JCM
Journals.ASM.org

Bacterial Whole-Genome Sequencing Revisited: Portable, Scalable, and Standardized Analysis for Typing and Detection of Virulence and Antibiotic Resistance Genes

Shana R. Leopold,[a]* Richard V. Goering,[b] Anika Witten,[c] Dag Harmsen,[d] Alexander Mellmann[a]

Institute of Hygiene, University of Münster, Münster, Germany[a]; Department of Medical Microbiology and Immunology, Creighton University School of Medicine, Omaha, Nebraska, USA[b]; Leibniz Institute for Arteriosclerosis, University of Münster, Münster, Germany[c]; Department of Periodontology, University of Münster, Münster, Germany[d]

Multidrug-resistant nosocomial pathogens present a major burden for hospitals. Rapid cluster identification and pathogen profiling, i.e., of antibiotic resistance and virulence genes, are crucial for effective infection control. Methicillin-resistant *Staphylococcus aureus* (MRSA), in particular, is now one of the leading causes of nosocomial infections. In this study, whole-genome sequencing (WGS) was applied retrospectively to an unusual spike in MRSA cases in two intensive care units (ICUs) over the course of 4 weeks. While the epidemiological investigation concluded that there were two separate clusters, each associated with one ICU, *S. aureus* protein A gene (*spa*) typing data suggested that they belonged to single clonal cluster (all cases shared *spa* type t001). Standardized gene sets were used to extract an allele-based profile for typing and an antibiotic resistance and toxin gene profile. The WGS results produced high-resolution allelic profiles, which were used to discriminate the MRSA clusters, corroborating the epidemiological investigation and identifying previously unsuspected transmission events. The antibiotic resistance profile was in agreement with the original clinical laboratory susceptibility profile, and the toxin profile provided additional, previously unknown information. WGS coupled with allelic profiling provided a high-resolution method that can be implemented as regular screening for effective infection control.

Staphylococcus aureus is responsible for a wide range of diseases, including skin, soft tissue, and bone infections and septicemia (1). Since its emergence in the 1960s, methicillin-resistant *S. aureus* (MRSA) has become a major burden for hospitals worldwide and is now one of the leading causes of nosocomial infections (1, 2). With widespread antibiotic use, a variety of multidrug-resistant nosocomial MRSA strains have emerged in recent years (2).

MRSA can be easily transmitted in a hospital setting, from patient to patient, via staff, or from environmental contamination. Focused infection control measures guided by epidemiological investigations and genotypic fingerprinting results are necessary to prevent nosocomial transmission. Pulsed-field gel electrophoresis (PFGE) is considered the gold standard for fingerprinting to confirm a suspected outbreak, but the results are subjective and difficult to interpret. Partial *S. aureus* protein A gene (*spa*) gene sequencing (*spa* typing) emerged because of the ease of interpreting its DNA sequencing results (3, 4). Both methods, however, have sometimes been found to provide insufficient resolution under circumstances in which a limited number of predominant clones are circulating, e.g., United Kingdom clone EMRSA-15 (5) or U.S. clone USA300. This limited discriminatory power is especially problematic when tracking a clonal pathogen, such as MRSA, in a nosocomial setting in which only a few nucleotide changes, which may be missed by *spa* typing or PFGE, distinguish between genetically similar but epidemiologically unrelated strains.

Initial studies have demonstrated the high discriminatory power and information content of whole-genome sequencing (WGS) (5–9), which is now readily accessible with recent technological advances. However, the broader use of WGS in infection control is currently hampered by the lack of a universal nomenclature, which would enable a straightforward comparison with historical isolates and among different laboratories. To enable the translation of WGS into the clinical setting for infection control, we established a standardized core genome allele-based typing scheme and a resistance and virulence profiling gene set. We subsequently used this approach to epidemiologically characterize a 27-day cluster of MRSA cases that occurred mainly in two different intensive care units (ICUs) of a tertiary care hospital for which the *spa* typing and epidemiological investigation results were in conflict.

MATERIALS AND METHODS

Setting and bacterial isolates. All detected MRSA isolates from infections and surveillance cultures at the University Hospital Münster, Germany, a 1,480-bed tertiary care hospital, were prospectively *spa* typed since 2002. From 25 August to 20 September 2003, a sharp increase in *spa* type t001 (a type only rarely isolated in 2003) MRSA cases were detected, triggering a cluster investigation.

