

Supplementary Materials

Table S1. List of snRNA genes in all analyzed insect genomes (MS Excell file)

Table S2. The Honeybee genes containing U12-dependent introns.

Figure S1. Models of secondary structure of the honeybee snRNAs.

Figure S2. Phylogenetic trees of insect snRNA genes.

Table S2. The honeybee genes containing U12-dependent introns.

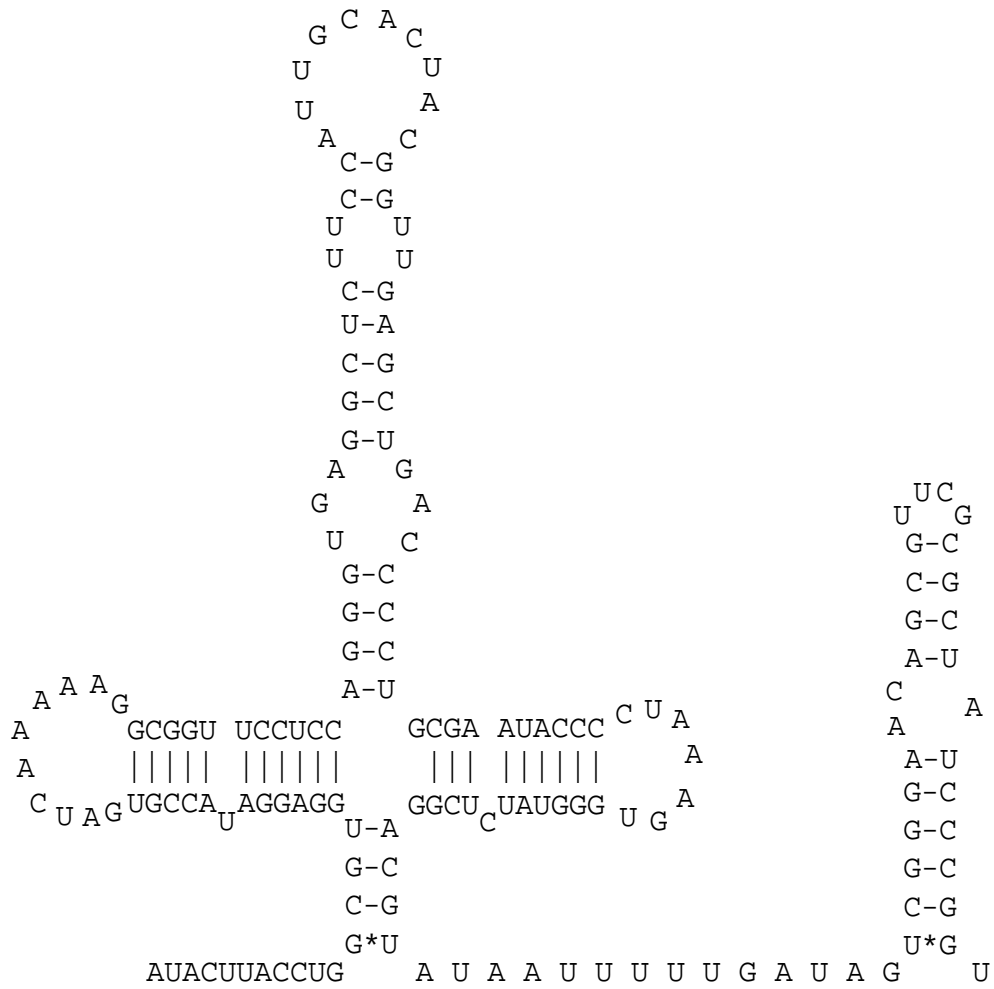
<i>Gene ID</i>	<i>Representative Transcript ID</i>	<i>U12 Intron Position*</i>	<i>5' Termini</i>	<i>3' Termini</i>	<i>Gene Family ID</i>	<i>Consensus annotation of member proteins</i>
ENSAPMG00000011219	ENSAPMT00000035154	1 (2)	AT	AC	ENSF00000002258	Actin related 2/3 complex subunit 5
ENSAPMG00000005355	ENSAPMT00000009291	6 (7)	GT	AG	ENSF00000001313	Activator 1.40 kda subunit replication factor c 40 kda
ENSAPMG00000009676	ENSAPMT00000025222	2 (2)	GT	AG	ENSF00000000116	Adp ribosylation factor
ENSAPMG00000005952	ENSAPMT00000010283	1 (1)	GT	AG	ENSF000000008621	Adult male small intestine cdna product:metallothionein
ENSAPMG00000000664	ENSAPMT00000025045	15 (15)	GT	AG	ENSF000000002670	Arsenite resistance 2
ENSAPMG00000007792	ENSAPMT00000017164	12 (15)	GT	AG	ENSF00000000151	Calpain subunit calcium activated neutral proteinase canp
ENSAPMG00000004291	ENSAPMT00000007421	1 (2)	GT	AG	ENSF000000001726	Cyclin t1 cyclin t cyct1
ENSAPMG00000008380	ENSAPMT00000014700	8 (10)	GT	AG	ENSF000000000851	Diaphanous related formin
ENSAPMG00000006735	ENSAPMT00000011716	1 (63)	GT	AT	ENSF000000000023	Dynein heavy chain
ENSAPMG00000001271	ENSAPMT00000002177	3 (3)	AT	AC	ENSF000000000783	Peptidyl prolyl cis trans isomerase ppiase rotamase kda fkbp
ENSAPMG00000005491	ENSAPMT00000009536	8 (8)	GT	AG	ENSF000000001529	Heterogeneous nuclear ribonucleoprotein m
ENSAPMG00000002935	ENSAPMT00000005122	15 (16)	GT	AG	ENSF000000003875	Huntington's disease
ENSAPMG00000007720	ENSAPMT00000017650	1 (15)	GT	AG	ENSF000000000659	Lethal 2 giant larvae
ENSAPMG00000001302	ENSAPMT00000002229	1 (19)	GT	AG	ENSF000000000659	Lethal 2 giant larvae
ENSAPMG00000003087	ENSAPMT00000005403	1 (4)	GT	AG	ENSF000000007301	Lipoyltransferase 1 mitochondrial precursor ec 6 lipoate ligase
ENSAPMG00000002182	ENSAPMT00000022797	1 (20)	GT	AG	ENSF000000001376	Niemann pick c1 precursor
ENSAPMG00000006414	ENSAPMT00000032748	12 (15)	GT	AG	ENSF000000001061	Phosphatidylinositol 4 5 bisphosphate 5
ENSAPMG00000000861	ENSAPMT00000024222	5 (5)	GT	AG	ENSF000000000329	Probable atp dependent rna helicase
ENSAPMG00000000269	ENSAPMT00000000440	3 (4)	AT	AC	ENSF000000001322	Prohibitin b cell receptor associated
ENSAPMG00000015926	ENSAPMT00000028003	1 (3)	GT	AG	ENSF000000007532	Retinal rod rhodopsin sensitive cgmp 3' 5' cyclic phosphodiesterase delta subunit
ENSAPMG00000000712	ENSAPMT00000001173	2 (3)	GT	AG	ENSF000000002958	Ring finger 5
ENSAPMG00000016898	ENSAPMT00000010083	2 (2)	GT	AG	ENSF000000004407	Sarcoma amplified sequence homolog
ENSAPMG00000007542	ENSAPMT00000013176	6 (8)	GT	AG	ENSF000000000852	Serine/threonine phosphatase 2a 55 kda regulatory subunit b

