# A novel family of plasmid-transferred anti-sense ncRNAs

**Sven Findeiß[a], Cornelius Schubert[b], Peter F. Stadler[a,c−f], and Ulla Bonas[b]**

[a] Bioinformatics Group, Department of Computer Science; and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, D-01407, Leipzig, Germany
[b] Institute of Biology, Department of Genetics, Martin-Luther-University Halle-Wittenberg, Weinbergweg 10, D-06120 Halle, Germany
[c] Max-Planck-Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany
[d] Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria
[e] Fraunhofer Institute for Cell Therapy und Immunology, Deutscher Platz 5e, D-04103 Leipzig, Germany
[f] Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM-87501, USA

The genome of *Xanthomonas campestris* pv. *vesicatoria* encodes a constitutively expressed small RNA, which we designate PtaRNA1, "Plasmid transferred anti-sense RNA". It exhibits all hallmarks of a novel RNA antitoxin that proliferates by frequent horizontal transfer. It shows an erratic phylogenetic distribution with occurrences on chromosomes in a few individual strains distributed across both beta- and gamma-proteobacteria. Moreover, a homologous gene located on plasmid pMATVIM-7 of *Pseudomonas aeruginosa* is found. All *ptaRNA1* homologs are located anti-sense to a putative toxin, which in turn is never encountered without the small RNA. The secondary structure of PtaRNA1, furthermore, is very similar to that of the FinP anti-sense RNA found on F-like plasmids in *Escherichia coli*.

*Key words:* small RNA, anti-sense RNA, antitoxin, plasmid

## 1 Introduction

Several toxin-antitoxin systems of type 1, in which the toxin is a short protein and the antitoxin an anti-sense RNA and of type 2, where both elements are proteins, are frequently found in both prokaryotic chromosomes and plasmids [1, 2, 3, 4]. The paradigmatic example for type 1 is the plasmid encoded *hok/sok* system in *Escherichia coli* and its close relatives. The toxin-encoding stable mRNA encodes a protein that rapidly leads to cell-death unless its translation is suppressed by a short-lived small RNA. The plasmid encoded module prevents the growth of plasmid-free offsprings thus ensuring the persistence of the plasmid in the population: After cell division, plasmid-free cells still contain the stable toxin mRNA, while the comparably unstable antitoxin is quickly depleted. It is poorly understood how the chromosomally encoded systems function. Interestingly, the SOS-induced genes *tisB* and *symE* are expressed under very specific stress conditions. The corresponding antitoxins (SymR and Sib) are constitutively expressed.

Although distinct toxin-antitoxin systems have been found in widely separated bacterial groups (e.g. *hok/sok* in *E. coli* and *txpA/ratA* in *Bacillus subtilis* [5]), each of the known examples exhibits a very narrow phylogenetic distribution.

In this contribution we characterize by computational means a small RNA that has all the hallmarks of the known type 1 toxin-antitoxin systems but shows a rather wide spread erratic phylogenetic distribution that hints at frequent horizontal gene transfers.

## 2 Results

The founding member, PtaRNA1 ("Plasmid transferred anti-sense RNA"), of the family was detected in a library of pyrosequencing data of *Xanthomonas campestris* pv. *vesicatoria* strain 85-10 (*Xcv*) that was prepared and analyzed for unrelated purposes. The superposition of the individual reads revealed a small RNA encoded adjacent to the *trbL* gene. Expression and approximate size of the small RNA was verified by Northern blot (Fig. 1). These analyses revealed a constitutive expression with respect to the tested growth phases. Interestingly, two bands which indicate procession of the full length PtaRNA1 are detected in the exponential but not in the stationary growth phase.

Chromosomally encoded homologs of *ptaRNA1* were found in beta-proteobacteria (*Nitrosomonas eutropha* C91, *Azoarcus* sp. EbN1, *Verminephrobacter eiseniae* EF01-2, *Burkholderia cenocepacia* J2315, *B. pseudomallei* K96243, *B. pseudomallei* 9, *B. pseudomallei* 91, and *Acidovorax* JS42) as well as gamma-proteobacteria (*X. campestris* pv. *vasculorum* NCPPB702, *Shigella flexneri* 2a 2457T, *Acinetobacter baumannii* ATCC 17978, *Marinobacter aquaeolei* VT8, and *Pseudomonas aeruginosa*

UCBPP-PA14). Two *ptaRNA1* copies were found in *N. eutropha* C91 and are named *ptaRNA1*–a and *ptaRNA1*–b. This observation is in accordance with various reported insertion, duplication and rearrangement events [6] within this species.