Overall, 18 isolates were included in our analyses (Table 1; see also Table S1 in the supplemental material). Thirteen *spa* type t001 isolates were collected during the cluster time frame. Additionally, five temporally unrelated t001 isolates (from patient 9 [P9], P10, P13, P14, and staff mem-

Received 27 January 2014 Returned for modification 25 February 2014
Accepted 14 April 2014
Published ahead of print 23 April 2014
Editor: Y.-W. Tang
Address correspondence to Alexander Mellmann, mellmann@uni-muenster.de.
D.H. and A.M. are co-senior authors.
* Present address: Shana R. Leopold, Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.00262-14.

doi:10.1128/JCM.00262-14
The authors have paid a fee to allow immediate free access to this article.

TABLE 1 Epidemiological data and *spa* type for MRSA isolates (all *spa* type t001) in this study

Isolate[a]	Isolation date (day-mo-yr)	Ward(s)	Source (clinical impact)[b]
P1	25-Aug-2003	ICU-B	Wound (I)
P2	29-Aug-2003	ICU-B	Wound (I)
P3	5-Sep-2003	ICU-A	Lung aspirate (C)
P4	8-Sep-2003	ICU-A	Nose (C)
S1	8-Sep-2003	ICU-A and ICU-B	Nose (C)
P5	9-Sep-2003	Ward-A	Wound (I)
P6	9-Sep-2003	ICU-A	Nose (C)
S2	10-Sep-2003	ICU-B	Nose (C)
E1	10-Sep-2003	ICU-B	Computer keyboard (NA)[c]
P7	19-Sep-2003	ICU-B	Skin (C)
P8	20-Sep-2003	Intermediate care-A	Skin (C)
P9	24-Aug-2001	Ward-B	Nose (C)
S3	19-Dec-2002	ICU-A and ICU-B	Nose (C)
P10	24-Apr-2003	Ward-C	Wound (I)
P11	25-Aug-2003	Ward-D	Wound (I)
P12	7-Sept-2003	ICU-C	Nose (C)
P13	8-Nov-2004	ICU-B	Nose (C)
P14	10-May-2006	ICU-A	Deep respiratory material (C)

[a] Isolates named by source: P, patient; S, staff member; E, environment.
[b] I, infection; C, colonization; NA, not applicable.
[c] The isolate was obtained from a computer keyboard in the staff room of the ward. Environmental sampling was initialized after the identification of an index patient (P1) and two further patients with *spa* type t001 (P2 and P3) due to suspicion of a cluster of nosocomial transmissions.

TABLE 2 Comparison of phenotypic and genotypic profiles for all MRSA t001 cluster isolates

Trait	Phenotype[a]	Gene(s)[b]	Genotype[b]
Antibiotic susceptibility			
Clindamycin	R	*ermA*, *ermC*	+, −
Erythromycin	R	*msrA*, *msrB*	+, +
Gentamicin and tobramycin	R	*aac6′-aph2″*	+
Linezolid	S	*cfr*	−
Methicillin	R	*mecA*	+
Mupirocin	S	*mupA*	−
Vancomycin	S	*vanA*	−
Toxins			
Toxic shock syndrome toxin	NA	*tst*	−
Exfoliative toxin A	NA	*eta*	+
Panton-Valentine leukocidin	NA	*lukF*, *lukS*	−, −
Species identification			
Catalase	+	*katA*	+

[a] R, resistant; S, susceptible, +, present.
[b] In the event that two genes conferred a single phenotypic trait, a comma is used to separate the genes and their respective presence or absence. −, absent; NA, not applicable.

ber 3 [S3]), collected between 2001 and 2006, were included for comparison. Analyses were performed as a proof-of-principle study of the application of WGS for molecular typing, for which ethical clearance was obtained (Ethical Committee of the University of Muenster, project 2013-302-f-S).

Phenotypic characterization and classical genotypic typing. For the identification of *S. aureus*, every isolate was tested with the API Staph ID 32 (bioMérieux, Marcy l'Étoile, France) and for the presence of free coagulase. The presence of the *mecA* gene responsible for methicillin resistance was confirmed using PCR (10). Further antibiotic susceptibility testing was carried out using the Vitek 2 automated system (bioMérieux).

spa typing was carried out in accordance with a published protocol (4), and Ridom StaphType software version 1.0 (Ridom GmbH, Münster, Germany) was used to assign *spa* types (11).

Whole-genome sequencing, assembly, and data analyses. We applied whole-genome shotgun sequencing to 18 MRSA isolates. The sequencing libraries were prepared using NextEra XT chemistry (Illumina, Inc., San Diego, CA, USA) for either a 100-bp paired-end sequencing run on an Illumina HiScanSQ sequencer or a 250-bp paired-end sequencing run on an Illumina MiSeq sequencer. The samples were sequenced to aim for a minimum coverage of 75-fold (12). After sequencing, the reads were quality trimmed using the CLC Genomics Workbench software version 6.0 (CLC bio, Aarhus, Denmark), with the following parameters: "removal of low quality sequence (limit, 0.05)" and "removal of ambiguous nucleotides: maximal 2 nucleotides allowed." Subsequently, the trimmed reads were *de novo* assembled with the CLC Genomics Workbench software, using the default settings but with a single modification ("length fraction, 0.8"). The resulting assembly files were exported as ACE files and imported into SeqSphere$^+$ software version 1.0 (Ridom GmbH).

