and GAL4-*ptc*, and *w1118* (controls), all segments in the thoracic leg including femur, tibia, 5 subsegments in tarsus, and a distal claw can be observed (Fig. 3A, Casares and Mann 2001), but in both GAL4-*Dll>miR-309-6* and GAL4-*ptc>miR-309-6* animals, the tarsal segments were shortened and reduced (Fig. 3B-C). This phenotype is not observed when expressing the two UAS-miR-309-6 cluster lines driven by any of the following promoters, including GAL4-*ap*, GAL4-*Aug21*, GAL4-*Bx*, GAL4-*dpp*, GAL4-*Kr-h1*, GAL4-*mirr*, GAL4-*sd*, GAL4-*tubP*, nor P{GawB}69B (data not shown), suggesting the phenotypic change is specific to the crosses. Further, to eliminate the possibilities that the phenotypes were caused by the disruption of CG32264 or CG10420 during transgenic insertion of UAS-miR-309-6-I and UAS-miR-309-6-II (as no functions are currently known for these two genes), UAS-CG32264-RNAi and UAS-CG10420-RNAi were crossed with either GAL4-*Dll* or GAL4-*ptc* to knockdown the expression of CG32264 and CG10420. RNAi knockdown of these genes did not result in loss of the tarsal leg segments as well (data not shown). Flies with shortened leg segments due to overexpression of the whole microRNA cluster were fertile and able to mate. Nevertheless, in mutants with shortened leg appendages, the mobility of both males and females are reduced. Moreover, during courtship, more effort is required for males and females to copulate, and the penetration time is dramatically reduced compared to wild type.

To dissect how many microRNA(s) in the cluster was/ were responsible for causing the phenotypic change in the mutants, we have generated a homozygous line UAS-*miR-309/3/286/4/5* (UAS-*miR-309-5*), and together with UAS-*miR-309*, UAS-*miR-3*, UAS-*miR-286*, UAS-*miR-286/4/5*, and UAS-*miR-6-1/6-2/6-3*, they were crossed with either GAL4-*Dll* or GAL4-*ptc* respectively. Surprisingly, *Dll* or *ptc* driven overexpression of either *miR-309-5* or *miR-6-1/6-2/6-3* (Fig. 3D, E, J and O, Supplementary information S4) recapitulated the phenotype created by overexpression

of the entire cluster, while all others showed a normal leg phenotype (Fig. 3 F-I and K-N). The number of tarsal segments were counted for different mutants, and only *Dll* or *ptc* driven overexpression of either *miR-309-5* or *miR-6-1/6-2/6-3* showed reduced tarsal segment numbers (Supplementary information S5). In addition to the aberrant tarsal segments, another phenotype, loss of the anterior wing cross vein was also observed in GAL4-*ptc>miR-309-6*, GAL4-*ptc>miR-309-5* and GAL4-*ptc>miR-6-1/6-2/6-3* flies (Fig. 3Q, R and W), while other miRNA members did not result in wing vein loss (Fig. 3S-V). UAS-microRNA-sponge lines that act as competitive inhibitors of the individual microRNAs including *miR-309, miR-3, miR-286, miR-4, miR-5* and *miR-6-1/6-2/6-3* were then crossed with GAL4-*Dll* and GAL4-*ptc*. Nevertheless, none of them generated the aberrant tarsal leg segments or wing vein loss phenotypes suggesting that the loss of any individual miRNA is insufficient to affect the development of leg and wing (Supplementary data S7). Moreover, in our unpublished data (Qu and Hui, personal communications), CRISPR-Cas9 deleted miR-6-1/2/3 resulted in larval lethality before developing into adults, which cannot be directly used to compare the adult leg phenotypes generated above. These data showed that the upregulation of either *miR-309-5* cluster or *miR-6-1/6-2/6-3* could cause the loss of tarsal segments and anterior cross vein.

To understand which target genes are regulated by this microRNA cluster, total RNA was extracted from the leg discs of pupariating third instar larvae of GAL4-*ptc>miR-309-6* and GAL4-*ptc* (control), and subjected to Illumina Hi-Seq2500 sequencing. The third instar larvae were chosen as this is the developmental stage for differentiation of the leg tarsal segments which originate from the leg disc (Kojima 2004). Differentially expressed genes were shown in Supplementary data S8, and the expression levels of CG32264 and CG10420 were similar in both the GAL4-*ptc>miR-309-6* and control further eliminating the possibility that the phenotypic change was caused by the disruption of these genes.

The formation of appendages in *Drosophila* has been extensively studied, genes associated with leg development were down regulated revealed by our transcriptome data (Supplementary data S8). Quantitative PCR was carried out to validate the gene expression changes in L3 leg discs of GAL4-*ptc>miR-309-6*, GAL4-*ptc>miR-309-5* and GAL4-*ptc>miR-6-1/6-2/6-3* (Fig. 4A-G). Our data showed that several leg patterning genes (such as *zfh-2*, *Sp1*, *Egfr*, *dysf*) were significantly down regulated in mutant leg discs, suggesting various cluster miRNA mutants generated short leg phenotypes were due to the repression of leg developmental genes. Subsequently, UAS-RNAi lines of these down-regulated genes were crossed to GAL4-*ptc* and

GAL4-*Dll* respectively, and we found that *ptc>zfh-2*-RNAi, *Dll>zfh-2*-RNAi, *Dll>Sp1*-RNAi, *Dll>dysf*-RNAi, *ptc>Egfr*-RNAi *ptc>dpp*-RNAi and *Dll>dpp*-RNAi resulted in tarsal segment deformities including loss of segment, joint boundaries and claws (Fig. 4H-N). *zfh-2* is a zinc finger homeodomain-2 transcription factor known to be involved in proximal-distal patterning of appendage (Guarner et al 2014), while the transcription factors *Sp1* and *dysf* were regulators of appendage growth and tarsal joint formation in insects (Córdoba and Estella 2014; Córdoba et al 2016). *Egfr* and *dpp* were also well known to be vital for leg patterning (Galindo et al 2002). These data indicate that the miR-309-5 cluster and miR-6-1/2/3 target similar gene cassettes to regulate leg development.

To determine if there are other genes affecting the leg development, two sets of differentially expressed genes were screened for further analyses, including 1) genes that were expressed in both mutants and controls but with the most significant changes in expression, and 2) genes not expressed in the mutants but highly expressed in the controls. Twenty-four genes were identified as differentially expressed including *Arc1, Ag5r, Ag5r2, CG5084, CG5506, CG6933, CG7017, CG7252, CG7714, CG14300, CG17826, Eig71Eb, Hsp68, Hsp70Bb, Hsp70Bc, Mtk, Muc96D, Peritrophin-15a, Sgs3, Sgs5, stv, Obp99a, obst-I,* and *w* (Supplementary data S8). Genes that were absent or downregulated in the GAL4-*ptc>miR-309-6* compared to controls were further tested by generating GAL4-*ptc* or *Dll* >UAS-RNAi lines for each gene. These lines were then analysed as to whether a short leg phenotype was observed. Similarly, GAL4-*ptc* or *Dll* >UAS-lines were generated for each gene that was upregulated in GAL4-*ptc>miR-309-6.* However, none of these individual manipulations were found to cause the shortening of leg appendage or loss of tarsal segments (Supplementary data S9).

