Final Report for Chris Amemiya, Richard Myers and Frank Ruddle, "Genomic identification and analysis of shared cis-regulatory elements in a developmentally critical homeobox cluster"

This was a collaborative project between the laboratories of Richard M. Myers (Stanford University School of Medicine, Stanford, CA), Francis H. Ruddle (Yale University, New Haven, CT), and Chris T. Amemiya (Benaroya Research Institute at Virginia Mason, Seattle, WA). Three separate awards were made to the collaborators. We have presented this work (poster presentations) at the DOE Genome Contractors meeting in Oakland (1/02) and at the Cold Spring Harbor meeting on Evolution of Developmental Diversity (4/02); a manuscript describing preliminary functional characterization has been published (Proc. Natl. Acad. Sci. 100: 4030-4034, 2003; see Appendix V). A manuscript describing the major comparative genomics findings of the study is currently being prepared.

The goals of our project were to isolate, characterize, and sequence the Dlx3/Dlx7 bigene cluster from twelve different species of mammals. The Dlx3 and Dlx7 genes are known to encode homeobox transcription factors involved in patterning of structures in the vertebrate jaw as well as vertebrate limbs. Genomic sequences from the respective taxa will subsequently be compared in order to identify conserved non-coding sequences that are potential cis-regulatory elements. Based on the comparisons we will fashion transgenic mouse experiments to functionally test the strength of the potential cis-regulatory elements. A goal of the project is to attempt to identify those elements that may function in coordinately regulating both Dlx3 and Dlx7 functions.

Specific Aims:

1. Using BAC (bacterial artificial chromosome) technology, we will isolate and characterize specific *Dlx* gene clusters (*Dlx7/Dlx3*) in 12 selected mammals.
2. We will sequence the respective *Dlx7/Dlx3* clusters from the 12 species (six 150-kb BACs per year), and compare and contrast their organization between and among all known *Dlx7/Dlx3* clusters in order to identify putative *cis*-regulatory elements (conserved noncoding motifs).
3. We will functionally analyze putative *cis*-regulatory elements by utilizing mouse transgenic and knockout/knock-in technology and implementing both molecular and morphological analyses. We seek to identify and characterize those elements that may be utilized for coordinate expression of both *Dlx7* and *Dlx3* genes, and make inferences with regard to their developmental, medical, and evolutionary implications.

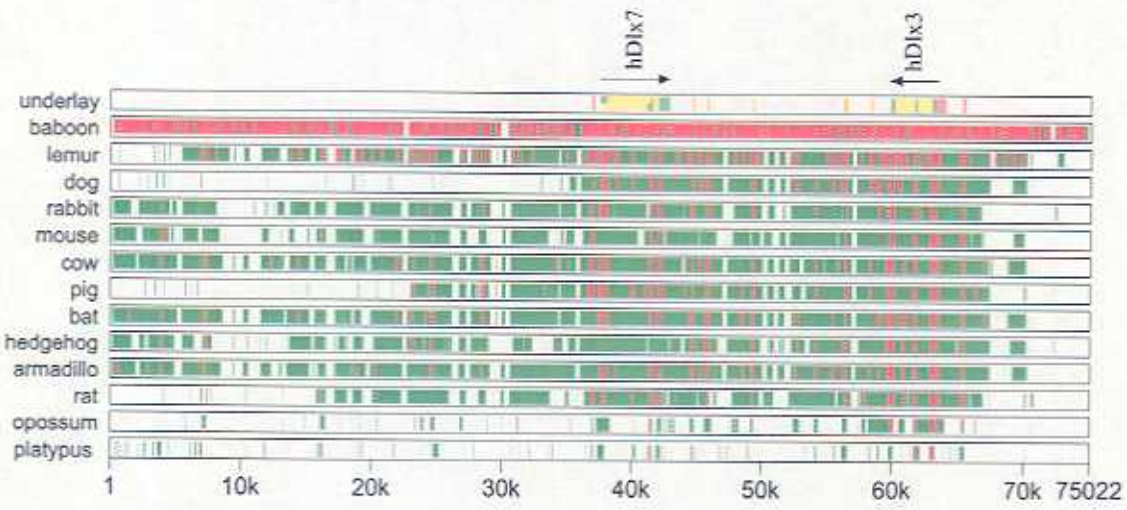
BAC clones encompassing the *Dlx7-Dlx3* bigene cluster were isolated from BAC libraries of 12 different mammalian taxa: baboon, lemur, rabbit, mouse, cow, pig, bat, hedgehog, armadillo, rat, opossum, and platypus. These species represent a broad spectrum of mammalian taxa. BAC clones were characterized for their *Dlx* gene content by restriction mapping and PCR and were subsequently sequenced at the Stanford Human Genome Center. The sequences have been deposited in GenBank and are being analyzed for publication.

