

The ribonuclease A superfamily of mammals and birds: identifying new members and tracing evolutionary histories[☆]

Soochin Cho^a, Jaap J. Beintema^b, Jianzhi Zhang^{a,*}

^a*Department of Ecology and Evolutionary Biology, University of Michigan, 3003 Natural Science Building, 830 North University Avenue, Ann Arbor, MI 48109, USA*

^b*Department of Biochemistry, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands*

Received 2 July 2004; accepted 13 October 2004

Available online 7 December 2004

Abstract

The RNase A superfamily has been important in biochemical, structural, and evolutionary studies and is believed to be the sole vertebrate-specific enzyme family. To understand the origin and diversification of the superfamily, we here determine its entire repertoire in the sequenced genomes of human, mouse, rat, and chicken. We report a previously unnoticed gene cluster in mouse chromosome 10 and a number of new genes, including mammalian RNases 11–13, which are close relatives of the recently identified RNases 9 and 10. Gene expression data imply male-reproductive functions for RNases 9–13, although their sequences suggest the lack of ribonucleolytic activities. In contrast to the presence of 13–20 functional genes in mammals, chicken has only 3 RNase genes, which are evolutionarily close to mammalian RNase 5, like other nonmammalian RNases. This and other evidence suggests that the RNase A superfamily originated from an RNase 5-like gene and expanded in mammals. Together with the fact that multiple lineages of the superfamily, including RNases 2, 3, 5, and 7, have antipathogenic activities, we suggest that the superfamily started off as a host-defense mechanism in vertebrates. Consistent with this hypothesis, all members of the superfamily exhibit high rates of amino acid substitution as is commonly observed in immunity genes.

© 2004 Elsevier Inc. All rights reserved.

Keywords: RNase; Gene family; Eosinophil-associated ribonuclease; Angiogenin; Positive selection; Gene sorting

For over half a century since the bovine pancreatic ribonuclease (also known as RNase A) was first purified [1], the RNase A superfamily, of which RNase A is the prototype, has been one of the most intensively studied protein superfamilies in biochemistry, structural biology, enzymology, and molecular evolution [2–8]. This superfamily constitutes a group of homologous proteins isolated from many vertebrates, but it has not been found outside vertebrates [4,5]. The ribonucleolytic activity originated multiple times during evolution and nonvertebrate organisms possess ribonucleases that are nonhomologous to vertebrate RNase A. In fact, the RNase A superfamily is

believed to be the sole enzyme family that is vertebrate-specific [9], although this hypothesis requires further scrutiny. For simplicity, we will use RNases to refer to RNase A superfamily members unless otherwise noted.

RNase proteins are typically composed of a signal peptide of about 25 amino acids and a mature peptide of about 130 amino acids. They have 3 catalytic residues (1 lysine and 2 histidines) at proper positions and 6 to 8 cysteines that form three to four disulfide bonds. Except for these conserved residues, RNases are quite divergent, with sequence identities varying from 20 to nearly 100%. They also exhibit diverse expression patterns and possess various catalytic activities against specific RNA substrates. A wide variety of physiological functions are known for RNases, including degradation of dietary RNAs in the digestive gut, angiogenesis, and innate immunity (reviewed in [4,5]).

[☆] Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AY665804–AY665837.

* Corresponding author. Fax: +1 (734) 763 0544.

E-mail address: jianzhi@umich.edu (J. Zhang).

Among nonmammalian vertebrates, RNases have been purified and sequenced from chicken, turtle, iguana, and frogs (see [5] and references therein). Among mammals, RNases are most extensively investigated in primates, rodents, and ruminants. The RNase A superfamily is extremely dynamic, with high rates of gene duplication and gene loss, resulting in variable numbers of genes in different species. Until recently, there were eight known human RNases, including pancreatic ribonuclease (RNase 1), eosinophil-derived neurotoxin (EDN, or RNase 2), eosinophil-cationic protein (ECP, or RNase 3), RNase 4, angiogenin (RNase 5), RNase 6 (or k6), RNase 7 [10,11], and RNase 8 [12]. Recently, Penttinen et al. reported the *in silico* discovery of mouse RNases 9 and 10 [13] and Castella et al. purified and identified RNase 10 (named Train A in [14]) from porcine epididymis [14]. We were able to locate genes encoding these two new RNases in the human genome, thus adding them to the growing inventory of the human RNase A superfamily. The completion or near completion of the human, mouse, rat, and chicken genome projects has opened the door for comprehensive identifications and comparisons of all members of the RNase A superfamily. We here describe the complete catalogs of the RNase A superfamily in the above four species, including a number of newly identified superfamily members. We also report a previously unknown RNase gene cluster on mouse chromosome 10. Our phylogenetic analyses provide several novel insights on the origin and physiological roles of this highly diversified protein superfamily of vertebrates.

Results and discussion

Chromosomal organizations of RNase A genes in human, mouse, rat, and chicken

From whole-genome database searches based on TBLASTN, we identified all RNase genes and pseudogenes from the genome sequences of human, mouse, rat, and chicken. We also conducted BLASTN-based nucleotide sequence searches to identify pseudogenes that had not been found using TBLASTN. The entire catalogs of RNase genes identified in this study are listed in Supplemental Tables 1–4, and their chromosomal locations are drawn to scale in Fig. 1. For pseudogenes, start and stop codons were determined according to their closest functional homologs. Below we first provide a general overview of the RNase genes identified in the four genomes and then present their novel features in a more detailed and specific manner.

Human RNase genes

Zhang et al. [12] reported that human RNase 1–8 genes are located on chromosome 14q11.2, forming a cluster of ~368 kb. This is confirmed by the present analysis (Fig. 1, Supplemental Table 1). Briefly, in order from the centromere to the telomere, angiogenin (RNase 5), RNase 4,

RNase 6, RNase 1, ECP (RNase 3), EDN pseudogene, EDN (RNase 2), RNase 7, and RNase 8 are separated from each other by a 6- to 90-kb interval. RNase 9 and RNase 10, two recently reported members of this superfamily, are also found within this cluster, located ~137 and ~183 kb away from angiogenin toward the centromere, respectively. In addition to these previously reported members, we identified three novel open reading frames (ORFs) within this cluster, which share a number of common features with other RNases (see below) and are thus named RNases 11–13. RNase 11 and RNase 12 are located between RNase 9 and angiogenin and are separated by only ~6 kb from each other. RNase 13 is ~9 kb away from RNase 7 toward the centromere and has a transcriptional direction opposite to that of RNases 7 and 8. Our extensive searches of the human genome sequence yielded no other ORFs with significant sequence similarity to previously reported RNase genes. Therefore, it is likely that all human RNase A superfamily members have been identified.

Rat RNase genes

As in human, all rat RNase genes are located on one chromosome (15p14), forming a single cluster of ~370 kb, which is shorter than the human cluster (~550 kb) (Fig. 1, Supplemental Table 2). Using the human–rat homologous chromosome map (<http://www.genboree.org>), we found that this region in the rat genome is syntenic to the human RNase gene cluster. Despite its small size, the rat cluster contains essentially every member of the superfamily in the same order and transcriptional direction as in human, with two exceptions. First, lineages of RNase 1 (RNase1 β , RNase1 γ , and RNase1 δ), eosinophil-associated RNase (EAR) (R15-17, ECP, R-pseudogene, and Ear3), and angiogenin (Ang1 and Ang2) have undergone expansion in the rat [15–17]. The duplications of rat RNase 1 genes were previously reported before the availability of the genome sequence [15] and the expansions of rat EARs and angiogenins will be discussed in detail later in this report. Second, orthologs of human RNases 7 and 8 are not present in the rat genome. This leaves RNase 13 at the centromeric boundary of the cluster. Intriguingly, we could not locate many of the rat RNase genes previously reported [16–18], including R1, R2, R4, R5, R7, R8, R12, and R14 (Supplemental Table 5).

