

Supplemental material

Figure Legends

Figure S1: Schematic of the antisense small RNAs that map to the EHI_197520 gene.

471 unique sRNAs map antisense (AS) to the coding-region of EHI_197520 with a 5'-distribution bias. The EHI_197520 gene length is 1233bp; in orange and green are the areas designated as the 5' trigger and 3' trigger, respectively.

Figure S2: Schematic of constructs used. **(A)** The first 132bp of the coding regions of genes (EHI_197520, EHI_048600, and EHI_025700) are cloned downstream of their endogenous promoter and fused to the firefly luciferase gene. **(B)** Shorter sections (102bp, 75bp, 48bp) of EHI_197520 coding region are cloned downstream of the EHI_197520 endogenous promoter and fused to the firefly luciferase gene. **(C) (i)** The first 132bp of the coding regions of genes (EHI_197520, EHI_048600, and EHI_025700) are cloned downstream of the cysteine synthase (CS) promoter and fused to the firefly luciferase gene. **(ii)** The first 132bp of the coding region of EHI_197520 is fused to the 3'-end of the luciferase gene. **(D) (i)** The last 132bp of the coding region of the EHI_197520 gene is cloned downstream of the CS promoter and fused to the firefly luciferase gene. **(ii)** The last 450bp of the coding region of EHI_197520 is fused to the 5'-end of the luciferase gene. **(iii)** The last 132bp of the coding region of EHI_197520 is fused to the 3'-end of the luciferase gene. **(E)** Constructs used to generate the stable cell lines with the 132bp EHI_197520-trigger fused to the 5'-end of the full-length genes **(i)** EhROM1, EhAgo2-1, EhAgo2-2, EhAgo2-3, EhRdRP1 and EhRNaseIII **(ii)** EhMyb **(iii)** EhROM1 but lacking a promoter **(iv)** the last 450bp of the coding region of EHI_197520 is fused to the 5'-end of the EhROM1 gene. **(v)** Constructs used to evaluate stable cell lines for functional AS sRNAs. They contain either 132bp of EhROM1, 132bp of EhROM1 recoded or 132bp of EhAgo2-2 as a "trigger" and are cloned downstream of the CS promoter and fused to the firefly luciferase gene **(F) (i)** The Firefly luciferase gene with no promoter. **(ii)** The Firefly luciferase gene driven by the CS promoter. **(iii)** The Renilla luciferase gene driven by the CS promoter.

Figure S3: EhROM1 silencing is specific and maintained after drug selection is removed. Evaluation of *E. histolytica* HM-1:IMSS parasites stably transfected with the 19 trigger-ROM construct with and without drug pressure. EhROM1 mRNA level and EhROM AS sRNA abundance was analyzed after two and three month of drug release. **(A)** RT-PCR of the 19 trigger-ROM and untransfected cell lines reveals that the EhROM1 knockdown persists two months after plasmid removal. **(B)** RT-PCR of the 19 trigger-ROM and untransfected cell lines showed that the EhROM1 knockdown also persists three months after plasmid removal. RT-PCR of the EHRM2 gene (EHI_060330, 56% sequence identity) showed unaffected mRNA level under all conditions tested.

Table legends

Table S1: Genes with strain-specific patterns of sRNA and mRNA abundance.

Genes selected that have abundant antisense small RNAs and low mRNA abundance in a strain-specific manner for *E. histolytica* strains HM-1:IMSS and Rahman. The selection was based on microarray data (1,2) and the sRNA library of HM-1:IMSS and Rahman (3). For all genes the number of unique AS sRNAs is indicated. For the gene EHI_197520 AS sRNAs were detected in HM-1:IMSS (see Figure SM1 for schematic) and no mRNA expression (mRNA expression value: 0.07). For the same gene no sRNA were detected in the Rahman strain, but mRNA expression was detected (mRNA expression value: 6.75). For the gene EhSTIRP/EHI_025700 AS sRNAs were detected in Rahman but no mRNA expression (mRNA expression value: 0.05). For the same gene no sRNAs in the HM-1:IMSS strain, but mRNA expression was detected (mRNA expression value: 28.0). For the gene EHI_048600 AS sRNAs were detected in HM-1:IMSS and Rahman and no mRNA expression is detected in either strain (mRNA expression value: HM-1:IMSS 0.04; Rahman 0.04). RNA sequencing results showed that a conserved EHI_048600 gene is moderately expressed in *E. invadens* (G. Ehrenkauffer, N. Hall and U. Singh; unpublished data). N/A is not available.

