Comparative ncRNA Detection in Archaea

Sarah J. Berkemer\textsuperscript{1,2,4,*}, Christian Höner zu Siederdissen\textsuperscript{1,2,4}, Fabian Amman\textsuperscript{2}, Axel Witsch\textsuperscript{3,4}, Sebastian Will\textsuperscript{1,4}, Ivo L. Hofacker\textsuperscript{2,5,6}, Sonja J. Prohaska\textsuperscript{3,4}, Peter F. Stadler\textsuperscript{1,2,4,6-9,*}

\textsuperscript{1} Bioinformatics Group, Department of Computer Science, Univ. Leipzig, Germany
\textsuperscript{2} Institute for Theoretical Chemistry, Univ. Vienna, Austria
\textsuperscript{3} Computational EvoDevo Group, Department of Computer Science, Univ. Leipzig, Germany
\textsuperscript{4} Interdisciplinary Center for Bioinformatics, University of Leipzig, Germany
\textsuperscript{5} Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, Univ. Vienna, Austria
\textsuperscript{6} Center for non-coding RNA in Technology and Health, Univ. Copenhagen, Denmark
\textsuperscript{7} Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
\textsuperscript{8} Fraunhofer Institut für Zelltherapie und Immunologie, Leipzig, Germany
\textsuperscript{9} Santa Fe Institute, Santa Fe, USA
E-mail: studla@bioinf.uni-leipzig.de
Supplementary Figure 1. 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.

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

<table>
<thead>
<tr>
<th>RNA-seq L. junc.</th>
<th>RNA-seq R. junc.</th>
<th>#Count</th>
<th>cm Search Rank Bit Score</th>
<th>Genomic Surrounding ncbi locus tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>69,921</td>
<td>69,985</td>
<td>2,065</td>
<td>–</td>
<td>7,845.</td>
</tr>
<tr>
<td>91,822</td>
<td>91,904</td>
<td>2,992</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>205,318</td>
<td>205,387</td>
<td>12,182</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>219,317</td>
<td>219,379</td>
<td>2,412</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>271,148</td>
<td>271,216</td>
<td>10,072</td>
<td>3,546.</td>
<td>25nt 77nt⇒MK0075</td>
</tr>
<tr>
<td>361,063</td>
<td>361,125</td>
<td>1,809</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>384,798</td>
<td>384,945</td>
<td>634</td>
<td>–</td>
<td>205nt 77nt⇒MK0075</td>
</tr>
<tr>
<td>459,461</td>
<td>459,532</td>
<td>9,830</td>
<td>9,039.</td>
<td>7,7</td>
</tr>
<tr>
<td>519,288</td>
<td>519,358</td>
<td>22,145</td>
<td>15,966.</td>
<td>8.1</td>
</tr>
<tr>
<td>520,778</td>
<td>520,845</td>
<td>13,484</td>
<td>–</td>
<td>25nt 77nt⇒MK0075</td>
</tr>
<tr>
<td>755,163</td>
<td>755,226</td>
<td>304</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>790,521</td>
<td>790,585</td>
<td>1,161</td>
<td>19,971.</td>
<td>MK0074 ←25nt⇒MK0075</td>
</tr>
<tr>
<td>993,172</td>
<td>993,238</td>
<td>50,802</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1,104,263</td>
<td>1,104,330</td>
<td>6,223</td>
<td>17,300.</td>
<td>MK0128 ←573nt⇒MK1129</td>
</tr>
<tr>
<td>1,238,104</td>
<td>1,238,178</td>
<td>9</td>
<td>4,630.</td>
<td>9.2</td>
</tr>
<tr>
<td>1,417,298</td>
<td>1,417,370</td>
<td>9,371</td>
<td>–</td>
<td>MK1390 ←36nt⇒MK1391</td>
</tr>
<tr>
<td>1,417,378</td>
<td>1,417,443</td>
<td>5,658</td>
<td>14,942.</td>
<td>MK1390 ←116nt⇒MK1391</td>
</tr>
<tr>
<td>1,444,847</td>
<td>1,444,927</td>
<td>2</td>
<td>–</td>
<td>MK1479 ←88nt⇒MK1480</td>
</tr>
<tr>
<td>1,500,955</td>
<td>1,501,112</td>
<td>2,648</td>
<td>–</td>
<td>MK1479 ←88nt⇒MK1480</td>
</tr>
<tr>
<td>1,506,611</td>
<td>1,506,673</td>
<td>8,396</td>
<td>6,984.</td>
<td>8.4</td>
</tr>
<tr>
<td>1,506,611</td>
<td>1,506,673</td>
<td>8,396</td>
<td>6,984.</td>
<td>8.4</td>
</tr>
</tbody>
</table>
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 was conducted with the GotohScan program [3]. The “Start” and “End” columns refer to position in the genomes NC_002754 and NC_007181, respectively.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sulfolobus solfataricus</th>
<th>Sulfolobus acidocaldarius</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RNA-seq cm Search</td>
<td>RNA-seq cm Search</td>
</tr>
<tr>
<td></td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>5S rRNA/SSOr02</td>
<td>77,945 78,067 863.</td>
<td>1,293,914 1,294,035 –</td>
</tr>
<tr>
<td>16S rRNA/SSOr03</td>
<td>871,658 873,216 –</td>
<td>1,108,641 1,107,094 –</td>
</tr>
<tr>
<td>23S rRNA/SSOr04</td>
<td>873,334 876,429 –</td>
<td>1,106,947 1,103,875 –</td>
</tr>
<tr>
<td>tRNA-Val/SSOt04</td>
<td>72,767 72,831 1,321. 5.9</td>
<td>49,197 49,262 548. 6.4</td>
</tr>
<tr>
<td>tRNA-Lys/SSOt07</td>
<td>138,386 138,407 1,420. 5.7</td>
<td>607,138 607,204 11. 13.0</td>
</tr>
<tr>
<td>tRNA-Met/SSOt11</td>
<td>184,817 184,841 13. 13.3</td>
<td>– – –</td>
</tr>
<tr>
<td>tRNA-Pro/SSOt42</td>
<td>898,313 898,333 242. 8.9</td>
<td>1,096,702 1,096,684 145. 8.3</td>
</tr>
<tr>
<td>tRNA-Ser/SSOt33</td>
<td>640,978 641,001 1,094. 6.2</td>
<td>512,691 512,669 62. 9.6</td>
</tr>
<tr>
<td>C/D box sR106</td>
<td>285,707 285,760 –</td>
<td>2,179,509 2,179,560 –</td>
</tr>
<tr>
<td>C/D box Sso-180</td>
<td>302,308 302,369 –</td>
<td>669,556 669,612 –</td>
</tr>
<tr>
<td>C/D box sR133</td>
<td>442,392 442,417 –</td>
<td>– – –</td>
</tr>
<tr>
<td>C/D box sR102</td>
<td>563,241 563,296 –</td>
<td>1,388,934 1,388,984 –</td>
</tr>
<tr>
<td>C/D box Sso-sR8</td>
<td>647,883 647,903 300. 5.6</td>
<td>1,185,917 1,185,967 –</td>
</tr>
<tr>
<td>C/D box Sso-sR4</td>
<td>666,143 666,186 1,779. 5.3</td>
<td>– – –</td>
</tr>
<tr>
<td>C/D box Sso-sR10</td>
<td>794,186 794,210 –</td>
<td>1,152,443 1,152,394 –</td>
</tr>
<tr>
<td>C/D box Sso-207</td>
<td>816,021 816,075 –</td>
<td>– – –</td>
</tr>
<tr>
<td>C/D box SSo02</td>
<td>829,352 829,405 –</td>
<td>1,117,732 1,117,685 –</td>
</tr>
<tr>
<td>C/D box Sso-sR12</td>
<td>2,189,397 2,189,456 –</td>
<td>– – –</td>
</tr>
<tr>
<td>C/D box sR105</td>
<td>2,237,915 2,237,962 –</td>
<td>217,040 217,087 –</td>
</tr>
<tr>
<td>H/ACA box sR109</td>
<td>59,5510 595,579 308. 8.3</td>
<td>458,983 459,052 351. 7.0</td>
</tr>
<tr>
<td>ncRNA</td>
<td>442,786 442,854 –</td>
<td>– – –</td>
</tr>
<tr>
<td>ncRNA</td>
<td>722,538 722,578 3005. 4.4</td>
<td>417,691 417729 –</td>
</tr>
<tr>
<td>Sso-117</td>
<td>1,576,633 1,576,671 –</td>
<td>– – –</td>
</tr>
<tr>
<td>Sso-109</td>
<td>1,927,228 1,927,258 –</td>
<td>– – –</td>
</tr>
<tr>
<td>7S rRNA/SSOr01</td>
<td>49,977 50023 839. 6.6</td>
<td>72,370 72,326 763. 5.9</td>
</tr>
<tr>
<td>Sso-214</td>
<td>105,148 105,181 –</td>
<td>– – –</td>
</tr>
<tr>
<td>RNase P</td>
<td>224,732 224,765 –</td>
<td>586,242 586,211 –</td>
</tr>
<tr>
<td>Sso-83</td>
<td>581,818 581,860 –</td>
<td>– – –</td>
</tr>
<tr>
<td>ncRNA</td>
<td>1,275,500 127,5567 –</td>
<td>– – –</td>
</tr>
<tr>
<td>SSO0393</td>
<td>343,138 343,264 –</td>
<td>650,082 650,206 –</td>
</tr>
<tr>
<td>Intergenic region</td>
<td>871,573 871657 1,107. 6.2</td>
<td>1,108,730 1,108,647 –</td>
</tr>
<tr>
<td>Intergenic region</td>
<td>873,215 873331 –</td>
<td>1,107,091 1,106,972 –</td>
</tr>
<tr>
<td>SSO0389</td>
<td>335,563 335,635 –</td>
<td>– – –</td>
</tr>
<tr>
<td>SSO0845</td>
<td>725,923 726,085 –</td>
<td>– – –</td>
</tr>
<tr>
<td>SSO2359</td>
<td>2,154,297 2,154,322 –</td>
<td>782,915 782,892 –</td>
</tr>
<tr>
<td>SSO2619</td>
<td>2,385,872 2,385,901 –</td>
<td>738,654 738,681 –</td>
</tr>
<tr>
<td>SSO2642</td>
<td>2,404,146 2,404,146 –</td>
<td>2,114,638 2,114,694 –</td>
</tr>
</tbody>
</table>
Supplementary Table 4. circRNA candidates of *M. kandleri* and *S. acidolarius* 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.

