Non-contiguous finished genome sequence of *Staphylococcus capitis* CR01 (pulsetype NRCS-A)

H. Lemriss[1,2], Martins Simões P[2,3], S. Lemriss[1], M. Butin[3], A. Ibrahimi[4], S. El Kabbaj[1], JP Rasigade[2,3], F. Laurent[2,3,5]

[1]**Department of Biosecurity PCL3, Laboratory of Research and Medical Analysis of the Fraternal of Gendarmerie Royale, Rabat, Morocco.**
[2]**International Centre for Research in Infectious diseases, INSERM U1111, University of Lyon, Lyon, France.**
[3]**Department of Clinical Microbiology, Northern Hospital Group, Hospices Civils de Lyon, Lyon, France.**
[4]**Medical Biotechnology lab. (MedBiotech), Medical and Pharmacy School, University Mohammed V Souissi, Rabat, Morocco.**
[5]**National Reference Center for Staphylococci, Hospices Civils de Lyon, Lyon, France.**

Correspondence: frederic.laurent@chu-lyon.fr

Staphylococcus capitis is a coagulase-negative staphylococci (CoNS) commonly found in the human microflora. Recently, a clonal population of *Staphylococcus capitis* (denominated NRCS-A) was found to be a major cause of late-onset sepsis (LOS) in several neonatal intensive care units in France. Here, we report the complete genome sequence and annotation of the prototype *Staphylococcus capitis* NCRS-A strain CR01. The 2,504,472 bp long genome (1 chromosome and no plasmids) exhibits a G+C content of 32.81%, and contains 2,468 protein-coding and 59 tRNA genes and 4 rRNA genes.

Abbreviations: EMBL- European Molecular Biology Laboratory, NCBI- National Center for Biotechnology Information (Bethesda, MD, USA), RDP- Ribosomal Database Project (East Lansing, MI, USA)

Introduction

A frequent cause of low-weight newborns mortality and morbidity in Neonatal Intensive Care Units (NICUs) are late-onset sepsis (LOS), that are defined as sepsis occurring after 3 days of age. The most frequently encountered pathogens are coagulase-negative staphylococci (CoNS) and within those *Staphylococcus epidermidis* has been shown to be the most prevalent [1,2]. However, a few studies have reported the emergence of *Staphylococcus capitis* as a main CoNS- and LOS- causative pathogen in NICU settings [2-4]. A study in French NICUs [2] has demonstrated the spread of a single clonal population of methicillin-resistant *S. capitis* (pulsotype NRCS-A) associated to reduced susceptibility to vancomycin, the first line of antibiotics used in cases of LOS. Moreover, this clone has also been recently identified in NICUs in Belgium, United Kingdom and Australia, which suggests a worldwide distribution. In contrast, in adult bac-

teremia, *S. capitis* are rarely found and when detected, it presents a bigger diversity in terms of genotypes as well as antimicrobial susceptibility profiles than neonates bacteremia.

In order to elucidate the molecular mechanisms behind the wide spreading of the *S. capitis* NRCS-A clone in NICUs throughout the world, we sequenced a prototype strain (CR01).

Classification and information

A strain belonging to the clonal population of *Staphylococcus capitis* NCRS-A pulsetype (Table 1) was isolated from the blood culture of a preterm infant with LOS, hospitalized in the NICU of the Northern Hospital Group Center (Hospices Civils de Lyon, Lyon, France) and suffering of LOS.

Species identification of the bacterial isolates and antimicrobial susceptibility testing (AST) were performed, respectively, using Vitek MS

(bioMérieux, Marcy l'Etoile), 16S rDNA sequencing, the automated BD Phoenix system (Becton Dickinson, Sparks, MD) and with Shimadzu-MALDI-TOF MS system (Shimadzu Corporation), as implemented on [21].

The strain was identified as being a*Staphylococcus capitis by* VITEK MS with 99.9% and at 93.7% by the MALDI-TOF MS, using the Shimadzu Launchpad software program and the SARAMIS database application (AnagnosTec GmbH) for automatic measurement and identification (Figure 1). Based on the information provided by the manufacture, when the score is ≥70%, identification is considered of high confidence.

