A COMPLETE INVENTORY OF *S. AUREUS* SMALL RNAs AND THEIR TEMPORAL EXPRESSION DURING GROWTH PHASE

Marie Beaume[1*], David Hernandez[1], Laurent Farinelli[2], Jacques Schrenzel[1], Patrice François[1]

1 Genomic Research Laboratory - University Hospital of Geneva, Switzerland (www.genomic.ch)
2 Fasteris SA, Plan-les-Ouates, Switzerland (www.fasteris.com)

*e-mail: marie.beaume@genomic.ch

Overview

Severity of staphylococcal infections varies drastically from benign skin infections to life-threatening diseases. *Staphylococcus aureus* is responsible for both nosocomial- and community-acquired infections which can be either acute or chronic. However, neither epidemicity nor virulence is yet clearly linked to the genome content.
We describe here the first transcriptional analysis of a clinical isolate of *S. aureus* using high-throughput sequencing. Samples collected at different times of the growth phase were sequenced and quantified to detect unknown transcripts. Approximately 100 small RNA molecules were identified in regions considered formerly as intergenic regions. Transcript sizes are generally around 150 bp and RNA larger than 300 are rare. These 100 transcripts are not co-localised at specific sites but dispersed along the chromosome and the plasmid of the bacterium. Some of the small RNAs are localized in biologically or clinically relevant regions; between key metabolic genes or pathogenicity islands as well as in the methicillin resistance element and in factors involved in bacterial virulence. Most of these small transcripts are highly conserved across the different published genomes and their secondary structure prediction is well correlated with their potential role in regulatory processes. We are currently working on the elucidation of the role of some of these small RNAs on the virulence of *S. aureus*.

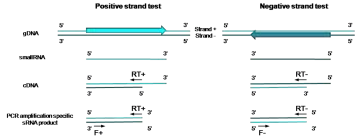
Material & Methods

Bacterial strain and culture conditions.
Bacteria strain used for RNA extraction was *S. aureus* N315 at growth times : 2 hours, 4h, 6h, 8h, in Muller Hinton Broth.

RNA extraction and purification.
Total RNA of N315 was extract with RNeasy kit (QIAGEN) and mirVana isolation kit (AMBION). The extracts were treated with MICROBExpress kit (AMBION) to limit contamination by structural 16S and 23S ribosomal RNAs.

RNA sequencing (Illumina-Solexa).
RNA samples were subjected to Illumina Genome Analyzer. Briefly, 2-10 ug of enriched RNA was physically fragmented by nebulization or by zinc treatment respectively after and before cDNA synthesis. After end-reparation and ligation of the adaptors, the products were purified on agarose gel to recover fragments around 150-bp. These samples were then used to generate DNA colonies for sequencing.
The standard procedure for transcriptome sequencing analysis consists in mapping the sequence reads onto the genomic sequence (RefSeqNC_2745 and NC_3140). From this mapping, a coverage profile is computed by counting the number of times each base in the genome is covered by a sequence read.

Identification of coding strand by Reverse Transcriptase-PCR.
sRNA specific primers have been designed for determining the coding strand of this sRNA from total RNA extract. The two possibility (positive strand and negative strand) have been studied in order to have a single final amplification corresponding to the coding strand (see below).



5' RACE-PCR.
The smart RACE cDNA amplication kit (Clontech) allowed us to perform 5'-rapid amplification and sequencing of cDNA ends from RNA extract containing the sRNAs. We have used putative sRNA specific primers to determine the 5' sRNA extremity and the orientation.

Results

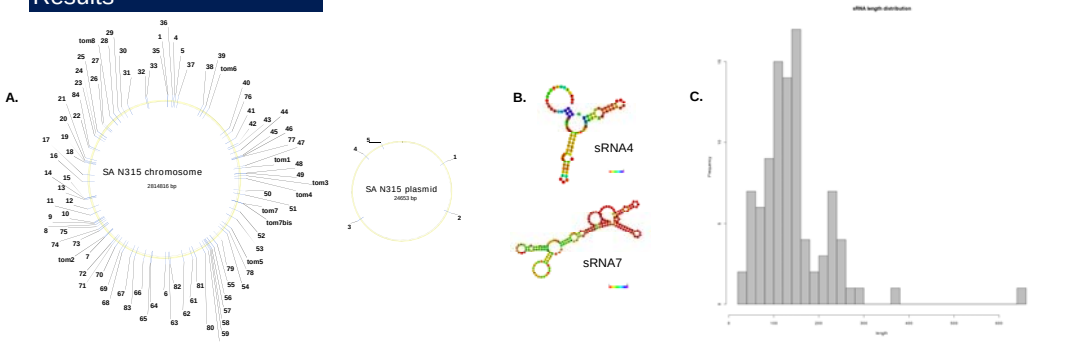


Figure 1 : sRNAs in *S. aureus* N315 chromosome and plasmid
A.Total RNA sequencing allowed to identify 95 sRNAs in the 2.8 Mb chromosome and 5 sRNAs in the 24 kb plasmid of strain N315. They have a homogenous distribution around the genome and some of them are localized in biologically or clinically relevant regions such as metabolic genes or virulence genes.
B. Among them, 24 sRNAs were found in the Génoscope database (France) for their secondary structure. RNA fold web server was also used to predict putative secondary structures. This figure shows the structure drawing encoding base-pair probabilities (on the color-scale, red corresponds to highly probable structures whereas blue are used for low probability). These predicted structures confirmed previous findings from other identified sRNA.
C. sRNAs length distribution shows that most of the sRNAs are included in the range of 30 to 300 pb.