<table>
<thead>
<tr>
<th>circRNA candidate locus</th>
<th>Number of Seqs</th>
<th>RNAz class probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC_003551.1 1500955 1501112</td>
<td>10</td>
<td>0.9992</td>
</tr>
<tr>
<td>NC_007181.1 1214595 1214683</td>
<td>4</td>
<td>0.680306</td>
</tr>
<tr>
<td>NC_007181.1 1254692 1254799</td>
<td>4</td>
<td>0.938321</td>
</tr>
<tr>
<td>NC_007181.1 1107137 1107281</td>
<td>39</td>
<td>0.998730</td>
</tr>
<tr>
<td>NC_007181.1 1803656 1803770</td>
<td>3</td>
<td>0.722219</td>
</tr>
<tr>
<td>NC_007181.1 183648 183733</td>
<td>5</td>
<td>0.703805</td>
</tr>
<tr>
<td>NC_007181.1 1995955 1996059</td>
<td>3</td>
<td>0.619096</td>
</tr>
<tr>
<td>NC_007181.1 553923 554080</td>
<td>4</td>
<td>0.799874</td>
</tr>
<tr>
<td>NC_007181.1 753148 753230</td>
<td>4</td>
<td>0.982281</td>
</tr>
<tr>
<td>NC_007181.1 766362 766509</td>
<td>4</td>
<td>0.646306</td>
</tr>
<tr>
<td>NC_007181.1 773268 773364</td>
<td>4</td>
<td>0.644060</td>
</tr>
<tr>
<td>NC_007181.1 86425 86509</td>
<td>3</td>
<td>0.554138</td>
</tr>
</tbody>
</table>
**Supplementary Table 5.** Box C/D snoRNA of *M. kandleri* from [4] 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 [5]. 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.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>1815</td>
<td>1879</td>
<td>+</td>
<td>C/D box snoRNA 1</td>
<td>y</td>
</tr>
<tr>
<td>42955</td>
<td>43022</td>
<td>-</td>
<td>C/D box snoRNA 2</td>
<td>y</td>
</tr>
<tr>
<td>44728</td>
<td>44789</td>
<td>+</td>
<td>C/D box snoRNA 3</td>
<td>y</td>
</tr>
<tr>
<td>69897</td>
<td>69962</td>
<td>+</td>
<td>C/D box snoRNA 4</td>
<td>n</td>
</tr>
<tr>
<td>111152</td>
<td>111214</td>
<td>+</td>
<td>C/D box snoRNA 6</td>
<td>n</td>
</tr>
<tr>
<td>137488</td>
<td>137553</td>
<td>+</td>
<td>C/D box snoRNA 7</td>
<td>y</td>
</tr>
<tr>
<td>139707</td>
<td>139764</td>
<td>+</td>
<td>C/D box snoRNA 8</td>
<td>n</td>
</tr>
<tr>
<td>146560</td>
<td>146637</td>
<td>+</td>
<td>C/D box snoRNA 9</td>
<td>y</td>
</tr>
<tr>
<td>147613</td>
<td>147672</td>
<td>+</td>
<td>C/D box snoRNA 10</td>
<td>n</td>
</tr>
<tr>
<td>170672</td>
<td>170746</td>
<td>+</td>
<td>C/D box snoRNA 11</td>
<td>n</td>
</tr>
<tr>
<td>170856</td>
<td>170923</td>
<td>+</td>
<td>C/D box snoRNA 12</td>
<td>y</td>
</tr>
<tr>
<td>172650</td>
<td>172747</td>
<td>-</td>
<td>C/D box snoRNA 13</td>
<td>n</td>
</tr>
<tr>
<td>181467</td>
<td>181526</td>
<td>+</td>
<td>C/D box snoRNA 14</td>
<td>y</td>
</tr>
<tr>
<td>203475</td>
<td>203540</td>
<td>+</td>
<td>C/D box snoRNA 15</td>
<td>y</td>
</tr>
<tr>
<td>205387</td>
<td>205458</td>
<td>-</td>
<td>C/D box snoRNA 17</td>
<td>y</td>
</tr>
<tr>
<td>227948</td>
<td>228036</td>
<td>+</td>
<td>C/D box snoRNA 19</td>
<td>y</td>
</tr>
<tr>
<td>262147</td>
<td>262212</td>
<td>+</td>
<td>C/D box snoRNA 20</td>
<td>y</td>
</tr>
<tr>
<td>271071</td>
<td>271136</td>
<td>-</td>
<td>C/D box snoRNA 21</td>
<td>y</td>
</tr>
<tr>
<td>271151</td>
<td>271217</td>
<td>-</td>
<td>C/D box snoRNA 22</td>
<td>y</td>
</tr>
<tr>
<td>325495</td>
<td>325560</td>
<td>-</td>
<td>C/D box snoRNA 24</td>
<td>n</td>
</tr>
<tr>
<td>325521</td>
<td>325579</td>
<td>-</td>
<td>C/D box snoRNA 25</td>
<td>y</td>
</tr>
<tr>
<td>327758</td>
<td>327832</td>
<td>+</td>
<td>C/D box snoRNA 26</td>
<td>y</td>
</tr>
<tr>
<td>346543</td>
<td>346604</td>
<td>-</td>
<td>C/D box snoRNA 28</td>
<td>y</td>
</tr>
<tr>
<td>360535</td>
<td>360607</td>
<td>-</td>
<td>C/D box snoRNA 29</td>
<td>y</td>
</tr>
<tr>
<td>383186</td>
<td>383259</td>
<td>+</td>
<td>C/D box snoRNA 31</td>
<td>n</td>
</tr>
<tr>
<td>384878</td>
<td>384947</td>
<td>+</td>
<td>C/D box snoRNA 32</td>
<td>y</td>
</tr>
<tr>
<td>409873</td>
<td>409936</td>
<td>+</td>
<td>C/D box snoRNA 33</td>
<td>y</td>
</tr>
<tr>
<td>459462</td>
<td>459533</td>
<td>-</td>
<td>C/D box snoRNA 34</td>
<td>n</td>
</tr>
<tr>
<td>462530</td>
<td>462593</td>
<td>+</td>
<td>C/D box snoRNA 35</td>
<td>y</td>
</tr>
<tr>
<td>515532</td>
<td>515599</td>
<td>-</td>
<td>C/D box snoRNA 37</td>
<td>y</td>
</tr>
<tr>
<td>518923</td>
<td>518988</td>
<td>-</td>
<td>C/D box snoRNA 38</td>
<td>y</td>
</tr>
<tr>
<td>519204</td>
<td>519281</td>
<td>-</td>
<td>C/D box snoRNA 39</td>
<td>y</td>
</tr>
<tr>
<td>519290</td>
<td>519359</td>
<td>-</td>
<td>C/D box snoRNA 40</td>
<td>y</td>
</tr>
<tr>
<td>520779</td>
<td>520847</td>
<td>+</td>
<td>C/D box snoRNA 41</td>
<td>y</td>
</tr>
<tr>
<td>524051</td>
<td>524115</td>
<td>+</td>
<td>C/D box snoRNA 43</td>
<td>y</td>
