Supplementary Figures

CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III

This file contains:

Figure	Title	Page
1	Genomic organization of type II (Nmeni/CASS4) CRISPR/Cas loci in *S. pyogenes*	3
2	Characterization of CRISPR01 pre-crRNA and crRNAs in *S. pyogenes* SF370	5
3	RT-PCR analysis of tracrRNA, CRISPR01 crRNA and CRISPR02 crRNA expression in *S. pyogenes* SF370	7
4	Conservation and expression of tracrRNA in *S. pyogenes*	8
5	*In silico* co-foldings of tracrRNA and crRNA species	11
6	Role of endogenous RNase III in tracrRNA-mediated maturation of pre-crRNA *in vivo*	13
7	Compensatory base-pair exchanges confirm tracrRNA-mediated crRNA maturation *in vitro*	14
8	Role of Cas proteins in the processing of tracrRNA and pre-crRNA *in vivo*	16
9	tracrRNA, pre-crRNA, RNase III and Csn1 confer immunity against acquisition of a lysogenic phage-protospacer gene	19
10	Genomic organization of type II (Nmeni/CASS4) CRISPR/Cas systems from selected bacterial species	21
11	Alignment of predicted tracrRNA homologues from selected bacterial species	23
12	tracrRNA homologue and pre-crRNA are expressed and processed in *Streptococcus mutans* UA159	24
13	tracrRNA homologue and pre-crRNA are expressed and processed in *Streptococcus thermophilus* LMD-9 (1st locus)	26

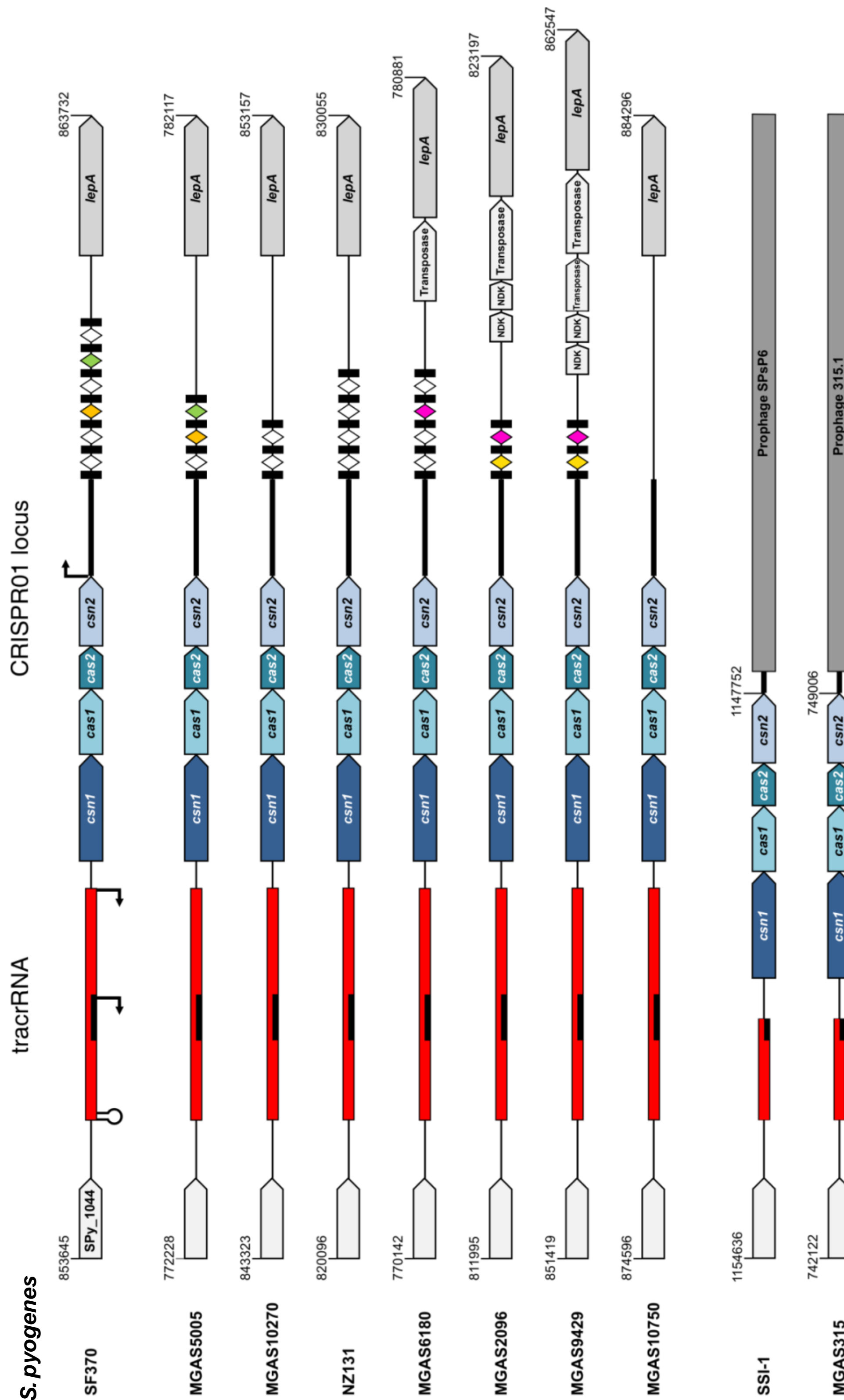
14 tracrRNA homologue and pre-crRNA are expressed and processed in *S. thermophilus* LMD-9 (2nd locus) 27

15 tracrRNA homologue and pre-crRNA are expressed and processed in *Listeria innocua* Clip11262 29

16 tracrRNA homologue and pre-crRNA are expressed and processed in *Neisseria meningitidis* Z2491 31

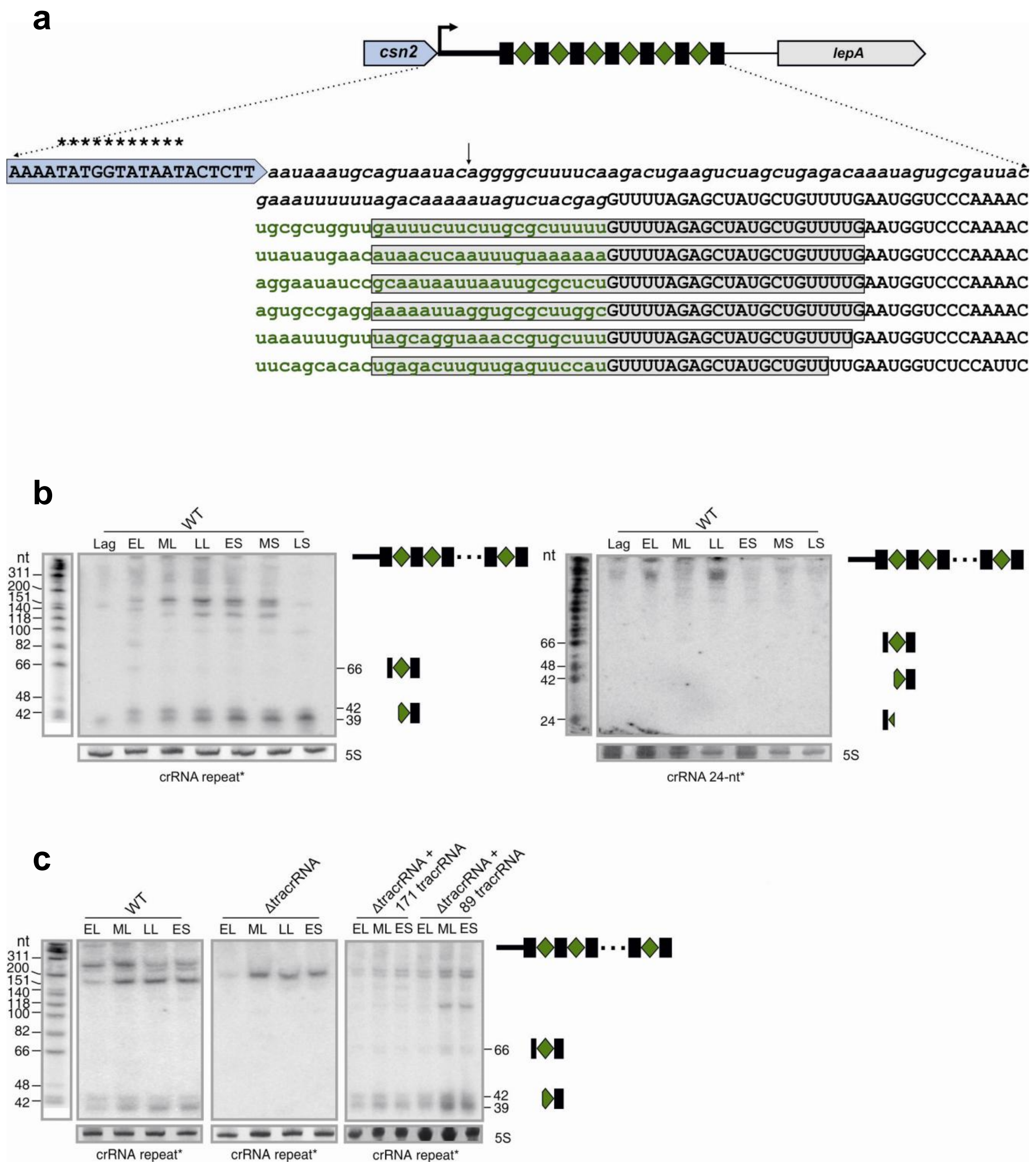
17 pre-crRNA but no tracrRNA homologue is expressed and processed in *Staphylococcus epidermidis* RP62a 33

References 35



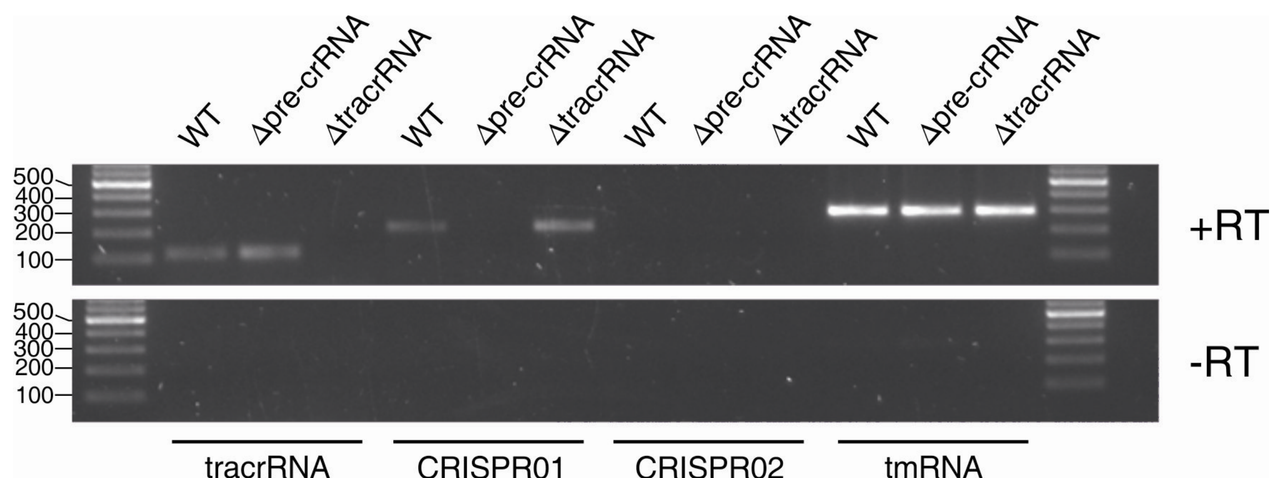
Supplementary Figure 1 | Genomic organization of type II (Nmeni/CASS4) CRISPR/Cas loci in *S. pyogenes*. A putative tracrRNA homologue (red) is encoded upstream of the CRISPR01/Cas locus in available sequenced genomes of *S. pyogenes*. The black bar within tracrRNA represents the sequence complementary to each CRISPR01 repeat. The CRISPR01 Cas genes (*csn1*-*cas1*-*cas2*-*csn2*) are followed by the

CRISPR01 leader sequence and repeat-spacer region (Supplementary Tables 1-5). Black rectangles, CRISPR01 repeats; diamonds, CRISPR01 spacers. Spacers of the same color are of identical sequence. White-filled spacers are unique in sequence. Transcriptional start sites and a Rho-independent transcription terminator are indicated. Note that in MGAS6180, MGAS2096 and MGAS9429 genomes, a transposase-encoding gene is inserted between the CRISPR01 repeat-spacer array and the *lepA* gene. The presence of transposons might be an indication for their previous involvement in the acquisition of CRISPR01-associated genes and/or sequences by *S. pyogenes* and their propagation by horizontal gene transfer among *S. pyogenes* and/or to other bacterial species. Although the MGAS10750 genome contains the CRISPR01-associated genes and leader sequence, no repeat-spacer array is present. In SSI-1 and MGAS315 genomes, a prophage sequence is inserted just downstream of the CRISPR01 leader sequence and no repeat-spacer array is present. Variations in the architecture of CRISPR loci among clinical isolates suggest rapid CRISPR evolution in this pathogen. In total considering both CRISPR01 and CRISPR02 (Supplementary Tables 1- 4), 23 spacers were found in the *S. pyogenes* analyzed genomes, from which 14 are unique and 4 are present in more than one genome. Six CRISPR loci contain one or two spacers with 80 to 100% sequence identity to endogenous target protospacers, suggesting a complex mechanism for self versus non-self discrimination in the pathogen.



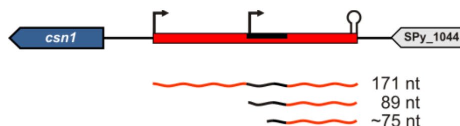
Supplementary Figure 2 | Characterization of CRISPR01 pre-crRNA and crRNAs in *S. pyogenes* SF370. a, pre-crRNA sequence. A leader sequence precedes the repeat-spacer region. Blue-highlighted, 3' end of the *csn2* gene; *, putative -10 promoter sequence of pre-crRNA; ↓, 5' end determined experimentally by head-to-tail circularization (data not shown); italic, leader sequence; lowercase, repeats; uppercase, spacers; grey-

highlighted sequence, longest and most relevant processed crRNA species observed by *in vivo* dRNA-seq (Fig. 1a,b). The sequence represented is from coordinates 860709 to 861256 (Accession number: NC_002737, SF370 genome). The leader sequence of the CRISPR01 locus is conserved among *S. pyogenes* genomes with available complete sequence containing CRISPR01 (Supplementary Tables 1,2). **b,** CRISPR01 crRNA is expressed as a pre-crRNA that is further processed into mature crRNAs. Northern blot analysis of crRNA expression in *S. pyogenes* SF370 cultures at different time points during growth: Lag; EL, early-logarithmic; ML, mid-logarithmic; LL, late-logarithmic; ES, early-stationary; MS, mid-stationary; LS, late-stationary. The blot was hybridized with probes specific to the CRISPR01 repeat (OLEC1049) (crRNA repeat*) (left panel) and to a 24-nt 1st repeat fragment (last 13 nt at the 3' end of the repeat)–2nd spacer fragment (first 11 nt at the 5' end of the spacer) (OLEC1684) (crRNA 24-nt*) (right panel). 5S rRNA (OLEC288) was used as loading control (Supplementary Table 10). pre-crRNA is processed into 66 nt intermediate repeat-spacer-repeat crRNAs by a first cleavage within each repeat (1st processing event; only observed for the 1st repeat-spacer in dRNA-seq data and expressed at very low level as shown here by northern blot analysis). This is followed by a second cleavage within the spacers resulting in mature 39-42 nt spacer (~20 nt)-repeat (~19-22 nt) crRNAs (2nd processing event). Note that the 2nd processing event occurs at a specific distance from the 1st cleavage within the repeats. Considering that spacer sequences are not identical among each other, it is thus likely that the 2nd processing event within the spacers is distance-dependent rather than sequence-dependent. Low abundance of 66 nt crRNAs suggests efficient, concomitant and rather fast 1st and 2nd processings into shorter forms. Note as well that the leader-proximal crRNA was the most abundant species. The third form corresponds to the remaining ~24 nt repeat-spacer, presumably an inactive species. This form, which should be generated from the 2nd processing event, was neither observed in the dRNA-seq data nor by northern blot analysis (see right panel), most probably because of instability or rapid degradation. **c,** tracrRNA directs the maturation of pre-crRNA. Northern blot analysis of crRNA expression in *S. pyogenes* cultures at different time points during growth (refer to b). Strains: wild-type (WT) (SF370), ΔtracrRNA, ΔtracrRNA + 89 tracrRNA (ΔtracrRNA complemented with 89 nt tracrRNA form) and ΔtracrRNA + 171 tracrRNA (ΔtracrRNA complemented with 171 nt tracrRNA form). Probe: crRNA repeat* (OLEC1049). Processing of pre-crRNA into mature crRNA forms (39-42 nt) is abrogated in ΔtracrRNA. Complementing ΔtracrRNA *in trans* with the 171 or 89 nt tracrRNA form restores the processing. Co-processing of tracrRNA and pre-crRNA occurs constitutively throughout growth (refer to Supplementary Fig. 4c).

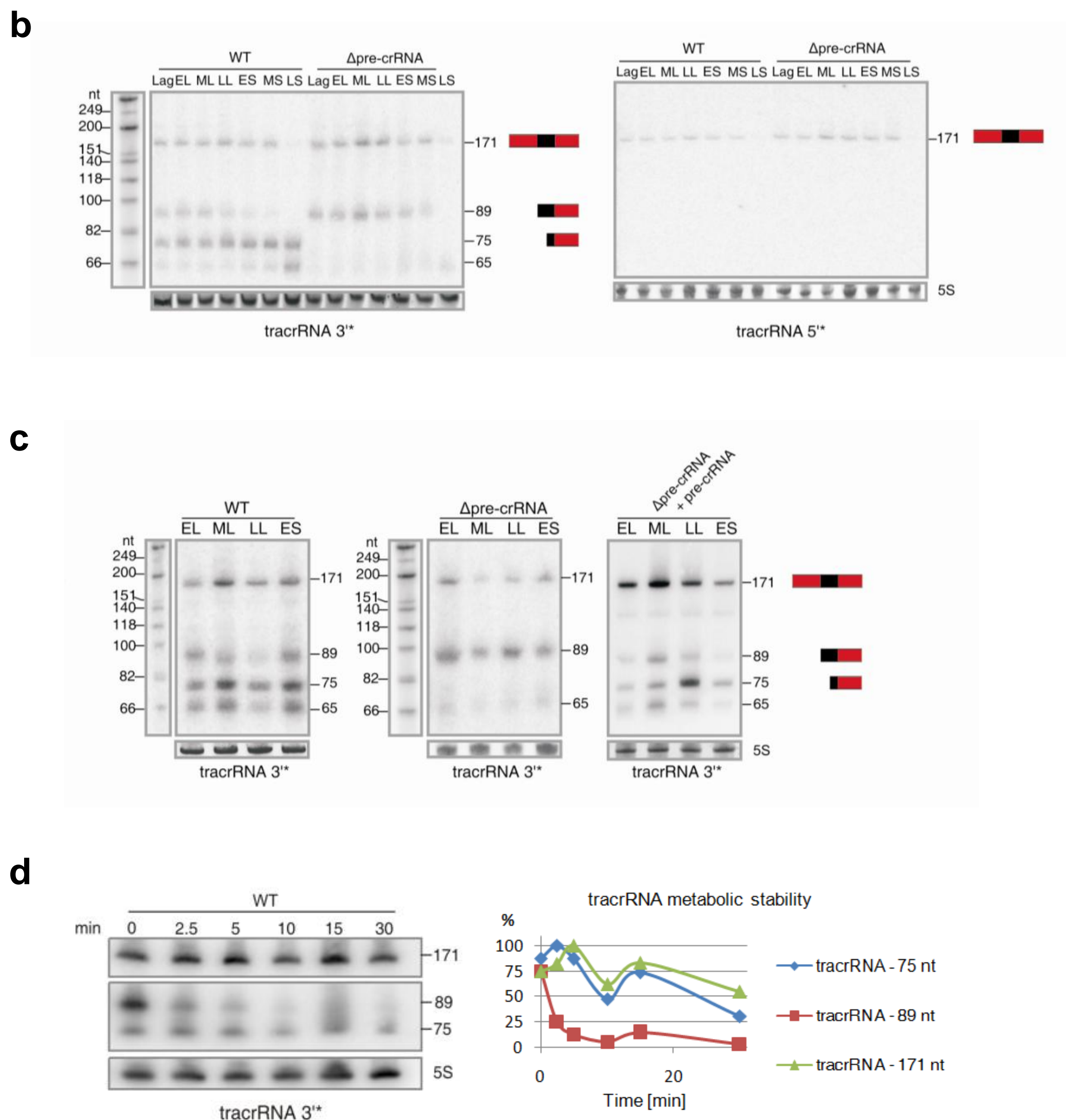


Supplementary Figure 3 | RT-PCR analysis of tracrRNA, CRISPR01 crRNA and CRISPR02 crRNA expression in *S. pyogenes* SF370. Total RNA from *S. pyogenes* cultures grown until mid-logarithmic phase was used for reverse-transcription followed by PCR using primers specific for tracrRNA (OLEC1299+OLEC1522), CRISPR01 crRNA (OLEC1140+OLEC1141), CRISPR02 crRNA (OLEC1219+OLEC1220) and tmRNA (OLEC140+OLEC449) (Supplementary Table 10). PCR-generated fragments were analyzed by agarose gel electrophoresis. RT-PCR analysis reveals tracrRNA expression in the wild-type (WT) (SF370) strain and Δpre-crRNA (CRISPR01) deletion mutant, and CRISPR01 pre-crRNA expression in the wild-type (WT) (SF370) strain and ΔtracrRNA mutant. However, no expression of CRISPR02 crRNA in WT, ΔtracrRNA and Δpre-crRNA (CRISPR01) strains could be detected. Additionally, no CRISPR02 pre-crRNA or crRNA expression in *S. pyogenes* WT could be observed by northern blot analysis (data not shown). These data do not rule out the possibility that CRISPR02 crRNA could be expressed and active at very low level under the conditions tested or under specific growth conditions, which remain to be determined.

