The Genome Sequence of a Type ST239 Methicillin-Resistant *Staphylococcus aureus* Isolate from a Malaysian Hospital

LS Lee[1,2], LK Teh[1], ZF Zainuddin[2] and MZ Salleh[1*]

[1]Integrative Pharmacogenomics Centre, Faculty of Pharmacy, Universiti Teknologi MARA Malaysia, 42300 Bandar Puncak Alam, Selangor, Malaysia.
[2]School of Health Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

* Corresponding author: MZ Salleh (zakisalleh.mzs@gmail.com)

Keywords: *Staphylococcus aureus*, MRSA, Malaysia, Genomics

We report the genome sequence of a healthcare-associated MRSA type ST239 clone isolated from a patient with septicemia in Malaysia. This clone typifies the characteristics of ST239 lineage, including resistance to multiple antibiotics and antiseptics.

Introduction

Antibiotic resistance in *S. aureus* is a major concern, as an increasing number of infections are caused by methicillin-resistant *S. aureus* (MRSA). Figure 1 shows the phylogenetic position of *S. aureus* in relation to other staphylococci. In Malaysia, the incidence of MRSA-related infections is a cause of concern in hospitals country-wide. Health-associated MRSA (HA-MRSA) has been dominated by a few lineages in Southeast Asia, particularly ST239. Sequence type 239 is an international healthcare-associated (HA) MRSA lineage prevalent in Asia, South America and Eastern Europe, which includes EMRSA-1, -4, -7, and -11 and the Brazilian, Portuguese, Hungarian, and Viennese clones. Strains of type ST239 are typically resistant to multiple classes of antibiotics and antiseptics such as β-lactam antibiotics.

Classification and features

We have chosen a representative of an MRSA strain, termed MRSA PR01 isolated from a patient with septicemia, isolated from a hospital in Kuala Lumpur. Table 1 indicates general information gathered on MRSA PR01. The MRSA PR01 strain has been identified as sequence type 239 (ST239) by multilocus sequence typing (MLST). Initial disc susceptibility tests showed that the strain is resistant to β-lactam antibiotics oxacillin, ampicillin, cefuroxime, ceftriaxone, gentamicin, erythromycin, ciprofloxacin and co-trimoxazole.

Genome sequencing information

Genome project history

This organism was selected for sequencing as a representative of MRSA infection in a local Malaysian hospital. The genome sequences of this organism were deposited in GenBank (WGS database). Sequencing, finishing and annotation were performed at the Pharmacogenomics Centre (PROMISE), UiTM. Table 2 presents the project information and its association with MIGS version 2.0 compliance [14].

Growth conditions and DNA isolation

MRSA PR01 was grown overnight under aerobic conditions in Tryptic Soy Broth at 37°C. DNA extraction was performed using MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Madison, USA) as per manufacturer's instructions. The concentration and purity of resultant DNA was assessed by UV spectrophotometry (Nanodrop, Thermo Scientific). 5 µg of genomic DNA ($A_{260/280}$ = 1.88) was used for library preparation.

Genome sequencing and assembly

The genome sequence was obtained using 104 Mb of paired-end (300 bp spacing) data from the Illumina GAII_x platform (Illumina, San Diego, CA) with 36-bp reads. Sequence data were assembled using CLCBio Genomics Workbench (CLC bio, Aar-

hus, Denmark). One hundred and ninety-five contigs (N50: 13,272 bp) were generated, and were overlaid with the reference sequence Mu50 using OSLay. Fourteen supercontigs were generated as a result. Gaps were closed using Sanger sequencing.

