Evolution of Hox Post-Transcriptional Regulation by Alternative Polyadenylation and MicroRNA Modulation Within 12 Drosophila Genomes

Pedro Patraquim,1 Maria Warnefors,1 and Claudio R. Alonso*1

1John Maynard Smith Building, School of Life Sciences, University of Sussex, Brighton, United Kingdom

*Corresponding author: E-mail: c.alonso@sussex.ac.uk.

Associate editor: Claudia Kappen

Abstract

Hox genes encode a family of transcriptional regulators that operate differential developmental programs along the anteroposterior axis of bilateral animals. Regulatory changes affecting Hox gene expression are believed to have been crucial for the evolution of animal body plans. In Drosophila melanogaster, Hox expression is post-transcriptionally regulated by microRNAs (miRNAs) acting on target sites located in the 3′ untranslated regions (3′ UTRs) of Hox mRNAs. Notably, recent work has shown that during D. melanogaster development Hox genes produce mRNAs with variable 3′ UTRs (short and long forms) in different sets of tissues as a result of alternative polyadenylation; importantly, Hox short and long 3′ UTRs contain very different target sites for miRNAs. Here, we use a computational approach to explore the evolution of Hox 3′ UTRs treated with especial regard to miRNA regulation. Our work is focused on the 12 Drosophila species for which genomic sequences are available and shows, first, that alternative polyadenylation of Hox transcripts is a feature shared by all drosophilids tested in the study. Second, that the regulatory impact of miRNAs is evolving very fast within the Drosophila group. Third, that in contrast to the low degree of primary sequence conservation, Hox 3′ UTR regions within the group show very similar RNA topology indicating that RNA structure is under strong selective pressure. Finally, we also demonstrate that Hox alternative polyadenylation can remodel the control regions seen by miRNAs by at least two mechanisms: via adding new cis-regulatory sequences—in the form of miRNA target sites—to short 3′ UTR forms as well as by modifying the regulatory impact of miRNA target sites in short 3′ UTR forms through changes in RNA secondary structure caused by the use of distal polyadenylation signals.

Key words: Hox genes, post-transcriptional regulation, alternative polyadenylation, microRNA evolution, Drosophila evolution, RNA secondary structure.

Introduction

The regulation of mRNA levels in time and space lies at the heart of the genetic programs controlling development. Such control of RNA expression levels relies on both transcriptional and post-transcriptional mechanisms (Britten and Davidson 1969; Alonso and Wilkins 2005; Alonso 2008). Current molecular models for gene expression indicate that information contained in mRNA 3′ untranslated regions (3′ UTRs) is interpreted by the cell so as to determine patterns of mRNA decay, localization, and rates of protein translation (Moore 2005). Although the exact molecular mechanisms that link particular 3′ UTR sequences to different mRNA behaviors are not fully understood, physical contacts between cis-regulatory elements located in transcript 3′ UTRs (Bartel 2004; Bartel and Chen 2004) and trans-acting regulators including RNA-binding proteins and small RNAs, such as microRNAs (miRNAs), are essential for RNA regulation to occur.

Hox genes encode a family of transcriptional regulators that pattern animal bodies along the anteroposterior axis (Lewis 1978; McGinnis and Krumlauf 1992; Alonso 2002). Evolutionary changes affecting Hox expression patterns and functions are thought to have mediated the evolution of animal body plans (Holland and Garcia-Fernández 1996; Pearson et al. 2005), but the molecular nature of such regulatory changes affecting Hox expression has yet not been completely resolved (Alonso and Wilkins 2005; Alonso 2008).

In Drosophila melanogaster, Hox genes are regulated by miRNAs via miRNA target sequences located in Hox 3′ UTRs (Ronshaugen et al. 2005; Bender 2008; Stark et al. 2008; Tyler et al. 2008). Intriguingly, recent work described that during D. melanogaster development Hox genes produce mRNAs with variable 3′ UTRs—that is, short and long forms—in different tissues as a result of alternative polyadenylation; notably, short and long 3′ UTRs contain very different target sites for miRNAs converting each mRNA species into substantially different miRNA targets (Thomsen et al. 2010).

