scientific reports

OPEN

The embryonic transcriptome of *Parhyale hawaiensis* reveals different dynamics of microRNAs and mRNAs during the maternal-zygotic transition

Llilians Calvo, Maria Birgaoanu, Tom Pettini, Matthew Ronshaugen✉ & Sam Griffiths-Jones✉

Parhyale hawaiensis* has emerged as the crustacean model of choice due to its tractability, ease of imaging, sequenced genome, and development of CRISPR/Cas9 genome editing tools. However, transcriptomic datasets spanning embryonic development are lacking, and there is almost no annotation of non-protein-coding RNAs, including microRNAs. We have sequenced microRNAs, together with mRNAs and long non-coding RNAs, in *Parhyale* using paired size-selected RNA-seq libraries at seven time-points covering important transitions in embryonic development. Focussing on microRNAs, we annotate 175 loci in *Parhyale*, 88 of which have no known homologs. We use these data to annotate the microRNAome of 37 crustacean genomes, and suggest a core crustacean microRNA set of around 61 sequence families. We examine the dynamic expression of microRNAs and mRNAs during the maternal-zygotic transition. Our data suggest that zygotic genome activation occurs in two waves in *Parhyale* with microRNAs transcribed almost exclusively in the second wave. Contrary to findings in other arthropods, we do not predict a general role for microRNAs in clearing maternal transcripts. These data significantly expand the available transcriptomics resources for *Parhyale*, and facilitate its use as a model organism for the study of small RNAs in processes ranging from embryonic development to regeneration.

Parhyale hawaiensis has emerged as a key crustacean model for studies ranging from regeneration to comparative developmental biology. Available genomics tools include a sequenced genome[1], transcriptome annotation[2,3], and successful application of CRISPR-Cas9 approaches[4]. Detailed description of embryonic developmental landmarks[5] such as the segmentation cascade[6], Hox gene expression[7] and cell lineage tracing studies have also been established[8,9]. However, there remain a number of missing tools in this expanding repertoire. A key omission is publicly available transcriptome data across the developmental time-course. Studies using pooled embryos from diverse stages have provided some insight into the *Parhyale* gene and transcript annotation, but there is no genome-wide temporal resolution or information about dynamic expression of transcripts. Existing annotations are limited in sequencing depth and replication, and annotation of small RNAs (including microRNAs) is limited to highly conserved sequences[2,3].

MicroRNAs are short non-coding RNAs of ~22 nucleotides (nt) in length that regulate gene expression at a post-transcriptional level in metazoans and plants. In animals, microRNAs target the 3'UTRs of mRNAs by partial base-pairing complementarity with target mRNAs[10] inducing either translation inhibition or deadenylation and decay of these target mRNAs[11]. Since their discovery, microRNAs have been found to regulate many biological processes, and their importance in development has been demonstrated repeatedly. For example, at the maternal-zygotic transition (MZT), zygotic microRNAs have been found to be involved in clearance of maternally-deposited mRNAs in several invertebrate species including *Drosophila melanogaster*[12], *Tribolium castaneum*[13] and *Blattella germanica*[14]. Similar results have been found in vertebrates such as *Danio rerio*[15] and

Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M13 9PT, UK. ✉email: matthew.ronshaugen@manchester.ac.uk; sam.griffiths-jones@manchester.ac.uk

Xenopus laevis[16], although the microRNAs involved do not appear to be conserved between vertebrates and invertebrates. Interestingly, this early developmental function for microRNAs has not been found in *Caenorhabditis elegans*[17] nor in mice, where it is suggested that microRNAs do not play an essential role in the clearance of maternal mRNAs[18]. Indeed, a recent study has shown that *C. elegans* embryonic development progresses up to gastrulation in the absence of the microRNA processing machinery, and that only two microRNAs, miR-51 and miR-35[19], are essential for complete development. These studies suggest two principles: first, that microRNAs are key regulators of embryonic development, whether facilitating clearance of maternal transcripts or acting later during diverse developmental processes; and second, that variation in microRNA function between species can be significant, and their developmental roles must therefore be examined on a species-by-species basis.

In this study, we have annotated and quantified the expression of mRNAs, long non-coding RNAs and microRNAs across 7 stages of embryonic development in *P. hawaiensis*[5]. Focusing on microRNA expression during *Parhyale* embryogenesis, we have increased the number of annotated microRNAs in this organism from 51 highly conserved sequences[1,3] to a total of 175 microRNA precursors, 88 of which have not been previously described in any organism. We have used the microRNA repertoire of *Parhyale* to provide a comprehensive similarity-based annotation of crustacean microRNAs in 37 species. We find that the core crustacean microRNA complement numbers around 61 families. Finally, the expression dynamics of microRNAs and target mRNAs through development suggests that zygotic genome activation (ZGA) occurs in two waves, with microRNA expression largely restricted to the second wave.

Results

Parhyale hawaiensis size-separated RNA sequencing and small RNA annotation.

To develop a comprehensive developmental transcriptome of *Parhyale*, we selected embryos from seven different key time-points spanning the whole of embryogenesis. The time-points were chosen to capture key transcriptional changes during important developmental transitions (Fig. 1A). The first time-point covers the 1 to 8 cell stages (S1–4) – at this time the zygote is still transcriptionally inactive[20], and therefore has exclusively maternally-loaded RNAs. The second time-point comprises stage 5 and stage 6 (S5–6)[5], which includes the 32-cell stage, described in the literature as the maternal-to-zygotic transition[20]. During the third time-point, stages 7 to 11 (S7–11), several events take place: embryonic cells migrate and segregate from the yolk cells, the germ disc condenses and the germband rows appear[5]. The next two time-points were built using precisely-staged embryos at stage 14 (S14) and stage 17 (S17) during the period of germ band extension. The final two libraries span stages 21 to 23 (S21–23) and 24 to 30 (S24–30), which represent wide windows of time encompassing multiple developmental events such as limb bud formation and morphogenetic movement (Fig. 1A). To facilitate comparison of microRNA and mRNA expression profiles during embryogenesis, we built paired "small" (< 150nt) and "large" (> 150nt) libraries from the same samples for each time-point (Fig. 1B).

The small RNA reads obtained from the sequencing were cleaned (adaptors removed) and selected to retain 18–26 nt reads (Fig. 1C). Reads that mapped to the genome but failed to map to *Parhyale* tRNAs or crustacean rRNAs were considered potential microRNAs and used for miRDeep2 microRNA prediction (Fig. 1D). Manual inspection of miRDeep2 predictions yielded a total of 175 high confidence microRNA precursor loci, and 349 distinct mature sequences. 27 of the hairpin precursor sequences map perfectly to more than one location in the genome. While even the best assembled genomes contain evidence of duplicated microRNA sequences, these may represent heterozygosity or errors in the genome assembly. 87 of the precursor loci were conserved among other metazoans, and 88 were previously unreported. As expected, the majority of the reads mapping to the predicted microRNAs were 22 nt long (Fig. 1E) and 5' uracil biased (Fig. 1F).

Annotation of predicted *Parhyale* microRNAs in crustacean genomes.

MicroRNAs in crustaceans are poorly annotated. Only *Daphnia pulex*, *Marsupenaeus japonicus* and *Triops cancriformis* have any published microRNA sequences, and the level of coverage and completeness is variable. In order to address this underlying sampling problem, we used the 175 microRNA precursors identified from our sequencing data in *Parhyale* to predict microRNA homologs in the genomes of 37 crustacean species available in the NCBI Assembly database (Supplementary Table 1). In *Parhyale*, the 175 identified precursors belong to 105 different microRNA families; 51 families have been previously annotated, and are therefore conserved in other metazoans, whereas 54 families were novel. 124 out of 175 precursors, belonging to 79 different families, were present in the genome of at least one other crustacean species surveyed (see Fig. 2), with 18 families not conserved outside of the Malacostraca. We therefore suggest that the core crustacean microRNA set is comprised of around 61 families. A total of 49 out of 175 precursors, belonging to 26 of the novel families, had no significant match in any other crustacean genome, and are therefore lineage-specific '*Parhyale* unique' microRNAs (28% of all precursor sequences). The 37 crustacean species tested include four species in the same order as *Parhyale* (Amphipoda). Divergence times among these five amphipod species is not well determined, but all belong to the Talitroid clade, sharing a common ancestor ~ 60 million years ago, therefore indicating that these 49 '*Parhyale* unique' precursors have evolved more recently than ~ 60 million years ago[21].

