

RESEARCH HIGHLIGHT

Irreconcilable differences: divorcing geographic mutation and recombination rates within a global MRSA clone

Todd J Treangen^{1,2} and Adam M Phillippy^{1,2*}

See related Research: <http://genomebiology.com/2012/13/12/R126>

Abstract

A growing resource of methicillin-resistant *Staphylococcus aureus* (MRSA) genomes uncovers intriguing phylogeographic and recombination patterns and highlights challenges in identifying the source of these phenomena.

MRSA phylogeography revisited

Microbes possess remarkable adaptability, including the capacity to rapidly acquire antibiotic resistance, threatening our ability to fight infectious disease epidemics [1]. Understanding the mechanisms of resistance and spread is critical for defense against such threats, but traditional genotyping methods lack sufficient resolution and speed. Recent improvements in throughput and cost have made whole-genome sequencing a promising alternative [2]. In 2010 Bentley and colleagues [3] published a groundbreaking survey of 63 temporally and geographically diverse isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) clone ST239, demonstrating pathogen surveillance across four decades at the resolution of single-nucleotide polymorphisms (SNPs). In this issue of *Genome Biology*, Feil and colleagues [4] build on this seminal study by sequencing 102 additional ST239 isolates and analyzing the recombination trends of this important pathogen. Their updated sampling and analyses confirm the previously reported phylogeographic clustering, but also raise important new questions and highlight the challenge of accurately quantifying bacterial recombination rates.

Sources of diversity: recombination, horizontal transfer, and mutation

Feil and colleagues [4] meticulously address the question of MRSA diversity by applying a population genomics approach to 165 global isolates. Specifically, the authors report variation in recombination rates between phylogeographically distinct subgroups of MRSA clone ST239. The key metric presented is the ratio of SNPs caused by recombination relative to mutation (r/m), and this value is observed to vary significantly across the three subgroups analyzed: South America, Asia, and Turkey. This variation is most apparent when including mobile genetic elements (which are either manually annotated or defined as any sequence more than 1 kb long not present in all isolates), but it is also apparent in the core genome (sequences conserved in all isolates, excluding mobile genetic elements). The authors speculate about genomic characteristics, population characteristics, or transmission dynamics as possible sources, but the true cause of the observed variation remains an intriguing open question.

The three described subgroups are apparent from the core-genome phylogeny, with deep, well-supported branches separating them from the rest of the phylogenetic tree. The authors [4] argue that this reflects discrete introductions from Europe in the 1980s and 1990s, followed by region-specific diversification of the founding clones. In addition to these top-level phylogeographic groups, there is evidence of hierarchical population structure on multiple regional scales, from individual cities to countries to continents. This ability to resolve evolutionary and transmission dynamics across such a wide temporal and geographic range reinforces an optimistic outlook that future epidemics can be tracked, and countered, in real time with the help of whole-genome sequencing.

This technique of whole-genome typing depends on the identification of high-quality core-genome SNPs from conserved, non-recombined regions of the genome. Thus, it is critically important that the SNPs selected for tree building stem from unique regions of vertical

*Correspondence: phillippya@nbacc.net

¹National Biodefense Analysis and Countermeasures Center, Frederick, MD 21702, USA

²Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742, USA

inheritance and not from duplicated, recombined, or horizontally transferred sequence. To accomplish this, Feil and colleagues [4] chose a careful approach involving multiple techniques, including the manual annotation of non-core elements and the computational segmentation of recombined sequences using both BRATNextGen [5] and an approach similar to ClonalFrame [6]. Highlighting the importance of these approaches, 53% of all SNPs were identified as having been introduced by recombination and excluded from the tree reconstruction. In a more extreme case, a previous study of *Streptococcus pneumoniae* showed 88% of SNPs as resulting from recombination [7].

It is clear from Feil and colleagues' results [4], and from previous work, that any attempt to trace transmission history without first identifying recombination will be prone to error. In addition, the aggressiveness of this segmentation process can directly affect both the phylogenetic tree and the value of r/m - too strict a segmentation process may bias the value of r/m , and too relaxed may bias the tree. Because of this, and other challenges outlined below, it is important to approach such analyses with a degree of caution.

Sources of bias: a call for caution

Feil and colleagues [4], along with other recent studies, lay the framework for pathogen surveillance using whole-genome sequencing. With these approaches becoming more widespread and destined to inform public health strategies, the authors are rightly cautious in acknowledging and controlling for potential sources of bias. We feel it is important to emphasize these points so that future studies may follow their lead and use improvements in technology to increase understanding of these complex phenomena. Here, we note three sources of potential bias and how they were addressed in this study: SNP filtration bias, reference bias, and sampling bias.

One source of bias lies in segmenting the genome by provenance into vertically inherited, horizontally transferred, and recombined regions. The statistical models for distinguishing simple mutation from foreign sources rely on identifying genomic regions with a higher SNP density than the background mutation rate. This approach assumes that allelic recombination and gene transfer affect only a small fraction of the genome. However, when this is not the case, as in *S. pneumoniae* [7], it becomes difficult to estimate the background mutation rate. Alternatively, it is difficult to distinguish recombined sequences when the source is closely related because there may not be a detectable difference in SNP frequency. Feil and colleagues [4] provide an excellent blueprint on how to perform this segmentation by focusing on the core genome and combing manual annotation of mobile genetic elements with two redundant methods for recombination detection.