Here, we use publicly available genome sequences from 12 Drosophila species to investigate the evolution of 3′ UTR regions in the Drosophila Hox gene Ultrabithorax (Ubx), searching for variation affecting primary sequence and RNA secondary structure. We focus our analysis on the distribution of discrete cis-regulatory modules including polyadenylation signals (PASs) and miRNA target sites, and RNA structural features affecting local and global topology of 3′ UTRs predicted to impact the recognition of Hox mRNAs by RNA regulators.
Materials and Methods

*Drosophila Ubx* 3′ UTR Sequences

3′ UTR sequences homologous to *D. melanogaster Ubx* 3′ UTR regions were retrieved from the University of California–Santa Cruz (UCSC) Genome Browser. Realignments were carried out using the LAGAN algorithm embedded in the VISTA tools (Brudno et al. 2003).

*Ubx* miRNA Target Site Predictions and Evolutionary Analyses

The Probability of Interaction by Target Accessibility (PITA) algorithm was used for miRNA target site predictions in all 12 *Drosophila Ubx* 3′ UTRs (http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html) (Kertesz et al. 2007). This software uses miRNA seed–mRNA target site complementarity, as well as target site accessibility, to compute an energy-based prediction of miRNA–mRNA interactions (ΔΔG). PITA parameters used included miRNA seeds of 6–8 nt in size, single G:U wobbles allowed as well as single mismatches in the case of 8 nt seed sequences, and accessibility of flanking regions (Kertesz et al. 2007). To define the minimal set of miRNA species suitable for an evolutionary comparison within the drosophilids, we applied three different criteria: evolutionary conservation of miRNA genes, developmental coexpression with *Ubx* in *D. melanogaster* (as inferred from Northern blot data in Ruby et al. 2007), and strong miRNA–mRNA predicted interactions (using as baseline the DDG values for miR-rib-4/miR-rib-8, given that interactions between *Ubx* and these miRNA species have been experimentally validated Ronshaugen et al. 2005; Bender 2008; Stark et al. 2008; Tyler et al. 2008; Thomsen et al. 2010). Regarding the evolutionary conservation of miRNA genes, information was retrieved from http://www.MiRBASE.org, Ruby et al. 2007 and Blast searches (http://flybase.org/blast/) for both the *D. melanogaster* pri-miRNA and seed sequences. Low Blast E values were folded using the RNA-Fold algorithm WebServer to validate the typical stem-hairpin miRNA structure and its minimum free energy value of —20.00 kcal/mol (Hofacker 2003; Ruby et al. 2007). For the miRNAs selected for evolutionary analysis, *Ubx* 3′ UTR miRNA targeting predictions were grouped as follows: *D. melanogaster, D. simulans, D. sechelia, D. yakuba,* and *D. erecta* in the first group; *D. ananassae, D. persimilis,* and *D. pseudoobscura* in the second group; and *D. wilistoni, D. mojavensis, D. virilis,* and *D. grimshawi* in the third group. One-way analyses of variance (ANOVA) were subsequently performed to examine evolutionary differences in miRNA targeting among these groups.

RNA Accessibility Predictions for *Drosophila Ubx* 3′ UTRs

*Drosophila Ubx* 3′ UTRs were submitted to PITA to obtain whole-sequence accessibility values (ΔΔG values). Because all these independently evolving sequences showed multiple “indels” and substitutions, a correspondence table was generated with homologous regions across these sequences at the primary sequence level, using the VISTA-LAGAN alignment produced previously. For this, 200 bp-windows of homology were ascribed across drosophilid *Hox* 3′ UTRs as compared with the baseline *D. melanogaster* sequence. This enlarged window allowed for the comparison of *Ubx* 3′ UTR accessibilities by ascribing overall homologous regions despite significant nucleotide divergence. The average accessibility values were calculated, for all of the 200 bp windows, across the whole 3′ UTR of each species, and the 12 results were plotted simultaneously. A scrambled set of sequences produced by reshuffling of the original *Ubx* 3′ UTR sequence alignments was used as a negative control. The results were further confirmed by performing a t-test for differences in accessibility variance between the *Ubx* 3′ UTRs multiple alignment and the negative control.

