Chapter 2

The Zebrafish HoxDb cluster has been reduced to a single microRNA

Joost M. Woltering & Antony J. Durston

_A modified version of this work has been published in Nature Genetics, volume 38, 6, 601-602 (2006)_

Abstract

We present the discovery of a new Zebrafish Hox cluster that has lost all functional sequences outside of the _miR-10d_ microRNA, hereby completing the number of postulated Teleost Hox clusters. In the process of loosing its genes, the size of the HoxDb cluster has shrunk an approximate 30 times to 8.1kb. Characterization of the expression of the _miR-10d_ microRNA with locked nucleic acid (LNA) and precursor probes shows an expression pattern similar to that of other _miR-10_ microRNAs, suggesting that the degenerated cluster has retained its original mode of regulation. We speculate about the reasons for the conservation of multiple, apparently redundant, genomic copies of microRNAs in the context of either presence or absence of gene dosage effects of both Hox genes and microRNAs.
Chapter 2

The metazoan Hox clusters consist of clustered homeodomain transcription factors involved in patterning of the body axes and were formed by tandem duplications (1). Non-vertebrates possess one cluster containing a maximum of 14 genes. After the major genome duplications in the vertebrate lineage the clusters have stayed intact and the Tetrapod land vertebrates possess four Hox clusters (A, B, C and D). Between each of today’s clusters there is an unequal number of genes which results from increased freedom to mutate after duplication. This has also allowed them to diverge and to adopt novel developmental roles (2). Fish (Teleosts) underwent an additional, more recent genome duplication and thus possess a theoretical number of 8 Hox clusters (named Aa, Ab, Ba, Bb, etc.) (3). After the fish specific duplication, a similar relaxation-mutation process took place during which whole clusters are believed to have disappeared; all diploid Teleosts in which the Hox clusters have been described seem to possess only 7 instead of 8 clusters (3, 4). The Medaka (Oryzias latipes) and the two Pufferfish species (Takifugu rubripes, Tetraodon nigroviridis) have only one HoxC cluster homologue and in the Zebrafish only one HoxD cluster has been described. In addition to the protein coding genes in the Hox clusters, the miR-10 and miR-196 families of Hox specific microRNAs have been described (5, 6). In Zebrafish, miR-10 copies are present in the Hox Ba, Bb, Ca and Da cluster and one copy located in a genomic position that seems at first sight Hox unrelated (fig.1a). This latter isoform has the same sequence as present in the Takifugu and Tetraodon HoxDb cluster. Annotation in Ensembl shows that it is positioned in an 8.1 kb ‘empty’ region flanked 3’ by mtx2 and 5’ by a Zebrafish lunapark homologue (Q6PFM4_BRARE) and ATP5G3 (fig.1a). This genomic location corresponds to the syntenic region of the HoxD cluster as it is conserved among vertebrates (fig.1c). Further extensive analysis of the region in between lunapark and mtx2 with Blast, translated Blast and Lagan didn’t reveal any additional homologues sequences shared with Tetrapod or Teleosts Hox clusters outside the ~100 nucleotide microRNA precursor sequence (fig.1b). From the genomic homology it is clear that the position of the miR-10 gene corresponds to the degenerated Zebrafish HoxDb, cluster but analysis shows that apparently all Hox related sequences were lost, with the exception of the miR-10 microRNA. This microRNA isoform was previously cloned from Zebrafish (7) and is listed as dre-miR-10d (MI0001889 and MI0001890) in miRBase (8). RT-PCR and sequencing further confirm embryonic expression of the microRNA precursor (data not shown).
The Zebrafish HoxDb cluster has been reduced to a single microRNA.