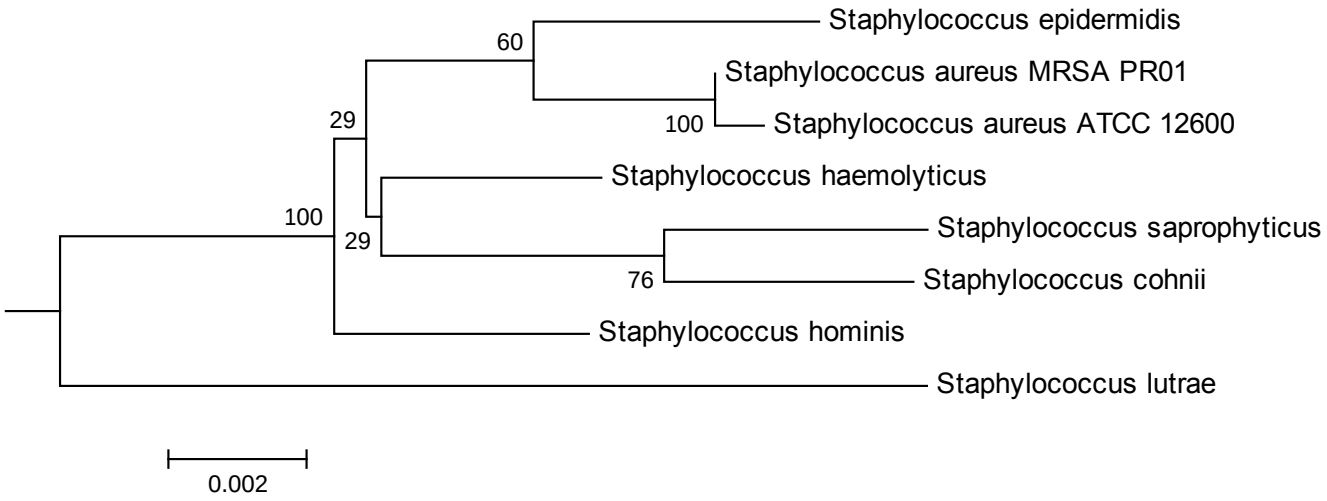

Figure 1. Phylogenetic tree highlighting the position of *Staphylococcus aureus* strain PR01 relative to other type strains within the *Staphylococcaceae*. The strains and their corresponding GenBank accession numbers for 16S rRNA genes are: *S. aureus* strain ATCC 12600, L36472; *S. saprophyticus* strain ATCC 15305, AP008934; *S. epidermidis* strain ATCC 14990, D83363; *S. hominis* strain DSM 20328, X66101; *S. haemolyticus* strain CCM2737, X66100; and *S. cohnii* strain ATCC 49330, AB009936. The tree uses sequences aligned by the RDP aligner, and uses the Jukes-Cantor corrected distance model to construct a distance matrix based on alignment model positions without the use of alignment inserts, and uses a minimum comparable position of 200. The tree is built with RDP Tree Builder, which uses Weighbor [1] with an alphabet size of 4 and length size of 1000. The building of the tree also involves a bootstrapping process repeated 100 times to generate a majority consensus tree [2]. *Staphylococcus lutrae* (X84731) was used as an outgroup.

Genome properties

The MRSA PR01 genome consists of a 2,725,110-bp circular chromosome with a GC content of 32.6% (Table 3). The MRSA PR01 genome contains 2668 CDs with 19 rRNA features (). A total of 1722 (64.5%) of protein coding genes were assigned to COGs, and a breakdown of the functional assignment of COG-assigned genes is shown in Table 4. Plasmid sequences were only partially sequenced. Figure 2 depicts genomic regions of interest found in the preliminary analysis of the MRSA PR01 genome.

Initial analysis of the genome revealed several key features. This genome has a typical SCCmec type III cassette, containing cadmium resistance genes. SCCmec type III is a composite element that is comprised of SCCmec and SCCmercury. In the MRSA PR01 genome, like others, this region harbors *ccrC*, pI258 and Tn554 as well as the genes involved in cadmium resistance. The MRSA PR01 genome contains two pathogenicity islands, and several resistance features were identified such as the *qacA* gene, which confers resistance to antiseptics such as cationic biocides, quaternary ammonium salts, and diamidines via an export-mediated mechanism, and the *norA* gene which confers resistance to hydrophilic quinolones such as norfloxacin and ciprofloxacin. There were 9 regions defined as prophage regions by PHAST [17] with one complete prophage region.genes were identified in the genome. A total of 2,267 genes (72.66%) were assigned a putative function. The remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Table 3. The distribution of genes into COGs and KEGG functional categories is presented in Table 4.

