

1 Article

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

4 João Paulo Pereira de Almeida ^{1,\$}, Ricardo Z. N. Vêncio ^{2,\$}, Alan P. R. Lorenzetti ¹, Felipe ten-
5 Caten ¹, José Vicente Gomes-Filho ¹ and Tie Koide ^{1,*}

6 ¹ Department of Biochemistry and Immunology, Ribeirão Preto Medical School, Universidade de São Paulo,
7 Brazil

8 ² Department of Computation and Mathematics, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto,
9 Universidade de São Paulo, Brazil

10 \$ authors contributed equally

11 * Correspondence: tkoide@fmrp.usp.br; Tel.: +55-16-3315-3107

12

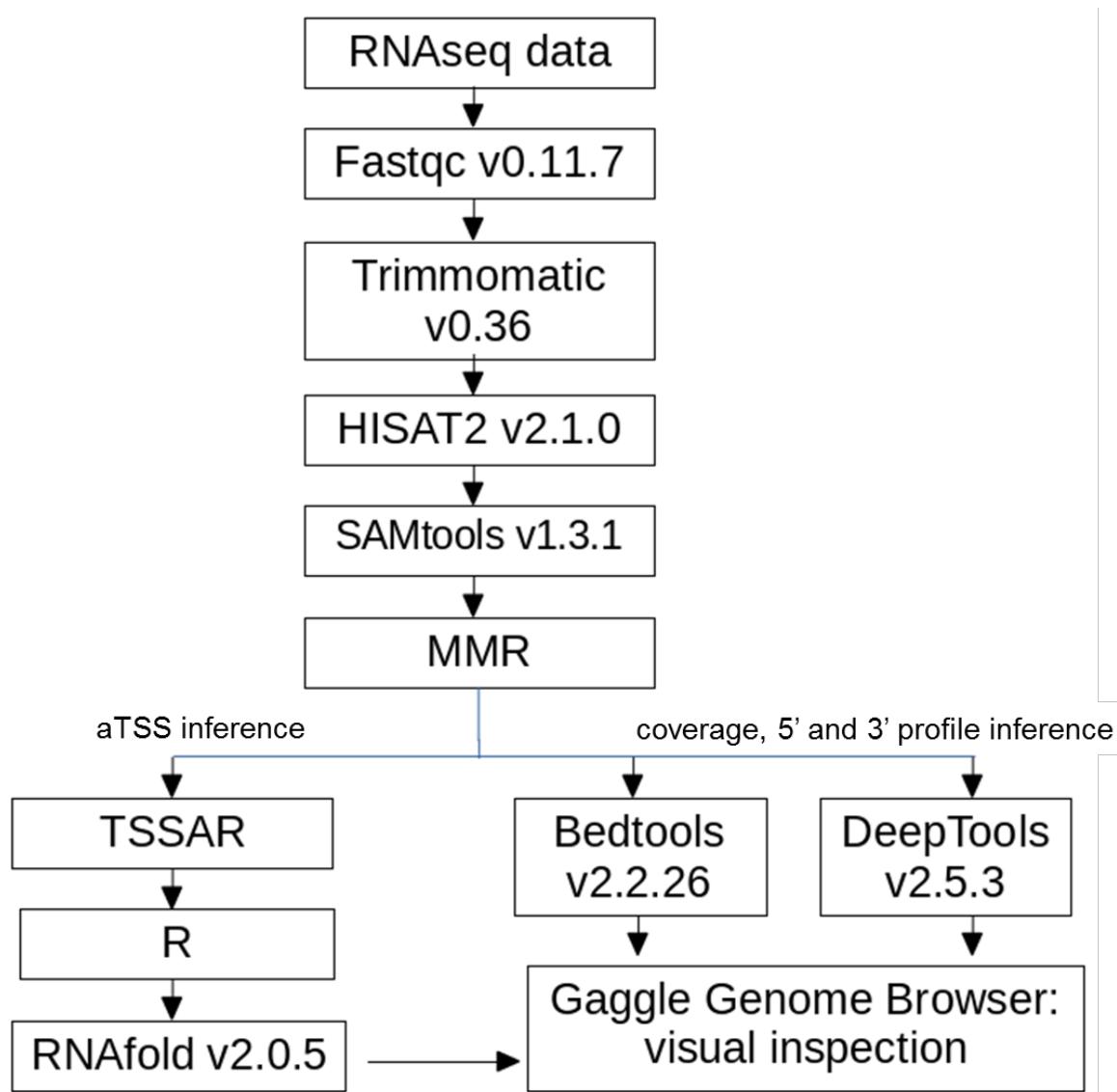
13

14 **Supplemental Figures**

15

16

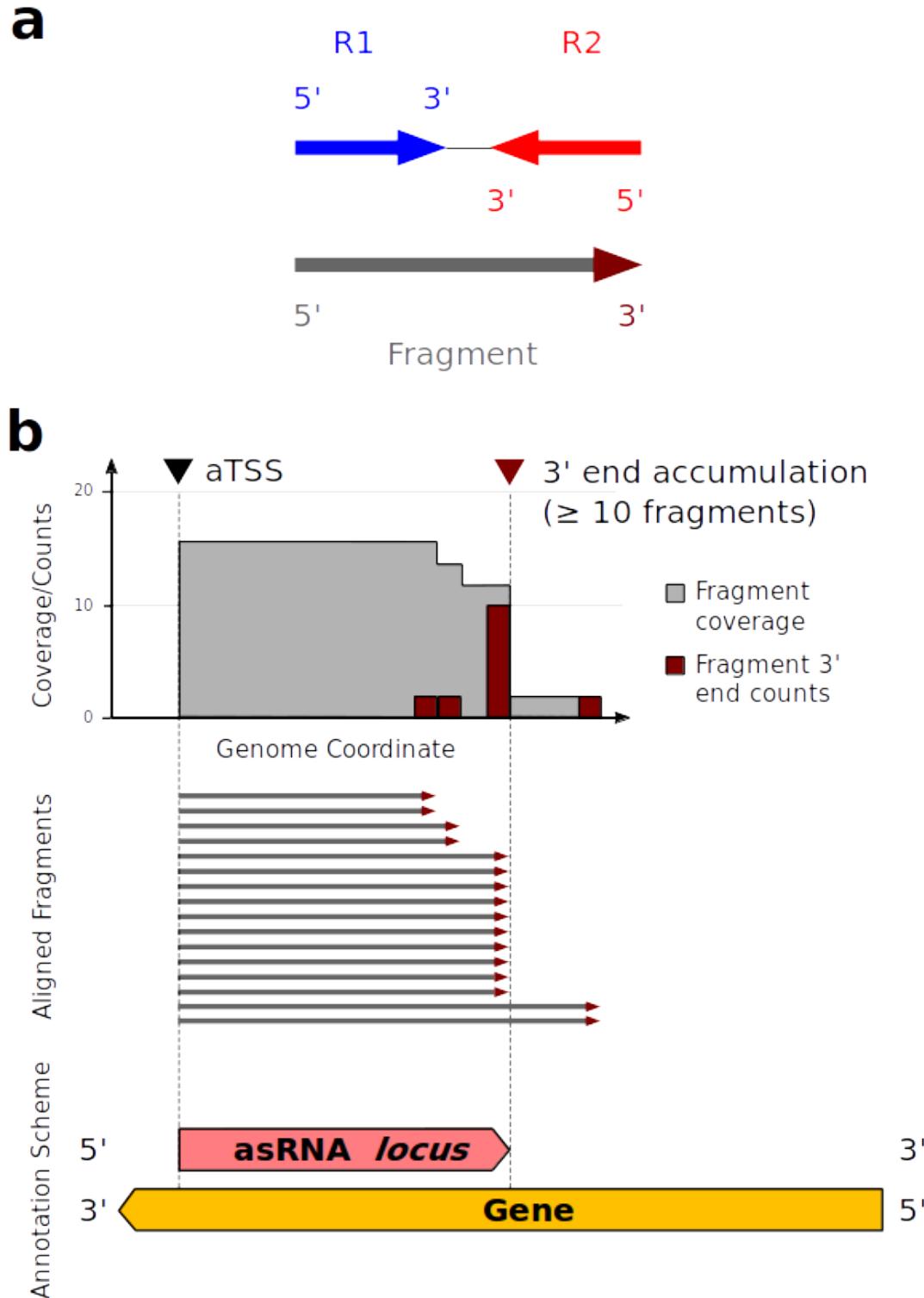
17



19Figure S1. Pipeline for asRNA annotation.

20

21



23Figure 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.

31

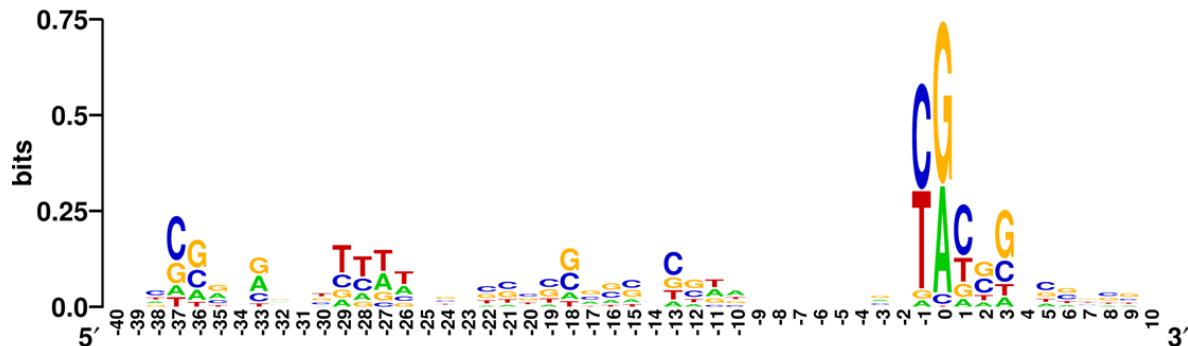
32

33

34

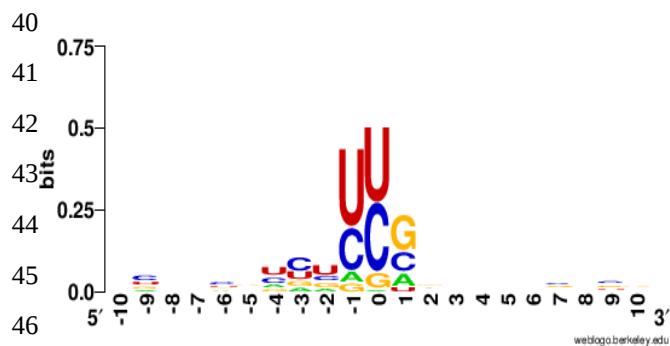
35

36(a)



38

39(b)



48

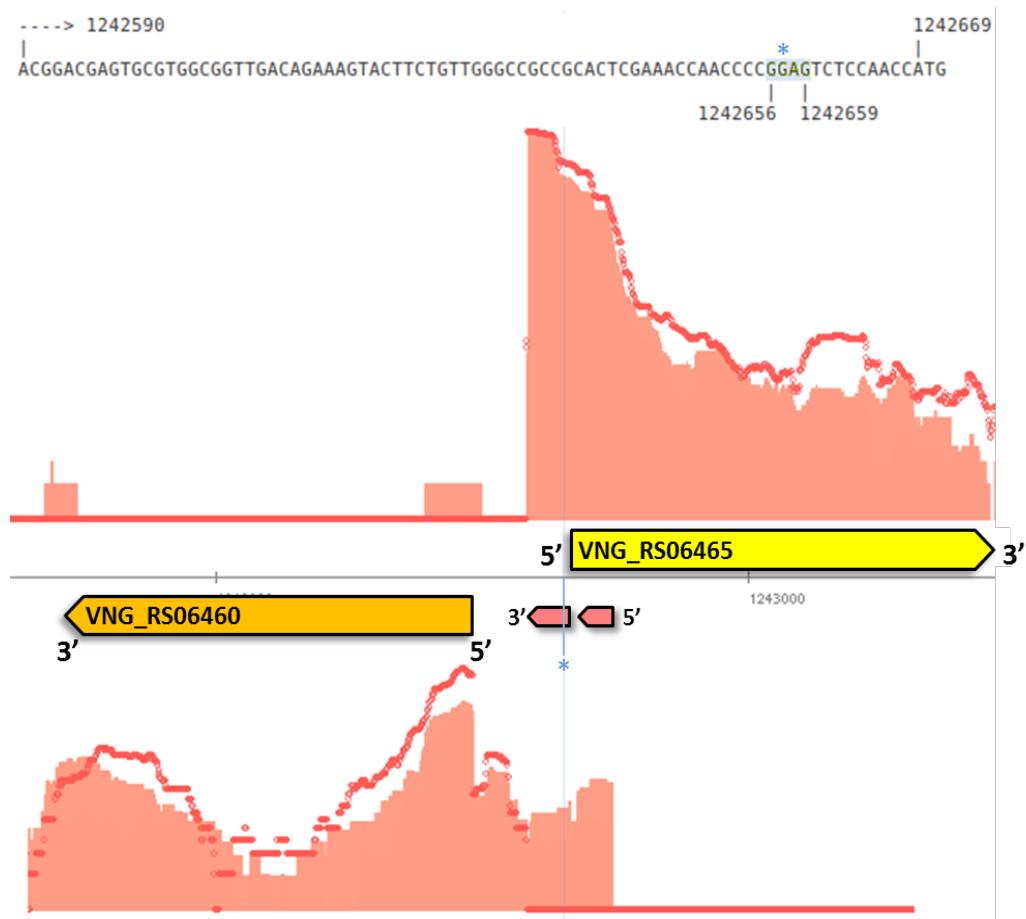
49 nt downstream of aTSS (position 0). (b) Frequency of nucleotides at the 3' end of asRNAs. Position 0
50 is the last nucleotide.

