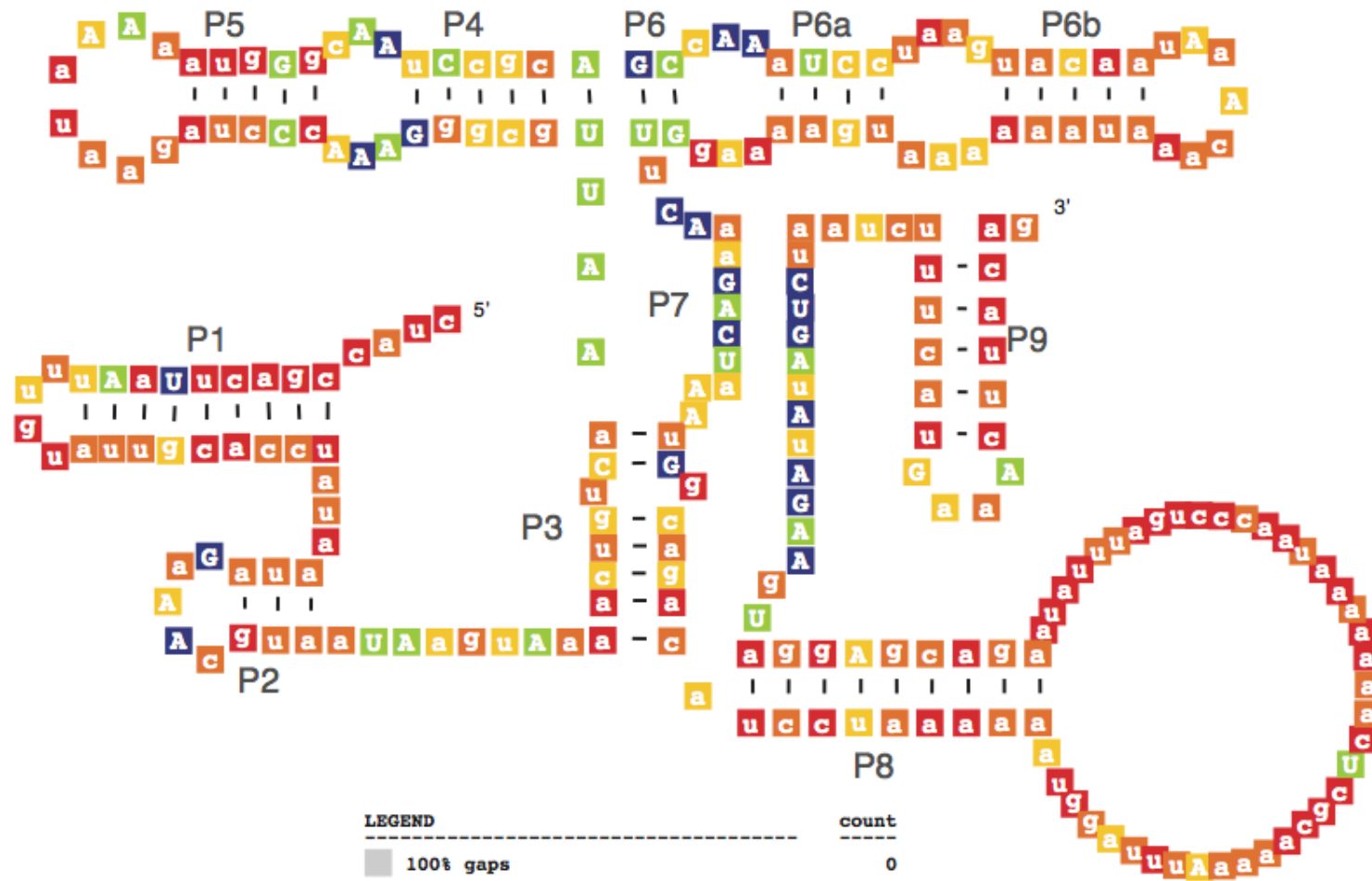


Searching for novel group I introns

02.03.15

model	#pos	#bps	#seqs	description
Group I Intron	251	55	12	information content per position

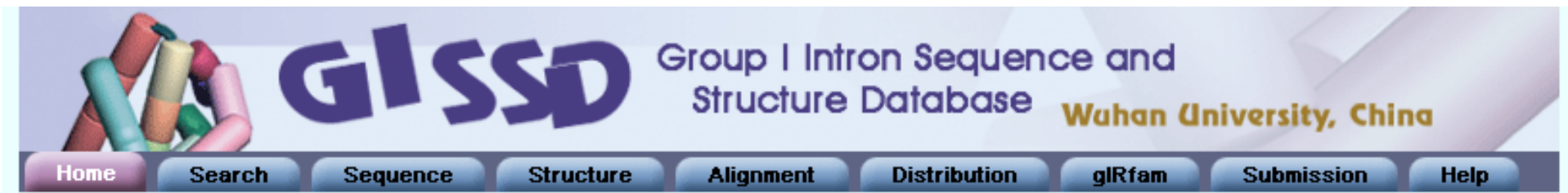


LEGEND	count
100% gaps	0
information content (bits):	
[0.000-0.400]	71
[0.400-0.800]	88
[0.800-1.200]	45
[1.200-1.600]	26
[1.600-1.990]	0
[1.990-2.000]	21

Consensus nucleotides (nt) are displayed, defined as the most frequent nt at each position.
Capitalized nts occur in ≥ 0.75 fraction of sequences that do not have a gap at the position.