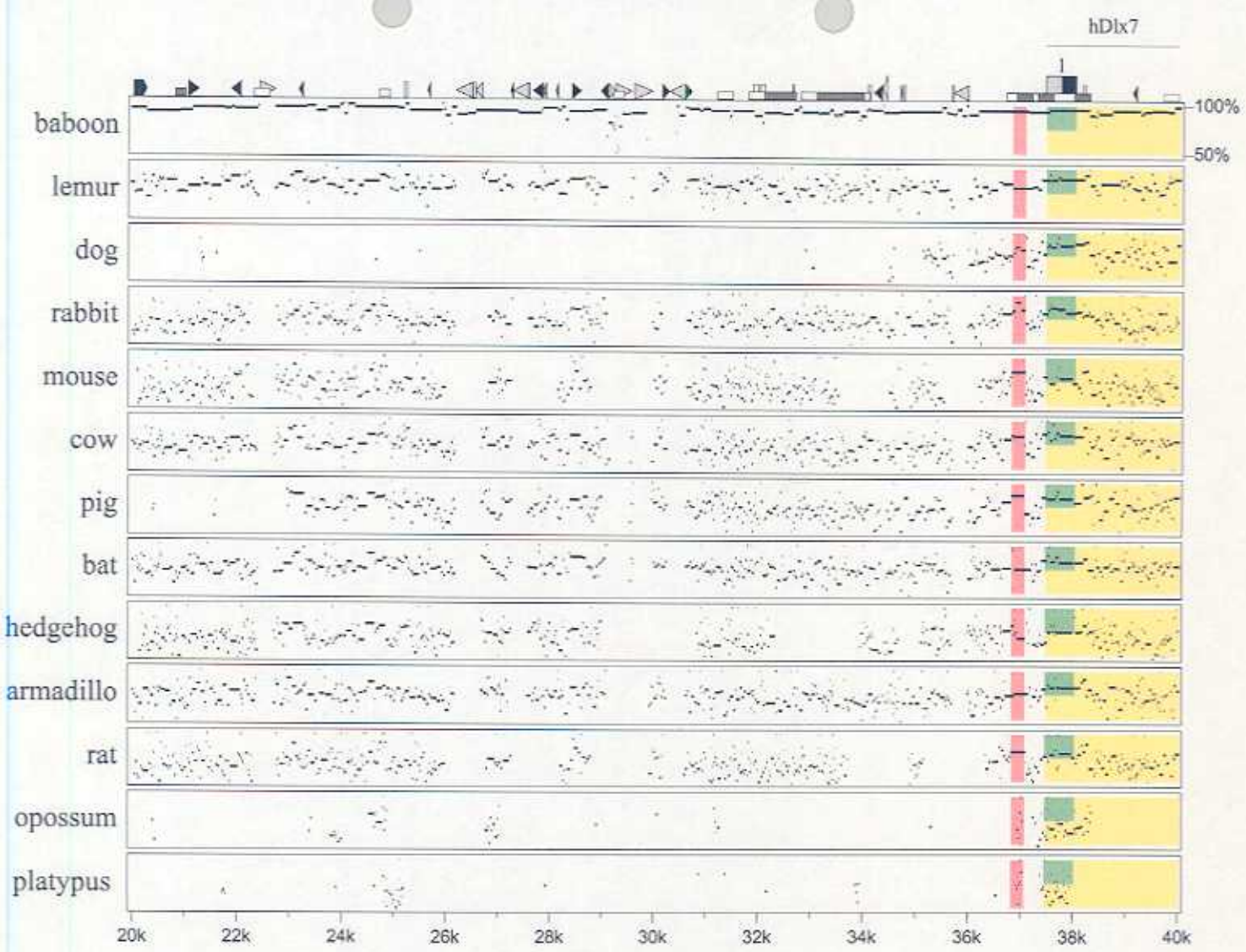
Global alignments have been performed on the cluster sequences from all taxa, including mouse and human. Visualization of the alignments was done using Multi-PIPmaker and is provided in Appendix I. The coding sequences of the *Dlx* genes allowed very good global alignments and showed the presence of several conserved blocks in the intergenic region between the *Dlx7* and *Dlx3* genes. Those blocks whose conservation levels were highest have been labeled I37-1 to I37-5 (see Appendix III).