31

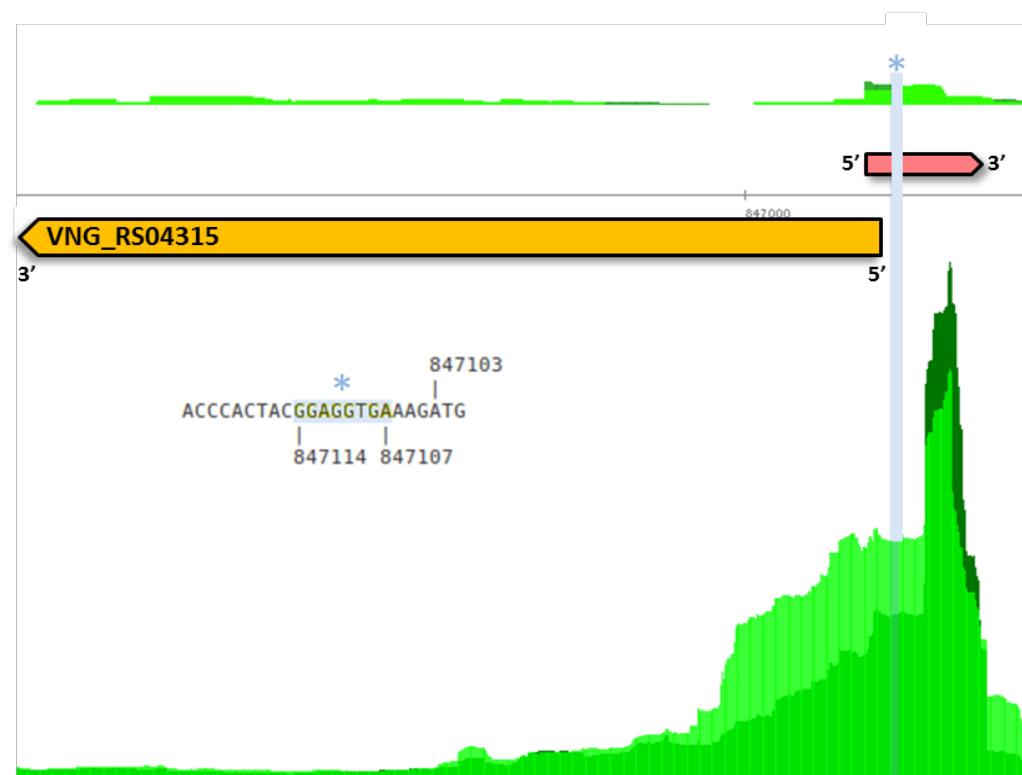
52

55

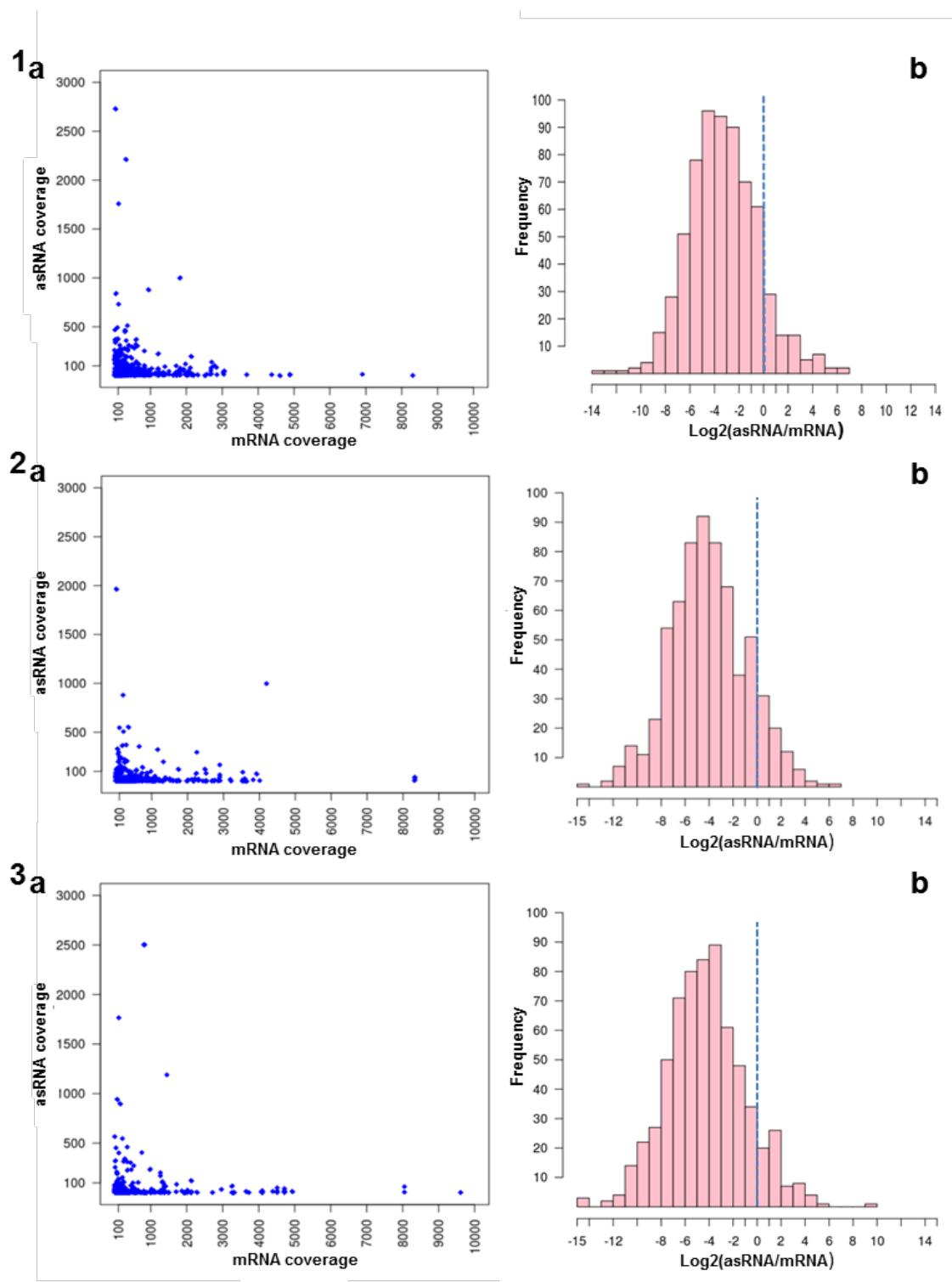
54 (a)



56 (b)