Secondary Structure Predictions for *Ubx* mRNA Transcripts

Full *D. melanogaster* mRNA sequences for *Ubx* were retrieved from http://www.Flybase.org, in the form of two 3′ UTR isoforms: one including only the proximal 3′ UTR region up to PAS1 (i.e., *Ubx* short 3′ UTR) and another including all proximal and distal 3′ UTR regions up to PAS2 (i.e., *Ubx* long 3′ UTR). The RNA secondary structure prediction tool RNAFold (on default settings) was then used to generate a visual interpretation of the most stable RNA structure for these two mRNA isoforms. Predicted RNA secondary landmarks were then superimposed onto *Ubx* mRNA sequences (coding and noncoding regions) according to information on *Ubx* transcripts in Flybase (http://flybase.org/reports/FBgn0003944.html).

Results and Discussion

Evolution of Alternative Polyadenylation of *Hox* Genes: Conservation and Plasticity

To identify the evolutionary constraints that determine the final size of *Hox* 3′ UTRs, we examined the level of conservation of *Ubx* PASs within the 12 drosophilids for which genomic sequences are available. We focus on the *Ubx* gene as this is the system for which alternative polyadenylation and miRNA regulation are known in highest detail (Ronshaugen et al. 2005; Bender 2008; Stark et al. 2008; Tyler et al. 2008; Thomsen et al. 2010).

Sequences were retrieved from the UCSC genome browser and aligned using “mVISTA-LAGAN” software (Brudno et al. 2003). Our analysis shows that PASs leading to the production of (at least) two alternative transcripts of distinct length are conserved throughout the group; however, the exact position of the PASs varies within the group indicating a certain level of plasticity in the mechanism of *Ubx* alternative polyadenylation (fig. 1A). We also find that 3′ UTR total length was approximately maintained within the group: in those species where the proximal 3′ UTR is shorter, the distal 3′ UTR is extended (fig. 1B). These observations indicate that the basic alternative polyadenylation patterns found in *D. melanogaster* are conserved throughout the group suggesting that *Ubx* alternative polyadenylation is likely a feature present in the common ancestor of the group.
miRNA Regulation Shows Dynamic Evolutionary Patterns Within the Drosophila

Two main factors determine the likelihood of a given miRNA to regulate a target mRNA via 3′ UTR sequences: primary sequence composition at target sites and local RNA topology (Kertesz et al. 2007; Long et al. 2007; Li et al. 2010). Therefore, the combined computation of sequence elements and RNA accessibility in target 3′ UTRs allows for an accurate calculation of the regulatory impact of miRNAs on mRNA targets (Kertesz et al. 2007; Long et al. 2007). To investigate the evolution of Hox miRNA regulation, we focused once again on Ubx, submitting Ubx 3′ UTR sequences from all 12 drosophilids to the prediction tool PITA (Kertesz et al. 2007), which computes target sequence and RNA topology simultaneously. PITA represents predicted regulatory strength in the form of an energy-based score termed ΔΔG, that is ascribed to a given miRNA–mRNA interaction; this value is calculated by subtracting the free energy lost by opening locally paired RNA structures (within the target mRNA) to the free energy gained by the formation of a specific miRNA–mRNA duplex. To maximize the stringency of our analysis, we focused on those miRNAs which are temporally coexpressed with Ubx during D. melanogaster development and for which miRNA seeds are known to be ultraconserved within the 12 Drosophila genomes (Ruby et al. 2007). This approach identified 14 miRNAs that were used in further analyses (figs. 2 and 3). Regulatory analysis of each one of the 14 miRNAs within the 12 Drosophila species shows that the evolution of miRNA regulation varies according to miRNA: eight miRNAs show a dynamic evolutionary pattern with significant change in targeting strength across Drosophila species (ascribed using one-way ANOVAs—fig. 2, top panel), whereas the remaining six miRNAs show “stasis” with no statistically significant shifts in regulatory predictions across Drosophila (fig. 2, bottom panel). Interestingly, individual miRNA species derived from the single genomic locus iab4/iab8 (namely: miRNA-iab-4-3p, miR-iab-4-5p, miR-iab-8-3p, and miR-iab-8-5p) display distinct evolutionary trends. miR-iab-4-3p and miR-iab-8-5p display dynamic evolutionary patterns of Ubx regulation with marked tendencies, albeit with different polarity (fig. 2, top panel); miR-iab-4-5p shows no significant change across Drosophila (fig. 2, bottom panel); finally, miR-iab-8-3p is not predicted to target Ubx at all.