Conclusion

This study is the first to report on the whole genome sequence of a Malaysian MRSA isolate. Preliminary analysis of the genome has highlighted the genetic determinants that are responsible for the organism to adapt easily to selective pressures. Further research is being conducted to provide insight on the adaptive power of this healthcare-associated strain to attain high resistance to antibiotics.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ANPO00000000. The version described in this paper is the first version, ANPO01000000.

Table 1. Classification and general features of *Staphylococcus aureus* MRSA PR01

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain *Bacteria*	[3]
		Phylum *Firmicutes*	[4-7]
		Class *Bacilli*	[8,9]
		Order *Bacillales*	[6,10]
		Family *Staphylococcaceae*	[9,11]
		Genus *Staphylococcus*	[6,12]
		Species *Staphylococcus aureus*	[6,12]
		Type strain MRSA PR01	TAS
	Gram stain	Positive	TAS
	Cell shape	Coccus	TAS
	Motility	Non-motile	TAS
	Sporulation	Non-sporulating	TAS
	Temperature range	Mesophile	TAS
	Optimum temperature	30-37°C	TAS
	Carbon source	Glucose	TAS
	Energy source	Chemoorganotrophic	
	Terminal electron receptor		
MIGS-6	Habitat	Human respiratory tract, skin	TAS
MIGS-6.3	Salinity		
	Oxygen		
MIGS-22		Facultative anaerobe	TAS
MIGS-15	Biotic relationship		
MIGS-14	Pathogenicity	Opportunistic pathogen	TAS
MIGS-4	Geographic location	Malaysia	
MIGS-5	Sample collection time	May 2009	
MIGS-4.1	Latitude	4.1936°N	
MIGS-4.2	Longitude	103.7249°E	
MIGS-4.3	Depth	Not reported	
MIGS-4.4	Altitude	Not reported	

[a]Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature). These evidence codes are from the Gene Ontology project [19].

Table 2. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Non-contiguous Finished
MIGS-28	Libraries used	One 350bp Illumina GAIIx genomic library
MIGS-29	Sequencing platforms	Illumina GAIIx, Sanger
MIGS-31.2	Fold coverage	>200×
MIGS-30	Assemblers	CLCBio Genomics Workbench
MIGS-32	Gene calling method	Glimmer and GeneMark
	Genbank ID	ANPO01000000
	Genbank Date of Release	January 11, 2014
	GOLD ID	Gi0037576
MIGS-13	Project relevance	Medical, Tree of life

