genes

Article

The primary antisense transcriptome of *Halobacterium salinarum* NRC-1

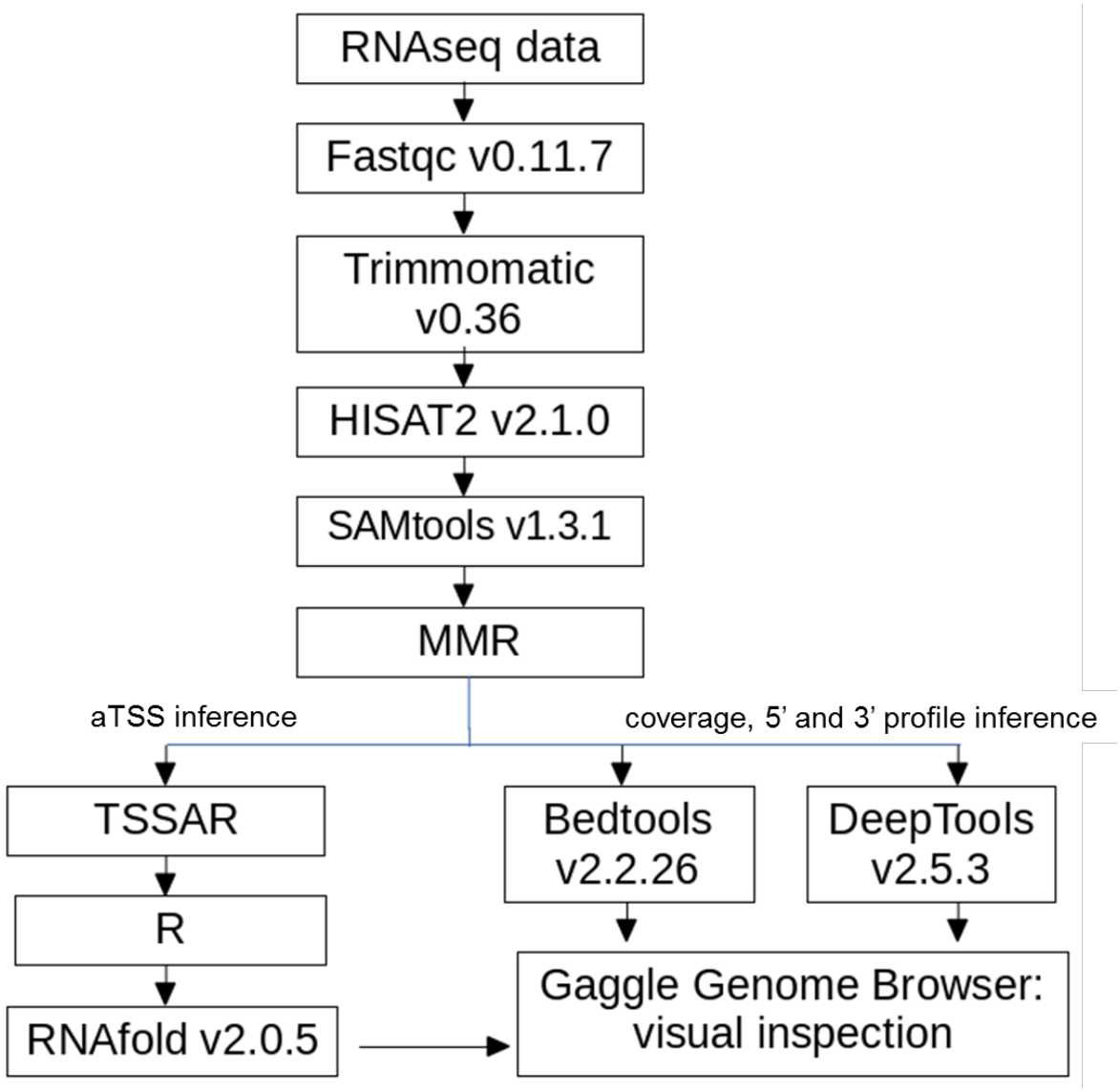
João Paulo Pereira de Almeida [1,$], Ricardo Z. N. Vêncio [2,$], Alan P. R. Lorenzetti [1], Felipe ten-Caten [1], José Vicente Gomes-Filho [1] and Tie Koide [1,*]

[1] Department of Biochemistry and Immunology, Ribeirão Preto Medical School, Universidade de São Paulo, Brazil
[2] Department of Computation and Mathematics, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Brazil
$ authors contributed equally
* Correspondence: tkoide@fmrp.usp.br; Tel.: +55-16-3315-3107

Supplemental Figures

16

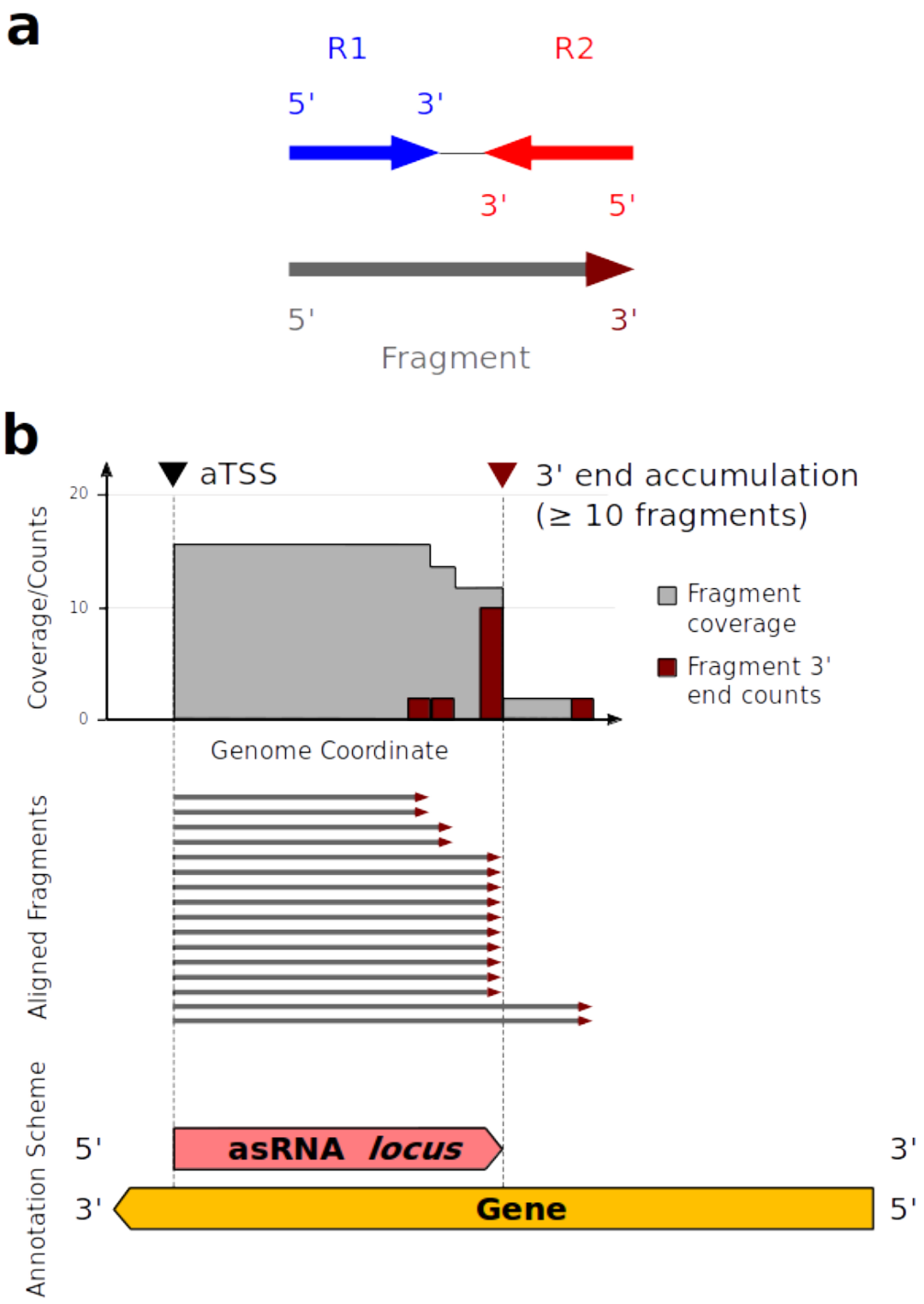
17



19**Figure S1.** Pipeline for asRNA annotation.