Mouse RNase genes

Unlike the human and rat genomes, in which RNase genes are located in a single cluster, we identified two RNase gene clusters on mouse chromosome 14qB–qC1 (“cluster A”) and 10qB1 (“cluster B”) (Fig. 1, Supplemental Table 3). Cluster A is syntenic to the human and rat clusters and is essentially identical to the rat cluster in gene content and order except for massive expansions of the EAR and angiogenin gene repertoires (see below). The chromosomal region homologous to cluster B is not present in human or rat, suggesting that this region emerged in mouse

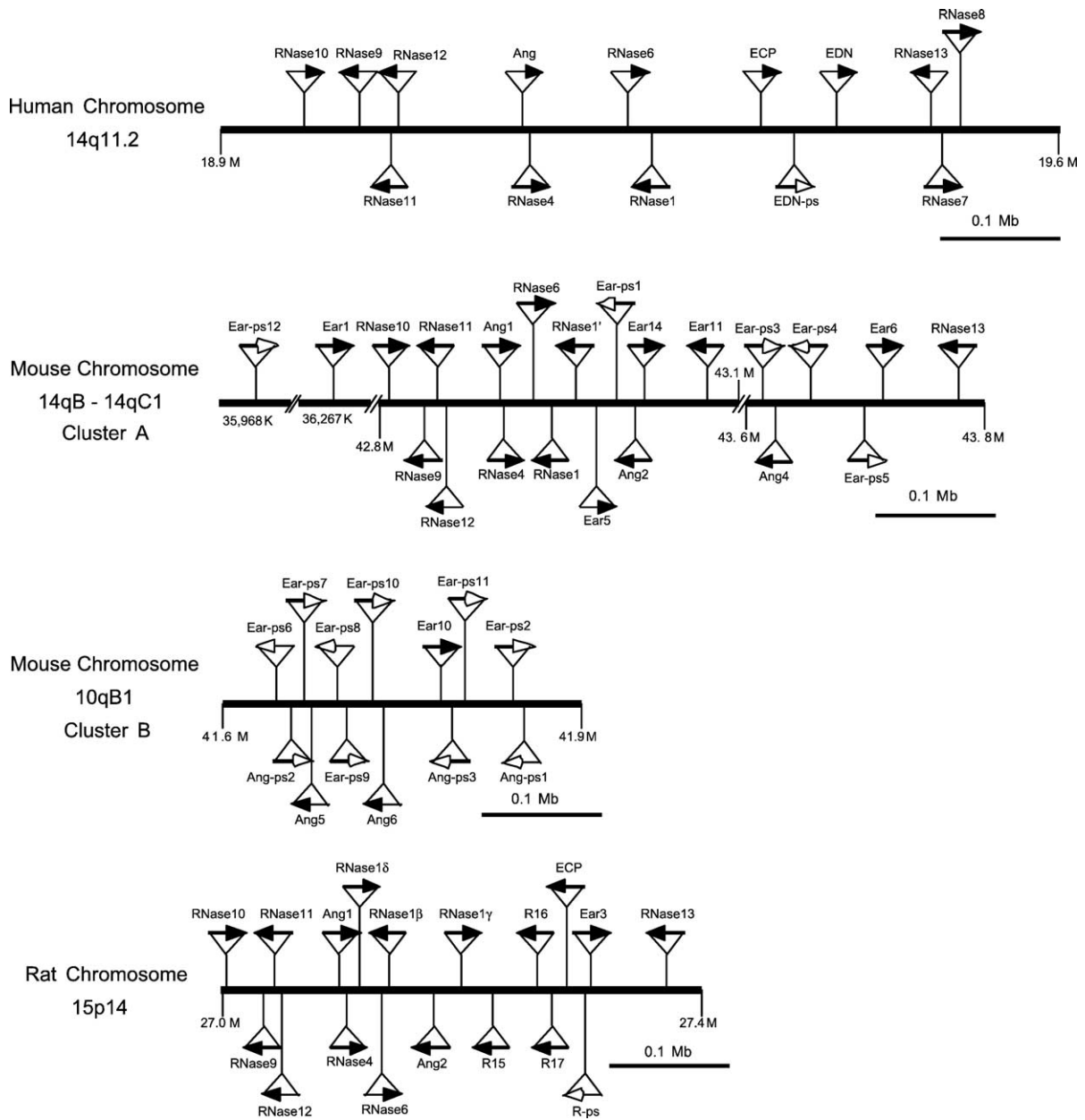


Fig. 1. Chromosomal locations of all human, mouse, and rat RNase genes. Gene locations are drawn to scale. Arrows indicate transcriptional directions. Filled arrows show functional genes and open arrows show pseudogenes. In human chromosome 14, RNase 10 is closer to the centromere, and RNase 8 is closer to the telomere. In mouse chromosome 14, Ear-ps12 is closer to the centromere, and RNase 13 is closer to the telomere. In mouse chromosome 10, Ear-ps6 is closer to the centromere, and Ang-ps1 is closer to the telomere. In rat chromosome 15, RNase 13 is closer to the centromere, and RNase 10 is closer to the telomere.

after the mouse–rat separation. Cluster B contains only genes and pseudogenes that belong to EAR and angiogenin families. Mouse cluster A (~7.9 Mb) is much longer than the human and rat clusters mostly due to the huge interval (~6.5 Mb) that separates Ear-ps12 and Ear1 from the other genes. The distance between RNase 10 and RNase 13 in mouse (~1.0 Mb) is also much greater than that in human (~550 kb) or rat (~400 kb). Mouse cluster B spans only 250 kb but contains 3 genes and 10 pseudogenes. We found two cases in which more than two sequences in different genomic loci are 100% identical (Supplemental Table 3). First, mouse Ear-ps6, ps11, and ps12 are identical to one another. Ear-ps6

and ps11 are more than 150 kb apart, with many other Ear genes interposed between them, and Ear-ps12 is on a separate chromosome. Therefore, it is unlikely that their sequences have been homogenized by gene conversion. Second, two copies of the RNase 1 gene were found on chromosome 14, separated by ~20 kb from each other. Tentatively, we named these two sequences RNase 1 and RNase 1'. These two identical sequences were generated by a tandem repeat of an ~20-kb sequence. At this point, we cannot tell whether they resulted from a very recent duplication event, a sequence assembly error, or gene conversion. As RNase 1 duplicated several times in a rat

ancestor, we may tentatively hypothesize that RNase 1 was already duplicated in the ancestor of mouse and rat, with the duplicates remaining identical in mouse by gene conversion. As in rat, we could not identify many previously reported RNase genes [16,19–27] from the assembled genome sequence, including Ear-2, 3, 4, 7, 8, 9, 12, 13, and Ang-3 (Supplemental Table 5).

Chicken RNase genes

We identified three chicken ORFs that show significant sequence similarity to known RNase genes (Supplemental Table 4). The one on chromosome 6 is identical to RNase CL2, a previously identified RNase from chicken liver, except for the presence of four additional amino acid residues at the C-terminus [28]. The other two ORFs are on chromosome 4 and are identical to the previously identified RNases from chicken bone marrow (RSFR [29]) and myelomonocytic cells (Clone 462 [30]), respectively. No novel chicken RNase genes were found. RSFR and Clone 462 are separated by only 10 kb and have the same transcriptional direction. They are also more similar to each other (85% identical in protein sequence) than either is to CL2 (~40% identical).

Missing RNase genes in the assembled mouse and rat genome sequences

As mentioned before, many previously reported rodent EAR genes and one mouse angiogenin gene (Ang3) are missing from the whole genome sequence assemblies (Supplemental Table 5), for which there can be at least four possible explanations. First, the assembled mouse and rat genome sequences are not complete and some of the previously reported genes can fall into the unsequenced regions. We could identify none of those missing genes from HTGS (High Throughput Genomic Sequence) entries either. Second, high levels of sequence similarities among EARs and among angiogenins might have caused errors in the sequence identification or contig assembly processes. For example, the nucleotide sequence of mouse Ear7, which is one of those missing genes, is 98.6% identical to mouse Ear6. Third, it could be due to genetic variations among different strains used. Finally, some experimental errors could have happened, such as sequencing errors or mutations incorporated during polymerase chain reaction and/or reverse transcription reactions. Notably, some of the genes were isolated using cDNA library screening or PCR amplification of genomic DNA, as is summarized in [25]. Furthermore, in some previous studies, primer sequences were not removed from the sequences submitted to GenBank, introducing additional errors [16]. Notwithstanding, independent experimental evidence has been provided for the existence of some of the missing genes listed in Supplemental Table 5. For example, mouse Ear-2 and Ear-7 have been identified by additional independent methods such as sequencing genomic λ clones and directly sequencing proteins purified from the eosinophil secondary granules [25].

New members of the RNase A superfamily

One of the major findings of this study is three novel RNase genes (RNases 11–13) in mammals, in addition to the recently identified RNase 9 and RNase 10. Fig. 2 shows the alignment of human RNases 1–13. Although the amino acid sequences of RNases 9–13 are only 15–30% identical to the canonical RNases (RNases 1–8) (Supplemental Table 6), several characteristics suggest that all these proteins share a common ancestry. First, the RNase genes are all closely linked on the chromosome (Fig. 1), an indication of origin by tandem gene duplication. Second, for all the RNase genes, the entire ORF is contained within a single exon. Third, most of the 6 to 8 cysteine residues and also several other residues that are important for folding and structure in canonical RNases are conserved in RNases 9–13 (Fig. 2 and Supplemental Fig. 1). Fourth, the ORFs of these proteins indicate that they have a signal peptide at the N-terminus and that the mature peptides have a two-domain structure with different N-terminal domains (with insertions of 40–50 residues in RNases 9, 10, and 11), but similar C-terminal ones of about 100 amino acid residues with four conserved disulfide bonds in most of them (the S-protein core).

In addition to the above shared features, canonical RNases invariably have the catalytic triad (H₁₂-K₄₁-His₁₁₉, numbers according to human RNase 1) that is required for ribonuclease activity [3] and the signature motif of CKXXN_TF, which includes the catalytic lysine residue K₄₁. In RNases 9 and 10, however, these two specific features are not conserved [13,14]. RNases 11–13 are also quite variable at these sites (Table 1).

