Regulatory RNAs in the light of *Drosophila* genomics

Antonio Marco

Abstract

Many aspects of gene regulation are mediated by RNA molecules. However, regulatory RNAs have remained elusive until very recently. At least three types of small regulatory RNAs have been characterized in *Drosophila*: microRNAs (miRNAs), piwi-interacting RNAs and endogenous siRNAs. A fourth class of regulatory RNAs includes known long non-coding RNAs such as roX1 or bxd. The initial sequencing of the *Drosophila melanogaster* genome has served as a scaffold to study the transcriptional profile of an animal, revealing the complexities of the function and biogenesis of regulatory RNAs. The comparative analysis of 12 *Drosophila* genomes has been crucial for the study of microRNA evolution. However, comparative genomics of other RNA regulators is confounded by technical problems: genomic loci are poorly conserved and frequently encoded in the heterochromatin. Future developments in genome sequencing and population genomics in *Drosophila* will continue to shed light on the conservation, evolution and function of regulatory RNAs.

Keywords: *Non-coding RNA; miRNA; piRNA; siRNA; transposable elements; gene regulation*

REGULATORY RNAs

Early models of gene expression envisioned a system of transcriptional regulation mediated by RNA molecules [1, 2]. This regulatory role of RNA molecules was largely abandoned as transcription factors were characterized, leading to a transcription–factor–centered view of gene regulation [3, 4]. After the discovery of RNA interference (RNAi) in eukaryotes (reviewed earlier [5]), the idea of regulatory RNAs was resurrected in a different form: some RNA molecules may be down-regulating other RNA molecules by sequence complementarity. This type of antisense RNA-mediated regulation had been already described in prokaryotes [6]. When microRNAs (miRNAs) were first observed in the roundworm *Caenorhabditis elegans*, a mechanism of gene down-regulation by RNA–RNA complementarity in eukaryotes became apparent [7, 8]. We currently know that multiple types of RNAs have important regulatory functions in the cell, and that they are widespread in animal genomes. Current models of gene regulation integrate the RNA component, providing a much more complex picture than we had two decades ago.

Drosophila melanogaster has dominated the field of genetics for over a century. Not surprisingly, genes regulating animal development were first discovered in this species [9]. Early investigations by Ed Lewis showed that multiple loci controlling the fly body patterning were closely linked in a single genomic region, the bithorax complex (BX-C, see [10] and references therein). These loci are located in the genome in the same order as they are spatially expressed in the fly, and they were named after the anatomic region affected in their mutants (Figure 1). Lewis initially characterized 8 genes in the BX-C complex, but only three of them coded for proteins: *Ubx*, *abd-A* and *Abd-B* [11]. Transcripts from the other loci were identified much later [12]. We currently know that three of these transcripts are regulatory RNAs: one long non-coding RNA, bxd and two miRNAs, iab-4 and iab-8 (Figure 1). The pioneering work by Ed Lewis on the BX-C complex in *Drosophila*, therefore, represented the

Corresponding author. Antonio Marco, Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, UK. Tel: +44 (0) 1612751565; Fax: +44 (0) 1612751505; E-mail: antonio.marco@manchester.ac.uk

Antonio Marco is a Postdoctoral Research Fellow at the University of Manchester. He obtained his PhD at the University of Valencia and postdoctoral training at Arizona State University. His research interests are in gene regulation and evolutionary genomics.
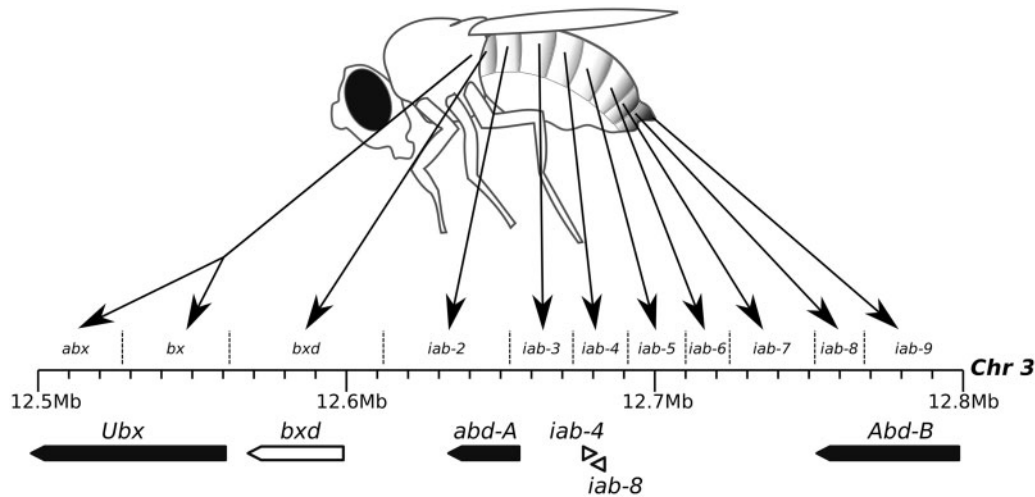
Figure 1: *Drosophila* Bithorax Complex and associated loci. Genetic loci associated with the two thoracic and nine abdominal *Drosophila* segments from early genetic experiments. Boxes depict genes annotated in FlyBase. Black boxes are protein-coding genes, and white boxes are non-protein-coding genes.

