INFECTIOUS DISEASE

Pathogen Microevolution in High Resolution

Ramy K. Aziz[1] and Victor Nizet[2]*

Published 27 January 2010; Volume 2 Issue 16 ps4

Microbial genomics has revolutionized infectious diseases and epidemiology research and is facilitating the tracking and containment of emerging biological threats. Among the most serious contemporary infectious agents are multiple antibiotic–resistant strains of the human pathogen *Staphylococcus aureus*, which present a formidable public health challenge that is no longer limited to hospitalized patients. To address key hypotheses regarding microbial strain evolution or virulence, conventional genotyping methods do not offer enough power to resolve minor changes between closely related strains. The application of next-generation high-throughput genotyping technologies, as illustrated in a recent analysis of a highly resistant *S. aureus* strain, can provide new clues about the geographical origin and intrahospital spread of important microbial pathogens.

In October 2009, fewer than 15 years after the first bacterial genome was sequenced in its entirety (*1*), the National Center for Biotechnology Information announced the attainment of the 1000th complete microbial genome sequence; this wealth of information adds to the more than 2000 viral genomes that have been fully sequenced and deposited in public databases. With recent unparalleled advances in sequencing technologies (*2*), the scientific community can soon expect 10,000 microbial genomes to be solved, broadly covering taxa from all branches in the entire tree of life (*3*). Full-fledged efforts also have been initiated to sequence and publish numerous human genomes (*4*) and microbiomes (*5*), cataloging variations among individuals and resident microflora and their potential link to infectious or inflammatory disease pathologies. Here we discuss a study by Harris *et al.* in a recent issue of *Science* (*6*) that has examined the power and functionality of high-resolution genotyping applied to bacterial molecular epidemiology.

Microbial genomics is already being translated in meaningful ways to public health epidemiology and the clinical practice of infectious disease medicine (*7*). One success story is the etiological identification, tracking, and containment of the severe acute respiratory syndrome coronavirus in 2002–2003 (*8*), a near-pandemic that otherwise could have claimed thousands more lives. Current global surveillance of swine-origin H1N1 influenza provides another example (*9*). Contemporary genome-wide approaches are being applied in the development of rapid diagnostic tools against bioterror threats and emerging infectious disease agents (*10*), microbial forensics (*11*), and reverse vaccinology platforms targeting leading human bacterial pathogens (*12*).

Genomic fingerprinting provides a basis for understanding the relatedness of microbial strains. DNA comparison is the core methodology of molecular epidemiologists, who study the emergence and geographical spread of virulent or drug-resistant clones and seek to decipher the reservoirs and transmission modes of pathogens during regional disease outbreaks. In recent years, techniques such as multilocus sequence typing (MLST) (*13*) have begun to supplant pulse-field gel electrophoresis and related DNA-banding subtyping methods, yielding greater insight into the population structure and evolutionary dynamics of leading human bacterial pathogens. However, MLST allows only a crude estimate of the total variation present in bacterial genomes, which typically exceed 2 Mb in size. MLST is further constrained by the preselection bias of the handful of genes to be analyzed, classically encoding housekeeping enzymes rather than virulence determinants central to the infectious disease process. Logic dictates that a snapshot of the entire genome of each strain isolate would provide the ultimate information trove for infectious disease detectives. The ever-increasing speed and cost-efficiency of next-generation sequencing technologies are now making this a reality.

The recent *Science* paper by Harris *et al.* (*6*) focuses on high-resolution genotyping of methicillin-resistant *Staphylococcus aureus* (MRSA) as the showcase pathogen. *S. aureus* can produce clinical disease throughout the body and in all patient groups and has the proven capacity to develop resistance to all current classes of antimicrobial agents. MRSA presents a critical global challenge to the public health (*14*), with incidence rates ranging from less than 1% of *S. aureus* disease isolates in the Netherlands and Sweden (*15*) to more than 50% in the United States (*16*), Japan (*17*), and Portugal (*15*). The U.S. Centers for Disease Control estimated that more than 90,000 Americans developed invasive MRSA in 2005 and that more than 18,000 patients died during their hospital stay from these serious infections (*18*). In recent years, severe community-acquired MRSA infections, often in healthy individuals without medical care risk factors, have been identified with alarming frequency (*19*).