In active RNases, His₁₁₉ is located in a conserved VPVH sequence in the C-terminus of the molecule. As is evident from inspection of Fig. 2 and Supplemental Fig. 1 there are no sequence similarities in the C-terminus between active RNase 1 and RNases 9–13 or among the latter. Few of the RNases 9–13 have a histidine in this region and it is difficult to judge whether the histidine is the catalytic histidine. No homology is apparent in the sequences preceding the first conserved cysteine residue, including His₁₂, suggesting that the S-peptide structure of active RNases is not present in RNases 9–13. Only in RNase 13 may there be such a structure, with His₁₂ present in mouse and rat, but not human. In the signature motif CKXXN_TF, the structurally important C and F are conserved in the majority of sequences. But the active site Lys₄₁ is replaced (in a few cases by Arg) or shifted by one position. Thr₄₅, which is involved in the active site, is not conserved either. These sequence features suggest that RNases 9–13 may not possess ribonucleolytic activities. But the conservation between orthologous sequences of the three species suggests functionality of the proteins.

Among members of the superfamily, RNase 4 shows the highest sequence similarity among orthologs in human, mouse, and rat [31], and EARs are most divergent [32], as is

```

HsRNase1 ---MALEKSL VRLLLLVLLI LVLGWVQPSL GKE-----
HsRNase2 --MVPKLFYS QIC...L.G. .AVEGSLHVK PPQF-----
HsRNase3 --MVPKLFYS QIC...L.G. MGVEGSLHAR PPQF-----
HsRNase4 ---...QRTH SL...L.T. .G...L...Y .QD-----
HsRNase5 -----MVM GLGV..LVFV .G..LTP.T. AQDN-----
HsRNase6 -----MVLG FP...L.V. WGPVCPHAW P.RL-----
HsRNase7 --MAP.RAGFC PL...L.G. W.AEIPVSAK P.GM-----
HsRNase8 --MAP.RAGCC PL...L.G. W.AEVLVRAK P.DM-----
HsRNase9 --MRT.ITTH PLP...LPQQ .LQLVQFQEV DTFD-----
HsRNase10 MKNLNVQIFF ML.M..LGLG MG..LGLHMA TAVLEESDQP LNEFWSSDSQ DKAEATEEGD GTQTTETLVL S-NKEVVQPG WPEDPILGED
HsRNase11 METFP.LLLS LG.V.AEASE STMKIIEEF TD.E-----
HsRNase12 -----MIIMV IIF.V.LFWE NEV-----
HsRNase13 ---.PAVTR LLF.Q...GP TLVMDIKMQI .SRN-----

```



```

HsRNase1 -----SR AKKFRQHM DSSPSSST --YC-NQMR R--RNMTQGR CKPVNTFVHE PLVDVQVCF QEK-VTCNG --QGNCYKSN 180
HsRNase2 -----TW .QW.EF..IN MT.QQ----- --TNA.Q V--I.NY.R. .NQ...LLT TFAN.V...G NPN-M...PSN KTRK..HH.G
HsRNase3 -----T. .QW.AI..IS LNPPR----- --TIA.A A--I.NYRW. .NQ...LRT TFAN.V...G NQS-IR.PHN RTLN..HR.R
HsRNase4 -----GM YQR.L...VH PEETGG.DR----- --L.L.Q --K..LYH .RF...I.. DIWNIRSI.S TTN-IQ.... --KM..HEG-
HsRNase5 -----YTH.LT..Y AKPQGRDR----- --ESI.. --GL.S-P .DI...I.G NKRSIKAI.E NKN-GNPHRE --NLRIS..-
HsRNase6 -----TK .HW.EI..IQ PSPLQ----- --RA.S G--I.NYQH .HQ...L.D SFQN.AA.D LLS-IV...R --RH..HQ.S
HsRNase7 -----TS SQW.KI..Q PSP-Q----- --A..SA.K N--I.KHTK. .DL...L.. .FSS.AAT.Q TP..IA.... --DK..HQ.H
HsRNase8 -----TS SQW.KT..VQ PSP-Q----- --A..SA.S I--I.KYTE. .DL...L.. .FSS.AIT.Q TPN-IA...S --CK..HQ.H
HsRNase9 -----PGMPLN.IE -----HEIM G-KNVVYKH. WVAEHY.LLM QYDEL.KI.Y NRF..P... --IRK.NR.K
HsRNase10 EVGGNKMLRA SAL..SNKDY LRLDQFDR-- --E..D..A HK-MKEPSQS .IAQYA.I.. D.NT.KA..N SPV-IA.ELK --G.K.H..S
HsRNase11 -----MS.T LLT.RSL.YN DPKGN..GND K-E.C.D.TV WRKVSEAN.S .WS.N.IRS STEVMRR.HR APS---.FV QNP.ISSCES
HsRNase12 -----AV MSTLEHL.V YPQNDVPVPA R---.H.II QRVIREPDHT ..KEHV.I.. RPRKINGI.I SP.K.A.Q.L S-AIF.FQ.E
HsRNase13 -----YTLSTI. YPRVNYPRGF RG...GL.S YMRGK.QNSD .PKIHYVI.A .WKAT.KF.K YSD-SF.E.Y N--EY.TLTO

```



```

HsRNase1 SSMHITDCLR TNGS--RYPN ↓ CAYRTSPKER HIIVACEG-- ----- --SPYVPHVF DASVEDST-- --- 253
HsRNase2 .QVPLIH.N .TP.PQNIS. .R.AQT.ANM FY...DNR- ----DQRRDP PQY.V...L .RII-----
HsRNase3 FRVPLIH.D I.PGAQNIS. .R.ADR.GR. FYV...DNR- ----DPR-DS PRY.V...L .TTI-----
HsRNase4 -VVKV...D .GS...A.. .R..AIAS. RVVI...D- ----N.Q...L .G-----
HsRNase5 -.FQV.T.K.HG. ---PW.P .Q..ATAGF. NVV...N- ----G-L...L .Q.IFRFP--
HsRNase6 KPVNM...S. ---K.K.Q .R.SAAQYK FF...DP- ----PQKSDP P-YKL...L .SIL-----
HsRNase7 GAVSL.M.K. .S. ---KH.. .R.KEKRONK SYV...KP- ----PQKDS QQFHL...L .RVL-----
HsRNase8 GP.SL.MGE. .S. ---K... .R.KEKHLNT PY...DP- ----PQGGDP G-Y.L...L .KV-----
HsRNase9 GLVEGVY.N .EAF--EI.A .K.ESLYRKG YVLIT.SQON EMQKRIPIHTI NDLVEP.E.R SFLS..GVFV ISP
HsRNase10 RPPDL.L.E. SQPD-QVT.. .N.L..VIKK . .IT.ND- ----MKRQL.TGQ -----
HsRNase11 LELEN.V.QF .T.K--QF.R .Q.HSVTSLE K.LTVLT.-- ----HSLMSWLW CG.KL-----
HsRNase12 TKFKM.V.Q. IE.T--.A .R.HY..T.G FVL.T.DD-- ----LR.DS. LGY.K-----
HsRNase13 D.LP..V.S. SHQQ--PPTS .Y.NSTLTNQ KLYLL.SR-- ----KYEAD.IGI AGLYSGI---

```

Fig. 2. Amino acid sequence alignment of human RNases 1–13. Dashes show alignment gaps and dots represent the same amino acids as in human RNase 1. The eight structural cysteines are indicated by arrows, whereas the three catalytic residues are indicated by “*”. The conserved CKXXNTEF motif is highlighted in a box. The first amino acid residue of the mature part is underlined. For RNases 9–13, we used the SignalP 3.0 server (<http://www.cbs.dtu.dk/services/SignalP>) to determine the locations of signal peptide cleavage sites. For the other genes, we followed Zhang et al. [12].

shown in Supplementary Fig. 2. Interestingly, we found that RNases 9 and 11 are as divergent as EARs, while RNases 10, 12, and 13 are moderately divergent (Supplemental Fig. 2). Most noticeably, human RNase 9 is only ~35% identical to its rodent orthologs, which is even lower than the percentage identities between human and rodent EARs.

What are the physiological functions of RNases 9–13? Mouse RNases 9 and 10 are exclusively expressed in the epididymis [13,14] and pig RNase 10 is the most abundant compound secreted in the anterior part of the epididymis, where sperm maturation and activation take place [14]. Interestingly, the isoelectric point (pI) of RNases 9 and 10 is lower than that for any other mammalian RNase (Supplemental Table 7). The relatively high pI is believed to be important for the ribonuclease activity, conferring affinity of the enzyme to its negatively charged RNA substrate [3]. Though not sufficient, the positive charges in antibacterial EARs and angiogenins might also be critical for the disruption of bacterial cell membranes, which are negatively charged [26,33–35]. Therefore, RNases 9 and 10 may have acquired functions related to male reproduction, but may lack ribonuclease or antibacterial activities. The absence of the catalytic residues in these proteins also supports this hypothesis. The functions of RNases 11–13 are still a mystery. Except for the sequence similarity to other RNases

and the presence of the signal sequence, these proteins do not have any known functional motifs or domains. Their pI values are neither high nor low (Supplemental Table 7). One interesting clue comes from the gene expression profile database (<http://symatlas.gnf.org/SymAtlas/>) generated by

Table 1

Survey of the important amino acid residues in RNases 9–13 of human, mouse, and rat

Gene name	Species	H ₁₂	CK ₄₁ XXNTEF	VPVH ₁₁₉	No. conserved cysteines
RNase 9	Human	–	WVAEHY F	IPHT	7
	Mouse	–	CYPEHYI	IPDH	8
	Rat	–	CYPERYI	IPNY	8
RNase 10	Human	–	CIAQY A F	GQ–	8
	Mouse	–	CINQY T F	K–	8
	Rat	–	CINDY T F	K–	8
RNase 11	Human	+	CKWSN N F	LTGH	6
	Mouse	–	CQLGN D F	LASN	5
	Rat	–	CQLGN N F	LTSH	6
RNase 12	Human	+	CKKEH V F	RPDS	8
	Mouse	+	CRKVH V F ^a	GPVD	8
	Rat	+	CRKVH V F ^a	GPVN	8
RNase 13	Human	–	CPK I HYV ^a	–KY	8
	Mouse	+	CPK I HYV ^a	–KH	8
	Rat	+	CPK I HYV ^a	–KH	8

^a K₄₁ shifted downstream by one amino acid.