		Group I Introns?	
		previously known	Infernal v1.1 predictions
EUKARYOTA	insects	-	+
	flatworms	-	+
	vertebrates	-	+
	jellyfish	+	+
	Choanoflagellata	-	+
	fungi	+	+
	plants	+	+
	ciliates	+	+
ARCHAEA	Euryarchaeota	-	-
	Crenarchaeota	-	+
	Thaumarchaeota	-	+
BACTERIA	Proteobacteria	+	+
	Cyanobacteria	+	+
	Aquifex	-	+
	Bacteroidetes	-	+
	Firmicutes	+	+
	Actinobacteria	-	+

GISSD: Group I Intron Sequence and Structure Database



```
<[GISSD]> esl-alistat -1 GISSD-plus-RF00028.sto
```

#	idx	name	format	nseq	alen	nres	small	large	avlen	%id
#	---	-----	-----	-----	-----	-----	-----	-----	-----	---
1		IA1	Stockholm	76	1471	44356	216	1261	583.6	45
2		IA2	Stockholm	15	503	4152	228	313	276.8	38
3		IA3	Stockholm	56	602	15808	205	374	282.3	46
4		IB1	Stockholm	42	503	12515	213	407	298.0	72
5		IB2	Stockholm	18	502	4360	163	320	242.2	39
6		IB3	Stockholm	7	373	1944	235	345	277.7	52
7		IB4	Stockholm	89	662	25124	203	392	282.3	44
8		IC1	Stockholm	837	1408	364902	251	615	436.0	39
9		IC2	Stockholm	32	452	10247	279	348	320.2	66
10		IC3	Stockholm	328	1584	83888	196	675	255.8	67
11		ID	Stockholm	17	386	4123	197	328	242.5	53
12		IE1	Stockholm	38	714	13763	308	479	362.2	60
13		IE2	Stockholm	56	710	22396	265	480	399.9	54
14		IE3	Stockholm	110	706	44653	252	481	405.9	57
15		RF00028	Stockholm	12	891	4378	206	488	364.8	34

Measuring specificity of GISSD models against training sequences using RNAVORE

		IA1	IA2	IA3	IB1	IB2	IB3	IB4	IC1	IC2	IC3	ID	IE1	IE2	IE3	RF0002
IA1	(76)	1.303	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IA2	(15)	-	1.000	-	-	-	-	-	-	-	-	-	-	-	-	-
IA3	(56)	-	-	1.000	-	-	-	-	-	-	-	-	-	-	-	-
IB1	(42)	-	-	-	0.952	0.024	-	0.024	-	-	-	-	-	-	-	-
IB2	(18)	-	-	-	-	0.944	-	-	-	0.111	-	-	-	-	-	-
IB3	(7)	-	-	-	-	-	1.000	-	-	-	-	-	-	-	-	-
IB4	(89)	-	-	-	0.011	0.022	0.011	0.978	-	-	-	-	-	-	-	-
IC1	(837)	-	-	-	-	-	-	-	0.994	-	-	-	-	0.006	-	-
IC2	(32)	-	-	-	-	-	-	-	-	1.000	-	-	-	-	-	-
IC3	(328)	-	-	-	-	-	-	-	-	-	0.997	-	-	-	-	0.003
ID	(17)	-	-	-	-	-	-	-	-	-	-	1.000	-	-	-	-
IE1	(38)	-	-	-	-	-	-	-	-	-	-	-	1.000	-	-	-
IE2	(56)	-	-	-	-	-	-	-	-	-	-	-	-	1.000	-	-
IE3	(110)	-	-	-	-	-	-	-	-	-	-	-	-	-	1.018	-
RF00028	(12)	0.250	0.083	-	-	-	-	0.083	0.250	-	0.167	-	-	-	-	0.167

Searching Rfamseq with GISSD models

type	# RF00028 seed seqs	# hits total	# hits common	# hits unique	total CPU hours
IA1	3	814	385	425	1076
IA2	1	1722	823	899	50
IA3		958	401	557	14
IB1		3949	1033	2916	32
IB2		1861	467	1394	31
IB3		479	136	343	40
IB4	1	5717	2400	3317	39
IC1	3	8475	5385	3090	24
IC2		4870	3858	1012	22
IC3	4	72692	66033	6659	136
ID		572	0	572	29
IE1		1305	10	1295	12
IE2		1377	8	1369	12
IE3		1379	1	1378	13
total	12	106170*	80940*	25226	1530
RF00028	-	71421	71421	-	125

* contains overlaps

Phylogenetic distribution of GISSD hits - viruses, bacteria and archaea

[illegible]