The antimicrobial susceptibility test (AST) results were analyzed according to the recommendations of the French Microbiology Society [22]. The *S. capitis* bacteremia was considered positive based on a single positive blood culture [2,23]. The *S. capitis* NCRS-A isolate CR01, as all isolates from this clone, is resistant to penicillin, methicillin, gentamicin, rifampicin, hetero-resistant to vancomycin and sensitive to fusidic acid and fluoroquinolones.

Table 1, Figure 2 and Figure 3 show detailed information concerning general features of *Staphylococcus capitis* strain (CR01) and position within the genus *Staphylococcus*.

Table 1. Classification and general features of *Staphylococcus capitis* strain CR01, pulsetype-NRCS-A according the MIGS recommendation [5].

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain *Bacteria*	TAS [6]
		Phylum *Firmicutes*	TAS [7,8]
		Class *Bacilli*	TAS [9,10]
		Order *Bacillales*	TAS [11,12]
		Family *Staphylococcaceae*	TAS [13,14]
		Genus *Staphylococcus*	TAS [11,15,16]
		Species *Staphylococcus capitis*	TAS [17]
		Strain CR01, pulsetype-NRCS-A	TAS [2,18]
	Gram stain	Positive	TAS [19]
	Cell shape	Coccoid	TAS [17]
	Motility	Non-motile	TAS [19]
	Sporulation	Non-sporuating	TAS [19]
	Temperature range	Mesophilic	IDA
	Optimum temperature	37°C	TAS [17]
	Carbon source	Carbohydrates, (glucose, sacharose, fructose, manitol, mannose)	TAS [17]
	Energy source	Chemoorganotropic	TAS [17]
	Terminal electron receptor	O_2	TAS [17]
MIGS-6	Habitat	Skin of humans	TAS [17]
MIGS-6.3	Salinity	Physiological	TAS [17]
MIGS-22	Oxygen	Facultative anaerobes	TAS [17]
MIGS-15	Biotic relationship	Free-living	TAS [17]
MIGS-14	Pathogenicity	Opportunistic pathogen (Nosocomial bacteremia in premature neonates)	TAS [2]
MIGS-4	Geographic location	NICU Lyon, France	TAS [2]
MIGS-5	Sample collection time	2007	IDA
MIGS-4.1 MIGS-4.2	Latitude – Longitude	45° 45' 35" N 4° 50' 32" E	IDA
MIGS-4.3	Depth	Not applicable	IDA
MIGS-4.4	Altitude	162 m	IDA

a) Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [20].
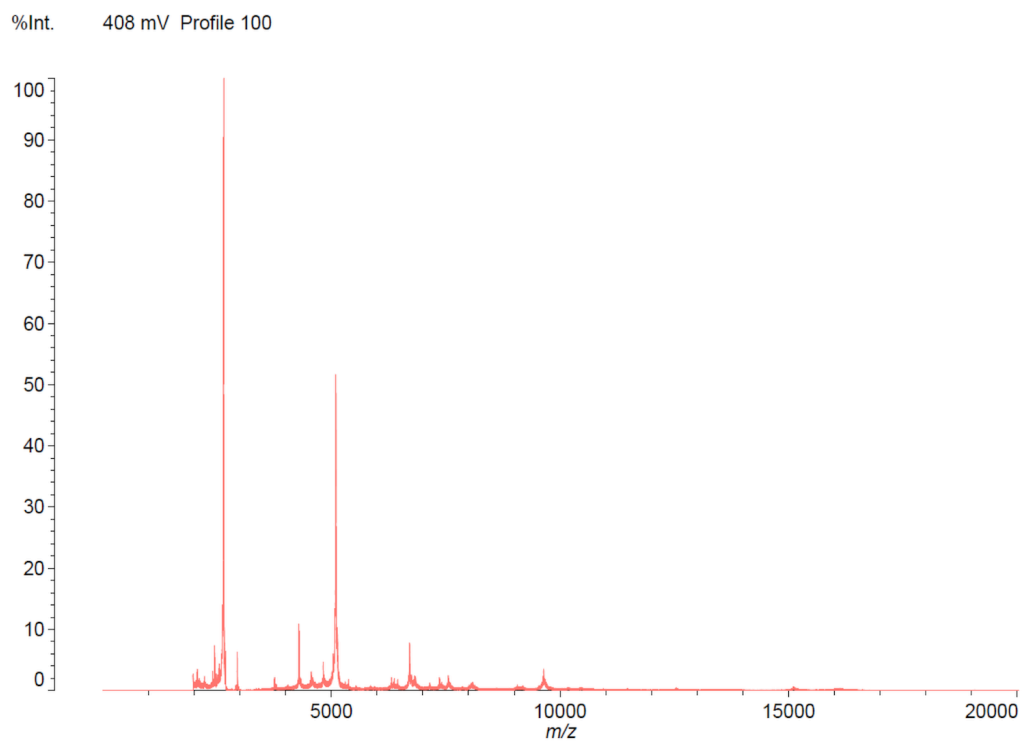
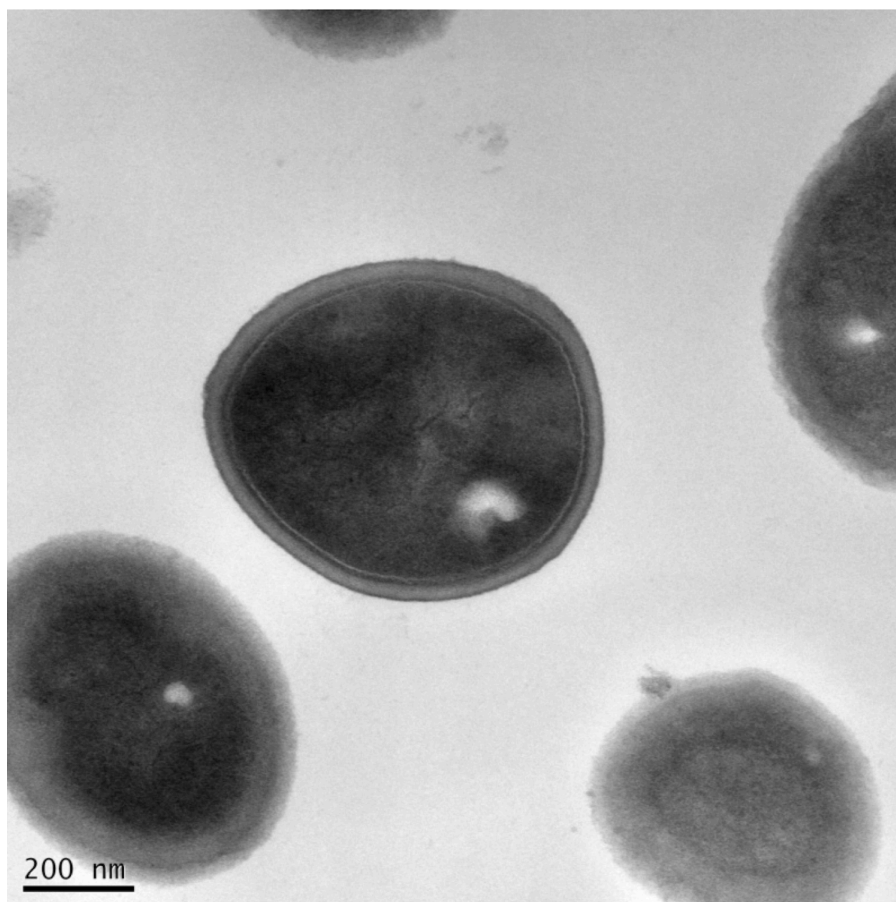**Figure 1.** Reference mass spectrum from *Staphylococcus capitis* strain (CR01).