Figure 1) Genomics and expression of miR-10d and the HoxDb cluster. (a) Location of the five miR-10 copies in the Zebrafish Hox clusters (3’ direction is to the left). (b) Lagan Vista plot of the Zebrafish HoxDb cluster with the Human HoxD cluster, the peak corresponds to the miR-10d precursor sequence. Below: peak sequence alignment, the mature miR-10d microRNA sequence is marked blue. (c) Genomic structure of the human (Hsa) HoxD and Zebrafish (Dre) HoxDb cluster and surrounding region. (d) Double in situ hybridization on 48 hf Zebrafish embryos. Red: Engrailed-2 staining at the midbrain-hindbrain boundary. Purple: upper panel, miR-10c; middle panel, miR-10d; lower panel, miR-10d 1MM (one mismatch control, no staining).
In situ hybridization with microRNA specific LNA probes able to distinguish between the different miR-10 isoforms (9) show a similar Hox like expression pattern in the posterior spinal cord for miR-10d as for the other miR-10 members (fig.1d). Probes corresponding to the precursor sequences confirm this (fig.2). This suggests that the Zebrafish miR-10d has kept its original function within this domain of expression. The region between lunapark and mtx2 containing the HoxD cluster has a size of 267, 268 × 118 and ~80 kb in humans, mouse, Zebrafish HoxDa (incomplete contig) and Tetraodon HoxDa respectively. The mouse HoxD cluster is under control of conserved global cis-regulatory elements located in positions between lunapark and ATP5G3 (10). Interestingly, this region has a respective size of 741, 667 and ~80 kb in human, mouse and Tetraodon HoxDa cluster, but comprises only 3 kb in the Zebrafish HoxDb cluster, which strongly suggests that the regulatory elements have mutated and disappeared together with the cluster itself. The discovery of the HoxDb cluster makes Zebrafish the first Teleosts in which the fate of all the original 8 Hox clusters is known and brings the number of described clusters in line with the model for a complete genome duplication in the fish lineages. The most likely scenario for the loss of most of the Zebrafish HoxDb cluster and its control regions is that these have been lost gradually, probably after inactivation by mutation; the miR-10d gene and the surrounding genes are still intact, excluding a one step disappearance process (e.g. by excision). Also the presence of pseudo-genes in other fish Hox clusters and an ‘intermediately degenerated’ HoxDb cluster in the two Pufferfish supports this view. The graded character of the process may indicate that, in the absence of selective pressure, small deletions form a major force in shaping genomes and that there is a strong tendency for compaction by removal of non-essential sequences. Hox clusters are believed to experience high levels of purifying selection and are from an evolutionary point of view very robust structures. The fact that, in absence of selective pressure, a cluster can virtually disappear, suggests that the many gene deserts in the genomes for which no function is known are actually maintained by positive selection instead of being simply mutated non functional regions. Further, microRNA’s are possibly the most conserved metazoan genes, their functions during life and development are still poorly understood though. The paralogue 9/10 associated microRNA miR-196 has been shown to target 7 and 8 paralogue group Hox genes (11, 12) but results of knockouts or mutants have to be awaited to determine its requirement during development. Not much of the function of miR-10 is
known, but it also seems to target more anterior Hox genes. In the case of the original HoxDb cluster at least many ten thousands of bases have disappeared. The fact that the only functional element remaining is a microRNA that is already redundantly present in 4 other copies in the genome (in the Zebrafish HoxBa, Bb, Ca and Da clusters) raises questions around the selective pressures maintaining these elements throughout evolution.

The evolutionary selection for a single copy in the HoxDb cluster without the accompanying Hox genes may indicate a quantitative mode of miR-10 microRNA regulation on Hox genes wherein a certain ratio is favored. In this context it is also worth highlighting that the preservation of a miR-10b copy in the Zebrafish HoxCa clusters correlates with the presence of Hox genes anterior to HoxC4a, while both the miR-10 and these anterior Hox genes are absent in the Tetrapod HoxC clusters.

