U7 snRNAs: A Computational Survey

Manja Lindemeyer\textsuperscript{a}, Axel Mosig\textsuperscript{b,c}, Bärbel M. R. Stadler\textsuperscript{c}, Peter F. Stadler\textsuperscript{a,e,f,d,g,*},

\textsuperscript{a}Bioinformatics Group, Department of Computer Science, University of Leipzig, Härterstraße 16-18, D-04107 Leipzig, Germany
\textsuperscript{b}Department of Combinatorics and Geometry (DCG), MPG/CAS Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences (SIBS) Campus, Shanghai, China
\textsuperscript{c}Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany
\textsuperscript{d}Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria
\textsuperscript{e}Interdisciplinary Center for Bioinformatics, University of Leipzig, Härterstraße 16-18, D-04107 Leipzig, Germany
\textsuperscript{f}Fraunhofer Institut für Zelltherapie und Immunologie — IZI, Deutscher Platz 5e, D-04103 Leipzig, Germany
\textsuperscript{g}Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

Manuscript 13 April 2007
Abstract

U7 snRNA sequences have been described only for a handful of animal species in the past. Here we describe a computational search for functional U7 snRNA genes throughout vertebrates which included the upstream sequence elements characteristic for snRNAs transcribed by pol-II. Based on the results of this search, we discuss the high variability of U7 snRNAs in both sequence and structure and we report on an attempt to find U7 snRNA sequences in basal deuterostomes and non-Drosophilid insect genomes based on a combination of sequence, structure, and promoter features. Due to the extremely short sequence and the high variability in both sequence and structure, no unambiguous candidates were found. These results cast doubt on putative U7 homologs in even more distant organisms which are reported in the most recent release of the Rfam database.

Key words: U7 snRNA, Noncoding RNA, RNA Secondary Structure, evolution

1 Introduction

The U7 snRNA is the smallest polymerase II transcript known to-date, with a length ranging from only 57nt (sea urchin) to 70nt (fruit-flies). Its expression level of only a few hundred copies per cell in mammals is at least three orders of magnitude smaller than the abundance of other snRNAs. It is part of the U7 RNP, which plays a crucial role in the 3'end processing of histone mRNAs (1). Restricted to metazoans, replication-dependent histone genes are the only eukaryotic protein-coding mRNAs that are not polyadenylated ending instead in a conserved stem-loop sequence, see (2) for a recent review.

The 5’ region of the U7 snRNA is complementary to the “Histone downstream element” (HDE), located just downstream of the conserved hairpin. The interaction of the U7 RNP with the HDE is crucial for the correct processing of the histone 3’ elements (1). The 3’ part of the U7 is occupied by a modified binding domain for the survival of motor neurons (SMN) protein complex. The binding domain consists of a deviant SMN-binding sequence and an adjacent stem-loop motif, see e.g. (3). The U7 RNP binds a distinct set of seven Sm-proteins, five of which are shared with the spliceosomal snRNAs, while the remaining two, Lsm10 and Lsm11, are probably restricted to the U7 snRNP (4; 5; 6). This difference is likely to be associated with the differences in the SMN-binding sequence. Recently, the U7 snRNP has not only received considerable attention from a structural biology point of view, see e.g. (7; 8), but it has also been investigated as a means of modifying splicing dys-regulation. In particular, U7 snRNA-derived constructs which target a mutant dystrophin gene were explored as a gene-therapy approach to Duchenne muscular dystrophy (9; 10).
Given the attention received by histone RNA 3’end processing and the protein components of the U7 snRNP, it may come as a surprise that the U7 snRNA itself has received little attention in the last decades. In fact, the only two experimentally characterized mammalian U7 RNAs are those of mouse (11; 12; 13; 14) and human (1; 15), while most of the earliest work on U7 snRNPs concentrated on the sea urchin Psammechinus miliaris (16; 17; 18; 19) and Xenopus species (20; 21; 22). More recently, the U7 RNA sequences have been reported for Drosophila melanogaster (23) and fugu (24).