We clustered the set of crustacean species based on the presence and absence of microRNA families in their genomes. Using only these characters, with no consideration of sequence similarity, it is interesting to note that the resulting tree clearly reproduces aspects of the established phylogeny of the crustacea (Fig. 2). For example, *Parhyale* has more microRNAs in common with other members of the class Malacostraca than it does with species belonging to the more distant Branchiopoda and Hexanauplia classes. Similarly, we observe strong correspondence in microRNA presence among species within the same order as *Parhyale*, the Amphipoda.
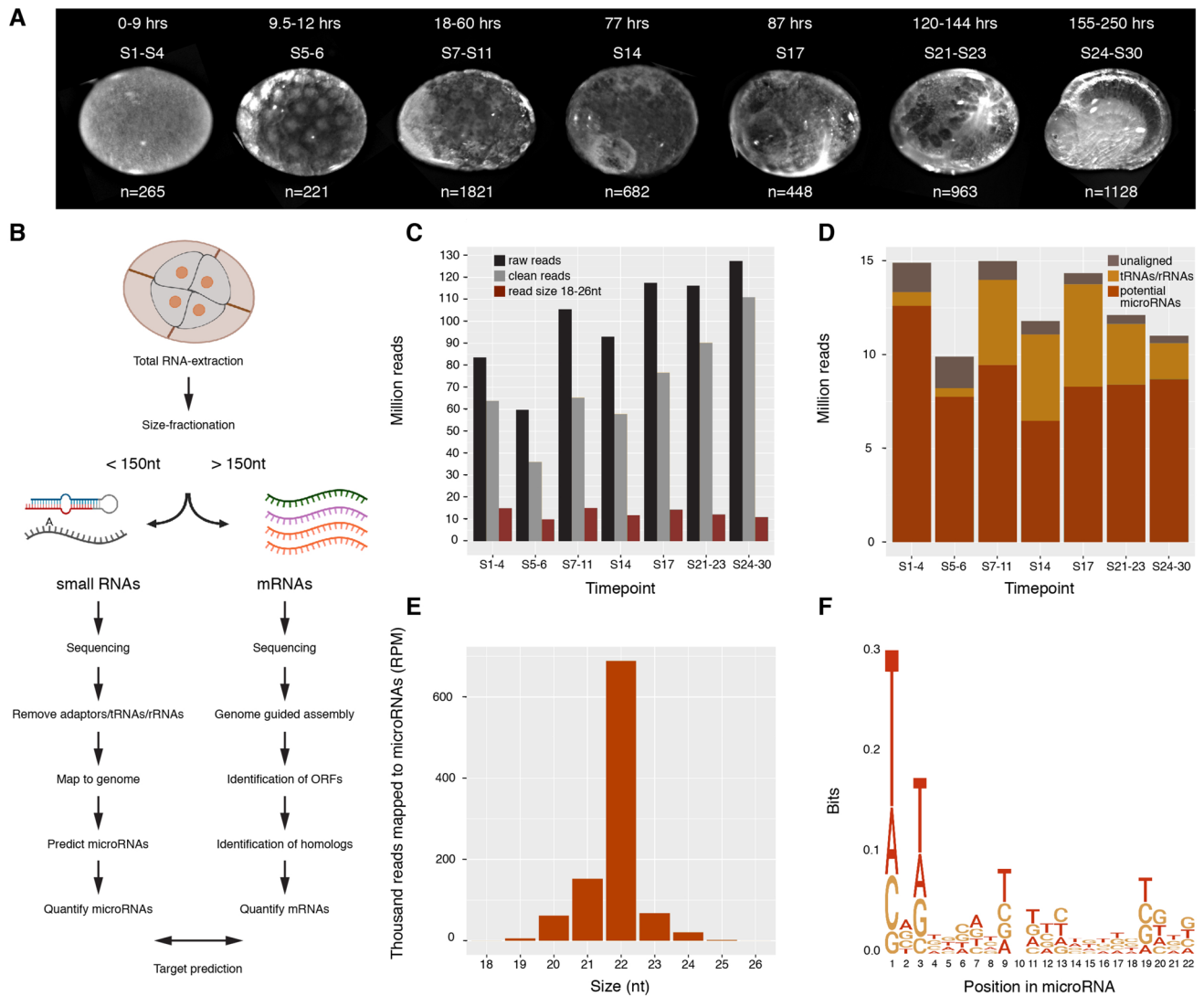
Figure 1. Library preparation and microRNA annotation in *Parhyale* development. (**A**) Brightfield images of embryo stages selected for building libraries. All pictures are lateral views. Developmental stages, the corresponding number of hours post-fertilization, and the number of embryos used for each time-point are indicated. (**B**) Diagram of workflow for size-separated library preparation and analysis. (**C**) Absolute abundance of sequence reads per time-point, total sequences reads (black), clean reads remaining after adaptor removal (grey), reads remaining after size selection (brown). (**D**) Distribution of size selected reads following mapping to the genome and to tRNAs/rRNAs database. (**E**) Size distribution of reads mapping to predicted microRNAs. (**F**) Sequence logo of the first 22 nt of non-redundant reads mapping to microRNAs.

Parhyale microRNA arm switching in development and evolution.

Each microRNA hairpin precursor is processed to produce two possible mature products, often of unequal abundance. Historically, the less abundant product was termed the miR* sequence, and was presumed to be degraded. Recently, this view has been abandoned with the discovery that for some animal microRNAs, arm dominance can differ between tissues, developmental times or species. Additionally, studies have shown that each arm can have many different targets, and that both arms can be functional. Arm switching therefore has the potential to diversify microRNA function[22,23].

We have examined developmental and evolutionary arm switching of all the predicted *Parhyale* microRNAs (Fig. 3A). Almost all microRNAs showed the same dominant arm throughout the course of development. However, a small proportion of microRNAs exhibit developmental arm switching (Fig. 3A). For some microRNAs, a pronounced switch in dominance was observed across the short timescale of adjacent time-points, for example mir-14127 and mir-14149a-1. For many microRNAs, approximately equal proportions of 5p and 3p arms were detected at specific time-points (Fig. 3A, white tiles), suggesting that both potential sequences may function to target different mRNAs at the same stages[22].

By comparing arm dominance between datasets for different species, we also identified some cases of arm switching through evolution (Fig. 3B–D). For example, miR-71 is consistently 3'-biased in the flour beetle, spider and honeybee (Fig. 3B–D respectively), but 5'-biased in *Parhyale* (Fig. 3E), suggesting that miR-71 switched arms