Figure 2. Transmission electron microscopy of *Staphylococcus capitis* strain (CR01) using a JEOL 1400. The scale bar represents 200 nm.
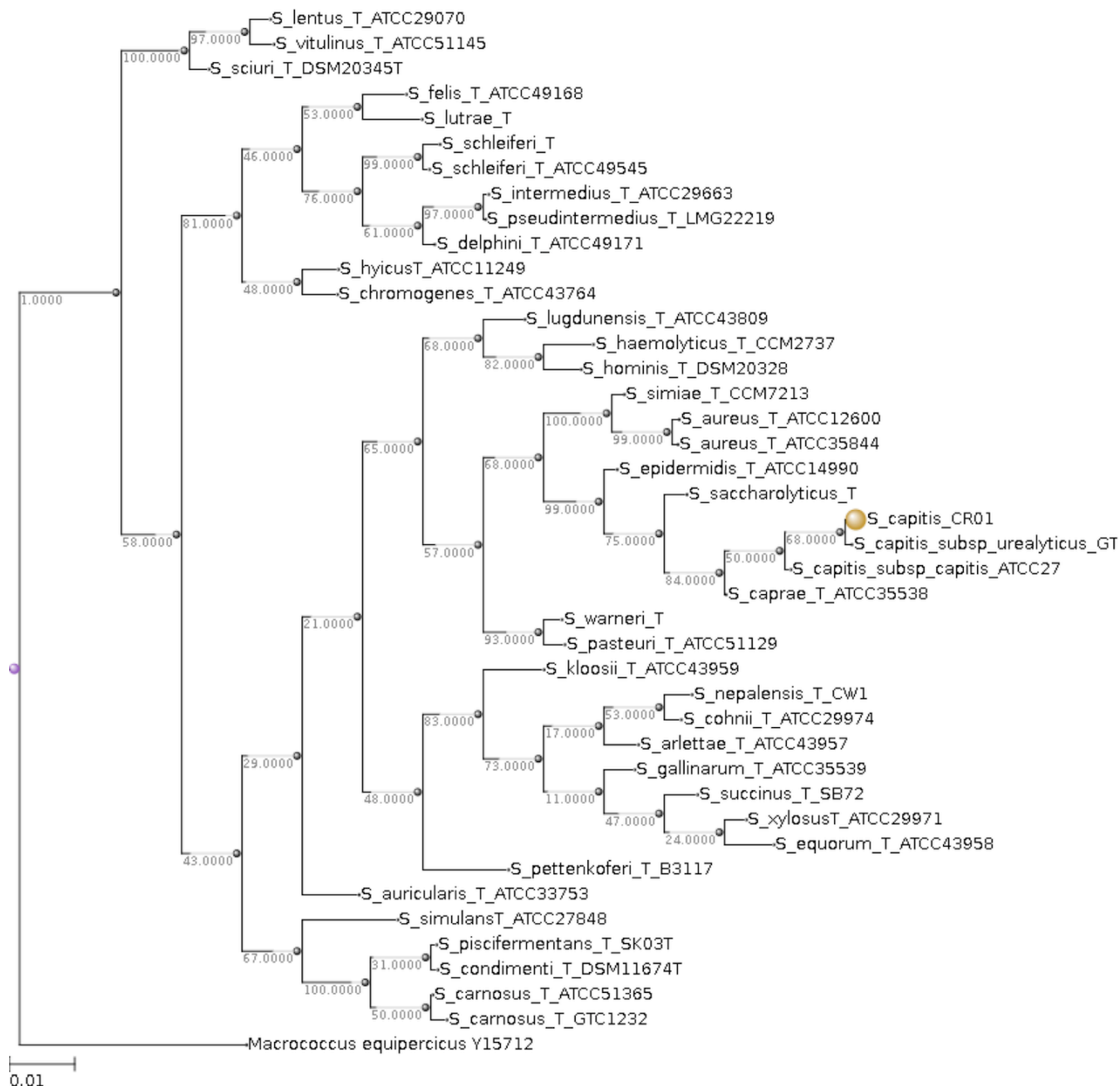
Figure 3. 16S rRNA Phylogenetic tree highlighting the position of *Staphylococcus capitis* strain CR01 (indicated by the yellow circle) relative to other type strains within the genus *Staphylococcus*. All 16S rRNA sequences were obtained from the RDP database using as filtering criteria: sequences with more than 1200 nt and classified as "good" quality sequences. The tree uses sequences aligned with the MUSCLE software, with the default parameters as implemented on Seaview version 4 [24], and a tree was inferred based on 1285 sites using the distance model of observed divergence, as implemented in the BioNJ algorithm.

The 16S rRNA sequences were aligned using the MUSCLE software, with the default parameters as implemented on Seaview version 4 [24], and a tree was inferred based on 1285 sites using the distance model of observed divergence, as implemented in the BioNJ algorithm, and a bootstrapping process repeated 500 times.

The final tree was rooted using the 16S rRNA sequence of *Macrococcus equipercicus* Type strain that belongs to a closely-related sister genus.