We are aware of only two studies that considered U7 snRNA from a bioinformatics point of view. In (25), the U7 snRNA is used as an example for the application of Construct to compute consensus secondary structures, and (26) briefly reports on a blast based homology search which uncovered candidate sequences for chicken and two teleost fishes.

The U7 snRNP-dependent mode of histone end processing is a metazoan innovation (4; 2). Nevertheless, the most recent release of the Rfam database (27) [Version 8.0; Feb. 2007] lists sequences from eukaryotic protozoa, plants, and even bacteria. This discrepancy prompted us to critically assess the available information on U7 snRNAs.

2 Materials and Methods

The experimentally known sequences snRNA sequences were retrieved from Genbank. Starting from the known functional mouse gene (Genbank X54748.4) we used the built-in blast search function of ENSEMBL (release 43) to retrieve homologous regions in other mammalian genomes and the chicken genome. Parameters were set to “distance homologies” and repeat-masking was disabled. The resulting sequences were downloaded and aligned using both dialign2 (28) and clustalw (29) to determine whether the characteristic upstream and downstream elements were present. In order to check for consistency we compared these alignments with the ENSEMBL genomic alignments of the homologous human locus. In all cases, ENSEMBL data and our own search gave consistent results. The fugu U7 snRNA sequence described in (24) was used as starting point for searching the teleost fish genomes.

Drosophilid sequences, with the exception of Drosophila melanogaster, were obtained from the website of the Drosophila Comparative Genomics Consortium http://rana.lbl.gov/drosophila/caf1.html. Homologs of the single Drosophila melanogaster U7 snRNA region were used as blast queries, resulting again in unique hits in the other Drosophilid genomes that exhibit the characteristic upstream elements, together with at most one likely pseudogene in some species.
Sequence alignments of U7 sequences were generated separately for mammals, sauropsids, teleosts, frogs, sea urchins, and fruit flies using clustalw. These alignments were combined manually using the ralee mode (30) for Emacs.

Consensus secondary structure for a given sequence alignment are computed using RNAalifold (31).

We expanded the aln2pattern, the component of the fragrep distribution (32) that generates a collection of PWMs as search patterns with a “Sequence-Logo” style output derived from the WebLogo PostScript code (33). This provides a convenient way of generating graphical representations of sequence patterns that consist of collections of local motifs from a single multiple sequence alignment.

In addition to purely sequence-based methods we also searched for more distant homologies based on combined sequence/structure patterns using Sean Eddy’s rnabob software. We constructed search patterns comprising the most conserved motif of the histone binding site, the SMN binding motif, and a stem-loop structure at the 3’ end which is enclosed by two GC pairs. In order to increase specificity, we additionally included a species-specific model of the PSE element, which was derived from the upstream regions of the spliceosomal snRNAs U1, U2, U4, U5, U4atac, U11, and U12. These RNAs are larger and better conserved than the U7 snRNAs and hence were straightforward to find also in most metazoan genome where they were not annotated previously. The rnabob descriptors are listed in the electronic supplement, http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/07-010/.

3 Results

3.1 Bona fide U7 snRNA Sequences

The results of the blast-based searches are summarized in Tab. 1. In most species only a single gene with clear snRNA-like upstream elements was found. In addition blast identified several pseudogenes. Clusters of U7 snRNAs as previously described for sea urchin and Xenopus were otherwise only found in zebrafish, Fig. 1.

The short length and the substantial divergence of the U7 snRNA sequences make it impossible to distinguish functional U7 snRNAs from pseudogenes based on the U7 sequence alone. To make this distinction, it is necessary to

1 Downloaded from ftp://ftp.genetics.wustl.edu/pub/eddy/software/rnabob-2.1.tar.Z
In several genomes we were not able to find an unambiguous candidate for a functional U7 snRNA, although we found sequences that clearly derive from U7 but are not accompanied by a recognizable PSE. Examples include *Sorex araneus* and platypus. Most likely, these *blast* hits are pseudogenes, although
many of them are annotated with ENSEMBL gene IDs. This annotation derives from sequence homology with the examples stored in the Rfam database. In Fig. 3 and Tab. 1 (Appendix) we compile the results of our blast-based homology search, which contains only sequences which are either experimentally known to be expressed or which are predicted to be functional genes based on the presence of conserved upstream elements.