Figure 2. Homologs of *Parhyale* microRNAs in crustacean genomes from NCBI. Heatmap representing *Parhyale* microRNA families in the genomes of 37 crustaceans; greyscale indicates the number of members per microRNA family found in each species. Clustering analysis based on microRNA presence/absence was used to generate a species tree (left). The class and order of each species is indicated by the colour-coded ribbon.
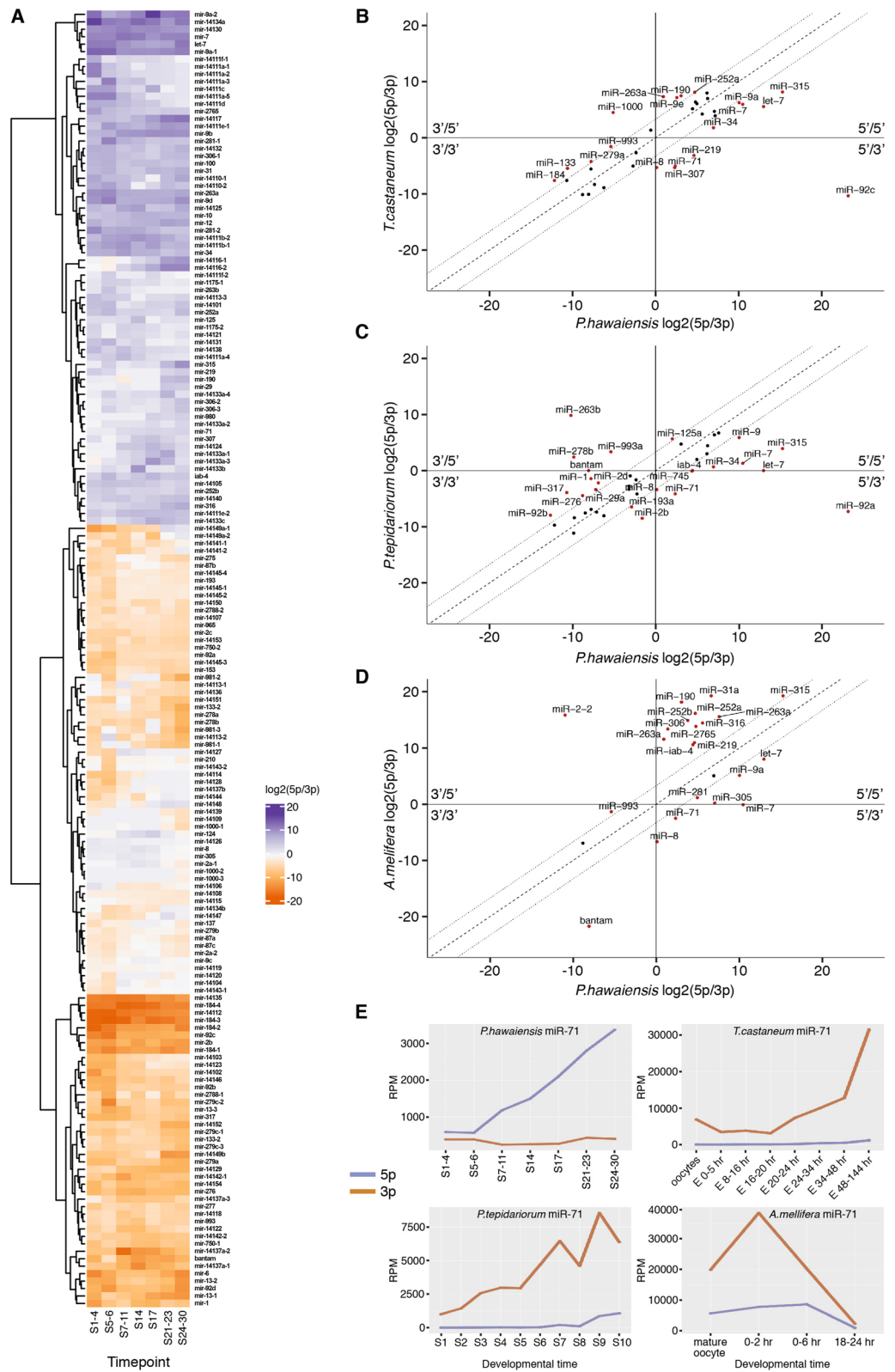
Figure 3. MicroRNA arm switching through development and evolution. (**A**) Heatmap showing relative arm usage changes across seven time-points of *Parhyale* embryonic development. Purple indicates 5' dominance, orange 3' dominance. Comparison of the relative arm usage of microRNA homologs between *Parhyale* and *Tribolium* (**B**), spider (**C**) and honeybee (**D**). All three show microRNAs that have undergone arm switching (5'/3' or 3'/5' quadrants). Dotted lines show the tenfold difference boundary, name labels are shown for each microRNA exceeding this tenfold change in arm usage. (**E**) miR-71-5p and -3p arm expression data through development for the four species analysed.

in evolution during arthropod diversification. miR-8 is also switched in *Parhyale* when compared to the other three species, although less dramatically, changing from a 3' bias in all three species, to approximately equal arm usage in *Parhyale* (Fig. 3B–D).

MicroRNA expression dynamics in embryogenesis.   We have used normalised read counts to quantify the expression of all 349 predicted *Parhyale* mature microRNAs throughout embryogenesis. Principle component analysis (PCA—Fig. 4A) using the mature read counts confirmed high similarity among replicates within each time-point, and also showed high similarity between the first two time-points (S1–4 and S5–6). These findings were confirmed with Spearman correlation tests for all pairwise combinations of expression profiles among the seven time-points (Fig. 4B). Previous estimates of ZGA place the timing between S4 and S6, and therefore large-scale expression differences might be expected between our first two time-points. Our observation of similar profiles between S1–4 and S5–6 suggests that zygotic transcription of microRNAs has not yet begun at S5–6. In contrast, the S7–11 time-point is clearly separated from the earlier stages in the PCA analysis, and the correlation coefficient between S5–6 and S7–11 is the lowest of any pair of adjacent time-points (Fig. 4B). We therefore suggest that the onset of zygotic microRNA expression occurs between S6 and S7. The mid-stages of embryogenesis (S7–11, S14, S17) show a similar microRNA composition (r > 0.9), which is distinct from the early stages. The microRNA composition by the end of embryogenesis (S24–30) is markedly different from both early and mid-embryogenesis (r < 0.7). The penultimate time-point (S21–23) is transitionary between mid and late embryogenesis.

Of the 349 mature microRNAs annotated, 234 (67%) had relatively high expression levels (≥ 10 normalized counts) in at least 1 time-point (see Fig. 4C). We find that at least 172 mature microRNAs are maternally provided in the fertilized egg (S1–4), whereas the number of microRNAs present in the early embryo drops slightly to 167 by S5–6. This drop is likely due to degradation without additional microRNA transcription, consistent with the suggestion that accumulation of zygotic microRNA transcripts does not occur until after S6. The number of expressed microRNAs is steady throughout mid-embryogenesis (S7-S11: 163; S14: 164; S17: 165), with a slight increase in the last two time-points (S21–23: 174; S24–30: 178).

MicroRNAs with similar expression profiles across the time course are likely to be involved in similar developmental processes. We therefore used a fuzzy c-means clustering approach to group microRNAs with similar expression dynamics. Unlike k-means, fuzzy c-means clustering assigns a membership coefficient to each microRNA, such that each data point belongs to a greater or lesser degree to each cluster[24]. Using this approach with the 234 mature microRNAs highly expressed in at least one stage, we identified four expression clusters (Fig. 4C). Expression profiles of the most significant microRNAs for each cluster (membership cut-off 0.6) are shown in Fig. 4D. Cluster 1 (26% of the 234 mature microRNAs) comprises exclusively maternally-loaded microRNAs, which are expected to function in the early embryo even before the onset of ZGA. Cluster 2 is composed of the first microRNAs to be expressed during ZGA (15%), cluster 3 represents microRNAs expressed predominantly during mid embryogenesis (20%), and cluster 4 includes the microRNAs expressed almost exclusively at late embryogenesis (39%) (Fig. 4D). We see that the largest number of microRNAs are expressed in late embryogenesis. This finding is similar to results reported in other species including zebrafish[25], *Drosophila virilis*[26] and *Tribolium*[13] where more microRNAs were found to be expressed at later stages, but different from findings in mice[27]. The high number of microRNAs with peak expression in later stages correlates with the increase in the number and variety of differentiated cell-types.

Comparing the distribution of conserved versus newly annotated microRNAs in each cluster revealed that cluster 1 (expressed during early embryogenesis) contains a disproportionately high number of newly annotated microRNAs (42 new, 19 conserved; X^2 (1, N = 61) = 14.95, p = 1.10 × 10^{-4}), whereas cluster 4 (expressed late in embryogenesis) contains a significantly higher proportion of conserved microRNAs (20 new: 72 conserved, X^2 (1, N = 92) = 19.40, p = 1.05 × 10^{-5}) (Fig. 4E). The abundance of evolutionarily young, lineage-specific microRNAs in early embryonic stages has also been described in other arthropod species such as *Drosophila virilis*[26], *Tribolium*[13] and *Blatella germanica*[28].

mRNA expression dynamics in embryogenesis.   To compare mRNA and small RNA expression dynamics in *Parhyale,* we use Trinity-based pipeline to perform genome-guided annotation of the developmental transcriptome using poly-A selected RNA-seq datasets collected across the same samples as above. We annotated a total of 49,532 protein-coding transcripts from 31,087 Trinity genes. Details of the transcriptome annotation pipeline and statistics are shown in Supplementary Table 2.

PCA analysis of transcript expression profiles defined by read counts clearly shows good agreement between the two replicates per time-point (Fig. 5A). As with the microRNA analysis, both PCA analysis and Spearman's correlation coefficients (Fig. 5B) show that the mRNA content at the two earliest stages of embryogenesis (S1–4 and S5–6) is similar, but markedly different from later time-points. Correlation scores show that the final stage (S24–30) is also very different from all preceding time-points, indicating that a distinct set of mRNAs is engaged at the end of embryogenesis, presumably in establishing the final RNA profiles of adult tissues (Fig. 5B).

