Resource

A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach

Evgeny A. Glazov,[1,3] Pauline A. Cottee,[2,3] Wesley C. Barris,[1] Robert J. Moore,[2] Brian P. Dalrymple,[1] and Mark L. Tizard[2,4]

[1]CSIRO Livestock Industries, Queensland Bioscience Precinct, St. Lucia QLD 4067, Australia; [2]CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong VIC 3220, Australia

MicroRNA (miRNA) and other types of small regulatory RNAs play a crucial role in the regulation of gene expression in eukaryotes. Several distinct classes of small regulatory RNAs have been discovered in recent years. To extend the repertoire of small regulatory RNAs characterized in chickens we used a deep sequencing approach developed by Solexa (now Illumina Inc.). We sequenced three small RNA libraries prepared from different developmental stages of the chicken embryo (days five, seven, and nine) to produce over 9.5 million short sequence reads. We developed a bioinformatics pipeline to distinguish authentic mature miRNA sequences from other classes of small RNAs and short RNA fragments represented in the sequencing data. Using this approach we detected almost all of the previously known chicken miRNAs and their respective miRNA* sequences. In addition we discovered 449 new chicken miRNAs including 88 miRNA candidates. Of these, 430 miRNAs appear to be specific to the avian lineage. Another six new miRNAs had evidence of evolutionary conservation in at least one vertebrate species outside of the bird lineage. The remaining 13 putative miRNAs appear to represent chicken orthologs of known vertebrate miRNAs. We discovered 39 additional putative miRNA candidates originating from miRNA generating intronic sequences known as mirtrons.

[Supplemental material is available online at www.genome.org. The raw sequence data from this study have been submitted to Gene Expression Omnibus under accession no. GSE10686.]

MicroRNAs (miRNAs) are small 21–23 nucleotide regulatory RNAs that modulate gene expression in animals and plants. In animals, regulation of gene expression by miRNAs is achieved by sequence-specific targeting of the 3′ untranslated regions of messenger RNAs by the RNA induced silencing complex (RISC). This results in translational repression of protein synthesis and, in some cases, destabilization of messenger RNA (He and Hannon 2004). The number of newly discovered miRNAs is growing rapidly (Griffiths-Jones et al. 2006). Moreover, several other classes of small regulatory RNAs, distinguished by their origin and biological function, have been identified in recent years (He and Hannon 2004; Chapman and Carrington 2007). These include small interfering RNAs (siRNAs), encompassing *trans*-acting siRNAs (tasiRNAs) and natural antisense transcript derived siRNAs (natsiRNAs), repeat-associated siRNAs (rasiRNAs, also referred to as PIWI-interacting RNAs or piRNAs) (for review see Chapman and Carrington 2007); and a recently identified class of small RNAs associated with gene promoters (PASRs) and 3′ termini (TASRs) (Kapranov et al. 2007).

Identification of comprehensive sets of miRNAs and other small regulatory RNAs in different organisms is a critical step to facilitate our understanding of genome organization, genome biology, and evolution. The chicken embryo continues to be a classic model of vertebrate developmental biology that has been used for many decades (Brown et al. 2003). A rapidly developing embryo is an excellent biological system to study the repertoire

and expression dynamics of small regulatory RNAs. The developmental processes of the embryo are well defined (Hamburger and Hamilton 1992) and live embryos can be readily manipulated in ovo (Brown et al. 2003). Moreover, the recently sequenced chicken genome (red jungle fowl, *Gallus gallus*) is the first non-mammalian amniote genome that represents the *Diapsida* lineage of vertebrate evolution and provides a resource for comparative genomics studies that help our understanding of mammalian evolution (Supplemental Fig. S1) (Hillier et al. 2004). The latest release of the miRNA database (miRBase 10.1) contains 149 chicken miRNA genes that code for 121 distinct mature miRNAs and three sequences originating from the RNA hairpin arm opposite to the annotated mature miRNA containing arm, the so-called miRNA* (Griffiths-Jones et al. 2006). Many of these miRNAs have been identified based only on sequence similarity to known miRNA orthologs and have never been confirmed experimentally (Hillier et al. 2004; Griffiths-Jones et al. 2006). Furthermore, the total number of chicken miRNA genes is currently lower than that identified in mouse (442 miRNA genes), human (533 miRNA genes), or even zebrafish (337 miRNA genes). The latter suggests that there are still many undiscovered miRNAs in birds.

We aimed to extend the known repertoire of small regulatory RNAs expressed in the chicken embryo by using a deep sequencing approach developed by Solexa (now Illumina Inc.). We constructed three small RNA libraries, which were prepared from chicken embryos collected at day five, day seven, and day nine of incubation (hereafter refred to as CE5, CE7, and CE9, respectively). These represent the chicken embryonic developmental stages 25–27, 30–31, and 35, which cover major morphological changes (e.g., limb and craniofacial development) as well as several underlying developmental processes including vasculogen-

---

esis, myogenesis, and osteogenesis (Hamburger and Hamilton 1992). Each library was sequenced individually and generated more than three million short sequence reads resulting in a total of over 9.5 million sequence reads. We developed a bioinformatics pipeline to distinguish authentic mature miRNA sequences from other small RNAs and short RNA fragments represented in the sequencing data. The following manuscript presents a detailed analysis of this sequence data and its interpretation.