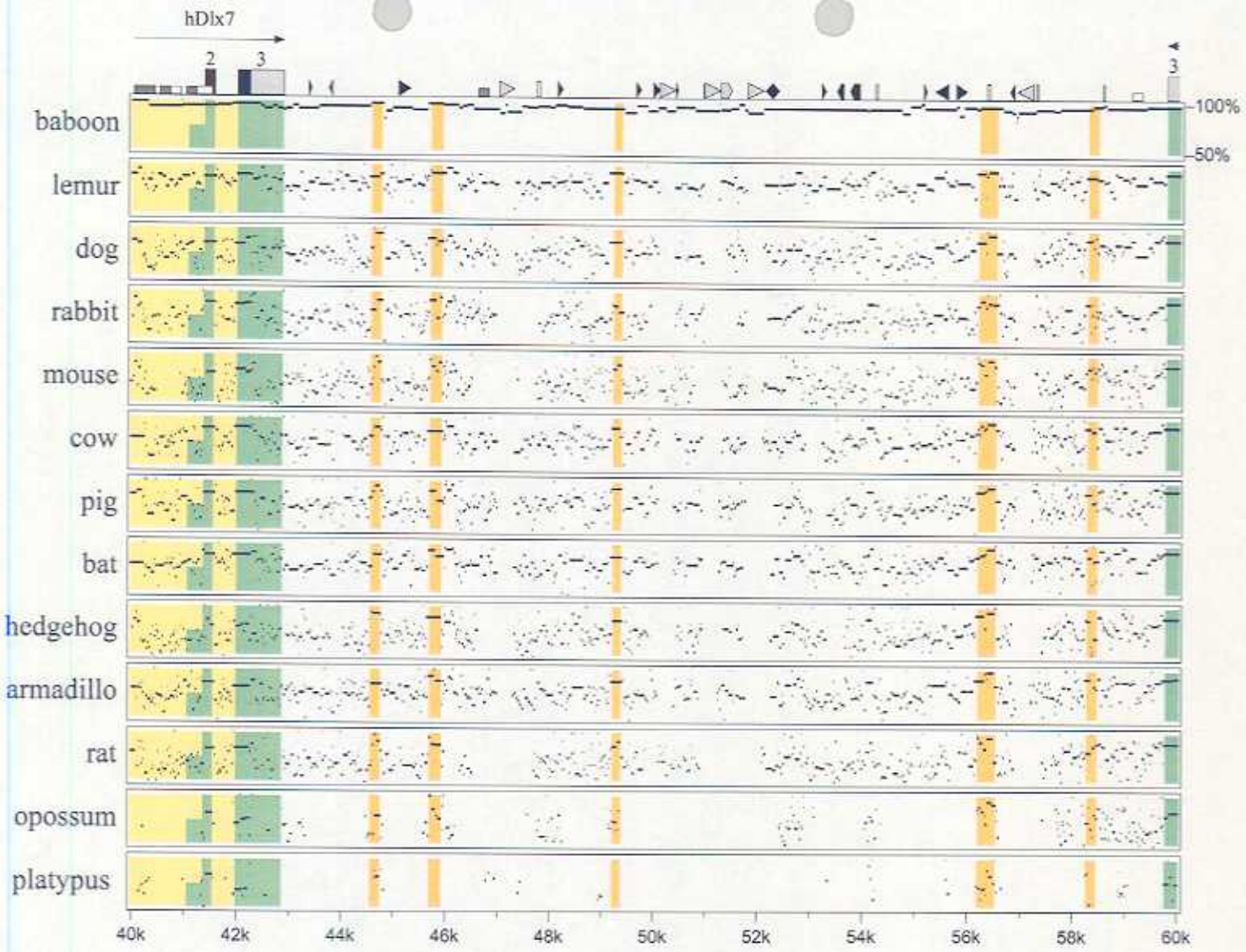
Studies using a mouse transgenic BAC-reporter system were initiated. We have analysed the functional properties of the cis-element candidates by using transgenic mouse technology. We have cloned the large genomic sequence of 80 Kb mouse Dlx7-3 cluster into pClasper, a yeast-bacterial shuttle vector. We inserted a LacZ reporter gene into the coding region of *Dlx3* and modified this clone by using yeast homologous recombination. We have conducted functional analysis based on this system. The element showing highest conservation between human and mouse is named I37-1 (see Appendix II). It has 90% similarity over 350 bp. When this element was deleted from the 80 kb construct, reporter gene expression diminished in the limb bud apical ectodermal ridge. The second most conserved element (I37-2) was also deleted from the construct. In this case, reporter gene expression in first and second visceral arches was completely lost. The I37-2 element alone (250 bp) was able to drive expression in visceral arches when fused to a [basal] hsp68 promoter and a lacZ reporter gene. Closer examination of the I37-2 element across 13 mammalian species (see Appendix III) revealed those positions that showed no mutations among species; these have been shown empirically to serve as DNA binding sites for transcription factors.