Table 1: Non-protein-coding RNAs annotated to the *Drosophila melanogaster* genome

	Number of loci*
Transfer RNAs	318
Small nuclear RNAs	43
Small nucleolar RNAs	281
Ribosomal RNAs	116
MicroRNAs	240
Other non-coding RNAs	577

*Annotated in FlyBase, 2 March 2012.

first functional analysis of regulatory RNAs in animals.

The *D. melanogaster* genome sequence has been particularly useful to study regulatory sequences [13]. FlyBase [14] catalogues about 1500 non-protein coding loci (Table 1). miRNAs are the only class of regulatory RNAs indexed in FlyBase. Other known and putative regulatory RNAs are included in the long non-coding RNA category. Genetic loci encoding other short regulatory RNAs such as piwi-interacting RNAs (piRNAs) or endogenous small interfering RNAs (siRNAs) are currently not even catalogued. This review focuses on how *Drosophila* genomics has contributed to the analysis of regulatory RNAs, and how future developments will provide a better understanding of their function and evolution.

microRNAs

miRNAs are key regulators of gene expression at the post-transcriptional level. They bind to target transcripts by sequence complementarity inducing either degradation or translational repression [15, 16]. miRNA biogenesis is well understood [Figure 2 (top-left)]. A miRNA locus is transcribed into a primary miRNA, which is processed by the RNase complex Drosha/Pasha producing a precursor hairpin [16]. Precursor hairpins are further cleaved in the cytoplasm by DCR-1 and LOQS (Table 2), the products of the genes *Dicer*-1 and *loquacious* [17]. The result is a double-stranded RNA molecule (miRNA duplex in Figure 2) with an approximate length of 21 nt. One of the arms of the miRNA duplex typically becomes the mature sequence. Partial complementary between the mature miRNA and its target mediates the translational repression in association with Argonaute 1 (AGO1). When the complementarity between the miRNA and the target is perfect, the miRNA enters the RNAi pathway, and the targeted transcript is instead, degraded by Argonaute 2 (AGO2) [18].

The first miRNA ever characterized was lin-4 in *C. elegans* [7, 8]. Lin-4 remained as a unique type of regulator until, a few years later, a second miRNA was characterized: let-7. Like lin-4, let-7 was first identified in *C. elegans* [19]. However, by that time, the genome of *Drosophila* was already available [20], and let-7 was identified by sequence similarity in this species, as well as in other animals with ongoing
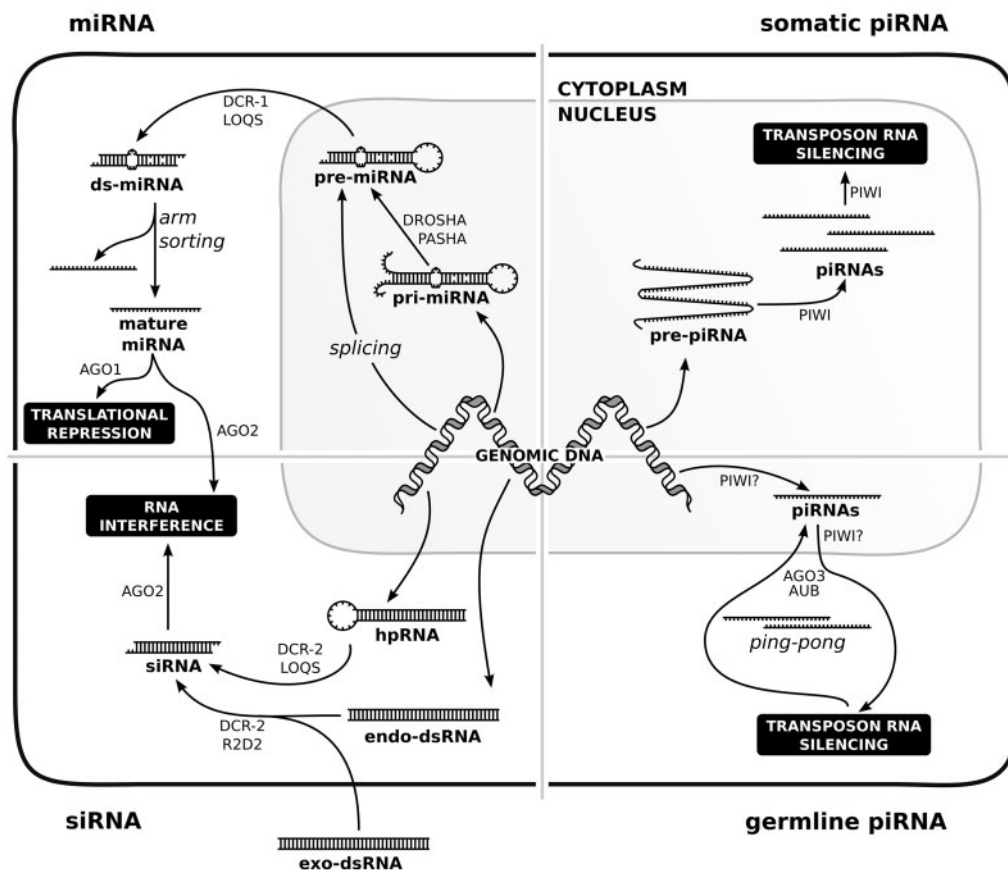
Figure 2: Biogenesis of *Drosophila* small regulatory RNAs. miRNA: primary microRNAs (pri-miRNA) are transcribed from the genome and processed by DROSHA/PASHA into precursor hairpins (pre-miRNA). Some miRNAs (mirtrons) are spliced from introns by the spliceosome machinery bypassing the action of DROSHA/PASHA. pre-miRNAs are processed in the cytoplasm by DCR-I/LOQS producing double-stranded miRNAs (ds-miRNA), from which one of the arms in sorted and loaded into AGOI or AGO2 inducing either translational repression or RNA interference, respectively. endo-siRNA: long endogenous double-stranded RNAs (endo-dsRNA) are encoded in transposon-rich genomic locations, and they are processed by DCR-2/R2D2 into double stranded siRNA. Exogenous dsRNAs follow the same path as endo-siRNAs. Other siRNAs are produced from the processing of genome encoded long hairpins (hpRNA) by DCR-2/LOQS. siRNAs trigger the RNA interference response in association with AGO2. Somatic piRNA: Long piRNA clusters are transcribed into precursors (pre-piRNA), which are cleaved by PIWI generating small piRNAs. PIWI/piRNA complexes mediate the silencing of RNA transposons in the nucleus. Germline piRNA: AGO3/AUB mediate the cleavage of genomic encoded piRNAs and RNA transposons in the cytoplasm in a feed-back loop called the ping-pong mechanism. PIWI is required in this pathway, but its role has not been clarified so far.