A total of 36,119 mRNAs had relatively high expression levels (≥ 10 normalized counts) in at least one time-point (Fig. 5C). As with microRNAs, we clustered mRNA expression profiles using fuzzy c-means clustering[24] yielding five different clusters. Expression profiles of the most significant mRNAs for each cluster (membership cut-off 0.6) are shown in Fig. 5D. Cluster 2 represents mRNAs with peak expression at S5–6 (Fig. 5C,D), immediately after the time of ZGA previously reported in *Parhyale*[20]. This cluster is absent from the microRNA expression profiles (Fig. 4C,D), suggesting that microRNAs are not generally transcribed during the first stages of ZGA. Indeed, 47% of the mRNA transcripts belong to clusters 1 and 2, representing peak expression at the first two time-points, whereas only 26% of microRNAs belong to the cluster that includes high expression at the
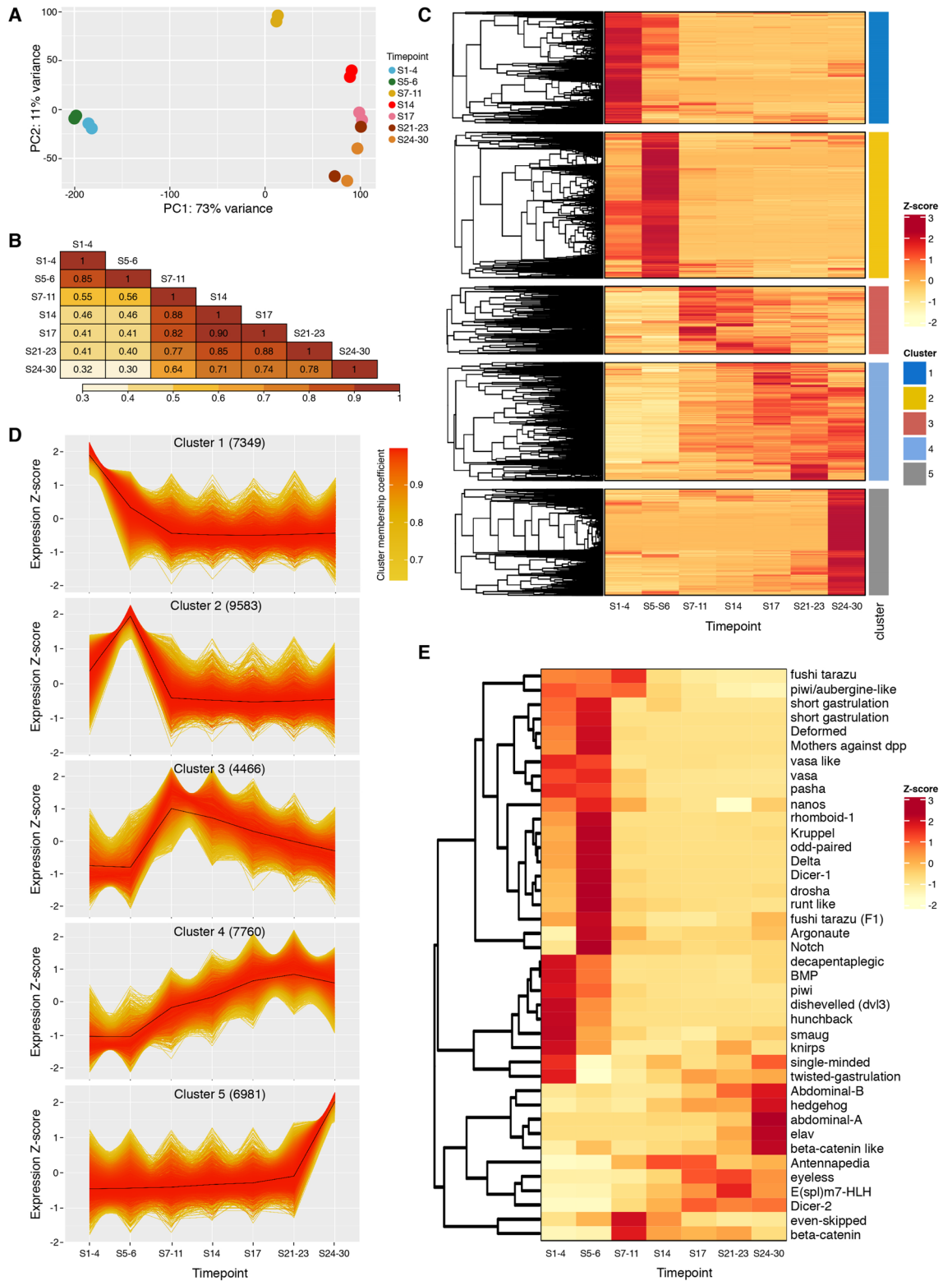
Figure 4. Differential expression analysis of microRNAs during development. (**A**) Principal component analysis (PCA) of microRNA expression in each replicate and time-point. Four replicates are shown for each time-point. (**B**) Heatmap of all-versus-all pairwise Spearman correlation coefficients between time-points. Numbers in tiles are r values, and heatmap colour coding is based on r value. (**C**) Heatmap showing z-score calculated for expression of each microRNA through embryonic development. Each microRNA is classified into an expression cluster, indicated by the colour coded ribbon. (**D**) Expression profiles for microRNAs with membership scores ≥ 0.6 for each cluster; the number indicated in parentheses is the total number of microRNAs belonging to each cluster. (**E**) Composition of the 4 expression clusters in newly annotated and previously annotated microRNAs. Chi-squared tests were performed for each cluster. Cluster 1 shows a significantly higher proportion of newly annotated microRNAs than expected, X^2 (1, $N = 61$) = 14.95, p = 1.10×10^{-4}, whereas cluster 4 contains an unexpectedly high proportion of conserved microRNAs, X^2 (1, $N = 92$) = 19.40, p = 1.05×10^{-5}, p-value significance levels are indicated by asterisks.

Figure 5. Differential expression analysis of mRNAs during development. (**A**) Principal component analysis (PCA) of mRNA expression levels in each replicate and time-point. Two replicates are shown for each time-point. (**B**) Heatmap of all-versus-all pairwise Spearman correlation coefficients calculated between time-points. Numbers in tiles are r values, and heatmap colour coding is based on r value. (**C**) Heatmap showing z-score calculated for expression of each mRNA through embryonic development. Each mRNA is classified into an expression cluster, indicated by the colour coded ribbon. (**D**) Expression profiles for mRNAs with membership scores ≥ 0.6 for each cluster; the number indicated in parentheses is the total number of mRNAs belonging to each cluster. (**E**) Heatmap showing z-score of a subset of known developmental genes extracted from (D).
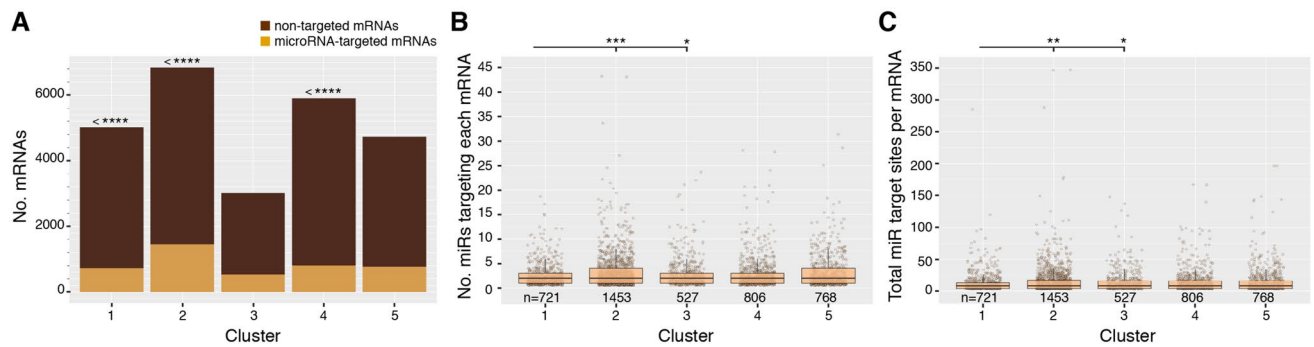
Figure 6. Differential targeting of mRNAs by microRNAs through development. (**A**) The number of mRNAs in each expression cluster that are predicted to be targeted by microRNAs versus those that are non-targeted. Hypergeometric tests were performed to compare the observed numbers with expected values, calculated based on the overall proportion of targeted mRNAs within the entire mRNA population. Significance values are indicated by asterisks. Cluster 1 is under-enriched for targeted mRNAs, $padj = 1.62 \times 10^{-9}$; cluster 2 is over-enriched for targeted mRNAs, $padj = 6 \times 10^{-45}$; and cluster 4 is under-enriched for targeted mRNAs, $padj = 8.25 \times 10^{-19}$. (**B**) The number of different microRNAs targeting each mRNA in each expression cluster. Pairwise Mann–Whitney-Wilcoxon tests were performed between all clusters; significant comparisons are highlighted ($padj < 0.05$): Cluster 1 vs 2, $padj = 5.76 \times 10^{-4}$; cluster 1 vs 3, $padj = 2.45 \times 10^{-2}$. (**C**) The number of microRNA targeting sites per mRNA 3'UTR in each expression cluster. Pairwise Mann–Whitney-Wilcoxon tests were performed between all clusters; significant comparisons are shown ($padj < 0.05$): Cluster 1 vs 2, $padj = 4.48 \times 10^{-3}$; cluster 1 vs 3, $padj = 3.31 \times 10^{-2}$.