To genotypically mirror the traditional identification, phenotypic profiling (of antibiotic resistance and virulence), and *spa* typing performed in the clinical laboratory, we defined dedicated target gene sets in SeqSphere$^+$ software. For species identification, the presence of the catalase gene (*katA*) and a partial 16S rRNA gene sequence were probed as described previously (13). To perform antibiotic resistance testing and virulence profiling as a proof of principle, we queried the WGS data for the presence of resistance and toxin-encoding genes, which were chosen by focusing on traits for which their presence or absence was known to be important for the resulting phenotype. Specifically, these comprised genes coding for clindamycin, erythromycin, gentamicin, tobramycin, linezolid, methicillin, mupirocin, and vancomycin resistance, as well as genes encoding exfoliative toxin A, Panton-Valentine leukocidin (PVL), and toxic shock syndrome toxin (TSST) (Table 2). For *spa* typing, the repeat region of the *S. aureus* protein A gene (*spa*) was extracted and analyzed.

A core genome multilocus sequence typing (cgMLST) (named MLST$^+$ within the SeqSphere$^+$ software) target set was determined using all finished *S. aureus* genomes available in GenBank (http://www.ncbi.nlm.nih.gov/GenBank/index.html) as of June 2013 (n = 40), with the exception of the taxonomic outlier *S. aureus* strain MSHR1132 (GenBank accession no. NC_016941); *S. aureus* strain COL (GenBank accession no. NC_002951) was used as a reference (see Table S2 in the supplemental material). To determine the cgMLST target gene set, a gene-by-gene comparison was performed using the MLST$^+$ target definer function of SeqSphere$^+$, with the default parameters. These parameters comprise the following filters for the reference genome (*S. aureus* COL) genes that are excluded from the cgMLST scheme: a minimum length filter that discards all genes <50 bp, a start codon filter that discards all genes that contain no start codon at the beginning of the gene, a stop codon filter that discards all genes that contain no stop codon, more than one stop codon, or if the stop codon is not at the end of the gene, a homologous gene filter that discards all genes that have fragments that occur in multiple copies within a genome (with identity 90% and more >100-bp overlap), and a gene overlap filter that discards the shorter gene from the cgMLST scheme if the affected two genes overlap >-4 bp. The remaining genes were then used in a pairwise comparison using BLAST (14) with the 39 query genomes (see Table S2 in the supplemental material). All genes of the reference genome that were common in all query genomes with a sequence identity of ≥90% and 100% overlap formed the final cgMLST scheme, consisting of 1,861 genes (see Table S3 in the supplemental material).

To validate the applicability of the *S. aureus* cgMLST target gene set, a test set of *S. aureus* isolates representing the most predominant clonal complexes (CCs) (CCs containing >10 sequence types [STs] with available raw read data) was compiled. eBURST version 3 (http://saureus.mlst.net/eburst) was used to cluster all *S. aureus* STs from the multilocus