To further explore the genes being controlled by individual members of this miR-309-6 cluster, we transfected different combinations of the cluster miRNAs into the *Drosophila* S2 cells and subjected the resultant transcripts to Illumina sequencing. As shown in Fig 5, there were 178 differentially expressed genes in total when comparing to the controls, and among these genes, 113 genes (~63.5%) and 65 genes (~36.5%) are commonly and uniquely regulated by individual members, respectively (Supplementary data S10). The gene ontology (GO) enrichment analysis was carried out between the whole cluster versus the new members (miR-6-1/2/3), and revealed that functions of the common and unique targets are similar (Supplementary data S11). A question that has been frequently asked within the field is whether it is crucial and important for protein coding genes to be regulated by microRNAs. In some views,

given that microRNAs can theoretically bind to hundreds of transcripts (e.g. Bartel 2009; Betel et al 2010; Reczko et al 2012), it has been proposed that the effect of microRNAs on targets would therefore be diluted, and thus most target repressions would be weak and biologically irrelevant. Considering the evolutionary conservation of microRNA target sites, fitness consequence, and their phenotypic control, it has been argued that this view is flawed (Zhao et al 2017). In addition, some of these studies usually do not consider the expression pattern of microRNAs and targets in the cell, i.e. whether both products do indeed coexist in the same cell at the same time, thus carrying out the biological functions therein, such that microRNAs are likely to have fewer targets in the responsive cells in carrying out specific biological functions. In other views with wider acceptance, based on the fact that few microRNAs have been found to strongly repressing targets and can result in phenotypic changes, it has also been argued that many microRNAs function are to repress target transcription noises (e.g. Flynt and Lai 2008; Pinzon et al 2017). In a line to this idea expanding to system-level, stabilisation of gene regulatory networks (GRN) via motifs or involvement in RNA cross-talks in canalisation has also been proposed (e.g. Hornstein and Shomron 2006; Peterson et al 2009; Wu et al 2009; Ebert and Sharp 2012; Posadas and Carthew 2014). Considering that a microRNA locus can produce two mature microRNA arms which can swap their dominance in expression at different cell status, physiological or pathological status (i.e. microRNA arm switching, e.g. Griffiths-Jones et al 2011), also given their highly conserved sequences and targets (e.g. Marco et al 2010; Hui et al 2013; or named as twin-miRs in Zhao et al 2017), the functional importance of microRNAs in biological systems is also suggested. In the data provided here *in vitro* and *in vivo*, expression of a polycistronic transcript containing eight microRNAs organised in a genomic cluster (miR-309-6) can result in phenotypic and gene expression changes (Fig. 5 and 6A), together with the fact that individual microRNA alone cannot recapitulate this phenotype, further supporting that microRNAs (cluster) as one of the important functional key components in the gene regulatory network and system.

Another question is whether microRNAs need to be clustered. The conventional view focus on the functional significance of shared targets by all microRNA members in a cluster, and with the governance of a shared promoter, these microRNAs need to regulate their targets in the same cell and developmental time. Hence, there is selection force on the flanking sequences (i.e. polycistronic promoter) of microRNA cluster as members of the microRNA cluster must cooperatively function together due to weak target repression (see introduction, Fig 6B). Our data supported this view. For instance, if there is mutation occurred in the promoter sequence resulting in overexpression of

this polycistronic transcript, this would then lead to phenotypes such as the shortening of legs that confer potential lower fitness, and thus implying the presence of negative selection force occurring on the promoter sequence such that members of microRNA cluster tend to cooperatively work for coordinated regulation.

In addition, our study also provided new understanding to microRNA clustering. In cases of expressing only miR-309, miR-3, or miR-286-4-5, no phenotypic effects are observed, while the summation of expressing all of them (miR-309-5) will result in phenotypic changes (Fig. 6A). These results suggested that a new selection force on the promoter sequence was generated in restraining its expression level when these microRNAs became a genomic cluster (Fig. 6B). Interesting, expression of miR-6-1/2/3 will also result in similar phenotypes (Fig. 6A). Considering that these microRNAs are novel or the youngest members gained in this microRNA cluster, all these data suggested that new microRNA members when being recruited into the microRNA cluster (either via *de novo* formation/tandem duplication), indeed, also possessed and increased the negative selective pressure acting on the existing microRNA cluster. Another possibility is that individual microRNA of a microRNA cluster can only target the leg patterning genes weakly, and a phenotype can only be resulted when multiple leg genes have been downregulated. Regardless, there are certainly varying selection forces acting by both old and new members of a microRNA cluster on the flanking sequence (i.e. the polycistronic promoter in controlling the expression level of the cluster, Fig. 6B). Plasticity-first evolution has been proposed as a dominant mechanism in nature given the undefined traits (Levis and Pfennig 2016), and microRNAs have also been postulated as the “missing link” in such by providing fine-tuning re-regulation of expression network for adaptation to frequent environmental change (Voskarides 2016). The beauty of a microRNA cluster will then be a balance of sequence mutations on its promoter that control the spatiotemporal aspect, and the functions of target genes either commonly or uniquely regulated by microRNAs inside the cluster on its expression level. Such beauty with both functional governance and flexibility probably contributes to phenotypic novelty during evolution. In summary, at least in the miR-6/5/4/286/3/309 cluster, varying functional constraints do indeed exist in holding microRNAs as a genomic cluster.

Methods

Genome-wide target prediction

Mature miRNA sequences were retrieved from the public repository for published microRNA sequences at the miRBase database (http://www.mirbase.org). Eukaryotic 3' untranslated region (UTR) sequences were retrieved from the public repository for published mRNA sequences at the FlyBase (ftp://ftp.flybase.net/releases/FB2018_01/). All mature miRNAs were then used to predict for targets in their respective genomes using the miRanda algorithm (Enright et al 2003) with parameters (i.e. -sc S Set score threshold to S 140 (from 140 to 772.00); -en -E Set energy threshold to -E kcal/mol (from -78.37 to -5.28); and -strict Demand strict 5' seed pairing). For *D. melanogaster*, target prediction was also performed by Targetscanfly (Agarwal et al 2018).

