

Published in final edited form as:

*Science*. 2010 January 22; 327(5964): 469–474. doi:10.1126/science.1182395.

## Evolution of MRSA During Hospital Transmission and Intercontinental Spread

Simon R. Harris<sup>1,\*</sup>, Edward J. Feil<sup>2,\*</sup>, Matthew T. G. Holden<sup>1</sup>, Michael A. Quail<sup>1</sup>, Emma K. Nickerson<sup>3,4</sup>, Narisara Chantratita<sup>3</sup>, Susana Gardete<sup>5,6</sup>, Ana Tavares<sup>5</sup>, Nick Day<sup>3,7</sup>, Jodi A. Lindsay<sup>8</sup>, Jonathan D. Edgeworth<sup>9,10</sup>, Hermínia de Lencastre<sup>5,6</sup>, Julian Parkhill<sup>1</sup>, Sharon J. Peacock<sup>3,4</sup>, and Stephen D. Bentley<sup>1,†</sup>

<sup>1</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. <sup>2</sup>Department of Biology and Biochemistry, University of Bath, South Building, Claverton Down, Bath BA2 7AY, UK. <sup>3</sup>Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. <sup>4</sup>Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. <sup>5</sup>Laboratory of Molecular Genetics, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, 2780-156 Oeiras, Portugal. <sup>6</sup>Laboratory of Microbiology, The Rockefeller University, New York, NY 10065, USA. <sup>7</sup>Centre for Clinical Vaccinology and Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 7LJ, UK. <sup>8</sup>Centre for Infection, Department of Cellular and Molecular Medicine, St. George's, University of London, Cranmer Terrace, London SW17 0RE, UK. <sup>9</sup>Department of Infectious Diseases, King's College London, Guy's, King's, and St. Thomas' Medical School, Guy's Hospital, London SE1 9RT, UK. <sup>10</sup>Directorate of Infection, Guy's and St. Thomas' National Health Service Foundation Trust, London SE1 7EH, UK.

### Abstract

Current methods for differentiating isolates of predominant lineages of pathogenic bacteria often do not provide sufficient resolution to define precise relationships. Here, we describe a high-throughput genomics approach that provides a high-resolution view of the epidemiology and microevolution of a dominant strain of methicillin-resistant *Staphylococcus aureus* (MRSA). This approach reveals the global geographic structure within the lineage, its intercontinental transmission through four decades, and the potential to trace person-to-person transmission within a hospital environment. The ability to interrogate and resolve bacterial populations is applicable to a range of infectious diseases, as well as microbial ecology.

The development of molecular typing techniques has been instrumental in studying the population structure and evolution of bacterial pathogens. Sequence-based approaches, such as multilocus sequence typing (MLST) (1), have resulted in large searchable databases of the most clinically important species. However, MLST defines variation within a very small sample of the genome and cannot distinguish between closely related isolates. Full-genome sequencing provides a complete inventory of microevolutionary changes, but this approach is impractical for large population samples. The use of next-generation sequencing technologies, such as Illumina Genome Analyzer, bridges this gap by mapping genome-wide

Copyright 2010 by the American Association for the Advancement of Science; all rights reserved.

<sup>†</sup>To whom correspondence should be addressed. [sdb@sanger.ac.uk](mailto:sdb@sanger.ac.uk).

\*These authors contributed equally to this work.

Supporting Online Material

[www.sciencemag.org/cgi/content/full/327/5964/469/DC1](http://www.sciencemag.org/cgi/content/full/327/5964/469/DC1) Materials and Methods Figs. S1 to S4 Tables S1 to S4

single-nucleotide polymorphisms (SNPs) and insertions or deletions (indels) to a reference sequence. The use of index adapters to create individually tagged genomic libraries provides the means to generate data for multiple bacterial isolates on a single sequencer lane and makes it feasible to rapidly generate whole-genome DNA sequence data for large population samples of bacteria.