Su et al. using high-density oligonucleotide arrays [36]. In their data, human RNase 11 is expressed in the testis at an outstandingly high level, compared with other tissues. Thus RNases 9–11 are all specifically or highly expressed in male reproductive organs. Because RNases 9–13 are evolutionarily closely related (see below), it is possible that the entire lineage including RNases 9–13 might have specialized in male reproductive functions. In addition, some EST clones made from adult human brain (medulla) match RNase 11 and RNase 13, suggesting their potential functions in the brain. We are currently in the process of systematic examination of the expression patterns of RNases 11–13 using Northern hybridization.

Expansion of the rodent EAR family

To date, 13 mouse EAR genes (Ear1–13) and 2 pseudogenes (Ear-ps1 and Ear-ps2) have been reported, and 10 rat EAR genes (R1, R2, R4, R5, R7, R8, R12, R14, ECP, Ear3) but no pseudogenes are known [16,18,20,21,23,25,34,37]. Our analysis revealed 1 new EAR gene (Ear14) and 10 new pseudogenes (Ear-ps3–ps12) in mouse and 3 new genes (R15–17) and 1 pseudogene (R-ps) in rat (Supplementary Tables 2 and 3). In human, EAR genes (ECP, EDN, and EDN pseudogene) are located between RNase 1 and RNase 13 on chromosome 14. This chromosomal organization is conserved in mouse and rat, except that, in mouse, several EAR genes and pseudogenes are mapped to cluster B (chromosome 10), and Ear1 and Ear-ps12 are located upstream of RNase 10 (Fig. 1). As mentioned, 8 mouse EARs and 8 rat EARs that were previously reported could not be identified in the genome sequences. In other words, the genome sequences contain only 6 mouse and 5 rat EAR genes. Nevertheless, our results confirm previous observations that the EAR genes have duplicated multiple times during mouse and rat evolution.

To have a detailed look at the evolution of rodent EAR genes, a phylogenetic tree was made with all the EAR genes and pseudogenes in mouse and rat. The tree (Fig. 3) reconfirms the separate groupings of mouse and rat EARs [16,23]. Furthermore, our discovery of 10 new mouse pseudogenes and placement of them on the phylogenetic tree demonstrate that not only gene duplication, but also extensive gene inactivation has shaped the current repertoire of the mouse EARs. In particular, at least two inactivation events occurred very recently. Two mouse pseudogenes, Ear-ps2 and Ear-ps9, show 98.4% sequence identity to Ear10 and 98.5% identity to Ear1, respectively. These results suggest that gene inactivation is an ongoing process in mice. Interestingly, there is only 1 pseudogene in the rat genome, being dwarfed by the presence of more than 10 pseudogenes in mouse, which suggests that the rules governing the evolution of EAR genes might be different between these two rodents. One possible explanation for this difference is that the rat EAR diversification is a relatively recent event as indicated by short terminal

branches in Fig. 3 and Supplemental Fig. 4, thus too short a time has lapsed for many pseudogenization events to occur. Alternatively, it might be due to differences in the functional requirements of the EAR genes between the two rodent species. It is possible that some of the mouse EARs became dispensable, rendering them pseudogenes, whereas most of the rat EARs retained their functional importance for a longer time.

What is the driving force behind the diversification of the mouse and rat EARs? We tested the role of positive Darwinian selection by computing the synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site by comparing pairs of paralogous genes within individual species (Supplemental Fig. 3). We confined our analysis of those genes found in the genome sequences to minimize the unwanted contribution of experimental errors to the results. Between the mouse paralogous genes, all 15 pair-wise comparisons result in $d_N > d_S$, suggesting the operation of positive selection on these genes. Indeed, the average pair-wise d_N (0.215 ± 0.059) among these genes is significantly greater than the average pair-wise d_S (0.141 ± 0.047) ($p < 0.01$, one-tailed t test). However, we could not find strong evidence for positive selection on rat EARs. Only half of the 10 pair-wise comparisons in rat yield $d_N/d_S > 1$. The average pair-wise d_N (0.130 ± 0.050) among the rat paralogs is indistinguishable from the average pair-wise d_S (0.130 ± 0.043). These results suggest that the contribution of positive selection to the evolution of EAR genes may be different between the two rodents.

Zhang et al. [23] demonstrated that the diversification of rodent EARs has been driven by rapid gene sorting and positive selection, where gene sorting is defined as a process leading to differential retention of ancestral genes or gene lineages in different species. They included in their phylogenetic analysis 36 potentially functional EAR genes and 23 pseudogenes from four *Mus* species (*M. musculus*, *M. caroli*, *M. saxicola*, and *M. pahari*) and 18 genes and 1 pseudogene identified from rat (*Rattus norvegicus*), Mongolian gerbil (*Meriones unguiculatus*), and Chinese hamster (*Cricetulus griseus*) (see Table 1 of Ref. [23]). The inclusion of these pseudogenes was important for drawing their conclusion, that is, the placement of the pseudogenes in many places where functional genes are absent indicated the frequent gene birth-and-death process combined with differential retention of the ancestral genes in *Mus* species. However, at that time, only one EAR pseudogene (Ear-ps1) was known in *M. musculus*, and it was not included in their analysis, rather limiting the application of the model to this species. Since then, another mouse EAR pseudogene (Ear-ps2) has been found [25], and 10 additional pseudogenes are identified in this study. To test whether these additional genes and pseudogenes support the rapid gene sorting model, we made a neighbor-joining tree with all currently available rodent EAR genes and pseudogenes, particularly focusing on those 12 *M. musculus* pseudogenes (Supplemental Fig. 4). Interestingly, many of these pseudogenes are