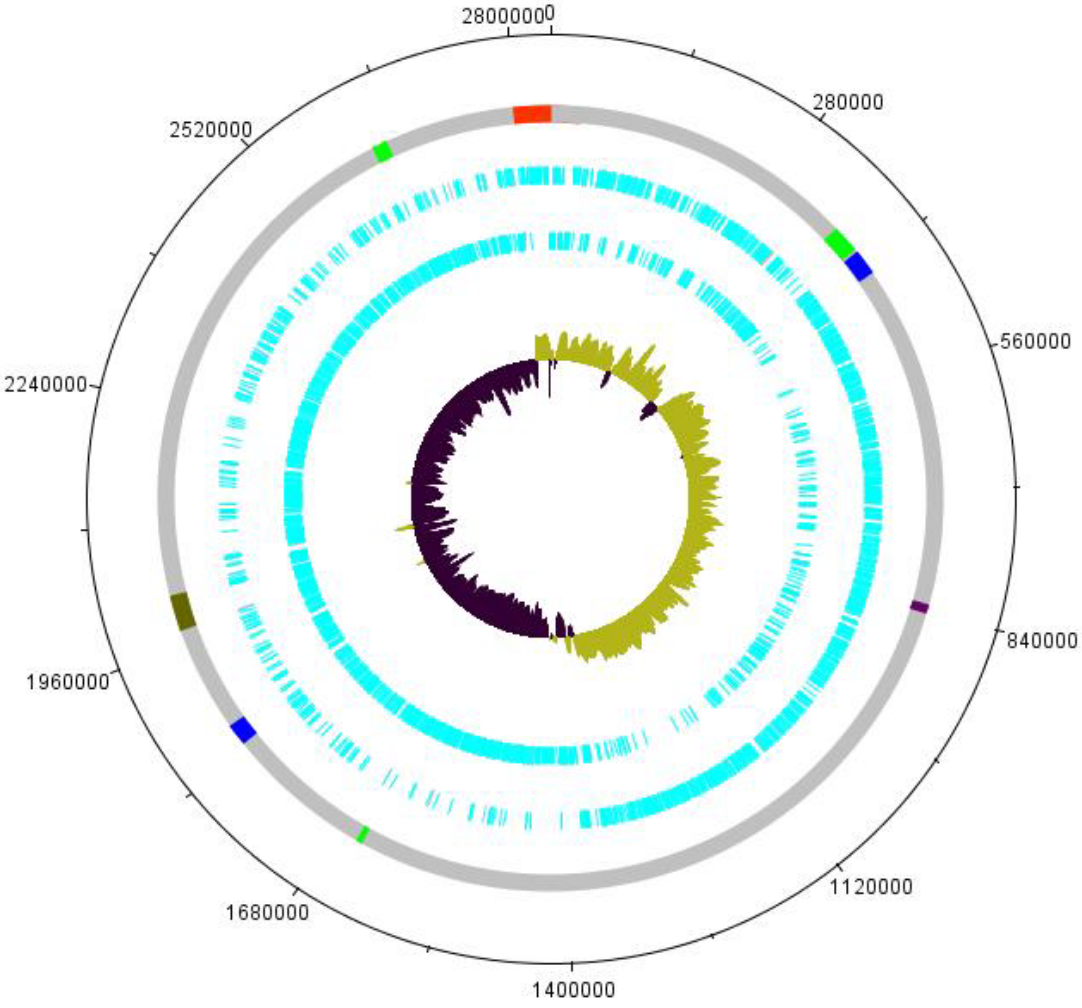

Figure 2. Visual representation of the MRSA PR01 genome. From outer to inner tracks: Scale (in bases); annotated CDSs colored according to predicted function (red, SCC element; blue, genomic island; green, transposon/ integrative conjugative element; purple, *S. aureus* pathogenicity island [SaPI], brown, prophage); forward strand CDS; reverse strand CDS; GC skew.

Table 3. Nucleotide content and gene count levels of the MRSA PR01 genome

Attribute	Value	% of total[a]
Genome size (bp)	2,725,110	100
DNA G+C content (bp)	888,386	32.6
DNA Coding region (bp)	2,555,544	90.03
Total genes	2687	100
RNA genes	19	0.7
Protein-coding genes	2668	99.3
Genes with protein function prediction	2,267	72.66
Genes assigned to COGs	1722	64.5

[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

Table 4. Number of genes associated with the 25 general COG functional categories

Code	Value	%age[a]	Description
J	140	5.247	Translation
A	-	-	RNA processing and modification
K	127	4.760	Transcription
L	126	4.723	Replication, recombination and repair
B	-	-	Chromatin structure and dynamics
D	23	0.862	Cell cycle control, mitosis and meiosis
Y	-	-	Nuclear structure
V	-	-	Defense mechanisms
T	47	1.762	Signal transduction mechanisms
M	91	3.411	Cell wall/membrane biogenesis
N	4	0.150	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	0	0	Intracellular trafficking and secretion
O	72	2.699	Posttranslational modification, protein turnover, chaperones
C	106	3.973	Energy production and conversion
G	129	4.835	Carbohydrate transport and metabolism
E	186	6.972	Amino acid transport and metabolism
F	68	2.549	Nucleotide transport and metabolism
H	83	3.111	Coenzyme transport and metabolism
I	62	2.324	Lipid transport and metabolism
P	123	4.610	Inorganic ion transport and metabolism
Q	23	0.862	Secondary metabolites biosynthesis, transport and catabolism
R	193	7.234	General function prediction only
S	119	4.460	Function unknown
-	946	35.457	Not in COGs

[a]The total is based on the total number of protein coding genes in the annotated genome.