Health care-associated, methicillin-resistant *Staphylococcus aureus* (HA-MRSA) is a globally important human pathogen. Current typing methods resolve the majority of HA-MRSA isolates into a small number of widely disseminated clonal lineages (2). One such clone, defined by MLST as sequence type 239 (ST239), is multiply antibiotic-resistant and accounts for at least 90% of HA-MRSA throughout China (3), Thailand (4), Turkey (5), and probably much of mainland Asia (6). ST239 has been detected in South America (7, 8) and is currently circulating in Eastern Europe (9-11). Variants of ST239 correspond to the epidemic MRSA(1)-1, -4, -11, Brazilian, Portuguese, Hungarian, and Viennese clones, which are distinguished on the basis of variation within the large type III SCC*mec* element, *spa* data, and subtle differences by pulsed-field gel electrophoresis (PFGE). Despite this variation, current typing methods provide little discriminatory power for subtyping ST239 isolates within a given region because single variants that undergo clonal expansion can dominate in hospitals throughout a large geographic area.

To investigate the utility of a second-generation DNA sequencing platform for high-resolution genotyping and investigation of the microevolutionary events within MRSA, we analyzed 63 ST239 isolates (table S1) from two distinct samples (12). The first sample, consisting of 43 isolates from a global collection recovered between 1982 and 2003, provides a snapshot of the global ST239 population. One of these isolates (TW20) was sequenced to completion to provide a reference for analysis. The second sample of 20 isolates, derived from patients at the Sappasithprasong hospital in northeast Thailand within a 7-month period, provides a very closely related group, potentially linked via a chain of transmission.

Mapping reads for each isolate against TW20 (table S2) identified 6714 high-quality SNPs. These SNPs had a markedly uneven distribution across the genome (fig. S1A), largely related to whether the SNP resided in the core (present in all sample isolates) or accessory regions of the genome. The accessory genome primarily comprised mobile genetic elements (MGEs) such as phage, transposons, SCC*mec*, and genomic islands that are known to constitute a major source of variation between *S. aureus* genomes (13). Because MGEs have an inherent potential for horizontal transfer between isolates, which could confound phylogenetic interpretations, we distinguished between the “core” and “noncore” genome for subsequent analysis.

The maximum likelihood phylogeny presented in Fig. 1 was reconstructed by using the 4310 variable sites in the core genome (table S3). We are confident that our approach has resulted in a robust tree. First, we noted little evidence of homoplasy (convergent evolution); of the 4310 sites that exhibited a SNP, only 38 (0.88%) were homoplastic (cannot be explained without convergence when mapped onto the tree) (Table 1). Notably, many of the homoplastic SNPs were in genes involved in drug resistance, with 10 corresponding to mutations known to confer resistance. Secondly, the tree showed a striking consistency with geographic source (Fig. 1). The South American isolates, with one exception, clustered tightly within a highly distinct and uniform clade, which may reflect a recent expansion of a single variant throughout the continent. Similarly, the Thai and Chinese isolates formed a single, although more diverse, Asian clade. The European isolates were more diverse still, with most positioned basally on the tree, consistent with a possible European origin for ST239. Within the European isolates, there was also evidence of geographical clustering.

There were several exceptions to this geographical structure that illustrate the intercontinental spread of MRSA. Two PFGE-distinguishable clones of ST239 are known to have dominated in Portuguese hospitals during the 1990s: the Portuguese clone in the early 1990s and the Brazilian clone that appeared in 1997. All seven Portuguese clone isolates recovered between 1990 and 1993 clustered together, whereas the three Brazilian clone isolates clustered within the South American clade, strongly supporting the hypothesis that this second wave in Portugal resulted from the introduction of a South American variant.

More intriguing were two European isolates that clustered within the Thai clade: DEN907, isolated in Denmark, and TW20, from a large 2-year outbreak at a London hospital (14). In addition to the core SNPs, both isolates contain the  $\phi$ SP $\beta$ -like (TW20) prophage characteristic of the Asian clade (fig. S1B). Records for the Danish isolate indicated that the patient was Thai, consistent with its position on the tree. The position of TW20 is less readily explained and potentially points to a single intercontinental transmission event, most likely from southeast Asia, that sparked the London outbreak.

