# Comparative Detection of Processed Small RNAs in Archaea

Christian Höner zu Siederdissen[1], Sarah Berkemer[2], Fabian Amman[2,3], Axel Wintsche[3,4], Sebastian Will[2,3], Sonja J. Prohaska[3,4], and Peter F. Stadler[1,2,3,5,6,7,8]

[1] Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria.
[2] Bioinformatics Group, Department of Computer Science University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany.
[3] Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany.
[4] Computational EvoDevo Group, Department of Computer Science, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany.
[5] Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany.
[6] Fraunhofer Institut für Zelltherapie und Immunologie, Perlickstraße 1, D-04103 Leipzig, Germany.
[7] Center for non-coding RNA in Technology and Health, University of Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg C, Denmark.
[8] Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501.
E-mail:studla@bioinf.uni-leipzig.de

**Abstract.** Enzymatic splicing in Archaeal tRNAs is guided by bulge-helix-bulge structural elements, while much less seems to be known about splicing in other small RNAs (sRNAs). We conduct a genome-wide analysis of several archaeal genomes to identify putative BHB elements and compare our findings with available RNA-seq data. We also provide an analysis of the viability of using pattern-based and stochastic structural scanning algorithms for in silico studies of the occurrence of BHB motifs. Furthermore, we comment on splicing motifs in other small RNAs, which mostly do not fit the pattern of bulge-helix-bulge motifs.

Appendix and supporting files available at:
http://www.bioinf.uni-leipzig.de/publications/supplements/14-001

**Keywords:** circular RNA, Archaea, splicing, structure-based motif search

## 1 Introduction

Until recently, not much was known about archaeal small non-coding RNAs (sRNAs) in general, at least in part because of the high level of sequence divergence between the available genomes hampers homology-based annotation. They

share ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and the RNA components of RNase P and the signal recognition particle (7S RNA) with the other two domains of life. A wide variety of further small non-coding RNA species have been described for individual species. Only two RNA classes recognizable within this diversity are well understood, however: With Eubacteria Archaea share CRISPR-Cas adaptive immune systems [1]. Like Eukarya, they use a diversity of box C/D and box H/ACA snoRNAs to direct chemical modifications of rRNAs and other non-coding RNAs [2] but they lack a spliceosomal splicing machinery.

Instead, enzymatic splicing is a common feature in archaeal RNA processing. The molecular mechanism, which is closely related to tRNA splicing in Eukarya, involves a specific endonuclease that recognizes and cleaves the so-called bulge-helix-bulge (BHB) structure and a specific ligase which joins the exons and circularizes the intron [3–5]. Intriguingly, archaeal tRNAs may have multiple introns [6]. Furthermore, the same mechanism implements a form of trans-splicing that composes tRNAs from two or three independently encoded fragments [7–11]. The maturation of the ribosomal RNAs makes use of the same machinery [12], cutting 16S and 23S rRNA from a longer precursor transcript, which again is guided by a BHB element formed over long distance. The BHB elements of tRNA introns and rRNAs have been studied in quite some detail already [13–17]. They are well-conserved across the Crenarchaeota phylum [13]. Also mRNA, namely the pre-mRNA of CBF5, the archaeal homolog of dyskerin, was reported to contain an intron with a BHB element in many crenarchaeal species [18].

Circularized forms of small RNAs are also abundant in many Archaea. At least some box C/D snoRNAs appear predominantly or even exclusively as circular RNAs [19–21]. This is also true for assorted other small RNA species, among them the 5S rRNA [20]. Little is known, however, about their biogenesis. BHB-element-dependent splicing has been reported as a circularization mechanism for a snoRNA only in the special case of the box C/D snoRNA processed from a long intron in the tRNA-Trp of *Pyrococcus* species [22].

Here we explore two interrelated questions: (1) Can BHB-element-dependent splicing explain all or at least most of the observed circularized or permuted small RNAs, and (2) to what extent can BHB elements be used in their own right as means of detecting novel small RNAs and/or likely sites of RNA processing. This is of particular interest since some of the RNA processing products involved cannot be directly observed in RNA-seq data for a variety of reasons: (1) Circularized RNAs are depleted in most RNA-seq protocols unless specifically enriched e.g. by RNase R treatment [20]. (2) Spliced tRNAs cannot be detected in cases where an unspliced paralog is present in the same genome [21]. (3) Circularized introns e.g. of tRNAs are often too short to be detectable by sequencing in their own right. For instance 5 of 8 introns listed in [23] have a length of 26 nt or less.