61Figure S4. Examples of putative RBS occlusion by an asRNA. (a) *cdcH*, encoding an AAA-type ATPase
62(VNG_RS06465). VNG_RS06465 *locus* (yellow arrow) is in forward strand (5'→3' left to right), neighbor
63gene VNG_RS06460 (orange arrow) is in reverse strand (5'→3' right to left). Aligned reads coverage along
64genomic coordinates for TEX+ libraries at exponential and stationary phases are shown in light red
65(solid) and red (dots), respectively (log₂ counts jointly normalized and arbitrarily scaled).
66VNG_as06465_925 asRNA (pink arrow) encompasses Shine-Dalgarno-like signature (* light blue
67highlight). (b) *rpl1* encoding the 50S ribosomal protein L1. VNG_RS04315 *locus* (orange arrow) in reverse
68strand (5'→3' right to left). dRNA-seq read coverage signal is shown in dark and light green for TEX+ and
69TEX- libraries, respectively. VNG_as04315_654 asRNA (pink arrow) encompasses Shine-Dalgarno-like
70signature (* 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 log₂ 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

Log2FC

	Log2FC	Log2FC	
aTSS_2145	3.6	1.2	VNG_RS12020
aTSS_617	3.6	1.3	VNG_RS04115
aTSS_1347	2.7	3.1	VNG_RS09530
daTSS_80	2.3	2.3	VNG_OE5069A1
● aTSS_1949	2	-1	VNG_OE5075F
● aTSS_1426	1.8	3.3	VNG_OE4600F
daTSS_68	1.2	2.6	VNG_RS12750
aTSS_49	1.2	-1.3	VNG_RS00175
● aTSS_810	-1	1.4	VNG_RS05605
daTSS_3	-1.4	2.7	VNG_RS00140
aTSS_390	-2	-1.5	VNG_RS02565
aTSS_1736	-2	-1.3	VNG_RS12765
daTSS_82	-2.1	1.9	VNG_RS13545
aTSS_84	-2.2	-2.2	VNG_RS00235
aTSS_32	-2.4	-1.1	VNG_RS00115
● aTSS_742	-2.6	1.2	VNG_RS05060
aTSS_184	-2.7	1.2	VNG_RS00870
aTSS_231	-2.8	1	VNG_RS01200
aTSS_20	-2.8	1.5	VNG_RS00105
aTSS_1884	-2.9	-1.4	VNG_RS11055
aTSS_701	-2.9	-1.6	VNG_RS04790
aTSS_1735	-3	-1.3	VNG_RS12765
aTSS_908	-3.1	1.8	VNG_RS06345
● aTSS_1321	-3.6	-3.2	VNG_RS09335
● aTSS_539	-3.6	1.7	VNG_RS03715
aTSS_33	-5.9	-1.1	VNG_RS00115

82

83

84

85

86(b)

87

Log2FC

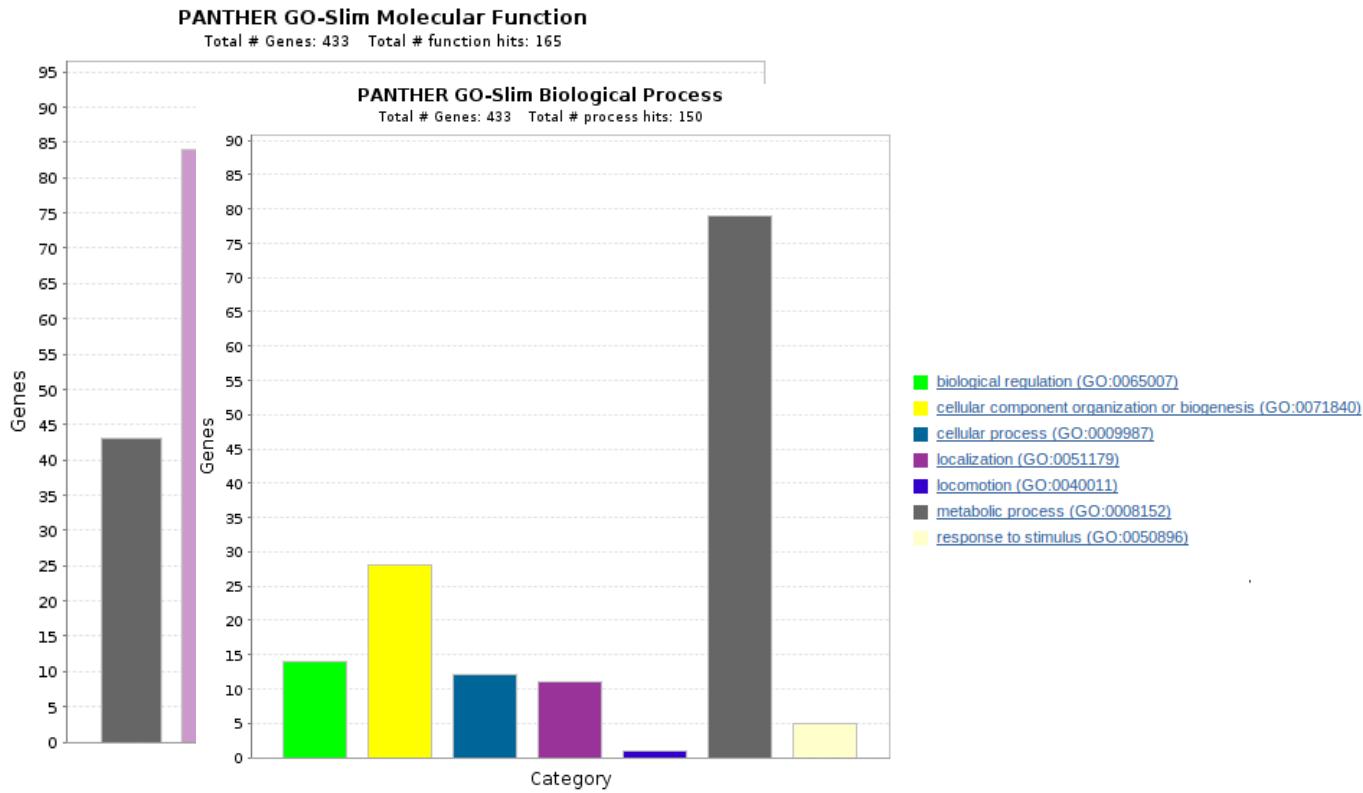
			Log2FC	
aTSS_1484	5.8	4.3	VNG_RS10400	transposase
aTSS_999	4.9	-1.2	VNG_RS07090	CTP synthetase
aTSS_203	4.2	1.3	VNG_RS00885	hypothetical protein
aTSS_62	4.2	1.9	VNG_RS13210	hypothetical protein
● aTSS_297	4.1	2.6	VNG_RS01695	sodium:proton antiporter
aTSS_2057	3.8	2.7	VNG_RS13595	hypothetical protein
aTSS_571	3.8	1.7	VNG_RS03850	integrase
aTSS_573	3.8	2.4	VNG_RS03865	hypothetical protein
aTSS_919	3.3	1.3	VNG_RS06405	hypothetical protein
aTSS_122	3	1.3	VNG_RS00420	restriction endonuclease subunit S
aTSS_1898	3	1.4	VNG_RS11080	hypothetical protein
aTSS_1762	3	2	VNG_RS12815	pyridoxamine 5'-phosphate oxidase family
● aTSS_78	2.9	2.4	VNG_RS00215	hypothetical protein
aTSS_1763	2.7	2	VNG_RS12815	pyridoxamine 5'-phosphate oxidase family
● aTSS_826	2.5	2.9	VNG_RS05750	hypothetical protein
aTSS_1485	2.5	4.3	VNG_RS10400	hypothetical protein
aTSS_86	2.4	1.4	VNG_RS00240	transposase
● aTSS_2050	2.4	2.9	VNG_RS13590	hypothetical protein
aTSS_187	2.3	2	VNG_RS13265	hypothetical protein
aTSS_362	2.3	1.1	VNG_RS02300	DUF4352 domain-containing protein
aTSS_1731	2.2	2	VNG_RS12755	cytochrome b
aTSS_922	2.1	1.8	VNG_RS13355	hypothetical protein
aTSS_65	2.1	1.9	VNG_RS13210	ATP-binding protein
aTSS_1732	1.9	2	VNG_RS12755	hypothetical protein
aTSS_1923	1.8	1.4	VNG_RS11165	hypothetical protein
● aTSS_2013	1.8	1.9	VNG_RS13565	chromosome segregation protein SMC
● aTSS_1915	1.7	1.6	VNG_RS11110	hypothetical protein
● aTSS_373	1.2	1.1	VNG_RS02420	hypothetical protein
aTSS_363	-2.3	1.1	VNG_RS02300	cytochrome b
aTSS_2056	-3.3	1.2	VNG_RS11690	IS4 family transposase ISH30