ENSAPMG00000016097	ENSAPMT00000023791	3 (7)	GT	AG	ENSF00000000307	Serine/threonine phosphatase
ENSAPMG00000005017	ENSAPMT00000032946	1 (2)	GT	AG	ENSF000000003845	Signal recognition particle 68 kda srp68
ENSAPMG00000012152	ENSAPMT00000023346	7 (11)	GT	AG	ENSF000000000973	Sodium/hydrogen exchanger
ENSAPMG00000016266	ENSAPMT00000026176	7 (14)	GT	AG	ENSF000000003210	Soluble adenylyl cyclase
ENSAPMG00000008977	ENSAPMT00000015684	3 (5)	GT	AG	ENSF000000006387	Syntaxin 8
ENSAPMG00000015758	ENSAPMT00000019878	5 (10)	GT	AG	ENSF000000000321	T complex 1 subunit
ENSAPMG00000008384	ENSAPMT00000014704	1 (4)	GT	AG	ENSF000000000688	Testis specific y encoded
ENSAPMG00000002844	ENSAPMT00000004980	3 (6)	GT	AG	ENSF000000000646	Thioredoxin reductase
ENSAPMG00000000225	ENSAPMT00000000377	1 (13)	GT	AG	ENSF000000002063	Transportin 1 importin beta 2 karyopherin beta 2
ENSAPMG00000010953	ENSAPMT00000035878	2 (2)	GT	AG	ENSF000000004666	Trna pseudouridine synthase a trna uridine isomerase i
ENSAPMG00000006044	ENSAPMT00000034395	5 (11)	GT	AG	ENSF000000002303	Tyrosine phosphatase
ENSAPMG00000009780	ENSAPMT00000000282	1 (3)	GT	AG	ENSF000000009013	U6 snrna associated sm lsm8
ENSAPMG00000003749	ENSAPMT00000023279	1 (1)	GT	AG	ENSF000000005974	Ubiquitin conjugating enzyme e2 h
ENSAPMG00000007222	ENSAPMT00000031526	1, 15 (36)	GT	AG	ENSF000000000132	Voltage gated calcium channel alpha subunit
ENSAPMG00000005438	ENSAPMT00000023419	2 (23)	GT	AG	ENSF000000000206	Xanthine dehydrogenase
ENSAPMG00000000193	ENSAPMT00000000318	1 (6)	GT	AG	ENSF000000002686	Zinc finger 207
ENSAPMG00000014345	ENSAPMT00000028945	2 (5)	GT	AG	ENSF000000001553	Zinc finger dhhc domain containing
ENSAPMG00000010717	ENSAPMT00000016959	2 (2)	GT	AG	ENSF000000003987	Zinc finger zpr1 zinc finger 259
ENSAPMG00000012828	ENSAPMT00000025313	9 (11)	GT	AG	ENSF000000003657	Ambiguous
ENSAPMG00000014507	ENSAPMT00000031128	2 (11)	GT	AG	ENSF000000004395	Ambiguous
ENSAPMG00000006695	ENSAPMT00000011646	5 (8)	GT	AG	ENSF000000003174	Ambiguous
ENSAPMG00000000108	ENSAPMT00000022083	9 (9)	GT	AG	ENSF000000003748	Ambiguous
ENSAPMG00000010627	ENSAPMT00000022301	1 (10)	GT	AG	ENSF000000006955	Ambiguous
ENSAPMG00000002502	ENSAPMT00000004365	3 (3)	GT	AG	ENSF000000002868	Ambiguous
ENSAPMG00000005057	ENSAPMT00000017056	1 (3)	GT	AG	ENSF000000007776	Ambiguous
ENSAPMG00000015393	ENSAPMT00000023238	2 (3)	GT	AG	ENSF000000002114	Ambiguous
ENSAPMG00000003349	ENSAPMT00000005878	1 (4)	AT	AC	ENSF000000004412	Ambiguous
ENSAPMG00000006764	ENSAPMT00000011768	2 (2)	GT	AG	ENSF000000009016	Unknown
ENSAPMG00000014306	ENSAPMT00000027863	1 (5)	TT	AG	ENSF000000036271	Unknown
ENSAPMG00000002604	ENSAPMT00000035832	2 (2)	GT	AG	ENSF000000038215	Unknown
ENSAPMG00000011976	ENSAPMT00000030378	1 (2)	GT	AG	ENSF000000036883	Unknown
ENSAPMG00000000031	ENSAPMT00000000060	4 (4)	AT	AC	ENSF000000032479	Unknown
ENSAPMG00000012683	ENSAPMT00000035873	6 (11)	GT	TG	ENSF000000038225	Unknown

*Because all Ensembl gene predictions are based on evidence derived from known protein, cDNA, and EST sequences (Curwen et al. 2004), in many cases genes are partially annotated. The positions of U12 introns presented in this table are based on number of exons that have been annotated for that transcript, and thus, do not necessarily reflect the position of U12 introns in the complete transcript. In parentheses the total number of introns inferred from annotated exons is presented.