Although the current isolate collection did not permit a robust temporal analysis, a linear regression of root-to-tip distances against the year of sampling showed a strong correlation, with older isolates positioned more basally (fig. S2). The estimated mutation rate for the isolate collection was  $3.3 \times 10^{-6}$  [95% confidence interval (CI) from  $2.5 \times 10^{-6}$  to  $4.0 \times 10^{-6}$ ] per site per year and would date the most recent common ancestor of ST239 to the mid to late 1960s, a period contemporaneous with the emergence of MRSA in Europe (15). This rate is about 1000 times faster than the canonical substitution rate estimate for *E. coli* (16) but more in line with recent rate estimates based on analyses of more closely related bacterial genomes (17, 18). Potential explanations for this could include a reduction in effective population size, leading to increased accumulation of mutations (although we have no evidence of this), or the possibility that some of the core SNPs were transferred by recombination, although the low level of homoplasy suggests that recombination has been rare. Alternatively, it may be that the greater resolution of our analysis allows us to determine the rate of mutation in the population before selection has had time to purify out those that are detrimental. This explanation implies that purifying selection acts on all mutations, including intergenic and synonymous sites, but over longer time periods, as suggested by Moran *et al.* (17) and shown for nonsynonymous mutations by Rocha *et al.* (19).

In addition to providing evidence for intercontinental transmission of ST239 variants, these data also hold the promise of revealing fine-scale transmission events between or within single hospitals. Our data included 20 isolates collected over 7 months at a single hospital in Thailand. These isolates were surprisingly divergent when compared with the South American clade (which encompasses isolates from Brazil, Chile, Argentina, and Uruguay). However, five isolates were differentiated by only 14 SNPs: four isolates (S21, S24, S39, and S42) obtained within a 16-day period and the remainder (S81) isolated 11 weeks later. These times of isolation are consistent with our estimated mutation rate of one core SNP every 6 weeks. We examined the possibility of an epidemiological link between these five isolates and noted that the patients were located in wards in adjacent blocks of the hospital and that these wards were not represented in the more divergent isolates. This result has important implications for infection control and generates invaluable information for interventions to target MRSA transmission.

Typing methods, such as *spa* and PFGE, are routinely used for epidemiological studies of *S. aureus* and other bacteria and can distinguish between different ST239 variants. We explored the extent to which the variation assayed by these methods is consistent with the high-resolution SNP data. Overall, we found high levels of consistency between *spa* type

and phylogenetic position (Fig. 2), with only a single example of a *spa* type being shared by unrelated isolates (GRE317 and HU25). This finding contrasts with the study of Nübel *et al.* (20), who noted inconsistencies between the *spa* data and SNP data for the ST5 lineage. One possible explanation for this discrepancy is that there has been insufficient time to accumulate numerous *spa* homoplasies within the younger ST239 clone.

PFGE data for the isolates (excluding the Thai isolates) divided the collection into 10 clusters (fig. S3). Again, there was a large degree of consistency between the PFGE clusters and the tree (Fig. 2). However, there were some incompatibilities. For example, cluster 6 was found in unrelated European and Asian isolates. Although certain prophage and MGEs are associated with specific clades [e.g.,  $\phi$ SP $\beta$ -like (TW20) prophage with the Asian clade], the inconsistencies here are likely to be due to the frequent gain and loss of MGEs, which can have dramatic effects on PFGE patterns.

By analyzing whole-genome data of a collection of MRSA ST239, we have gained new insights into fundamental processes of evolution in an important human pathogen. By creating a precise and robust phylogeny for the collection, we now have a highly informative perspective on the evolution of the clone.