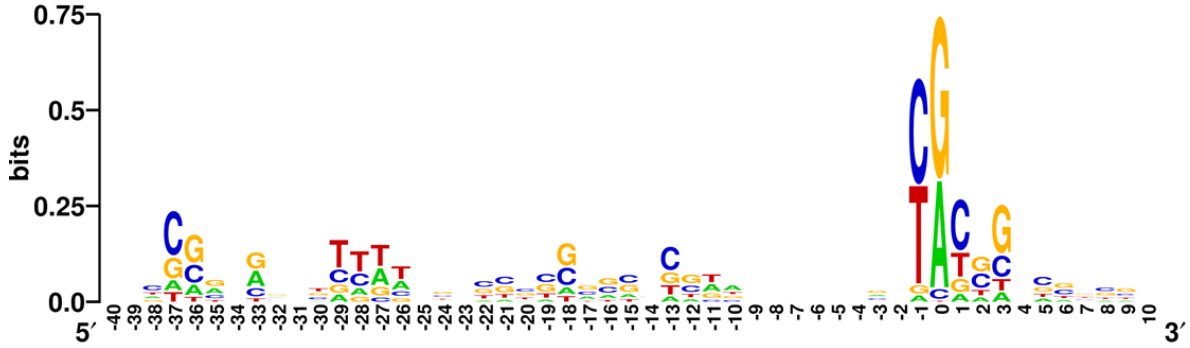
20

21



23**Figure S2**. asRNA *loci* inference approach. (a) Reads aligned as a pair (R1-R2; blue and red arrows, 24respectively) are a representation of an existent RNA fragment (grey line with brown arrowhead), 25since our RNA library preparation protocol ligates adapters directly to 5′ and 3′ ends of RNA 26molecules. Furthermore, we can extrapolate that orphan R2 reads aligned using single-end mode 27also have implicit information about the ending position of a fragment. (b) In our approach, the 28minimum size of an asRNA locus (pink arrow) can be inferred by mapping its beginning (black 29triangle), given by an aTSS position, and its minimum ending position (brown triangle), given by 30the accumulation of at least 10 observations of fragment 3′ ends.

36(a)



39(b)

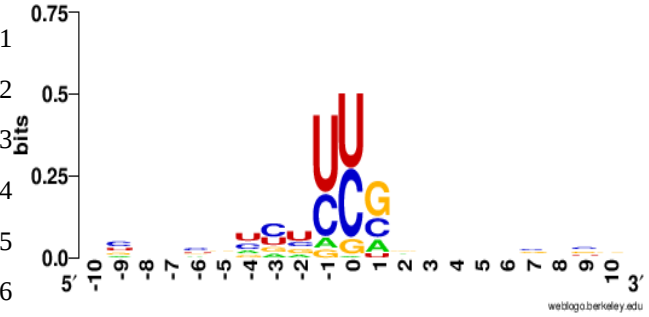
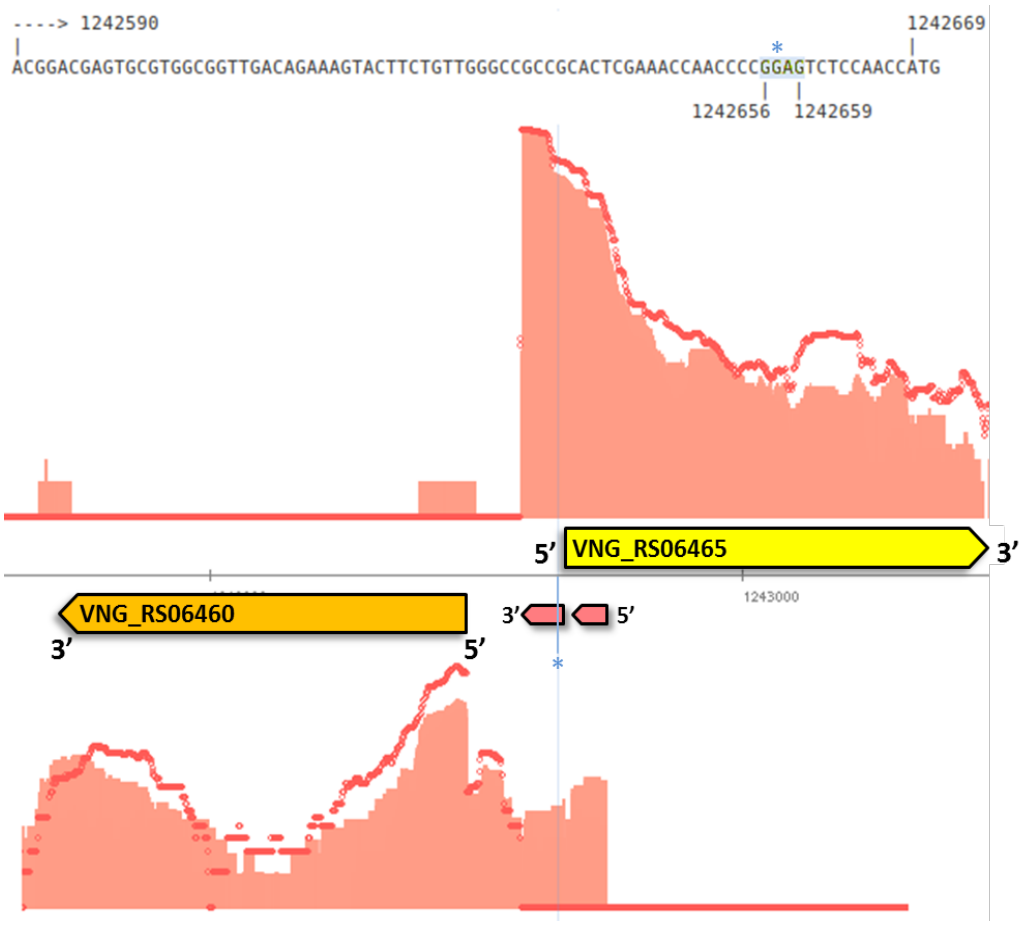


Figure S3. asRNA properties. (a) Promoter analysis: frequency of nucleotides 40 nt upstream and 10 nt downstream of aTSS (position 0). (b) Frequency of nucleotides at the 3′ end of asRNAs. Position 0 is the last nucleotide.

