BMC
Genomics

PROCEEDINGS   **Open Access**

An approach to identifying drug resistance associated mutations in bacterial strains

Michal Wozniak[1,2*], Jerzy Tiuryn[1], Limsoon Wong[2]

Abstract

Background: Drug resistance in bacterial pathogens is an increasing problem, which stimulates research. However, our understanding of drug resistance mechanisms remains incomplete. Fortunately, the fast-growing number of fully sequenced bacterial strains now enables us to develop new methods to identify mutations associated with drug resistance.

Results: We present a new comparative approach to identify genes and mutations that are likely to be associated with drug resistance mechanisms. In order to test the approach, we collected genotype and phenotype data of 100 fully sequenced strains of *S. aureus* and 10 commonly used drugs. Then, applying the method, we re-discovered the most common genetic determinants of drug resistance and identified some novel putative associations.

Conclusions: Firstly, the collected data may help other researchers to develop and verify similar techniques. Secondly, the proposed method is successful in identifying drug resistance determinants. Thirdly, the *in-silico* identified genetic mutations, which are putatively involved in drug resistance mechanisms, may increase our understanding of the drug resistance mechanisms.

Introduction

The problem of bacterial drug resistance did not exist in 1930s, when antibiotics were introduced to treat bacterial infections. Since then, due to various factors–such as irresponsible dosage of antibiotics, naturally occurring mutations, transmission of drug-resistant strains, etc.– drug resistance has become a serious health problem. This has drawn the attention of WHO (World Health Organization), ECDC (European Centre for Disease Prevention and Control) and CDC (Centers for Disease Control and Prevention), which monitor and report the spreading of drug-resistant pathogens in the world. As a consequence, for example, WHO launched in 2006 a new global program "Stop TB Strategy" to fight the spreading of *M. tuberculosis* (MTB). A recent WHO report on

MTB estimates that the bacteria was responsible for around 1.7 million deaths world-wide in 2009 [1]. According to the report, 3.3% of new MTB cases in 2009 were multi-drug resistant (MDR). Moreover, 58 countries reported cases of extensively-drug-resistant (XDR) isolates of the bacteria. Very recently, ECDC reported that as high as 58% of all *Staphylococcus aureus* isolates tested in Malta was methicillin resistant (MRSA) [2].

The emergence of drug resistance is appalling, because it is often not economically justifiable for pharmaceutical companies to develop new drugs against it [3]. One promising approach to address the problem is to use old drugs that were designed for treating other diseases and are also effective against pathogens [4]. An effort in this direction was recently undertaken in a research study on *M. tuberculosis* [5]. The authors used three-dimensional docking to identify *in-silico* some putative drug-target interactions. For example, they predicted Comtan, a drug

* Correspondence: m.wozniak@mimuw.edu.pl
[1]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland
Full list of author information is available at the end of the article

BioMed Central

used in treating Parkinson's disease, as potentially effective against *M. tuberculosis* infections.

The need of more efficient strategies to develop new drugs stimulates research to better understand drug resistance mechanisms. Several drug resistance mechanisms have been discovered so far. They can be categorized as: (i) drug target modification; (ii) drug molecule modification by specialized enzymes; (iii) reduced accumulation of the drug inside a bacteria cell by decreased cell wall permeability or by pumping out the drug; and (iv) alternative metabolic pathways [6]. Moreover, there are known genes and mutations responsible for most of the drug resistance mechanisms. While genomics can be used on samples before and after drug resistance emerges to identify the likely associated mutations, most of the known mutations and genes associated with drug resistance were discovered by analyzing *a priori* candidates such as drug target genes or genes located on plasmids. Some information on drug target genes is available in the drugbank.ca database [7]; and some lists of genes known to be responsible for drug resistance (specific to bacterial species and drugs) are available in the ARDB (Antibiotic Resistance Genes Database) database [8].

Despite the above mentioned achievements, our understanding of drug resistance mechanisms is still incomplete. For example, there are reports of *S. aureus* isolates with atypical drug resistance profiles [9-11], which have not been explained yet. We hypothesize that these atypical drug resistance profiles might be due to genomic mutations in genes which are not *a priori* suspected of being involved in drug resistance mechanisms.

In this work, we use whole-genome sequences to identify and associate genetic mutations with drug resistance phenotype for bacterial strains (within *S. aureus*). Thus, conceptually our approach is similar to Genome-Wide Association Study (GWAS) approaches, which have been successfully applied to identify SNPs associated with human diseases [12,13]. We hypothesize that similar approaches, when applied to bacteria, should bring interesting results. However, it may not make sense to directly transfer this methodology to bacteria, because, for example, horizontal gene transfer (HGT) plays an important role in the evolution of bacteria. Besides mutations which would explain the reported atypical drug resistance profiles, we expect, by applying our approach, to identify also mutations that can be interpreted as compensatory mutations. These compensatory mutations are not directly involved in drug resistance, but they are important to neutralizing the deleterious effect (caused by mutations directly responsible for the resistance mechanisms) on bacterial fitness [14-16].

There are published studies, based on comparative analysis of whole-genome sequences, associating genetic mutations with drug resistance [17-21]. However, the methodologies used in these studies are simple and were applied to a relatively small number of strains. In our opinion, this is caused by two main problems: first, the number of fully sequenced bacterial strains within the same species have not been sufficiently large until recently; second, phenotype data with respect to drug susceptibility tests are spread throughout the literature and are not easy to collect.

In this work, we collected genotype data for 100 fully sequenced *S. aureus* strains and addressed the second problem by a careful search of the literature for results of drug susceptibility tests of the strains considered. We also developed and tested a new approach to associate mutations and genes with drug resistance.

Materials and methods

Below we present details of our methodology including the problem setting, collection of data and subsequent steps of the identification of drug resistance associated genetic features. These subsequent steps comprise:

- unification of protein-coding gene annotations of bacterial strains and determination of gene families;
- computing multiple alignments for the gene families and reconstructing the consensus phylogenetic tree;
- identification of genetic features, possibly associated with drug resistance, such as point mutations and gene gain/losses, based on the multiple alignments and the determined gene families; and
- association of genetic features with drug resistance phenotypes.

Problem setting

We consider a set $\mathcal{S}$ of bacterial strains and their response to the application of a given drug. The response, which we called *drug resistance profile*, is represented by a vector $v : \mathcal{S} \rightarrow \{'S', 'R', '?'\}$, where by 'S' and 'R' we denote respectively drug-susceptible and drug-resistant strain phenotypes, by '?' we indicate that the phenotype is unknown. Additionally, we denote by $\mathcal{S}_v^S$ and $\mathcal{S}_v^R$ the sets of drug-susceptible and drug-resistant strains for a given drug resistance profile, respectively.