Results

An overview of the sequencing results

To simplify the sequencing data, all identical sequence reads in each small RNA library were grouped and converted into sequence tags—unique sequences with associated counts of the individual sequence reads. Although the total numbers of sequence reads in all three RNA libraries were approximately the same (~3.2 million; Supplemental Table S2), the numbers and fractions of the unique sequence tags were substantially different between the libraries (Supplemental Table S3). The CE5 library had the highest number of unique sequence tags; this was followed by the CE7 and then the CE9 libraries. The apparent decrease in the diversity of the small RNA populations from day five to day nine of chick embryo development is also evident in the relative increase in the fraction of sequence tags that were sequenced multiple times in a CE7 and CE9 sample (Supplemental Table S4). These data highlight differences in the overall complexity of the steady-state small RNA pools between the different developmental stages.

To gain insight into possible mechanisms underlying observed changes in the sequence diversity of the small RNA libraries we analyzed sequence tags associated with known RNA classes by mapping sequence tags onto the chicken genome assembly using BLAT (Kent 2002). We found that known chicken miRNAs accounted for ~60% of all sequence reads in all three embryonic libraries (Supplemental Table S2). This result indicates that our small RNA libraries were highly enriched in mature miRNA sequences. However, when we analyzed numbers of the unique sequence tags, we discovered that these 60% of sequence reads derived from known miRNAs represented only a relatively small fraction (4%–12%) of the total numbers of the unique sequence tags. The highest fraction of the unique sequence tags (52%–73%) was attributed to unclassified small RNAs (Supplemental Table S3). We conclude that while known miRNAs might be the most abundant class of small regulatory RNAs in the chicken

embryo, there is a less abundant but much more diverse pool of small RNAs that may represent new miRNAs and other classes of regulatory RNA.

Known miRNAs

To assess the efficiency of the deep sequencing procedure for miRNA detection we analyzed the numbers and distribution of the sequence tags associated with known chicken miRNA genes. We found that out of 121 known chicken miRNAs 118 were expressed in at least two of our three embryonic small RNA libraries, and 117 miRNAs were detected in all three RNA libraries (Supplemental Table S1). These corresponded to 144 and 145 miRNA genes respectively. These results illustrate that a rapidly developing chicken embryo employs almost the entire repertoire of the known miRNAs. The high sensitivity of the deep sequencing approach was clearly evident when we looked at the distributions and numbers of sequence tags originating from different arms of the miRNA precursors (pre-miRNAs). We identified sequence tags representing miRNA* sequences for 136 out of 145 expressed miRNA genes, a large increase over the three reported previously (Supplemental Table S9). As expected, in most of the cases total sequence read counts were heavily skewed toward the RNA hairpin arm containing the annotated miRNA. This observation is consistent with the current knowledge of miRNA biogenesis and strand selection. The dsRNA-specific endonuclease known as DICER excises a 21–23 bp small RNA duplex from a hairpin-like miRNA precursor, producing an equimolar ratio of miRNA/miRNA*. Subsequently, the strand with lower thermodynamic stability in its 5′ end is preferentially incorporated into the RISC complex and, unlike the other strand, is protected from rapid degradation (He and Hannon 2004). In many cases, miRNA* cannot be detected by conventional methods due to their rapid turnover; however, the deep sequencing procedure allows many of them to be identified.

Although the majority of miRNA genes display strand bias, some of them give rise to miRNA and miRNA* with similar 5′ end stability that leads to equal incorporation of either strand into the RISC and their protection from degradation. These miRNA genes are usually characterized by similar expression and/or cloning recovery rates of miRNA and miRNA* (He and Hannon 2004). A few such miRNA genes have been predicted and validated in different species (Aravin et al. 2003; Schwarz et al. 2003). Our analysis of the sequence tags originating from the known miRNA genes identified eight miRNA genes that appeared to encode miRNAs on both arms of the pre-miRNA (Table 1). These genes demonstrated a nearly equal number of sequence reads originat-

Table 1. miRNAs expressed from both arms of the miRNA precursor

miRNA ID	CE5		CE7		CE9	
	No. of 5p reads	No. of 3p reads	No. of 5p reads	No. of 3p reads	No. of 5p reads	No. of 3p reads
gga-mir-135a-2	**3341**	2193	**1306**	1539	**437**	1436
gga-mir-30e	**8636**	2182	**4205**	2555	**3249**	4183
gga-mir-219	**2783**	6081	**1555**	2506	**2679**	542
gga-mir-187	127	**295**	139	**226**	145	**233**
gga-mir-455	**275**	178	**568**	513	**1003**	428
gga-mir-301	1408	**351**	685	**58**	731	**27**
gga-mir-302a	311	**125**	41	**23**	4	**3**
gga-mir-30c-1	**1088**	1060	**859**	719	**308**	626

Small RNA libraries prepared from day five, day seven, and day nine chicken embryos are referred to as CE5, CE7, and CE9, respectively. Numbers indicate a total sequence read count associated with 5′ (5p) and 3′ (3p) arms of the miRNA precursor. Numbers corresponding to the mature miRNAs currently annotated in the miRBase 10.1 are shown in bold.

ing from the 5′ and 3′ arms of the miRNA hairpin precursor (5p and 3p, respectively). Moreover, six of the eight genes had a higher number of sequence reads originating from the arm of the pre-miRNA hairpin opposite to the annotated mature miRNA containing arm in at least one of the RNA libraries (Fig. 1A; Table 1). Furthermore, four of these miRNAs (miR-135a-2, miR-30e, miR-219, and miR-30c) showed a reversal in the ratios of the 5p- and 3p-derived sequence tags across the three RNA libraries (Table 1). The latter cases may indicate developmental switching in preferential use of mature miRNAs originating from different arms of the pre-miRNA precursor and suggest additional levels of complexity in miRNA processing which remain to be uncovered.

