Cartography of methicillin-resistant S. aureus transcripts: detection, orientation and temporal expression during growth phase and stress conditions

BEAUME, Marie-Emilie, et al.

Abstract

BACKGROUND: Staphylococcus aureus is a versatile bacterial opportunist responsible for a wide spectrum of infections. The severity of these infections is highly variable and depends on multiple parameters including the genome content of the bacterium as well as the condition of the infected host. Clinically and epidemiologically, S. aureus shows a particular capacity to survive and adapt to drastic environmental changes including the presence of numerous antimicrobial agents. Mechanisms triggering this adaptation remain largely unknown despite important research efforts. Most studies evaluating gene content have so far neglected to analyze the so-called intergenic regions as well as potential antisense RNA molecules.

PRINCIPAL FINDINGS: Using high-throughput sequencing technology, we performed an inventory of the whole transcriptome of S. aureus strain N315. In addition to the annotated transcription units, we identified more than 195 small transcribed regions, in the chromosome and the plasmid of S. aureus strain N315. The coding strand of each transcript was identified and structural analysis enabled classification of [...]
Cartography of Methicillin-Resistant \textit{S. aureus} Transcripts: Detection, Orientation and Temporal Expression during Growth Phase and Stress Conditions

Marie Beaume$^{1*}$, David Hernandez$^1$, Laurent Farinelli$^2$, Cécile Deluen$^2$, Patrick Linder$^3$, Christine Gaspin$^4$, Pascale Romby$^5$, Jacques Schrenzel$^1$, Patrice Francois$^1$

$^1$Genomic Research Laboratory, Infectious Diseases Service, Geneva University Hospitals, Geneva, Switzerland, $^2$Fasteris SA, Plan-les-Ouates, Switzerland, $^3$Department of Microbiology and Molecular Medicine, University Medical Center, Genève, Switzerland, $^4$INRA, UBI and Plateforme GenoToul Bioinfo, UR 875, Castanet-Tolosan, France, $^5$Architecture et Réactivité de FARN, Université de Strasbourg, CNRS, IBMC, Strasbourg, France

Abstract

\textbf{Background:} \textit{Staphylococcus aureus} is a versatile bacterial opportunist responsible for a wide spectrum of infections. The severity of these infections is highly variable and depends on multiple parameters including the genome content of the bacterium as well as the condition of the infected host. Clinically and epidemiologically, \textit{S. aureus} shows a particular capacity to survive and adapt to drastic environmental changes including the presence of numerous antimicrobial agents. Mechanisms triggering this adaptation remain largely unknown despite important research efforts. Most studies evaluating gene content have so far neglected to analyze the so-called intergenic regions as well as potential antisense RNA molecules.

\textbf{Principal Findings:} Using high-throughput sequencing technology, we performed an inventory of the whole transcriptome of \textit{S. aureus} strain N315. In addition to the annotated transcription units, we identified more than 195 small transcribed regions, in the chromosome and the plasmid of \textit{S. aureus} strain N315. The coding strand of each transcript was identified and structural analysis enabled classification of all discovered transcripts. RNA purified at four time-points during the growth phase of the bacterium allowed us to define the temporal expression of such transcripts. A selection of 26 transcripts of interest dispersed along the intergenic regions was assessed for expression changes in the presence of various stress conditions including pH, temperature, oxidative shocks and growth in a stringent medium. Most of these transcripts showed expression patterns specific for the defined stress conditions that we tested.

\textbf{Conclusions:} These RNA molecules potentially represent important effectors of \textit{S. aureus} adaptation and more generally could support some of the epidemiological characteristics of the bacterium.

Introduction

\textit{Staphylococcus aureus} is a ubiquitous Gram-positive bacterium responsible for a wide spectrum of infections ranging from localized to life-threatening diseases. Since the worldwide emergence and spread of methicillin-resistant \textit{S. aureus} (MRSA), this bacterium has revealed a particular capacity to survive and adapt to drastic environmental changes [1]. Until recently, \textit{S. aureus} was considered to be the prototype of a nosocomial pathogen but it has now also clearly shown capacity for outbreaks in the community, affecting young and previously healthy people [2–4], thus suggesting a major epidemiological evolution. \textit{S. aureus} colonizes approximately 30% of the general population without causing any clinical manifestation [3,6]. The bacterial genome and its plasticity facilitate the evolution of virulent and drug-resistant strains [7–13] in response to major and ever-changing clinical challenges. The microbial strategies also include a unique ability to hide in protected “niches” that are shielded from antimicrobials and host defense components. Several studies have indeed shown that \textit{S. aureus} is able to survive for prolonged periods of time in the cytoplasm of non-phagocytic cells [14,15]. Thus, the high diversity of diseases caused by \textit{S. aureus} and the breadth of environments colonized by this bacterium strongly relies on the expression regulation of numerous genes, such as virulence and/or metabolism genes. Confronted with these epidemiological changes, there is a marked need to better understand the pathogenicity and virulence of this organism, and to more precisely define the emergence of epidemic clones, as well as their adaptation during the infection phase and their spreading in the hospital or the community. The pathogenicity, virulence or emergence of epidemic clones within MRSA populations is not clearly defined at a genetic level, despite several attempts to identify common
molecular features between strains sharing similar epidemiological and/or virulence behavior. These studies included pattern profiling analyses such as MLVA [16], AFLP [17], MLST [18], or microarrays [19]. To date, all approaches failed to unravel direct correlations between the presence of any molecular determinants and clinical outcomes. We hypothesize that bacterial behavior potentially resides in other parts of the genome that were overlooked by previous studies, such as the so-called intergenic regions.

The regulation of the genes encoding virulence factors occurs at multiple levels and in a temporal manner, according to the stress [20,21]. This suggests that the production of virulence factors is precisely regulated during infection. In this context, RNAs are recognized as major regulators of gene expression [22–24]. The RNA regulators, named small RNAs (sRNAs) in prokaryotic organisms, constitute a heterogeneous group of molecules that act by various mechanisms to modulate a wide range of physiological responses. RNAIII, for example, is the largest and best studied sRNA in S. aureus that controls the temporal expression of numerous virulence genes encoding exoproteins and cell wall-associated proteins [25]. Overall, sRNAs can modulate transcription, translation, mRNA stability, DNA maintenance or silencing. Thus, these RNA regulators can play a major role in the complex process of epidemicity and virulence as well as for adjusting to diverse environmental stresses. Various regulatory mechanisms involving sRNA have been reported and include changes in RNA conformation using riboswitches, protein binding, base pairing with mRNAs, and interaction with DNA [26,27].