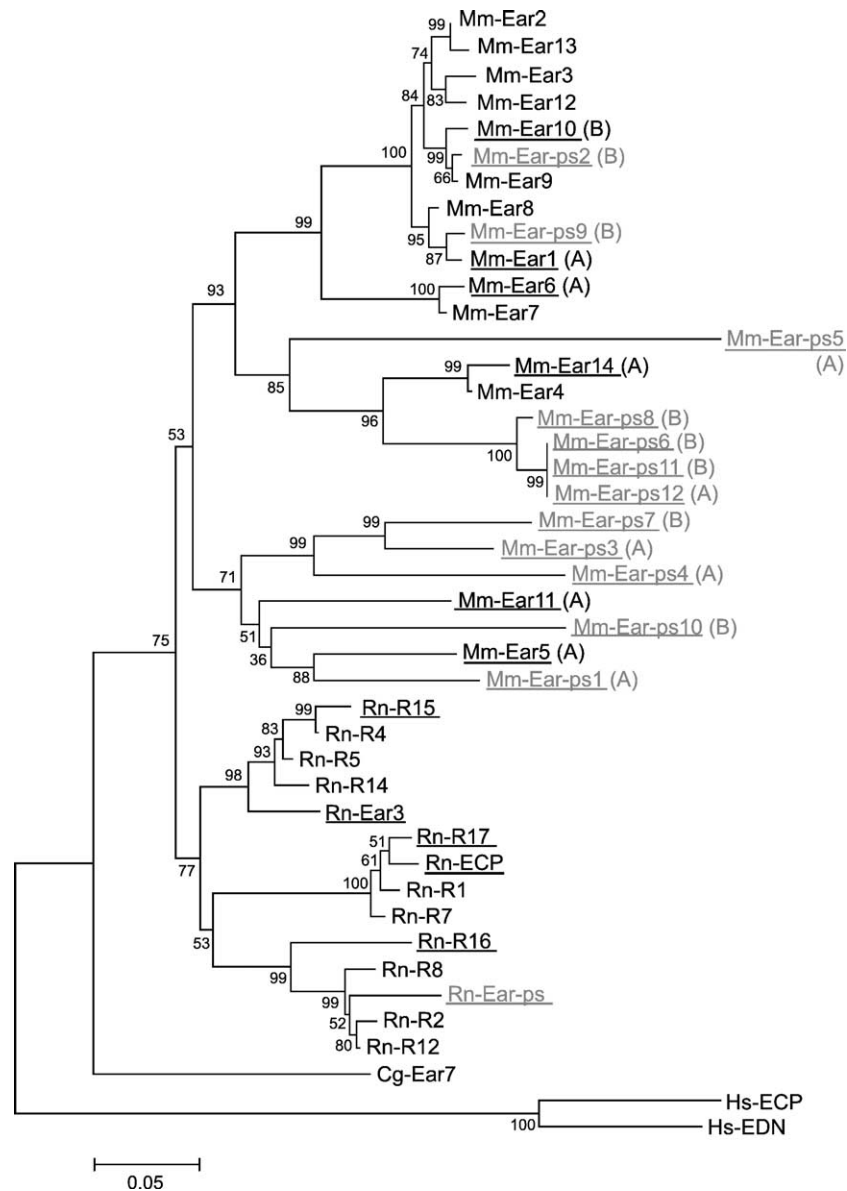


Fig. 3. Phylogenetic tree of functional genes and pseudogenes of the mouse and rat EAR family. A total of 435 nucleotide sites were used in tree-making. Underlined genes are the ones that are identified in the genome sequences, while other genes were previously reported but missing in the genome sequences. For mouse genes, the genomic cluster (A or B) to which each underlined gene belongs is marked in parentheses. Pseudogenes are labeled in gray. Chinese hamster (*Cricetulus griseus*) Ear7, human ECP, and human EDN are used as outgroups to root this tree. The neighbor-joining method with Kimura's two-parameter distance was used to make the tree. We used 2000 bootstrap replications to test the reliability of the tree.

found in a number of branches in which previously no *M. musculus* genes were identified. This result verified the rapid gene sorting model [23], which postulates that there were already many EAR genes in the ancestor of the *Mus* genus and subsequent differential gene duplication and inactivation resulted in different EAR repertoires in different species.

The rapid gene birth-and-death and gene sorting is a common evolutionary theme among the vertebrate host-defense gene families, such as the major histocompatibility complex, immunoglobulin, and T cell receptor gene families [38–41]. Among members of the RNase A superfamily, in vitro experiments suggested that human ECP and EDN are

involved in host defense [42], and at least two rat EARs (Ear1 and Ear2) have in vitro bactericidal activities [34]. It is therefore tempting to conjecture that the rapid gene sorting and positive selection observed in rodent EARs reflect host responses to diverse and ever-changing pathogens [23].

Expansion of the rodent angiogenin (RNase 5) family

Thus far, four mouse angiogenin genes (Ang1–4) and two mouse angiogenin pseudogenes (Ang-ps1–ps2) have been reported [43]. Our perusal of the genome sequences revealed two additional angiogenin genes (Ang5 and 6) and one additional pseudogene (Ang-ps3) in mouse. Ang6 has

insertions of two amino acid residues between the first two conserved cysteine residues, which may have influence on the properties of the protein. We also identified two angiogenin genes (Ang1 and 2) but no pseudogenes from rat. Ang3, one of the previously reported mouse angiogenins [27], however, could not be found in the genome sequence. All newly identified mouse angiogenin genes and pseudogenes are located in cluster B, whereas previously known genes are in cluster A. In the evolutionary tree of all angiogenin genes and pseudogenes from the mouse and rat (Fig. 4), separate groupings of mouse and rat genes are found, similar to the situation in mouse and rat EAR genes. This indicates that multiple recent gene duplications have occurred in the angiogenin family since the mouse–rat separation.

Is positive selection playing a role in the divergence of rodent angiogenin genes as in EAR genes? We computed d_S and d_N between paralogous angiogenin genes (Ang1–6) of the individual species (Supplemental Fig. 5). Between the two rat angiogenins, d_S (0.047 ± 0.019) is significantly greater than d_N (0.013 ± 0.007) ($p < 0.05$, Fisher's exact test), suggesting that purifying selection has been acting on these genes after the duplication. In mouse, however, most of the 10 pair-wise comparisons show higher d_N than d_S . Furthermore, the average pair-wise d_N (0.142 ± 0.024) is significantly greater than the average pair-wise d_S (0.116 ± 0.023) ($p < 0.05$, one-tailed t test), suggesting the operation of positive selection on these genes. Because positive selection results from differences in fitness, our results suggest that these paralogous mouse angiogenins might have divergent functions. Interestingly, two previously characterized mouse angiogenins (Ang1 and Ang4) are known to exhibit species-selective bactericidal activities [26], suggesting an intriguing hypothesis that the functional

diversification of the angiogenin family in mouse may also be subject to the gene sorting process found in EARs.

Block duplication of the RNase gene cluster in mouse

One unexpected finding of this study is a new RNase gene cluster in mouse (Fig. 1). This cluster (cluster B) contains genes and pseudogenes of only two lineages of the superfamily, EARs and angiogenins, suggesting that this cluster was generated by a block duplication that included at least these genes. Because neither the EARs nor the angiogenins in cluster B form a monophyletic group (Figs. 3 and 4), the initial duplication event must have included at least two EAR genes and two angiogenin genes. After the block duplication, subsequent tandem duplications within cluster B generated more genes (Figs. 1, 3, and 4). When did the initial block duplication take place? Because of the complicated evolutionary history of this gene cluster and the rapid gene duplication and pseudogenization in mouse EARs (see below and [23]), it is difficult to estimate the time of the block duplication by molecular dating. However, it would be most parsimonious to infer that the block duplication occurred in the mouse lineage after the mouse–rat separation, because in human and rat, all RNase genes and pseudogenes are mapped within a single gene cluster. Furthermore, this scenario is supported by the previous [16,23] and present results (Fig. 3) showing that mouse and rat EARs form two separate phylogenetic groups.

Origin and evolution of the RNase A superfamily

What are the evolutionary relationships among the members of the RNase A superfamily? To address this

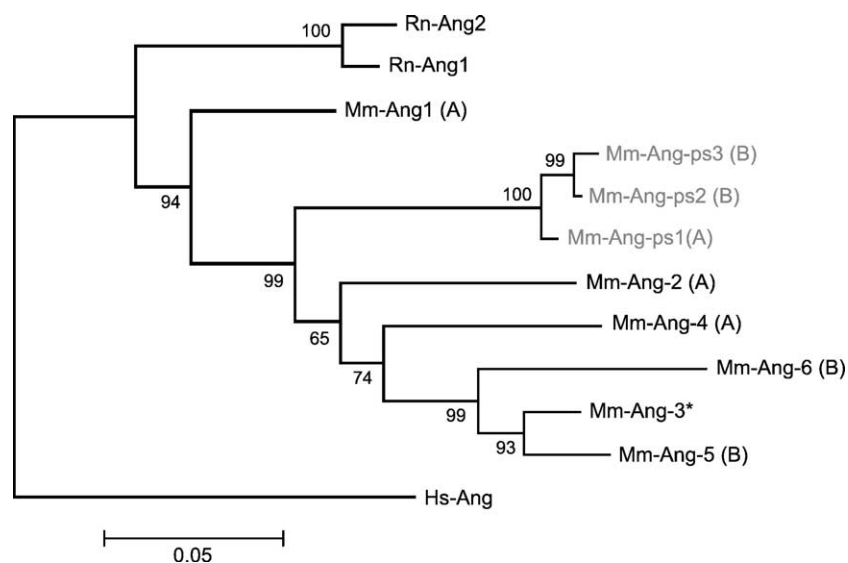


Fig. 4. Phylogenetic tree of functional genes and pseudogenes of the human, mouse, and rat angiogenin family. A total of 435 nucleotide sites were used in tree-making. For mouse genes, the genomic cluster (A or B), to which each gene belongs, is marked in parentheses. Mouse Ang3 (labeled by “*”) was not identified in the genome database. Pseudogenes are labeled in gray. The neighbor-joining method with Kimura’s two-parameter distance was used to make the tree. We used 2000 bootstrap replications to test the reliability of the tree.

question, we reconstructed a phylogenetic tree using all the RNase genes identified from the genome sequences of the human, rat, mouse, and chicken (Fig. 5). This unrooted tree reveals several features of the superfamily. First, the orthologs of mammalian RNases 9–13 form gene-specific clusters, confirming the orthologous relationships. Furthermore, these five RNases form a cluster with a moderate bootstrap support (66%), defining a new family within the RNase A superfamily. Second, the five mouse angiogenins do not form a monophyletic group, in contrast to the tree in Fig. 4. But the present tree (Fig. 5) uses fewer sites and has lower bootstrap percentages than the previous tree (Fig. 4), thus it may be less reliable. Third, neither the mouse genome nor the rat genome has genes orthologous to human RNases 7 and 8, and RNase 6 is sister to these two primate-specific RNases [10,12]. The phylogeny suggests that the common ancestor of the RNase 7/8 gene pair emerged from the duplication of the RNase 6 gene prior to the primate–rodent divergence and that this ancestral gene was subsequently lost in rodents. Fourth, although chicken CL2 has only ~40% identity to the other two chicken RNases (RSFR and Clone 462), the three chicken RNases form a monophyletic group with 91% bootstrap support (Fig. 5). This chicken RNase group is sister to mammalian angiogenins, indicating a close relationship between the chicken RNases and the mammalian angiogenins. However, this clustering has only weak bootstrap support (53%), suggesting that the relationship cannot be established unequivocally. Even so, it is noteworthy that all other sequenced nonmammalian vertebrate RNases also group with the mammalian angiogenins in previous phylogenetic studies, and they, including the chicken RNases, do not have two of the eight conserved cysteine residues, which are also missing in all mammalian angiogenins ([20,28] and unpublished data).