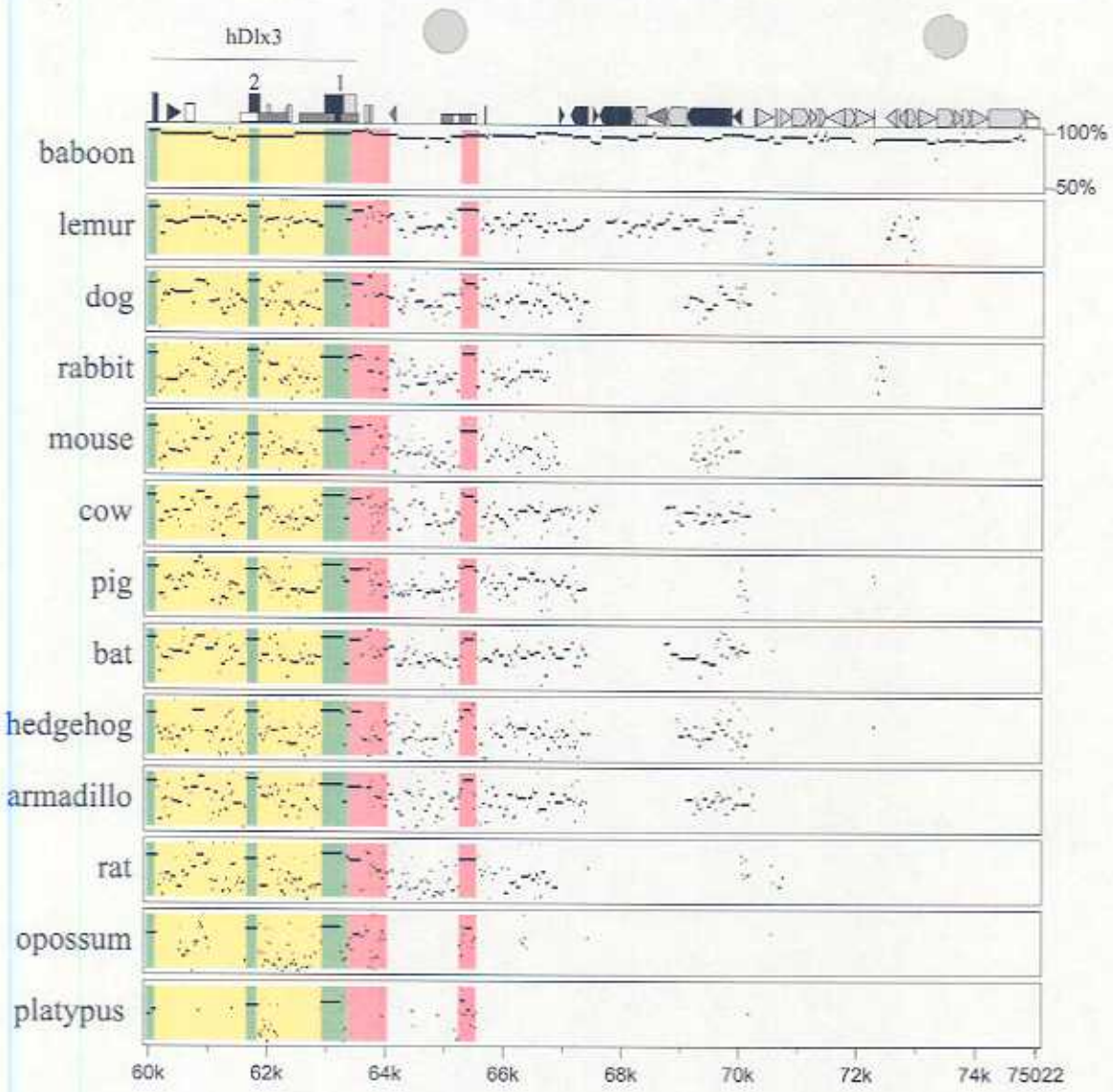
The appeal of the comparative genomics approach is that it reveals trends in the sequences that can guide functional characterization. Besides selection of highly conserved targets for assessment of enhancer activity, "differences" can also be highly revealing. For example, elements I37-2 and -5, which are highly conserved, are