Fly culture, mutant construction, and insertion site checking

To prepare the overexpression constructs of *D. melanogaster* miRNA cluster miR-309-6 and miR-309-5, the correspond stem-loop with flanking sequences was amplified and cloned into the GAL4-inducible vector pUAST (primer information is provided in Supplementary information S12). Constructs were sequenced prior to injection into *D. melanogaster* w^{1118} embryos. Flies were screened and crossed to generate stable homozygous transformants. Insertion site of the *UAS-miR-309-6* cluster transgene was checked with Splinkerette PCR (Potter and Luo 2010). Various GAL4 drivers, UAS-gene and UAS-RNAi lines were obtained from the Bloomington *Drosophila* Stock Center (BDSC) (Supplementary information S6). miR-309-6 whole cluster deletion line was ordered from BDSC donated by Prof. Stephen Cohen (#58922, Chen et al 2014). UAS-CG32264-RNAi and UAS-CG10420-RNAi were donated by Transgenic RNAi Project. UAS-miR-6-1/6-2/6-3, UAS-miR-286, UAS-miR-309 and UAS-miR-286/4/5 were donated by Stephen Cohen and Eric Lai, and UAS-miR-3 was obtained from the Zurich ORFeome Project. UAS-microRNA-sponge lines including miR-3, miR-4, miR-5, miR-6-1/2/3, miR-286 and miR-309 were donated by David Van Vactor. All flies were maintained on standard yeast-cornmeal-agar medium at 25˚C. Males and virgin females from each fly line were randomly collected for crossings. For each crossing of GAL4 and UAS fly lines, three random males and three random virgin females were used, and reciprocal crosses were carried out. At least three separate crossings were performed for each GAL4 and UAS pair.

MicroRNA expressing vector construction and cell transfection

MicroRNAs were amplified from *D. melanogaster* (primer information shown in Supplementary information S12). Amplicons were cloned into pAC5.1 vector

(Invitrogen). All constructs were sequenced to confirm their identities. *Drosophila* S2 cells (DRSC) were kept at 23°C in Schneider *Drosophila* medium (Life Technologies) with 10% (v/v) heat-inactivated fetal bovine serum (Gibco, Life Technologies) and 1:100 Penicillin-Streptomycin (Gibco, Life Technologies). The pAC5.1-miRNA (300 ng) was transfected into *Drosophila* S2 cells using Effectene (Qiagen) per the manufactures' instructions. RNA was isolated at 48 h post-transfection.

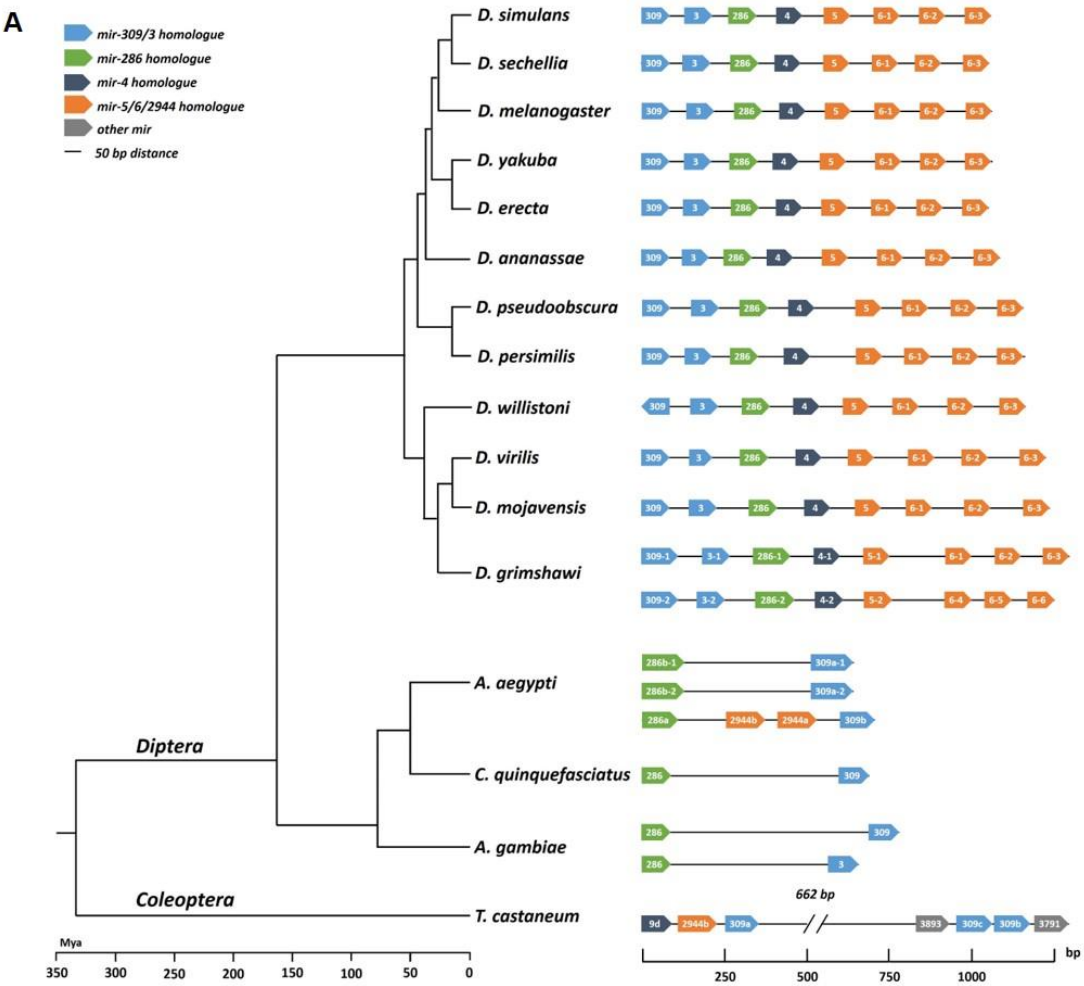
Transcriptome and small RNA sequencing

RNA was extracted from leg discs of pupariating 3rd instar larvae of GAL4-*ptc>miR-309-6*, GAL4-*ptc>miR-309-5*, GAL4-*ptc>miR-6-1/6-2/6-3* and GAL4-*ptc* (control) lines, S2 cells expressing different combination of microRNAs in the cluster and S2 (control) cells. Wildtype leg discs RNA was subjected to BGI for HiSeq Small RNA library construction and 50 bp single-end (SE) sequencing. The expressed miRNAs were quantified by the "quantifier.pl" module of the mirDeep2 package with parameter "-g 0 -e 0 -f 0", and the clean reads were aligned to hairpin sequence from miRBase (22 release) by bowtie with parameter "-l 18 -v 0 -a --best --norc --strata". Transcriptome libraries were constructed using the TruSeq stranded mRNA LT sample prep kit, and sequenced on an Illumina HiSeq2500 platform (BGI Hong Kong). Raw reads were filtered using Trimmomatic and mapped to Flybase v.6.14 using Cufflinks (Trapnell et al 2012). Differential gene expressions were evaluated using Cuffdiff.

Taqman microRNA assays and real-time PCR

Expression of microRNAs was measured via Taqman microRNA assays (Applied Biosystems™) per the manufactures' instruction. For detection of differential gene expression, RNAs from the respective crosses were reverse-transcribed into cDNA using the iScript™ cDNA synthesis Kit (BioRad). Real-time PCR was conducted in three biological replicates using the CFX96 Touch™ Real-Time PCR Detection System (BioRad), with a cycle of denaturation at 95°C for 3 min followed by 40 cycles of 95°C/ 10s, 55°C/ 10s and 72°C/ 15s. PCRs were run with half iTaq™ Universal SYBR® Green Supermix (BioRad) and 0.2 µM of each primer pair (primer information is listed in Supplementary information S12).