These observations point to a limited number of successful intercontinental transmission events and expansion of subclonal variants that in some cases have become dominant in their new geographical region. The potential to detect these new introductions and target heightened infection control interventions, as occurred in the London TW20 outbreak, has clear public health implications and highlights the need for more informed global surveillance strategies. Equally important is the achievement of absolute discrimination of isolates within a single clinical setting, even those recovered only days apart, and the ability to use this SNP data to inform epidemiological analysis. Multiple additional costly infection control interventions are often used to reduce MRSA transmission supported by patient, staff, and environmental screening programs. The estimated rate of core genome divergence (1 SNP per ~6 weeks) should provide sufficient diversity to separate recent from distant transmission events, thereby dramatically improving contact tracing in endemic and outbreak settings and allowing targeting of diagnostics and interventions according to need. The additional variation from noncore regions provides supplementary discriminatory power and may inform the design of bespoke typing schemes for specific clones and locales.

From these data, we have described an estimated time frame for the emergence of a bacterial pathogen clone and how it has subsequently evolved. Of particular importance is the observation that over a quarter (28.9%) of the homoplasies detected can be directly related to evolution of resistance to antibiotic drugs currently in use (21-26), confirming clinical practice as a major driver of pathogen evolution and lending heightened importance to understanding the relevance of other homoplasies. Such insights inform future surveillance strategies for the detection of emerging clones and management of epidemic spread. We fully anticipate that, as the technology and analytical methods improve, the approach described here will underpin the next wave of molecular data for epidemiological and microevolutionary studies in bacteria.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

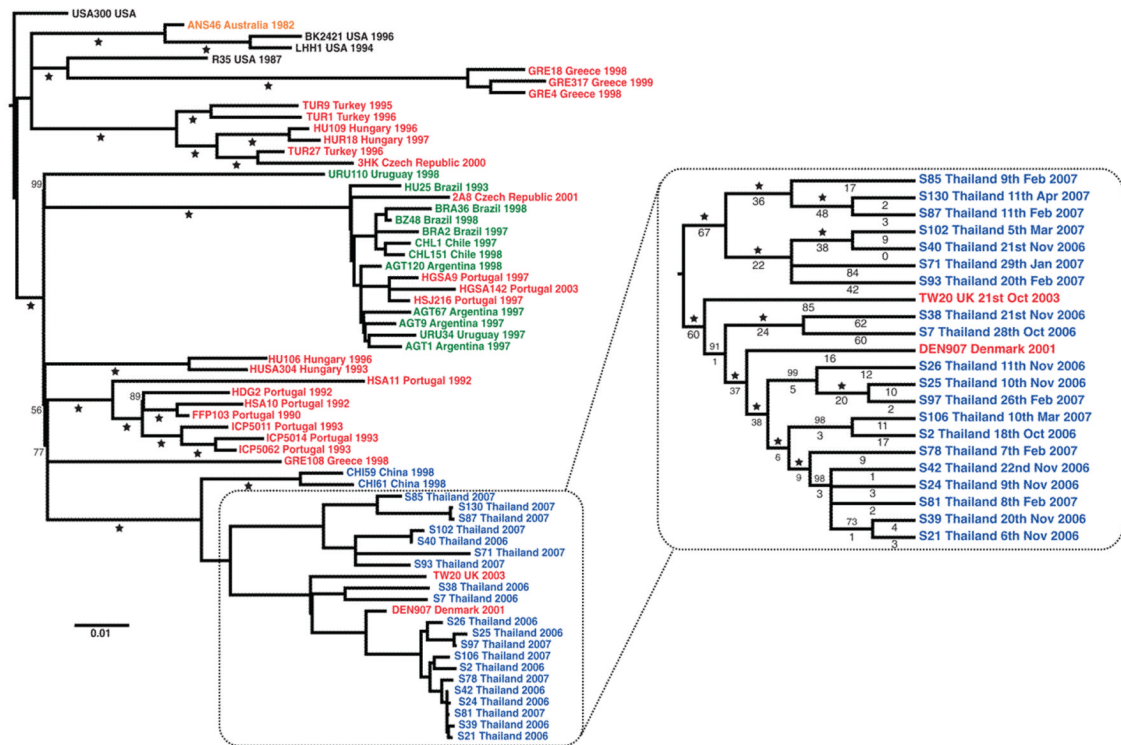
## Acknowledgments

The Sanger Institute is core funded by the Wellcome Trust. We thank C. Milheiro and J. D. Cockfield for preparation of genomic DNA and G. Dougan and the Sanger Institute Sequencing and Informatics groups for