first two time-points. Conversely, only 19% of mRNAs fall into cluster 5 (peak expression at S24–30), whereas the equivalent cluster 4 for microRNAs contained 39% of the microRNAs. These data clearly suggest that zygotic expression of many mRNAs is initiated before zygotic microRNAs are expressed, and that mRNAs have functional roles early in embryogenesis independent of microRNAs, while a high proportion of microRNAs function late.

We examined the mRNA abundance of a number of specific genes known to play important roles during embryogenesis or in microRNA biogenesis in other species (Fig. 5E). For example, mRNAs including *nanos*, *hunchback*, *dishevelled* and *smaug* are all known to be maternally-loaded in other species, and the mRNAs of predicted homologs were also present at high levels in the *Parhyale* early embryo. In *Drosophila*, the RNA-binding protein Smaug is an important player during MZT, responsible for the degradation of hundreds of maternally-loaded mRNAs[29]. In *Parhyale*, normalized sequencing data suggests that the mRNA of the *smaug* homolog is more abundant during the time points S1–4, consistent with a conserved biological function.

In accordance with other studies in *Parhyale*, our analysis failed to identify an unambiguous *zelda* ortholog. However, we identified a homolog of *odd-paired* mRNA, a pioneering factor suggested to be a key player during ZGA[30]. The *odd-paired* homolog is maternally-loaded and its relative abundance increases at S5–6, therefore showing the same behaviour as *Dmel zld*. Expression of homologs of other *Drosophila* pair-rule genes (*eve*, *ftz* and *runt*) also increased during ZGA. Conservation of temporal expression was also observed for mRNAs known to be expressed during late embryogenesis, such as *eyeless*, *elav* and *E(spl)m7-HLH* involved in eye development, axon guidance and neurogenesis respectively. Expression of *piwi* and *vasa*, components of the piRNA processing pathway are predominantly expressed in the early embryo, consistent with previous studies[31], whereas *Dicer-2* (implicated in siRNA processing) is primarily expressed at mid to late stages. This hints that piRNAs are likely to play important roles in the early embryo, whereas siRNA function may be more prominent later in development. Interestingly, relative abundance of mRNAs encoding known microRNA processing proteins such as *Dicer-1*, *drosha* and *pasha* all peak early at S5–6 when the zygotic genome first becomes active.

Comparative expression dynamics of microRNAs and their predicted targets.
Our paired, size-separated libraries allow the analysis of temporal expression of microRNAs in combination with their targets (mRNAs). Target predictions were performed using the SeedVicious algorithm and potential interactions then filtered (see methods) to produce a list of putative microRNA-mRNA interactions. To assess the degree to which mRNAs are targeted by microRNAs through development, we analysed the proportion of mRNAs in each expression cluster that are predicted to be targeted by microRNAs (Fig. 6A). Of the total 36,119 mRNAs assigned to expression clusters, 21,243 (59%) had 3' UTRs, and of these, 4,275 (12% of total expressed mRNAs) were predicted to be targeted by microRNAs. Using a hypergeometric test, we find that clusters 1 and 4 were significantly under-enriched for mRNAs targeted by microRNAs, whereas cluster 2 was significantly over-enriched for targeted mRNAs (Fig. 6A). Cluster 1 primarily contains maternally-loaded mRNAs, whereas cluster 2 spans the beginning of zygotic transcription.

We also compared the number of different microRNAs targeting each mRNA per expression cluster (Fig. 6B), and the number of microRNA target sites in each 3' UTR per cluster (Fig. 6C). Pairwise Mann–Whitney–Wilcoxon tests between the five clusters in all combinations (with Bonferroni correction) revealed that mRNAs in cluster 1 were targeted by significantly fewer different microRNAs, and also had significantly fewer microRNA target sites in their 3' UTRs than mRNAs in clusters 2 and 3 (see Fig. 6B). These data therefore suggest that
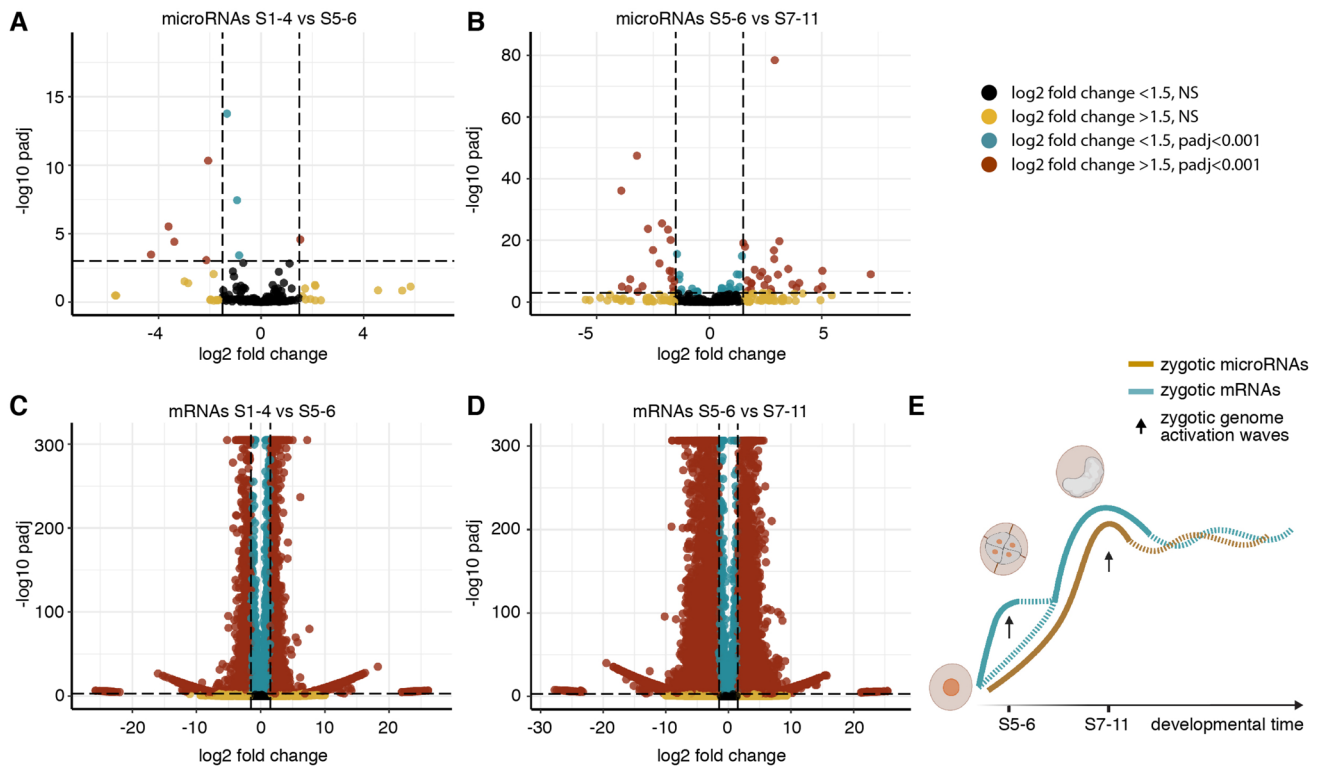
Figure 7. Differential expression analysis during zygotic genome activation. (**A,B**) Volcano plots showing log2 fold change in expression (x-axis) versus the p-value (y-axis), for each microRNA expressed between the first two time-points S1–4 to S5–6 (A) and S5–6 to S7–11 (B). (**C,D**) Volcano plot showing log2 fold change in expression (x-axis) versus the p-value (y-axis) for each mRNA expressed between the first two time-points S1–4 to S5–6 (**C**) and S5–6 to S7–11 (**D**). Only red dots (log2 fold change ≤ -1.5 or ≥ 1.5 with padj ≤ 0.001 are considered significant. (**E**) Model of zygotic genome activation occurring in two different waves of expression for the mRNAs; onset of microRNA expression occurs only in the second wave.

globally, microRNAs are more involved in regulating zygotically-expressed genes than in clearing maternally-loaded transcripts.