</tr>
<tr>
<td>524115</td>
<td>524177</td>
<td>+</td>
<td>C/D box snoRNA 44</td>
<td>y</td>
</tr>
<tr>
<td>524175</td>
<td>524238</td>
<td>+</td>
<td>C/D box snoRNA 45</td>
<td>y</td>
</tr>
<tr>
<td>537777</td>
<td>537843</td>
<td>+</td>
<td>C/D box snoRNA 46</td>
<td>y</td>
</tr>
<tr>
<td>561189</td>
<td>561251</td>
<td>+</td>
<td>C/D box snoRNA 47</td>
<td>y</td>
</tr>
<tr>
<td>582279</td>
<td>582357</td>
<td>+</td>
<td>C/D box snoRNA 48</td>
<td>y</td>
</tr>
<tr>
<td>603202</td>
<td>603273</td>
<td>+</td>
<td>C/D box snoRNA 49</td>
<td>y</td>
</tr>
<tr>
<td>605153</td>
<td>605234</td>
<td>+</td>
<td>C/D box snoRNA 50</td>
<td>y</td>
</tr>
<tr>
<td>627727</td>
<td>627790</td>
<td>-</td>
<td>C/D box snoRNA 52</td>
<td>y</td>
</tr>
<tr>
<td>700325</td>
<td>700386</td>
<td>-</td>
<td>C/D box snoRNA 53</td>
<td>n</td>
</tr>
<tr>
<td>729573</td>
<td>729639</td>
<td>+</td>
<td>C/D box snoRNA 54</td>
<td>y</td>
</tr>
<tr>
<td>751600</td>
<td>751664</td>
<td>+</td>
<td>C/D box snoRNA 55</td>
<td>y</td>
</tr>
<tr>
<td>755168</td>
<td>755227</td>
<td>-</td>
<td>C/D box snoRNA 56</td>
<td>y</td>
</tr>
<tr>
<td>766722</td>
<td>766794</td>
<td>-</td>
<td>C/D box snoRNA 57</td>
<td>y</td>
</tr>
<tr>
<td>772696</td>
<td>772766</td>
<td>+</td>
<td>C/D box snoRNA 58</td>
<td>y</td>
</tr>
<tr>
<td>776072</td>
<td>776149</td>
<td>+</td>
<td>C/D box snoRNA 59</td>
<td>y</td>
</tr>
<tr>
<td>779256</td>
<td>779317</td>
<td>+</td>
<td>C/D box snoRNA 60</td>
<td>n</td>
</tr>
<tr>
<td>779315</td>
<td>779376</td>
<td>-</td>
<td>C/D box snoRNA 61</td>
<td>n</td>
</tr>
<tr>
<td>790520</td>
<td>790586</td>
<td>-</td>
<td>C/D box snoRNA 62</td>
<td>y</td>
</tr>
<tr>
<td>825703</td>
<td>825775</td>
<td>+</td>
<td>C/D box snoRNA 63</td>
<td>y</td>
</tr>
<tr>
<td>830254</td>
<td>830325</td>
<td>+</td>
<td>C/D box snoRNA 64</td>
<td>y</td>
</tr>
<tr>
<td>839773</td>
<td>839862</td>
<td>-</td>
<td>C/D box snoRNA 65</td>
<td>y</td>
</tr>
</tbody>
</table>
Supplementary Table 6. Box C/D snoRNA of *M. kandleri* from [4] 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 [5]. 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.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>843279</td>
<td>843335</td>
<td>+</td>
<td>C/D box snoRNA 66</td>
<td>n</td>
</tr>
<tr>
<td>858313</td>
<td>858382</td>
<td>+</td>
<td>C/D box snoRNA 67</td>
<td>y</td>
</tr>
<tr>
<td>879751</td>
<td>879813</td>
<td>+</td>
<td>C/D box snoRNA 68</td>
<td>y</td>
</tr>
<tr>
<td>879938</td>
<td>879993</td>
<td>-</td>
<td>C/D box snoRNA 69</td>
<td>y</td>
</tr>
<tr>
<td>892485</td>
<td>892553</td>
<td>+</td>
<td>C/D box snoRNA 70</td>
<td>y</td>
</tr>
<tr>
<td>900585</td>
<td>900653</td>
<td>+</td>
<td>C/D box snoRNA 71</td>
<td>y</td>
</tr>
<tr>
<td>915062</td>
<td>915133</td>
<td>-</td>
<td>C/D box snoRNA 73</td>
<td>y</td>
</tr>
<tr>
<td>961063</td>
<td>961126</td>
<td>+</td>
<td>C/D box snoRNA 74</td>
<td>y</td>
</tr>
<tr>
<td>963110</td>
<td>963171</td>
<td>-</td>
<td>C/D box snoRNA 75</td>
<td>y</td>
</tr>
<tr>
<td>993170</td>
<td>993242</td>
<td>+</td>
<td>C/D box snoRNA 76</td>
<td>y</td>
</tr>
<tr>
<td>1003081</td>
<td>1003142</td>
<td>+</td>
<td>C/D box snoRNA 77</td>
<td>n</td>
</tr>
<tr>
<td>1022562</td>
<td>1022623</td>
<td>-</td>
<td>C/D box snoRNA 78</td>
<td>y</td>
</tr>
<tr>
<td>1022628</td>
<td>1022698</td>
<td>-</td>
<td>C/D box snoRNA 79</td>
<td>y</td>
</tr>
<tr>
<td>1023968</td>
<td>1024031</td>
<td>+</td>
<td>C/D box snoRNA 80</td>
<td>y</td>
</tr>
<tr>
<td>1043750</td>
<td>1043777</td>
<td>-</td>
<td>C/D box snoRNA 81</td>
<td>n</td>
</tr>
<tr>
<td>1048100</td>
<td>1048194</td>
<td>+</td>
<td>C/D box snoRNA 82</td>
<td>y</td>
</tr>
<tr>
<td>1065992</td>
<td>1066050</td>
<td>-</td>
<td>C/D box snoRNA 84</td>
<td>n</td>
</tr>
<tr>
<td>1067504</td>
<td>1067564</td>
<td>-</td>
<td>C/D box snoRNA 85</td>
<td>n</td>
</tr>
<tr>
<td>1073361</td>
<td>1073433</td>
<td>-</td>
<td>C/D box snoRNA 86</td>
<td>y</td>
</tr>
<tr>
<td>1077825</td>
<td>1077895</td>
<td>+</td>
<td>C/D box snoRNA 87</td>
<td>y</td>
</tr>
<tr>
<td>1094668</td>
<td>1094726</td>
<td>-</td>
<td>C/D box snoRNA 88</td>
<td>n</td>
</tr>
<tr>
<td>1104271</td>
<td>1104344</td>
<td>-</td>
<td>C/D box snoRNA 89</td>
<td>y</td>
</tr>
<tr>
<td>1118640</td>
<td>1118700</td>
<td>+</td>
<td>C/D box snoRNA 90</td>
<td>n</td>
</tr>
<tr>
<td>1164281</td>
<td>1164371</td>
<td>+</td>
<td>C/D box snoRNA 92</td>
<td>n</td>
</tr>
<tr>
<td>1165264</td>
<td>1165330</td>
<td>-</td>
<td>C/D box snoRNA 93</td>