On the other hand, an explanation for the persistence of high microRNA genomic copy numbers in the context of the genome duplications could be that, in contrast to many protein coding genes, microRNAs may not be functioning in a quantitative manner. It has been suggested that after a tetraploidization event there is a strong selection against genes with high dosage effects and that this is also the case for Hox genes (13). A pure negative selection (i.e. in the case were the genes would have an unambiguous deleterious effect) against the genes in the HoxDb cluster seems, however, unlikely for the following reasons: the disappearance of the HoxDb cluster likely was a graded event and the individuals that gave rise to the Zebrafish lineage still possessed coding parts of the HoxDb cluster. These individuals were at least healthy and fertile enough to produce progeny. These same fish also are the ancestors of other fish lineages like Pufferfish and Medaka, which nowadays still possess Hox coding genes in the HoxDb cluster. A deleterious effect caused by the disappeared HoxDb genes per se is therefore difficult to envisage. Assuming that, as proposed, Hox genes indeed have dosage effects, the following explanation for the disappearance of the HoxDb genes and the persistence of the miR-10d microRNA may be thought of. Selectively loosing specific Hox genes may cause changes of a more subtle nature and offer species a mechanism for adaptation and speciation. In this scenario, selective loss of a normal Hox gene could be an advantageous event allowing for instance niche differentiation. On the other hand, loss of a single microRNA copy would be a neutral event that is under neither positive nor negative selection and would therefore be less likely to get fixed during evolution.
In situ hybridization on 24hpf embryos for \textit{miR-10b-1}, \textit{miR-10b-2}, \textit{miR-10c} and \textit{miR-10d} using LNA and precursor probes. In all cases similar staining is observed for LNA and precursor probe. Since \textit{miR-10b-1} and \textit{miR-10b-2} have identical mature sequences they are detected by the same LNA probe.

\textbf{Material and methods}

Alignments were performed using BLAST at the NCBI and Lagan (14). In situ hybridizations were performed according to standard procedures and Wienholds et al. (15). LNA probes were obtained from Exiqon (Denmark), hybridization temperature used was 58°C.

The 4 \textit{miR-10} isoforms differ from each other at 1-3 positions (see table 1 below). The closest isoform (and so the most probable to give cross reactivity) to \textit{miR-10d} is \textit{miR-10b} which differs at only one position by a T to C substitution. The control probe was designed to contain a substitution at this position that should recognize a G, thereby creating a probe against a none existing \textit{miR-10} isoform differing at the same position from \textit{miR-10b} as \textit{miR-10d} does.

Precursor in situ hybridization experiments were done at 55°C. Precursor probes were synthesized from pGEM-TE vectors containing the ~100 nt pre-miRNA sequences of \textit{miR-10b-1}, \textit{miR-10b-2}, \textit{miR-10c} and a ~1000bp PCR fragment containing the \textit{pri-miR-10d} RNA using T7 or Sp6 polymerase. The \textit{miR-10d} probe was hydrolyzed to obtain an average
The Zebrafish HoxDb cluster has been reduced to a single microRNA precursor in situs in a vertebrate.

Zebrafish HoxDb is located in gbank accession: BX546447 and in Ensembl Zebrafish v36 dec. 2005: chromosome 6, 6193163-6201262.

miR-10 isoform nomenclature used is according to miRBase. In miRBase two accession numbers are present for miR-10d which apparently both correspond to the same microRNA and result from erroneous assembly of part of the Zebrafish genome.

The sequences of Zebrafish miR-10b-1, Tetraodon nigroviridis miR-10d and Takifugu rubiceps miR-10d have been deposited in miRBase under the following accession numbers:
dre-miR-10b-1: MI0001364
tni-miR-10d: MI0004966
tru-miR-10d: MI0004967

<table>
<thead>
<tr>
<th></th>
<th>miR-10a</th>
<th>miR-10b</th>
<th>miR-10c</th>
<th>miR-10d</th>
<th>control probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-10a</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>miR-10b</td>
<td>1</td>
<td>-</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>miR-10c</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>miR-10d</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>control probe</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1 nucleotide substitutions between miR-10 isoforms

Acknowledgements: We would like to thank dr. Dana Zivkovic for the Engrailed-2 construct, Hans Jansen for help with the Pufferfish genomic sequence and dr. Axel Meyer for useful discussion.
Chapter 2

References

5 Lagos-Quintana M. RNA 9, 175 (2003)
6 Lim, L. P. Science 299, 1540 (2003)
3 27-29 (2006)
11 Yekta S., Shih I.H., Bartel D.P. Science 304, 594-596
Evol) 308B (2007)