genome projects [21]. Since both lin–4 and let–7 control developmental timing, they were classified as small temporal RNAs (stRNAs). In a collective effort, three groups cloned multiple stRNAs from *D. melanogaster*, *C. elegans* and humans [22–24], and introduced the term microRNA.

The initial cloning of miRNAs from 22 *Drosophila* loci [22] showed early that miRNAs are often clustered in the genome. The comparative analysis of miRNAs in *Drosophila* was crucial to establish the basis of the computational prediction of small RNAs [25]. By first screening the genome for potential miRNA loci, the cloning experiments became more specific (i.e. less expensive). Likewise, the prediction of miRNA targets was first modelled in *D. melanogaster* using this initial set [26, 27]. Both prediction of miRNA loci and targets had relied on conservation in a second available *Drosophila* genome sequence: *D. pseudoobscura*. Because of the small size of miRNAs and their target sites, the proper study of miRNAs required a more extensive collection of closely related genomes. This opportunity came

Table 2: *Drosophila melanogaster* loci encoding for enzymes involved in small RNA biogenesis

Gene	Protein	Small RNAs associated
Dicer-1	DCR-1	microRNAs
Dicer-2	DCR-2	siRNAs
Loquacious	LOQS	microRNAs, siRNAs
r2d2	R2D2	siRNAs
Drosha	DROSHA	microRNAs
Partner of drosha	PASHA	microRNAs
Argonaute-1	AGOI	microRNAs
Argonaute 2	AGO2	siRNAs, microRNAs
Argonaute 3	AGO3	germline piRNAs
aubergine	AUB	germline piRNAs
piwi	PIWI	somatic and germline piRNAs

with the sequencing, assembly and comparative analysis of the 12 *Drosophila* genomes [28]. Additionally, the breakthrough of high–throughput sequencing allowed small RNAs characterization without the need for cloning. The combination of computational prediction of miRNAs based on comparative genomics and the fast validation of candidates by deep sequencing resulted in a dramatic expansion in the number of known miRNAs in *Drosophila* (Figure 3, [29–31]). These analyses revealed additional miRNA features: (i) as suspected, the mature functional sequence of a miRNA is more conserved than the precursor hairpin [28]; (ii) some miRNAs (mirtrons) bypass the action of Drosha during their biogenesis, being processed as introns by the splicing machinery [32, 33]; (iii) the comparison of closely related species improves the identification of functional miRNA target sites [34]. More recently, as a part of the modEncode project [35], the profile of small RNAs has been thoroughly investigated in multiple tissues and developmental stages, permitting the discovery of additional miRNAs [36]. miRBase [37], the repository for all miRNAs sequences, currently catalogues 240 loci encoding miRNAs in *D. melanogaster.*

The systematic characterization of miRNAs in multiple *Drosophila* genomes has provided an excellent opportunity to study the evolutionary dynamics of these tiny regulators [38, 39]. Within the *Drosophila* lineage, miRNAs appear to have high turnover rates [38, 40]. Comparison with other species also shows that only a few miRNAs are conserved among the animals [41, 42]. However, a number of striking observations have been made

from the deep sequencing of miRNAs from multiple species: (i) Highly conserved miRNAs can change their function during evolution by modifying their Dicer/Drosha cleavage sites [42, 43]; (ii) functional changes can also occur by changing the arm of the precursor that will produce the mature miRNA [43–45]. Specifically, in *D. melanogaster,* ∼20% of the conserved miRNAs produce a different mature sequence than their *Tribolium castaneum* orthologue [43]; (iii) Clusters of co-transcribed miRNAs change dynamically during evolution [43, 46]. All these changes are likely to affect the miRNA function. Undoubtedly, the analysis of more arthropods will provide a clearer picture of miRNA functional evolution.

