Arthropod 7SK RNA

Andreas R. Gruber,* Carsten Kilgus,†** Axel Mosig,†‡ Ivo L. Hofacker,* Wolfgang Hennig,†** and Peter F. Stadler§‖¶*#

*Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria; †Department of Combinatorics and Geometry, CAS–MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, Shanghai, China; ‡Max Planck Insitute for Mathematics in the Sciences, Leipzig, Germany; §Bioinformatics Group, Department of Computer Science, University of Leipzig, Leipzig, Germany; ‖Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany; ¶Fraunhofer Institut für Zelltherapie und Immunologie—IZI Perlickstrasse 1, Leipzig, Germany; #Santa Fe Institute; and **Institut für Genetik, Johannes Gutenberg-Universität Mainz, Mainz, Germany

The 7SK small nuclear RNA (snRNA) is a key player in the regulation of polymerase (pol) II transcription. The 7SK RNA was long believed to be specific to vertebrates where it is highly conserved. Homologs in basal deuterostomes and a few lophotrochozoan species were only recently reported. On longer timescales, 7SK evolves rapidly with only few conserved sequence and structure motifs. Previous attempts to identify the Drosophila homolog thus have remained unsuccessful despite considerable efforts. Here we report on the discovery of arthropod 7SK RNAs using a novel search strategy based on pol III promoters, as well as the subsequent verification of its expression. Our results demonstrate that a 7SK snRNA featuring 2 highly structured conserved domains was present already in the bilaterian ancestor.

Introduction

The 7SK small nuclear RNA (snRNA) is one of the most highly abundant noncoding RNAs (ncRNAs) in vertebrate cells. Due to its abundance, it has been known since the 1960s. Its function as a transcriptional regulator, however, has only recently been discovered: 7SK mediates the inhibition of transcription elongation factor P-TEFb, a critical regulator of RNA polymerase (pol) II transcription which stimulates the elongation phase (Nguyen et al. 2001; Yang et al. 2001; Michels et al. 2004; Blazek et al. 2005; Egloff et al. 2006; Peterlin and Price 2006; Krueger et al. 2008). In addition, 7SK RNA suppresses the deaminase activity of APOBEC3C and sequesters this enzyme in the nucleolus (He et al. 2006).

The pol III transcript with a length of about 330 nt (Krüger and Benecke 1987; Murphy et al. 1987) is highly conserved in vertebrates (Gürsoy et al. 2000). In contrast to the nearly perfect sequence conservation in jawed vertebrates, the 7SK RNA from the lamprey *Lampetra fluviatilis* differs in more than 30% of its nucleotide positions from its mammalian counterpart (Gürsoy et al. 2000). Based on several unsuccessful attempts to clone 7SK homologs, the molecule has long been believed to be vertebrate specific. In a recent contribution (Gruber et al. 2008), however, we reported on the computational detection and experimental verification of 7SK sequences from several basal deuterostomes as well as a few Lophotrochozoa. Direct experimental evidence is available for the hagfish *Myxine glutinosa*, the lancet *Branchiostoma lanceolatum*, and the snail *Helix pomata*. In contrast, neither experimental cloning procedures nor computational homology search revealed a plausible 7SK candidate in *Drosophila melanogaster* or any other sequenced genome of an ecdysozoan.

In this contribution, we report on the computational discovery of the 7SK snRNA homologue in Drosophilidae and other arthropod genomes, on its bioinformatical characterization, and its subsequent verification in *D. melanogaster*.

Materials and Methods
Sequence Data

Genomic sequences were downloaded from EN-SEMBL (version 48, http://www.ensembl.org), the Joint Genome Institute (http://www.jgi.doe.gov), and the Broad Institute (http://www.broad.mit.edu) Web sites. Details on the assemblies used here are listed in the Electronic Supplement online. In all, 100 nt upstream regions of the annotated U6 (CR31379, CR32867, and CR31539) and U6atac (CR32989) RNAs were retrieved from FlyBase (http://www.flybase.org). Previously described 7SK sequences and their alignment were taken from Gruber et al. (2008).