Figures and legends



B

	dme-miR-286							
dme-miR-286	760	dme-miR-3						
dme-miR-3	178	479	dme-miR-309					
dme-miR-309	165	649	559	dme-miR-4				
dme-miR-4	286	276	255	1823	dme-miR-5			
dme-miR-5	172	179	173	308	312	dme-miR-6-1		
dme-miR-6-1	150	164	160	295	792	437	dme-miR-6-2	
dme-miR-6-2	197	173	149	279	822	781	352	dme-miR-6-3
dme-miR-6-3	157	129	111	218	766	781	772	189

Unique targets: 66.43%
Common targets: 33.53%

Fig. 1 miR-309-6 cluster (miR-6/5/4/286/3/309) in various insects. A) Genomic organisation of miR-309-6 cluster in insects (drawn to scale). MicroRNA homologues are present in same colour (Mohammed et al 2014; Ninova et al 2014, 2016). B) Number of *in silico*-predicted targets shared by microRNAs in the cluster in *Drosophila melanogaster* by miRanda. Unique targets are highlighted in yellow. For other species, please refer to supplementary data S2.

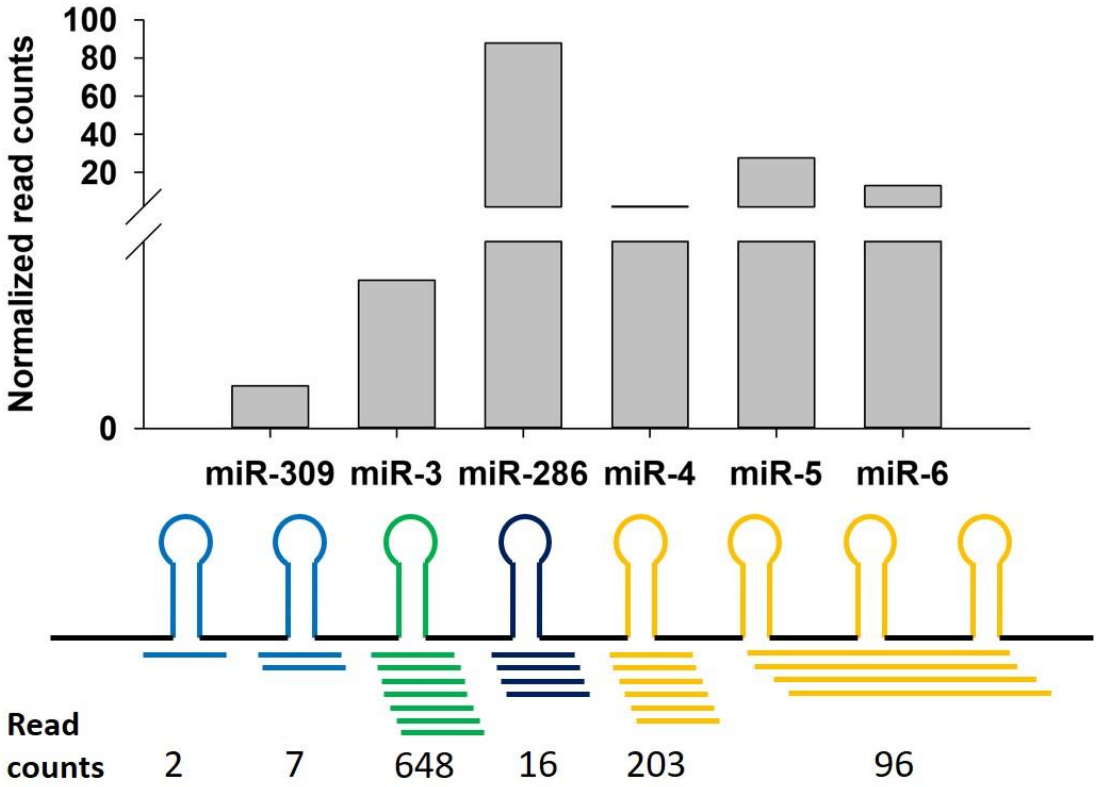


Fig. 2 Small RNA sequencing revealed the expression of miR-309-6 cluster in *D. melanogaster* L3 larvae leg discs. The upper panel showed normalized read counts quantified by mirDeep2. Below panel showed the clean read counts mapped by bowtie.