a



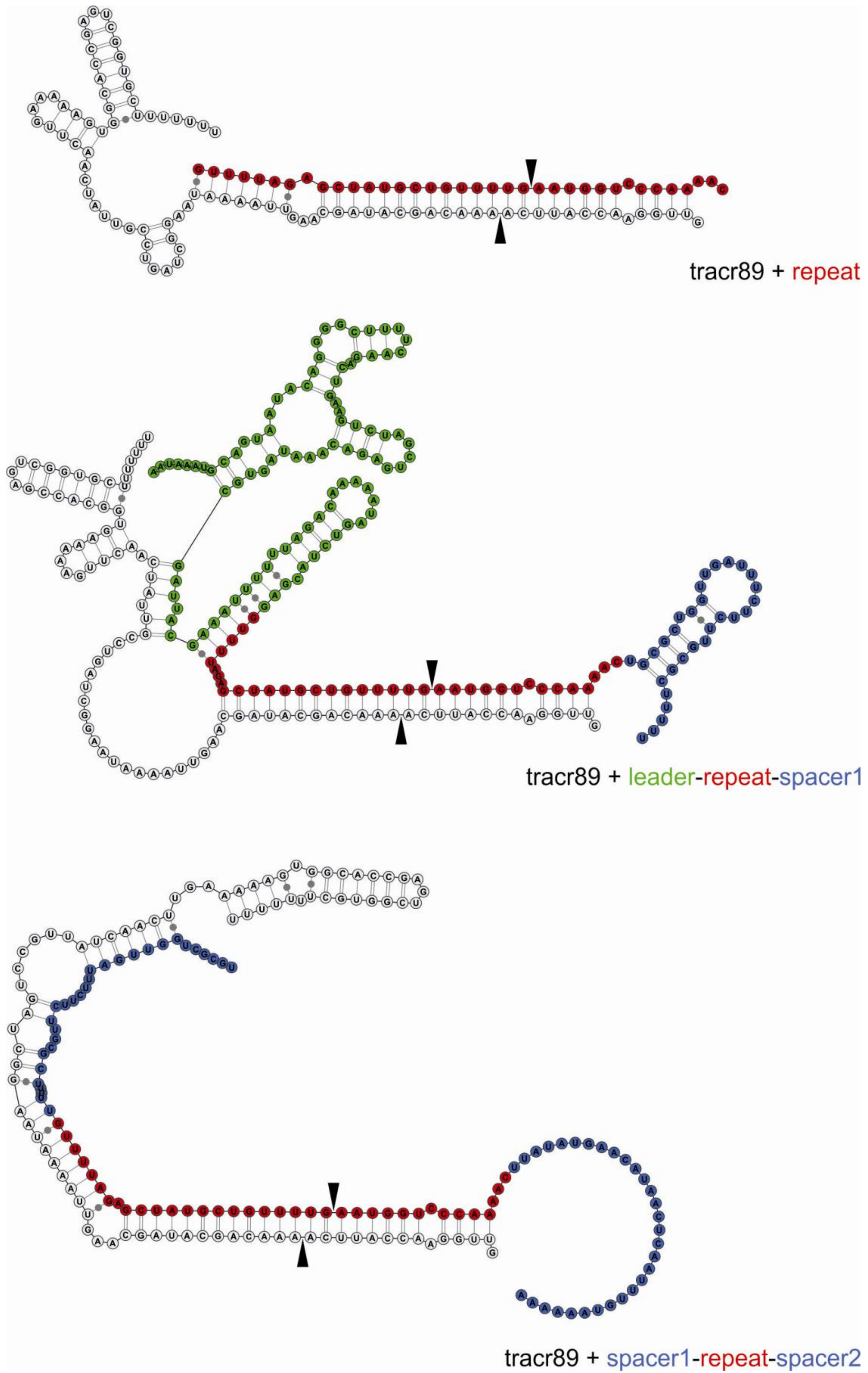

```
SF370-M1         TCCGAAATTAGTTTAATATGCTTTAATTTTTCTTTTTCAAAATATCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATT
MGAS5005-M1      TCCGAAATTAGTTTAATATGCTTTAATTTTTCTTTTTCAAAATATCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATT
MGAS10270-M2     TCCGAAATTAGTTTAATATGCTTTAATTTTTCTTTTTCAAAATATCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATT
MGAS10750-M4     TCCGAAATTAGTTTAATATGCTTTAATTTTTCTTTTTCAAAATATCTCTTCAAAAAATATTACCTAATACTTAATAATGAATAAATTATAACACAAAATC
MGAS2096-M12     TCCGAAATTGGTTTAATATGCTTTAATTTTTCTTTTTCAAAATATCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATT
MGAS9429-M12     TCCGAAATTGGTTTAATATGCTTTAATTTTTCTTTTTCAAAATATCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATT
MGAS6180-M28     TCCGAAATTGGTTTAATATGCTTTAATTTTTCTTTTTCAAAATATCTCTTCAAAAAATATTACCCAATACTTAATAATAAATAGATTATAACACAAAATT
MGAS315-M3       TCCGAAATTAGTTTAATATGCTTTAATTTTTCTTT-----------------------------------------------------------------
SSI1-M3          TCCGAAATTAGTTTAATATGCTTTAATTTTTCTTT-----------------------------------------------------------------
Manfredo-M5      ----------------------------------------------------------------------------------------------------
MGAS10394-M6     ----------------------------------------------------------------------------------------------------
MGAS8232-M18     ----------------------------------------------------------------------------------------------------


SF370-M1         CTTTTAAAAAGTAGTTTATTTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAA
MGAS5005-M1      CTTTTAAAAAGTAGTTTATTTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAA
MGAS10270-M2     CTTTTAAAAAGTAGTTTATTTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAA
MGAS10750-M4     CTTTTAAAAAGTAGTTTATTTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAA
MGAS2096-M12     CTTTTAAAAAGTAGTTTATTTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAA
MGAS9429-M12     CTTTTAAAAAGTAGTTTATTTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAA
MGAS6180-M28     CTTTTAAAAAGTAGTTTATTTTGTTATCATTCTATAGTATTAAGTATTGTTTTATGGCTGATAAATTTCTTTGAATTTCTCCTTGATTATTTGTTATAAA
MGAS315-M3       ----------------------------------------------------------------------------------------------------
SSI1-M3          ----------------------------------------------------------------------------------------------------
Manfredo-M5      ----------------------------------------------------------------------------------------------------
MGAS10394-M6     ----------------------------------------------------------------------------------------------------
MGAS8232-M18     ----------------------------------------------------------------------------------------------------


                                                                                            >>>>>>>>            <<<<<
SF370-M1         AGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS5005-M1      AGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS10270-M2     AGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS10750-M4     AGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS2096-M12     AGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS9429-M12     AGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS6180-M28     AGTTATAAAATAATCTTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS315-M3       ---------------------------TTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
SSI1-M3          ---------------------------TTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
Manfredo-M5      --------------------GAGACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATTAACTTGAAAAAGTGGCACCGATTCGGTGCT
MGAS10394-M6     ------------------TGGAGACCATTCAAAACAGCATAGCCAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT
MGAS8232-M18     ------------------TTGGGACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCT


                 <<<<<<
SF370-M1         TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACCAAAAAAACAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
MGAS5005-M1      TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACCAAAAAAACAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
MGAS10270-M2     TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACTAAAAAA-CAAGCGCTTTCAAAACGCTTGTTCTATCATTTTTAGGGAAATTAATCTCTTAATCC
MGAS10750-M4     TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACTAAAAAA-CAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
MGAS2096-M12     TTTTTTTGATACTTCTATTCTACTCTGACTGCAAACTAAAAAA-CAAGCGCTTTCAAAACGCTTGTTTTATCATT---------------------------
MGAS9429-M12     TTTTTTTGATACTTCTATTCTACTCTGACTGCAAACTAAAAAA-CAAGCGCTTTCAAAACGCTTGTTTTATCATT---------------------------
MGAS6180-M28     TTTTTTTGATACTTCTATTCTACTCTGACTGCAAACTAAAAAA-CAAGCGCTTTCAAAACGCTTGTTTTATCATT---------------------------
MGAS315-M3       TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACT-AAAAAACAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
SSI1-M3          TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACT-AAAAAACAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
Manfredo-M5      TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACT-AAAAAACAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
MGAS10394-M6     TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACCCAAAAAACAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
MGAS8232-M18     TTTTTT-GATACTTCTATTCTACTCTGACTGCAAACC-GAAAAACAAGCGCTTTCAAAACGCTTGTTTTATCATTTTTAGGGAAATTAATCTCTTAATCC
```



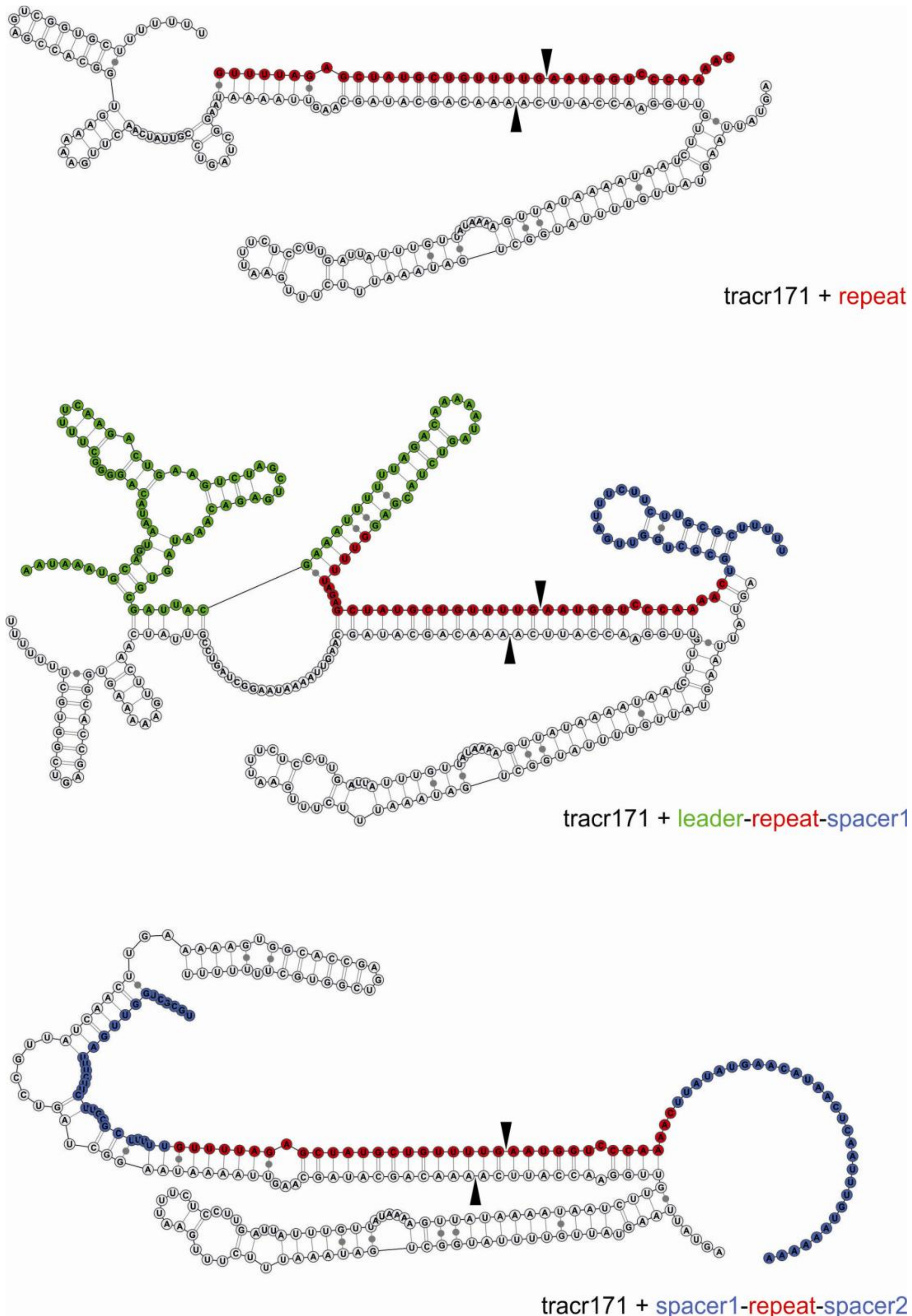
Supplementary Figure 4 | Conservation and expression of tracrRNA in *S. pyogenes*. **a,** Sequence alignment of tracrRNA loci from *S. pyogenes* genomes with available complete sequence. The sequence is given from 5' (coordinate 854677) to 3' (coordinate 854279) (minus strand of the SF370 genome; Accession number: NC_002737). Blue, -35 and -10 promoter regions; green and facing arrows, Rho-independent transcriptional terminator; red, alternative 5' ends as determined by primer extension and head-to-tail circularization (data not shown); grey, mismatches. Full-length tracrRNA is present and conserved among *S. pyogenes* genomes that contain the CRISPR01-associated genes and CRISPR01 leader-repeat-spacer array (Supplementary Fig. 1; Supplementary Table 1). In the MGAS10750 genome, which only contains the CRISPR01-associated genes, full-length tracrRNA-encoding DNA is also present (Supplementary Fig. 1; Supplementary Table 1). In the SSI-1 and MGAS315 genomes, which also only contain the CRISPR01-

associated genes and in the Manfredo, MGAS8232 and MGAS10394 genomes, which do not contain both the CRISPR01-associated genes and CRISPR01 leader-repeat-spacer array, the tracrRNA locus is truncated in its 5' end (Supplementary Fig. 1; Supplementary Table 1) and thus lacks the promoter region. This indicates a close interconnection between the presence of tracrRNA and a functional CRISPR01 locus. **b,** tracrRNA is observed as four species expressed throughout growth. Northern blot analysis of tracrRNA expression in *S. pyogenes* cultures at different time points during growth (refer to Supplementary Fig. 2b). Strains: wild-type (WT) (SF370) and ∆pre-crRNA. Probes: tracrRNA 5'* (OLEC1698), tracrRNA 3'* (OLEC1697 or OLEC1014) and 5S rRNA (OLEC288) as loading control (Supplementary Table 10). tracrRNA is an abundant small non-coding RNA that is observed as four species: 171, 89, ~75 and 65 nt. The ~75 nt band corresponds to a processed form generated from the longer tracrRNA forms (171 and 89 nt) (Fig. 1a,b). **c,** pre-crRNA is essential for the processing of tracrRNA. Northern blot analysis of tracrRNA expression in *S. pyogenes* cultures at different time points during growth (refer to b). Strains: wild-type (WT) (SF370), ∆pre-crRNA and ∆pre-crRNA + pre-crRNA (∆pre-crRNA complemented with pre-crRNA). Probe: tracrRNA 3'* (OLEC1014) (Supplementary Table 10). Processing of tracrRNA is abrogated in ∆pre-crRNA. Complementing ∆pre-crRNA *in trans* with pre-crRNA restores the processing. Co-processing of tracrRNA and pre-crRNA occurs constitutively throughout growth (refer to Supplementary Fig. 2c). **d,** Metabolic stability of tracrRNA. *S. pyogenes* wild-type (WT) (SF370) cultures at mid-logarithmic phase of growth were treated with rifampicin for the indicated times to stop transcription. Total RNA was then prepared and analyzed by northern blot analysis. Probes: tracrRNA 3'* (OLEC1697) and 5S rRNA (OLEC288) as loading control (Supplementary Table 10). The half-life is defined as the point at which 50% of the RNA was degraded. y axis, relative amount of tracrRNA; x axis, time (min.). Data are representative of three independent experiments. Note that the 89 nt form is the least stable, suggesting its rapid processing. Together with the signal intensities of 171, 89 and ~75 nt bands observed on the northern blot (refer to b), the 89 nt form seems to be preferentially processed compared to the 171 nt form.

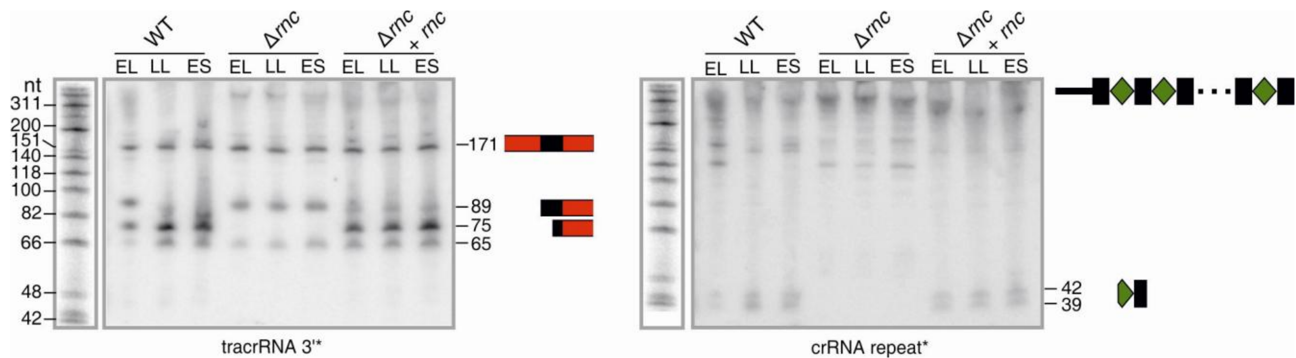
a



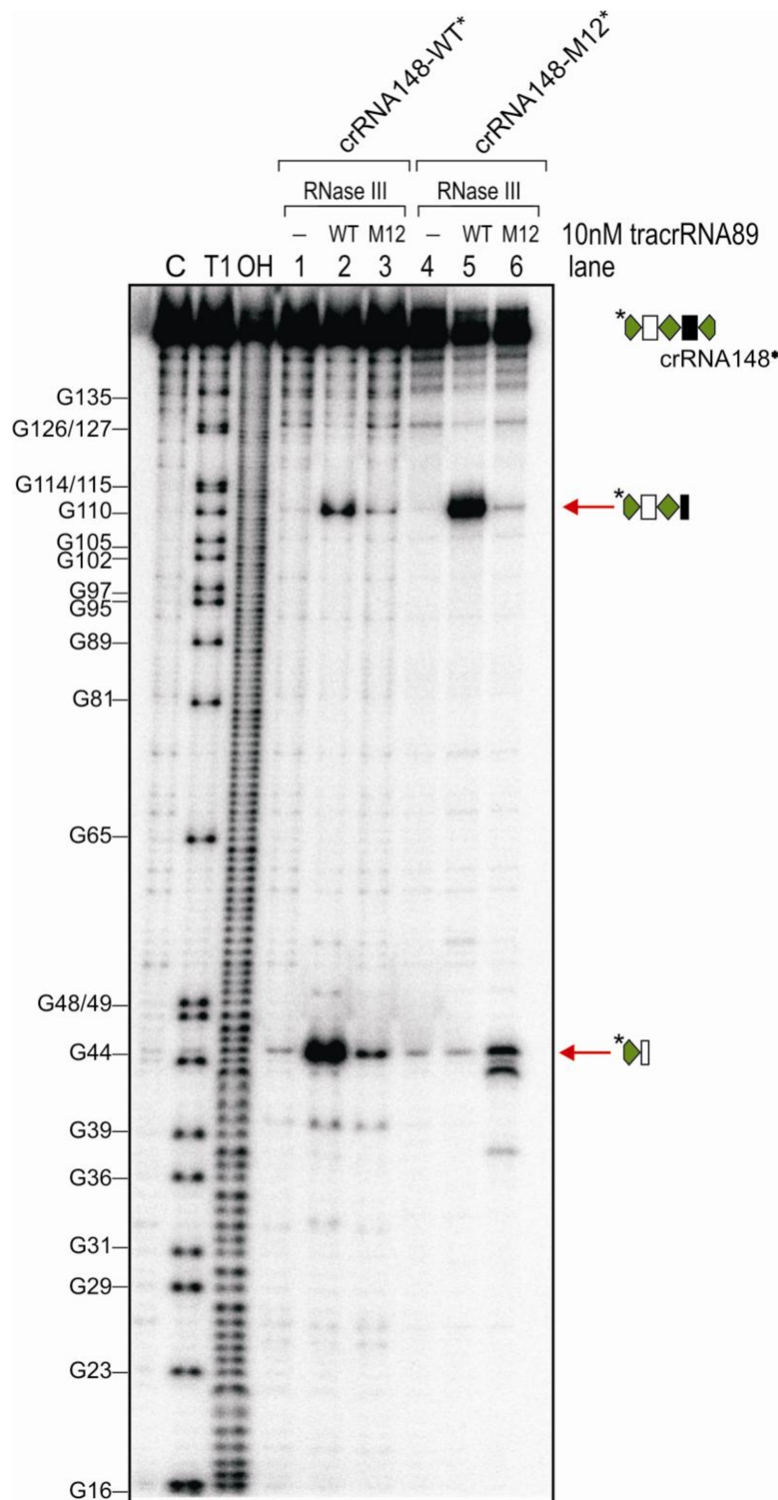
b



Supplementary Figure 5 | *In silico* co-foldings of tracrRNA and crRNA species. Both 171 nt and 89 nt tracrRNA species basepair with the CRISPR01 repeat sequence regardless of the surrounding regions (leader or spacers), thus forming a duplex RNA. Cleavage sites identified by the *in vivo* dRNA-seq analysis are indicated with arrows. The generation of tracrRNA and crRNA processed forms with short overhangs at the 3' ends suggests the involvement of RNase III to dice the RNAs.



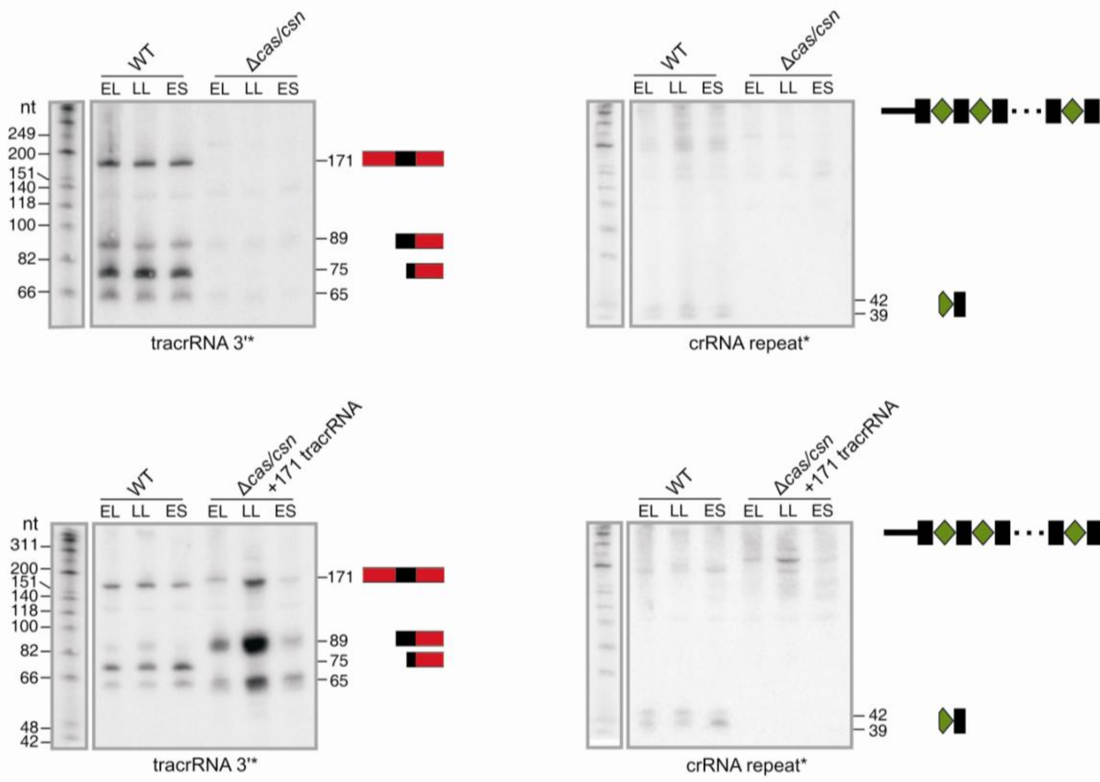
Supplementary Figure 6 | Role of endogenous RNase III in tracrRNA-mediated maturation of pre-crRNA *in vivo*. Northern blot analyses of tracrRNA (left panel) and crRNA (right panel) expression in *S. pyogenes* cultures at different time points during growth (refer to Supplementary Fig. 2b). Strains: wild-type (WT) (SF370), Δ*rnc*, Δ*rnc* + *rnc* (Δ*rnc* complemented with *rnc*). The gene *rnc* encoding RNase III is the first gene of an operon consisting of *rnc*-*smc* (*smc*, encoding the SMC protein involved in structural maintenance of chromosome) (RT-PCR analysis, data not shown). Note that the control WTts strain (not shown here, refer to Supplementary Methods) displayed the same phenotype as the WT strain. Probes: tracrRNA 3'* (OLEC1014) and crRNA repeat* (OLEC1049). Processing of both tracrRNA into ~75 nt species and pre-crRNA into mature crRNA forms (39-42 nt) is abolished in Δ*rnc*. Complementing Δ*rnc in trans* with *rnc* expressed from its own promoter restores the processing events. This indicates that the endogenous RNase III is essential for the maturation of both tracrRNA and pre-crRNA.