<td>n</td>
</tr>
<tr>
<td>1166529</td>
<td>1166612</td>
<td>+</td>
<td>C/D box snoRNA 94</td>
<td>y</td>
</tr>
<tr>
<td>1178363</td>
<td>1178424</td>
<td>-</td>
<td>C/D box snoRNA 95</td>
<td>n</td>
</tr>
<tr>
<td>1182419</td>
<td>1182494</td>
<td>+</td>
<td>C/D box snoRNA 96</td>
<td>y</td>
</tr>
<tr>
<td>1183490</td>
<td>1183549</td>
<td>-</td>
<td>C/D box snoRNA 97</td>
<td>n</td>
</tr>
<tr>
<td>1187688</td>
<td>1187763</td>
<td>+</td>
<td>C/D box snoRNA 98</td>
<td>n</td>
</tr>
<tr>
<td>1204769</td>
<td>1204834</td>
<td>+</td>
<td>C/D box snoRNA 99</td>
<td>n</td>
</tr>
<tr>
<td>1221349</td>
<td>1221417</td>
<td>-</td>
<td>C/D box snoRNA 100</td>
<td>y</td>
</tr>
<tr>
<td>1231253</td>
<td>1231340</td>
<td>+</td>
<td>C/D box snoRNA 101</td>
<td>n</td>
</tr>
<tr>
<td>1231699</td>
<td>1231762</td>
<td>+</td>
<td>C/D box snoRNA 102</td>
<td>y</td>
</tr>
<tr>
<td>1232371</td>
<td>1232433</td>
<td>+</td>
<td>C/D box snoRNA 103</td>
<td>n</td>
</tr>
<tr>
<td>1233144</td>
<td>1233210</td>
<td>+</td>
<td>C/D box snoRNA 104</td>
<td>y</td>
</tr>
<tr>
<td>1233411</td>
<td>1233470</td>
<td>-</td>
<td>C/D box snoRNA 105</td>
<td>y</td>
</tr>
<tr>
<td>1263413</td>
<td>1263479</td>
<td>+</td>
<td>C/D box snoRNA 107</td>
<td>y</td>
</tr>
<tr>
<td>1383364</td>
<td>1383421</td>
<td>-</td>
<td>C/D box snoRNA 108</td>
<td>y</td>
</tr>
<tr>
<td>1386498</td>
<td>1386561</td>
<td>+</td>
<td>C/D box snoRNA 109</td>
<td>n</td>
</tr>
<tr>
<td>1415602</td>
<td>1415673</td>
<td>-</td>
<td>C/D box snoRNA 110</td>
<td>n</td>
</tr>
<tr>
<td>1417300</td>
<td>1417370</td>
<td>-</td>
<td>C/D box snoRNA 111</td>
<td>y</td>
</tr>
<tr>
<td>1417379</td>
<td>1417442</td>
<td>-</td>
<td>C/D box snoRNA 112</td>
<td>y</td>
</tr>
<tr>
<td>1417448</td>
<td>1417512</td>
<td>-</td>
<td>C/D box snoRNA 113</td>
<td>y</td>
</tr>
<tr>
<td>1492448</td>
<td>1492515</td>
<td>+</td>
<td>C/D box snoRNA 114</td>
<td>y</td>
</tr>
<tr>
<td>1493125</td>
<td>1493192</td>
<td>+</td>
<td>C/D box snoRNA 115</td>
<td>y</td>
</tr>
<tr>
<td>1495330</td>
<td>1495399</td>
<td>+</td>
<td>C/D box snoRNA 116</td>
<td>y</td>
</tr>
<tr>
<td>1506611</td>
<td>1506674</td>
<td>+</td>
<td>C/D box snoRNA 117</td>
<td>y</td>
</tr>
<tr>
<td>1508733</td>
<td>1508798</td>
<td>-</td>
<td>C/D box snoRNA 118</td>
<td>y</td>
</tr>
<tr>
<td>1510490</td>
<td>1510576</td>
<td>-</td>
<td>C/D box snoRNA 119</td>
<td>n</td>
</tr>
<tr>
<td>1531085</td>
<td>1531148</td>
<td>+</td>
<td>C/D box snoRNA 120</td>
<td>y</td>
</tr>
<tr>
<td>1561232</td>
<td>1561294</td>
<td>+</td>
<td>C/D box snoRNA 121</td>
<td>n</td>
</tr>
<tr>
<td>1587746</td>
<td>1587811</td>
<td>-</td>
<td>C/D box snoRNA 122</td>
<td>y</td>
</tr>
<tr>
<td>1605995</td>
<td>1606060</td>
<td>+</td>
<td>C/D box snoRNA 123</td>
<td>n</td>
</tr>
<tr>
<td>1607219</td>
<td>1607287</td>
<td>-</td>
<td>C/D box snoRNA 124</td>
<td>y</td>
</tr>
<tr>
<td>1637333</td>
<td>1637398</td>
<td>+</td>
<td>C/D box snoRNA 126</td>
<td>y</td>
</tr>
</tbody>
</table>
Supplementary Table 7. Box C/D snoRNA of *M. kandleri* from [4] 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 [5]. 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.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>91821</td>
<td>91905</td>
<td>+</td>
<td>C/D box snoRNA5</td>
<td>y</td>
</tr>
<tr>
<td>205310</td>
<td>205380</td>
<td>-</td>
<td>C/D box snoRNA16</td>
<td>n</td>
</tr>
<tr>
<td>219316</td>
<td>219380</td>
<td>+</td>
<td>C/D box snoRNA18</td>
<td>n</td>
</tr>
<tr>
<td>274089</td>
<td>274184</td>
<td>+</td>
<td>C/D box snoRNA23</td>
<td>y</td>
</tr>
<tr>
<td>333110</td>
<td>333179</td>
<td>+</td>
<td>C/D box snoRNA27</td>
<td>y</td>
</tr>
<tr>
<td>361062</td>
<td>361156</td>
<td>-</td>
<td>C/D box snoRNA30</td>
<td>n</td>
</tr>
<tr>
<td>509783</td>
<td>509880</td>
<td>+</td>
<td>C/D box snoRNA36</td>
<td>y</td>
</tr>
<tr>
<td>520852</td>
<td>520933</td>
<td>+</td>
<td>C/D box snoRNA42</td>
<td>y</td>
</tr>
<tr>
<td>605300</td>
<td>605373</td>
<td>+</td>
<td>C/D box snoRNA51</td>
<td>y</td>
</tr>
<tr>
<td>902800</td>
<td>902861</td>
<td>-</td>
<td>C/D box snoRNA72</td>
<td>n</td>
</tr>
<tr>
<td>105548</td>
<td>1055621</td>
<td>+</td>
<td>C/D box snoRNA83</td>
<td>n</td>
</tr>
<tr>
<td>1157340</td>
<td>1157415</td>
<td>+</td>
<td>C/D box snoRNA91</td>
<td>y</td>
</tr>
<tr>
<td>1234309</td>
<td>1234398</td>
<td>-</td>
<td>C/D box snoRNA106</td>
<td>n</td>
</tr>
<tr>
<td>1607250</td>
<td>1607305</td>
<td>+</td>
<td>C/D box snoRNA125</td>
<td>y</td>
</tr>
</tbody>
</table>