Homology Search

From the 100 nt upstream regions of the *D. melanogaster* U6 and U6atac snRNAs, we generated a multiple sequence alignment using MAFFT (Katoh et al. 2002). Guided by previous findings (Hernandez et al. 2007; Mount et al. 2007), we selected the search pattern such that it contained the conserved promoter region, 2 conserved thymidine residues to guarantee distinguishability from pol II recognized PSE elements and the TATA-box. Then we scanned the *D. melanogaster* genome using fragrep (version 2) (Mosig et al. 2007) in position weight matrix mode. The sequence conservation pattern downstream of the resulting hits was visually inspected in the UC Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu). Neoptera species were searched iteratively using the Blast front end at the FlyBase Web site, using previously identified hits as additional queries. In addition, we searched GenBank using National Center for Biotechnology Information's (NCBI) Web interface (http://www.ncbi.nlm.nih.gov/blast). Sean Eddy's rnabob (http://selab.janelia.org/software.html) was used for pattern-based RNA structure searches.
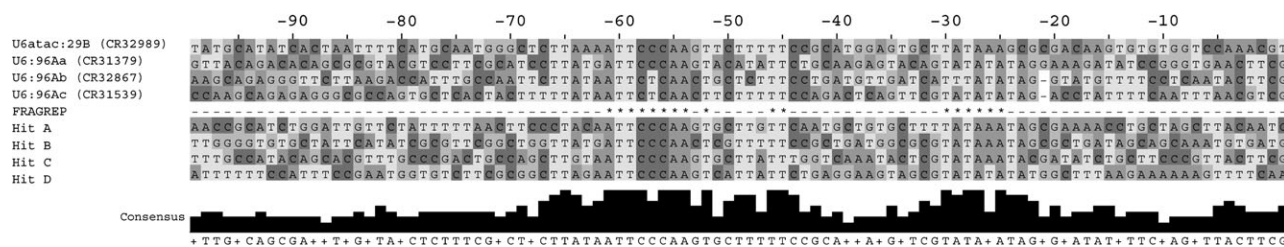
Fig. 1.—Upstream regions of known U6 and U6atac genes and the 4 candidate sequences listed in table 1. The fragrep pattern used to scan the genome of *Drosophila melanogaster* is indicated in the middle.

Sequence–Structure Alignments

Initial alignments were generated using ClustalW (Thompson et al. 1994), dialign2 (Morgenstern 1999; Morgenstern et al. 2006), and MAFFT (Katoh et al. 2002). Initial structure annotation was produced using RNAalifold (Hofacker et al. 2002). This information was used as the basis for a semimanual alignment edited in emacs using the ralee mode (Griffiths-Jones 2005). The 5′ and 3′ domains were realigned using locarna (Will et al. 2007).

Northern Blot

Total RNA was isolated from *D. melanogaster* (Canton S) flies according to Chomczynski and Sacchi (1987). For northern blots, 15 μg of total RNA were separated in 2% agarose–formaldehyde gels and blotted onto Hybond-N membrane (Roche, Palo Alto, CA; catalog number 1417240) according to Sambrook et al. (2001). A DIG-labeled probe of 344 nt of the 7SK RNA was obtained by amplification of the respective fragment on genomic DNA of *D. melanogaster* (Canton S) with the primers CGATATTCAGG-TAACTGCATCTG (positions 35–58 in the predicted transcript) and CGAAAATCCGAAGCTAAGCTACT (positions 356–379) and the polymerase chain reaction (PCR) DIG-labeling mix (Roche, catalog number 11636090910). Hybridization was carried out in 5× standard saline citrate, 0.1%N-lauryl-sarcosine, 1% milk powder, and 0.02% sodium dodecyl sulfate at 65 °C overnight. The membranes were washed with 0.1 M Tris–HCl (pH 7.5), 0.15 M NaCl, and 0.3% Tween 20. The same buffer with additional 1% milk powder was used for the blocking. For detection, we used the alkaline phophatase-conjugated anti-DIG-antibody (Roche, catalog number 11093657910) in a dilu-tion of 1:7500 in the same buffer at room temperature for 2 h. For detection, 7-ml AP buffer (0.1 M Tris–HCl, 0.1 M NaCl, 5 mM MgCl$_2$, pH 9.5) was freshly mixed with 14 μl nitroblue tetrazolium chloride (NBT; 100 mg/ml) and 21 μl 5-bromo-4-chloro-3 p-toluidine salt (BCIP; 50 mg/ml). The substrate reaction was stopped when a signal appeared (after 20–30 min) by adding ddH$_2$O to decrease the pH value.

Results
Initial Search

Because direct homology search had failed previously, we employed a different strategy. The snRNAs, including the 7SK snRNA, exhibit a characteristic promoter structure (Hernandez, 2001) that is fairly well conserved in evolution. The spliceosomal snRNAs had recently been studied in great detail in Drosophilidae (Hernandez et al. 2007; Mount et al. 2007), and their promoter sequence motifs are known in detail for most of the 12 sequenced drosophilid fly species. The 7SK snRNA has a canonical pol III type 3 promoter in vertebrates, see Bannister et al. (2007) and the references therein. We thus derived a search pattern for canonical type 3 pol III promoters, figure 1, using a region of 100 nt upstream of the U6 and U6atac snRNAs as template.

The pattern was used to search the *D. melanogaster* genome. In addition to recovering the U6 and U6atac snRNAs, we uncovered 4 hits, summarized in table 1. One of them belongs to a known small nucleolar RNA previously described in Huang et al. (2005). Pol III regulated expression of snoRNAs has not been described in Drosophilidae so far. The observation is not unexpected, however, because pol III transcription of snoRNAs has been observed previously in *Saccharomyces cerevisiae* (Moqtaderi and Struhl 2004). The candidate located on the X chromosome shows no

Table 1
Characterization of Conserved Loci with Putative U6-Like snRNA Promoter Motifs