Reference

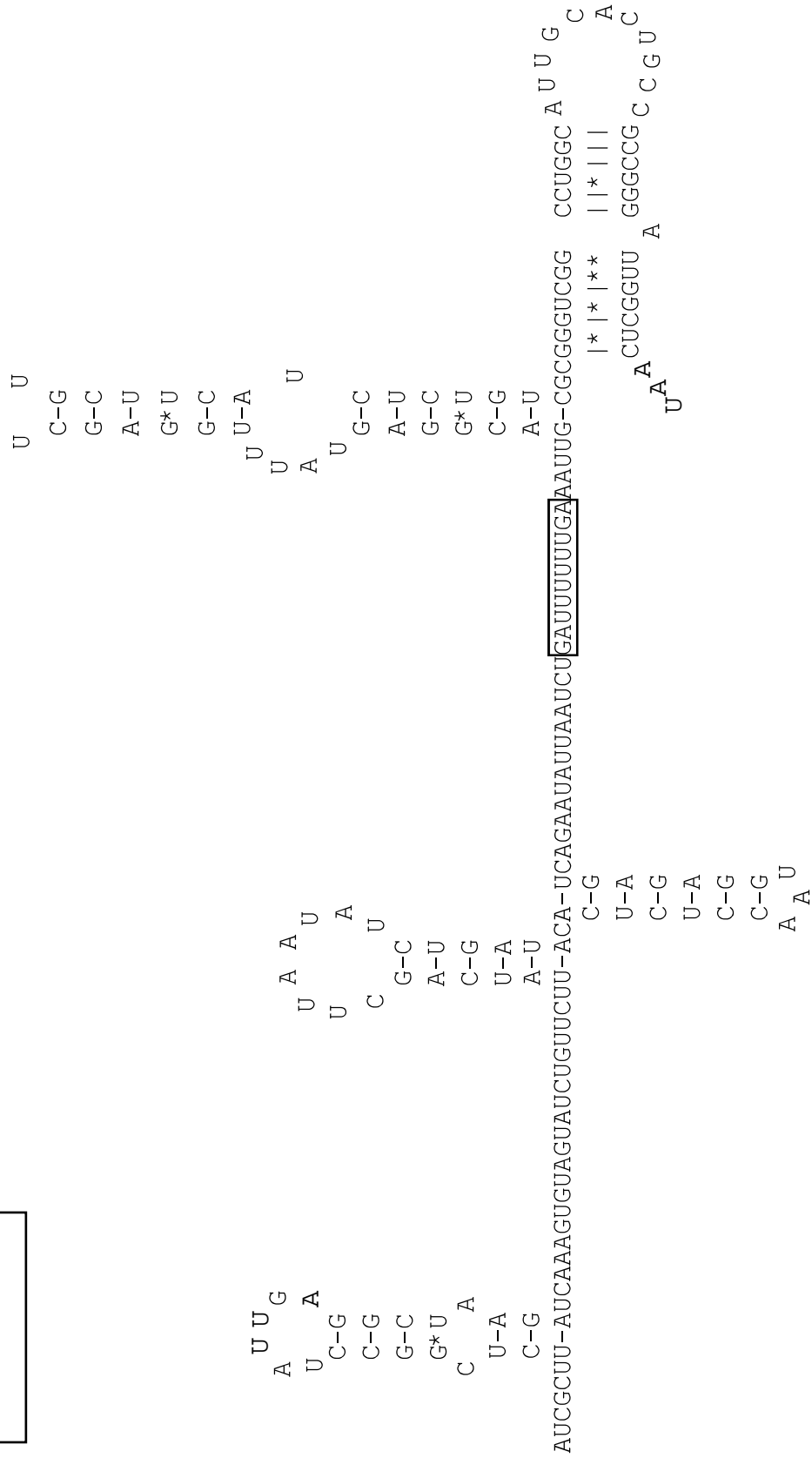
Val Curwen, Eduardo Eyras, T. Daniel Andrews, Laura Clarke, Emmanuel Mongin, Steven M.J. Searle, and Michele Clamp. 2004. The Ensembl Automatic Gene Annotation System. *Genome Res.* May; 14(5):942-950.



Apis mellifera U1 snRNA

Figure S1A

Figure S1B



Apis mellifera U2 snRNA

GACACGCAAAAUCGGUAGGCGUUCACACAUUUUU 3'

U U C

C G

G G

U C

U-A

C-G

A-U

U-A

G-A

C-G

A-U

U-A

U U C

C G

G G

U C

U-A

C-G

A-U

U-A

G-A

C-G

A-U

U-A

5' G-CAUAUACUAAAAUUGGAACGGAUACAGA G

U6

3'

G-CAAUGGUUUUUAAC-GCCCCAUAACCAUUUUCAAAAG

G-C

G-C

G-C

C

A A

C U

3'

G-CAAUGGUUUUUAAC-GCCCCAUAACCAUUUUCAAAAG

G-C

G-C

G-C

C

A A

C U

G A

C U

U U

C-G

G-A

G-C

C-G

U-A

A-U

C-G

U U

A G

A A

A A

5' G-CAUAUACUAAAAUUGGAACGGAUACAGA G

A A

U U A

G * U

A * U

U - A

C - G

G - C

U - A

C - G

G - C

U - A

A U G

G

U U U

U - A

A - U

G - C

C - G

G * U

A

U - A

G - C

G - C

G - C

G - C

U U C

C G

G G

U C

U-A

C-G

A-U

U-A

G-A

C-G

A-U

U-A

5' G-CAUAUACUAAAAUUGGAACGGAUACAGA G

A A

U U A

G * U

A * U

U - A

C - G

G - C

U - A

C - G

G - C

U - A

A U G

G

U U U

U - A

A - U

G - C

C - G

G * U

A

U - A

G - C

G - C

G - C

G - C

5'

