International Journal of Medical Microbiology 308 (2018) 558–568

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

*Aureo*Wiki– The repository of the *Staphylococcus aureus* research and annotation community

Stephan Fuchs[a,c,1], Henry Mehlan[b,1], Jörg Bernhardt[c], André Hennig[d], Stephan Michalik[b], Kristin Surmann[b], Jan Pané-Farré[c], Anne Giese[c], Stefan Weiss[b], Linus Backert[d], Alexander Herbig[d,2], Kay Nieselt[d], Michael Hecker[c,e], Uwe Völker[b,e], Ulrike Mäder[b,*]

[a] *FG13 Nosocomial Pathogens and Antibiotic Resistance, Robert Koch Institute, Wernigerode, Germany*
[b] *Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany*
[c] *Institute for Microbiology, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany*
[d] *Center for Bioinformatics Tübingen, University of Tübingen, Tübingen, Germany*
[e] *ZIK FunGene, Ernst-Moritz-Arndt-University Greifswald and University Medicine Greifswald, Greifswald, Germany*

ARTICLE INFO

Keywords:
S. aureus
Database
Gene and protein information
Pan-genome
Unified gene names

ABSTRACT

In light of continuously accumulating data and knowledge on major human pathogens, comprehensive and up-to-date sources of easily accessible information are urgently required. The *Aureo*Wiki database (http://aureowiki.med.uni-greifswald.de) provides detailed information on the genes and proteins of clinically and experimentally relevant *S. aureus* strains, currently covering NCTC 8325, COL, Newman, USA300_FPR3757, and N315. By implementing a pan-genome approach, *Aureo*Wiki facilitates the transfer of knowledge gained in studies with different *S. aureus* strains, thus supporting functional annotation and better understanding of this organism. All data related to a given gene or gene product is compiled on a strain-specific gene page. The gene pages contain sequence-based information complemented by data on, for example, protein function and localization, transcriptional regulation, and gene expression. The information provided is connected via links to other databases and published literature. Importantly, orthologous genes of the individual strains, which are linked by a pan-genome gene identifier and a unified gene name, are presented side by side using strain-specific tabs. The respective pan-genome gene page contains an orthologue table for 32 *S. aureus* strains, a multiple-strain genome viewer, a protein sequence alignment as well as other comparative information. The data collected in *Aureo*Wiki is also accessible through various download options in order to support bioinformatics applications. In addition, based on two large-scale gene expression data sets, *Aureo*Wiki provides graphical representations of condition-dependent mRNA levels and protein profiles under various laboratory and infection-related conditions.

1. Introduction

The major human pathogen *Staphylococcus aureus* causes infections that range from superficial skin infections to life-threatening diseases (Lowy, 1998). This Gram-positive bacterium is also a common component of skin and mucosal flora, colonizing about 20% of the human population (van Belkum et al., 2009). A growing problem is the emergence of antibiotic-resistant strains, such as methicillin-resistant *S. aureus* (MRSA) (de Kraker et al., 2011). Thus, *S. aureus* is not only the most common causative agent of nosocomial infections, but also a leading cause of death in hospitalized patients (Otto, 2013). However, since the late 1990s MRSA infections have been reported in healthy individuals without connection to health care institutions (Chambers, 2001), mainly with the highly virulent USA300, the predominant clone of community-associated (CA)-MRSA in the United States. The development of CA-MRSA strains is associated with the acquisition of various mobile genetic elements (MGEs), in particular phage phiSA2 carrying the Panton-Valentine leucocidin (PVL) genes (Vandenesch et al., 2003).

The pan-genome of a species consists of the core genome, the dispensable genome comprising genes present in a subset of strains, and genes unique to single strains (Medini et al., 2005). For *S. aureus*, comparative genomics revealed high genetic diversity and a clonal

* Corresponding author.
E-mail address: ulrike.maeder@uni-greifswald.de (U. Mäder).
[1] Contributed equally.
[2] Present address: Max Planck Institute for the Science of Human History, Jena, Germany.

https://doi.org/10.1016/j.ijmm.2017.11.011
Received 8 September 2017; Received in revised form 20 November 2017; Accepted 24 November 2017

structure of the species. More than ten years ago, Lindsay et al. (2006) reported that each clonal lineage carries a specific combination of a large number of "core variable" genes. Based on these observations, *S. aureus* genes were classified into a core genome (approx. 75% of the genes), a core variable genome (approx. 10%), and MGEs (approx. 15%) (reviewed in Lindsay, 2010; Lindsay, 2014). The highly conserved core genome comprises genes associated with essential metabolic and other housekeeping functions. The core variable genome includes mainly genes encoding surface proteins and virulence regulators such as the accessory gene regulator (Agr) and differs between lineages in variant regions within genes and the presence/absence of genes or gene clusters. MGEs such as phages, pathogenicity islands, and the staphylococcal cassette chromosomes (SCC) encode virulence factors and proteins required for antibiotic resistance. In addition, strains can significantly differ in the expression of virulence factor genes. For example, expression of cytolytic toxins, including α-toxin and phenol-soluble modulins (PSMs), whose genes are present in all sequenced *S. aureus* strains, is higher in CA-MRSA compared to hospital-associated (HA)-MRSA (Wang et al., 2007).

Whole-genome sequencing data constitute an essential foundation of today's research. However, in-depth functional annotation requires the integration of genomic data with existing knowledge and experimental data, in particular from large-scale functional genomics studies. Easy access to all this information is essential to improve our understanding of cellular physiology. For the model organism of Gram-positive bacteria, *Bacillus subtilis*, we developed *Subti*Wiki (Lammers et al., 2010; Mäder et al., 2012) to collect all available information on this bacterium. Besides comprehensive information on the individual genes and proteins of *B. subtilis*, it provides presentations of metabolic and regulatory pathways, manually curated protein-protein interaction diagrams as well as gene expression data from a large-scale transcriptome study and from absolute quantification of cytoplasmic proteins (Michna et al., 2014). *Subti*Wiki has developed into an integrated database with different forms of presentations and interactive tools, thus enabling novel insights about gene or protein functions and interactions. This database has become one of the most complete collections of knowledge on a living organism in one single resource (Michna et al., 2016). A recently developed resource dedicated to the closely related genus *Listeria* focuses on omics data on the pathogen *L. monocytogenes* and other *Listeria* species (Bécavin et al., 2017). It integrates all published *Listeria* genome, transcriptome, and proteome data sets and provides tools for interactive visualization and exploration of this data.

Pathogens are among the bacteria attracting intensive research and subject to large genome sequencing projects. For example, with more than 100 completely sequenced strains, *S. aureus* belongs to the group of organisms for which the largest number of genome sequences is available. Due to its considerable importance, a huge amount of knowledge and experimental data on *S. aureus* has been gained. Consequently, several databases cover information on *S. aureus*, which are often part of species-independent efforts (e.g. MicrobesOnline, RegPrecise, PubMLST, DEG, and VFDB) covering specific research areas such as transcriptional regulation, multilocus sequence typing, essential genes, or virulence factors of bacterial pathogens (Enright and Spratt, 1999; Novichkov et al., 2013; Luo et al., 2014; Chen et al., 2016). In addition, several *S. aureus* databases exist that are dedicated to specific datasets, including transcriptome data (SATMD – *Staphylococcus aureus* transcriptome meta-database), stress-related proteome signatures (*Aureolib*), and regulatory RNAs (SRD – Staphylococcal regulatory RNA database) (Nagarajan and Elasri, 2007; Fuchs et al., 2013; Sassi et al., 2015). Easy access to the different data sources and the available knowledge about an organism is pivotal for the interpretation of experimental data, in particular from genome-scale experiments. Therefore, a comprehensive and clearly presented collection of genomic data together with the different aspects of functional annotation is required that interconnects information from scattered sources.

In this report we present *Aureo*Wiki, a manually curated database providing detailed information – supplemented by links to external data sources and relevant publications – on the genes and proteins of clinically and experimentally relevant *S. aureus* strains, currently covering NCTC 8325, COL, Newman, USA300_FPR3757, and N315. *Aureo*Wiki was developed to support the analysis and interpretation of experimental data, in particular from genome-scale studies, with the overarching goal of supporting research to better understand the molecular mechanisms of *S. aureus* pathogenicity and virulence factor production. *S. aureus* research is conducted with different strains including numerous clinical isolates, with a number of well-characterized strains being used in the majority of studies. For example, the NCTC 8325 lineage includes prototype strains for basic research on gene regulation and physiology (Herbert et al., 2010), whereas strains such as Newman or USA300_FPR3757 are often used to study staphylococcal virulence and pathogenesis. Hence, *Aureo*Wiki was designed to combine gene/gene product information of individual strains based on the *S. aureus* pan-genome. Orthologous genes of the individual strains are linked by a common identifier and species-wide unified gene names that are in line with the scientific literature. Use of consistent terminology of gene identifiers, gene names and protein functions helps researchers to transfer knowledge gained in studies with different strains and supports functional annotation of *S. aureus*.

