# Comparative genomics analysis of two Helcococcus kunzii strains co-isolated with Staphylococcus aureus from diabetic foot ulcers

Benjamin A. R. N Durand, Alex Yahiaoui Martinez, Damien Baud, Patrice François, Jean-Philippe Lavigne, Catherine Dunyach-Rémy



Genomics 114 (2022) 110365

Contents lists available at ScienceDirect

Genomics

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

# Comparative genomics analysis of two *Helcococcus kunzii* strains co-isolated with *Staphylococcus aureus* from diabetic foot ulcers

Benjamin A.R.N. Durand [a,1], Alex Yahiaoui Martinez [b,1], Damien Baud [c], Patrice François [c], Jean-Philippe Lavigne [a,*], Catherine Dunyach-Remy [a]

[a] *Bacterial Virulence and Chronic Infections, INSERM U1047, University of Montpellier, Department of Microbiology and Hospital Hygiene, University Hospital Nîmes, 30908 Nîmes, France*
[b] *Department of Microbiology and Hospital Hygiene, University Hospital Nîmes, University of Montpellier, 30029 Nîmes, France*
[c] *Department of Infectious Diseases, Genomic Research Laboratory, Geneva University Hospitals, 1205 Geneva, Switzerland*

ARTICLE INFO

*Keywords:*
Autoinducer-like protein
Diabetic foot ulcer
Genomic characterization
*Helcococcus kunzii*
Iron export
Phage

ABSTRACT

*Helcococcus kunzii* is a commensal Gram-positive bacterial species recovered from the human skin microbiota and considered as an opportunistic pathogen. Although little is known about its clinical significance, its increased abundance has been reported in infected wounds, particularly in foot ulcers in persons with diabetes. This species is usually detected in mixed cultures from human specimens and frequently isolated with *Staphylococcus aureus*. Modulation of staphylococci virulence by *H. kunzii* has been shown in an infection model of *Caenorhabditis elegans*. The aim of this study was to compare the genomes of two *H. kunzii* strains isolated from foot ulcers -isolate H13 and H10 showing high or low impact on *S. aureus* virulence, respectively- and the *H. kunzii* ATCC51366 strain. Whole genome analyses revealed some differences between the two strains: length (2.06 Mb (H13) and 2.05 Mb (H10) bp), GC content (29.3% (H13) and 29.5% (H10)) and gene content (1,884 (H13) and 1,786 (H10) predicted genes). The core-proteome phylogenies within the genus characterised *H. kunzii* H13 and H10 as genetically similar to their ancestor. The main differences between the strains were mainly in sugar-associated transporters and various hypothetical proteins. Five targets were identified as potentially involved in *S. aureus* virulence modulation in both genomes: the two-component iron export system and three autoinducer-like proteins. Moreover, H13 strain harbours a prophage inserted in 1,261,110-1,295,549 (*attL-attR*), which is absent in H10 strain. The prophage PhiCD38_2 was previously reported for its ability to modulate secretion profile, reinforcing the autoinducer-like hypothesis. In the future, transcriptomics or metaproteomics approaches could be performed to better characterize the H13 strain and possibly identify the underlying mechanism for *S. aureus* virulence modulation.

## 1. Introduction

The genus *Helcococcus* includes catalase-negative, facultatively anaerobic, Gram-positive lipophilic cocci and belongs to the *Peptoniphilaceae* family, *Tissierellales* order, *Tissierellia* class within the *Firmicutes* phylum. To date, five species have been reported: *H. kunzii*, *Helcococcus massiliensis*, *Helcococcus ovis*, *Helcococcus seattlentis*, and *Helcococcus sueciencis*. *H. kunzii* is a commensal bacterium of the human skin microbiota. This species is also considered as an opportunistic pathogen in invasive infections, such as bacteraemia, pleural emphysema, implantable cardiac device infection, prosthetic joint infection and brain abscess [1–5]. First isolated from a human wound in 1993 [6], *H. kunzii* has mainly been described in case-reports and microbiota description studies. The American type culture collection (ATCC) contains a reference isolate ATCC51366 identified in 1997 [7] and sequenced more recently. Its general antibiotic resistance profile was determined

*Abbreviations:* AIP, AutoInducers Peptide; ANI, Average Nucleotide Identity; BLAST, Basic Local Alignment Search Tool; BRIG, BLAST Ring Image Generator; CDS, Coding DNA sequences; DDH, DNA-DNA Hybridization; FUd, Foot ulcers in persons with diabetes; MALDI-TOF MS, Matrix Assisted Laser Desorption Ionization-Time of Flight coupled with Mass Spectrometry system; NCBI, National Center for Biotechnology Information; PGAP, Prokaryotic Genome Annotation Pipeline; PATRIC, PAThosystems Resource Integration Center; PHASTER, PHAge Search Tool Enhanced Release; SNP, Single Nucleotide polymorphisms; TCS, Tetra Correlation Search.

* Corresponding author.
*E-mail address:* jean.philippe.lavigne@chu-nimes.fr (J.-P. Lavigne).
[1] These authors contributed equally to this paper.

https://doi.org/10.1016/j.ygeno.2022.110365
Received 15 December 2021; Received in revised form 1 March 2022; Accepted 6 April 2022
Available online 9 April 2022

recently [8].

Clinically, *H. kunzii* has mainly been described in the context of infected wounds, notably in infected foot ulcers in persons with diabetes (FUd) [7]. FUd pathogenesis is complex due to the combination of glycaemic imbalance, peripheral neuropathy, and arteriopathy. It comprises different diseases ranging from a superficial skin lesion with progression into deeper tissue and even into the bone, which in rare cases can lead to bloodstream infections, but more frequently to osteomyelitis and amputation [9–11]. In cooler climates, Gram-positive aerobic cocci are the main microorganisms responsible for infected FUd, and *S. aureus* is the most commonly isolated bacteria [12,13]. Interestingly, a previous publication demonstrated that *H. kunzii* was exclusively co-isolated with *S. aureus* among samples obtained from infected FUd [8]. Moreover, our team observed that some *H. kunzii*, represented by the H13 strain, altered *S. aureus* virulence in a *Caenorhabditis elegans* model with a significant decrease of the mortality of the worms when co-infected by the two species. In contrast, other *H. kunzii*, represented by H10 strain, showed lower impact on *S. aureus* virulence. This modulation was shown to be mediated through a downregulation of the Staphylococcal *agr* system [14], for which activation was mediated by AutoInducers Peptide (AIP) [15]. This resulted in two main hypotheses for the molecular basis of this interaction due to: i) an iron competition and/or ii) an extracellular interference with the quorum sensing *agr* loop, auto-inducers-like (Ail) following the model described by Horswill et al. [15].

The aim of this study was to genotypically describe and compare two representative *H. kunzii* strains isolated from infected FUd showing high (*H. kunzii* H13) or low (*H. kunzii* H10) impact on *S. aureus* virulence and to determine the basis of virulent traits. The genomes of these strains were also compared to the *H. kunzii* ATCC51366, the only sequenced strain available so far. We also wanted to validate the taxonomic assignment of the newly sequenced strains, assess homogeneity within the species and characterize the genomic specific traits.

## 2. Materials and methods

### 2.1. Bacterial strains and growth conditions

*H. kunzii* H13 and H10 were isolated in 2013 at Cahors General Hospital (France) from samples taken from infected FUd [8]. Both strains were isolated on Trypticase Soy agar +5% sheep blood plate media (BioMérieux, Marcy l'Etoile, France) at 37 °C under anaerobic growth condition after 48 h. Identification was performed with Matrix Assisted Laser Desorption Ionization-Time of Flight coupled with Mass Spectrometry system (MALDI-TOF MS, VITEK® MS Version MS-3.2 CE.0, BioMérieux) [16]. The isolates were further grown on Columbia blood agar plate media (BioMérieux) at 37 °C under 5% $CO_2$ for 48 h.

### 2.2. Molecular biology and genome sequencing

The genomic DNA of *H. kunzii* H10 and H13 was extracted using a DNeasy UltraClean Microbial Kit (Qiagen France SAS, Courtaboeuf, France) according to manufacturer instructions. Genome sequencing was conducted by the Genomics Platform iGE3 (Geneva, Switzerland). Genomic DNA was quantified with a Qubit fluorimeter (Life Technologies, Illkirch, France) and fragmented with a Covaris S2. The TruSeq DNA-nano kit from Illumina was used for the library preparation with 200 ng of DNA as input. Library molarity and quality were assessed on the Qubit and the Tape Station using a DNA High sensitivity chip (Agilent Technologies, Basel, Switzerland). Libraries were loaded on a HiSeq 2500 with the reagent kit v2. Paired-end reads of 250 bp were generated.

### 2.3. Genome assembly, annotation and comparative genome analysis

De novo contig assembly was performed using SPAdes v3.10.0, (BioProject: PRJNA600407) and the scaffolding of the contigs were then assembled using CONTIGuator webserver with *H. kunzii* ATCC51366 (GenBank accession number: GCA_000245755.1) as the reference. The genomes were annotated using National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP, v4.11) [17–19]. Briefly, this annotation service provided by the NCBI uses a combination of prediction algorithms (gene prediction and homology methods) to generate annotations related to submitted genomes. The genome sequences of *H. kunzii* H13 and H10 can be accessed through the GenBank database (accession numbers SAMN13819011 for H13 and SAMN14604363 for H10).

Direct comparisons between *H. kunzii* H13 and H10 and between *H. kunzii* ATCC51366 and H10 were generated using BLAST Ring Image Generator (BRIG) [20]. The origin of replication was determined using GenSkew software (http://genskew.csb.univie.ac.at/).

Genomes were screened for plasmid identification using the webtool PlasmidFinder 2.0, (https://cge.cbs.dtu.dk/services/PlasmidFinder/) [21]. The gene plotting was performed on PGAP annotated files (GeneBank format) allowing the circular representation of the *H. kunzii* H10 and H13 strains using DNAPlotter software [22].

Taxonomy assignation was corroborated through determination of Tetra Correlation Search (TCS), Average Nucleotide Identity (ANI) and DNA-DNA Hybridization (DDH) values that were assessed in silico using online tools. TCS- and pairwise ANIb- values were obtained using JSpeciesWS (http://jspecies.ribohost.com/jspeciesws /#analyse) [23]. Briefly, ANIb is a genome ~1kbp in silico fragmentation combined with BLASTn (nucleotide Basic Local Alignment Search Tool) tool to evaluate and compare species specific signatures. TCS is a regression tool allowing occurrence frequencies comparison of tetranuclotide combination against a genomic database (JSpeciesWS) [24].

Assembled genomes of strain H13 and H10 were uploaded in the distance calculation form of the Genome to Genome Distance Calculator web server (GGDC v2.1, http://ggdc.dsmz.de/ggdc.php#) with the recommended local alignment tool BLAST+ and compared with all available *Helcococcus* representative complete genomes (Table 1) to obtain DDH-values. The statistic comparison (logistic regression) used a significant probability value of DDH >70 or 79% (Supplementary material 2). Briefly, in silico DDH were calculated on intergenomic, without prior genome hashing (performed through BLASTn hit), distance allowing species assignation [25–27].

