

Comparative ncRNA Detection in Archaea

Sarah J. Berkemer^{1,2,4,7,*}, Christian Höner zu Siederdisen^{1,2,4}, Fabian Amman², Axel Wintsche^{3,4}, Sebastian Will^{1,4}, Ivo L. Hofacker^{2,5,6}, Sonja J. Prohaska^{3,4}, Peter F. Stadler^{1,2,4,6-9,*}

¹ Bioinformatics Group, Department of Computer Science, Univ. Leipzig, Germany

² Institute for Theoretical Chemistry, Univ. Vienna, Austria

³ Computational EvoDevo Group, Department of Computer Science, Univ. Leipzig, Germany

⁴ Interdisciplinary Center for Bioinformatics, University of Leipzig, Germany

⁵ Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, Univ. Vienna, Austria

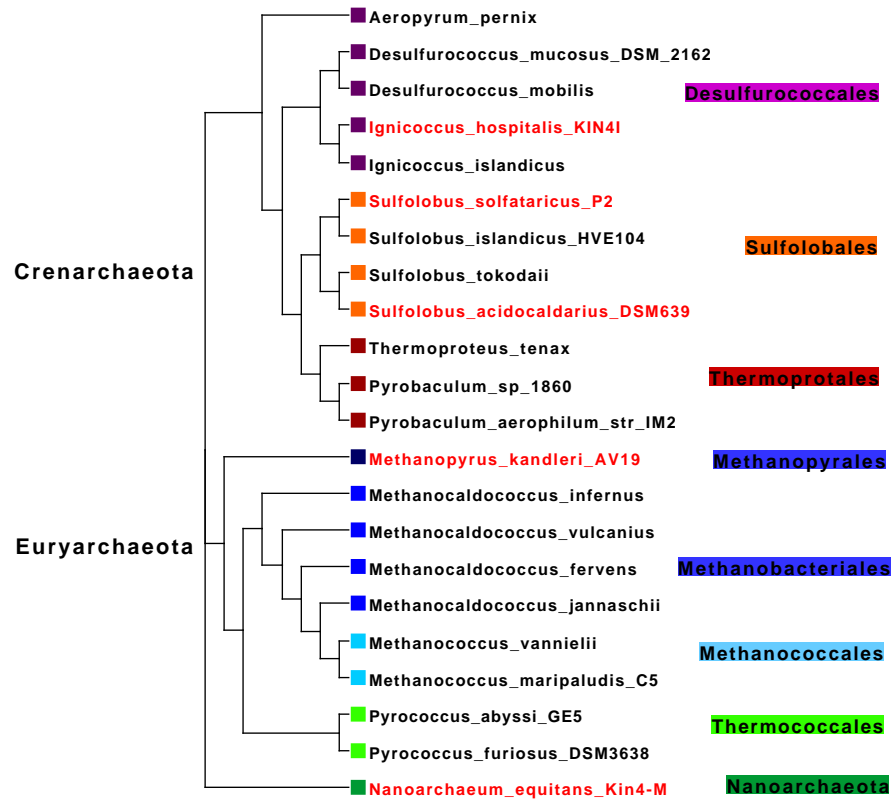
⁶ Center for non-coding RNA in Technology and Health, Univ. Copenhagen, Denmark

⁷ Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

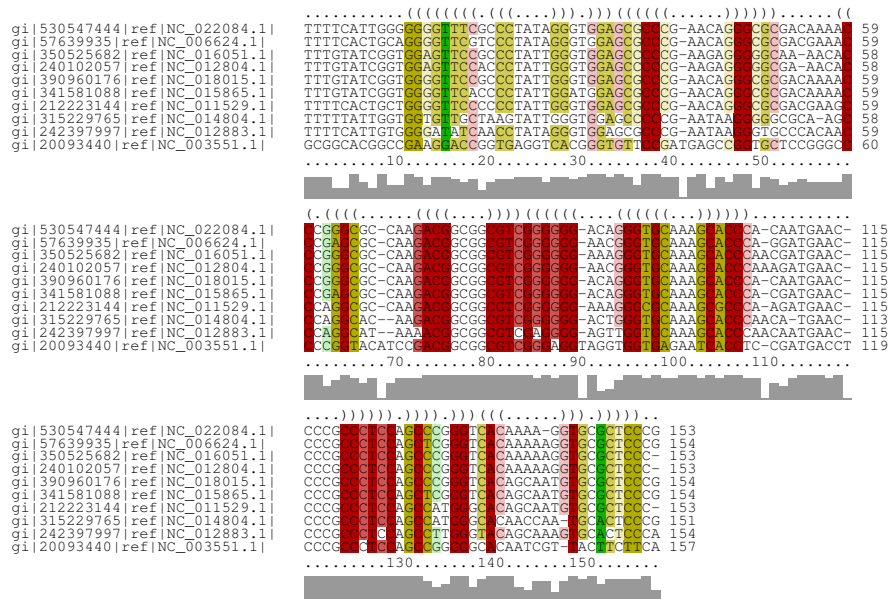
⁸ Fraunhofer Institut für Zelltherapie und Immunologie, Leipzig, Germany

⁹ Santa Fe Institute, Santa Fe, USA

E-mail: studla@bioinf.uni-leipzig.de



Supplementary Figure 1. Bayesian phylogenetic tree of life of archaeal species. Red labels indicate the archaeal species used in this study. All species are classified into the domains Crenarchaeota and Euryarchaeota. Colors indicate the subdomains.