Phylogenetic distribution of GISSD hits - eukarya

domain	group	Rfam11	Rfam12	GISSD	IA-all	IB-all	IC-all	ID	IE-all	
eukaryota	nematoda	-	-	-	-	-	-	-	-	(1 at E=0.1 cutoff)
eukaryota	hexapoda	-	10	8	2	-	6	-	-	
eukaryota	crustacea	-	2	2	2	-	-	-	-	
eukaryota	chelicerata	-	-	-	-	-	-	-	-	(2 at E=1 cutoff)
eukaryota	platyhelminthes	-	48	458	-	-	458	-	-	
eukaryota	tunicata	-	-	-	-	-	-	-	-	(1 at E=1 cutoff)
eukaryota	craniata	6	52	19	4	7	7	1	-	
eukaryota	cnidaria	-	3	52	-	51	-	-	1	
eukaryota	placozoa	-	7	20	-	20	-	-	-	
eukaryota	ctenophora	-	-	1	-	-	-	1	-	
eukaryota	demospongiae	-	6	9	-	9	-	-	-	
eukaryota	choanoflagellida	-	1	3	1	-	-	-	2	
eukaryota	ichthyosporea	-	17	18	-	16	1	1	-	
eukaryota	fungi	1971	5010	9533	574	1533	5788	491	1147	
eukaryota	apusozoa	-	-	1	-	-	-	-	1	
eukaryota	stramenopiles	20	49	74	1	10	55	6	2	
eukaryota	alveolata	16	17	17	-	-	17	-	-	
eukaryota	rhodophyta	79	392	480	-	-	479	-	1	
eukaryota	viridiplantae	71943	71457	74198	330	2137	71524	40	167	
eukaryota	glaucocystophyceae	2	4	4	-	-	4	-	-	
eukaryota	euglenozoa	26	27	27	-	-	27	-	-	
eukaryota	heterolobosea	61	56	86	-	-	83	-	3	
eukaryota	amoebozoa	198	280	331	65	26	215	4	21	
eukaryota	rhizaria	37	60	67	-	-	62	-	5	
eukaryota	environmental samples	7	8	18	-	-	8	-	10	

Strategy for investigation of vertebrate hits

- Align each hit to its best matching GISSD model.
- Use that single sequence alignment as a SEED for a search against Rfamseq.
- Possible outcomes:
 - Novel hit, unlike all known group Is: most top hits are in related organisms
 - Contamination, misannotated: many nearly identical hits far away on tree

First group to investigate: vertebrates

An example hit in human

#	# bits	evalue	seqLabel	name	overlap	ncbiId	species	taxString
353.9	4.2e-109	FULL	AADC01089394.1	-	9606	Homo sapiens (human)	Eukaryota; Metazoa; Chordata	
353.9	4.2e-109	FULL	AADB02028502.1	-	9606	Homo sapiens (human)	Eukaryota; Metazoa; Chordata	
338.4	5.7e-104	FULL	ACBA01000116.1	-	565660	Enterococcus faecium 1,231,410	Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus	
335.5	5e-103	FULL	ADMM01000064.1	-	791161	Enterococcus faecium PC4.1	Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus	
332.8	4.1e-102	FULL	ABRY01000198.1	-	544875	Enterococcus faecium E1636	Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus	
203.0	3.1e-59	FULL	ACU001000009.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
202.1	6.1e-59	FULL	ACDC02000037.1	-	469599	Fusobacterium sp. 2_1_31	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
201.9	7.5e-59	FULL	ADEE01000005.1	-	469601	Fusobacterium sp. 21_1A	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
201.6	9.3e-59	FULL	ADEE01000038.1	-	469601	Fusobacterium sp. 21_1A	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
201.5	1e-58	FULL	ACU001000042.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
201.3	1.1e-58	FULL	ACU001000007.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
200.9	1.6e-58	FULL	ACU001000013.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
200.0	3.2e-58	FULL	ACU001000049.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
196.4	4.9e-57	FULL	ACU001000003.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
195.3	1.1e-56	FULL	ADEE01000034.1	-	469601	Fusobacterium sp. 21_1A	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
194.4	2.2e-56	FULL	AARG01000011.1	-	393480	Fusobacterium nucleatum subsp. polymorphum ATCC 10953	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
192.0	1.4e-55	FULL	ACU001000008.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
191.3	2.3e-55	FULL	ACU001000039.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
191.3	2.3e-55	FULL	ACU001000013.1	-	457403	Fusobacterium sp. 11_3_2	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
190.0	6.3e-55	FULL	ADDB01000034.1	-	693991	Fusobacterium nucleatum subsp. polymorphum F0401	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
190.0	6.3e-55	FULL	ADDB01000018.1	-	693991	Fusobacterium nucleatum subsp. polymorphum F0401	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
190.0	6.3e-55	FULL	ADDB01000017.1	-	693991	Fusobacterium nucleatum subsp. polymorphum F0401	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
190.0	6.3e-55	FULL	ADDB01000009.1	-	693991	Fusobacterium nucleatum subsp. polymorphum F0401	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
190.0	6.3e-55	FULL	ADDB01000008.1	-	693991	Fusobacterium nucleatum subsp. polymorphum F0401	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
189.3	1.1e-54	FULL	ADEE01000031.1	-	469601	Fusobacterium sp. 21_1A	Bacteria; Fusobacteriia; Fusobacteriaceae; Fusobacterium	
188.9	1.5e-54	FULL	AJ269494.1	-	1496	Clostridium difficile	Bacteria; Firmicutes; Clostridia; Clostridiaceae; Clostridium	
188.6	1.8e-54	FULL	ADLJ01000044.1	-	742733	Clostridium citroniae WAL-17108	Bacteria; Firmicutes; Clostridia; Clostridiaceae; Clostridium	
188.3	2.4e-54	FULL	FN668941.1	-	699034	Clostridium difficile BI1	Bacteria; Firmicutes; Clostridia; Clostridiaceae; Clostridium	