G - C

G

G * U

A - U

A - U

C - G

G - C

C - G

G - C

U - A

C - G

C U

C - G

C - G

G - C

U - A

C - G

G - C

A U G

G

U U U

U - A

A - U

G - C

C - G

G * U

A

U - A

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

G - C

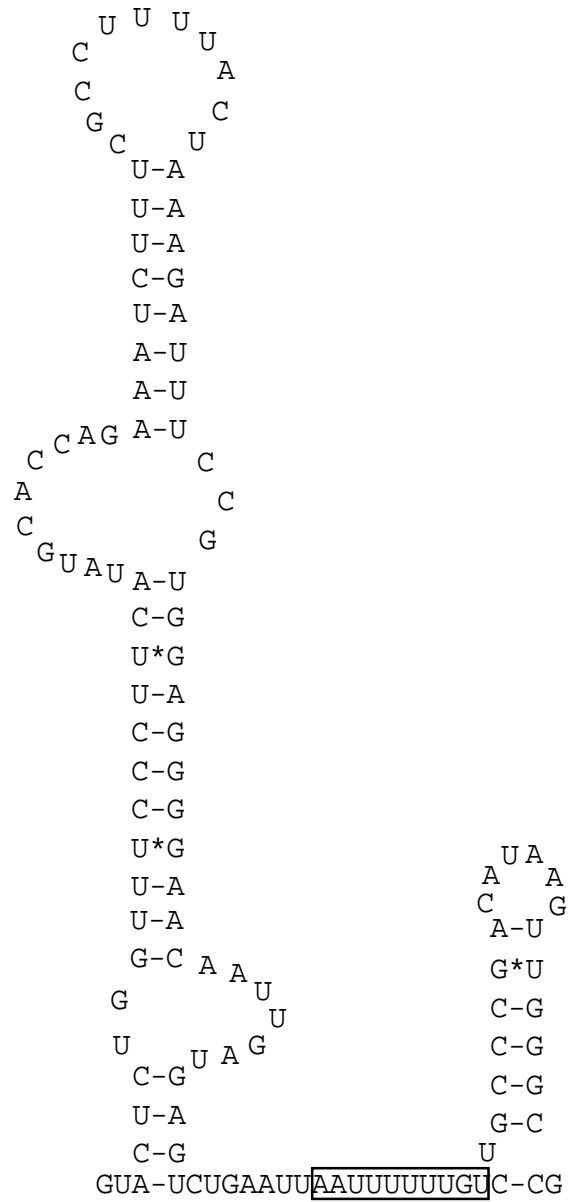
G - C

U4

G A U G A G
C C A U C U U G