Conspicuously, *ptaRNA1* was not found in other closely related genomes, e.g. other strains of *Burkholderia*, *Pseudomonas*, or *Xanthomonas*. This distinguishes PtaRNA1 from most other bacterial small RNAs, such as the cyanobacteria-specific Yfr RNAs [7]. In addition to the chromosomal loci listed above, we found a *ptaRNA1* homolog in the *P. aeruginosa* plasmid pMATVIM-7, adjacent to the transfer region, a gene cluster that encodes proteins of unknown function and the plasmid stabilization protein ParE. Figure 2a depicts the alignment and the resulting consensus secondary structure of all detected PtaRNA1 homologs.

Phylogenetic analysis of the PtaRNA1 sequences (Fig. 3) shows that the phylogeny of the PtaRNA1 sequences is not congruent with the phylogeny of their "host" species. This indicates that the proliferation of *ptaRNA1* depends on frequent horizontal transfer, presumably by means of plasmids.

The *ptaRNA1* gene in *Xcv* is located anti-sense to a so far uncharacterized protein coding gene (*XCV2162*). The gene is adjacent to *trbL*, encoding a type IV secretion system protein. The small overlap of both genes strongly suggests that PtaRNA1 is an anti-sense regulator of *XCV2162* (Fig. 4). We therefore searched the complete set of eubacterial genomes for homologs of *XCV2162* and found that *ptaRNA1* and *XCV2162* co-occur in all cases, indicating their functional linkage. This is in particular also the case in the *P. aeruginosa* plasmid pMATVIM-7, whose uncharacterized gene *p07-406.22* is an ortholog of *XCV2162*, as in *Xcv* adjacent to *trbL* (Fig. 4). According to `MEMSAT3` prediction [8], XCV2162 contains a trans-membrane domain (Fig. 2b), as it is also the case for many reported toxic proteins [3, 4].

Furthermore, the gene phylogeny of the XCV2162 proteins (not shown) is congruent with the phylogeny of PtaRNA1 sequences, indicating that they are transferred together. We observed a frequent co-occurrence of *ptaRNA1*/*XCV2162* and *trbL* homologs, albeit *trbL* was detected in at least 155 eubacterial genomes, suggesting that *trbL* might have a role in the frequent chromosomal insertions of the *ptaRNA1*/*XCV2162* system. In *V. eiseniae* EF01-2 we found a truncated *XCV2162* homolog. *Verminephrobacter* is also the only *ptaRNA1* encoding species in which no *trbL* homolog was found.

Analysis of the putative *ptaRNA1* promoter regions revealed the existence of two highly conserved sequence motifs of eight nucleotides. The first one starts between 42 and 36 nt upstream and the other 13/12 nt upstream of the transcription start corresponding to the -35 and -10 box, respectively (Fig. 4). In the upstream region of *XCV2162*, an ultra conserved `AG`-rich motif was found (Fig. 4), probably representing the Shine-Dalgarno sequence of the mRNA. This motif is entirely covered by the complementary *ptaRNA1* sequence (Fig. 2a).

The consensus secondary structure of PtaRNA1 consists of a 5'-stem loop and a long 3' stem which presumably acts as terminator hairpin, (Fig. 2a). This structure is very similar to that of FinP found in *E. coli* (data not shown). Interestingly, FinP is the anti-sense regulator of TraJ, a transcriptional activator required for expression of various conjungative protein components [9]. Thus, FinP is only one component in a complex network of several interacting molecules.

## 3 Discussion

All evidence available for PtaRNA1 suggests that *XCV2162*/*ptaRNA1* is a novel toxin-antitoxin system: both genes are only found as combined cluster and in fact do not appear as single genes; *XCV2162* encodes a relatively short protein that shows the typical topology of toxins with a trans-membrane domain; the presence on a plasmid in combination with the erratic phylogenetic distribution of the system indicates a frequent horizontal gene transfer. It is known that type 2 systems, where both molecules are proteins, show an erratic phylogenetic distribution [4]. We assume that this might also be the case for type 1 systems, such as the one presented here.