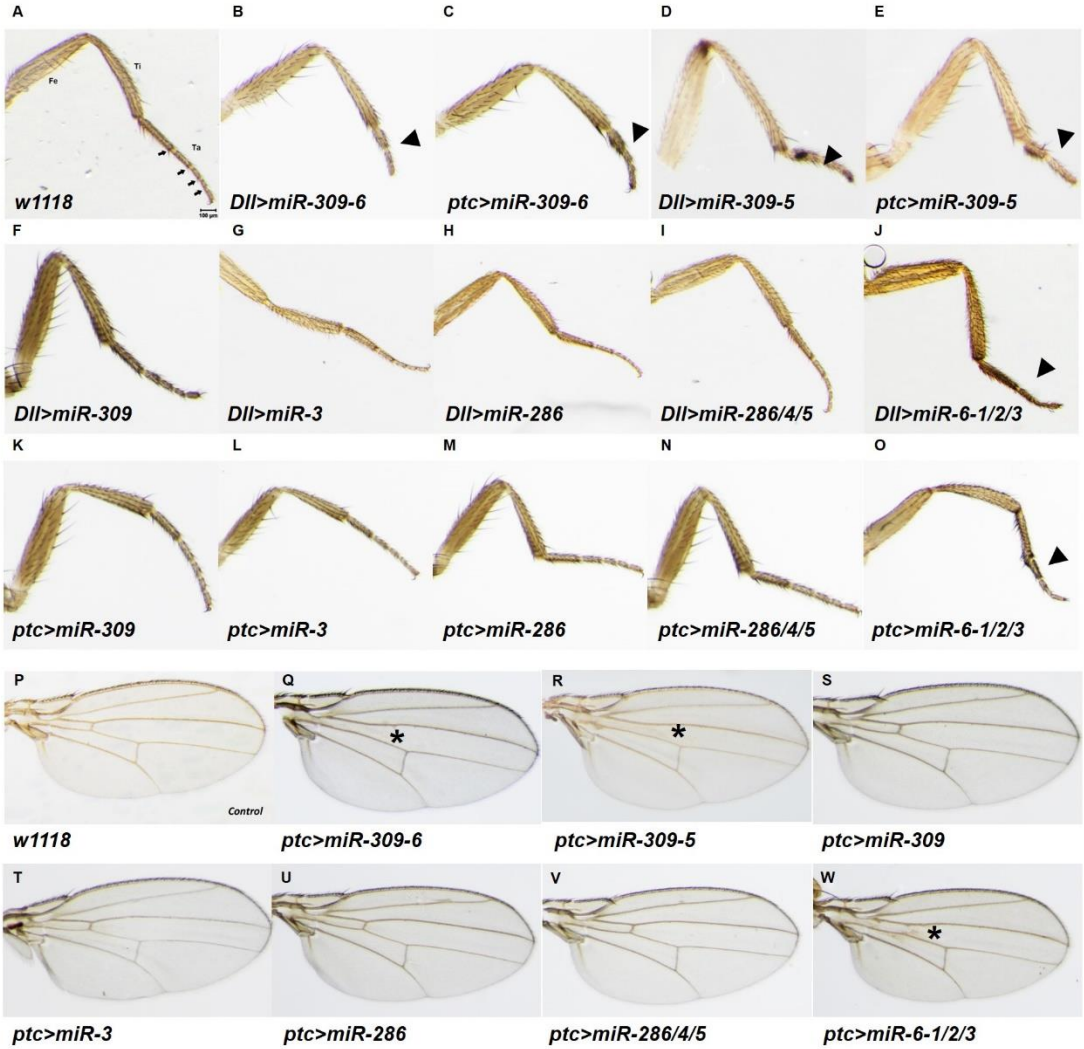


Fig. 3 Expression of the whole miR-309-6 cluster, miR-309-5 cluster or miR-6-1/2/3 alone results in shortening of leg tarsus and loss of anterior cross vein, where expression of other microRNAs from the cluster does not cause these phenotypic changes. Leg pictures of A) w1118 (control), B-E) expressing of whole miR-309-6 cluster and miR-309-5 cluster in either the *Distal-less* or *patched* expressing cells, F-O) expression of individual microRNA from the cluster in either the *Distal-less* or *patched* expressing cells; Wing pictures of P) w1118, Q-W) expressing of whole miR-309-6 cluster, miR-309-5 cluster or individual microRNA in *patched* expressing cells. Abbreviations: Fe: femur; Ti: tibia; Ta: tarsus; Arrow indicates shorted tarsal segments; star indicates the loss of anterior wing vein.

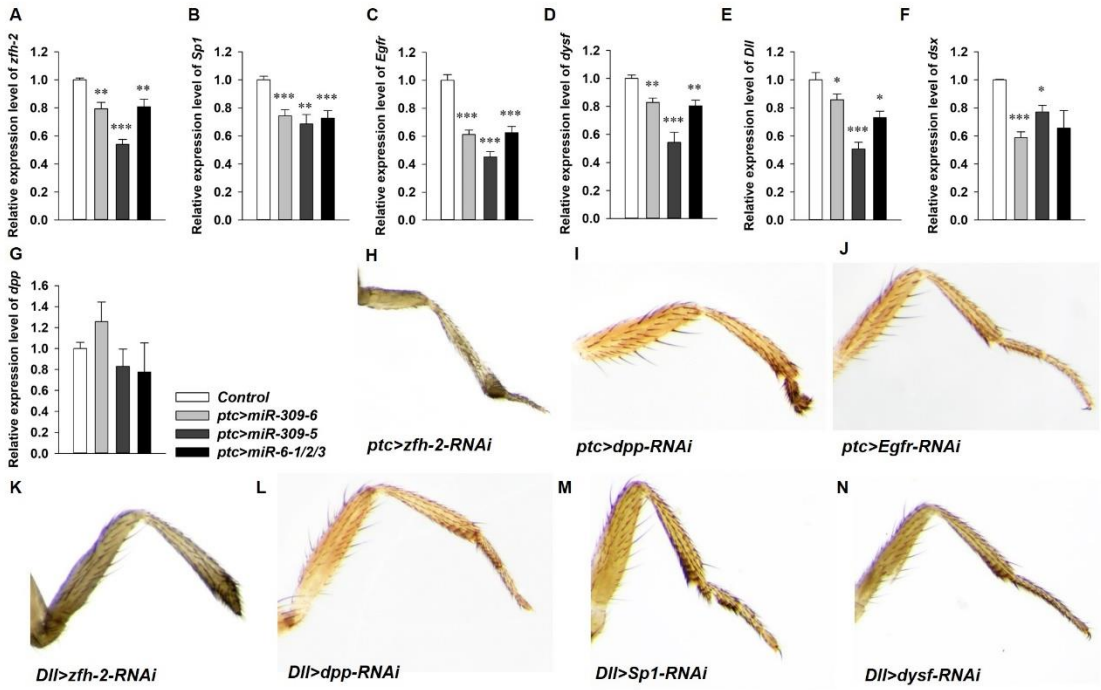


Fig. 4 Expression of whole miR-309-6 cluster, miR-309-5 cluster and miR-6-1/2/3 regulate similar gene regulatory networks of leg development. A-G) Relative expression levels of *zfh-2*, *Sp1*, *Egfr*, *dysf*, *Dll*, *dsx* and *dpp* in leg discs of *ptc>miR-309-6*, *ptc>miR-309-5* and *ptc>miR-6-1/2/3*. Values represent mean ± S.E.M, * $p<0.05$, ** $p<0.005$ and *** $p<0.001$. H-J) Aberrant tarsal leg phenotypes created by *ptc* GAL4 driven expression of *zfh-2*-RNAi, *dpp*-RNAi and *Egfr*-RNAi. K-N) Tarsal leg deformities created by *Dll* GAL4 driven expression of *zfh-2*-RNAi, *dpp*-RNAi, *Sp1*-RNAi and *dysf*-RNAi.