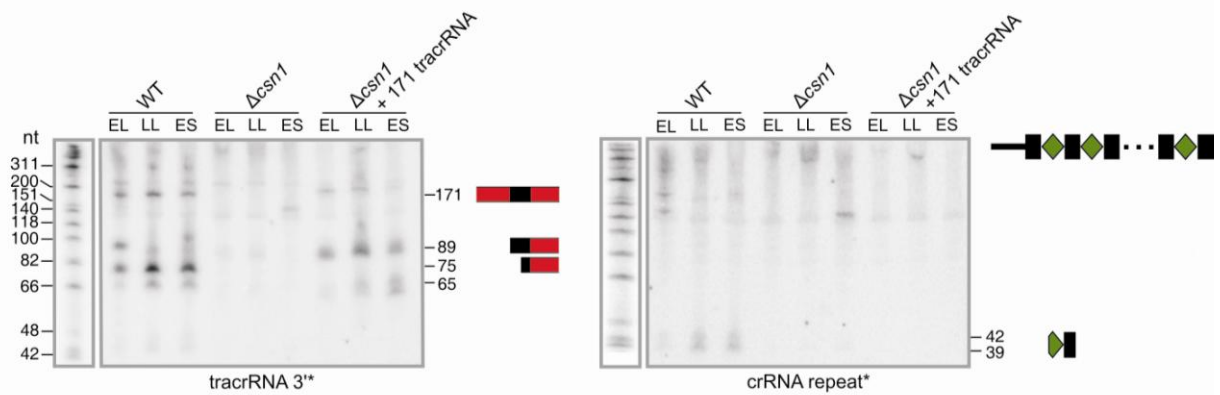
Supplementary Figure 7 | Compensatory base-pair exchanges confirm tracrRNA-mediated crRNA maturation *in vitro*. Identification of tracrRNA binding sites on a 148-nt long fragment of pre-crRNA (crRNA148*). 5' end-labeled crRNA148* wild-type (WT) or mutated (M12) (~10 nM) was subjected to RNase III cleavage in absence (lanes 1, 4) and presence of cold tracrRNA89 (~10 nM) (WT, lanes 2, 5; compensatory mutations to crRNA148 mutated (M12), lanes 3, 6). A stretch of mutations upstream of the RNase III cleavage site in the first repeat (white box) of crRNA148 (crRNA148-M12; 5'-GCUAUGCUGUUU-3' → 5'-CGAUACGACAAA-3') and cognate compensatory mutations

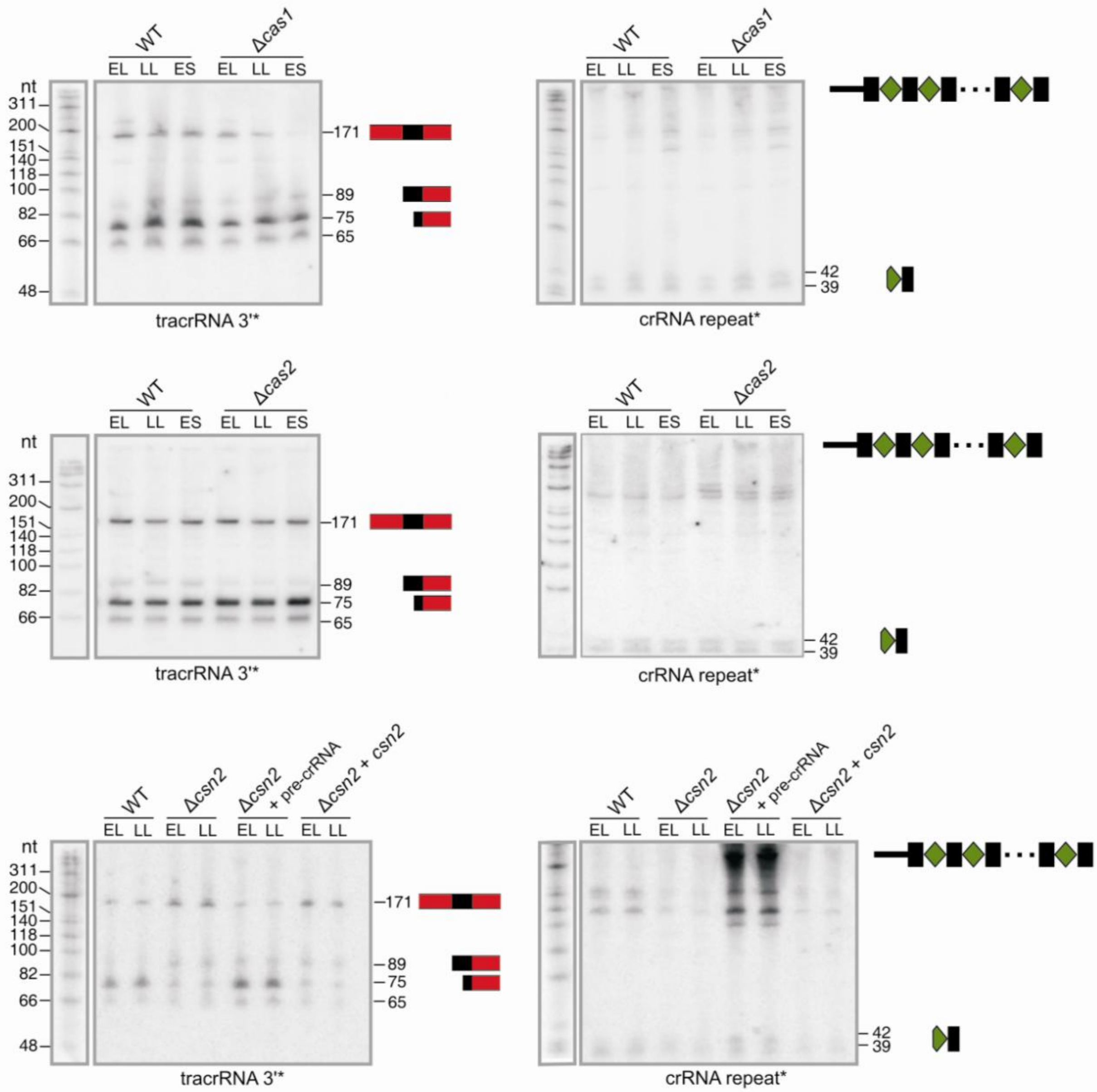
in tracrRNA89 (tracrRNA89-M12; 5’-AAACAGCAUAGC-3’ → 5’ UUUGUCGUAUCG-3’) were introduced. Lane C: untreated crRNA148*; Lane T1: RNase T1 digest of crRNA148* under denaturating conditions; Lane OH: alkaline ladder. Cleaved G residues are labeled. Vertical bars: crRNA148 region protected by tracrRNA89. Arrows denote specific RNase III cleavages in the two repeat regions of crRNA148 in the presence of tracrRNA89. tracrRNA89 directs the processing of crRNA148. Mutations in tracrRNA89 hinder tracrRNA89-directed cleavage of crRNA148 (lane 3). The second repeat of crRNA in crRNA148-M12, which corresponds to the WT repeat, is cleaved upon addition of tracrRNA WT (lane 5), whereas the first repeat with the twelve mutated nucleotides is not cleaved anymore. However, cleavage in the mutated first repeat is fully restored when the compensatory tracrRNA89 and crRNA148-M12 are paired (lane 6).

a

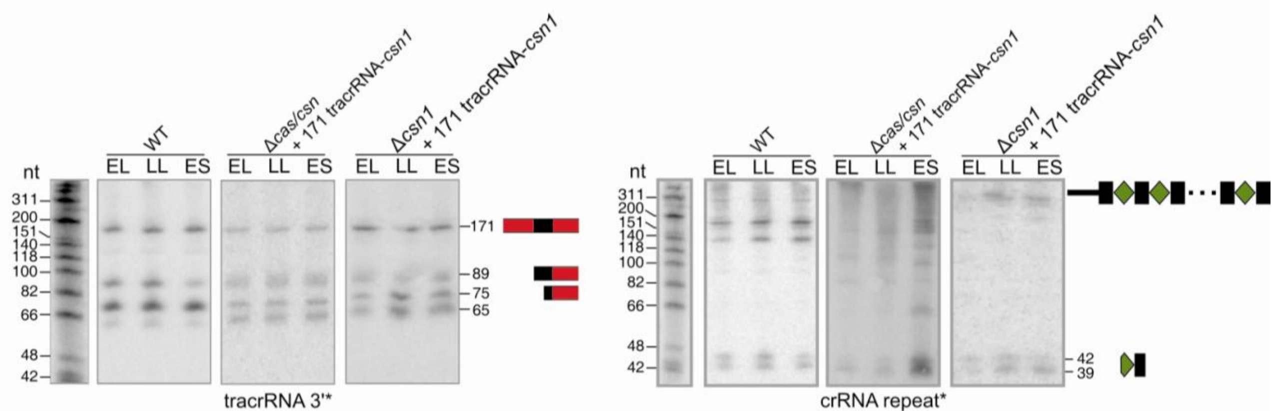


b

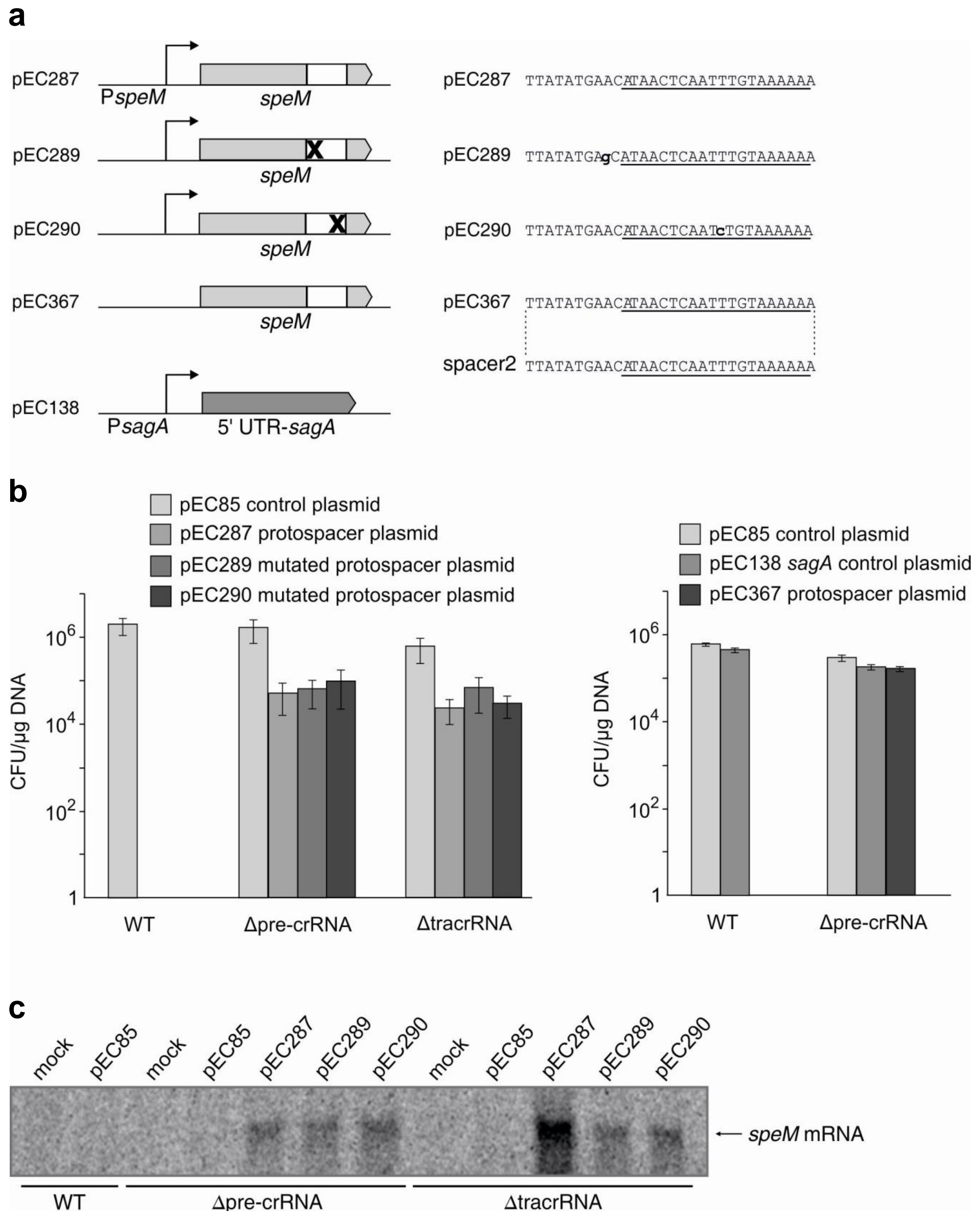



c
WT
Δcas1
EL LL ES EL LL ES
nt
311
200
151
140
118
100
82
66
48
171
89
75
65
tracrRNA 3'*
42
39
crRNA repeat*
Δcas2
Δcsn2
Δcsn2 + pre-crRNA
Δcsn2 + csn2
EL LL

d

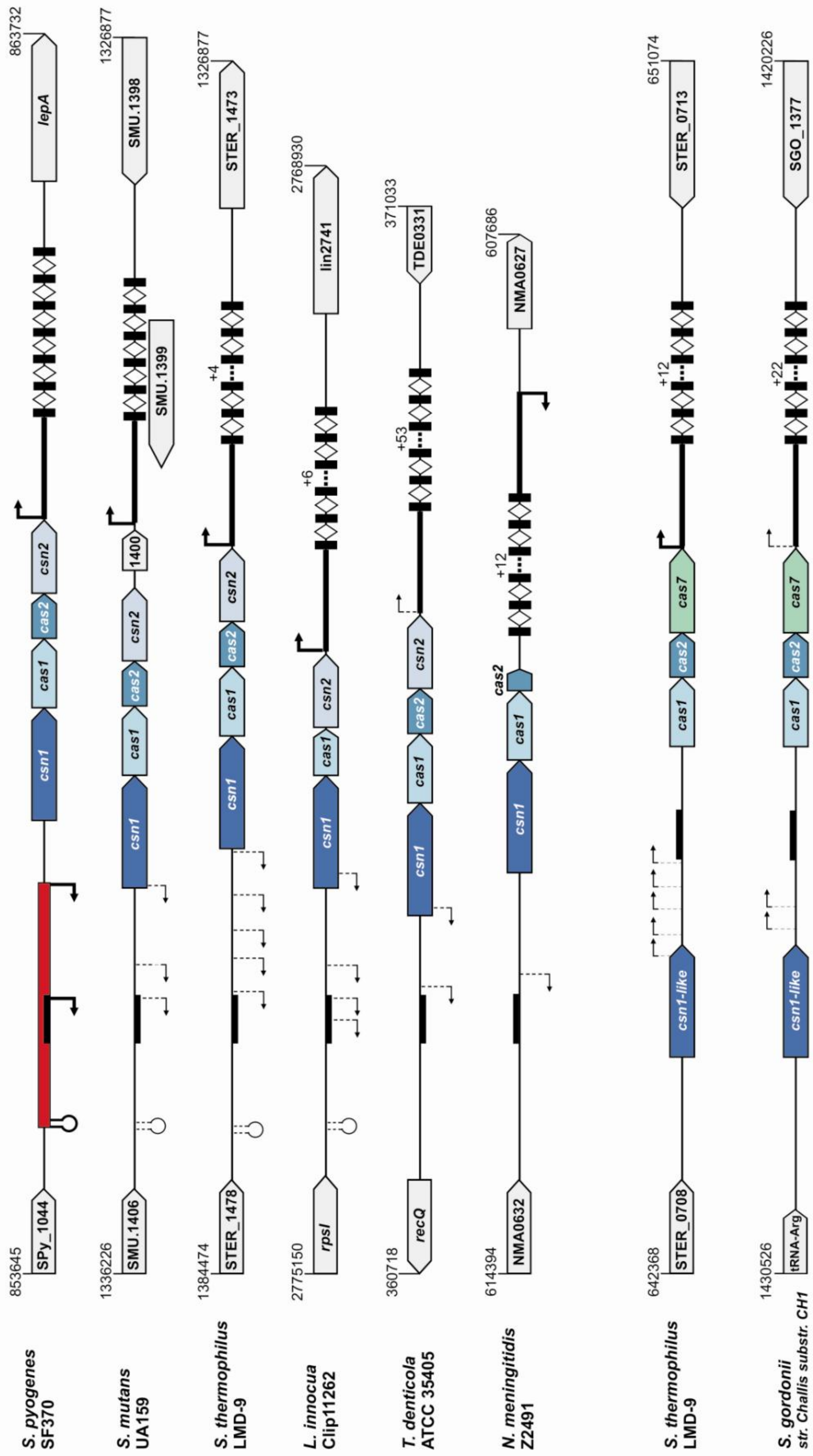


Supplementary Figure 8 | Role of Cas proteins in the processing of tracrRNA and pre-crRNA *in vivo*. Northern blot analysis of tracrRNA (left panel) and crRNA (right panel) expression in *S. pyogenes* cultures at different time points during growth (refer to Supplementary Fig. 2b). Strains: wild-type (WT) (SF370), Δ*cas/csn* is Δ*csn1-cas1-cas2-csn2*, Δ*cas/csn* + 171 tracrRNA, Δ*cas/csn* + 171 tracrRNA-*csn1*, Δ*cas/csn* + *cas1*, Δ*cas/csn* + *cas2*, Δ*cas/csn* + *csn2*, Δ*csn1*, Δ*csn1* + 171 tracrRNA, Δ*csn1* + 171 tracrRNA-*csn1*, Δ*cas1*, Δ*cas2*, Δ*csn2*, Δ*csn2* + pre-crRNA (Δ*csn2* expressing pre-crRNA *in trans*), and Δ*csn2* + *csn2*. Complementation with *csn* or *cas* was done with gene expression under the control of the *cas/csn* promoter (P*csn1*). The *csn1-cas1-cas2-csn2* genes consist of an operon that is expressed as a single mRNA from a promoter located upstream of *csn1* (P*csn1*) (RT-PCR analysis, data not shown). Note that the phenotypes of the control WTts strains (not shown here, refer to Supplementary Methods) were identical to that of the WT strain. Probes: tracrRNA 3'* (OLEC1014) and crRNA repeat* (OLEC1049). **a,b,** Csn1 is required for the processing of tracrRNA and pre-crRNA. Compared to WT, processing of both tracrRNA into ~75 nt species and pre-crRNA into mature crRNA forms (39-42 nt) is abrogated in the Δ*cas/csn* **(a)** and Δ*csn1* mutants **(b)**. Note that tracrRNA is less abundant in the Δ*cas/csn* **(a)** and Δ*csn1* **(b)** mutants. We argue that Csn1 is likely to be required for binding to and/or stabilization of tracrRNA. To confirm that deletion of *cas/csn* or *csn1* affects the processing events, we introduced *in trans* the 171 nt tracrRNA form in both Δ*cas/csn* **(a)** and Δ*csn1* **(b)** mutants. As expected, northern blot analysis in these strains confirmed the deficiency in RNA processing of the mutants. **c,** Cas1, Cas2 and Csn2 do not seem to be involved in tracrRNA-mediated maturation of pre-crRNA. Processing of both tracrRNA and pre-crRNA is not affected in Δ*cas1*, Δ*cas2* and Δ*csn2* mutants. Accordingly, complementing the Δ*cas/csn* mutant with *cas1*, *cas2* or *csn2 in trans* did not rescue the loss-of-function phenotype observed in the mutant. Because expression of pre-crRNA was low in the Δ*csn2* mutant, we complemented Δ*csn2* with *csn2* (pEC307) *in trans* (strain Δ*csn2* + *csn2*) and Δ*csn2* expressing pre-crRNA (pEC299) *in trans* (strain Δ*csn2* + pre-crRNA). Northern blot analysis of these strains confirms that the RNA processing events are not affected in Δ*csn2*. **d,** In both Δ*cas/csn* and Δ*csn1* mutants, complementation with *csn1 in trans* restores the processing events.



Supplementary Figure 9 | tracrRNA, pre-crRNA, RNase III and Csn1 confer immunity against acquisition of a lysogenic phage-protospacer gene. In *S. pyogenes*, temperate phages are vectors for the horizontal transfer of genes encoding virulence

factors such as the superantigens, phospholipases and streptodornases, which play key roles during *S. pyogenes* infection of humans [1]. Here, we analyzed the role of CRISPR01 in the immunity against acquisition of the protospacer gene, *speM* (encoding superantigen; targeted by spacer 2, Spyo1h_002 (Supplementary Table 2)). **a,** Plasmids used to transform *S. pyogenes* (Supplementary Table 9). Left panel: *speM*, superantigen-encoding gene (protospacer gene targeted by CRISPR01, source: MGAS8232 (M18)); *sagA*, P*sagA*-5'UTR*sagA* (not targeted by CRISPR01, acting as negative control); white, protospacer; cross, mutation. Right panel: WT and mutated *speM* protospacer sequences; underlined, sequence corresponding to spacer 2 (*speM*) in processed crRNA; bold lowercases, mutations. **b,** Transformation efficiencies of *S. pyogenes* with protospacer containing plasmids. Left panel**:** Analysis of mismatch mutations in the protospacer. A mismatch mutation in the 5' end of the protospacer that is not recognized by the mature 42 nt crRNA (pEC289) and a mistmatch mutation in the 3' end of the protospacer that is recognized by the mature 42 nt crRNA (pEC290) were analyzed. WT, ΔtracrRNA, Δpre-crRNA strains were transformed with pEC289 (pEC85ΩP*speM*-*speM* (A2773G)) or pEC290 (pEC85ΩP*speM*-*speM* (T2785C)). Both plasmids were tolerated by the ΔtracrRNA and Δpre-crRNA mutants. Each mutation was not sufficient to abolish immunity of the WT strain to DNA entry. This is consistent with the observation that sequence homology between mobile DNA protospacers and CRISPR01 spacers is not always 100% identical (Supplementary Table 3) and indicates that CRISPR01 evolved to allow plasticity in target recognition. Right panel: Plasmid pEC85 (backbone vector) was used as reference. Note that the transformation efficiencies with pEC287, pEC289, pEC290 and pEC367 were lower than that with pEC85 (left panel and Fig. 5). An additional control, pEC138 (Supplementary Table 9), containing an insert that is not targeted by CRISPR01 but of length similar as that of pEC367, was used. As expected, pEC138 was tolerated by both WT and Δpre-crRNA strains. The transformation efficiency with pEC138 was similar to that with pEC367 and lower than that with pEC85. Thus in our experimental set-up, transformation efficiency seems sensitive to increasing plasmid size. Plasmid pEC367 containing the *speM* gene without its promoter was not tolerated by wild-type (WT) (SF370) but was maintained in the Δpre-crRNA mutant, similarly to pEC287, indicating that protospacer transcription is not required for CRISPR01 interference. Graph bars represent the mean values of colony forming units (CFU) per µg of plasmid DNA. Error bars represent the standard deviation (SD), n≥3. **c,** The *speM* protospacer is expressed in ΔtracrRNA and Δpre-crRNA transformed with pEC287. Expression of *speM* in the indicated transformed strains was confirmed by northern blot analysis using a *speM*-specific probe (OLEC1558) (Supplementary Table 10).