Summarizing vertebrate single sequence searches

#idx #---	type ---	self-score -----	#2 score -----	self-#2 -----	#2 taxonomy -----
1	IA1	270.9	270.9	0.0	Eukaryota; Fungi; Dikarya;
2	IA2	353.9	353.9	0.0	Eukaryota; Metazoa; Chordata; Craniata; ... Homo;
3	IA2	353.9	353.9	0.0	Eukaryota; Metazoa; Chordata; Craniata; ... Homo;
4	IB1	198.7	198.7	0.0	Eukaryota; Fungi; Dikarya;
5	IB2	258.7	239.7	19.0	Eukaryota; Fungi; Dikarya;
6	IB4	262.4	67.4	195.0	Eukaryota; Metazoa; Chordata; Craniata; ... Rhamphochromis;
7	IB4	277.5	95.0	182.5	Eukaryota; Fungi; Dikarya;
8	IB4	310.3	61.0	249.3	Eukaryota; Metazoa; Chordata; Craniata; ... Rhamphochromis;
9	IB4	275.3	275.3	0.0	Eukaryota; Fungi; Dikarya;
10	IB4	371.0	237.7	133.3	Eukaryota; Fungi; Fungi incertae sedis;
11	IC1	481.5	463.8	17.7	Eukaryota; Fungi; Dikarya;
12	IC2	168.1	153.4	14.7	Eukaryota; Fungi; Dikarya;
13	IC3	649.4	649.4	0.0	Eukaryota; Viridiplantae; Streptophyta;
14	IC3	550.3	550.3	0.0	Eukaryota; Viridiplantae; Streptophyta;
15	IC3	312.3	280.4	31.9	Bacteria; Proteobacteria;
16	IC3	281.7	227.2	54.5	Bacteria; Proteobacteria;
17	ID	370.6	234.5	136.1	Eukaryota; Fungi; Fungi incertae sedis;

All but four IB4 candidates and one ID candidates are almost certainly contamination.

Remaining 5 hits are in the cichlid *Rhamphochromis esox*

- These are all probably contamination:
 - one was 100% identical to a plant sequence
 - two were clearly mitochondrial fungal sequence (protein homology)
 - two were judged probably fungal due to similarity to other fungal group Is.

2nd group to investigate: archaea

		Group I Introns?	
		previously known	Infernal v1.1 predictions
EUKARYOTA	insects	-	+
	flatworms	-	+
	vertebrates	-	+
	jellyfish	+	+
	Choanoflagellata	-	+
	fungi	+	+
	plants	+	+
	ciliates	+	+
ARCHAEA	Euryarchaeota	-	-
	Crenarchaeota	-	+
	Thaumarchaeota	-	+
BACTERIA	Proteobacteria	+	+
	Cyanobacteria	+	+
	Aquifex	-	+
	Bacteroidetes	-	+
	Firmicutes	+	+
	Actinobacteria	-	+

Annotation of crenarchaeota candidate

AB201308.1 Unclutured Candidatus Nitrosocaldus sp. DNA, fosmid clone: No.45-H-12.
(Archaea; Crenarchaeota; Candidatus Nitrosocaldus; environmental samples.)

SSU_rRNA_archaea	RF01959	AB201308.1	-	cm	1	1477	19565	21032	+ no	1 0.61	47.8	1535.8	0 !	-
LSU_rRNA_archaea	RF02540	AB201308.1	-	cm	3	2067	21195	23185	+ 5'&3'	5 0.59	67.4	1802.6	0 !	-
IB4	-	AB201308.1	-	cm	2	147	23176	23333	+ 5'&3'	5 0.44	0.0	47.1	4.9e-11 !	-
HEG (NCBI)							23335	23853						
IB4	-	AB201308.1	-	cm	135	201	23847	23926	+ 5'&3'	5 0.46	0.0	31.6	1.1e-06 !	-
LSU_rRNA_archaea	RF02540	AB201308.1	-	cm	2013	2978	23924	24897	+ 5'&3'	5 0.61	30.5	932.7	1.6e-293 !	-

~Automated annotation:

21000 22000 23000 24000 25000
|-----|-----|-----|-----|-----|

~Automated annotation:

|-----|-----|-----|-----|-----|
21192 LSU 23185 IB4 23333 HEG 23853 IB4 23926 LSU 24897
-10 +1 -7 -3

Manually tweaked

|-----|-----|-----|-----|-----|
21192 LSU 23184* IB4 23334* HEG 23853 IB4 23941* LSU 24897
0 0 0 0

Annotation of thaumarchaeota candidate

>AP011865.1 Candidatus Caldiarchaeum subterraneum DNA, partial sequence, fosmid clone: JFF029_A12.
Archaea; Thaumarchaeota; unclassified Thaumarchaeota; Candidatus Caldiarchaeum.