Furthermore, the phylogenetic distribution of the *XCV2162*/*ptaRNA1* pair indicates a very rapid loss of the chromosomal homologs: the erratic distribution suggests that we only see very recent chromosomal insertions. The homolog in *Verminephrobacter* with its truncated *XCV2162* coding sequence might represent the first step towards the complete loss of the system.

This apparent evolutionary instability further supports the hypothesis that *XCV2162*/*ptaRNA1* is a toxin-antitoxin pair. Only chromosomal integration of the toxin-antitoxin pair makes the plasmid dispensable. Thus, cell-death is prevented by chromosomal integration of the system. Plasmid-loss and subsequent destruction of the toxin *XCV2162* then leaves the chromosomal copy of the antitoxin PtaRNA1 without function, so it is also rapidly removed from the genome.

## 4 Materials and Methods

For Northern blot analysis *Xanthomonas campestris* pv. *vesicatoria* strain 85-10 was cultivated at 30 °C in nutrient-yeast-glycerol medium [10]. Cells were harvested at exponential and stationary growth phase at $OD_{600}$ of 0.6 and 1.5, respectively. RNA was extracted as described in [11]. Northern blots were done according to [12] with the following modifications: 20 $\mu$g RNA were separated on 8.3 M urea - 6 % polyacrylamide gels. For detection of PtaRNA1 and 5S rRNA membranes were incubated for 1 h at 42 ° C with Rapid-hyb™ Buffer (GE Healthcare) containing $^{32}$P 5' end-labeled oligodeoxyribonucleotides NB39 (5'-`ATGGAGAGGTGAATCATGGC`-3') and NB5S (5'-`ATGACCTACTCTCGCATGGC`-3'), respectively.

Homology searches were based on scans of the bacterial NCBI genome database (`ftp://ftp.ncbi.nih.gov/genomes/Bacteria/` downloaded 06/12/2009) as well as the plasmid genome database (`http://www.genomics.ceh.ac.uk/plasmiddb/` downloaded 06/12/2009). Homologs of protein coding genes were searched using `tblastn` of the Blast package [13]. Since non-coding RNAs may vary in sequence but still fold into the same secondary structure a semi-global alignment implementation, `GotohScan` [14], was used to scan for *ptaRNA1* homologs. The microbial web Blast (`http://www.ncbi.nlm.`

nih.gov/sutils/genom_table.cgi) was used to search for additional homologs especially in unfinished genome projects.

Alignments were calculated with ClustalW [15] and locARNATE [16] for sequence structure alignments, respectively. The consensus structure model was calculated with RNAalifold [17].

Using, MEME [18] we analyzed the 100 nt upstream region of all homologous *ptaRNA1* loci. MEME searches for similarities among the given sequences and calculates descriptors for these motifs.

To search for known regulatory sites within the 100 nt upstream region the PRODORIC database was queried using the Virtual Footprint v3.0 web tool [19].

## 5 Submitted Data

This manuscript documents the seed alignment ptaRNA1.seed.stk. A short summary is available as Wikipedia Entry at http://en.wikipedia.org/wiki/User:SveFinBioInf/ptaRNA1


### Acknowledgments

This work was supported by a grant from the DFG (German Research Foundation, SPP 1258) to U.B. and P.F.S.


In order to keep the list of references at reasonable length, we had to give priority to reviews and recent publications.

### References


[1] Gerdes K, Wagner EGH. RNA antitoxin. Cur Op Microbiol 10, 2007:117–124.

[2] Weaver KE. Emerging plasmid-encoded antisense RNA regulated systems. Curr Opin Microbiol 10, 2007:110–116.

[3] Fozo EM, Hemm MR, Storz G. Small toxic proteins and the antisense RNAs that repress them. Microbiol Mol Biol Rev 72, 2008:579–589.

[4] Makarova KS, Wolf YI, Koonin EV. Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. Biol Direct 4, 2009:19.