## 2 Circularized sRNAs with and without BHB elements

Most of the BHB elements characterized in the literature derive from tRNAs [13, 20, 23]. They have been catalogued into three classes in [13], see Figure 1, of which two can be seen as relaxed versions of the structurally most complex group. We extended the alignments of [13] by adding the tRNA associated BHB elements from ref. [23] as well as the sequences of a subset of the box C/D snoRNAs reported in [23] that conform to the established BHB patterns according to visual inspection. After extensive manual curation we obtained a multiple alignment that highlights a very well-conserved secondary structure pattern comprising two bulges, which contain the cleavage sites, separated by a stable helix. Although there is no clear sequence consensus, the helix contains many GC base pairs, see Fig. 1. We therefore used the multiple alignment to build both, `RNAbob` [24] pattern descriptors and `Infernal` [25] covariance models to search archaeal species (*M. kandleri*, *S. solfataricus*, *S. acidocaldarius*) for putative BHB elements.

After reviewing the number of putative elements predicted by each method (i.e in *S.acidocaldarius*: 89 310 `RNAbob` hits vs. 2 534 `Infernal` hits; *M.kandleri*: 483 829 `RNAbob` hits vs. 24 934 `Infernal` hits; models based on Fig. 1) we chose to use only `Infernal`-based models. Here, the generality of the consensus sequence of the multiple alignment is mirrored by the generality of the `RNAbob` pattern.

Our choice of `Infernal` models allows for a more fine-grained control between sensitivity and specificity. The `RNAbob` patterns mostly constrain the length of the individual helices, the two bulges, and the intron (marked *loop* in Fig. 1). `Infernal` covariance models (CMs) capture sequence and structural statistical information, and thereby offer two advantages over `RNAbob` patterns: (i) CMs deal gracefully with degrading matches that can only partially match the given model; and (ii) each hit is accompanied by a measure of how well it fits the model. The second property is especially useful compared to the simple hit or miss events of `RNAbob` as it allows us to *rank* hits by how well they match the consensus model as we need around 10–50 times more `RNAbob` hits (more general patterns) to achieve the same level of sensitivity than with `Infernal` – leading to much lower specificity.

Information on circular sRNAs was collected from the literature for *Sulfolobus* [20]. In addition, we re-analyzed publicly available RNA-seq data for *Methanopyrus kandleri* [26] following the strategy outlined in [21]. In brief, we retrieved the data sets SRR769472 to SRR769505 from the short read archive, pooled the different runs into a single library, quality-trimmed the reads and mapped them to the genome of *M. kandleri* (NC_003551) using `segemehl` [27, 28]. The option `--splits` forces mapping reads, if possible, across splice sites, so that a seed of at least 11 nt maps to each side of the split. 80.3% of the 61,129,675 reads were mapped. Split reads indicating a circularization event were extracted using in-house python scripts. Valid circularization sites were defined as all sites covered by at least two circularized reads, spanning at most 200 nt, so that for each of the two single circularization sites the majority of split reads must be involved in this particular circularization event. Supplementary Table 2 gives an overview of the characterized loci in *M. kandleri*.