Harris *et al.* conducted high-resolution genotyping using Illumina index tagging paired to a genomic sequence analyzer (*20*) to dissect subtle genomic differences among MRSA isolates from a prominent clonal lineage (sequence type 239, or ST239, by MLST) that spread around the globe over a two-decade period. The authors examined a geographically and temporally dispersed sample (43 isolates) to understand the ST239 global population structure, while a second sample (20 isolates) from a single facility in Thailand was used to study the potential chain of transmission in a hospital-infection control setting. With this technique, genome-wide information is acquired, yet without the need for complete genome sequencing of each individual isolate; in effect, single-nucleotide polymorphisms (SNPs) or insertions and deletions (indels) are mapped back to a reference genome. The approach delivers a few new insights regarding intercontinental spread and local epidemiology, as well as the microevolution of a bacterial lineage in response to antibiotic selection pressures.

Geographical structuring of SNPs among the isolates in the MRSA ST239 strain collection suggested a model to represent the evolutionary history of the clone (Fig. 1). Isolates from countries including Turkey, Portugal, and Hungary showed the greatest diversity and were most often situated toward the root of the phylogenetic tree, pointing to a possible European origin for

[1]Department of Microbiology and Immunology, Faculty of Pharmacy, Cairo University, Cairo 11562, Egypt. E-mail: ramy.aziz@salmonella.org [2]Department of Pediatrics and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA.
*Corresponding author. E-mail: vnizet@ucsd.edu

the clone. On the other hand, South American and Southeast Asian isolates formed more tightly clustered and uniform clades, indicating more recent expansion of a unique variant on each continent. Exceptions to the geographical structure suggested subsequent reintroduction of these new variants to the European continent. These included one pandemic clone of South American origin that was the most prevalent MRSA isolate in Portuguese hospitals in the late 1990s (21), and another clone of Southeast Asian origin that produced a sustained outbreak of vascular device–associated bacteremia in a London intensive care unit (22). Overall, the calculated mutation rate among the collected ST239 MRSA strains (3.3×10^{-6} per site per year) corroborated recent MLST-based analyses (23) and suggested a clonal origin of ST239 in the late 1960s, coincident with the first appearance of MRSA in European hospitals.

Overall, less than 1% of the analyzed SNPs were described as homoplasic; that is, suggestive of convergent evolution— the attainment of a particular SNP in unrelated lineages. More than a quarter of these homoplasies occurred in genes known to play a role in resistance to currently used antibiotics such as trimethoprim, rifampin, mupiricin, fusidic acid, and quinolones. These findings appear to confirm that the clinical practice of antibiotic selection and use can be a major driver of pathogen microevolution. Recent studies have shown how the overuse of certain classes of broad-spectrum antibiotics can promote MRSA nasal carriage or influence hospital MRSA infection rates (24). The application of new-generation sequencing analysis to such investigations may provide further definitive bases for optimizing hospital formulary policies.

When a collection of MRSA ST239 isolates from the Southeast Asian clade recovered in the same hospital during a 7-month period was analyzed, a surprising amount of SNP variation was identified using the next-generation sequencing

Fig. 1. Phylogeography of methicillin-resistant *S. aureus* strain ST239. A map showing the geographical spread of MRSA ST239 as deduced by Harris *et al.* (6).

platform; such absolute discrimination of clinical isolates in the infection control setting could not be achieved by prior methods. Five closely related isolates could be linked to adjacent wards of the hospital that were not associated with the larger, more divergent group, thus suggesting an epidemiological link of transmission. The discriminating power of the technology promises to provide more accurate clues regarding an emerging epidemic clone and could thereby facilitate targeted interventions. Ultimately, one can envision extending the analysis to probe the complete ecology of MRSA transmission dynamics. This would entail the evaluation of non–disease-associated staphylococcal isolates, such as those producing transient colonization of patients or health care workers or those contaminating fomites or hospital environmental surfaces.

In their analysis, the authors made a key distinction between two components of the MRSA genomes (Fig. 2). The core genome represents genes shared by most strains. These are thought to be of ancestral *S. aureus* chromosomal origin, and because of their older age are subjected to higher variability. The variability of genes within the core genome is uneven and is dependent on whether individual SNPs or indels produce functional alterations in their encoded products and whether such changes are under positive, negative, or neutral selection pressure. The non-core or accessory genome contains mostly horizontally transferred genes, encoded on mobile genetic elements (MGEs) such as bacteriophages, transposons, pathogenicity islands (enriched in superantigen genes), plasmids, and the so-called staphylococcal cassette chromosome (SCC) elements (25, 26). Sequences of these genes tend to be less variable among strains in which they are present, because they were recently acquired. Among the most notorious of MGEs within MRSA accessory genomes are SCC*mec*, which encodes the altered penicillin-binding protein 2a (PBP2a') with reduced affinity for methicillin and related β-lactam antibiotics, and the bacteriophage ϕSa2(MW2), which contains the genes that encode the subunits of the Panton-Valentine leukocidin (PVL). Epidemiologically, PVL expression is strongly associated with highly virulent community-acquired MRSA clones such as USA300 (27), although mechanistic linkages of the toxin to clinical outcomes or to pathogenesis in animal models remain controversial (28–30).