We also assume to have a given set of genetic mutations among the considered bacterial strains. Analogous to drug resistance profiles, we represent mutations as vectors $m : \mathcal{S} \rightarrow \sum \cup \{'?'\}$, where Σ denotes an alphabet of possible states, such as amino acids in the strain sequence corresponding to a given position in the multiple alignment. By '?' in the mutation profile we denote strains which are not present in the aligned gene family corresponding to the considered point mutation.

Then, the problem is mainly to identify a subset of genetic mutations associated with a given drug resistance profile and, secondarily, to use the identified mutations to predict unknown places in the given drug resistance profile (marked by '?').

Genotype data

We collected genotype data (genome sequences and annotations) for the following 100 fully sequenced strains of S. aureus from the GenBank [22] and PATRIC databases [23]. Additionally, genotype data for strain EMRSA-15 were downloaded from the Wellcome Trust Sanger Institute website. At the time of writing, 31 out of the 100 S. aureus strains had "completed" sequencing status. For the remaining strains whose genomes are still being assembled, contig sequences (covering around 90% of the genomes) and annotations are provided.

We unify the original annotations employing our previously published method, called CAMBer [24]. Briefly, CAMBer iteratively extends the protein-coding annotations by homology transfer, until the transitive closure of a given homology relation is computed. This homology relation defines the consolidation graph. In this graph, there is an edge between a pair of genes if there was an accepteble BLAST hit between them.

Then, we determine gene families as connected components in the consolidation graph. However, we additionally extend the consolidation graph by edges coming from BLAST amino-acid queries. More formally, we add an edge between a pair of genes to the consolidation graph if the percent of identity (calculated as the number of identities over the length of the longer gene) of the BLAST hit between them exceeds a threshold $P(L)$ given by the HSSP curve formula [25]:

$$P(L) = \begin{cases} 100 & L \leq 11 \\ c + 480.L^{-0.32.(1+e^{-L/1000})} & 11 < L \leq 450 \\ c + 19.5 & L > 450 \end{cases} \quad (1)$$

Here, c is set to 40.5 and L is the number of aligned amino acid residues.

Each connected component in the consolidation graph corresponds to a gene family [24]. We compute multiple alignments using MUSCLE [26] for all these gene families. Then, we consider two kinds of genetic variations:

- gene gain/loss,
- amino acid point mutations.

Intuitively, we represent the considered genetic variations as 0 - 1 vectors, indexed by strains, where 0 denotes the reference state and 1 denotes some change. We call vectors of these genetic variations as gain/loss profiles and point mutation profiles.

Gene gain/loss profiles are transformed from gene families which do not span the set $\mathcal{S}$ of all considered strains. For each such gene family, we transform it into a vector representation $g : \mathcal{S} \to \{'G','L'\}$ as follows: for a given strain i, we define $g(i)$ = 'G' if the gene family contains at least one gene in that gene family for strain i; otherwise we set $g(i)$ = 'L'.

Similarly, point mutation profiles are transformed from columns in multiple alignments computed for gene families with elements present in at least $|\mathcal{S}| - 1$ strains. We take into account only columns which contain at least two different characters (ignoring '?'). For each such column (in the multiple alignment), we transform it into a vector representation $m : \mathcal{S} \to \sum_{AA} \cup \{'-','?'\}$ as follows: for a given strain i, we set $m(i)$ = 'x' if the character 'x' is present (one of 20 amino-acids or '-') in the row corresponding to strain i; and set $m(i)$ = '?' if strain i is not present in the aligned gene family.

Phylogenetic tree of the strains

We compute the phylogenetic tree of the input strains using a consensus method with majority rule implemented in the PHYLIP package [27]. We apply the consensus method to trees constructed for all gene families with exactly one element in each strain. The trees are constructed using the maximum likelihood approach implemented in the PHYLIP package [27].

Phenotype data (drug susceptibility)

Drug susceptibility data were collected from the following sources: (i) publications issued together with the fully sequenced genomes: USA300_TCH1516 and USA300_TCH959 [28], MRSA252 and MSSA476 [29], 04-02981 [30], T0131 [31], ST398 [32], COL [33], JKD6008 [34], 16 K [35], TW20 [36], Newman [37], RN4220 [38], O46 and O11 [39], RF122 [40], Mu3 [41], MRSA252 [29], CF-Marseille [42], N315 and Mu50 [43], MW2 [44], MSHR1132 [45], ECT-R_2 [46], JH1 and JH9 [17], ED98 [47], JKD6159 [48], LGA251 [49], ED133 [50]; (ii) NARSA project http://www.narsa.net; (iii) email exchange with the authors of publications related to strains ST398 and TW20; and (iv) other publications found by searching of related literature [21,23,33,37,51-98]. The complete collected phenotype data are available in the supplementary table (additional file 1).

We represent the collected information as a set of drug resistance profiles, defined for each drug separately.

Essential mutations

For a given drug resistance vector v we introduce a function r_v which describes the reference state of a given point mutation or gene gain/loss profile p. We

define it as the most often-occurring state in drug-susceptible strains (ignoring '?'), i.e.:

$$r_v(p) = \arg\max_{x \in \Sigma} \sum_{i \in \mathcal{S}_v^{\mathcal{S}}} [p(i) = x] \tag{2}$$

Here, and in all the following equations, square brackets are used for Iverson's notation.

In the current implementation, in the case when there is a multiple number of states present in the same maximal number of strains, the function r_v returns the first state in the lexicographical order. Note that this is just a technical assumption, since such mutations will not be considered as associated with drug resistance.

We say that a point mutation m is *present* in a strain i if $m(i) \notin \{r_v(m),'?'\}$; otherwise we say that the point mutation m is *absent* in strain i.

Then, we distinguish two categories of gene gain/loss and point mutation profiles depending on how they correspond to a given drug resistance profile. We categorize a given mutation profile m as:

- *essential mutation*, when m is absent in all drug-susceptible strains,
- *conflict mutation*, when m is present in at least one drug-susceptible strain.

Further, we distinguish *neutral mutations* as a subclass of *essential mutations*, these are *essential mutations* that are not present in any of drug-resistant strains.

Analogously, we transfer the above introduced concepts to gene/loss profiles, defining *essential*, *neutral* and *conflict* gain/loss profiles.