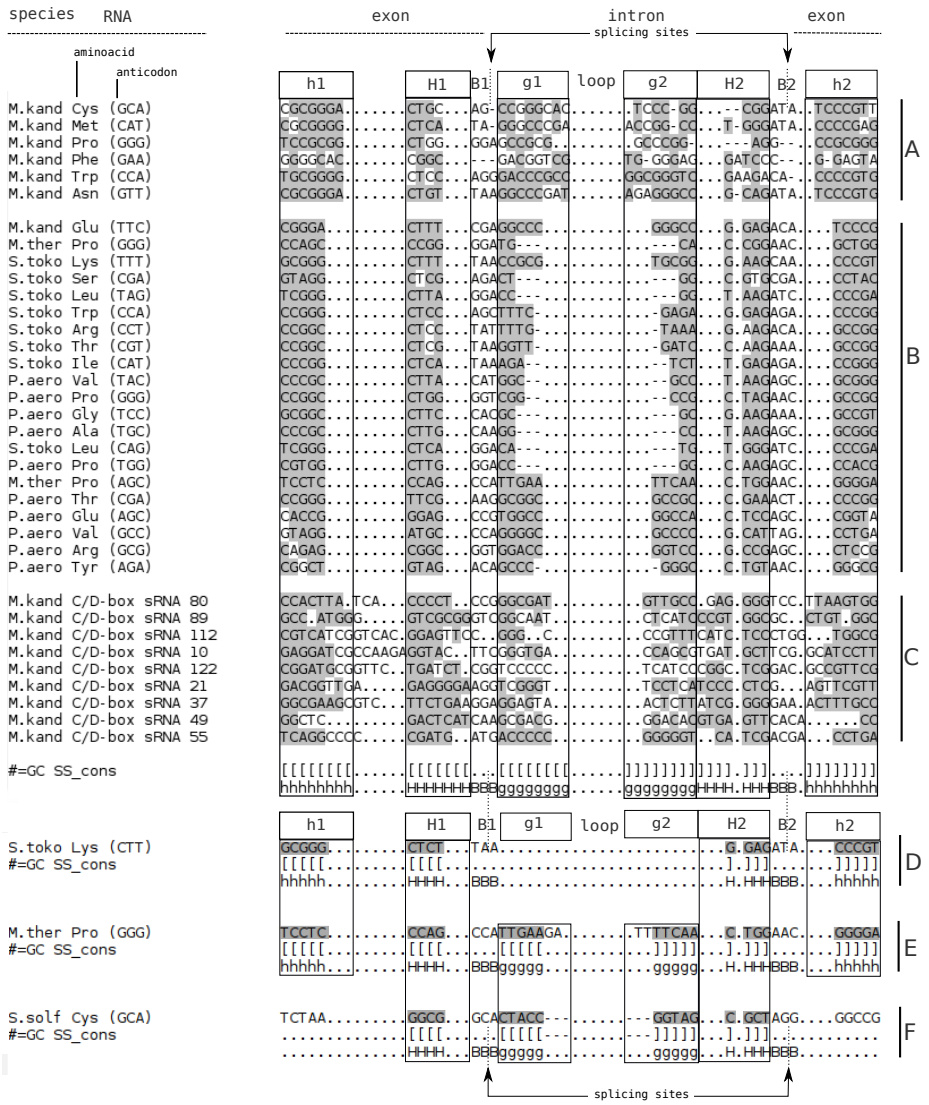


**Fig. 1.** Consensus multiple alignment of BHB structural elements. The two cleavage sites in the bulges (labeled B1 and B2 and marked by BBB at the bottom) are indicated by arrows. Base pairs are denoted by `[,]`. The three helical regions are highlighted and labelled `H`, `h`, and `g`, resp. Sequence blocks A (tRNA sequences) and C (box C/D snoRNAs) are taken from [23], the tRNAs B are from [13]. Below, the three three variants of consensus structures for tRNA BHB elements [13] are shown with a paradigmatic example: The three-stem structure E most closely resembles our consensus. Both D and F are degenerate versions, missing the inner and outer stem, respectively, with the stem II denoted 'H' being conserved.

In analogy we processed the RNA-seq reads from *S. acidocaldarius* pooled from [29, 30]. There, after quality control, 26,023,157 reads remained, whereof 88% could be mapped to the reference genome (NC_007181). In this way, we collected 20 circular RNA (circRNA) candidates in *M. kandleri* and 65 candidates in *S. acidocaldarius*. For a more detailed comparative inspection of these candidates, we located potential homologs by a blast search against all publicly available archaeal genomes from NCBI genbank. For each locus, we determined all corresponding sequences up to tolerant e-values of 0.01 for *M. kandleri*, where no close relatives exist in the database, and much more conservative $10^{-30}$ for *S. acidocaldarius*. For *M. kandleri*, such potential homologs existed in 6 cases and resulted in one to nine potential homologs per candidate; for *S. acidocaldarius*, we identify at least 2 homologs for all candidates. Finally, we evaluated `ClustalW` alignments of the sequences with `RNAz` [31] to detect potential RNA structure. For *M. kandleri*, this indicated a stable and evolutionary conserved structure of the circRNA candidate for only one locus `1500955-1501112` (Suppl. Fig. 2); there, with a very high RNA class probability of 0.9992. A search of the `Rfam` database could not assign the structural circRNA candidate to any known non-coding RNA family. For *S. acidocaldarius*, RNAz predicts RNA structure in 9 cases with RNAz class probabilities $\geq 0.5$ (Suppl. Table 4).