LSU_rRNA_archaea	RF02540	AP011865.1	-	cm	3	1223	18823	17588	- 5'&3'	5 0.63	73.7	1008.7	0 !	-
HEG (NCBI)							17584	16859						
LSU_rRNA_archaea	RF02540	AP011865.1	-	cm	1223	2067	16896	16078	- 5'&3'	5 0.63	36.7	715.6	4.8e-225 !	-
IB4	-	AP011865.1	-	cm	2	122	16087	15967	- 5'&3'	5 0.51	0.0	49.7	5.5e-12 !	-
HEG (NCBI)							16047	14959						
IB4	-	AP011865.1	-	cm	114	235	15083	14931	- 5'&3'	5 0.51	0.0	39.5	3.9e-09 !	-
LSU_rRNA_archaea	RF02540	AP011865.1	-	cm	2030	2547	14945	14419	- 5'&3'	5 0.63	15.5	453.7	3.1e-142 !	-
HEG (NCBI)							14457	13819						
LSU_rRNA_archaea	RF02540	AP011865.1	-	cm	2544	2980	13818	13344	- 5'&3'	5 0.65	20.2	435.7	1.5e-136 !	-

19000 18000 17000 16000 15000 14000 13000
|-----|-----|-----|-----|-----|-----|

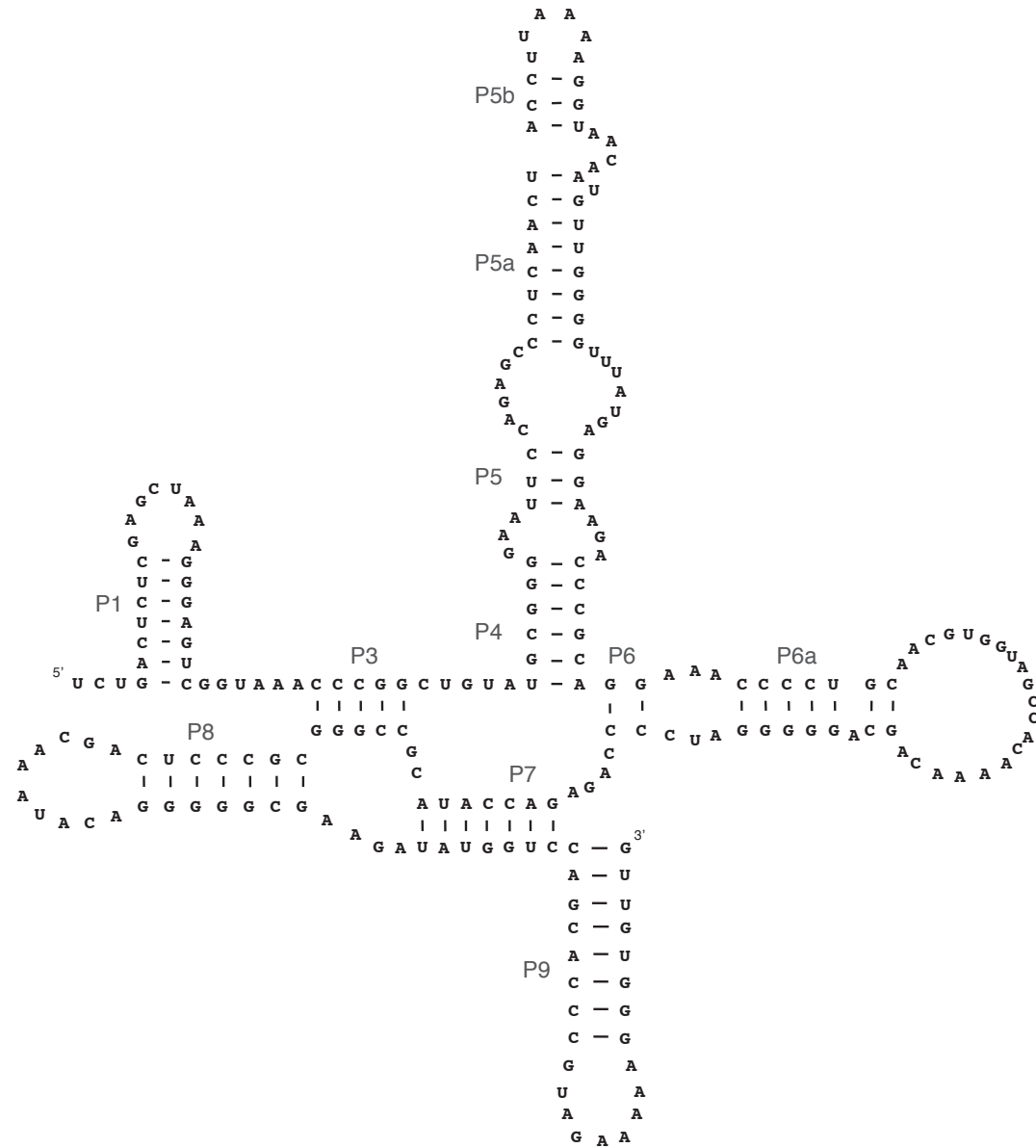
~Automated annotation:

-----	LSU	-----	HEG	-----	LSU	-----	IB4	-----	HEG	-----	IB4	--	LSU	--	HEG	---	LSU	----	
18823		17588		16859		16078	15967				14959	14931		14419			13819		13344
		+5		-38		-10	-81				-125	-15		-39			0		