A phylogenetic analysis was carried out using a predicted core proteome with all available *Helcococcus* representative complete genomes (Table 1) performed using Orthofinder v2.2.7 [28,29] (Supplementary material 3). The phylogenetic tree was built using all *Helcococcus* genomes available on GenBank. FigTree v1.4.4 was used to draw the phylogenetic distance calculated between each species protein of the core genome (http://tree.bio.ed.ac.uk/software/figtree/). Single Nucleotide polymorphisms (SNP) were studied using Snippy [30] pipeline (version 4.5.0, Supplementary material 4), on Galaxy server (https://usegalaxy.org/) to explore genomic variation within *H. kunzii* clade against the reference strain ATCC51366. Briefly, Snippy maps the Illumina reads onto a haploid reference genome (here the ATCC51366 one) and finds substitution, insertion, or deletion sites. Combination of the results with other comparisons for the same reference generates a core SNP alignment. Potential Ail were tracked using a recognition search pattern written in R (Supplementary material 1, **Fig. S1. A**), and using characteristic motifs of AIP, the receptor docking hydrophobic ring and activator tail [31].

Side-by-side comparisons were performed using the PAThosystems Resource Integration Center webtool (PATRIC, Supplementary material 5) [32]. H13 and H10 continuous fasta files were submitted for annotation on the platform. The resulting annotations were compared (two-by-two) using the proteome comparison service. The obtained excel files were filtered with a 0.9 threshold for identity percentage for the Venn diagram constitution.

Genomes were screened for phage insert using PHASTER (PHAge

Search Tool Enhanced Release) webtool [33] (Supplementary material 6). Study of most abundant phage species was performed using linear comparison of the three *Clostridium* phages and *H. kunzii* H13 prophage region, as illustrated through Easyfig [34]. Missing rRNAs were addressed through BLASTn, and iron export system by BLASTp. Conserved domains comparisons were performed using the BLAST for either nucleotides or protein sequences provided by NCBI [35,36].

A flowchart summarising these different steps is presented in Fig. 1.

## 3. Results

### 3.1. Genomic sequencing analysis of H. kunzii H10 and H13

The genomes of *H. kunzii* H13 and H10 strains were assembled in only four contigs (169,814 bp; 577,790 bp; 810,572 bp; 524,317 bp for H13 and 153,583 bp; 523,723 bp; 805,181 bp; 569,384 bp for H10). After genome polishing, the final genome sizes obtained were 2,062,793 bp for H13 and 2,052,171 bp for H10, and their origins of replication were found at 1,870,235 bp and 1,797,553 bp, respectively (**Fig. S3**).

The two studied *H. kunzii* strains exhibited a similar GC content compared to ATCC51366 (Table 2) with 29.33%, 29.61% and 29.50% for H13, H10 and ATCC51366, respectively. The numbers of coding DNA sequences (CDS) (and total gene number prediction) were 1852 (1888) for H13 and 1810 (1846) for H10, compared to 1862 (1903) for ATCC51366.

With a coverage varying between 88 and 92%, strong identity (>99%) was found between all strains. However, H13 and H10 genomes were, respectively, 33,671 bp and 44,293 bp shorter compared to the ATCC51366, predominantly due to the gene content. Indeed, whereas H13 had only 15 fewer genes than ATCC51366, H10 lacked 57 genes compared to ATCC51366.

### 3.2. Phylogenetic relationship of the three H. kunzii to other Helcococcus species

Initially identified by MALDI-TOF, an investigation of the species level assignation for H13 and H10 was carried out at the genomic level. TCS was first used to obtain their closest-related organisms, confirming the close relatedness with the *H. kunzii* ATCC51366 strain with z-score of 0.99941 and 0.9986 for H13 and H10, respectively (**Table S1**). This result was then refined through an ANIb including all the *Helcococcus* species present in the genome database of JSpeciesWS and H13 and H10 strains (Table 1). Once again, the ATCC51366 strain was the closest *Helcococcus* with 99.00% nucleotide identity on 82.12% and 80.33% of aligned nucleotides for H13 and H10, respectively. The nucleotide identity percentages decreased as percentage of aligned nucleotides increased. When using H13 genome as template, the percentage of aligned nucleotides for ATCC51366 increased to 84.41%, but the average identity decreased to 98.85% (for H10 strain 82.43% of aligned nucleotides and 98.90% identity were obtained). Interestingly H13 and H10 strains showed better identity (99.01% [82.35% aligned] and 99.09% [82.76% aligned] identity with H13 or H10 respectively as template) compared to the values obtained upon comparison with the ATCC51366 (**Table S2**). The species level assignation was validated using DDH, resulting in a probability of 96.32% for H13 belonging to the same species as *H. kunzii* ATCC51366 (i.e DDH ≥70%), and 96.36% for H10. Their DDH values were 91.60% and 91.80%, respectively (**Table S3**).

A phylogenetic tree was built to gain better insight into the classification of the *Helcococcus* genera. *H. kunzii* H13 and H10 strains were compared with different strains belonging to *Helcococcus* species. Out of the total of 14,263 encoded proteins, 13,832 (96.98%) were assigned to an orthogroup. Six species were retrieved in at least 50% of final 2188 orthogroups. In regard of overlapping orthogroups and, as suggested, by the mean number of species by orthogroup (6.3), most of them were shared across species. 759 orthogroups were retrieved from all explored species, constituting the working core proteome. 34.69% protein conversation among the genera (759/2188) included 680 single copy orthogroups (**Table S4**).

The *Helcococcus* genus was clearly separated into three clades and four clusters: the *H. sueciensis* and *H. massiliensis* clades showed the shortest evolutionary distance from the *Helcococcus* common ancestor and the second clade containing the *H. ovis* and the *H. kunzii* clusters (Fig. 2). In this second cluster, H13 and H10 were closest relatives of *H. kunzii* ATCC51366. The H13 strain appeared to be closer to the common ancestor of the *H. kunzii* species from whom the ATCC51366 and H10 strains seemed to emerge. Finally, the *H. kunzii* H10 strain showed the most evolutionary distance within this clade. SNPs analysis identified 6226 and 6739 divergent nucleotides between H10 and ATCC51366 and H13 and ATCC51366, respectively, corresponding to 0.30% (6226/2,052,171) and 0.32% (6739/2,062,793) of the genome, respectively (**Table S5**). A test of given proportion was used to evaluate differences in SNPs rate. No significant result was observed between two proportions 0.30 and 0.32 ($p\text{-value} = 0.621$), indicating similar SNPs rate between H10 and H13 upon ATCC51366 comparison. None of these mutations affected genes classified as virulence factor-encoding genes.

### 3.3. Comparative genome analysis of ATCC51366 and the two studied strains

The genomic comparison between the *H. kunzii* H10 and H13 and the ATCC51366 strains showed a good coverage between H13 and ATCC51366 genomes with a high identity (92% coverage and 99.82% identity), similar genome size (2.06 Mb vs 2.09 Mb), similar GC content and total number of CDS (Table 2).

Compared to H13, some unmatched regions with variable sizes were observed, most often shared by ATCC51366 and H10 strains. We focused on regions 15 Kbp or more, and particularly regions presenting low identity of H10 towards H13 (meaning that the ATCC31566 sequence was largely preserved in those regions) resulting in five regions of interest. They were further subdivided into two groups: i) three predominantly specific to H10 and ii) two shared between H10 and ATCC51366. From the three large sequence areas belonging to H13 and only partially found specifically in H10 strain (group i)), we identified: a 33,401 Kbp region located between 343,164 and 376,566 bp, a 35,718 Kbp region located 1260,001-1295,719 bp and a 173,132 Kbp region located

**Table 1**
Summary of *Helcococcus* isolates compared against *Helcococcus kunzii* H13.

| Genus | Species | Strains | BioSample accession number | RefSeq accession number | Contigator | Ortho-Finder | BRIG | DNA-plotter | DDH | TCS/ANI |
|---|---|---|---|---|---|---|---|---|---|---|
| *Helcococcus* | *kunzii* | ATCC51366 | SAMN02596728 | GCA_000245755.1 | yes | yes | yes | yes | yes | yes |
| *Helcococcus* | *kunzii* | H10 | SAMN14604363 | – | yes | yes | yes | yes | yes | – |
| *Helcococcus* | *ovis* | KG-40 | SAMN10734408 | GCA_004524815.1 | – | yes | – | – | yes | – |
| *Helcococcus* | *ovis* | KG-39 | SAMN10734407 | GCA_004524755.1 | – | yes | – | – | yes | – |
| *Helcococcus* | *ovis* | KG-37 | SAMN10734406 | GCA_004524765.1 | – | yes | – | – | yes | – |
| *Helcococcus* | *ovis* | KG-36 | SAMN10734405 | GCA_004524775.1 | – | yes | – | – | yes | – |
| *Helcococcus* | *sueciensis* | DSM 172443 | SAMN02441168 | GCA_000423145.1 | – | yes | – | – | yes | yes |
| *Helcococcus* | *massilensis* | Marseille-P4590 | SAMEA104557165 | GCA_900258485.1 | – | yes | – | – | yes | yes |

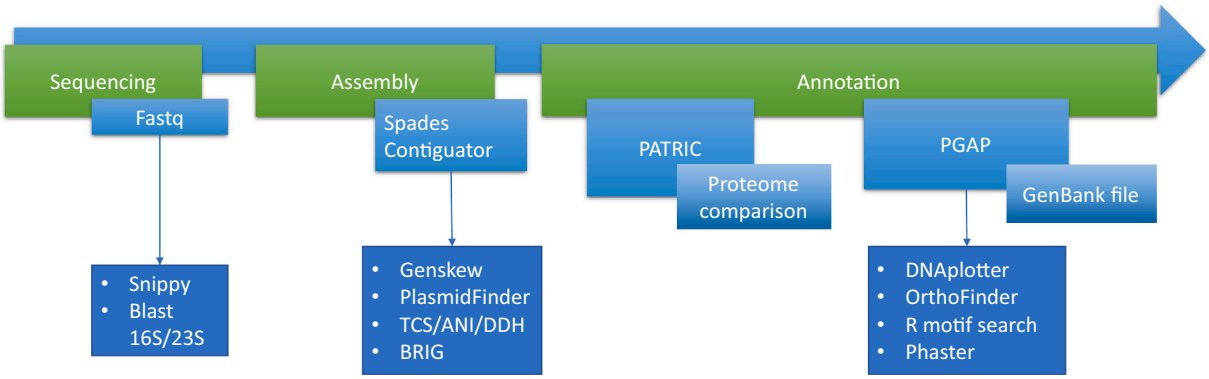


**Fig. 1.** Flowchart illustrating entries and tools used in the course of this study.

**Table 2**
Main characteristics of *Helcococcus kunzii* genomes.

| Feature | H13 | ATCC 51366 | H10 |
|---|---|---|---|
| Genome size (bp) | 2,062,793 | 2,096,464 | 2,052,171 |
| Number contigs | 4 | 39 | 4 |
| Largest contig (bp) | 810,572 | 292,748 | 805,181 |
| N50 length[1] (bp) | 577,790 | 98,606 | 569,384 |
| Number Scaffolds | 1 | 2 | 1 |
| Number of genes | 1888 | 1903 | 1846 |
| Number of CDS | 1852 | 1862 | 1810 |
| Number of genes (coding) | 1824 | 1841 | 1786 |
| Number of rRNA genes | 2 (5S) | 2, 2, 2 (5S, 16S, 23S) | 2 (5S) |
| Number of tRNA genes | 31 | 32 | 31 |
| DNA G + C content (mol%) | 29.33 | 29.50 | 29.61 |
| BLASTn comparison[2] (%coverage; % identity) | | | |
| vs H10 | 89; 99.76 | 88; 99.81 | 100; 100 |
| vs ATCC 51366 | 92; 99.82 | 100; 100 | 90; 99.81 |
| vs H13 | 100; 100 | 90; 99.82 | 89; 99.76 |
| Date of collection | Aug. 2013 [8] | Jul. 1993 [6] | Aug. 2013 [8] |
| Location | France and Switzerland | Faculty of Sciences, Ghent University, Belgium | France and Switzerland |
| Date of sequencing | Feb. 2017 | Jan. 2012 | Feb. 2017 |