Surprisingly, a large fraction of sRNAs with evidence for circularization are not flanked by BHB elements and thus cannot be understood as products of the canonical archaeal RNA processing machinery. Although circular RNAs are depleted in RNA-seq data it is possible to detect some circularization products as well as co-linear splicing products. However, PCR artifacts and other technical difficulties make it hard to distinguish *bona fide* sRNAs from non-biological noise based on the deep sequencing data alone. Hence the absence of BHB elements from a putative splicing or circularization product may hint at an experimental artefact. We therefore consider only circular RNAs without BHB elements as examples of BHB-unrelated circularization that have also been reported in the literature.

## 3 Survey for BHB elements

The BHB elements are short (12-23 base pairs) compared to the models of non-coding RNAs (ncRNAs) collected in the `Rfam` database [32]. The situation is further complicated by the inherent non-locality of BHB-elements, which span an intron or even the entire functional ncRNA. We therefore construct `Infernal` models that can handle the large intronic inserts. As a consequence the sensitivity of the models is limited, making it necessary to disable all heuristic pre-filters that boost the performance of `Infernal` and to reduce the cutoffs at which a hit is reported. The small size of the archaeal genomes nevertheless leaves us with a tractable problem.

In principle we would like to score *just* the BHB structure and completely ignore the large intronic insert. Such a pattern, however, cannot be handled by the current implementation of `Infernal`. As a workaround, we explicitly

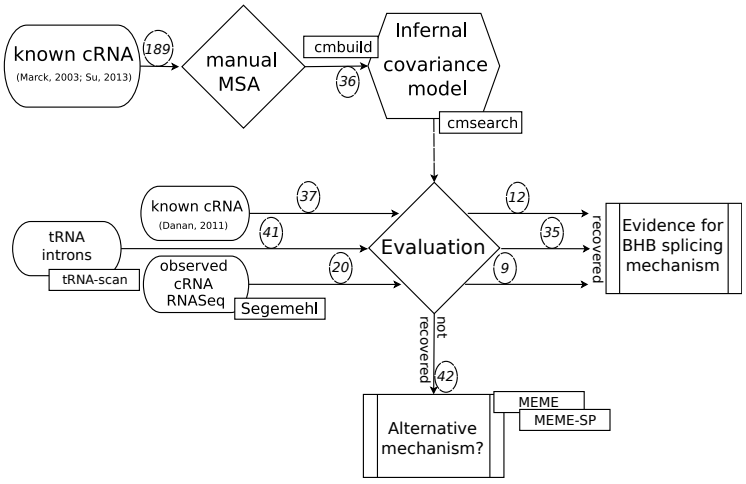


**Fig. 2.** Workflow and summary of results for genome-wide BHB survey and circular or spliced RNA survey. From a cache of known circRNA sequences, we created a curated multiple-sequence alignment (MSA), which serves as the basis for both, a pattern-based descriptor as well as a stochastic `Infernal` model. Evaluation is done using data from RNA-seq experiments and putative BHB elements founds in a number of archael genomes. Together this data yields 56 BHB candidates with strong in-silico evidence. Known circularized circRNA without BHB elements were further subjected to sequence and structure motif discovery using `MEME` and `MEME-SP`.

model the insert as a long stretch of `N` characters in the input alignment for `cmbuild` so that the covariance model then allows large insertions relative to the structural consensus to occur with very small cost but only at the denoted position. As evidenced by the results given in Supplementary Tables 1–3, this is enough to recover a large number of putative BHB sites in archaeal genomes with acceptable running times in the order of 1–5 hours (a task that is parallelized by `Infernal`). Due to the body of literature available for the previously mentioned genomes (*M. kandleri, S. solfataricus, S. acidocaldarius*) we chose those as well as two additional genomes (*P. furiosus*, *N. equitans*, data not shown) for this survey.

The results of the computational survey is summarized in Fig. 2. Based on a subset of 189 known circRNA, we created a manually curated multiple-sequence alignment. This alignment serves as a basis for genome-wide scans in several archaeal genomes. The putative BHB elements discovered in those scans were evaluated using additional evidence gathered from RNA-seq data. In total, we accept 56 BHB structures that show strong in silico evidence. A further 42 sequences lack evidence for BHB structures and were analysed using the sequence and structural motif discovery programs `MEME` [33] and `MEME-SP` [34]. For details see Sec. 4.

