

The whole-genome sequence data generated on three separate runs provided an average of 2.2 million reads (80 Mbp) per isolate. Each read was then mapped to the TW20 reference genome, giving an average of 23X depth of coverage (two runs) for more than 92.1% of the reference genome. "The sequencing was done over the span of several weeks, which is a rapid turnaround for such a large-scale project," Dr. Bentley added. "Without the speed of the Genome Analyzer and its ability to multiplex, it would have taken years."

A total of 6,714 sites were found to contain a high-quality SNP in at least one of the mapped isolates, but they were not evenly distributed across the genome. As with all bacteria, *S. aureus* possesses a compound genome made up of two parts, the core and the accessory. While the core genome of *S. aureus* is stable, the accessory genome includes regions that are horizontally acquired from other strains within the species and even other species, and is therefore inherently more diverse. To refine the SNP list to include only those SNPs that resided on the core genome, the team defined the core variable sites as those which were present in all members of the sample. This reduced the SNP list to 4,310.

Results Yield Clear Phylogeny of MRSA Samples

Using this refined SNP data, phylogenetic trees were generated, providing a precise representation of how the ST239 clone evolved. "We were surprised and very pleased with these results," said Dr. Bentley. "In the past when SNP typing was done, you might have seen some geographical structuring of the data, but it was not nearly as clear as what we saw in this research. By using whole-genome information, rather than a panel of SNPs, we could track intercontinental transmissions."

For example, there were two European isolates that clustered within the Thai group, one previously identified in Denmark and the other from a large two-year outbreak at a London hospital. Records for the Danish isolate indicated that the patient was Thai, consistent with its position on the tree. The position of the other isolate may be explained by a single intercontinental transmission event, most likely from Southeast Asia, that sparked the London outbreak.

"Whole-genome sequencing also provided us with clarity regarding the mutations that were occurring in response to antibiotic drugs," Dr. Bentley said. "While it made sense that the pathogen was responding to human clinical practices, it was remarkable to be able to see that this was going on actively. Rather than the progenitor of the clone possessing resistance and all of its descendants having exactly the same mutation, we could see that the same mutations arose at different time periods, in different locations around the world."

The team also saw evidence of convergent evolution or homoplasies that linked with no known mutations. These mutations might be in response to clinical practice and could identify bacterial survival mechanisms that are not yet understood. "Out of the 38 cases of homoplasies that we saw, 11 were variations that had already been shown to be associated with antibiotic resistance," Dr. Bentley added. "So the question is, what about the others? In one instance, there was a group of SNPs that were upstream of a gene cluster encoding for a manganese transport system. There has been some evidence in the literature about the importance of manganese in the ability to colonize on mucosal surfaces. That might explain why HA-MRSA thrives in hospital environments."

Developing a New Bacterial Pathogen Database

Dr. Bentley and his team have moved on to the next phase of the project and are using the Genome Analyzer to assemble a large database of whole-genome information for many major bacterial human disease pathogens. This will include loci which cover the current typing schemes to align the data with these typing standards, albeit at much higher resolution. "We'll also be capturing virulence factors and antimicrobial loci," Dr. Bentley said. "We'll be able to quickly sequence new strains and compare them to this database. Based on how they match up, we can determine which pathogen it is, what virulence factors it has, how invasive it is, how likely it is to cause severe disease, and what are its antimicrobial resistances."

The ability to use next-generation sequencing platforms to quickly and inexpensively sequence multiple genomes will empower the use of whole-genome sequencing to perform real-time bacterial pathogen tracking of hospital transmission and intercontinental spread. In the middle of an outbreak, whether hospital-contained or globally wide-

"Multiplexing on the Genome Analyzer meant that we could compare multiple genomes within a species in large enough numbers that we could take an epidemiologically relevant sample size and be able to sequence all those very quickly."

spread, speed will be of the essence. "We have already demonstrated proof-of-principle that we can use whole-genome sequencing to map transmission chains in a hospital environment. For the samples collected from the hospital in Thailand, two were taken 14 days apart and were clearly related. The interpretation was that these had been transmitted from one person to another within that ward. Even though the two samples were taken only 14 days apart, we could differentiate them by six SNPs. So those SNPs had been generated just by mutation over that period of time. Once we have this type of information in real time, we can inform clinical interventions to reduce that kind of transmission."

The availability of the Genome Analyzer and the HiSeq™ 2000 for this type of analysis and the ability to multiplex whole genomes will enable more samples per experiment to yield statistically robust data. "For example, we can look at what happens during a chronic infection, such as when someone is infected with tuberculosis," Dr. Bentley said. "We'll be able to take samples over a long time period and sequence each one to see how the organism is changing. We could sequence bacterial isolate transcriptomes as well. With the cost of sequencing becoming lower, the challenge for scientists will be to spot innovative uses of the technology that will benefit health care."