ENDOGENOUS siRNAs
The injection of double-stranded RNAs to induce targeted gene silencing has been used extensively in the genetic analysis of plants and animals [5]. This mechanism, called RNAi, is now well understood [5, 47]. Long exogenous double-stranded RNAs are cleaved in the cell into double-stranded RNA molecules of about 21 nt, known as siRNAs. This cleavage is mediated by the DCL-2 Dicer family member in *Drosophila* [Figure 2 (bottom-left)]. siRNAs bind to full complementary sequences within the target inducing their degradation. In *Drosophila,* this degradation is mediated by AGO2. The first endogenous (endo-) siRNAs (i.e. encoded in the genomic sequence) in animals were found in *C. elegans* [48], followed 2 years later by their discovery in *Drosophila* [49–51]. Strikingly, experiments in *Drosophila* revealed the existence of two independent genomic sources of endo-siRNAs [Figure 2 (bottom-left)]. Some siRNAs are generated from long double-stranded RNA molecules (endo-dsRNAs) and are processed by the same enzymes known to cleave exogenous siRNAs: DCR-2 and R2D2 [49, 50]. Endo-dsRNAs are mainly composed of transposon–derived sequences. Other siRNAs are derived from long RNA hairpins (hpRNAs), and instead of R2D2, the processing is mediated by LOQS, the partner of DCR-1 in the miRNA pathway [49, 51]. miRNA and siRNA pathways are thus intertwined, sharing at least two proteins: AGO2 and LOQS [Table 2; Figure 2 (left)].

Unlike miRNAs, endo-siRNAs are mostly derived from repetitive regions. Their detection therefore, requires the mapping of short sequenced reads to highly repetitive genomic regions. Perhaps,
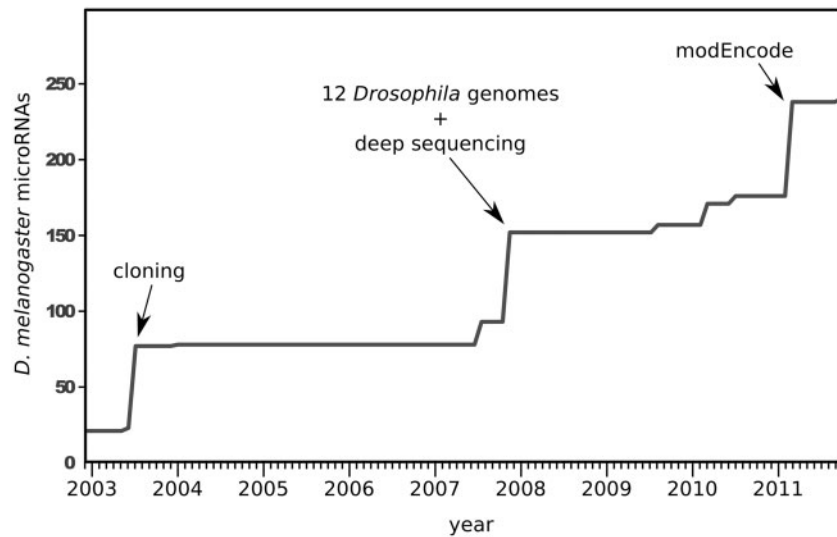
Figure 3: The number of *D. melanogaster* miRNAs annotated in miRBase.

for that reason, the characterization of *Drosophila* endogenous siRNAs (and piRNAs, see below) occurred after the annotation of the heterochromatic regions of the genome, which are largely composed of nested transposable elements [52]. Heterochromatin sequences of other *Drosophila* species have been identified, but an assembled heterochromatic genome only exixts, so far, for *D. melanogaster* [52, 53]. This imposes a limit on the study of small RNAs other than miRNAs. Consequently, the identification of orthologous siRNA loci among drosophilids has not been very successful. A couple of exceptions can be found for siRNA loci overlapping conserved protein coding genes, such as *cis*-NAT (antisense to *tkv*) [54] and hp-CG4068 (antisense to CG4068) [51]. The conservation is, however, limited to closely related species. The study of neighbouring genes to detect orthologous siRNA loci and a better characterization of heterochromatin across the 12 *Drosophila* genomes will tell us more about the origin of these small regulators.

