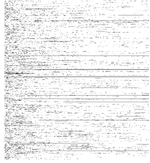

Research in Microbiology 164 (2013) 913–922

www.elsevier.com/locate/resmic

Short repeats in the *spa* gene of *Staphylococcus aureus* are prone to nonsense mutations: stop codons can be found in strains isolated from patients with generalized infection

Vladislav Victorovich Khrustalev [a], Ehsanollah Ghaznavi-Rad [b,*], Vasanthakumari Neela [c], Mariana-Nor Shamsudin [c], Alireza Amouzandeh-Nobaveh [b], Eugene Victorovich Barkovsky [a]

[a] *Department of General Chemistry, Belarusian State Medical University, Dzerzinskogo 83, Minsk, Belarus*
[b] *Department of Medical Microbiology and Immunology, Faculty of Medicine, Arak University of Medical Science, Arak, Basij Square 38481-7-6941, Iran*
[c] *Department of Medical Microbiology and Parasitology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang 43400, Malaysia*

Received 18 September 2012; accepted 8 April 2013
Available online 13 July 2013

Abstract

Fifteen sequences with stop codons have been obtained in the course of standard methicillin-resistant *Staphylococcus aureus* (MRSA) *spa* typing. In nine of those sequences, stop codons occurred due to nonsense G–T and A–T transversions. G–T transversions would appear to be frequent in the *spa* gene, mostly due to symmetric mutational AT-pressure in the whole *S. aureus* genome and due to replication-associated mutational pressure characteristic of lagging strands of the "chromosome". A–T transversions would appear to be frequent in the *spa* gene mostly due to transcription-associated mutational pressure. Relative to other *S. aureus* genes, short repeats in *spa* are enriched by nonsense sites for G–T and A–T transversions; the probability of being nonsense for A–T transversion is high in that part of *spa* coding region. 13 out of 15 (87%) of the sequences with stop codons were obtained from strains isolated from patients with generalized *S. aureus* infection. Truncation of *spa* at its C-terminus is predicted to result in a protein that possesses functional IgG binding domains unable to be linked to the cell wall. This is discussed in light of the known fact that extracellular *spa* is a strong virulence factor involved in immune evasion.
© 2013 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Preterminal codons; Truncation; Mutational pressure

1. Introduction

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a human pathogen that causes both healthcare-associated and community-associated infections. Localized MRSA infection results in the inflammation of skin and soft tissues. Generalization of MRSA infection leads to bacteremia and inflammation of inner organs. Generalized MRSA infection is known to be associated with considerable morbidity and mortality (Cosgrove et al., 2005). At the present time, such molecular typing methods as multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE) and staphylococcal protein A typing (*spa*-typing) are available for source tracing and comparison of MRSA isolates from different regions. In general, ''band-based'' and ''sequence-based'' methods can be used to investigate the genetic background of MRSA. However, ''sequence-based'' methods like *spa*-typing are preferred, because the data are exchangeable between laboratories from different countries (Harmsen et al., 2003).

Harmsen et al. (2003) presented a novel computerized tool for rapid determination of *spa* repeats. The greatest advantage of the tool lies in the fact that this software automatically recognizes repeats, classifies them and submits the data to the database "Ridom SpaServer" which is freely accessible via internet (http://www.spaServer.ridom.de). Although less

* Corresponding author. Department of Medical Microbiology, Faculty of Medicine, Arak University of Medical Sciences, Arak, Basij Square 38481-7-6941, Iran. Tel.: +98 8614173526.
E-mail address: e.ghaznavirad@arakmu.ac.ir (E. Ghaznavi-Rad).

0923-2508/$ - see front matter © 2013 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.resmic.2013.07.003

discriminatory power is shown for *spa*-typing compared to PFGE, it showed good strain differentiation, which is sufficient for determination of patient to patient transmission or transmission among wards in short epidemiological studies (Harmsen et al., 2003).

The Spa protein is encoded by a 2.15 kb gene. There are several structural parts in the Spa protein: 1) a signal peptide; 2) from four to five long repeats, each of which functions as an IgG binding domain (each of them is encoded by a 159 bp region); 3) several short repeats usually eight amino acids in length known as the polymorphic X-region; 4) a single copy of a LysM domain followed by an "LPXTG" sorting signal responsible for the linking of Spa to the cell wall; and 5) a C-terminus removed by sortase (Brigido Mde et al., 1991; Shopsin et al., 1999). It is widely accepted that Spa is one of the most important *S. aureus* virulence factors (Yung et al., 2011).

IgG binding domains bind Fc fragments of immunoglobulins allowing *S. aureus* to escape phagocytosis (Shopsin et al., 1999). Bacteria become coated with IgG in an inappropriate conformation which cannot be recognized by Fc receptors on neutrophils. However, IgG binding domains are able to bind other proteins as well. It was shown that each of the IgG binding domains is able to bind and activate tumor necrosis factor α (TNF-α) receptor 1, finally provoking anti-inflammatory consequences (Gómez et al., 2006). Moreover, Spa can bind TNF-α receptors on pre-osteoblasts and promote apoptosis of those cells, preventing new bone formation during osteomyelitis (Widaa et al., 2012). It was also shown that Spa increases histamine, leukotriene, and prostaglandin D(2) levels due to mast cell degranulation in mucous tissues from nasal polyps after 30 min incubation (Patou et al., 2008). An inflammatory response in human corneal epithelial cells can also be induced by Spa (Kumar et al., 2007). The full-length recombinant IgG binding region of Spa was shown to be able to bind von Willebrand factor and consequently to help *S. aureus* to adhere to platelets (O'Seaghdha et al., 2006).

The exact function of the X-region is unknown. It is thought that this region may serve to extend the N-terminal IgG-binding portion of the protein through the cell wall (Shopsin et al., 1999). Since the number of short repeats in that region is highly variable and amino acid substitutions are frequently observed, the region coding for that domain has been chosen for typing *S. aureus* strains (Frenay et al., 1996).

In this work, we report on sequences coding for the Spa X-region that display terminal codons. We also discuss the four following questions. 1) Is a *S. aureus* strain carrying truncated Spa able to cause generalized infection? 2) What is the most frequent cause of truncations among sequences from this study? 3) Is the *spa* gene more prone to truncations than other genes from the *S. aureus* genome? 4) Is there positive selection, which can be detected by the distance-based method in the whole *spa* gene or in the part encoding the X-region?