Manually tweaked:

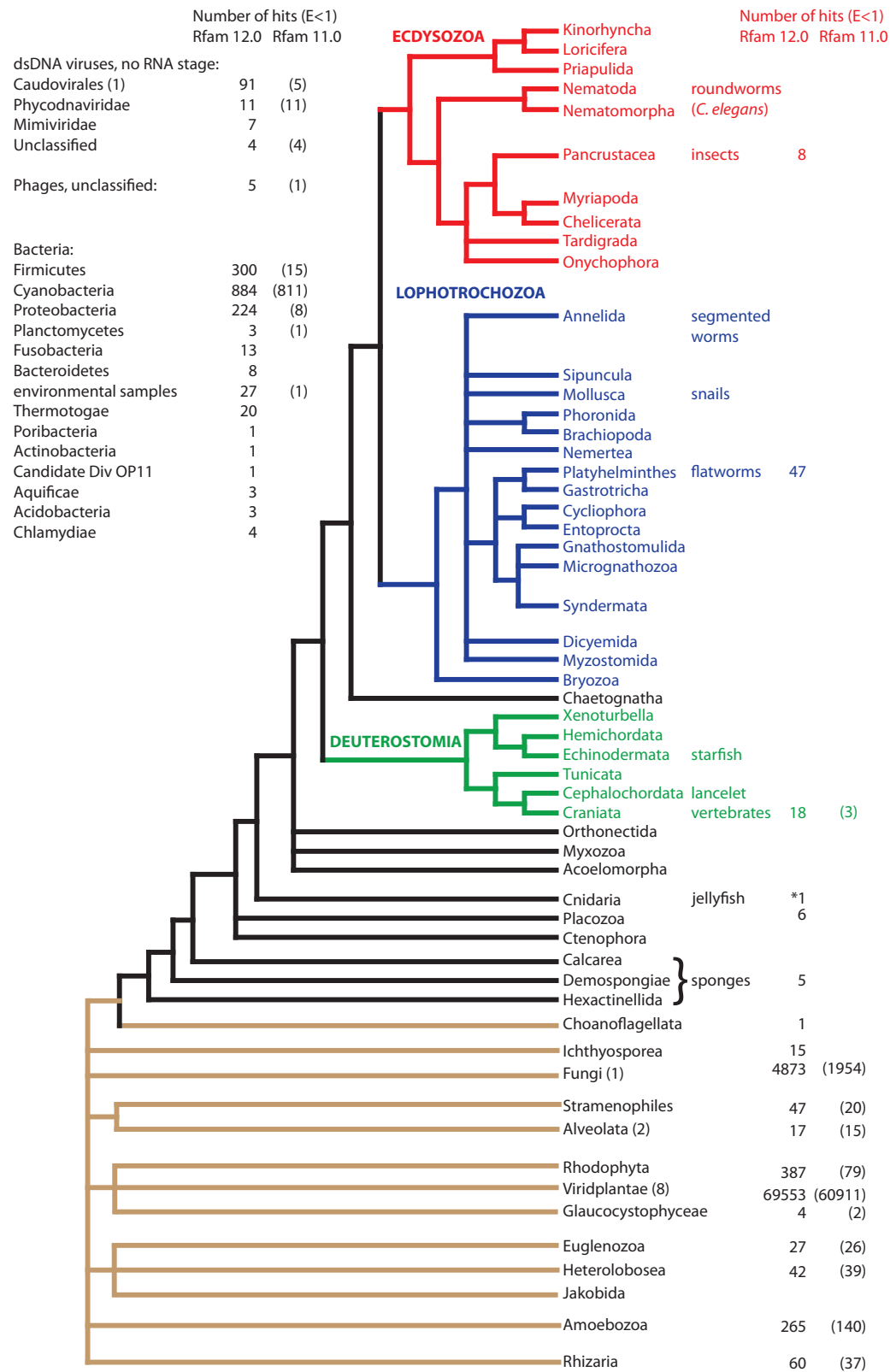
-----	LSU	-----	HEG	-----	LSU	-----	IB4	-----	HEG	-----	IB4	--	LSU	--	HEG	---	LSU	----	
18821*		17588		16859		16078	15952*				15045*	14946*		14419			13819		13344
		+5		-38		-10	0				0	0		0			0		