[5] Silvaggi JM, Perkins JB, Losick R. Small untranslated RNA antitoxin in *Bacillus subtilis*. J Bacteriol 187, 2005:6641–6650.

[6] Stein LY, et al. Whole-genome analysis of the ammonia-oxidizing bacterium, *Nitrosomonas eutropha* C91: implications for niche adaptation. Environ Microbiol 9, 2007:2993–3007.

[7] Voss B, Gierga G, Axmann IM, Hess WR. A motif-based search in bacterial genomes identifies the ortholog of the small RNA Yfr1 in all lineages of cyanobacteria. BMC Genomics 8, 2007:375.

[8] Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics 10, 2009:159.

[9] Arthur DC, Ghetu AF, Gubbins MJ, Edwards RA, Frost LS, Glover JNM. FinO is an RNA chaperone that facilitates sense-antisense RNA interactions. EMBO J 22, 2003:6346–6355.

[10] Daniels MJ, Barber CE, Turner PC, Sawczyc MK, Byrde RJW, Fielding AH. Cloning of genes involved in pathogenicity of *Xanthomonas campestris* pv. campestris using the broad host range cosmid pLAFR1. EMBO J 3, 1984:3323–3328.

[11] RK Hartmann AS A Bindereif, Westhof E. *Handbook of RNA Biochemistry*. Wiley-VCH, 2005.

[12] Urban JH, Vogel J. Translational control and target recognition by Escherichia coli small RNAs *in vivo*. Nucleic Acids Res 35, 2007:1018–1037.

[13] Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 1997:3389–3402.

[14] Hertel J, et al. Non-coding RNA annotation of the genome of trichoplax adhaerens. Nucleic Acids Res 37, 2009:1602–1615.

[15] Larkin MA, et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2007:2947–2948.

[16] Otto W, Will S, Backofen R. Structure local multiple alignment of RNA. In *Proceedings of the German Conference on Bioinformatics (CGB 08)*, volume P-136 of *LNI*. GI, 2008, 178–188.

[17] Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 9, 2008:474.

[18] Bailey TL, Elkan C. The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 3, 1995:21–29.

[19] Münch R, et al. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. Bioinformatics 21, 2005:4187–4189.

Figure 1:
Expression of PtaRNA1and 5S RNA in exponential and stationary growth phase of *Xcv* analysed by Northern blot. The size of corresponding marker bands is indicated on the left.

Figure 2:
a) Consensus secondary structure model of PtaRNA1 based on the depicted seed alignment. The structure is highly stable (minimum free energy $-37.06 \frac{kcal}{mol}$) and supported by various compensatory mutations within the stem on the right-hand side. Marked in blue is the region complementary to the putative Shine-Dalgarno sequence of the *XCV2162* mRNA.
b) Amino acid alignment of XCV2162 homologs. The alignment shows various totally (indicated by '$*$') and by substitutions (indicated by ':' and '.') supported and therefore conserved columns. The protein topology of a trans-membrane domain, predicted by MEMSAT3 [8] is indicated as well. '+' marks inside loop, '$\sim$' outside loop, 'O' outside helix cap, 'X' central trans-membrane helix segment and 'I' inside cap. The truncated *Verminephrobacter* sequence was not used for the calculation of the conservation track.

Figure 3:
Phylogenetic tree based on PtaRNA1 alignment (similar for XCV2162 alignment, data not shown). Class of the "host" species is shown by the symbols on the right hand side. Numbers indicate bootstrap values of the inner nodes.

Figure 4:
Surrounding genomic location of the *ptaRNA1* gene in *Xcv*. On the plus strand the coding sequences of *trbL* and *XCV2162* are indicated in blue. In front of *XCV2162* an ultra conserved AG rich motif, the putative Shine-Dalgarno sequence is shown. *ptaRNA1* is encoded on the minus strand and indicated in green. A conserved -10 as well as -35 box (sequence logos) was found directly upstream of this gene.