Similar to other deep sequencing studies, we observed heterogeneity at the 5′ and 3′ ends of the sequenced tags (Fig. 1A) (Ruby et al. 2006, 2007b; Landgraf et al. 2007; Morin et al. 2008). In some cases sequence read counts for these miRNA isoforms were higher than the corresponding sequence read counts for known miRNA sequences reported in the miRBase 10.1. We suggest that these most frequently sequenced miRNA isoforms should be utilized to refine miRBase annotations of chicken miRNAs (Supplemental Table S10).

In 15 cases we observed sequence tags originating from the terminal loop region of the pre-miRNA precursor (Fig. 1B; Supplemental Figs. S3, S4). For 12 of them the total numbers of small RNA sequence reads derived from the terminal hairpin loop were very small as compared to numbers of sequence reads derived from mature miRNAs. Examples like these have been reported before (Ruby et al. 2006), and may be explained by the larger volume of sequence data occasionally detecting pre-miRNA processing intermediates. However, in the three remaining cases (mir-18b, mir-451, and mir-218-1) we found that the numbers of sequence reads derived from the terminal hairpin loop were relatively high (Fig. 1B; Supplemental Figs. S3, S4). The discrete clustering of sequence tags along the miRNA hairpin precursor (Fig. 1B) suggests that the detected sequence tags are genuine products of pre-miRNA processing rather than random degradation products of the unprocessed pre-miRNA. These data indicate a complex post-transcriptional processing in these three miRNA genes.

As reflected by the total counts of miRNA-derived sequence reads, known miRNAs had a very broad range of expression which varied from hundreds of thousands of sequence reads for the most abundant miRNAs to zero for the three previously known chicken miRNAs that have not been detected (Supplemental Table S1). Similarly to the observations made by Takada and colleagues in the mouse embryo (Takada et al. 2006), we found that 10 of the most abundant known miRNAs account for approximately half of all miRNA-derived sequence tags in all three libraries (Supplemental Table S1).

Although our experiment was not designed for direct comparison of miRNA abundance between the three different developmental stages, we applied a statistical evaluation procedure enabling us to identify potentially significant changes in relative miRNA abundance between the three embryonic small RNA libraries (see Supplemental Methods). Table 2 lists the known chicken miRNAs that show statistically significant ($P < 0.01$) changes in their relative abundance between samples from embryonic stages at days five, seven, and nine. (The full list of the miRNAs and the detailed analysis is available in Supplemental Table S1.) Seven out of 32 miRNAs listed in the Table 2 appeared to be up-regulated in the order of CE5 < CE7 < CE9, while 22 miRNAs seemed to be down-regulated in the same order (CE5 > CE7 > CE9).

Newly identified miRNAs

To identify novel miRNAs in the sequencing data from the three small RNA libraries, we used the following criteria: (1) genomic loci annotated as known chicken miRNAs or as other classes of noncoding RNA were excluded; (2) to be considered for further analysis an individual locus had to be supported by at least two independent sequence reads originating from at least two small RNA libraries; (3) the loci lacking hairpin-like RNA secondary structures including the positions of the small RNA tags were eliminated. The resulting set of sequences and their respective RNA structures were analyzed further to distinguish genuine miRNA precursors from other RNAs that contain similar RNA structures (e.g., tRNA-derived repeat elements; Supplemental Methods).

The resulting data set was composed of 361 unique sequences identified as novel chicken miRNAs (Supplemental Table S7; http://www.livestockgenomics.csiro.au/courses/chicken/). In addition we identified another 88 miRNA candidates that met all of our inclusion criteria except that these were detected in only one miRNA library (Supplemental Table S13). Sequence comparisons between the new chicken miRNA candidates and other vertebrate miRNAs present in miRBase (miRBase 10.1) revealed that 13 of 361 of these new chicken miRNAs were orthologous to miRNAs identified in other vertebrate species (Supplemental File S1; Supplemental Table S5). To investigate evolutionary conservation of the remaining 348 chicken miRNAs, we searched for highly similar DNA sequences in the human, dog, opossum, zebrafish, xenopus, and lizard genome assemblies. Sequences returned by sequence similarity searches were then confirmed as orthologous miRNA candidates by analysis of their predicted RNA structures. We found six new chicken miRNA genes that were conserved in at least one of the analyzed vertebrate genomes (Supplemental File S1; Supplemental Table S5). To identify potential avian-specific miRNAs we used a similar approach to search the NCBI archive for sequence traces of the zebra finch (*Taeniopygia guttata*). Traces with high sequence identity scores were analyzed for the presence of evolutionarily conserved hairpin-like structures. We identified 22 chicken miRNA sequences that are also present in zebra finch. This number is likely to be an underestimate of the avian-specific miRNAs, as our searches were limited to zebra finch sequences available in the NCBI trace archive at the time of this study. It is not yet possible to give an accurate estimate of what portion of the zebra finch genome has been covered by the NCBI sequence entries.