Supplementary Table 8. Box C/D snoRNA sequences of *N. equitans* from [5] used to build the covariance model for detection 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 with (*) couldn’t be predicted in [5] such that now name was given. We kept the order of sequences and so the name consists of a number in between.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>46922</td>
<td>46979</td>
<td>-</td>
<td>C/D box snoRNA1b*</td>
<td>y</td>
</tr>
<tr>
<td>54028</td>
<td>54085</td>
<td>-</td>
<td>C/D box snoRNA2</td>
<td>n</td>
</tr>
<tr>
<td>90898</td>
<td>90954</td>
<td>-</td>
<td>C/D box snoRNA3</td>
<td>y</td>
</tr>
<tr>
<td>113577</td>
<td>113628</td>
<td>-</td>
<td>C/D box snoRNA4</td>
<td>n</td>
</tr>
<tr>
<td>144953</td>
<td>145004</td>
<td>-</td>
<td>C/D box snoRNA5</td>
<td>y</td>
</tr>
<tr>
<td>145962</td>
<td>146019</td>
<td>-</td>
<td>C/D box snoRNA6</td>
<td>y</td>
</tr>
<tr>
<td>164113</td>
<td>164168</td>
<td>-</td>
<td>C/D box snoRNA7</td>
<td>y</td>
</tr>
<tr>
<td>248187</td>
<td>248239</td>
<td>-</td>
<td>C/D box snoRNA7b*</td>
<td>n</td>
</tr>
<tr>
<td>282915</td>
<td>282973</td>
<td>-</td>
<td>C/D box snoRNA8</td>
<td>y</td>
</tr>
<tr>
<td>315990</td>
<td>316045</td>
<td>-</td>
<td>C/D box snoRNA10</td>
<td>y</td>
</tr>
<tr>
<td>323041</td>
<td>323094</td>
<td>-</td>
<td>C/D box snoRNA11b*</td>
<td>n</td>
</tr>
<tr>
<td>328897</td>
<td>328949</td>
<td>-</td>
<td>C/D box snoRNA13</td>
<td>y</td>
</tr>
<tr>
<td>337448</td>
<td>337505</td>
<td>-</td>
<td>C/D box snoRNA15</td>
<td>y</td>
</tr>
<tr>
<td>371860</td>
<td>371917</td>
<td>-</td>
<td>C/D box snoRNA16</td>
<td>y</td>
</tr>
<tr>
<td>375354</td>
<td>375406</td>
<td>-</td>
<td>C/D box snoRNA17</td>
<td>y</td>
</tr>
</tbody>
</table>
Supplementary Table 9. Box C/D snoRNA sequences of *N. equitans* from [5] 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 [5] such that now name was given. We kept the order of sequences and so the name consists of a number in between.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>35779</td>
<td>45831</td>
<td>+</td>
<td>C/D box snoRNA1</td>
<td>n</td>
</tr>
<tr>
<td>297634</td>
<td>297687</td>
<td>+</td>
<td>C/D box snoRNA9</td>
<td>n</td>
</tr>
<tr>
<td>318915</td>
<td>318968</td>
<td>+</td>
<td>C/D box snoRNA11</td>
<td>y</td>
</tr>
<tr>
<td>334552</td>
<td>334611</td>
<td>+</td>
<td>C/D box snoRNA14</td>
<td>y</td>
</tr>
<tr>
<td>359412</td>
<td>359470</td>
<td>+</td>
<td>C/D box snoRNA15a*</td>
<td>y</td>
</tr>
<tr>
<td>362755</td>
<td>362808</td>
<td>+</td>
<td>C/D box snoRNA15b*</td>
<td>y</td>
</tr>
<tr>
<td>382458</td>
<td>382509</td>
<td>+</td>
<td>C/D box snoRNA18</td>
<td>y</td>
</tr>
<tr>
<td>401223</td>
<td>401280</td>
<td>+</td>
<td>C/D box snoRNA18b*</td>
<td>y</td>
</tr>
<tr>
<td>403464</td>
<td>403522</td>
<td>+</td>
<td>C/D box snoRNA19</td>
<td>n</td>
</tr>
</tbody>
</table>