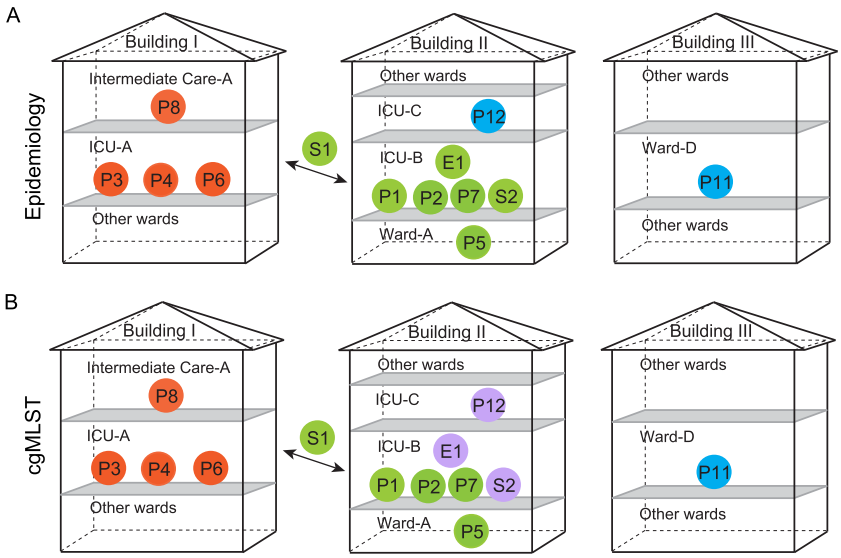

FIG 1 MRSA t001 cluster schematic. A comparison of the cluster characterization of the 13 t001 isolates collected between 25 August and 20 September 2003, based on epidemiology and on cgMLST, is shown. Each isolate is represented by a circle with the isolate name indicated (P, patient; S, staff member; E, environment), and isolates belonging to the same cluster are indicated by color. Only affected wards are shown with respect to the building in which they were located; however, as this is only a cartoon model, the true structure and geographic relationship between wards occupying the same building is not depicted here. Orange, cluster 1; green, cluster 2; blue, unrelated isolates. (A) Schematic based on epidemiological investigation. While isolate S1 was found in a staff member who was assigned to both ICU-A and ICU-B at the time of the cluster, the epidemiological data suggested that the MRSA isolate was related to ICU-B. (B) Schematic based on cgMLST data. Orange, cluster 1; green, cluster 2; purple, cluster 3; blue, unrelated isolates.

sequence typing (MLST) database into clonal complexes (CC). The European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena/) accession numbers for the WGS of a representative isolate from each CC of interest were extracted from the Bacterial Isolate Genome Sequence Database (BIGSdb) (15), and the corresponding fastq files were downloaded from the ENA. A *de novo* assembly was created with the CLC Genomics Workbench software. The SeqSphere$^+$ software was used to evaluate the presence of cgMLST targets in each isolate.

For the 18 MRSA isolates, the defined cgMLST sequences were extracted from each assembly and assessed for quality, i.e., the absence of premature stop codons and ambiguous nucleotides, and a minimum sequence coverage of ≥10-fold over the whole gene, with a minimum substitution frequency in the reads of 75%. If a gene fulfilled all of these quality criteria, its complete sequence was analyzed in comparison to that in *S. aureus* COL, and a numerical allele type was assigned by SeqSphere$^+$. The combination of all alleles in each strain formed an allelic profile, i.e., a typing result, which was used for the subsequent generation of an unweighted-pair group method using average linkages (UPGMA) tree, in which the numbers of differing alleles were given as a scale bar.

RESULTS

Epidemiology and *spa* typing. Starting on 25 August 2003, an increase in *spa* type t001 MRSA cases was detected within two intensive care units (ICU-A and ICU-B), located in different buildings, over a 27-day period. At that time, the average frequency of MRSA cases on these wards was approximately 1/month. Epidemiological investigations, screening, and environmental sampling were initialized after identifying an index patient (P1 in ICU-B) and two further patients with *spa* t001 (P2 in ICU-B, and P3 in ICU-A) due to the suspicion of a cluster of nosocomial transmissions. Eleven epidemiologically related *spa* type t001 MRSA cases were identified in connection with ICU-A, ICU-B, and their neighboring wards (intermediate care-A and ward-D, respectively) until 20 September 2003 (Table 1 and Fig. 1A). The epidemiological investigation concluded that there were two different clusters, each associated with a different ICU. Two t001 strains were identified on additional separate wards in the hospital (ICU-C and ward-D) during this cluster time frame, but the epidemiological investigation concluded that they were unrelated to any other *spa* type t001 isolate collected at the same time (Table 1 and Fig. 1A). These spatially separate isolates were included in our investigation, for a total of 13 t001 MRSA isolates collected in the hospital between 25 August and 20 September 2003. Five additional *spa* type t001 MRSA isolates that were collected at different times from our hospital (temporally separated) were also included for comparison, bringing the total to 18 t001 isolates included in all WGS analyses (Table 1).

WGS data results: all 18 t001 MRSA strains were sequenced and assembled *de novo*. To reflect the normal sequence of identification in medical microbiology diagnostics, we first extracted sequence information for species identification, antibiotic resistance, toxin gene presence, and classical genotypic typing (*spa* typing) from the 13 temporally related isolates, demonstrating as a proof of principle the broad applicability of genomic data (Fig. 2). All isolates shared the same 16S rRNA gene sequence, which was identical to those of *S. aureus* COL and other *S. aureus* strains. As an example of a typical phenotypic diagnostic test, the presence of catalase (presence of *katA*) was genotypically confirmed. All 13 isolates had the same antibiotic resistance gene composition; genes encoding clindamycin, erythromycin, gentamicin, tobramycin, and methicillin resistance were present, with an identical sequence type. These were in agreement with the susceptibility phenotype determined by the clinical laboratory at the time of the collection. In all isolates, we also detected genes encoding exfoliative toxin A but not genes encoding Panton-Valentine leukocidin or toxic shock syndrome toxin (Fig. 2, Table 2). The *spa* type extracted from the WGS data (t001) confirmed our previous results obtained by Sanger sequencing.