(a)


----> 1242590
1242669
ACGGACGAGTGCGTGGCGGTTGACAGAAAGTACTTCTGTTGGGCCGCCGCACTCGAAACCAACCCCGGAGTCTCCAACCATG
1242656
1242659
5′
VNG_RS06465
3′
1243000
VNG_RS06460
3′
3′
5′
5′

(b)

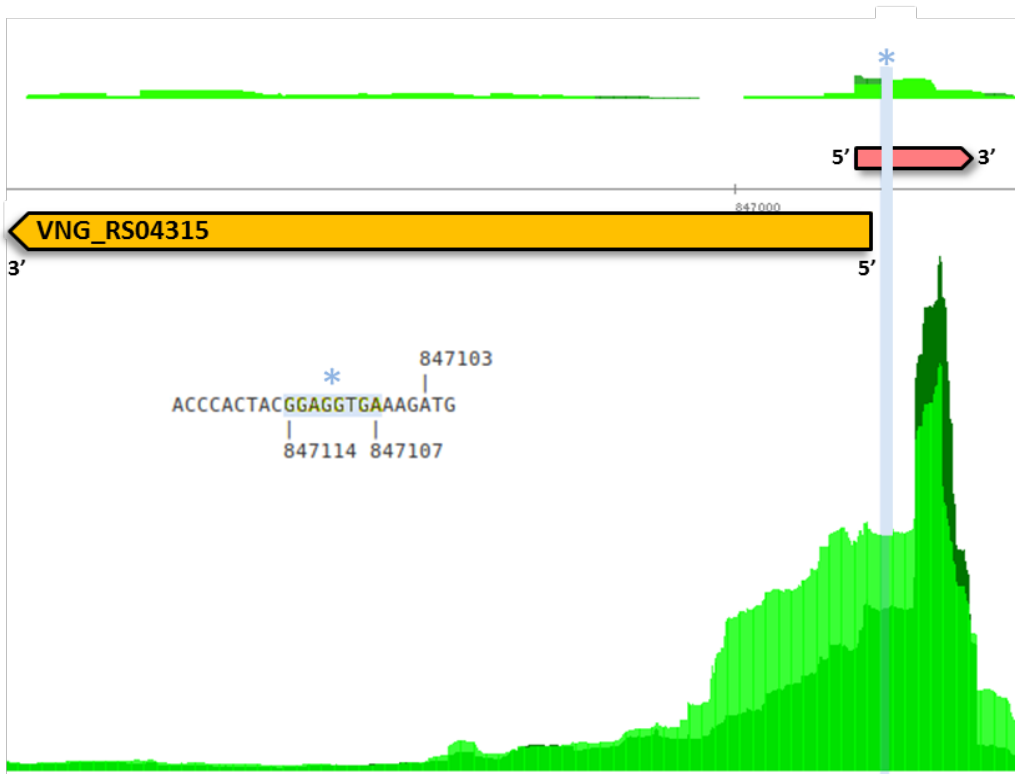
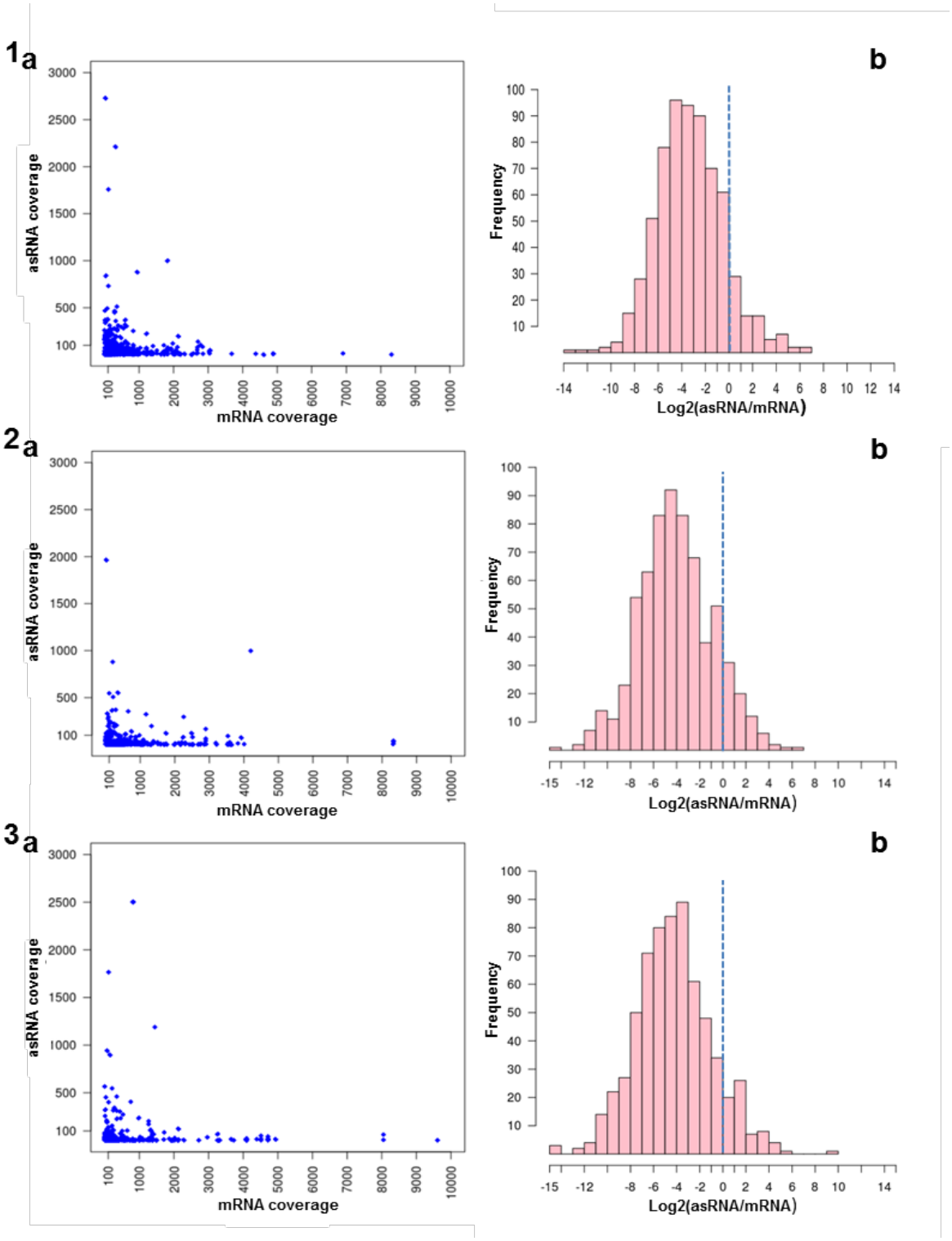

5′
3′
847000
VNG_RS04315
3′
5′
847103
ACCCACTACGGAGGTGAAAGATG
847114
847107

Figure S4. Examples of putative RBS occlusion by an asRNA. (a) *cdcH*, encoding an AAA-type ATPase (VNG_RS06465). VNG_RS06465 *locus* (yellow arrow) is in forward strand (5′→3′ left to right), neighbor gene VNG_RS06460 (orange arrow) is in reverse strand (5′→3′ right to left). Aligned reads coverage along genomic coordinates for TEX+ libraries at exponential and stationary phases are shown in light red (solid) and red (dots), respectively (log_2 counts jointly normalized and arbitrarily scaled). VNG_as06465_925 asRNA (pink arrow) encompasses Shine-Dalgarno-like signature (* light blue highlight). (b) *rpl1* encoding the 50S ribosomal protein L1. VNG_RS04315 *locus* (orange arrow) in reverse strand (5′→3′ right to left). dRNA-seq read coverage signal is shown in dark and light green for TEX+ and TEX- libraries, respectively. VNG_as04315_654 asRNA (pink arrow) encompasses Shine-Dalgarno-like signature (* light blue highlight in genome coordinates and zoomed in sequence).