Despite the importance of sRNAs, their study in the Staphylococcus aureus genome is relatively recent, notably due to the lack of appropriate tools to detect them. Most have been detected by in silico predictions, such as sgr and rsa small RNAs [26,28], and their involvement in the regulation of virulence factors or metabolism was then reported [26]. Nowadays, the development of massively parallel analysis methods including high-throughput sequencing (HTS) and tilling arrays allows for the development of massively parallel analysis methods including high-throughput sequencing (HTS) and tilling arrays allows for the detection of sRNAs. In theory, the MiRvana isolation kit provides a procedure allowing an enrichment of RNAs from various types of RNA preparations. We also evaluated the fragmentation methods on ribosomal-free RNA samples by performing a mass of RNA used (ng) for each experiment. The numbers of reads reached 95 and 91%, with or without zinc-fragmentation, respectively. Overall, the numbers of reads successfully mapped on the sequences of N315 were similar and reached 95 and 91 %, with or without zinc-fragmentation, respectively (Table 1).

Table 1. Summary of cDNA samples sequenced by Illumina RNA-Seq and mapped on N315 S. aureus genome.

<table>
<thead>
<tr>
<th>Detection run</th>
<th>Orientation run</th>
<th>Expression run</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zinc-fragmentation</td>
<td>Without Zinc-fragmentation</td>
</tr>
<tr>
<td>Mass of RNA used (ng)</td>
<td>300</td>
<td>276</td>
</tr>
<tr>
<td>Read size (nt)</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>Number of reads</td>
<td>3'528'913</td>
<td>3'936'842</td>
</tr>
<tr>
<td>Mapped reads (%)</td>
<td>90</td>
<td>95</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0010725.t001

Results

Analysis of the Illumina RNA-Seq coverage profiles

In order to analyze the S. aureus N315 transcriptome obtained by Illumina-HTS in depth, and particularly to detect the non-annotated regions, we applied the following three criteria to characterize each transcript: (i) detection, (ii) orientation and (iii) expression during S. aureus N315 growth in rich medium (MHB). For the two first experiments, RNAs were purified after 4 hours of growth using RNasy or MirVana isolation kits for comparison purposes. In theory, the MirVana isolation kit provides a procedure allowing an enrichment of RNAs<200 bases. Mapping of the resulting reads to the annotated S. aureus N315 genome revealed only marginal quantitative differences between the two types of RNA preparations. We also evaluated the fragmentation methods on ribosomal-free RNA samples by performing a chemical treatment (presence of Zn+) or using a protocol without zinc-fragmentation. Overall, the numbers of reads successfully mapped on the sequences of N315 were similar and reached 95 and 91 %, with or without zinc-fragmentation, respectively (Table 1).

Mapping of cDNA sequence reads to the annotated S. aureus N315 genome

We manually inspected the whole RNA-Seq coverage profile along the genome of S. aureus strain N315 and its plasmid to identify all transcription units including non-annotated ones. As expected, the expression profile showed a strong correlation with the positions of previously annotated genes. Among all annotated CDS, 86% displayed a significant coverage (cut-off value: 0.5 average of sequencing coverage), in at least one of the experiments.
In addition, the transcriptional profile revealed numerous signals located in intergenic regions, some of which may correspond to potential open reading frames or to 5' - or 3' - untranslated regions of genes.

All detected transcripts are listed in Table S1 and named “Teg”, for “Transcript from Experimental method from Geneva” according to their discovery order. Total RNA sequencing allowed for identification of 150 intergenic transcripts within the 2.9 Mbp of the chromosome and 9 on the 24 kbp of the plasmid of strain N315. These transcripts are not localized to specific regions but show a rather homogenous distribution along the genome and plasmid of strain N315. A total of 9 transcripts (Teg15, 16, 31, 33, 34, 54, 103, 111, 128), not yet annotated in the genome of strain N315, cover small putative ORFs by mapping large regions without any stop codon and some display a particularly unusual length (Table S1). Overall, we detected 160 intergenic transcripts in the genome of N315, encompassing almost 10% of the intergenic regions with relevant size (i.e. larger than the length of reads (Table 1)). The total size of these newly discovered transcripts represents >30'000 bases, i.e. around 4% of the coding sequences annotated in the genome of S. aureus strain N315.

Illumina-HTS was then performed with the orientation protocol, after ligation of specific linkers on each strand of the purified RNA preparation. This method allowed us to determine the orientation of all previously identified transcripts in a single run. We validated this experiment by controlling the adequate orientation of the previously annotated transcripts. Strand orientation was unambiguous for all detected transcripts and showed the same read orientation as the reference genome (Figure S1). As only minor differences were observed when using the two dir-mRNA-seq protocols (i.e. with or without zinc fragmentation), we decided to pool these two runs in order to increase the sequencing depth without altering the results. Considering all detected transcripts, the repartition of the transcript coding strands along the circular bacterial genome followed what is known for CDS orientation. Indeed, we observed that the intergenic transcripts localized in the first half of the genome are transcribed by the positive coding strand (21 positive strand transcripts and 15 negative strand transcripts in the first genome quarter; 31 positive strand transcripts and 11 negative strand transcripts in the second genome quarter) whereas the transcripts encoded by the second half of the genome are mostly in the minus stand (12 positive strand transcripts and 15 negative strand transcripts in the fourth genome quarter). Some of these transcript profiles are depicted in Figure 1 (A to E) and their corresponding Teg described in Table S1.

sRNA annotation and classification

Bioinformatics predictions have been performed to classify the detected transcripts according to their potential regulation mechanism (Table S1). Small RNAs constitute a heterogeneous group of molecules acting by various mechanisms to regulate gene expression. In this context, we have found 14 sRNAs, which correspond to cis-regulator riboswitches. Riboswitches are characteristic structures located in the 5’-region of mRNAs that can bind intracellular metabolites. In our study, many classes of such regulators are represented, such as the S-box [35–37]. We also identified 17 cis-acting regulatory regions that are specifically recognized by a variety of regulatory proteins (t-proteins, GtL), or by tRNAs (T-box). Otherwise, we have categorized 57 bona fide small RNAs which could act in trans. Among the small transcripts, 21 sRNAs were detected in the opposite direction of their two flanking genes, and many of them contained a typical Rho-independent transcription terminator hairpin. Thus, these transcripts are most likely synthesized independently of their flanking genes. These observations were used to classify them in the bona fide small RNAs group. We have also recognized the stable housekeeping RNAs, such as 4.3S, 6S, RNase P and mRNA [27,29]. Evaluation of promoters and terminators, start and stop codons, as well as Shine-Dalgarno sequences allowed us to identify 9 potential CDSs, yet not annotated in RefSeq NC_002743. The annotation software may miss small CDSs and/or CDSs containing alternative start codons. Indeed, five of these nine transcripts (Teg13, 31, 33, 34, 54) show homology with hypothetical protein in other S. aureus strains and/or others Staphylococcus species, by using Basic Local Alignment Search Tool [38] (BLAST). Finally, 32 signals juxtaposed to one of their flanking genes were classified as 5’- or 3’-UTR of this gene. These regions have been identified and classified in this transcriptomic cartography because of their unusual length for a UTR region. No particular structure or bioinformatics prediction was observed for these 5’/3’ UTR but this observation does not preclude these regions to be involved in regulatory processes. Overall, most of these putative regulatory RNAs present significant secondary structures as exemplified in Figure 2. This analysis unravels conserved regions often showing internal base-pairing that reflects their regulatory function [39]. Note that these sequences are likely of major importance considering their high degree of conservation in the staphylococci (Table S1).