Location		RNAz	Pol III	Note	Reference
A	*3L*:7 632 840-7 632 900(+)	—	—	CR34703 C/D snoRNA Me18S-A1806-RA	(Huang et al. 2005)
B	*3R*:3 300 270-3 300 900(−)	1077	TRF1 (1582) BRF (1580)	CR33925 snmRNA:331-RA	(Yuan et al. 2003)
C	*3R*:19 555,800-19 556 250(−)	7371, 7372	BRF (9494) BRF (9495)	CR33682 snmRNA:342 RNAse MRP	(Yuan et al. 2003; Piccinelli et al. 2005; Woodhams et al. 2007)
D	X:21 308 600-21 308 750(+)	—	—	—	—

NOTE.—Evidence for evolutionarily conserved secondary structure is taken from a recent RNAz-based survey. The numbers refer to the loci listed in the Supplemental Material of Rose et al. (2007). Evidence from ChIP-on-chip data for binding of TRF1 and BRF refers to the loci listed in the Supplemental Material of Isogai et al. (2006).
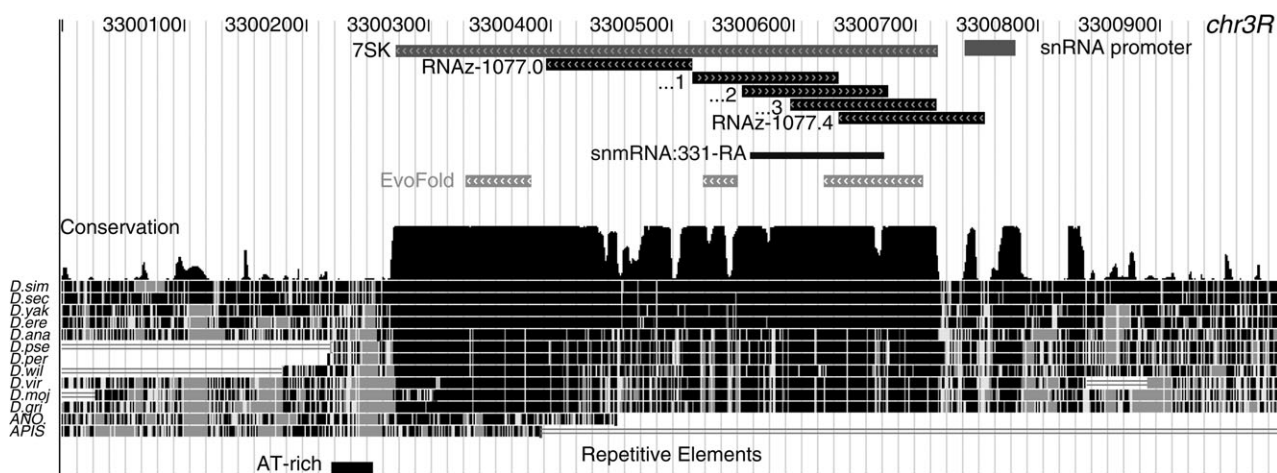
Fig. 2.—Detailed genomic view of the 7SK candidate B at *3R*(3.3M). Adapted from a UCSU Genome Browser view. The upper bar indicates the predicted 7SK transcript along with the snRNA promoter element. The predicted transcript overlaps the conserved secondary structures reported by RNAz and evofold as well as the fragment cloned in Yuan et al. (2003).

direct evidence for pol III transcription in the study (Isogai et al. 2006).

Two candidates on chromosome *3R* overlap small nonmessenger RNAs cloned in an experimental survey of small nonmessenger RNAs (snmRNAs) in *D. melanogaster* (Yuan et al. 2003). While both Woodhams et al. (2007) and Piccinelli et al. (2005) list candidate C as RNAse MRP, no further annotation is available for candidate B. A comparison with a recent computational survey of structure conserved ncRNAs in flies shows that both loci have been detected by RNAz (Rose et al. 2007). Furthermore, there is direct evidence that these regions are transcribed by pol III. Isogai et al. (2006) showed that unlike in most other eukaryotes, TRF1/BRF binding appears responsible for the initiation of all classes of pol III transcription and they have mapped TRF1- and BRF-binding sites in the respective sites.

Homology Search

Candidate B, located on chromosome *3R* at 3.3M, figure 2, shows strong evidence for pol III transcription, strong evidence for an evolutionarily well-conserved secondary structure, and a characteristic T-rich region indicative of a pol III terminator. With an overall length of about 450 nt, the conserved sequence is only slightly longer than the previously known 7SK snRNAs (Gruber et al. 2008). Note, however, that the ends of the transcripts cannot be predicted accurately. In *D. melanogaster*, an AT-rich low-complexity region is located immediately downstream of the annotated conserved region, which could be (partially) transcribed. The human 7SK, for instance, shows some variability in the exact position of its 3′ end, which consists of a short U-rich tail of length 5–7. In addition, a fraction of the human transcripts are adenylated posttranscriptionally (Sinha et al. 1998). For the bioinformatic analysis, we defined the 3′ end of the arthropod candidate sequence before the low-complexity region.