A G - C A A U
A G A

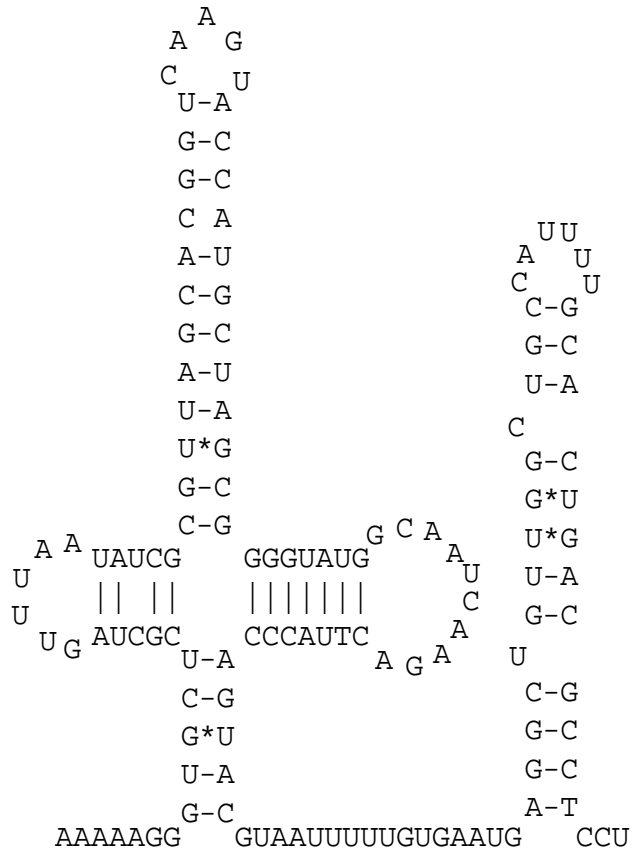
Figure S1C

Apis mellifera U6/U4 snRNAs



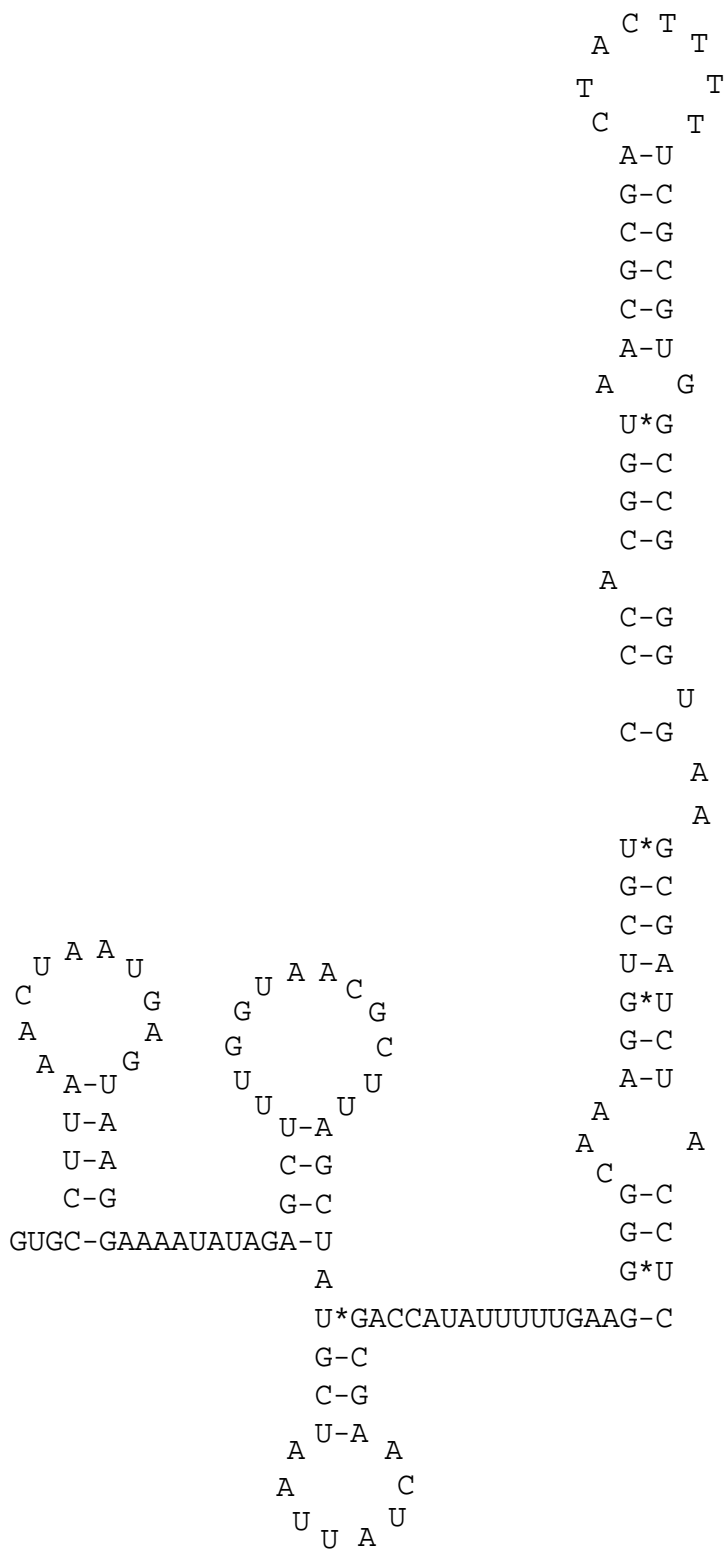
Apis mellifera U5 snRNA

Figure S1D



Apis mellifera u11 snRNA

Figure S1E



Apis mellifera U12 snRNA

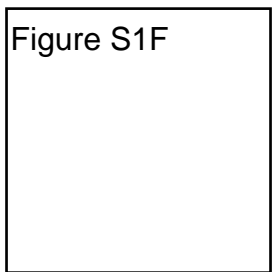
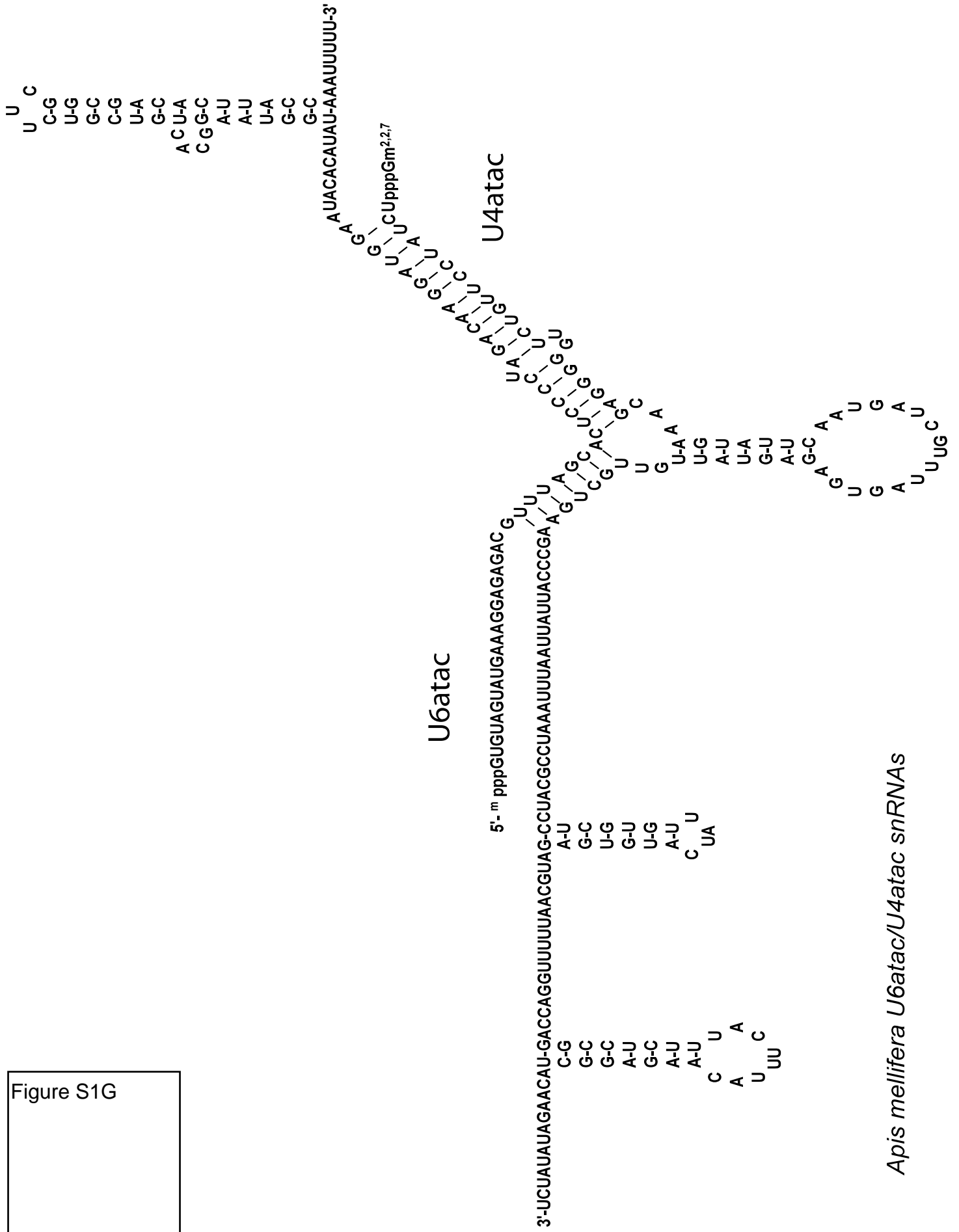