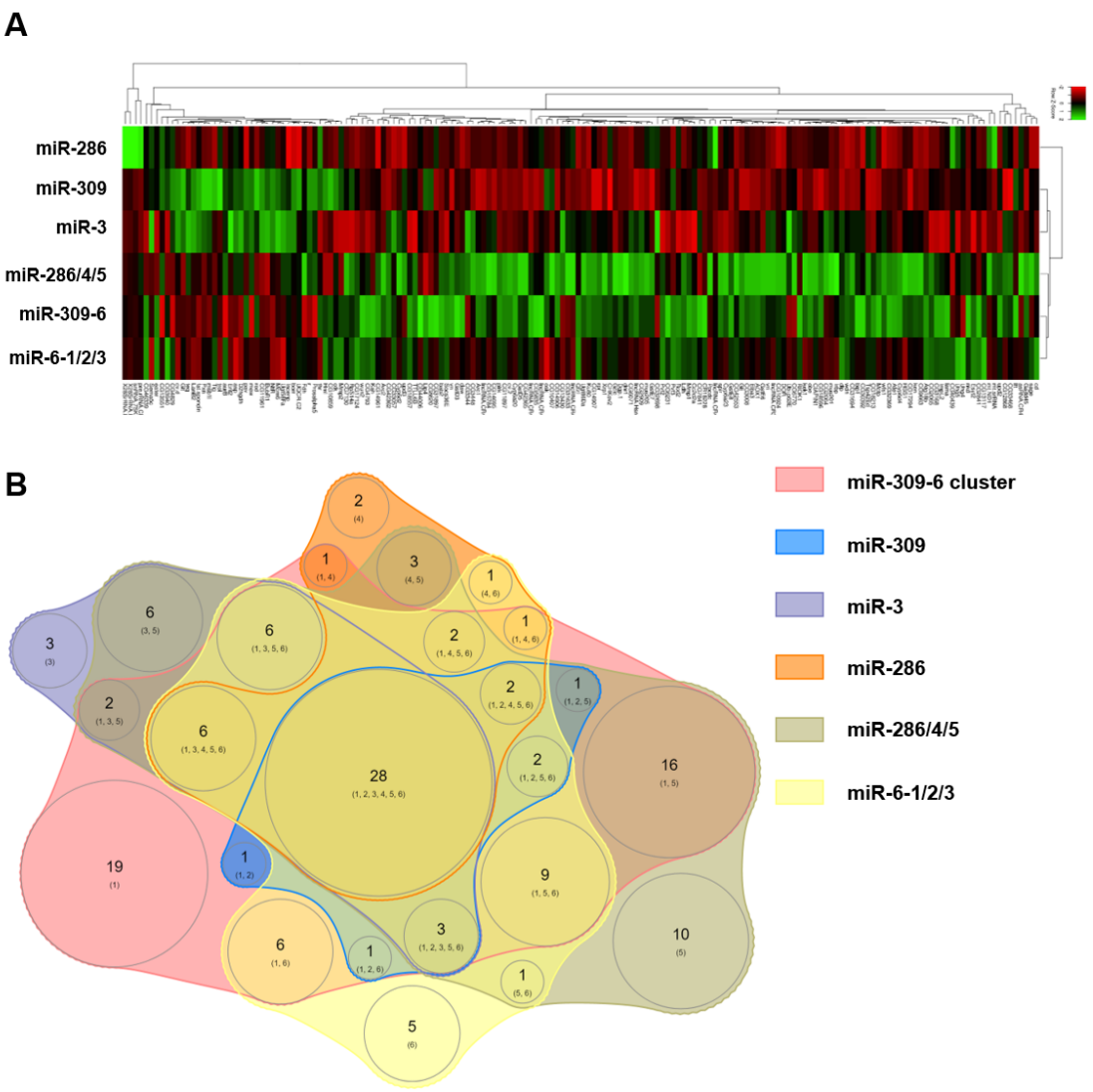


Fig. 5 Cell transfection assays reveal common and unique targets of microRNAs in the miR-309-6 cluster. A) Transcriptome analyses of differential gene expression after transfection of different combinations of microRNAs from the cluster to *Drosophila* S2 cells. B) Venn-diagram showing the genes commonly and uniquely regulated by different combinations of microRNAs in the cluster. Numbers shown within the circles represent the common or unique genes being regulated by microRNA cluster members. Details of the gene list and their expression comparing to the controls are listed in Supplementary data S10.

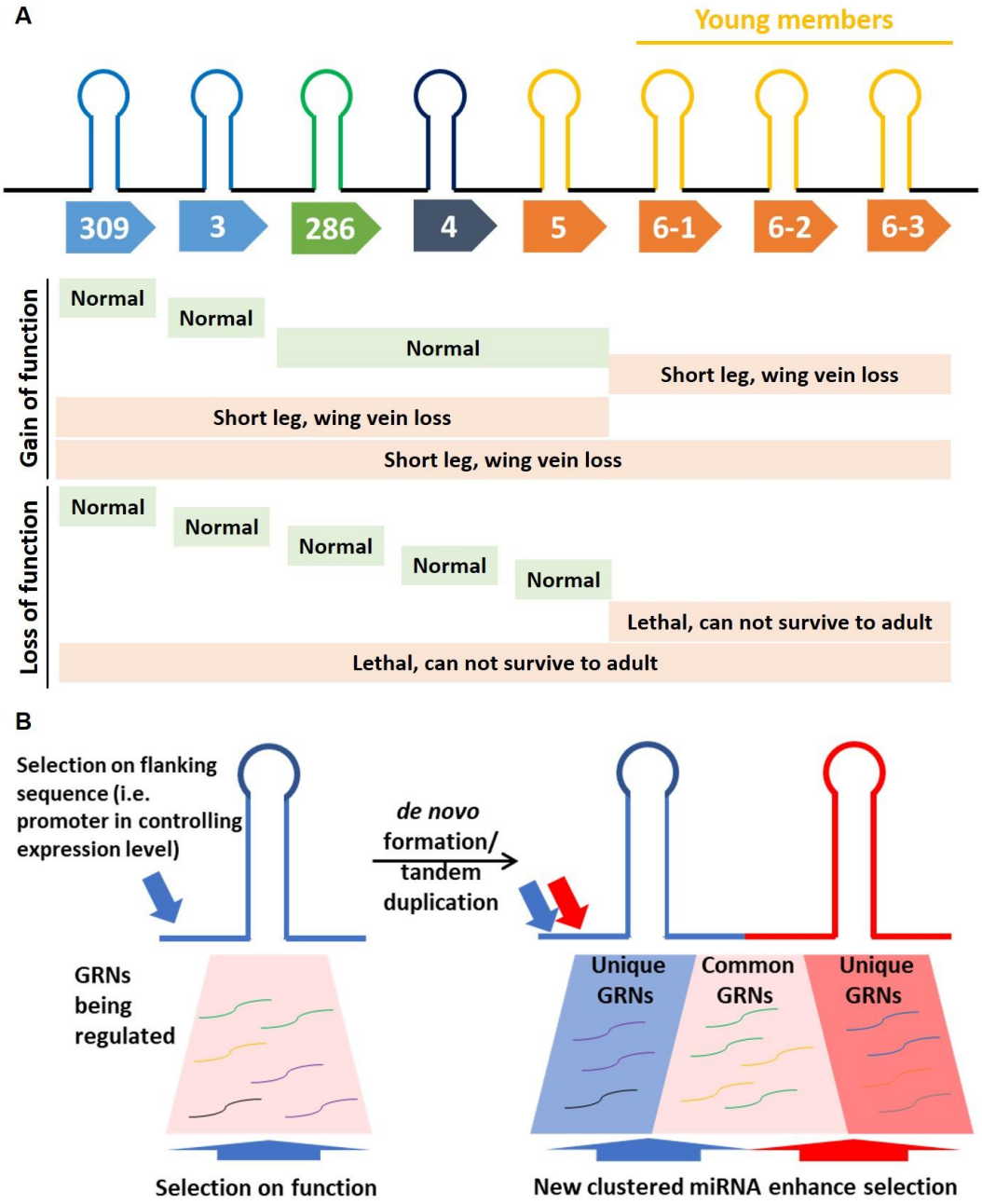


Fig. 6 Varying selection forces acting by new and old microRNAs in a genomic cluster. A) Summary of the phenotypic results obtained in the gain-of-function and loss-of-function experiments of miR-309-6 cluster microRNAs in *Drosophila melanogaster*. B) During the formation of new microRNA in a genomic cluster, negative selection forces are reinforced. For details, please refer to main text. The same colour decodes mRNAs involved in the same gene regulatory network (GRN). The arrows represent the selection forces on either sequence/function.