piRNAs
Both mature miRNAs and siRNAs have a size of ~21 nt. During a systematic cloning of small RNAs in *Drosophila*, a class of RNAs slightly longer than a miRNA was noticed [55]. These sequences were described as repeat-associated small interfering RNAs (rasiRNAs). rasiRNAs were soon identified as a particular type of piwi-interacting RNAs (piRNAs) [56]. piRNAs are transcribed from genomic piRNA-clusters, mostly composed of inactive transposable elements (TEs, reviewed in [57]). Before the discovery of piRNAs, co-suppression of TEs had been already described [58, 59], and a role of ancestral transposon insertions in silencing novel TEs from the same family was also proposed [59, 60]. Moreover, it has been suggested that clusters of TEs may form a co-suppression network that down-regulates the expression of other TEs [61]. Indeed, piRNAs are known now to be an important defensive mechanism against transposons [62–65]. This suggests that, most likely, a complex TE co-suppression network based on piRNAs does exist.

The discovery of the piRNA pathway has revealed the nature of two previously known phenomena in *Drosophila*. First, the maternal effect locus *flamenco* induces the silencing of *gypsy* transposons [66]. The *flamenco* locus is actually a piRNA-cluster [63]. Second, hybrid dysgenesis in *Drosophila* is produced by the massive mobilization of P-elements in the germ line [67]. The piRNA-mediated response is behind this classic phenomena [68]. There may be, however, two independent piRNA pathways [64, 69, 70] [Figure 2 (right)]. A somatic pathway happens in the nucleus of the follicle cells of the ovary. In somatic cells, large piRNA-clusters are transcribed and then processed by PIWI into small piRNAs [Figure 2 (top-right)]. PIWI/piRNA complexes directly target transposable elements. The *flamenco* locus is of this kind. An independent germ-line pathway occurs mainly in the cytoplasm of the nurse cells [Figure 2 (bottom-right)]. In this case, genomic piRNA transcripts and targeted

(transposon-derived) sequences induce the degradation of each other through the proteins AUB and AGO3. A feed-back loop, called 'ping-pong', is established, generating a characteristic pattern of sense/antisense piRNAs overlapping each other by 10 nucleotides. Functional analyses showed that PIWI is also involved in this second pathway, but its role is not yet clear [64].

The piRNA pathway is conserved in animals (reviewed previously [47]). Consequently, other *Drosophila* species should code for piRNAs. However, the comparative analysis of conserved piRNAs between drosophilids is, as in the case of endo-siRNAs, problematic. The *flamenco* locus is conserved between *D. melanogaster*, *D. yakuba* and *D. erecta* and it encodes anti-sense piRNAs that target transposons of the *gypsy* family across the *Drosophila* lineage, although the specific transposons that are targeted vary from species to species [69]. Clusters of TEs, from which piRNAs derive, tend to be located in the heterochromatin [57]. As in the case of endogenous siRNAs, the analysis of the heterochromatic part of the genome is crucial to further investigate the origin and evolutionary dynamics of piRNAs.

Although the primary function of piRNAs is the defence against TEs, a role in chromatin regulation has also been proposed [56, 71]. Interestingly, piRNAs are likely to regulate *Drosophila* telomeric chromatin, which is mostly composed of retrotransposons (reviewed in [72]). In agreement with these observations, specific piRNA targets in the *Drosophila* telomeric retrotransposon HeT-A have been identified [73]. These targets are conserved in other *Drosophila* species [73]. Other instances of piRNA-mediated chromatin regulation involving TEs are still unknown.

LONG NON-CODING RNAs

Long non-coding RNAs (lncRNAs) are often defined as transcripts longer than 200 nt with little or no protein-coding capacity [74]. In practice, an RNA molecule is considered to be a lncRNA if it cannot be ascribed to any other class of non-protein coding RNAs (Table 1). As discussed in the previous sections, small regulatory RNAs show signatures of enzymatic processing in their mature products (mainly conserved size and RNase cleavage sites), facilitating their identification in the genome. However, lncRNAs have no recognizable signatures and their characterization has been based, almost

Table 3: *Drosophila melanogaster* loci producing long non-coding RNAs

Gene (Symbol)	RNA length[a]	Function	Reference
bxd	1755	Ubx regulation	[12]
Hsrω	14 084	Heat–shock stress response	[77]
pgc	1167	Germ cell transcriptional inhibition	[75]
roX1	3748	Dosage compensation	[76]
roX2	1368	Dosage compensation	[76]
sphinx	1280	Courtship behaviour	[78]
yar	1488	Sleep behaviour	[79]

Note: [a]Length of the longest RNA transcript.

exclusively, on transcriptional analyses. The first characterized lncRNA in *Drosophila* was bxd (Figure 1), a non-protein-coding transcript that regulates the expression of *Ubx*. Paradoxically, when bxd transcripts were first characterized, it was proposed that they encode small regulatory proteins rather than acting as long regulatory RNAs [12]. Soon after, other lncRNAs were identified in *Drosophila* (Table 3). The successful differentiation of female germline cells in the ovary requires the RNA product of the gene *pgc* [75]. Dosage compensation in males is also mediated by two RNA molecules: roX1 and roX2 [76]. Even the heat–shock stress response is regulated by non–protein–coding RNAs form the *Hsrω* gene [77].