When we looked at the relative abundance of the new miRNAs (as reflected by the total counts of the most frequently sequenced reads), we found that evolutionarily conserved miRNAs were often among the most abundant ones (Supplemental Tables S5, S6). This result is consistent with many other studies demonstrating some correlation between evolutionary conservation of miRNAs and their expression levels (for the most recent studies see Berezikov et al. 2006a; Landgraf et al. 2007; Ruby et al. 2007b). Interestingly several new miRNAs identified in our study that appeared to be chicken specific had expression levels comparable to some evolutionarily conserved miRNAs (Supplemental Tables S5, S6).

Mirtrons

Three recent studies have described an alternative miRNA processing pathway that uses intron splicing machinery instead of the DROSHA (RNASEN) endonuclease to generate miRNA precur-
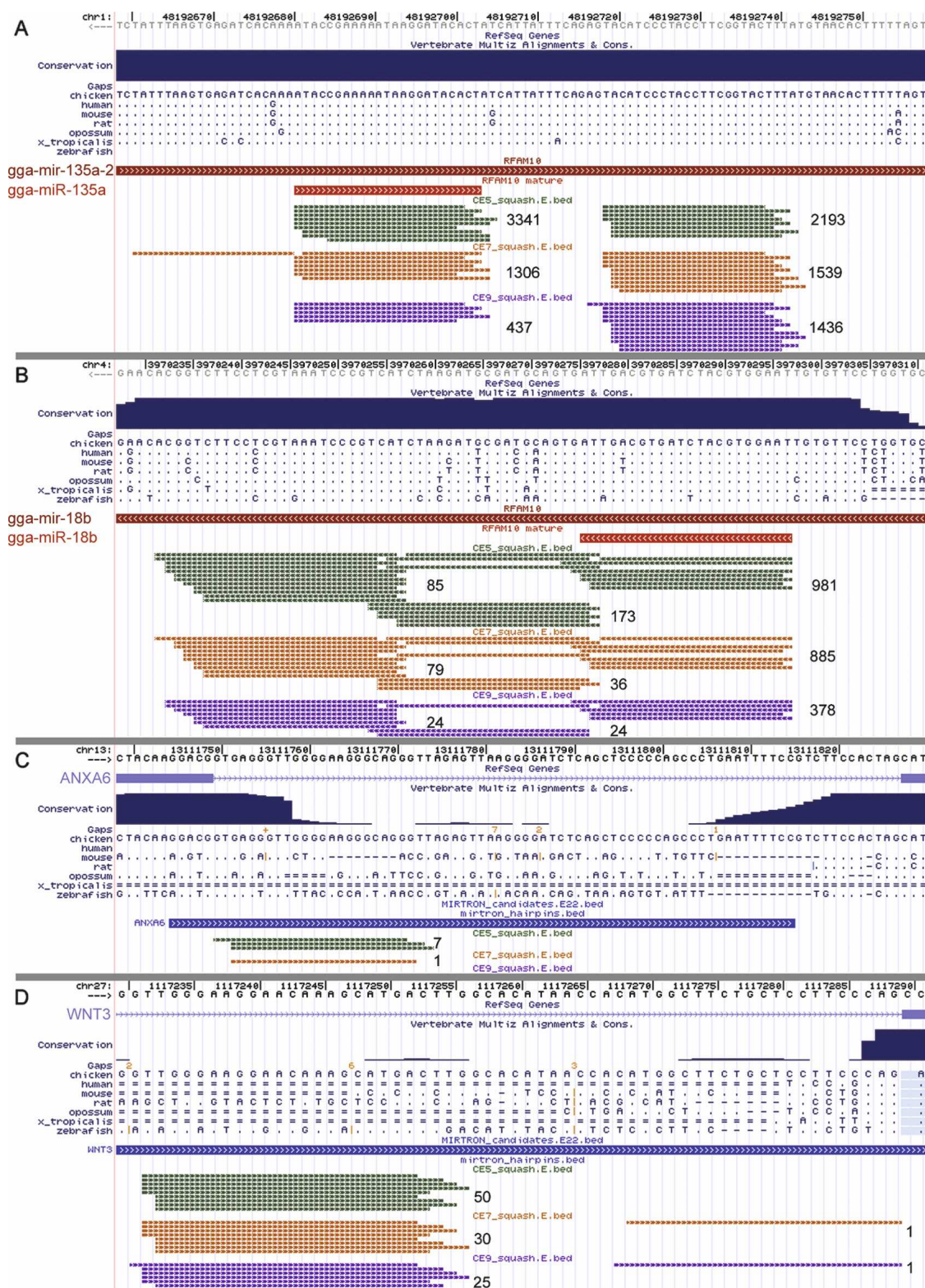
Glazov et al.

Figure 1. The figure shows UCSC genome browser screens displaying chicken miRNAs gga-mir-135a-2 (*A*), gga-mir-18b (*B*), a mirtron within the *ANXA6* gene (*C*), and an atypical mirtron within the *WNT3* gene (*D*). Sequence tags originating from these loci are shown as thin lines of three different colors representing each of the small RNA libraries: CE5, green; CE7, orange; and CE9, magenta. Numbers to the *right* of the sequence tag clusters indicate total number of reads originating from this cluster. Arrow directions indicate positive or negative strand of genomic DNA. Identical bases in the multiple sequence alignment are shown as dots.