Antisense RNAs

The orientation protocol allowed us to identify 23 antisense RNAs that showed perfect complementarity with target mRNAs encoded from the same DNA locus (Figure 1F and 1G, for example). Four other transcripts were observed as antisense of the UTR region or antisense of previously identified ncRNAs, such as sact. Interestingly, a recent bioinformatics analysis suggested that type I toxin–antitoxin modules are widely distributed among bacteria [40]. This system consists in small hydrophobic peptides whose expression is repressed by a small antisense RNA. Among these predicted modules, Teg152 and Teg154 are predicted to encode for type I antitoxins in S. aureus N315 strain (Table S1). These two ncRNAs are antisense to Teg8 (SprA) and SprG, respectively which are predicted to encode type I toxins [40]. These four RNAs (Teg151, 152, 153, 154) were however classified in the bona fide small RNAs group for their intergenic characteristics. In addition, we detected nine antisense RNAs from repeated regions in the genome and three antisense molecules which correspond to previously annotated genes (bbDQ, SA0752, SA0833) but whose 3’ UTR appears as antisense of the flanking gene (Teg39as, 40as, 41as) (Table S1). Most of the antisense RNAs are found in pathogenicity islands. These antisense molecules are putatively acting in at least six functional categories (Clusters of Orthologous Groups; COG) but they display a particularly high frequency in the categories “cell-wall and cell envelopes” (M) and “replication, recombination and repair” (L), which are essential biological functions. Another group of cis-encoded antisense RNAs may modulate the expression of genes in a potential operon, i.e. between two genes, such as Teg7as (Figure 1F), Teg8as, Teg15as, Teg16as and Teg19as. Based on their location on the mRNA, these transcripts may act as potential inhibitors of translation initiation, as mRNA degradation and/or as mRNA transcription termination.

Antisense transcripts detected in this study correspond typically to cis-encoded antisense RNAs, i.e. encoded in the same DNA locus as their complementary targets. Some are predominant in comparison to their antisense gene (Teg6as, 10as, 14as, 15as, 17as, 21as, 22as, 23as, 24as and its duplicated regions, 39as, 40as, 41as),
some are less expressed (Teg5as, 6as, 9as, 16as, 19as, 20as, 25as, 26as, 27as, 36as, 37as, 10aspl) and five have almost the same expression levels (Teg7as, 18as, 28as, 38as). Mechanisms of these antisense RNAs may be variable and yield to transcription attenuation, translation inhibition, or mRNA degradation. Several antisense RNAs are complementary to the 3' end or 5' end of transposase genes (Figure 3). We were not surprised to find that these transcripts exist in pairs (represented by the same color in Table S1) as they are likely duplicated by the transposase encoded by the associated transposase gene. Potential promoter regions and Shine-Dalgarno sequences have been identified at two locations in the insertion sequence, in the vicinity of antisense transcripts. The genome of N315 contains approximately 25 transposases from 5 major types (for Tn554, IS1181, IS232, or for other poorly characterized IS elements).

Additionally, we have detected 4 antisense RNAs which map to parts of the transcripts identified in intergenic regions. Teg152, Teg153 and Teg154 are complementary to non-coding RNAs previously described by Pichon and Felden [28]. Teg151 has been identified as antisense to the 3' untranslated region of SA1344 mRNA; we also find that Teg51 is highly expressed in our Illumina-HTS data. These four antisense RNAs have been classified as bona fide small RNA, not in the antisense sub-category, due to their intergenic location (see experimental section).

Orientation of transcripts

The oriented RNA-seq protocol produces reads that are oriented according to the RNA molecules they originate from. The mapping orientation is thus informative, as opposed to regular RNA-seq experiments where the strand information is lost during the cDNA synthesis step. The read mapping procedure revealed that 45% and 55% of the reads map onto the direct and reverse strand of the S. aureus N315 genome, respectively (RefSeq NC_2745). This unequal repartition is due to the highly expressed tRNAs and rRNAs that are more abundant on the reverse strand of the genomic sequence. Housekeeping RNAs are not fully captured during the sample purification procedure and they represent 65% of the total sequenced reads. The remaining 35% of the reads show a mapping repartition of 50.6% and 49.4% for the direct and reverse strand respectively. Investigating the
mapping orientation within CDSs revealed that 97% of the reads are oriented according to the coding strands and thus agree with the presence of mRNA transcripts. The remaining 3% correspond either to non-coding RNAs, false-positive read matches or possible gDNA contamination. We observed that reads mapping the non-coding strands are not uniformly spread among the CDSs. To quantify this observation, we computed the non-coding mapping ratio for each CDS displaying a sequencing coverage larger than 10x. This ratio is calculated as nR/(nR+nD), where nR and nD are the coverage values on the non-coding and coding strand, respectively. Thus, CDSs only covered on the coding strand have a non-coding coverage ratio equal to 0, while higher values reveal the presence of reads that also map the non-coding strand. The CDSs displaying larger non-coding mapping ratios correspond to the antisense sRNAs we manually identified from the coverage signals (details are provided in Table S2).

Transcript sequence characteristics and conservation

Transcripts detected in our study were compared across phylogeny using the Basic Local Alignment Search Tool [38] (BLAST). All detected transcripts are conserved across all sequenced *S. aureus* strains with a requirement for 80% similarity (Table S1). None of the identified transcripts was unique to strain N315. Some of them are conserved at the bacterial genus, family, or even at the organism class level. Moreover, a few of the RNAs were found to be partially or fully repeated in the genome of *S. aureus* strain N315. These RNAs potentially control important regulatory functions, and this may explain their duplication. These transcripts are listed in Table S1 under “repeated sequences”. Indeed, these duplications in the genome split the assignment of the sequencing coverage for each copy and make it impossible to map a precise expression signal at a defined position.