After the sequencing of the *Drosophila* genome, the first systematic screenings of *Drosophila* lncRNAs detected 52 putative loci [80, 81]. Whole-genome tiling arrays have facilitated the detection of potential lncRNAs [82], although their validation requires further experimental confirmation. More recently, it has been estimated that around 5000 loci may encode non-protein-coding transcripts in *Drosophila* [83]. However, the number of functional regulatory lncRNAs is still to be determined.

In some cases, the detection of orthologous lncRNAs among drosophilids requires analysis of their secondary structures, in addition to their primary sequences. For instance, *roX1* and *roX2* sequences diverged so fast that their detection in the 12 *Drosophila* genomes was based on the conservation of structural features of their RNA products [28]. According to FlyBase [14], *bxd*, *Hsrω* and *pgc* are also conserved across drosophilids. However, both sequence and structural conservation is often

restricted to a small part of the RNA molecule [74], making the detection of homologous lncRNAs difficult even between closely related species. Consequently, a comprehensive evolutionary analysis of *Drosophila* lncRNAs is still missing. It is expected that the sequencing of complete genomes from different populations of *Drosophila* will help us to understand the evolutionary origin of these enigmatic sequences.

FUTURE PROSPECTS

Comparative genomics has been particularly useful for the detection of non-protein-coding RNAs [84]. However, prediction of small regulatory RNAs is based on the structure of their precursors due to the size and unstructured nature of the mature sequences. Recently, identification of novel mature small RNAs has proceeded almost exclusively by transcriptional profiling of small RNAs. The combination of deep sequencing and comparative genomics in *Drosophila* has permitted the identification and evolutionary analyses of miRNAs, but the study of piRNAs and endo-siRNAs has additional issues. First, both piRNAs and endo-siRNAs are likely to vary with the transposable element content of the host genome, and the comparison between even close species is difficult. Also, siRNAs and piRNAs are often located in heterochromatic regions [57], which has been extensively studied in *D. melanogaster*, but not as much in other fly species. The sequencing and assembling of heterochromatic DNA from the other 11 *Drosophila* genomes will create an opportunity to study the conservation of these RNAs within the *Drosophila* genus.

The identification of lncRNAs is also a challenge, particularly as we do not know of any universal features of all lncRNAs. The comparative analyses of lncRNAs could be improved by using indirect strategies to identify homologues. For instance, the study of syntenic blocks has been very helpful to annotate orthologous transfer-RNAs in *Drosophila* [85]. Similar approaches may be successfully applied to lncRNAs (and other non-protein-coding sequences).

Population genetics is particularly useful to study the evolutionary dynamics of fast evolving genes (e.g. [86]). The study of regulatory RNAs in populations has been mostly restricted to miRNAs [87, 88], although piRNAs have recently captured the attention of population geneticists [65]. With the development of deep sequencing, the characterization of entire genomes from hundreds of different populations is becoming a reality (http://dpgp.org/). Population genomics of non-coding RNAs shows particular promise for the near future.

Do any large classes of regulatory RNA remain unidentified? Are there genomic signatures that would allow us to detect non-coding RNAs without having transcriptional information? Do small RNAs have other, yet unknown, biological functions? These are some of the most important questions in the RNA biology field. The *D. melanogaster* genome and its close relatives will have a lot to say.

Key Points

- The sequencing and analysis of *Drosophila* genomes have had a big impact on the study of regulatory RNAs.
- The comparison of 12 *Drosophila* genomes revealed important aspects of the evolutionary dynamics of miRNA sequences.
- Comparative genomics analysis of regulatory RNAs other than miRNAs has been, so far, less successful.
- Population genomics and heterochromatin sequencing in other *Drosophila* species are promising areas to investigate the nature of regulatory RNAs.

References

1. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 1961;**3**:318–56.
2. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. *Science* 1969;**165**:349–57.
3. Davidson EH. *Genomic Regulatory Systems: Development and Evolution*. San Diego: Academic Press, 2001.
4. Carroll S, Grenier J, Weatherbee S. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden: Blackwell Publishing Ltd, 2005.
5. Tomari Y, Zamore PD. Perspective: machines for RNAi. *Genes Dev* 2005;**19**:517–29.
6. Eguchi Y, Itoh T, Tomizawa J. Antisense RNA. *Annu Rev Biochem* 1991;**60**:631–52.

7. Lee RC, Feinbaum RL, Ambros V. The C. elegans hetero-chronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993;**75**:843–54.

8. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 1993;**75**: 855–62.

9. Lawrence PA. *The Making of a Fly: The Genetics of Animal Design*. Oxford: Blackwell Science, 1992.

10. Lewis EB. A gene complex controlling segmentation in Drosophila. *Nature* 1978;**276**:565–70.

11. Martin CH, Mayeda CA, Davis CA, et al. Complete sequence of the bithorax complex of Drosophila. *Proc Natl Acad Sci USA* 1995;**92**:8398–402.