Supplementary Figure 2. Novel putative circRNA from *M. kandleri* (at 1500955-1501112). Alignment and consensus RNA secondary structure with homolog sequences in other archaea; the homologs were identified by blast search at e-value cut-off 0.01 as described in the main text. The figure furthermore reports the genome accession codes of the homolog sequences. The consensus structure and the output figure were generated using RNAalifold [1].

Supplementary Table 1. Comparison between tRNA introns according to tRNAscan results for *Methanopyrus kandleri*, *Sulfolobus solfataricus*, and *Sulfolobus acidocaldarius* and cm search results.

Species	tRNAscan			cm Search	
	tRNA Type	Intron Begin	Intron End	Rank	Bit Score
Methanopyrus kandleri	Trp	55,108	55,183	5.	21.5
Methanopyrus kandleri	Pro	1,499,308	1,499,322	124.	14.7
Methanopyrus kandleri	Pseudo	1,659,640	1,659,691	2.	25.3
Methanopyrus kandleri	Phe	1,639,150	1,639,119	not found	
Methanopyrus kandleri	Cys	1,062,337	1,062,317	79.	15.4
Methanopyrus kandleri	Asn	881,764	881,738	3.	25.2
Methanopyrus kandleri	Met	382,127	382,092	4.	22.2
Sulfolobus solfataricus	Asn	49,381	49,394	23.	12.5
Sulfolobus solfataricus	Met	466,263	466,279	209.	9.1
Sulfolobus solfataricus	Leu	637,204	637,218	not found	
Sulfolobus solfataricus	Leu	837,058	837,073	146.	10.0
Sulfolobus solfataricus	Ile	913,737	913,726	19.	12.6
Sulfolobus solfataricus	Pro	898,333	898,313	242.	8.9
Sulfolobus solfataricus	Thr	789,727	789,713	172.	9.4
Sulfolobus solfataricus	Tyr	642,512	642,500	not found	
Sulfolobus solfataricus	Ser	641,001	640,978	801.	6.7
Sulfolobus solfataricus	Arg	290,939	290,927	566.	7.3
Sulfolobus solfataricus	Arg	249,046	249,032	219.	9.0
Sulfolobus solfataricus	Thr	206,385	206,373	156.	9.7
Sulfolobus solfataricus	Met	184,841	184,817	13.	13.3
Sulfolobus solfataricus	Lys	138,407	138,386	1420.	5.7
Sulfolobus solfataricus	Lys	122,617	122,595	2.	17.3
Sulfolobus solfataricus	Trp	72,831	72,767	1321.	5.9
Sulfolobus acidocaldarius	Ser	512,669	512,693	62.	9.6
Sulfolobus acidocaldarius	Leu	512,819	512,833	5.	14.4
Sulfolobus acidocaldarius	Met	515,240	515,257	140.	8.3
Sulfolobus acidocaldarius	Lys	608,795	608,816	1.	16.7
Sulfolobus acidocaldarius	Pro	1,096,684	1,096,704	145.	8.3
Sulfolobus acidocaldarius	Met	1,166,860	1,166,879	10.	13.4
Sulfolobus acidocaldarius	Asn	2,181,266	2,181,254	129.	8.5
Sulfolobus acidocaldarius	Gly	2,160,121	2,160,107	988.	5.5
Sulfolobus acidocaldarius	Arg	1,241,011	1,240,995	4.	14.8
Sulfolobus acidocaldarius	Thr	1,188,440	1,188,425	73.	9.4
Sulfolobus acidocaldarius	Leu	716,510	716,493	12.	13.0
Sulfolobus acidocaldarius	Cys	610,584	610,569	not found	
Sulfolobus acidocaldarius	Lys	607,184	607,157	11.	13.0
Sulfolobus acidocaldarius	Thr	563,576	563,550	8.	13.9
Sulfolobus acidocaldarius	Phe	458,889	458,872	not found	
Sulfolobus acidocaldarius	Gly	458,680	458,666	not found	
Sulfolobus acidocaldarius	Arg	138,765	138,749	9.	13.7
Sulfolobus acidocaldarius	Trp	49,256	49,197	21.	11.4

Supplementary Table 2. Comparison between circularized RNA according to RNA-seq read analysis and predicted BHB elements for *Methanopyrus kandleri*. The first two columns give the genomic position of the left and right circularizing bases. “Read Count” gives the number of reads supporting this particular circularization event. For each locus, which was reported to be associated with an BHB element, the Rank in the genomic screen and its bit score is provided. The last column describes the genomic neighborhood. If it is within an annotated gene, its locus tag is given. For loci in intergenic regions the distance to the upstream and downstream gene is given.