2. Methods

2.1. Computation of the S. aureus pan-genome

The *S. aureus* pan-genome was computed as described previously (Herbig et al., 2012; Hennig et al., 2015). In brief, RefSeq files of 32 *S. aureus* strains sequenced and fully annotated in 2012 were obtained from the NCBI FTP site. For the computation of the pan-genome, first a global DNA alignment of the 32 genomes was performed using progressiveMauve (Darling et al., 2010). Based on this alignment, a system of common gene coordinates was set up, called the SuperGenome (Herbig et al., 2012). Subsequently, orthologous gene groups were extracted from genes that overlap in the coordinate system of the SuperGenome and iteratively refined by using gene descriptions, functional classification (TIGRFAMs) and further analysis of sequence similarity on amino acid level.

2.2. Database implementation

*Aureo*Wiki is based on a MediaWiki engine with an adapted layout. The data and information collected about *S. aureus* is stored in a database and embedded in *Aureo*Wiki *via* user-defined tags. The database enables rapid updates in case of changes of external data or addition of new types of data to the *Aureo*Wiki. Users can add comments and additional content to each Wiki page through a WYSIWYG editor or through a standard Wiki editor.

2.3. Protein function assignment and protein localization

Functional assignments have been generated as follows: (i) For enzymes the catalytic activity is provided by the EC number (extracted from NCBI RefSeq and UniProt databases), complemented by the corresponding enzyme name and reaction equation (extracted from ExPASy at http://enzyme.expasy.org/). (ii) The assignment of protein sequences to TIGRFAMs protein families (Haft et al., 2013) is based on TIGRFAM Hidden Markov Models (HMM) using hmmscan of the HMMER3 software package (Finn et al., 2011). The display of TIGRFAMs is ordered according to their HMM scores as a significance measure of the assignment and possesses a tree like structure including the TIGR role categories (main role and sub role) and an added meta level summarizing the TIGR main roles. The color code of the genome viewer is based on these meta roles (orange brown – Metabolism, blue

shades – Genetic information processing, green shades – Signal Processing, pink – Cell Wall and Envelope, red – Cellular Processes, black – RNA Genes, grey – Hypothetical and Proteins with unknown function). (iii) As described for TIGRFAMs, the assignment of sequences to Pfam protein families (Finn et al., 2016) is based on HMMs and uses the HMMER package. Pfams with the highest HMM scores are shown first. A large part of Pfams is grouped into clans (evolutionary related families), which are displayed on top of the Pfam annotation. (iv) Assignment of predicted protein functions is obtained from the SEED database (Overbeek et al., 2005), a comparative genomics database based on expert annotation of subsystems (sets of related functional roles).

The data on predicted protein subcellular localization is calculated or downloaded using the following online tools: PSORTb (http://psort.org/psortb/index.html), LocateP (http://www.cmbi.ru.nl/locatep-db/locatepdb/0.129849.txt), SignalP (http://www.cbs.dtu.dk/services/SignalP/), and TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/). If prediction by PSORTb yields equal scores for all four localizations (cytoplasmic, cytoplasmic membrane, cell wall, extracellular), the PSORTb-based localization is summarized by the term “unknown (no significant prediction)”.

3. Results and discussion

3.1. The key feature of AureoWiki: the pan-genome approach

The basic concept of *Aureo*Wiki is to combine gene/gene product information of individual *S. aureus* strains by an interlinked presentation of orthologous genes. As a prerequisite, the *S. aureus* pan-genome based on 32 strains fully annotated in 2012 was constructed as outlined in the Methods section. The pan-genome comprises 6471 genes, of which 2115 genes are present in at least 31 strains and 2032 genes in only one of the analyzed strains. However, in particular the number of the latter group (orphans) might be overestimated and would most probably decrease by manual curation of the pan-genome computation (Hennig et al., 2015).

Orthologous genes are linked by a common identifier, the *pan locus tag*, and designated with a unified gene name, the *pan gene symbol*, which results from a manual curation effort as described below. For a given *S. aureus* gene, gene pages for the individual strains (i.e. strains for which detailed information is currently available, see next section) and the corresponding pan-genome page are presented side by side using tabs (Fig. 1), which allows the user to easily switch between pages of orthologous genes. In addition, the locus tags of the orthologous genes in all 32 strains are displayed on the pan-genome pages and are available for download.

Importantly, the interlinked presentation of orthologous genes facilitates the comparison of results obtained in studies with different strains and the transfer of knowledge between *S. aureus* strains. These aspects are specifically supported in two ways: First, data that are to some extent strain-specific, such as those on gene regulation and gene expression patterns, are presented on the respective gene page with direct accessibility from the pages of all orthologous genes (Fig. 2). Assignment of genes to regulation by transcription factors or the alternative RNA polymerase sigma factor SigB was conducted in numerous genome-wide studies involving different strains, including four transcriptome studies mapping the SigB regulon using strains COL, Newman, HG001 and GP268, the latter two being derivatives of NCTC 8325 (Bischoff et al., 2004; Pané-Farré et al., 2006; Schulthess et al., 2011; Mäder et al., 2016). As shown in the upper part of Fig. 2, mutually supportive results are combined by adding the text “other strains” to the sigma factor information, which can be activated by mouse-over to show the locus tags of the orthologous genes identified as SigB targets in other *S. aureus* strains and allows direct access to the corresponding gene pages.

Second, the interlinked view provides access to functional information about proteins with poorly characterized or unknown function by (i) displaying all (putative) functions available in the RefSeq annotations of the 32 strains on the pan-genome page and (ii) presenting results from studies that newly assigned protein functions and gene symbols. Such information cannot be found in the NCBI RefSeq database or other databases automatically retrieving RefSeq data. For example, components and a major substrate, the nuclease toxin EsaD, of the type VII secretion system were newly identified in two recent studies (Anderson et al., 2011; Cao et al., 2016). Whereas the products of the respective genes are annotated as hypothetical proteins and DUF5081 domain-containing protein, respectively, in all *S. aureus* strains, the *Aureo*Wiki pan genome symbols reflect the newly assigned gene names and the relevant publications are shown on the gene pages.

3.2. Strain-specific pages for the individual genes and proteins

*Aureo*Wiki is centered on the individual genes and gene products of clinically and experimentally relevant *S. aureus* strains. Strain-specific pages for each gene provide a wide range of genome-based information complemented by specific data including, for instance, gene function and regulation or condition-dependent gene expression, together with links to external data sources such as various databases and published literature. Currently, detailed information is provided for five model strains: NCTC 8325, COL, Newman, USA300_FPR3757, and N315.

These strains have been widely used for studies in laboratory settings, and their genomes were among the first *S. aureus* genomes to be sequenced. In 2001, the first *S. aureus* genome sequences, namely of MRSA strains N315 and Mu50, became available (Kuroda et al., 2001). Seven years later, with the completion of the sequence of strain Newman, 12 *S. aureus* genomes were completely sequenced (Baba et al., 2008). Strains COL, N315, and USA300_FPR3757 are MRSA strains, of which COL is one of the earliest HA-MRSA isolates from the 1960s and USA300_FPR3757 is the CA-MRSA prototype isolated in 2002. NCTC 8325 and Newman, human clinical isolates from 1943/1952, are frequently used model strains for basic research and in animal infection models, respectively. With respect to clonal relationship, strains Newman, COL, NCTC 8325, and USA300_FPR3757 belong to the same clonal complex (CC8) (Enright et al., 2000), whereas N315 is more distantly related. It belongs to CC5, which is the predominant lineage of HA-MRSA in the United States and of vancomycin-resistant *S. aureus* (VRSA) (Kos et al., 2012).

An example of an *Aureo*Wiki gene page is shown in Fig. 1. The *Summary* section at the top of each page contains the locus tag, the gene name and function of the gene product from RefSeq annotation as well as the pan locus tag and the pan gene symbol. In the following *Genome View*, condensed genome information is provided, initially aligned to the position of the respective gene. The genome position in the genome browser can be changed by dragging the slider. By clicking on gene arrows, the user is transferred to the corresponding gene page, thus enabling a page by page walking through the genome. Colors correspond to the gene functional categories as described in the Methods section. Finally, the genome browser combines for each strain the well-established RefSeq annotation and the new RefSeq annotation introduced in 2015 (Tatusova et al., 2015). This view, by allowing a direct comparison, facilitates the transition to the new annotations and locus tags (Fig. 3).