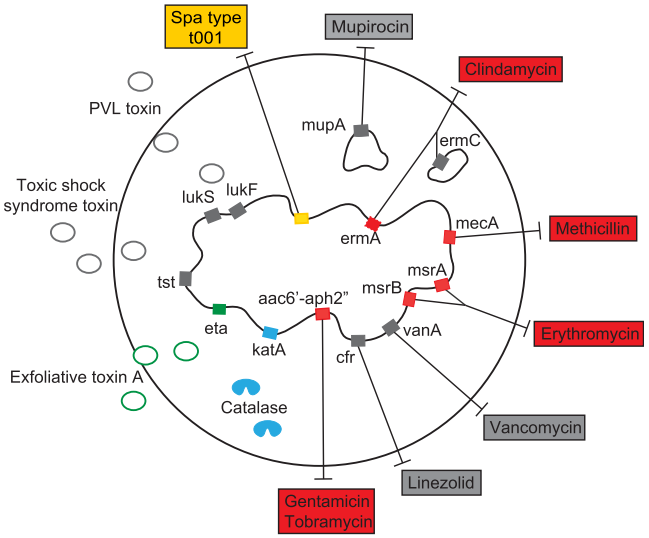

FIG 2 Genotype to phenotype diagram. Species identification, *spa* type, antibiotic susceptibility profile, and presence of toxins can be rapidly determined by query of the WGS data. The colored squares represent genes potentially present on the chromosome and/or plasmids. The presence of genes in our cluster isolates are indicated by color: red, antibiotic resistance genes; green, toxin genes; blue, catalase-encoding *katA*; yellow, f *spa* gene; gray, genes that were queried but not found.

For cgMLST schema evaluation, 16 CCs were identified containing >10 STs, with raw read data available from the European Nucleotide Archive (ENA) (see Table S4 in the supplemental material). Moreover, we added for the largest CC (CC5) eight major subgroup founders to consider the diversity of this CC. For each representative isolate, >97% of cgMLST genes were present, with a mean of 99.1% cgMLST genes present for all isolates.

The cgMLST targets were queried against the resulting whole-genome shotgun assemblies of all 18 sequenced isolates. Of the 1,861 target genes, 1,714 were present in all isolates (mean, 98.8% cgMLST targets present in each isolate) and included in further analyses. The 18 t001 isolates exhibited up to 18 differing alleles only. UPGMA dendrograms were generated to visualize the relatedness among the 13 temporally related isolates (Fig. 3A) and their relationship to the 5 temporally unrelated t001 isolates (Fig. 3B). cgMLST confirmed the results of the epidemiological investigation for the majority of cases by grouping nine of the 11 ICU-A- and ICU-B-related t001 isolates into clusters 1 and 2, with one cluster associated with each ICU (Fig. 1B). In contrast to the epidemiological data, two (environmental 1 [E1] and S1) of these 11 isolates had a different genotype, which was also unexpectedly shared with one (P12) of the two spatially different t001 isolates (from a patient on ICU-C), forming cluster 3 (Fig. 3). The second spatially different t001 isolate (P11, from ward-D) was genotypically unrelated from all other t001 isolates collected at that time (Fig. 3). Further analysis of the cgMLST allelic profiles showed that there were three and eight cluster-specific variant alleles for clusters 1 and 2, respectively, with a total of 11 allelic variants differentiating the two clusters (see Table S5 in the supplemental material). The allelic profiles within each cluster were identical (clusters 2 and 3) or nearly so (cluster 1 had two isolates [P4 and P6] with one unique allele each). Cluster 3 was identified by cgMLST only, sharing 18 variant alleles in comparison to clusters 1 and 2 (Fig. 3A; see also Table S5).

DISCUSSION

In this study, we developed an *S. aureus* core genome allele-based typing method in combination with an antibiotic resistance and toxin profile based on a standardized analysis of whole-genome sequences, thereby enabling a universally applicable comparison of recent and historical isolates. We applied this approach to an MRSA cluster event in which *spa* typing suggested a scenario of spread that was in conflict with the epidemiological data in order to elucidate the mechanisms of nosocomial transmission and demonstrate genotype-to-phenotype profiling. cgMLST resolved three distinct clusters, one each in ICU-A and ICU-B in accordance with epidemiological data, as well as a previously unsuspected transmission event, which identified a patient isolate from ICU-C as part of a third cluster within ICU-B and ICU-C (Fig. 1 and 3). Furthermore, an antibiotic resistance and toxin profile was generated solely from the *de novo* assembled contigs, providing a phenotypic profile that mirrored the tests performed by the clinical laboratory.