Support
We aim to identify genetic variations which are likely to be associated with drug resistance. Intuitively, such mutations or gained genes should often be present in drug-resistant strains and rarely in drug-susceptible strains. To reflect this intuition we assign a score, which we call a *support*, to all point mutation and gene gain/loss profiles. For a given point mutation or gene gain/loss profile p and drug resistance profile v, the *support* (s_v) is defined as the number of drug-resistant strains with the mutation present (or gene gained) minus the number of drug-susceptible strains with the mutation present (or gene gained):

$$s_v(p) = \sum_{i \in \mathcal{S}_v^R} [p(i) \neq r_v(p)] - \alpha_v \sum_{i \in \mathcal{S}_v^{\mathcal{S}}} [p(i) \neq r_v(p)] \tag{3}$$

Here, α_v is a weight which we use to punish mutations for their presence in drug-susceptible strains. It is defined as the proportion of the number of drug-resistant to the number of drug-susceptible strains, so that occurrences of a mutation are given equal emphasis in

drug-resistant and drug-susceptible strains. More formally:

$$\alpha_v = \frac{|\mathcal{S}_v^R|}{|\mathcal{S}_v^{\mathcal{S}}|} \tag{4}$$

Weighted support
Although the *support* is a simple and intuitive score, it does not incorporate any phylogenetic information. For example, let us assume there are two point mutations with the same *support* 3, where the first mutation covers only drug-resistant strains within one subtree of the phylogenetic tree, whereas the second mutation covers the same number of strains but spread throughout the whole tree. The first mutation is likely to be associated with the phylogeny, driven by some environmental changes. This suggests that the second mutation should have a greater score as it has to be acquired a few times independently during the evolution process.

We propose *weighted support* as a score to account for the above situation. For a given phylogenetic tree T and gene gain/loss or point mutation profile p, *weighted support* (ws_v) is defined as follows:

$$ws_v^T(p) = \sum_{i \in S} w_i^T [p(i) \neq r_v(p)] \tag{5}$$

where w_i^T are weights assigned to each cell in a given drug resistance profile.

In all our experiments we assign weights in the following way: all drug-susceptible strains are assigned weight $-\alpha_v$ (defined as above); each drug-resistant strain i is assigned a weight $\frac{1}{n}$, where n is the number of drug-resistant strains in the subtree (containing strain i) determined by its highest parental node, such that the subtree does not contain any drug-susceptible strain in its leaves. All strains without drug resistance information are assigned weights 0.

Note that the *support* score can also be expressed as *weighted support*, where w_i are assigned as $-\alpha_v$, 1, 0 for drug-susceptible, drug-resistant and strains without drug resistance information, respectively.

Figure 1 illustrates the concept of *support* and *weight-support*.

In order to make the support scores more comparable between drugs, we introduce normalized versions of the scores, *normalized support* and *normalized weighted support* which denote the respective support value divided the maximal possible *support* or *weighted support*, respectively.

Odds ratio
For a given drug resistance profile v and mutation p, we calculate *odds ratio* using the formula:

Figure 1 Support and weighted support. A schematic example of classification of genetic variation profiles and computation of their supports. Point mutations 1 and 4 are *essential*, mutation 2 is *conflict* and mutation 3 is *neutral*. Light blue circles mark nodes which appear in the definition of *weighted support*. These are nodes the highest parental nodes (for the leaf nodes corresponding to drug-resistant strains), that their subtrees do not contain any drug-susceptible strains in leaves. The scores (a) *support* and (b) *weighted support* are assigned to these mutations. For this drug-resistance profile, the ratio α_v equals $\frac{5}{3}$.

$$odds\_ratio_v\left(p\right) = \frac{n_{R1} \cdot n_{S0}}{max\left(1, n_{R0}\right) \cdot max\left(1, n_{S1}\right)} \quad (6)$$

Here, n_{R1}, n_{S0}, n_{R0} and n_{S1} denote the number of drug-resistant strains with mutation p, drug-susceptible strains without mutation p, drug-resistant strains without mutation p and drug-susceptible strains with mutation p, respectively.

The same formula is used to calculate odds ratio for gene gain/loss profiles.

Statistical significance

In order to assess statistical significance of the associations we calculate their *p-value*.

More precisely, for a given drug resistance profile *v*, let X be the random variable giving *support* of a random mutation. Then, for a given observed mutation with *support = c*, its *p-value* is defined by the following formula:

$$\mathbb{P}\left(X \geq c\right) = \sum_{n=1}^{|\mathcal{S}|} \mathbb{P}\left(X \geq c | N = n\right) \cdot \mathbb{P}\left(N = n\right) \quad (7)$$

Here, N is a random variable which denotes the number of mutated strains in a random mutation. For each n the probability $\mathbb{P}(N = n)$ of observing a mutation present in n strains is estimated (as the number of mutations present in n strains to the total number of considered mutations) from the data for point mutation and gene gain/loss profiles separately. The details follow. Assume that weights, for a given drug resistance profile v, take k different values: $l_1, l_2, ..., l_k$. For $1 \leq j \leq k$, let m_j be the number of strains which take value l_j. Clearly we have $m_1 + m_2 + ... + m_k = |S|$. Then, the probability $\mathbb{P}(X \geq c | N = n)$ (from the equation 7) is given by the formula:

$$\sum_{\substack{0 \leq n_1 \leq m_1 \\ 0 \leq n_2 \leq m_2 \\ \cdots \\ 0 \leq n_k \leq m_k \\ n_1 + n_2 + \cdots + n_k = n}} \frac{\prod_{j=1}^{k} \binom{m_j}{n_j}}{\binom{|\mathcal{S}|}{n}} \left[\sum_{j=1}^{k} n_j \cdot l_j \geq c \right] \quad (8)$$

Here we describe our algorithm for calculating the p-value. It should be clear that the problem reduces to computing $\mathbb{P}\,(X \geq c | N = n) = \dfrac{t_c\,(n)}{\binom{|\mathcal{S}|}{n}}$ for each $0 \leq n \leq |S|$, where $t_c(n)$ denotes the number of ways for distributing n ones over $|S|$ strains, such that the corresponding sum of weights is greater or equal than c. The term $\binom{|\mathcal{S}|}{n}$ is the total number of possible ways for distributing n ones over $|S|$ strains. Thus, the problem reduces to calculating $t_c(n)$ for each $0 \leq n \leq |S|$. Additionally, without any loss of generality, we may assume that the weight levels are strictly decreasing: $l_1 > l_2 > ... > l_k$, where $l_k < 0$ and $l_{k-1} \geq 0$.

The algorithm iteratively generates partial combinations (without n_k) starting from the partial combination $(n_1 = m_1, ..., n_{k-1} = m_{k-1})$ in the following manner: if j is the highest index of the non-zero n_i in the current partial combination, the next partial combination will be $(n_1, ..., n_j - 1, n_{j+1} = m_{j+1}, ..., n_{k-1} = m_{k-1})$. The algorithms terminates generating partial combinations when two following partial combinations have their corresponding sum of weights below the level of c. At each step of the algorithm, all possible full combinations $(n_1, ... n_{k-1}, n_k)$ are generated from the current partial combination $(n_1, ... n_{k-1})$. If for the full combination its corresponding sum of weights is greater or equal $c\left(\sum_{i=1}^{k} n_i \cdot l_i \geq c\right)$, then we increment the value $t_c(n)$ by $\prod_{i=1}^{k} \binom{m_i}{n_i}$, where $n = n_1 + ... + n_k$. As the outcome, we obtain $t_c(n)$ and, thus, also $\mathbb{P}(X \geq c | N = n)$ for each n.