These questions were answered in the following way. 1) Most strains with mutations in regions coding for short repeats of *spa*, which should lead to truncation at the C-terminus of the resulting protein, were isolated from blood (73%), as well as from cerebrospinal fluid (7%) and peritoneal fluid (7%). This means that those strains caused generalized infection. 2) Most of the sequences were truncated due to nonsense G–T and A–T transversions. 3) We showed that the *spa* region that encodes for short repeats is prone to nonsense G–T and especially A–T transversions compared to most other genes from the *S. aureus* genome. G–T and A–T transversions are frequent in the whole genome of *S. aureus*, whereas usage of nonsense sites for them is high, particularly in the short repeats from *spa*. 4) Positive selection was not detected by the Kumar method in *spa* or in its part encoding the X-region (Nei and Kumar, 2000). This means that the general rate of non-synonymous mutation fixation is significantly lower than the rate of synonymous mutation fixation in *spa*. On the other hand, positive selection of bacterial cells possessing nonsense mutations in the region of *spa* encoding the X-region may take place during periods of acute generalized infection.

Possible consequences of truncated *spa* translation have been discussed. It is possible that an extracellular Spa protein might be more dangerous for the host than a Spa protein covalently attached to the bacterial cell wall (Yung et al., 2011).

2. Materials and methods

2.1. Materials

To determine the molecular epidemiology of MRSA in a central teaching hospital of Kuala Lumpur, 389 clinical isolates were collected from October 2007 to September 2008 (Ghaznavi-Rad et al., 2010). All isolates were subjected to *spa* typing according to the method of Shopsin et al. (1999).

For calculation of synonymous and non-synonymous evolutionary distances by the Kumar method (Nei and Kumar, 2000), sequences coding for full-length Spa proteins from 15 completely sequenced *S. aureus* genomes were used: Mu 50 (NC_002758); TCH60 (NC_017342.1); MRSA252 (NC_002952.2); M013 (NC_016928.1); ED133 (NC_017337.1); LGA251 (NC_017349.1); 11819-97 (NC_017351.1); MW2 (NC_003923.1); str. Newman (NC_009641.1); USA300_TCH1516 (NC_010079.1); TW20 (NC_017331.1); COL (NC_002951.2); N315 (NC_002745.2); 04-02981 (NC_017340.1); Mu3 (NC_009782.1). Calculation of bioinformatic indices is described in Section 2.5.

2.2. Amplification of the spa gene

The short sequence repeats (SSR) located in the X region of *spa* were amplified using the following primers (Shopsin et al., 1999): 5′-AGACGATCCTTCGGTGAGC-3 (forward); 5′-GCTTTTGCAATGTCATTTACTG-3′ (reverse). A ready-to-use PCR master mixture containing Taq DNA polymerase 0.1 unit/ul, dNTPs 1 mM, glycerol 10%, $MgCl_2$ 7.5 mM (iDNA Biotechnology Sdn Bhd, Malaysia) was used for PCR amplification. PCR amplification was carried out as described by Harmsen et al. (2003) using the following cycling parameters: initial denaturation of 10 min at 95 °C, followed by 30 cycles of denaturation for 30 s at 95 °C, 30 s annealing at

60 °C and 45 s extension at 72 °C and a final extension at 72 °C for 10 min. ATCC 25923 was used as positive control and sterile deionized water as negative control. Amplified PCR products were visualized and interpreted after agarose gel electrophoresis (60 V, 500 mA, duration of 60 min) on 1% agarose gel. Repeated negative results were categorized as non-typeable strains by *spa* typing.

2.3. PCR purification and sequencing

PCR purification was done to obtain a clean amplicon for sequencing purposes. PCR purification was carried out using the GF-1 PCR Clean-up kit (Vivantis, BiosynTech Sdn Bhd, Malaysia) according to instructions of the manufacturer.

Purified PCR products were sequenced (DNA sequencer ABI, model: 3730-XL) commercially (First base Laboratories Sdn Bhd) using the *spa* primers mentioned above. The sequences obtained were further subjected to in silico analyses.

2.4. Analyses of sequences

The usage of nonsense sites for G–T and A–T transversions and probabilities for being nonsense for G–T and A–T transversions were calculated in sliding windows 100 codons in length along the *spa* coding region from the *S. aureus* strain Mu50 complete genome (Kuroda et al., 2001) with the help of the "VVK Sliding Window" algorithm (Khrustalev et al., 2012) (www.barkovsky.hotmail.ru). Nucleotide usage in three codon positions for each gene from the complete genome of *S. aureus* Mu50 strain was calculated with the help of the "Chore Viewer" algorithm (Khrustalev and Barkovsky, 2011) (www.barkovsky.hotmail.ru). That algorithm used a list with quantities of each codon in each gene for that completed genome from the Codon Usage Database (Nakamura et al., 2000) (www.kazusa.or.jp/codon) as an input.

Indices including (i) usage of nonsense sites for G–T transversions, (ii) usage of nonsense sites for A–T transversions, (iii) the probability of being nonsense for the G–T mutation and (iv) the probability of being nonsense for A–T mutation were calculated for each coding region from the 15 completely sequenced genomes of *S. aureus* strains listed in Section 2.1. To check whether those indices are distributed normally among genes from each genome, we applied the set of normality tests (Shapiro–Wilk; Jarque–Bera; Chi-square) available in the PAST program (http://folk.uio.no/ohammer/past).

Since the distribution was not normal for all indices in each of the 15 genomes, medians were calculated for them. Interestingly, 15 genomic medians of the four indices were distributed normally. Those genomic medians were compared with the help of a paired *t*-test with subsequent indices calculated for parts of *spa* genes encoding X-regions.

Regions coding for short repeats of the *spa* were separated from the rest of the sequences according to the description of Mu50 *spa* (P0A015) from the Uniprot (www.uniprot.org) database. Use of nonsense sites for G–T and A–T transversions, as well as probabilities of being nonsense for G–T and A–T transversions, were calculated in those sequences with the help of the "VVK Protective Buffer" algorithm (Khrustalev et al., 2012) (www.barkovsky.hotmail.ru) which uses nucleotide sequences as an input.