The next section contains the information about the gene (*Gene* section, see Fig. 1). It covers basic information as in the *Summary* section, complemented by the gene coordinates, gene length, essentiality, DNA sequence, and gene-specific database entries. The largest section of the page, the *Protein* section, is devoted to the encoded protein. Here, the protein length, the molecular weight and isoelectric point, catalyzed reaction, protein function assignments based on TIGRFAMs, Pfam, and The SEED (see Methods section), protein interaction partners, subcellular localization as well as other information are shown. As for the

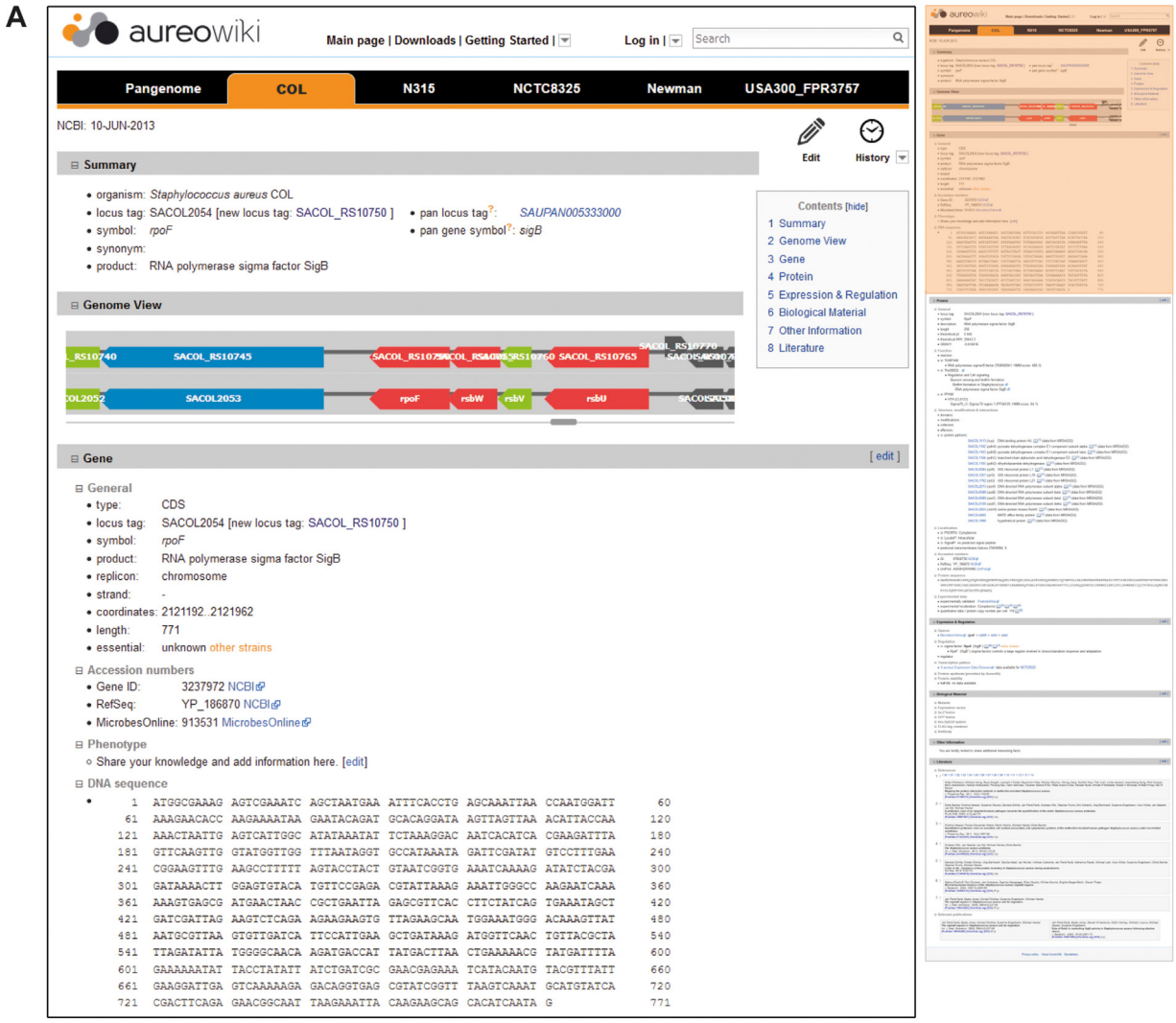

Fig. 1. Example of a strain-specific gene page of *Aureo*Wiki. Each gene page contains detailed information about the selected gene **(A)** and its protein product **(B)** as well as gene expression and regulation, supplemented by links to other resources and relevant publications **(C)**. **(A)** The tabs on top of the page allow switching to the pages of the orthologous genes of other *S. aureus* model strains and to the corresponding pan-genome page. The page starts with the *Summary* section providing general gene information such as locus tag, gene name, function of the gene product based on NCBI-RefSeq annotation as well as the pan locus tag and the pan gene symbol which is followed by the strain-specific genome browser and the *Gene* section. **(B)** The *Protein* section contains detailed information about the encoded protein including protein function assignments (based on TIGRFAMs, Pfam, and The SEED), predicted subcellular localization, and experimental data. **(C)** The following sections of the gene page are devoted to gene expression and its regulation, additional information and relevant literature.

gene section, the protein section is concluded with database links (NCBI Protein database and UniProt). The function of up to one third of the proteins even of well characterized microorganisms is still unknown. Therefore, function predictions based on complementary classification algorithms (TIGRFAMs, Pfam, and The SEED) independent of the RefSeq annotation pipeline, which can support the elucidation of protein functions, are integrated in the *Aureo*Wiki pages as part of the protein related information. For the sake of conciseness, by default the lists of assigned predicted functions are collapsed and show only the hit with the highest score, but can be expanded by clicking the plus sign. The last part of the *Protein* section contains experimental data including protein localization (Becher et al., 2009; Dreisbach et al., 2010; Hempel et al., 2011) and absolute quantification of cytoplasmic proteins (Zühlke et al., 2016).

The following section of the gene page provides information related to regulation and gene expression, including the predicted operon structure obtained from MicrobesOnline (Dehal et al., 2010), transcriptional regulation by alternative RNA polymerase sigma factors and transcription factors as well as gene expression profiles. Importantly, besides the protein function assignments included in the *Protein* section, these types of information can help to uncover the physiological role of a protein with poorly characterized or unknown function. Data on transcription factor regulons was retrieved from the RegPrecise database (Novichkov et al., 2013). Target genes of SigB and SigH were extracted from published literature. Gene expression data from two large-scale studies (Fuchs et al., 2013; Mäder et al., 2016) covering a wide range of growth, stress and infection-related conditions were included in *Aureo*Wiki by graphical representations of condition-dependent mRNA levels and protein induction profiles on the gene pages of strains NCTC 8325 and COL, respectively (see Fig. 2). Online resources related to these studies, which provide various search and clustering options, are directly accessible from the gene pages, thus integrating research tools that extend beyond the scope of *Aureo*Wiki.

All data are provided with links to the external data sources, including various databases and published literature. References are indicated by the book symbol. Details of the corresponding publication are displayed by mouse-over. A list of references is found at the bottom of the page.

Information preceded by a filled bullet point is obtained from the *Aureo*Wiki database and cannot be changed by the user. In particular,