The last step is to calculate formula 7 using these calculated probabilities.

Note that, since *support* is a special case of *weighted support*, the same formula and algorithm can be used to compute its corresponding p-values.

Results and discussion

We verify the usability of our approach by trying to re-identify the known drug resistance determinants. In this experiment, we compare our proposed scoring methods –*support* and *weighted support* –to *odds ratio*, which is a popular measure used in genome-wide association studies. Table 1 shows rankings of the gene gain/loss profiles corresponding to genes which are known drug resistance determinants. The experiment suggests that *weighted support* identifies putative associations better than *support* and *odds ratio*, both of which do not incorporate additional information about phylogeny.

This experiment also reveals that the amount of the collected drug resistance information is not sufficient to correctly identify drug resistance associated genes. However, the high consistency of drug resistance profiles corresponding to the collected information and the presence of drug resistance determinants (summing over drugs, there are 117 drug resistant strains, where only 4 of them do not have any known drug resistance determinants; and there are 112 drug-susceptible strains, where only 8 of them have at least one drug resistance determinant) suggests that we can use the determinants to predict drug resistance in the strains without drug resistance information available. It is perhaps questionable to predict drug resistance in those strains for which the whole-genome sequence is not determined yet. So we do prediction only for those strains with completed sequencing or at least information on their plasmids

Table 1 Rankings of known drug resistance determining genes

gene id.	drug name	Rankings before prediction			Rankings after prediction		
		S	WS	OR	S	WS	OR
tet	Tetracycline	54.5	2.5	43.7	1.5	1.5	1.5
tetM	Tetracycline	14.5	11.5	7.5	4	4	4
mecA	Methicillin	1	1	1	1	1	1
mecA	Oxacillin	3	4	2	1	2	1
ermA1	Clindamycin	5.5	5.5	5.5	1	1	1
ermC	Clindamycin	907	471	907	414.5	11	191.5
ermA1	Erythromycin	3	3	4	1	1	1
ermC	Erythromycin	1527	3994.5	1006.5	413.5	28	214.5
aacA-aphD	Gentamicin	72	34	34	1	1	1
blaZ	Penicillin	163	66	223	1.5	1	2.5
mecA	Penicillin	163	8	223	11	5	52
Average ranking (excluding ermC):		53.27	**15.05**	60.411	2.55	**1.94**	7.22

Rankings of the known drug resistance determinants obtained by employing three different methods to score gene gain/loss profiles: *support* (S), *weighted support* (WS) and *odds ratio* (OR). Since some of the gene gain/loss profiles are assigned with the same score, we calculate their rankings as the arithmetic mean of positions of the profiles with the same score on the list sorted according to the scores; thus some of the rankings are not round numbers. The rankings were computed before and after prediction of drug resistance, which is based on the presence of the drug resistance determinants. We excluded the gene ermC from the calculations of the rankings since none of the methods were able to pull it out into the top 100 before prediction.

(which often carry the drug resistance determinants). Nevertheless, we predict drug resistance also for those strains that are not yet fully sequenced, provided the presence of drug resistance determining genes has been confirmed for them. Moreover, we predict drug resistance to rifampicin and ciprofloxacin for all 100 strains, as the drug resistance for rifampicin and ciprofloxacin is determined by point mutations in genes *rpoB, gyrA* and *grlA* (synonymous name to *parC*), which are sequenced in all strains. More precisely, we predicted as rifampicin-resistant all strains with any mutation present in the rifampicin resistance determining region (RRDR). We defined the RRDR as the amino-acid range from 463 to 530 in the *rpoB* gene sequence (according to [94]). Analogously, we predicted as ciprofloxacin-resistant all strains with any point mutation in the quinolone resistance determining region (QRDR). We defined QRDR as the amino-acid ranges from position 68 to 107 and from position 64 to 103 in the *grlA* and *parC* gene sequences, respectively (according to [65]). Figure 2 shows the complete information about drug susceptibility after prediction.

Then, we applied our approach to the dataset supplemented by the predicted information about drug susceptibility for the following drugs: tetracycline, β-lactames (penicillin, oxacillin, methicillin), erythromycin, gentamicin, vancomycin, ciprofloxacin and rifampicin.

We discuss in the subsections below the results of our approach applied separately to the following drugs: tetracycline, β-lactames (penicillin, methicillin), erythromycin, gentamicin, vancomycin, ciprofloxacin. We do not discuss here results for oxacillin and clindamycin, since they have very similar drug resistance profiles to methicillin and erythromycin, respectively. All other drugs were excluded from the analysis due to the low number of strains with available drug resistance information on these drugs.

Tables 2 and 3 present the top-scored gene gain/loss, and point mutation profiles for the dis-cussed drugs, respectively. The genes presented in the tables were selected according to the following procedure: for each drug we construct a function, which gives for each gene (listed in descending order with respect to *normalized weighted support*) the minus logarithm of p-value (-*log*(p-value)) of this score. Then, we report genes which correspond to the portion of the graph of this function before it gets flattened. Complete results for all the drugs are provided in supplementary Excel tables (additional files 3 and 4).

Tetracycline

Tetracycline acts by binding to the 30S ribosomal subunit (rpsS, 16S rRNA are its direct targets), preventing binding of tRNA to the mRNA-ribosome complex, and thus inhibiting protein synthesis [7].

The most common drug resistance mechanism to tetracycline in *S. aureus* is mediated by ribosome protection proteins (RPPs) such as *tet* and *tetM*, which bind to the ribosome complex, thus preventing the binding of tetracycline [99,100].

Proteins *tet* and *tetM* mediating the mechanism cover all drug-resistant strains except MW2. This may be caused by errors in the drug susceptibility tests, errors in sequencing, or by some other not yet known drug resistance mechanism. The inconsistent information about strain MW2's tetracyline susceptibility (see supporting Table 1) and the lack of identified drug resistance determinants suggest that the strain is possibly drug susceptible. In our experiment we initially assumed that the tetracycline resistance information is not available for strain MW2.

Our method shows that, besides *tet* and *tetM*, there are a few more genes that have highly scored gene gain/loss profiles. Especially interesting are the following genes which are not gained by any of the drug susceptible strains: *repC, pre, thiI, int, clfB* (see Table 2). There are studies reporting the significance of these *clfB* and *repC* genes in drug resistance [101,102]. Interestingly, the gene *repC* seems to co-evolve with *tet* (correlated gene gain/loss profiles).