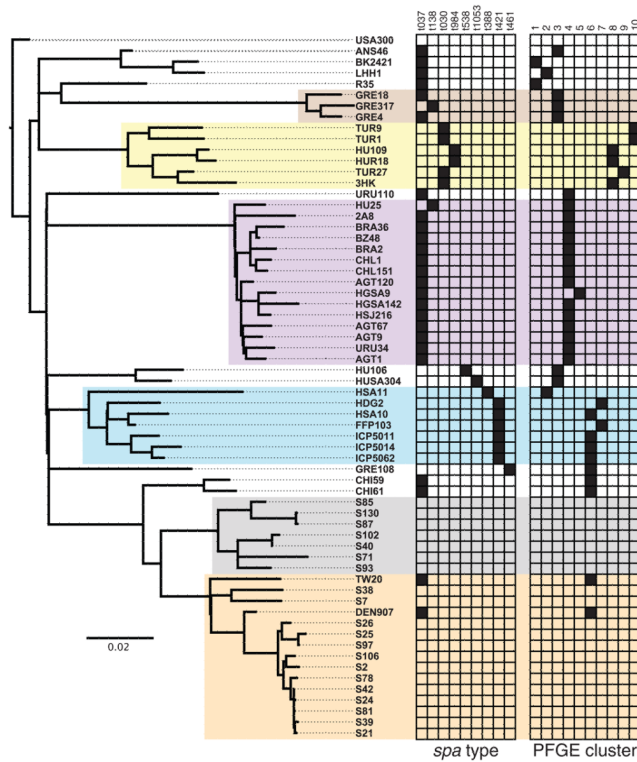
general support. S.G. and A.T. were supported by grants SFRH/BPD/25403/2005 and SFRH/BD/44220/2008, respectively, from Fundação para a Ciência e Tecnologia, Portugal. E.K.N., N.C., N.D., and S.J.P. were funded by the Wellcome Trust. Funding for the sequencing of the TW20 genome was provided by Guy's and St. Thomas' Charity. J.D.E. receives funding from the Department of Health via the National Institute for Health Research's comprehensive Biomedical Research Centre award to Guy's and St. Thomas' National Health Service Foundation Trust in partnership with King's College London. The Illumina Genome Analyzer reads are deposited in the Short Read Archive (National Center for Biotechnology Information) under the accession no. ERA000102. The annotated chromosome of TW20 has been submitted to European Molecular Biology Laboratory with the accession number FN433596.