When and how did the RNase A superfamily originate? Thus far, the superfamily has been found only in vertebrates [5]. Database searches against the genomes of *Caenorhabditis elegans* and *Drosophila melanogaster* did not yield any significant matches. In addition, no RNase A genes could be identified from the completed genome sequences of *Ciona intestinalis* and *Ciona savignyi*, the ascidian urochordates, further confirming the vertebrate specificity of the superfamily. The tree shown in Fig. 5 presents six relatively well-established groups in the superfamily: (i) RNase 4, (ii) RNase 1, (iii) EARs and RNases 6–8, (iv) RNases 9–13, (v) angiogenins, and (vi) chicken RNases. Yet, the phylogenetic relationships among these six groups are poorly resolved, as indicated by the low bootstrap percentages for the interior branches connecting them. Furthermore, since the position of the root in the tree is not known, we cannot unambiguously determine the most primitive (or basal) group or the order in which these groups emerged during evolution. Nevertheless, the fact that only angiogenin/RNase 5-like RNases have been reported outside the class Mammalia suggests that the angiogenin/

RNase 5 group is probably the most ancient form of this superfamily and all other members arose during mammalian evolution [5,43]. The presence of only angiogenin-like genes in the entire chicken genome, as revealed in this study, fortifies this idea. Our preliminary survey of the incomplete genome sequence of the zebrafish (*Danio rerio*) and the EST database of the salmon (*Salmo salar*) identified only angiogenin/RNase 5-like RNases, further supporting this thought. Future verifications of this scenario would require a comprehensive survey of RNase genes in basal vertebrate lineages such as the hagfish, lamprey, and shark. It should be emphasized that none of the RNases in nonmammalian organisms are known to possess the angiogenic activity. It is likely that the first RNases were similar to the mammalian angiogenins at the sequence or structural level, but the angiogenic activity was acquired later in evolution.

The fact that the ancestral RNase A resembled mammalian angiogenin/RNase 5 at the structural and sequence level suggests an intriguing hypothesis that the superfamily started off as a host-defense mechanism during early vertebrate evolution and that its expansion in mammals led to their current functional diversity. There are two lines of observations that support this hypothesis. First, multiple lineages within the superfamily exhibit bactericidal or antiviral activities, including mouse Ang1 and Ang4 [26], human angiogenin [26], rat EARs [34], human ECP and EDN [44], human RNase 7 [10,11], and frog RNases [4]. Second, the *pI* of the three avian RNases are among the highest of all RNases known thus far (Supplemental Table 7), and high *pI* has been associated with antibacterial activities in the superfamily [11,33]. Interestingly, d_N (0.085 ± 0.018) is significantly higher than d_S (0.049 ± 0.021) between RSFR and Clone 462, indicating the role of positive selection in their divergence, as in other bactericidal RNases (see above and [16,23,33,45]). Experimental determination of the bactericidal activity of chicken and other nonmammalian RNases is necessary to test this hypothesis.

Next, to test whether different members of the superfamily have experienced different selective forces during rodent evolution, we compared the d_N and d_S between the orthologous genes of the mouse and rat (Fig. 6). The d_S values of the 10 gene pairs range from 0.151 to 0.232 with an average of 0.19, which is identical to the average d_S between over 11,000 orthologous gene pairs between mouse and rat [46]. However, the d_N values generate a wider distribution, ranging from 0.049 to 0.206, demonstrating that nucleotide substitutions altering protein sequences are selected to different extents among the members. In parallel with this observation, the d_N/d_S ratio also varies substantially among the members. The highest d_N/d_S ratio is detected in EARs (1.01, the average of all possible orthologous EAR pairs), and the smallest in RNase 13 (0.33), with an average of 0.62. All 10 values are greater than the average d_N/d_S ratio (0.11) found in ~11,000

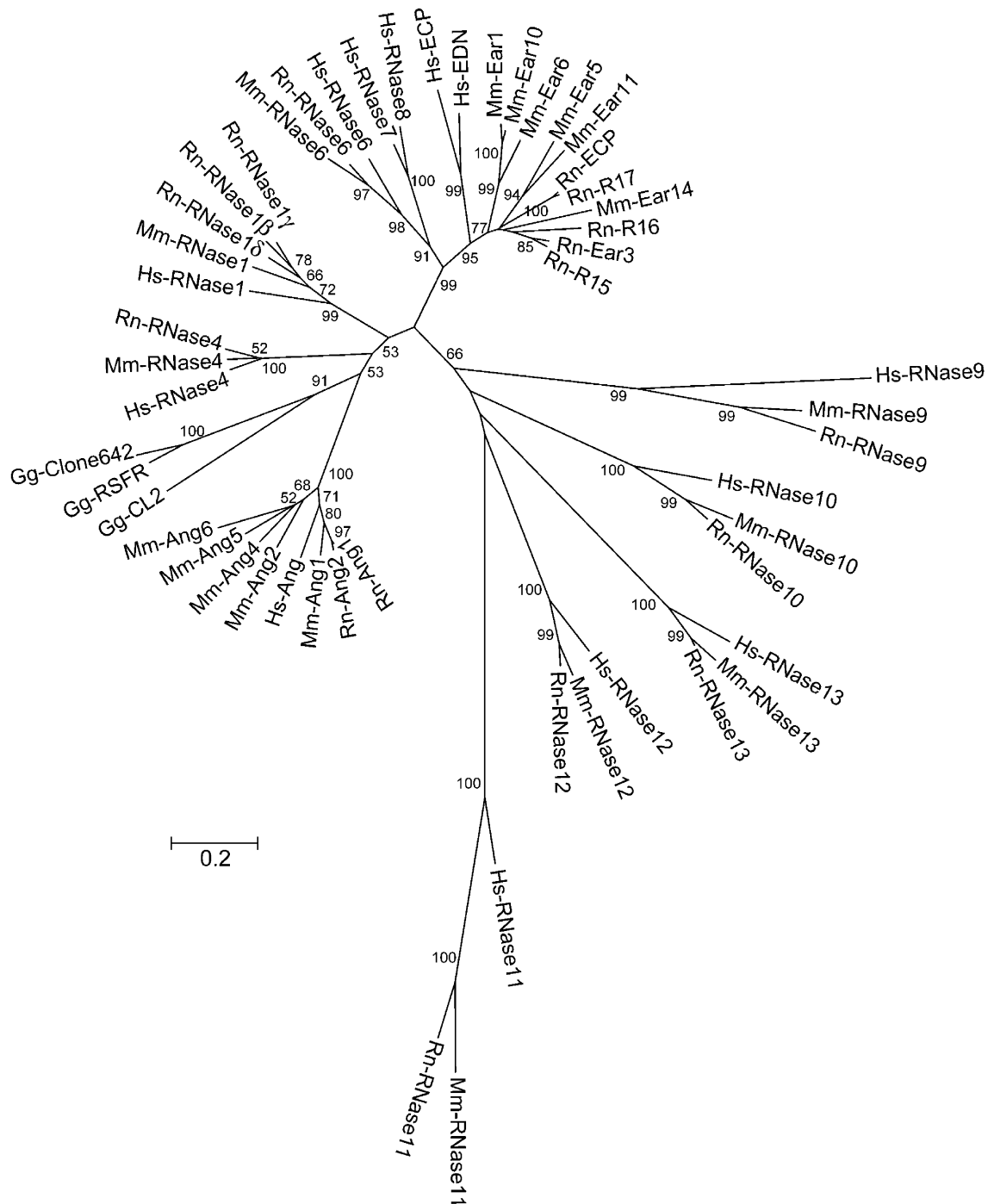


Fig. 5. Unrooted phylogenetic tree of all RNases identified in human, mouse, rat, and chicken genomes. A total of 89 amino acid sites were used in tree-making. The neighbor-joining method with Poisson-corrected distance was used to make the tree. We used 2000 bootstrap replications to test the reliability of the tree. Bootstrap percentages ≥ 50 are shown on interior branches.

orthologous gene pairs of mouse and rat and fall within the 10% upper tail of the distribution; the average value (0.62) falls even within the 0.1% upper tail. These results indicate that, although the overall d_N/d_S values are equal to or less than 1 in RNases, the rates of nonsynonymous nucleotide substitutions are elevated in all members of the superfamily in comparison to average genes of mouse and rat. This is consistent with the roles of many RNase genes in host

defense, as host-defense genes show higher d_N/d_S than other genes.