[1] Contig length at which 50% of the total base in final genome were in contig of that length or greater.
[2] Concatenated supercontig of ATCC51366 was used. Only the higher identity results were reported.

between 1,889,246 and 2,062,378 bp. They accounted for 50, 52 and 179 CDS in H13, respectively (Fig. 3A). Interestingly, the third, and longest, unmatched region was found following the origin of replication (**Fig. S3**). The two other unmatched regions (group ii) were located at 78,087–94,786 bp (16,699 Kbp, 16 CDS) and 1,696,421-1,736,592 bp (40,171 Kbp, 39 CDS).

The circularised maps of the *H. kunzii* H10 and H13 strains based on the predicted CDS illustrated their positioning and GC over-abundance along both genomes (Fig. 3B and C). Each genome presented two rRNA (duplication of the 5S) in addition to a set of 31 tRNA for both variants, compared to the ATCC51366 (with 6 rRNA (duplication of the 5S, 16S, and 23S) and 32 tRNA) (Table 2).

### 3.4. Genomic structure and gene content

To investigate the genomic structure and gene content of *H. kunzii*, we performed pan-genome analysis using the protein sequences of the three *H. kunzii* strains. Our analysis showed that *H. kunzii* had 1551 core genes (66.57%) within its pan-genome, which consisted of 2330 genes (Fig. 4). The H13 and ATCC51366 strains shared 101 identical proteins.

Both newly sequenced genomes showed a codifying potential of 2157 genes, predominantly distributed between core genes ($n = 1614$; 74.83%) compared to accessory genes ($n = 543$; 25.17%). Moreover, the two genomes presented a comparable number of hypothetical proteins: $n = 522$ for H10 and $n = 582$ for H13, representing 28.09% and 30.42% of genomic coding abilities, respectively.

Built-in protein family assignation conducted via PATRIC annotation allowed partial classification of genes into various subsystems, linking 36% of H13 coding ability and 37% of H10 to function repartition. There was a similar repartition of protein processing and metabolism, accounting for >50% of subsystem assignations, with 384 genes for both H13 and H10 (Fig. 5). More specifically, of the 120 metabolism pathways retrieved, only six presented variations between the strains, principally affecting lipid metabolism ($n = 3$), followed by xenobiotic metabolism (n = 1), a co-factors/vitamins metabolism (n = 1), and immune system (n = 1) (**Table S6**).

### 3.5. Bacterial resistance, defence, and pathogenicity

The subsystem classification demonstrated that H13 exhibited 32 genes related to resistance towards antibiotics and toxic compounds, increased to 34 for H10 (a choloylglycine hydrolase and a modification associated with macrolide resistance). The 10 predicted resistance markers covered various families of antibiotics: aminoglycosides, fluoroquinolones, tetracyclines, peptide antibiotics; most of which were predicted active, except for mupirocin and fusidic acid, which presented an uncertain level of activity. Regarding this last antibiotic family, a resistance gene was specifically observed in H10 and H13 (via the large subunit ribosomal protein L9e) but without certainty about the impact on the fusidic acid resistance (**Table S7**).

No known virulence factor was found in the H13 genome, whereas a DUF1706 domain-containing protein was detected in H10; this protein had an 83% protein identity with a conserved hypothetical protein of *Streptococcus pneumoniae* TIGR4 involved in lung infection in a murine model [37]. In addition, ATCC51366 harboured two supplementary genes: a transcriptional regulator of AraC family and an ABC efflux transporter for a total of three virulence factors.

Regarding iron-related metabolism genes, a bicomponent system of iron export was predicted in the three genomes. Briefly, this system is composed by two proteins: FetA, an ATPase, and its associated membrane channel, FetB. Between H13 and H10, only *fetA* sequence differed: three nucleotides are substituted in H10 compared to H13, resulting in P14S, D132Y and V154E mutations. These mutations are not in an important region and do not affect the protein (**Fig. S4**).

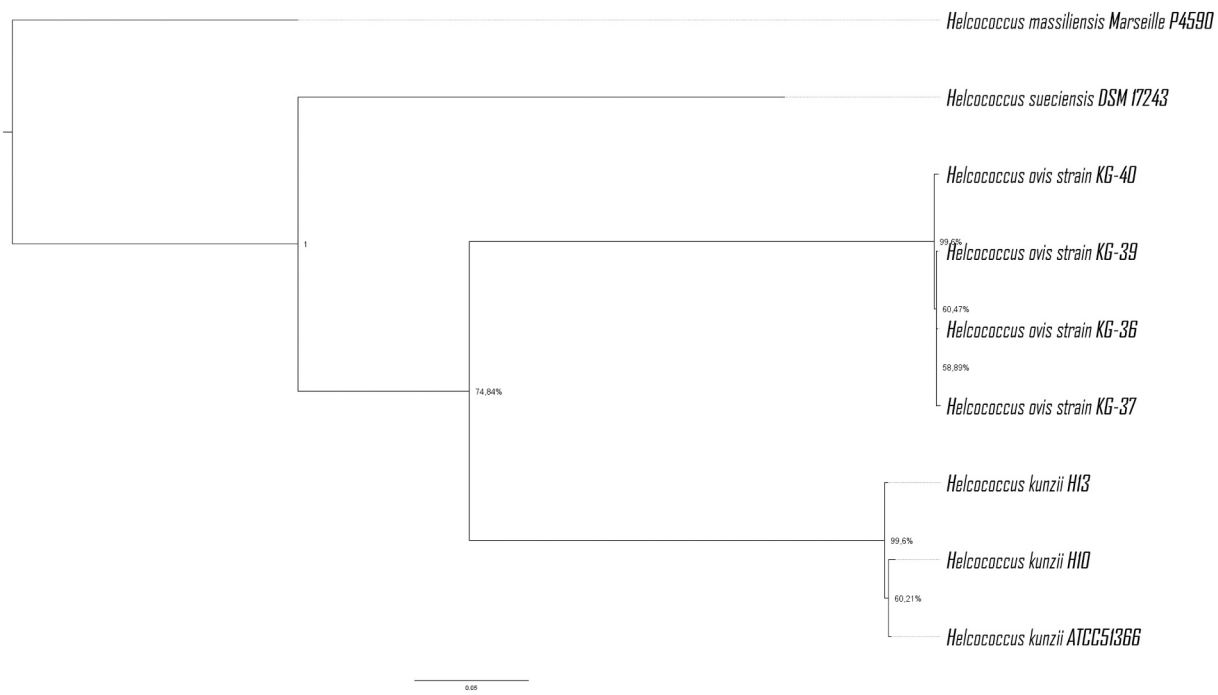


**Fig. 2.** Phylogenetic position of *Helcococcus kunzii* strain H13 and H10 according to the core predicted proteome. This phylogenetic tree was based on 759 core predicted proteins of 9 species. The scale represented an average of 0.05 substitutions per amino acid site rate, each node labelled with its STAG support value percentage (summarise the fraction of trees made according to each orthogroup that support each bipartition).

### *3.6. Comparative genomics analysis between H10 and H13*

A side-by-side proteome comparison showed that the unmatched regions (Fig. 2A) were either low homology (corresponding to uncoloured regions of the green outer ring) or gene absence (the three regions described above, mostly specific to H10). Poorly homologous regions were mainly associated with sugar transport (PTS system and PTS associated protein) and, surprisingly, two regions (343–376 Kbp and 1260–1295 Kbp) contained CDS encoding for hypothetical proteins and genes found in phages. The three genomes were submitted to PHASTER screening for phage identification. All three strains exhibited an incomplete phage insertion corresponding to the 343–376 Kbp region; H10 had fewer CDS with 9 associated either with the phage (7 CDS) or hypothetical proteins (2 CDS) accounting for 9 Kbp. Interestingly, both ATCC51366 and H13 had a second phage insertion at the 1260–1295 Kbp region, classed as a 'questionable' phage sequence (90 scores for H13). In H13, this sequence of approximately 41.3 Kbp included a mosaic of 57 CDS (phage and hypothetical proteins). Among the phage-associated proteins, 23 phage species were found, the most abundant of which was phiCD38–2 with 9 CDS (clustered in 1–4-4). Phage phiCD38–2 matched with two other *Clostridium* phages, phiCD146–2 and phiCD111 (Table 3). To gain more insight into the relationship of these 9 CDS and their associated assignation, a linear genome comparison was performed (**Fig. S2**) resulting in a non-strain specificity. As only phiCD38–2 was already described in the literature, we used this assignation.

Due to the downregulation of the Staphylococcal *agr* system observed in our previous study [6], a pattern research was performed on both strains for AIP motif. This analysis was performed *in fine* with 4 minimal patterns leading to a match. They were for: type I autoinducer YSTC (Tyrosine-Serine-Threonin-Cystein) for the tail part, CDFIM (Cysteine-Aspartic Acid-Phenylalanine-Isoleucine-Methionine) for the docking one; CSSLF (Cysteine-Serine-Serine-Leucine-Phenylalanine) for the docking part of the type II and YFIM (Tyrosine-Phenylalanine-Isoleucine-Methionine) for the docking part of type IV [15]. Cysteine was absent from this last motif, preventing closure of the cycle. This pattern was thus dropped. Finally, three potential targets for Ail, which were present in both H13 and H10 genomes, were identified: an ABC type 2 transporter, a ribosomal sub-unit B and an alcohol dehydrogenase (**Fig. S1. B**).