To determine the likelihood that these three clusters were part of a larger long-term outbreak, we examined the number of intra- and intercluster allelic differences and compared those with spatially or temporally unrelated t001 isolates. Clusters 1 and 2 were the most closely related, differing by 11 alleles, whereas cluster 3 differed by 18 alleles. No isolate differed by more than one allele level in each cluster, indicating a high level of conservation. Additionally, the epidemiological investigation revealed that the index patient of cluster 1 was already colonized with MRSA at the time of admission, which occurred >6 days after the first two isolates from cluster 2 were collected, providing further evidence that the clusters in our hospital were unrelated. Applying the estimated mutation rate for MRSA of one nucleotide change per 6 weeks (6), we calculated that the lineages leading to clusters 1 and 2 diverged approximately 36 weeks before the MRSA cluster event. Taken together, these data suggest that the close relationship between clusters 1 and 2 is likely due to the probable recent introduction of MRSA t001 to the region (4) rather than being the product of a larger ongoing outbreak within the hospital. The setting points to the fact that the resolution of WGS typing is limited to the evolution rate of the investigated pathogen, which in the case of MRSA is one point mutation per 6 weeks. This has been shown by others in larger-scale studies (5, 6), but in our short-term study, there was not sufficient passage of time to accrue informative mutations to give a temporal signal for the delineation of transmission direction. The transmission direction, however, does not play a major role in day-to-day cluster detection, as the control response is applied to all cluster cases. Most importantly, the identification of all cases belonging to a cluster and the separation of unrelated cases are the major goals of infection control.

Moreover, we investigated the feasibility of using genomic data not only for genotypic typing but also as a proof of principle for the simultaneous detection of antibiotic resistance and toxin genes (Fig. 2) (16). The genotypically derived antibiotic resistance profile was in concordance with the profile obtained at the time of the cluster by the clinical microbiology lab (Table 2). Nevertheless, the absence of genomic information coding for resistance should always be judged cautiously and might require verification with an independent method, e.g., using a specific PCR, as genes might be falsely absent due to incorrect assembly or incomplete coverage of the respective gene. We also successfully extracted the