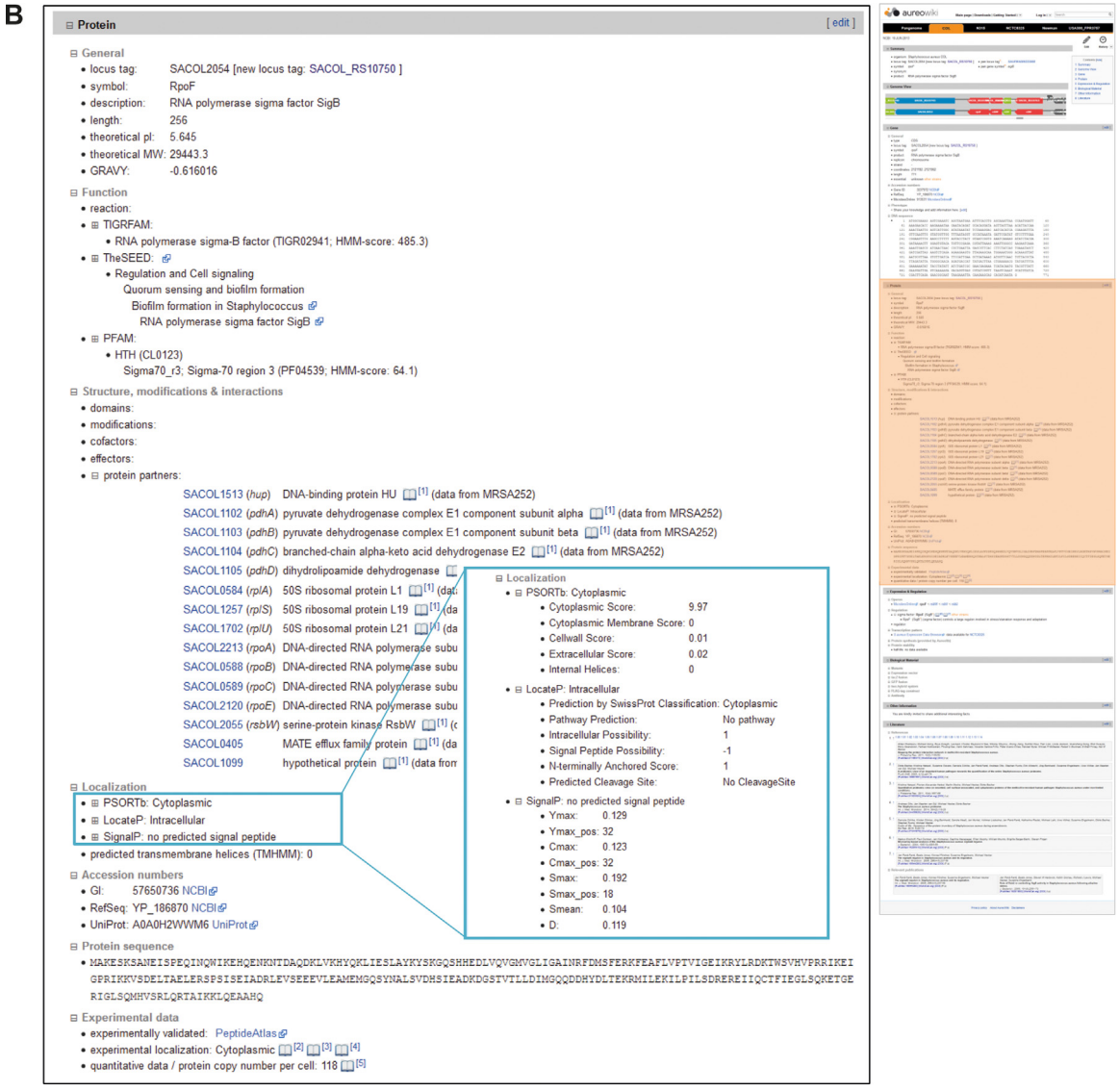

Fig. 1. (*continued*)

all sequence based information needs to be consistently maintained in accordance with the RefSeq annotations or manual annotation updates in the case of *S. aureus* NCTC 8325 (see below). The same holds true for links to other databases. Users can add comments and additional content such as experimental data or references using the [Edit] link. Specifically, placeholders were inserted to ask for user input in order to complement the already available information about the respective gene or protein, for example regarding the phenotypes of gene knockout or overexpression mutants. User added information is indicated by open bullet points.

Because of the variation in the occurrence and expression of virulence genes, the group of strains commonly used to study staphylococcal pathophysiology is clearly not restricted to the five model strains covered so far and includes, for example, UAMS-1 and 6850 (Herbert et al., 2010; Fraunholz et al., 2013). In future efforts to increase the number of strains covered by *Aureo*Wiki, content of gene pages can be imported using already developed scripts.

3.3. The pan-genome pages

Each *S. aureus* gene present in at least one of five model strains (NCTC 8325, COL, Newman, USA300_FPR3757, and N315) is represented by strain-specific gene pages and a corresponding pan-genome page, which was established to provide comparative information. Each pan-genome page (Fig. 4) starts with a *Summary* section containing the pan locus tag, the pan gene symbol, (putative) functions of the encoded protein extracted from the RefSeq annotations of all 32 strains covered by our current pan-genome, the multiple genome alignment coordinates (for the SuperGenome concept see Herbig et al., 2012), and the gene occurrence frequency expressed as percentage of 32 strains. In the following *Orthologs* section, all 32 *S. aureus* strains are listed in a fixed order and, if the gene is present in the respective strain, the locus tag is shown together with the strain-specific gene name, if assigned, from the RefSeq annotation. The next two sections, the *Genome Viewer* and the *Alignments*, refer to the five *S. aureus* strains for which strain-specific information is provided in *Aureo*Wiki. The multiple-strain genome viewer has the same functionality and uses the same color scheme based on functional categories as its strain-specific counterparts. Finally, a protein sequence alignment generated by the MAFFT program (Katoh and Standley, 2013) is provided, which can be displayed using different color schemes such as coloring of the amino acids based on their chemical properties.

3.4. Additional gene pages for re-annotated RefSeq genomes

As outlined before, all bacterial RefSeq genomes were re-annotated

Fig. 1. (*continued*)

using an improved prokaryotic genome annotation pipeline introduced by the NCBI RefSeq project in 2015 (Tatusova et al., 2015). Accordingly, for all *S. aureus* strains, except for NCTC 8325, the well-established gene identifiers (locus tags) used in most of the scientific literature and in the various databases relevant for *S. aureus* research are no longer supported by the NCBI resources. In the new genome annotations, protein coding genes are linked to so-called non-redundant protein sequences (designated with prefix WP_). These are annotated to orthologous genes of different genomes if the encoded proteins possess identical protein sequences. The new pipeline is used to annotate all prokaryotic RefSeq genomes with the exception of a small number of so-called reference genomes, one of which is *S. aureus* NCTC 8325. For all other strains, new locus tags were assigned.

Importantly, *Aureo*Wiki provides gene pages for the genes of the original annotations as well as all newly annotated genes for *S. aureus* strains COL, Newman, USA300_FPR3757, and N315. If the locus in the new annotation is a replacement of the original gene, the user can directly switch between the corresponding gene pages by clicking on the second locus tag found in the *Summary* and *Gene* sections (see Fig. 1A). Linking of new and old locus tags can be retrieved from NCBI RefSeq complete genome records where gene annotations provide an "old_locus_tag" qualifier along with the new "locus tag". When searching the NCBI databases with the old locus tags or Gene IDs, the Gene records discontinued in 2015 can still be accessed, but the information on locus tag replacement and assigned non-redundant protein sequence has not been updated according to RefSeq annotation versions released after 2015.

In addition, the strain-specific *Aureo*Wiki genome viewer (Fig. 3) was set up to combine old and new RefSeq annotations, because the direct comparison best supports comprehension of the differences in gene content and/or coordinates. The combined genome viewer is particularly helpful to recognize the assignment in genomic regions where genes in the new annotation are not a direct replacement of the original genes.

3.5. Manual annotation updates of the S. aureus NCTC 8325 genome

The genome of *S. aureus* NCTC 8325 (Gillaspy et al., 2006) belongs to the RefSeq reference genomes and was therefore not subject to the re-annotation effort described before. Nevertheless, periodic updates have

