International Journal of Medical Microbiology 301 (2011) 614–618

Contents lists available at SciVerse ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.de/ijmm

Mini Review

From types to trees: Reconstructing the spatial spread of *Staphylococcus aureus* based on DNA variation

Ulrich Nübel*, Birgit Strommenger, Franziska Layer, Wolfgang Witte

Robert Koch-Institut, Burgstr. 37, 38855 Wernigerode, Germany

ARTICLE INFO

Keywords:
MRSA
MLST
PFGE
spa typing
MLVA
SNPs
Phylogeography

ABSTRACT

Tracing the spatial spread of pathogens is a key objective of molecular infectious disease epidemiology. Accordingly, a wide range of genotyping approaches have been used to monitor the dissemination of *Staphylococcus aureus* strains, from localized outbreaks to global spread. We provide a critical review of available methods, revealing that molecular markers currently in use for typing *S. aureus* acquire changes so slowly that they monitor evolutionary change over timescales that are largely irrelevant to epidemiology. Moreover, the more variable markers frequently do not reflect the pathogen's evolutionary history and, hence, provide potentially misleading information about spread. More recent work has demonstrated that staphylococcal evolution proceeds sufficiently fast that the dynamics of *S. aureus* spatial spread can be elucidated at great detail on the basis of genome-wide single-nucleotide polymorphisms.

Introduction

Staphylococcus aureus is both, a widespread commensal colonizer of the human anterior nose and a notorious pathogen causing a range of infectious diseases. Resistance against multiple classes of antibiotics is frequently observed, and the high prevalence of methicillin-resistant *S. aureus* (MRSA) in many places is particularly worrisome. Ever since the ability of *S. aureus* was demonstrated to cause outbreaks of disease and to disseminate within and among hospitals (Williams, 1959), a great interest has existed to monitor the spatial distribution and spread of this pathogen. Accordingly, considerable effort has been invested over the decades to apply and continuously refine a multitude of methods to differentiate and track peculiar strains of *S. aureus* over spatial scales ranging from localized outbreaks to global spread (Shopsin and Kreiswirth, 2001). Nevertheless, however, the extent and the preferential routes of spatial spread of *S. aureus* and the best options for curbing it are not well understood to date (Nübel et al., 2008). Recent work has demonstrated that staphylococcal evolution proceeds sufficiently fast that the dynamics of *S. aureus* spatial spread can be elucidated on the basis of DNA sequence variation, provided the sequences are long enough (Harris et al., 2010; Nübel et al., 2010). Here, we review molecular typing methods for *S. aureus* with respect to their utility for tracing the pathogen's spatial spread. We discuss present limitations and provide a hopeful outlook in the light of increasingly affordable genome sequencing.

Multilocus sequence typing

For multilocus sequence typing (MLST) of *S. aureus*, nucleotide sequences from seven metabolic housekeeping genes are determined. Strains with identical sequences at all seven genetic loci are assigned unique 'sequence types' (ST), and clusters of closely related STs are called 'clonal complexes' (Enright et al., 2002). MLST provides excellent inter-laboratory reproducibility and data portability. Sequences can be submitted to a central, curated database (available at http://saureus.mlst.net/), which in turn enables the identification of alleles and STs via queries over the internet. This database currently contains 1851 STs based on 3655 entries (as of 28th January 2011).