Figure S1G



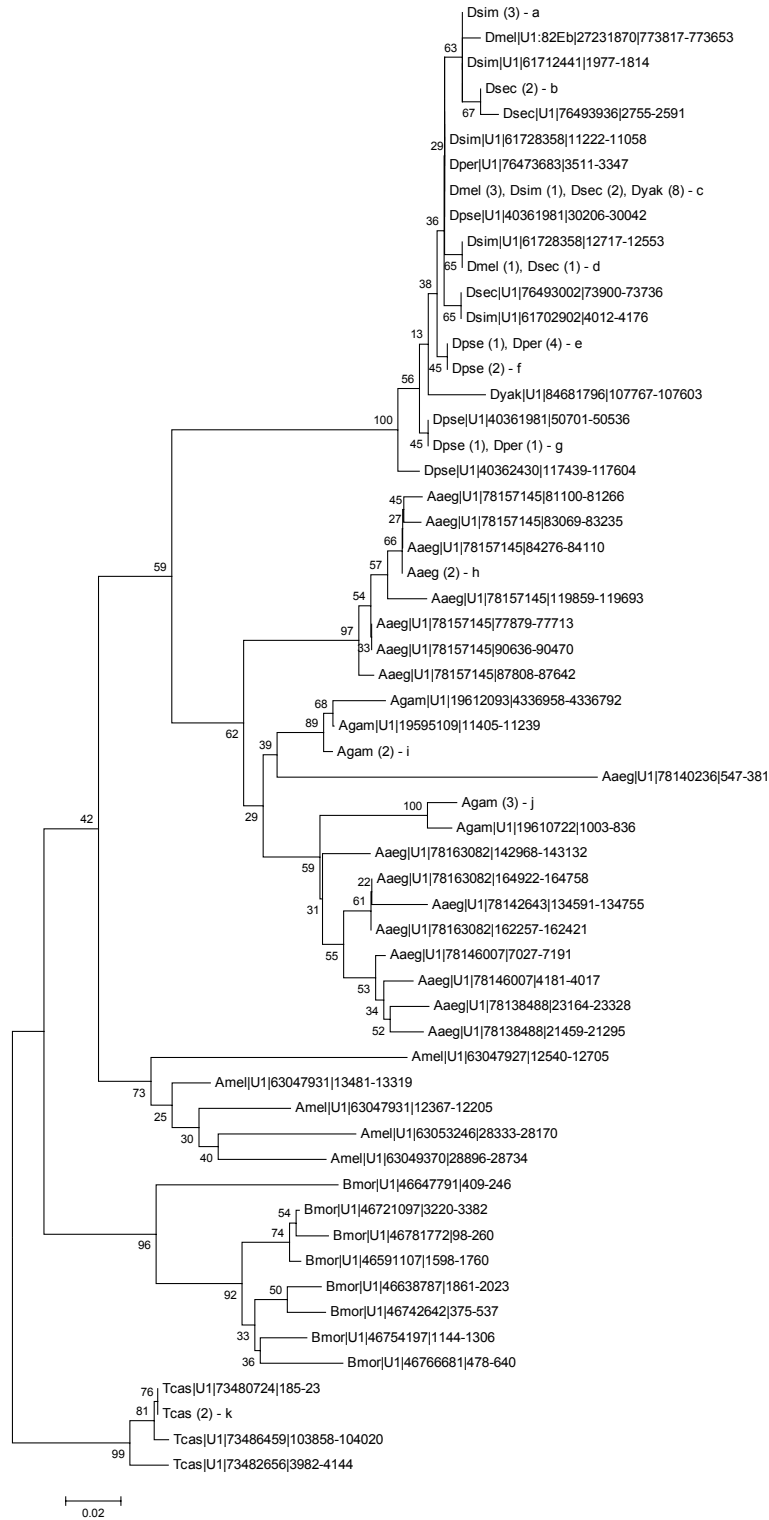


Figure S2.A. Neighbor-joining tree of U1 snRNA genes. Kimura two-parameter model with complete gap deletion, 157 sites, and 1000 bootstrap replicates were used to construct the tree. **Labels:** **a** - Dsim|U1|61699979|698-534 || Dsim|U1|61708433|767-603 || Dsim|U1|61709518|3559-3723; **b** - Dsec|U1|76493933|5388-5224 || Dsec|U1|76493940|8152-8316; **c** - Dmel|U1:21D|27232011|894003-893839 || Dmel|U1:95Cb|27232095|16076124-16076288 || Dmel|U1:95Ca|27232095|16107563-16107399 || Dsim|U1|61728361|4214-4378 || Dsec|U1|76495092|129629-129793 || Dsec|U1|76495092|144532-144368 || Dyak|U1|84680576|2211-2047 || Dyak|U1|84680578|27515-27679 || Dyak|U1|84680579|14246-14410 || Dyak|U1|84681616|65326-65490 || Dyak|U1|84681616|80561-80397 || Dyak|U1|84681616|82094-81930 || Dyak|U1|84681697|67313-67477 || Dyak|U1|84681697|68378-68214; **d** - Dmel|U1:95Cc|27232095|16074588-16074752 || Dsec|U1|76495092|128146-128310; **e** - Dpse|U1|40361981|54719-54554 || Dper|U1|76473680|2789-2624 || Dper|U1|76473680|6775-6610 || Dper|U1|76473683|4478-4643 || Dper|U1|76473683|6507-6672; **f** - Dpse|U1|40361981|31135-31300 || Dpse|U1|40361981|32822-32987; **g** - Dpse|U1|40361981|52625-52460 || Dper|U1|76473680|5026-4861; **h** - Aaeg|U1|78157145|74313-74147 || Aaeg|U1|78157145|75798-75632; **i** - Agam|U1|19595109|16783-16949 || Agam|U1|19612317|8915972-8915806; **j** - Agam|U1|19612007|689846-689681 || Agam|U1|19612007|701022-701187 || Agam|U1|19638324|369-204; **k** - Tcas|U1|73486304|72322-72484 || Tcas|U1|73486343|78341-78503