clearly missing only in the platypus (see Appendix II). Element I37-2 is known to drive visceral arch expression and these data suggest that lack of the I37-2 element could be involved in the unusual jaw morphology of the platypus. This hypothesis is certainly testable and we are making the requisite constructs for mouse transgenic analysis in order to ascertain whether or not the loss of such a genomic region can have far reaching developmental consequences. Data from one such transgenic construct is given in Appendix IV. Moreover, we continue to analyze the sequences using different "reference" sequences. This is being done to determine if indeed there are other trends in the data that are not obvious due to our alignment scheme.

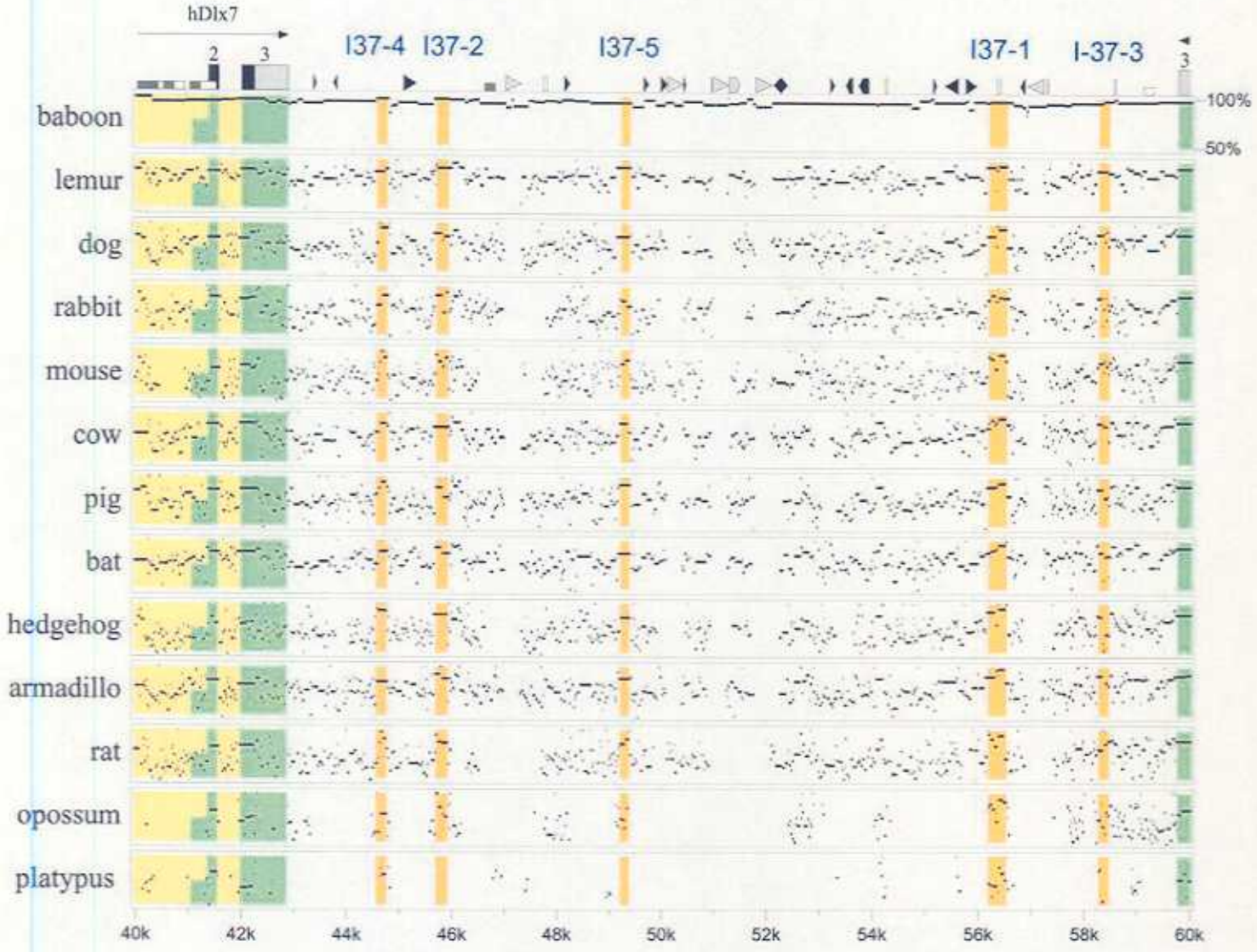

hDlx7
hDlx3
underlay
baboon
lemur
dog
rabbit
mouse
cow
pig
bat
hedgehog
armadillo
rat
opossum
platypus
1
10k
20k
30k
40k
50k
60k
70k
75022


hDlx7
100%
50%
baboon
lemur
dog
rabbit
mouse
cow
pig
bat
hedgehog
armadillo
rat
opossum
platypus
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k