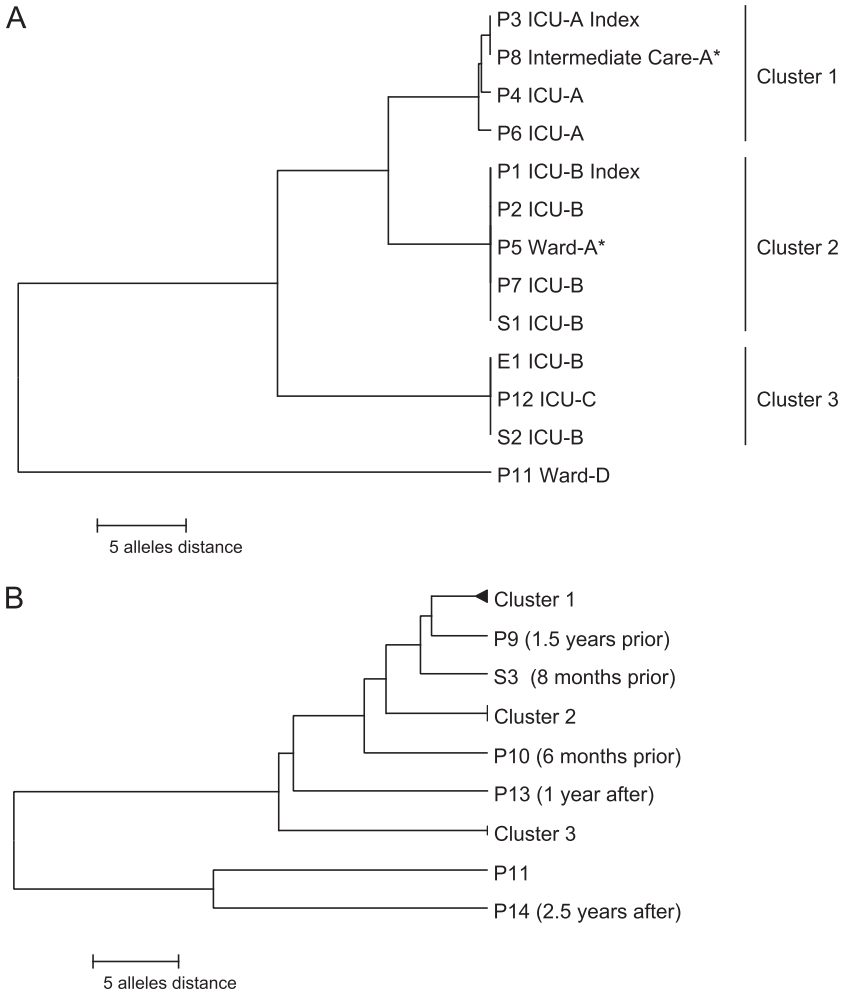

FIG 3 Clonal relationship of t001 isolates. A phylogenetic dendrogram (UPGMA) was generated for all *spa* type t001 isolates based on the allelic profiles of 1,714 cgMLST target genes. The scale bars indicate the number of differing alleles comprising the calculated distance. (A) Only temporally related isolates ($n = 13$) are shown. The ward in which each isolate was isolated is noted next to the isolate name. The index patients for clusters 1 and 2 are labeled. Asterisks indicate a ward that is next to and shares the same patients as the ICU within each respective cluster. (B) cgMLST clusters 1, 2, and 3 are collapsed, and isolates differing in time and/or location of collection are included for phylogenetic placement. The approximate time of isolation is shown relative to the August to September 2003 cluster.

spa types from the genomic data, enabling backwards compatibility. For species identification, the catalase gene and a partial 16S rRNA sequence were extracted (13).

Here, we have shown an allele-based approach for cluster detection, which is adaptable for use with any bacterial pathogen. A major advantage of such allele-based typing systems is the possibility of easily storing and curating allelic data in a central database, which is a prerequisite for ensuring a universal and expandable nomenclature. This is similar to *spa* typing, for which the central SpaServer (www.spaserver.ridom.de) hosts the nomenclature. Additionally, allele-based comparisons have an advantage over single nucleotide polymorphisms (SNP) because both SNP and single recombination events (which likely results in several nucleotide changes) are treated correctly as one evolutionary event. The very same argument for an allelic-based nomenclature was and still is a major factor in the success of classical MLST (17). Such allele-based typing has also recently been proposed by Maiden et al. (18) for WGS data. However, in contrast to the approach proposed by Maiden and colleagues (18), which utilizes an *ad hoc* shared, i.e., variable core genome gene set for only the isolates involved in the immediate study (whole-genome MLST), our approach always analyzes all genes present in the same set of species-specific core genes to facilitate standardization, which is crucial in infection control to enable comparisons, for example, those with historical isolates.

While WGS provides a high-resolution unambiguous genetic typing for cluster identification, there are limitations that still need to be addressed for the successful widespread adoption of this method for interhospital comparisons and (global) public health programs. Whereas we showed that cluster detection is achievable by WGS within a single laboratory, the interlaboratory exchange of data is currently limited due to a lack of standardized nomenclature and a central repository. However, initiatives (e.g., Global Microbial Identifier [19]) to remedy this are under way, and automated curated databases have already been established for *spa* typing (www.spaserver.ridom.de), paving the way for adaptation to WGS typing to create a publicly available central nomenclature service.

In conclusion, the use of a standardized analysis of WGS data enabled us to definitively detect and define MRSA clusters and to extract the phenotype from the genotype in a way that dramatically facilitates interlaboratory comparisons of data. While we have applied our approach to MRSA, it is in principle adaptable for use with every bacterial pathogen, e.g., we applied it prospectively during the large enterohemorrhagic *Escherichia coli* (EHEC) O104:H4 outbreak in Germany in 2011 (20). Fostered by the constantly decreasing prices of WGS (today <$150 on benchtop machines) and the availability of affordable benchtop sequencers with a rapid turnaround time from culture to analyzed sequence of 2 to 3 days (12, 21), even small- and medium-sized clinical laboratories are now in the position to implement WGS for routine testing. Future studies will focus on the delineation of species-specific thresholds for cluster identification, based upon the population structure and infection dynamics (e.g., incubation period and mode of transmission) of each species. Moreover, one future challenge will be the integration of transcriptomic and proteomic data for the full transition from genotype to phenotype.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (grant number ME3205/2-1 to A.M.), the European Community's Seventh Framework Program (grant FP7/2007-2013 to D.H. and A.M.) under grant 278864 in the framework of the European Union Patho-NGen-Trace project, and the medical faculty of the University of Münster (grant BD9817044 to A.M.).

We thank Thomas Boeking, Isabell Höfig, and Ursula Keckevoet for their skillful technical assistance, and the Core Facility of LIFA for their support.

D.H. is one of the developers of the Ridom SeqSphere$^+$ software mentioned in the article, which was a development of the company Ridom GmbH (Münster, Germany), which is partially owned by him. The other authors have declared no competing interests.