The high level of sequence conservation in Drosophilidae promoted us to search for homologs in additional arthropod genomes. In Neoptera species, these could easily be retrieved by iterative Blast searches. As Blast failed to recover a homologue in *Ixodes scapularis*, we constructed a fragrep pattern from already identified arthropod sequences (see Electronic Supplement online at http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/08-008/).

We recovered candidate sequences from most of the available arthropod genomes, with the notable exceptions of the lepidopteran *Bombyx mori* and the aphid *Acyrthosiphon pisum* and the crustacean *Daphnia pulex*, see figure 3 and Electronic Supplement online. In these cases, it is plausible to assume that no candidate was found due to the quality of the current draft assemblies, although we cannot rule out that the sequence is too derived to be recognizable by our search methods.

In addition to genomic DNA, we also searched the NCBI expressed sequence tag (EST) database using all the genomic hits as Blast queries. This resulted in some evidence for expression of the 7SK candidates beyond the fragments reported in Yuan et al. (2003).

A Blast search of the NCBI NR and EST collections revealed additional evidence for transcription of this locus in several species, namely *Culex pipiens* (multiple ESTs from an unpublished EST project), *Armigeres subalbatus* (a single EST from the reference Aliota et al. [2007]), *Gryllus bimaculatus* (a single unpublished cDNA), and *Mesobuthus gibbosus* (5 ESTs from an unpublished EST project). GenBank accession numbers are listed in the Electronic Supplement online.

A multiple sequence alignment (Electronic Supplement online) shows that the candidate sequences have 2 well-conserved domains located at the 5′ and the 3′ termini, whereas the intermediate portion appears to evolve rapidly and contains large insertions and deletions, see also figure 3. Overall, this organization conforms the observations for the known 7SK sequences (Gruber et al. 2008): the highest sequence conservation among the known 7SK snRNAs is also observed in the 5′ and 3′ hairpin regions.
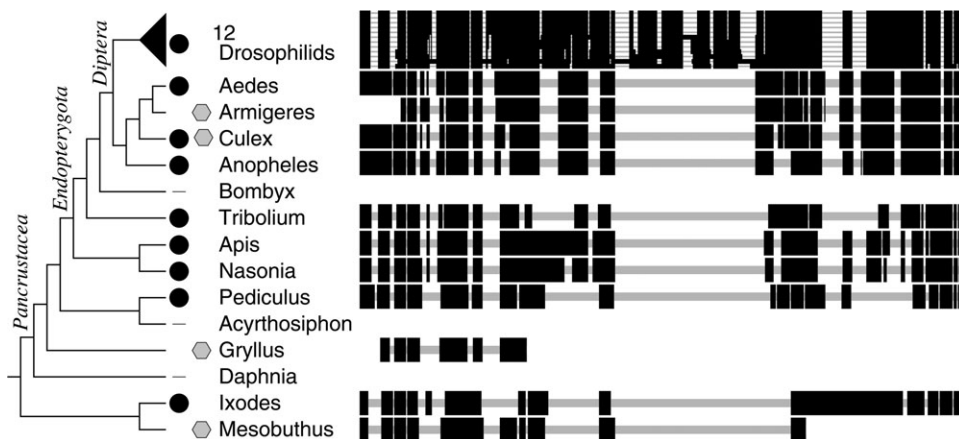
Fig. 3.—Phylogenetic distribution of 7SK candidate sequences. A bullet indicates a match in the genomic sequence, the hexagons for *Armigeres*, *Culex*, *Gryllus*, and *Mesobuthus* refer to partial ESTs. For complete genomic sequences, a sketch of the alignment structure highlights the large insertion domains in Pancrustacea and in Drosophilidae in particular. Aligned blocks are shown in black, gray bars indicate gaps in the alignment, and missing sequence data adjacent to EST regions appear white. The underlying tree is composed from the genome-wide near intron positions (Krauss et al. 2008), a phylogeny of mosquitoes (Harbach and Kitching 1998), and 2 recent studies of arthropod phylogeny (Cameron et al. 2004; Kjer 2004), for the relationships outside the Endopterygota.

Figure 4 demonstrates substantial similarities between the 7SK snRNAs reported in Gruber et al. (2008) and the candidate sequences discovered in this contribution. The domains with similar sequences are located in a similar structural context, see below.

Structural Characterization

We therefore constructed a structural consensus model of the arthropod sequences and compared this with the structural models derived in Gruber et al. (2008). Two distinct secondary structure elements are highly conserved throughout vertebrates (Egloff et al. 2006): a 5′-terminal hairpin structure that binds both HEXIM1 and P-TEFb and a 3′-terminal hairpin that interacts with P-TEFb only. A recent study (Krueger et al. 2008) revealed that 7SK snRNA is stably associated with LARP7, a close relative of La, which is associated with many nascent pol III transcripts, including 7SK snRNA (Hogg and Collins 2007). It is unknown, however, how LARP7 binds to 7SK. Interestingly, LARP7 has a well-known homologue in *D. melanogaster* (Krueger et al. [2008]), namely *mxc* (multi

sexcomb reduced), a member of the "polycomb" group regulating gene expression during development (Rajasekhar and Begemann 2007).