Applying our method to point mutations we have identified two highly scored (and essential) point mutations in ribosomal complex proteins: $K_{101}R$ in *rpsL* and $K_{57}M$ in *rpsJ*. According to our knowledge, this is the first report on the significance of the point mutations for drug resistance in *S. aureus*. However, we found a study associating mutations in *rpsJ* with tetracycline resistance in another bacteria *Neisseria gonorrhoeae* [103].

Beta-lactams

Beta-lactams are a broad class of antibiotics, which possess (by definition) the β-lactam ring in their structure. The ring is capable of binding transpeptidase proteins (also known as Penicillin Binding Proteins – PBPs) [7], which are important to synthesis of the peptidoglycan layer of bacterial cell wall. PBPs with attached drug molecules are no longer able to synthesize peptidoglycan, leading to bacterial death [104]. In our case study we consider three β-lactam antibiotics: penicillin, oxacillin and methicillin. However, since the drug resistance profile and drug resistance mechanisms for oxacillin and methicillin are very similar we discuss results only for methicillin.

There are two common β-lactames resistance mechanisms in *S. aureus* [104,105]. The first one is mediated by β-lactamase enzymes, which bind drug molecules and break the β-lactam ring, thus deactivating the drug molecules. This mechanism is effective against

	Vancomycin	Penicillin	Methicillin	Oxacillin	Tetracycline	Erythromycin	Clindamycin	Gentamicin	Ciprofloxacin	Rifampicin
H19	?	r	?	?	?	?	?	?	s	s
D139	?	?	?	?	?	?	?	?	s	s
M013	?	r	r	r	s	s	s	s	s	s
RF122	?	s	s	s	s	s	s	s	s	s
JKD6159	S	r	R	r	S	S	S	s	S	S
21235	?	?	?	?	?	?	?	?	s	s
ED133	?	s	s	s	s	s	s	s	s	s
LGA251	?	R	R	R	S	S	S	S	S	S
21269	?	r	?	?	?	?	?	?	s	s
O11	?	?	?	?	?	?	?	?	s	s
O46	?	?	?	?	?	?	?	?	s	s
ST398	S	R	R	R	R	S	S	s	R	S
TCH60	?	r	R	r	s	s	s	s	s	s
WBG10049	?	r	R	r	?	?	?	?	s	s
C427	?	r	?	?	?	?	?	?	s	s
A01793497	?	r	R	r	?	?	?	?	s	s
WW270397	?	r	R	r	?	?	?	?	s	s
CGS00	?	r	?	?	r	r	?	?	s	s
Btn1260	?	r	S	?	?	?	?	?	s	s
MRSA252	S	R	R	R	S	R	R	S	R	S
EMRSA16	?	r	R	r	?	r	r	?	r	s
MN8	S	R	S	S	S	S	S	S	S	s
C160	?	r	?	?	?	?	?	?	s	s
21195	?	?	?	?	?	?	?	?	s	s
552053	?	r	S	?	?	?	?	?	s	s
M1015	?	r	S	?	?	?	?	?	s	s
M809	?	r	S	?	?	?	?	?	s	s
C101	?	r	S	?	r	?	?	?	s	s
E1410	?	R	S	?	S	?	?	?	s	s
M876	?	r	S	?	?	?	?	?	s	s
M899	?	r	S	?	r	?	?	?	s	s
58-424	?	R	?	?	R	R	r	?	s	s
65-1322	?	R	?	?	R	R	r	?	s	s
68-397	?	R	?	?	R	R	r	?	s	s
MSHR1132	?	r	r	R	s	s	s	s	s	s
21200	?	r	?	?	?	?	?	?	s	s
A9635	S	r	?	?	?	?	?	?	s	s
EMRSA15	?	r	r	r	s	r	r	s	r	s
21310	?	r	?	?	?	?	?	r	s	s
MR1	?	r	R	r	r	?	?	?	s	s
ED98	?	s	s	s	r	s	s	s	r	s
A9299	S	?	?	?	?	?	?	?	s	s
21201	?	r	?	?	?	?	?	?	s	s
A8117	S	?	?	?	r	?	?	?	s	s
A8115	S	?	?	?	r	?	?	?	s	s
CF-Marseille	?	r	R	r	R	R	r	S	r	s
Mu50-omega	S	r	R	R	R	R	R	R	R	r
Mu50	R	R	R	R	R	R	R	R	R	R
Mu3	S	R	R	R	R	R	R	R	R	S
N315	S	R	R	r	S	R	R	S	S	S

	Vancomycin	Penicillin	Methicillin	Oxacillin	Tetracycline	Erythromycin	Clindamycin	Gentamicin	Ciprofloxacin	Rifampicin
21318	?	r	?	?	?	?	?	?	s	s
ECT-R_2	?	r	S	s	s	R	R	s	r	s
A9763	S	r	r	r	?	r	r	?	r	s
A5937	?	r	r	r	?	r	r	?	r	s
A6224	S	r	r	r	?	r	r	?	r	r
04-02981	S	R	R	R	S	R	R	S	R	s
A6300	?	r	r	r	?	r	r	?	r	r
A10102	S	r	r	r	?	r	r	?	r	s
A8796	S	r	r	r	?	r	r	?	r	s
21172	?	r	r	r	?	r	r	?	r	s
A8819	S	r	r	r	?	r	r	?	r	s
CGS03	?	r	r	r	?	r	r	?	r	s
JH9	Y	r	R	S	S	R	R	r	r	R
JH1	S	r	R	S	S	R	R	r	r	S
A9781	S	r	r	r	?	r	r	?	r	s
A9719	S	r	r	r	?	r	r	?	r	s
21305	?	r	?	?	r	?	?	?	s	s
21193	?	?	?	?	?	?	?	?	s	s
USA300_TCH959	S	R	S	S	S	S	S	S	S	S
11819-97	?	r	R	r	s	s	s	s	s	s
NCTC_8325	S	S	S	S	S	S	S	S	S	S
21189	?	?	?	?	?	?	?	?	s	s
VC40	R	s	s	s	s	r	r	s	s	s
RN4220	S	S	S	S	S	S	S	S	S	S
A5948	S	r	r	r	?	r	r	?	s	s
132	?	?	?	?	?	?	?	?	r	s
D30	?	?	?	?	?	?	?	?	s	s
MRSA131	S	r	R	r	?	?	?	S	S	S
MRSA177	S	r	R	r	r	?	?	S	S	S
CGS01	?	r	r	r	?	r	r	?	s	s
USA300_TCH1516	S	R	R	R	S	R	S	S	S	S
A9754	S	r	r	r	r	?	?	?	r	r
USA300_FPR3757	S	R	R	R	R	R	R	s	R	s
930918-3	?	r	?	?	?	r	r	?	s	s
A9765	S	r	r	r	r	r	r	r	r	r
COL	S	R	R	R	R	S	S	S	S	s
Newman	S	S	S	S	S	S	s	s	S	s
TW20	S	R	R	r	R	R	r	R	R	S
T0131	?	r	R	r	r	r	r	s	r	r
16K	?	r	r	r	r	s	s	r	r	R
JKD6009	S	r	R	r	R	r	r	R	R	S
JKD6008	R	r	R	R	R	r	r	R	R	S
ATCC_BAA-39	?	r	r	r	r	r	r	r	r	s
ATCC_51811	?	r	?	?	?	?	?	?	s	s
TCH70	?	r	r	r	?	r	r	?	s	s
MW2	S	R	R	R	S	S	S	S	S	s
MSSA476	S	R	S	S	S	S	S	S	S	S
21259	?	?	?	?	r	?	?	?	s	s
TCH130	?	r	?	?	r	r	r	?	s	s
21266	?	r	?	?	?	?	?	?	s	s