RFAM allowed us to predict a large majority of the 14 riboswitches and the 17 cis-acting regulatory elements found within the bacterial chromosome of strain N315, primarily located in the 5’-untranslated regions (UTR) of mRNAs. Several of these regulatory regions are known to respond to the intracellular concentration of various metabolites (purine, pre-Q, glycine, lysine, Glucosamine-6P, thiamine pyrophosphate, S-adenosyl-methionine, flavin mononucleotide). Furthermore, twelve of the sRNAs contain a putative T-box, and a tandem T-box was found upstream SA1199. Three of them (Teg67, Teg69, Teg73) were not predicted in RFAM. We manually folded the putative T-boxes in order to find their target tRNA (Figure 2A, B, C). It is remarkable that three of these new T-boxes do not contain the consensus signature [41]. The NMR structure of the specifier loop was recently shown to adopt a loop E motif with stacked non canonical basepairs allowing an optimal presentation of the specifier triplet nucleotides for pairing with the anticodon of the sensing tRNA [42]. This specific signature was neither found in Teg67, Teg69 and Teg73 although alternative base pairing internal loop was predicted for Teg69 with lower reliability. Teg67 harbors a particularly short specifier hairpin containing a Gly codon and a very long antiterminator moiety, but its structure appears quite different from the consensus. Another example is constituted by Teg73 that regulates the expression of leucyl-tRNA synthetase. In that example, the folding of the specifier hairpin that should position a leucine “codon” in an appropriate configuration appears not optimal, raising the question of its functionality (Figure 2A, B, C).

Approximately 10 of these sequences harbor GC content significantly higher than the average GC content of the genome. Among all of these cis-acting putative regulatory regions, 14 are transcribed on the positive strand according to RefSeq NC_002745, whereas 17 are transcribed on the minus strand. The genes located downstream of these putative regulatory regions belong to two functional COG categories: “translation, ribosomal structure and biogenesis” and “amino acid transport and metabolism”. Of interest, these regions were found to be highly conserved in sequenced *S. aureus* strains with ≥96% homology. In addition, significant homology levels, generally ≥80%, were observed in several coagulase-negative staphylococcal species such as *S. homolyticus*, *S. epidermidis* or *S. saprophyticus* (Table S1). Decreasing the stringency of the homology cut-off allows us to

Figure 3. Schematic representation of antisense transcripts in the *tnp* region. Among the transposases annotated in the genome of *S. aureus* N315, antisense RNAs were identified 8 times at the extremities of the *tnp* transcripts, suggesting an important functional role. This example shows Teg24as and Teg17 (pink italic font) and their complementary *tnp* gene (green). In the sequences detailed below, the potential Shine-Dalgarno sequences appear in bold fonts.

doi:10.1371/journal.pone.0010725.g003
obtain partial homology. BLAST searches using these criteria uncovered the presence of numerous transcripts conserved in Bacilli or more frequently in Firmicutes (Table S1).

A total of 57 bona fide sRNAs were identified in the bacterial genome on the basis of bioinformatics prediction or based on the literature. Only eight of these sRNAs harbor a GC content that is significantly higher than that of the bacterial chromosome. Similarly, nine sRNAs harbor a GC content significantly lower than that of the genome. In this category, four transcripts show a GC content lower than 25% and reveal partial homology to sequences found in low-GC Gram-negative organisms (e.g., Campylobacter, Wolbachia). Interestingly, most of the previously described transcripts uncovered by bioinformatics strategies were found in this category of sRNAs, such as rsa and spr [26,28]. Note also that 26% (n = 14) were observed on documented pathogenicity islands.

Discovered sRNAs are generally in single copy in the genome of strain N315 with the exception of five transcripts that were found in multiple copies (listed as repeated sequences, Table S1). Most of these duplicated transcripts are located in the vicinity of transposases that exist in multiple copies in the genome of strain N315. These transcripts appear to be either species-specific or markers of mobile genetic elements found in other staphylococcal species and are highly conserved (Table S1). Interestingly, one transcript (Teg12) classified in this sub-category shows repeated sequences in low-GC Gram-negative organisms considering their respective category, with a median length varying from 285 nt for the antisense RNAs, 108 nt for the 3' or 5' UTR, 148 nt for the sRNAs, 141 nt for the riboswitches or 203 nt for the cis-acting regulators. Finally, the multicopy transcripts were the smallest molecules with a median size of only 95 nt. The size average of the potential CDS regions is 136 nt (excluding the unclassifiable Teg34).

mRNA transcript expression during growth-phase by Illumina-HTS

Temporal expression of transcripts at the genome level during the growth-phase was obtained by HTS (Figure 4) whereas precise expression of a subset of 26 sRNAs identified in this study was assessed by RT-qPCR. Expression levels of annotated genes known to contribute to the S. aureus pathogenicity were evaluated in the specific context of strain N315 which is agr-deficient. Obviously, hld was not expressed during the growth kinetics and consequently, RNAIII is missing. The agr locus is an important global regulator controlling the temporal expression of several S. aureus virulence factors [48]. Nevertheless, strain N315 has an inactive agr-system which is not rare in clinical isolates. In fact, N315 is one of the most widespread S. aureus lineages responsible for human infections [49,50]. Expression of virulence factors is also regulated by other partners, acting independently from the agr-system. Previous observations regarding sRNA expression suggest that other factors, such as small RNA molecules could also contribute to the regulation mechanisms of virulence factors [28]. A similar absence of expression has been observed in N315 for the whole ica operon that contributes to biofilm formation [51] and the cap genes encoding for the production of the bacterial capsule [52]. On the other hand, vpl or sps genes encoding for ribosomal proteins followed previously described kinetics [53]. Bacterial adhesins are important proteins contributing to the attachment to host cells or tissues [54] and the regulation of some of these factors has been previously studied in detail [55–58]. In our study, fibronectin-binding proteins A and B were detectable mainly during the early phase of growth, but remained only poorly expressed during the entire growth kinetics in this agr-deficient background [57]. Clumping factors A and B which provide the bacterium with the capability to interact with human fibrinogen [59,60] showed similar expression profiles during the growth-phase. In accordance with previous reports, we noticed high levels of expression of clfB at a very early phase of growth [61] which is then followed by a significant decrease. However, in this strain background, clfB expression appeared maximal during the later growth times. Finally, the expression of spa, the gene encoding the protein which binds the major Fc fragment of immunoglobulins, appeared to be similar to previously published studies [62]. This observation suggests that agr acts as an amplifier of spa transcript level during the complete growth kinetics. Agr is not the driver of spa expression, as the synthesis kinetics appear similar irrespective of the presence of an active agr. Similar observations are available for other important virulence factors such as hla and hlb genes encoding for haemolysins, in the absence of a functional agr. Based on the mapping obtained from multiple timepoints in our kinetic
Table 2. Comparison of transcripts detected in intergenic regions in regards to previous analysis.