## References and Notes

1. Maiden MC, et al. *Proc. Natl. Acad. Sci. U.S.A.* 1998; 95:3140. [PubMed: 9501229]
2. Enright MC, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2002; 99:7687. [PubMed: 12032344]
3. Xu BL, et al. *J. Hosp. Infect.* 2009; 71:245. [PubMed: 19157644]
4. Feil EJ, et al. *J. Clin. Microbiol.* 2008; 46:1520. [PubMed: 18234867]
5. Alp E, et al. *J. Infect.* 2009; 58:433. [PubMed: 19446883]
6. Ko KS, et al. *J. Clin. Microbiol.* 2005; 43:421. [PubMed: 15635004]
7. Oliveira DC, Tomasz A, de Lencastre H. *Microb. Drug Resist.* 2001; 7:349. [PubMed: 11822775]
8. Vivoni AM, et al. *J. Clin. Microbiol.* 2006; 44:1686. [PubMed: 16672394]
9. Bartels MD, et al. *Eur. J. Clin. Microbiol. Infect. Dis.* 2008; 27:757. [PubMed: 18481123]
10. Szczepanik A, Koziol-Montewka M, Al-Doori Z, Morrison D, Kaczor D. *Eur. J. Clin. Microbiol. Infect. Dis.* 2007; 26:29. [PubMed: 17180608]
11. Melter O, et al. *J. Clin. Microbiol.* 2003; 41:4998. [PubMed: 14605130]
12. Materials and methods are available as supporting material on Science Online.
13. Lindsay JA, Holden MTG. *Trends Microbiol.* 2004; 12:378. [PubMed: 15276614]
14. Edgeworth JD, et al. *Clin. Infect. Dis.* 2007; 44:493. [PubMed: 17243050]
15. de Lencastre, H.; Tomasz, A. *Evolutionary Biology of Bacterial and Fungal Pathogens*. Baquero, F.; Nombela, C.; Cassell, GH.; Gutierrez, JA., editors. American Society for Microbiology; Washington, DC: 2008. p. 333-346.
16. Ochman H, Elwyn S, Moran NA. *Proc. Natl. Acad. Sci. U.S.A.* 1999; 96:12638. [PubMed: 10535975]
17. Moran NA, McLaughlin HJ, Sorek R. *Science.* 2009; 323:379. [PubMed: 19150844]
18. Falush D, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2001; 98:15056. [PubMed: 11742075]
19. Rocha EP, et al. *J. Theor. Biol.* 2006; 239:226. [PubMed: 16239014]
20. Nübel U, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:14130. [PubMed: 18772392]
21. Besier S, Ludwig A, Brade V, Wichelhaus TA. *Mol. Microbiol.* 2003; 47:463. [PubMed: 12519196]
22. Dale GE, et al. *J. Mol. Biol.* 1997; 266:23. [PubMed: 9054967]
23. Hurdle JG, O'Neill AJ, Chopra I. *J. Antimicrob. Chemother.* 2004; 53:102. [PubMed: 14657089]
24. Tanaka M, Wang T, Onodera Y, Uchida Y, Sato K. *J. Infect. Chemother.* 2000; 6:131. [PubMed: 11810552]
25. Vickers AA, Potter NJ, Fishwick CW, Chopra I, O'Neill AJ. *J. Antimicrob. Chemother.* 2009; 63:1112. [PubMed: 19383727]
26. Wichelhaus TA, et al. *Antimicrob. Agents Chemother.* 2002; 46:3381. [PubMed: 12384339]
27. Harmsen D, et al. *J. Clin. Microbiol.* 2003; 41:5442. [PubMed: 14662923]



**Fig. 1.** Phylogenetic evidence for intercontinental spread and hospital transmission of ST239 isolates. Maximum likelihood phylogenetic tree based on core genome SNPs of ST239 isolates, annotated with the country and year of isolation. The continental origin of each isolate is indicated by the color of the isolate name: blue, Asia; black, North America; green, South America; red, Europe; and yellow, Australasia. Bootstrap values are shown below each branch, with a star representing 100% bootstrap support. The scale bar represents substitutions per SNP site. A cladogram of the Thai clade is displayed for greater resolution with bootstrap values (above the branch), number of distinguishing SNPs (below the branch), and isolates labeled with date of isolation, where known.



**Fig. 2.** Comparison of phylogeny with traditional typing techniques. Maximum likelihood phylogenetic tree based on core genome SNPs of ST239 isolates, annotated with *spa* typing databased on the RIDOM scheme (27), and PFGE typing databased on BioNumerics (version 4.0, Applied Maths, Ghent, Belgium) clustering (excluding the Thai hospital isolates and USA300, which had not been typed). The most common *spa* type was t037, which accounted for all but one of the isolates corresponding to the South American clade but was also represented among a scattering of isolates from Europe and Asia, suggesting that t037 represents the ancestral ST239 *spa* type (the plesiomorphic state). Solid boxes in the appropriate column indicate the respective *spa* type (left grid) and PFGE cluster (right grid) of the strain. Major clades in the tree are shaded for clarity.

**Table 1**

Homoplasies identified in the core regions of ST239 isolates. The SNP substitutions listed relate to the predicted forward strand of the TW20 chromosome. Isolates where homoplasies were detected are indicated, and where the isolates share the same node as illustrated in the phylogenetic tree in Fig. 1 they are in parentheses. bp indicates base pairs.