Fifteen full-length *spa* sequences from complete genomes had been aligned by the PAM algorithm included in the MEGA5.1 program (Tamura et al., 2011). Then non-synonymous (DN) and synonymous (DS) evolutionary distances between each pair of those sequences were calculated by the Kumar method (Nei and Kumar, 2000). The hypothesis that DN is significantly higher than DS was tested for each pair of sequences by the codon-based *Z*-test of selection. The average ratio between DS and DN (DS/DN) was calculated for all pairs of sequences in which DN was higher than zero. The same kind of analysis was performed for the alignment containing just parts of *spa* genes encoding X-regions only.

Average levels of nucleotide usage in fourfold- (T4f; A4f; C4f and G4f) and twofold-degenerated sites (T2f3p; A2f3p; C2f3p and G2f3p) from third codon positions had been calculated in 15 *spa* sequences from complete genomes of *S. aureus* strains with the help of the "VVK Protective Buffer" algorithm (Khrustalev et al., 2012) and compared with each other by a two tailed *t*-test.

SPSS 16.0 was used to perform Chi-square tests in cross tables.

3. Results

3.1. Features of Spa sequences containing stop codons

The frequencies of MRSA isolates obtained from different types of specimens are summarized in Fig. 1. Fifteen strains had mutations in the region coding for short repeats of the Spa protein, which should lead to synthesis of the product truncated at its C-terminus. The number of strains without stop codon mutations in the part of *spa* gene encoding the X-region was equal to 374. Most strains without stop codons in partial *spa* sequences were isolated from patients suffering from localized infection, while most strains with stop codons were isolated from patients with already developed generalized infection (Fig. 1). Two distributions represented in Fig. 1 were significantly different from each other according to the results of the Chi-square test ($P < 0.001$). Indeed, the frequency of strains isolated from blood according to the Chi-square test in the cross table was significantly higher ($P < 0.001$) for the group of strains with stop codons (73.33%) than for the group of strains without stop codons (13.37%) in the part of the gene encoding the X-region of Spa.

An example of the short repeat eight amino acids in length is shown in Fig. 2. An example of the nucleotide sequence encoding this short repeat is also shown in that figure.

Seven sequences had a single terminal codon that occurred due to G–T transversion. This transversion may be nonsense only in codon GAA (see the third codon in Fig. 2) coding for glutamic acid. Four of those sequences were obtained from strains isolated from blood. Three other sequences were from strains isolated from an abscess, peritoneal fluid and conjunctiva, respectively.

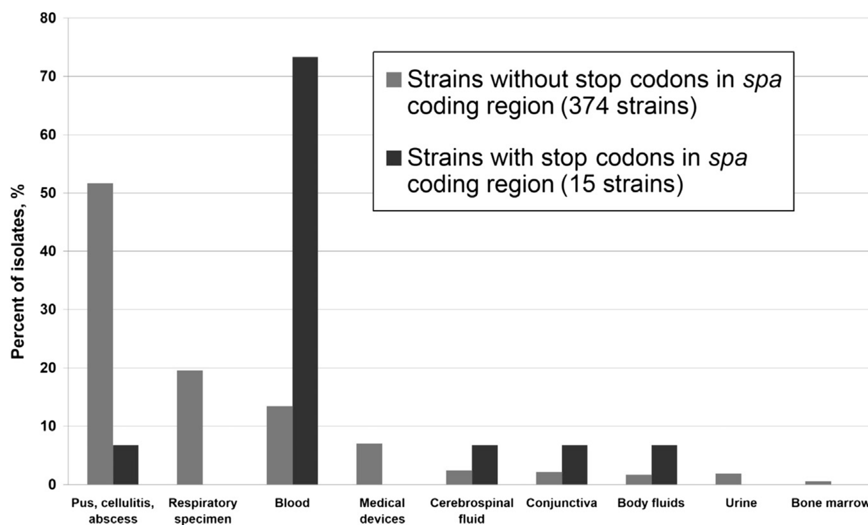

Fig. 1. Comparison between sources of isolation for MRSA with and without stop codons in the region of *spa* coding for short repeats. Types of specimens: pus, cellulitis, abscess, respiratory specimens (pulmonary secretion, sputum, and tracheal aspirate), medical devices (central nervous system tip and catheter), cerebrospinal fluid (CSF), conjunctiva, body fluids (peritoneal and synovial) and bone marrow. The difference between the two distributions is significant ($P < 0.001$) according to the Chi-square test.

In two sequences, multiple stop codons were found. In one of them, there was a stop codon mutation which occurred due to A–T transversion in codon AAA coding for lysine (see the second codon in Fig. 2), while two stop codons occurred due to G–T transversions. They took place in the second, third and fourth subunits of the repeat, respectively. Five other subunits of the repeat contained no stop codons. In the second sequence, four stop codons were found. Two of them appeared to be due to A–T transversions in the first position of codon AAA coding for lysine, which occupied the seventh position in the repeat from Fig. 2. Two other stop codons appeared to be due to G–T transversions. Thus, there existed four subunits of the repeat containing stop codons and four subunits without them. The first sequence was obtained from the strain isolated from blood, while the second sequence was obtained from the strain isolated from cerebrospinal fluid.

Six sequences obtained from strains isolated from blood also contained several stop codons, while their origin was not associated with nonsense nucleotide mutations. In one of them, deletion of the single nucleotide (T) occurred in the GGT codon coding for glycine, which is the first amino acid in the repeat from Fig. 2. In the second sequence, insertion of a single nucleotide (T) occurred in the same first codon of one of the repeat subunits. In the third sequence, four nucleotides (GACA) from one of the repeat subunits were replaced by three nucleotides (TTT). The fourth sequence, unlike three

G	K	E	D	G	N	K	P
GGC	**AAA**	**GAA**	GAC	GGC	AAC	**AAG**	CCT
	T	T				T	

Fig. 2. Example of the short repeat subunit from the *spa* X-region (*S. aureus* Mu50 strain, from amino acid #285 to amino acid #292). The nucleotide sequence encoding this subunit is provided on the second line. Pre-terminal codons are in bold and underlined. Mutations leading to formation of stop codons are shown.

previously described ones, contained multiple indels. The fifth sequence contained a relatively long insertion of unknown origin with a single stop codon. The sixth sequence also contained a relatively long insertion with a single stop codon that occurred due to an inverse repeat.

Theoretically, in case of a single ribosome "jump" in the region containing a premature stop codon, the *spa* protein containing its C-terminus with a sortase motif may be translated only from mRNAs which belong to three strains with single frameshift events. In case of a single ribosome "jump" on mRNAs from other strains, one closed open reading frame will be changed to another closed one.

3.2. Nucleotide usage biases in the spa *gene*

Here we should state that strains of *S. aureus* with truncated *spa* (or with no product from a mutated *spa* gene) may be isolated from patients suffering from generalized infection, and the main causes of those truncations are G–T and A–T nonsense mutations. To find out whether those two types of transversions are frequent in *spa*, we analyzed nucleotide usage biases of this gene.

In Table 1, use of thymine in fourfold-degenerated sites (T4f) was extremely high (56.43 ± 0.42%) in *spa*. The level of adenine usage in fourfold-degenerated sites (A4f) was also relatively high (27.52 ± 0.55%). Use of cytosine in fourfold-degenerated sites (C4f) is equal to just 10.76 ± 0.44%. The use of guanine in fourfold-degenerated sites (G4f) was even lower (5.30 ± 0.25%).

According to the data from Table 1, we suggest that transversions leading to thymine in place of guanine and adenine occur with higher frequency than transversions leading to thymine replacement by guanine and adenine (Sueoka, 1988). Moreover, G–T transversions would be especially frequent, since G4f is significantly lower than A4f.

Table 1
Average nucleotide use in fourfold- (A4f; T4f; G4f; C4f) and twofold-degenerated sites (A2f3p, T2f3p, G2f3p, C2f3p) from third codon positions of *spa* genes from 15 *S. aureus* genomes.

Nucleotide use in 4-fold-degenerated sites	A4f	T4f	G4f	C4f
	27.52 ± 0.55%	56.43 ± 0.42%	5.30 ± 0.25%	10.76 ± 0.44%
Nucleotide use in 2-fold-degenerated sites from third codon positions	A2f3p 48.81 ± 0.31%	T2f3p 17.96 ± 0.44%	G2f3p 5.52 ± 0.25%	C2f3p 27.70 ± 0.45%

The use of thymine in fourfold-degenerated sites may also have been high due to frequent cytosine to thymine transitions. To test this hypothesis, we calculated nucleotide usage biases in twofold-degenerated sites from third codon positions of *spa* (see Table 1).

As it can be seen in Table 1, the use of thymine in twofold-degenerated sites from third codon positions (T2f3p) was significantly lower (17.96 ± 0.44% vs. 27.70 ± 0.45%) than the use of cytosine (C2f3p). This indicates high T–C transitions rates: they would have to be higher than the rates of C–T transitions to produce the kind of bias represented in Table 1. Thus, the hypothesis of T4f growth due to C–T transitions can be rejected.

G–A transitions make a contribution to the decrease in G4f, since the use of guanine in twofold-degenerated sites from third codon positions (G2f3p) is much lower than the use of adenine (A2f3p), as can be seen in Table 1 (5.52 ± 0.25% vs. 48.81 ± 0.31%). In twofold-degenerated sites from third codon positions, A–T and G–T transversions (as well as all other transversions) are non-synonymous, so they are much less frequently fixed there than in fourfold-degenerated sites. This is why the use of T4f is extremely high, while the use of T2f3p is lower than levels of A2f3p and C2f3p.

In general, it can be stated that the most frequent types of transversions in *spa* would be G–T and A–T, and the most frequent types of transitions would be T–C and G–A. There are no nonsense sites for T–C transitions in the universal genetic code (Barkovsky and Khrustalev, 2009). There exist two sites for nonsense G–A transitions which are both in the TGG codon encoding tryptophan in the universal genetic code (Barkovsky and Khrustalev, 2009). Tryptophan is a rare amino acid which may be absent in many proteins such as *S. aureus* Spa. In other words, there are no nonsense sites for G–A transitions in the *spa* gene.

As one can see in Fig. 3, there exist nonsense sites for both G–T and A–T transversions in *spa*. Interestingly, the highest level of their use is characteristic to sliding windows 100 codons in length that include regions coding for 11 short repeats. Use of nonsense sites for A–T transversions is approximately two times higher than that of nonsense sites for G–T transversions at their highest peak (see Fig. 3). There exists another index necessary for estimating the risk of nonsense mutations. This index is the probability of being nonsense for a certain type of nucleotide mutation (Khrustalev and Barkovsky, 2011).

The probability of being nonsense for A–T transversion has its peak (about 18%) in the region coding for short repeats of Spa (see Fig. 4). This means that the region coding for short repeats is more prone to nonsense A–T mutations than other regions of that gene, due to both the more frequent use of nonsense sites for A–T transversions and the higher probability of being nonsense for that kind of transversion.

The probability of being nonsense for G–T transversion has its peak near codon #190, and not in the region coding for short repeats. On the other hand, the probability of being nonsense for G–T transversion (about 15%) is higher in the

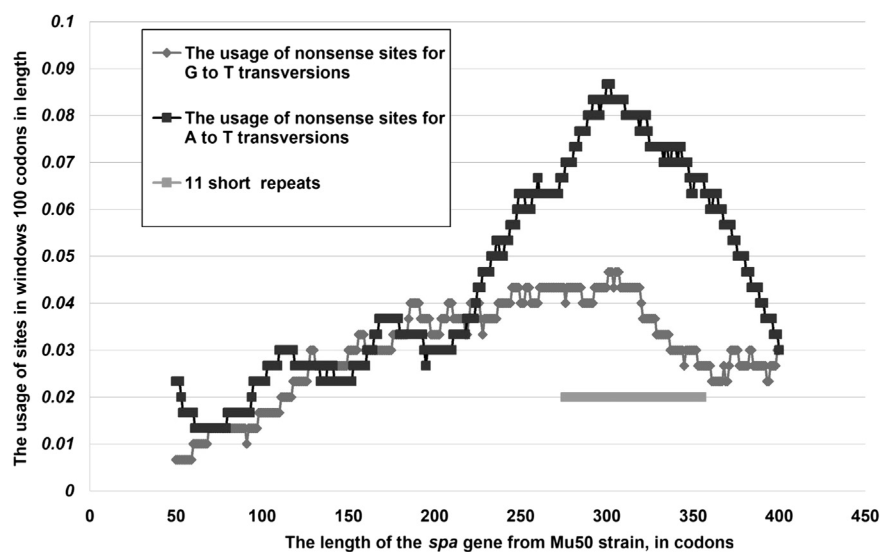

Fig. 3. Use of nonsense sites for G–T and A–T transversions along the length of *spa* from the *S. aureus* Mu50 strain. The length of a sliding window is equal to 100 codons; each step is equal to a single codon. Coordinates of the region encoding short repeats are shown.

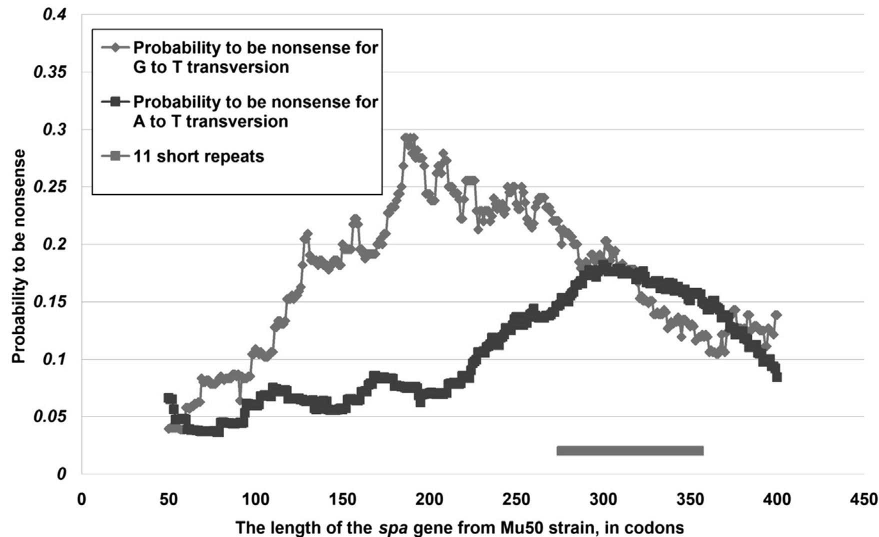

Fig. 4. Probability of being nonsense for G–T and A–T transversions along the length of *spa* from the *S. aureus* Mu50 strain. The length of a sliding window is equal to 100 codons and each step is equal to a single codon. Coordinates of the region encoding short repeats are shown.

region coding for short repeats than for the 5′-end of *spa* (see Fig. 4). Nevertheless, the region coding for short repeats demonstrates a high level of nonsense sites for G–T transversion use, while the probability of being nonsense for those transversions is not the highest with respect to the region coding for third and fourth long repeats.

3.3. *Strong mutational AT-pressure in the* S. aureus *genome*

There is strong symmetrical (strand-non-specific) mutational AT pressure in the genome of *S. aureus*, since GC content in third codon positions (3GC) is almost equally low (22.8 ± 0.2%) for all genes (Khrustalev et al., 2012). This is why 3GC shows weak correlation on G + C ($R = 0.293$) for them (see Fig. 5). In contrast, GC content in the first and second codon positions (1GC and 2GC, respectively) correlates with G + C (coefficients of correlation are equal to 0.726 and 0.721, respectively). Both 1GC and 2GC are usually higher than 3GC in genes from the *S. aureus* genome: negative selection eliminates most non-synonymous mutations, preventing decrease of 1GC and 2GC, and does not eliminate synonymous mutations which decrease 3GC. An especially high level of 1GC is characteristic of *spa* (see Fig. 5). This could be the consequence of tandem duplications: four out of eight codons from the short repeat subunit begin from guanine (see Fig. 2).

3.4. *Risk of nonsense mutation is higher for the part of the* spa *gene encoding the X-region than for most* S. aureus *genes*

The rates of G–T transversions are high in the whole genome of *S. aureus* because of general symmetric mutational

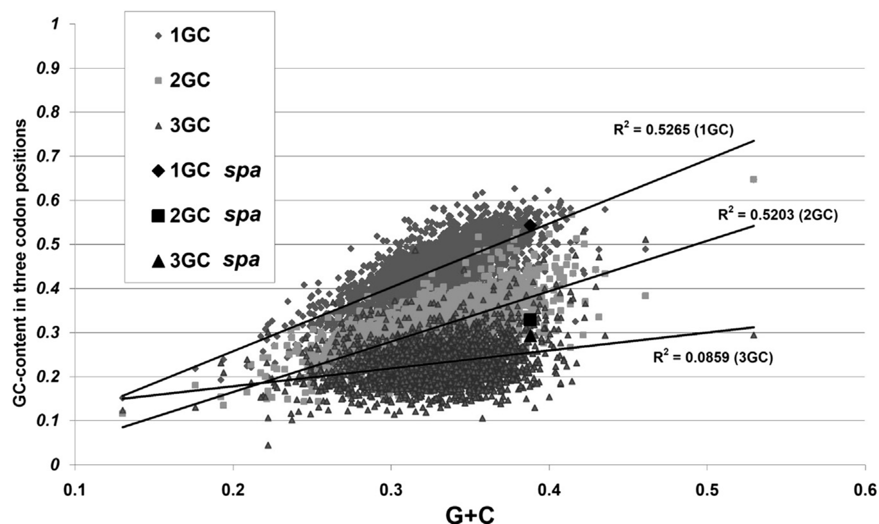

Fig. 5. Dependency between GC content in three codon positions (1GC, 2GC and 3GC) and total GC content (G + C) for all coding regions from the completely sequenced genome of the *S. aureus* Mu50 strain. Levels of 1GC, 2GC and 3GC for *spa* are shown.

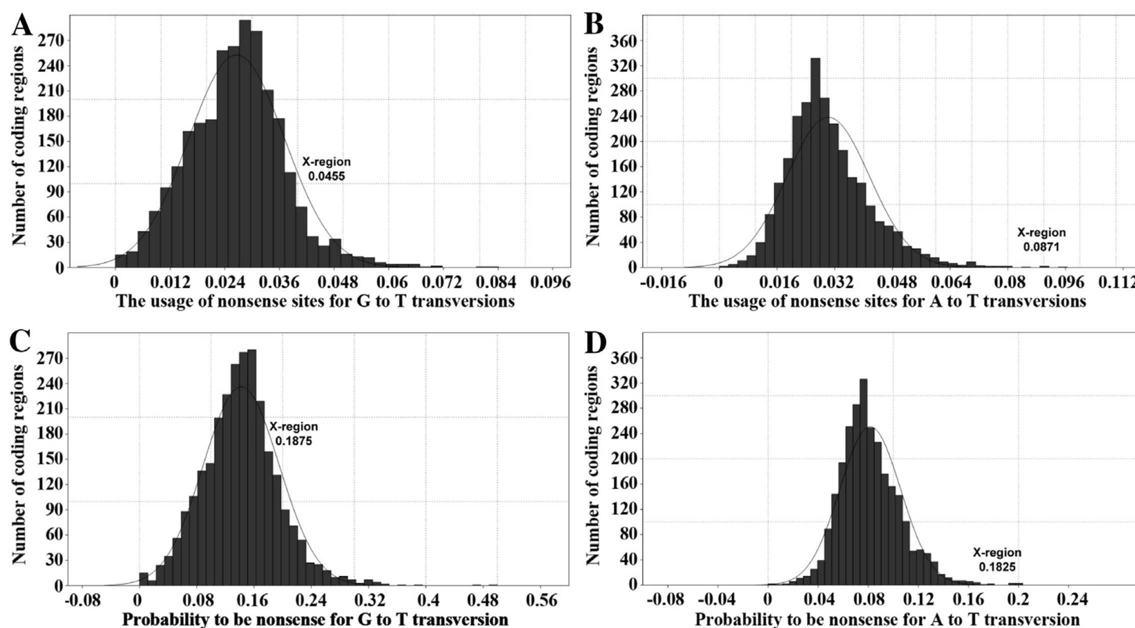

Fig. 6. Distribution of the use of nonsense sites for G–T transversions (A), use of nonsense sites for A–T transversions (B), probability of being nonsense for G–T transversion (C) and for A–T transversion (D) for all genes (including *spa*) from the completely sequenced genome of *S. aureus* Mu50. The above-mentioned indexes for sequences encoding short repeats from *spa* were added to the graphs.

AT pressure. Moreover, they are elevated in genes from lagging strands due to replication-associated mutational pressure (Khrustalev and Barkovsky, 2011). The rates of A–T transversion are high due to transcription-associated mutational pressure acting on all genes from *S. aureus* (Khrustalev and Barkovsky, 2011). Fig. 6A shows a histogram representing use of nonsense sites for G–T transversions. The distribution of this index is not normal for all genes from the *S. aureus* Mu50 genome, nor for genomes of the other fourteen strains used in this study. Use of nonsense sites for G–T transversions for the region encoding the X-region of *spa* (0.0459 ± 0.0003) is 1.7 times higher than the average median for all the genomes (0.0269 ± 0.0001). The paired difference between them is significant ($P < 0.001$).

Distribution of the use of nonsense sites for A–T transversion also cannot be considered normal (see Fig. 6B). Use of nonsense sites for A–T transversions for the region encoding the X-region of *spa* (0.0854 ± 0.0017) is 3.0 times higher than the average median (0.0283 ± 0.0001), and the paired difference is significant ($P < 0.001$).

Distributions of probabilities of being nonsense for the G–T transversion and for the A–T transversion in all genes of the complete genome of *S. aureus* strain Mu50 are shown in Fig. 6C and D, respectively. Neither distribution is normal. The probability of being nonsense for G–T transversion for the part of the *spa* gene encoding the X-region (0.1922 ± 0.0054) is 1.4-times higher than the average median for this index (0.1421 ± 0.0006). The probability of being nonsense for A–T transversion for the part of *spa* encoding the X-region (0.1782 ± 0.0024) is 2.3 times higher than the average median for this index (0.0783 ± 0.0002).

The risk of nonsense G–T transversion is elevated in the region of *spa* encoding eleven short repeats relative to other genes, mostly due to the significantly higher level of nonsense sites for G–T transversion usage.

The risk of nonsense A–T transversion is high in the region of *spa* encoding eleven short repeats compared to other genes because of the high use of nonsense sites for A–T transversions and because of the higher probability of being nonsense for those mutations.

The higher the number of short repeats in *spa* from a given strain, the higher the number of hot spots for nonsense mutations.

3.5. *Calculation of the DS/DN ratio for* spa *genes from completely sequenced* S. aureus *genomes*

According to results of the codon-based *Z*-test of selection, non-synonymous (DN) distances between full length genes coding for *spa* from 15 completely sequenced *S. aureus* genomes are no higher than synonymous (DS) distances. The average ratio between DS and DN for the full-length *spa* gene is equal to 9.41 ± 1.59. This provides evidence of strong negative selection acting on *spa*: the rates of fixation of non-synonymous mutations are nine times lower than the rates of that of synonymous mutations, even though the probability of synonymous GC–AT mutation occurrence is decreased in *spa* due to strong mutational AT pressure.

We also performed a codon-based *Z*-test for positive selection on alignment of *spa* X-regions from 15 strains of *S. aureus*. There were 55 amino acid positions without gaps in that alignment. Once again, DN was not significantly higher than DS; the average DS/DN ratio was equal to 6.23 ± 0.83. However, these results cannot enable us to reject the hypothesis that certain non-synonymous mutations (i.e. nonsense mutations) are being fixed by positive selection during

generalized infection. In our opinion, nonsense mutations in the X-region may become subjects for positive selection during the acute phase of infection or, at least, may escape negative selection during that period of the life cycle. In other words, mutants with stop codons in the region coding for the *spa* X-region should be able to survive in the infected organism during the acute phase of generalized infection. Moreover, those mutants may produce extracellular *spa* which should promote generalization of the infection.

4. Discussion

In the bioinformatics part of this study, we demonstrated that regions coding for short repeats of *spa* from *S. aureus* are more prone to nonsense G–T and A–T mutations than most other genes from that genome. Results of this part of the study enabled us to understand that truncated *spa* may often be found in different strains of *S. aureus* due to the high risk of nonsense mutations occurring under the influence of mutational pressure characteristic of this microorganism.

According to our data, lagging strands of *S. aureus* genomes accumulate more G–T than C–A transversions, in contrast to leading strands (Khrustalev and Barkovsky, 2010). Since *spa* is situated on the lagging strand, some part of G–T transversions in this gene may be caused by replication-associated mutational pressure (Khrustalev and Barkovsky, 2012). Using our methodology we showed that transcription-associated mutational pressure leads to an increase in T4f compared to A4f in genes from the *S. aureus* genome (Khrustalev and Barkovsky, 2011). According to these data, A–T transversions usually occur during transcription of *spa*. Both replication-associated and transcription-associated mutational pressures should introduce significant deviation from the normal distribution for indexes, which depend on nucleotide use biases such as for those represented in Fig. 6.

The cause of the increased probability of nonsense G–T and A–T mutations in the part of the gene encoding the X-region of *spa* is the presence of tandem repeats inside it. Biases in nucleotide composition characteristic to the part of the gene which somehow had been duplicated were inherited by other repeated subunits. The ongoing process of deletions and duplications of repeat subunits should make its contribution to stabilization of these nucleotide use biases: mutated subunits are often deleted. On the other hand, the presence of several repeated subunits may play an important role in the function of *spa*. One of the consequences of the existence of tandem repeats prone to nonsense mutations in *spa* may even be helpful for *S. aureus*. It may help this pathogen to establish generalized infection. Indeed, MRSA with truncated *spa* can be significantly more frequently found in samples from blood, bone marrow, cerebrospinal and other body fluids than MRSA with normal *spa* (86.67% vs. 17.91%, $P < 0.001$ according to Chi-square test).

Mortality among patients infected by MRSA with stop codon(s) in *spa* (33.33%) is not significantly higher than mortality among patients infected by MRSA without stop codons in that gene (14.97%), according to the Chi-square test ($P = 0.055$). At the time of sample collection, 13 patients were suffering from generalized infection caused by MRSA with truncated *spa* and 67 patients had generalized infection caused by MRSA with normal *spa*. The difference in mortality between these groups (38.46% vs. 16.42%) was also insignificant according to the Chi-square test ($P = 0.069$). One cannot state that MRSA with truncated *spa* gene is better tolerated by the host than MRSA with normal *spa*. It is also not clear whether infection with MRSA possessing truncated *spa* more often leads to a lethal outcome.

Since only partial sequences of *spa* were obtained in this study, it cannot be affirmed that there were no other stop codons in that gene among the 15 strains studied. In case of complete loss of *spa*, the corresponding *S. aureus* strain should be less pathogenic than others (Shopsin et al., 1999). However, most strains from our study were isolated from patients suffering from generalized infection. One possible explanation lies in the existence of some undiscovered compensatory mechanism (or mechanisms) making *spa*-deficient strains as pathogenic as *spa*-possessing ones. Another explanation is based on the fact that truncation of the studied Spa proteins from pathogenic strains occurred at their C-terminal parts and might not alter IgG binding domains. Theoretically, a Spa protein containing functional IgG binding domains, but lacking a sortase motif may be excreted by bacterial cells (since it cannot be linked to their cell walls).

Extracellular Spa protein has been detected in clinical samples from patients infected with *S. aureus* and in samples from a mouse model of skin abscess caused by MRSA (Yung et al., 2011). It was shown that 31 out of 42 host immunoregulatory chemokines (including the CXCL9 host chemokine) are responsible for cleavage of Spa from the bacterial cell wall (Yung et al., 2011). Therefore, this chemokine was added to the list of rare host signals beneficial to bacteria during infection. Release of a strong virulence factor such as Spa was thought to play an important role in immune evasion by *S. aureus* (Yung et al., 2011). Interestingly, synthetic Spa (which is not covalently anchored to the cell wall) induces biofilm development of *S. aureus* (Merino et al., 2009).

Two *S. aureus* strains isolated from *Bos taurus* were shown to possess truncated Spa due to nonsense G–T transversion in their last IgG binding domain (Stutz et al., 2011). Indeed, according to our data, the region coding for the last IgG binding domain is prone to G–T nonsense mutations due to the high probability of being nonsense for that type of transversion. Interestingly, the resulting protein was shown to have significantly impaired immunoglobulin binding affinity (Stutz et al., 2011). Theoretically, the IgG binding activity of truncated Spa encoded by genes from the 15 strains described in the present study should not be impaired in cases where they have no additional stop codons in regions coding for IgG binding domains.

There exist 564 types of short Spa repeats available via the "Ridom SpaServer" (Harmsen et al., 2003) (http://www.spaServer.ridom.de). Among them, there are 24 types of short repeats with terminal codons. There are 17 types of repeats possessing nonsense G–T transversions in codon #3

(See Fig. 2), four types of repeats with nonsense A–T transversions in codon #2 and three types of repeats with nonsense A–T transversions in codon #7. These data confirm that nonsense mutations in those short repeats are present in strains sequenced in other countries by other groups of researchers. However, possible Spa truncations due to those nonsense mutations and the molecular causes of such mutations have not been previously described or interpreted.

The most interesting feature of short tandem repeats is their ability to provoke frequent deletions and duplications. Once again, it should be emphasized that the number of repeat subunits in *spa* is highly variable among different strains. Hypothetically, a single repeat subunit possessing a nonsense mutation has a high probability of being spontaneously deleted or duplicated. This means that the length of translated Spa may vary for the same bacterial strain during the period of infection.

Finally, the hypothetical scenario of generalized *S. aureus* infection has been formulated. Infection may be caused by the strain with full-length Spa. During the time of infection, nonsense mutations occurring due to mutational AT pressure may cause truncation of Spa in certain bacterial cells. Those mutant cells may begin to produce extracellular Spa, making a substantial contribution to the spread and severity of the infection. In that period of time, positive selection may be favorable to those bacterial cells that produce truncated *spa* or negative selection may become weaker for them. Then, some of the progeny of those strains producing truncated Spa may return to synthesis of cell wall-anchored Spa due to spontaneous deletion of a mutated repeat subunit (or subunits). Since bacterial cells without anchored Spa are not protected from phagocytosis, the remaining extracellular Spa "producers" may be eliminated from the population of *S. aureus* by the immune system of the host during the recovery period.

Acknowledgments

The authors would like to thank Dr. Philip M. Murphy, Chief of the Laboratory of Molecular Immunology of the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA for his interest in the study and for useful corrections.

References

Barkovsky, E.V., Khrustalev, V.V., 2009. Inverse correlation between the GC content of bacterial genomes and their level of preterminal codon use. Mol. Gen. Microbiol. Virol. 24, 17–23.

Brigido Mde, M., Barardi, C.R., Bonjardin, C.A., Santos, C.L., Junqueira, M.L., Brentani, R.R., 1991. Nucleotide sequence of a variant protein A of *Staphylococcus aureus* suggests molecular heterogeneity among strains. J. Basic Microbiol. 31, 337–345.

Cosgrove, S.E., Qi, Y., Kaye, K.S., Harbarth, S., Karchmer, A.W., Carmeli, Y., 2005. The impact of methicillin resistance in *Staphylococcus aureus* bacteremia on patient outcomes: mortality, length of stay, and hospital charges. Infect. Contr. Hosp. Epidemiol. 26, 166–174.

Frenay, H.M.E., Bunschoten, A.E., Schouls, L.M., Van Leeuwen, W.J., Vandenbroucke-Grauls, C., Verhoef, J., Mooi, F.R., 1996. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. Europ. J. Clin. Microbiol. Infect. Dis. 15, 60–64.

Ghaznavi-Rad, E., Shamsudin, M.N., Sekawi, Z., Khoon, L.Y., Aziz, M.N., Hamat, R.A., Othman, N., Chong, P.P., van Belkum, A., Ghasemzadeh-Moghaddam, H., Neela, V., 2010. Predominance and emergence of clones of hospital-acquired methicillin-resistant *Staphylococcus aureus* in Malaysia. J. Clin. Microbiol. 48, 867–872.

Gómez, M.I., O'Seaghdha, M., Magargee, M., Foster, T.J., Prince, A.S., 2006. *Staphylococcus aureus* protein A activates TNFR1 signaling through conserved IgG binding domains. J. Biol. Chem. 281, 20190–20196.

Harmsen, D., Claus, H., Witte, W., Rothganger, J., Turnwald, D., Vogel, U., 2003. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. J. Clin. Microbiol. 41, 5442–5448.

Khrustalev, V.V., Barkovsky, E.V., 2010. The probability of nonsense mutation caused by replication-associated mutational pressure is much higher for bacterial genes from lagging than from leading strands. Genomics 96, 173–180.

Khrustalev, V.V., Barkovsky, E.V., 2011. Bioinformatical approaches for studies on replication-associated and transcription-associated mutational pressure, interpretations and applications. In: Urbano, K.V. (Ed.), 2011. Advances in Genetics Research, vol. 6. NOVA Publishers, New York, pp. 1–128.

Khrustalev, V.V., Barkovsky, E.V., 2012. A blueprint for a mutationist theory of replicative strand asymmetries formation. Curr. Genomics 13, 55–64.

Khrustalev, V.V., Arjomandzandegan, M., Barkovsky, E.V., Titov, L.P., 2012. Low rates of synonymous mutations in sequences of *Mycobacterium tuberculosis* gyrA and katG genes. Tuberculosis 92, 333–344.

Kumar, A., Tassopoulos, A.M., Li, Q., Yu, F.S.X., 2007. *Staphylococcus aureus* protein A induced inflammatory response in human corneal epithelial cells. Biochem. Biophys. Res. Commun. 354, 955–961.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., Hiramatsu, K., 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. Lancet 357, 1225–1240.

Merino, N., Toledo-Arana, A., Vergara-Irigaray, M., Valle, J., Solano, C., Calvo, E., Lopez, J.A., Foster, T.J., Penadés, J.R., Lasa, I., 2009. Protein A-mediated multicellular behavior in *Staphylococcus aureus*. J. Bacteriol. 191, 832–843.

Nakamura, Y., Gojobori, T., Ikemura, T., 2000. Codon use tabulated from the international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 28, 292.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

O'Seaghdha, M., van Schooten, C.J., Kerrigan, S.W., Emsley, J., Silverman, G.J., Cox, D., Lenting, P.J., Foster, T.J., 2006. *Staphylococcus aureus* protein A binding to von Willebrand factor A1 domain is mediated by conserved IgG binding regions. FEBS J. 273, 4831–4841.

Patou, J., Gevaert, P., van Zele, T., Holtappels, G., van Cauwenberge, P., Bachert, C., 2008. *Staphylococcus aureus* enterotoxin B, protein A, and lipoteichoic acid stimulations in nasal polyps. J. Allergy Clin. Immunol. 121, 110–115.

Shopsin, B., Gomez, M., Montgomery, S.O., Smith, D.H., Waddington, M., Dodge, D.E., Bost, D.A., Riehman, M., Naidich, S., Kreiswirth, B.N., 1999. Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J. Clin. Microbiol. 37, 3556–3563.

Stutz, K., Stephan, R., Tasara, T., 2011. SpA, ClfA, and FnbA genetic variations lead to staphaurex test-negative phenotypes in bovine mastitis *Staphylococcus aureus* isolates. J. Clin. Microbiol. 49, 638–646.

Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA 85, 2653–2657.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum

likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731−2739.

Widaa, A., Claro, T., Foster, T.J., O'Brien, F.J., Kerrigan, S.W., 2012. *Staphylococcus aureus* protein A plays a critical role in mediating bone destruction and bone loss in osteomyelitis. PLoS ONE 7, e40586.

Yung, S.C., Parenti, D., Murphy, P.M., 2011. Host chemokines bind to *Staphylococcus aureus* and stimulate protein A release. J. Biol. Chem. 286, 5069−5077.