<table>
<thead>
<tr>
<th>Study</th>
<th>Discovered transcripts</th>
<th>In common with our study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In silico prediction tested</td>
<td>Experimental validation</td>
</tr>
<tr>
<td>This paper</td>
<td>139</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roberts 2006 [45]</td>
<td>42</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marchais 2009 [46]</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abu-Qatouseh 2010 [47]</td>
<td>65</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers and names of non overlapping transcripts found in intergenic regions are given for each study and compared to the present study. The locus of each transcript has been considered. Three studies [26,28,46] used in silico prediction in first intention and confirmed the existence of transcripts by experimental validation such as Northern blot or Race-PCR. Only the sRNAs which have been experimentally tested have been taken into account. Otherwise, our study as those of Roberts [45], Anderson [44] and Abu-Qatouseh [47] are based on direct experimental approaches as Affymetrix microarrays, or cloning/sequencing strategy. The number of detected transcripts varies according to the methodologies and the strain used by the author. The study of Geissmann [26] also predicted the known cis-acting elements such as the metabolite-sensing riboswitches, the T-box motifs and several protein-mediated antitermination events.

doi:10.1371/journal.pone.0010725.t002
run (see experimental section), the kinetics of the signals we obtained showed impressive reproducibility, both for the annotated genes and the newly discovered transcripts (Figure 4).

RT-qPCR validation of HTS data for a subset of sRNA genes expressed during stress

We used RT-qPCR to complete the RNA-Seq expression run with an assessment of the expression of 26 transcripts that are representative of most sRNA sub-categories (Figure 5). The experiments were first performed in rich MHB medium in order to obtain baseline values. Most RNAs are highly expressed in the stationary phase and typically show a significant increase during the transition between the mid-exponential phase and the stationary phase (Figure 5).

Various stress conditions were then tested to mimic environmental changes encountered by the bacteria during the infection, such as pH or temperature variations as well as oxidative or nutrient stresses. Gene expression was evaluated during the early or late exponential phase of growth (2h and 6h). The main results are depicted in Figure 5. In summary, we observe that at 2 hours, thermal, oxidative and pH stresses induce the expression of most of the evaluated transcripts. This induction seems to be particularly important at the two pH values tested during the early (2h) timepoint we observed for several transcripts (Figure 5).

On the other hand, during the late-exponential phase of growth, expression variation was systematically lower except for Teg17, Teg47, Teg55, Teg72 and Teg76 (Figure 5). We also found spectacular inductions of expression during the oxidative stress response of 11 small transcripts at 2h (Teg1, 4, 21, 24, 26, 28, 42, 61, 69, 73 and 91). Early responses (2h) to temperature increase or decrease were observed for 8 different transcripts. In contrast, many transcripts were repressed at 6 hours, at both of the tested temperatures (Figure 5). Expression during growth in stringent medium was also remarkable for numerous tested transcripts. At 6h, the expression of the vast majority of our target genes was either not influenced or significantly decreased during growth in minimal medium. Yet some of the tested genes showed increased expression during the early phase of growth (Teg1, 18, 21, 38, 42, 45). Among them, the kinetics of Teg42 (4.5S) illustrates the ubiquitous role of signal recognition particles (SRP), which appear tightly regulated during the growth phase but also during various stress situations such as oxidative or pH shocks as well as nutrient stress.

Figure 4. Temporal expression of selected RNAs obtained by RNA-Seq Illumina-HTS. Transcript expression levels are shown at 2 hours (red), 4h (green), 6h (blue), and 8h (black). The X-axis corresponds to N315 genomic coordinate and Y-axis symbolizes the normalized coverage. We smoothed the coverage profiles by using a sliding average window of size 41. The expression profile of control genes has been used for experimental validation. General profiles are consistent with published studies with respect to the temporal expression of these transcripts. The experiment shows that the expression of most small transcripts identified in this study are regulated during the growth phase.

doi:10.1371/journal.pone.0010725.g004
starvation. Note that Teg42 (4.5S) as well as other detected sRNAs appeared particularly abundant (Figure 4) and exposure to stress conditions yielded to impressive fold change values or copy number (Figure 5).

Discussion

*S. aureus* is an important human pathogen with a spectacular capability to regulate its metabolism in order to survive a variety of changing environmental conditions. Considering the human body as the main replication site for the bacterium, several lines of evidence demonstrate that temperature or pH changes, the presence of reactive oxygen species, nutrient starvation as well as the exposure to various families of antibiotics have an impact on its growth phenotype, but they also unravel a surprising capacity to rapidly adopt behavior allowing survival and persistence. Mechanisms underlying pathogenicity, virulence and the emergence of epidemic clones of *S. aureus*, including MRSA, are not clearly understood yet. Benefiting from HTS capacity, we describe here the largest dataset representing the *S. aureus* transcriptome obtained to date. RNA-Seq allowed us to provide a complete pattern of small RNAs expressed under specific conditions in *S. aureus* strain N315, and to enrich previous bioinformatics approaches, that are currently limited in ability to detect low structured sRNAs.

We identified all transcripts expressed in strain N315 at different time points. This includes the measure of predicted CDS but also the discovery of 35 antisense RNAs and 160 small RNAs expressed from 'non-coding' or not (yet) annotated regions. Some of these molecules show unusual consensus sequences that preclude detection by classical algorithms. In the genome of strain N315, we discovered that almost 10% of the expressed transcripts were in fact ignored as of today, permitting us to now properly assess their role in virulence or metabolism. The method used in RNA-Seq allowed us to provide a complete pattern of small RNAs expressed under specific conditions in *S. aureus* strain N315, and to enrich previous bioinformatics approaches, that are currently limited in ability to detect low structured sRNAs.
our study is “observational” and because we know that all detected transcripts were indeed expressed, we know they constitute functional transcription units of the considered organism. Despite their considerable merit, predictive bioinformatics tools or massively parallel methods such as microarrays remain limited by the need for knowledge of consensus or sequence which may not yet exist, and are sometime affected by false-positive [63-65].

HTS also allows access to operon visualization, but in this study we focused specifically on the cartography of RNA regulators. A recent study of Abu-Qatouseh [47] in S. aureus small colony variants identified 78 RNA candidates transcribed in antisense orientation, representing 55% of detected non-protein coding RNAs. Compared to our findings, this result demonstrated that expression of these antisense RNA molecules is strongly regulated according to the different conditions. A similar predominance for antisense sRNAs was described in cyanobacteria [71]. In humans, antisense RNAs appear to be widespread with estimates that they comprise up to 25% of the transcripts [72,73]. This number is generally lower in bacteria and particularly in this S. aureus N315 strain where we identified this type of organization for only 1.3% of mRNAs, considering our cut-off. Note however that the presence of an antisense RNA in the capsule operon as well as in the staphylococcal secretary antigen (soa) may have important functions since this envelope structure is critical for the survival of the bacteria in human phagocytic cells [74]. Indeed, SsaA appears highly expressed during the course of an infection and is apparently involved in bacterial susceptibility against specific antibiotic families.