Genome sequencing information

The genome sequence of *S. capitis* strain CR01 was determined by high-throughput sequencing performed on a Genome Sequencer FLX + system (454 Life Sciences/Roche) using FLX Titanium reagents according to the manufacturer's protocols and instructions, with approximately 47-fold coverage of the genome. This platform provides longer read lengths than other sequencing platforms to obtain raw sequences. *De novo* assemblies were performed using the Roche Newbler (v 2.7) software package.

Genome project history

Table 2 presents the project information and its association with MIGS version 2.0 compliance [5].

Growth conditions and DNA isolation

The sample was prepared for sequencing by growing *S. capitis* CR01, aerobically at 37°C in Blood Agar for 24-48 hours. Genomic DNA was extracted using the PureLink™ genomic DNA kit (Invitrogen™) according to the manufacturer's recommended protocol. The quantity of DNA obtained was determined using a NanoVue ™ Plus (HVD Life Sciences), and 1 µg of DNA was used for sequencing of whole-genome of this strain.

Genome sequencing and assembly

The isolated DNA of *S. capitis* CR01, was used to create 454-shotgun libraries following the GS Rapid library protocol (Roche 454, Roche). The resulting 454 DNA libraries were sequenced using a whole-genome shotgun strategy by GS FLX Titanium sequencing kit XL+ [25] (202,108 reads totaling 2.5 Mb, X48 fold coverage of the genome). Genome sequences were processed by Roche's sequencing software according to the manufacturer's instructions (454 Life Science). The resulting shotgun reads were assembled *de novo* using the Roche Newbler assembly software 2.7 (454 Life

Science) and 26 large contigs (Contig00001 to Contig00026) were obtained. The *N*50 was 176239 bp.

Genome annotation

An automatic syntactic and functional annotation of the draft genome was performed using the MicroScope platform pipeline [26,27]. The syntactic analysis combines a set of programs including AMIGene [28], tRNAscan-SE [29], RNAmmer [30], Rfam scan [31] and Prodigal software [32] to predict genomic objects that are mainly CDSs and RNA genes. More than 20 bioinformatics methods are then used for functional and relational analyses: homology search in the generalist databank UniProt [33] and in more specialized databases as COG [34], InterPro [35], PRIAM profiles for enzymatic classification [36], prediction of protein localization using TMHMM [37], SignalP [38] and PsortB [39] tools.

Genome properties

The genome includes one circular chromosome of 2,504,472 bp (32.81% GC content). A total of 2,565 genes were predicted with 2,453 being protein-coding genes, 59 tRNA-encoding genes, 4 rRNA-encoding genes (including 2 copies of 5S rRNA, 1 copy of both the large and the small-subunits, respectively, 23S and 16S rRNA) and 34 other RNA related ORFs. No plasmid was detected.

Of the 2,453 protein-coding genes, 1,892 genes (76.7%) were assigned to a putative function with the remaining annotated as hypothetical proteins. The predicted coding density in *S. capitis* strain CR01 *was 86%.*

Table 3 and Figure 4 detailed description of the properties and the statistics of *Staphylococcus capitis* strain CR01 genome. The distribution of the genes into COGs functional categories is presented in Table 4.

Table 2. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Non-contiguous finished
MIGS-28	Libraries used	454 pyrosequence rapid library
MIGS-29	Sequencing platforms	454 GS FLX+
MIGS-31.2	Fold coverage	47.0 × pyrosequence
MIGS-30	Assemblers	Newbler Assembler 2.7
	GenBank	CBUB000000000.1

Table 3. Nucleotide content and gene count levels of the genome

Attribute	Value	% of total[a]
Genome size (bp)	2.504.472	100.00%
DNA G+C content (bp)	821.717	32.81%
DNA coding region (bp)	2.158.855	6.2%
Number of Scaffolds	26	-
Total genes[b]	2566	100.00%
RNA genes	97	4.00%
tRNA-enconding genes	59	2.30%
rRNA-encoding genes	4	0.20%
Protein-coding genes (CDS)	2454	96.00%
Genes assigned to COGs	1999	81.00%
Genes of unknown function	561	23.34%
Genes with transmembrane helices[c]	630	25.70%
CRISPR repeats	1	-

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.
b) Total number of genes includes CDS, RNA genes and pseudogenes.
c) Detection of transmembrane helices was performed using TMHMM v. 2.0 [40]

Table 4. Number of genes associated with general COG functional categories.