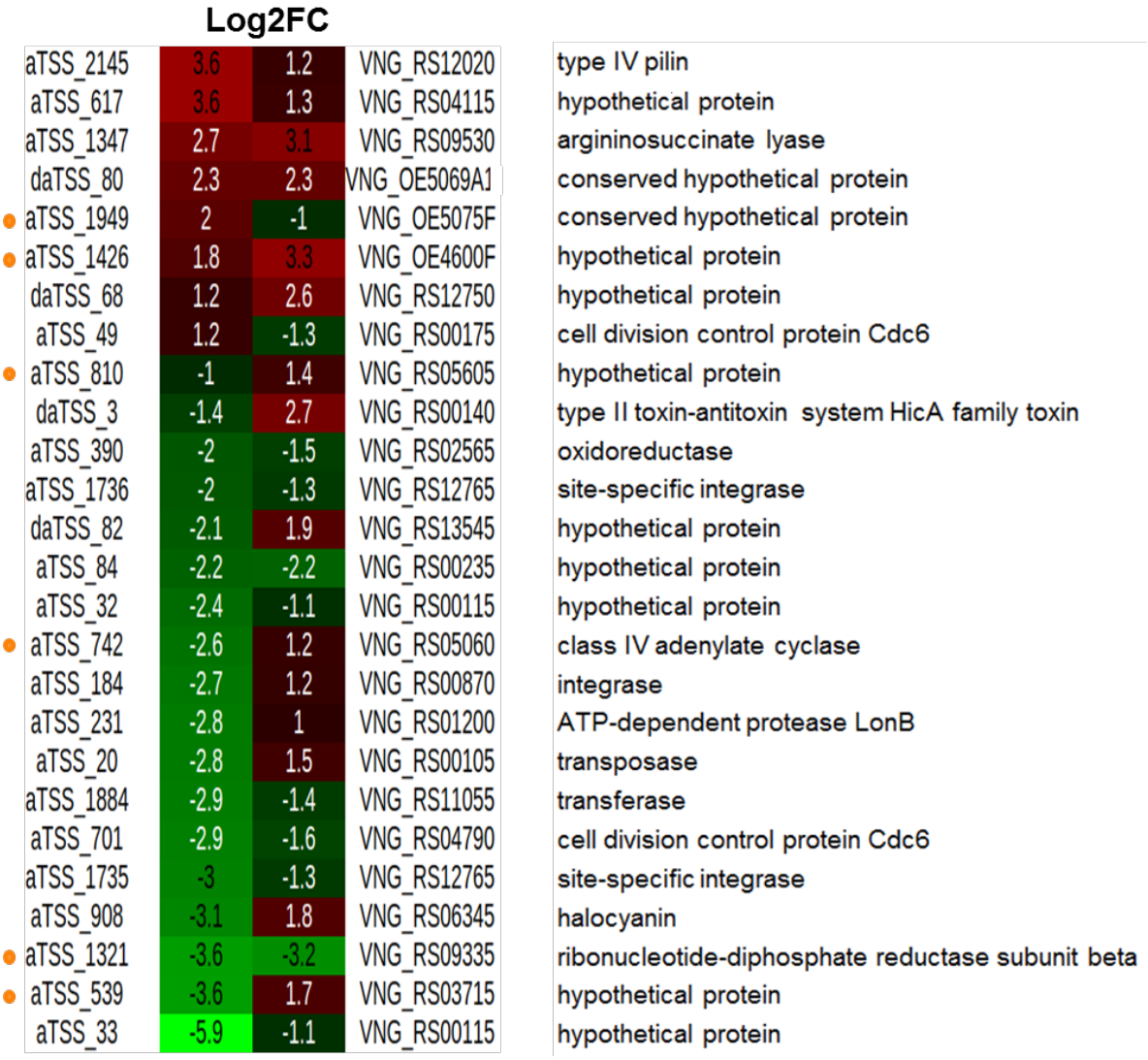
73 **Figure S5.** Coverage of asRNAs relative to mRNAs on the opposite strand. (a) Coverage of mRNAs
74 (x-axis) and coverage of asRNAs (y-axis). (b) Histogram of $\log_2$ fold change of asRNA coverage
75 relative to mRNA on the opposite strand; vertical dotted lines mark 1:1 expression levels. (1)
76 stationary phase -17 h, (2) beginning of exponential phase -37 h, (3) gas vesicles release phase – 86 h.

77

78

79(a)

80



82

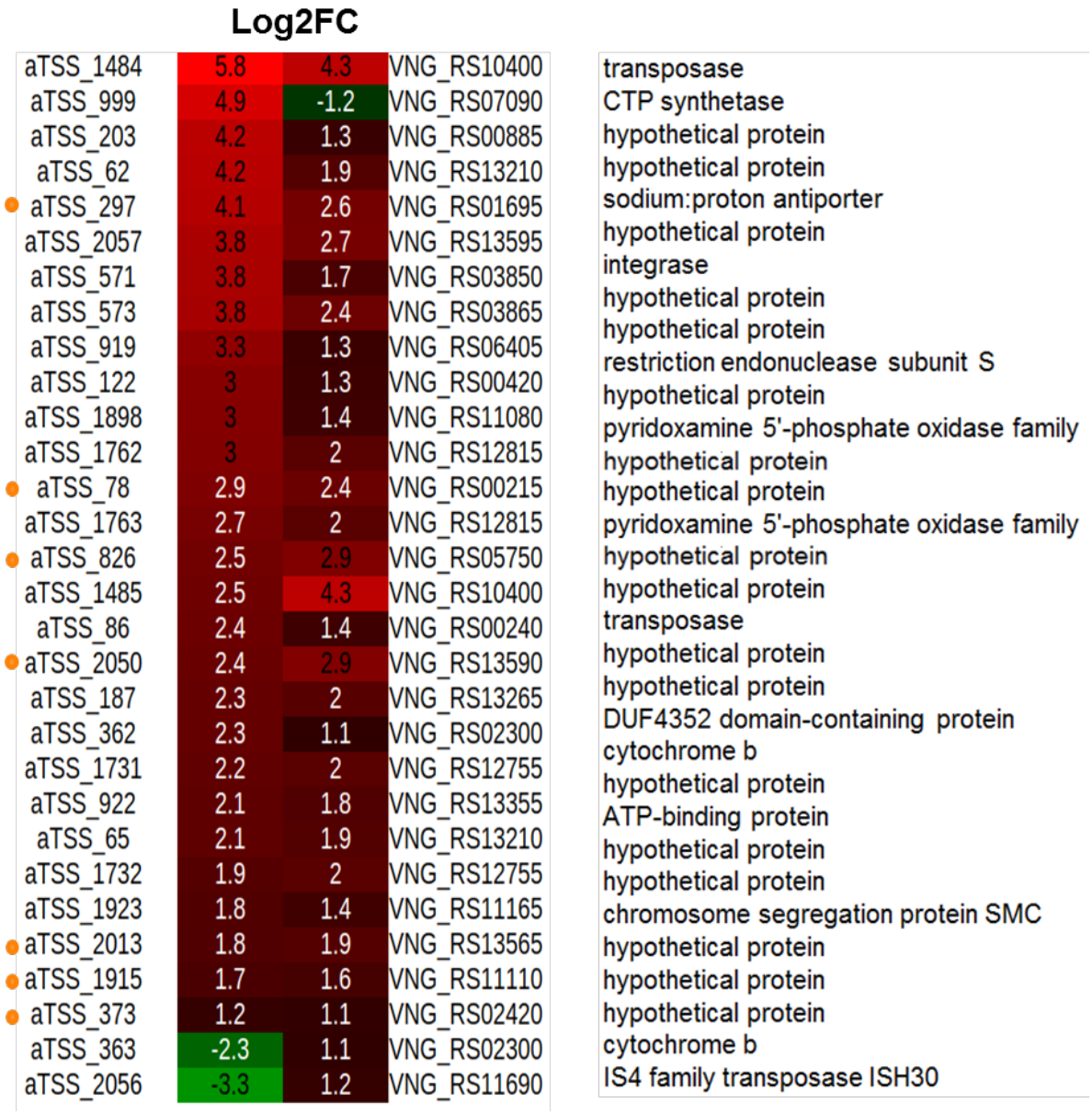
83

84

85

86(b)