Figure 2 The collected dataset of phenotypes with predictions. The collected dataset of phenotypes put together with results of our drug resistance predictions based on the presence of known drug resistance determinants. Due to the high number of strains the table is split into two panels. Columns represent drugs, rows represent *S. aureus* strains included in the study in the order corresponding to the reconstructed phylogenetic tree of strains. Green, yellow and red cell colors represent susceptible, intermediate resistant and resistant phenotypes, respectively. Analogously, light green and light red cell colors represent predicted susceptible and resistant phenotypes, respectively. White cell color represents unknown (not determined by experiments or prediction) drug resistance phenotypes.

penicillin (which is β-lactamase sensitive) and not effective against methicillin and oxacillin (which are β-lactamase resistant) [106]. The second β-lactam resistance mechanism is mediated by proteins which are capable of functionally substituting for PBPs, but have much smaller affinity to β-lactam molecules. This mechanism is effective against penicillin, methicillin and oxacillin.

Penicillin

In our dataset all strains resistant to penicillin possess proteins responsible for one of the two mechanisms. More precisely, there are 69 drug-resistant strains (with available drug resistance information), which possess *BlaZ* – the standard β-lactamase protein (note that its regulators *BlaR1* and *blaI* do not always co-occur).

Table 2 The top scored gene gain/loss profiles

Gene identifier	NS	NWS	OR	p-value	Gene functional annotation
Penicillin (NWS-threshold: 0.58)					
SAR1831(blaZ)	0.84	0.81	37.15	1.15e-06	beta-lactamase
SAR1829(blaI)	0.84	0.74	37.15	5.24e-06	transcriptional repressor
SAR1830(blaR1)	0.82	0.73	31.27	7.09e-06	beta-lactamase regulatory protein blar1
SAR0056	0.63	0.71	12.13	1.03e-05	conserved hypothetical protein
SAR0039(mecA)	0.61	0.70	10.94	1.28e-05	penicillin-binding protein pbp2a, methicillin resistance determinant mecA, transpeptidase
SAR0060(ccrA)	0.61	0.63	10.94	4.40e-05	resolvase, n-terminal domain protein
SAR0061(yycG)	0.61	0.63	10.94	4.40e-05	putative membrane protein
NWMN 0025	0.57	0.63	9.40	4.41e-05	conserved domain protein
SAR0037(ugpQ)	0.60	0.63	10.39	5.08e-05	glycerophosphoryldiester phosphodiesterase
SAR0038(maoC)	0.60	0.63	10.39	5.08e-05	dehydratase
SAR0057	0.57	0.59	9.40	9.78e-05	conserved hypothetical protein
Methicillin (NWS-threshold: 0.68)					
SAR0039(mecA)	1.00	1.00	950.00	4.48e-20	penicillin-binding protein pbp2a, methicillin resistance determinant mecA, transpeptidase
SAR0037(ugpQ)	0.98	0.94	931.00	6.77e-15	glycerophosphoryldiester phosphodiesterase
SAR0038(maoC)	0.98	0.94	931.00	6.77e-15	dehydratase
SAR0056	0.95	0.85	900.00	7.55e-12	conserved hypothetical protein
SAR0036	0.64	0.80	33.78	5.77e-11	putative membrane protein
SAR0057	0.85	0.75	162.00	6.47e-10	conserved hypothetical protein
SAR0060(ccrA)	0.91	0.73	432.00	1.40e-09	resolvase, n-terminal domain protein
SAR0061(yycG)	0.91	0.73	432.00	1.40e-09	putative membrane protein
MW0028(ebpS)	0.54	0.71	22.30	2.76e-09	hmg-coa synthase
Tetracycline (NWS-threshold: 0.32)					
SAAV_b3(repC)	0.54	0.64	27.69	5.70e-08	plasmid replication protein
SATW20_00660(tet)	0.54	0.64	27.69	5.70e-08	tetracycline resistance protein
SATW20_00670(pre)	0.50	0.50	24.00	3.51e-06	plasmid recombination enzyme type 3
SATW20_04620(tetM)	0.46	0.37	20.80	7.54e-05	tetracycline resistance protein tetM
SATW20_08990(virE)	0.42	0.37	19.93	7.67e-05	pathogenicity island protein
SATW20_09000	0.42	0.37	19.93	7.67e-05	pathogenicity island protein
SATW20_09010(lipA)	0.42	0.37	19.93	7.67e-05	putative protein in superantigen-encoding pathogenicity islands
SATW20_04610(thil)	0.43	0.35	18.00	1.32e-04	putative transcriptional regulator
MW0745(int)	0.25	0.32	8.00	2.28e-04	site-specific recombinase, phage integrase family
MW0747	0.25	0.32	8.00	2.28e-04	DNA-binding helix-turn-helix protein
Erythromycin (NWS-threshold: 0.27)					
SAR0050(ermA1)	0.80	0.58	76.00	1.36e-06	rRNA adenine n-6-methyltransferase
CGSSa03 12660	0.47	0.44	17.19	2.98e-05	conserved hypothetical protein
SAR0054(tnpA1)	0.75	0.39	72.00	8.12e-05	transposase for transposon
SAR1734	0.75	0.39	72.00	8.12e-05	methylase
SAR1736(spc2)	0.75	0.39	72.00	8.12e-05	spectinomycin 9-o-adenylyltransferase
SaurJH9_1711(radC)	0.72	0.38	62.00	8.83e-05	predicted protein
SAUSA300_pUSA030006	0.20	0.35	4.75	1.65e-04	replication and maintenance protein
SAR1737(tnpC2)	0.72	0.34	62.00	1.89e-04	Unknown
SAR1529	0.33	0.33	9.15	2.43e-04	conserved hypothetical protein
SATW20_04860(recF_1)	0.23	0.30	5.52	3.67e-04	recombinational DNA repair ATPase
SAR1738(tnpB2)	0.70	0.29	54.00	4.39e-04	transposase B from transposon Tn554
SauraJ_010100009720	0.23	0.27	5.52	6.60e-04	conserved domain protein
Gentamicin (NWS-threshold: 0.83)					
SaurJH1_2806(aacA-aphD)	0.83	0.90	150.00	9.38e-11	bifunctional acetyltransferase/phosphotransferase GNAT family acetyltransferase
SaurJH1_2805	0.75	0.83	90.00	2.95e-09	