12. Lipshitz HD, Peattie DA, Hogness DS. Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev* 1987;**1**:307–22.

13. Ashburner M, Bergman CM. Drosophila melanogaster: a case study of a model genomic sequence and its consequences. *Genome Res* 2005;**15**:1661–7.

14. McQuilton P, St Pierre SE, Thurmond J, et al. FlyBase 101 – the basics of navigating FlyBase. *Nucleic Acids Res* 2012;**40**: D706–14.

15. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;**116**:281–97.

16. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 2010;**11**:597–610.

17. Saito K, Ishizuka A, Siomi H, et al. Processing of pre-microRNAs by the Dicer-1–Loquacious complex in Drosophila cells. *PLoS Biol* 2005;**3**:e235.

18. Förstemann K, Horwich MD, Wee L, et al. Drosophila microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell* 2007;**130**:287–97.

19. Reinhart BJ, Slack FJ, Basson M, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 2000;**403**:901–6.

20. Adams MD, Celniker SE, Holt RA, et al. The genome sequence of Drosophila melanogaster. *Science* 2000;**287**: 2185–95.

21. Pasquinelli AE, Reinhart BJ, Slack F, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 2000;**408**:86–9.

22. Lagos-Quintana M, Rauhut R, Lendeckel W, et al. Identification of novel genes coding for small expressed RNAs. *Science* 2001;**294**:853–8.

23. Lau NC, Lim LP, Weinstein EG, et al. An Abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* 2001;**294**:858–62.

24. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. *Science* 2001;**294**:862–64.

25. Lai EC, Tomancak P, Williams RW, et al. Computational identification of Drosophila microRNA genes. *Genome Biol* 2003;**4**:R42.

26. Enright A, John B, Gaul U, et al. MicroRNA targets in Drosophila. *Genome Biol* 2003;**5**:R1.

27. Stark A, Brennecke J, Russell RB, et al. Identification of Drosophila microRNA targets. *PLoS Biol* 2003;**1**:E60.

28. Clark AG, Eisen MB, Smith DR, et al. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 2007; **450**:203–18.

29. Ruby JG, Stark A, Johnston WK, et al. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res* 2007;**17**:1850–64.

30. Stark A, Kheradpour P, Parts L, et al. Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. *Genome Res* 2007;**17**:1865–79.

31. Stark A, Lin MF, Kheradpour P, et al. Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* 2007;**450**:219–32.

32. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. *Nature* 2007; **448**:83–6.

33. Okamura K, Hagen JW, Duan H, et al. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* 2007;**130**:89–100.

34. Kheradpour P, Stark A, Roy S, et al. Reliable prediction of regulator targets using 12 Drosophila genomes. *Genome Res* 2007;**17**:1919–31.

35. Roy S, Ernst J, Kharchenko PV, et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* 2010;**330**:1787–97.

36. Berezikov E, Robine N, Samsonova A, et al. Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. *Genome Res* 2011;**21**:203–15.

37. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;**39**:D152–7.

38. Lu J, Shen Y, Wu Q, et al. The birth and death of microRNA genes in Drosophila. *Nat Genet* 2008;**40**:351–5.

39. Berezikov E, Liu N, Flynt AS, et al. Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. *Nat Genet* 2010;**42**:6–9.

40. Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in Drosophila species. *Genome Biol Evol* 2010;**2**:180–9.

41. Sempere LF, Cole CN, McPeek MA, et al. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zool B Mol Dev Evol* 2006;**306**:575–88.

42. Wheeler B, Heimberg A, Moy V, et al. The deep evolution of metazoan microRNAs. *Evol Dev* 2009;**11**:50–68.

43. Marco A, Hui JHL, Ronshaugen M, et al. Functional shifts in insect microRNA evolution. *Genome Biol Evol* 2010;**2**: 686–96.

44. de Wit E, Linsen SEV, Cuppen E, et al. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res* 2009;**19**:2064–74.

45. Griffiths-Jones S, Hui JHL, Marco A, et al. MicroRNA evolution by arm switching. *EMBO Rep* 2011;**12**:172–7.

46. Marco A, Hooks K, Griffiths-Jones S. Evolution and function of the extended miR-2 microRNA family. *RNA Biol* 2012;**9**:242–8.

47. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet* 2009;**10**:94–108.

48. Ruby JG, Jan C, Player C, *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* 2006;**127**:1193–207.

49. Czech B, Malone CD, Zhou R, *et al.* An endogenous small interfering RNA pathway in Drosophila. *Nature* 2008;**453**: 798–802.

50. Chung W-J, Okamura K, Martin R, *et al.* Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr Biol* 2008;**18**:795–802.

51. Okamura K, Chung W-J, Ruby JG, *et al.* The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 2008;**453**:803–6.

52. Smith CD, Shu S, Mungall CJ, *et al.* The release 5.1 annotation of Drosophila melanogaster Heterochromatin. *Science* 2007;**316**:1586–91.

53. Hoskins RA, Carlson JW, Kennedy C, *et al.* Sequence finishing and mapping of Drosophila melanogaster heterochromatin. *Science* 2007;**316**:1625–8.