Thaumarchaeota group I structure



Phylogenetic distribution of GISSD hits - eukarya

domain	group	Rfam11	Rfam12	GISSD	IA-all	IB-all	IC-all	ID	IE-all	
eukaryota	nematoda	-	-	-	-	-	-	-	-	(1 at E=0.1 cutoff)
eukaryota	hexapoda	-	10	8	2	-	6	-	-	
eukaryota	crustacea	-	2	2	2	-	-	-	-	
eukaryota	chelicerata	-	-	-	-	-	-	-	-	(2 at E=1 cutoff)
eukaryota	platyhelminthes	-	48	458	-	-	458	-	-	
eukaryota	tunicata	-	-	-	-	-	-	-	-	(1 at E=1 cutoff)
eukaryota	craniata	6	52	19	4	7	7	1	-	
eukaryota	cnidaria	-	3	52	-	51	-	-	1	
eukaryota	placozoa	-	7	20	-	20	-	-	-	
eukaryota	ctenophora	-	-	1	-	-	-	1	-	
eukaryota	demospongiae	-	6	9	-	9	-	-	-	
eukaryota	choanoflagellida	-	1	3	1	-	-	-	2	
eukaryota	ichthyosporea	-	17	18	-	16	1	1	-	
eukaryota	fungi	1971	5010	9533	574	1533	5788	491	1147	
eukaryota	apusozoa	-	-	1	-	-	-	-	1	
eukaryota	stramenopiles	20	49	74	1	10	55	6	2	
eukaryota	alveolata	16	17	17	-	-	17	-	-	
eukaryota	rhodophyta	79	392	480	-	-	479	-	1	
eukaryota	viridiplantae	71943	71457	74198	330	2137	71524	40	167	
eukaryota	glaucocystophyceae	2	4	4	-	-	4	-	-	
eukaryota	euglenozoa	26	27	27	-	-	27	-	-	
eukaryota	heterolobosea	61	56	86	-	-	83	-	3	
eukaryota	amoebozoa	198	280	331	65	26	215	4	21	
eukaryota	rhizaria	37	60	67	-	-	62	-	5	
eukaryota	environmental samples	7	8	18	-	-	8	-	10	



Platyhelminthes

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
1	AAWT01055253.1	1.4e-29	IC1	-----	-----	361.9	317.9	44.0	Eukaryota; Metazoa; Platyhelminthes;
2	AAWT01074732.1	5e-24	IC1	-----	-----	378.0	324.5	53.5	Eukaryota; Metazoa; Platyhelminthes;
3	AAWT01061738.1	5.6e-29	IC1	-----	-----	415.1	407.5	7.6	Eukaryota; Metazoa; Platyhelminthes;
4	AAWT01045369.1	2.1e-17	IC1	-----	-----	238.8	236.2	2.6	Eukaryota; Metazoa; Platyhelminthes;
5	AAWT01070677.1	3.9e-30	IC1	-----	-----	414.2	405.1	9.1	Eukaryota; Metazoa; Platyhelminthes;
6	AAWT01028849.1	4.6e-25	IC1	-----	-----	374.0	352.9	21.1	Eukaryota; Metazoa; Platyhelminthes;
7	AAWT01019700.1	5.3e-20	IC1	-----	-----	390.1	331.6	58.5	Eukaryota; Metazoa; Platyhelminthes;

Arthropoda

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
1	ADND01290505.1	9.6e-20	IA2	-----	-----	381.4	175.1	206.3	unclassified sequences; metagenomes;
2	AAAB01004644.1	6.1e-19	IA2	-----	-----	349.9	219.8	130.1	Eukaryota; Choanoflagellida; Codonosigidae;
3	ADND01071445.1	2.3e-09	IA2	-----	-----	251.8	160.0	91.8	Eukaryota; Metazoa; Arthropoda; Crustacea; Maxillop
4	ABLF02001275.1	5.4e-07	IA2	-----	-----	440.2	96.0	344.2	Viruses; dsDNA viruses, no RNA stage;
5	AFFH01005192.1	1.5e-27	IC3	-----	-----	334.2	327.9	6.3	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta;
6	AFFF01004110.1	1.4e-27	IC3	-----	-----	335.2	329.6	5.6	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta;
7	AFFE01003852.1	3.5e-27	IC3	-----	-b---	334.6	331.7	2.9	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta;
8	AFPQ01002822.1	6e-22	IC3	-----	-----	337.0	196.9	140.1	Bacteria; Proteobacteria;
9	AFFG01001574.1	1.4e-20	IC3	-----	-----	337.2	337.2	0.0	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta;
10	AFFF01006283.1	1.4e-20	IC3	-----	-----	337.2	337.2	0.0	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta;

Cnidaria

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
----	-----	-----	----	-----	-----	-----	-----	-----	-----
1	AB289575.1	3.7e-08	IB4	-----	-----	241.4	241.4	0.0	Eukaryota; Metazoa; Cnidaria;
2	FJ345448.1	3.2e-09	IB4	-----	-----	241.8	241.8	0.0	Eukaryota; Metazoa; Cnidaria;
3	U36784.1	8.3e-06	IB1	-----	-----	394.8	394.8	0.0	Eukaryota; Metazoa; Cnidaria;
4	DQ831341.1	1.8e-14	IB1	-----	-----	261.7	215.6	46.1	Eukaryota; Metazoa; Cnidaria;
5	AF000023.1	8.3e-06	IB1	-B----	-B---	394.8	394.8	0.0	Eukaryota; Metazoa; Cnidaria;

Placozoa

#		GISSD				self	#2	diff		
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy	
1	DQ889456.1	2e-21	IB4	-B----	-B---	490.9	195.3	295.6	Eukaryota; Metazoa; Placozoa; unclassified Placozoa.	
2	DQ112541.1	1.4e-11	IB4	-B----	-B---	614.7	187.3	427.4	Eukaryota; Metazoa; Placozoa; unclassified Placozoa.;;	
3	DQ889456.1	8.1e-07	IB4	-B----	-B---	309.2	158.8	150.4	Eukaryota; Metazoa; Placozoa; unclassified Placozoa.	
4	DQ889457.1	1.1e-11	IB4	-B----	A----	612.7	568.2	44.5	Eukaryota; Metazoa; Placozoa; unclassified Placozoa.	
5	AB538868.2	9.4e-10	IB1	-B----	A----	611.6	590.4	21.2	Eukaryota; Metazoa; Placozoa; unclassified Placozoa.	