Final remarks

In this study, we took a comparative genomic approach in the study of the RNase A superfamily. This approach

confirms, complements, and corrects previous findings that were based on experimental methods of gene hunting. First, it provides equal representation for every gene without experimental bias. Due to the artificial nature of primer or probe selection, experimental approaches based on PCR or hybridization methods might generate overrepresentation of some genes, with some others underrepresented or missing. However, it should be noted that the genomic databases used in this study do not represent the entire genomic content, possibly leaving some genes undetected. For example, the National Center for Biotechnology Information (NCBI) mouse and rat genome builds are ~90% complete (see Materials and methods). Second, pseudogenes can be more easily identified. Because pseudogenes tend to diverge faster than functional genes, they sometimes escape from experimental gene-hunting methods. For example, in this study, we could identify 10 new mouse pseudogenes, in contrast to only 2 experimentally identified ones. Third, chromosomal locations of genes and syntenic information are readily obtained. Finally, it is immune to many experimental errors such as PCR errors. This is particularly important for the RNase A superfamily because of the presence of many recent duplicates whose sequences are nearly identical. As more and more vertebrate genomes are being sequenced, the comparative genomic approach as taken here will help answer important questions on the evolution of this rapidly changing enzyme superfamily.

Materials and methods

Nomenclature

In this paper, we use “human” for *Homo sapiens* (Hs for short), “mouse” for the house mouse *M. musculus* (Mm), “rat” for Norway brown rat *R. norvegicus* (Rn), and “chicken” for the red jungle fowl *Gallus gallus* (Gg), unless otherwise specified. By “functional gene,” we mean that the homologous sequence under investigation is contained in an uninterrupted ORF, and, conversely, a sequence is considered a “pseudogene” if the ORF is interrupted by a premature stop codon anywhere in the ORF or by frame-shifting insertions/deletions. Pseudogenes are distinguished from functional genes by having “-ps” between the gene name and the number (e.g., Ear-ps1). Genes and pseudogenes are named according to the order in which they were identified.

Identification of RNase genes

Using all currently available human RNase sequences (RNases 1–10) as queries, TBLASTN and BLASTN searches (<http://www.ncbi.nlm.nih.gov/BLAST/>) were carried out on the human, mouse, rat, and chicken genome sequence databases available at the NCBI. We used $1e-10$ as the cutoff for our TBLASTN and BLASTN searches. For human genome sequence, we used the NCBI Build 34, which has been sequenced to an accuracy of 99.99% and

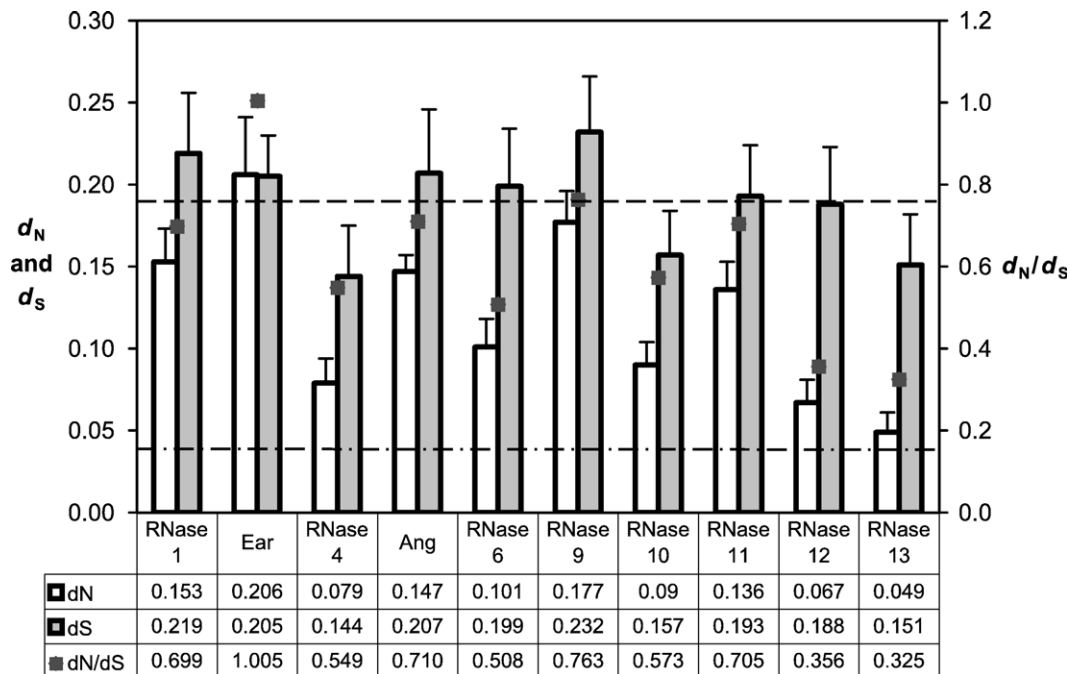


Fig. 6. Synonymous (d_S) and nonsynonymous (d_N) nucleotide distances between orthologous gene pairs of mouse and rat for each RNase family. For RNase 1, Ear, and angiogenins, which have multiple representatives in mouse or rat, the average distance of all possible orthologous combinations was taken. The average d_S (= 0.19) and d_N/d_S ratio (= 0.11) for over 11,000 mouse–rat orthologous gene pairs is marked by the dashed horizontal line and the dash-and-dot line, respectively.

covers ~99% of the gene-containing regions in the genome (<http://www.ncbi.nlm.nih.gov/genome/guide/human/>). Our mouse genome searches were based on NCBI Build 32, which is estimated to place more than 90% of the mouse genome on chromosomes (<http://www.ncbi.nlm.nih.gov/genome/guide/mouse/>). We used the NCBI rat genome Build 2 (<http://www.ncbi.nlm.nih.gov/genome/guide/rat/index.html>) and the first assembled draft chicken genome sequence produced by the Genome Sequencing Center at Washington University (<http://www.genome.wustl.edu/projects/chicken/>). The rat genome is over 90% complete, and the chicken genome has been sequenced to a 6.63 coverage and anchors approximately 88% of the sequence to chromosomes.

Because some of the previously identified rodent RNase genes show very high sequence identity to each other (e.g., mouse Ear6 and 7 are 98.6% identical in DNA sequence), we used a strict criterion that two sequences compared may represent a single gene only when they share >99% identity in DNA sequence. Based on this, we did not follow the NCBI annotation for two mouse genes, Ear4 and Ang3, because the actual genomic sequences of the ORFs at these loci are only 97.9 and 96.8% identical to the reference sequences used for the NCBI annotation, respectively. We instead assigned each of them a new name, Ear14 and Ang5.

The RNase gene sequences reported in this paper have been deposited with GenBank under Accession Nos. AY665804–AY665837.

Sequence alignment and evolutionary analysis

Protein and nucleotide sequence alignments were made by ClustalX [47] with manual adjustments. MEGA2 [48] was used for evolutionary analyses. Phylogenetic trees were reconstructed using the neighbor-joining method [49] with 2000 bootstrap replications [50]. We used the complete deletion option for all trees. Numbers of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site between homologous DNA sequences were computed by the modified Nei-Gojobori method [33].

Acknowledgments

J.J.B. thanks Leonard Bosgraaf for his help in analyzing the mouse and rat genomes. This work was supported by a startup fund from the University of Michigan and National Institutes of Health Grant GM67030 to J.Z.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, on the online version, at [doi:10.1016/j.ygeno.2004.10.008](https://doi.org/10.1016/j.ygeno.2004.10.008).