Table 2 The top scored gene gain/loss profiles (Continued)

			Ciprooxacin (NWS-threshold: 0.4)			
SATW20_04610(thiI)	0.35	0.45	36.00	1.33e-07	putative transcriptional regulator
SATW20_04650(cap8J)	0.32	0.40	31.57	8.25e-07	lipoprotein
SATW20_04670(capL)	0.32	0.40	31.57	8.25e-07	putative ATP/GTP-binding protein
SATW20_04780	0.32	0.40	31.57	8.25e-07	conjugation related protein
SATW20_04800	0.32	0.40	31.57	8.25e-07	replication initiation factor
SATW20_04810	0.32	0.40	31.57	8.25e-07	DNA segregation ATPase FtsK/SpoIIIE
SATW20_04830	0.32	0.40	31.57	8.25e-07	conjugative transposon protein

Summarizing information for the top scored gene gain/loss profiles. The consequent columns refer to: gene identifier of the corresponding gene family; *normalized support* (NS); *normalized weighted support* (NWS); *odds ratio* (OR); p-value and the gene functional annotation. Thresholds for *weighted support* are provided in brackets for each drug. Colored gene gain/loss profiles are provided in the supplementary the (additional file 3). Complete results for all the drugs are provided in the supplementary Excel table (additional files 2).

Table 3 The top scored point mutation profiles, only for essential mutations

Gene identifier	desc.	NS	NWS	OR	p-value	Gene functional annotation
			Penicillin (NWS-threshold: 0.4)			
SAR0023(sasH)	$G_{723}D$	0.55	0.63	8.51	1.87e-05	virulence-associated cell-wall-anchored protein; 5'-nucleotidase
SAR0023(sasH)	$T_{725}A$	0.54	0.62	8.11	2.23e-05	virulence-associated cell-wall-anchored protein; 5'-nucleotidase
SAR0304	$V_{295}I$	0.39	0.49	4.48	3.25e-04	acid phosphatase
SAR2791	$V_{182}M$	0.46	0.46	6.05	5.41e-04	transcriptional regulator, Xre family
SAR2700	$N_{493}KD$	0.52	0.45	7.72	6.16e-04	ABC transporter permease protein
SAR0233(hmp)	$Q_{333}K$	0.44	0.44	5.48	7.21e-04	avohemoprotein (nitric oxide dioxygenase)
SAR0318(sbnA)	$N_{25}HK$	0.44	0.43	5.48	8.36e-04	alpha/beta family hydrolase
SAR2664	$V_{282}AT$	0.44	0.43	5.48	8.36e-04	probable monooxygenase
SAR2779	$S_{48}G$	0.44	0.43	5.48	8.36e-04	n-hydroxyarylamine o-acetyltransferase
SAR0318(sbnA)	$T_{138}IM$	0.43	0.43	5.21	8.36e-04	alpha/beta family hydrolase
SAR0318(sbnA)	$T_{139}AQ$	0.43	0.43	5.21	8.36e-04	alpha/beta family hydrolase
SAR0023(sasH)	$A_{749}TG$	0.41	0.43	4.96	8.44e-04	virulence-associated cell-wall-anchored protein; 5'-nucleotidase
SAR0318(sbnA)	$R_{130}CG$	0.41	0.43	4.96	8.72e-04	alpha/beta family hydrolase
SAR0322(folC)	$H_{201}YQE$	0.41	0.43	4.96	8.72e-04	macro domain, possibly adp-ribose binding module
SAR0233(hmp)	$K_{323}ET$	0.40	0.42	4.71	9.08e-04	avohemoprotein (hemoglobin-like protein)
SAR2750(icaC)	$I_{21}V$	0.40	0.42	4.71	9.46e-04	polysaccharide intercellular adhesin (PIA) biosynthesis protein
SAR0233(hmp)	$S_{309}RN$	0.39	0.42	4.48	9.46e-04	avohemoprotein (hemoglobin-like protein)
			Methicillin (NWS-threshold: 0.25)			
SAR0198(oppF)	$T_{287}IK$	0.10	0.29	2.11	1.41e-04	putative glutathione transporter, ATP-binding component
SAR0420	$I_{72}F$	0.10	0.29	2.11	1.41e-04	membrane protein
SAR2508(sbi)	$S_{219}AT$	0.10	0.29	2.11	1.41e-04	IgG-binding protein Sbi
SAR2508(sbi)	$N_{222}QK$	0.10	0.29	2.11	1.41e-04	IgG-binding protein Sbi
SAR2508(sbi)	$K_{224}SDN$	0.10	0.29	2.11	1.41e-04	IgG-binding protein Sbi
			Tetracycline (NWS-threshold: 0.2)			
SAR1840	$D_{291}YS$	0.18	0.23	5.22	7.09e-04	NAD(FAD)-utilizing dehydrogenases
SAR2336(rpsJ)	$K_{57}M$	0.29	0.23	9.60	7.32e-04	SSU ribosomal protein S10P (S20E)
SAR0550(rpsL)	$K_{113}R$	0.36	0.20	13.33	1.14e-03	SSU ribosomal protein S12P (S23E)
			Erythromycin (NWS-threshold: 0.2)			
SAR0576	$A_{68}EV$	0.07	0.21	1.54	8.89e-04	phosphoglycolate phosphatase
			Gentamicin (NWS-threshold: 0.21)			
SAR1840	$L_{289}IW$	0.33	0.29	15.00	1.43e-03	NAD(FAD)-utilizing dehydrogenases
SAR1840	$D_{291}YS$	0.33	0.29	15.00	1.43e-03	NAD(FAD)-utilizing dehydrogenases
SAR1840	$H_{327}RF$	0.33	0.29	15.00	1.43e-03	NAD(FAD)-utilizing dehydrogenases
SAR1167(ylmH)	$K_{215}N$	0.25	0.29	10.00	1.43e-03	RNA-binding S4 domain-containing protein
SAR1167(ylmH)	$R_{216}V$	0.25	0.29	10.00	1.43e-03	RNA-binding S4 domain-containing protein

Table 3 The top scored point mutation profiles, only for essential mutations (Continued)