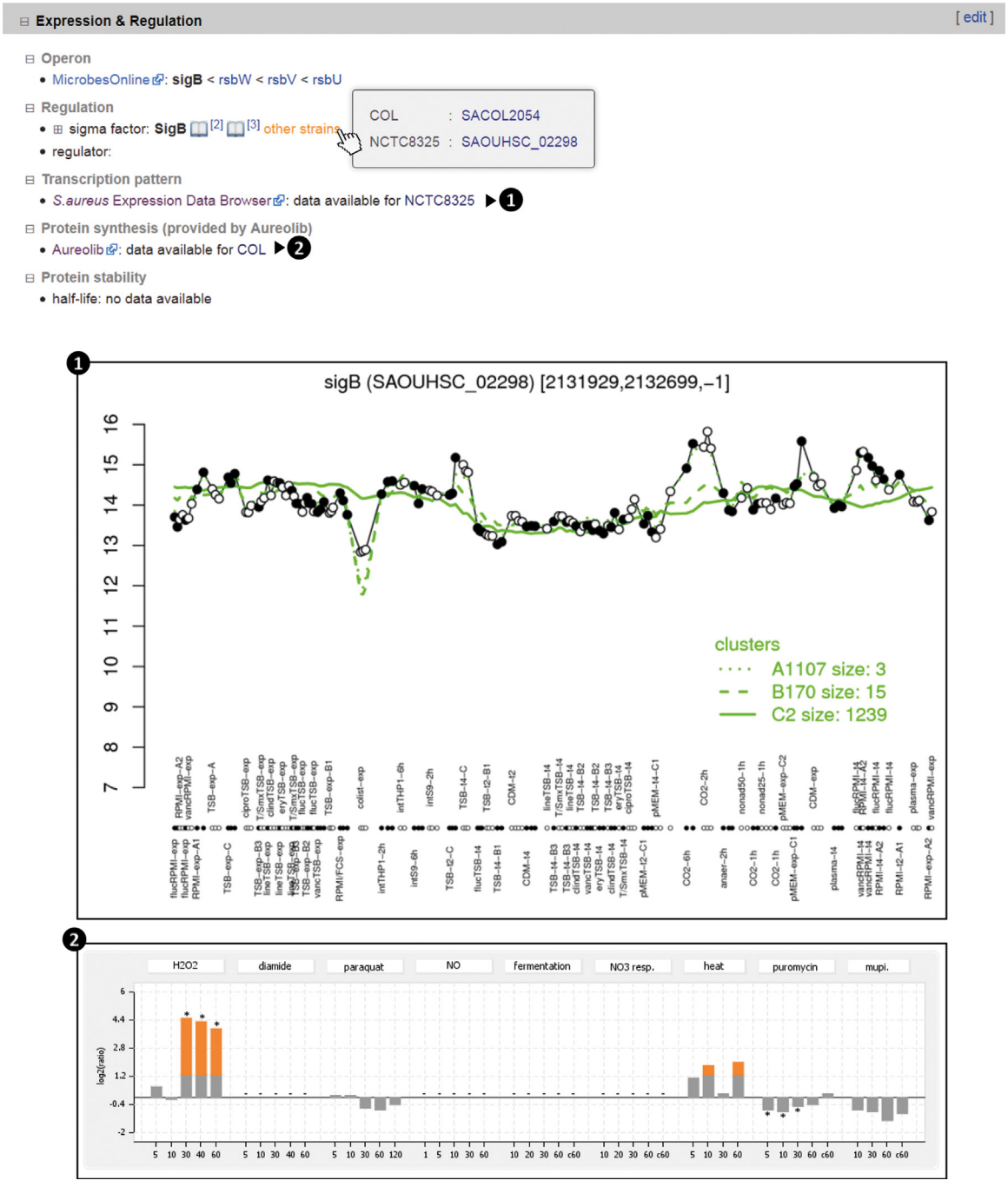

Fig. 2. Presentation of gene expression patterns and data on transcriptional regulation combines information from different *S. aureus* strains based on the pan-genome approach. The example shows a sub-section of the page for the *sigB* gene of *S. aureus* Newman. Gene-specific expression levels for strain NCTC 8325 (*S. aureus* Expression Data Browser) and protein induction profiles for strain COL (Aureolib) obtained under various laboratory and infection-related conditions, are accessible from the corresponding gene pages of the other strains (numbered bullet points). As can be seen in the upper part of the figure, supportive regulatory information is indicated by the field “other strains”, which upon activation by mouse-over shows the corresponding locus tags and allows direct access to their gene pages.

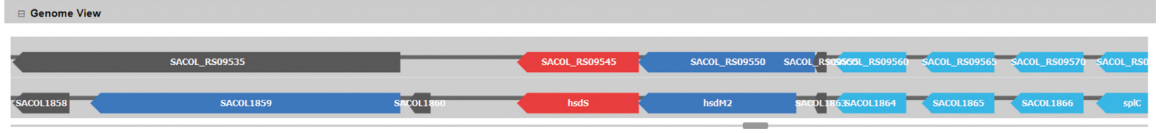

Fig. 3. The strain-specific genome browser of *Aureo*Wiki provides a direct comparison of the well-established RefSeq annotation and the new annotation introduced in 2015 for *S. aureus* strains COL, N315, Newman, and USA300_FPR3757. On the gene pages, the displayed genome section is initially centered on the respective gene, but can be changed by dragging the slider. Colors correspond to functional categories of the encoded proteins. In the selected genomic region of *S. aureus* COL, one previously annotated gene (SACOL1860) was discarded, two genes (SACOL1858 and SACOL1859) were merged into a single locus (SACOL_RS09535) and for one gene (SACOL1862, hsdM2) the start position was changed.

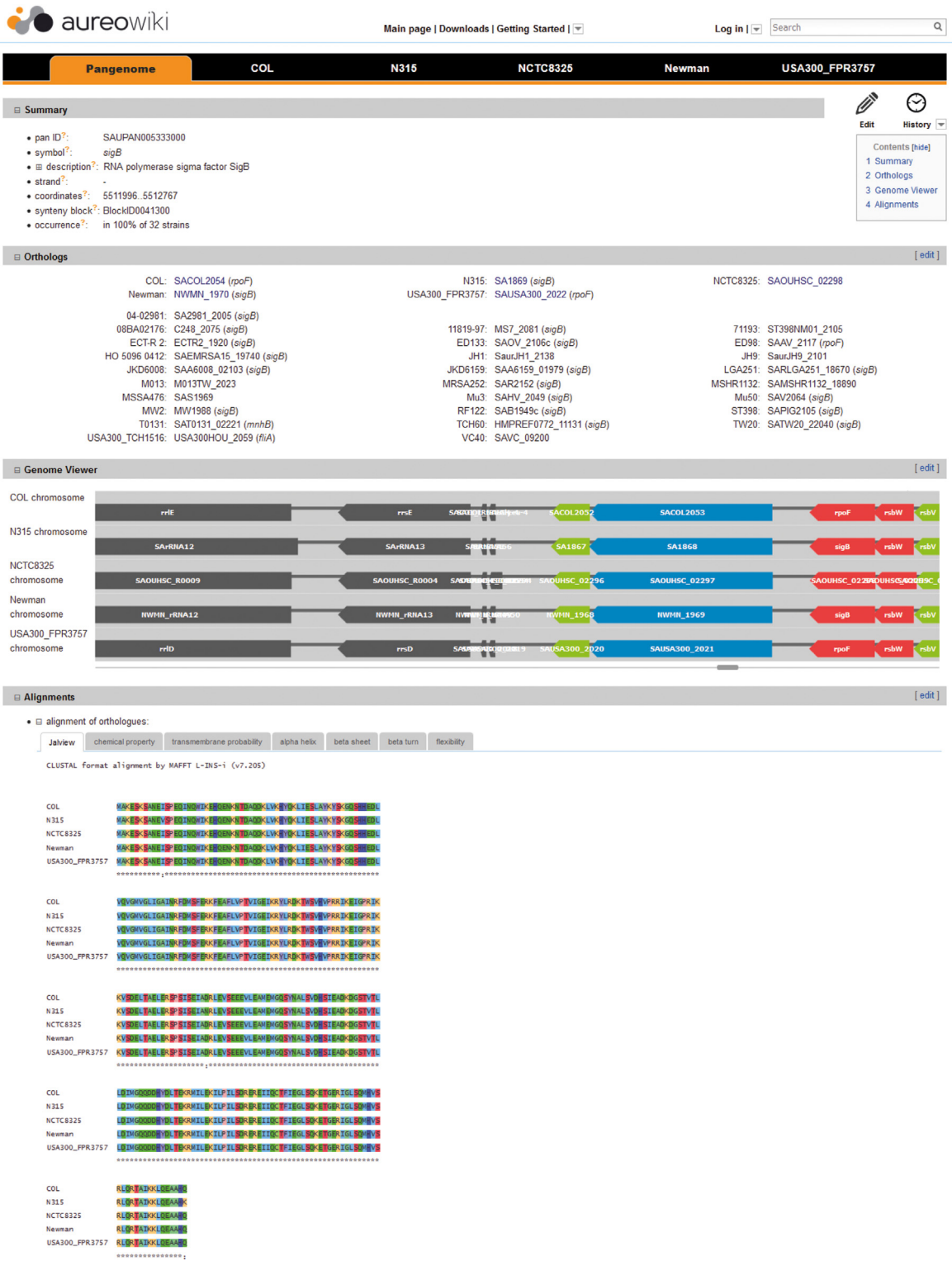

Fig. 4. Example of a pan-genome page of *Aureo*Wiki. The pan-genome pages contain combined information for the respective group of orthologous genes of 32 *S. aureus* strains including the pan-genome identifier (pan ID), the species-wide unified gene name (symbol), the orthologue table, the multiple-strain genome browser and a protein sequence alignment. Of note, orthologous genes were frequently named differently in genome annotations of individual *S. aureus* strains. On the pan-genome pages, these strain-specific names from the NCBI-RefSeq annotations are displayed in parentheses together with the respective locus tag. Unified gene names (*sigB* in the example shown) were assigned and implemented as pan gene symbol, which is provided in the *Summary* section of the pan-genome page as well as of the strain-specific gene pages.

been made to the RefSeq annotation of the NCTC 8325 genome, including addition and removal of gene annotations. Of 104 genes that have been removed with respect to the original annotation considered in our pan-genome, the majority was kept in *Aureo*Wiki, except for those now being part of a larger gene or located on the antisense strand of a newly annotated gene. When searching *Aureo*Wiki for one of the removed genes, a page opens that explains the removal and provides a link to the page of the new gene. We have added 28 newly annotated RefSeq genes and also six genes of NCTC 8325 only reported in UniProt (*bshC*, *mnhF2*, and *psmα1* to *psmα4*) for which complete gene pages are

provided. For instance, the α-type PSM peptides and their encoding genes were described in a study by Wang et al. published in 2007 (Wang et al., 2007), but are still missing in the annotations of staphylococcal genomes. Cytolytic peptides, namely α- and β-type PSMs as well as δ-toxin, are major determinants of *S. aureus* virulence.

Finally, the sequence of 22 genes including housekeeping genes like *rpmF*, *rpsB*, *infC*, *mnmA*, *groEL*, *ezrA*, and *murA1* were corrected based on a resequencing analysis performed by Berscheid et al. (2012). In this study, sequencing errors in the original *S. aureus* NCTC 8325 genome sequence were identified by comparison to the genome sequences of its derivatives RN4220 and RN4220 Δ*mutS* (Nair et al., 2011; Berscheid et al., 2012). Of the 22 corrections, nine resulted in the replacement of an apparent pseudogene by a functional coding sequence.