Most archaeal tRNAs, including the spliced ones, can be detected efficiently in the genome using `tRNAscan-SE` [35] albeit with the notable exception of split and permuted tRNAs. Thus, we ran `tRNAscan-SE -A` for all genomes to check whether BHB-based candidate loci recover the already known tRNA genes. We obtained a recall of 85% for the tRNAs with introns of *Methanopyrus kandleri* (6/7), *Sulfolobus solfataricus* (14/16), and *Sulfolobus acidocaldarius* (15/18). The well-described BHB elements flanking the 16S and 23S are not recovered as expected, since these insertions are much too long for the CM-based approach (see Supplementary Table 1).

Only two annotated box C/D snoRNAs (Sso-sR8 and Sso-sR4) in *S. solfataricus* are flanked by BHB elements. Of these only Sso-sR8 has a homolog in *S. acidocaldarius* (homology search conducted with `GotohScan` [36]), albeit without a recognizable BHB structure. The single box H/ACA snoRNA sR109 and the 7S RNA are the only two known sRNAs with a conserved BHB element in both *Sulfolobus* species. The best BHB hit within coding regions in *Sulfolobus* is the previously known intron of the CBF5 pre-mRNA [37]. As expected from the analysis in [18] the CBF5 intron is absent in *M. kandleri*. In *S. solfataricus* we find BHB elements for one sRNA that has not been further characterized in [20] and one intergenic location.

Most of the BHB candidates detected with the CM do not coincide with known RNA processing sites. We therefore investigated to what extent they match circularized RNA-seq reads. A comparison of the top 2 500 candidates for BHB elements with the `Rfam` data base did not overlap snoRNAs in *M. kandleri*, while 8 and 14 snoRNAs were recovered in *S. solfataricus* and *S. acidocaldarius*, respectively, see Supplemental Figure 1.

For *M. kandleri* we identified 20 circularized products (besides the tRNA introns) from the RNA-seq data of which 9 match candidate BHB elements. Of these, 6 are located within annotated protein coding genes. Conversely, of the 9 intergenic circular RNAs only three are associated with BHB elements. Of the 11 introns within ORFs, 7 preserve the reading frame, suggesting that enzymatic splicing may lead to functional isoforms.

In interpreting the relatively small overlap between circular RNAs and BHB elements one has to keep in mind, however, that many of the introns, notably those of tRNAs, have a length not much longer than the theoretical detection limit of 22 nt for a circularizing junction. Such short introns thus need to be identified by considering the spliced reads directly.

## 4 Circular RNAs without BHB elements

No RNA processing mechanism other than the BHB-element directed splicing has been described in Archaea that could explain the abundant circularized RNA species. Hence we searched for possible sequence and/or secondary structure elements that could be involved in circularization. To this end we extended all circular RNAs without recognizable BHB elements by 50nt at both circularization sites. For *Sulfolobus solfataricus* a map of transcription start sites is

available [38]. We pruned the 5' extensions at annotated start sites, since we expect any processing signals to reside within the RNA transcript. In total, we used 107 published sRNA sequences from *M. kandleri*, 11 published sRNAs from *S. solfataricus* and 50 circularized sequences from our analysis of RNA-seq data (*M. kandleri*: 11, *S. acidocaldarius*: 16, and *S. solfataricus*: 23) for motif discovery. These include a set of sequences known from RNA-seq to be circularized in addition to those sequences where no BHB structural element was found.

We used `MEME` (version 4.8.1) [33] with parameters `-mod zoops -minw 4 -maxw 10` as well as `MEME-SP` [34], an extension of `MEME` to search for combined patterns of sequence and RNA secondary structure. `MEME-SP` uses the same expectation maximization framework as `MEME` to learn from sequences annotated with secondary structure profiles that specify for each position of each input sequence the probability that the base is paired upstream, downstream, or unpaired. This secondary structure information is computed from all predicted locally stable sub-structures using `RNALfold` [39] with parameters `-T 70` (to account for the fact that *Sulfolobus solfataricus* is a hyperthermophile) and a window size that encompasses the entire input sequence.