Conclusion

Description of *Sulfurimonas hongkongensis* sp. nov.

Sulfurimonas hongkongensis (hong.kong.en'sis. N.L. fem. adj. *hongkongensis* pertaining to Hong Kong, the city where the type strain was isolated).

Strain AST-10^T is rod-shaped with size of 0.2-0.4 µm x 0.5-1.2 µm. It is an obligate anaerobe and occurs singly. The temperature range for growth is 15-35^{o}C, optimum at 30^{o}C. The pH range for growth is 6.5-8.5, optimum at 7.0-7.5. The salinity range for growth is 10-60 g L^{-1}, and optimum at 30 g L^{-1}. Strictly chemolithoautotrophic growth occurs with H_2, HS^- or $S_2O_3^{2-}$ as an electron donor and with nitrate as an electron acceptor. Nitrate is reduced to N_2, and reduced sulfur compounds are oxidized into S^0 or SO_4^{2-} (depending on molar ratio of $S_2O_3^{2-}/NO_3^-$). The major cellular fatty acids are $C_{14:0}$, $C_{16:0}$, 2-OH $C_{16:0}$, $C_{16:1}$, $C_{18:0}$, and $C_{18:1}$, with $C_{16:0\ 2\text{-}OH}$ as a unique fatty acid different from other species in the genus *Sulfurimonas*.

The type strain AST-10^T = DSM 2096^T = JCM 18418^T, was isolated from coastal sediment at the Kai Tak Approach Channel connected to Victoria Harbour in Hong Kong, China. The GC content of the genome is 34.9%. The genome sequence has been deposited at DDBJ/EMBL/GenBank under accession number AUPZ00000000.

Acknowledgments

Dr. Lin Cai thanks The University of Hong Kong for the Postdoctoral Fellowship. This study was financially supported by the Research Grants Council of Hong Kong (HKU7201/11E).

References

1. Inagaki F, Takai K, Kobayashi H, Nealson KH, Horikoshi K. *Sulfurimonas autotrophica* gen. nov., sp. nov., a novel sulfur-oxidizing epsilonproteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol* 2003; **53**:1801-1805. PubMed http://dx.doi.org/10.1099/ijs.0.02682-0
2. Takai K, Suzuki M, Nakagawa S, Miyazaki M, Suzuki Y, Inagaki F, Horikoshi K. *Sulfurimonas paralvinellae* sp. nov., a novel mesophilic, hydrogen- and sulfur-oxidizing chemolithoautotroph within the *Epsilonproteobacteria* isolated from a deep-sea hydrothermal vent polychaete nest, reclassification of *Thiomicrospira denitrificans* as *Sulfurimonas denitrificans* comb. nov. and emended description of the genus *Sulfurimonas*. *Int J Syst Evol Microbiol* 2006; **56**:1725-1733. PubMed http://dx.doi.org/10.1099/ijs.0.64255-0
3. Labrenz M, Grote J, Mammitzsch K, Boschker HT, Laue M, Jost G, Glaubitz S, Jurgens K. *Sulfurimonas gotlandica* sp. nov., a chemoautotrophic and psychrotolerant epsilonproteobacterium isolated from a pelagic Baltic Sea redoxcline, and an emended description of the genus *Sulfurimonas*. *Int J Syst Evol Microbiol* 2013; **63**:4141-4148. PubMed http://dx.doi.org/10.1099/ijs.0.048827-0
4. Grote J, Schott T, Bruckner CG, Glockner FO, Jost G, Teeling H, Labrenz M, Jurgens K. Genome and physiology of a model *Epsilonproteobacterium* responsible for sulfide detoxification in marine oxygen depletion zones. *Proc Natl Acad Sci USA* 2012; **109**:506-510. PubMed http://dx.doi.org/10.1073/pnas.1111262109
5. Sievert SM, Scott KM, Klotz MG, Chain PS, Hauser LJ, Hemp J, Hugler M, Land M, Lapidus A, Larimer FW, *et al*. Genome of the epsilonproteobacterial chemolithoautotroph *Sulfurimonas denitrificans*. *Appl Environ Microbiol* 2008; **74**:1145-1156. PubMed http://dx.doi.org/10.1128/AEM.01844-07
6. Sikorski J, Munk C, Lapidus A, Ngatchou Djao OD, Lucas S, Glavina Del Rio T, Nolan M, Tice H, Han C, Cheng JF, *et al*. Complete genome sequence of *Sulfurimonas autotrophica* type strain (OK10). *Stand Genomic Sci* 2010; **3**:194-202. PubMed
7. Bruckner CG, Mammitzsch K, Jost G, Wendt J, Labrenz M, Jurgens K. Chemolithoautotrophic denitrification of *Epsilonproteobacteria* in marine pelagic redox gradients. *Environ Microbiol* 2013; **15**:1505-1513. PubMed http://dx.doi.org/10.1111/j.1462-2920.2012.02880.x
8. Shao M, Zhang T, Fang HH. Autotrophic denitrification and its effect on metal speciation during marine sediment remediation. *Water Res* 2009; **43**:2961-2968. PubMed http://dx.doi.org/10.1016/j.watres.2009.04.016
9. Zhang M, Zhang T, Shao MF, Fang HH. Autotrophic denitrification in nitrate-induced marine sediment remediation and *Sulfurimonas denitrificans*-like bacteria. *Chemosphere* 2009; **76**:677-682. PubMed http://dx.doi.org/10.1016/j.chemosphere.2009.03.066
10. Shao MF, Zhang T, Fang HH, Li X. The effect of nitrate concentration on sulfide-driven autotrophic denitrification in marine sediment. *Chemosphere* 2011; **83**:1-6. PubMed