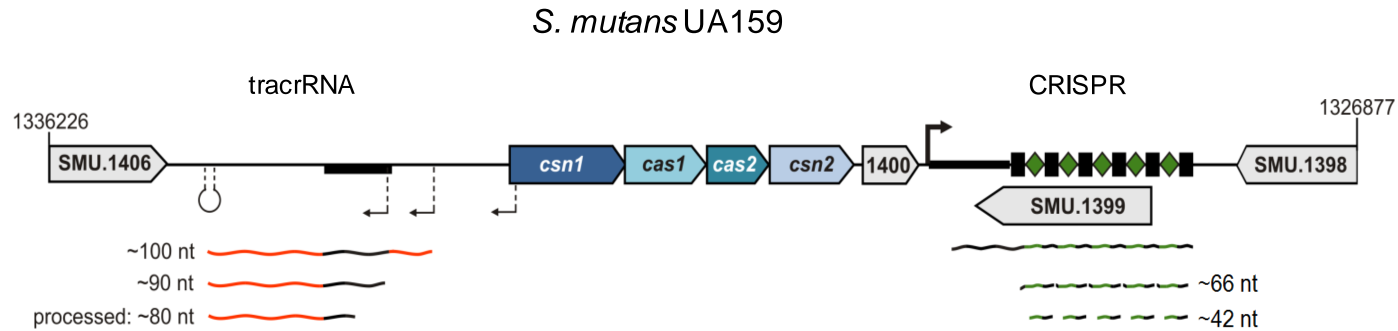

S. pyogenes
SF370
853645
SPy_1044
csn1
cas1
cas2
csn2
lepA
863732
S. mutans
UA159
1336226
SMU.1406
csn1
cas1
cas2
csn2
1400
SMU.1399
SMU.1398
1326877
S. thermophilus
LMD-9
1384474
STER_1478
csn1
cas1
cas2
csn2
+4
STER_1473
1326877
L. innocua
Clip11262
2775150
rpsI
csn1
cas1
csn2
+6
lin2741
2768930
T. denticola
ATCC 35405
360718
recQ
csn1
cas1
cas2
csn2
+53
TDE0331
371033
N. meningitidis
Z2491
614394
NMA0632
csn1
cas1
cas2
+12
NMA0627
607686
S. thermophilus
LMD-9
642368
STER_0708
csn1-like
cas1
cas2
cas7
+12
STER_0713
651074
S. gordonii
str. Challis substr. CH1
1430526
tRNA-Arg
csn1-like
cas1
cas2
cas7
+22
SGO_1377
1420226

Supplementary Figure 10 | Genomic organization of type II (Nmeni/CASS4) CRISPR/Cas systems from selected bacterial species. The type II CRISPR/Cas loci from *L. innocua* Clip11262, *N. meningitidis* Z2491, *S. gordonii str. Challis substr.* CH1, *S. mutans* UA159, *S. pyogenes* SF370, *S. thermophilus* LMD-9 and *T. denticola* ATCC 35405 are represented. The DNA region encoding tracrRNA in *S. pyogenes* SF370 is represented in red. The black bar within each tracrRNA represents the sequence complementary to each CRISPR repeat (Supplementary Table 7). The CRISPR-associated genes (*csn1*-*cas1*-*cas2*-*csn2* or *cas7*) are followed by the CRISPR leader sequence and repeat-spacer region. Rectangles, CRISPR repeats; diamonds, CRISPR spacers. A DNA region encoding a putative small RNA (tracrRNA homologue) is found upstream of all represented CRISPR loci. Validated (plain lines) and predicted (dashed lines) transcription start sites and Rho-independent transcription terminators are indicated. Intriguingly, in the *S. thermophilus* LMD-9 (2nd CRISPR locus) and *S. gordonii* CH1 genomes, the tracrRNA homologue is located within the *csn1*-*cas1* intergenic region and transcribed in the same direction as the *cas*/*csn* operon. Furthermore, in *N. meningitidis*, pre-crRNA is encoded on the opposite strand of the *cas* genes.

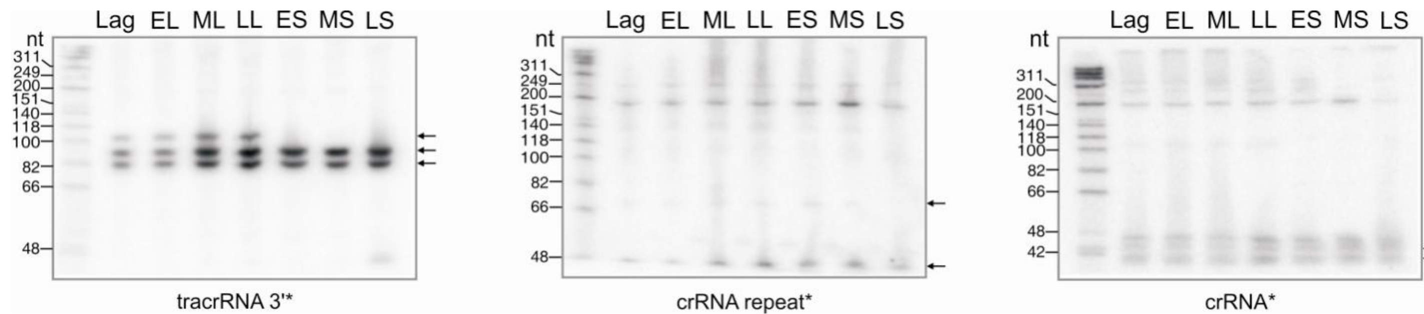
```
  Spy     (1) AGUAUUAAGUAUUGUUUUAUGGCUGAUAAAUUUCUUUGAAUUUCUCCUUGAUUAUUUGUUAUAAAAGUUA
  Lin     (1) ----------------------------------------------------------------------
  Smu     (1) ----------------------------------------------------------------------
 Sth1     (1) ----------------------------------------------------------------------
  Nme     (1) ----------------------------------------------------------------------
 Sth2     (1) -----------------------------------------------------------UAAUAAUAGUG

                         Anti-repeat region
  Spy    (71) UAAAAUAAUCUUGUUGGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGC-------U-------AG
  Lin     (1) ----------AUAUUGUUAGUAUUCAAAAUAACAUAGCAAGUUAAAAUAAGGC-----UUU--------G
  Smu     (1) -----------UGUUGGAAUCAUUCGAAACAACACAGCAAGUUAAAAUAAGGCAGUGAUUUUUAAUCCAG
 Sth1     (1) ------UUGUGGUUUGAAACCAUUCGAAACAACACAGCGAGUUAAAAUAAGGC-----UU--------AG
  Nme     (1) -------CAUAUUGUCGCACUGCGAAAUGAGAACCGUUGCUACAAUAAGGCCGUC----UGAAAAGAUGU
 Sth2    (12) UAAGGGACGCCUUACACAGUUACUUAAAUCUUGCAGAAGCUACAAAGAUAAGGC-----UU------CAU

                                                         STOP
  Spy   (127) UCCGUUAUCAACU-----UGAAA--AAGUG-GCACCGAGUCGGUGCUUUUUUU
  Lin    (48) UCCGUUAUCAACU-----UUUAAUUAAGUA-GCGCUGUUUCGGCGCUUUUUUU
  Smu    (60) UCCGUACACAACU-----UGAAA--AAGUGCGCACCGAUUCGGUGCUUUUUUA
 Sth1    (52) UCCGUACUCAACU-----UGAAA--AGGUG-GCACCGAUUCGGUGUUUUUUUU
  Nme    (60) GCCGCAACGCUCUGCCCCUUAAAGCUUCUGCUUUAAGGGGCAUCGUUUAUUUC
 Sth2    (71) GCCGAAAUCAACACCC--UGUCAUUUUAUG-GCAGGGUGUUUUCGUUAUUUAA
```

Supplementary Figure 11 | Alignment of predicted tracrRNA homologues from selected bacterial strains. Spy, *S. pyogenes* SF370; Lin, *L. innocua* Clip11262, Smu, *S. mutans* UA159; Sth1, S*. thermophilus* LMD-9 1st locus; Sth2, *S. thermophilus* LMD-9 2nd locus; Nme, *N. meningitidis* Z2491. Putative promoters and transcriptional terminators were predicted bioinformatically. The arrow indicates the transcriptional start site of the 89-nt long tracrRNA form associated with *S. pyogenes* CRISPR01. The lengths of the predicted RNA species correspond to those of tracrRNA homologues observed by northern blot analysis (Supplementary Figs 12-16). Sequences were aligned using AlignX (Invitrogen). Black-highlighted, perfect homology; grey-highlighted, imperfect homology. tracrRNA homologues show limited but significant homology between the examined species. The highest degree of similarity is observed among tracrRNAs of *S. pyogenes*, *S. mutans*, *S. thermophilus* (1st CRISPR locus) and *L. innocua*.

a



b



c

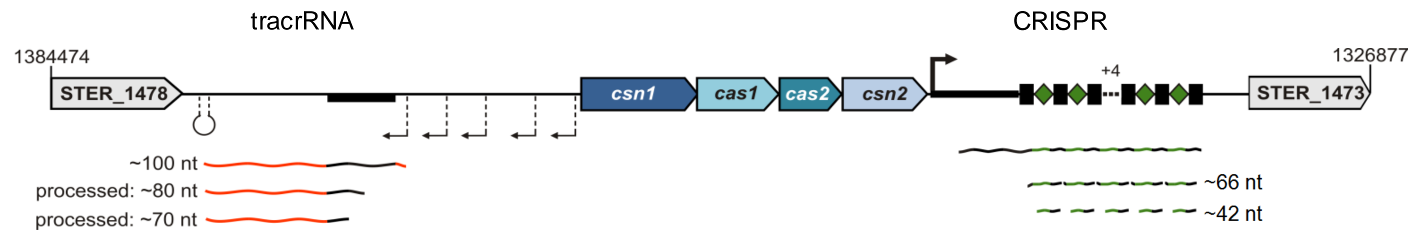
```
5′ .. GUUUUAG--AGCUGUGUUGUUUCGAAUGGUUCCAAAAC... 3′  CRISPR repeat
         ||||•    |||||||||||||||||•||||||  ||
3′ ...   AAAUUGAACGACACAACAAAGCUUACUAAGGUUGUG... 5′  tracrRNA
```

Supplementary Figure 12 | tracrRNA homologue and pre-crRNA are expressed and processed in *S. mutans* UA159. a, Genomic organization of tracrRNA and CRISPR locus (type II, Nmeni/CASS4) in *S. mutans* UA159. The tracrRNA homologue is encoded upstream of the CRISPR-associated genes and on the opposite strand. Putative promoters and Rho-independent

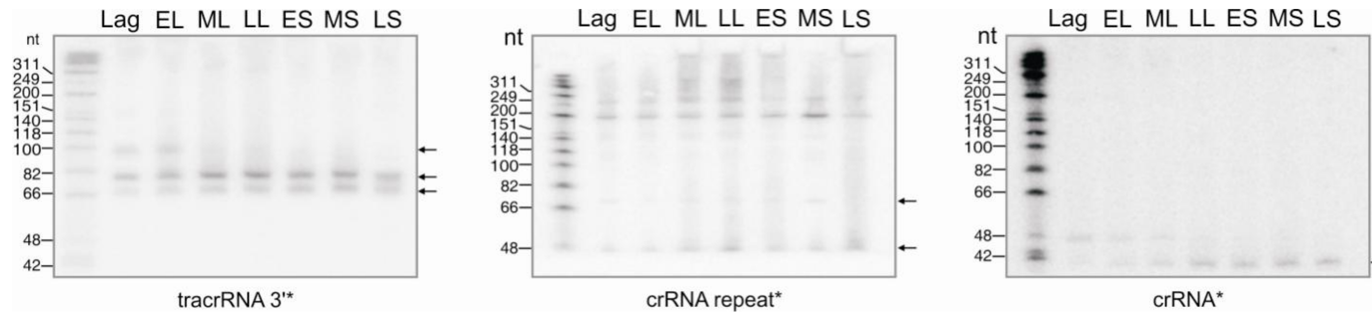
terminator are indicated. The black bar within tracrRNA represents the sequence complementary to each CRISPR repeat. Sizes of tracrRNA species (red-black) correspond to those observed by northern blot analysis. pre-crRNA is encoded downstream of the CRISPR-associated genes. Predicted sizes of CRISPR crRNAs are indicated. Rectangles, CRISPR repeats; diamonds, CRISPR spacers. **b,** Northern blot analysis of total RNA isolated from wild-type (WT) cultures grown to different time points (Refer to Supplementary Fig. 2b). Probes: tracrRNA homologue* (OLEC1678), crRNA repeat* (OLEC1679) and crRNA* (OLEC1680) (Supplementary Table 10). Expression and processing of both tracrRNA homologue and pre-crRNA are observed in *S. mutans*. According to the lengths of detected RNAs and bioinformatic predictions (Supplementary Fig. 11), we suggest that the shortest tracrRNAs result from tracrRNA/pre-crRNA co-processing. **c,** Predicted base-pairing of the CRISPR repeat sequence with the anti-repeat sequence within the tracrRNA homologue is represented.

S. thermophilus LMD-9
1st locus

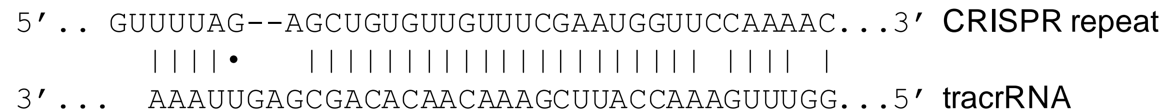
a



b



c



Supplementary Figure 13 | tracrRNA homologue and pre-crRNA are expressed and processed in *S. thermophilus* LMD-9 (1st locus). Same as Supplementary Fig. 12. Probes: tracrRNA homologue* (OLEC1672), crRNA repeat* (OLEC1673) and crRNA* (OLEC1674) (Supplementary Table 10).

S. thermophilus LMD-9
2nd locus

a

tracrRNA CRISPR
642368 651074
STER_0708 csn1 cas1 cas2 cas7 STER_0713
+12
~110 nt
processed: ~80 nt
processed: ~70 nt
66 nt
42 nt

b

Lag EL ML LL ES MS LS
nt
311 249 200 151 140 118 100 82 66 48
tracrRNA 3'*

Lag EL ML LL ES MS LS
nt
311 249 200 151 140 118 100 82 66 48 42
crRNA*

c

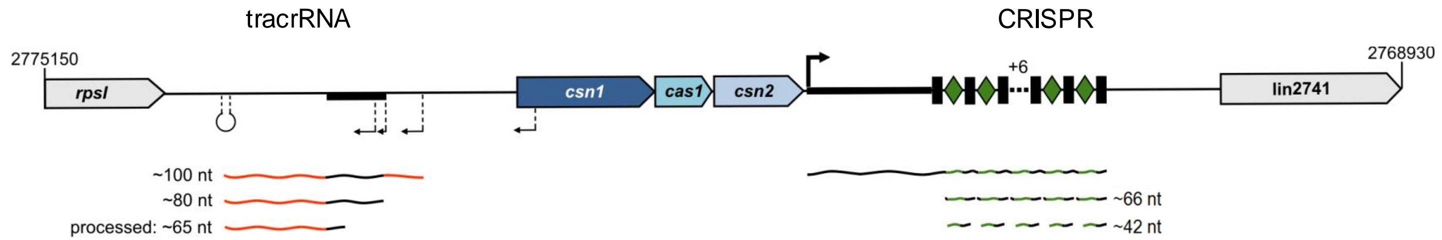
```
5′.. GUUUUUGUACUCU-CAAGAUUUAAGUAACUGUACAAC...3′ CRISPR repeat
       |||  |•    |||  |||||||||||||||||   ||
3′... AAACAUCGAAGACGUUCUAAAUUCAUUGACACAUUC...5′ tracrRNA
```

Supplementary Figure 14 | tracrRNA homologue and pre-crRNA are expressed and processed in *S. thermophilus* LMD-9 (2nd locus). Same as Supplementary Fig. 12. Probes: tracrRNA homologue* (OLEC1669) and crRNA* (OLEC1671)

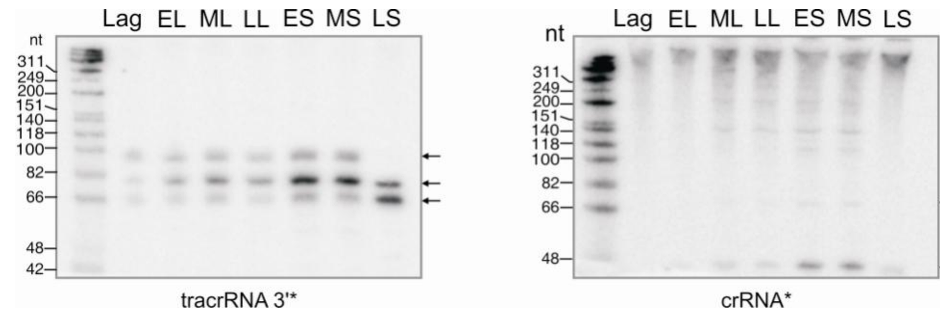
(Supplementary Table 10). Note that for the 2nd locus of this strain, the tracrRNA homologue is located within the *csn1*-*cas1* intergenic region. In addition, the *csn2* gene is absent and replaced by the *cas7* gene. In a previous study, Cas7 was shown to be dispensable for phage resistance and was suggested to be involved in the synthesis and/or insertion of new spacers and additional repeats [2]. Therefore, the absence of *csn2* in this CRISPR-associated gene locus and its replacement by the *cas7* gene, which does not seem to be required for phage resistance, implies that Csn2 in type II CRISPR/Cas systems does not play a role in the RNA processing.

L. innocua Clip11262

a



b



c

```
5′.. GUUUUAG--AGCUAUGUUAUUUUGAAUGCUAACAAAAC...3′ CRISPR repeat
         ||||•    ||||||||||||||||•|||||||  |
3′.....AAAUUGAACGAUACAAUAAAACUUAUGAUUGUUAUA...5′ tracrRNA
```

Supplementary Figure 15 | tracrRNA homologue and pre-crRNA are expressed and processed in *L. innocua* Clip11262. Same as Supplementary Fig. 12. Probes: tracrRNA homologue* (OLEC1675) and crRNA* (OLEC1677) (Supplementary Table 10).

Note that the *cas2* gene is missing in this CRISPR locus of *L. innocua* Clip11262. Because the mature crRNA species and the putative processed forms of tracrRNA homologue are observed in this strain, we argue that Cas2 in type II CRISPR/Cas systems is unlikely to be involved in the RNA processing.

a

N. meningitidis Z2491

tracrRNA

CRISPR

614394

NMA0632

csn1 *cas1* *cas2*

+12

607686

NMA0627

~110 nt

processed: ~90 nt

~66 nt

~42 nt

b

0.2 0.395 0.43 0.2 0.395 0.43 OD_{550}

881 692 501 404 331 242 190 147 110 67 34

tracrRNA 3'*

crRNA repeat*

c

```
5'... GUUGUAGCUCCCUUUCUCAUUUCGCAGUGCUACAAU ...3' CRISPR repeat
      •|||||||   |   |||||||||||||||||  |||||
3'... UAACAUCGUUGCCAAGAGUAAAGCGUCACGCUGUUA ...5' tracrRNA
```

Supplementary Figure 16 | tracrRNA homologue and pre-crRNA are expressed and processed in *N. meningitidis* Z2491. Same as Supplementary Fig. 12. Probes: tracrRNA homologue* (JVO-4862) and crRNA repeat* (JVO-4863) (Supplementary Table 10). Note that the *csn2* gene is missing in this CRISPR locus of *N. meningitidis* Z2491. The mature crRNA species and the putative processed forms of tracrRNA homologue are observed in this strain. Therefore, Csn2 in type II CRISPR/Cas systems is unlikely to be involved in the RNA processing. Note that in contrast to the other represented CRISPR loci, pre-crRNA in this locus is transcribed from the opposite strand of the *csn1-cas1-cas2* operon.

S. epidermidis RP62A

a

b

c

```
5'...GGAUCGAUACCCACCCCGAAGAAAAGGGGACGAGAAC...3' CRISPR repeat
     |• |||||| ||||||||  ||||||||||||||||
3'...CUAAGCUAUUGGUGGGGCGGCUUUUCCCCUGCUCUUG...5'degenerated repeat
```

Supplementary Figure 17 | pre-crRNA but no tracrRNA homologue is expressed and processed in *S. epidermidis* RP62a. Same as Supplementary Fig. 12. Probes: crRNA repeat* (OLEC1712) and crRNA* (OLEC1714) (Supplementary Table 10). The CRISPR locus of *S. epidermidis* RP62a belongs to the type III (Mtube/CASS6) CRISPR/Cas system [3]. The repeat-spacer array is located upstream of the *cas/csm* genes. Two additional degenerated repeats interspaced with a 30 nt sequence are located

downstream of the *cas/csm* operon. Expression and processing of pre-crRNA is observed in *S. epidermidis* RP62a, consistent with a recent study [4]. However, expression of a putative tracrRNA homologue (that would be encoded by the additional repeats) could not be detected (data not shown). Processing of pre-crRNA without expression of a tracrRNA homologue can be explained by the presence of the *cas6* gene in this CRISPR locus. Cas6 of *P. furiosus* was already described to be responsible for processing of pre-crRNA *in vitro* by cleavage within the repeat region [5]. Thus in *S. epidermidis* RP62a, as already suggested in the literature [6,7], Cas6 would be responsible for processing of the pre-crRNA using a mechanism similar to that described for Cas6 of *P. furiosus*. No *trans*-activating small RNA is required for the maturation process.

References to Supplementary Figures

1. Vojtek, I. *et al.* Lysogenic transfer of group A Streptococcus superantigen gene among streptococci. *J. Infect. Dis.* 197, 225-234 (2008).
2. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712 (2007).
3. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11, 181-190 (2010).
4. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463, 568-571 (2010).
5. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22, 3489-3496 (2008).
6. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845 (2008).
7. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* 34, 401-407 (2009).