The whole set of genetic variations present in a species population undergoing microevolutionary change is often referred to as a variome. To meet their goals of elucidating patterns of intercontinental or intrahospital spread of MRSA ST239 clones, Harris *et al.* focused incisively on SNPs that reside in the core variome, because horizontal gene transfer events could confound the accuracy of the deduced phylogenetic tree.
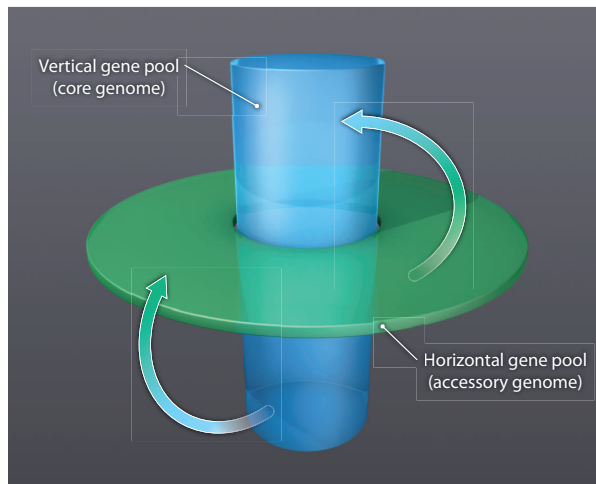
Fig. 2. Bidimensional model of a microbial pathogen variome. The variome of *S. aureus* and other pathogens can be conceptualized in a bidimensional architecture consisting of vertical and horizontal gene pools. The central "cylinder" represents the core genome and its somehow steady temporal evolution. Phylogenies and typing methods based on core genome elements (for example, restriction fragment length polymorphism and MLST) reflect only the vertical evolution of the species, whereas elements of the accessory genomes (orange halo) can produce more rapid evolutionary leaps via horizontal gene transfer. Both genomic dimensions need to be taken into consideration to produce accurate, clinically relevant phylogenies. Arrows reflect continuous bidirectional genetic exchange between the core and accessory gene pools.

We are now appreciating that when major changes are observed in the epidemiology, clinical disease spectrum, or antibiotic resistance profile of an established pathogen, the responsible genetic changes are transpiring in the accessory genome through horizontal gene transfer.

A deeper understanding of the dual dimensionality of microbial genotypes, reflected in their core and accessory genomes, is required for microbiologists and infectious disease physicians to come to grips with a chronic problem of typing and classification. For over a century, researchers have documented a contribution of horizontal gene transfer to microbial genotypes and phenotypes. However, the impact of horizontal gene transfer was greatly underestimated until the flood of new microbial genome sequences (*31*) revealed the major effects of MGEs on microbial genomes' architecture and characteristics. Two phylogenetically closely related strains can be dramatically different in their pathogenic potential, which is dependent on the recent acquisition or loss of a critical MGE. Prior strain classification schemes, including the most advanced ones such as MLST, have traditionally focused on the vertical lineage of a strain. New-generation sequencing technologies, while maintaining and enhancing analysis of the core genome for vertical phylogeny, can also provide a comprehensive content inventory of the accessory genome. This second dimension, which reflects the content of phage and other MGEs, can help predict whether the strain is associated with invasive disease potential and whether it will be resistant to certain antibiotics (even if its closest ancestor is not).

The impending application of next-generation sequencing to the molecular epidemiology of microbial pathogens will provide a veritable deluge of candidate SNPs in search of meaning. Facing potential information overload, the infectious disease research community will need to separate the proverbial wheat from the chaff. Beyond refining phylogenetic analyses as illustrated by the work of Harris *et al.*, we note that the technology may be applied fruitfully to evolutionary processes that occur during infection of the individual patient. Most leading human pathogens, including *S. aureus*, commonly exist in healthy individuals as a commensal of epithelial surfaces (for example, the upper respiratory mucosa, skin, and gut). In the course of invasive infection, as it spreads to the deep tissues and bloodstream, the organism encounters intense natural selection from phagocytic cells and other components of the innate immune system. The variant that is optimized for systemic survival may differ from that best suited for colonization, in which competition against resident microflora for epithelial adherence and nutrients produces the strongest selective pressures. This is well illustrated by the globally disseminated M1T1 clone of *Streptococcus pyogenes*, where the application of next-generation sequencing technology (*32*) showed that invasive disease follows specific mutations in the *covR/S* regulatory locus, leading to hyperencapsulation and up-regulation of toxins and immune evasion proteins (*33, 34*). Genome-wide analysis of SNPs between paired strains isolated from the bloodstream or nasal mucosa of the same patient, or between input and output strains generated in a relevant animal model, could reveal similar genetic switch mechanisms critical for invasive disease by MRSA and other disease agents.