Code	Value	%age[a]	Description
J	178	7.21	Translation
K	184	7.46	Transcription
L	169	6.85	Replication, recombination and repair
D	29	1.18	Cell cycle control, mitosis and meiosis
V	85	3.44	Defense mechanisms
T	89	3.61	Signal transduction mechanisms
M	113	4.58	Cell wall/membrane biogenesis
N	23	0.9	Cell motility
W	1	0.04	Extracellular structures
U	33	1.34	Intracellular trafficking and secretion
O	86	3.48	Posttranslational modification, protein turnover, chaperones
C	144	5.83	Energy production and conversion
G	210	8.51	Carbohydrate transport and metabolism
E	370	14.99	Amino acid transport and metabolism
F	92	3.73	Nucleotide transport and metabolism
H	109	4.42	Coenzyme transport and metabolism
I	92	3.73	Lipid transport and metabolism
P	270	10.94	Inorganic ion transport and metabolism
Q	56	2.27	Secondary metabolites biosynthesis, transport and catabolism
R	425	17.22	General function prediction only
S	203	8.23	Function unknown
-	469	19.00	Not in COGs

a) The total is based on the total number of protein coding genes in the annotated genome.
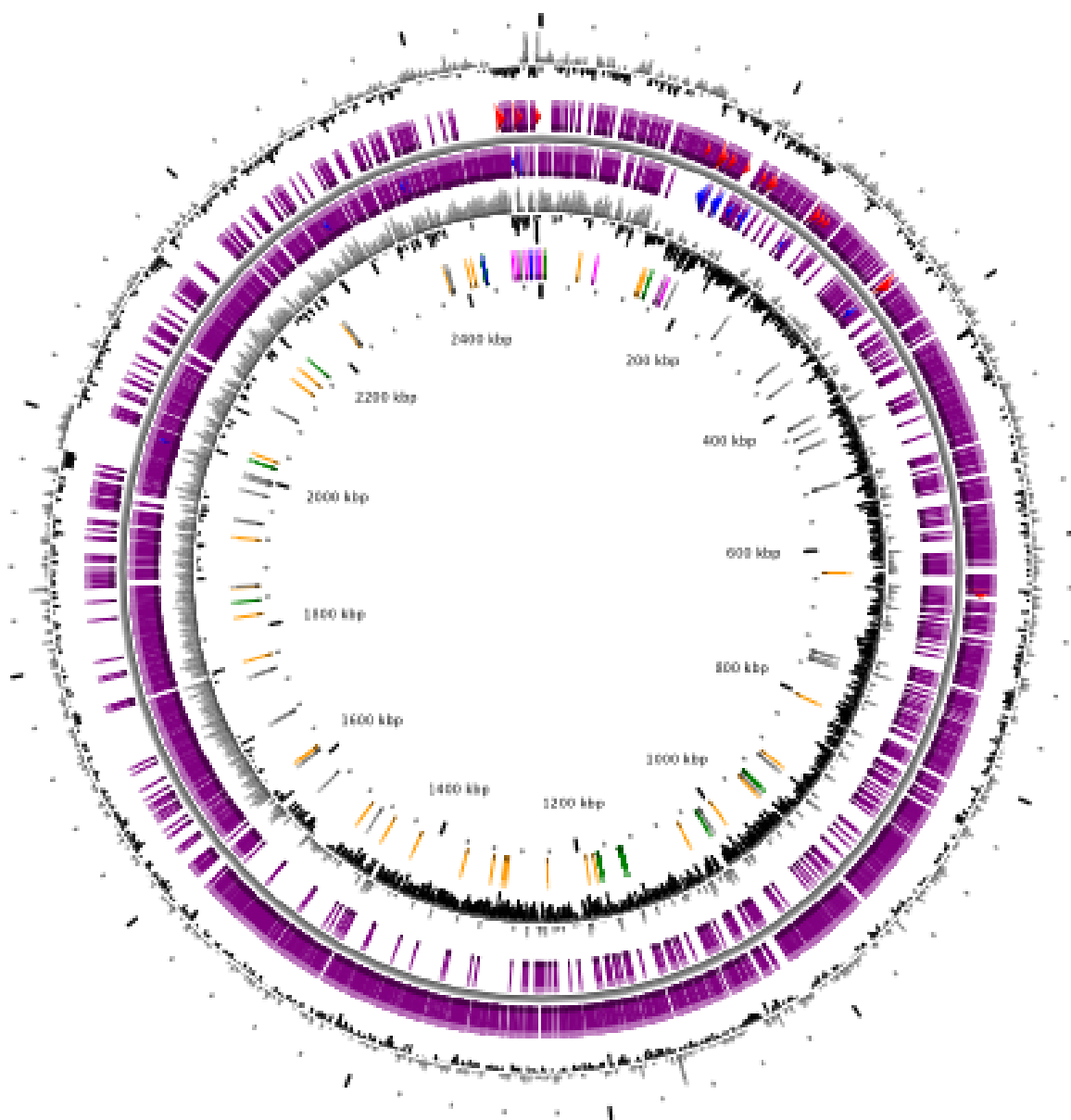
Figure 4. Graphical circular map of the chromosome. From outside to the center: Genes on the forward strand (colored by COG categories), genes on the reverse strand colored by COG categories), RNA genes (tRNAs green, rRNAs blue), GC content, and GC skew