Separate multiple sequence alignments of Amniots, Teleosts, Xenopus, sea urchins, and flies reveal strong conservation of the SMN-binding motif, consisting of the deviant SMN-binding site AUUUNUC and the hairpin 3’ structure. Furthermore, the histone-binding region contains a universally conserved box UCUUU (37). Using these features as anchors, one obtains the alignment in Fig. 3, which highlights the differences between major clades. Notable variations within the vertebrates are in particular the A-rich 5’ and the reduced stem in teleosts, and their A-rich sequence in the hairpin loop. The hairpin region is very poorly conserved at sequence level between vertebrates, sea urchins, and flies, although its structural variation is limited in essence to the length of the stem and a few short interior loops or single-nucleotide bulges.

3.2 More Distant Homologs?

The U7 snRNA sequences evolve rather fast. Together with the short sequence length, this limits the power of sequence-based approaches to distant homology search. The consensus pattern in Fig. 3 indicates quite clearly that such methods are bound to fail outside the four groups with experimentally known sequences (tetrapoda, teleosts, echinoderms, fruit-flies). Indeed, both blast and fragrep did not provide additional candidates that could be unambiguously classified as U7 snRNAs based on sequence information alone.

The comparison of the U7 hairpins in the different clades, Fig. 4, reveals significant differences in the secondary structures of invertebrates and vertebrates: vertebrate have smaller stem-loop structures with smaller or no interior loops or bulges. The stem in teleosts, furthermore, is systematically shorter than in tetrapods. These structural differences between clades has to be taken into account for homology search. In fact, as a consensus rule, we can only deduce that the stem-loop structure has a total of 8-15 base pairs, that it is nearly symmetric, and that it is enclosed by an uninterrupted stem of length at least 5 with two GC pairs at its base.

Even combined with with the conserved sequence motives in the 5’ part of the molecule, this yields only a rather loose definition of a U7. Release 8.0 of the Rfam database (27) lists several sequences in its U7 RNA section that are surprising. Neither contained in the literature nor contained in the manu-
sequences. Below we display sequence logos for the partial alignment comprising the SMN binding site, and the histone-binding domains are highlighted. The 5' most

![Sequence Logo](image)

Fig. 3. Manually curated alignment of functional U7 snRNA sequence. The 3' stem, the SMN binding site, and the histone-binding domains are highlighted. The 5' most part of the histone-binding region is not aligned between vertebrate and Drosophilid sequences. Below we display sequence logos for the partial alignment comprising only tetrapods, teleosts, sea urchins, or flies, respectively, as well as the consensus pattern arising from combining all data.

Manually curated U7 “seed-set”, these candidate sequences have been found using a homolog search based on infernal (38) and the seed alignment. While the Danio rerio sequences are identical with the sequences we identified in work starting from the much closer homolog in fugu, the candidates reported for Caenorhabditis elegans, and Girardia tigrina raise serious doubts. The
Fig. 4. Comparison of U7 hairpin structures. Consensus secondary structures are computed using RNAalifold using the manual improved alignments of tetrapods, teleost fishes, sea urchins, and fruit-flies, respectively. Circles indicate consistent and compensatory mutations which leave the structure intact. Gray letters indicate that one or two of the aligned sequences cannot form the base pair.

_Caenorhabditis elegans_ sequence, although ostensibly well conserved in comparison with the deuterostome sequences, has no recognizable homologs in any one of the other three sequenced Caenorhabditis species, (_C. briggsae_, _C. remanei_, "C. sp.4". The _Girardia tigrina_ sequence is located in the 3' UTR of the _Dthox-E-Hox_ gene (_X95413_). Both sequences furthermore do not share the consensus SMN-binding motive UUUNUC. Several additional candidates were reported for plants, protozoans, and even bacteria. Since these organisms do not have replication-dependent metazoan-style histone 3' end processing (_4; 2_), and since these histone genes are apparently the only mRNAAs that are processed in this way (_39_), it would be extremely surprising if true homologs of U7 snRNAs were found outside the metazoans. These examples show once again that at least for very short ncRNAs, the results from homology searches have to be taken with caution, in particular when they are not corroborated by additional supporting evidence.