MLST has provided an extremely useful, basic understanding of the population structure of *S. aureus*. Used in conjunction with the characterization of SCC*mec* variability, it has revealed the evolutionary origins of major MRSA clones (Enright et al., 2002; Robinson and Enright, 2003). Even though a large number of STs has been discovered, a limited number of clonal complexes (CC1, 5, 8, 15, 22, 30, 45, 59, 80, 97, 121) appear to predominate the *S. aureus* population. These clonal complexes display a worldwide distribution, as they have been found ubiquitously through local and national surveys in many countries on every continent. Interestingly, most clonal complexes prevail among both, local populations of methicillin-susceptible *S. aureus* from non-diseased, nasal carriers (Ruimy et al., 2009, investigated samples from France, Algeria, Moldova, Cambodia; Fan et al., 2009, China; Sakwinska et al., 2009, Switzerland; Ko et al., 2008, Korea; Ruimy et al., 2008, Mali; Feil et al., 2003, England; Grundmann et al., 2002, England; Kuehnert et al., 2006, USA) and MRSA, where most surveys have been targeted

* Corresponding author. Tel.: +49 3943 679 338.
E-mail address: nuebelu@rki.de (U. Nübel).

1438-4221/$ – see front matter © 2011 Elsevier GmbH. All rights reserved.
doi:10.1016/j.ijmm.2011.09.007

at human infections (examples: Ko et al., 2005, 12 countries in Asia; Holtfreter et al., 2007, Germany; Coombs et al., 2004, Australia; Limbago et al., 2009, USA). Notable exceptions are CC15, 97, and 121, which have rarely been detected among MRSA. Commonly, these clonal complexes each are represented by single STs that are abundant, plus a number of rare variants (Feil et al., 2003; Ruimy et al., 2008; Sakwinska et al., 2009). It follows that a small number (not much larger than the number of major CCs listed above) of STs make up a large proportion of the *S. aureus* population, both locally and globally. As a consequence, the import of these widely distributed STs from external sources into a local setting cannot be identified. Further, the discriminatory power of MLST does not suffice to track spatial spread over local scales (Willems et al., 2011). Application of the recently estimated rate of (short-term) evolution for the MRSA core genome (Harris et al., 2010; Nübel et al., 2010) predicts that contemporary STs on average are many years old and that newly emerging and spreading strains will rarely be associated with novel STs. Striking differences between the frequencies of individual clonal complexes at different localities have been observed quite commonly, but it is usually impossible to infer whether these differences may reflect historic migration patterns of the pathogen or very recent expansions of local populations. The latter has been reported for a sample of methicillin-susceptible isolates from Oxfordshire, UK (Fraser et al., 2005) and for carriage populations in individual kindergartens within Chengdu city, China (Fan et al., 2009).

To summarize, MLST in most cases will be of limited use for tracking the spread of individual *S. aureus* clones, due to the insufficient discriminatory power of the method. It has to be noted, however, that most of the data supporting our conclusion represent the situation in industrialised countries, as large regions of the world remain undersampled. Exceptions have also been reported; for example, ST93 had long been considered unique to Australia, and ST93-MRSA-IV (the 'Queensland clone') was only recently discovered in the UK in association with long-distance travel. Additional examples include ST152, which has been reported from several countries in West Africa, but is rarely found elsewhere (Ghebremedhin et al., 2009; Ruimy et al., 2008), and hospital-associated ST22-MRSA, which to date is virtually non-existent in North America (Limbago et al., 2009) and got introduced into Portugal (Amorim et al., 2007) and Singapore (Hsu et al., 2005) only recently.

Also of note, the fact that the global diversity of *S. aureus* as resolved by MLST is not much greater than the diversity in any representative sample from a local population, is not reflected by the dataset currently available from the central MLST database (http://saureus.mlst.net/). Instead, rare STs are over-represented in the database, because it is mostly used for ST assignment, and strain information typically gets submitted only for novel STs. Moreover, there are only few studies where population samples have been fully typed by MLST. Instead, isolates commonly are typed by DNA macrorestriction or *spa* typing (see below), and only selected isolates representing the respective, differentiated groups subsequently get MLST-typed. As a consequence, unfortunately, the MLST database is not a representative population sample.