Using `MEME` credible sequence motifs were found only in the circular C/D-box snoRNA sequences published in [23]. As expected, the recovered pattern matched the box C and box D sequences. No further conserved motifs could be found except short recurring stretches of guanines or cytosines, which are scattered across the entire input sequence, however. `MEME-SP`, in contrast, predicted motifs that show conservation in their base-pairing patterns rather than sequence. Two kinds of structural patterns were observed corresponding to either 5' parts of helices or 3' parts of helices, again without specific localization. Thus there is no indication for a specific pattern associated with a putative BHB-element-independent circularization mechanism.

## 5 Conclusions

Although some of the circular RNAs in Archaea are most likely produced by BHB-element-dependent enzymatic splicing, our analysis shows that this cannot be the only mechanism. While tRNAs, rRNAs (including 5S rRNA), 7S RNA, as well as a few of the snoRNAs in *Sulfolobus* are associated with recognizable BHB elements, this is not the case in general. In fact the majority of the box C/D snoRNAs and most of the unclassified sRNAs are not associated with BHB elements. A search for sequence motifs as well as combined sequence-structure patterns did not reveal any indication for a common alternative processing pathway, however.

We have introduced here a workflow for identifying putative archaeal sRNAs based on a structural stochastic matcher in the form of an `Infernal` covariance model. BHB elements are difficult to identify because they do not rely on distinctive sequence patterns but are defined largely by a non-local secondary structure element. The survey presented here, therefore, can serve only as a starting point

for a more detailed investigation into the realm of archaeal sRNAs and their processing.

Nevertheless, the combination of a hand-curated multiple-sequence alignment, genome-wide scans and validation with RNA-seq data allowed us to recover known intron-containing circRNAs that are spliced using the BHB structural pattern, as well as to discover novel candidates that present strong evidence for BHB-based splicing. For tRNAs we confirm the observation that the BHB pattern is strongly conserved. While most tRNA introns can not be recovered from RNA-seq data, in silico methods for tRNA discovery are very successful (i.e. using `tRNAscan-SE` [35]), so that tRNA BHB discovery can serve as a general test for the applicability of our method.

For non-tRNA, the question of how successful splice-site detection via BHB element discovery becomes more intriguing, since the BHB structural pattern is only evident in a subset of the sequences. Again, we were able to recover known introns, as well as discover novel candidates. The absence of the BHB motif for some other known circularized RNAs suggests that another splicing mechanism is involved as well.

The work presented here poses a number of open questions that we will need to address in the future. The BHB motif is an excellent example of a *non-local* motif that consists of two external regions bracketing an internal region, namely the intron to be spliced out. Covariance models [40] as used by `Infernal` are, however, "local" in that every nucleotide aligned to the model is scored. For any normal RNA family this does not constitute a major problem as the whole sequence "belongs" to the model. In our case, this is different in that the intron should not be considered or scored at all.

While it is not possible to handle arbitrary insertions in complete independence of the scoring and alignment mechanism of `Infernal`, we were still able to design families that allow for sufficiently large insertions of basically random intronic regions by careful model construction, disabling of any sequence-based pre-filters and the consideration of a large set of BHB candidates. In the future we hope to ameliorate this with the construction of a specialised stochastic machinery. That such machines can be constructed with reasonable effort has recently been shown [41, 42]. Earlier work on more principled construction of stochastic context-free grammars [43] will also be helpful in this case.

Detection of archaeal splicing sites also provides another challenge in the form of *trans splicing*. Mechanistically, this is achieved by ligating transcripts produced independently from two distinct genomic loci. Computationally, we have to deal with "spliced out" regions that can be of arbitrary (even genome-length) size. Currently, no computational approach is available to handle such cases and while an algorithm following the ideas above might be able to detect suitable candidates with any intronic size, the resulting running times would be prohibitive. Allowing genome-size intronic regions to occur effectively squares the genome size for any scanning algorithm pushing the running time from the order of CPU hours to the order of CPU years for a single genome and model. How to handle such cases will also be subject of future research. We note that in

the special case of split tRNAs computational approaches become feasible since the parts of known, highly conserved tRNA sequences included in each transcript can be used as anchor points. This is exploited in the `SPLITS` tool [44]. Of course this idea does not generalize to the case of unknown or rapidly evolving sRNA genes.

**Acknowledgments**

This work was funded, in part, by the Austrian FWF, project "SFB F43 RNA regulation of the transcriptome", the German ministry of science (0316165C as part of the e:Bio initiative). AW was funded by a PhD stipend from the European Social Fund. We thank Udo Bläsi and Sonja-Verena Albers for sharing RNA-seq data.