Supplementary Table 10. Box C/D snoRNA sequences of *N. equitans* from [5] 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 [5] such that now name was given. We kept the order of sequences and so the name consists of a number in between.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>325371</td>
<td>325424</td>
<td>+</td>
<td>C/D box snoRNA12</td>
<td>n</td>
</tr>
<tr>
<td>384218</td>
<td>384278</td>
<td>+</td>
<td>C/D box snoRNA18a*</td>
<td>y</td>
</tr>
</tbody>
</table>

Supplementary Table 11. 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 [5] such that now name was given. We kept the order of sequences and so the name consists of a number in between.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>362293</td>
<td>362384</td>
<td>+</td>
<td>C/D box snoRNA 180</td>
<td>y</td>
</tr>
<tr>
<td>794171</td>
<td>794255</td>
<td>+</td>
<td>C/D box snoRNA 10</td>
<td>y</td>
</tr>
<tr>
<td>829337</td>
<td>829420</td>
<td>+</td>
<td>C/D box snoRNA 02</td>
<td>y</td>
</tr>
<tr>
<td>218932</td>
<td>218947</td>
<td>+</td>
<td>C/D box snoRNA 12</td>
<td>n</td>
</tr>
<tr>
<td>127459</td>
<td>127487</td>
<td>-</td>
<td>C/D box snoRNA n2*</td>
<td>n</td>
</tr>
<tr>
<td>200352</td>
<td>200368</td>
<td>+</td>
<td>C/D box snoRNA 67</td>
<td>n</td>
</tr>
<tr>
<td>211132</td>
<td>211146</td>
<td>-</td>
<td>C/D box snoRNA 67b*</td>
<td>n</td>
</tr>
</tbody>
</table>
**Supplementary Table 12.** 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 [5] such that now name was given. We kept the order of sequences and so the name consists of a number in between.