Ctenophora

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
----	-----	-----	----	-----	-----	-----	-----	-----	-----
1	AGCP01017868.1	6e-07	ID	-----	-----	471.9	312.5	159.4	Eukaryota; Metazoa; Ctenophora;

Demospongiae

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
1	HM032741.1	2.2e-19	IB4	-----	-----	265.0	265.0	0.0	Eukaryota; Metazoa; Porifera; Demospongiae;
2	HQ269356.1	2.3e-16	IB4	-B-----	-B---	271.9	266.0	5.9	Eukaryota; Metazoa; Porifera; Demospongiae;
3	EU237487.1	9.4e-18	IB4	-B-----	-B---	267.3	185.7	81.6	Eukaryota; Metazoa; Porifera; Demospongiae;
4	HQ269352.1	2.3e-17	IB4	-B-----	-B---	270.6	266.3	4.3	Eukaryota; Metazoa; Porifera; Demospongiae;
5	HM032739.1	2.2e-19	IB4	-----	-----	265.0	265.0	0.0	Eukaryota; Metazoa; Porifera; Demospongiae;

Choanoflagellata

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
1	AC198640.1	3.2e-20	IA2	-----	-----	345.0	214.3	130.7	Eukaryota; Metazoa; Arthropoda;
2	HQ026767.1	1.1e-44	IE1	-----	--E--	482.0	128.6	353.4	Eukaryota; Viridiplantae; Chlorophyta;
3	EU011942.1	9.4e-45	IE2	-----	--E--	667.2	167.5	499.7	Eukaryota; Fungi; Dikarya;

Ichthyosporea

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
1	AF538047.1	8.4e-25	IB1	-----	-----	419.8	216.5	203.3	Eukaryota; Ichthyosporea; Ichthyophonida;
2	AF538043.1	2.7e-12	IB1	-----	-----	431.4	79.3	352.1	Eukaryota; Viridiplantae; Streptophyta;
3	AF538044.1	8.8e-35	IB1	-B----	-----	374.4	191.3	183.1	Eukaryota; Ichthyosporea; Ichthyophonida;
4	AF538047.1	8e-32	IB1	-----	-----	393.3	196.6	196.7	Eukaryota; Ichthyosporea; Ichthyophonida;
5	AF538044.1	5.7e-20	ID	-B----	-----	295.0	116.4	178.6	Eukaryota; Fungi; Dikarya;

Apusozoa

#		GISSD				self	#2	diff	
#idx	sequence	E-value	type	SSU?	LSU?	score	score	self-#2	#2 taxonomy
----	-----	-----	----	-----	-----	-----	-----	-----	-----
1	GU001168.1	2.2e-53	IE2	--E---	--E--	626.8	163.5	463.3	Eukaryota; Fungi; Dikarya;

Phylogenetic distribution of GISSD hits - viruses, bacteria and archaea

[illegible]

Phylogenetic distribution of GISSD hits - eukarya

domain	group	Rfam11	Rfam12	GISSD	IA-all	IB-all	IC-all	ID	IE-all	
eukaryota	nematoda	-	-	-	-	-	-	-	-	(1 at E=0.1 cutoff)
eukaryota	hexapoda	-	10	8	2	-	6	-	-	
eukaryota	crustacea	-	2	2	2	-	-	-	-	
eukaryota	chelicerata	-	-	-	-	-	-	-	-	(2 at E=1 cutoff)
eukaryota	platyhelminthes	-	48	458	-	-	458	-	-	
eukaryota	tunicata	-	-	-	-	-	-	-	-	(1 at E=1 cutoff)
eukaryota	craniata	6	52	19	4	7	7	1	-	
eukaryota	cnidaria	-	3	52	-	51	-	-	1	
eukaryota	placozoa	-	7	20	-	20	-	-	-	
eukaryota	ctenophora	-	-	1	-	-	-	1	-	
eukaryota	demospongiae	-	6	9	-	9	-	-	-	
eukaryota	choanoflagellida	-	1	3	1	-	-	-	2	
eukaryota	ichthyosporea	-	17	18	-	16	1	1	-	
eukaryota	fungi	1971	5010	9533	574	1533	5788	491	1147	
eukaryota	apusozoa	-	-	1	-	-	-	-	1	
eukaryota	stramenopiles	20	49	74	1	10	55	6	2	
eukaryota	alveolata	16	17	17	-	-	17	-	-	
eukaryota	rhodophyta	79	392	480	-	-	479	-	1	
eukaryota	viridiplantae	71943	71457	74198	330	2137	71524	40	167	
eukaryota	glaucocystophyceae	2	4	4	-	-	4	-	-	
eukaryota	euglenozoa	26	27	27	-	-	27	-	-	
eukaryota	heterolobosea	61	56	86	-	-	83	-	3	
eukaryota	amoebozoa	198	280	331	65	26	215	4	21	
eukaryota	rhizaria	37	60	67	-	-	62	-	5	
eukaryota	environmental samples	7	8	18	-	-	8	-	10	