The resulting genome annotation of *S. aureus* NCTC 8325 including all updates and sequence corrections described in this section is also accessible in the form of FASTA files for genes and proteins by selecting the link to gene-specific information on the *Downloads* page.

3.6. Unified gene names in line with the scientific literature

S. aureus genes were frequently named differently in genome annotations of individual strains, which has led to the use of multiple names for the same (orthologous) gene. Clearly, this situation is a source of confusion, complicating the exchange of results and the knowledge flow in the scientific community. Another problem consists in the fact that assigned gene symbols are often ambiguous, even within a single strain. In addition, in some *S. aureus* strains only a relatively small number of annotated genes is provided with a gene name. Notably, of the 32 strains considered in the current pan-genome, in nine strains (including NCTC 8325) less than 500 genes have a gene symbol, whereas on average 900 genes are annotated with symbols in the remaining 23 strains. In order to support effective scientific communication, a manual curation effort was initiated in the context of *Aureo*Wiki, which aims to provide species-wide unified and unique gene names, implemented as pan gene symbols, for all *S. aureus* genes. Descriptive gene names are given preference to using the common identifier (pan locus tag). In brief, the pan gene symbol is assigned based on strain-specific gene names from NCBI-RefSeq and/or supported by relevant publications or, alternatively, the name of the orthologous *B. subtilis* gene. While annotation of *S. aureus* strain NCTC 8325 serves as the reference, pan gene symbols will also be assigned to all genes not present in its genome, so far realized for two groups of genes, namely those encoding superantigens (Lina et al., 2004) and lipoproteins (Shahmirzadi et al., 2016).

The following rules were applied for pan gene symbol assignment: (1) Use of the most common gene name if it occurs in at least 20 of the 32 strains; (2) Use of an NCBI-RefSeq gene name assigned in one or more of the 32 strains if it is supported by relevant publications (in case of differences in published gene names, the symbol is assigned that was used when the gene/function was first reported); (3) If no RefSeq gene name is supported by relevant publications or no RefSeq gene name is assigned, use of the gene name first introduced in a relevant publication; (4) Use of the name of the orthologous *B. subtilis* gene except for "y"-names; (5) Use of the gene name provided in UniProt if originating from manual assertion.

As a result, almost 600 *S. aureus* genes with multiple (strain-specific) names used in genome annotations and publications have received unified and unique names. Moreover, gene symbols were newly assigned to more than 70 genes, primarily extracted from published studies that suggested gene names based on elucidated or predicted protein functions, but were not considered in the *S. aureus* genome annotations.

In some functional categories particular confusion is caused by differences in published names for the same gene, as for example in the case of wall teichoic acid (WTA) biosynthesis. The enzymes responsible for the first steps of WTA biosynthesis, which are conserved in many Gram-positive bacteria, have been given the names *tagO* and *tagAHGBXD* in *S. aureus* genome annotations and in many publications, but were sporadically renamed as *tar* genes in bacteria producing poly-ribitol-phosphate (RboP) WTAs like most *S. aureus* strains. As suggested by Xia and Peschel (2008), in *Aureo*Wiki the original *tag* names were assigned to the genes encoding the initial highly conserved steps and *tar* was only used for the *tarFIJLS* genes (Qian et al., 2006; Brown et al., 2008; Brown et al., 2012), which are specific for poly-RboP WTA biosynthesis.

3.7. Data accessibility and updates

In order to make the huge amount of information on *S. aureus* genes and their functional characterization easily accessible to researchers, *Aureo*Wiki was developed with user-friendly presentations and intuitive genome browsers. Data accessibility is facilitated by various download options to meet the requirements of bioinformatics applications, particularly implemented in genome-scale studies. By selecting the link to gene-specific information on the *Downloads* page, most of the strain-specific information provided on the *Aureo*Wiki gene pages can be downloaded in tabular form, currently for five *S. aureus* strains (NCTC 8325, COL, Newman, USA300_FPR3757, and N315) and including two RefSeq annotations for each of the strains except for NCTC 8325 as described before. The user can choose between all available information (more than 50 entries per gene) and selected columns, such as the unified gene name (pan gene symbol) or all kinds of protein information including protein function assignments and localization prediction. The second link on the *Downloads* page leads to the orthologue table. It contains in the first column the common identifiers (pan locus tags) of all 6471 genes of the *S. aureus* pan-genome followed by separate columns for each of the 32 strains showing the corresponding strain-specific locus tags. This orthologue mapping information can be downloaded for all 32 strains or for any number of user-selected strains.

The content of *Aureo*Wiki requires constant actualization in order to comply with updates of RefSeq annotations and other databases interrogated for information on *S. aureus* genes and proteins. Furthermore, the functional annotation of gene products, in particular those of so far unknown function, is continuously updated as more information becomes available, leading to the assignment of new pan gene symbols and the addition of relevant literature references. The same holds true for newly published studies on assignments of genes to transcription factor regulons.

3.8. Perspectives

To provide the community with interconnected and up-to-date information on the genes of *S. aureus* and their functional annotation, we created the Wiki-based repository *Aureo*Wiki. It represents the basis for further developments toward a comprehensive coverage of *S. aureus* physiology, including, amongst others, metabolic pathways and all aspects of transcriptional regulation collected from extensive literature search. The envisaged high level of data integration has already been successfully implemented in the *Subti*Wiki database by accompanying modules that graphically present metabolic and regulatory pathways, protein-protein interactions and gene expression data (Michna et al., 2016). With respect to transcriptional regulation in *S. aureus*, an extensive review of published data has been performed. Manual evaluation of over 250 published global transcription and regulatory studies of *S. aureus* wild type and mutant strains exposed to various *in vivo* and *in vitro* growth and stress conditions resulted in a comprehensive list of over 53.000 individual regulatory events and 2000 protein-DNA interactions, representing the transcriptional landscape and regulatory network of *S. aureus* (Fig. 5). Since gene expression may vary considerably between strains and the collected data is derived from many different genetic backgrounds, including model laboratory strains as well as less well characterized clinical isolates, all information will be

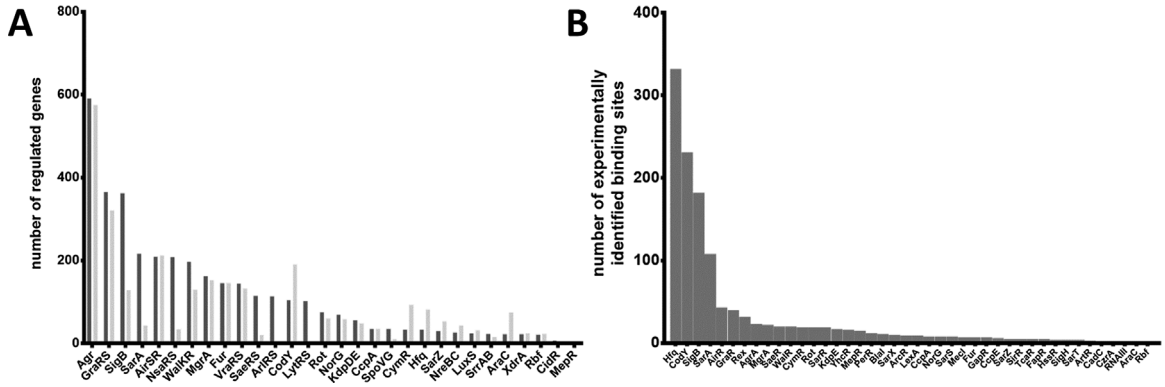

Fig. 5. Transcriptional and regulatory information extracted from literature. The data shown is from experiments aimed at the elucidation of regulons controlled by different transcription factors and the RNA polymerase sigma factor SigB. If transcription of a particular gene was affected in more than one study investigating the same regulator, the gene was counted only once to eliminate redundancy in the data. **A)** Number of up-regulated (green) and down-regulated (red) genes identified in studies comparing the transcription profiles of wild type strains and isogenic mutants lacking a specific regulator. Of all genes assigned to transcription factor regulation, ~50% are subject to multiple regulation. **B)** Number of genes for which a binding site for a regulatory protein was reported in the corresponding up-stream region. Large differences between gene numbers in A and B, e.g. for global regulators such as SigB, are explained by the different methodology used, namely transcriptional profiling vs. regulatory site mapping by e.g. primer extension or EMSA.

summarized on the pan-genome pages. From the pan-genome page, regulatory information can be traced back to the respective strain from which the data was obtained. In addition to the respective literature reference, information will also be provided on the type of experiment used to generate the experimental data, such as ChIP-on-chip (chromatin immunoprecipitation followed by microarray hybridization), primer extension, EMSA (electrophoretic mobility shift assay), and microarray hybridization. As exemplified for SigB, for genes for which supportive regulatory information is available from other genetic backgrounds the text “other strains” will be used to indicate this fact and to provide direct access to the corresponding gene pages. Finally, to integrate regulatory information at the network level, regulatory interaction between genes will be visualized separately on a network page that can be accessed from a regulated gene or regulator if regulatory connections to other genes were detected.