RNA-seq			cm Search		Genomic Surrounding
L. junc.	R. junc.	#Count	Rank	Bit Score	ncbi locus tag
69,921	69,985	2,065	–	–	MK0074 \Leftarrow 25nt 77nt \Rightarrow MK0075
91,822	91,904	2,992	7,845.	8.1	\Leftarrow MK0099 \Rightarrow
205,318	205,387	12,182	–	–	MK0213 \Leftarrow 21nt 102nt \Rightarrow MK0214
219,317	219,379	2,412	–	–	\Leftarrow MK0233 \Rightarrow
271,148	271,216	10,072	3,546.	9.8	\Leftarrow MK0280 \Rightarrow
361,063	361,125	1,809	–	–	\Leftarrow MK0371 \Rightarrow
384,798	384,945	634	–	–	MK0403 \Leftarrow 665nt 201nt \Rightarrow MK0404
459,461	459,532	9,830	9,039.	7.7	\Leftarrow MK0498 \Rightarrow
519,288	519,358	22,145	15,966.	6.1	\Leftarrow MK0556 \Rightarrow
520,778	520,845	13,484	–	–	MK0557 \Leftarrow 6nt 269nt \Rightarrow MK0558
755,163	755,226	304	–	–	\Leftarrow MK0794 \Rightarrow
790,521	790,585	1,161	19,971.	5.2	MK0830 \Leftarrow 6nt 142nt \Rightarrow MK0831
993,172	993,238	50,802	–	–	\Leftarrow MK1033 \Rightarrow
1,104,263	1,104,330	6,223	17,300.	5.7	MK1128 \Leftarrow 573nt 260nt \Rightarrow MK1129
1,238,104	1,238,178	9	4,630.	9.2	\Leftarrow MK1253 \Rightarrow
1,417,298	1,417,370	9,371	–	–	MK1390 \Leftarrow 36nt 224nt \Rightarrow MK1391
1,417,378	1,417,443	5,658	14,942.	6.3	MK1390 \Leftarrow 116nt 144nt \Rightarrow MK1391
1,444,847	1,444,927	2	–	–	\Leftarrow MK1415 \Rightarrow
1,500,955	1,501,112	2,648	–	–	MK1479 \Leftarrow 88nt 205nt \Rightarrow MK1480
1,506,611	1,506,673	8,396	6,984.	8.4	\Leftarrow MK1486 \Rightarrow

Supplementary Table 3. Circulare RNA in *Sulfolobus solfataricus* [2] are tested for recovery in the cm screen using the consensus model in `glocal` mode. Additionally, the analysis was redone using the homology loci, if available, in *Sulfolobus acidocaldarius*. The homology search eas conducted with the `GotohScan` program [3]. The “Start” and “End” columns refer to position in the genomes NC_002754 and NC_007181, respectively.

Name	Sulfolobus solfataricus				Sulfolobus acidocaldarius			
	RNA-seq Start	RNA-seq End	cm Search Rank	cm Search Bit Score	RNA-seq Start	RNA-seq End	cm Search Rank	cm Search Bit Score
5S rRNA/SSOr02	77,945	78,067	863.	6.5	1,293,914	1,294,035	–	–
16S rRNA/SSOr03	871,658	873,216	–	–	1,108,641	1,107,094	–	–
23S rRNA/SSOr04	873,334	876,429	–	–	1,106,947	1,103,875	–	–
tRNA-Trp/SSOt04	72,767	72,831	1,321.	5.9	49,197	49,262	548.	6.4
tRNA-Lys/SSOt07	138,386	138,407	1,420.	5.7	607,138	607,204	11.	13.0
tRNA-Met/SSOt11	184,817	184,841	13.	13.3	–	–	–	–
tRNA-Pro/SSOt42	898,313	898,333	242.	8.9	1,096,702	1,096,684	145.	8.3
tRNA-Ser/SSOt33	640,978	641,001	1,094.	6.2	512,691	512,669	62.	9.6
C/D box sR106	285,707	285,760	–	–	2,179,509	2,179,560	–	–
C/D box Sso-180	362,308	362,369	–	–	669,556	669,612	–	–
C/D box sR133	442,392	442,417	–	–	–	–	–	–
C/D box sR102	563,241	563,296	–	–	1,388,934	1,388,984	–	–
C/D box Sso-sR8	647,783	647,833	601.	7.1	1,885,917	1,885,967	–	–
C/D box Sso-sR4	666,143	666,186	1,779.	5.3	–	–	–	–
C/D box Sso-sR10	794,186	794,240	–	–	1,152,443	1,152,394	–	–
C/D box Sso-207	816,021	816,075	–	–	–	–	–	–
C/D box SSOs02	829,352	829,405	–	–	1,117,732	1,117,685	–	–
C/D box Sso-sR12	2,189,397	2,189,456	–	–	–	–	–	–
C/D box sR105	2,237,915	2,237,962	–	–	217,040	217,087	–	–
H/ACA box sR109	59,5510	595,579	308.	8.3	458,983	459,052	351.	7.0
ncRNA	442,786	442,854	–	–	–	–	–	–
ncRNA	722,538	722,578	3005.	4.4	417,691	417729	–	–
Sso-117	1,576,633	1,576,671	–	–	–	–	–	–
Sso-109	1,927,228	1,927,258	–	–	–	–	–	–
7S rRNA/SSOr01	49,977	50023	839.	6.6	72,370	72,326	763.	5.9
Sso-214	105,148	105,181	–	–	–	–	–	–
RNase P	224,732	224,765	–	–	586,242	586,211	–	–
Sso-83	581,818	581,860	–	–	–	–	–	–
ncRNA	1,275,500	127,5567	–	–	–	–	–	–
SSO0393	343,138	343,264	–	–	650,082	650,206	–	–
Intergenic region	871,573	871657	1,107.	6.2	1,108,730	1,108,647	–	–
Intergenic region	873,215	873331	–	–	1,107,091	1,106,972	–	–
SSO0389	335,563	335,635	–	–	–	–	–	–
SSO0845	725,923	726,085	–	–	–	–	–	–
SSO2359	2,154,297	2,154,322	–	–	782,915	782,892	–	–
SSO2619	2,385,872	2,385,901	–	–	738,654	738,681	–	–
SSO2642	2,404,146	2,404,146	–	–	2,114,638	2,114,694	–	–