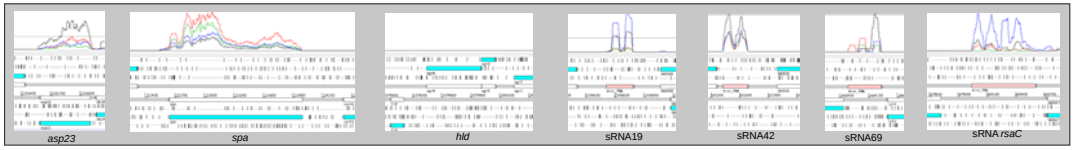


Figure 2 : Sequencing coverage of some sRNAs during the *S. aureus* N315 growth.
Total RNA sequencing allowed to quantify RNA expression during growth phase: 2 hours (red), 4h (green), 6h (blue), 8h (black).
We can visualize the 6 potential frame-shifts with the associated stop codon (black bars), the annotated genes and the intergenic region showing transcription, a signal in this region correspond to a potential small RNA.
The expression profile of control genes have been inspected for experimental validation. For example, *asp23*, a target gene reflecting *sig B* expression, and *hld* (delta-heamolysin) which is not express in N315, or *spa* (protein A) expression are depicted on the figure. The general profiles are accurate with published studies related to the temporal expression of these genes. The experiment shows that the expression of most sRNAs identified in this study is regulated during the growth phase.

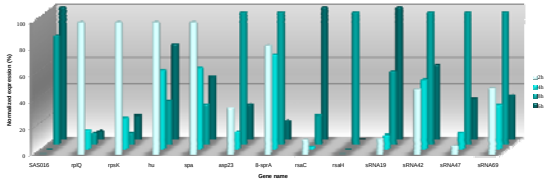


Figure 3 : Relative expression quantification of some sRNAs during the *S. aureus* N315 growth. This histogram shows the normalized genes and sRNAs relative expression which are depicted in the Figure 2.

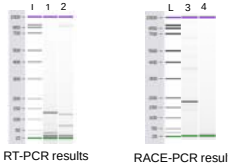


Figure 4 : Evaluation of sRNA1 orientation by RT-PCR and RACE-PCR using Bioanalyzer . These two sRNA1-specific experiments allowed determining the coding strand and consequently the functional sequence. The sRNA sequence corresponding to the positive strand according to RefSeq NC_002745.
L: ladder, **1**: cDNA PCR amplification with theorical positive strand sRNA1 primers, **2**: cDNA PCR amplification with theorical negative strand sRNA1 primers, **3**: cDNA 5'-RACE amplification with theoretical positive strand sRNA1 primers, 4: cDNA 5'-RACE amplification with theoretical negative strand sRNA1 primers

Conclusions and Outlooks

- High-throughput sequencing of RNA allows quantifying the transcriptome expression and identify small RNAs in a single experiment.
- Numerous transcripts were identified in regions described as "non-coding" or intergenic.
- We have detected in the genome of *S. aureus* N315 approximately 100 putative sRNA molecules which are transcript at least at one ofthe studied time of growth.
- These 100 sRNAs are distributed homogenously around the chromosome but not clustered at specific locations.
- The identification of small RNAs is not trivial : some transcripts correspond to potential (not yet annotated) CDS or to potential operon (see below).

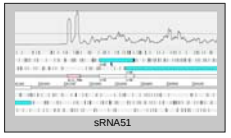


Figure 5 : Illustration of the difficulty to evaluate some of putative sRNAs from RNA Sequencing.

- RT-PCR and RACE-PCR allow determining sRNA coding strand. For example, sRNA1 sequence corresponds to the positive strand according to RefSeq NC_002745.
- Studies are ongoing to evaluate the expression of these sRNAs at the genome–scale, in various environmental conditions (stress, media, selection pressures).
- Finally, we plan to determine the role of these sRNAs notably in the context of *S. aureus* virulence (regulation of virulence factor expression, environmental adaptation, …)

Literature cited

- Pichon C, Felden B. Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. Proc Natl Acad Sci U S A. 2005 Oct 4;102(40):14249-54
- Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. 2008 May;18(5):802-9

Acknowledgments

This work was supported by the SNF (3100A0-116075).

FNSNF
FONDS NATIONAL SUISSE DE LA RECHERCHE SCIENTIFIQUE