89

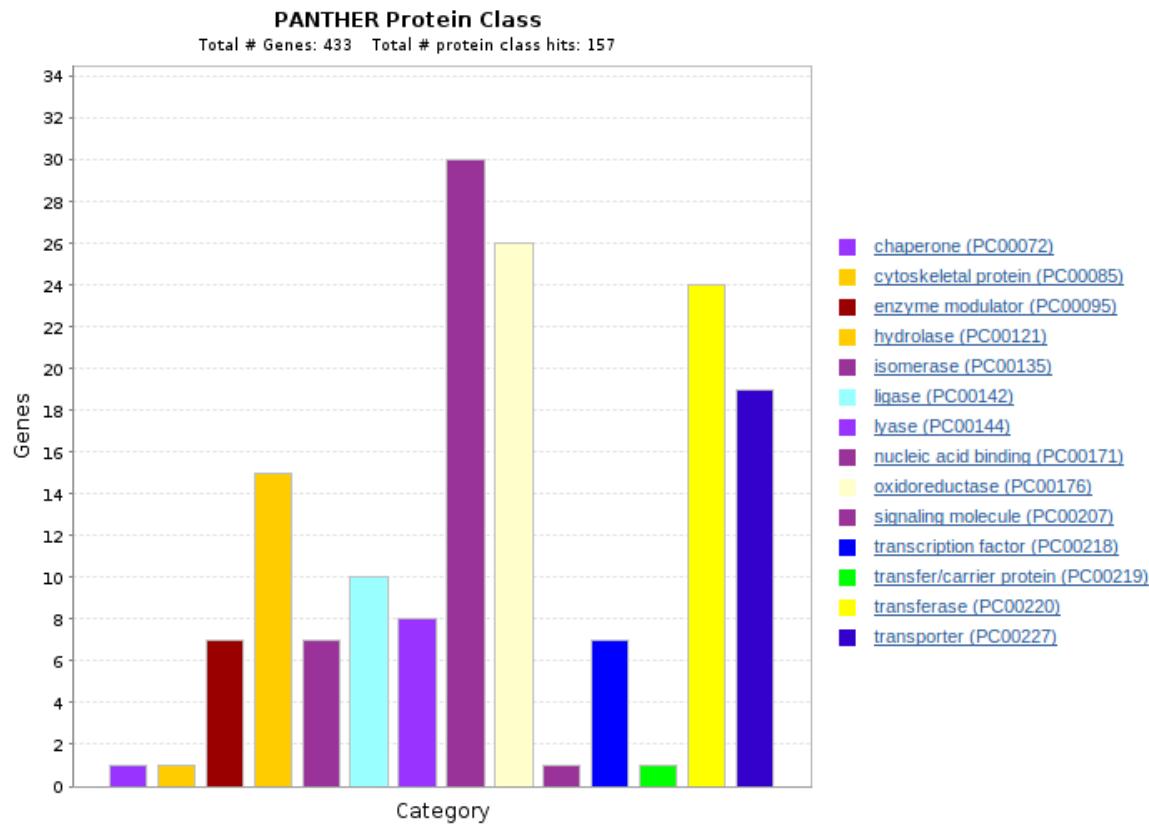
90**Figure S6.** Differential expression of asRNAs and mRNAs on the opposite strand. Pairs asRNA/mRNA that
 91 presented either positive or negative correlation are shown. Up-regulated genes have its log fold change
 92 ($\log_2\text{FC}$) inside red cells and down-regulated genes inside green cells. (a) $\log_2\text{FC}$ for 37 h/17 h (b) $\log_2\text{FC}$ for 86
 93 h/37 h. Orange dots highlight asRNAs overlapping the 5' UTR of the mRNA.