Supplementary Table 4. Comparison between bit scores for generic BHB models (g), and two more strict models (s_{dp} , s_{zp}) that conform more to the canonical BHB element. The models were tested with sequences from *N. equitans* confirmed as box C/D snoRNAs [4].

#	g	s_{dp}	s_{zp}
1	31.14	25.96	-32.28
3	20.53	-9.27	-69.67
4	28.08	1.86	-58.19
8	30.27	-0.83	-51.5
9	36.21	26.09	-6.51
10	43.96	23.66	-23.38
11	29.72	16.35	-51.46
12	36.43	4.87	-47.21
13	18.22	16.35	-44.68
15	13.23	16.56	-20.78
16	31.7	16.48	-42.31
17	20.22	18.35	-25.63
18	22.22	6.13	-67.05
19	15.49	11.34	-41.4
20	25.26	8.26	-64.38
21	34.06	19.65	-50.62
23	15.51	7.73	-25
24	-15.44	-11.13	-69.78
25	17.66	11.44	-53.01
26	6.92	12.72	-44.16
intron	-47.03	-48.4	-97.97
mean	19.7	8.3	-46.99
sd	19.8	16.5	20.8
% $b > 0$	0.9	0.81	0.0

Supplementary Table 5. circRNA candidates of *M. kandleri* and *S. acidolaris* with putatively conserved stable secondary structures as predicted by RNAz. As described in the main text, circRNA candidates were identified by mapping RNA-Seq data, homologs were located in all archaeal genomes, potential homologs were aligned, and subsequently evaluated by RNAz. The table lists the candidates that are predicted as putative structural RNAs together with the number of homologous sequences in the locus alignment and the assigned RNAz class probability.

circRNA candidate locus			Number of Seqs	RNAz class probability
Accession	Start	Stop		
NC_003551.1	1500955	1501112	10	0.9992
NC_007181.1	1214595	1214683	4	0.680306
NC_007181.1	1254692	1254799	4	0.938321
NC_007181.1	1107137	1107281	39	0.998730
NC_007181.1	1803656	1803770	3	0.722219
NC_007181.1	183648	183733	5	0.703805
NC_007181.1	1995955	1996059	3	0.619096
NC_007181.1	553923	554080	4	0.799874
NC_007181.1	753148	753230	4	0.982281
NC_007181.1	766362	766509	4	0.646306
NC_007181.1	773268	773364	4	0.644060
NC_007181.1	86425	86509	3	0.554138

Supplementary Table 6. Box C/D snoRNA of *M. kandleri* from [5] showing evidence for a BHB element after aligning to the covariance model for box C/D snoRNA sequences based on sequences of *N. equitans* from [4]. Start and stop enclose just the box C/D snoRNA sequences. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences.