In current efforts, information on non-coding RNAs is being included in *Aureo*Wiki. RNA-mediated regulation in *S. aureus* is well studied and many small regulatory RNAs (sRNAs) have been characterized since the early discovery of the sRNA RNAIII, the main intracellular effector of the Agr quorum-sensing system, in 1993 (for review, see Felden et al., 2011; Tomasini et al., 2013). However, a unified nomenclature and curation of experimental non-coding RNA annotations are lagging behind, which is reflected by the recent establishment of the *Staphylococcal regulatory RNA database* (SRD) (Sassi et al., 2015). In *Aureo*Wiki, gene pages for non-coding RNAs are being generated and connected to the corresponding SRD entries and the *S. aureus* Expression Data Browser (Mäder et al., 2016). In addition, comprehensive annotation of non-coding RNAs for *S. aureus* strain HG001, a derivative of NCTC 8325 (Herbert et al., 2010), which was recently provided by Caldelari et al. (2017), will be used to further improve the annotation of *S. aureus* NCTC 8325.

Finally, current efforts concern peptide data derived from mass spectrometry experiments (Depke et al., 2015; Michalik et al., 2017), which provide proteomic evidence for gene expression and can be accessed from the *Aureo*Wiki gene pages via the PeptideAtlas link. The *S. aureus* PeptideAtlas is a mass spectrometry centered database consisting of high confident peptide spectra, which supports the identification of proteins/peptides that are detectable by proteomic mass spectrometry approaches and aids to find suitable proteotypic peptides for targeted proteomics workflows like SRM (selected reaction monitoring). The spectral library search (e.g. SpectraST) based on high-confidence peptide identifications, which are provided by the PeptideAtlas, clearly outperforms established sequence-based database searches in terms of speed and false-discovery rate (Lam et al., 2007).

Acknowledgments

We thank S. Engelmann, H. Kusch, F. Schmidt and S. Holtfreter for helpful discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft in the framework of SFB/TRR34 (to JPF, MH, SF, UM and UV).

References

Anderson, M., Chen, Y.H., Butler, E.K., Missiakas, D.M., 2011. EsaD, a secretion factor for the Ess pathway in *Staphylococcus aureus*. J. Bacteriol. 193, 1583–1589.
Bécavin, C., Koutero, M., Tchitchek, N., Cerutti, F., Lechat, P., Maillet, N., Hoede, C., Chiapello, H., Gaspin, C., Cossart, P., 2017. Listeriomics: an interactive web platform for systems biology of *Listeria*. mSystems 2, e00186–16.
Baba, T., Bae, T., Schneewind, O., Takeuchi, F., Hiramatsu, K., 2008. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. J. Bacteriol. 190, 300–310.
Becher, D., Hempel, K., Sievers, S., Zühlke, D., Pane-Farre, J., Otto, A., Fuchs, S., Albrecht, D., Bernhardt, J., Engelmann, S., Völker, U., van Dijl, J.M., Hecker, M., 2009. A proteomic view of an important human pathogen–towards the quantification of the entire Staphylococcus aureus proteome. PLoS One 4, e8176.
Berscheid, A., Sass, P., Weber-Lassalle, K., Cheung, A.L., Bierbaum, G., 2012. Revisiting the genomes of the *Staphylococcus aureus* strains NCTC 8325 and RN4220. Int. J. Med. Microbiol. 302, 84–87.
Bischoff, M., Dunman, P., Kormanec, J., Macapagal, D., Murphy, E., Mounts, W., Berger-Bachi, B., Projan, S., 2004. Microarray-based analysis of the *Staphylococcus aureus* sigmaB regulon. J. Bacteriol. 186, 4085–4099.
Brown, S., Zhang, Y.H., Walker, S., 2008. A revised pathway proposed for *Staphylococcus aureus* wall teichoic acid biosynthesis based on in vitro reconstitution of the intracellular steps. Chem. Biol. 15, 12–21.
Brown, S., Xia, G., Luhachack, L.G., Campbell, J., Meredith, T.C., Chen, C., Winstel, V., Gekeler, C., Irazoqui, J.E., Peschel, A., Walker, S., 2012. Methicillin resistance in *Staphylococcus aureus* requires glycosylated wall teichoic acids. Proc. Natl. Acad. Sci. U S A 109, 18909–18914.
Caldelari, I., Chane-Woon-Ming, B., Noirot, C., Moreau, K., Romby, P., Gaspin, C., Marzi, S., 2017. Complete genome sequence and annotation of the *Staphylococcus aureus* strain HG001. Genome Announc. 5, e00783–17.
Cao, Z., Casabona, M.G., Kneuper, H., Chalmers, J.D., Palmer, T., 2016. The Type VII Secretion System of *Staphylococcus Aureus* Secretes a Nuclease Toxin That Targets Competitor Bacteria.
Chambers, H.F., 2001. The changing epidemiology of *Staphylococcus aureus*? Emerg. Infect. Dis. 7, 178–182.
Chen, L., Zheng, D., Liu, B., Yang, J., Jin, Q., 2016. VFDB 2016: hierarchical and refined dataset for big data analysis –10 years on. Nucleic Acids Res. 44, D694–D697.
Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: multiple genome alignment with gene gain loss and rearrangement. PLoS One 5, e11147.
Dehal, P.S., Joachimiak, M.P., Price, M.N., Bates, J.T., Baumohl, J.K., Chivian, D., Friedland, G.D., Huang, K.H., Keller, K., Novichkov, P.S., Dubchak, I.L., Alm, E.J., Arkin, A.P., 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 38, D396–D400.
Depke, M., Michalik, S., Rabe, A., Surmann, K., Brinkmann, L., Jehmlich, N., Bernhardt, J., Hecker, M., Wollscheid, B., Sun, Z., Moritz, R.L., Völker, U., Schmidt, F., 2015. A peptide resource for the analysis of *Staphylococcus aureus* in host-pathogen interaction studies. Proteomics 15, 3648–3661.
Dreisbach, A., Hempel, K., Buist, G., Hecker, M., Becher, D., van Dijl, J.M., 2010. Profiling