Conclusion

Here, we described a new genome sequence of *Staphylococcus capitis* (strain CR01 belonging to NRCS-A clone) as a first step toward comparing its content with other sequenced *Staphylococcus capitis* genomes as well as CoNS genomes of species associated with late-onset sepsis. Detailed analyses are in progress to identify virulence factors and mobile genetic elements (MBE), such as the staphylococcal chromosome cassette (SCCmec) [18], potentially related to the high specificity of the NRCS-A clone to the NICU environment.

Acknowledgements

References

1. Klingenberg C, Rønnestad A, Anderson AS, Abrahamsen TG, Zorman J, Villaruz A, Flægstad T, Otto M. Sollid, Ericson J. Persistent strains of coagulase-negative staphylococci in a neonatal intensive care unit: virulence factors and invasiveness. *Clin Microbiol Infect* 2007; **13**:1100-1111. PubMed http://dx.doi.org/10.1111/j.1469-0691.2007.01818.x

2. Rasigade JP, Raulin O, Picaud JC, Tellini C, Bes M, Grando J, Ben Saïd M, Claris O, Etienne J, Tigaud S, Laurent F. Methicillin-Resistant*Staphylococcus capitis* with Reduced Vancomycin Susceptibility Causes Late-Onset Sepsis in Intensive Care Neonates. *PLoS ONE* 2012; **7**:e31548. PubMed http://dx.doi.org/10.1371/journal.pone.0031548

3. Ng PC, Chow VC, Lee CH, Ling JM, Wong HL, Chang RCY. Persistent *Staphylococcus capitis* septicemia in a preterm infant. *Pediatr Infect Dis J* 2006; **25**:652-654. PubMed http://dx.doi.org/10.1097/01.inf.0000225785.32137.d3

4. de Silva GDI, Kantzanou M, Justice A, Massey RC, Wilkinson AR, Day NPJ, Peacock SJ. The ica Operon and Biofilm Production in Coagulase-Negative Staphylococci Associated with Carriage and Disease in a *Neonatal Intensive Care* Unit. *J Clin Microbiol* 2002; **40**:382-388. PubMed http://dx.doi.org/10.1128/JCM.40.02.382-388.2002

5. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

6. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eukarya*. Proc Natl Acad Sci USA 1990; 87:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

7. Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of *Bacteria*. *Int J Syst Bacteriol* 1978; **28**:1-6; http://dx.doi.org/10.1099/00207713-28-1-1.

8. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), Bergey's Manual of Systematic Bacteriology, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.

9. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469-472. http://dx.doi.org/10.1099/ijs.0.022855-0

10. Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 19-20.

11. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

12. Prévot AR. In: Hauderoy P, Ehringer G, Guillot G, Magrou. J., Prévot AR, Rosset D, Urbain A (eds), Dictionnaire des Bactéries Pathogènes, Second Edition, Masson et Cie, Paris, 1953, p. 1-692.

13. List Editor. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469-472. http://dx.doi.org/10.1099/ijs.0.022855-0

14. Schleifer KH, Bell JA. Family VIII.*Staphylococcaceae* fam. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 392.

15. Rosenbach FJ. In: Bergmann JF (ed), Microorganismen bei den Wund-Infections-Krankheiten des Menschen., Wiesbaden, 1884, p. 1-122.