## 4. Discussion

Initially collected from a human wound in 1993 by Collins et al. [6], the archetype of *H. kunzii* genus was sequenced in 2012. Two decades later, new *H. kunzii* strains were identified from infected FUd, co-isolated with *S. aureus* [6]. Our previous study explored virulence of different *S. aureus* strains in combination, or not, with some of these *H. kunzii* strains using the in vivo nematode killing assay. We noted that *H. kunzii* strains could be classified as wide or narrow spectrum modulators. We highlighted that the group of *H. kunzii* represented by the H13 strain exhibited an inhibitory activity of the virulence of various tested *S. aureus* strains. In contrast, another group of *H. kunzii* represented by the H10 strain only showed activity against colonising and avirulent strains [6]. In parallel, the ATCC51366 reference strain was sequenced (SAMN02596728), but its virulence potential remains unknown. In this study, the relationship between *H. kunzii* clinical isolates (H10 and H13 strains) and the ATCC51366 strain was highlighted through genomic comparison, functional annotation, and phylogenetic approach. Furthermore, a focus was made on H10 and H13 strains as potential targets mediating virulence attenuation of *S. aureus* by motif search and characterization of genomic prediction ability (particularly related to iron-metabolism).

In an initial general comparison, we observed that H10 and H13 genomes showed high identity and similar genome sizes. H13 and ATCC51366 strains showed the closest total length (2,062,593 bp and 2,083,191 bp, respectively) and gene number (1884 and 1888, respectively) (Table 2), whereas H10 strain showed the largest evolutionary distance within this clade (Fig. 2). One of the limits of this study was the absence of complete rRNA operon for the 16S and the 23S in final polished genomes. Those sequences have been found in the raw reads (BLASTn of *H. kunzii* ATCC51366 rRNAs against raw reads) but were not

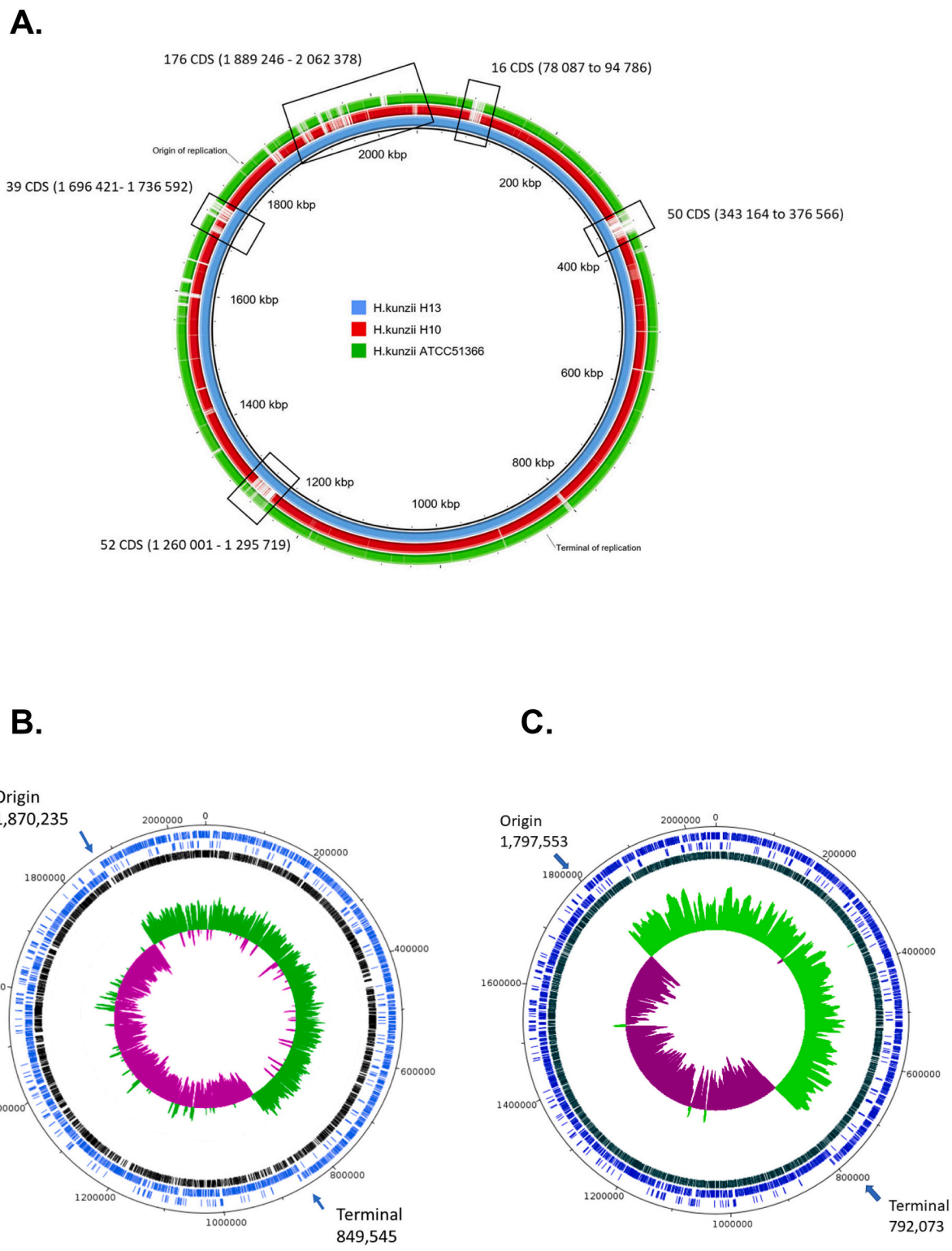


**Fig. 3.** *Helcoccus kunzii* genomes analysis. **A**: *Helcococcus kunzii* strains comparative analysis. *H. kunzii* strains comparative analysis was done using BRIG software [20]. The inner blue ring represents the H13 genome; the middle red one shows the H10 genome after BLASTn match; while the outer green ring corresponds to ATCC51366 (1st scaffold). Only regions with >90% nucleotide identity are coloured. Lower identity percentage or no match are represented by blank spaces in each ring. The black boxes indicate unmatched regions equal or superior to 15 Kbp. B, C: Circularised maps of *H. kunzii* H13 (B) and H10 (C) strains. The graph is composed of 4 concentric rings. CDS are mapped in blue segments on the two outer rings. The first outer ring corresponds to the positive strand and the second ring to the negative strand. All assigned CDS were mapped through the inner ring and coloured in black. The inner bicolor ring corresponds to the GC skew of *H. kunzii* isolates. Green profile indicates overabundant of GC nucleotides whereas purple corresponds to the opposite. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

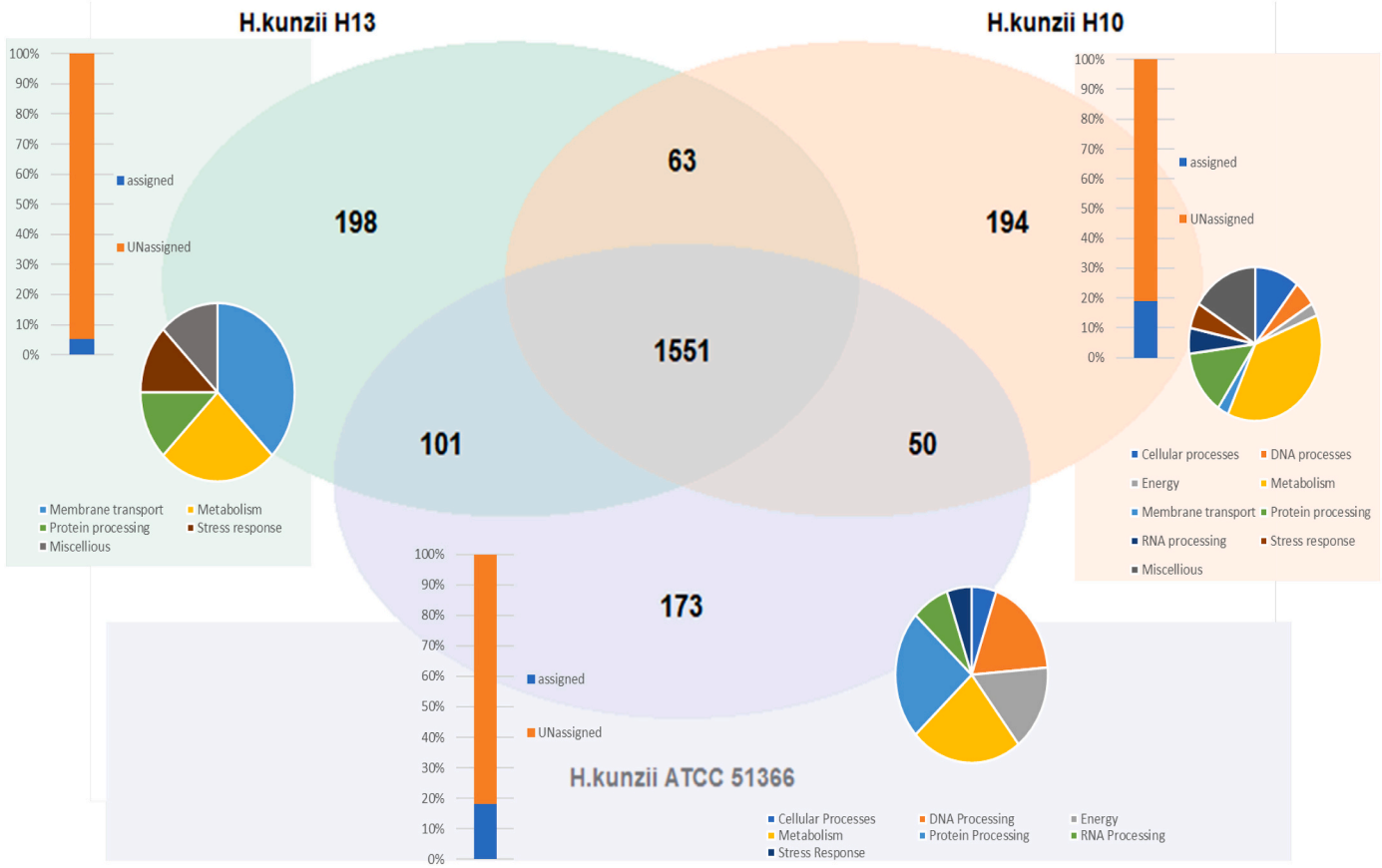


**Fig. 4.** Venn diagram of the three *Helcococcus kunzii* isolates. Venn diagram was built to compare the core proteome between all *H. kunzii* strains using R VennDiagram package. Subsystem repartitions of strain specific assigned proteins were determined using Patric software.

implemented during SPAdes assembly probably due to both i) highly repeated sequences of these genomic portions, and ii) the length of reads obtained through Illumina sequencing [38–40].