Table S2: Primer for Construct cloning. All cloning primers used in this study are listed.

Restriction enzymes in the primer cloning sites are mentioned in the name of the primer and highlighted in bold in the primer sequence.

Table S3: Primer for Northern Blot. All primers used for Northern Blot analyses are listed.

Table S4: Primer for RT-PCR. All primers used for RT-PCR are listed.

1. MacFarlane, R.C. and Singh, U. (2006) Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species: potential implications for amebic pathogenesis. *Infect Immun*, **74**, 340-351.
2. Vicente, J.B., Ehrenkaufer, G.M., Saraiva, L.M., Teixeira, M. and Singh, U. (2009) *Entamoeba histolytica* modulates a complex repertoire of novel genes in response to oxidative and nitrosative stresses: implications for amebic pathogenesis. *Cellular microbiology*, **11**, 51-69.
3. Zhang, H., Ehrenkaufer, G.M., Hall, N. and Singh, U. (2013) Small RNA pyrosequencing in the protozoan parasite *Entamoeba histolytica* reveals strain-specific small RNAs that target virulence genes. *BMC Genomics*, **14**, 53.

Figure S1

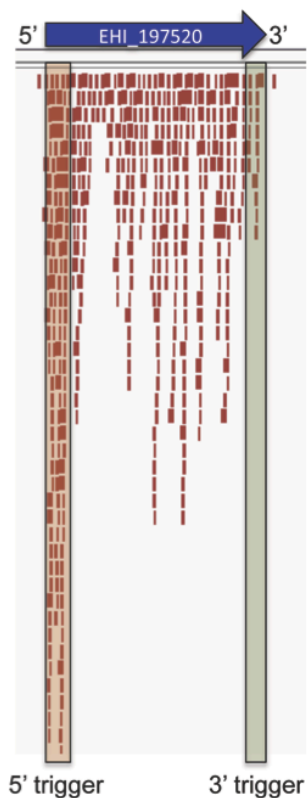


Figure S2

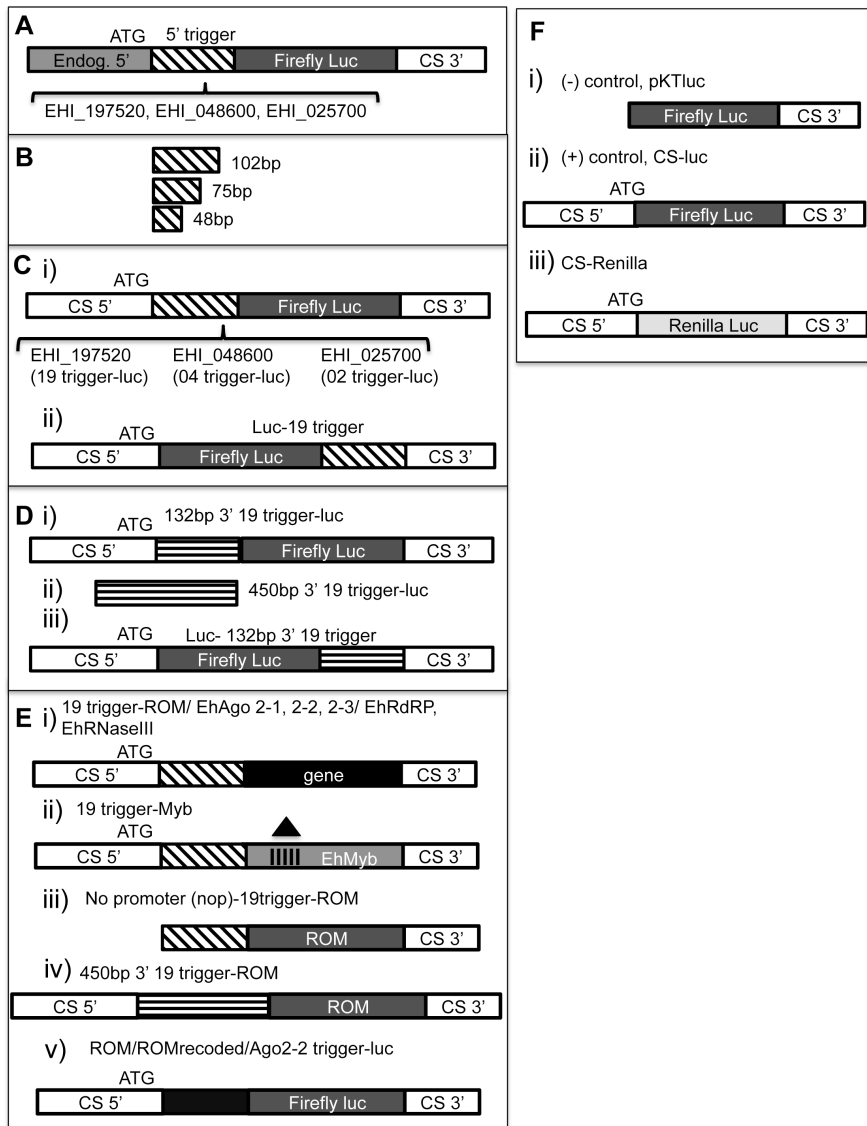
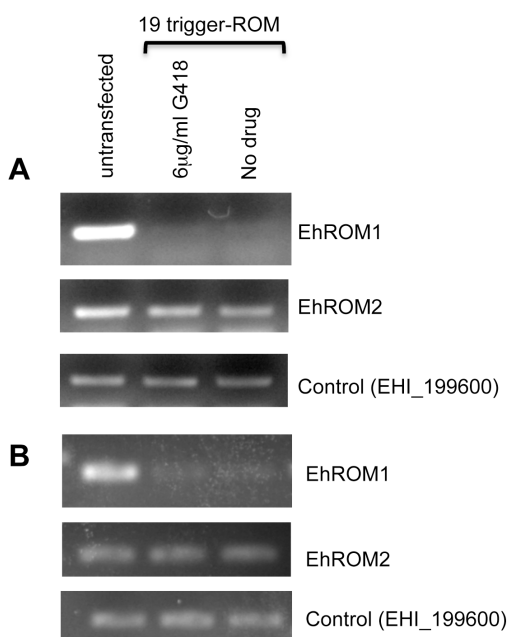


Figure S3



<i>Entamoeba</i> species and strains		EHI_197520 (Hypothetical protein)	EHI_025700 (Serine,Threonine, and Isoleucine rich protein (EhSTIRP))	EHI_048600 (Hypothetical protein)
<i>E. histolytica</i> HM-1:IMSS (virulent)	# of unique AS small RNAs	484	5	684
	mRNA expression	not detectable	high	not detectable
<i>E. histolytica</i> Rahman (non virulent)	# of unique AS small RNAs	zero	602	89
	mRNA expression	medium	not detectable	not detectable
<i>E. invadens</i> IP-1	# of unique AS small RNAs	N/A	N/A	N/A
	mRNA expression	N/A	N/A	medium