Supplementary Tables

CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III

This file contains:

Table	Title	Page
1	Distribution of CRISPR loci in *Streptococcus pyogenes* clinical isolates with complete genome sequence available	2
2	Features of spacer sequences from CRISPR01 present in M1 serotype SF370 and their BLAST candidates	3
3	Features of spacer sequences from CRISPR02 present in M1 serotype SF370 and their BLAST candidates	4
4	Features of spacer sequences from CRISPR01 and CRISPR02 present in *S. pyogenes* genomes other than the M1 serotype SF370 genome and their best blast candidates (highest identities)	5
5	List of CRISPR-associated genes in *S. pyogenes* clinical isolates with complete genome sequence available	7
6	Differential RNA sequencing (raw) data of intergenic small RNAs expressed in *S. pyogenes* SF370 SPY- (total) and SPY+ (enriched) libraries	8
7	Basepairings between CRISPR repeats and tracrRNA homologues from various bacterial species	12
8	Strains used in the study	13
9	Plasmids used in the study	15
10	Primers used in the study	16
References		22

Supplementary Table 1. Distribution of CRISPR loci in *S. pyogenes* clinical isolates with complete genome sequence available.

STRAIN[a]	M TYPE	ACCESSION NUMBER	LOCUS TAG	PHAGES	CRISPR01[b]	CRISPR02[c]	CRISPR DB[d]	CRISPR SPACER[e]	REPEATS	REFERENCE
SF370	M1	NC_002737	SPy_xxxx	Φ370.1-4	✓		NC_002737_1	Spyo1h	7	[1]
						✓	NC_002737_4	Spyo2h	4	
MGAS5005	M1	NC_007297	M5005_Spy_xxxx	Φ5005.1-3	✓		NC_007297_1	Spyo1e	4	[2]
						✓	NC_007297_4	Spyo2e	5	
MGAS10270	M2	NC_008022	MGAS10270_Spyxxxx	Φ10270.1-5	✓		NC_008022_1	Spyo1b	3	[3]
						✓	NC_008022_3	Spyo2b	4	
MGAS315	M3	NC_004070	SpyM3_xxxx	Φ315.1-6			-	-	-	[4]
SSI-1	M3	NC_004606	SPsxxxx	ΦSPsP1-6			-	-	-	[5]
MGAS10750	M4	NC_008024	MGAS10750_Spyxxxx	Φ10750.1-4		✓	NC_008024_2	Spyo2c	6	[3]
Manfredo	M5	NC_009332	SpyM5xxxx	ΦMan.1-5			-	-	-	[6]
MGAS10394	M6	NC_006086	M6_Spyxxxx	Φ10394.1-6			-	-	-	[7]
MGAS2096	M12	NC_008023	MGAS2096_SPyxxxx	Φ2096.1-2	✓		-	Spyo1d	3	[3]
						✓	NC_008023_2	Spyo2d	7	
MGAS9429	M12	NC_008021	MGAS9429_Spyxxxx	Φ9429.1-3	✓		-	Spyo1g	3	[3]
						✓	NC_008021_2	Spyo2g	8	
MGAS8232	M18	NC_003485	spyM18_xxxx	Φ8232.1-5			-	-	-	[8]
ATCC 10782	M24	NZ_AEEO000 00000	HMPREF0841_xxxx	ND	ND	ND	-	-	-	D. Muzny
MGAS6180	M28	NC_007296	M28_Spyxxxx	Φ6180.1-4	✓		NC_007296_2	Spyo1f	5	[9]
						✓	NC_007296_3	Spyo2f	2	
NZ131	M49	NC_011375	Spy49_xxxx	ΦNZ131.1-3	✓		NC_011375_1	-	5	[10]
						✓	NC_011375_3	-	6	
M49 591	M49	NZ_AAFV000 00000	SpyoM0100xxxx	ND	ND	ND	-	-	-	A. Podbielski

[a] *S. pyogenes* M49 591 and ATCC10782, unfinished sequence, whole genome shotgun sequencing project. Note that MGAS315, SSI-1 and MGAS10750 genomes do not contain the CRISPR01 repeat-spacer array but do contain CRISPR01-associated genes (Refer to Supplementary Fig. 1 and Supplementary Table 5).
[b] CRISPR01 locus belongs to the type II (Nmeni/CASS4) CRISPR/Cas system. CRISPR01 repeat sequences (GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC) are identical in all genomes except MGAS2096 and MGAS9429 genomes. The two latter genomes contain questionable CRISPR01 structures with shorter repeats (AATAATTGGTATAGTCTAATTATA).
[c] CRISPR02 locus belongs to the type I-C (Dvulg/CASS1) CRISPR/Cas system. CRISPR02 repeat sequences (ATTTCAATCCACTCACCCATGAAGGGTGAGAC) are identical in all genomes.
[d] Nomenclature is according to the CRISPR database (CRISPR DB) (http://crispr.u-psud.fr/).
[e] Nomenclature is according to [11].
ND, not determined.
Noteworthy, from these data, no clear correlation between the absence or presence of CRISPR loci and the polylysogeny stage of *S. pyogenes* clinical isolates can be assessed. This is exemplified by strain MGAS10270, the genome of which contains both CRISPR01 and CRISPR02 loci but contains also 5 prophages.

Supplementary Table 2. Features of spacer sequences from CRISPR01 present in M1 serotype SF370 and their BLAST candidates.

CRISPR01 SPACER[a]	SPACER SEQUENCE	BLAST CANDIDATE			
		STRAIN (M type)[b]	PHAGE	GENE[c]	FUNCTION
Spyo1h_001	TGCGCTGGTTGATTTCTTCTTGCGCTTTTT	MGAS315 (M3)	Φ315.2	SpyM3_0930	endopeptidase
		MGAS315 (M3)	Φ315.4	SpyM3_1215	endopeptidase
		SSI-1 (M3)	ΦSPsP3	SPs0647	endopeptidase
		SSI-1 (M3)	ΦSPsP5	SPs0926	endopeptidase
		NIH1 (M3**)**	ΦNIH1.1	NIH1.1_43	endopeptidase
		MGAS10394 (M6)	Φ10394.3	M6_Spy0994	endopeptidase
		MGAS10394 (M6)	Φ10394.6	M6_Spy1349	endopeptidase
		MGAS8232 (M18)	Φ8232.5	spyM18_0769	endopeptidase
		MGAS6180 (M28)	Φ6180.2	M28_Spy1234	endopeptidase
		S. equi subsp. equi 4047	ΦP9	P9_gp42	endopeptidase
		SF 370 (M1)	Φ370.1	SPy_0700*	endopeptidase
		MGAS5005(M1)	Φ5005.3	M5005_Spy1424*	endopeptidase
		MGAS2096 (M12)	Φ2096.1	MGAS2096_Spy0592*	endopeptidase
		MGAS2096 (M12)	Φ2096.2	MGAS2096_Spy1450*	endopeptidase
		MGAS9429 (M12)	Φ9429.3	MGAS9429_Spy1426*	endopeptidase
Spyo1h_002	TTATATGAACATAACTCAATTTGTAAAAAA	MGAS8232 (M18)	Φ8232.3	*speM*	superantigen
Spyo1h_003	AGGAATATCCGCAATAATTAATTGCGCTCT	MGAS10270 (M2)	Φ10270.1	MGAS10270_Spy0563	adenine-specific methyltransferase
		MGAS10750 (M4)	Φ10750.1	MGAS10750_Spy0588	adenine-specific methyltransferase
		MGAS8232 (M18)	Φ8232.5	spyM18_0742	adenine-specific methyltransferase
Spyo1h_004	AGTGCCGAGGAAAAATTAGGTGCGCTTGGC	MGAS315 (M3)	Φ315.3	SpyM3_1101	hyaluronidase (EC 3.2.1.35)
		SSI-1 (M3)	ΦSPsP4	SPs0763	hyaluronidase (EC 3.2.1.35)
		MGAS10750 (M4)	Φ10750.3	MGAS10750_Spy1285	hyaluronidase (EC 3.2.1.35)
		MGAS10394 (M6)	Φ10394.5	M6_Spy1203	hyaluronidase (EC 3.2.1.35)
		MGAS9429 (M12)	Φ9429.2	MGAS9429_Spy0843	hyaluronidase (EC 3.2.1.35)
		M49	ΦH4489A	*hylP*	hyaluronidase (EC 3.2.1.35)
		MGAS8232 (M18)	Φ8232.3	spyM18_1254*	hyaluronidase (EC 3.2.1.35)
		NZ131 (M49)	ΦNZ131.2	Spy49_0785*	hyaluronidase (EC 3.2.1.35)
Spyo1h_005	TAAATTTGTTTAGCAGGTAAACCGTGCTTT	MGAS10270 (M2)	Φ10270.1	MGAS10270_Spy0546	phage protein
		MGAS10270 (M2)	Φ10270.2	MGAS10270_Spy0804	phage protein
		MGAS315 (M3)	Φ315.2	SpyM3_0965	phage protein
		MGAS315 (M3)	Φ315.5	SpyM3_1347	phage protein
		SSI-1 (M3)	ΦSPsP2	SPs0517	phage protein
		SSI-1 (M3)	ΦSPsP5	SPs0888	phage protein
		MGAS10750 (M4)	Φ10750.2	MGAS10750_Spy0839	phage protein
		NZ131 (M49)	ΦNZ131.3	Spy49_1511c	phage protein
Spyo1h_006	TTCAGCACACTGAGACTTGTTGAGTTCCAT	-	-	-	-

[a] Spacer code is according to [11].
[b] All strains refer to *S. pyogenes* except for the indicated *S. equi*.
[c] ND, no data available.
All spacers have 100% sequence identity to known prophage-encoded genes (exception for Spyo1h_001 and Spyo1h_004 that anneal with 1 mismatch to the indicated genes with an asterix*). Only the Spyo1h_006 spacer did not show any sequence similarity with sequences in the NCBI database. Sequences of spacers Spyo1h_001, Spyo1h_003, Spyo1h_004 and Spyo1h_005 anneal to the coding strands of the respective genes. Sequence of spacer Spyo1h_002 anneals perfectly to the non-coding strand of superantigen SpeM-encoding gene. One spacer (Spyo1h_001) anneals with one mismatch to the endogenous prophage gene, Spy_0700. Note the presence of spacers that are specific to *S. equi* genomes. This suggests that transfer of phages between the human pathogen *S. pyogenes* and the animal pathogen *S. equi* might have occurred either directly or indirectly via other streptococcal species [12].

Supplementary Table 3. Features of spacer sequences from CRISPR02 present in M1 serotype SF370 and their BLAST candidates.

CRISPR02 SPACER[a]	SPACER SEQUENCE	BLAST CANDIDATE				
		STRAIN (M type)[b]	PHAGE	GENE	FUNCTION[c]	ALIGNMENT[d]
Spyo2h_001	AAAAAGCATATCACGAAAATCACCAATTACATCAG	MGAS10270 (M2)	Φ10270.1	MGAS10270_Spy0562	methyltransferase	AAAAGGCATATCACGAAAATCACCAATTACATCAG
		MGAS10750 (M4)	Φ10750.1	MGAS10750_Spy0587	methyltransferase	AAAAGGCATATCACGAAAATCACCAATTACATCAG
		Manfredo (M5)	ΦMan.4	SpyM51298	methyltransferase	AAAAGGCATATCACGAAAATCACCAATTACATCAG
		MGAS8232 (M18)	Φ8232.2	spyM18_0741	methyltransferase	AAAAGGCATATCACGAAAATCACCAATTACATCAG
		MGAS315 (M3)	Φ315.2	SpyM3_0953	methyltransferase	AAAAGGCATGTCACGAAAATCACCAATTACATCAG
		SSI-1 (M3)	ΦSPsP5	SPs0903	methyltransferase	AAAAGGCATGTCACGAAAATCACCAATTACATCAG
		NZ131 (M49)	ΦNZ131.3	Spy49_1494c	methyltransferase	AAAAGGCATGTCACGAAAATCACCAATTACATCAG
Spyo2h_002	AAGTCATTTTCTGCTTCTGCTAGGTTTGCTTTA	*S. equi subsp. equi* 4047	ΦSeq2	SEQ_0825	Φ capsid protein	AAGTCATTTTCTGCTTCTGCTAGGTTTGCTTTA
		MGAS10394 (M6)	Φ10394.3	M6_Spy1006	Φ capsid protein	AAGTCGTTTTCTGCTTCTGCTAGGTTCGCTTTA
		MGAS315 (M3)	Φ315.2	SpyM3_0941	Φ capsid protein	AAGTCGTTTTCTGCTTCTACTAGGTTCGCTTTA
		SSI-1 (M3)	ΦSPsP5	SPs0914	Φ capsid protein	AAGTCGTTTTCTGCTTCTACTAGGTTCGCTTTA
Spyo2h_003	TTTTTACTTTGATTACATCCCGCAATGTCACAGC	MGAS10270 (M2)	Φ10270.3	MGAS10270_Spy1333	Φ HP	TTTTTGCTTTGATTACATTCCGCAATGTCACAGC
		MGAS10270 (M2)	Φ10270.2	MGAS10270_Spy0821	Φ HP	TTTTTGCTTTGATTACATTCCGCAATGTCACAGC
		MGAS315 (M3)	Φ315.4	SpyM3_1239	Φ HP	TTTTTGCTTTGATTACATTCCGCAATGTCACAGC
		NIH1 (M3)	ΦNIH1.1	ND	ND	TTTTTGCTTTGATTACATTCCGCAATGTCACAGC
		MGAS10750 (M4)	Φ10750.2	MGAS10750_Spy0857	Φ HP	TTTTTGCTTTGATTACATTCCGCAATGTCACAGC
		Manfredo (M5)	ΦMan.1	SpyM50500	Φ HP	TTTTTGCTTTGATTACATTCCGCAATGTCACAGC
		MGAS6180 (M28)	Φ6180.2	M28_Spy1257	Φ HP	TTTTTGCTTTGATTACATTCCGCAATGTCACAGC

[a] Spacer code is according to [11].
[b] All strains refer to *S. pyogenes* except for the indicated *S. equi*.
[c] ND, no data available.
[d] All spacers have imperfect sequence identity to known prophage-encoded genes. Sequences of spacers anneal to the coding strands of the respective genes. Mismatches with respect to spacer sequence are highlighted in black boxes. Blast candidates are ranked according to the increasing number of mismatches in similarities to the consensus spacer sequence.

Supplementary Table 4. Features of spacer sequences from CRISPR01 and CRISPR02 present in *S. pyogenes* genomes other than the M1 serotype SF370 genome and their best BLAST candidates (highest identities).

STRAIN (M type)	CRISPR SPACER[a]	SPACER SEQUENCE	BLAST CANDIDATE[b]	% IDENTITY
MGAS5005 (M1)	Spyo1e_001	GGGTGGTTGGCTGACGCATCGCAATATTAA	SPy_0701; MGAS10270_Spy0590; MGAS10270_Spy1310; SPs0648; SpyM3_0929; SpyM3_1214; MGAS10750_Spy0615; SpyM50524; SpyM51031; SpyM51271; M6_Spy1550; MGAS9429_Spy0586; spyM18_0770; HylP1, *Streptococcus* phage P9; HylP, *Streptococcus* phage NIH1	100
	Spyo1e_002	AGGAATATCCGCAATAATTAATTGCGCTCT	MGAS10270_Spy0563; MGAS10750_Spy0588; SpyM18_0742	100
	Spyo1e_003	TAAATTTGTTTAGCAGGTAAACCGTGCTTT	MGAS10270_Spy0546; MGAS10270_Spy0804; SPs0517; SPs0888; SpyM3_0965; SpyM3_1347; MGAS10750_Spy0839; Spy49_1511c	100
	Spyo2e_001	AAAAAGCATATCACGAAAATCACCAATTACATCAG	MGAS10270_Spy0562; MGAS10750_Spy0587; SpyM51298; spyM18_0741	97
	Spyo2e_002	AAGTCATTTTCTGCTTCTGCTAGGTTTGCTTTA	M6_Spy1006; (SEQ_0825: 100%)	93
	Spyo2e_003	TTTTTACTTTGATTACATCCCGCAATGTCACAGC	MGAS10270_Spy0821; MGAS10270_Spy1333; SpyM3_1239; MGAS10750_Spy0857; SpyM50500; M28_Spy1257	94
	Spyo2e_004	TGTCCGCATACCTTGATTTGAGCGAGTAAACTC	SpyM51030	100
MGAS10270 (M2)	Spyo1b_001	AAGCCAAACGCTAGGTTTAACCGATGTTGT	spyM18_1292; SAK_0616; M28_Spy1015	100
	Spyo1b_002	TAAAAAAACTCATACTAAAAAAATAGATTA	SPy_0701; SPy_1445; SpyM3_1101; SpyM3_1214; SPs0648; SPs0763; MGAS10750_Spy1285; SpyM50680; SpyM51031; M6_Spy1203; spyM18_1455; hylP Bacteriophage H4489A; hylP2; SEQ_2045	100
	Spyo2b_001	CTGCATGATACCCTGTTCCCACATCATGTTAGC	SPy_1460; M5005_Spy_1189; SpyM3_1115; SPs0749; MGAS10750_Spy1299; SpyM50666; M6_Spy1217; spyM18_1474; SAG0585; SAK_0635; (SDEG_1129, 100%)	93
	Spyo2b_002	AATAGAGAAAAAACTATATGAATACAAAGTCTATG	spyM18_0718; SpyM51320	100
	Spyo2b_003	ACCTCTCATAATATCTTTAAGTTCTCTTACCTCAC	SpyM3_0729; SPs1123; SEQ_2041; spyM18_0389	100
MGAS10750 (M4)	Spyo2c_001	TACTAGCGTTGAGATAGCCGGTATTAATCTTACC	SpyM50681	73
	Spyo2c_002	ATTTCGTCACCTCCTCAATCAATAATAGAGTCA		
	Spyo2c_003	CTAAGATGATACCAGTTACAATACCGATTTTAAGC	spyM18_1491; (MGAS10750_Spy1310, 80%)	100
	Spyo2c_004	AAATAATTACATGACTAGCCAATACACCCACATA	SpyM50471; SpyM3_1354; SPs0507; spyM18_1808; SDEG_1665	100
	Spyo2c_005	TTCATGAGCTTCTTTACTCTCAAAGTAAGAGTG	SPy_1475 (MGAS10750_Spy1314, 93%)	100
MGAS2096 (M12)	Spyo1d_001	GATTGACCACAACATCCAACGCTTAGGTTAT	Spy_1460, M5005_Spy1189; SpyM3_1115; SPs0749; SpyM50666; MGAS10750_Spy1299; M6_Spy1217; spyM18_1474; (SDEG_1129, 100%)	96
	Spyo1d_002	AATAGAGCTAGACAAAAAATTGAGTTTGAC	SPs0927; M6_Spy1348; spyM18_1254; spyM18_1757; HylP, *Streptococcus* phage P9; SEQ_2045	100
	Spyo2d_001	ATCGATTTTGCAGATAAAAGGAAACATAGAGTTC	SpyM3_0710; SPs1142; spyM18_0369	100
	Spyo2d_002	AATTTCATCTCCTGCATCTTGATCAGTTAGGGTTAC	SpyM3_0702; SPs1150; spyM18_0361; *Streptococcus* phage phi3396	100
	Spyo2d_003	AGATTCGGCTAGATATTCTAAAAAATCGATAAAGC	SpyM51029	94
	Spyo2d_004	TCCAACTTTTGTAAAAGTAGAATTTGCTACGTTTG	SpyM51321; spyM18_0717	100
	Spyo2d_005	ATTATTTTATTAAAGTTATCACGTAAATTTTGCAA	M5005_Spy_1029; SPs0902; SpyM50494; spyM18_1789; M28_Spy1005	100
	Spyo2d_006	CATTATATGAACAATGCCTTTGCGGAATTAGTT	SPy_0008; M5005_Spy_0006; MGAS10270_Spy0006; SpyM3_0006; SPs0006; SpyM50006; MGAS10750_Spy0006; M6_Spy0006; MGAS2096_Spy0006; MGAS9429_Spy0006, M28_Spy0006; Spy49_0006; (spyM18_0007, 93%)	96
MGAS9429 (M12)	Spyo1g_001	GATTGACCACAACATCCAACGCTTAGGTTAT	Spy_1460, M5005_Spy_1189; SpyM3_1115; SPs0749; SpyM50666; MGAS10750_Spy1299; M6_Spy1217; spyM18_1474; (SDEG_1129, 100%)	96
	Spyo1g_002	AATAGAGCTAGACAAAAAATTGAGTTTGAC	SPs0927; M6_Spy1348; spyM18_1254; spyM18_1757; HylP, *Streptococcus* phage P9; SEQ_2045	100
	Spyo2g_001	ATCGATTTTGCAGATAAAAGGAAACATAGAGTTC	SpyM3_0710; SPs1142; spyM18_0369	100
	Spyo2g_002	AATTTCATCTCCTGCATCTTGATCAGTTAGGGTTAC	SpyM3_0702; SPs1150; spyM18_0361; *Streptococcus* phage phi3396	100
	Spyo2g_003	ATGAACTAGGACATATTGAACATGATTCTGGCCAAT	M5005_Spy_1220	100
	Spyo2g_004	AGATTCGGCTAGATATTCTAAAAAATCGATAAAGC	SpyM51029	94
	Spyo2g_005	TCCAACTTTTGTAAAAGTAGAATTTGCTACGTTTG	SpyM51321; spyM18_0717	100
	Spyo2g_006	ATTATTTTATTAAAGTTATCACGTAAATTTTGCAA	M5005_Spy1029; SPs0902; SpyM50494; spyM18_1789; M28_Spy1005	100
	Spyo2g_007	CATTATATGAACAATGCCTTTGCGGAATTAGTT	SPy_0008; M5005_Spy_0006; MGAS10270_Spy0006; SpyM3_0006; SPs0006; SpyM50006; MGAS10750_Spy0006; M6_Spy0006; MGAS2096_Spy0006; MGAS9429_Spy0006, M28_Spy0006; Spy49_0006; (spyM18_0007, 93%)	96
MGAS6180 (M28)	Spyo1f_001	TGCTCCAGATGGATTTTTAAACTGATTATT		
	Spyo1f_002	TAGAAATTGACAATGCTTTTTCTTTTGTCT	SPy_0986; MGAS9429_Spy0833	100
	Spyo1f_003	AATAGAGCTAGACAAAAAATTGAGTTTGAC	SPs0927; M6_Spy1348; spyM18_1254; spyM18_1757; HylP, *Streptococcus* phage P9; SEQ_2045	100

STRAIN (M type)	CRISPR SPACER[a]	SPACER SEQUENCE	BLAST CANDIDATE[b]	% IDENTITY
	Spyo1f_004	GACATATTAACGTCCTTTCTCCTGCTTTCC	SPy_1459; M5005_Spy1188; SpyM3_1114; SPs0750; MGAS10750_Spy1298; M6_Spy1216; spyM18_1472; SDEG_1128, SAG0586; SAK_0636	100
	Spyo2f_001	AAATAATTACATGACTAGCCAATACACCCACATA	SpyM50471; SpyM3_1354; SPs0507; spyM18_1808; SDEG_1665	100
NZ131 (M49)	Spyo1a_001	AGCTGCATCGCTGTAGTATTTACCAATATA	SPy_1464; M5005_Spy1193; SpyM3_1120; SPs0744; MGAS10750_Spy1302; SpyM50661; M6_Spy1222; spyM18_1480	100
	Spyo1a_002	ACTGGGAAATTGATAAAATCGGCAATGCCCG	SPy_0700; SpyM3_1215; SpyM50523; M6_Spy1349; MGAS2096_Spy0592; spyM18_0769, M28_Spy1234	100
	Spyo1a_003	CAAGTTGTTTAATCGAAGAATTTCCCGTTG	SpyM3_1234; SPs0627; spyM18_1782; M28_Spy1253	100
	Spyo1a_004	GTGCCTGTGGAGGAATTGATGAACATGCCT	Spy49_1465c	90
	Spyo2a_001	AATCAATCGGCGCTTTTTGTCGATTATAAGGGA	M6_Spy1550; HylP, *Streptococcus* phage P9	100
	Spy02a_002	TCCATTTTTTACTTTCTCATGTGGCAAAACATAAT	SpyM50535; M6_Spy1540	100
	Spy02a_003	CGCTAATCGACTTATCGACTAAAGCAACTGTTATC	SPy_1455; M5005_Spy1185; spyM18_1469; SDEG_1125	97
	Spy02a_004	CAAAAATATCTGAAGAAGAATCCATTTTAAGATCT		
	Spy02a_005	TTTTAAATCGACATGTGCAGTACCAGTGTCATCTT	*Streptococcus* phage phi3396	94

[a] Spacer code is according to [11].