The close resemblance between the three genomes was further explored in BLAST analysis of ATCC51366 and H10 against H13 as reference genome. An almost complete coverage of H13 by ATCC51366 and, to a lesser extent, H10 was observed (Fig. 3A). Lower identity patches of ATCC51366 were mainly shared with the H10 strain and could have arisen from evolutionary process. Interestingly, the different genomes presented a considerable arsenal of resistance genes but very few virulence factor-encoding genes, supporting the commensal role of this species and their absence of clear virulence [14]. As *H. kunzii* forms part of the human skin microbiota, notably of the lower extremities [7,8], and was mainly isolated from infected FUd, its high resistance profile (**Table S7**) could be due to the difficulties in diagnosing infection of this chronic wound and the currently inappropriate use of antibiotics, involving the emergence of multidrug resistant bacteria [10].

Although *H. kunzii* strains display a commensal profile, our team has previously observed that some members of this species could modulate the virulence of a pathogen commonly isolated from infected foot ulcers, *S. aureus* [14]. This inhibition potential was highlighted through an interference with the *agr* system of *S. aureus* [14]. We hypothesised that a secreted protein produced by *H. kunzii* could act as a possible antagonist for AgrC. This predicted protein identified by motif search is expected to be secreted and could possess the two structural characteristics of autoinducing peptides (AIPs): the hydrophobic cycle (responsible for the receptor docking) and the 2 to 4 amino acid tail (responsible for the receptor activation) [31]. These Ail are signalling molecules that inhibit quorum sensing and virulence gene expression and act as AgrC-AIP interaction inhibitors [15]. Thus, the H10 and H13 strains could have the potential to decrease *S. aureus* virulence by secreting an Ail protein. We investigated the presence of an AIP-like motif in the H10 and H13 genomes and identified three candidate proteins: Alcohol dehydrogenase 2 (Adh2), Ribosomal sub-unit B (RsbB) and ABC type 2 family (ABC-2fam). As the Adh2 met the two criteria (being predicted as a secreted protein and having the same docking sequence as AIP (str. I)), we suggest that this protein is the better candidate to explain the modulation of *S. aureus* virulence. RsbB and ABC-2fam only presented homologue sequences of the AIP tail.

One of the main differences between H10 and H13 genomes was the presence of phiCD38–2-like prophage, exclusively present in the H13 strain. This prophage, a temperate phage of the *Siphoviridae* family, has already been described as stimulating bacterial secretion by up-regulation and affecting the virulence. For example, in a *Clostridium difficile* NAP1/027 model, this phage up-regulates the PaLoc (Pathogenicity Locus) island resulting in an overproduction and secretion of the two toxins, TcdA and TcdB [41]. We could hypothesise that this prophage positively affects the production of the Ail protein in H13 strain by an up-regulation of *adh2* gene and an interaction with AgrC of *S. aureus*. This could explain the wide spectrum of inhibition presented by the H13 strain. Conversely, in the H10 strain, the absence of this phage would not affect the Ail production, which would be at a basal level, explaining the reduced impact of the H10 strain largely affecting colonising bacterial species (and marginally the virulent ones).

Another hypothesis explaining the inhibitory effect of some *H. kunzii* on *S. aureus* virulence could be the iron availability modulation. Iron is essential for bacterial growth and plays a vital role in bacterial infection, acting as a virulence-associated cofactor [42,43]. FetA and FetB increase resistance to oxidative stress in the presence of iron, modulating iron homeostasis in *E. coli* [44]. The *E. coli* FetA protein (NP_415023.1)

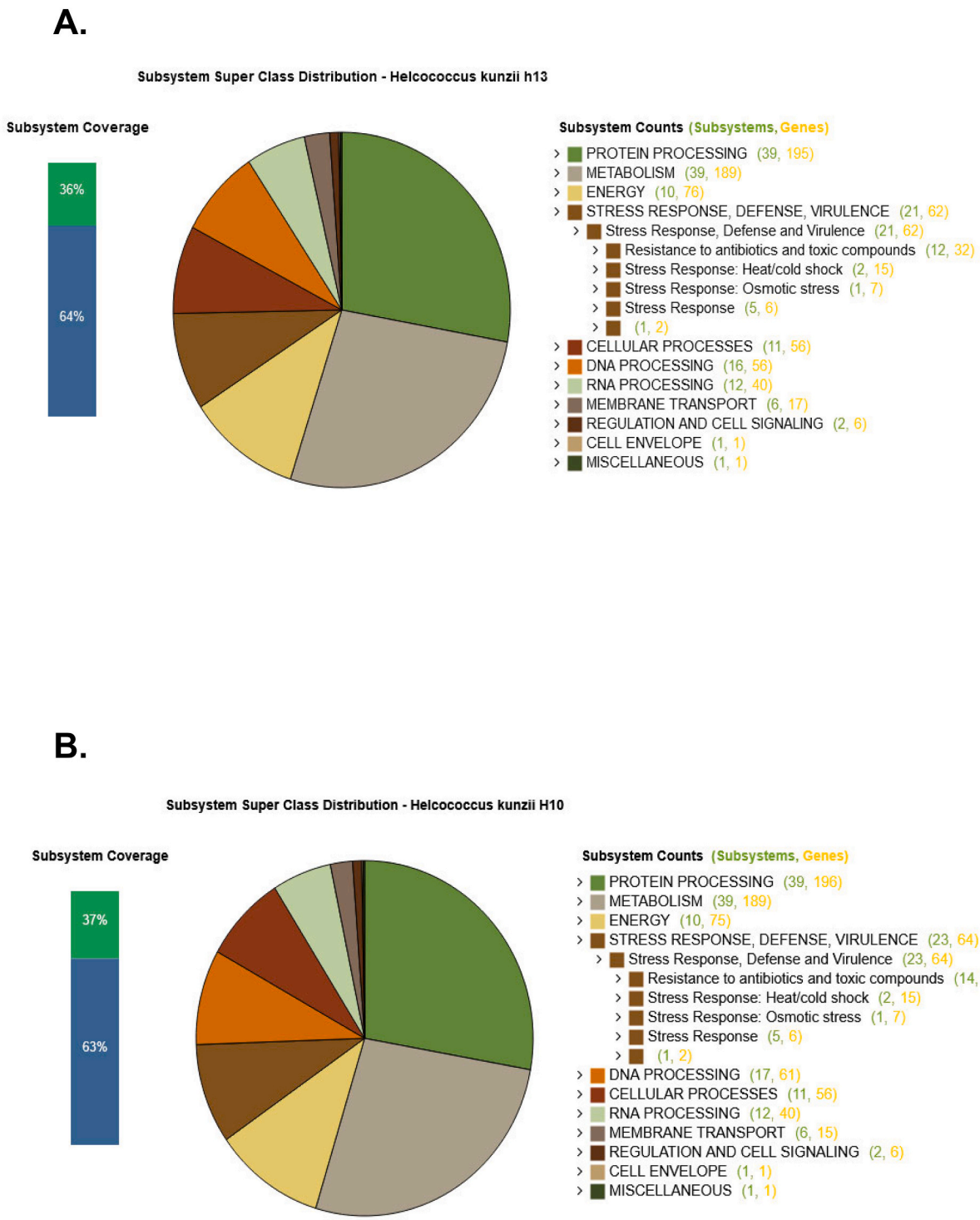


**Fig. 5.** Subsystems distribution and counts for (**A**) H13 and (**B**) H10. The coverage represented in green is the percentage of the genomic features assigned to a subsystem, whilst blue is the portion not assigned, includeing hypothetical proteins. The breakdown of the super class "stress response, defence, virulence" allows visualisation of gene counts related to its various subsystems. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

presents 44% identity with H13 upon BLASTp and 43% identity with H10. We could hypothesise that FetA and FetB membrane transporters from the H10 strain support *S. aureus* virulence through iron release. In the H13 strain, the export system functionality could be altered and thus induces iron sequestration either by a non-functional protein (in regard to the FetA point mutations towards H10), or by a membrane transporter negative regulation. All these hypotheses must be further investigated.

The lower phylogenetic distance and identity coverage would indicate an equivalent virulence modulation between H10 and ATCC51366. Our genomic study alone cannot explain the underlying mechanism. The *H. kunzii*-dependant virulence inhibitor phenotype observed by Ngba Essebe et al. [14] should be explored through a transcriptomic and/or proteomic approach considering the large proportion of hypothetical proteins found in these genomes, and a possible threshold effect in the

**Table 3**
Phaster (A) summary and identification (B) of CDS present in the low/absence homology area of H13.

**A.**

| Region | Length (Kb) | Completness (score) | Specific keyword | Position | Total protein number | Phage hit protein number | Hypothetical protein number | *Att* site | Phage species number | Most common phage(hit_genes_count) | first most common phage | First most common phage percentage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 33 | Questionable 80 | integrase, portal, head, capsid, tail | 342975-376012 | 40 | 24 | 16 | yes | 13 | PHAGE_Bacill_BCJA1c_NC_006557(6), PHAGE_Entero_EFC_1_NC_025453(6), PHAGE_Paenib_Vegas_NC_028767(6), PHAGE_Lactob_iA2_NC_028830(5), PHAGE_Lactob_PLE2_NC_031036(5), PHAGE_Bacill_vB_BhaS_171_NC_030904 (4), PHAGE_Clostr_phiCD506_NC_028838(3), PHAGE_Paenib_PG1_NC_021558(3), PHAGE_Staphy_phiPV83_NC_002486(2), PHAGE_Clostr_phiMMP04_NC_019422(2), PHAGE_Lactoc_PLgT_1_NC_031016(2), PHAGE_Clostr_phiCD481_1_NC_028951(2), PHAGE_Lactob_Lj965_NC_005355(1), PHAGE_Lactoc_1358_NC_027120(1), PHAGE_Staphy_Ipla35_NC_011612(1), PHAGE_Weisse_WCP30_NC_031101(1), PHAGE_Staphy_tp310_3_NC_009763(1), PHAGE_Staphy_JS01_NC_021773(1), PHAGE_Arthro_vB_ArtM_ArV1_NC_026606 (1), PHAGE_Lactoc_ul36_NC_004066(1), PHAGE_Actino_phiAsp2_NC_005885(1), PHAGE_Strept_SpSL1_NC_027396(1), PHAGE_Bdello_phi1422_NC_019525(1), PHAGE_Thermu_OH2_NC_021784(1), PHAGE_Bacill_phBC6A51_NC_004820(1), PHAGE_Lister_P35_NC_009814(1), PHAGE_Lactob_Lj928_NC_005354(1), PHAGE_Strept_315.2_NC_004585(1), PHAGE_Staphy_phi5967PVL_NC_019921 (1), PHAGE_Clostr_phiCDHM11_NC_029001 (1), PHAGE_Staphy_phiPVL_CN125_NC_012784 (1) ,PHAGE_Lactoc_KSY1_NC_009817(1), PHAGE_Strept_YDN12_NC_028974(1), PHAGE_Strept_T12_NC_028700(1), PHAGE_Lactob_LfeSau_NC_029068(1), PHAGE_Strept_phiARI0923_NC_030946(1), PHAGE_Bacill_BM5_NC_029069(1), PHAGE_Staphy_vB_SauS_phi2_NC_028862 (1), PHAGE_Rhizob_16_3_NC_011103(1), PHAGE_Lactoc_Tuc2009_NC_002703(1), PHAGE_Bacill_1_NC_009737(1), PHAGE_Lister_LP_101_NC_024387(1), PHAGE_Entero_phiFL4A_NC_013644(1), PHAGE_Arthro_PrincessTrina_NC_042053 (1), PHAGE_Staphy_tp310_1_NC_009761(1), PHAGE_Thermo_THSA_485A_NC_018264 (1), PHAGE_Lactoc_28201_NC_031013(1), PHAGE_Lister_P40_NC_011308(1), PHAGE_Idioma_Phi1M2_2_NC_025471(1), PHAGE_Geobac_GBSV1_NC_008376(1) | 6 | 15% |
| 2 | 41.3 | Questionable 90 | tail, capsid, integrase | 1256729-1298127 | 57 | 35 | 22 | yes | 23 | PHAGE_Clostr_phiCD38_2_NC_015568(9), PHAGE_Clostr_phiCD111_NC_028905(9), PHAGE_Clostr_phiCD146_NC_028958(9) ,PHAGE_Entero_EFC_1_NC_025453(6), PHAGE_Paenib_Vegas_NC_028767(6), PHAGE_Bacill_BCJA1c_NC_006557(6), PHAGE_Lister_vB_LmoS_188_NC_028871 (3), PHAGE_Clostr_phiCTP1_NC_014457(3), PHAGE_Strept_20617_NC_023503(2), PHAGE_Bacill_PBC1_NC_017976(2), PHAGE_Bacill_IEBH_NC_011167(2), | 9 | 15.78% |