87

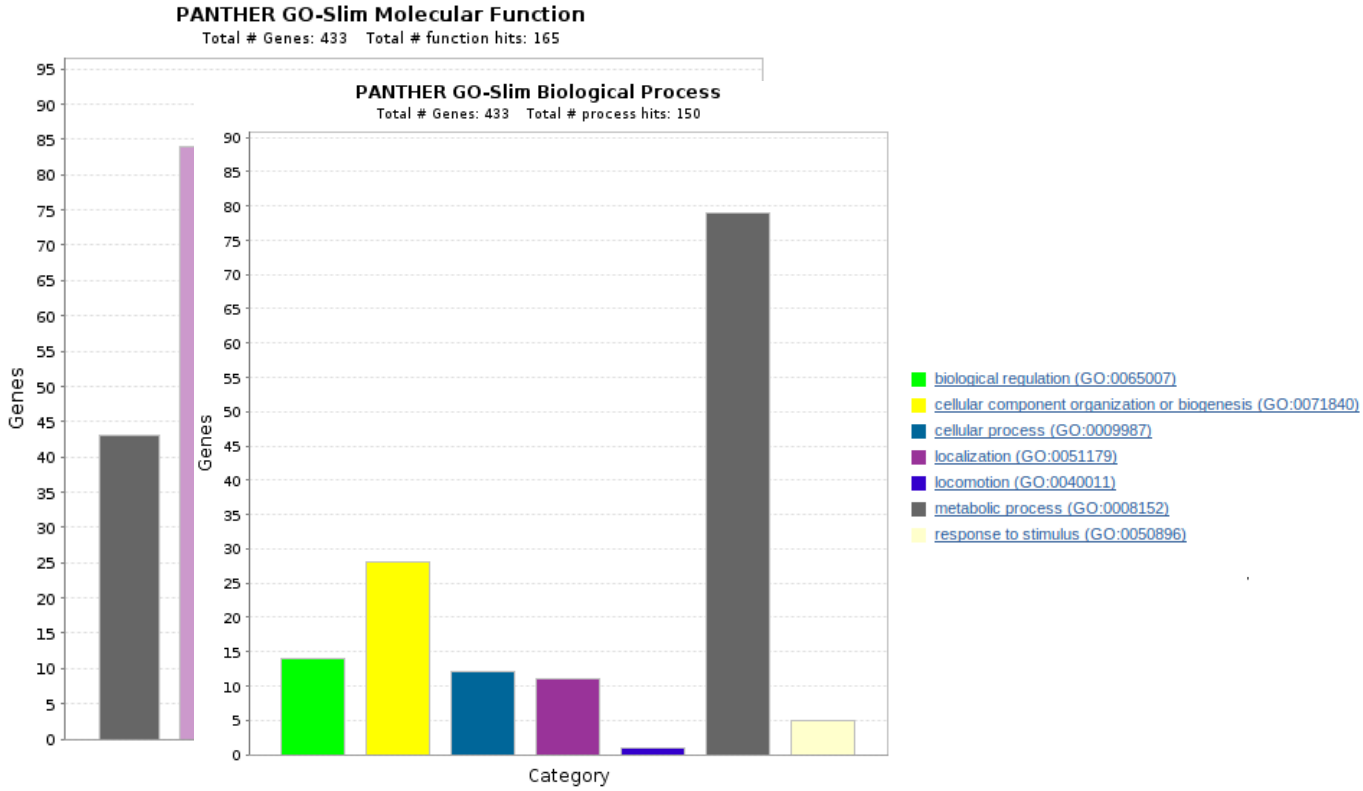


89

90**Figure S6.** Differential expression of asRNAs and mRNAs on the opposite strand. Pairs asRNA/mRNA that 91presented either positive or negative correlation are shown. Up-regulated genes have its log fold change 92(log_2FC) inside red cells and down-regulated genes inside green cells. (a) log_2FC for 37 h/17 h (b) log_2FC for 86 93h/37 h. Orange dots highlight asRNAs overlapping the 5′ UTR of the mRNA.

94

95



97

98

99

100

101

102

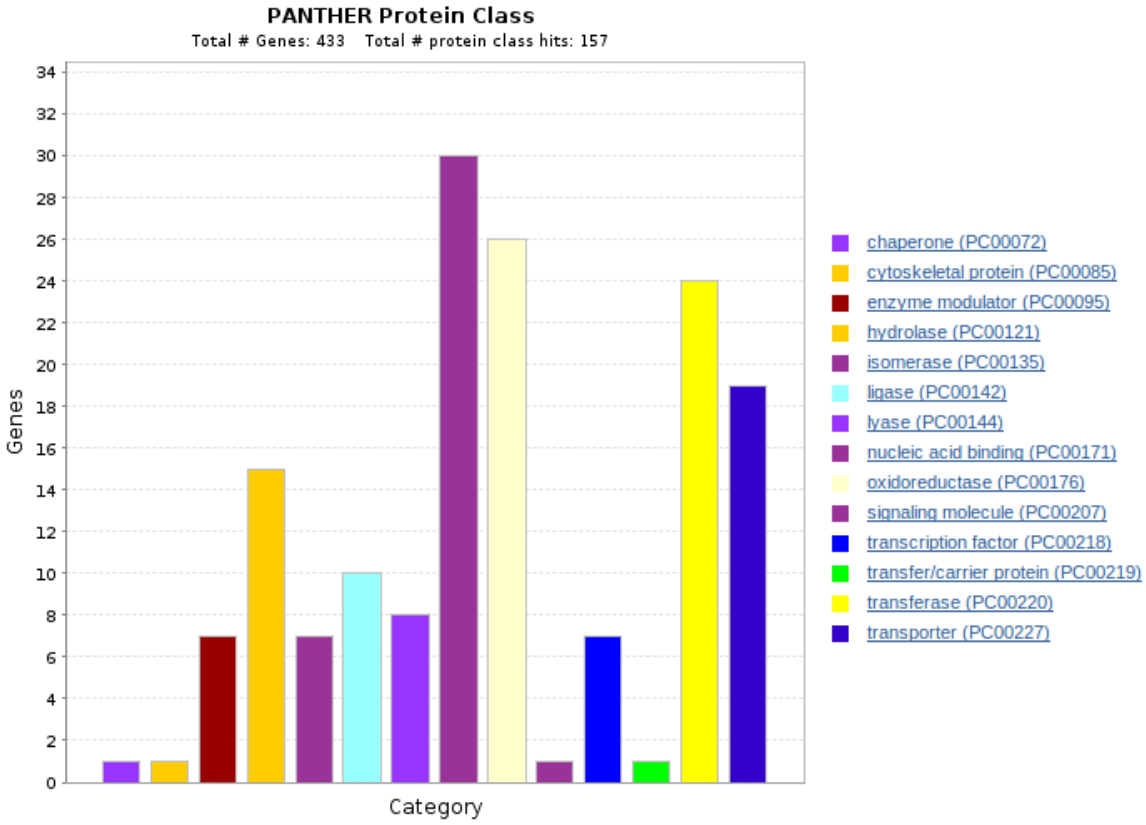


Figure S7. Functional categorization of genes presenting asRNAs according to Gene Ontology (GO). PANTHER tool was able to classify 433 genes out of the 613 submitted. Panels show molecular function, biological processes and protein classes, respectively.