Structural alignments of identified candidate sequences based on previously published 7SK sequence data (Gruber et al. 2008) were generated for both the 5′ region and the terminal 3′ region. Independent models were generated for the 5′ regions of Drosophilidae, Neoptera, and all Arthropoda, respectively. Using also the previously published sequence data on 7SK (Gruber et al. 2008), we furthermore constructed combined models for Arthropoda + Lophotrochozoa and Vertebrata + Cephalochordata. Their combination was then used to suggest a consensus model.

Overall, the secondary structure of the 5′-stem region of arthropods is quite similar to its vertebrate and lophotrochozoan counterpart. Although the lower part of the stem-loop structure is very similar in all know sequences, the closing hairpin loop varies considerably in size and base composition. In drosophilid flies, this stem is extended by a helical element consisting of 5 bp (supported by several compensatory mutations), whereas otherwise the terminal loop consists of 8–15 nt. The hairpin loop is closed by

Fig. 4.—Sequence logos (Schneider and Stephens 1990) representing an alignment of the 7SK consensus sequences reported in Gruber et al. (2008) (upper sequences) and the consensus of the arthropod candidates (lower sequences). The logos were computed using aln2pattern (Mosig et al. 2007) separately for the 2 groups of sequences. Common well-conserved motifs are shaded.
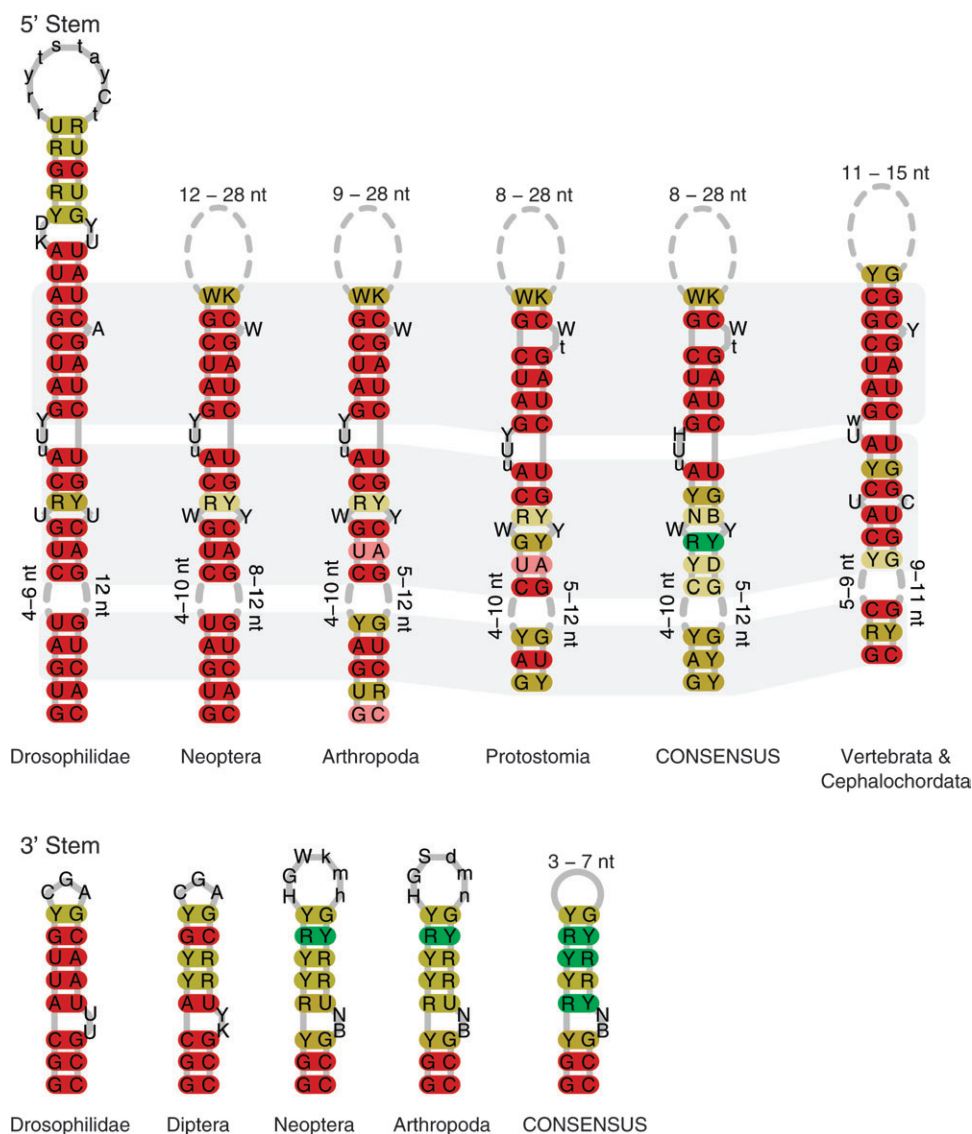
FIG. 5.—Comparison of structural motifs of 7SK snRNAs. Conserved nucleotides in stems are shown in red, whereas ocher (green) indicates 2 (3) different supporting compensatory mutations. Pale colors indicate that a base pair cannot be formed by all the sequences. Lower case letters imply a deletion in some sequences. The variable size regions close to the hairpin loops, which have no clear consensus folds, are drawn as dashed ellipses. Correspondences of helices are highlighted by a gray background.

a stem that is highly conserved in both sequence and structure. This stacked region is only interspersed by a positionally conserved bulge loop. The outer part of this stem comprises the GAUC–GAUC motif enclosed by positionally conserved bulge loops. The functional importance of this motif is discussed in detail in Egloff et al. (2006). The consensus model shows that there exists only a structural, not a strong sequence constraint on the other elements of the 5′-stem region.