Start	Stop	Strand	Name	transrealign
1815	1879	+	C/D box snoRNA 1	y
42955	43022	-	C/D box snoRNA 2	y
44728	44789	+	C/D box snoRNA 3	y
69897	69962	+	C/D box snoRNA 4	n
111152	111214	+	C/D box snoRNA 6	n
137488	137553	+	C/D box snoRNA 7	y
139707	139764	+	C/D box snoRNA 8	n
146560	146637	+	C/D box snoRNA 9	y
147313	147372	+	C/D box snoRNA 10	n
170072	170146	-	C/D box snoRNA 11	n
170850	170923	+	C/D box snoRNA 12	y
172650	172747	-	C/D box snoRNA 13	n
181467	181526	+	C/D box snoRNA 14	y
203475	203540	+	C/D box snoRNA 15	y
205387	205458	-	C/D box snoRNA 17	y
227948	228036	+	C/D box snoRNA 19	y
262147	262212	+	C/D box snoRNA 20	y
271071	271136	-	C/D box snoRNA 21	y
271151	271217	-	C/D box snoRNA 22	y
325495	325560	-	C/D box snoRNA 24	n
325521	325579	-	C/D box snoRNA 25	y
327758	327832	+	C/D box snoRNA 26	y
346543	346604	-	C/D box snoRNA 28	y
360535	360607	-	C/D box snoRNA 29	y
383186	383259	+	C/D box snoRNA 31	n
384878	384947	+	C/D box snoRNA 32	y
409873	409936	+	C/D box snoRNA 33	y
459462	459533	-	C/D box snoRNA 34	n
462530	462593	+	C/D box snoRNA 35	y
515532	515599	-	C/D box snoRNA 37	y
518923	518988	-	C/D box snoRNA 38	y
519204	519281	-	C/D box snoRNA 39	y
519290	519359	-	C/D box snoRNA 40	y
520779	520847	+	C/D box snoRNA 41	y
524051	524115	+	C/D box snoRNA 43	y
524115	524177	+	C/D box snoRNA 44	y
524175	524238	+	C/D box snoRNA 45	y
537577	537643	+	C/D box snoRNA 46	y
561189	561251	+	C/D box snoRNA 47	y
582279	582357	+	C/D box snoRNA 48	y
603202	603273	+	C/D box snoRNA 49	y
605153	605234	+	C/D box snoRNA 50	y
627727	627790	-	C/D box snoRNA 52	y
700325	700386	-	C/D box snoRNA 53	n
729573	729639	+	C/D box snoRNA 54	y
751600	751664	+	C/D box snoRNA 55	y
755168	755227	-	C/D box snoRNA 56	y
766722	766794	-	C/D box snoRNA 57	y
772696	772766	+	C/D box snoRNA 58	y
776072	776149	+	C/D box snoRNA 59	y
779256	779317	+	C/D box snoRNA 60	n
779315	779376	-	C/D box snoRNA 61	n
790520	790586	-	C/D box snoRNA 62	y
825703	825775	+	C/D box snoRNA 63	y
830254	830325	+	C/D box snoRNA 64	y
839773	839862	-	C/D box snoRNA 65	y

Supplementary Table 7. Box C/D snoRNA of *M. kandleri* from [5] showing evidence for a BHB element after aligning to the covariance model for box C/D snoRNA sequences based on sequences of *N. equitans* from [4]. Start and stop enclose just the box C/D snoRNA sequences. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences.

Start	Stop	Strand	Name	transrealign
843279	843335	+	C/D box snoRNA 66	n
858313	858382	+	C/D box snoRNA 67	y
879751	879813	+	C/D box snoRNA 68	y
879938	879993	-	C/D box snoRNA 69	y
892485	892553	+	C/D box snoRNA 70	y
900585	900653	+	C/D box snoRNA 71	y
915062	915133	-	C/D box snoRNA 73	y
961063	961126	+	C/D box snoRNA 74	y
963110	963171	-	C/D box snoRNA 75	y
993170	993242	+	C/D box snoRNA 76	y
1003081	1003142	+	C/D box snoRNA 77	n
1022562	1022623	-	C/D box snoRNA 78	y
1022628	1022698	-	C/D box snoRNA 79	y
1023968	1024031	+	C/D box snoRNA 80	y
1043677	1043750	-	C/D box snoRNA 81	n
1048100	1048194	+	C/D box snoRNA 82	y
1065992	1066050	-	C/D box snoRNA 84	n
1067504	1067564	-	C/D box snoRNA 85	n
1073361	1073433	-	C/D box snoRNA 86	y
1077825	1077895	+	C/D box snoRNA 87	y
1094668	1094726	-	C/D box snoRNA 88	n
1104271	1104334	-	C/D box snoRNA 89	y
1118640	1118700	+	C/D box snoRNA 90	n
1164281	1164371	+	C/D box snoRNA 92	n
1165264	1165330	-	C/D box snoRNA 93	n
1166529	1166612	+	C/D box snoRNA 94	y
1178363	1178424	-	C/D box snoRNA 95	n
1182419	1182494	+	C/D box snoRNA 96	y
1183490	1183549	-	C/D box snoRNA 97	n
1187688	1187763	+	C/D box snoRNA 98	n
1204769	1204834	+	C/D box snoRNA 99	n
1221349	1221417	-	C/D box snoRNA 100	y
1231253	1231340	+	C/D box snoRNA 101	n
1231699	1231762	+	C/D box snoRNA 102	y
1232371	1232433	+	C/D box snoRNA 103	n
1233144	1233210	+	C/D box snoRNA 104	y
1233411	1233470	-	C/D box snoRNA 105	y
1263413	1263479	+	C/D box snoRNA 107	y
1383364	1383424	-	C/D box snoRNA 108	y
1386498	1386561	+	C/D box snoRNA 109	n
1415602	1415673	-	C/D box snoRNA 110	n
1417300	1417370	-	C/D box snoRNA 111	y
1417379	1417442	-	C/D box snoRNA 112	y
1417448	1417512	-	C/D box snoRNA 113	y
1492448	1492515	+	C/D box snoRNA 114	y
1493125	1493192	+	C/D box snoRNA 115	y
1495330	1495399	+	C/D box snoRNA 116	y
1506611	1506674	+	C/D box snoRNA 117	y
1508733	1508798	-	C/D box snoRNA 118	y
1510490	1510576	-	C/D box snoRNA 119	n
1531085	1531148	+	C/D box snoRNA 120	y
1561232	1561294	+	C/D box snoRNA 121	n
1587746	1587811	-	C/D box snoRNA 122	y
1605995	1606060	+	C/D box snoRNA 123	n
1607219	1607287	-	C/D box snoRNA 124	y
1633733	1633798	+	C/D box snoRNA 126	y