Selection of a reference genome is a second potential source of bias. Using a single reference genome ignores mobile genetic elements present in the population that are absent from the reference. As a result, the diversity of the non-core genome will be underestimated, making the statistics regarding mobile genetic elements difficult to interpret. Feil and colleagues [4] acknowledge the effect of a reference bias when accounting for mobile genetic elements, but note that ST239 genomes are highly similar because of this clone's recent emergence and that the core genome analysis is unaffected by selection of a reference.

Sampling bias is a third and important potential source of error to be addressed. For a true summary of the population, samples must be randomly selected across both temporal and geographic domains. However, this is not feasible as microbial sampling is typically opportunistic and diverse samples, particularly from healthy individuals, are often difficult to obtain. Thus, sampling bias is often unintentionally introduced, as has been previously discussed in the context of human influenza A [8]. The authors [4] mitigated this bias by including over 150 strains across three continents to capture an impressive range of diversity of ST239. As evidenced by this and previous studies, Feil and colleagues are leaders in the politically challenging realm of global sampling, and we stress the importance of these types of international collaborations along with open data sharing.

Paving the way

This study [4] offers an exciting result - that recombination to mutation ratios seem to differ by geographic subtype. It adds to the growing knowledge of MRSA evolution and leads the way for future studies of bacterial recombination. Following the blueprint it provides, we advocate a cautious approach in light of the potential biases. Forthcoming advances in sequencing technology and bioinformatics promise to address these challenges further. New algorithms, scalable to many genomes, continue to be developed for the detection and management of recombination and horizontal gene transfer. The emergence of third-generation sequencing promises the affordable closure of bacterial genomes [9], which would eliminate reference bias and enable a greater understanding of the non-core genome. Lastly, the continued plummeting of sequencing costs will help dampen the effects of sampling bias by enabling systematic sampling approaches to include latent microbial reservoirs in both the natural and built environments. Ideally, future sequencing technologies will feed a universally deployed sensor network, capable of providing a comprehensive view of pathogen population diversity [10]. The remarkable population sequencing studies of today, such as this one [4], continue to predict a bright future in the fight against infectious disease.

Abbreviations

MRSA, methicillin-resistant *Staphylococcus aureus*; r/m , ratio of SNPs caused by recombination relative to mutation; SNP, single-nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

Published: 27 December 2012

References

1. Moran GJ, Krishnadasan A, Gorwitz RJ, Fosheim GE, McDougal LK, Carey RB, Talan DA: **Methicillin-resistant *S. aureus* infections among patients in the emergency department.** *N Engl J Med* 2006, **355**:666-674.
2. Relman DA: **Microbial genomics and infectious diseases.** *N Engl J Med* 2011, **365**:347-357.
3. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD: **Evolution of MRSA during hospital transmission and intercontinental spread.** *Science* 2010, **327**:469-474.
4. Castillo-Ramírez S, Marttinen P, Aldeljawi M, Hanage W, Westh H, Boye K, Gulay Z, Bentley SD, Parkhill J, Holden MT, Feil EJ: **Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus* (MRSA).** *Genome Biol* 2012, **13**:R126.
5. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, Corander J: **Detection of recombination events in bacterial genomes from large population samples.** *Nucleic Acids Res* 2012, **40**:e6.
6. Didelot X, Falush D: **Inference of bacterial microevolution using multilocus sequence data.** *Genetics* 2007, **175**:1251-1266.
7. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD: **Rapid pneumococcal evolution in response to clinical interventions.** *Science* 2011, **331**:430-434.
8. Bush RM, Smith CB, Cox NJ, Fitch WM: **Effects of passage history and sampling bias on phylogenetic reconstruction of human influenza A evolution.** *Proc Natl Acad Sci USA* 2000, **97**:6974-6980.
9. Bashir A, Klammer AA, Robins WP, Chin CS, Webster D, Paxinos E, Hsu D, Ashby M, Wang S, Peluso P, Sebra R, Sorenson J, Bullard J, Yen J, Valdovino M, Mollova E, Luong K, Lin S, LaMay B, Joshi A, Rowe L, Frace M, Tarr CL, Turnsek M, Davis BM, Kasarskis A, Mekalanos JJ, Waldor MK, Schadt EE: **A hybrid approach for the automated finishing of bacterial genomes.** *Nat Biotechnol* 2012, **30**:701-707.
10. Schatz MC, Phillippy AM: **The rise of a digital immune system.** *GigaScience* 2012, **1**:4.

doi:10.1186/gb-2012-13-12-181

Cite this article as: Treangen TJ, Phillippy AM: **Irreconcilable differences: divorcing geographic mutation and recombination rates within a global MRSA clone.** *Genome Biology* 2012, **13**:181.