REFERENCES

1. **Lowy FD.** 1998. *Staphylococcus aureus* infections. N. Engl. J. Med. **339:** 520–532. http://dx.doi.org/10.1056/NEJM199808203390806.
2. **Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG.** 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc. Natl. Acad. Sci. U. S. A. **99:**7687–7692. http://dx.doi.org/10.1073/pnas.122108599.
3. **Frenay HM, Bunschoten AE, Schouls LM, van Leeuwen WJ, Vandenbroucke-Grauls CM, Verhoef J, Mooi FR.** 1996. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. Eur. J. Clin. Microbiol. Infect. Dis. **15:**60–64. http://dx.doi.org/10.1007/BF01586186.
4. **Mellmann A, Friedrich AW, Rosenkötter N, Rothgänger J, Karch H, Reintjes R, Harmsen D.** 2006. Automated DNA sequence-based early warning system for the detection of methicillin-resistant *Staphylococcus aureus* outbreaks. PLoS Med. **3:**e33. http://dx.doi.org/10.1371/journal.pmed.0030033.
5. **Holden MT, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, Skov R, Westh H, Zemlicková H, Coombs G, Kearns AM, Hill RL, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramírez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensen DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley SD, Nübel U.** 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. Genome Res. **23:** 653–664. http://dx.doi.org/10.1101/gr.147710.112.
6. **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.** 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science **327:** 469–474. http://dx.doi.org/10.1126/science.1182395.
7. **Nübel U, Nachtnebel M, Falkenhorst G, Benzler J, Hecht J, Kube M, Bröcker F, Moelling K, Bührer C, Gastmeier P, Piening B, Behnke M, Dehnert M, Layer F, Witte W, Eckmanns T.** 2013. MRSA transmission on a neonatal intensive care unit: epidemiological and genome-based phylogenetic analyses. PLoS One **8:**e54898. http://dx.doi.org/10.1371/journal.pone.0054898.
8. **Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CL, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TE, Walker AS, Crook DW.** 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. BMJ Open **2:**e001124. http://dx.doi.org/10.1136/bmjopen-2012-001124.
9. **Köser CU, Bryant JM, Becq J, Török ME, Ellington MJ, Marti-Renom MA, Carmichael AJ, Parkhill J, Smith GP, Peacock SJ.** 2013. Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. N. Engl. J. Med. **369:**290–292. http://dx.doi.org/10.1056/NEJMc1215305.
10. **Murakami K, Minamide W, Wada K, Nakamura E, Teraoka H, Watanabe S.** 1991. Identification of methicillin-resistant strains of staphylococci by polymerase chain reaction. J. Clin. Microbiol. **29:**2240–2244.
11. **Harmsen D, Claus H, Witte W, Rothgänger J, Claus H, Turnwald D, Vogel U.** 2003. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. J. Clin. Microbiol. **41:**5442–5448. http://dx.doi.org/10.1128/JCM.41.12.5442-5448.2003.
12. **Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J, Harmsen D.** 2013. Updating benchtop sequencing performance comparison. Nat. Biotechnol. **31:**294–296. http://dx.doi.org/10.1038/nbt.2522.
13. **Becker K, Harmsen D, Mellmann A, Meier C, Schumann P, Peters G, von Eiff C.** 2004. Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of *Staphylococcus* species. J. Clin. Microbiol. **42:**4988–4995. http://dx.doi.org/10.1128/JCM.42.11.4988-4995.2004.
14. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
15. **Jolley KA, Maiden MC.** 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics **11:**595. http://dx.doi.org/10.1186/1471-2105-11-595.
16. **Priest NK, Rudkin JK, Feil EJ, van den Elsen JM, Cheung A, Peacock SJ, Laabei M, Lucks DA, Recker M, Massey RC.** 2012. From genotype to phenotype: can systems biology be used to predict *Staphylococcus aureus* virulence? Nat. Rev. Microbiol. **10:**791–797. http://dx.doi.org/10.1038/nrmicro2880.
17. **Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.** 1998. 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. http://dx.doi.org/10.1073/pnas.95.6.3140.
18. **Maiden MC, van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND.** 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat. Rev. Microbiol. **11:**728–736. http://dx.doi.org/10.1038/nrmicro3093.
19. **Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, Koopmans M, Kroneman A, Lo Fo Wong D, Lund O, Palm D, Sawanpanyalert P, Sobel J, Schlundt J.** 2012. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerg. Infect. Dis. **18:**e1. http://dx.doi.org/10.3201/eid/1811.120453.
20. **Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H.** 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One **6:**e22751. http://dx.doi.org/10.1371/journal.pone.0022751.
21. **Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ.** 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. **30:**434–439. http://dx.doi.org/10.1038/nbt.2198.