Supplementary Table 8. Box C/D snoRNA of *M. kandleri* from [5] where no BHB element could be found after aligning to the covariance model for box C/D snoRNA sequences based on sequences of *N. equitans* from [4]. Start and stop enclose just the box C/D snoRNA sequences. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences.

Start	Stop	Strand	Name	transrealign
91821	91905	+	C/D box snoRNA5	y
205310	205380	-	C/D box snoRNA16	n
219316	219380	+	C/D box snoRNA18	n
274089	274184	+	C/D box snoRNA23	y
333110	333179	+	C/D box snoRNA27	y
361062	361156	-	C/D box snoRNA30	n
509783	509880	+	C/D box snoRNA36	y
520852	520933	+	C/D box snoRNA42	y
605300	605373	+	C/D box snoRNA51	y
902800	902861	-	C/D box snoRNA72	n
1055548	1055621	+	C/D box snoRNA83	n
1157340	1157415	+	C/D box snoRNA91	y
1234309	1234398	-	C/D box snoRNA106	n
1607250	1607305	+	C/D box snoRNA125	y

Supplementary Table 9. Box C/D snoRNA sequences of *N. equitans* from [4] used to build a the covariance model for cetection of box C/D snoRNA with a BHB element. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences. Sequences were enumerated following the enumeration in [4].

Start	Stop	Strand	Name	transrealign
46922	46979	-	C/D box snoRNA2	y
54028	54085	-	C/D box snoRNA3	n
90898	90954	-	C/D box snoRNA4	y
113577	113628	-	C/D box snoRNA5	n
144953	145004	-	C/D box snoRNA6	y
145962	146019	-	C/D box snoRNA7	y
164113	164168	-	C/D box snoRNA8	y
248187	248239	-	C/D box snoRNA9	n
282915	282973	-	C/D box snoRNA10	y
315990	316045	-	C/D box snoRNA12	y
323041	323094	-	C/D box snoRNA14	n
328897	328949	-	C/D box snoRNA16	y
337448	337505	-	C/D box snoRNA18	y
371860	371917	-	C/D box snoRNA21	y
375354	375406	-	C/D box snoRNA22	y

Supplementary Table 10. Box C/D snoRNA sequences of *N. equitans* from [4] where a BHB motif could be detected. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences. Sequences were enumerated following the enumeration in [4].

Start	Stop	Strand	Name	transrealign
45779	45831	+	C/D box snoRNA1	n
297634	297687	+	C/D box snoRNA11	n
318915	318968	+	C/D box snoRNA13	y
334552	334611	+	C/D box snoRNA17	y
359412	359470	+	C/D box snoRNA19	y
362755	362808	+	C/D box snoRNA20	y
382458	382509	+	C/D box snoRNA23	y
401223	401280	+	C/D box snoRNA25	y
403464	403522	+	C/D box snoRNA26	n

Supplementary Table 11. Box C/D snoRNA sequences of *N. equitans* from [4] where no clear BHB motif could be detected. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences. Sequences were enumerated following the enumeration in [4].

Start	Stop	Strand	Name	transrealign
325371	325424	+	C/D box snoRNA15	n
384218	384278	+	C/D box snoRNA24	y

Supplementary Table 12. Box C/D snoRNA sequences of *S. solfataricus* from [6, 7] where a BHB motif could be detected. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences. Sequences with (*) couldn't be predicted in [4] such that now name was given. We kept the order of sequences and so the name consists of a number in between.