Table 2. Changes in miRNA expression

miRNA ID	Total no. of sequence reads		
	CE5	CE7	CE9
gga-let-7b	3,460	14,843	38,845
gga-let-7d	379	1,216	1,964
gga-let-7i	7,440	13,179	26,834
gga-let-7k	1,077	4,164	5,833
gga-mir-124b	6	22	30
gga-mir-455	487	1,234	1,628
gga-mir-7b	8	70	61
gga-mir-101	47,877	35,917	13,863
gga-mir-106	11,979	4,322	2,048
gga-mir-10b	4,152	1,816	630
gga-mir-130b	15,567	8,196	4,308
gga-mir-17	12,868	4,806	2,360
gga-mir-18a	588	470	163
gga-mir-18b	1,244	1,003	378
gga-mir-19a	541	200	48
gga-mir-19b	9,098	2,595	901
gga-mir-200a	2,276	924	697
gga-mir-200b	1,844	591	537
gga-mir-20a	1,695	925	492
gga-mir-20b	7,113	2,508	1,099
gga-mir-218	1,589	1,024	236
gga-mir-302a	616	91	9
gga-mir-302b/d	413	53	10
gga-mir-302c	93	12	1
gga-mir-31	5,147	2,027	1,006
gga-mir-34a	54	22	15
gga-mir-34b/c	1,547	386	589
gga-mir-449	1,243	374	402
gga-mir-7	2,504	741	799
gga-mir-202	73	393	157
gga-mir-181a	23,484	14,882	50,376
gga-mir-181b	52,538	12,625	38,027

Known miRNAs that show statistically significant ($P < 0.01$) change in the total number of sequence reads between day five (CE5), day seven (CE7), and day nine (CE9) embryonic small RNA libraries. Spaces separate parts of the table listing miRNAs with increasing (top), decreasing (middle), and mixed (bottom) abundance. For miRNA families with multiple gene copy numbers, a single representative miRNA is shown. The analysis of all 145 known miRNAs detected in this study is available in Supplemental Table S3 and Supplemental Methods.

sors from short intronic sequences (Berezikov et al. 2007; Okamura et al. 2007; Ruby et al. 2007a). A distinct feature of such miRNA-generating introns is that the miRNA hairpin-like precursor is directly adjacent to the splice sites such that mature miRNA sequences often start directly at the 5′ terminus of the intron and/or end at its 3′ terminus. The few mirtrons identified to date originate from the diverse evolutionarily lineages of insect (*Drosophila sp.*), worms (*Caenorhabditis elegans*), and mammals (mostly represented by primates) (Berezikov et al. 2007; Okamura et al. 2007; Ruby et al. 2007a). A bioinformatics approach similar to the one described by Berezikov et al. (2007) was utilized to investigate the presence of mirtrons in the chicken genome. We found 12 mirtrons that were supported by multiple sequence reads (minimum two) in at least two embryonic libraries (Fig. 1C; Table 3; Supplemental Figs. S2a–S2l). None of these mirtrons were identified in previous studies. Having identified 12 chicken mirtrons we extended our search to larger (>200 nt) introns, looking for intronic sequence tags directly adjacent to splice sites. This search yielded another six mirtron candidates (Fig. 1D; Table 3; Supplemental Figs. S2m–S2r). Although we cannot call these candidates "typical" mirtrons, as the predicted pre-miRNA

precursor is anchored by only one of the two splice sites, these miRNAs otherwise have typical characteristics of the known mirtrons reported in literature (Berezikov et al. 2007). Sequence analysis of the 3′ and 5′ ends of the representative sequence tags uncovered sequence motifs similar to ones reported by Berezikov and colleagues for mammalian mirtrons (Supplemental Fig. S5). Surprisingly, we found more sequence reads supporting expression of the "atypical" mirtrons originating from the large introns than we did for the "typical" mirtrons derived from small introns (Table 3). Additional analyses of chicken introns identified a further 21 mirtron candidates that had some typical mirtron characteristics but which did not meet our filtering criteria and therefore are considered as mirtron candidates (Supplemental Table S8). Analysis of the evolutionary conservation of the newly identified mirtrons revealed that only one mirtron, located within the neurexin 1 gene, was evolutionarily conserved between mammals and birds. Another mirtron candidate was found within the chicken ortholog of the human DEAH-box RNA helicase gene *DHX30*, which has already been reported to contain a mirtron (Berezikov et al. 2007). However, examination of the microsynteny between chicken and human orthologs revealed that the chicken mirtron is located in a different *DHX30* intron compared to human. We did not find any small sequence reads supporting expression of the human mirtron ortholog in chicken.