We also explored the relationship between the evolution of individual target sites and the net miRNA regulatory effects on Ubx. For this, we took PITA outputs for all
predicted targets for the 14 miRNAs (see supplementary table S1, Supplementary Material online) and divided the resulting 358 miRNA target sites into three broad categories according to regulatory strength: "strong" ($\Delta \Delta G < -8$), "intermediate" ($-8 < \Delta \Delta G < -4$), and "weak" sites ($\Delta \Delta G > -2$). This analysis first revealed that the targeting of a given mRNA by miRNAs seems to depend on a predominant strong site, responsible for most of the net regulatory value, followed by "shadow" ("intermediate" and weak) miRNA target sites which when present, make a small contribution to net mRNA target regulation (fig. 3).

Secondly, 77.1% of the miRNA target sites are found within the same 3′ UTR isoform (long or short) in which the strong target site lies in D. melanogaster (fig. 3) ($P < 0.0002$). Third, strong sites tend to show a significantly slower evolutionary turnover, as 92% of the strong sites identified (35 of 38) are conserved within Ubx 3′ UTRs, in contrast to the 59% conservation level observed for shadow sites ($P < 0.0001$) (See supplementary fig. S4, Supplementary Material online). Fourth, we superimposed individual target site evolution to the different net miRNA–mRNA evolutionary patterns described above (dynamic vs. stasis) and found that stasis is linked to a higher level of conservation of shadow (intermediate and weak) sites ($P = 0.0334$). This analysis shows that strong miRNA target sites dominate mRNA regulation and that most auxiliary sites tend to remain in the same 3′ UTR location suggesting that miRNA regulation is isoform specific for mRNAs with alternative 3′ UTRs.

RNA Accessibility in Ubx 3′ UTRs is Ultraconserved Despite Significant Change at the Primary Sequence Level

The conservation of primary 3′ UTR sequences within the Drosophila group is limited to a series of scattered small ultraconserved elements (UCEs) (see conservation profiles in fig. 1A and supplementary fig. S3, Supplementary Material online). However, given that miRNA regulation of target genes relies on both primary sequence as well as on RNA secondary

Fig. 2. Quantitative evolution of miRNA regulation of Ubx. The figure shows the regulatory evolution of the 14 miRNAs analyzed illustrating a dynamic evolutionary pattern (top panel box plots), and a stasis pattern of no significant change in predicted regulatory effects (bottom panel box plots). For these statistical analyses, using one-way ANOVA analyses, species were grouped in clades displaying high levels of sequence conservation (see fig. 1A and supplementary fig. S2, Supplementary Material online). Drosophila melanogaster was grouped with D. simulans, D. sechelia, D. erecta, and D. yakuba. The second group comprised D. ananassae, D. pseudoobscura, and D. persimilis, whereas the third group included D. willistoni, D. mojavensis, D. virilis, and D. grimshawii. The different significances in evolutionary change are represented by * ($P$ value < 0.05), ** ($P$ value < 0.005), and *** ($P$ value < 0.0005). Notice that miRNAs produced from the iab-4/iab-8 locus—which have been experimentally shown to target Ubx mRNAs (Ronshaugen et al. 2005; Bender 2008; Stark et al. 2008; Tyler et al. 2008; Thomsen et al. 2010)—show significant evolutionary change regarding the targeting of Ubx 3′ UTR:long mRNAs.
Evolution of Hox Post-Transcriptional Regulation by Alternative Polyadenylation