SNP position	Region	Isolates	SNP	Substitution	Antibiotic
7254	DNA gyrase subunit A (GyrA)	HUSA304, (S85, S130, S87)	T→G	Ser <sup>84</sup> →Ala <sup>84</sup> (Ser84Ala)*	
7255	GyrA	(BK2421, LHH1), (GRE18, GRE317, GRE4), (TUR9, TUR1), (TUR27, 3HK), (URU110, HU25, 2A8, BRA36, BZ48, BRA2, CHL1, CHL151, HGSA9, HGSA142, HSI216, AGT67, AGT9, URU34, AGT1), HUSA304, GRE108, (CHI59, CHI61), (S102, S40, S71, S93, TW20, S38, S7, DEN907, S26, S25, S97, S106, S2, S78, S42, S24, S81, S39, S21)	C→T	Ser84Leu	Quinolone (23)
7266	GyrA	(HU109, HUR18), AGT120, HU106, (ICP5014, ICP5062)	G→A	Lys88Glu	Quinolone (23)
133864	Immunoglobulin G binding protein A precursor	HU25, GRE108	G→A	Synonymous	–
134787	92 bp upstream of immunoglobulin G binding protein A precursor	3HK, (HU25, BZ48, BRA2, CHL1, AGT120, HGSA142, HSI216, AGT67, AGT9, AGT1)	G→T	Intergenic	–
278498	129 bp upstream of putative acetyl-coenzyme A transferase	(ANS46, BK2421, LHH1, R35, GRE18, GRE317, GRE4, TUR9, TUR1, HU109, HUR18, TUR27, 3HK), GRE108	T→C	Intergenic	–
436474	34 bp upstream of putative dioxygenase	(TUR9, TUR1, HU109, HUR18, TUR27, 3HK), HSA10	C→T	Intergenic	–
594883	Tetrapyrrole (corrin/porphyrin) methylase family protein	(BK24210, LHH1), TUR9	C→T	Pro49Ser	–
657696	DNA-directed RNA polymerase beta chain protein (RpoB)	GRE4, HSI216, GRE108, HDG2	C→A	Asp471Glu	Rifampin (25)
657724	RpoB	(GRE18, GRE317, GRE4), (TUR9, TUR1, HU109, HUR18, TUR27, 3HK), (HU25, 2A8, BRA36, BZ48, BRA2, CHL1, CHL151, AGT120, HGSA9, HGSA142, HSI216, AGT67, AGT9, URU34, AGT1), (HDG2, HSA10, FFP103), (S85, S87, S130, S93, S71, S102, S40)	C→A	His481Asn	Rifampin (25)
657869	DNA-directed RNA polymerase beta chain protein RpoB	AGT67, (S93, S71, S102, S40)	C→T	Ser529Leu	Rifampin (25)
666536	Translation elongation factor G (FusA)	(GRE18, GRE317, GRE4), GRE108	T→A	Leu461Lys <sup>†</sup>	Fusidic acid (20)
666537	FusA	(GRE18, GRE317, GRE4), GRE108	T→A		
681826	48 bp upstream of serine-aspartate repeat-containing protein C	CHI61, (S26, S97, S2, S78, S39)	C→A	Intergenic	–