Discussion

The results presented here provide experimental and bioinformatic evidence supporting the discovery of 361 new miRNAs, 88 new miRNA candidates, 18 mirtrons (including six novel atypical mirtron candidates), and 21 mirtron candidates that are expressed in the chicken embryo. Together with the existing set of 121 known chicken miRNAs, this brings the total number of miRNAs in the chicken to 609. Analysis of the evolutionary conservation of the newly identified putative miRNAs revealed that only six of them are conserved in non-avian vertebrates, with the majority of the remaining miRNAs likely to be specific to the bird and/or chicken lineages. The small number of the newly identified miRNAs conserved in vertebrates is not surprising, given the results of other recent studies. Berezikov et al. (2006a) identified 447 new miRNA genes, most of which appeared to be specific to primates; with only 11 of 447 miRNA genes identified in other groups of vertebrates. It appears that while most of the miRNAs discovered in early studies were found to be highly conserved in evolution, more and more of the newly identified miRNAs are present in only a small group of organisms and in some cases in a single species (Bentwich et al. 2005; Berezikov et al. 2006a,b; Ruby et al. 2007b). This phenomenon can be explained in part by the frequently observed correlation between the level of evolutionary conservation of a given miRNA and its expression level, and hence a technical capacity for it to be detected. It is noteworthy that the oldest known miRNA, let-7 (Pasquinelli et al. 2000), and its family members, were sequenced more than one million times in our three embryonic libraries (Supplemental Table S1). Historically, this correlation between evolutionary conservation and expression level led to a widespread notion that all miRNAs are highly conserved. However, this notion was formulated before the abundant lineage-specific miRNAs were identified and hence might be misleading. Although, the functional significance of these evolutionarily divergent miRNAs has not been established experimentally, it has been hypothesized

Table 3. List of chicken mirtrons

Gene name	RefSeq ID	No. of reads	Position of reads	Splice site	Genomic coordinates	Strand
ANXA6	NM_204730	8	5p	D	chr13:13111744–13111815	+
CCNA2	NM_205244	7	3p	A	chr4:55473432–55473504	+
WBSCR27	BG625158*	4	3p	A	chr19:662469–662546	−
UBE2A	NM_204865	4	3p	A	chr4:16683254–16683323	−
SCLY	CR406321*	4	3p	A	chr9:1877568–1877643	−
PSMD3	NM_001031362	4	3p	A	chrUn_random:29463080–29463151	−
ERBB2	NM_001044661	3	5p	D	chrUn_random:29524889–29524961	−
TNRC5	CR524291*	3	3p	A	chr3:16952470–16952545	+
RRP12	NM_001012908	3	3p	A	chr6:23780662–23780736	+
NOL5A	NM_001031388	2	5p	D	chrUn_random:42527297–42527387	+
FURIN	NM_204715	2	3p	A	chr10:22271057–22271117	−
PARC	NM_015089*	2	3p	A	chr3:4279292–4279380	−
WNT3	NM_001081696	107	5p, 3p	A	chr27:1117229–1117291	+
ADAMTS10	NM_030957*	91	5p, 3p	A	chr28:1360196–1360279	−
TMEM63B	BU230031*	20	5p, 3p	A	chr3:31574849–31574928	−
SCMH1	NM_001031694	16	5p, 3p	A	chr23:1213445–1213505	+
CRSP3	DT655869*	14	5p, 3p	A	chr3:59307909–59308000	+
DHX30	NM_001012851	8	3p	A	chr2:625612–625684	−

The space separates typical (upper part of the table) and atypical (lower part of the table) mirtrons. The number of reads indicates the total number of sequence reads originating from all three embryonic libraries, and derived from either arm of the hairpin-like precursor. The splice site column indicates whether sequence reads are adjacent to donor (D) or acceptor (A) splice sites. Positions of the sequence reads within the pre-miRNA precursor are indicated as 5p and 3p arms, respectively. Chicken RefSeq gene IDs are provided if available; in other cases, indicated by an asterisk, refSeq gene IDs of the closest ortholog or GenBank IDs of chicken mRNA or EST are given. Genomic coordinates refer to the UCSC galGal3 genome assembly.

that these miRNAs might play a role in establishing and maintaining phenotypic diversity between different groups of organisms (Plasterk 2006; Sempere et al. 2006). It is plausible that the conserved miRNAs are responsible for control of the basic cellular and developmental pathways common to most eukaryotes (e.g., cell cycle) whereas the nonconserved miRNAs are involved in regulation of the lineage-specific pathways and functions.

As compared to the numbers of primate-specific miRNAs (Berezikov et al. 2006a), the number of potential avian-specific miRNAs identified in our study appears relatively low; only 22 of the newly identified miRNAs were found to have homologs in the available genomic sequence of zebra finch. However, as suggested earlier, this number is likely to be an underestimate, and poorly reflects the true extent of the evolutionary conservation of miRNAs among birds. There are at least two reasons to suggest this. Firstly, the availability of the zebra finch genome sequences is currently limited, and the full analysis of the miRNA conservation can only be performed after the whole-genome sequencing and assembly are complete for the zebra finch. Secondly, the estimated divergence time between the two bird genomes analyzed in this study is much higher than that of human and chimp genomes: ~105 million years in birds as compared to five million years in primates. In fact the estimated divergence time between the two bird lineages is comparable to the divergence of the human and dog lineages (Supplemental Fig. S1). Therefore, it is conceivable that many new species-specific miRNAs have evolved since the divergence of the two bird lineages.