94

95



97
98
99
100
101
102



104

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

108

109

110

111

112

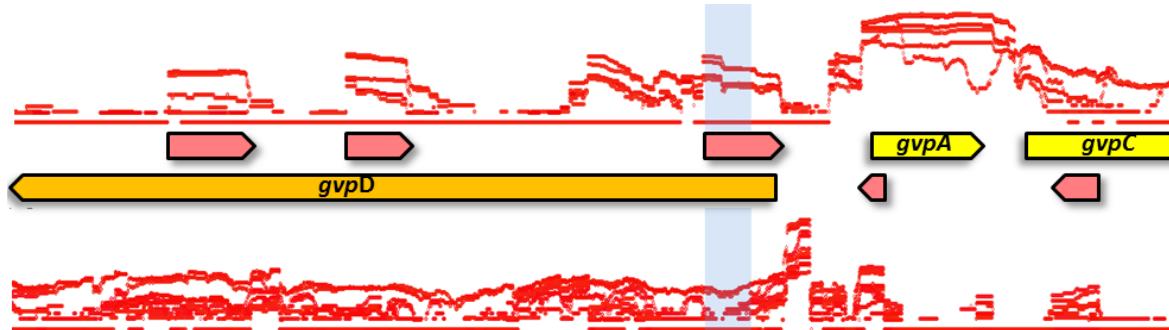
113

114

115

116

117

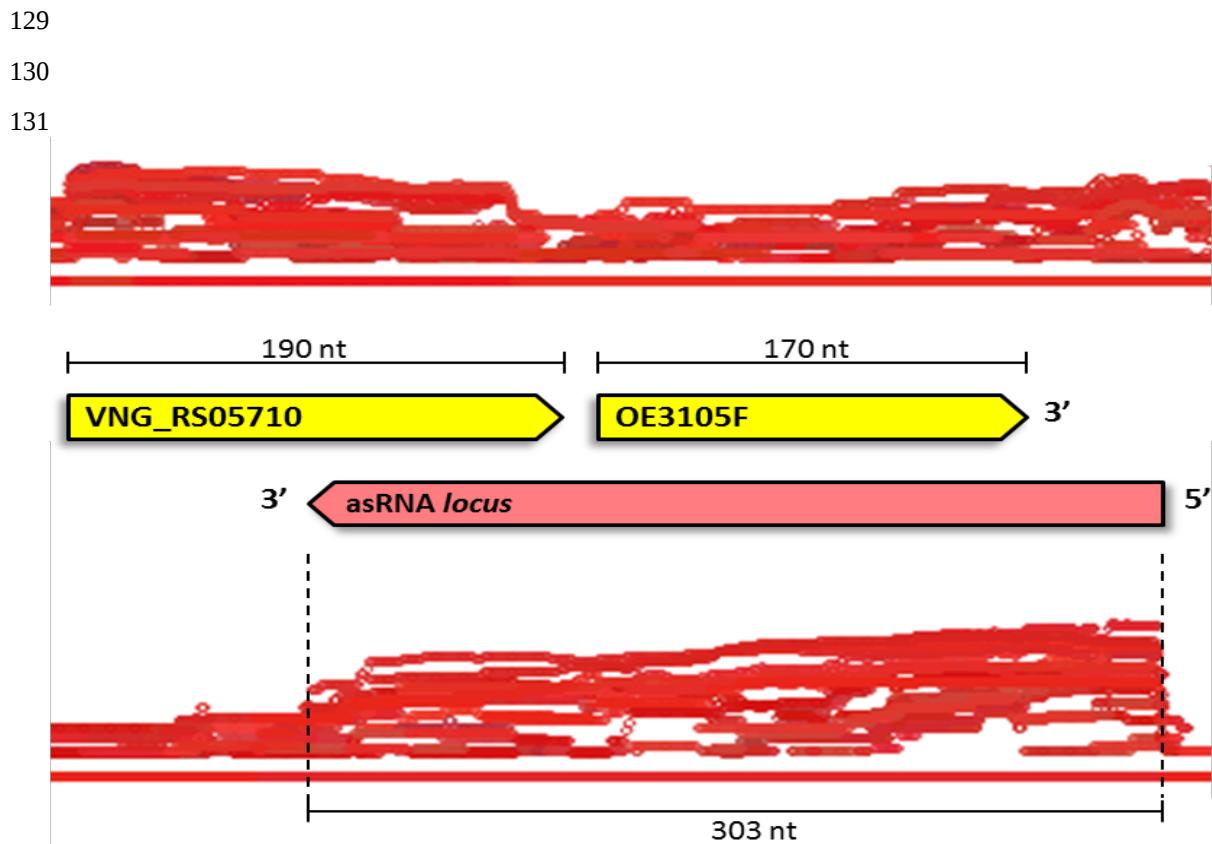


119

120 **Figure S8.** asRNAs in *gvp* genes. Upper panel shows RNA-seq log₂ counts for forward strand, and
121 lower panel shows signal for the reverse strand. All sequenced libraries are shown simultaneously
122 and normalized (profiles as red dots). Genes *gvpA* and *gvpC* in forward strand (yellow arrows) and
123 *gvpD* in reverse strand (orange arrow). Annotated asRNAs are represented by pink arrows.
124 asRNAs related to *gvp* genes showed mean expression levels lower than their cognate genes. Region
125 corresponding to the probe used for asRNA detection by Krüger and Pfeifer (1996) is highlighted in
126 light blue.

127

128



133

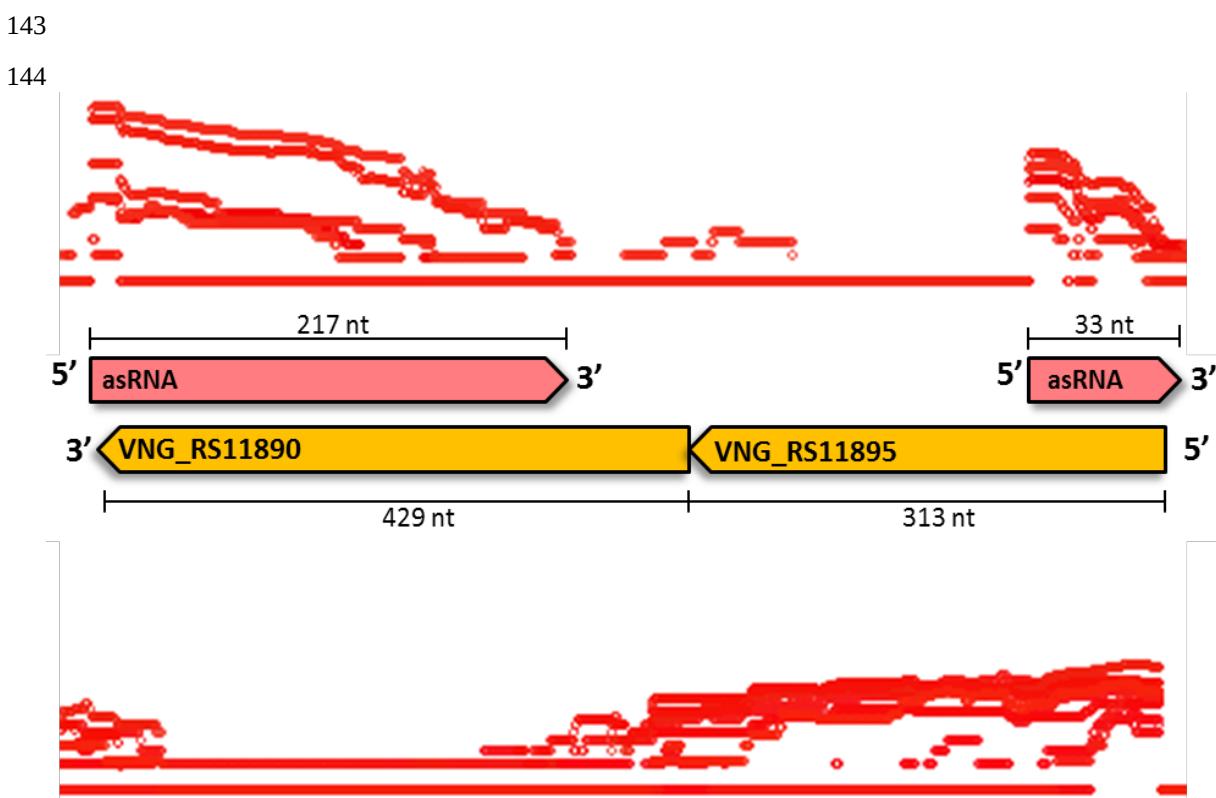
134 **Figure S9.** asRNAs in *brb* and *brz* genes. Upper panel shows RNA-seq \log_2 counts for forward
135 strand, and lower panel shows signal for the reverse strand. All sequenced libraries are shown
136 simultaneously and normalized (profiles as red dots). *brz* (VNG_RS05710) and *brb* (OE3105F)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

139

140

141

142



146 **Figure S10.** asRNAs in type II TA system. Upper panel shows RNA-seq log₂ counts for forward
147 strand, and lower panel shows signal for the reverse strand. All sequenced libraries are shown
148 simultaneously and normalized (profiles as red dots). Genes VNG_RS11890 (toxin) and
149 VNG_RS11895 (antitoxin) are on the reverse strand (orange arrows) and their respective annotated
150 asRNAs VNG_da11890_90 and VNG_as11895_2118 on the forward strand (pink arrows).