Table S1

Table S2

048600promoter-HindIII-s	ACAAAGCTTTGGTGTCTCCTGTAATGCA
048600promoter-AvrII-as	ACACCTAGGTTGTTATATTTAGAATTATTAGT
048600-21bp-AvrII-as	ACACCTAGGGAGGGTTAATTCAATTTCCA
048600-132bp-AvrII-as	ACACCTAGGCGTTGATGCTGCAATTTTTGT
132bpof04-AvrII-s	ACACCTAGGATGGAAATTGAATTAACCCT
025700promoter-HindIII-s	ACAAAGCTTTTGATACTTCGTTATTCTAAAATA
025700promoterplus21-AvrII-as	ACACCTAGGTATTAACAATCTTTTGACATACA
025700promoterplus132bp-AvrII-as	ACACCTAGGATTTGTATTTTCATTGTTCTCTTCA
025700-AvrII-s	ACACCTAGGATGTCAAAGATTGTTTAATAAACT
197520promoter-HindIII-s	ACAAAGCTTCCTTATTTCAATCTTTCTTTTTCT

197520promoter-AvrII-as	ACACCTAGGGAATAAATAAATGAATCTTGA
197520+21-AvrII-as	ACACCTAGGAGTTGGTTGAGCTGAAGACAT
197520promoter+132-AvrII-as	ACACCTAGGAGTTCCAGTACTTACTGGT
197520-132bp-AvrII-s	ACACCTAGGATGTCTTCAGCTCAACCAACT
197520+ 48bp-AvrII-as	ACACCTAGGGTATTGAGGTTGATTAGAATTAGT
197520+75bp-AvrII-as	ACACCTAGGATATTGTTGATTTGCTGCATCA
197520+102bp-AvrII-as	ACACCTAGGCATTGACTGTTGAATAGGA
197520-XhoI-s	ACACTCGAGATGTCTTCAGCTCAACCAACT
197520+132-XhoI-as	ACACTCGAGTTAAGTTCCAGTACTTACTGGT
Luc-s	ACACCTAGGATGGAAGACGCCAAAAACA
Luc-as noSTOP-XhoI	ACACTCGAGCAATTTGGACTTTCCGCCCT
197520+132-SmaI-as	ACAGTCGACTTAAGTTCCAGTACTTACTGGT
197520-3`-132bp-XhoI-s	ACACTCGAGGGAGCACCAAGATTTGATCA
197520-3`-132bp-XhoI-as	ACACTCGAGTTAAAAACCAGAGCCATAACT
197520-3`-132bp-noStop-AvrII-s	ACACCTAGGGGAGCACCAAGATTTGATCA
197520-3`-450bp-AvrII-s	ACACCTAGGGTTCAACAACAAATGACTAATGCT
ROMrec-132bp-AvrII-s	ACACCTAGGATGCACTCCCCACACAACAACATCCACA
ROMrec-132bp-AvrII-as	ACACCTAGGGGCCTTGCGGGTGCTGGT
ROM-AvrII-s	ACACCTAGGATGCATTCTCCACCACATAACA
ROM-132bp-AvrII-as	ACACCTAGGTTTAGCTGTTGAAGTAGGTGT
EHI_197980-s	ACACCCGGGATGAGTGTCACTCTTTC
EHI_197980 -as	ACACTCGAGTCAAATCAATGCTTGATTG

Table S3

Luciferase AS sRNAs	AGGCCCGGCGCCATTCTATCCTCTAGAGG
EhROM1 AS sRNAs	ATGGATGATGATAGAGAACCAACACCTACT
EhMyb AS sRNAs	ATGAGTGTCACTCTTTCTTCTACTGAAAAT
EhMyb AS sRNAs to the intron	GTTTCTTAGTGTCTTTTAAACCCCTTAAA
EhMyb AS sRNAs after the intron	TATTCAAACAACACTGGAGGAATTAACAAA
EhROM1 mRNA	TTCACCTTCCAGGCATTCTC

Table S4

ROM1-RTs	GGGAGACATGAGGTTTCGAGT
ROM1-RT-as	GGTGAACATGGTGGTGAAT
Eh_199600-s	ACGTCATGCTGAATTTGCTG
Eh_199600-as	CCTTTAAGCCCAGCCTTTCT
EhMybRT-s	CGAATTTACAAAACAAATTGAGATG
EhMybRT-as	CAATGCTTGATTGAAAAATGGA
ROM-060330RT-s	CCAATTGTTTGGGTCCAAC
ROM-060330RT-as	GGAAGAATCCTATTTGCCATT