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

**Supplementary Table 13.** 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.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>2179492</td>
<td>2179575</td>
<td>+</td>
<td>C/D box snoRNA 18</td>
<td>n</td>
</tr>
<tr>
<td>793941</td>
<td>794027</td>
<td>-</td>
<td>C/D box snoRNA 17</td>
<td>y</td>
</tr>
<tr>
<td>2196878</td>
<td>2196963</td>
<td>+</td>
<td>C/D box snoRNA 16</td>
<td>n</td>
</tr>
<tr>
<td>1400376</td>
<td>1400445</td>
<td>+</td>
<td>C/D box snoRNA 15</td>
<td>y</td>
</tr>
<tr>
<td>1117434</td>
<td>1117521</td>
<td>-</td>
<td>C/D box snoRNA 14</td>
<td>y</td>
</tr>
<tr>
<td>345672</td>
<td>345759</td>
<td>+</td>
<td>C/D box snoRNA 13</td>
<td>y</td>
</tr>
<tr>
<td>86078</td>
<td>86167</td>
<td>+</td>
<td>C/D box snoRNA 12</td>
<td>n</td>
</tr>
<tr>
<td>131771</td>
<td>131859</td>
<td>+</td>
<td>C/D box snoRNA 11</td>
<td>y</td>
</tr>
<tr>
<td>217020</td>
<td>217107</td>
<td>+</td>
<td>C/D box snoRNA 10</td>
<td>y</td>
</tr>
<tr>
<td>42115</td>
<td>42207</td>
<td>+</td>
<td>C/D box snoRNA 9</td>
<td>y</td>
</tr>
<tr>
<td>2175730</td>
<td>2175814</td>
<td>-</td>
<td>C/D box snoRNA 8</td>
<td>y</td>
</tr>
<tr>
<td>393323</td>
<td>393923</td>
<td>-</td>
<td>C/D box snoRNA 7</td>
<td>y</td>
</tr>
<tr>
<td>368172</td>
<td>368254</td>
<td>+</td>
<td>C/D box snoRNA 6</td>
<td>y</td>
</tr>
<tr>
<td>449443</td>
<td>449530</td>
<td>-</td>
<td>C/D box snoRNA 5</td>
<td>y</td>
</tr>
<tr>
<td>2217096</td>
<td>2217181</td>
<td>-</td>
<td>C/D box snoRNA 4</td>
<td>n</td>
</tr>
<tr>
<td>1117670</td>
<td>1117755</td>
<td>+</td>
<td>C/D box snoRNA 3</td>
<td>y</td>
</tr>
<tr>
<td>1075489</td>
<td>1075575</td>
<td>+</td>
<td>C/D box snoRNA 2</td>
<td>y</td>
</tr>
<tr>
<td>1220543</td>
<td>1220628</td>
<td>+</td>
<td>C/D box snoRNA 1</td>
<td>n</td>
</tr>
<tr>
<td>2179492</td>
<td>2179575</td>
<td>+</td>
<td>C/D box snoRNA 106</td>
<td>n</td>
</tr>
<tr>
<td>669541</td>
<td>669627</td>
<td>+</td>
<td>C/D box snoRNA 102</td>
<td>y</td>
</tr>
<tr>
<td>1388919</td>
<td>1388999</td>
<td>+</td>
<td>C/D box snoRNA 180</td>
<td>y</td>
</tr>
<tr>
<td>217025</td>
<td>217102</td>
<td>+</td>
<td>C/D box snoRNA 105</td>
<td>y</td>
</tr>
</tbody>
</table>
Supplementary Table 14. 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.

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Strand</th>
<th>Name</th>
<th>transrealign</th>
</tr>
</thead>
<tbody>
<tr>
<td>1152379</td>
<td>1152458</td>
<td>-</td>
<td>C/D box snoRNA 10</td>
<td>y</td>
</tr>
<tr>
<td>1117670</td>
<td>1117747</td>
<td>-</td>
<td>C/D box snoRNA 02</td>
<td>y</td>
</tr>
</tbody>
</table>

Supplementary Table 15. 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.

<table>
<thead>
<tr>
<th>Species</th>
<th>pooled reads</th>
<th>mapped reads</th>
<th>remapped reads</th>
<th>split reads</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. kandleri</em></td>
<td>19242863</td>
<td>13736113</td>
<td>2153772</td>
<td>128603</td>
</tr>
<tr>
<td><em>S. acidocaldarius</em></td>
<td>26023157</td>
<td>22762924</td>
<td>360912</td>
<td>108283</td>
</tr>
<tr>
<td><em>S. solfataricus</em></td>
<td>8976013</td>
<td>7764310</td>
<td>878127</td>
<td>74143</td>
</tr>
<tr>
<td><em>N. equitans</em></td>
<td>*</td>
<td>10728929</td>
<td>-</td>
<td>25721</td>
</tr>
<tr>
<td><em>I. hospitalis</em></td>
<td>*</td>
<td>5567812</td>
<td>-</td>
<td>42757</td>
</tr>
</tbody>
</table>
Supplementary Table 16. New splicesites with BHB elements found in *M. kandleri* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

<table>
<thead>
<tr>
<th>MSA1 Start</th>
<th>Stop</th>
<th>Strand</th>
<th>MSA2 Start</th>
<th>Stop</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>143515</td>
<td>143632</td>
<td>+</td>
<td>31858</td>
<td>31995</td>
<td>+</td>
</tr>
<tr>
<td>163201</td>
<td>163467</td>
<td>+</td>
<td>40439</td>
<td>40539</td>
<td>+</td>
</tr>
<tr>
<td>325320</td>
<td>325593</td>
<td>+</td>
<td>262091</td>
<td>262244</td>
<td>+</td>
</tr>
<tr>
<td>351364</td>
<td>351601</td>
<td>+</td>
<td>325412</td>
<td>325593</td>
<td>+</td>
</tr>
<tr>
<td>399364</td>
<td>399622</td>
<td>+</td>
<td>384725</td>
<td>384960</td>
<td>+</td>
</tr>
<tr>
<td>490225</td>
<td>490354</td>
<td>+</td>
<td>497434</td>
<td>497516</td>
<td>+</td>
</tr>
<tr>
<td>518084</td>
<td>518344</td>
<td>+</td>
<td>507442</td>
<td>507586</td>
<td>+</td>
</tr>
<tr>
<td>518229</td>
<td>518322</td>
<td>+</td>
<td>509798</td>
<td>509895</td>
<td>+</td>
</tr>
<tr>
<td>518263</td>
<td>518344</td>
<td>+</td>
<td>510730</td>
<td>510856</td>
<td>+</td>
</tr>
<tr>
<td>522017</td>
<td>522129</td>
<td>+</td>
<td>517708</td>
<td>517974</td>
<td>+</td>
</tr>
<tr>
<td>585393</td>
<td>585475</td>
<td>+</td>
<td>517864</td>
<td>518003</td>
<td>+</td>
</tr>
<tr>
<td>585629</td>
<td>585772</td>
<td>+</td>
<td>520831</td>
<td>520932</td>
<td>+</td>
</tr>
<tr>
<td>585946</td>
<td>586138</td>
<td>+</td>
<td>576556</td>
<td>576720</td>
<td>+</td>
</tr>
<tr>
<td>586198</td>
<td>586414</td>
<td>+</td>
<td>582234</td>
<td>582370</td>
<td>+</td>
</tr>
<tr>
<td>586751</td>
<td>586957</td>
<td>+</td>
<td>585979</td>
<td>586110</td>
<td>+</td>
</tr>
<tr>
<td>586869</td>
<td>587025</td>
<td>+</td>
<td>588030</td>
<td>588271</td>
<td>+</td>
</tr>
<tr>
<td>587307</td>
<td>587436</td>
<td>+</td>
<td>693682</td>
<td>693927</td>
<td>+</td>
</tr>
<tr>
<td>587746</td>
<td>587871</td>
<td>+</td>
<td>739727</td>
<td>739822</td>
<td>+</td>
</tr>
<tr>
<td>673889</td>
<td>673968</td>
<td>+</td>
<td>744880</td>
<td>745039</td>
<td>+</td>
</tr>
<tr>
<td>1019844</td>
<td>1020094</td>
<td>+</td>
<td>779240</td>
<td>779332</td>
<td>+</td>
</tr>
<tr>
<td>1022388</td>
<td>1022636</td>
<td>+</td>
<td>879679</td>
<td>879914</td>
<td>+</td>
</tr>
<tr>
<td>1112401</td>
<td>1112677</td>
<td>+</td>
<td>1183460</td>
<td>1183557</td>
<td>+</td>
</tr>
<tr>
<td>1221433</td>
<td>1221523</td>
<td>+</td>
<td>1219327</td>
<td>1219497</td>
<td>+</td>
</tr>
<tr>
<td>1243818</td>
<td>1244084</td>
<td>+</td>
<td>1232353</td>
<td>1232448</td>
<td>+</td>
</tr>
<tr>
<td>1343553</td>
<td>1343808</td>
<td>+</td>
<td>1252557</td>
<td>1252674</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1259272</td>
<td>1259422</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1279749</td>
<td>1279961</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1415483</td>
<td>1415647</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1540620</td>
<td>1540755</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1688642</td>
<td>1688778</td>
<td>+</td>
</tr>
</tbody>
</table>
**Supplementary Table 17.** New splicesites with BHB elements found in *S. acidocaldarius* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