16. Judicial Commission. Opinion 17. Conservation of the Generic name *Staphylococcus* Rosenbach, Designation of *Staphylococcus aureus* Rosenbach as the Nomenclatural Type of the Genus Staphylococcus Rosenbach, and Designation of the Neotype culture of *Staphylococcus aureus* Rosenbach. *Int Bull Bacteriol Nomencl Taxon* 1958; **8**:153-154.

17. Kloos WE, Musselwhite MS. Distribution and Persistence of *Staphylococcus* and *Micrococcus* Species and Other Aerobic Bacteria on Human Skin. *Appl Microbiol* 1975; **30**:381-385. PubMed

18. Martins Simões P, Rasigade JP, Lemriss H, Butin M, Ginevra C, Lemriss S, Goering RV, Ibrahimi A, Picaud JC, Vandenesch FEL, *et al*. Characterization of a novel staphylococcal chromosome cassette (SCCmec) within a composite SCC island in neonatal sepsis-associated *Staphylococcus capitis* pulsotype NRCS-A. *Antimicrob Agents*

Chemother 2013; **57**:6354. PubMed
http://dx.doi.org/10.1128/AAC.01576-13

19. Schleifer KH, Bell JA. Family VIII.*Staphylococcaceae* fam. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 392.

20. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

21. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, Francois P, Schrenzel J. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *J Clin Microbiol* 2010; **48**:1169-1175. PubMed http://dx.doi.org/10.1128/JCM.01881-09

22. French Society for Microbiology. Recommandations du Comite de l'Antibiogramme de la Société Française de Microbiologie. 2009. Available at: wwwsfmasofr/doc/downloadphp?doc=DiU8C&fic=casfm_2009pdf. Accessed 22 February 2009.

23. Hall KK, Lyman JA. Updated review of blood culture contamination. *Clin Microbiol Rev* 2006; **19**:788-802. PubMed http://dx.doi.org/10.1128/CMR.00062-05

24. Gouy M, Guindon S, Gascuel O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* 2010; **27**:221-224. PubMed http://dx.doi.org/10.1093/molbev/msp259

25. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, *et al*. Whole-genome random sequencing and assembly of*Haemophilus influenzae* Rd. *Science* 1995; **269**:496-512. PubMed http://dx.doi.org/10.1126/science.7542800

26. Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fèvre F, Longin C, Mornico D, Roche D, *et al*. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res* 2013; **41**:D636-D647. PubMed http://dx.doi.org/10.1093/nar/gks1194

27. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Médigue C. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res* 2006; **34**:53-65. PubMed http://dx.doi.org/10.1093/nar/gkj406

28. Bocs S, Cruveiller S, Vallenet D, Nuel G, Médigue C. AMIGene: Annotation of MIcrobial Genes. *Nucleic Acids Res* 2003; **31**:3723-3726. PubMed http://dx.doi.org/10.1093/nar/gkg590

29. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. PubMed http://dx.doi.org/10.1093/nar/25.5.0955

30. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. PubMed http://dx.doi.org/10.1093/nar/gkm160

31. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. Rfam: updates to the RNA families database. *Nucleic Acids Res* 2009; **37**:D136-D140. PubMed http://dx.doi.org/10.1093/nar/gkn766

32. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. PubMed http://dx.doi.org/10.1186/1471-2105-11-119

33. UnitProt Consortium. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res* 2009; **37**:D169-D174. PubMed http://dx.doi.org/10.1093/nar/gkn664

34. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, *et al*. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 2003; **4**:41. PubMed http://dx.doi.org/10.1186/1471-2105-4-41

35. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, *et al*. InterPro: the integrative protein signature database. *Nucleic Acids Res* 2009; **37**:D211-D215. PubMed http://dx.doi.org/10.1093/nar/gkn785

36. Claudel-Renard C, Chevalet C, Faraut T, Kahn D. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res* 2003; **31**:6633-6639. PubMed http://dx.doi.org/10.1093/nar/gkg847

37. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 1998; **6**:175-182. PubMed

38. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795. PubMed http://dx.doi.org/10.1016/j.jmb.2004.05.028

39. Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* 2005; **21**:617-623. PubMed http://dx.doi.org/10.1093/bioinformatics/bti057

40. Moller S, Croning MDR, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 2001; **17**:646-653. PubMed http://dx.doi.org/10.1093/bioinformatics/17.7.646