The poor sequence conservation between major groups highlighted in Fig. 3 suggest that purely sequence-based homology searches have little chance of success in insect or basal deuterostome genomes. Indeed, neither blast nor fragrep found convincing candidates. We therefore resorted to structure-based approaches and explicitly included the PSE in the search procedure (see Materials & Methods for details). We used rnaob with a non-restrictive pattern to find plausible initial candidates, which were then manually compared with the alignment in Fig. 3. The most plausible candidates are shown
Fig. 5. Best candidates from searches with `rnabob` in the lamprey *Petromyzon marinus*, *Branchiostoma floridae*, and *Bombyx mori*. In addition to the putative U7 RNA sequence shown here, these candidate sequences also have a putative PSE element associated with them.

in Fig. 5, albeit none of them is unambiguous. No convincing candidates were found in the fly *Anopheles gambiae*, and the honeybee *Apis melifera*.

4 Discussion

Since U7 snRNA has its primary function in histone 3’ maturation it is virtually certain that this class of non-coding RNAs is restricted to metazoan animals – after all, the process in which they play a crucial role is unknown outside multicellular animals. With its length of 70nt or less, U7 snRNA is the smallest known pol-II transcript. Each of its three major domains, the histone binding region, the SNM binding sequence, and the 3’ stem-loop structure exhibit substantial variation in both sequence and structural details, as can be seen from the detailed sequence alignments (Fig. 3) and the structural models of the terminal stem-loop structure (Fig. 4). As a consequence, our computational survey not only compiled a large number of previously undescribed U7 homologs from vertebrates and drosophilids, but also stresses the limits of current approaches to RNA homology search.

While `blast` already fails to unambiguously recognize teleost fish homology from mammalian queries and *vice versa*, even more sophisticated (and computationally expensive) methods have limited success when applied to basal deuterostomes or insect genomes. On the other hand, not only the limited sensitivity of current approaches poses a problem. Conversely, the most sensitive methods are fooled plant or bacterial sequences which are almost certainly false positives.

In summary, thus, this study calls both for more experimental data on U7 snRNAs – which, if any, of our U7 candidate sequence in lamprey, silk worm, are really U7 snRNAs in these species? – and for improved bioinformatics approaches for homology search that can deal with such small and rapidly evolving genes.
Supporting Online Material

Alignments of U7 sequences and other data can be downloaded in machine-readable form from http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/07-010/.

Acknowledgments

BMRS and PFS thank the PICB in Shanghai for its hospitality, where much of this work was performed in spring 2007. Financial support by the DFG-funded Graduierten Kolleg “Wissensrepräsentation” to ML, the DFG Bioinformatics Initiative to PFS is gratefully acknowledged.

Author’s Contributions

All authors collaborated in data analysis and homolgy search as well as in the interpretation of the data. AM and PFS conceived the study and wrote the manuscript. All authors read and approved the final manuscript.

Conflicts of Interests

None declared.

References

[1] K. Mowry and J. A. Steitz. Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNAs. Science, 238:1682–1687, 1987.
[2] W. F. Marzluff. Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. Curr. Opin. Cell. Biol., 17:274–280, 2005.
[3] T. J. Golembe, J. Yong, and G. Dreyfuss. Specific sequence features, recognized by the SMN complex, identify snRNAs and determine their fate as snRNPs. Mol. Cell Biol., 25:10989–11004, 2005.
[4] T. N. Azzouz and D. Schümperli. Evolutionary conservation of the U7 small nuclear ribonucleoprotein in Drosophila melanogaster. RNA, 9:1532–1541, 2003.
[5] R. S. Pillai, M. Grimmel, G. Meister, C. L. Will, R. Lührmann, U. Fischer, and D. Schümperli. Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes. Dev.*, 17:2321–2333, 2003.

[6] D. Schümperli and R. S. Pillai. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cell. Mol. Life Sci.*, 61:2560–2570, 2004.