Supplementary data / information

S1. Sequence alignment of microRNAs in miR-309-6 cluster.

S2. Prediction of microRNA-309-6 cluster targets in different *Drosophila* genomes.

S3. Expression of miR-309-6 cluster microRNAs in *D. melanogaster* L3 leg discs.

S4. Tarsal segments of *Dll* and *ptc* GAL4 driven expression of miR-309-6, miR-309-5 and miR-6-1/2/3.

S5. Tarsal segment numbers of *Dll* and *ptc* GAL4 driven expression of miR-309-6, miR-309-5 and miR-6-1/2/3 mutants.

S6. UAS-gene and UAS-RNAi fly lines used in this study.

S7. Leg and wing pictures of loss-of-function of individual microRNA from miR-309-6 cluster in either the *Distal-less* (A-F) or *patched* (G-R) expressing cells.

S8. Differential expression of genes with over 2-fold changes and leg patterning related genes between GAL4-ptc>miR-309-6 (mutant) and control (GAL4-ptc) leg discs.

S9. Phenotypes of gain- and loss- of-function of selected differentially expressed genes.

S10. List of genes being regulated by microRNA cluster members *in vitro*.

S11. Gene Ontology (GO) enrichment analysis

S12. Primer information used in this study.

Author Contributions

JHLH conceived and supervised the study. ZQ, WCY, HYY, CWCY, WN, AYPW, IHTL, JHLH carried out the experimental works. NWYW cloned the cluster into pUAST vector. FKMC obtained homozygous red-eye lines after pUAST vector injection by BestGene. ZQ, WCY, HYY, CWCW, WN, AYPW, IHTL, FKMC, TFC, KFL, SZ, KHC, DEKF, WGB, JHLH wrote the manuscript.

Acknowledgements

The authors thank KB Wong and HM Lam for discussion. This research was supported by the Hong Kong Research Grant Council GRF Grant (14103516), The Chinese University of Hong Kong Direct Grant (4053248), and TUYF Charitable Trust (6903957) (JHLH).

Competing interests

The authors declare no conflict of interests.

References

Agarwal, V; Subtelny, AO; Thiru, P; Ulitsky, I; Bartel, DP. Predicting microRNA targeting efficacy in *Drosophila*. Genome biology. 2018. 19, 152.

Altuvia, Y; Landgraf, P; Lithwick, G; Elefant, N; Pfeffer, S; Aravin, A; Brownstein, MJ; Tuschl, T; Margalit, H. Clustering and conservation patterns of human microRNAs. NUCLEIC ACIDS RESEARCH. 2005. 33(8), 2697-2706.

Bartel, DP. MicroRNAs: Target recognition and regulatory functions. CELL. 2009. 136(2), 215-233.

Bartel DP. Metazoan MicroRNAs. Cell. 2018. 173(1), 20-51.

Betel, D; Koppal, A; Agius, P; Sander, C; Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. GENOME BIOLOGY. 2010. 11(8), R90.

Biemar, F; Zinzen, R; Ronshaugen, M; Sementchenko, V; Manak, JR; Levine, MS. Spatial regulation of microRNA gene expression in the Drosophila embryo. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 2005. 102(44), 15907-15911.

Bushati, N; Stark, A; Brennecke, J; Cohen, SM. Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. CURRENT BIOLOGY. 2008. 18(7), 501-506.

Casares, F; Mann, RS. The ground state of the ventral appendage in Drosophila. SCIENCE. 2001. 293(5534), 1477-1480.

Córdoba, S; Estella, C. The bHLH-PAS Transcription Factor Dysfusion Regulates Tarsal Joint Formation in Response to Notch Activity during *Drosophila* Leg Development. PLoS genetics. 2014. 10, e1004621.

Córdoba, S; Requena, D; Jory, A; Saiz, A; Estella, C. The evolutionarily conserved transcription factor Sp1 controls appendage growth through Notch signaling. Development. 2016. 143, 3623-3631.

Chen, YW; Song, S; Weng, R; Verma, P; Kugler, JM; Buescher, M; Rouam, S; Cohen,

SM. Systematic study of Drosophila microRNA functions using a collection of targeted knockout mutations. DEVELOPMENTAL CELL. 2014. 31(6), 784-800.

Dambal, S; Shah, M; Mihelich, B; Nonn, L. The microRNA-183 cluster: the family that plays together stays together NUCLEIC ACIDS RESEARCH. 2015. 43(15), 7173-7188.

Deschamps, J; Duboule, D. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. GENES & DEVELOPMENT. 2017. 31(14), 1406-1416.

Doron, S; Melamed, S; Ofir, G; Leavitt, A; Lopatina, A; Keren, M; Amitai, G; Sorek, R. Systematic discovery of antiphage defense systems in the microbial pangenome. SCIENCE. 2018. 359(6379), eaar4120.

Ebert, MS; Sharp, PA. Roles for microRNAs in conferring robustness to biological processes. CELL, 2012. 149(3), 515-524.

Flynt, AS; Lai, EC. Biological principles of microRNA-mediated regulation: shared themes amid diversity. NATURE REVIEWS GENETICS. 2008. 9(1), 831-842.

Fromm, B; Billipp, T; Peck, LE; Johansen, M; Tarver, JE; King, BL; Newcomb, JM; Sempere, LF; Flatmark, K; Hovig, E; Peterson, KJ Bassler, BL. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. Annual Review of Genetics. 2015. 49.

Galindo, MI; Bishop, SA; Greig, S; Couso, JP. Leg Patterning Driven by Proximal-Distal Interactions and EGFR Signaling. Science 2002. 297, 256-259.

Griffiths-Jones, S; Hui, JHL; Marco, A; Ronshaugen, M. MicroRNA evolution by arm switching. EMBO REPORTS. 2011. 12(2), 172-177.

Guarner, A.; Manjón, C.; Edwards, K.; Steller, H.; Suzanne, M.; Sánchez-Herrero, E. The zinc finger homeodomain-2 gene of *Drosophila* controls Notch targets and regulates apoptosis in the tarsal segments. Developmental Biology 2014. 385, 350-365.

Hausser, J; Zavolan, M. Identification and consequences of miRNA-target interactions - beyond repression of gene expression. NATURE REVIEWS GENETICS. 2014.

15(9), 599-612.

Hornstein, E; Shomron, N. Canalization of development by microRNAs. NATURE GENETICS. 2006. 38(6), S20-S24.

Hui, JHL; Marco, A; Hunt, S; Melling, J; Griffiths-Jones, S; Ronshaugen, M. Structure, evolution and function of the bi-directionally transcribed iab-4/iab-8 microRNA locus in arthropods. NUCLEIC ACIDS RESEARCH. 2013. 41(5), 3352-3361.