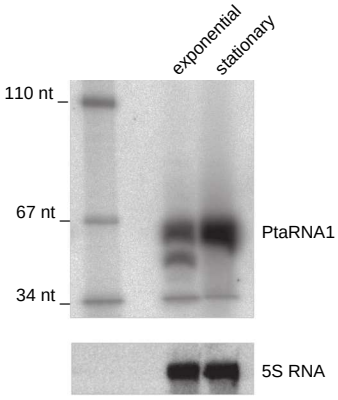


Figure 1:

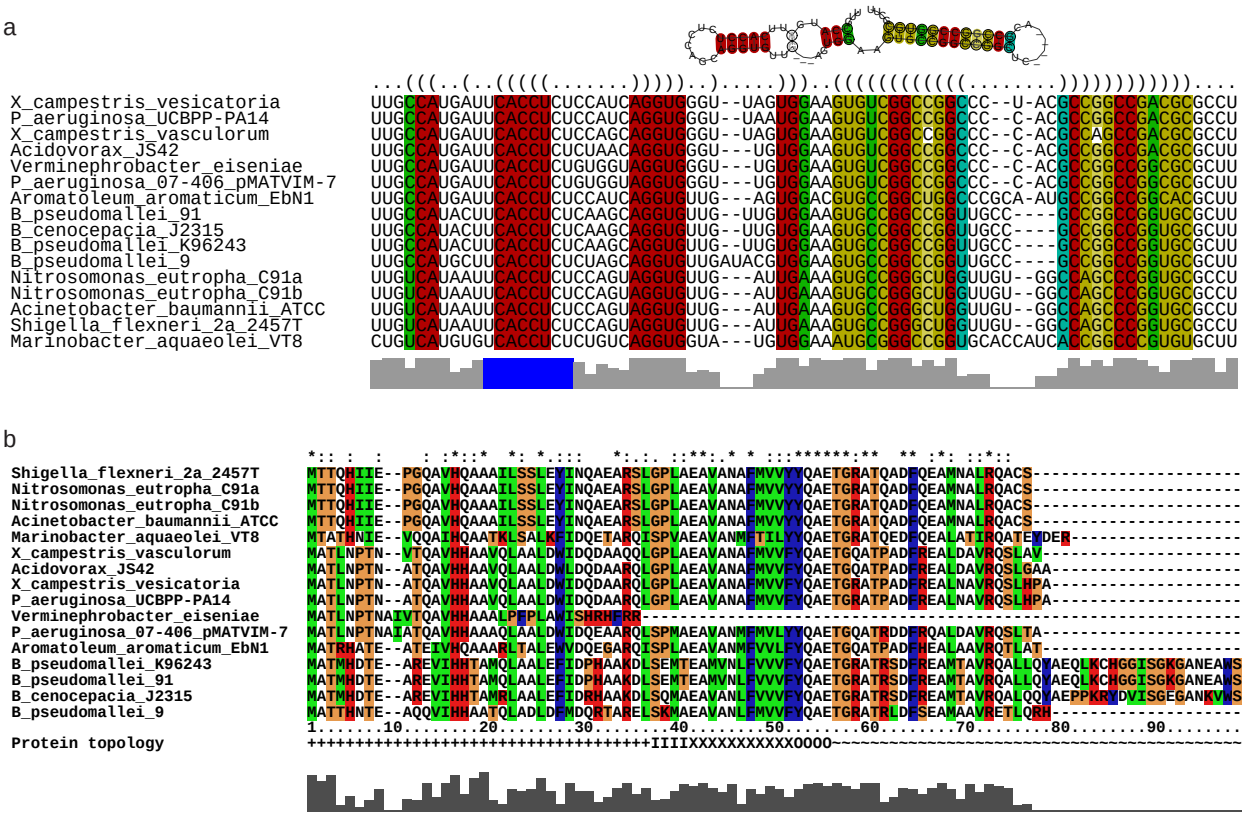
a
...(((..(..(((((.......)))))..)......))).((((((((((((.........))))))))))))....
X_campestris_vesicatoria
P_aeruginosa_UCBPP-PA14
X_campestris_vasculorum
Acidovorax_JS42
Verminephrobacter_eiseniae
P_aeruginosa_07-406_pMATVIM-7
Aromatoleum_aromaticum_EbN1
B_pseudomallei_91
B_cenocepacia_J2315
B_pseudomallei_K96243
B_pseudomallei_9
Nitrosomonas_eutropha_C91a
Nitrosomonas_eutropha_C91b
Acinetobacter_baumannii_ATCC
Shigella_flexneri_2a_2457T
Marinobacter_aquaeolei_VT8
b
Shigella_flexneri_2a_2457T
Nitrosomonas_eutropha_C91a
Nitrosomonas_eutropha_C91b
Acinetobacter_baumannii_ATCC
Marinobacter_aquaeolei_VT8
X_campestris_vasculorum
Acidovorax_JS42
X_campestris_vesicatoria
P_aeruginosa_UCBPP-PA14
Verminephrobacter_eiseniae
P_aeruginosa_07-406_pMATVIM-7
Aromatoleum_aromaticum_EbN1
B_pseudomallei_K96243
B_pseudomallei_91
B_cenocepacia_J2315
B_pseudomallei_9
Protein topology