Differential expression analysis of mRNAs and microRNAs identifies two waves of ZGA. To further explore the apparent developmental lag in zygotic expression of microRNAs with respect to mRNAs, we compared the expression levels of mRNAs and microRNAs between adjacent time-points, S1–4 with S5–6, and S5–6 with S7-S11, using DESeq2 (Fig. 7). Changes between S1–4 and S5–6 could be due to degradation of maternally-loaded RNAs, or onset of zygotic RNA production. We find that the expression of only one microRNA (mir-14137a-3p) increased significantly (log2 fold change ≥ 1.5, padj ≤ 0.001) between S1–4 and S5–6 (Fig. 7A). The overwhelming majority of microRNAs either decreased, likely signifying degradation without replacement, or did not significantly change. In contrast, a total of 27 different microRNAs were significantly upregulated between S5–6 and S7–11, representing 11.5% of all microRNAs expressed during development (Fig. 7B). An equivalent analysis of the mRNAs showed that 2,248 mRNAs (6.2% of all expressed mRNAs in development) were significantly upregulated (log2 fold change ≥ 1.5, padj ≤ 0.001) between S1–4 and S5–6 (Fig. 7C). Between S5–6 and S7–11, a total of 6,150 mRNAs (17.0%) were upregulated (log2 fold change ≥ 1.5, padj ≤ 0.001) (Fig. 7D). We therefore propose a model where a subset of protein-coding genes are activated in a first wave of ZGA between S1–4 and S5–6, with expression of microRNAs accompanying a larger number of mRNAs that are activated at a later point during the second (and biggest) wave of ZGA between S5–6 and S7–11 (Fig. 7E). Additional equivalent analysis of putative lncRNAs revealed a pattern of activation intermediate between micro-RNAs and mRNAs, with 2,485 lncRNAs (0.9% of all expressed lncRNAs) significantly upregulated (log2 fold change ≥ 1.5, padj ≤ 0.001) between S1–4 and S5–6, and 9,917 (3.6%) significantly upregulated between S5–6 and S7–11 (Supplementary Fig. 1). Gene set enrichment analysis suggests that mRNA transcripts upregulated in the first wave are enriched for Gene Ontology terms related to metabolic pathways and transcription activation, whereas mRNAs upregulated during the second wave are enriched for RNA binding activity and translation regulation (Supplementary Fig. 2).

Discussion

We have generated paired, size-separated libraries across different stages of *Parhyale* embryogenesis, providing both microRNA and mRNA transcriptomes for this crustacean model organism during embryonic development. We have identified a total of 349 mature microRNAs expressed from 175 microRNA precursors in the genome of *Parhyale*. Of the precursor loci, 87 have been previously described while 88 are unrelated to any previously identified microRNAs in any species. We have used this dataset to provide a first glimpse at the microRNAome of 37 other crustacean species, the majority of which have no microRNA expression data or annotation available. This work enables and accelerates investigation of crustacean microRNAs, which others have argued are excellent markers for phylogenetic inference[32–34]. However, our methodology is *Parhyale*-centric – we can only detect conservation of *Parhyale* microRNAs in other crustaceans, but not microRNAs that are novel in other crustacean clades or lost in *Parhyale*. More sampling of microRNAs in crustaceans is clearly required.

Analyses of microRNA expression in other animals have shown that the early embryo is a highly permissive environment, over-enriched for evolutionarily young, lineage-specific microRNAs, an observation that we can now extend from the Holometabolan insects to the Pancrustaceans[26]. This emerging theme suggests insights into microRNA birth, selection, and death processes. We also find that highly conserved microRNAs dominate the later stages of development, in agreement with previous studies[13,26].

Precise determination of the timing of MZT and ZGA by transcriptomics is complicated by a number of factors, including the relative abundance of maternal and zygotic transcripts, synthesis and degradation rates, and numbers of cells at different embryonic stages. Further work is required to clarify the exact timing of events surrounding MZT. However, our data clearly suggest that ZGA in *Parhyale* occurs in two waves. In the first wave, transcription is primarily associated with a set of early mRNAs and lncRNAs, while the onset of microRNA transcription is almost exclusively limited to the second wave, along with additional mRNAs and lncRNAs. This observation of two waves of ZGA has been reported in several other species[35]. In *Drosophila*, the first wave of transcription is widespread across the genome, producing short, inefficiently-processed transcripts[36,37]. The function of transcription at these loci could be to activate regions of the genome for later transcription of competent mRNAs. However, many of the short early transcripts have been found to be implicated in the sex-determination pathway, thus suggesting function beyond genome activation[37]. Furthermore, specific sets of genes are also known to be expressed in two waves in mice; for example, paternal genes are expressed preferentially during the minor ZGA. In the chicken, the opposite was observed with the paternal transcriptome only being activated during the second wave[38]. Also in chicken, microRNAs are predominantly transcribed during the second wave of ZGA, as we see in *Parhyale*[38].

Target prediction showed that maternally-loaded mRNAs are targeted by fewer microRNAs and have fewer microRNA target sites than later expressed mRNAs. We also see that a large proportion of microRNAs show peak expression at the very end of embryogenesis, suggesting that in *Parhyale*, microRNAs might be more active players during the late stages of development. This is in contrast to the mRNAs, almost half of which show peak expression at the earliest two time-points, while a relatively small proportion peak at late embryogenesis. These findings all point to microRNA regulation being more prevalent for zygotic genes than for maternally-loaded transcripts. This may reflect the importance of microRNAs in balancing and buffering active transcription[39], a role less necessary for maternal transcripts that are not being actively replenished, and which may instead degrade with time via other passive or active mechanisms.

For decades, *Drosophila* has held a virtual monopoly over transcriptomics studies in arthropods, and much of our knowledge today about development is thanks to the fly community. However, other organisms provide models for evolutionary questions that cannot be tackled in *Drosophila* alone. We provide annotation of the *Parhyale* transcriptome (both mRNAs and microRNAs) throughout embryonic development. To our knowledge this is the first publicly-available study in *Parhyale* that provides temporal resolution throughout embryogenesis with tightly spaced time-points, and representation of major developmental transitions such as ZGA, germ band extension, and morphogenesis. This work helps to establish *Parhyale* as a model for questions related to the evolution of crustaceans and insects and facilitates functional studies of microRNAs during crustacean development.

Materials and methods

Parhyale hawaiensis culture, sample preparation and library construction.

Wild-type *Parhyale* were kindly donated by Aziz Aboobaker's lab at Oxford University. Animals were reared in standard plastic aquarium tanks containing artificial sea water (aquarium salt and deionised water) at a salinity of 30 PPT, and kept at ~26°C. Cultures were aerated with aquarium pumps and airstones, water was changed once a week and animals were fed fish flakes and carrots. Embryos were manually collected from the ventral brood pouch of gravid females anaesthetised using clove oil (Sigma) diluted 1:10,000 in sea water. Embryos were washed in filtered seawater and manually staged using a Leica Stereo Fluorescence microscope. Isolated embryos were stored in RNAlater (Sigma) and total RNA was then extracted using the SPLIT RNA Extraction Kit (Lexogen) following the manufacturer's instructions. Small RNA libraries (4 replicates per time-point) were built using the Small RNA-Seq Library Prep Kit (Lexogen). Long library fragments and linker-linker artefacts were removed using a purification module with magnetic beads (Lexogen). Long mRNA libraries (2 replicates per time-point) were built using the TruSeq Stranded mRNA HT Sample Prep Kit (Illumina). Library concentration was assessed for all libraries using the Qubit fluorimetric system (Invitrogen) and quality was assessed using the Agilent 2200 TapeStation. Sequencing was performed at the University of Manchester Genomic Technologies Facility.

Small RNA-seq data analysis and microRNA prediction.

RNA-seq raw reads were trimmed using Cutadapt v. 1.18[40] and read length distribution was assessed using FastQC v0.11.8[41]. For microRNA predictions, reads ranging from 18 to 26 nt were retained. These reads were mapped to *Parhyale* tRNAs and rRNAs using

Bowtie (v1.1.1; parameters -p4 -v3 –un)[42]. *Parhyale* tRNAs were predicted using tRNAscan-SE (v2.0; option -e)[43,44] and crustacean rRNAs were downloaded from RNAcentral release 16[45]. Non tRNA/rRNA reads were then mapped to the *P. hawaiensis* genome (Phaw 5.0; GCA_001587735.2) using mapper.pl from the miRDeep2 suite (with options -h -i -j -m), and the mapped reads were used for microRNA annotation using the miRDeep2 tool (v 0.1.1)[46]. To run miRDeep2 we used all the metazoan microRNAs available on miRBase as references (v 22.1)[47]. Predicted microRNAs were manually filtered to keep microRNAs obeying the following criteria[48]: at least 10 reads for both the 5p and 3p mature sequences, minimum loop length of 8 nt, and at least 50% of the reads for each mature microRNA having the same 5' end. Exceptions were only made for highly conserved microRNAs that are confidently annotated in other species, predicted using BLASTN (v2.6.0+; -word_size 4 -reward 2 -penalty -3 -evalue 0.01 -perc_identity 70)[49] hits and verified by manual inspection.

Identification of *Parhyale* microRNA homologs in crustacean genomes.

We downloaded the genomes of 37 crustacean species available in the NCBI Assembly database (Supplementary Table 1) and searched for homologs of all the 175 predicted microRNAs in *Parhyale* using BLASTN (-word_size 4 -evalue 0.01 -reward 2 -penalty -3 -perc_identity 30). A presence/absence matrix of microRNA families and family copy number was plotted in R using the package pheatmap (v1.0.12)[50] with default clustering settings (Euclidian distance).

Relative arm usage.

Homologs of *Parhyale* precursors in three other species with expression data available – *Tribolium castaneum*, *Apis mellifera* and *Parasteatoda tepidariorum* – were identified by BLASTN (-task blastn-short -evalue 0.01) and manual inspection. Read counts for *T. castaneum* mature microRNAs were obtained from Ninova et al.[13], counts for *A. mellifera* from Pires et al.[51], and counts for *P. tepidariorum* were calculated in-house using methods and data from Leite et al.[52]. Relative arm usage was calculated using the method described in Marco et al.[53]: $\log_2(N5'/N3')$; where $N5'$ is the number of reads mapped to the -5p arm, and $N3'$ is the number of reads mapped to the -3p arm.

Transcriptome assembly and annotation.

Paired-end RNA-seq reads from each developmental library were mapped to the *Parhyale* genome (Phaw 5.0; GCA_001587735.2) using STAR (v2.7.2b)[54]. The mapped reads were then assembled using Trinity (v2.9.0)[55]. and the resulting transcripts were mapped back to the genome with gmap (version 2020–06-01)[56], and duplicates removed. Only these transcripts were used for further analysis. Transdecoder (v5.5.0) (https://github.com/TransDecoder/TransDecoder) was used to identify potential coding regions within these transcripts and only the longest ORF per transcript was kept. Using BLAST search (Uniprot release 2020_02, BLASTP version 2.9.0, e-value $\leq 10^{-6}$), we looked for ORFs with similarity to known proteins in 7 crustacean species with annotated transcriptomes (*Daphnia pulex*, *Daphnia magna*, *Penaeus vannamei*, *Armadillidium nasatum*, *Armadillidium vulgare, Portunus trituberculatus*, and *Tigriopus californicus*), as well as *Drosophila melanogaster* and *Apis mellifera*.

Functional annotation.

We searched for protein signatures in the Pfam database (version 33.1)[57] using hmmscan (HMMER version 3.3.2)[58]. Using BLAST (BLASTP version 2.9.0; e-value $\leq 10^{-3}$), we searched our peptide sequences against annotated Swissprot proteins (Uniprot release 2021_02). The Pfam and Uniprot hits were then loaded into a Trinotate sqlite database (v3.2.2)[59], which provided KEGG, EGGNOG and GO terms associated with each transcript.

Quantification and differential gene expression analysis.

To quantify small RNAs, reads not mapping to tRNAs/rRNAs were mapped to the predicted mature microRNAs, using Bowtie (v1.1.1; -v 1 -S -a) and mapped reads were quantified using salmon quant from salmon (v0.14.1)[60] in alignment mode. To quantify mRNAs, reads mapped to the annotated transcriptome were quantified using salmon quant from salmon (v0.14.1) in mapping base mode.

Quantifications were then used for differential expression analysis using the package DESeq2 (v1.28.1)[61] in R Studio (v1.3.1056)[62] R v4.0.2[63]. To group each mRNA or microRNA into expression clusters we applied fuzzy c-means clustering using the function cmeans from the R package e1071 (v1.7.4)[64] to the normalized counts computed using DESeq2. The number of clusters for each dataset was previously determined using the elbow method. Heatmaps were computed using the R package ComplexHeatmap (v2.5.5)[65]. The proportion of previously annotated vs newly annotated microRNAs belonging to each expression cluster was assessed by chi-squared tests, using the ratio within the total population of expressed microRNAs to generate expected values.

Target prediction.

Targets of our annotated microRNAs within the predicted UTRs of our annotated mRNA transcripts were predicted using Seedvicious (v1.3)[66] and filtered adhering to the following criteria: free energy below –10 kcal/mol, microRNAs and mRNAs expressed with at least 10 normalised counts, and each microRNA targets the same UTR more than once. Pairwise Mann–Whitney-Wilcoxon tests were performed between all possible pairs of the five different mRNA expression clusters, to compare the number of different microRNAs targeting each mRNA and the number of microRNA targeting sites per mRNA 3'UTR. Enrichment of microRNA-targeted mRNAs in each mRNA cluster was assessed using the phyper formula of the hypergeometric distribution in R as follows: phyper(q-1, m, n, k, lower.tail=FALSE), where q=successes in subset, m=successes in population, n=population total − successes in population, and k=subset. All p-values were adjusted using the Bonferroni method in R.

Gene Ontology (GO) annotation and GO enrichment analysis. Significantly upregulated mRNAs were subjected to gene set enrichment analysis using the TopGO package v2.42.0 in R[67]. The classic Fisher test was used to generate enrichment p-values, with the algorithm weight01 and a p-value cutoff of $p < 0.01$.

Data availability

All RNA sequencing data and quantifications were deposited in the Gene Expression Omnibus (GEO) database under accession number GSE178877.