SNP position	Region	Isolates	SNP	Substitution	Antibiotic
862898	Putative membrane protein	GRE4, (S87, S130)	A→C	Ser160Ala	-
1130135	63 bp upstream of FojD bifunctional protein	URU110, HGSA9	G→T	Intergenic	-
1138698	Phosphoribosylglycinamide formyltransferase (PurN)	(GRE18, GRE317, GRE4), (HUSA304, HUI106), (HSA10, FFP103)	T→A	Leu174Met	-
1172434	50 bp upstream of probable manganese transport protein	(TUR27, 3HK), (HU25, 2A8, BRA36, BZ48, BRA2, CHL1, CHL151, AGT120, HGSA9, HGSA142, HSI216, AGT67, AGT9, URU34, AGT1), (TW20, S38, S7, DEN907, S26, S97, S25, S2, S106, S78, S24, S81, S39)	T→G	Intergenic	-
1172436	52 bp upstream of probable manganese transport protein	(BK24210, LHH1), HSA11	T→C	Intergenic	-
1172444	60 bp upstream of probable manganese transport protein	(R35, GRE18, GRE317, GRE4), (HDG2, HSA10, FFP103, ICP5011, ICP5062)	C→G	Intergenic	-
1206826	ribonuclease HIII	(HU25, 2A8, BRA36, BZ48, BRA2, CHL1, CHL151, AGT120, HGSA9, HGSA142, HSI216, AGT67, AGT9, URU34, AGT1), (TW20)	C→T	Glu199Lys	-
1261219	Isoleucyl-tRNA synthetase	CHI59, TW20	G→T	Val588Phe	Mupirocin (22)
1448063	Topoisomerase IV subunit A (GIIA)	ANS46, R35, (HDG2, HSA10, FFP103), ICP5011	T→C	Ser80Phe	Quinolone (23)
1524413	Dihydrofolate reductase type I (DfrB)	GRE18, (URU110, HU25, 2A8, BRA36, BZ48, BRA2, CHL1, CHL151, AGT120, HGSA9, HGSA142, HSI216, AGT67, AGT9, URU34, AGT1)	T→C	His150Arg	Trimethoprim (24)
1524566	DfrB	(ANS46, BK2421, LHH1), (GRE18, GRE317, GRE4), (URU110, HU25, 2A8, BRA36, BZ48, BRA2, CHL1, CHL151, AGT120, HGSA9, HGSA142, HSI216, AGT67, AGT9, URU34, AGT1), (HU106, HUSA304, HDG2, HSA10, FFP103, ICP5011, ICP5014, ICP5062)	A→T	Phe99Tyr	Trimethoprim (21)
1524789	DfrB	LHH1, 2A8	G→A	Synonymous	-
1525796	Thymidylate synthase (ThyA)	LHH1, 2A8, GRE108	G→A	Synonymous	-
1525817	ThyA	LHH1, 2A8, GRE108	G→A	Synonymous	-
1525832	ThyA	LHH1, 2A8, GRE108	G→A	Synonymous	-
1640281	Glyoxalase/bleomycin resistance protein/dioxygenase superfamily protein	ICP5014, (CHI59, CHI61)	T→G	Synonymous	-
1689862	Putative transcriptional repressor (CcpN)	(BK24210, LHH1), (HU106, HUSA304)	C→T	Synonymous	-
1755814	Probable cell wall amidase (LytH)	HDG2, (S85, S87, S130, S93, S71, S102, S40, TW20, S38, S7, DEN907, S26, S25, S97, S106, S2, S78, S42, S24, S81, S39, S21)	A→G	Pro63Ser	-
1921379	Bifunctional riboflavin biosynthesis	ANS46, URU110	G→T	Asn208Lys	-

SNP position	Region	Isolates	SNP	Substitution	Antibiotic
2334865	protein (RibD)	TUR1, S40	G→A	Ser43Phe	-
2753531	Protein SprT-like	(BK24210, LHH1), GRE18	A→T	Intergenic	-
2828688	458 bp upstream of conserved hypothetical protein	(TUR9, TUR1, HU109, TUR27, 3HK), GRE108, CHI59	T→C	Intergenic	-
2828714	200 bp downstream of putative exported protein	(TUR9, TUR1, HU109, TUR27, 3HK), GRE108, CHI59, (S38, DEN907, S26, S25, S97, S106, S2, S78, S42, S24, S81, S39, S21)	G→T	Intergenic	-
2859765	226 bp downstream of putative exported protein	39 bp upstream of O-acetyltransferase (OatA)	C→T	Intergenic	-

\* Change from serine to alanine occurs due to accompanying SNP (7255) within the same codon.

<sup>†</sup> Change from leucine to lysine due to the presence of both SNPs (666536, 666537) within the same codon.