<table>
<thead>
<tr>
<th>MSA1 Start</th>
<th>MSA1 Stop</th>
<th>MSA1 Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>49182</td>
<td>49271</td>
<td>+</td>
</tr>
<tr>
<td>73211</td>
<td>72447</td>
<td>+</td>
</tr>
<tr>
<td>708452</td>
<td>708600</td>
<td>+</td>
</tr>
<tr>
<td>789407</td>
<td>789634</td>
<td>+</td>
</tr>
<tr>
<td>848243</td>
<td>848477</td>
<td>+</td>
</tr>
<tr>
<td>1013752</td>
<td>1013976</td>
<td>+</td>
</tr>
<tr>
<td>1029123</td>
<td>1029366</td>
<td>+</td>
</tr>
<tr>
<td>1030222</td>
<td>1030393</td>
<td>+</td>
</tr>
<tr>
<td>1270855</td>
<td>1270973</td>
<td>+</td>
</tr>
<tr>
<td>1285009</td>
<td>1285134</td>
<td>+</td>
</tr>
<tr>
<td>1366353</td>
<td>1366466</td>
<td>+</td>
</tr>
<tr>
<td>1785213</td>
<td>1785353</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSA2 Start</th>
<th>MSA2 Stop</th>
<th>MSA2 Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>49182</td>
<td>49271</td>
<td>+</td>
</tr>
<tr>
<td>73220</td>
<td>73336</td>
<td>+</td>
</tr>
<tr>
<td>97306</td>
<td>97393</td>
<td>+</td>
</tr>
<tr>
<td>267193</td>
<td>267304</td>
<td>+</td>
</tr>
<tr>
<td>450894</td>
<td>450975</td>
<td>+</td>
</tr>
<tr>
<td>677737</td>
<td>677855</td>
<td>+</td>
</tr>
<tr>
<td>624809</td>
<td>624949</td>
<td>+</td>
</tr>
<tr>
<td>639014</td>
<td>639153</td>
<td>+</td>
</tr>
<tr>
<td>753133</td>
<td>753241</td>
<td>+</td>
</tr>
<tr>
<td>808261</td>
<td>808346</td>
<td>+</td>
</tr>
<tr>
<td>913155</td>
<td>913261</td>
<td>+</td>
</tr>
<tr>
<td>986137</td>
<td>986328</td>
<td>+</td>
</tr>
<tr>
<td>1126119</td>
<td>1126204</td>
<td>+</td>
</tr>
<tr>
<td>1212710</td>
<td>1212788</td>
<td>+</td>
</tr>
<tr>
<td>1327997</td>
<td>1328130</td>
<td>+</td>
</tr>
<tr>
<td>1381538</td>
<td>1381622</td>
<td>+</td>
</tr>
<tr>
<td>1648994</td>
<td>1649075</td>
<td>+</td>
</tr>
<tr>
<td>2072492</td>
<td>2072665</td>
<td>+</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>MSA1 Start</th>
<th>MSA1 Stop</th>
<th>MSA1 Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>396266</td>
<td>396385</td>
<td>+</td>
</tr>
<tr>
<td>433125</td>
<td>433234</td>
<td>+</td>
</tr>
<tr>
<td>396539</td>
<td>396812</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSA2 Start</th>
<th>MSA2 Stop</th>
<th>MSA2 Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>403449</td>
<td>403535</td>
<td>+</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>MSA1 Start</th>
<th>MSA1 Stop</th>
<th>MSA1 Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>603277</td>
<td>603363</td>
<td>+</td>
</tr>
<tr>
<td>733481</td>
<td>733720</td>
<td>+</td>
</tr>
<tr>
<td>798681</td>
<td>798876</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSA2 Start</th>
<th>MSA2 Stop</th>
<th>MSA2 Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>3921</td>
<td>4027</td>
<td>+</td>
</tr>
<tr>
<td>281959</td>
<td>282165</td>
<td>+</td>
</tr>
<tr>
<td>602934</td>
<td>603198</td>
<td>+</td>
</tr>
<tr>
<td>731896</td>
<td>732076</td>
<td>+</td>
</tr>
<tr>
<td>733248</td>
<td>733446</td>
<td>+</td>
</tr>
<tr>
<td>945608</td>
<td>945702</td>
<td>+</td>
</tr>
<tr>
<td>1115636</td>
<td>1115739</td>
<td>+</td>
</tr>
</tbody>
</table>
**Supplementary Table 20.** New splicesites with BHB elements found in *S. solfataricus* with MSA1 and MSA2. Start and Stop positions include the 15nt flanking sequence.

<table>
<thead>
<tr>
<th></th>
<th>MSA1</th>
<th>MSA2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>Stop</td>
<td>Strand</td>
<td>Start</td>
</tr>
<tr>
<td></td>
<td>Stop</td>
<td>Strand</td>
<td>Stop</td>
</tr>
<tr>
<td>763512</td>
<td>763600</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>872878</td>
<td>873132</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>872918</td>
<td>873156</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>875970</td>
<td>876218</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>876228</td>
<td>876426</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2361384</td>
<td>2361469</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2939076</td>
<td>2939162</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
References