References

1. Kao, D. *et al.* The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *Elife* **5**, 20062 (2016).
2. Zeng, V. *et al.* De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean *Parhyale hawaiensis*. *BMC Genom.* **12**, 581 (2011).
3. Blythe, M. J. *et al.* High through-put sequencing of the *Parhyale hawaiensis* mRNAs and microRNAs to aid comparative developmental studies. *PLoS ONE* **7**, e33784 (2012).
4. Martin, A. *et al.* CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. *Curr. Biol.* **26**, 14–26 (2016).
5. Browne, W. E., Price, A. L., Gerberding, M. & Patel, N. H. Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiensis*. *Genesis* **42**, 124–149 (2005).
6. Hannibal, R. L., Price, A. L. & Patel, N. H. The functional relationship between ectodermal and mesodermal segmentation in the crustacean, Parhyale hawaiensis. *Dev. Biol.* **361**, 427–438 (2012).
7. Serano, J. M. *et al.* Comprehensive analysis of Hox gene expression in the amphipod crustacean *Parhyale hawaiensis*. *Dev. Biol.* **409**, 297–309 (2016).
8. Alwes, F., Hinchen, B. & Extavour, C. G. Patterns of cell lineage, movement, and migration from germ layer specification to gastrulation in the amphipod crustacean *Parhyale hawaiensis*. *Dev. Biol.* **359**, 110–123 (2011).
9. Gerberding, M., Browne, W. E. & Patel, N. H. Cell lineage analysis of the amphipod crustacean *Parhyale hawaiensis* reveals an early restriction of cell fates. *Development* **129**, 5789–5801 (2002).
10. Tang, G. siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.* **30**, 106–114 (2005).
11. Pasquinelli, A. E. Molecular biology. Paring miRNAs through pairing. *Science* **328**, 1494–1495 (2010).
12. Bushati, N., Stark, A., Brennecke, J. & Cohen, S. M. Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila*. *Curr. Biol.* **18**, 501–506 (2008).
13. Ninova, M., Ronshaugen, M. & Griffiths-Jones, S. MicroRNA evolution, expression, and function during short germband development in *Tribolium castaneum*. *Genome Res.* **26**, 85–96 (2016).
14. Ylla, G., Piulachs, M. D. & Belles, X. Comparative analysis of miRNA expression during the development of insects of different metamorphosis modes and germ-band types. *BMC Genom.* **18**, 774 (2017).
15. Giraldez, A. J. *et al.* Zebrafish miR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
16. Lund, E., Liu, M., Hartley, R. S., Sheets, M. D. & Dahlberg, J. E. Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA* **15**, 2351–2363 (2009).
17. Stoeckius, M. *et al.* Global characterization of the oocyte-to-embryo transition in *Caenorhabditis elegans* uncovers a novel mRNA clearance mechanism. *EMBO J.* **33**, 1751–1766 (2014).
18. Yang, Q. *et al.* Highly sensitive sequencing reveals dynamic modifications and activities of small RNAs in mouse oocytes and early embryos. *Sci. Adv.* **2**, e1501482 (2016).
19. Dexheimer, P. J., Wang, J. & Cochella, L. Two microRNAs are sufficient for embryonic patterning in *C. elegans*. *Curr. Biol.* **30**, 5058–5065 (2020).
20. Nestorov, P., Battke, F., Levesque, M. P. & Gerberding, M. The maternal transcriptome of the crustacean *Parhyale hawaiensis* is inherited asymmetrically to invariant cell lineages of the ectoderm and mesoderm. *PLoS ONE* **8**, e56049 (2013).
21. Copilas-Ciocianu, D., Borko, S. & Fiser, C. The late blooming amphipods: Global change promoted post-Jurassic ecological radiation despite Palaeozoic origin. *Mol. Phylogenet. Evol.* **143**, 1064 (2020).
22. Griffiths-Jones, S., Hui, J. H., Marco, A. & Ronshaugen, M. MicroRNA evolution by arm switching. *EMBO Rep.* **12**, 172–177 (2011).
23. Marco, A., MacPherson, J. I., Ronshaugen, M. & Griffiths-Jones, S. MicroRNAs from the same precursor have different targeting properties. *Silence* **3**, 8 (2012).
24. Futschik, M. E. & Carlisle, B. Noise-robust soft clustering of gene expression time-course data. *J. Bioinform. Comput. Biol.* **3**, 965–988 (2005).
25. Wienholds, E. *et al.* MicroRNA expression in zebrafish embryonic development. *Science* **309**, 310–311 (2005).
26. Ninova, M., Ronshaugen, M. & Griffiths-Jones, S. Fast-evolving microRNAs are highly expressed in the early embryo of *Drosophila virilis*. *RNA* **20**, 360–372 (2014).
27. Rahmanian, S. *et al.* Dynamics of microRNA expression during mouse prenatal development. *Genome Res.* **29**, 1900–1909 (2019).
28. Ylla, G., Fromm, B., Piulachs, M. D. & Belles, X. The microRNA toolkit of insects. *Sci. Rep.* **6**, 37736 (2016).
29. Benoit, B. *et al.* An essential role for the RNA-binding protein Smaug during the *Drosophila* maternal-to-zygotic transition. *Development* **136**, 923–932 (2009).
30. Soluri, I. V., Zumerling, L. M., Payan Parra, O. A., Clark, E. G. & Blythe, S. A. Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the *Drosophila* segmentation network. *Elife* **9**, 53916 (2020).
31. Ozhan-Kizil, G., Havemann, J. & Gerberding, M. Germ cells in the crustacean *Parhyale hawaiensis* depend on Vasa protein for their maintenance but not for their formation. *Dev. Biol.* **327**, 230–239 (2009).
32. Wheeler, B. M. *et al.* The deep evolution of metazoan microRNAs. *Evol. Dev.* **11**, 50–68 (2009).
33. Tarver, J. E. *et al.* Well-annotated microRNAomes do not evidence pervasive miRNA loss. *Genome Biol. Evol.* **10**, 1457–1470 (2018).
34. Fromm, B. *et al.* The metazoan microRNA complement: MirGeneDB 2.0. *Nucl. Acids Res.* **48**, D132–D141 (2020).
35. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
36. Abe, K. I. *et al.* Minor zygotic gene activation is essential for mouse preimplantation development. *Proc. Natl. Acad. Sci. USA* **115**, E6780–E6788 (2018).
37. Kwasnieski, J. C., Orr-Weaver, T. L. & Bartel, D. P. Early genome activation in *Drosophila* is extensive with an initial tendency for aborted transcripts and retained introns. *Genome Res.* **29**, 1188–1197 (2019).
38. Hwang, Y. S. *et al.* Zygotic gene activation in the chicken occurs in two waves, the first involving only maternally derived genes. *Elife* **7**, 39381 (2018).
39. Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149**, 515–524 (2012).

40. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
41. Andrews, S. FASTQC. A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
42. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
43. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**, 1–14 (2019).
44. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res.* **25**, 955–964 (1997).
45. RNAcentral Consortium. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucl. Acids Res.* **49**, D212-D220 (2021).
46. Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucl. Acids Res.* **40**, 37–52 (2012).
47. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: From microRNA sequences to function. *Nucl. Acids Res.* **47**, D155–D162 (2019).
48. Fromm, B. *et al.* A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microR-NAome. *Annu. Rev. Genet.* **49**, 213–242 (2015).
49. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
50. Kolde, R. in *pheatmap: pretty heatmaps. R package version 1.0.12.* https://CRAN.R-project.org/package=pheatmap (2019).
51. Pires, C. V., Freitas, F. C., Cristino, A. S., Dearden, P. K. & Simoes, Z. L. Transcriptome analysis of honeybee (*Apis mellifera*) haploid and diploid embryos reveals early zygotic transcription during cleavage. *PLoS One* **11**, e0146447 (2016).
52. Leite, D. J. *et al.* Pervasive microRNA duplication in Chelicerates: Insights from the embryonic microRNA repertoire of the spider *Parasteatoda tepidariorum*. *Genome Biol. Evol.* **8**, 2133–2144 (2016).
53. Marco, A., Hui, J. H., Ronshaugen, M. & Griffiths-Jones, S. Functional shifts in insect microRNA evolution. *Genome Biol. Evol.* **2**, 686–696 (2010).
54. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
55. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech.* **29**, 644–652 (2011).
56. Wu, T. D. & Watanabe, C. K. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
57. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucl. Acids Res.* **49**, D412–D419 (2021).
58. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucl. Acids Res.* **46**, W200–W204 (2018).
59. Bryant, D. M. *et al.* A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **18**, 762–776 (2017).
60. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
62. RStudio Team. RStudio: Integrated Development Environment for R. (2020).
63. R Team. R: A Language and Environment for Statistical Computing. (2020).
64. Meyer, D. *et al.* Package 'e1071'. *The R Journal* (2019).
65. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
66. Marco, A. SeedVicious: Analysis of microRNA target and near-target sites. *PloS One* **13**, e0195532 (2018).
67. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment analysis for Gene Ontology. https://bioconductor.org/packages/release/bioc/html/topGO.html (2020).

Author contributions

Concept and experimental design L.C., M.R. and S.G.J.; culture and embryo handling L.C. and T.P.; RNA-Seq L.C.; data analysis L.C.; microRNA annotation L.C.; transcriptome annotation M.B.; statistical analysis L.C. and T.P.; manuscript writing – L.C., T.P., M.R., and S.G.J.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-03642-9.

Correspondence and requests for materials should be addressed to M.R. or S.G.-J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.