[7] N. G. Kolev and J. A. Steitz. *In vivo* assembly of functional U7 snRNP requires RNA backbone flexibility within the Sm-binding site. *Nat. Struct. Mol. Biol.*, 13:347–353, 2006.

[8] S. Jaeger, F. Martin, J. Rudinger-Thirion, R. Giegé, and G. Eriani. Binding of human SLBP on the 3’-UTR of histone precursor H4-12 mRNA induces structural rearrangements that enable U7 snRNA anchoring. *Nucleic Acids Res.*, 34:4987–4995, 2006.

[9] C. Brun, D. Suter, C. Pauli, P. Dunant, H. Lochmüller, B. J.-M., D. Schümperli, and J. Weis. U7 snRNAs induce correction of mutated dystrophin pre-mRNA by exon skipping. *Cell. Mol. Life Sci.*, 60:557–566, 2003.

[10] A. Goyenvalle, A. Vulin, F. Fougerousse, F. Leturcq, J.-C. Kaplan, L. Garcia, and O. Danos. Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. *Science*, 306:1796–1799, 2004.

[11] D. Soldati and D. Schümperli. Structural and functional characterization of mouse U7 small nuclear RNA active in 3’ processing of histone pre-mRNA. *Mol. Cell Biol.*, 8:1518–1524, 1988.

[12] A. Gruber, D. Soldati, M. Burri, and D. Schumperli. Isolation of an active gene and of two pseudogenes for mouse U7 small nuclear RNA. *Biochim. Biophys. Acta*, 1088:151–154, 1991.

[13] S. C. Phillips and P. C. Turner. A transcriptional analysis of the gene encoding mouse U7 small nuclear RNA. *Gene*, 116:181–186, 1992.

[14] S. C. Phillips and P. C. Turner. Sequence and expression of a mouse U7 snRNA type II pseudogene. *DNA Seq.*, 1:401–404, 1991.

[15] Y.-T. Yu, W.-Y. Tarn, T. A. Yario, and J. A. Steitz. More Sm snRNAs from vertebrate cells. *Exp. Cell Res.*, 229:276–281, 1996.

[16] K. Strub, G. Galli, M. Busslinger, and M. L. Birnstiel. The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. *EMBO J.*, 3:2801–2807, 1984.

[17] M. De Lorenzi, U. Rohrer, and M. L. Birnstiel. Analysis of a sea urchin gene cluster coding for the small nuclear U7 RNA, a rare RNA species implicated in the 3’ editing of histone precursor mRNAs. *Proc. Natl. Acad. Sci. USA*, 83:3243–3247, 1986.

[18] G. M. Gilmartin, F. Schaufele, G. Schaffner, and M. L. Birnstiel. Functional analysis of the sea urchin U7 small nuclear RNA. *Mol. Cell Biol.*, 8:1076–1084, 1988.

[19] C. Southgate and M. Busslinger. *In vivo and in vitro* expression of U7
snRNA genes: cis- and trans-acting elements required for RNA polymerase II-directed transcription. *EMBO J.*, 8:539–549, 1989.

[20] S. C. Phillips and M. L. Birnstiel. Analysis of a gene cluster coding for the *Xenopus laevis* U7 snRNA. *Biochim. Biophys. Acta*, 1131:95–98, 1992.

[21] N. J. Watkins, S. C. Phillips, and P. C. Turner. The U7 small nuclear RNA genes of *Xenopus borealis*. *Biochem. Soc. Trans.*, 20:301S, 1992.

[22] C.-H. H. Wu and J. G. Gall. U7 small nuclear RNA in C snurposomes of the Xenopus germinal vesicle. *Proc. Natl. Acad. Sci. USA*, 90:6257–6259, 1993.

[23] Z. Dominski, X.-c. Yang, M. Purdy, and W. F. Marzluff. Cloning and characterization of the *Drosophila* U7 small nuclear RNA. *Proc. Natl. Acad. Sci. USA*, 100:9422–9427, 2003.

[24] E. Myslinksi, A. Krol, and P. Carbon. Characterization of snRNA and snRNA-type genes in the pufferfish *Fugu rubripes*. *Gene*, 330:149–158, 2004.