References

- [1] M. Kunitz, M.R. McDonald, Ribonuclease—crystalline—from beef pancrease, *Biochem. Preparat.* 3 (1953) 9–19.
- [2] J. Zhang, Evolution by gene duplication: an update, *Trends Ecol. Evol.* 18 (2003) 292–298.
- [3] P. Blackburn, S. Moore, Pancreatic ribonuclease, in: P.D. Boyer (Ed.), *The Enzymes*, vol. 15, Academic Press, New York, 1982, pp. 317–433.
- [4] J.J. Beintema, H.J. Breukelman, A. Carsana, A.A. Furia, Evolution of vertebrate ribonucleases: ribonuclease A superfamily, in: G. D'Alessio, J.F. Riordan (Eds.), *Ribonucleases—structures and functions*, Academic Press, New York, 1997, pp. 245–269.
- [5] J.J. Beintema, R.G. Kleineidam, The ribonuclease A superfamily: general discussion, *Cell. Mol. Life Sci.* 54 (1998) 825–832.
- [6] J. Zhang, Y.P. Zhang, H.F. Rosenberg, Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey, *Nat. Genet.* 30 (2002) 411–415.
- [7] T.M. Jermann, J.G. Opitz, J. Stackhouse, S.A. Benner, Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily, *Nature* 374 (1995) 57–59.
- [8] C.B. Anfinsen, Principles that govern folding of protein chains, *Science* 181 (1973) 223–230.
- [9] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al., Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [10] J. Harder, J.-M. Schroder, RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin, *J. Biol. Chem.* 277 (2002) 46779–46784.
- [11] J. Zhang, K.D. Dyer, H.F. Rosenberg, Human RNase 7: a new cationic ribonuclease of the RNase A superfamily, *Nucleic Acids Res.* 31 (2003) 602–607.
- [12] J. Zhang, K.D. Dyer, H.F. Rosenberg, RNase 8, a novel RNase A superfamily ribonuclease expressed uniquely in placenta, *Nucleic Acids Res.* 30 (2002) 1169–1175.
- [13] J. Penttinen, D.A. Pujianto, P. Sipila, I. Huhtaniemi, M. Poutanen, Discovery *in silico* and characterization *in vitro* of novel genes exclusively expressed in the mouse epididymis, *Mol. Endocrinol.* 17 (2003) 2138–2151.
- [14] S. Castella, S. Fouchecourt, A.P. Teixeira-Gomes, J. Vinh, M. Belghazi, F. Dacheux, J.L. Dacheux, Identification of a member of a new RNase A family specifically secreted by epididymal caput epithelium, *Biol. Reprod.* 70 (2004) 319–328.
- [15] J.Y.F. Dubois, P.A. Jekel, P.P.M.F.A. Mulder, A.P. Bussink, F.M. Catzefflis, A. Carsana, J.J. Beintema, Pancreatic-type ribonuclease 1 gene duplications in rat species, *J. Mol. Evol.* 55 (2002) 522–533.
- [16] N.A. Singhania, K.D. Dyer, J. Zhang, M.S. Deming, C.A. Bonville, J.B. Domachowske, H.F. Rosenberg, Rapid evolution of the ribonuclease A superfamily: adaptive expansion of independent gene clusters in rats and mice, *J. Mol. Evol.* 49 (1999) 721–728.
- [17] W. Zhao, Z. Kote-Jarai, Y. van Santen, J. Hofsteenge, J.J. Beintema, Ribonucleases from rat and bovine liver: purification, specificity and structural characterization, *BBA-Protein Struct. M.* 1384 (1998) 55–65.
- [18] T. Nittoh, M. Hirakata, S. Mue, K. Ohuchi, Identification of cDNA encoding rat eosinophil cationic protein/eosinophil-associated ribonuclease, *BBA-Gene Struct. Expression* 1351 (1997) 42–46.
- [19] H.F. Rosenberg, K.D. Dyer, Eosinophil cationic protein and eosinophil-derived neurotoxin—evolution of novel function in a primate ribonuclease gene family, *J. Biol. Chem.* 270 (1995) 21539–21544.
- [20] K.A. Larson, E.V. Olson, B.J. Madden, G.J. Gleich, N.A. Lee, J.J. Lee, Two highly homologous ribonuclease genes expressed in mouse eosinophils identify a larger subgroup of the mammalian ribonuclease superfamily, *Proc. Natl. Acad. Sci. USA* 93 (1996) 12370–12375.
- [21] D. Batten, K.D. Dyer, J.B. Domachowske, H.F. Rosenberg, Molecular

- cloning of four novel murine ribonuclease genes: unusual expansion within the ribonuclease A gene family, *Nucleic Acids Res.* 25 (1997) 4235–4239.
- [22] K.A. Larson, T.L. Mitchell, N.A. Lee, J.J. Lee, Characterization of a new murine eosinophil granule protein, eosinophil-associated ribonuclease-3 (mEAR-3), *J. Allergy Clin. Immun.* 103 (1999) S113.
- [23] J. Zhang, K.D. Dyer, H.F. Rosenberg, Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection, *Proc. Natl. Acad. Sci. USA* 97 (2000) 4701–4706.
- [24] K.D. Dyer, J.M. Moreau, H.F. Rosenberg, Transcriptional regulation of mouse eosinophil associated ribonuclease 2 (mEAR2) by intronic enhancer elements: common regulatory features of mouse EAR genes, *Mol. Biol. Cell.* 12 (2001) 99a.
- [25] S.A. Cormier, K.A. Larson, S.B. Yuan, T.L. Mitchell, K. Lindenberger, P. Carrigan, N.A. Lee, J.J. Lee, Mouse eosinophil-associated ribonucleases: a unique subfamily expressed during hematopoiesis, *Mamm. Genome* 12 (2001) 352–361.
- [26] L.V. Hooper, T.S. Stappenbeck, C.V. Hong, J.L. Gordon, Angiogenins: a new class of microbicidal proteins involved in innate immunity, *Nat. Immunol.* 4 (2003) 269–273.
- [27] X.Y. Fu, W.G. Roberts, V. Nobile, R. Shapiro, M.P. Kamps, mAngiogenin-3, a target gene of oncoprotein E2a-Pbx1, encodes a new angiogenic member of the angiogenin family, *Growth Factors* 17 (1999) 125–137.
- [28] K. Hayano, M. Iwama, H. Sakamoto, H. Watanabe, A. Sanda, K. Ohgi, M. Irie, Characterization of poly-C-preferential ribonuclease from chicken liver, *J. Biochem.* 114 (1993) 156–162.
- [29] E.M. Klenova, I. Botezato, V. Laudet, G.H. Goodwin, J.C. Wallace, V.V. Lobanenko, Isolation of a cDNA clone encoding the RNase-superfamily-related gene highly expressed in chicken bone-marrow cells, *Biochem. Biophys. Res. Commun.* 185 (1992) 231–239.
- [30] T. Nakano, T. Graf, Identification of genes differentially expressed in 2 types of v-myb-transformed avian myelomonocytic cells, *Oncogene* 7 (1992) 527–534.
- [31] J. Hofsteenge, A. Vicentini, O. Zelenko, Ribonuclease 4, an evolutionarily highly conserved member of the superfamily, *Cell. Mol. Life Sci.* 54 (1998) 804–810.
- [32] H.F. Rosenberg, K.D. Dyer, H.L. Tiffany, M. Gonzalez, Rapid evolution of a unique family of primate ribonuclease genes, *Nat. Genet.* 10 (1995) 219–223.
- [33] J. Zhang, H.F. Rosenberg, M. Nei, Positive Darwinian selection after gene duplication in primate ribonuclease genes, *Proc. Natl. Acad. Sci. USA* 95 (1998) 3708–3713.
- [34] K. Ishihara, K. Asai, M. Nakajima, S. Mue, K. Ohuchi, Preparation of recombinant rat eosinophil-associated ribonuclease-1 and -2 and analysis of their biological activities, *BBA-Mol. Basis Dis.* 1638 (2003) 164–172.
- [35] E. Carreras, E. Boix, H.F. Rosenberg, C.M. Cuchillo, M.V. Nogues, Both aromatic and cationic residues contribute to the membrane-lytic and bactericidal activity of eosinophil cationic protein, *Biochemistry* 42 (2003) 6636–6644.
- [36] A.I. Su, T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, et al., A gene atlas of the mouse and human protein-encoding transcriptomes, *Proc. Natl. Acad. Sci. USA* 101 (2004) 6062–6067.
- [37] M. Nakajima, M. Hirakata, T. Nittoh, K. Ishihara, K. Ohuchi, Expression and purification of recombinant rat eosinophil-associated ribonucleases, homologues of human eosinophil cationic protein and eosinophil-derived neurotoxin, and their characterization, *Int. Arch. Allergy Immunol.* 125 (2001) 241–249.
- [38] M. Nei, X. Gu, T. Sitnikova, Evolution by the birth-and-death process in multigene families of the vertebrate immune system, *Proc. Natl. Acad. Sci. USA* 94 (1997) 7799–7806.
- [39] T. Sitnikova, M. Nei, Evolution of immunoglobulin kappa chain variable region genes in vertebrates, *Mol. Biol. Evol.* 15 (1998) 50–60.
- [40] C. Su, M. Nei, Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences, *Mol. Biol. Evol.* 18 (2001) 503–513.
- [41] L.F. Cadavid, C. Shufflebotham, F.J. Ruiz, M. Yeager, A.L. Hughes, D.I. Watkins, Evolutionary instability of the major histocompatibility complex class I loci in New World primates, *Proc. Natl. Acad. Sci. USA* 94 (1997) 14536–14541.
- [42] H.F. Rosenberg, The eosinophil ribonucleases, *Cell. Mol. Life Sci.* 54 (1998) 795–803.
- [43] D.J. Strydom, The angiogenins, *Cell. Mol. Life Sci.* 54 (1998) 811–824.
- [44] R.I. Lehrer, D. Szklarek, A. Barton, T. Ganz, K.J. Hamann, G.J. Gleich, Antibacterial properties of eosinophil major basic-protein and eosinophil cationic protein, *J. Immunol.* 142 (1989) 4428–4434.
- [45] J. Zhang, H.F. Rosenberg, Diversifying selection of the tumor-growth promoter angiogenin in primate evolution, *Mol. Biol. Evol.* 19 (2002) 438–445.
- [46] R.G.S.P. Consortium, Genome sequence of the Brown Norway rat yields insights into mammalian evolution, *Nature* 428 (2004) 493–521.
- [47] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res.* 25 (1997) 4876–4882.
- [48] S. Kumar, K. Tamura, I.B. Jakobsen, M. Nei, MEGA2: molecular evolutionary genetics analysis software, *Bioinformatics* 17 (2001) 1244–1245.
- [49] N. Saitou, M. Nei, The neighbor-joining method—a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* 4 (1987) 406–425.
- [50] J. Felsenstein, Confidence-limits on phylogenies—an approach using the bootstrap, *Evolution* 39 (1985) 783–791.