Small RNAs are famous for their capacity to interact with the ribosome binding site of mRNAs encoding transposases and consequently inhibiting their mobility [22,70,75]. Interesting structures including antisense RNA molecules were frequently found in the vicinity of transposases in the genome of strain N315. The location of these antisense RNAs (Teg17, 30, 59, 98, 119, 142, 145) at the 5'-UTR of tnp suggests their capacity to sequester the transcription and translation initiation domains of tnp gene. We hypothesize that small RNAs may be implicated in regulation cascades by their ability to act on transposase gene expression. The interactions are sometimes complex as sRNAs act through transposition action, gene activation, gene inactivation and/or genetic rearrangement [76]. Interestingly, we have identified another group of antisense transcripts of tnp genes, which are complementary to the 3'-untranslated region of tnp (Teg24as, 29as, 31as, 32as, 33as, 34as, 35as, 30as). Surprisingly, these 3'-antisense RNAs have been localized near the transcription and translation initiation regions of the potential downstream gene. This structural organization suggests that this insertion sequence may activate downstream genes, by the incorporation of initiation sequences upstream of an ORF, and that antisense transcripts could regulate this mechanism (Figure 3). Finally, these processes may contribute to the evolution of the bacterium by affecting the capacity of S. aureus to acquire genetic elements implicated in virulence, by facilitating its adaptation to various environments [77] or by acting in trans to produce multicyclic inhibition [78]. The latter mechanism may be involved in either strict gene regulation or fast evolution of the bacterial genome during exposure to selective pressure such as that encountered during an infection [76].

Despite the fact that many fully complementary antisense sRNAs are currently known, the vast majority are found encoded on bacterial plasmids [70,79]. Many detected antisense RNAs have also been found in IS1181, which was initially identified in a methicillin-resistant S. aureus strain [80]. In this study, 10 antisense transcripts were localized in pathogenicity islands and could thus play a key role during S. aureus infections. These molecules are present in at least 6 COG categories and are particularly abundant...
in genes involved in cell wall and cell envelope biogenesis as well as in replication, recombination and repair. We speculate that these transcripts act though base-pairings with their mRNA targets leading to the inhibition of translation initiation, mRNA degradation and/or mRNA transcription termination.

Small RNAs category

Antisense RNAs recognizing non coding RNA (ncRNAs) were also found in our study. This type of interaction between homologous ncRNAs has been previously described in *Listeria monocytogenes*, another Gram-positive organism, suggesting regulatory networks that involve several sRNAs. The implication of such networks in the colonization and infection processes has been recently documented in that organism [22,61]. Two of these antisense-sense ncRNAs belong to the type I toxin-antitoxin systems [40] (Table S1).

Our study also reveals that the genome of *S. aureus* strain N315 contains >50 sRNAs located in pathogenicity islands or closely related to key metabolic genes previously identified as virulence genes. Whether these molecules trigger regulatory functions and their respective impact on the virulence of the bacterium remains to be precisely evaluated. Regarding *S. aureus* pathogenesis, the abundance of sRNA in pathogenicity islands (around 30% of the identified sRNA in our study), appears to be of paramount importance. Similar to rRNAs and tmRNA, downstream regions of small RNAs are identified as hotspots for foreign DNA integration [82]. For example, Ballottin has recently discovered that small RNAs could mediate prophage insertion and consequently integration of carried exogenous DNA [83]. This observation could explain the large amount of sRNAs frequently found near virulence genes, which are acquired by horizontal transfer during exposure to high selection pressure occurring, for example, in host infection.

We identified here 57 *bona fide* small RNAs here in specific regions of the bacterial genome such as pathogenicity islands, metabolic genes and flanking transposase genes. For these elements, the GC content was significantly higher than that of the average bacterial chromosome, suggesting their potential acquisition during horizontal gene transfer. Of interest, prediction of the secondary structure reveals that many of these sRNAs are characterized by several stable hairpin motifs, and 30% of them carry an unpaired UCCG motif. This motif has been recently shown to be a functional sequence that is used to target the ribosome binding sites of target mRNAs such as RNAIII [84] and RsaE [26]. One of these typical sRNAs is shown in Figure 2F.

The housekeeping sRNA molecule, the major constituent of the signal recognition particle (SRP), was identified in *S. aureus* [28]. In bacteria, SRP delivers proteins to the plasma membrane or participates in post-translational protein sorting [85]. SRP is ubiquitous and found in all living organisms, but its size and the characteristics that lead to secondary structures vary between the different phylogenetic groups. In *S. aureus*, the main loop of SRP sRNA involves 35 perfectly paired nucleotides. The sequence of this loop appears extremely conserved within *Staphylococcus* SRP sRNAs, and also well conserved in the *Firmicutes* phylum, indicating a common evolution which explains the conservation of the loop structure. This suggests a common ancestor preceding the divergence to lower levels of the phylogenetic tree, mainly accumulating punctual mutations in the non-paired regions. Expression of sRNAs involved in SRP formation is highly influenced by environmental conditions, including medium and physico-chemical composition [26] (See *TegF*2 in Figure 4 and Figure 5). In that sense, stress conditions have major impacts on the expression of the sRNA discovered in our study (Figure 5).

Note also that the expression study performed by RT-qPCR is representative of a large panel of RNA categories such as riboswitches (2/26 tested transcripts), cis-acting regulators (7/26), *bona fide* small RNAs (7/26), UTR regions (7/26), repeated sequences (2/26) and plasmid RNAs (1/26). Temperature as well as pH values, presence of oxidative stress and growth in minimal media allowed the identification of sRNAs specifically responsive to particular stress conditions. We conclude that stress conditions have a major impact on the transcript expression and therefore confirm that these RNAs could be implicated in gene regulation and bacterial adaptation to drastic environmental changes.

Our study is affected by potential technical limitations related to the utilization of i) enzymatic steps that could bias the sequence representation, ii) arbitrarily–defined cut-offs for the detection of transcripts. This large inventory, the largest to date, will likely increase when further experiments are performed in other growth conditions or using different *S. aureus* strains. This study already represents an important milestone in the characterization of the *S. aureus* transcriptome and genome, and provides a foundation for studies aimed at elucidating the biology and virulence of this ubiquitous human pathogen.

Conclusions

In the genome of *S. aureus* strain N315, we identified 160 small RNA molecules in regions formerly considered to be intergenic, and an additional 35 that were detected for their antisense properties against cognate CDS. Some of the small RNAs are localized in biologically or clinically relevant regions, between key metabolic or virulence genes and especially within pathogenicity islands as well as within the methicillin-resistance element. Most of these are temporally regulated during growth phase and affected by various stress conditions, suggesting their implication in the adaptability of this ubiquitous human pathogen. The sizeable amount of data generated by Illumina-HTS and their analysis contribute a broad repository for future investigations of sRNAs. We expect that a systematic analysis of such sRNA networks will shed new light on the virulence and epidemiology of *S. aureus*.

Materials and Methods

Bacterial strain, culture conditions and RNA preparation

*S. aureus* strain N315 was grown in fresh Mueller-Hinton Broth (MHB) at 37°C, after a 1:100 dilution of overnight cultures. For kinetics analysis, total RNA was purified at 2h, 4h, 6h and 8h of growth as previously described [15]. Total RNA of N315 was extracted with the RNeasy (Qiagen) or the MirVana isolation kit (Ambion). For HTS experiments, total RNA was treated with the MICROBExpress kit (Ambion) to limit contamination of the RNA molecules. Illumina-HTS and their analysis were performed by various stress conditions, suggesting their implication in the adaptability of this ubiquitous human pathogen. The sizeable amount of data generated by Illumina-HTS and their analysis contribute a broad repository for future investigations of sRNAs. We expect that a systematic analysis of such sRNA networks will shed new light on the virulence and epidemiology of *S. aureus*.

Bacterial strain, culture conditions and RNA preparation

*S. aureus* strain N315 was grown in fresh Mueller-Hinton Broth (MHB) at 37°C, after a 1:100 dilution of overnight cultures. For kinetics analysis, total RNA was purified at 2h, 4h, 6h and 8h of growth as previously described [15]. Total RNA of N315 was extracted with the RNeasy (Qiagen) or the MirVana isolation kit (Ambion). For HTS experiments, total RNA was treated with the MICROBExpress kit (Ambion) to limit contamination by multicopy structural rRNAs, following manufacturer’s instructions with slight modifications. The maximal amount of total RNA per tube was limited to 4–5 μg because previous experiments showed important contamination levels when using 10 μg. Between each purification step, RNA quality and quantity were determined by Bioanalyzer (Agilent) using the RNA Nano chips, and quantified using the ND-8000 (NanoDrop Technologies).

For the evaluation of transcript expression by quantitative RT-qPCR, *S. aureus* strain N315 was exposed to various stress conditions. Cultures were performed in the presence of 1 mM of paraquat (Sigma) for 30 minutes, or grown in the stringent medium NZM (10 g/l casein tryptone digested, Fluka 95039; 5 g/l NaCl; 2 g/l MgSO₄·7H₂O; pH 7.0). For thermal shocks, bacterial cells were subjected to either a cold (25°C) or a heat shock (42°C) during 1 hour after exponential growth. For the pH stress, cells were centrifuged 5 min at 3,000 g and resuspended in fresh MHB media.
buffered at pH 5.5 or pH 8.0 during 30 min. The cells were then rapidly centrifuged and stabilized in acetone/ethanol (1/1) [15]. Total RNA was extracted with CellsDirect One-Step RT-qPCR kit (Invitrogen) including an extensive DNase treatment (10× final concentration, as suggested by the manufacturer) for the analysis by quantitative RT-PCR.

Library construction and Illumina RNA-seq runs

We performed three different Illumina RNA-seq experiments; i) a detection, ii) an orientation and iii) a kinetic run. The detection experiment was aimed at revealing intergenic sRNAs. The orientation experiment aimed at showing the strand of transcription for all detected sRNAs and at identifying antisense sRNAs. The kinetic experiment consisted in a set of four RNA-seq runs to map the transcriptome at 2h, 4h, 6h, and 8h of growth in MHB. Briefly, the detection run was performed on double-stranded cDNAs prepared from purified DNAse-treated transcripts (MicrobExpress, Ambion) using random primers and RNaseH. The double stranded cDNAs were fragmented by nebulization. After ends-repair and purification, the fragments were ligated with Illumina genomic double-stranded adapters. After size-selection on agarose gel to recover fragments with inserts of approx. 30–150 bp, the library was amplified by PCR with Phusion polymerase. PCR cycling conditions were: i) enzyme activation: 98°C for 30 seconds, ii) denaturation: 98°C for 10 seconds, iii) annealing: 60°C for 30 seconds, iv) elongation: 72°C for 30 seconds, and v) final extension: 72°C for 5 seconds. Steps 2, 3, and 4 were repeated 17 times. The library was then sequenced on the Illumina Genome Analyzer GA for 36 cycles.

The orientation run was performed on transcripts previously purified by the MirVana isolation kit (Ambion) and prepared using a dir-mRNA-SEQ protocol starting either from fragmentation with zinc or from non-fragmented transcripts. After ends-reparation, RNAs were ligated with single-stranded small RNA adapters to preserve the strand orientation. The 3'-adapter was ligated first, followed by the bar-coded 5'-adapter. After cDNA synthesis and PCR amplification for 15 cycles, the library was size-selected on agarose gel to recover fragments with inserts of 15–100 bp.

The kinetic run was performed on purified transcripts (MirVana isolation kit) following fragmentation with zinc using illumina’s mRNA-SEQ kit. Double-stranded cDNA fragments were prepared with random primers and RNaseH. After ends-reparation and purification, fragments were ligated with Fasteris-designed bar-coded genomic adapters. After size-selection on agarose gel to recover fragments with inserts in the range of 100–200 bp, the library was amplified by PCR for 18 cycles. The libraries were then sequenced in one channel in the Illumina genome Analyzer GAII for 36 cycles.

Read mapping and visualization

The standard mapping for transcriptome analysis consists in mapping the reads onto the genomic sequence. For this purpose, we used the MAQ software [86] with default parameters for mapping the sequenced reads onto the Staphylococcus aureus strain N315 [34], considering both, the genomic and plasmid sequences (RSeq NC_2745 and NC_3140). This mapping allowed computing a sequencing coverage by counting the number of time each base of the genome was covered by a sequence read. This coverage profile reflects the transcriptional activity across the genome. For visualization and analysis of this coverage, we used the Artemis genome viewer [87] that allows studying the RNAs expression in the context of the annotated genes. For the orientation experiment, we computed separately the coverage profiles corresponding to the reads mapped onto the direct strand from the ones mapped onto the reverse complement strand of the reference sequence.

Quantification by high-throughput sequencing

Total RNA of N315 was extracted and purified as described above at 2, 4, 6 and 8h. Individual RNA preparations were fragmented and ligated with specific tags before reverse-transcription and ends-repair steps. Similar amounts of RNA from these 4 conditions were mixed and sequenced using Illumina. Simple normalization was applied following coverage obtained for the individual time points.

Quantitative real-time PCR (RT-qPCR) from RNA

Transcript-specific primers were developed for 26 intergenic transcripts and for the hU internal control gene [88] using the Vector NTI Advance™ 10 software (Invitrogen). Sequences and optimized concentrations of all primers used for these quantifications are shown in Table S3. S. aureus N315 total RNAs were amplified and quantified by using a one-step SYBR® Green Quantitative RT-PCR Kit (Sigma) in a final volume of 15 μl. The reactions were performed in a StepOne Real-Time PCR device (Applied Biosystems). For these experiments, we used a MMLV-RT (1/μl), SYBR Green Taq Ready Mix for RT-qPCR, and primers according to the concentrations mentioned in the Table S3. ROX was used for fluorescence normalization. Two μl of RNA CellsDirect extracts for the 26 transcripts of interest were added whereas 2 μl of a 1:1,000 dilution of the same sample was used for hU quantification. Results are average of two independent determinations using the following RT-PCR cycling conditions: i) reverse transcription: 42°C for 30 minutes, ii) enzyme activation: 94°C for 2 minutes, iii) cDNA denaturation: 94°C for 15 seconds, and iv) annealing and elongation: 60°C for 60 seconds. Steps 3–4 were repeated 40 times. Differences in Ct values between tested transcripts and hU signals were used for normalization purpose and based on the MHB medium condition at 2h as a reference. The fold change was expressed as the inverse exponential of the difference between MHB Ct (reference) and the stress condition Ct.

Bioinformatics analysis of intergenic and antisense small RNA transcripts

All the small RNA transcripts detected in intergenic regions by the mapping of all transcripts in the Artemis annotation environment were analyzed with regard to their location, their characteristics, their potential function and their previous identification in the literature. For each intergenic small RNA transcript, several analyses were performed. We identified the presence of a putative related ribosome binding site and/or a rho-independent terminator of transcription. The orientation of the neighbor genes was also considered to classify small RNA transcripts. We computed the GC content of each small RNA transcript, considering that the presence of a bias in the GC content would be an argument in favor of bona fide small RNA transcripts. Five classes of small RNA transcripts were therefore defined that included: putative small CDSs, cis-acting regulatory elements, riboswitches, bona fide small RNAs and antisense RNAs. We computed the GC content of each small RNA transcript, considering that the presence of a bias in the GC content would be an argument in favor of bona fide small RNA as it is the case for tRNA and rRNA in that species. The presence of such a bias has been evaluated by considering two standard deviations of the lower tRNA GC percentage in comparison with the average GC percentage of the S. aureus N315 genome (32.5%). The cut-off for such a higher GC percentage was set to 42%. For each category,
we scanned the transcript for the presence of a UCCC motif as described previously [26]. All small RNA transcripts were also compared to those identified by the five previously published studies [26,28,44–46]. In that case, only the locus was considered for all overlapping identified regions (not the strand orientation).

Data available in the study of Anderson and colleagues were mapped to the N315 genomic sequence by using the Basic Local Alignment Tool [38] with default parameters. The sequence of each small RNA transcript was enriched by annotation data whenever possible by considering both RFAM and Genbank databanks and the literature for spec-G sRNAs [28] and other stable rRNAs, [26]. NCBI and RFAM databases provided more than 110 annotated RNAs for the S. aureus N315 genome including essentially tRNAs and rRNAs at NCBI while other putative sRNAs and riboswitches were found annotated in the RFAM database. For all the transcripts that lacked an annotation, a comparative analysis was performed by using BLAST with default parameters. This analysis was completed by both the HMM approach previously described by Geissmann [26] to identify GC-rich regions and RNAsim comparative analysis [26] that included Wu-blast 2.0 pairwise comparisons of sequences for searching similarities, and QRNA [89] to identify base substitution patterns in pairwise alignments that could correspond to a conserved RNA secondary structure. The interpretation of our data was facilitated by the ApolloRNA program, which was used to assign and visualize all the available information on the annotated S. aureus N315 genome [90]. Structure prediction was performed as previously described [26].

Supporting Information

**Figure S1** Reads mapping onto the S. aureus genomic sequence from dir-mRNA-SEQ protocol.

Found at: doi:10.1371/journal.pone.0010725.s001 (1.51 MB TIF)

**Figure S2** Secondary structures of some sRNAs bona fide based on in silico analysis.

Found at: doi:10.1371/journal.pone.0010725.s002 (0.14 MB DOC)

**Table S1** List of transcripts discovered by Illumina-HTS. Categories and identification of all transcripts discovered by RNA-Seq Illumina-HTS in intergenic regions or in antisense (annotated as “as”) of previously annotated genes in S. aureus N315. Some of these transcripts are found in multiple copies, preventing precise quantification of the signal in each of the multiple locations. The bioinformatics predictions allowed us to classify the intergenic signals in various categories such as riboswitches, bona fide small RNAs, 5’ or 3’ units of transcription and small CDSs. Transcripts of plasmid origin are also reported and annotated as “pl”. a This method allows to map the approximate transcriptional start and end points of the small RNAs. Based on characterized CDS with carefully mapped origins, we evaluated the precision of this method at ±20 nt. b Coding strand symbol around represent flanking CDSs. The middle one indicates the orientation of the small transcripts. c Sequence conservation among publicly available bacterial sequences realized by BLAST. The strict sequence conservation was evaluated with a cut-off value of 80% on the sequence totality. The nomenclature indicates the conservation level such as S. aureus specie (sp), Staphylococcus genus (g), Bacillales order (o), Bacilli class (c), Firmicutes phylum (p) or Bacteria kingdom (k). Partial homology was also considered with a cut-off value of 90% of homology on >1/3 of the entire sequence. We have considered that S. cassolyticus, otherwise known as Macroccocus caseolyticus, as belonging to the Staphylococcus genus [91]. d HMM analysis and the sequence conservation among S. aureus, staphylococcal species, and firmicutes (CA) were previously performed [26]. e Transcripts in the same box have the same sequence found at multiple locations within the bacterial genome.

Found at: doi:10.1371/journal.pone.0010725.s003 (0.10 MB XLS)

**Table S2** Sequencing coverage of the coding and non-coding strands in the individual CDSs. The oriented RNA-seq protocol allows differentiating direct and reverse sequencing coverage. Thus, the reads coverage in the individual CDSs can be separated into coding and non-coding coverage. This table enumerates all CDSs of Staphylococcus aureus strain N315 having a coverage value of at least 10x. Columns 2 and 3 show, for each CDS, the coverage values for the coding and non-coding strand, respectively. The table is sorted according to the percentage of non-coding coverage (column 4). The CDSs displaying a high percentage of non-coding coverage are good candidates for being regulated by an antisense sRNA. Column 11 shows the corresponding Teg, whenever it has been manually identified from the transcriptomic profile (see Table S1). Some CDSs were not covered enough to allow a clear and reliable identification of a RNA signal from the transcriptional profile. However, the high number of CDSs displaying a significant level of non-coding coverage suggests the antisense RNAs to be an important regulation mechanism in the bacteria.

Found at: doi:10.1371/journal.pone.0010725.s004 (0.27 MB XLS)

**Table S3** Characteristics of oligonucleotides used in the RT-qPCR assay.

Found at: doi:10.1371/journal.pone.0010725.s005 (0.06 MB DOC)

Acknowledgments

The authors sincerely thank M. Girard for support in RT-qPCR experiments and K. Whiteson for help with editing the manuscript.

Author Contributions

Conceived and designed the experiments: PL JS PF. Performed the experiments: MB LF CD. Analyzed the data: MB DH CG PR PF. Contributed reagents/materials/analysis tools: LF CD. Wrote the paper: MB PF.

References