The large number of sequence tags originating from known chicken miRNAs in all three embryonic RNA libraries enabled us to analyze changes in miRNA expression and observe less common events in the miRNA biogenesis that have not been previously reported. Data on sequence reads originating from the terminal loop of the miRNA precursor, exemplified by chicken mir-18b (Fig. 1B), and apparent shifts in strand preference in some miRNA genes (Table 1) highlight the complexity of the post-transcriptional regulation of miRNA processing. Evidence for the regulated and/or alternative processing of pre-miRNAs is just be-

ginning to emerge. The processing of a human miRNA from the mir-18 family (hsa-mir-18a) was recently demonstrated to be dependent on the splicing repressor protein HNRNPA1 (Guil and Caceres 2007). In addition, a new sub-class of intronic miRNAs named mirtrons was reported to bypass the DROSHA-dependent step of miRNA processing by utilizing the splicing machinery to excise RNA hairpin precursors (Okamura et al. 2007; Ruby et al. 2007a). In this study we have reported 39 new chicken mirtrons including 21 mirtron candidates. The identification of six atypical mirtrons provides further evidence for the complex relationship between the splicing machinery and pre-miRNA processing. Analysis of the orthologous sequences for newly identified mirtrons revealed that none of these have been reported in other vertebrates. These results may indicate either rapid divergence and/or independent evolution of mirtrons in different evolutionary lineages.

Given that our small RNA libraries were prepared from whole-embryo extracts at relatively advanced stages of embryonic development, it is difficult to discuss observed changes in abundance of different miRNAs in the context of specific biological processes. However, certain trends appear to be clear. For example, an apparent decrease in the abundance of 22 miRNAs presented in Table 2 and potentially of some others (Supplemental Table S1) is likely to reflect progressive restriction of the expression of these miRNAs from broad embryonic domains such as endoderm, mesoderm, and ectoderm to specific regions, organs, and potentially cell types, as embryo development progresses to maturity. A similar trend was observed in zebrafish and chicken embryonic development using in situ hybridization and miRNA microarray techniques (Wienholds et al. 2005; Darnell et al. 2006).

Although the examples described above by no means represent all of the molecular events in miRNA biogenesis and expression, they illustrate the high value of the deep sequencing data for qualitative and potentially quantitative studies of small regulatory RNAs. The substantial proportion of unclassified small RNAs identified in this study suggests that there may be other

classes of small regulatory RNA in chicken embryos that were not covered by our analyses. Some of these may also include rare miRNAs that were represented by single sequence reads and thus did not pass our filtering criteria. These may be confirmed in future studies applying the deep sequencing approach to specialized tissues or cells.

Methods

Sources of sequences and assemblies

Draft genome assembly of the green lizard (*Anolis carolinensis*) was produced by the Broad Institute at MIT and Harvard (http://www.broad.mit.edu/). Draft genome assembly of the frog (*Xenopus tropicalis*) was produced by the DOE Joint Genome Institute (JGI) (http://genome.jgi-psf.org/Xentr4/Xentr4.home.html/). Draft genome assembly of the zebrafish (*Danio rerio*) was produced by The Wellcome Trust Sanger Institute in collaboration with the Max Planck Institute for Developmental Biology in Tuebingen, and the Netherlands Institute for Developmental Biology (http://www.sanger.ac.uk/). Zebra finch (*Taeniopygia guttata*) sequence data were produced by the Genome Sequencing Center at Washington University School of Medicine in St. Louis (http://genome.wustl.edu/genome_group_index.cgi). The NCBI Trace Archive accession nos. for zebra finch (*Taeniopygia guttata*) sequences used in this study are provided in supplemental data. The NCBI Taxonomy ID numbers for chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*) are 9031 and 59729, respectively. Human, chicken, and opossum genomes were produced by their respective genome sequencing consortiums (Lander et al. 2001; Hillier et al. 2004; Mikkelsen et al. 2007). Unless specified otherwise, sequences of miRNA precursors and mature miRNAs were obtained from the latest release of miRNA database (miRBase 10.1, December 2007, http://microrna.sanger.ac.uk/registry/) (Griffiths-Jones et al. 2006).

Chicken embryo collection and RNA isolation

Fertilized Ross-308 eggs were obtained from Barter Enterprises. The eggs were incubated in a Multiquip incubator (Multiquip) at 37.5°C with rotation every 6 h. Chick embryos were collected at days five, seven, and nine of incubation and selected to represent the embryonic developmental stages 25–27, 30–31, and 35, respectively (Hamburger and Hamilton 1992). After removal of the amnion, embryos were rinsed in $1\times$ PBS, and immediately processed for RNA isolation.

Low molecular weight RNA was extracted using *mir*Vana miRNA Isolation Kit (Ambion) following the manufacturer's protocol. In brief, embryos were dissected and homogenized in the lysis buffer supplied with the kit. The RNA concentration and purity were determined photometrically by measuring absorbance at 260 nm and A260/A280 ratio using the NanoDrop ND-1000 spectrophotometer (Nanodrop Technologies). RNA samples were stored at –80°C until further use.

Small RNA library construction and sequencing

For small RNA library construction and deep sequencing, RNA samples were prepared as follows: for each developmental stage equal quantities (7 μg) of small RNA isolated from three individual embryos were pooled. Approximately 20 μg of small RNA representing each developmental stage were submitted to Solexa (now Illumina Inc.) for sequencing.

In brief, the sequencing was performed as follows: RNA was purified by polyacrylamide gel electrophoresis (PAGE), to enrich for molecules in the range of 18–30 nt, and ligated with propri-

etary adapters to the 5′ and 3′ termini of the RNA. The samples were used as templates for cDNA synthesis. The cDNA was amplified with 18 PCR cycles to produce sequencing libraries that were subjected to Solexa's proprietary sequencing-by-synthesis method.