[25] R. Lück, S. Gräf, and G. Steger. Construct: A tool for thermodynamic controlled prediction of conserved secondary structure. *Nucl. Acids Res.*, 27:4208–4217, 1999.

[26] A. F. Bompfinewerer, C. Flamm, C. Fried, G. Fritzsch, I. L. Hofacker, J. Lehmann, K. Missal, A. Mosig, B. Müller, S. J. Prohaska, B. M. R. Stadler, P. F. Stadler, A. Tanzer, S. Washietl, and C. Witwer. Evolutionary patterns of non-coding rnas. *Th. Biosci.*, 123:301–369, 2005.

[27] S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, and A. Bateman. *Rfam*: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*, 33:D121–D124, 2005.

[28] B. Morgenstern. DIALIGN2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*, 15:211–218, 1999.

[29] J. D. Thompson, D. G. Higgins, and T. J. Gibson. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.*, 22:4673–4680, 1994.

[30] S. Griffiths-Jones. *RALEE*—RNA alignment editor in *Emacs*. *Bioinformatics*, 21:257–259, 2005.

[31] I. L. Hofacker, M. Fekete, and P. F. Stadler. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, 319:1059–1066, 2002.

[32] A. Mosig, K. Sameith, and P. F. Stadler. *fragrep*: Efficient search for fragmented patterns in genomic sequences. *Geno. Prot. Bioinfo.*, 4:56–60, 2005.

[33] G. E. Crooks, G. Hon, J. M. Chandonia, and S. E. Brenner. WebLogo: A sequence logo generator. *Genome Research*, 14:1188–1190, 2004.

[34] H. N. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem.*, 276:26733–26736, 2001.

[35] G. Hernandez Jr., F. Valafar, and W. E. Stumph. Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved
in determining promoter activity and RNA polymerase specificity. *Nucleic Acids Res.*, 35:21–34, 2007.

[36] D. Soldati and D. Schümperli. Structures of four human pseudogenes for U7 small nuclear RNA. *1990*, 95:305–306, 1990.

[37] Z. Dominski, X.-C. Yang, M. Purdy, and W. Marzluff. Differences and similarities between *Drosophila* and mammalian 3’ end processing of histone pre-mRNAs. *RNA*, 11:1835–1847, 2005.

[38] E. P. Nawrocki and S. R. Eddy. Query-dependent banding (QDB) for faster RNA similarity searches. *PLoS Comp. Biol.*, 3:e56, 2007. Doi:10.1371/journal.pcbi.0030056.