An argument can be made that for microbial pathogens, the truth is hidden in the genome. We are getting closer to the truth.

REFERENCES AND NOTES

1. R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillipps, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, J. C. Venter, Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
2. S. C. Schuster, Next-generation sequencing transforms today's biology. *Nat. Methods* **5**, 16–18 (2008).
3. D. Wu, P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N. N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B. J. Tindall, S. D. Hooper, A. Pati, A. Lykidis, S. Spring, I. J. Anderson, P. D'haeseleer, A. Zemla, M. Singer, A. Lapidus, M. Nolan, A. Copeland, C. Han, F. Chen, J. F. Cheng, S. Lucas, C. Kerfeld, E. Lang, S. Gronow, P. Chain, D. Bruce, E. M. Rubin, N. C. Kyrpides, H. P. Klenk, J. A. Eisen, A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056–1060 (2009).
4. J. Kaiser, DNA sequencing. A plan to capture human diversity in 1000 genomes. *Science* **319**, 395 (2008).
5. P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, J. I. Gordon, The human microbiome project. *Nature* **449**, 804–810 (2007).
6. S. Harris, E. J. Feil, M. T. Holden, M. A. Quail, E. K. Nickerson, N. Chantratita, S. Gardete, A. Tavares, N. Day, J. A. Lindsay, J. D. Edgeworth, H. de Lencastre, J. Parkhill, S. J. Peacock, S. Bentley, Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, TKTK (2010).
7. L. K. McNeil, R. K. Aziz, In silico reconstruction of the metabolic and pathogenic potential of bacterial genomes using subsystems. *Genome Dyn.* **6**, 21–34 (2009).
8. L. L. Poon, Y. Guan, J. M. Nicholls, K. Y. Yuen, J. S. Peiris, The aetiology, origins, and diagnosis of severe acute respiratory syndrome. *Lancet Infect. Dis.* **4**, 663–671 (2004).
9. G. Neumann, T. Noda, Y. Kawaoka, Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **459**, 931–939 (2009).
10. C. M. Fraser, A genomics-based approach to biodefence preparedness. *Nat. Rev. Genet.* **5**, 23–33 (2004).
11. A. Dance, Anthrax case ignites new forensics field. *Nature* **454**, 813 (2008).
12. R. Rappuoli, A. Covacci, Reverse vaccinology and genomics. *Science* **302**, 602 (2003).
13. M. C. Maiden, J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, B. G. Spratt, Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3140–3145 (1998).
14. N. Woodford, D. M. Livermore, Infections caused by

Gram-positive bacteria: A review of the global challenge. *J. Infect.* **59**, (Suppl 1), S4–S16 (2009).

15. European Antimicrobial Resistance Surveillance System (EARSS), MRSA Prevalence Figures (2008) (http://www.rivm.nl/earss/database).

16. D. J. Farrell, R. E. Mendes, J. E. Ross, R. N. Jones, Linezolid surveillance program results for 2008 (LEADER Program for 2008). *Diagn. Microbiol. Infect. Dis.* **65**, 392–403 (2009).

17. H. Goto, K. Shimada, H. Ikemoto, T. OguriStudy Group on Antimicrobial Susceptibility of Pathogens Isolated from Respiratory Infections, Antimicrobial susceptibility of pathogens isolated from more than 10,000 patients with infectious respiratory diseases: A 25-year longitudinal study. *J. Infect. Chemother.* **15**, 347–360 (2009).

18. R. M. Klevens, M. A. Morrison, J. Nadle, S. Petit, K. Gershman, S. Ray, L. H. Harrison, R. Lynfield, G. Dumyati, J. M. Townes, A. S. Craig, E. R. Zell, G. E. Fosheim, L. K. McDougal, R. B. Carey, S. K. FridkinActive Bacterial Core surveillance (ABCs) MRSA Investigators, Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *JAMA* **298**, 1763–1771 (2007).

19. H. F. Chambers, F. R. Deleo, Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat. Rev. Microbiol.* **7**, 629–641 (2009).