151

152

153

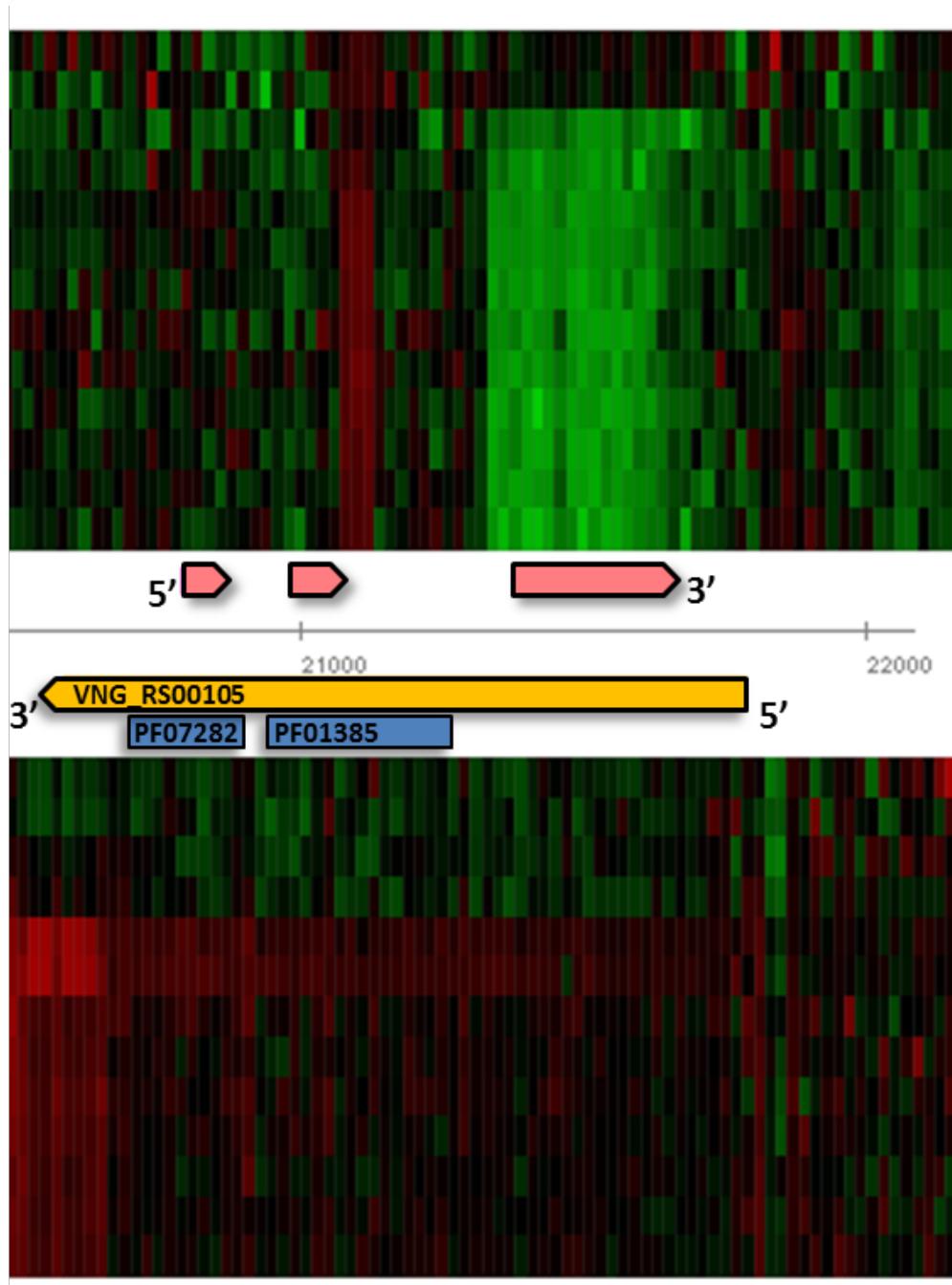
154

155

156

157

158



160

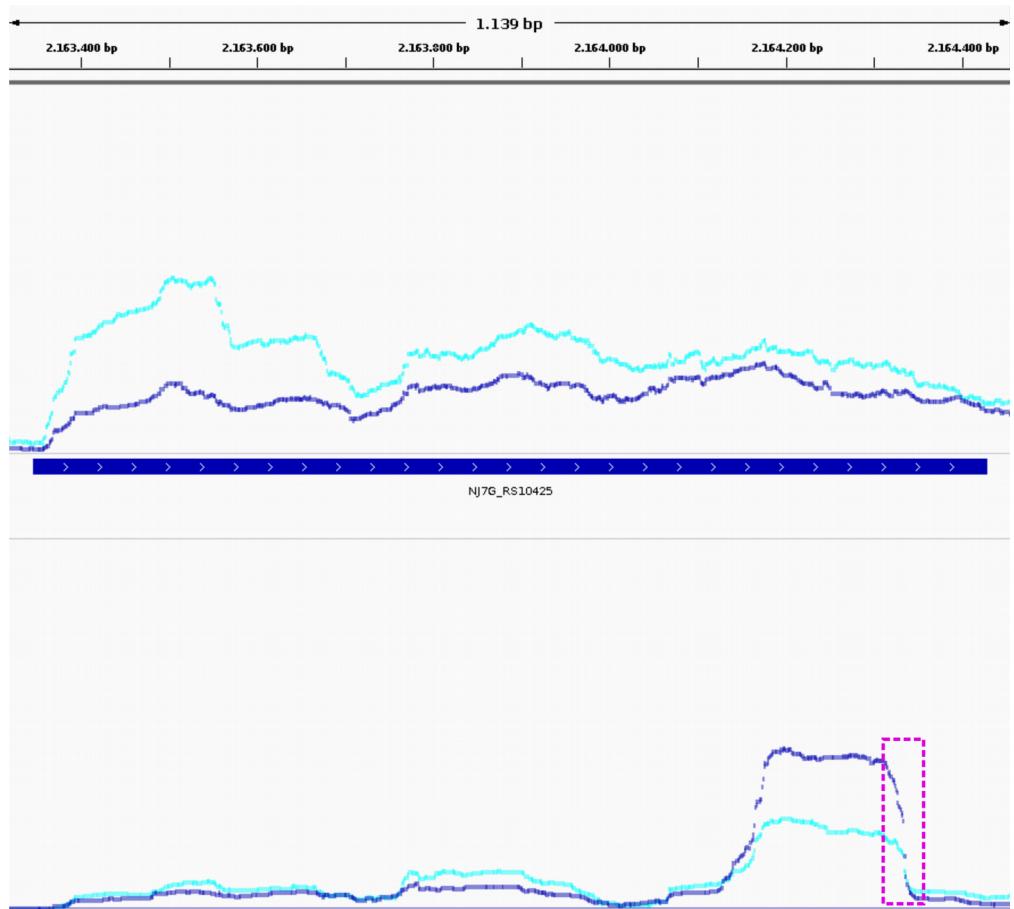
161 **Figure S11.** asRNAs antisense to a transposase differentially expressed along the growth curve.
162 VNG_RS00105 (orange arrow) encodes a transposase on the reverse strand (5'→ 3' right to left),
163 PFAM domains are shown in blue. asRNAs are shown in pink arrows on forward strand (5'→ 3' left
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