[b] Locus tags: SPy_xxxx, SF370 (M1); M5005_Spyxxxx (or M5005_Spy_xxxx), MGAS5005 (M1); MGAS10270_Spyxxxx, MGAS10270 (M2); SpyM3_xxxx, MGAS315 (M3); SPsxxxx, SSI-1 (M3); MGAS10750_Spyxxxx, MGAS10750 (M4); SpyM5xxxx, Manfredo (M5); M6_Spyxxxx, MGAS10394 (M6); MGAS2096_SPyxxxx, MGAS2096 (M12); MGAS9429_Spyxxxx, MGAS9429 (M12); spyM18_xxxx, MGAS8232 (M18); M28_Spyxxxx, MGAS6180 (M28); Spy49_xxxx, NZ131 (M49); SDEG_xxxx, *Streptococcus dysgalactiae subsp. equisimilis* GGS_124; SAGxxxx, *Streptococcus agalactiae* 2603 (serotype V); SAK_xxxx, *S. agalactiae* A909 (serotype Ia); SEQ_xxxx, *Streptococcus equi subsp. equi* 4047. Intriguingly, five spacers show 80 to 100% sequence identity to endogenous genes (MGAS10750_Spy1310, MGAS10750_Spy1314, MGAS2096_Spy0006, MGAS9429_Spy0006 and Spy49_1465c). The presence of self-targeting spacers (see also the self-targeting spacer (Spyo1h_001) of the SF370 genome in Supplementary Table 2) suggests a complex mechanism for self versus non-self targeting in *S. pyogenes*.

Supplementary Table 5. List of CRISPR-associated genes in *S. pyogenes* clinical isolates with complete genome sequence available. Panel A: CRISPR01-associated genes. Panel B: CRISPR02-associated genes.

A. CRISPR01

STRAIN[a]	CSN1		CAS1		CAS2		CSN2	
	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]
SF370	1046	100.0	1047	100.0	1048	100.0	1049	100.0
MGAS5005	0769	99.9	0770	100.0	0771	100.0	0772	100.0
MGAS10270	0886	99.6	0887	100.0	0888	100.0	0889	90.2
MGAS315	0677	98.8	0678	100.0	0679	99.1	0680	99.1
SSI-1	1176	98.8	1175	100.0	1174	99.1	1173	99.1
MGAS10750	0921	88.6	0922	97.9	0923	100.0	0924	97.3
Manfredo	-		-		-		-	
MGAS10394	-		-		-		-	
MGAS2096	0843	99.3	0844	100.0	0845	100.0	0846	99.1
MGAS9429	0885	99.2	0886	100.0	0887	100.0	0888	99.1
MGAS8232	-		-		-		-	
MGAS6180	0748	99.3	0749	100.0	0750	100.0	0751	99.1
NZ131	0823	98.4	0825	99.0	0826	100.0	0827	99.1
M49 591	0639	ND	0640	ND	0641	ND	0642	ND

B. CRISPR02

STRAIN[a]	CAS3		CAS5D (CSD5D)		CSD1		CSD2		CAS4		CAS1		CAS2	
	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]	GENE[b]	% IDENTITY[c]
SF370	1567	100.0	1566	100.0	1565	100.0	1564	100.0	1563	100.0	1562	100.0	1561	100,0
MGAS5005	1291	100.0	1290	100.0	1289	99.8	1288	100.0	1287	100.0	1286	100.0	1285	100,0
MGAS10270	1372	98.6	1371	99.6	1370	99.5	1369	99.3	1368	99.6	1366	90.2	1365	100,0
MGAS315	-		-		-		-		-		-		-	
SSI-1	-		-		-		-		-		-		-	
MGAS10750	1399	99.3[d]	1397	99.6	1396	98.6	1395	98.6	1394	99.6	1393	99.1	1392	100,0
Manfredo	-		-		-		-		-		-		-	
MGAS10394	-		-		-		-		-		-		-	
MGAS2096	1310	99.5	1309	99.6	1308	98.7	1307	99.3	1306	100.0	1305	99.7	1304	100,0
MGAS9429	1285	99.5	1284	99.6	1283	98.7	1282	99.3	1281	100.0	1280	99.7	1279	100,0
MGAS8232	-		-		-		-		-		-		-	
MGAS6180	1295	99.5	1294	100.0	1293	98.9	1292	99.3	1291	99.6	1290	99.1	1289	100,0
NZ131	1213	99.0	1212	99.6	1211	99.0	1210	99.6	1209	100.0	1208	100.0	1207	100,0
M49 591	0911	ND	0910	ND	0909	ND	0908	ND	ND	ND	ND	ND	0907	ND

[a] *S. pyogenes* M49 591, unfinished sequence, whole genome shotgun sequencing project.
[b] Gene number from strain locus tag is given.
[c] Protein identity is according to KEGG database (http://www.genome.jp/kegg/). Percent identities refer to SF370 proteins.
[d] C-terminus-truncated protein, does not contain the predicted helicase domain.

Supplementary Table 6. Differential RNA sequencing (raw) data of intergenic small RNAs expressed in *S. pyogenes* SF370 SPY- (total) and SPY+ (enriched) libraries.

CANDIDATE[a]	COORDINATES[b]		SIZE (nt)	ORIENTATION[c]	ADJACENT GENES[d]	LIBRARY SPY-[e]		LIBRARY SPY+[e]		ENRICHMENT[f]
						OCC.	%	OCC.	%	
SPync001	33952	33977	26	→→→	SPy_0021 / SPy_0022	-	-	2	0.114	-
SPync002	75867	75892	26	→→→	SPy_0074 / SPy_0075	13	0.505	36	2.050	4.06
SPync003	93359	93452	94	→→→	SPy_0097 / SPy_0098	20	0.776	22	1.253	1.61
SPync004	116781	116940	160	←→→	SPy_0124 / SPy_0127	7	0.272	5	0.285	1.05
SPync005	117005	117382	378	←→→	SPy_0124 / SPy_0127	7	0.272	-	-	-
SPync006	117449	117638	190	←→→	SPy_0124 / SPy_0127	3	0.116	2	0.114	0.98
SPync007	118243	118267	25	←→→	SPy_0124 / SPy_0127	2	0.078	-	-	-
SPync008	118454	118679	226	←→→	SPy_0124 / SPy_0127	2	0.078	-	-	-
SPync009	148619	148656	38	→→→	SPy_0163 / SPy_0164	8	0.311	26	1.481	4.77
SPync010	174217	174355	139	→→→	SPy_0185 / SPy_0186	2	0.078	-	-	-
SPync011	174317	174339	23	→←→	SPy_0185 / SPy_0186	4	0.155	-	-	-
SPync012	193039	193067	29	→→←	SPy_0215 / SPy_0216	6	0.233	-	-	-
SPync013	195120	195261	142	←←→	SPy_0216 / SPy_0217	22	0.854	19	1.082	1.27
SPync014	207419	207451	33	→→→	SPy_0237 / SPy_0238	2	0.078	-	-	-
SPync015	213738	213967	230	→→→	SPy_0245 / SPy_0246	7	0.272	2	0.114	0.42
SPync016	235753	235809	57	→→→	SPy_0269 / SPy_0271	3	0.116	9	0.513	4.40
SPync017	239475	239505	31	→→→	SPy_0273 / SPy_0274	4	0.155	8	0.456	2.93
SPync018	254646	254675	30	→→←	SPy_0290 / SPy_0292	2	0.078	-	-	-
SPync019	254861	254880	20	→←←	SPy_0290 / SPy_0292	3	0.116	-	-	-
SPync020	254987	255200	214	→←←	SPy_0290 / SPy_0292	2	0.078	-	-	-
SPync021	279245	279299	55	→→→	SPy_0316 / SPy_0317	2	0.078	-	-	-
SPync022	291636	291799	164	→→→	SPy_0334 / SPy_0337	26	1.009	13	0.740	0.73
SPync023	291895	291913	19	→←→	SPy_0334 / SPy_0337	3	0.116	-	-	-
SPync024	318404	318504	101	→→→	SPy_0370 / SPy_0371	9	0.349	7	0.399	1.14
SPync025	319190	319385	196	→→→	SPy_0371 / SPy_0373	11	0.427	6	0.342	0.80
SPync026	320611	320690	80	→→→	SPy_0374 / SPy_0376	2	0.078	-	-	-
SPync027	334719	334770	52	→→→	SPy_0393 / SPy_0395	-	-	6	0.342	-
SPync028	340985	341032	48	←→→	SPy_0393 / SPy_0395	4	0.155	-	-	-
SPync029	353014	353122	109	→→→	SPy_0422 / SPy_0425	-	-	5	0.285	-
SPync030	365319	365355	37	→→→	SPy_0441 / SPy_0442	2	0.078	-	-	-
SPync031	365567	365585	19	→→→	SPy_0441 / SPy_0442	2	0.078	-	-	-
SPync032	384429	384545	117	→→→	SPy_0469 / SPy_0470	2	0.078	5	0.285	3.67
SPync033	391409	391475	67	←→→	SPy_0478 / SPy_0479	9	0.349	-	-	-
SPync034	392307	392328	22	→→←	SPy_0480 / SPy_0481	8	0.311	-	-	-
SPync035	392989	393045	57	←→→	SPy_0481 / SPy_0484	3	0.116	2	0.114	0.98
SPync036	394513	394564	52	→→→	SPy_0486 / SPy_0488	4	0.155	-	-	-

CANDIDATE[a]	COORDINATES[b]		SIZE (nt)	ORIENTATION[c]	ADJACENT GENES[d]	LIBRARY SPY-[e]		LIBRARY SPY+[e]		ENRICHMENT[f]
						OCC.	%	OCC.	%	
SPync037	416968	417181	214	→ → →	SPy_0516 / SPy_0517	12	0.466	19	1.082	2.32
SPync038	450289	450338	50	→ → →	SPy_0556 / SPy_0558	3	0.116	-	-	-
SPync039	455300	455330	31	→ → ←	SPy_0563 / SPy_0565	2	0.078	-	-	-
SPync040	477702	477741	40	→ ← ←	SPy_0593 / SPy_0595	3	0.116	-	-	-
SPync041	480669	480696	28	→ ← ←	SPy_0596 / SPy_0598	4	0.155	-	-	-
SPync042	482816	482963	148	← ← ←	SPy_0600 / SPy_0601	24	0.932	30	1.708	1.83
SPync043	500054	500171	118	→ → →	SPy_0622 / SPy_0623	21	0.815	33	1.879	2.31
SPync044	516048	516126	79	→ → →	SPy_0642 / SPy_0643	-	-	2	0.114	-
SPync045	519042	519069	28	→ ← ←	SPy_0645 / SPy_0646	2	0.078	-	-	-
SPync046	531082	531162	81	← → ←	SPy_0655 / SPy_0656	9	0.349	-	-	-
SPync047	531117	531142	26	← ← ←	SPy_0655 / SPy_0656	2	0.078	-	-	-
SPync048	532743	532788	46	→ ← →	SPy_0658 / SPy_0659	2	0.078	7	0.399	5.13
SPync049	544208	544233	26	→ → →	SPy_0676 / SPy_0677	2	0.078	-	-	-
SPync050	544265	544313	49	→ ← →	SPy_0676 / SPy_0677	161	6.250	86	4.897	0.78
SPync051	546305	546568	264	→ → →	SPy_0679 / SPy_0680	4	0.155	-	-	-
SPync052	570184	570216	33	← ← →	SPy_0712 / SPy_0713	-	-	2	0.114	-
SPync053	598538	598590	53	→ → →	SPy_0737 / SPy_0738	2	0.078	-	-	-
SPync054	610456	610546	91	→ → →	SPy_0747 / SPy_0749	2	0.078	-	-	-
SPync055	611058	611099	42	→ → →	SPy_0749 / SPy_0751	3	0.116	-	-	-
SPync056	623726	623927	202	→ → →	SPy_0766 / SPy_0768	3	0.116	6	0.342	2.93
SPync057	661400	661517	118	→ → →	SPy_0803 / SPy_0804	12	0.466	36	2.050	4.40
SPync058	670424	670496	73	→ → ←	SPy_0813 / SPy_0814	6	0.233	-	-	-
SPync059	680209	680258	50	→ → →	SPy_0827 / SPy_0830	9	0.349	-	-	-
SPync060	694259	694308	50	→ → →	SPy_0841 / SPy_0843	2	0.114	-	-	-
SPync061	704620	704656	37	→ → →	SPy_0852 / SPy_0853	4	0.228	-	-	-
SPync062	717625	717809	185	→ → →	SPy_0867 / SPy_0870	-	-	6	0.342	-
SPync063	736913	736994	82	← ← →	SPy_0888 / SPy_0889	2	0.078	-	-	-
SPync064	736937	736980	44	← → →	SPy_0888 / SPy_0889	3	0.116	-	-	-
SPync065	744487	744599	113	→ → →	SPy_0899 / SPy_0900	5	0.194	23	1.310	6.75
SPync066	762841	762872	32	→ → →	SPy_0917 / SPy_0918	-	-	2	0.114	-
SPync067	779816	779906	91	← → ←	SPy_0937 / SPy_0938	36	1.398	6	0.342	0.24
SPync068	779835	779869	35	← ← ←	SPy_0937 / SPy_0938	6	0.233	-	-	-
SPync069	779921	780010	90	← ← ←	SPy_0937 / SPy_0938	-	-	8	0.456	-
SPync070	792418	792465	48	→ ← →	SPy_0965 / SPy_0967	21	0.815	67	3.815	4.68
SPync071	800300	800369	70	→ ← →	SPy_0978 / SPy_0979	14	0.543	17	0.968	1.78
SPync072	819045	819088	44	→ → →	SPy_1006 / SPy_1007	-	-	5	0.285	-
SPync073	826031	826112	82	← → →	SPy_1013 / SPy_1016	3	0.116	-	-	-
SPync074 / tracrRNA	**854346**	**854545**	**200**	→ ← →	**SPy_1044 / SPy_1046**	**472**	**18.323**	**139**	**7.916**	**0.43**
SPync075 / pre-crRNA	**860730**	**861239**	**510**	→ → →	**SPy_1049 / SPy_1052**	**84**	**3.261**	**4**	**0.228**	**0.07**

CANDIDATE[a]	COORDINATES[b]		SIZE (nt)	ORIENTATION[c]	ADJACENT GENES[d]	LIBRARY SPY-[e]		LIBRARY SPY+[e]		ENRICHMENT[f]
						OCC.	%	OCC.	%	
SPync076	882398	882493	96	← → →	SPy_1071 / SPy_1072	4	0.155	2	0.114	0.73
SPync077	883527	883559	33	→ → →	SPy_1073 / SPy_1075	2	0.078	-	-	-
SPync078	892663	892687	25	→ → →	SPy_1088 / SPy_1092	2	0.078	-	-	-
SPync079	900294	900327	34	→ → →	SPy_1100 / SPy_1101	4	0.155	-	-	-
SPync080	925914	925944	31	→ → ←	SPy_1129 / SPy_1131	6	0.233	-	-	-
SPync081	930739	930907	169	→ → →	SPy_1135 / SPy_1136	4	0.155	3	0.171	1.10
SPync082	950946	951202	257	→ → →	SPy_1155 / SPy_1156	3	0.116	7	0.399	3.42
SPync083	956690	956718	29	→ → →	SPy_1163 / SPy_1164	-	-	2	0.114	-
SPync084	995996	996147	152	← ← ←	SPy_1211 / SPy_1212	15	0.582	23	1.310	2.25
SPync085	1015415	1015638	224	← ← ←	SPy_1230 / SPy_1232	-	-	6	0.342	-
SPync086	1043592	1043669	78	← ← ←	SPy_1265 / SPy_1267	-	-	7	0.399	-
SPync087	1046345	1046592	248	← → →	SPy_1267 / SPy_1270	12	0.466	25	1.424	3.06
SPync088	1056325	1056373	49	← ← ←	SPy_1281 / SPy_1282	4	0.155	-	-	-
SPync089	1065375	1065433	59	→ → →	SPy_s01 / SPy_1290	4	0.155	-	-	-
SPync090	1110799	1110925	127	← ← →	SPy_1340 / SPy_1343	-	-	7	0.399	-
SPync091	1126627	1126903	277	← ← ←	SPy_1359 / SPy_1361	6	0.233	-	-	-
SPync092	1141311	1141334	24	← ← →	SPy_1373 / SPy_1374	4	0.228	-	-	-
SPync093	1177552	1177592	41	← ← ←	SPy_1416 / SPy_1419	2	0.078	-	-	-
SPync094	1182402	1182452	51	← → →	SPy_1423 / SPy_1424	3	0.116	-	-	-
SPync095	1186759	1186876	118	→ ← ←	SPy_1432 / SPy_1434	5	0.194	-	-	-
SPync096	1219348	1219391	44	← → ←	SPy_1484 / SPy_1485	27	1.048	7	0.399	0.38
SPync097	1221287	1221363	77	→ ← →	SPy_1487 / SPy_1488	4	0.155	-	-	-
SPync098	1221305	1221337	33	→ → →	SPy_1487 / SPy_1488	3	0.116	-	-	-
SPync099	1245098	1245204	107	← ← ←	SPy_1513 / SPy_1514	2	0.078	-	-	-
SPync100	1259958	1260092	135	← ← →	SPy_1531 / SPy_1532	23	0.893	14	0.797	0.89
SPync101	1267768	1267850	83	← ← ←	SPy_1539 / SPy_1541	2	0.078	-	-	-
SPync102	1274705	1274742	38	← ← ←	SPy_1547 / SPy_1548	2	0.114	-	-	-
SPync103	1295681	1295825	145	← ← ←	SPy_1570 / SPy_1571	6	0.233	26	1.481	6.36
SPync104	1329485	1329560	76	→ → →	SPy_1608 / SPy_1610	5	0.194	-	-	-
SPync105	1375180	1375200	21	← ← ←	SPy_1654 / SPy_1656	-	-	2	0.114	-
SPync106	1389798	1389821	24	← ← ←	SPy_1675 / SPy_1676	-	-	4	0.228	-
SPync107	1404702	1404773	72	← ← ←	SPy_1689 / SPy_1691	5	0.194	4	0.228	1.17
SPync108	1404854	1404921	68	← ← ←	SPy_1689 / SPy_1691	2	0.078	-	-	-
SPync109	1424598	1424669	72	← ← ←	SPy_1719 / SPy_1721	2	0.078	-	-	-
SPync110	1434502	1434542	41	→ → ←	SPy_1731 / SPy_1733	2	0.078	-	-	-
SPync111	1443306	1443489	184	→ → →	SPy_1741 / SPy_1742	2	0.078	7	0.399	5.13
SPync112	1471969	1472001	33	→ ← ←	SPy_1780 / SPy_1781	3	0.116	-	-	-
SPync113	1525734	1525787	54	← ← ←	SPy_1837 / SPy_1839	2	0.078	-	-	-
SPync114	1539483	1539578	96	← ← ←	SPy_1854 / SPy_1856	9	0.349	11	0.626	1.79

CANDIDATE[a]	COORDINATES[b]		SIZE (nt)	ORIENTATION[c]	ADJACENT GENES[d]	LIBRARY SPY-[e]		LIBRARY SPY+[e]		ENRICHMENT[f]
						OCC.	%	OCC.	%	
SPync115	1551946	1551986	41	←←→	SPy_1869 / SPy_1870	2	0.078	-	-	-
SPync116	1555688	1555887	200	←→→	SPy_1874 / SPy_1875	4	0.155	5	0.285	1.83
SPync117	1567929	1567948	20	←←←	SPy_1889 / SPy_t52	6	0.233	3	0.171	0.73
SPync118	1571078	1571135	58	←←←	SPy_1894 / SPy_1895	6	0.233	16	0.911	3.91
SPync119	1616478	1616535	58	←←←	SPy_1942 / SPy_1944	3	0.116	-	-	-
SPync120	1620414	1620465	52	←←←	SPy_1946 / SPy_1947	3	0.116	9	0.513	4.40
SPync121	1682437	1682459	23	→←←	SPy_2013 / SPy_2016	2	0.078	-	-	-
SPync122	1696383	1696460	78	←→←	SPy_2034 / SPy_2037	-	-	2	0.114	-
SPync123	1698116	1698270	155	←←←	SPy_2037 / SPy_2038	4	0.155	25	1.424	9.17
SPync124	1699865	1700109	245	←←←	SPy_2039 / SPy_2040	9	0.349	-	-	-
SPync125	1700636	1700687	52	→→→	SPy_2041 / SPy_2042	18	0.699	-	-	-
SPync126	1700505	1700699	195	→←→	SPy_2041 / SPy_2042	7	0.272	34	1.936	7.13
SPync127	1713445	1713593	149	←←←	SPy_2058 / SPy_2059	4	0.155	2	0.114	0.73
SPync128	1713551	1713600	50	←→←	SPy_2058 / SPy_2059	3	0.116	-	-	0.00
SPync129	1718004	1718024	21	→←←	SPy_2063 / SPy_2065	5	0.194	-	-	0.00
SPync130	1721493	1721621	129	→←←	SPy_2066 / SPy_2070	29	1.126	17	0.968	0.86
SPync131	1721510	1721598	89	→→←	SPy_2066 / SPy_2070	16	0.621	-	-	-
SPync132	1755319	1755445	127	→←←	SPy_2099 / SPy_2102	29	1.126	36	2.050	1.82
SPync133	1760101	1760323	223	←←←	SPy_2107 / SPy_2110	-	-	3	0.171	-
SPync134	1774735	1774778	44	←→←	SPy_2122 / SPy_2125	2	0.078	-	-	-
SPync135	1775001	1775122	122	←←←	SPy_2122 / SPy_2125	8	0.311	-	-	-
SPync136	1778711	1778807	97	→←→	SPy_2129 / SPy_2130	-	-	2		-
SPync137	1811913	1811977	65	←←←	SPy_2178 / SPy_2180	-	-	2	0.114	-
SPync138	1820514	1820593	80	←←←	SPy_2186 / SPy_2188	3	0.116	4	0.228	1.96
SPync139	1840683	1840712	30	←←←	SPy_2206 / SPy_2207	2	0.078	2	0.114	1.47
SPync140	1851575	1851593	19	→←→	SPy_2216 / SPy_2217	-	-	2	0.114	-

[a] Candidate names are given for *S. pyogenes* non-coding RNA candidate followed by an arbitrary number: SPyncXXX.
[b] The coordinates are according to the NCBI database (SF370 genome, accession number: NC_002737).
[c] The orientation of the sRNA candidates and surrounding genes is depicted with arrows: →, plus strand; ←, minus strand.
[d] The identity of the surrounding genes is according to KEGG database (http://www.genome.jp/kegg/).
[e] Library SPY- (untreated total RNA) contains 2576 blast reads. Library SPY+ (TEX-treated total RNA) contains 1756 blast reads. Blast reads (less than 20 nt) do not overlap tRNAs, rRNAs, SRP RNA, RNAse P RNA, tmRNA and all annotated ORFs specified in NC_002737. Occurrence of the transcripts (OCC.) and % of representation are given.
[f] Enrichment values represent the ratio between occurrence of transcripts in library SPY+ comparing to that in library SPY- (%SPY+/%SPY-).