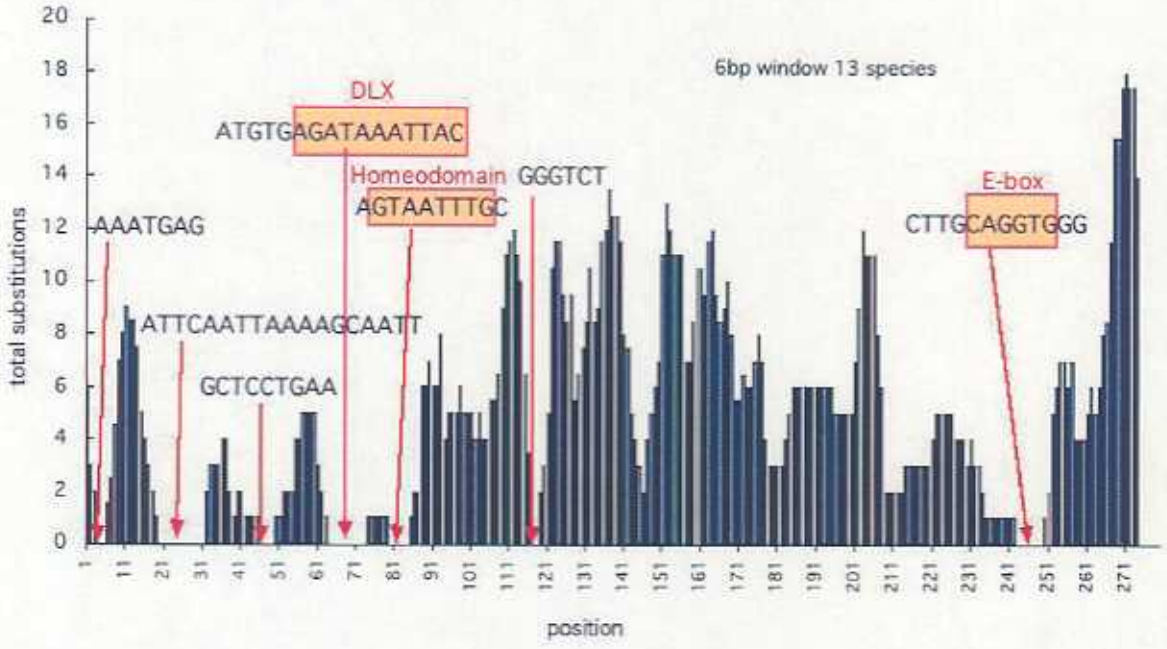

hDlx7
2
3
3
100%
50%
baboon
lemur
dog
rabbit
mouse
cow
pig
bat
hedgehog
armadillo
rat
opossum
platypus
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k


hDlx3
2
1
baboon
lemur
dog
rabbit
mouse
cow
pig
bat
hedgehog
armadillo
rat
opossum
platypus
100%
50%
60k
62k
64k
66k
68k
70k
72k
74k
75022

Appendix II. Intergenic region between Dlx7 and Dlx3 among 13 mammalian species (reference species is human). Orange boxes denote highly conserved regions.



Appendix III. Close analysis the I37-2 highly conserved region from 13 mammals. The data were analyzed using a sliding window method that generates a compilation of frequency histograms for each nucleotide position within the conserved region compared (Sumiyama K, Kim CB, Ruddle FH. 2001, "An efficient cis-element discovery method using multiple sequence comparisons based on evolutionary relationships," Genomics 71: 260-262). The regions of low substitution are potential sites for DNA binding proteins; known conserved motifs are outlined.



Appendix IV. Transgenic mouse analysis of PAC *Dlx3* LacZ reporter constructs. A mouse PAC clone containing a LacZ reporter gene inserted in-frame into *Dlx3* was used to generate a transgenic mouse (**left**, 13 dpc). This [wildtype control] embryo shows good expression in digits, nose (including vibrissae), mandible and outer ear. The embryo on the **right** shows a similar-staged transgenic embryo with the identical construct except that element I37-2 was removed. It appears that there is reduced overall expression, especially in the mandible, outer ear and nose area.