http://dx.doi.org/10.1016/j.chemosphere.2011.01.042
11. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *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
12. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576
13. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part B, Springer, New York, 2005, p. 1.
14. Validation List No. 107. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2006; **56**:1-6. PubMed http://dx.doi.org/10.1099/ijs.0.64188-0
15. Garrity GM, Bell JA, Lilburn T. Class V. *Epsilonproteobacteria* class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part C, Springer, New York, 2005, p. 1145.
16. Garrity GM, Bell JA, Lilburn T. Order I. *Campylobacterales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part C, Springer, New York, 2005, p. 1145.
17. Garrity GM, Bell JA, Lilburn T. Family II. *Helicobacteraceae* fam. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part C, Springer, New York, 2005, p. 1168.
18. 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. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556
19. Stackebrandt E, Goebel BM. A Place for DNA-DNA Reassociation and 16s Ribosomal-RNA Sequence-Analysis in the Present Species Definition in Bacteriology. *Int J Syst Bacteriol* 1994; **44**:846-849. http://dx.doi.org/10.1099/00207713-44-4-846
20. RossellóoMora R. Amann R. The species concept for prokaryotes. *FEMS Microbiol Rev* 2001; **25**:39-67. PubMed http://dx.doi.org/10.1016/S0168-6445(00)00040-1
21. Gevertz D, Telang AJ, Voordouw G, Jenneman GE. Isolation and characterization of strains CVO and FWKO B, two novel nitrate-reducing, sulfide-oxidizing bacteria isolated from oil field brine. *Appl Environ Microbiol* 2000; **66**:2491-2501. PubMed http://dx.doi.org/10.1128/AEM.66.6.2491-2501.2000
22. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 2000; **28**:33-36. PubMed http://dx.doi.org/10.1093/nar/28.1.33
23. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 2011; **12**:444. PubMed http://dx.doi.org/10.1186/1471-2164-12-444
24. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011; **39**:W29-W37. PubMed http://dx.doi.org/10.1093/nar/gkr367
25. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, *et al*. The Pfam protein families database. *Nucleic Acids Res* 2012; **40**:D290-D301. PubMed http://dx.doi.org/10.1093/nar/gkr1065
26. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 2011; **8**:785-786. PubMed http://dx.doi.org/10.1038/nmeth.1701
27. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; **305**:567-580. PubMed http://dx.doi.org/10.1006/jmbi.2000.4315