Both helices in the 3′-stem region are supported by many compensatory mutations. The position of the bulge loop as well as the position of the hairpin loop are highly conserved. Although both Vertebrata and Lophotrochozoa show a sequence constraint in the hairpin loop, this does not seem to be the case in arthropods. For Diptera, the hairpin loop is reduced to a minimal size of 3 nt. Based on the struc-

ture model for Deuterostomia and Lophotrochozoa suggested in Gruber et al. (2008) and the arthropod model derived here, we suggest a universal structural model of the 3′-terminal stem.

The sequence similarities, figure 4, the very similar structural organization of both the 5′ and the 3′ conserved domains, figure 5, and the fact that the *Drosophila* loci have the typical organization of a pol III transcript with a type 3 pol III promoter demonstrate beyond reasonable doubt that the *3R*(3.3M) locus **B** indeed harbors a 7SK homologue.

The conserved elements in figures 4 and 5 can in principle be used to construct sequence or sequence–structure patterns for further homology searches. Attempts to find a 7SK homologue in the shotgun traces of the *D. pulex* genome remained unsuccessful, however, with both fragrep and rnabob.
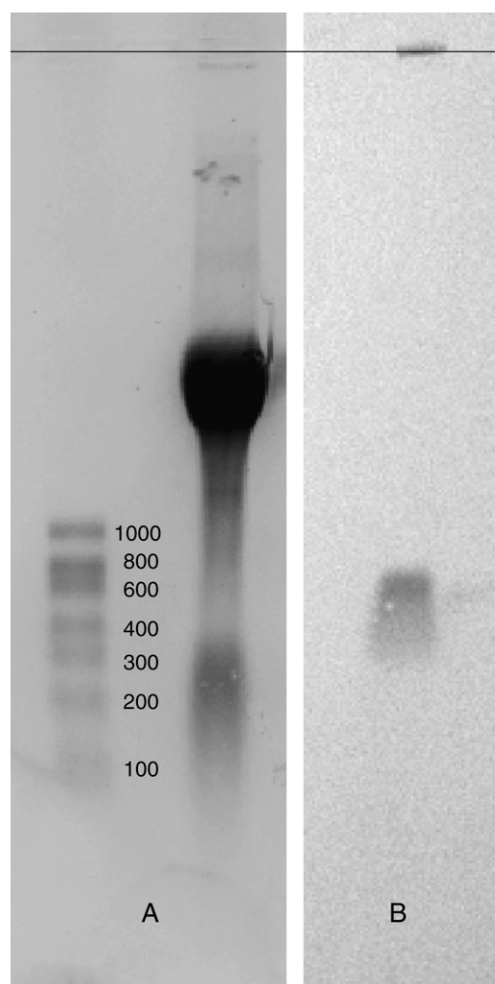
FIG. 6.—(*A*) Electrophoretic separation of total RNA from *Drosophila melanogaster* adults in a 2% agarose gel (ethidium bromide stained). (*B*) Northern blot hybridized with the 7SK DNA probe of length 344 bp.

Expression in *D. melanogaster*

In order to verify the expression of the 7SK locus, for which a previous study had already reported a partial transcript (Yuan et al. 2003), we performed a standard northern blot experiment. We used a 344-bp probe located between position 35 and 379 within the 445-nt long predicted 7SK gene.

The DIG-labeled PCR fragment was hybridized to a blot of total RNA from flies, separated on an agarose gel. For the detection of the hybrids, we used alkaline phosphatase-labeled anti-DIG antibody for the reaction with NBT/BCIP as substrate which yields a purple reaction product.

Figure 6 shows the electrophoretic separation of the total RNA and the northern blot, which resulted in a clear single band. Comparison between the marker in the gel and the blot shows that detected transcript appears somewhat larger than the predicted 7SK gene.

Discussion

Homology search for ncRNAs has turned out to be a surprisingly hard problem in bioinformatics. Standard methods of homology search often fail due to large varia-tions in sequence length and oftentimes extremely poor sequence conservation, see, for example, (Mosig et al. 2007; Gruber et al. 2008; Xie et al. 2008) for recent examples. Indeed, the arthropod 7SK RNAs reported in this contribution were not discovered by straightforward search but rather by an indirect strategy that uses the typical promoter structure of 7SKs (Bannister et al. 2007), experimental evidence for pol III transcripts in *D. melanogaster* (Isogai et al. 2006), sequence conservation (*Drosophila* 12 Genomes Consortium 2007), and de novo prediction of evolutionarily conserved RNA secondary structure (Rose et al. 2007). Once the representative sequences in Drosophilidae were found, conventional Blast-based searches revealed additional homologs, which could then be used as starting point for pattern-based searches that resulted in 7SK sequences spanning most of the arthropod tree.