the surfacome of *Staphylococcus aureus*. Proteomics 10, 3082–3096.
Enright, M.C., Spratt, B.G., 1999. Multilocus sequence typing. Trends Microbiol. 7, 482–487.
Enright, M.C., Day, N.P., Davies, C.E., Peacock, S.J., Spratt, B.G., 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J. Clin. Microbiol. 38, 1008–1015.
Felden, B., Vandenesch, F., Bouloc, P., Romby, P., 2011. The *Staphylococcus aureus* RNome and its commitment to virulence. PLoS Pathog. 7, e1002006.
Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.
Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285.
Fraunholz, M., Bernhardt, J., Schuldes, J., Daniel, R., Hecker, M., Sinha, B., 2013. Complete genome sequence of *Staphylococcus aureus* 6850, a highly cytotoxic and clinically virulent methicillin-sensitive strain with distant relatedness to prototype strains. Genome Announc. 1, e00775–13.
Fuchs, S., Zühlke, D., Pane-Farre, J., Kusch, H., Wolf, C., Reiss, S., Binh le, T.N., Albrecht, D., Riedel, K., Hecker, M., Engelmann, S., 2013. Aureolib ?a proteome signature library: towards an understanding of Staphylococcus aureus pathophysiology. PLoS One 8, e70669.
Gillaspy, A.F., Worrell, V., Orvis, J., Roe, B.A., Dyer, D.W., Iandolo, J.J., 2006. Staphylococcus aureus NCTC8325 genome. In: Fischetti, V., Novick, R., Ferretti, J., Portnoy, D., Rood, J. (Eds.), Gram- Positive Pathogens. ASM Press, Washington, DC, pp. 381–412.
Haft, D.H., Selengut, J.D., Richter, R.A., Harkins, D., Basu, M.K., Beck, E., 2013. TIGRFAMs and genome properties in 2013. Nucleic Acids Res. 41, D387–D395.
Hempel, K., Herbst, F.A., Moche, M., Hecker, M., Becher, D., 2011. Quantitative proteomic view on secreted, cell surface-associated, and cytoplasmic proteins of the methicillin-resistant human pathogen *Staphylococcus aureus* under iron-limited conditions. J. Proteome Res. 10, 1657–1666.
Hennig, A., Bernhardt, J., Nieselt, K., 2015. Pan-Tetris: an interactive visualisation for Pan-genomes. BMC Bioinf. 16 (11), S3.
Herbert, S., Ziebandt, A.K., Ohlsen, K., Schäfer, T., Hecker, M., Albrecht, D., Novick, R., Götz, F., 2010. Repair of global regulators in *Staphylococcus aureus* 8325 and comparative analysis with other clinical isolates. Infect. Immun. 78, 2877–2889.
Herbig, A., Jäger, G., Battke, F., Nieselt, K., 2012. GenomeRing: alignment visualization based on SuperGenome coordinates. Bioinformatics 28, 7–15.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kos, V.N., Desjardins, C.A., Griggs, A., Cerqueira, G., van Tonder, A., Holden, M.T., Godfrey, P., Palmer, K.L., Bodi, K., Mongodin, E.F., Wortman, J., Feldgarden, M., Lawley, T., Gill, S.R., Haas, B.J., Birren, B., Gilmore, M.S., 2012. Comparative genomics of vancomycin-resistant *Staphylococcus aureus* strains and their positions within the clade most commonly associated with Methicillin-resistant S. aureus hospital-acquired infection in the United States. MBio 3, e00112–12.
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., Hiramatsu, K., 2001. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. Lancet 357, 1225–1240.
Lam, H., Deutsch, E.W., Eddes, J.S., Eng, J.K., King, N., Stein, S.E., Aebersold, R., 2007. Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics 7, 655–667.
Lammers, C.R., Flórez, L.A., Schmeisky, A.G., Roppel, S.F., Mäder, U., Hamoen, L., Stülke, J., 2010. Connecting parts with pieces: *Subti*Wiki and *Subti*Pathways integrate gene and pathway annotation for *Bacillus subtilis*. Microbiology 156, 849–859.
Lina, G., Bohach, G.A., Nair, S.P., Hiramatsu, K., Jouvin-Marche, E., Mariuzza, R., 2004. International nomenclature committee for staphylococcal superantigens. Standard nomenclature for the superantigens expressed by staphylococcus. J. Infect. Dis. 189 (12), 2334–2336 (Jun 15).
Lindsay, J.A., Moore, C.E., Day, N.P., Peacock, S.J., Witney, A.A., Stabler, R.A., Husain, S.E., Butcher, P.D., Hinds, J., 2006. Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. J. Bacteriol. 188, 669–676.
Lindsay, J.A., 2010. Genomic variation and evolution of *Staphylococcus aureus*. Int. J. Med. Microbiol. 300, 98–103.
Lindsay, J.A., 2014. *Staphylococcus aureus* genomics and the impact of horizontal gene transfer. Int. J. Med. Microbiol. 304, 103–109.
Lowy, F.D., 1998. Staphylococcus aureus infections. N. Engl. J. Med. 339, 520–532.
Luo, H., Lin, Y., Gao, F., Zhang, C.T., Zhang, R., 2014. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Res. 42, D574–D580.
Mäder, U., Schmeisky, A.G., Flórez, L.A., Stülke, J., 2012. SubtiWiki –a comprehensive resource for the model organism Bacillus subtilis. Nucleic Acids Res. 40, D1278–D1287.
Mäder, U., Nicolas, P., Depke, M., Pané-Farré, J., Debarbouille, M., van der Kooi-Pol, M.M., Guerin, C., Derozier, S., Hiron, A., Jarmer, H., Leduc, A., Michalik, S., Reilman, E., Schaffer, M., Schmidt, F., Bessieres, P., Noirot, P., Hecker, M., Msadek, T., Völker, U., van Dijl, J.M., 2016. Staphylococcus aureus transcriptome architecture: from laboratory to infection-mimicking conditions. PLoS Genet. 12, e1005962.
Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pangenome. Curr. Opin. Genet. Dev. 15, 589–594.
Michalik, S., Depke, M., Murr, A., Gesell Salazar, M., Kusebauch, U., Sun, Z., Meyer, T.C., Surmann, K., Pförtner, H., Hildebrandt, P., Weiss, S., Palma Medina, L.M., Gutjahr, M., Hammer, E., Becher, D., Pribyl, T., Hammerschmidt, S., Deutsch, E.W., Bader, S.L., Hecker, M., Moritz, R.L., Mäder, U., Völker, U., Schmidt, F., 2017. A global Staphylococcus aureus proteome resource applied to the in vivo characterization of host-pathogen interactions. Sci. Rep. 7, 9718.
Michna, R.H., Commichau, F.M., Tödter, D., Zschiedrich, C.P., Stülke, J., 2014. SubtiWiki –a database for the model organism Bacillus subtilis that links pathway, interaction and expression information. Nucleic Acids Res. 42, D692–D698.
Michna, R.H., Zhu, B., Mäder, U., Stülke, J., 2016. SubtiWiki 2. 0–an integrated database for the model organism Bacillus subtilis. Nucleic Acids Res. 44, D654–D662.
Nagarajan, V., Elasri, M.O., 2007. SAMMD: staphylococcus aureus microarray meta-database. BMC Genomics 8, 351.
Nair, D., Memmi, G., Hernandez, D., Bard, J., Beaume, M., Gill, S., Francois, P., Cheung, A.L., 2011. Whole-genome sequencing of Staphylococcus aureus strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. J. Bacteriol. 193, 2332–2335.
Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R.A., Kazanov, M.D., Riehl, W., Arkin, A.P., Dubchak, I., Rodionov, D.A., 2013. RegPrecise 3. 0–a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics 14, 745.
Otto, M., 2013. Community-associated MRSA: what makes them special? Int. J. Med. Microbiol. 303, 324–330.
Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V., 2005. Nucleic Acids Res. 33, 5691–5702.
Pané-Farré, J., Jonas, B., Forstner, K., Engelmann, S., Hecker, M., 2006. The sigmaB regulon in Staphylococcus aureus and its regulation. Int. J. Med. Microbiol. 296, 237–258.
Qian, Z., Yin, Y., Zhang, Y., Lu, L., Li, Y., Jiang, Y., 2006. Genomic characterization of ribitol teichoic acid synthesis in Staphylococcus aureus: genes, genomic organization and gene duplication. BMC Genomics 7, 74.
Sassi, M., Augagneur, Y., Mauro, T., Ivain, L., Chabelskaya, S., Hallier, M., Sallou, O., Felden, B., 2015. SRD: a staphylococcus regulatory RNA database. RNA 21, 1005–1017.
Schulthess, B., Bloes, D.A., Francois, P., Girard, M., Schrenzel, J., Bischoff, M., Berger-Bachi, B., 2011. The sigmaB-dependent yabJ-spoVG operon is involved in the regulation of extracellular nuclease, lipase, and protease expression in Staphylococcus aureus. J. Bacteriol. 193, 4954–4962.
Shahmirzadi, S.V., Nguyen, M.T., Götz, F., 2016. Evaluation of *Staphylococcus aureus* lipoproteins: role in nutritional acquisition and pathogenicity. Front. Microbiol. 7, 1404.
Tatusova, T., Ciufo, S., Federhen, S., Fedorov, B., McVeigh, R., O'Neill, K., Tolstoy, I., Zaslavsky, L., 2015. Update on RefSeq microbial genomes resources. Nucleic Acids Res. 43, D599–D605.
Tomasini, A., François, P., Howden, B.P., Fechter, P., Romby, P., Caldelari, I., 2013. The importance of regulatory RNAs in Staphylococcus aureus. Infect. Genet. Evol. 21, 616–626.
Vandenesch, F., Naimi, T., Enright, M.C., Lina, G., Nimmo, G.R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.E., Etienne, J., 2003. Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerg. Infect. Dis. 9, 978–984.
Wang, R., Braughton, K.R., Kretschmer, D., Bach, T.H., Queck, S.Y., Li, M., Kennedy, A.D., Dorward, D.W., Klebanoff, S.J., Peschel, A., DeLeo, F.R., Otto, M., 2007. Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. Nat. Med. 13, 1510–1514.
Xia, G., Peschel, A., 2008. Toward the pathway of *S. aureus* WTA biosynthesis. Chem. Biol. 15, 95–96.
Zühlke, D., Dorries, K., Bernhardt, J., Maass, S., Muntel, J., Liebscher, V., Pane-Farre, J., Riedel, K., Lalk, M., Völker, U., Engelmann, S., Becher, D., Fuchs, S., Hecker, M., 2016. Costs of life – dynamics of the protein inventory of *Staphylococcus aureus* during anaerobiosis. Sci. Rep. 6, 28172.
de Kraker, M.E., Wolkewitz, M., Davey, P.G., Koller, W., Berger, J., Nagler, J., Icket, C., Kalenic, S., Horvatic, J., Seifert, H., Kaasch, A.J., Paniara, O., Argyropoulou, A., Bompola, M., Smyth, E., Skally, M., Raglio, A., Dumpis, U., Kelmere, A.M., Borg, M., Xuereb, D., Ghita, M.C., Noble, M., Kolman, J., Grabljevec, S., Turner, D., Lansbury, L., Grundmann, H., 2011. Clinical impact of antimicrobial resistance in European hospitals: excess mortality and length of hospital stay related to methicillin-resistant *Staphylococcus aureus* bloodstream infections. Antimicrob. Agents Chemother. 55, 1598–1605.
van Belkum, A., Verkaik, N.J., de Vogel, C.P., Boelens, H.A., Verveer, J., Nouwen, J.L., Verbrugh, H.A., Wertheim, H.F., 2009. Reclassification of *Staphylococcus aureus* nasal carriage types. J. Infect. Dis. 199, 1820–1826.