Start	Stop	Strand	Name	transrealign
362293	362384	+	C/D box snoRNA 180	y
794171	794255	+	C/D box snoRNA 10	y
829337	829420	+	C/D box snoRNA 02	y
2189382	2189471	+	C/D box snoRNA 12	n
1274759	1274879	-	C/D box snoRNA n2*	n
2003526	2003688	+	C/D box snoRNA 67	n
2111320	2111468	-	C/D box snoRNA 67b*	n

Supplementary Table 13. Box C/D snoRNA sequences of *S. solfataricus* from [2, 7] where no clear BHB motif could be detected. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences. Sequences with (*) couldn't be predicted in [4] such that now name was given. We kept the order of sequences and so the name consists of a number in between.

Start	Stop	Strand	Name	transrealign
285692	285775	+	C/D box snoRNA 106	n
442377	442432	+	C/D box snoRNA 133	n
563226	563311	+	C/D box snoRNA 102	y
816006	816090	+	C/D box snoRNA 207	y
2237900	2237977	+	C/D box snoRNA 105	y
22518	22604	+	C/D box snoRNA 195	n
163939	164015	+	C/D box snoRNA 65	n
442348	442449	+	C/D box snoRNA n1*	n
590308	590410	-	C/D box snoRNA 101	n
837791	837886	-	C/D box snoRNA 125	n
885646	885840	-	C/D box snoRNA 110	y
1401961	1402198	-	C/D box snoRNA n3*	y
2369460	2369600	+	C/D box snoRNA 93	n

Supplementary Table 14. Box C/D snoRNA sequences of *S. acidocaldarius* from [6] [2] where a BHB motif could be detected. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences.

Start	Stop	Strand	Name	transrealign
2179492	2179578	+	C/D box snoRNA 18	n
793941	794027	-	C/D box snoRNA 17	y
2196878	2196963	+	C/D box snoRNA 16	n
1400376	1400445	+	C/D box snoRNA 15	y
1117434	1117521	-	C/D box snoRNA 14	y
345672	345759	+	C/D box snoRNA 13	y
86078	86167	+	C/D box snoRNA 12	n
131771	131859	+	C/D box snoRNA 11	y
217020	217107	+	C/D box snoRNA 10	y
42115	42207	+	C/D box snoRNA 9	y
2175730	2175814	-	C/D box snoRNA 8	y
393832	393923	-	C/D box snoRNA 7	y
368172	368254	+	C/D box snoRNA 6	y
449443	449530	-	C/D box snoRNA 5	y
2217096	2217181	-	C/D box snoRNA 4	n
1117670	1117755	+	C/D box snoRNA 3	y
1075489	1075575	+	C/D box snoRNA 2	y
1220543	1220628	+	C/D box snoRNA 1	n
2179494	2179575	+	C/D box snoRNA 106	n
669541	669627	+	C/D box snoRNA 102	y
1388919	1388999	+	C/D box snoRNA 180	y
217025	217102	+	C/D box snoRNA 105	y

Supplementary Table 15. Box C/D snoRNA sequences of *S. acidocaldarius* from [6] [2] where no clear BHB motif could be detected. For the alignment, 15 nucleotides of flanking sequences were added on both sides of the box C/D snoRNA sequences.

Start	Stop	Strand	Name	transrealign
1152379	1152458	-	C/D box snoRNA 10	y
1117670	1117747	-	C/D box snoRNA 02	y

Supplementary Table 16. Details of the RNASeq analysis. Sequences were mapped to the reference genome with **segemehl** [8, 9] and remapped with **lack**, another program of the segemehl suite as well as **transrealign** which was used to extract the split reads. As *N. equitans* and *I. hospitalis* live in a parasymbiotic manner, the RNASeq data(*) (in total 16020851 reads) was mapped together to both reference genomes at the same time and splitted afterwards.

Species	pooled reads	mapped reads	remapped reads	split reads
<i>M. kandleri</i>	19242863	13736113	2153772	128603
<i>S. acidocaldarius</i>	26023157	22762924	360912	108283
<i>S. solfataricus</i>	8976013	7764310	878127	74143
<i>N. equitans</i>	*	10728929	-	25721
<i>I. hospitalis</i>	*	5567812	-	42757

Supplementary Table 17. New splicesites with BHB elements found in *M. kandleri* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