Supplementary Table 7. Basepairings between CRISPR repeats and tracrRNA homologues from various bacterial species.

<table>
<tr><th>STRAIN</th><th>NUMBER OF CRISPRs</th><th>CRISPRs associated with tracrRNA[a]</th><th>CRISPR REPEAT:tracrRNA BASEPAIRING[b]</th></tr>
<tr><td>S. pyogenes
SF370</td><td>2</td><td>NC_002737_1</td><td>5′GUUUUAG--AGCUAUGCUGUUUUGAAUGGUCCCAAAAC 3′
||||• |||||||||||||||||||| |||| |
3′ AAAUUGAACGAUACGACAAAACUUACCAAGGUUGUU 5′</td></tr>
<tr><td>S. mutans
UA159</td><td>1</td><td>NC_004350_1</td><td>GUUUUAG--AGCUGUGUUGUUUCGAAUGGUUCCAAAAC
||||• |||||||||||||||||•||||||| ||
AAAUUGAACGACACAACAAAGCUUACUAAGGUUGUG</td></tr>
<tr><td rowspan="2">S. thermophilus
LMD-9</td><td rowspan="2">3</td><td>NC_008532_5</td><td>GUUUUAG--AGCUGUGUUGUUUCGAAUGGUUCCAAAAC
||||• |||||||||||||||||||||| |||| |
AAAUUGAGCGACACAACAAAGCUUACCAAAGUUUGG</td></tr>
<tr><td>NC_008532_2[c]</td><td>GUUUUUGUACUCU-CAAGAUUUAAGUAACUGUACAAC
||| |• ||| |||||||||||||||||| ||
AAACAUCGAAGACGUUCUAAAUUCAUUGACACAUUC</td></tr>
<tr><td>L. innocua
Clip11262</td><td>1</td><td>NC_003212_2</td><td>GUUUUAG--AGCUAUGUUAUUUUGAAUGCUAACAAAAC
||||• |||||||||||||||||•||||||| |
AAAUUGAACGAUACAAUAAAACUUAUGAUUGUUAUA</td></tr>
<tr><td>Treponema denticola
ATCC35405</td><td>1</td><td>NC_002967_1</td><td>GUUUGAG--AGUUGUGUAAUUUAAGAUGGAUCUCAAAC
||||• |||||||||||||||||||||||| |||
AACUUGAGCAACACAUUAAAUUCUACCUAGAAUUUA</td></tr>
<tr><td>N. meningitidis
Z2491</td><td>2</td><td>NC_003116_10</td><td>GUUGUAGCUCCCUUUCUCAUUUCGCAGUGCUACAAU
•||||||| | |||||||||||||||||| |||||
UAACAUCGUUGCCAAGAGUAAAGCGUCACGCUGUUA</td></tr>
<tr><td>Streptococcus gordonii
str. Challis substr. CH1</td><td>1</td><td>NC_009785_2[c]</td><td>GUUUUUGUACUCU-CAAGAUUUAAGUAACUGUACAAC
||| |• |||| |||||||||||||||||| ||
AAACAUCGAAGACGUUCUAAAUUCAUUGACACAUUC</td></tr>
</table>

[a] The CRISPR loci belong to the type II (Nmeni/CASS4) CRISPR/Cas system. Nomenclature is according to the CRISPR database (CRISPR DB) (http://crispr.u-psud.fr/). Note that *S. thermophilus* LMD-9 contains two CRISPR loci with an organization similar to that of CRISPR01 from *S. pyogenes*. Anti-CRISPR repeats located in the vicinity of type II CRISPR/Cas loci, that could be encoded within tracrRNA homologues, were also found in genomes from other bacterial species including: *Streptococcus agalactiae*, *Streptococcus dysgalactiae subsp. equisimilis*, *Streptococcus equi subsp. zooepidemicus*, *Corynebacterium dyphteriae, Campylobacter jejuni* and *Pasteurella multocida*.

[b] Upper sequence, CRISPR repeat consensus sequence; lower sequence, tracrRNA homologue fragment annealing to the repeat; grey-highlighted, bases that differ from the *S. pyogenes* SF370 CRISPR01 repeat and anti-repeat sequences. Although the tracrRNA and CRISPR sequences differ significantly in the analyzed genomes, CRISPR repeat sequences show a certain degree of similarity, especially in the terminal regions and around the putative cleavage site. Notably, despite sequence differences, sequence complementarity in the CRISPR repeat:tracrRNA homologue base pairing remains conserved. A base substitution in a tracrRNA homologue leads to a compensatory substitution in the cognate repeat. This indicates that both sequences co-evolved and further demonstrates that the CRISPR repeat:tracrRNA base pairing is crucial for the functionality of type II CRISPR/Cas systems. The duplexes also differ in the position and length of mismatches which, according to the northern blot data does not interfere with the co-processing (Supplementary Figs 12-16).

[c] For these two CRISPR loci, the tracrRNA homologue-encoding DNA is located between the *csn1* and *cas1* genes.

Supplementary Table 8. Strains used in the study.

Strain	Relevant characteristics	Source
Streptococcus pyogenes		
WT		
EC904	SF370 (M1 serotype)	ATCC 700294
EC1246	MGAS8232 (M18 serotype)	ATCC BAA-572
ΔtracrRNA		
EC1412	SF370 (pEC264) (replicating at 28°C)	This study
EC1414	SF370ΩpEC264 (integration at 37°C)	This study
EC1500	SF370ΔtracrRNA	This study
EC1502	SF370 WT-TṢ after loss of pEC264)	This study
EC1807	EC1500 complemented with pEC268	This study
EC1808	EC1500 complemented with pEC269	This study
Δpre-crRNA		
EC1472	SF370 (pEC273) (replicating at 28°C)	This study
EC1475	SF370ΩpEC273 (integration at 37°C)	This study
EC1479	SF370Δpre-crRNA	This study
EC1484	SF370 WT-TṢ after loss of pEC273)	This study
EC1809	EC1479 complemented with pEC299	This study
Δ*csn1-cas1-cas2-csn2*		
EC1625	SF370 (pEC298) (replicating at 28°C)	This study
EC1626	SF370ΩpEC298 (integration at 37°C)	This study
EC1630	SF370Δ*csn1-cas1-cas2-csn2*	This study
EC1631	SF370 WT-TṢ after loss of pEC298)	This study
EC1691	EC1630 complemented with pEC306	This study
EC1692	EC1630 complemented with pEC308	This study
EC1693	EC1630 complemented with pEC307	This study
EC1686	EC1630 complemented with pEC268	This study
EC1685	EC1630 complemented with pEC269	This study
EC1828	EC1630 complemented with pEC368	This study
Δ*csn1*		
EC1783	SF370 (pEC315) (replicating at 28°C)	This study
EC1785	SF370ΩpEC315 (integration at 37°C)	This study
EC1788	SF370Δ*csn1*	This study
EC1791	SF370 WT-TṢ after loss of pEC315)	This study
EC1801	EC1788 complemented with pEC268	This study
EC1824	EC1788 complemented with pEC368	This study
Δ*cas1*		
EC1654	SF370 (pEC312) (replicating at 28°C)	This study
EC1661	SF370ΩpEC312 (integration at 37°C)	This study
EC1675	SF370Δ*cas1*	This study
EC1681	SF370 WT-TṢ after loss of pEC312)	This study
Δ*cas2*		
EC1658	SF370 (pEC313) (replicating at 28°C)	This study
EC1667	SF370ΩpEC313 (integration at 37°C)	This study
EC1793	SF370Δ*cas2*	This study
EC1794	SF370 WT-TṢ after loss of pEC313)	This study
Δ*csn2*		
EC1656	SF370 (pEC314) (replicating at 28°C)	This study
EC1664	SF370ΩpEC314 (integration at 37°C)	This study
EC1796	SF370Δ*csn2*	This study
EC1798	SF370 WT-TṢ after loss of pEC314)	This study
EC1812	EC1796 complemented with pEC299	This study
EC1810	EC1796 complemented with pEC307	This study
Δ*rnc*		
EC1620	SF370 (pEC297) (replicating at 28°C)	This study

Strain	**Relevant characteristics**	**Source**
EC1633	SF370ΩpEC297 (integration at 37°C)	This study
EC1636	SF370Δ*rnc*	This study
EC1637	SF370 WT-TS after loss of pEC297)	This study
EC1804	EC1636 complemented with pEC366	This study
Streptococcus mutans		
EC1293	UA159 (WT)	Pierre Renault
Streptococcus thermophilus		
EC810	LMD-9 (WT)	Pierre Renault
Staphylococcus epidermidis		
EC1617	RP62a (WT)	Knut Ohlsen, Mohammad Shahrooei
Listeria innocua		
EC917	Clip11262 (WT)	Thomas Decker
Neisseria meningitidis		
Z2491	A Z2491 (WT)	Thomas Meyer
E. coli		
TOP10	Host for cloning	Invitrogen
XL10	Host for cloning	Stratagene

Supplementary Table 9. Plasmids used in the study.

Plasmids	Relevant characteristics	Source
Vectors for *S. pyogenes*		
pEC85	*repDEG*-pAMβ1, pJH1-*aphIII*, ColE1	Lab strain collection
pEC214	*repAts*-pWV01, pJH1-*aphIII*, P*clpB-ßgaB*, ColE1	This study
Plasmids for gene-deletion mutants in *S. pyogenes*		
pEC264	pEC214ΩΔtracrRNAup-ΔtracrRNAdw	This study
pEC273	pEC214ΩΔpre-crRNAup-Δpre-crRNAdw	This study
pEC298	pEC214ΩΔ*csn1-cas1-cas2-csn2*up-Δ*csn1-cas1-cas2-csn2*dw	This study
pEC315	pEC214ΩΔ*csn1*up-Δ*csn1*dw	This study
pEC312	pEC214ΩΔ*cas1*up-Δ*cas1*dw	This study
pEC313	pEC214ΩΔ*cas2*up-Δ*cas2*dw	This study
pEC314	pEC214ΩΔ*csn2*up-Δ*csn2*dw	This study
pEC297	pEC214ΩΔ*rnc*up-Δ*rnc*dw	This study
Complementation plasmids in *S. pyogenes*		
pEC268	pEC85Ω171 tracrRNA (171 nt form)	This study
pEC269	pEC85Ω89 tracrRNA (89 nt form)	This study
pEC299	pEC85Ωpre-crRNA	This study
pEC368	pEC85Ω171 tracrRNA–P*csn1-csn1*	This study
pEC306	pEC85ΩP*csn1-cas1*	This study
pEC308	pEC85ΩP*csn1-cas2*	This study
pEC307	pEC85ΩP*csn1-csn2*	This study
pEC366	pEC85ΩP*rnc-rnc*	This study
Plasmids for CRISPR01 transformation read-out in *S. pyogenes*		
pEC138	pEC85ΩP*sagA*-5'UTR*sagA*	This study
pEC367	pEC85Ω*speM*	This study
pEC287	pEC85ΩP*speM-speM*	This study
pEC289	pEC85ΩP*speM-speM* (A2773G)	This study
pEC290	pEC85ΩP*speM-speM* (T2785C)	This study

Supplementary Table 10. Primers used in the study.

Purpose	Primer code	Sequence 5'-3'[a]	F/R[b]	Usage[c]
Sequencing of CRISPR01 locus in *S. pyogenes* SF370				
lepA	OLEC1142	*GAAATTACGAATCTTCTCCTGACG*	R	PCR
SPy_1559	OLEC1143	*GAATCATCTGCATCCAGTAAAATG*	F	PCR
cas2	OLEC1144	*GGAAAAATTGGCAGAGGCGT*	R	PCR
tracrRNA expression in *S. pyogenes* SF370				
tracrRNA mapping	OLEC702	*ATACTTCTATTCTACTCTGAC*	F	SL
	OLEC703	*CGACTCGGTGCCACTTTTTCA*	R	SL, PE
tracrRNA analysis	OLEC1698	*CAGCCATAAAACAATACTTAATAC*	F	NB (5' probe)
	OLEC1014	*GGACTAGCCTTATTTTAACTTG*	R	NB (3' probe)
	OLEC1697	GAAAT**TAATACGACTCACTATAGGG***CACCGACTCGGTGCCACTTTTTC*	R	NB (3' probe)
	OLEC1299	GAAAT**TAATACGACTCACTATAGG***TTGGAACCATTCAAAACAGCATAGC*	F	RT-PCR
	OLEC1522	*AAAAAAAGCACCGACTCGGTGCCAC*	R	RT-PCR
crRNA (CRISPR01) expression in *S. pyogenes* SF370				
pre-crRNA mapping	OLEC1426	*GACAAATAGTGCGATTAC*	F	SL
	OLEC1425	*AGCTAGACTTCAGTCTTG*	R	SL
	OLEC1427	*AATAAATGCAGTAATACAGG*	F	SL, RT-PCR
	OLEC1428	*CTCGTAGACTATTTTTGTC*	R	SL, RT-PCR
crRNA expression analysis	OLEC1227	*TCTCAGCTAGACTTCAGTCTTGAA*	R	NB (leader)
	OLEC1050	*GCACTATTTGTCTCAGCTAGACTTCAGTCT*	R	NB (leader)
	OLEC1140	*GCAGTAATACAGGGGCTTTTCAAG*	F	NB (use with OLEC1493), RT-PCR
	OLEC1049	*GGACCATTCAAAACAGCATAGCTCTAAAAC*	R	NB (repeat)
	OLEC1228	*CAAAACAGCATAGCTCTAAAAC*	R	NB (repeat)
	OLEC1684	*TGTTCATATAAGTTTTGGGACCAT*		NB (crRNA 24-nt)
	OLEC1493	*GAATAGGAAGGTATCCGACTGCTGG*	R	NB (use with OLEC1140)
	OLEC1141	*CAAATTGAGTTATGTTCATATAAG*	R	RT-PCR
crRNA (CRISPR02) expression in *S. pyogenes* SF370				
CRISPR02 leader	OLEC1219	*GCTTTCACAGAAACACCTTGC*	F	NB, RT-PCR
CRISPR02 spacer 1	OLEC1220	*TTGGTGATTTTCGTGATATGC*	R	NB, RT-PCR
CRISPR02 plus strand	OLEC1180	*CAAAAATAACTTAAAAAAGAAGCGAAATGG*	R	NB, RT-PCR
	OLEC1182	*CATGGGTGAGTGGATTGAAAT*	R	NB, RT-PCR
CRISPR02 minus strand	OLEC1181	*ATTTCAATCCACTCACCCATG*	F	NB, RT-PCR
	OLEC1179	*AACGGATGGCACACATAGAGGTGTTTTAAG*	F	NB, RT-PCR
CRISPR01 Cas gene expression in *S. pyogenes* SF370				
CRISPR01 *csn1*	OLEC1480	*CAGTGCGCAAAGTATTGTCCATGC*	F	RT-PCR
	OLEC1193	*GATTAAGTCTTTTTTAACTTCCTTATATCC*	R	RT-PCR
CRISPR01 *cas1*	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	RT-PCR
	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	RT-PCR
CRISPR01 *cas2*	OLEC1477	*ATACTTATGTTTGATATGCCGACGG*	F	RT-PCR
	OLEC1478	*ACTGTTTTTCCGTGACCGTTAGTAA*	R	RT-PCR
	OLEC1191	*AATAGAAAATTGATGCATGATAAACCCTTC*	R	RT-PCR
CRISPR01 *csn2*	OLEC1476	*GGCGGTACAATTCTTGTGCTCGAAGATG*	F	RT-PCR
	OLEC1190	*CATTTTCTAAGCATTCAAAGACAATCAGTT*	R	RT-PCR
CRISPR02 Cas gene expression in *S. pyogenes* SF370				
CRISPR02 *cas3*	OLEC1487	*GACTAATCTTAGTATGCGTTATGCG*	F	RT-PCR
	OLEC1189	*GCAGGGATTGAACTTCATCTAGAATCACAA*	R	RT-PCR
CRISPR02 *cas5d (csd5d)*	OLEC1486	*GACAGGCACTGAATGGTATCGTTGATG*	F	RT-PCR
	OLEC1188	*AACATTCTCTTGTACCCAAAAACACATCTC*	R	RT-PCR
CRISPR02 *csd1*	OLEC1484	*GCAAGCTAATGAGCGTATTGAAGATG*	F	RT-PCR
	OLEC1186	*CATTGGTACTACGCGTAATTTGAAGGCTGG*	R	RT-PCR
CRISPR02 *csd2*	OLEC1485	*GGAACAGCTTACCGATAAGCATAGAGG*	F	RT-PCR
	OLEC1187	*TATAACAGGAGGCGTATCGTCATCATCATC*	R	RT-PCR
CRISPR02 *cas4*	OLEC1483	*CAATGGCTTGATAATGAAGCGACAGCG*	F	RT-PCR
	OLEC1185	*GTATTCCACAACTTTTGGTAACCATTTCCC*	R	RT-PCR
CRISPR02 *cas1*	OLEC1482	*GTCTGCTTTGGTGAAGTTATGTACGG*	F	RT-PCR

Purpose	Primer code	Sequence 5’-3’[a]	F/R[b]	Usage[c]
	OLEC1184	*AAAGAGTCTTTATTATCTGCTGCCTGAACC*	R	RT-PCR
CRISPR02 *cas2*	OLEC1481	*GACACCTTTATCTGGGTCATAGCTG*	R	RT-PCR
	OLEC1183	*AAAACAGAATTTTGAACACGTTGCCCATAG*	R	RT-PCR
	OLEC1498	*TTGCCAAACTCTGTGTGGACTATG*	F	RT-PCR
Loading controls for Northern blots				
5S rRNA	OLEC287	*AGTTAAGTGACGATAGCCTAG*	F	NB
	OLEC288	*CTAAGCGACTACCTTATCTCA*	R	NB
tmRNA	OLEC448	*TTTGGGGTTGTTACGGATTC*	F	NB
	OLEC449	*AACATATTTGTCTACGTCCA*	R	NB, RT-PCR
	OLEC140	*TTCGACAGGCATTATGAG*	F	RT-PCR
EC1500 (SF370ΔtracrRNA)				
Up fragment (pEC264)	OLEC823	TAGACTCTGCAG*GACTGATATCATGGCAGGTAT*	F	Up
	OLEC824	CTGTAAGAGCTC*TTCTATTCTACTCTGACTGCA*	R	Up
Down fragment (pEC264)	OLEC821	GACAGCGAGCTC*ATAGAATGATAACAAAATAAA*	F	Dw
	OLEC822	AAGTGCGAATTC*ATCAGCATATTGATCTCCAAT*	R	Dw
PCR checking	OLEC486	*TAAAACAAGCGTTTTGAAAGC*	F	PCR
	OLEC487	*TTAATGACCTCCGAAATTAGT*	R	PCR, SB
	OLEC839	*ACCTGTTTTTATTGGCAAGC*	F	PCR, SB
	OLEC840	*GGAGCCTTAGTTATTTCAGT*	R	PCR
	OLEC823	TAGACTCTGCAG*GACTGATATCATGGCAGGTAT*	F	
	OLEC673	*TGTTGGAACCATTCAAAACAG*	F	PCR
SB checking	OLEC703	*CGACTCGGTGCCACTTTTTCA*	R	Internal probe
	OLEC998	CGAGCGGGATCC*GAATTTCTCCTTGATTATTTG*	F	Internal probe
	OLEC836	*TTACTATGTAGCTGGGTATG*	F	Dw probe
	OLEC837	*CAGAATCTTTCAAGGGATCA*	R	Dw probe
Complementation (pEC268)	OLEC997	GCGTCGGGATCC*TTAATGACCTCCGAAATTAGT*	F	PCR
Complementation (pEC269)	OLEC998	CGAGCGGGATCC*GAATTTCTCCTTGATTATTTG*	F	PCR
Complementation (pEC268/pEC269)	OLEC996	GCATCGGAATTC*TGCAGTCAGAGTAGAATAGAA*	R	PCR
EC1479 (SF370Δpre-crRNA)				
Up fragment (pEC273)	OLEC1132	ATGCAGGGATCC*TGAGTTATAGATATATGAGAATG*	F	Up
	OLEC1135	*CTCAAGTTATCATCGGCAATGTTGGTACCAAGAGTAT TATACCATATTTTTAG*	R	LM-PCR
Down fragment (pEC273)	OLEC1134	*CTAAAAATATGGTATAATACTCTTGGTACCAACATTGC CGATGATAACTTGAG*	F	LM-PCR
	OLEC1133	CTGCATGAATTC*TCAATAAGGTGGAAAATGTAGG*	R	Dw
PCR checking	OLEC1136	*CAAAGGATGAAGTGGCTAGTTTAC*	F	PCR
	OLEC1137	*GCTTAACCCCATATGGCTTAATC*	R	PCR
	OLEC1138	*ATGAAGGGAAAGGAGTTCCTG*	F	PCR
	OLEC1139	*TAAAGTCCACATGCCCTGGGG*	R	PCR
	OLEC1140	*GCAGTAATACAGGGGCTTTTCAAG*	F	PCR
	OLEC1141	*CAAATTGAGTTATGTTCATATAAG*	R	PCR, RT-PCR
SB checking	OLEC1140	*GCAGTAATACAGGGGCTTTTCAAG*	F	Internal probe
	OLEC1141	*CAAATTGAGTTATGTTCATATAAG*	R	Internal probe
	OLEC1197	*GAACCGATTCCATTAAGAGGC*	F	Up probe
	OLEC1198	*TCGTAGCAACCAATTTGTCAA*	R	Up probe
	OLEC1199	*TTAAGCCATATGGGGTTAAGC*	F	Dw probe
	OLEC1200	*CGCAAAGAGCATCCTTAGGA*	R	Dw probe
Complementation (pEC299/pEC300)	OLEC1749	GGTGGTGGATCC*CCACGTGAACTATATGATTTTCCGC*	F	PCR
Complementation (pEC299)	OLEC1750	GGTGGTGAATTC*CAACAAGTCTCAGTGTGCTGAAG*	R	PCR
Complementation (pEC300)	OLEC1751	GGTGGTGAATTC*CAATTAATTATTGCGGATATTCCT*	R	PCR
EC1630 (SF370Δ*csn1-cas1-cas2-csn2*)				
Up fragment (pEC298)	OLEC1686	ATGCAGCTGCAG*ATGTGAAGTGGCTTTTTCAC*	F	Up
	OLEC1689	*GAAATAATCTTCATCTAAAATATACTGGTACCGCCGA TATCTAAGCCTATTGAG*	R	LM-PCR
Down fragment (pEC298)	OLEC1688	*CTCAATAGGCTTAGATATCGGCGGTACCAGTATATTT TAGATGAAGATTATTTC*	F	LM-PCR
	OLEC1687	CTGCATGAATTC*CCAGGAGCCTGAGCAAAATTC*	R	Dw
	OLEC1690	*GCTGTGGTATCCTCTTGAGGTGGAC*	F	PCR