20. Next-generation sequencing technologies are becoming more widely available and cost-efficient. Platforms by other vendors, including FLX(454) (454 Life Sciences, a Roche Company), ABI SOLID (Life Technologies), and Helicos HeliScope (Helico Bioscience Corporation), could provide similar functionalities as Agilent/Illumina for genome-wide microbial strain characterization.

21. M. Aires-de-Sousa, B. Correia, H. de LencastreMulti-laboratory Project Collaborators, Changing patterns in frequency of recovery of five methicillin-resistant *Staphylococcus aureus* clones in Portuguese hospitals: Surveillance over a 16-year period. *J. Clin. Microbiol.* **46**, 2912–2917 (2008).

22. J. D. Edgeworth, G. Yadegarfar, S. Pathak, R. Batra, J. D. Cockfield, D. Wyncoll, R. Beale, J. A. Lindsay, An outbreak in an intensive care unit of a strain of methicillin-resistant *Staphylococcus aureus* sequence type 239 associated with an increased rate of vascular access device-related bacteremia. *Clin. Infect. Dis.* **44**, 493–501 (2007).

23. D. S. Smyth, L. K. McDougal, F. W. Gran, A. Manoharan, M. C. Enright, J. H. Song, H. de Lencastre, D. A. Robinson, F. R. DeLeo, Population structure of a hybrid clonal group of methicillin-resistant *Staphylococcus aureus*, ST239-MRSA-III. *PLoS ONE* **5**, e8582 (2010).

24. M. A. Aldeyab, D. L. Monnet, J. M. López-Lozano, C. M. Hughes, M. G. Scott, M. P. Kearney, F. A. Magee, J. C. McElnay, Modelling the impact of antibiotic use and infection control practices on the incidence of hospital-acquired methicillin-resistant *Staphylococcus aureus*: a time-series analysis. *J. Antimicrob. Chemother.* **62**, 593–600 (2008).

25. J. A. Lindsay, M. T. Holden, *Staphylococcus aureus*: Superbug, super genome? *Trends Microbiol.* **12**, 378–385 (2004).

26. J. A. Lindsay, M. T. Holden, Understanding the rise of the superbug: Investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct. Integr. Genomics* **6**, 186–201 (2006).

27. F. C. Tenover, R. V. Goering, Methicillin-resistant *Staphylococcus aureus* strain USA300: Origin and epidemiology. *J. Antimicrob. Chemother.* **64**, 441–446 (2009).

28. I. G. Bae, G. T. Tonthat, M. E. Stryjewski, T. H. Rude, L. F. Reilly, S. L. Barriere, F. C. Genter, G. R. Corey, V. G. Fowler, Jr., Presence of genes encoding the panton-valentine leukocidin exotoxin is not the primary determinant of outcome in patients with complicated skin and skin structure infections due to methicillin-resistant *Staphylococcus aureus*: Results of a multinational trial. *J. Clin. Microbiol.* **47**, 3952–3957 (2009).

29. M. Labandeira-Rey, F. Couzon, S. Boisset, E. L. Brown, M. Bes, Y. Benito, E. M. Barbu, V. Vazquez, M. Höök, J. Etienne, F. Vandenesch, M. G. Bowden, *Staphylococcus aureus* Panton-Valentine leukocidin causes necrotizing pneumonia. *Science* **315**, 1130–1133 (2007).

30. J. Bubeck Wardenburg, A. M. Palazzolo-Ballance, M. Otto, O. Schneewind, F. R. DeLeo, Panton-Valentine leukocidin is not a virulence determinant in murine models of community-associated methicillin-resistant *Staphylococcus aureus* disease. *J. Infect. Dis.* **198**, 1166–1170 (2008).

31. N. Ahmed, J. E. Stajich, A flood of microbial genomes-do we need more? *PLoS ONE* **4**, e5831 (2009).

32. P. Sumby, A. R. Whitney, E. A. Graviss, F. R. DeLeo, J. M. Musser, Genome-wide analysis of group a streptococci reveals a mutation that modulates global phenotype and disease specificity. *PLoS Pathog.* **2**, e5 (2006).

33. M. J. Walker, A. Hollands, M. L. Sanderson-Smith, J. N. Cole, J. K. Kirk, A. Henningham, J. D. McArthur, K. Dinkla, R. K. Aziz, R. G. Kansal, A. J. Simpson, J. T. Buchanan, G. S. Chhatwal, M. Kotb, V. Nizet, DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. *Nat. Med.* **13**, 981–985 (2007).

34. R. K. Aziz, M. Kotb, Rise and persistence of global M1T1 clone of *Streptococcus pyogenes*. *Emerg. Infect. Dis.* **14**, 1511–1517 (2008).

35. The authors declare no conflicts of interest.