Analysis of sequencing data

Individual sequence reads with the base quality scores were produced by Illumina/Solexa. All identical sequences were counted and eliminated from the initial data set. The resulting set of the unique sequences with associated read counts is referred to as sequence tags. A mirror of the UCSC genome browser and database was created with the *Gallus gallus* v.2.1 genome sequence and annotations (galGal3, May 2006) (Kent et al. 2002; Karolchik et al. 2003). After trimming the 3′ adaptor sequence, sequence tags were mapped onto the chicken genome assembly using BLAT software (Kent 2002). To identify sequence tags originating from coding exons, repeats, rRNA, tRNA, snRNA, and snoRNA, we used UCSC RefGene, RepeatMasker, and NCBI RefSeq data (Pruitt and Maglott 2001; Karolchik et al. 2003; Kuhn et al. 2007), as well as our own sets of ncRNA annotations compiled from the NCBI GenBank data (http://www.ncbi.nlm.nih.gov/). To identify novel miRNA genes we identified all hairpin-like RNA structures encompassing small RNA sequence tags using RNAfold (Hofacker 2003); then we analyzed sequence and structural features of the predicted hairpin-like structures to distinguish genuine miRNA precursors from other RNA classes that may contain similar RNA structures (e.g., tRNA-derived repeat elements; see Supplemental Methods for details). All sequences identified as new miRNA precursors and the most frequently sequenced tags for each mature miRNA were submitted to the microRNA database (miRBase at the Sanger center http://microrna.sanger.ac.uk/registry/).

Acknowledgments

References

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5:** 337–350.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* **37:** 766–770.

Berezikov, E., Thuemmler, F., van Laake, L.W., Kondova, I., Bontrop, R., Cuppen, E., and Plasterk, R.H.A. 2006a. Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.* **38:** 1375–1377.

Berezikov, E., van Tetering, G., Verheul, M., van de Belt, J., van Laake, L., Vos, J., Verloop, R., van de Wetering, M., Guryev, V., Takada, S., et al. 2006b. Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res.* **16:** 1289–1298.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. 2007. Mammalian mirtron genes. *Mol. Cell* **28:** 328–336.

Brown, W.R., Hubbard, S.J., Tickle, C., and Wilson, S.A. 2003. The chicken as a model for large-scale analysis of vertebrate gene function. *Nat. Rev. Genet.* **4:** 87–98.

Chapman, E.J. and Carrington, J.C. 2007. Specialization and evolution

of endogenous small RNA pathways. *Nat. Rev. Genet.* **8:** 884–896.

Darnell, D.K., Kaur, S., Stanislaw, S., Konieczka, J.H., Yatskievych, T.A., and Antin, P.B. 2006. MicroRNA expression during chick embryo development. *Dev. Dyn.* **235:** 3156–3165.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. 2006. miRBase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34:** D140–D144.

Guil, S. and Caceres, J.F. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.* **14:** 591–596.

Hamburger, V. and Hamilton, H.L. 1992. A series of normal stages in the development of the chick embryo. 1951. *Dev. Dyn.* **195:** 231–272.

He, L. and Hannon, G.J. 2004. MicroRNAs: Small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5:** 522–531.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A.M., Delany, M.E., et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695–716.

Hofacker, I.L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31:** 3429–3431.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., et al. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316:** 1484–1488.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res.* **31:** 51–54.

Kent, W.J. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* **12:** 656–664.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. 2002. The human genome browser at UCSC. *Genome Res.* **12:** 996–1006.

Kuhn, R.M., Karolchik, D., Zweig, A.S., Trumbower, H., Thomas, D.J., Thakkapallayil, A., Sugnet, C.W., Stanke, M., Smith, K.E., Siepel, A., et al. 2007. The UCSC genome browser database: Update 2007. *Nucleic Acids Res.* **35:** D668–D673.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129:** 1401–1414.

Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L.,

Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447:** 167–177.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., et al. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18:** 610–621.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130:** 89–100.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408:** 86–89.

Plasterk, R.H.A. 2006. Micro RNAs in animal development. *Cell* **124:** 877–881.

Pruitt, K.D. and Maglott, D.R. 2001. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29:** 137–140.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127:** 1193–1207.

Ruby, J.G., Jan, C.H., and Bartel, D.P. 2007a. Intronic microRNA precursors that bypass *Drosha* processing. *Nature* **448:** 83–86.

Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., and Lai, E.C. 2007b. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* **17:** 1850–1864.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115:** 199–208.

Sempere, L.F., Cole, C.N., McPeek, M.A., and Peterson, K.J. 2006. The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. *J. Exp. Zoolog. Part B. Mol. Dev. Evol.* **306B:** 575–588.

Takada, S., Berezikov, E., Yamashita, Y., Lagos-Quintana, M., Kloosterman, W.P., Enomoto, M., Hatanaka, H., Fujiwara, S., Watanabe, H., Soda, M., et al. 2006. Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Res.* **34:** e115. doi: 10.1093/nar/gkl653.

Wienholds, E., Kloosterman, W.P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H.R., Kauppinen, S., and Plasterk, R.H.A. 2005. MicroRNA expression in zebrafish embryonic development. *Science* **309:** 310–311.

A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach

Evgeny A. Glazov, Pauline A. Cottee, Wesley C. Barris, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2008/05/12/gr.074740.107.DC1.html
References	This article cites 34 articles, 13 of which can be accessed free at: http://genome.cshlp.org/content/18/6/957.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions