J. Asiat. Soc. Bangladesh, Sci. 39(1): 9-20, June 2013

BIOINFORMATICS ANALYSIS OF NOVEL NON-CODING MOTIFS IN PATHOGENIC BACTERIAL GENOME

NISHAT SHAYALA[1] AND ISHTIAQUE RASHID

Department of Life Sciences, North South University, Dhaka, Bangladesh
[1]Department of Biotechnology and Genetic Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh

Abstract

Antibiotic resistance of MRSA (Methicillin Resistant *Staphylococcus aureus*) has been evolved through the rapid and diversified changes in the genetic structure of the bacterial strains. The onset of new resistant strains makes the diagnosis, prognosis, treatment and control process more difficult. The study was started with MRSA252, the most diverse strain of *S. aureus* and then elaborated to 14 other strains. By screening out the complete genome sequences of MRSA, 274 regions of 362,792bp non-coding unique sequences were found. Among those, sequences of less than 500bp length are mostly important to use in diagnostic purposes. Functional analysis and comparison with few other pathogens were done to find correlations.

Key words: *Staphylococcus aureus*, Non-coding DNA, Antibiotic resistance, Sequence analysis, Functional analysis

Introduction

Methicillin Resistant *Staphylococcus aureus* (MRSA) is one of the major 'superbugs' in nosocomial and community acquired infections. The extent of infection may range from simple to life threatening and this has become of great concern in hospitals (Am J Infect Control 1999). Although they are named after the antibiotic methicillin, MRSA strains show resistance against a number of antibiotics, including the Penicillin family (methicillin, dicloxacillin, nafcillin, oxacillin, etc.), tetracycline, minocycline, vancomycin, streptomycin and to some toxic metals (Neu1992).The most common mechanism of resistance against penicillin involves the production of penicillinase or β-lactamase, which breaks the β-lactam ring of the penicillin molecule (Lyon and Skurray 1987). Another mechanism introduces the product from *MecA* gene, which synthesizes penicillin binding proteins (PBP2a and PBP2') and upon binding inactivates penicillin and methicillin (Ponting *et al.* 2009). Other mechanisms include production of protonated amide or hydroxyl groups by interacting with 30S ribosomal RNA and efflux the drugs out from the cell (Hiramatsu 2004). Bacterial strains having plasmid, and/or genomic islands may carry resistance genes to different antibiotics. Yet most of the mechanisms are not properly understood (Neu1992 and Strommenger *et al.* 2003). These bacteria can confer resistance through mutation, gene alteration or gene transfer (i.e. horizontal gene

[1] Corresponding author: E-mail: nshayala@yahoo.com

transfer). Naturally evolving bacteria may also gain or lose functions that could turn them into resistant strains (Neu1 992 and Hiramatsu 2004). One of the most fascinating characteristics of these bacteria is their ability to change their non-coding region of the genome as per necessity e.g. under stressful condition (Ponting *et al.* 2009).

MRSA transmission occurs through skin contact and as it can be highly infectious, prevention, control and treatment is necessary as soon as it is detected. Early detection of the pathogens and their resistance can prevent the infection from becoming worse. It is preferable to have a diagnostic system that needs a small amount of sample and can give an accurate and specific result in the shortest possible time (Holden *et al.* 2004). Detection of MRSA is commonly done by a micro-dilution assay of a bacterial broth culture which gives phenotypic antibiotic resistance data and takes a minimum of 2-3 days incubation time. ELISA and scanning electrochemical microscopy (SECM) coupled with ELISA can be good methods for detecting MRSA (Thornsbery *et al.* 1983, J Hosp Infection 1998 and Dequaire *et al.* 1999). The latter needs a sample of 5.25pg/mL and takes a day for incubation. Molecular methods are now sophisticated and popular because of the reliability. Polymerase chain reaction (PCR) and multiplex PCR are the fastest ways of identifying MRSA infections. Multiplex PCR can identify genes in different groups of antibiotics. These procedures can take 2.5 to 6 hours for detection with the smallest amount of DNA. The main problems with these systems are the presence of inhibitors and the amplification of non-targets, which can give false negative results (Kasai *et al.* 2000 and Strommenger *et al.* 2003). Detection of *MecA* gene is also used in many hospitals which can be done by either by radio labeling, PCR amplification or simply by enrichment culture broths (Kitagawa *et al.* 1996).But this method is not very well established as it can only detect nasal and wound infections(Brown *et al.* 2005 and Oberdorfer *et al.*2006). Another DNA screening can be done to find the Panton-Valentine leukocidin (*PVL*) gene, which is also a toxin producer in MRSA infection (Shrestha *et al.* 2002).

The non-coding DNA which were previously referred to as “junk sequences” are now found to carry great value as they have regulatory roles, can play an important role in protein folding and can be promoter or operator binding sites as well. They also give information regarding evolution (Oberdorfer *et al.* 2006).In recent times thousands of different types of RNAs which are transcribed from non-coding regions of the genome have been found to play regulatory roles (Locey and White 2010). These non-coding RNAs are also found to be associated with disease, from microbes to humans. Compared to eukaryotes, prokaryotic genomes contain a very small amount of non-coding DNA, which mainly serves as the inter-genic region and typically has regulatory functions. Studies have found that approximately 6-14% of most of the bacteria and archaeal genome has non-coding part and has a positive selective pressure on the evolutionary process (Mercer *et al.* 2009). MRSA has a highly gene dense genome (84% coding region) containing approximately 2.9Mb DNA (Rogozin *et al.* 2002). Analysis of the

genome sequence using bioinformatics tools may provide valuable information before heading the laboratory for experimentations. This study was done with an aim of finding novel motifs of MRSA non-coding sequences with significance, which can be used for diagnostic purposes.

Materials and Methods

The complete genome of bacteria and other related sequences were downloaded from NCBI nucleotide database (*www.ncbi.nlm.nih.gov › NCBI › DNA and RNA*) and the Welcome Trust Sanger Institute's own genome database (www.sanger.ac.uk/ *resources/databases/*). For sequences comparison and functional study, different online tools and software were used (Johnson *et al.* 2008). Artemis, ACT (Artemis Comparison Tool, www.sanger.ac.uk/resources/software/artemis), Inter ProScan (http:// www. ebi.ac.uk/Tools/pfa/iprscan/), AmiGO (http:// amigo.geneontology.org/cgi-bin/amigo/ go.cgi), BLAST, BLASTn, BLASTx, (*blast.ncbi.nlm.nih.gov,* version 2.2.23) were the major tools used for this study purpose (Altschul *et al.* 1990, Rutherford *et al.* 2000, Korf *et al.*2003 and Carver *et al.* 2005). PERL and BioPerl (www.bioperl.org) were used for the bioinformatics analysis (Tisdall 2001).

Fourteen *Staphylococcus aureus* strains were taken under consideration as per their availability on reliable databases and the relatedness with the diseases symptoms (Data retrieve date: June to August, 2011). Table 1 represents one strain susceptible to the antibiotic Methicillin and other commonly used antibiotics (MSSA476) and other 13 strains which are resistant to several antibiotics (including Methicillin).

The complete process of the method of research was organised as follows: 1. Getting sequences of MRSA252 where no MSSA476 common sequence is present; 2. Screening out (Using non-redundant databases) the unique non-coding MRSA252 sequences; 3. Finding conserved non-coding sequences in 14 strains of MRSA (Table 1) which have no match with MSSA476; 4. Comparing MRSA non-coding sequences with other pathogenic bacteria and 5. Sequence to function analysis.

Windows and Linux operating system were used for convenience. Bioperl and Perl were used for the analysis. In Bioperl, there are already prepared commands, available to use in biological study. The major commands of Perl used here was the 'subroutines', 'hash', 'formatdb', 'makeblastdb' etc.

The obtained sequences were in different sizes. Those were then taken to compare with some other pathogenic bacteria randomly. From 100% BLASTn match, few organisms were taken and their sequences were downloaded and crosschecked with NCBI prokaryotic genome database (Data retrieve date: June to August, 2011 by Nishat Shayala). Overview of the sequence searching and matching process is presented in Fig. 1. Detailed information of the pathogens was used for functional analysis. Again Perl was used for comparison and finding the non-coding genes around the coding region.

Table 1. List of *Staphylococcus aureus* strains taken for this study.

Name of strain	Accession no	Compete genome Size	Source
Staphylococcus aureus 04-02981	CP001844.2	2,821,452bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* COL chromosome	NC_002951.2	2,809,422bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* JH9	CP000703.1	2,906,700bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* MRSA252 chromosome	NC_002952.2	2,902,619bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* MSSA476	NC_002953.3	2,799,800bp circular DNA	Sanger Institute
Staphylococcus aureus subsp. *aureus* Mu50	NC_002758.2	2,878,529bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* MW2	NC_003923.1	2,820,462bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* N315	NC_002745.2	2,814,816bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* NCTC 8325 chromosome	NC_007795.1	2,821,361bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* str. Newman chromosome	NC_009641.1	2,878,897bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* T0131	CP002643.1	2,913,900bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* TCH60	CP002110.1	2,802,675bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* USA300_FPR3757	CP000255.1	2,872,769bp circular DNA	NCBI (nucleotide database)
Staphylococcus aureus subsp. *aureus* USA300_TCH1516	CP000730.1	2,872,915bp circular DNA	NCBI (nucleotide database)

Searching nearby genes and protein products of non-coding sequences and analysing their functions were done to find if there was any impact of that region on the coding sequences. ARTEMIS and ACT not only give the sequence data but also statistical and graphical presentations and compare sequences from any format. MRSA252 was taken and compared with 3 other *Staphylococcus* species.

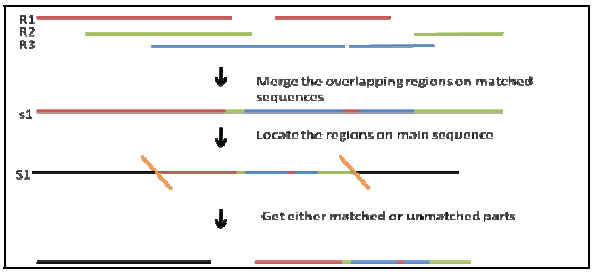

Fig.1. Overview of the working process. R1, R2 and R3 are the regions of the matched sequences after the BLAST search. Their overlapping regions are merged in the next step. Finally that portion is either removed or taken out of the main sequence to get the expected output. In every step after the BLAST search this basic procedure is used with the help of the Perl programming language.

InterProScan is as simple as BLAST and with only a sequence input can search for every major domain and protein family database, looking for single peptides, finding transmembrane domains and showing low complexity regions. It also assigns sequences to InterPro family and tells about the gene ontology terms that apply to it. The names of proteins found in three species that matched sequences with MRSA252 were taken for function prediction. From Artemis, the amino acid sequences in FASTA for the genes/locus were taken to search in InterProScan. From the InterProScan result for each protein, their GO (Gene Ontology) was searched through the GO link and through "AmiGO" for the cellular component, biological process and molecular functions (Quevillon *et al.*2005).

Results and Discussion

Uncommon in other prokaryotic pathogens *Staphylococcus aureus* has a complex genome structure with combination of genes and mobile genetic elements which can also be found on their plasmids (Weigel *et al.*2003). This research work started by looking into the non-coding sequence of *S. aureus* which was found to be 16% of their genome in the previous study (Rogozin *et al.* 2002).

Results from the search for unique sequence of MRSA252 strain which excluded common MSSA476 sequences came out with 492,331bp unique MRSA252 sequences of the total 2.9Mb genome (Table 2). From this large unique region 363,767bp sequences were found to be non-coding in MRSA252 (Table 2). Conserved sequences among the 14MRSA strains were found to have a total sequence length of 362,792 bp. A total of 274 sequence pool was formed with variant sequence length ranges. Among those, 139 regions of sequences had lengths less than 500 bp, 36 had lengths between 500 to 1000 bp, 29 had lengths between 1000 to 1500 bp and remaining had lengths above 1500 bp (Fig. 2). The mean length of conserved sequences was 1324.058 bp (Table 2).

Table 2. Unique non-coding sequence among the MRSA252, MSSA476 and 14 other MRSA strains.

Outputs	MRSA252-MSSA476	Non coding MRSA252	Conserved non coding 14 MRSA strains
Unique sequence < 500 bp	92	225	139
Unique sequence length between 500-1000 bp	25	63	36
Unique sequence length between 1000-1500 bp	10	29	29
Unique sequence length > 1500 bp	53	75	70
Total :	180	392	274
Mean length:	2735.17bp	1010.46bp	1324.06bp

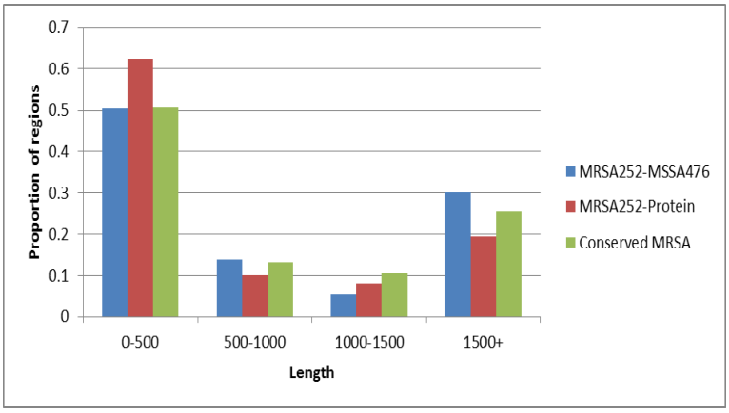

Fig. 2. Vertical data on bar diagram represent the number of regions in percentage and the horizontal data represent the corresponding length. This is a comparative analysis of how the sequence length and region number varied during the screen out process. Blue bar shows the result of first analysis where only MRSA252 sequences were present without MSSA476 common sequences. Red bar shows the non-coding sequences of MRSA252 and the green bar shows the conserved non-coding sequences of MRSA strains.

Elaborate search for conserved MRSA non-coding sequences, 362,792bp sequences with 100% identity showed that these sequences were almost the same as the MRSA252 non-coding sequences with very small differences in lengths and also were consistent with the previously published data. An interesting aspect of this study result was the length of the non-coding unique sequences, which were more than the average length found in other bacteria. Compared to eukaryotes, prokaryotes contain very less number of non-coding sequences to avoid complexity in the inter-genic regions and use the maximum spaces in the cellular compartment. The non-coding sequence can be deleted or modified depending on the necessity of the organism which includes the actions in case of selective pressure

for antibiotic resistance (Comeron 2001). More than average non-coding sequences found in this study suggest that these sequences might act as regulatory elements and/or can perform roles on protein folding (Taft *et al.* 2010).

In this study most of the strains matching with MRSA conserved sequences were from *Staphylococcus* family. Three species were selected from the matching sequences for functional analysis. Of the three selected species two belonged to different species (*Salmonella enterica* subsp. *entericaserovarSchwarzengrund* str. CVM19633, and *Streptococcus gallolyticus* UCN34) and one was *Staphylococcusaureus* subsp. *aureus* MSSA476 plasmid pSAS.

Results from ACT (Artemis Comparison Tool) show that the BLAST search matched (Fig. 3) MRSA252 in region between 39,658 bp to 97,899 bp and in *Salmonella enteric* subsp. *enteric serovarSchwarzengrund* str. CVM19633 in region between 46,011 bp to 46,588 bp of their complete genome with 100% identity. This matching comes about just after the *MecA* gene of MRSA252 and after SeSa_B0001 locus in *Salmonella enteric* CVM19633.

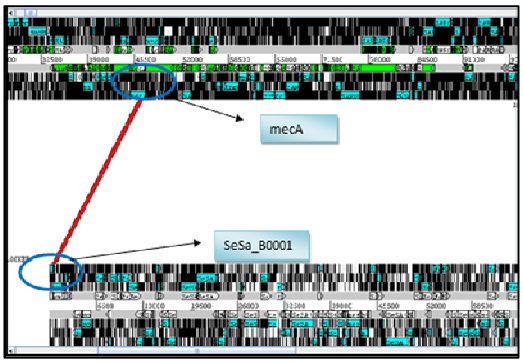

Fig. 3. *Salmonella enteric* CVM19633 genome and MRSA252 genome compared with BLAST result. Bluecircles show the nearby genes of the target non-coding sequence.

Blast result for MRSA252 matched with MSSA476pSAS plasmid with 100% identity (Fig. 4) in regions extending from 743,760 bp to 787,721 bp. The regions are from 3,729 bp to 3,770 bp and 7,735 bp to 7,766 bp with 100% identity and score 42 and 32 correspondingly. MSSA476 plasmid with a nearby gene locus pSAS04 was found to match with the first sequence whereas no known gene locus was found close to match with MRSA252 strain. The second match between 7,735 bp to 7,766 bp of the non-coding sequence of MSSA476 plasmid had the gene locus pSAS13 nearby and that matched with MRSA252 non-coding regions near to SAR0007 locus (Fig. 4).

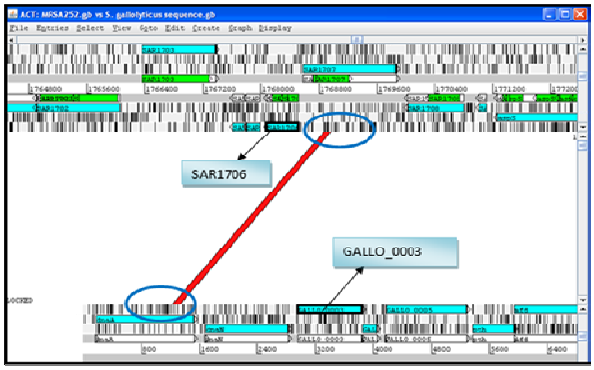

Fig. 4. Comparison between the genomes of MSSA476 pSAS and MRSA 252. Blue circles indicate the nearby locus of the non-coding matching sequences.

Match between MRSA252 and *Streptococcus gallolyticus* found in region 37,077 bp to 93,774 bp and in regions between 1,220 bp and 1,320 bp.Gallo_0003 gene loci was found in the close proximity of *S. gallolyticus* and in MRSA252 the nearby locus found was SAR1706 (Fig. 5).

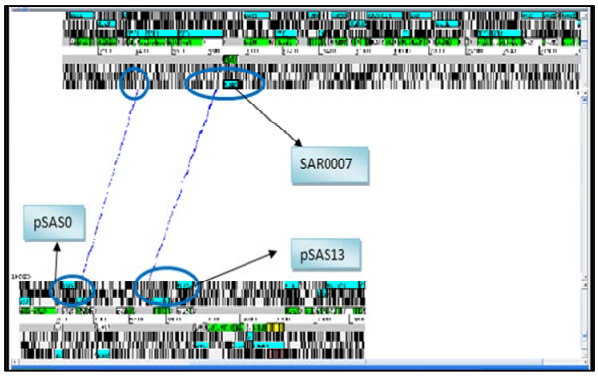

Fig. 5. Comparison between *S. gallolyticus* genome and MRSA252 genome. Blue circles indicate the nearby locus of the non-coding matching sequences.

A simple tabular comparative presentation (Table 3) of the three selected species with MRSA252 shows their gene/locus name, accession number with protein products.

InterProScan analysis results show the biological functions of the genes found at selected regions with the GO (Gene Ontology) terms. Table 4 represents a list of gene/loci involved in biological processes, molecular functions, cellular processes and name of other species where these genes can be found.

Table 3. Comparative matching list of three pathogens with MRSA252.

Query Species: *S. aureus* subsp. MRSA252				Matched Species: *Salmonella enteric* subsp. enteric *serovarschwarzengrund* str. CMV19633 plasmid pCMV19633_110			
Accession	Region	Locus/ gene	Protein product	Accession	Region	Locus/gene	Protein product
NC_002952	39658-97899	*MecA*	Penicillin binding protein	NC_011092	46011-6588	SeSa_B0001	protein SamA
				Matched Species: S. *aureus* subsp. *aureus* MSSA476 plasmid pSAS,			
	743760-787721	-	-	NC_005951	3729 - 3770	Nearby locus: pSAS04	Similar to *Neisseria gonorrhoeae* replication
	743760-787721	SAR0007	Similar to Lactococcuslactis hypothetical protein ycfG		7735-7766	Nearby locus: pSAS13	hypothetical protein
				Matched species: *Streptococcus gallolyticus* UCN34, complete genome			
	37077-93774	SAR1706	Similar to Bacillus halodurans hypothetical protein BH1259	NC_013798	1220-1320	Nearby locus: GALLO_0003	diacylglycerol kinase

Table 4. List of functions of the genes found near to the non-coding regions.

Gene/locus	Protein product	Biological process	Molecular function	Cellular component	Species
MecA	Penicillin binding protein	GO: 0009273: peptidoglycan-based cell wall biosynthesis, GO:00046677: response to antibiotics	GO:0008658: Penicillin binding	-	*E.coli*, *S.aureus*
SeSa_B0001	SamA	-	GO:0003677: DNA binding,	-	*S. enteric* CMV19633
pSAS04	Replication protein	GO:0006270: DNA-dependent DNA replication	GO:0003887: DNA-directed polymerase activity	GO: 0005727: Extra chromosomal circular DNA	MSSA476 plasmid pSAS
pSAS13	Hypothetical protein				MSSA476 plasmid pSAS
SAR0007	ycfG	Carbohydrate related kinase			*Lactococcuslactis*,
SAR1706	Hypothetical protein BH1259				*Bacillus halodurans*,
GALLO_0003	Diacylglycerol kinase	Activation of protein kinase C activity by G-coupled receptor protein signalling pathway	Diacylglycerol kinase activity		*Streptococcus gallolytcus*

This result indicates that most of the unique sequences found in MRSA252 belong to the non-coding region and the small percentage that is found in the coding region code for the *MecA* gene, the major gene responsible for the resistance to antibiotics. Other than the *MecA* gene in the coding region, this study did not find any known function of the non-coding regions which might be associated with other genes or gene products for defined functional purposes.

Functional information provided by the Inter ProScan and GO depicts *MecA* gene as a penicillin binding protein which changes the conformation of protein-drug complex upon binding with antibiotics from the penicillin family. Other function of *MecA* includes peptidoglycan-based cell wall synthesis in organisms like *E. coli*. Protein product locus SeSa_B0001 found to be matched in *Salmonella enteric* plasmid codes for SamA which is a DNA binding protein that can bind with single or double stranded DNA.

Two matched areas with 100% identity between MRSA252 and *Staphylococcus aureus* subsp. *aureus* MSSA476 plasmid pSAS show that even though there were matched sequences they did not share any functional similarities between species. The first match pSAS04, which was found nearby the gene locus of a non-coding region in MSSA plasmid, was defined as a replication protein having polymerase activity in DNA directed manner and resided in an extra-chromosomal circular DNA. For this matched region no noticeable gene locus was found in MRSA252 strain. pSAS13 locus was the second matched region that was found in MSSA476 pSAS plasmid and no defined function was reported from that region. In case of MRSA252 strain the matched region was SAR0007 which coded the protein ycfG. The function of this protein was found to be a carbohydrate kinase in different *Lactococcuslactis* strains. Through BLAST analysis a single match was found between MRSA252 and *Streptococcus gallolyticus* with 100% identity where Gallo_0003 gene locus of *S. gallolyticus* and *SAR1706* locus of MRSA252 strain, respectively. Gallo_0003 has molecular function as a diacylglycerol kinase which is a secondary messenger in signaling pathway. It activates protein kinase C by coupling with G protein. SAR1706 locus is defined as transcription regulator Rrf-2 with no further information.

Though the non-coding sequence matched with different regions among the selected species, in every case they were found near genes which are involved in different molecular and cellular functions. The result from these comparisons among the different species suggests that there is no functional relationship among them and also with MRSA252 strain. Although non-coding sequences compared within the MRSA strains found similar functions, no relation was found with other pathogens. It would have been more useful if common non-coding sequences were found in other species and as well as in MRSA strains; implying a specific and universal biological or regulatory function of that specific region.

Even though no regular functional correlation was found among the pathogens, these unique non-coding sequences can be used for detection and diagnostic purposes with higher specificity and accuracy. Further investigation of these novel non-coding sequences might reveal unknown information about the MRSA species. Studies searching open reading frames and then comparing them for functional analysis might provide new insight about the functions and pathogenicity as well. Understanding and information from the in-silico laboratory can ease the process of investigation of these pathogens in wet laboratory where DNA microarray based studies could be another approach for studying Methicillin Resistant *Staphylococcus aureus* (MRSA).

References

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.

Am J Infect Control. 1999. National Nosocomial Infection Surveillance (NNIS) System. *NNIS System report, data from January 1990–May 1999*. **27**:520–532.

Brown, Derek F. J., D. I. Edwards, P. M. Hawkey, G. L. Ridgway, K. J. Towner, Michael W. D. Wren and D. Morrison. 2005. Guidelines for the laboratory diagnosis and susceptibility testing of methicillin-resistant *Staphylococcus aureus* (MRSA). *J. Antimicrob. Chemotherapy*. **56**(6): 1000–1018.

Carver T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell and J. Parkhill. 2005. *ACT: the Artemis Comparison Tool. Bioinformatics*. **21**(16): 3422-3423.

Comeron, J.M. 2001. What controls the length of non-coding DNA? Curr.Opin.Genet. Dev. **11**: 652-659.

Dequaire, M., C. Degrand, and B. Limoges. 1999. An Immunomagnetic Electrochemical Sensor Based on a Perfluorosulfonate-Coated Screen-Printed Electrode for the Determination of 2, 4-Dichlorophenoxyacetic Acid. *Anal.Chem.* **71** (13): 2571–2577.

Hiramatsu, K. 2004. Elucidation of the mechanism of antibiotic resistance acquisition of methicillin-resistant *Staphylococcus aureus* (MRSA) and determination of its whole genome nucleotide sequence. *J.M.A.J.* **47**(4): 153–159.

Holden, M. T. G., E. J. Feil, J. S. Peacock, N. P. J. Day, M. C. Enright, T. J. Foster, C.E. Moore, L. Hurst, R. Atkin, A. Barron, N. Bason, S. D. Bentley, D., S, C. Chillingworth, T. Chillingworth, C. Churcher, L. Clark, C. Corton, A. Cronin, J. Doggett, L. Dowd, T. Feltwell, Z. Hance, B. Harris, H. Hauser, S. Holroyd, K. Jagels, K. D. James, N. Lennard, A. Line, R. Mayes, S. Moule, K. Mungall, D. Ormond, M. A. Quail, E. Rabbinowitsch, K. Rutherford, M. Sanders, S. Sharp, M. Simmonds, K. Stevens, S. Whitehead, B. G. Barrell, B. G. Spratt, J. Parkhill and J. A. Lindsay. 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *PNAS*. 9786–9791. **101**(26): 9786-9791.

J Hosp Infection. 1998. *Working Party Report: Revised guidelines for the control of methicillin-resistant Staphylococcus aureus infection in hospitals*. **39**:253–290.

Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T. L. Madden. 2008. NCBI BLAST: a better web interface. *Oxford Journals Life Sciences Nucleic Acids Research*. **36**: W5-W9.

Kasai, S., H. Zhou, M. Nishizawa, K. Niwa, T. Onouchi, T. Matsue and Y. Akiko. 2000. Immunoassay of the MRSA-related toxic protein, leukocidin, with scanning electrochemical microscopy. *Analytical Chemistry*. **72(**23): 5761-5765.

Kitagawa, Y., M. Ueda, N. Ando, K. Ishibiki, Y. Kobayashi, T. Arai, M. Kitajima and M. Endo. 1996. Rapid diagnosis of methicillin-resistant *Staphylococcus aureus* bacteremia by nested polymerase chain reaction. *Ann Surg*. **224**(5): 665–671.

Korf, I., M. Yandell, J. Bedell, 2003. BLAST. *O'Reilly & Associates, Inc. Sebastopol*, CA, USA.

Locey, K. J. and E. P. White. 2010. Simple structural differences between coding and non-coding DNA. *PLoS ONE.* **6**(2): 14651.

Lyon B. R. and R. Skurray. 1987.Antimicrobial resistance of *Staphylococcus aureus*: genetic basis. *Microbiol. Rev*. **51**(1):88–134.

Mercer, T. R., M. E. Dinger and J. S. Mattick. 2009. Long non-coding RNAs: insights into functions. *Nature reviews*, *Genetics*. **10**:155-159.

Neu, H., C. 1992. The Crisis of antibiotic resistance. *Science*. **257**: 1064- 1073.

Oberdorfer, K., S. Pohl, M. Frey, K. Heeg and C. Wendt. 2006. Evaluation of a single locus real-time polymerase chain reaction as a screening test for specific detection of methicillin-resistant *Staphylococcus aureus* in ICU patients. *Eur. J. Clin. Microbiol. Infect. Dis*. **25**:657–663.

Ponting, C. P., P.L. Oliver, W. Reik. 2009. Evolution and functions of long noncoding RNAs. *Cell*. **136(**4)**:** 629–641.

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler and R. Lopez. 2005. InterProScan: protein domains identifier. *Oxford Journals Life Sciences & Nucleic Acids Research*. **33**. Issue suppl 2Pp. W116-W120.

Rogozin, I. B., K. S. Makarova, D. A. Natale, R. L. Tatusov, Y. I. Wolf, J. Yin, E. V. Koonina and A. N . Spiridonov. 2002. Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. *Nucleic Acids Research.* **30**(19): 4264-4271.

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream and B. Barrell. 2000. *Bioinformatics* (Oxford, England); **16**(10): 944-945; see also: (www.sanger.ac.uk/ resources/software/artemis/)

Shrestha, N. K., M. J. Tuohy, G. S. Hall, C. M. Isada and G. W. Procop. 2002. Rapid identification of *Staphylococcus aureus* and the *mecA*Gene from BacT/ ALERT blood culture bottles by using the lightcycler system. *J. Clin. Microbiol*. **40** (7): 2659–2661.

Strommenger B., K. Christiane, W. Guido and W. Wolfgang. 2003. Multiplex PCR assay for simultaneous detection of nine clinically relevant antibiotic resistance genes in *Staphylococcus aureus*. *J. Clin. Microbiol.* **41**: 4089–4094.

Tisdall, J., 2001. *Beginning Perl for Bioinformatics: An Introduction to Perl for Biologists*. Publisher: O'Reilly Media. p. 384.

Thornsbery, C., L. K. McDougal and J. Clin. 1983. *Microbiol*. **18**: 1084−1091.

Taft, R. J., K. C. Pang, C. K. Pang, T. R. Mercer, M. Dinger and J. S. Mattick. 2010. Non-coding RNAs: regulators of disease. *J. Patho*. **220**: 126–139.

Weigel, L. M., D. B. Clewell, S. R. Gill, L. K. McDougal1, E. S. Flannagan, J. F. Kolonay, J. Shetty, E. G. Killgore, C. F. Tenover, and N. C. Clark. 2003. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureu*s. *Scienc*e. **302:** 1569-1571.

(Received revised manuscript on 26 May 2013)