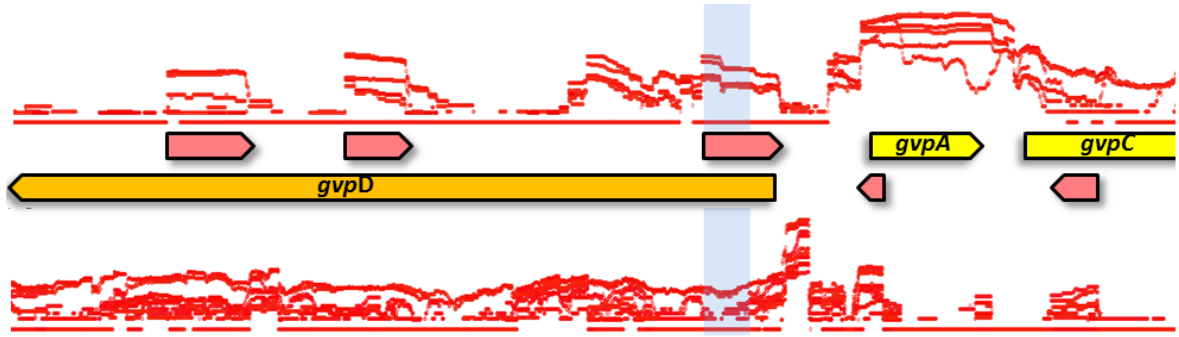


Figure S8. asRNAs in *gvp* genes. Upper panel shows RNA-seq $\log_2$ counts for forward strand, and lower panel shows signal for the reverse strand. All sequenced libraries are shown simultaneously and normalized (profiles as red dots). Genes *gvpA* and *gvpC* in forward strand (yellow arrows) and *gvpD* in reverse strand (orange arrow). Annotated asRNAs are represented by pink arrows. asRNAs related to *gvp* genes showed mean expression levels lower than their cognate genes. Region corresponding to the probe used for asRNA detection by Krüger and Pfeifer (1996) is highlighted in light blue.

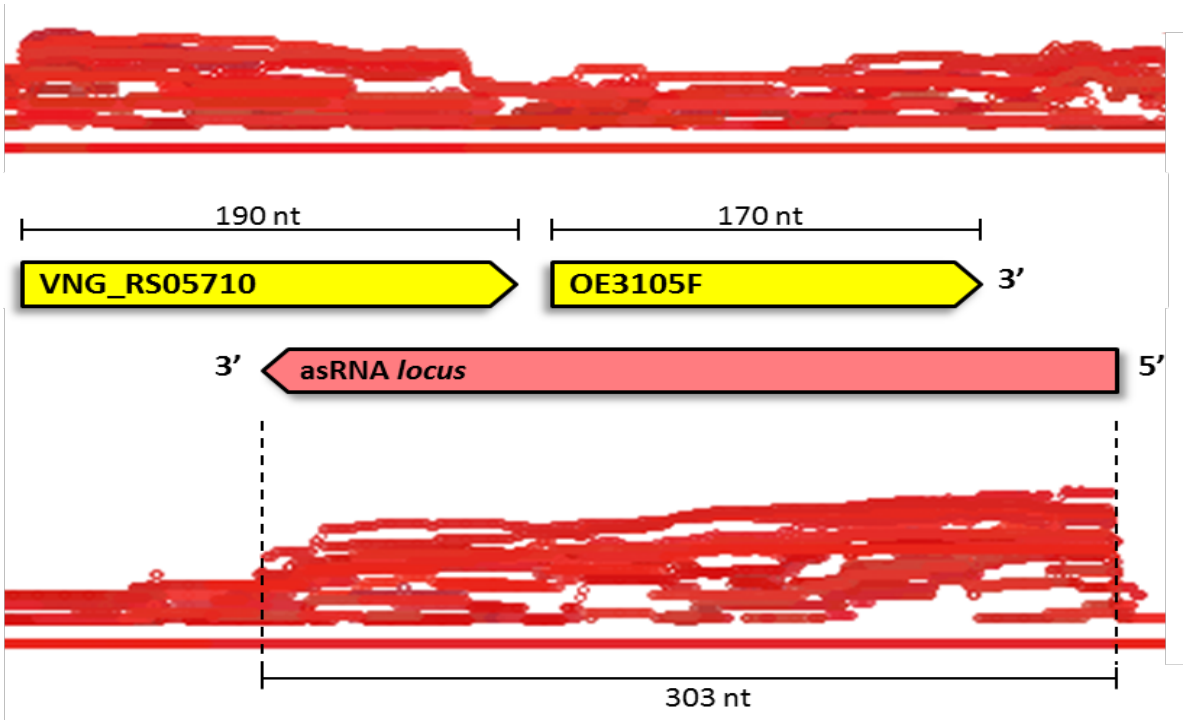


Figure S9. asRNAs in *brb* and *brz* genes. Upper panel shows RNA-seq $\log_2$ counts for forward strand, and lower panel shows signal for the reverse strand. All sequenced libraries are shown simultaneously and normalized (profiles as red dots). *brz* (VNG_RS05710) and *brb* (OE3105F) are on the forward strand (yellow arrows) and annotated asRNA (VNG_da3105F_36) on the reverse strand (pink arrow).