DNA macrorestriction

DNA macrorestriction typing of *S. aureus* isolates is based on the comparison of banding patterns resulting from pulsed-field gel electrophoresis (PFGE) of *Sma*I-digested genomic DNA (Ichiyama et al., 1991; Prevost et al., 1991). PFGE was the first molecular typing tool for MRSA that was widely applied and considerable effort has been spent to standardize assays internationally, in order to overcome limitations of interlaboratory reproducibility (Chung et al., 2000; Murchan et al., 2003). Presently, PFGE is no longer considered the gold standard for typing *S. aureus* isolates (Cookson et al., 2007). It was replaced by MLST in this regard in the new millennium, because the unambiguous nature of DNA sequences is advantageous with respect to data reproducibility, interpretation, and portability. However, PFGE continues to be the most popular method for typing *S. aureus*, particularly outside Europe. Groupings based on PFGE banding patterns in most cases are concordant with MLST at the level of clonal complexes (Cookson et al., 2007; Grundmann et al., 2002; Strommenger et al., 2006). PFGE provides higher discriminatory power than MLST, as band patterns vary considerably within MLST sequence types (e.g., by up to 5 band differences in an early MLST study (Enright et al., 2000)). This variation of patterns within CCs is not always concordant with the phylogenetic relationships among isolates (Harris et al., 2010; Nübel et al., 2008), however, which is not surprising as changes in band patterns may result from a variety of genetic events, including the gain or loss of mobile genetic elements and intrachromosomal recombination (Tenover et al., 1995). Specific MLST sequence types frequently cannot be identified on the basis of PFGE band patterns (Grundmann et al., 2002; Nübel et al., 2010).

PFGE has been applied mostly for studies of local epidemiology and has proven very useful for identifying MRSA strains during outbreaks of disease (Tenover et al., 1995). One limitation is, however, that individual strains recognized by PFGE quite commonly dominate local populations, in which case the differentiation of short-term outbreaks from endemic situations may be very difficult (Ghebremedhin et al., 2007; McDougal et al., 2003; Roberts et al., 1998; Tenover et al., 1995). At regional and national scales, PFGE frequently enabled the recognition of newly emerging MRSA clones (Deplano et al., 2000; Hookey et al., 1998; Witte et al., 1994a,b).

When MRSA isolates from international sources were compared, PFGE grouped many of them according to geographic origin (Diekema et al., 2000). Accordingly, regional groupings commonly were assigned names such as 'Japan', 'New York', 'Iberian', 'South German', etc. (Murchan et al., 2003). Remarkably, however, some identical band patterns have been found in multiple countries on several continents (Aires de Sousa et al., 1998; Diekema et al., 2000; Murchan et al., 2003), suggesting the widespread occurrence of a small number of MRSA strains (Oliveira et al., 2002). Indeed, the extremely close relatedness of full genome sequences from isolates affiliated to the 'Brazilian' MRSA clone (ST239) recently confirmed its intercontinental spread within few decades (Harris et al., 2010). In contrast, MRSA have emerged multiple times within ST5 (encompassing PFGE types 'New York', 'EMRSA-3', 'Rhein-hessen', and others). Individual MRSA clones within ST5 tend to have a more regional distribution, and PFGE analysis lumps strains that are more distantly related (Nübel et al., 2008).

Genetic lineages identified by PFGE that are prevalent among MRSA have also repeatedly been found among methicillin-sensitive *S. aureus* (MSSA), both contemporary and historic (i.e., revived from strain archives pre-dating the era of MRSA), and these susceptible strains may be equally widespread geographically (Crisóstomo et al., 2001; Hallin et al., 2007; Kuehnert et al., 2006; Witte et al., 1994a,b). As a consequence, it may be difficult to ascertain whether a particular MRSA strain has evolved locally or been imported from another region in recent times, especially as the rate at which PFGE band patterns may change over time is unknown at present.

Spa typing

Spa typing is based on sequence analyses of the 'X-region' of the *spa* (staphylococcal protein A) gene, which consists of a highly polymorphic succession of short, sequence-variable tandem repeats (Koreen et al., 2004). Isolates with different *spa* sequences