54. Okamura K, Balla S, Martin R, *et al.* Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. *Nat Struct Mol Biol* 2008;**15**:581–90.

55. Aravin AA, Lagos-Quintana M, Yalcin A, *et al.* The small RNA profile during Drosophila melanogaster development. *Dev Cell* 2003;**5**:337–50.

56. Yin H, Lin H. An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature* 2007;**450**:304–8.

57. Senti K-A, Brennecke J. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* 2010; **26**:499–509.

58. Jensen S, Gassama M-P, Heidmann T. Taming of transposable elements by homology-dependent gene silencing. *Nat Genet* 1999;**21**:209–12.

59. Jensen S, Gassama M-P, Dramard X, *et al.* Regulation of I-transposon activity in Drosophila: evidence for cosuppression of nonhomologous transgenes and possible role of ancestral I-related pericentromeric elements. *Genetics* 2002; **162**:1197–209.

60. Desset S, Meignin C, Dastugue B, *et al.* COM, a heterochromatic locus governing the control of independent endogenous retroviruses from Drosophila melanogaster. *Genetics* 2003;**164**:501–9.

61. Bergman CM, Quesneville H, Anxolabéhère D, *et al.* Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. *Genome Biol* 2006;**7**:R112.

62. Gunawardane LS, Saito K, Nishida KM, *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* 2007;**315**: 1587–90.

63. Brennecke J, Aravin AA, Stark A, *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* 2007;**128**:1089–1103.

64. Li C, Vagin VV, Lee S, *et al.* Collapse of germline piRNAs in the absence of argonaute3 reveals somatic piRNAs in flies. *Cell* 2009;**137**:509–21.

65. Lu J, Clark AG. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. *Genome Res* 2010;**20**:212–27.

66. Prud'homme N, Gans M, Masson M, *et al.* Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. *Genetics* 1995;**139**:697–711.

67. Engels WR. P elements in Drosophila melanogaster. In: Berg DE, Howe MM (eds). *Mobile DNA*. Washington: American Society for Microbiology, 1989.

68. Brennecke J, Malone CD, Aravin AA, *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 2008;**322**:1387–92.

69. Malone CD, Brennecke J, Dus M, *et al.* Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* 2009;**137**:522–35.

70. Lau N, Robine N, Martin R, *et al.* Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res* 2009;**19**:1776–85.

71. Klattenhoff C, Theurkauf W. Biogenesis and germline functions of piRNAs. *Development* 2008;**135**:3–9.

72. Shpiz S, Kalmykova A. Role of piRNAs in the Drosophila telomere homeostasis. *Mob Genet Elements* 2011;**1**:274–8.

73. Petit N, Piñero D, López-Panadès E, *et al.* HeT-A_pi1, a piRNA target sequence in the Drosophila telomeric retrotransposon HeT-A, is extremely conserved across copies and species. *PLoS One* 2012;**7**:e37405.

74. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;**10**: 155–9.

75. Nakamura A, Amikura R, Mukai M, *et al.* Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. *Science* 1996;**274**:2075–9.

76. Franke A, Baker BS. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. *Mol Cell* 1999;**4**: 117–22.

77. Lakhotia SC, Sharma A. The 93D (hsr-omega) locus of Drosophila: non-coding gene with house-keeping functions. *Genetica* 1996;**97**:339–48.

78. Wang W, Brunet FG, Nevo E, *et al.* Origin of sphinx, a young chimeric RNA gene in Drosophila melanogaster. *Proc Natl Acad Sci USA* 2002;**99**:4448–53.

79. Soshnev AA, Ishimoto H, McAllister BF, *et al.* A conserved long noncoding RNA affects sleep behavior in Drosophila. *Genetics* 2011;**189**:455–468.

80. Inagaki S, Numata K, Kondo T, *et al.* Identification and expression analysis of putative mRNA-like non-coding RNA in Drosophila. *Genes Cells* 2005;**10**:1163–73.

81. Tupy JL, Bailey AM, Dailey G, *et al.* Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. *Proc Natl Acad Sci USA* 2005;**102**: 5495–500.

82. Manak JR, Dike S, Sementchenko V, *et al.* Biological function of unannotated transcription during the early development of Drosophila melanogaster. *Nat Genet* 2006;**38**: 1151–8.

83. Li Z, Liu M, Zhang L, *et al.* Detection of intergenic non-coding RNAs expressed in the main developmental stages in Drosophila melanogaster. *Nucleic Acids Res* 2009; **37**:4308–14.

84. Eddy SR. Computational genomics of noncoding RNA genes. *Cell* 2002;**109**:137–40.

85. Rogers HH, Bergman CM, Griffiths-Jones S. The evolution of tRNA genes in Drosophila. *Genome Biol Evol* 2010;**2**: 467–77.

86. Levine MT, Jones CD, Kern AD, *et al*. Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci USA* 2006;**103**:9935–9.

87. Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 2006;**38**:1452–6.

88. Saunders MA, Liang H, Li W-H. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci USA* 2007;**104**:3300–5.