MSA1				MSA2			
Start	Stop	Strand	Ann	Start	Stop	Strand	Ann
143515	143632	+	CDS	31858	31995	+	-
163201	163467	+	-	40439	40539	+	-
325320	325593	+	CDS	262091	262244	+	CDS
351364	351601	+	CDS	325412	325593	+	-
399364	399622	+	CDS	384725	384960	+	-
490225	490354	+	CDS	497434	497516	+	-
518084	518344	+	-	507442	507586	+	CDS
518229	518322	+	rRNA	509798	509895	+	-
518263	518344	+	rRNA	510730	510856	+	rRNA
522017	522129	+	-	517768	517974	+	-
585393	585475	+	-	517864	518003	+	-
585629	585772	+	-	520831	520932	+	-
585946	586138	+	rRNA	576556	576720	+	-
586198	586414	+	rRNA	582234	582370	+	-
586751	586957	+	rRNA	585979	586110	+	-
586869	587025	+	rRNA	588030	588271	+	-
587307	587436	+	rRNA	693682	693927	+	-
587746	587871	+	-	739727	739822	+	-
673889	673968	+	-	744880	745039	+	CDS
1019844	1020094	+	-	779240	779332	+	-
1022388	1022636	+	-	879679	879914	+	-
1112401	1112677	+	-	1183460	1183557	+	-
1221433	1221523	+	CDS	1219327	1219497	+	-
1243818	1244084	+	-	1232353	1232448	+	-
1343553	1343808	+	-	1252557	1252674	+	-
				1259272	1259422	+	CDS
				1279749	1279961	+	-
				1415483	1415647	+	-
				1540620	1540755	+	CDS
				1688642	1688778	+	-

Supplementary Table 18. New splicesites with BHB elements found in *S. acidocaldarius* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

MSA1				MSA2			
Start	Stop	Strand	Ann	Start	Stop	Strand	Ann
49182	49271	+	-	49182	49271	+	-
72311	72447	+	-	73220	73336	+	-
708452	708600	+	-	97306	97393	+	ncRNA
789407	789634	+	-	267193	267304	+	CDS
848243	848477	+	CDS	450894	450975	+	-
1013752	1013976	+	-	677737	677855	+	-
1029123	1029366	+	CDS	624809	624949	+	-
1030222	1030393	+	CDS	639014	639153	+	-
1270855	1270973	+	-	753133	753241	+	CDS
1285009	1285134	+	-	808261	808346	+	-
1366353	1366466	+	-	913155	913261	+	-
1785213	1785353	+	-	986137	986328	+	-
				1126119	1126204	+	-
				1212710	1212788	+	-
				1327997	1328130	+	-
				1381538	1381622	+	CDS
				1648994	1649075	+	ncRNA
				2072492	2072665	+	-

Supplementary Table 19. New splicesites with BHB elements found in *N. equitans* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

MSA1				MSA2			
Start	Stop	Strand	Ann	Start	Stop	Strand	Ann
396266	396385	+	-	403449	403535	+	-
433125	433234	+	-				
396539	396812	+	CDS				

Supplementary Table 20. New splicesites with BHB elements found in *I. hospitalis* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

MSA1				MSA2			
Start	Stop	Strand	Ann	Start	Stop	Strand	Ann
603277	603363	+	-	3921	4027	+	-
733481	733720	+	-	281959	282165	+	-
798681	798876	+	CDS	602934	603198	+	CDS
				731896	732076	+	rRNA
				733248	733446	+	rRNA
				945608	945702	+	-
				1115636	1115739	+	-

Supplementary Table 21. New splicesites with BHB elements found in *S. solfataricus* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

MSA1				MSA2			
Start	Stop	Strand	Ann	Start	Stop	Strand	Ann
				763512	763600	+	-
				872878	873132	+	-
				872918	873156	+	-
				875970	876218	+	-
				876228	876426	+	rRNA
				2361384	2361469	+	-
				2939076	2939162	+	-

References

1. Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., Stadler, P.F.: *RNAalifold*: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9** (2008) 474
2. Danan, M., Schwartz, S., Edelheit, S., Sorek, R.: Transcriptome-wide discovery of circular RNAs in archaea. *Nucleic Acids Res.* **40** (2012) 3131–3142
3. Hertel, J., de Jong, D., Marz, M., Rose, D., Tafer, H., Tanzer, A., Schierwater, B., Stadler, P.F.: Non-coding RNA annotation of the genome of *Trichoplax adhaerens*. *Nucleic Acids Res.* **37** (2009) 1602–1615
4. Randau, L.: RNA processing in the minimal organism *Nanoarchaeum equitans*. *Genome Biology* **13** (2012)
5. Su, A.A.H., Tripp, V., Randau, L.: RNA-Seq analyses reveal the order of tRNA processing events and the maturation of C/D box and CRISPR RNAs in the hyperthermophile *Methanopyrus kandleri*. *Nucleic Acids Res.* **41** (2013) 6250–6258
6. Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., Dennis, P.P.: Homologs of small nucleolar rnas in archaea. *Science* **288**(5465) (2000) 517–522
7. Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B.A., Sorek, R.: A single-base resolution map of an archaeal transcriptome. *Genome Res.* **20** (2010) 133–141
8. Hoffmann, S., Otto, C., Kurtz, S., Sharma, C.M., Khaitovich, P., Vogel, J., Stadler, P.F., Hackermüller, J.: Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comp. Biol.* **5** (2009) e1000502
9. Hoffmann, S., Otto, C., Doose, G., Tanzer, A., Langenberger, D., Christ, S., Kunz, M., Holdt, L.M., Teupser, D., Hackermüller, J., Stadler, P.F.: A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. *Genome Biology* **15** (2014) R34