A detailed analysis of sequence motifs and the construction of RNA secondary models based on a combination of thermodynamic folding and sequence covariation demonstrates that our candidate sequences share key features, namely the 5′-stem and 3′-stem regions, with deuterostome and lophotrochozoan 7SK RNAs, demonstrating that we have indeed found the 7SK snRNA.

A search of EST and cDNA data revealed evidence for transcription of the 7SK locus in several species across the Arthropoda. We furthermore performed a northern blot to verify the 7SK in Drosophilidae directly. The resulting transcript is somewhat longer than expected. There is, however, an AT-rich repetitive region immediately downstream of the 7SK RNA which may be at least partially transcribed. Human 7SK ends are known to be heterogeneous (Sinha et al. 1998). Furthermore, an extension of pol III transcripts beyond a putative 3′ end inferred from homology search was for instance observed in mouse and rat vault RNAs (compared with most other mammalian vault RNAs) (Vilalta et al. 1994; Kickhoefer et al. 2003). The smear observed in the northern blot below the major signal might indicate the presence of a series of smaller transcripts due to earlier termination.

Our results demonstrate that a 7SK snRNA featuring 2 highly structured conserved domains was present already in the bilaterian ancestor. This suggests that also the function of the 7SK snRNA is evolutionary conserved despite a recent report that the inhibition of P-TEFb by the peptide Pgc is RNAse insensitive in primordial germ cells (Hanyu-Nakamura et al. 2008). The hypothesis of functional conservation is further supported by the observation that all major protein components of the human 7SK snRNP (P-TEFb, HEXIM, and LARP7) have homologs in *D. melanogaster* (P-TEFb, CG3508, and *mxc*, respectively). More generally, the presumably ancient origin of 7SK snRNA and the ubiquitous role of 6S RNA as transcriptional regulator in bacteria (Barrick et al. 2005) suggest that the recently uncovered variety of ncRNAs regulating the transcriptional machinery (Goodrich and Kugel 2006; Barrandon et al. 2008) may also be evolutionary ancient (Lu et al. 2008).

Supplementary Material

An electronic supplement is located online at http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/

08-008/ which compiled sequence data, primers, alignments in machine-readable form, and fragrep2 search patterns. Supplementary materials are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals. org/).

Acknowledgments

Literature Cited

Aliota MT, Fuchs JF, Mayhew GF, Chen CC, Christensen BM. 2007. Mosquito transcriptome changes and filarial worm resistance in *Armigeres subalbatus*. BMC Genomics. 8:463.

Bannister SC, Wise TG, Cahill DM, Doran TJ. 2007. Comparison of the chicken 7SK and U6 RNA polymerase III promoters for hairpin RNA expression. BMC Biotech. 7:79.

Barrandon C, Spiluttini B, Bensaude O. 2008. Non-coding RNAs regulating the transcriptional machinery. Biol Cell. 100:83–95.

Barrick JE, Sudarsan N, Weinberg Z, Ruzzo WL, Breaker RR. 2005. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. RNA. 11:774–784.

Blazek D, Barboric M, Kohoutek J, Oven I, Peterlin BM. 2005. Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. Nucleic Acids Res. 33:7000–7010.

Cameron SL, Miller KB, D'Haese CA, Whiting MF, Barker SC. 2004. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Arthropoda). Cladistics. 20:534–557.

Chomczynski P, Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 162:156–159.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 450:203–218.

Egloff S, Van Herreweghe E, Kiss T. 2006. Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. Mol Cell Biol. 26:630–642.

Goodrich JA, Kugel JF. 2006. Non-coding-RNA regulators of RNA polymerase II transcription. Nat Rev Mol Cell Biol. 7:612–616.

Griffiths-Jones S. 2005. RALEE—RNA ALignment editor in Emacs. Bioinformatics. 21:257–259.

Gruber AR, Koper-Emde D, Marz M, Tafer H, Bernhart S, Obernosterer G, Mosig A, Hofacker IL, Stadler PF, Benecke B-J. 2008. Invertebrate 7SK snRNAs. J Mol Evol. 107–115:66.

Gürsoy H-C, Koper D, Benecke B-J. 2000. The vertebrate 7SK RNA separates hagfish (*Myxine glutinosa*) and lamprey (*Lampetra fluviatilis*). J Mol Evol. 50:456–464.

Hanyu-Nakamura K, Sonobe-Nojima H, Tanigawa A, Lasko P, Nakamura A. 2008. *Drosophila* Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature. 451:730–733.

Harbach RE, Kitching IJ. 1998. Phylogeny and classification of the Culicidae (Diptera). Syst Entomol. 23:327–370.

He WJ, Chen R, Yang Z, Zhou Q. 2006. Regulation of two key nuclear enzymatic activities by the 7SK small nuclear RNA. Cold Spring Harb Symp Quant Biol. 71:301–311.