(*continued on next page*)

**Table 3** (*continued*)

**A.**

| Region | Length (Kb) | Completeness (score) | Specific keyword | Position | Total protein number | Phage hit protein number | Hypothetical protein number | *Att* site | Phage species number | Most common phage(hit_genes_count) | first most common phage | First most common phage percentage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | PHAGE_Strept_EJ_1_NC_005294(2), PHAGE_Lister_A006_NC_009815(2), PHAGE_Staphy_55_NC_007060(2), PHAGE_Bacill_BM5_NC_029069(2), PHAGE_Lactoc_Tuc2009_NC_002703(2), PHAGE_Clostr_phiSM101_NC_008265(2), PHAGE_Strept_Dp_1_NC_015274(2), PHAGE_Staphy_CNPH82_NC_008722(2), PHAGE_Lister_B054_NC_009813(2), PHAGE_Clostr_phiC2_NC_009231(1), PHAGE_Lactob_phig1e_NC_004305(1), PHAGE_Lister_A118_NC_003216(1), PHAGE_Pseudo_phiPsa374_NC_023601(1), PHAGE_Strept_MM1_NC_003050(1), PHAGE_Strept_SM1_NC_004996(1), PHAGE_Bacill_phi4J1_NC_029008(1), PHAGE_Arthro_vB_ArtM_ArV1_NC_026606(1), PHAGE_Strept_phiARI0131_1_NC_031901(1), PHAGE_Staphy_StB12_NC_020490(1), PHAGE_Bdello_phi1422_NC_019525(1), PHAGE_Entero_vB_IME197_NC_028671(1), PHAGE_Lactoc_PLgT_1_NC_031016(1), PHAGE_Strept_phiNJ2_NC_019418(1), PHAGE_Bacill_Bobb_NC_024792(1), PHAGE_Strept_Abc2_NC_013645(1), PHAGE_Strept_PH10_NC_012756(1), PHAGE_Geobac_TP_84_NC_041918(1), PHAGE_Lactob_LL_H_NC_009554(1), PHAGE_Bacill_1_NC_009737(1), PHAGE_Staphy_Ipla5_NC_018281(1), PHAGE_Arthro_PrincessTrina_NC_042053(1), PHAGE_Cellul_phi14:2_NC_021806(1), PHAGE_Strept_P9_NC_009819(1), PHAGE_Lactob_phiJB_NC_022775(1), PHAGE_Entero_phiFL1A_NC_013646(1), PHAGE_Staphy_tp310_2_NC_009762(1), PHAGE_Lactoc_TP901_1_NC_002747(1), PHAGE_Lister_LP_030_3_NC_024384(1), PHAGE_Bacill_Aurora_NC_031121(1), PHAGE_Strept_315.5_NC_004588(1), PHAGE_Lactob_jlb1_NC_024206(1), PHAGE_Actino_phiAsp2_NC_005885(1), PHAGE_Entero_EFLK1_NC_029026(1), PHAGE_Entero_phiEF24C_NC_009904(1), PHAGE_Clostr_phiMMP02_NC_019421(1), PHAGE_Roseob_2_NC_041958(1), PHAGE_Entero_ECP3_NC_027335(1), PHAGE_Clostr_phiCD505_NC_028764(1), PHAGE_Bacill_AR9_NC_031039(1), PHAGE_Cellul_phi38:1_NC_021796(1), PHAGE_Lactoc_KSY1_NC_009817(1), PHAGE_Strept_YDN12_NC_028974(1), PHAGE_Lactob_LfeSau_NC_029068(1), PHAGE_Erysip_SE_1_NC_029078(1), PHAGE_Halocy_JM_2012_NC_017975(1), PHAGE_Clostr_phiCT19406C_NC_029006(1), PHAGE_Strept_PH15_NC_010945(1), PHAGE_Strept_A25_NC_028697(1), PHAGE_Bacill_SPP1_NC_004166(1), PHAGE_Rhizob_16_3_NC_011103(1), PHAGE_Roseob_1_NC_015466(1), PHAGE_Staphy_Twort_NC_007021(1), PHAGE_Entero_phiFL2A_NC_013643(1), PHAGE_Bacill_0305phi8_36_NC_009760(1), PHAGE_Clostr_phiCT9441A_NC_029022(1), | | |

(*continued on next page*)

**Table 3** (*continued*)

A.

| Region | Length (Kb) | Completness (score) | Specific keyword | Position | Total protein number | Phage hit protein number | Hypothetical protein number | *Att* site | Phage species number | Most common phage(hit_genes_count) | first most common phage | First most common phage percentage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | PHAGE_Bacill_250_NC_029024(1), PHAGE_Strept_phiBHN167_NC_022791(1), PHAGE_Staphy_StauST398_3_NC_021332(1), PHAGE_Thermo_THSA_485A_NC_018264(1), PHAGE_Bacill_PBS1_NC_043027(1), PHAGE_Clostr_phiMMP01_NC_028883(1), PHAGE_Strept_315.6_NC_004589(1), PHAGE_Idioma_Phi1M2_2_NC_025471(1), PHAGE_Bacill_phBC6A52_NC_004821(1), PHAGE_Bacill_phiAGATE_NC_020081(1), PHAGE_Staphy_StauST398_5_NC_023500(1) | | |

B.

| Position | Assignation | Function | CDS number | Corresponding phage protein | e value |
|---|---|---|---|---|---|
| complement (1256729..1257454) | PHAGE_Roseob_2_NC_041958 | hypothetical protein | PP_01167 | phage(gi100032) | 2.46e-09 |
| complement (1257694..1259322) | PHAGE_Cellul_phi142_NC_021806 | chaperonin GroEL | PP_01168 | phage(gi526178751) | 3.53e-100 |
| complement (1259339..1259611) | PHAGE_Bacill_0305phi8_36_NC_009760 | GroES | PP_01169 | phage(gi156564025) | 9.85e-07 |
| complement (1260001..1260192) | hypothetical | | PP_01170 | | N/A |
| complement (1260202..1261248) | PHAGE_Bacill_Aurora_NC_031121 | replication and recombination DNA helicase | PP_01171 | phage(gi100033) | 1.58e-11 |
| 1261110..1261121 | *attL* | | | N/A | |
| complement (1261361..1261687) | hypothetical | | PP_01172 | N/A | |
| complement (1261702..1261920) | hypothetical | | PP_01173 | N/A | |
| complement (1261930..1263933) | PHAGE_Strept_Abc2_NC_013645 | tail protein | PP_01174 | phage(gi281416402) | 2.16e-19 |
| complement (1263933..1265942) | PHAGE_Strept_20617_NC_023503 | phage capsid and scaffold protein | PP_01175 | phage(gi588295130) | 6.39e-17 |
| complement (1265944..1266627) | PHAGE_Clostr_phiSM101_NC_008265 | hypothetical protein | PP_01176 | phage(gi110804045) | 5.30e-06 |
| complement (1266627..1269578) | PHAGE_Strept_Dp_1_NC_015274 | TMP | PP_01177 | phage(gi327198366) | 1.26e-168 |
| complement (1269632..1269793) | hypothetical | | PP_01178 | N/A | |
| complement (1269765..1270325) | PHAGE_Clostr_phiCD146_NC_028958 | hypothetical protein | PP_01179 | phage(gi971820169) | 4.53e-34 |
| complement (1270329..1270742) | hypothetical | | PP_01180 | N/A | |
| complement (1270790..1271320) | hypothetical | | PP_01181 | N/A | |
| complement (1271322..1271738) | PHAGE_Clostr_phiCD146_NC_028958 | hypothetical protein | PP_01182 | phage(gi971820166) | 1.19e-20 |
| complement (1271735..1272118) | PHAGE_Clostr_phiCD146_NC_028958 | hypothetical protein | PP_01183 | phage(gi971820165) | 1.96e-21 |
| complement (1272111..1272509) | PHAGE_Clostr_phiCD146_NC_028958 | hypothetical protein | PP_01184 | phage(gi971820164) | 1.64e-13 |
| complement (1272493..1272900) | PHAGE_Clostr_phiCD146_NC_028958 | hypothetical protein | PP_01185 | phage(gi971820163) | 2.60e-18 |
| complement (1272911..1273177) | hypothetical | | PP_01186 | N/A | |
| complement (1273226..1274092) | PHAGE_Lactob_LfeSau_NC_029068 | capsid protein | PP_01187 | phage(gi985757732) | 3.13e-59 |
| complement (1274107..1274715) | PHAGE_Bacill_SPP1_NC_004166 | hypothetical protein | PP_01188 | phage(gi22855064) | 1.76e-18 |
| complement (1275062..1275274) | PHAGE_Strept_EJ_1_NC_005294 | hypothetical protein | PP_01189 | phage(gi39653716) | 1.21e-16 |
| complement (1275264..1276859) | PHAGE_Clostr_phiCD146_NC_028958 | Minor capsid protein | PP_01190 | phage(gi971820159) | 1.49e-52 |
| | PHAGE_Clostr_phiCD146_NC_028958 | Minor capsid protein | PP_01191 | phage(gi971820158) | |

(*continued on next page*)