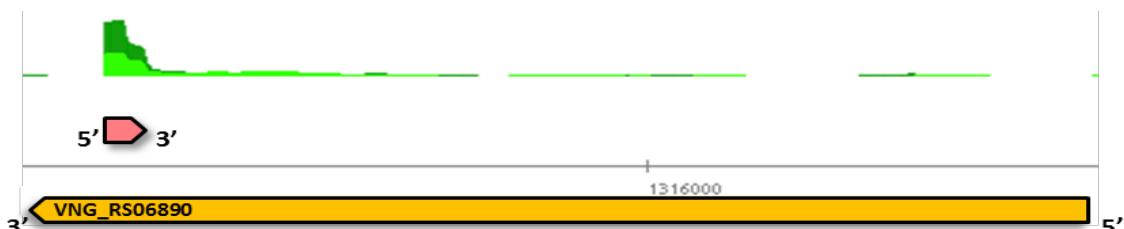
170(a)

171



173(b)

174

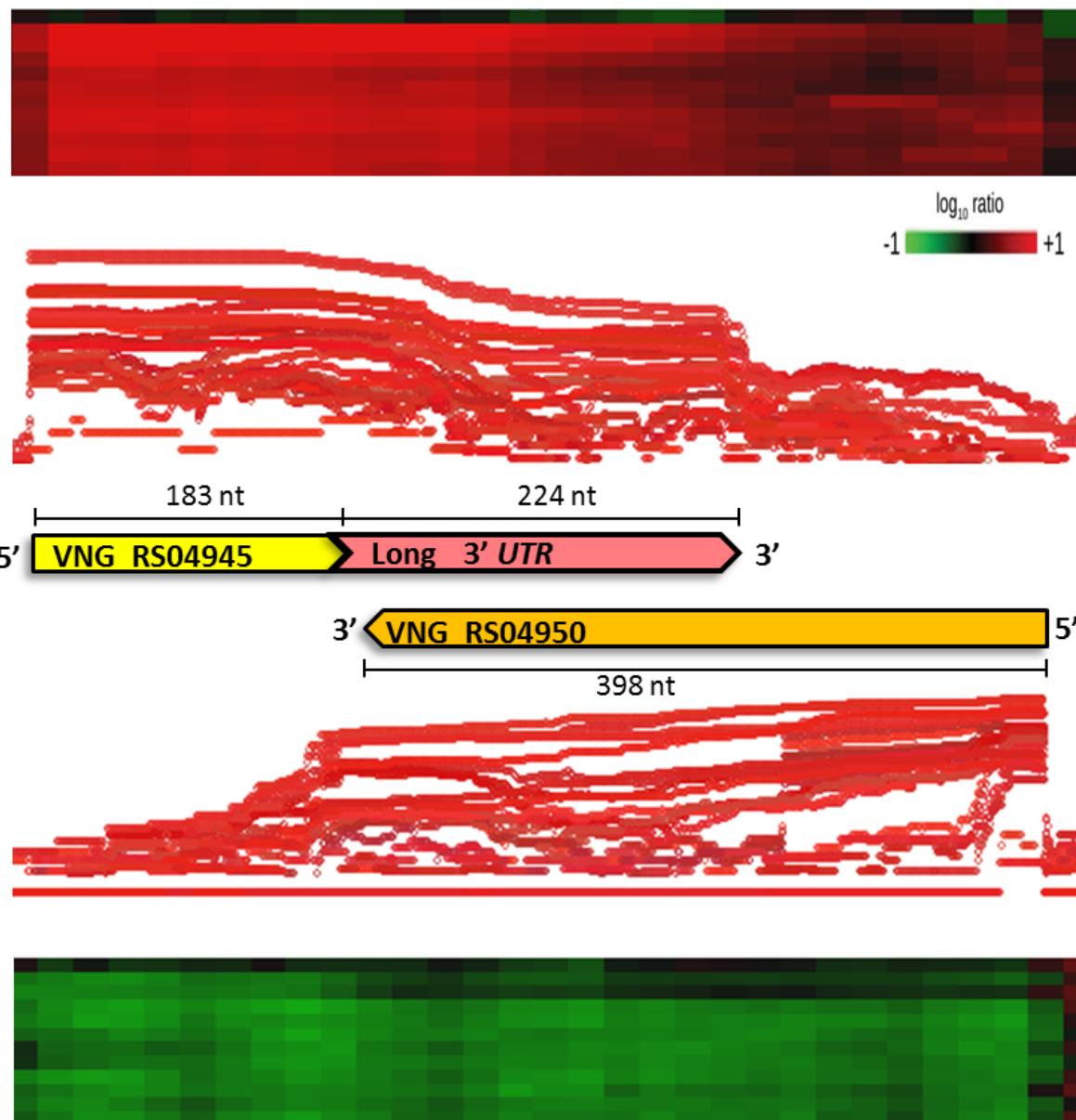


176 **Figure S12.** Differential expression of putative asRNAs antisense to the *nirH* gene in *Natrinema* sp.
 177 J7-2. (a) Genomic region showing normalized read coverage for low (15% NaCl; light blue) and high
 178 (30% NaCl, dark blue) salt concentrations, respectively. The y-axes range from 0 to 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

187



189

190 **Figure S13.** Long 3' UTR of a coding sequence acting as an asRNA. Pink arrow indicates a possible
191 long 3' UTR of the gene VNG_RS04945 (yellow arrow, annotated as hypothetical protein) which
192 could act as an asRNA of the gene VNG_RS04950 (orange arrow, annotated as a subunit of a
193 translation initiation factor). Upper panel shows tilling array expression data for the forward strand
194 and lower panel for the reverse (Koide *et al.* 2009). Red lines show RNA-seq signal (log₂ counts per
position).

195

196

197

198



199 © 2019 by the authors. Submitted for possible open access publication under the terms
200 and conditions of the Creative Commons Attribution (CC BY) license
201 (<http://creativecommons.org/licenses/by/4.0/>).