Hernandez N. 2001. Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. J Biol Chem. 276:26733–26736.

Hernandez G Jr, Valafar F, Stumph WE. 2007. 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.

Hofacker IL, Fekete M, Stadler PF. 2002. Secondary structure prediction for aligned RNA sequences. J Mol Biol. 319:1059–1066.

Hogg JR, Collins K. 2007. RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. RNA. 13:868–880.

Huang ZP, Zhou H, He HL, Chen CL, Liang D, Qu LH. 2005. Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. RNA. 11:1303–1316.

Isogai Y, Takada S, Tjian R, Keles S. 2007. Novel TRF1/BRF target genes revealed by genome-wide analysis of Drosophila pol III transcription. EMBO J. 26:79–89.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Kickhoefer VA, Emre N, Stephen AG, Poderycki MJ, Rome LH. 2003. Identification of conserved vault RNA expression elements and a non-expressed mouse vault RNA gene. Gene. 309:65–70.

Kjer KM. 2004. Aligned 18S and insect phylogeny. Syst Biol. 53:506–514.

Krauss V, Thümmler C, Georgi F, Lehmann J, Stadler PF, Eisenhardt C. 2008. Near intron positions are reliable phylogenetic markers: an application to Holometabolous insects. Mol Biol Evol. doi:10.1093/molbev/msn013

Krueger BJ, Jeronimo C, Roy BB, et al. (11 co-authors). 2008. LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res. 36:2219–2229.

Krüger W, Benecke BJ. 1987. Structural and functional analysis of a human 7 S K RNA gene. J Mol Biol. 195:31–41.

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI. 2008. The birth and death of microRNA genes in *Drosophila*. Nat Genet. 40:351–355.

Michels AA, Fraldi Q, Li A, Adamson TE, et al. (11 co-authors). 2004. Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. EMBO J. 23:2608–2619.

Moqtaderi Z, Struhl K. 2004. Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. Mol Cell Biol. 24:4118–4127.

Morgenstern B. 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics. 15:211–218.

Morgenstern B, Prohaska SJ, Pohler D, Stadler PF. 2006. Multiple sequence alignment with user-defined anchor points. Algo Mol Biol. 1:6.

Mosig A, Chen JL, Stadler PF. 2007. Homology search with fragmented nucleic acid sequence patterns. In: Giancarlo R, Hannenhalli S, editors. Algorithms in Bioinformatics (WABI 2007). Vol. 4645. *Lecture Notes in Computer Science*. Berlin (Germany): Springer Verlag. p. 335–345.

Mount SM, Gotea V, Lin C-F, Hernandez K, Makalowski W. 2007. Spliceosomal small nuclear RNA genes in 11 insect genomes. RNA. 13:5–14.

Murphy S, Di Liegro C, Melli M. 1987. The *in vitro* transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell. 51:81–87.

Nguyen VT, Kiss T, Michels AA, Bensaude O. 2001. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature. 414:322–325.

Peterlin BM, Price DH. 2006. Controlling the elongation phase of transcription with P-TEFb. Mol Cell. 23:297–305.

Piccinelli P, Rosenblad MA, Samuelsson T. 2005. Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes. Nucleic Acids Res. 33:4485–4495.

Rajasekhar VK, Begemann M. 2007. Concise review: roles of polycomb group proteins in development and disease: a stem cell perspective. Stem Cells. 25:2498–2510.

Rose DR, Hackermüller J, Washietl S, Findeiß S, Reiche K, Hertel J, Stadler PF, Prohaska SJ. 2007. Computational RNomics of Drosophilids. BMC Genomics. 8:406.

Sambrook J, Russell D. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18:6097–6100.

Sinha KM, Gu J, Chen Y, Reddy R. 1998. Adenylation of small RNAs in human cells: development of a cell-free system for accurate a adenylation on the 3′-end of human signal recognition particle RNA. J Biol Chem. 273:6853–6859.

Thompson JD, Higgs DG, Gibson TJ. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Vilalta A, Kickhoefer VA, Rome LH, Johnson DL. 1994. The rat vault RNA gene contains a unique RNA polymerase III promoter composed of both external and internal elements that function synergistically. J Biol Chem. 269:29752–29759.

Will S, Missal K, Hofacker IL, Stadler PF, Backofen R. 2007. Inferring non-coding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput Biol. 3:e65.

Woodhams MD, Stadler PF, Penny D, Collins LJ. 2007. RNase MRP and the RNA processing cascade in the eukaryotic ancestor. BMC Evol Biol. 7(Suppl 1): S13.

Xie M, Mosig A, Qi X, Li Y, Stadler PF, Chen JJ-L. 2008. Size variation and structural conservation of vertebrate telomerase RNA. J Biol Chem. 283:2049–2059.

Yang Z, Zhu Q, Luo K, Zhou Q. 2001. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature. 414:317–322.

Yuan G, Klämbt C, Bachellerie JP, Brosius J, Hüttenhofer A. 2003. RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. Nucleic Acids Res. 31:2495–2507.