**Table 3** (*continued*)

| B. Position | Assignation | Function | CDS number | Corresponding phage protein | e value |
|---|---|---|---|---|---|
| complement (1276852..1278213) | | | | | 8.32e-126 |
| complement (1278259..1279575) | PHAGE_Clostr_phiCD146_NC_028958 | Terminase large subunit | PP_01192 | phage(gi971820157) | 6.10e-169 |
| complement (1279559..1280239) | PHAGE_Clostr_phiCD146_NC_028958 | Terminase small subunit | PP_01193 | phage(gi971820156) | 3.55e-39 |
| complement (1280392..1280832) | hypothetical | | PP_01194 | N/A | |
| complement (1280842..1281006) | hypothetical | | PP_01195 | N/A | |
| complement (1281082..1281273) | hypothetical | | PP_01196 | N/A | |
| complement (1281336..1281599) | hypothetical | | PP_01197 | N/A | |
| complement (1281574..1281786) | hypothetical | | PP_01198 | N/A | |
| complement (1281783..1282505) | hypothetical | | PP_01199 | N/A | |
| complement (1282732..1283133) | PHAGE_Entero_phiFL2A_NC_013643 | YopX superfamily protein | PP_01200 | phage(gi281416516) | 9.34e-18 |
| complement (1283149..1283331) | PHAGE_Entero_EFLK1_NC_029026 | hypothetical protein | PP_01201 | phage(gi971766788) | 1.76e-09 |
| complement (1283392..1283826) | PHAGE_Pseudo_phiPsa374_NC_023601 | hypothetical protein | PP_01202 | phage(gi589287223) | 3.10e-05 |
| complement (1283842..1284042) | hypothetical | | PP_01203 | N/A | |
| complement (1284060..1284191) | hypothetical | | PP_01204 | N/A | |
| complement (1284255..1284425) | hypothetical | | PP_01205 | N/A | |
| complement (1284422..1284610) | hypothetical | | PP_01206 | N/A | |
| complement (1284610..1284903) | hypothetical | | PP_01207 | N/A | |
| complement (1285122..1285553) | PHAGE_Bacill_BCJA1c_NC_006557 | rus | PP_01208 | phage(gi56694892) | 5.17e-35 |
| complement (1285853..1288084) | PHAGE_Paenib_Vegas_NC_028767 | DNA primase | PP_01209 | phage(gi971741650) | 0.0 |
| complement (1288074..1289672) | PHAGE_Entero_EFC_1_NC_025453 | DEAD/DEAH box helicase | PP_01210 | phage(gi725950530) | 0.0 |
| complement (1289672..1290097) | PHAGE_Bacill_BCJA1c_NC_006557 | hypothetical protein | PP_01211 | phage(gi56694883) | 1.69e-26 |
| complement (1290110..1291249) | PHAGE_Paenib_Vegas_NC_028767 | AAA domain ATPase | PP_01212 | phage(gi971741646) | 1.90e-98 |
| complement (1291249..1292577) | PHAGE_Paenib_Vegas_NC_028767 | chromosome segregation protein SMC | PP_01213 | phage(gi971741645) | 6.55e-163 |
| complement (1292588..1292788) | hypothetical | | PP_01214 | N/A | |
| complement (1292943..1293071) | PHAGE_Strept_EJ_1_NC_005294 | hypothetical protein | PP_01215 | phage(gi39653688) | 1.21e-10 |
| complement (1293075..1293236) | hypothetical | | PP_01216 | N/A | |
| complement (1293246..1293527) | hypothetical | | PP_01217 | N/A | |
| complement (1293566..1293760) | PHAGE_Staphy_CNPH82_NC_008722 | putative cro-like repressor | PP_01218 | phage(gi119953714) | 1.49e-07 |
| 1293925..1294314 | PHAGE_Staphy_55_NC_007060 | ORF044 | PP_01219 | phage(gi66396156) | 4.22e-10 |
| 1294314..1294739 | PHAGE_Thermo_THSA_485A_NC_018264 | protein of unknown function DUF955 | PP_01220 | phage(gi397912661) | 1.39e-11 |
| 1294742..1295719 | hypothetical | | PP_01221 | N/A | |
| 1295538..1295549 | *attR* | | | N/A | |
| 1295856..1296929 | PHAGE_Strept_PH15_NC_010945 | putative integrase | PP_01222 | phage(gi190151416) | 4.88e-18 |
| 1297234..1298127 | PHAGE_Clostr_phiC2_NC_009231 | putative abortive infection bacteriophage resistance protein ORF 37 | PP_01223 | phage(gi134287370) | 2.62e-47 |

CDS belonging to the phage phiCD38–2 have been highlighted with grey shadow for legibility purpose.

gene expression. Regarding the phylogenetic relationship and the collection date (Table 2), it is possible that the narrowed spectrum of *S. aureus* virulence inhibition could have arisen from a co-evolution process as i) *H. kunzii* strains were mostly co-isolated with *S. aureus* and ii) the H10 strain shows the greatest evolutionary distance within this clade.

## 5. Conclusion

Following our previous studies, we were motivated to understand the differential action of *H. kunzii* on *S. aureus* virulence modulation [14]. This study focused on examining the genomic differences of two new clinical isolates, as previously performed for other genus [45]. This approach has been used to understand the genomic modulation of *Acinetobacter* sp. during its adaptation to several ecological niches [46,47]. Our study highlights the genomes of *H. kunzii* genus, through pioneering molecular and genomic study of the evolution and virulence potential of *H. kunzii*. As previously illustrated by Horswill et al., the *S. aureus* quorum sensing system represents a major therapeutic potential for antagonist molecules against pathogen virulence [15]. Based on sequence activity and on the model of pattern search [48], we identified the Adh2 protein, an AIP-like protein, as a candidate potentially able to modulate the virulence of *S. aureus*. This result could provide new promising approaches for the development of alternatives to conventional antibiotic treatment. Further in vitro functional studies are needed to confirm this bioinformatics work with transcriptomic or proteomic experiments.

## Availability of data and materials

Raw reads of both newly sequenced strains of *H. kunzii* have been deposited under the BioProject No. PRJNA600407. *H. kunzii* strains H13 and H10 are available under accession no. SAMN13819011 and SAMN14604363, respectively.

## Data availability statement

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary material Materials. Additional data related to this paper may be requested from the authors.

## Funding

This study was supported by the University Hospital of Nîmes. This University hospital provided its structural, human, and financial support through the award obtained by our team during the internal call for tenders “Thématiques phares”. B.A.R.N.D.’s PhD was funded by a grant from the University Hospital of Nîmes.

## CRediT authorship contribution statement

**Benjamin A.R.N. Durand:** Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft. **Alex Yahiaoui Martinez:** Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft. **Damien Baud:** Data curation, Investigation, Writing – review & editing. **Patrice François:** Data curation, Resources, Writing – review & editing. **Jean-Philippe Lavigne:** Conceptualization, Funding acquisition, Resources, Writing – review & editing. **Catherine Dunyach-Remy:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing.

## Declaration of Competing Interest

The authors declare no conflict of interest.

## Acknowledgements

We thank Sarah Kabani for her editing assistance. J.-P.L. and C.D.-R. belong to the FHU InCh (Federation Hospitalo Universitaire Infections Chroniques, Aviesan).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2022.110365.

## References

[1] P.C.Y. Woo, H. Tse, S.S.Y. Wong, C.W.S. Tse, A.M.Y. Fung, D.M.W. Tam, S.K.P. Lau, K. Yuen, Life-threatening invasive *Helcococcus kunzii* infections in intravenous-drug users and *ermA*-mediated erythromycin resistance, J. Clin. Microbiol. 43 (2005) 6205–6208, https://doi.org/10.1128/JCM.43.12.6205-6208.2005.

[2] S. McNicholas, B. McAdam, M. Flynn, H. Humphreys, The challenges of implantable cardiac device infection due to *Helcococcus kunzii*, J. Hosp. Infect. 78 (2011) 337–338, https://doi.org/10.1016/j.jhin.2011.04.010.

[3] C. Pérez-Jorge, J. Cordero, M. Marin, J. Esteban, Prosthetic joint infection caused by *Helcococcus kunzii*, J. Clin. Microbiol. 50 (2012) 528–530, https://doi.org/10.1128/JCM.01244-11.

[4] S. Sridhar, J.F.W. Chan, K.-Y. Yuen, First report of brain abscess caused by a satelliting phenotypic variant of *Helcococcus kunzii*, J. Clin. Microbiol. 52 (2014) 370–373, https://doi.org/10.1128/JCM.02550-13.

[5] J.H. Park, B.-M. Woo, S.K. Hong, E.-C. Kim, First Korean case of *Helcococcus kunzii* bacteremia in a patient with diabetes, Ann. Lab. Med. 34 (2014) 484–486, https://doi.org/10.3343/alm.2014.34.6.484.

[6] M.D. Collins, R.R. Facklam, U.M. Rodrigues, K.L. Ruoff, Phylogenetic analysis of some *Aerococcus*-like organisms from clinical sources: description of *Helcococcus kunzii* gen. nov., sp. nov, Int. J. Syst. Evol. Microbiol. 43 (1993) 425–429, https://doi.org/10.1099/00207713-43-3-425.

[7] J. Haas, S.L. Jernick, R.J. Scardina, J. Teruya, A.M. Caliendo, K.L. Ruoff, Colonization of skin by *Helcococcus kunzii*, J. Clin. Microbiol. 35 (1997) 2759–2761. http://jcm.asm.org/content/35/11/2759.

[8] A. Vergne, F. Guérin, R. Lienhard, A. Le Coustumier, C. Daurel, C. Isnard, N. Marty, C. Poyart, V. Cattoir, In vitro antimicrobial susceptibility of *Helcococcus kunzii* and molecular analysis of macrolide and tetracycline resistance, Eur. J. Clin. Microbiol. Infect. Dis. 34 (2015) 2057–2061, https://doi.org/10.1007/s10096-015-2451-5.

[9] D.G. Armstrong, A.J.M. Boulton, S.A. Bus, Diabetic foot ulcers and their recurrence, N. Engl. J. Med. 376 (2017) 2367–2375, https://doi.org/10.1056/NEJMra1615439.

[10] B.A. Lipsky, E. Senneville, Z.G. Abbas, J. Aragón-Sánchez, M. Diggle, J.M. Embil, S. Kono, L.A. Lavery, M. Malone, S.A. van Asten, V. Urbančič-Rovan, E.J.G. Peters, International Working Group on the Diabetic Foot (IWGDF), Guidelines on the diagnosis and treatment of foot infection in persons with diabetes (IWGDF 2019 update), Diabetes Metab. Res. Rev. 36 (2020), https://doi.org/10.1002/dmrr.3280 e3280.

[11] N. Tentolouris, Introduction, in: Atlas Diabet. Foot, John Wiley & Sons, Ltd, 2019, pp. 1–10, https://doi.org/10.1002/9781119255314.ch1.

[12] C. Dunyach-Remy, C. Ngba Essebe, A. Sotto, J.-P. Lavigne, *Staphylococcus aureus* toxins and diabetic foot ulcers: role in pathogenesis and interest in diagnosis, Toxins. 8 (2016) 209, https://doi.org/10.3390/toxins8070209.