Fig. 3. Evolution of individual miRNA target sites within drosophilid Ubx 3' UTR sequences. (A) The diagram shows the evolution of miRNA target sites within Ubx 3' UTR sequences for 14 miRNAs. miRNA target sites for each miRNA species are depicted according to their regulatory strength (full square, strong sites; full circle, intermediate sites; and empty circles, weak sites). Notice the recent acquisition of strong conserved sites for miR-iab-4-3p within the melanogaster subgroup (including D. melanogaster, D. simulans, D. sechellia, D. yakuba, and D. erecta), pointing to a likely recent regulatory novelty. Notice also that most novel miRNA targeting tends to occur within one of the alternative 3' UTR isoforms (see text for further details).
structure, we decided to test to what extent 3’ UTR secondary structure had evolved within the Drosophila group. For this, we first divided the D. melanogaster Ubx long 3’ UTRs into 200 bp homology windows and then used “mVISTA-LAGAN” primary sequence alignments (fig. 1A) to define homologous regions for each D. melanogaster window in all species. We then calculated average accessibility values for each window in all 12 drosophilids analyzed using one of the intermediate outputs of PITA (ΔG_{open}) (fig. 4). Strikingly, this analysis revealed that patterns of RNA accessibility within Ubx 3’ UTRs are significantly conserved, specially around the first PAS (fig. 4A) suggesting that the maintenance of a particular RNA topology is likely to be under strong selective pressure and that RNA accessibility might play a role in the polyadenylation mechanism itself. As negative control, we used a scrambled set of sequences produced by reshuffling of the original Ubx 3’ UTR sequence alignments and found out that for this region there is very little conservation of accessibility patterns among the 12 drosophilids (fig. 4B). We further validated these differences by looking at the profiles of variance in accessibility values between these two cases (fig. 4C). A t-test for differences between the accessibility of the Ubx 3’ UTR alignments and the negative control yielded a P value of 0.0007. These results suggest that target RNA secondary structure may play a very significant role for the evolution of 3’ UTR regions.

We also looked at the putative role of alternative polyadenylation as a trigger of an RNA secondary structure switch, using for this “RNAFold” software (Vienna package) (Hofacker 2003) to fold Ubx mRNA sequences including only the proximal 3’ UTR segment (use of PAS1) (fig. 5A) or the full 3’ UTR (fig. 5B). We observed that sequences located at the end of the proximal 3’ UTR are predicted to change in secondary structure when the distal segment is also present in the molecule (fig. 5C). To establish the effects of the change in RNA folding experienced by this region (Alternative Folding Region or AFR), we compared accessibility values (the PITA intermediate output ΔG_{open}) for both long and short Ubx 3’ UTRs, which showed a region with 3’ UTR isoform

FIG. 4. RNA accessibility is conserved across Hox 3’ UTRs, unlike primary sequence. (A) RNA accessibility alignments for Ubx 3’ UTRs. Homologous regions were ascribed to 200 bp homology windows using Drosophila melanogaster as a baseline. A measure of RNA accessibility (ΔG_{open}) is plotted versus Ubx 3’ UTR length. Low ΔG_{open} values indicate low accessibility. Despite significant divergence at the level of primary sequence, which is limited to a series of scattered ultraconserved elements (see supplementary fig. S3, Supplementary Material online), the accessibility of homologous regions of the Ubx 3’ UTR remains generally constant. (B) ΔG_{open} values for a control sequence generated by complete reshuffling of Ubx 3’ UTR alignments; note the high level of variation in ΔG_{open} values observed in this case. (C) Variance analysis of Ubx and the control reshuffled Ubx 3’ UTR; although Ubx variance in accessibility values remains fairly unchanged across the Ubx 3’ UTR, variance for the control segment shows distinct peaks revealing lack of conservation in accessibility predictions (t-test; P = 0.0007).
context-dependent accessibility starting around 80 bp upstream of the first PAS.

We also compared Ubx:short vs. Ubx:long PITA miRNA target site predictions for the AFR region and found that the regulatory strength of miRNA sites located within the proximal 3'UTR segment for miR-92a and miR-92b is predicted to decrease when the long 3'UTR segment is also included in the 3'UTR (fig. 5D). These results point to a previously unknown post-transcriptional mechanism by which the regulated addition of a nucleotide stretch to the 3'UTR of a transcript changes the RNA secondary structure of the proximal constitutive 3'UTR. This implies that alternative polyadenylation can remodel the regulatory landscape of an mRNA molecule by at least two mechanisms: by adding new miRNA target sites.
downstream of a 3′ UTR form and by a nonlinear mechanism that modifies the regulatory value of multiple miRNA target sites simultaneously through changes in RNA secondary structure. In sum, our work shows a complex interplay between RNA secondary structure, alternative polyadenylation, and miRNA regulation during the evolution of 3′ UTRs of a key developmental gene.

Supplementary Material
Supplementary table S1 and figure S1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
We thank members of the Alonso Lab for discussions that helped shaping this study. We also wish to thank the constructive criticisms of two anonymous reviewers whose comments improved the quality of this article. This work was supported by The UK Biotechnology and Biological Sciences Research Council (BBSRC) grants BB/E01173X/1 and BB/E01173X/2 to C.R.A.

References