## References

1. Barrangou, R.: CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip Rev RNA **4** (2013) 267–278
2. Dennis, P.P., Omer, A.: Small non-coding RNAs in Archaea. Curr. Op. Microbiol. **8** (2005) 685–694
3. Thompson, L.D., Daniels, C.J.: A tRNA (trp) intron endonuclease from halobacterium volcanii. Unique substrate recognition properties. Journal of Biological Chemistry **263**(34) (1988) 17951–17959
4. Kjems, J., Garrett, R.A.: Novel splicing mechanism for the ribosomal RNA intron in the archaebacterium Desulfurococcus mobilis. Cell **54**(5) (1988) 693–703
5. Salgia, S.R., Singh, S.K., Gurha, P., Gupta, R.: Two reactions of *Haloferax volcanii* RNA splicing enzymes: Joining of exons and circularization of introns. RNA **9** (2003) 319–330
6. Sugahara, J., Yachie, N., Arakawa, K., Tomita, M.: In silico screening of archaeal tRNA-encoding genes having multiple introns with bulge-helix-bulge splicing motifs. RNA **13** (2007) 671–681
7. Randau, L., Münch, R., Hohn, M.J., Jahn, D., Söll, D.: *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'-and 3'-halves. Nature **433**(7025) (2005) 537–541
8. Randau, L., Söll, D.: Transfer RNA genes in pieces. EMBO Rep. **9** (2008) 623–628
9. Fujishima, K., Sugahara, J., Kikuta, K., Hirano, R., Sato, A., Tomita, M., Kanai, A.: Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. Proc Natl Acad Sci USA **106** (2009) 2683–2687
10. Sugahara, J., Fujishima, K., Morita, K., Tomita, M., Kanai, A.: Disrupted tRNA gene diversity and possible evolutionary scenarios. Journal of molecular evolution **69**(5) (2009) 497–504
11. Richter, H., Mohr, S., Randau, L., et al.: C/D box sRNA, CRISPR RNA and tRNA processing in an archaeon with a minimal fragmented genome. Biochemical Society Transactions **41**(part 1) (2013)
12. Baserga, S.J.: Ribonucleoproteins in archaeal pre-rRNA processing and modification. Archaea **2013** (2013)

13. Marck, C., Grosjean, H.: Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. RNA **9**(12) (2003) 1516–1531
14. Ghosh, Z., Chakrabarti, J., Mallick, B., Das, S., Sahoo, S., Sethi, H.S.: tRNA-isoleucine-tryptophan composite gene. Biochemical and biophysical research communications **339**(1) (2006) 37–40
15. Tocchini-Valentini, G.D., Fruscoloni, P., Tocchini-Valentini, G.P.: Processing of multiple-intron-containing pretRNA. Proceedings of the National Academy of Sciences **106**(48) (2009) 20246–20251
16. Chan, P.P., Cozen, A.E., Lowe, T.M.: Discovery of permuted and recently split transfer RNAs in Archaea. Genome Biol **12**(4) (2011) R38
17. Yamazaki, S., Yoshinari, S., Kita, K., Watanabe, Y., Kawarabayasi, Y.: Identification of an entire set of tRNA molecules and characterization of cleavage sites of the intron-containing tRNA precursors in acidothermophilic crenarchaeon *Sulfolobus tokodaii* strain7. Gene **489** (2011) 103–110
18. Yokobori, S., Itoh, T., Yoshinari, S., Nomura, N., Sako, Y., Yamagishi, A., Oshima, T., Kita, K., Watanabe, Y.: Gain and loss of an intron in a protein-coding gene in Archaea: the case of an archaeal RNA pseudouridine synthase gene. BMC Evol Biol **9** (2009) 198
19. Starostina, N.G., Marshburn, S., Johnson, L.S., Eddy, S.R., Terns, R.M., Terns, M.P.: Circular box C/D RNAs in *Pyrococcus furiosus*. Proc. Natl. Acad. Sci. U.S.A. **101** (2004) 14097–14101
20. Danan, M., Schwartz, S., Edelheit, S., Sorek, R.: Transcriptome-wide discovery of circular RNAs in archaea. Nucleic Acids Res. **40** (2012) 3131–3142
21. Doose, G., Alexis, M., Kirsch, R., Findeiß, S., Langenberger, D., Machné, R., Mörl, M., Hoffmann, S., Stadler, P.F.: Mapping the RNA-seq trash bin: Unusual transcripts in prokaryotic transcriptome sequencing data. RNA Biology **10** (2013) 1204–1210
22. Clouet d'Orval, B., Bortolin, M.L., Gaspin, C., Bachellerie, J.P.: Box C/D RNA guides for the ribose methylation of archaeal tRNAs: the tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. Nucleic Acids Res. **29** (2001) 4518–4529
23. Su, A.A., 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 research (2013)
24. Eddy, S.: RNABOB: a program to search for RNA secondary structure motifs in sequence databases (1996)
25. Nawrocki, E.P., Kolbe, D.L., Eddy, S.R.: Infernal 1.0: inference of RNA alignments. Bioinformatics **25**(10) (2009) 1335–1337
26. Su, A.A., 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 research (2013)
27. 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 computational biology **5**(9) (2009) e1000502
28. Hoffmann, S., Otto, C., Doose, G., Tanzer, A., Langenberger, D., Christ, S., Kunz, M., Holdt, L., Teupser, D., Hackermüller, J., Stadler, P.F.: A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. Genome Biol. (2014) in press.