Hui, JHL; McDougall, C; Monteiro, AS; Holland, PWH; Arendt, D; Balavoine, G; Ferrier, DEK. Extensive chordate and annelid macrosynteny reveals ancestral homeobox gene organization. MOLECULAR BIOLOGY AND EVOLUTION. 2012. 29(1), 157-165.

Kim, VN; Nam, JW. Genomics of microRNA. TRENDS IN GENETICS. 2006. 22(3), 165-173.

Kim, YK; Yu, J; Han, TS; Park, SY; Namkoong, B; Kim, DH; Hur, K; Yoo, MW; Lee, HJ; Yang, HK; Kim, VN. Functional links between clustered microRNAs: suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer. NUCLEIC ACIDS RESEARCH. 2009. 37(5), 1672-1681.

Kojima, T. The mechanism of Drosophila leg development along the proximodistal axis. DEVELOPMENT GROWTH & DIFFERENTIATION. 2004. 46(2), 115-129.

Kozomara, A; Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data NUCLEIC ACIDS RESEARCH. 2014. 42(D1), D68-73.

Lagos-Quintana, M; Rauhut, R; Lendeckel, W; Tuschl, T. Identification of novel genes coding for small expressed RNAs. SCIENCE. 2001. 294(5543), 853-858.

Levis, NA; Pfennig, DW. Evaluating 'Plasticity-First' evolution in nature: key criteria and empirical approaches. TRENDS IN ECOLOGY & EVOLUTION. 2016. 31(7), 563-574.

Marco, A; Hui, JHL; Ronshaugen, M; Griffiths-Jones, S. Functional shifts in insect

microRNA evolution. GENOME BIOLOGY AND EVOLUTION. 2010. 2, 686-696.

Marco, A; Ninova, M; Ronshaugen, M; Griffiths-Jones, S. Clusters of microRNAs emerge by new hairpins in existing transcripts. NUCLEIC ACIDS RESEARCH. 2013. 41(16), 7745-7752.

Mohammed, J; Siepel, A; Lai, EC. Diverse modes of evolutionary emergence and flux of conserved microRNA clusters RNA. 2014. 20(12), 1850-1863.

Ninova, M; Ronshaugen, M; Griffiths-Jones, S. Fast-evolving microRNAs are highly expressed in the early embryo of Drosophila virilis. RNA. 2014. 20(3), 360-372.

Ninova, M., Ronshaugen, M; Griffiths-Jones, S. MicroRNA evolution, expression, and function during short germband development in *Tribolium castaneum*. Genome Research. 2016. 26, 85-96.

Nojima, M; Matsui, T; Tamori, A; Kubo, S; Shirabe, K; Kimura, K; Shimada, M; Utsunomiya, T; Kondo, Y; Iio, E; Naito, Y; Ochiya, T; Tanaka, Y. Global, cancer-specific microRNA cluster hypomethylation was functionally associated with the development of non-B non-C hepatocellular carcinoma. MOLECULAR CANCER. 2016. 15, 31.

Peterson, KJ; Dietrich, MR; McPeek, MA. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. BIOESSAYS. 2009. 31 (7), 736-747.

Pinzon, N; Li, B; Martinez, L; Sergeeva, A; Presumey, J; Apparailly, F; Seitz, H. microRNA target prediction programs predict many false positives. GENOME RESEARCH. 2017. 27(2), 234-245.

Posadas, DM; Carthew, RW. MicroRNAs and their roles in developmental canalization CURRENT OPINION IN GENETICS & DEVELOPMENT. 2014. 27(1), 6.

Potter, CJ; Luo, LQ. Splinkerette PCR for Mapping transposable elements in Drosophila. PLOS ONE. 2010. 5(4), e10168.

Qu, Z; Bendena, WG; Nong, WY; Siggens, KW; Noriega, FG; Kai, ZP; Zang, YY; Koon, AC; Chan, HYE; Chan, TF; Chu, KH; Lam, HM; Akam, M; Tobe, SS; Hui, JHL. MicroRNAs regulate the sesquiterpenoid hormonal pathway in Drosophila and other arthropods. PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES. 2017. 284, 1869.

Reczko, M; Maragkakis, M; Alexiou, P; Grosse, I; Hatzigeorgiou, AG. Functional microRNA targets in protein coding sequences. BIOINFORMATICS. 2012. 28(6), 771-776.

Tanzer, A; Stadler, PF. Molecular evolution of a microRNA cluster. JOURNAL OF MOLECULAR BIOLOGY. 2004. 339(2), 327-335.

Trapnell, C; Roberts, A; Goff, L; Pertea, G; Kim, D; Kelley, DR; Pimentel, H; Salzberg, SL; Rinn, JL; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks NATURE PROTOCOLS. 2012. 7(3), 562-578.

Ventura, A; Young, AG; Winslow, MM; Lintault, L; Meissner, A; Erkeland, SJ; Newman, J; Bronson, RT; Crowley, D; Stone, JR; Jaenisch, R; Sharp, PA; Jacks, T. Targeted deletion reveals essential and overlapping functions of the miR-17 similar to 92 family of miRNA clusters. CELL. 2008. 132(5), 875-886.

Voskarides K. 'Plasticity-First' evolution and the role of miRNAs: A comment on Levis and Pfennig. Trends Ecol Evol. 2016. 31(11), 816-817.

Wang, J; Haubrock, M; Cao, KM; Hua, X; Zhang, CY; Wingender, E; Li, J. Regulatory coordination of clustered microRNAs based on microRNA-transcription factor regulatory network. BMC SYSTEMS BIOLOGY. 2011. 5, 199.

Wang, YR; Luo, JJ; Zhang, H; Lu, J. microRNAs in the same clusters evolve to coordinately regulate functionally related genes. MOLECULAR BIOLOGY AND EVOLUTION. 2016. 33(9), 2232-2247.

Wu, CI; Shen, Y; Tang, T. Evolution under canalization and the dual roles of microRNAs-A hypothesis. GENOME RESEARCH. 2009. 19(5), 734-743.

Yuan, XY; Liu, CN; Yang, PC; He, SM; Liao, Q; Kang, SL; Zhao, Y. Clustered microRNAs' coordination in regulating protein-protein interaction network. BMC SYSTEMS BIOLOGY. 2009. 3, 65.

Zhao, YX; Shen, X; Tang, T; Wu, CI. Weak regulation of many targets is cumulatively powerful - An evolutionary perspective on microRNA functionality. MOLECULAR BIOLOGY AND EVOLUTION. 2017. 34(12), 3041-3046.

Zhou, L; Lim, M.Y.T; Kaur, P; Saj, A; Bortolamiol-Becet, D; Gopal, V; Tolwinski, N; Tucker-Kellogg, G; Okamura, K. Importance of miRNA stability and alternative primary miRNA isoforms in gene regulation during *Drosophila* development. eLife. 2018. *7*, e38389.