[13] M. Loesche, S.E. Gardner, L. Kalan, J. Horwinski, Q. Zheng, B.P. Hodkinson, A. S. Tyldsley, C.L. Franciscus, S.L. Hillis, S. Mehta, D.J. Margolis, E.A. Grice, Temporal stability in chronic wound microbiota is associated with poor healing, J. Invest. Dermatol. 137 (2017) 237–244, https://doi.org/10.1016/j.jid.2016.08.009.

[14] C. Ngba Essebe, O. Visvikis, M. Fines-Guyon, A. Vergne, V. Cattoir, A. Lecoustumier, E. Lemichez, A. Sotto, J.-P. Lavigne, C. Dunyach-Remy, Decrease of *Staphylococcus aureus* virulence by *Helcococcus kunzii* in a *Caenorhabditis elegans* model, Front. Cell. Infect. Microbiol. 7 (2017), https://doi.org/10.3389/fcimb.2017.00077.

[15] A.R. Horswill, C.P. Gordon, Structure–activity relationship studies of small molecule modulators of the staphylococcal accessory gene regulator, J. Med. Chem. 63 (2020) 2705–2730, https://doi.org/10.1021/acs.jmedchem.9b00798.

[16] J.O. Lay, R.D. Holland, Rapid identification of bacteria based on spectral patterns using MALDI-TOF MS, in: J.R. Chapman (Ed.), Mass Spectrom. Proteins Pept. Mass Spectrom. Proteins Pept, Humana Press, Totowa, NJ, 2000, pp. 461–487.

[17] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, Nucleic Acids Res. 44 (2016) 6614–6624, https://doi.org/10.1093/nar/gkw569.

[18] D.H. Haft, M. DiCuccio, A. Badretdin, V. Brover, V. Chetvernin, K. O’Neill, W. Li, F. Chitsaz, M.K. Derbyshire, N.R. Gonzales, M. Gwadz, F. Lu, G.H. Marchler, J. S. Song, N. Thanki, R.A. Yamashita, C. Zheng, F. Thibaud-Nissen, L.Y. Geer, A. Marchler-Bauer, K.D. Pruitt, RefSeq: an update on prokaryotic genome annotation and curation, Nucleic Acids Res. 46 (2018) D851–D860, https://doi.org/10.1093/nar/gkx1068.

[19] S. Ciufo, S. Kannan, S. Sharma, A. Badretdin, K. Clark, S. Turner, S. Brover, C. L. Schoch, A. Kimchi, M. DiCuccio, Using average nucleotide identity to improve

taxonomic assignments in prokaryotic genomes at the NCBI, Int. J. Syst. Evol. Microbiol. 68 (2018) 2386–2392.
[20] N.-F. Alikhan, N.K. Petty, N.L. Ben Zakour, S.A. Beatson, BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons, BMC Genomics 12 (2011) 402, https://doi.org/10.1186/1471-2164-12-402.
[21] A. Carattoli, E. Zankari, A. García-Fernández, M. Voldby Larsen, O. Lund, L. Villa, F. Møller Aarestrup, H. Hasman, In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing, Antimicrob. Agents Chemother. 58 (2014) 3895–3903, https://doi.org/10.1128/AAC.02412-14.
[22] T. Carver, N. Thomson, A. Bleasby, M. Berriman, J. Parkhill, DNAPlotter: circular and linear interactive genome visualization, Bioinformatics. 25 (2009) 119–120, https://doi.org/10.1093/bioinformatics/btn578.
[23] M. Richter, R. Rosselló-Móra, F. Oliver Glöckner, J. Peplies, JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison, Bioinformatics. 32 (2016) 929–931, https://doi.org/10.1093/bioinformatics/btv681.
[24] M. Richter, R. Rosselló-Móra, Shifting the genomic gold standard for the prokaryotic species definition, Proc. Natl. Acad. Sci. 106 (2009) 19126–19131, https://doi.org/10.1073/pnas.0906412106.
[25] A.F. Auch, M. von Jan, H.-P. Klenk, M. Göker, Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison, Stand. Genomic Sci. 2 (2010) 117, https://doi.org/10.4056/sigs.531120.
[26] J.P. Meier-Kolthoff, A.F. Auch, H.-P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, BMC Bioinformatics. 14 (2013) 60, https://doi.org/10.1186/1471-2105-14-60.
[27] J.P. Meier-Kolthoff, R.L. Hahnke, J. Petersen, C. Scheuner, V. Michael, A. Fiebig, C. Rohde, M. Rohde, B. Fartmann, L.A. Goodwin, O. Chertkov, T. Reddy, A. Pati, N. N. Ivanova, V. Markowitz, N.C. Kyrpides, T. Woyke, M. Göker, H.-P. Klenk, Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy, Stand. Genomic Sci. 9 (2014) 2, https://doi.org/10.1186/1944-3277-9-2.
[28] D.M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy, Genome Biol. 16 (2015) 157, https://doi.org/10.1186/s13059-015-0721-2.
[29] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (2019) 238, https://doi.org/10.1186/s13059-019-1832-y.
[30] T. Seemann, Snippy: Fast Bacterial Variant Calling from NGS Reads. https://github.com/tseemann/snippy, 2015.
[31] J.S. Wright, G.J. Lyon, E.A. George, T.W. Muir, R.P. Novick, Hydrophobic interactions drive ligand-receptor recognition for activation and inhibition of staphylococcal quorum sensing, Proc. Natl. Acad. Sci. 101 (2004) 16168–16173, https://doi.org/10.1073/pnas.0404039101.
[32] A.R. Wattam, J.J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, E. M. Dietrich, T. Disz, J.L. Gabbard, S. Gerdes, C.S. Henry, R.W. Kenyon, D. Machi, C. Mao, E.K. Nordberg, G.J. Olsen, D.E. Murphy-Olson, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, M. Shukla, V. Vonstein, A. Warren, F. Xia, H. Yoo, R. L. Stevens, Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center, Nucleic Acids Res. 45 (2017) D535–D542, https://doi.org/10.1093/nar/gkw1017.
[33] D. Arndt, J.R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, D.S. Wishart, PHASTER: a better, faster version of the PHAST phage search tool, Nucleic Acids Res. 44 (2016) W16–W21, https://doi.org/10.1093/nar/gkw387.
[34] M.J. Sullivan, N.K. Petty, S.A. Beatson, Easyfig: a genome comparison visualizer, Bioinformatics. 27 (2011) 1009–1010, https://doi.org/10.1093/bioinformatics/btr039.
[35] Z. Zhang, S. Schwartz, L. Wagner, W. Miller, A greedy algorithm for aligning DNA sequences, J. Comput. Biol. J. Comput. Mol. Cell Biol. 7 (2000) 203–214, https://doi.org/10.1089/10665270050081478.
[36] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402, https://doi.org/10.1093/nar/25.17.3389.
[37] D.L. Hava, A. Camilli, Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors, Mol. Microbiol. 45 (2002) 1389–1406.
[38] S.D. Brown, S.M. Utturkar, T.S. Magnuson, A.E. Ray, F.L. Poole, W.A. Lancaster, M. P. Thorgersen, M.W.W. Adams, D.A. Elias, Complete genome sequence of *Pelosinus* sp. strain UFO1 assembled using single-molecule real-time DNA sequencing technology, Genome Announc 2 (2014), https://doi.org/10.1128/genomeA.00881-14.
[39] S.M. Utturkar, D.M. Klingeman, M.L. Land, C.W. Schadt, M.J. Doktycz, D. A. Pelletier, S.D. Brown, Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences, Bioinformatics 30 (2014) 2709–2716, https://doi.org/10.1093/bioinformatics/btu391.
[40] S.L. Salzberg, A.M. Phillippy, A. Zimin, D. Puiu, T. Magoc, S. Koren, T.J. Treangen, M.C. Schatz, A.L. Delcher, M. Roberts, G. Marçais, M. Pop, J.A. Yorke, GAGE: a critical evaluation of genome assemblies and assembly algorithms, Genome Res. 22 (2012) 557–567, https://doi.org/10.1101/gr.131383.111.
[41] O. Sekulovic, M. Meessen-Pinard, L.-C. Fortier, Prophage-stimulated toxin production in *Clostridium difficile* NAP1/027 Lysogens, J. Bacteriol. 193 (2011) 2726–2734, https://doi.org/10.1128/JB.00787-10.
[42] M. Johnson, M. Sengupta, J. Purves, E. Tarrant, P.H. Williams, A. Cockayne, A. Muthaiyan, R. Stephenson, N. Ledala, B.J. Wilkinson, R.K. Jayaswal, J. A. Morrissey, Fur is required for the activation of virulence gene expression through the induction of the *sae* regulatory system in *Staphylococcus aureus*, Int. J. Med. Microbiol. 301 (2011) 44–52, https://doi.org/10.1016/j.ijmm.2010.05.003.
[43] S.M. Zughaier, P. Cornelis, Editorial: role of iron in bacterial pathogenesis, Front. Cell. Infect. Microbiol. 8 (2018) 344, https://doi.org/10.3389/fcimb.2018.00344.
[44] S.A. Nicolaou, A.G. Fast, E. Nakamaru-Ogiso, E.T. Papoutsakis, Overexpression of *fetA* (*ybbL*) and *fetB* (*ybbM*), encoding an iron exporter, enhances resistance to oxidative stress in *Escherichia coli*, Appl. Environ. Microbiol. 79 (2013) 7210–7219, https://doi.org/10.1128/AEM.02322-13.
[45] L. Bai, S. Zhang, Y. Deng, C. Song, G. Kang, Y. Dong, Y. Wang, F. Gao, H. Huang, Comparative genomics analysis of *Acinetobacter haemolyticus* isolates from sputum samples of respiratory patients, Genomics. 112 (2020) 2784–2793, https://doi.org/10.1016/j.ygeno.2020.03.016.
[46] D. Vallenet, P. Nordmann, V. Barbe, L. Poirel, S. Mangenot, E. Bataille, C. Dossat, S. Gas, A. Kreimeyer, P. Lenoble, S. Oztas, J. Poulain, B. Segurens, C. Robert, C. Abergel, J.-M. Claverie, D. Raoult, C. Médigue, J. Weissenbach, S. Cruveiller, Comparative analysis of acinetobacters: three genomes for three lifestyles, PLoS One 3 (2008), e1805, https://doi.org/10.1371/journal.pone.0001805.
[47] H. Yakkala, D. Samantarrai, M. Gribskov, D. Siddavattam, Comparative genome analysis reveals niche-specific genome expansion in *Acinetobacter baumannii* strains, PLoS One 14 (2019), e0218204, https://doi.org/10.1371/journal.pone.0218204.
[48] O. Avsian-Kretchmer, A.J.W. Hsueh, Comparative genomic analysis of the eight-membered ring Cystine knot-containing bone morphogenetic protein antagonists, Mol. Endocrinol. 18 (2004) 1–12, https://doi.org/10.1210/me.2003-0227.