29. Märtens, B., Amman, F., Manoharadas, S., Zeichen, L., Orell, A., Albers, S.V., Hofacker, I., Bläsi, U.: Alterations of the transcriptome of Sulfolobus acidocaldarius by exoribonuclease aCPSF2. PloS one **8**(10) (2013) e76569
30. Reimann, J., Esser, D., Orell, A., Amman, F., Pham, T.K., Noirel, J., Lindås, A.C., Bernander, R., Wright, P.C., Siebers, B., et al.: Archaeal signal transduction: Impact of protein phosphatase deletions on cell size, motility, and energy metabolism in Sulfolobus acidocaldarius. Molecular & Cellular Proteomics **12**(12) (2013) 3908–3923
31. Gruber, A.R., Findeiß, S., Washietl, S., Hofacker, I.L., Stadler, P.F.: `RNAz 2.0`: improved noncoding RNA detection. Pac. Symp. Biocomput. **15** (2010) 69–79
32. Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., et al.: Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Research **39**(suppl 1) (2011) D141–D145
33. Bailey, T.L., Elkan, C.: Fitting a mixture model by expectation maximization to discover motifs in bipolymers (1994)
34. Wintsche, A., Stadler, P.F., Prohaska, S.J.: MEME-SP: de novo prediction of short RNA motifs. in preparation
35. Lowe, T.M., Eddy, S.R.: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research **25**(5) (1997) 0955–964
36. 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 research **37**(5) (2009) 1602–1615
37. Watanabe, Y.i., Yokobori, S.i., Inaba, T., Yamagishi, A., Oshima, T., Kawarabayasi, Y., Kikuchi, H., Kita, K.: Introns in protein-coding genes in Archaea. FEBS letters **510** (2002) 27–30
38. 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
39. Hofacker, I.L., Priwitzer, B., Stadler, P.F.: Prediction of locally stable RNA secondary structures for genome-wide surveys. Bioinformatics **20**(2) (2004) 186–190
40. Eddy, S.R., Durbin, R.: RNA sequence analysis using covariance models. Nucleic Acids Res. **22** (1994) 2079–2088
41. Höner zu Siederdissen, C., Hofacker, I.L., Stadler, P.F.: How to Multiply Dynamic Programming Algorithms. In: Brazilian Symposium on Bioinformatics (BSB 2013). Volume 8213 of Lecture Notes in Bioinformatics., Springer, Heidelberg (2013) 82–93
42. Höner zu Siederdissen, C., Hofacker, I.L., Stadler, P.F.: Product Grammars for Alignment and Folding. submitted (2014)
43. Giegerich, R., Höner zu Siederdissen, C.: Semantics and Ambiguity of Stochastic RNA Family Models. IEEE/ACM Transactions on Computational Biology and Bioinformatics **8**(2) (2011) 499–516
44. Sugahara, J., Yachie, N., Sekine, Y., Soma, A., Matsui, M., Tomita, M., Kanai, A.: SPLITS: a new program for predicting split and intron-containing tRNA genes at the genome level. In Silico Biol. **6** (2006) 411–418