Figure 2:

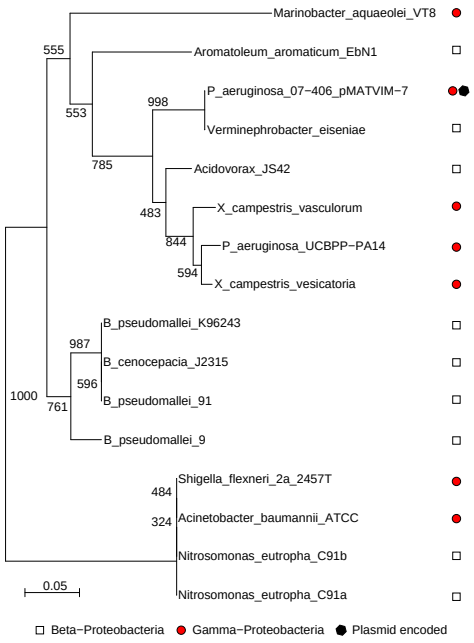
Marinobacter_aquaeolei_VT8
Aromatoleum_aromaticum_EbN1
P_aeruginosa_07−406_pMATVIM−7
Verminephrobacter_eiseniae
Acidovorax_JS42
X_campestris_vasculorum
P_aeruginosa_UCBPP−PA14
X_campestris_vesicatoria
B_pseudomallei_K96243
B_cenocepacia_J2315
B_pseudomallei_91
B_pseudomallei_9
Shigella_flexneri_2a_2457T
Acinetobacter_baumannii_ATCC
Nitrosomonas_eutropha_C91b
Nitrosomonas_eutropha_C91a
555
553
998
785
483
844
594
987
596
1000
761
484
324
0.05
Beta−Proteobacteria
Gamma−Proteobacteria
Plasmid encoded

Figure 3:

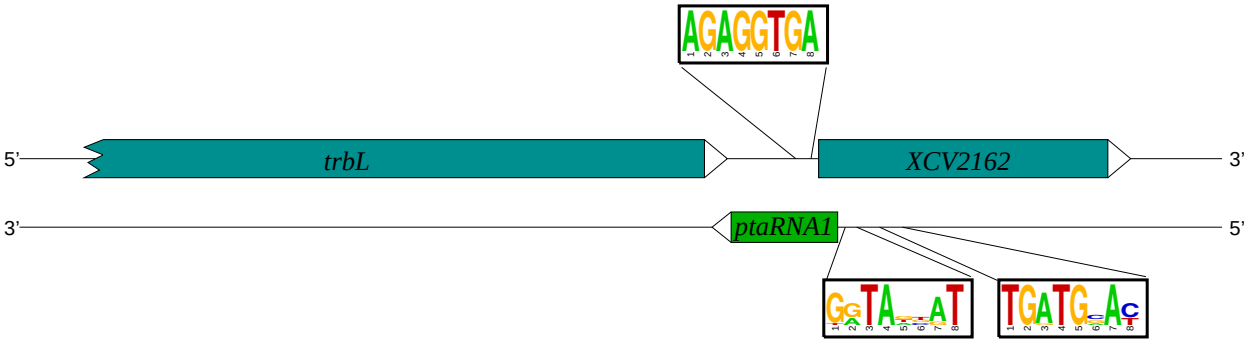


Figure 4: