Complete Genome Sequence of the MRSA Isolate HC1335 from ST239 Lineage Displaying a Truncated AgrC Histidine Kinase Receptor


1Laboratório de Biologia Molecular de Bactérias, Instituto de Microbiologia Paulo de Góes, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

2Laboratório Nacional de Computação Científica, Petrópolis, Brazil

*Corresponding author: E-mail: agnes@micro.ufrj.br.

Accepted: September 9, 2016

Data deposition: CP012012.

Abstract

Methicillin-resistant Staphylococcus aureus (MRSA) is still one of the most important hospital pathogen globally. The multiresistant isolates of the ST239-SCCmecIII lineage are spread over large geographic regions, colonizing and infecting hospital patients in virtually all continents. The balance between fitness (adaptability) and virulence potential is likely to represent an important issue in the clonal shift dynamics leading the success of some specific MRSA clones over another. The accessory gene regulator (agr) is the master quorum sensing system of staphylococci playing a role in the global regulation of key virulence factors. Consequently, agr inactivation in S. aureus may represent a significant mechanism of genetic variability in the adaptation of this healthcare-associated pathogen. We report here the complete genome sequence of the methicillin-resistant S. aureus, isolate HC1335, a variant of the ST239 lineage, which presents a natural insertion of an IS256 transposase element in the agrC gene encoding AgrC histidine kinase receptor.

Key words: methicillin-resistant Staphylococcus aureus, ST239 lineage, agrC dysfunction, IS256 insertion, complete genome sequence.

Introduction

Despite the recent improvements on control infection measures, methicillin-resistant Staphylococcus aureus (MRSA) is still a major health problem concern worldwide (Figueiredo and Ferreira 2014). The isolates of the ST239(CC8)-SCCmecIII lineage—one of the most largely disseminated healthcare-associated MRSA (HA-MRSA)—are endemic in several hospitals and have been described in outbreaks around the world (Li et al. 2012; Figueiredo and Ferreira 2014; Baines et al. 2015). In Brazil, MRSA isolates of a multiresistant S. aureus clone—the Brazilian epidemic clone (BEC) of the lineage ST239-SCCmecIII—are disseminated in hospitals of different regions (Teixeira et al. 1995; Campos et al. 2012; Martins et al. 2014). However, the factors involved in the predominance of a MRSA clone over another remains unclear (Figueiredo and Ferreira 2014). The accessory gene regulator (Agr) is the main S. aureus quorum-sensing system in the control of assorted virulence factors and key mechanisms associated with the pathogenesis of S. aureus infections, including biofilm development (Novick et al. 2000; Ferreira et al. 2012).

The agr locus has two divergent promoters, P2 and P3. P2 promoter guides the transcription of the genes agrBDCA whereas P3 promoter controls the transcription of the main agr effector molecule, the RNAIII. The AgrD protein is processed—to produce a small auto-inducing peptide (AIP)—and secreted by the AgrB protein. In the bacterial cell membrane, AIP binds to and activates the AgrC receptor histidine kinase. The activated AgrC phosphorylates the response
regulator AgrA that binds to P2 and P3 promoter regions to activate the transcription of the whole agr locus. RNAIII directly regulates some surface anchored proteins involved in bacterial–host interactions such as protein A (Spa), coagulase (Coa) and fibrinogen-binding protein (Fib), and indirectly modulates—via Rot inactivation—the expression of a number of exoproteins including serine proteases SspA and SspB, α-hemolysin (Hla) and lipase Geh (Novick et al. 2000; Boisset et al. 2007). Thus, the Agr system plays an important role in the severity of the staphylococcal infections (Le and Otto 2015). In addition to RNAIII, the AgrA response regulator can also regulate phenol-soluble modulins (PSM) \( \alpha \) and \( \beta \) by directly binding the promoter regions (Queck et al. 2008).

\( \text{agr} \)-dysfunctional \( \text{S. aureus} \) isolates have shown increased competitive fitness compared with \( \text{agr} \)-functional isolates (Paulander et al. 2012). In addition, the bacterial exposure to selected antimicrobials caused increased \( \text{agr} \) expression and consequent reduction in bacterial fitness (Paulander et al. 2012).

Here we report the complete genome sequence of the MRSA isolate HC1335 (GenBank accession number CP012012.1), a ST239 isolate, presenting a natural \( \text{agr} \) dysfunction, obtained from nasal colonization of a patient in a home care system, in 2001, as part of a genome project with ST239 isolates collected from infection and colonization cases in Brazil.

The genome of HC1335 consists of one circular chromosome with 2,976,370 bp and GC content of 32.89% (fig. 1). Functional annotation showed 2,740 protein-coding sequences (CDS)—2,330 of them were assigned to known functions and 410 to unknown categories—and a total of 39 pseudogenes were annotated. In addition, the genome harbors four copies of 16S rRNA, four of 23S rRNA, and five of 5S rRNA, and 57 tRNA genes. We have previously reported the complete genome sequence of an \( \text{agr} \)-functional ST239 isolate named BMB9393, collected from bloodstream infection in a hospitalized patient (Costa et al. 2013; GenBank accession number CP005289.1). Those two isolates share a core
genome with 2,417 coding sequences (CDS). HC1335 genome has 13 exclusive CDS, the vast majority (11) located in a phage region related to phage φNM25 at coordinates 1,169,991 to 1,192,826 (supplementary table S1, Supplementary Material online). In addition, HC1335 has an oxide nitric reductase, quinol dependent (qNOR) that lacks the cytochrome c component and uses quinol as the electron donor. This enzyme is present in several pathogens that are not denitrifiers, and the qNOR of HC1335 has 100% amino acid sequence identity with that of S. aureus Z172 (Uniprot: A0A0E1AHM6). The other exclusive CDS is an uncharacterized acyltransferase with 100% amino acid sequence identity with that of S. aureus Z172 (Uniprot: A0A0E1AGX5). The S. aureus Z172 is also a ST239-SCCmec III hospital-associated MRSA, isolated in Taiwan, whose genome sequence had already been deposited in the GenBank (accession numbers: CP006838, CP006839, and CP006840). Whole genome alignment revealed a single inversion of approximately 0.4 Mb between HC1335 and BMB9393 chromosomes, whose sequence is highly conserved between these genomes (fig. 1). Point mutation analysis revealed only three high quality single nucleotide polymorphisms (SNPs) in coding regions, all of them leading to non-conservative amino acid changes. The three predicted CDS affected by these substitutions showed putative functions of AraC family of transcriptional regulator (SAHC1335_00312; amino acid substitution Y25S at five amino acids from the sulfurase domain), carbamoyl-phosphate synthase large chain (SAHC1335_02130; amino acid substitution P289L in the phosphate transferase domain), and in a MOSC domain-containing protein (SAHC1335_00312; amino acid substitution Y25S at five amino acids from the sulfurase domain).

Bacteriophage analyses detected five phage elements in the HC1335 genome, the bacteriophages φNM3-like, YMC/09-like and φMR25-like as intact prophages, the PT1028-like as a questionable prophage and an incomplete phage region probably related with the staphylococcal phage 96. Phage φNM3-like, is a β-hemolysin converting prophage, and carries the genes chp, scn and scn, encoding for the chemotaxis-inhibiting protein–CHIPS (de Haas et al. 2004), staphylococcal complement inhibitor–SCIN (Rooijakkers et al. 2005) and staphylokinase–SAK (Sako and Tsuchida 1983), respectively. These genes are part of innate immune evasion cluster (IEC) and provides S. aureus with a unique mechanism to adapt to and counteract the human host (Van Wamel et al. 2006). However, the sea gene encoding the staphylococcal enterotoxin A—frequently carried by this phage—was absent in the HC1335 genome. None confirmed CRISPR repeat was found in the chromosome of this isolate.

The search for possible dysfunction in virulence regulation caused by transposase insertions revealed that HC1335 isolate has a transposase element—derived from the insertional element (IS) named IS256—inserted in the agrC gene (fig. 2A) at nucleotide 543 (fig. 2B), which was absent in the BMB9393 genome. This insertion truncated the agrC with consequent inhibition of the gene, affecting the regulation of the whole agr locus (fig. 2C), as expected and confirmed by the experiments of RNAIII expression (fig. 2D). The IS256 element was initially characterized by Byrne et al. (1989) as part of the Tn4001. This element has 1324bp in length and carries a single CDS (encoding a 390-amino acid transposase) preceded by a potential RBS (ribosome binding site) and −35 and −10 promoter sequences. Inverted repeat (IR) sequences were found upstream and downstream the transposase gene (trp) (Byrne et al. 1989). The IS256-derived element found disrupting agrC in HC1335 has the same structure of the IS256 described by Byrne et al. (1989) but lacks part of the upstream and all the downstream IRs.

It is remarkable that we also found an IS256 element truncating the agrC gene (fig. 2A) in the genome of the ST239-SCCmec III isolate T0131 (GenBank accession number: CP002643.1) recovered in 2016 from an 87-year-old patient in China (Li et al. 2011). In the case of the Chinese isolate, the complete IS256 element was inserted at nucleotide 507, i.e. 36nt upstream the insertion site observed in the HC1335 genome (fig. 2B). Notably, an IS256 element was also found inserted at nucleotide position 1114 of the agrC gene of another MRSA strain (Cafiso et al. 2007); however, in this case, the trp was located in the agrC antisense strand (fig. 2B). Additionally, other authors have reported the insertion of IS256 elements in the agrC of staphylococci presenting distinct genetic backgrounds including non-ST239 S. aureus and Staphylococcus epidermidis isolates. However, neither agr locus sequence nor the IS256 position in the agrC were informed in those studies (Vuong et al. 2004; Shopsin et al. 2010, McEvoy et al. 2013). Taken together, these data support the idea of a convergent evolution toward agrC inactivation because the same genetic event—agrC inactivation via IS256 truncation—have paralleled occurred at different integration sites in distinct S. aureus lineages and even in different staphylococcal species (S. aureus and S. epidermidis), which, although have evolved in the same direction—leading to the inactivation of the whole Agr system—have most likely acquired the same characteristic independently. Therefore, this mechanism of agrC inactivation was likely driven by staphylococcal pathoadaptation, because Agr system is a global regulator of a plethora of virulence genes.

The presence of multiple copies of the IS256 is a common feature in MRSA genomes (McEvoy et al. 2013; Benson et al. 2014). In addition to the IS256-derived element inserted in the agrC, 20 intact copies of the IS256 element were detected in the chromosome of HC1335. In fact, Kwon et al. (2008) have positively correlated the prevalence IS256 in S. aureus genome with high-level multiresistance and biofilm capability. In addition, studies have detected the emergence of intermediate vancomycin resistance—probably driven by the selective pressure caused by the vancomycin treatment—due to IS256 insertions in the untranslated region (5’ UTR) of the autolysin.
regulator walKR, which reduced the regulator activity by 50% (Howden et al. 2010; McEvoy et al. 2013).

Benson et al. (2014) have reported an IS256 element in the genome of another MRSA clone, the so-called USA500, close to the region of the promoter sequence of rot regulator gene encoding the repressor of toxin (Rot). Those authors concluded that the insertion of the IS256 element in rot promoter region increases the expression of cytotoxins, including α-hemolysin (hla), with consequent increased virulence in an animal model (Benson et al. 2014).

Accordingly, the inactivation of rot and agr by the IS256 suggests that the insertion of this element in master gene regulators of successful MRSA clones (ST239 and USA500) may represent a customary mechanism of bacterial adaptability to different environmental and epidemiological scenario. Thus, taken together these comparative genomic studies have revealed a new role for the IS256 that by causing small change in MRSA genomes leads to an immediate reprogramming of the bacterial virulence repertory, with predictable consequences for bacterial colonization, spread, host-interaction and virulence potential.

Despite the importance of Agr in the pathogenesis of staphylococcal diseases, isolates lacking detectable Agr activity have also been collected from cases of S. aureus bacteremia (Fowler et al. 2004; Traber et al. 2008; Shopsin et al. 2010; Butterfield et al. 2011; Ferreira et al. 2012). Indeed, agr-dysfunctional S. aureus are more likely to cause persistent infections than the agr-functional isolates, resulting in increased rate of secondary infections and mortality (Fowler et al. 2004; Traber et al. 2008; Valour et al. 2015). Finally, the increased fitness observed among agr-dysfunctional isolates (Paulander et al. 2012) may represent a compensatory mechanism for the biological cost resulted from the acquisition of high number of multiresistant traits by ST239 hospital isolates. Actually, agr-dysfunction was also described in high percentage among ST239 hospital isolates from the State of Victoria, Australia (Baines et al. 2015).

**Fig. 2.** Disruption of the agrC gene in MRSA isolates by IS256 insertions. (A). Schematic representation of the agr locus of the MRSA isolates HC1335 (this work), T0131 (Li et al. 2011) and MRSAcvc No 17 (Cafiso et al. 2007). The agrBDCA, malla, and IS256 trp coding sequences (CDSs) are represented by red, orange, green, purple, blue, and pink arrows, respectively. These isolates have an IS256 sequence inserted in the agrC CDS at different insertion sites, and in opposite direction in the case of the isolate MRSAcvc No 17. (B) Green arrow represents the agrC CDS and small black arrows indicate the positions of the insertion of the IS256 element in the isolates HC1335, T0131, and MRSAcvc No 17. (C) Schematic representation of the Agr system in S. aureus based on Novick et al. (2000). Activation of the P2 promoter leads to the transcription of the RNAII region that includes the genes agrBDCA. AgrD encodes the AgrD protein (orange square) that is transported and processed by the AgrB (red rectangle) in small auto-inducer peptides, the AIP (yellow circles). Outside the cell, AIP binds to the histidine kinase receptor AgrC (green rectangle), which phosphorylates the response regulator AgrA (purple ellipse). The phosphorylated AgrA (P-AgrA) binds to both P2 and P3 promoters activating the transcription of the RNAIII—the main regulatory molecule of Agr system. Note that an insertion of a transposase element (black triangle) in the agrC (green rectangle) will produce a non-functional (X marked) AgrC protein, blocking the whole agr auto-induced circuit. (D) Expression of agr-RNAIII transcripts in the isolates BMB9393 (agr functional) and HC1335 (agr dysfunctional). The RNA preparation of the isolate BMB9393 was used as the calibrator sample for normalizing the experiment. RQ: relative quantification.

Botelho et al.  

Downloaded from https://academic.oup.com/gbe/article-abstract/8/10/3187/2939561 by guest on 28 February 2019
Genomic DNA was prepared from 50 mL tryptase soy broth (TSB) culture—37 °C for 18 h under vigorous shaking (250 RPM)—using phenol extraction and ethanol precipitation (Sambrook et al. 1989). The concentration and purity of the DNA preparation were assessed using a Qubit® 2.0 fluorometer (Invitrogen; Eugene, Oregon, USA) and 5 μg genomic DNA was used to prepare the library. Whole-genome sequencing was performed using a 454 GS FLX titanium (3-kb paired-end library) approach (Roche Diagnostics Corporation, Indianapolis, IN, USA). The assembly, based on 488,230 reads corresponding to 82,521,347 bp (27-fold coverage), was performed using Newbler v 2.6 (Roche) (Margulies et al. 2005) and Celera genome assembly v 6.1 from JCV Institute (Myers et al. 2000). High-quality adapter-free reads were used for genome assembly. Optical duplicates were removed using the wrapper extract replicates that employs CD-HIT. After de-duplication, 315.034 reads were assembled using Newbler v 2.6 (Roche). Assembled reads have mean size of 169pb, and 95% of the bases on the assembled genome have Phred quality >40, calculated by Newbler. Gaps within scaffold folds, resulting from repetitive sequences, were resolved by in silico gap filling. In order to perform this procedure, gaps intra- and inter-scaffold folds were filled by individual assemblies of the reads present in both frontiers of the gap. This was accomplished by selecting reads that formed the end of contigs adjacent to each gap stretches. Those reads were assembled separately using Newbler v 2.6 (Roche). Contiguous sequences generated by this approach that were able to complete the gap and anchor on the two adjacent contigs were added to the sequence, and thus closing the gap. This approach generated a gap- and error-free chromosome with only six ambiguous bases on the same homopolimeric-T stretch.

For gene prediction, it was applied the prediction programs Glimmer GeneMark and Prodigal (http://prodigalornl.gov; last accessed July 1, 2016) including a correction routine for start codon from the multiple alignment of “n” similar proteins identified by BLAST against NR database (http://www.ncbi.nlm.nih.gov/protein; last accessed July 1, 2016). CDSs identified after this process were annotated (Gene, product, EC number etc.) using UniProt, Kegg and NR databases. In this process, it was only considered the CDSs over 70 aa with almost one BLASTP hit against any of the three databases mentioned above. Next, a comparative analysis was done for gene content between HC1335 and BMB9393 using the RNeasy Mini kit (Qiagen, Maryland, IN, USA) and quantified by Nanodrop Lite (Thermo Fisher Scientific, Wilmington, DE, USA). The RNA quality was analyzed using RNA gel electrophoresis. Real time quantitative reverse transcription polymerase chain reaction (real time qRT-PCR) was performed using the Power SYBRW Green RNA-to-CT TM 1-Step Kit (Applied Biosystems, Foster City, CA, USA) as recommended, and the amount of transcripts were estimated by the Ct comparative method. The primers used for agr-RNAIII and 16S rRNA (endogenous control) were described previously (Ferreira et al. 2012). The run was performed in a Step OneTM Real Time PCR System (Applied Biosystems) and analyzed using the Step One Software 2.2 (Applied Biosystems).

Bacteriophage analysis was performed using Phage Search Tool (PHAST) (Zhou et al. 2011). Comparative analysis was conducted using SABIA comparative tool (Almeida et al. 2004) and whole-genome alignment using Mauve (Darling et al. 2010). SNP calling was performed using the BMB9393 genome as reference. Briefly, the deep sequencing libraries files were quality checked with FASTQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc; last accessed July 29, 2016) and further trimmed with Fastx_toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html; last accessed July 29, 2016). The trimmed reads from HC1335 were mapped against the BMB9393 genome using the Bowtie 2 mapper with one mismatch per seed region (20 nt in length). The resulting mapping files were treated with SAMtools program (http://samtools.sourceforge.net/; last accessed July 29, 2016), only mapping reads with map quality (mapQ) above 30 were kept. The Picard mark duplicates tool (http://broad-institute.github.io/picard; last accessed July 29, 2016) was used in order to flag putative sequencing artifact, such as optical duplicates. The Genome Analysis Toolkit—GATK (https://software.broadinstitute.org/gatk; last accessed July 29, 2016) was used to call the variants using default parameters. The SNPs were annotated with the snpeff tool (http://snpeff.sourceforge.net/; last accessed July 29, 2016) and custom Python scripts. Only SNPs with coverage bigger than 10 reads were considered in this analysis.

For RNA preparations, bacterial cells grown in brain-heart infusion (BHI) broth (18 h; 37 °C; 250 rpm) were collected in the stationary phase. The total RNA was prepared using the RNeasy Mini kit (Qiagen, Maryland, IN, USA) and quantified by Nanodrop Lite (Thermo Fisher Scientific, Wilmington, DE, USA). The RNA quality was analyzed using RNA gel electrophoresis. Real time quantitative reverse transcription polymerase chain reaction (real time qRT-PCR) was performed using the Power SYBRW Green RNA-to-CT TM 1-Step Kit (Applied Biosystems, Foster City, CA, USA) as recommended, and the amount of transcripts were estimated by the Ct comparative method. The primers used for agr-RNAIII and 16S rRNA (endogenous control) were described previously (Ferreira et al. 2012). The run was performed in a Step OneTM Real Time PCR System (Applied Biosystems) and analyzed using the Step One Software 2.2 (Applied Biosystems).

Supplementary Material
Supplementary table S1 is available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments
This work was supported in part by the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ grant no.: E26/102.901/2011, E26/110.625/2011 and E26/111.663/2013); and the Conselho Nacional de
Desenvolvimento Científico e Tecnológico (CNPq grant no. 472034/2012-0). Botelho AMN was the recipient of a fellowship from Fundação CAPES/FAPERJ.

**Literature Cited**


Bacteriol. 188:1310–1315.