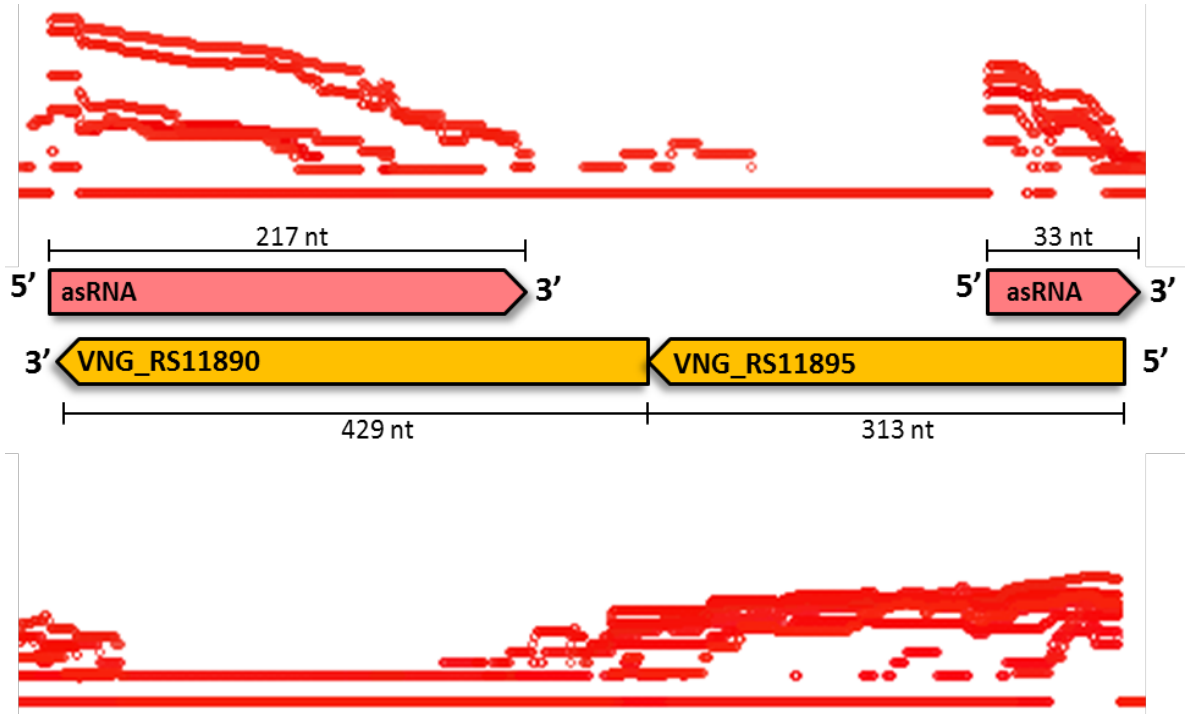
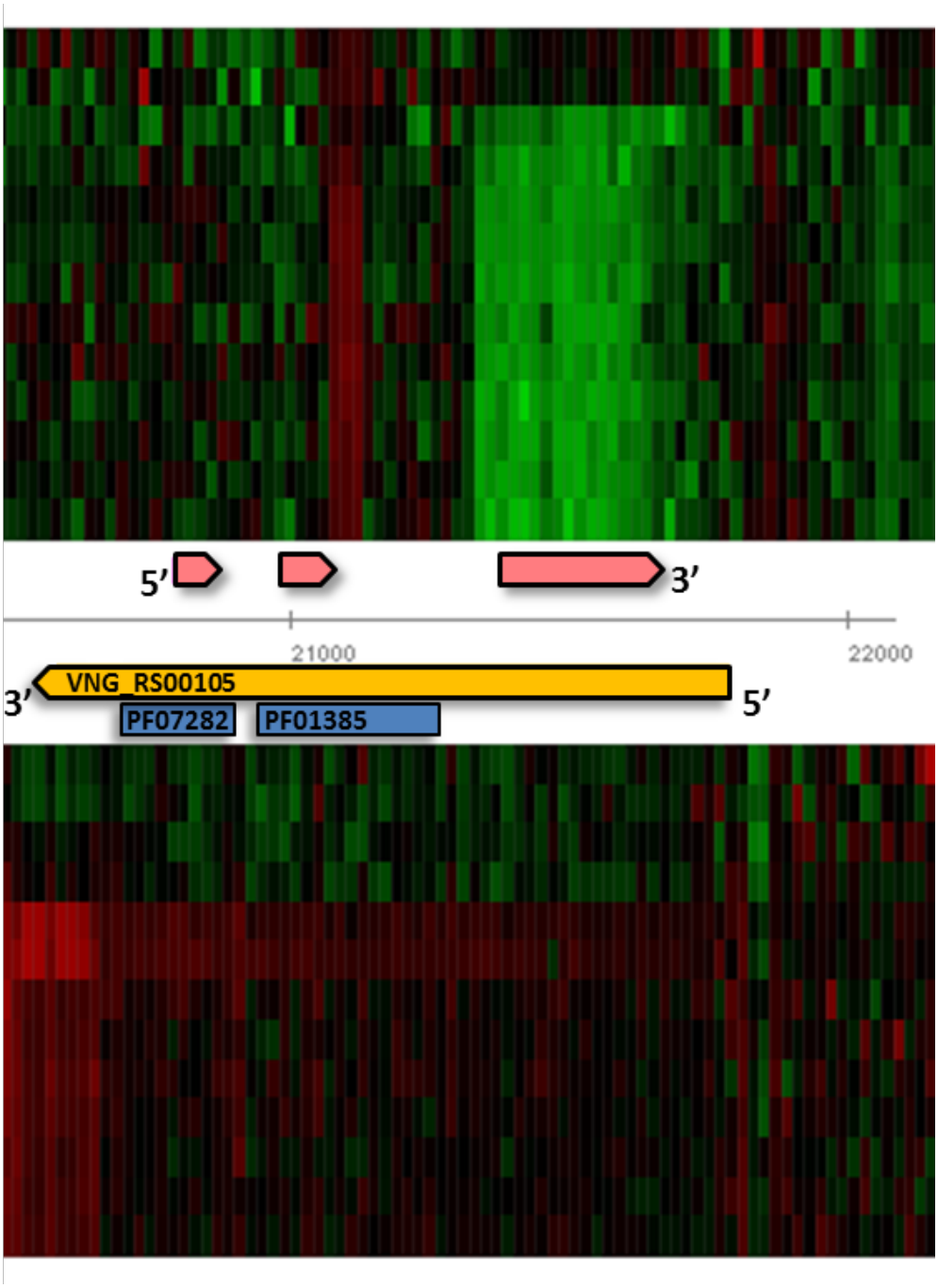


Figure S10. asRNAs in type II TA system. Upper panel shows RNA-seq $\log_2$ counts for forward strand, and lower panel shows signal for the reverse strand. All sequenced libraries are shown simultaneously and normalized (profiles as red dots). Genes VNG_RS11890 (toxin) and VNG_RS11895 (antitoxin) are on the reverse strand (orange arrows) and their respective annotated asRNAs VNG_da11890_90 and VNG_as11895_2118 on the forward strand (pink arrows).



160

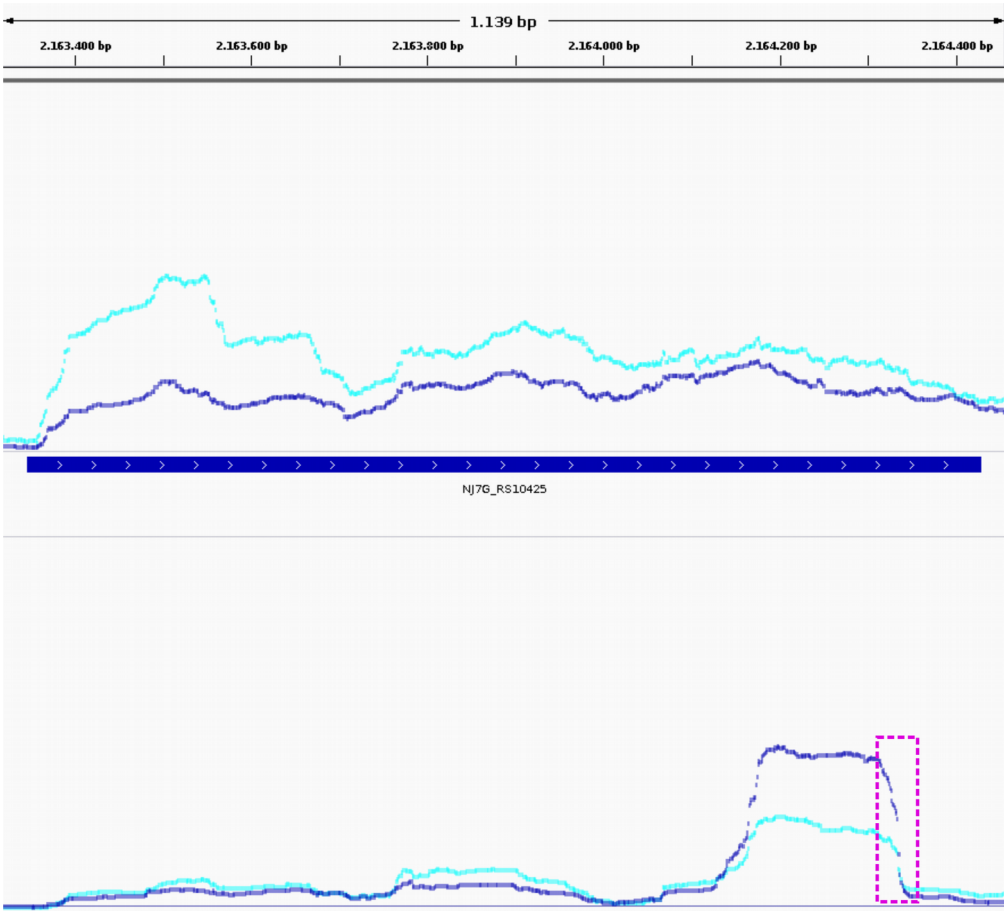
161 **Figure S11.** asRNAs antisense to a transposase differentially expressed along the growth curve.
162 VNG_RS00105 (orange arrow) encodes a transposase on the reverse strand (5′→ 3′ right to left),
163 PFAM domains are shown in blue. asRNAs are shown in pink arrows on forward strand (5′→ 3′ left
164 to right). Upper and lower panels show tiling array expression data over a growth curve (Koide *et*
165 *al.*, 2009) for forward and reverse strand, respectively. Each line represents one point in the growth
166 curve. Heat map is color coded: green shows down-regulation and red up-regulation relative to
167 reference condition.