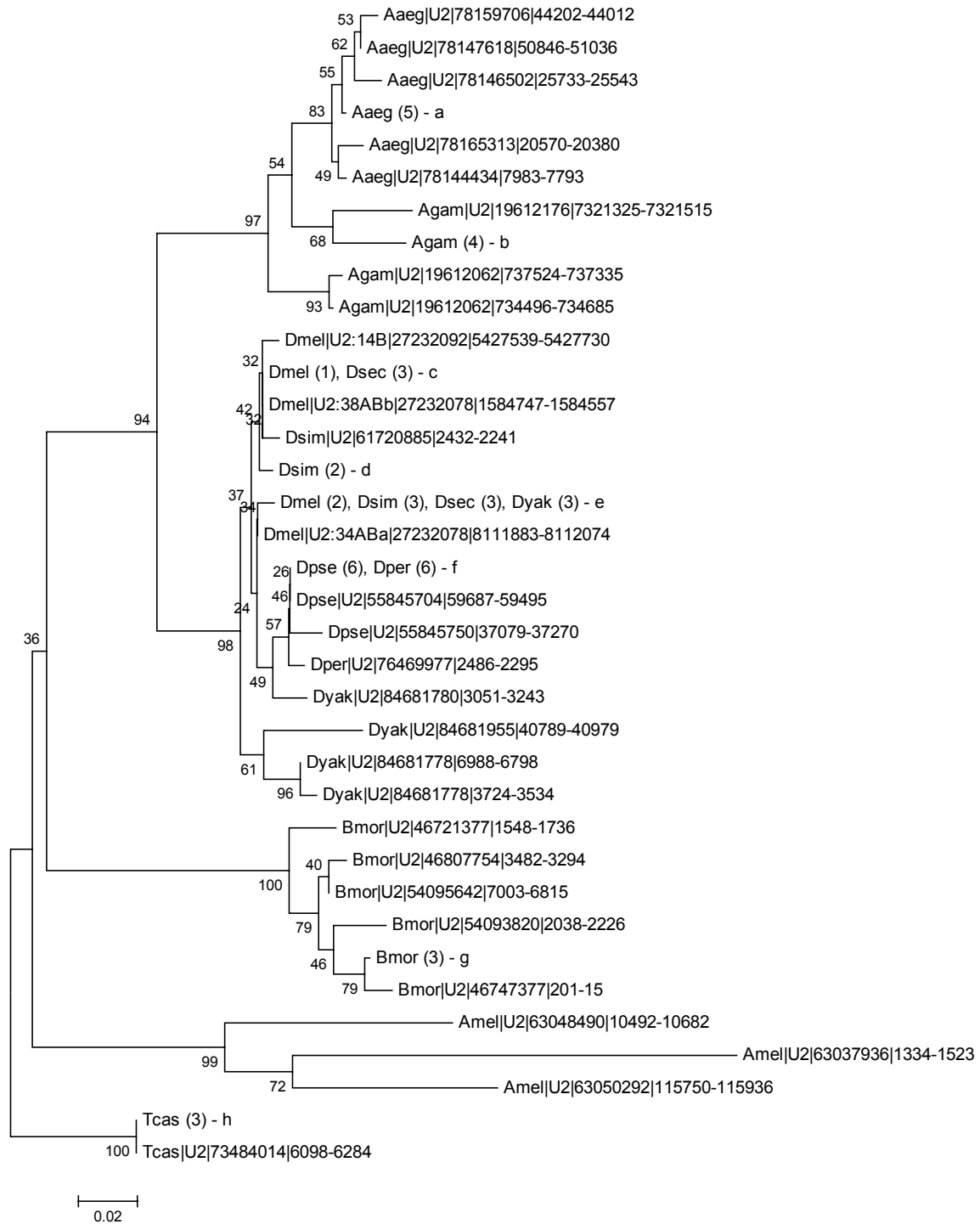


Figure S2.B. Neighbor-joining tree of U2 snRNA genes. Kimura two-parameter model with complete gap deletion, 181 sites, and 1000 bootstrap replicates were used to construct the tree. **Labels:** **a** - Aaeg|U2|78144433|2968-2778 || Aaeg|U2|78144434|5169-5359 || Aaeg|U2|78144435|3024-2834 || Aaeg|U2|78165313|11643-11833 || Aaeg|U2|78165313|17692-17882; **b** - Agam|U2|19610493|49504-49693 || Agam|U2|19610493|52866-52677 || Agam|U2|19612176|1258882-1259071 || Agam|U2|19612176|1261731-1261542; **c** - Dmel|U2:38ABa|27232078|1581588-1581779 || Dsec|U2|76491791|34481-34672 || Dsec|U2|76493802|411-602 || Dsec|U2|76493802|3796-3605; **d** - Dsim|U2|61721815|9492-9683 || Dsim|U2|61721815|12751-12560; **e** - Dmel|U2:34ABb|27232078|8108160-8107969 || Dmel|U2:34ABb|27232078|8108160-8107969 || Dsim|U2|61703534|912-1103 || Dsim|U2|61708251|413-604 || Dsim|U2|61708254|8366-8557 || Dsec|U2|76492803|12379-12188 || Dsec|U2|76492805|13157-12966 || Dsec|U2|76492805|16743-16934 || Dyak|U2|84681394|2008-1817 || Dyak|U2|84681737|509360-509551 || Dyak|U2|84681867|390179-389988; **f** - Dpse|U2|55838948|1994-1803 || Dpse|U2|55844316|8400-8591 || Dpse|U2|55844316|24870-25061 || Dpse|U2|55845750|39240-39431 || Dpse|U2|55845750|69357-69548 || Dpse|U2|55845750|72533-72342 || Dper|U2|76469969|16829-17020 || Dper|U2|76469969|19202-19393 || Dper|U2|76469969|50257-50448 || Dper|U2|76469969|55056-54865 || Dper|U2|76471473|25810-25619 || Dper|U2|76471473|42255-42064; **g** - Bmor|U2|46629215|1752-1940 || Bmor|U2|46705289|357-545 || Bmor|U2|46738190|1509-1697; **h** - Tcas|U2|73483022|7032-6846 || Tcas|U2|73484565|18776-18962 || Tcas|U2|73485930|35505-35691

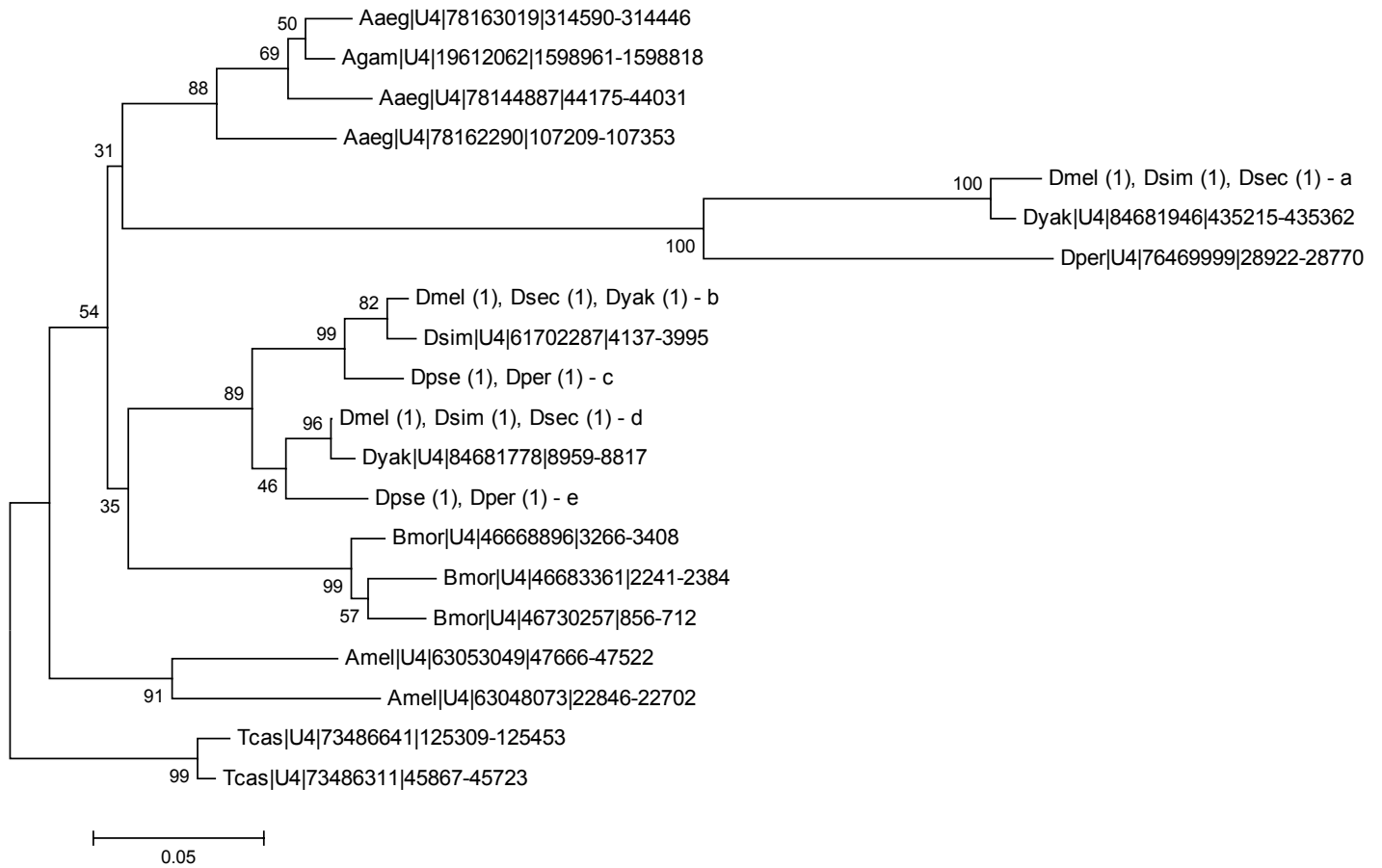


Figure S2.C. Neighbor-joining tree of U4 snRNA genes. Kimura two-parameter model with complete gap deletion, 139 sites, and 1000 bootstrap replicates were used to construct the tree. **Labels:** **a** - Dmel|U4:25F|27232078|15760002-15759855 || Dsim|U4|61722488|11691-11838 || Dsec|U4|76494020|8285-8432; **b** - Dmel|U4:39B|27232078|203181-203323 || Dsec|U4|76492647|25643-25501 || Dyak|U4|84681744|218060-217918; **c** - Dpse|U4|40360653|4903-5045|Dper|U4|76469999|49524-49382; **d** - Dmel|U4:38AB|27232078|1586659-1586517 || Dsim|U4|61721815|14639-14497 || Dsec|U4|76493803|64636-64778; **e** - Dpse|U4|40360900|493-351|Dper|U4|76469977|4883-4741