[39] W. D. Townley-Tilson, S. A. Pendergrass, W. F. Marzluff, and M. L. Whitfield. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein RNA. *RNA*, 12:1853–1867, 2006.
| Species                  | Assembly | Sequence from | to | ori | DB ID                      | ψ    |
|-------------------------|----------|---------------|----|-----|---------------------------|------|
| Mus musculus            | ensembl43| Chr.6         |    |     | 12470884 - 12470906      |      |
| Rattus norvegicus       | ensembl43| Chr.4         |    |     | 11816584 - 11816586      |      |
| Homo sapiens            | ensembl43| Chr.12        |    |     | 6923240 - 6923302        |      |
| Macaca mulatta          | ensembl43| Chr.11        |    |     | 7125496 - 7125557        |      |
| Otoloturn gurnettii     | PreEnsembl43| scaffold |   |     | 117572 - 117633           |      |
| Orcyllagocagus canalicus| ensembl43| GeneScaffold1693| | | 111485 - 111546 |      |
| Procyca capensis        | NCBI-TRACE| 177519230    |    |     | 275 - 396                |      |
| Locodonta africana      | ensembl43| scaffold60301|   |     | 425 - 431                |      |
| Echinops telfairi       | ensembl43| GeneScaffold2204| | | 10742 - 10803 |      |
| Felis catus             | ensembl43| GeneScaffold69|   |     | 192907 - 192909          |      |
| Rattus norvegicus       | ensembl43| Chr.X         |    |     | 11816584 - 11816586      |      |
| Macaca mulatta          | ensembl43| Chr.11        |    |     | 7125496 - 7125557        |      |
| Otoloturn gurnettii     | PreEnsembl43| scaffold |   |     | 117572 - 117633           |      |
| Orcyllagocagus canalicus| ensembl43| GeneScaffold1693| | | 111485 - 111546 |      |
| Procyca capensis        | NCBI-TRACE| 177519230    |    |     | 275 - 396                |      |
| Locodonta africana      | ensembl43| scaffold60301|   |     | 425 - 431                |      |
| Echinops telfairi       | ensembl43| GeneScaffold2204| | | 10742 - 10803 |      |
| Felis catus             | ensembl43| GeneScaffold69|   |     | 192907 - 192909          |      |
| Rattus norvegicus       | ensembl43| Chr.X         |    |     | 11816584 - 11816586      |      |
| Macaca mulatta          | ensembl43| Chr.11        |    |     | 7125496 - 7125557        |      |
| Otoloturn gurnettii     | PreEnsembl43| scaffold |   |     | 117572 - 117633           |      |
| Orcyllagocagus canalicus| ensembl43| GeneScaffold1693| | | 111485 - 111546 |      |
| Procyca capensis        | NCBI-TRACE| 177519230    |    |     | 275 - 396                |      |
| Locodonta africana      | ensembl43| scaffold60301|   |     | 425 - 431                |      |
| Echinops telfairi       | ensembl43| GeneScaffold2204| | | 10742 - 10803 |      |
| Felis catus             | ensembl43| GeneScaffold69|   |     | 192907 - 192909          |      |
| Rattus norvegicus       | ensembl43| Chr.X         |    |     | 11816584 - 11816586      |      |
| Macaca mulatta          | ensembl43| Chr.11        |    |     | 7125496 - 7125557        |      |
| Otoloturn gurnettii     | PreEnsembl43| scaffold |   |     | 117572 - 117633           |      |
| Orcyllagocagus canalicus| ensembl43| GeneScaffold1693| | | 111485 - 111546 |      |
| Procyca capensis        | NCBI-TRACE| 177519230    |    |     | 275 - 396                |      |
| Locodonta africana      | ensembl43| scaffold60301|   |     | 425 - 431                |      |
| Echinops telfairi       | ensembl43| GeneScaffold2204| | | 10742 - 10803 |      |
| Felis catus             | ensembl43| GeneScaffold69|   |     | 192907 - 192909          |      |
| Rattus norvegicus       | ensembl43| Chr.X         |    |     | 11816584 - 11816586      |      |
| Macaca mulatta          | ensembl43| Chr.11        |    |     | 7125496 - 7125557        |      |
| Otoloturn gurnettii     | PreEnsembl43| scaffold |   |     | 117572 - 117633           |      |
| Orcyllagocagus canalicus| ensembl43| GeneScaffold1693| | | 111485 - 111546 |      |
| Procyca capensis        | NCBI-TRACE| 177519230    |    |     | 275 - 396                |      |
| Locodonta africana      | ensembl43| scaffold60301|   |     | 425 - 431                |      |
| Echinops telfairi       | ensembl43| GeneScaffold2204| | | 10742 - 10803 |      |
| Felis catus             | ensembl43| GeneScaffold69|   |     | 192907 - 192909          |      |

**Notes:** ψ gives the number of paralog loci, most likely U7 pseudogenes, defined by a blast E-value less than 0.001 compared to the functional copy. CAF-1 refers to the genome freezes used Drosophila Comparative Genomics Consortium retrieved from [http://rana.lbl.gov/drosophila/caf1.html](http://rana.lbl.gov/drosophila/caf1.html). The Drosophila melanogaster sequence is the one used by the USCS browser (Release 4; Apr. 2004, UCSC version dm2). The sea urchin Genome BCM_Spur_v2.1 was obtained from [ftp://ftp.bgee.bcm.tmc.edu/pub/data/Spurpuratus/fasta/Spur_v2.1/linearScaff](ftp://ftp.bgee.bcm.tmc.edu/pub/data/Spurpuratus/fasta/Spur_v2.1/linearScaff).