168

169

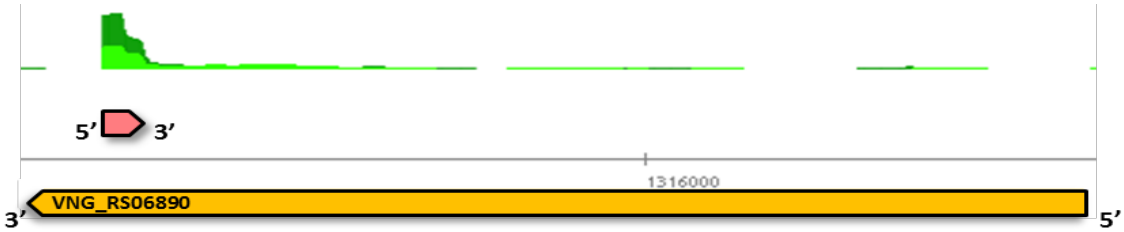
170(**a**)

171



173(**b**)

174



176 **Figure S12.** Differential expression of putative asRNAs antisense to the *nirH* gene in *Natrinema* sp.
177 J7-2. (a) Genomic region showing normalized read coverage for low (15% NaCl; light blue) and high
178 (30% NaCl, dark blue) salt concentrations, respectively. The y-axes range from 0 to 500 counts. CDS
179 in *Natrinema* sp. J7-2 (blue rectangle) is shown at forward strand (5′→3′ left to right). Signals above
180 and below the blue rectangle are for forward and reverse strands, respectively. A highlight box
181 (magenta) indicates the approximate asRNA *locus* position identified in *H. salinarum.* (b) *H.*
182 *salinarum* dRNA-seq read coverage signal is shown in dark and light green for TEX+ and TEX-
183 libraries, respectively (5′→3′ right to left). AsRNA *locus* position is highlighted (magenta). Coverage
184 signal above the axis is for forward strand.

185

186

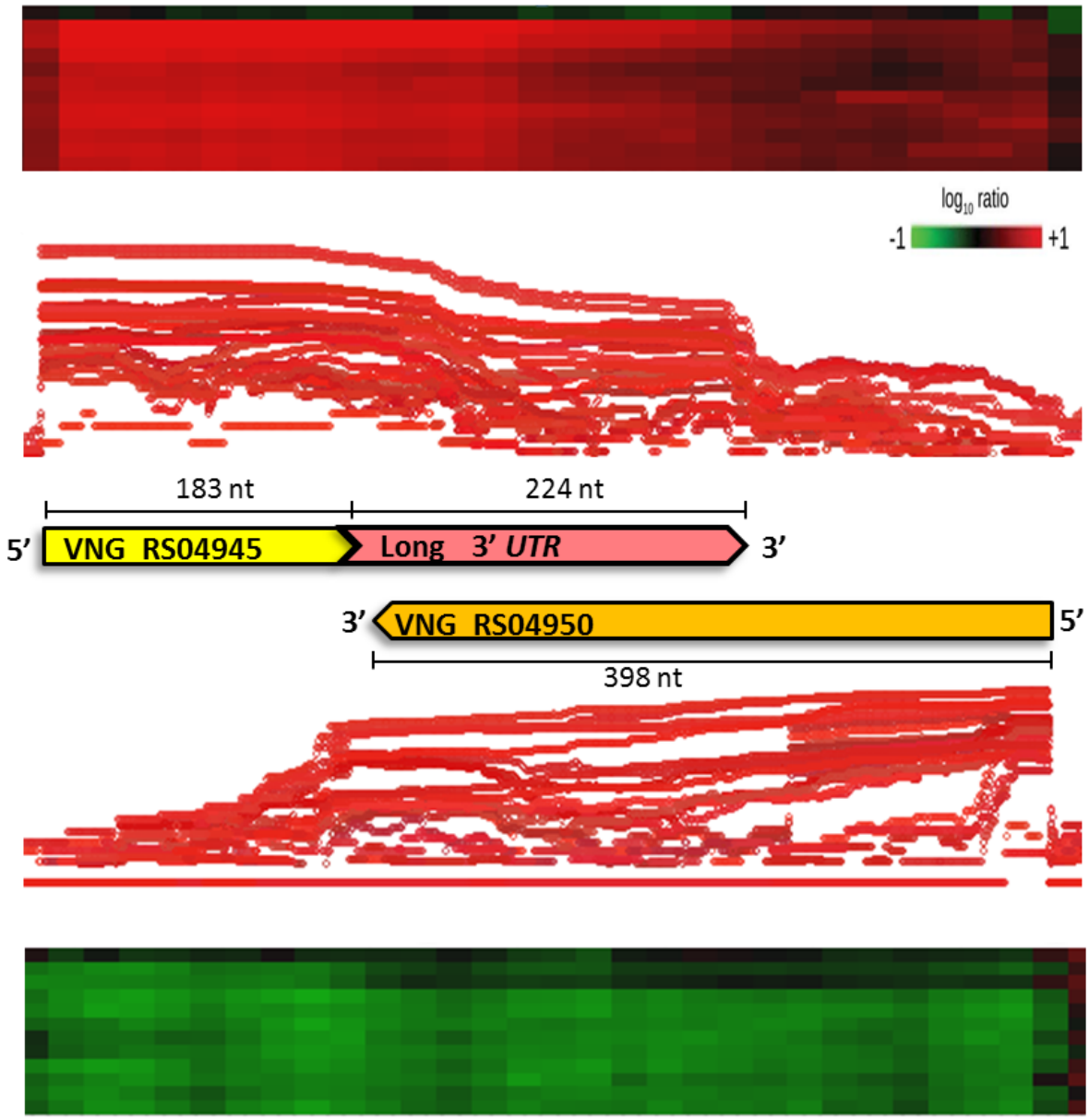


Figure S13. Long 3′ UTR of a coding sequence acting as an asRNA. Pink arrow indicates a possible long 3′ UTR of the gene VNG_RS04945 (yellow arrow, annotated as hypothetical protein) which could act as an asRNA of the gene VNG_RS04950 (orange arrow, annotated as a subunit of a translation initiation factor). Upper panel shows tilling array expression data for the forward strand and lower panel for the reverse (Koide *et al.* 2009). Red lines show RNA-seq signal (log_2 counts per position).

© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).