Purpose	Primer code	Sequence 5'-3'[a]	F/R[b]	Usage[c]
PCR checking	OLEC1705	*GAACCTTATATTCATCAGTG*	R	PCR
	OLEC1724	GGTGGTCATATG*AATCTTAATTTTTCCTTACTAG*	F	PCR
	OLEC1476	*GGCGGTACAATTCTTGTGCTCGAAGATG*	F	PCR
	OLEC1140	*GCAGTAATACAGGGGCTTTTCAAG*	F	PCR
	OLEC996	GCATCGGAATTC*TGCAGTCAGAGTAGAATAGAA*	R	PCR
	OLEC1493	*GAATAGGAAGGTATCCGACTGCTGG*	R	PCR
	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	PCR
	OLEC1478	*ACTGTTTTTCCGTGACCGTTAGTAA*	R	PCR
	OLEC1014	*GGACTAGCCTTATTTTAACTTG*	F	PCR
	OLEC1141	*CAAATTGAGTTATGTTCATATAAG*	R	PCR
	OLEC1142	*GAAATTACGAATCTTCTCCTGACG*	R	PCR
SB checking	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	Internal probe
	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	Internal probe
	OLEC998	CGAGCGGGATCC*GAATTTCTCCTTGATTATTTG*	F	Up probe
	OLEC996	GCATCGGAATTC*TGCAGTCAGAGTAGAATAGAA*	R	Up probe
	OLEC1692	*GCTAGTTCGAAATACTTGGCTAAG*	F	Dw probe
	OLEC1687	CTGCATGAATTC*CCAGGAGCCTGAGCAAAATTC*	R	Dw probe
EC1788 (SF370Δ*csn1*)				
Up fragment (pEC315)	OLEC1686	ATGCAGCTGCAG*ATGTGAAGTGGCTTTTTCAC*	F	Up
	OLEC1769	*CCACAACAGTACGCCAACCAGCCATTTTTGCCTCCTA AAATAAAAAGTTTA*	R	LM-PCR
Down fragment (pEC315)	OLEC1770	*TAAACTTTTTATTTTAGGAGGCAAAAATGGCTGGTTG GCGTACTGTTGTGG*	F	LM-PCR
	OLEC1771	CTGCATGAATTC*GGCTTTTCGTTCCTCAGCG*	R	Dw
PCR checking	OLEC1014	*GGACTAGCCTTATTTTAACTTG*	F	PCR
	OLEC1690	*GCTGTGGTATCCTCTTGAGGTGGAC*	F	PCR
	OLEC1478	*ACTGTTTTTCCGTGACCGTTAGTAA*	R	PCR
	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	PCR
	OLEC1774	*GAAATCATCAAACTCATTATGG*	F	PCR
	OLEC1775	*CTTATAGGATAATTTCGAGTGGG*	F	PCR
	OLEC1480	*CAGTGCGCAAAGTATTGTCCATGC*	F	PCR
	OLEC1193	*GATTAAGTCTTTTTTAACTTCCTTATATCC*	R	PCR
SB checking	OLEC1480	*CAGTGCGCAAAGTATTGTCCATGC*	F	Internal probe
	OLEC1193	*GATTAAGTCTTTTTTAACTTCCTTATATCC*	R	Internal probe
	OLEC996	GCATCGGAATTC*TGCAGTCAGAGTAGAATAGAA*	R	Up probe
	OLEC998	CGAGCGGGATCC*GAATTTCTCCTTGATTATTTG*	F	Up probe
	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	Dw probe
	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	Dw probe
Complementation (pEC368)	OLEC2066	GGTGGTCTGCAG*GTTTGCAGTCAGAGTAGAATAGAA G*	F	PCR
	OLEC1744	CTGCATCCCGGG*GGATAATTTCGAGTGGGTATTTAC*	R	PCR
Complementation (pEC369)	OLEC1718	GGTGGTGTCGAC*ATGGATAAGAAATACTCAATAGGC*	F	PCR
	OLEC1744	CTGCATCCCGGG*GGATAATTTCGAGTGGGTATTTAC*	R	PCR
EC1675 (SF370Δ*cas1*)				
Up fragment (pEC312)	OLEC1763	ATGCAGCTGCAG*GAAATTACACTTGCAAATGGAGAG*	F	Up
	OLEC1764	*CATTCTCATATATCTATAACTCATCAGTCACCTCCTAG CTGACTCAAATC*	R	LM-PCR
Down fragment (pEC312)	OLEC1765	*GATTTGAGTCAGCTAGGAGGTGACTGATGAGTTATAG ATATATGAGAATG*	F	LM-PCR
	OLEC1766	CTGCATGAATTC*AATTTCGTAATCGCACTATTTG*	R	Dw
PCR checking	OLEC1767	*CGTCGTTGGAACTGCTTTGATTAAG*	F	PCR
	OLEC1768	*CAATCCATCACTGGTCTTTATGA*	F	use with OLEC1478
	OLEC1478	*ACTGTTTTTCCGTGACCGTTAGTAA*	R	PCR
	OLEC1701	*GCGCAATTAATTATTGCGGATATTCC*	R	SEQ
	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	PCR
	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	PCR
SB checking	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	Internal probe
	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	Internal probe
	OLEC1480	*CAGTGCGCAAAGTATTGTCCATGC*	F	Up probe
	OLEC1193	*GATTAAGTCTTTTTTAACTTCCTTATATCC*	R	Up probe
	OLEC1476	*GGCGGTACAATTCTTGTGCTCGAAGATG*	F	Dw probe
	OLEC1198	*TCGTAGCAACCAATTTGTCAA*	R	Dw probe
	OLEC1746	*CACAACAGTACGCCAACCAGCCATTTTTGCCTCCTAA*	R	LM-PCR

Purpose	Primer code	Sequence 5'-3'[a]	F/R[b]	Usage[c]
Complementation (pEC306)		*AATAAAAAG*		Use with OLEC1743
	OLEC1747	*CTTTTTATTTTAGGAGGCAAAAATGGCTGGTTGGCGT ACTGTTGTG*	F	LM-PCR
	OLEC1748	CTGCATGAATTC*TCATATCCTAAATTCAGGAAC*	R	PCR
EC1793 (SF370Δ*cas2*)				
Up fragment (pEC313)	OLEC1776	ATGCAGCTGCAG*CATCAATCCATCACTGGTCT*	F	Up
	OLEC1777	*CTAGTAAGGAAAAATTAAGATTCATATCCTAAATTCAG GAACTCCTTTCCC*	R	LM-PCR
Down fragment (pEC313)	OLEC1778	*GGGAAAGGAGTTCCTGAATTTAGGATATGAATCTTAA TTTTTCCTTACTAG*	F	LM-PCR
	OLEC1779	CTGCATGAATTC*GCTCTAAAACAAAAAGCGCAAG*	R	Dw
PCR checking	OLEC1477	*ATACTTATGTTTGATATGCCGACGG*	F	Internal
	OLEC1478	*ACTGTTTTTCCGTGACCGTTAGTAA*	R	Internal
	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	Up-dw
	OLEC1198	*TCGTAGCAACCAATTTGTCAA*	R	Up-dw
	OLEC1773	*CGCTGCTTTTAAATATTTTGATAC*	F	Upup-dwdw
	OLEC1701	*GCGCAATTAATTATTGCGGATATTCC*	R	Upup-dwdw
	OLEC1476	*GGCGGTACAATTCTTGTGCTCGAAGATG*	F	SEQ
SB checking	OLEC1477	*ATACTTATGTTTGATATGCCGACGG*	F	Internal probe
	OLEC1478	*ACTGTTTTTCCGTGACCGTTAGTAA*	R	Internal probe
	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	Up probe
	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	Up probe
	OLEC1476	*GGCGGTACAATTCTTGTGCTCGAAGATG*	F	Dw probe
	OLEC1198	*TCGTAGCAACCAATTTGTCAA*	R	Dw probe
Complementation (pEC308)	OLEC1752	*GTATCATTCTCATATATCTATAACTCATTTTTGCCTCC TAAAATAAAAAG*	R	LM-PCR Use with OLEC1743
	OLEC1753	*CTTTTTATTTTAGGAGGCAAAAATGAGTTATAGATATA TGAGAATGATAC*	F	LM-PCR
	OLEC1754	CTGCATGAATTC*TTAAGATTCATCAAAAGCCTC*	R	PCR
EC1796 (SF370Δ*csn2*)				
Up fragment (pEC314)	OLEC1780	ATGCAGCTGCAG*GATTTTGAATCAATCTTGC*	F	Up
	OLEC1781	*GTATTACTGCATTTATTAAGAGTATTAATTAAGATTCA TCAAAAGCCTCCC*	R	LM-PCR
Down fragment (pEC314)	OLEC1782	*GGGAGGCTTTTGATGAATCTTAATTAATACTCTTAATA AATGCAGTAATAC*	F	LM-PCR
	OLEC1783	CTGCATGAATTC*CCAGGAGCCTGAGCAAAATTC*	R	Dw
PCR checking	OLEC1493	*GAATAGGAAGGTATCCGACTGCTGG*	R	SEQ
	OLEC1768	*CAATCCATCACTGGTCTTTATGA*	F	PCR
	OLEC1701	*GCGCAATTAATTATTGCGGATATTCC*	R	SEQ
	OLEC1476	*GGCGGTACAATTCTTGTGCTCGAAGATG*	F	PCR
	OLEC1198	*TCGTAGCAACCAATTTGTCAA*	R	PCR
	OLEC1477	*ATACTTATGTTTGATATGCCGACGG*	F	PCR
	OLEC1050	*GCACTATTTGTCTCAGCTAGACTTCAGTCT*	R	PCR
	OLEC1142	*GAAATTACGAATCTTCTCCTGACG*	R	PCR
SB checking	OLEC1476	*GGCGGTACAATTCTTGTGCTCGAAGATG*	F	Internal probe
	OLEC1198	*TCGTAGCAACCAATTTGTCAA*	R	Internal probe
	OLEC1192	*GAGAATTTATTTTAATACACTTTTTGGGAA*	F	Up probe
	OLEC1479	*CTGATTTCGATTTTCATAAACAATC*	R	Up probe
	OLEC1140	*GCAGTAATACAGGGGCTTTTCAAG*	F	Dw probe
	OLEC1701	*GCGCAATTAATTATTGCGGATATTCC*	R	Dw probe
Complementation (pEC307)	OLEC1755	*CATCTAGTAAGGAAAAATTAAGATTCATTTTTGCCTCC TAAAATAAAAAG*	R	LM-PCR Use with OLEC1743
	OLEC1756	*CTTTTTATTTTAGGAGGCAAAAATGAATCTTAATTTTT CCTTACTAGATG*	F	LM-PCR
	OLEC1757	CTGCATGAATTC*TTATACCATATTTTTAGTTA*	R	PCR
EC1636 (SF370Δ*rnc*)				
Up fragment (pEC297)	OLEC1654	ATGCAGGGATCC*TGTCAGTGATAGGATGACAGG*	F	Up
	OLEC1657	*CATTTTTAGCAGCATCTTGCTCCG*GGTACC*TTAAATTG AATATCAAAAGATG*	R	LM-PCR
Down fragment (pEC297)	OLEC1656	*CATCTTTTGATATTCAATTTAA*GGTACC*CGGAGCAAGA TGCTGCTAAAAATG*	F	LM-PCR

Purpose	Primer code	Sequence 5'-3'[a]	F/R[b]	Usage[c]
	OLEC1655	CTGCATGAATTC*CTGCTTGAAAACCATCCAATTG*	R	Dw
PCR checking	OLEC1658	*GTCAGCCATGACGCTATTGATCCTC*	F	PCR
	OLEC1659	*CTAATGCATTGCTCAATTCGTTTAAC*	R	PCR
	OLEC1661	*GATATGTCTAAGTTACGTTCCATG*	F	PCR
	OLEC1730	*GCATCTTGCTCCGCTAATTTTTTAG*	R	PCR
	OLEC1665	*GATGTTTAATAAATCACGAAAAATC*	F	PCR
	OLEC1666	*CCTACGACTGCAGTTACTCCCTTATC*	R	PCR
	OLEC1660	*GGCATATTTACCGAAATGGAGATAG*	F	SEQ
SB checking	OLEC1661	*GATATGTCTAAGTTACGTTCCATG*	F	Internal probe
	OLEC1730	CTGCATGAATTC*GCATGGGCAGGTCCTTTCTCAC*	R	Internal probe
	OLEC1663	*GTACCTACTGCTAAATCTGCCATAG*	R	Up probe
	OELC1654	ATGCAGGGATCC*TGTCAGTGATAGGATGACAGG*	F	Up probe
	OLEC1664	*CTAGTCAAAGATATTGATATAGCTC*	F	Dw probe
	OLEC1655	CTGCATGAATTC*CTGCTTGAAAACCATCCAATTG*	R	Dw probe
Complementation (pEC322)	OLEC1667	ATGCACGGATCC*CCTGTAGTTTTGGCTTGTCTGATC*	F	PCR
	OLEC1668	ATGCACGAATTC*CTTTTAAAAACATCTAAACCTCAC*	R	PCR
pre-crRNA and tracrRNA homologue expression and processing in *S. mutans* UA159				
crRNA repeat	OLEC1679	*GAACCATTCGAAACAACACAGCTCTAAAAC*		NB
crRNA (repeat-spacer)	OLEC1680	*GCTCTAAAACTCACCATATTAATTAATG*		NB
tracrRNA homologue	OLEC1678	*CACTGCCTTATTTTAACTTGCTGTG*		NB
pre-crRNA and tracrRNA homologue expression and processing in *S. thermophilus* LMD-9 (1st CRISPR locus)				
crRNA repeat	OLEC1673	*GAACCATTCGAAACAACACAGCTCTAAAAC*		NB
crRNA (repeat-spacer)	OLEC1674	*CAGCTCTAAAACTAAGTACTCGTAC*		NB
tracrRNA homologue	OLEC1672	*GCCTTATTTTAACTCGCTGTGTTG*		NB
pre-crRNA and tracrRNA homologue expression and processing in *S. thermophilus* LMD-9 (2nd CRISPR locus)				
crRNA repeat	OLEC1670	*GTTGTACAGTTACTTAAATCTTGAGAGTAC*		NB
crRNA (repeat-spacer)	OLEC1671	*GAGAGTACAAAAACGTTAGTAGATGAC*		NB
tracrRNA homologue	OLEC1669	*CGGCATGAAGCCTTATCTTTGTAGC*		NB
pre-crRNA and tracrRNA homologue expression and processing in *S. epidermidis* RP62a				
crRNA repeat	OLEC1712	*GTTCTCGTCCCCTTTTCTTCGGGGTGGGTATC*		NB
crRNA (repeat-spacer)	OLEC1714	*GGCTTTAAAGAACTATGATTCGATAACCACC*		NB
Additional repeats	OLEC1713	*GGTGGGTATCGATCCTTTGTACTGATG*		NB
pre-crRNA and tracrRNA homologue expression and processing in *L. innocua* Clip11262				
crRNA repeat	OLEC1676	*GTTAGCATTCAAAATAACATAGCTCTAAAAC*		NB
crRNA (repeat-spacer)	OLEC1677	*CATAGCTCTAAAACTATAATGAGTTAG*		NB
tracrRNA homologue	OLEC1675	*CGGACAAAGCCTTATTTTAACTTGC*		NB
pre-crRNA and tracrRNA homologue expression and processing in *N. meningitidis* Z2491				
crRNA (repeat-spacer)	JVO-4863	*CGAAATGAGAAAGGGAGCTACAAC*		NB
tracrRNA homologue	JVO-4862	*ATCTTTTCAGACGGCCTTATTGTAGC*		NB
***In vitro* tracrRNA and CRISPR01 crRNAs**				
T7-tracrRNAs	JVO-4831	GTTTTTTT**TAATACGACTCACTATAGG**G*TATTAAGTATTGTTTTATGGCTGATAAAT*	F	T7-tracrRNA (167 nt form)
	JVO-4832	GTTTTTTT**TAATACGACTCACTATAGG**G*TTGGAACCATTCAAAACAGC*	F	T7-tracrRNA (86 nt form)
	JVO-4833	*AAAAGCACCGACTCGGTG*	R	T7-tracrRNA (both forms)
T7-crRNAs	JVO-4837	GTTTTTTT**TAATACGACTCACTATAGG***TTGATTTCTTCTTGCGCTT*	F	T7-cr148 (1st spacer to 3rd spacer)
	JVO-4841	*CAATTAATTATTGCGGATATTCCT*	R	T7-cr148 (1st spacer to 3rd spacer)
	JVO-4836	GTTTTTTT**TAATACGACTCACTATAGG***AATAAATGCAGTAATACAGGGG*	F	T7-cr213 (leader to 2nd repeat-spacer)
	JVO-4840	*AGTTATGTTCATATAAGTTTTGGGAC*	R	T7-cr213 (leader to 2nd repeat-spacer)
Mismatch introduction	JVO-7371	TTCATTTGACGTATCGTCTAAAACAAAAAGCGCAA	R	cr148-M12-A fragment
	JVO-7374	*TAGACGATACGTCAAATGAATGGTCCCAAAACTTAT*	F	cr148-M12-B fragment
	JVO-7377	*CTTCGATACGACAAATGAATGGTTCCAACGGAGA*	R	tracrRNA-M12-

Purpose	Primer code	Sequence 5'-3'[a]	F/R[b]	Usage[c]
				A fragment
	JVO-7380	TCATTTGTCGTATCGAAGTTAAAATAAGGCTAGTCCG	F	tracrRNA-M12-B fragment
***speM* expression in *S. pyogenes* MGAS8232 (M18)**				
speM	OLEC1557	*TGAGAGTGTCTTTTCAGATGCTGTG*	F	RT-PCR
	OLEC1558	*AGCGGTATCTGTTCCCCAAAA*	R	RT-PCR, NB
Plasmids used for CRISPR01 transformation read-out in *S. pyogenes*				
pEC367	OLEC2071	ATGCAC<u>CTGCAG</u>*TGAAAAAAAATACCTTG*	F	Promoter less-*speM*
	OLEC1556	ATGCAC<u>GGATCC</u>*AGCTGCTCTAAGTCTACTAAAACC*	R	Promoter less-*speM*
pEC287	OLEC1555	ATGCAC<u>CTGCAG</u>*GATGCTGTTAAATAATTGTATAC*	F	P*speM*-*speM*
	OLEC1556	ATGCAC<u>GGATCC</u>*AGCTGCTCTAAGTCTACTAAAACC*	R	*speM*
pEC289	OLEC1593	ATGAGCGATAATAGAATAAAGTTATATGAGCATAACT CAATTTGTAAAAAAGGGTATTG	F	GAA->GAG
	OLEC1594	CAATACCCTTTTTTACAAATTGAGTTATGCTCATATAA CTTTATTCTATTATCGCTCAT	R	GAA->GAG
pEC290	OLEC1595	GCGATAATAGAATAAAGTTATATGAACATAACTCAATC TGTAAAAAAGGGTATTGGG	F	ATT->ATC
	OLEC1596	CCCAATACCCTTTTTTACAGATTGAGTTATGTTCATAT AACTTTATTCTATTATCGC	R	ATT->ATC
pEC138	OLEC132	ATCTAG<u>CTGCAG</u>*TTCAATTGAAAAGGCAAC*	F	*PsagA*-5'UTR-*sagA*
	OLEC219	TTTGAA<u>GGATCC</u>*AAGGTTTACCTCCTTATCTAATAAG*	R	*PsagA*-5'UTR-*sagA*
Verification of plasmid constructs by PCR or sequencing analysis				
ColE1	OLEC1560	*GAAGTACATCCGCAACTGTC*	F	SEQ
	OLEC1562	*GTGATGCTCGTCAGGGGG*	F	SEQ
repAts	oliRN228	*GGAACGAAAACTCACGTTAA*	R	SEQ
	oliRN229	*AGGTTCTTGATGCTGAAACG*	R	SEQ

[a] *italic*, sequence annealing to the template; <u>underlined</u>, restriction site; **bold**, T7 promoter.
[b] F, forward primer; R, reverse primer.
[c] LM-PCR, PCR-mediated ligation; NB, probe for Northern blot; PE, primer extension; SB, probe for Southern blot; SEQ, sequencing SL, self-ligation (RNA mapping by head-to-tail circularization).

References to Supplementary Tables

1. Ferretti, J. J. *et al.* Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4658-4663 (2001).
2. Sumby, P. *et al.* Evolutionary origin and emergence of a highly successful clone of serotype M1 group A Streptococcus involved multiple horizontal gene transfer events. *J. Infect. Dis.* **192**, 771-782 (2005).
3. Beres, S. B. *et al.* Molecular genetic anatomy of inter- and intraserotype variation in the human bacterial pathogen group A Streptococcus. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7059-7064 (2006).
4. Beres, S. B. *et al.* Genome sequence of a serotype M3 strain of group A Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc. Natl. Acad. Sci. U S A* **99**, 10078-10083 (2002).
5. Nakagawa, I. *et al.* Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Res.* **13**, 1042-1055 (2003).
6. Holden, M. T. *et al.* Complete genome of acute rheumatic fever-associated serotype M5 *Streptococcus pyogenes* strain manfredo. *J. Bacteriol.* **189**, 1473-1477 (2007).
7. Banks, D. J. *et al.* Progress toward characterization of the group A Streptococcus metagenome: complete genome sequence of a macrolide-resistant serotype M6 strain. *J. Infect. Dis.* **190**, 727-738 (2004).
8. Smoot, J. C., Korgenski, E. K., Daly, J. A., Veasy, L. G. & Musser, J. M. Molecular analysis of group A Streptococcus type emm18 isolates temporally associated with acute rheumatic fever outbreaks in Salt Lake City, Utah. *J. Clin. Microbiol.* **40**, 1805-1810 (2002).
9. Green, N. M. *et al.* Genome sequence of a serotype M28 strain of group A streptococcus: potential new insights into puerperal sepsis and bacterial disease specificity. *J. Infect. Dis.* **192**, 760-770 (2005).
10. McShan, W. M. *et al.* Genome sequence of a nephritogenic and highly transformable M49 strain of *Streptococcus pyogenes*. *J. Bacteriol.* **190**, 7773-7785 (2008).
11. Horvath, P. *et al.* Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.* **131**, 62-70 (2009).
12. Vojtek, I. *et al.* Lysogenic transfer of group A Streptococcus superantigen gene among streptococci. *J. Infect. Dis.* **197**, 225-234 (2008).