SAR1167(ylmH)	$V_{217}L$	0.25	0.29	10.00	1.43e-03	RNA-binding S4 domain-containing protein
SAR0547(rpoB)	$D_{471}YG$	0.17	0.21	6.00	4.61e-03	DNA-directed RNA polymerase beta subunit
SAR1833(trmB)	$T_{54}IK$	0.17	0.21	6.00	4.61e-03	tRNA (guanine46-n7-)-methyltransferase
Cirprooxacin (NWS-threshold: 0.12)						
SAR1367(grlA)	$S_{80}YF$	1.00	1.00	2244.00	6.03e-30	topoisomerase IV subunit a
SAR0006(gyrA)	$S_{90}AL$	0.94	0.88	1056.00	1.92e-18	DNA gyrase subunit a
SAR2449(lytT)	$V_{45}I$	0.21	0.20	17.11	2.06e-04	transcriptional regulator
SAR1840	$L_{289}IW$	0.12	0.20	8.80	4.56e-04	NAD(FAD)-utilizing dehydrogenases
SAR1793(thiI)	$A_{92}ET$	0.09	0.20	6.39	2.06e-04	thiamine biosynthesis protein thiI
SAR2212(murA2)	$A_{102}T$	0.06	0.20	4.12	2.06e-04	UDP-n-acetylglucosamine 1-carboxyvinyltransferase
SAR1367(grlA)	$E_{84}KG$	0.26	0.15	23.76	9.40e-04	topoisomerase IV subunit a
SAR0235(pstG 1)	$F_{401}LV$	0.09	0.13	6.39	2.21e-03	PTS system, maltose and glucose-specific IIC component
SAR0400(nfrA)	$R_{194}H$	0.09	0.13	6.39	2.21e-03	nitroreductase family protein

Summarizing information for the top scored point mutation profiles, only for essential mutations. The *conflict mutatations* were removed from the table for: tetracycline, erythromycin and gentamicin (for the rest of drugs there were no *conflict mutations* above the set thresholds). The consequent columns refer to: gene identifier of the corresponding gene family; corresponding position in the multiple alignment and changed amino acids; *normalized support* (NS); *normalized weighted support* (NWS); *odds ratio* (OR); p-value and the gene functional annotation. Thresholds for *weighted support* are provided in brackets for each drug. Colored gene gain/loss profiles are provided in the supplementary the (additional file 4). Complete results for all the drugs are provided in the supplementary Excel table (additional files 5).

All the remaining penicillin-resistant strains have *mecA*, which is an altered PBP. Table 2 provides information about the top-scored gene gain/loss profiles.

Applying our method we have also identified the uncategorized putative protein, *SAR0056*, as putatively associated with penicillin resistance (see Table 2). We suggest to examine further the role of that gene in β-lactams resistance.

Methicillin

Applying our approach to gene gain/loss profiles we identified (beside *mecA*) genes *ugpQ* and *maoC*. The correlation of gene profiles to the profile of *mecA* and their close proximity on the genomes suggests that these genes co-evolve (see Figure 3 for more details). This co-evolution may reflect some important role played by these genes in methicillin resistance. This calls for further study of the role of these two genes in methicillin resistance.

We have also identified a few point mutations that are putatively associated with methicillin resistance. Interestingly, two of the mutations in the top 10 essential mutations according to *weighted support* ($I_{72}F$ in *SAR0420* and $E_{208}QKD$ in *SAR0436*) are present in cell membrane proteins. This suggests some compensatory mechanism to the presence of *mecA*.

Ciprofloxacin

Ciprofloxacin belongs to a broad class of antibiotics, called fluoroquinolones, which are functional against bacteria by binding DNA gyrase subunit A (encoded by *gyrA*) and DNA topoisomerase 4 subunit A (encoded by *parC*), which are enzymes necessary to separate bacterial DNA, thereby inhibiting cell division [7]. The most common ciprofloxacin resistance mechanism is mediated by point mutations in the drug targets, *parC* and *gyrA*.

Applying our approach we identified (by highest support) two point mutations in ciprofloxacin target genes – S_{80} *FY* in *parC* and S_{90} *AL* in *gyrA* – which are located in QRDR and known to be responsible for the first mechanism of ciprofloxacin resistance [65]. The presence of these mutations is correlated with the ciprofloxacin resistance profile for strains with available drug resistance information. However, they differ for two strains ED98 and 16 K (only the mutation in *parC* is present). This may suggest intermediate drug resistance level for these strains. Unfortunately ciprofloxacin resistance information is not available for these strains.

Erythromycin

Erythromycin acts by binding the 23S rRNA molecule (in the 50S subunit) of the bacterial ribosome complex, leading to inhibition of protein synthesis [7].

There are three known erythromycin resistance mechanisms [107]. First – the most common mechanism – is by methylation (addition of two residues to the domain V of 23S rRNA) of the 23S rRNA molecule, which prevents the ribosome from binding with erythromycin. This methylation is mediated by enzymes from the *erm* gene family, the most common are *ermA* and *ermC*. The second mechanism is mediated by the presence of macrolide efflux pumps (encoded by *msrA* and *msrB*). The third mechanism is the inactivation of drug molecules by specialized enzymes such as *EreA* or *EreB* [107].

We found that none of the strains in our case study possess genes *EreA* or *EreB*. Genes encoding efflux

Figure 3 Genes related to methicillin resistance. Presence and relative genome coordinates of genes related to methicillin resistance (*mecA, mecR1, mecI, ccrA, ccrB, ccrC*), put together with the identified genes: *ugpQ* and *maoC*. The gene presence profiles are clustered with respect to the genes order. In this figure we include only these methicillin-resistant strains for which all the genes where located on the main genome and within the same sequence conting (in order to determine the relative positions).

pumps (*msrA* and *msrB*) are present also in drug-susceptible strains (for example, NCTC 8325 and Newman), which may suggest that the mechanism is inactive for the considered strains of *S. aureus* or the enzyme production rates are too small, which we are not able to account by our method. Using our approach we identified (by the highest support) the gene *ermA* responsible for the most common drug resistance mechanism.

Here, there is one erythromycin-susceptible strain, USA300_TCH959, which harbours the *ermA* gene. This may suggest disruption of the drug resistance mechanism in that strain, errors in drug susceptibility testing or errors in sequencing.

Interestingly, we identified gene *SAR1736(spc2)* (which is a known spectinomycin resistance determinant) as potentially associated with erythromycin resistance. This suggests that drug resistance to spectinomycin and erythromycin co-evolved, despite these two drugs belonging to different classes according to the ATC drug classification system [106].

Gentamicin

Gentamicin works by inhibition of protein synthesis by binding the 30S subunit of the ribosome complex [108].

Interestingly, strain USA300-FPR3757 exhibits intermediate drug resistance, which is correlated with the absence of *aacA-aphD* gene in its genome sequence. Since our method requires binary information on drug susceptibility, we marked this strain as drug susceptible for experiments.

The most common resistance mechanism responsible for high levels of Gentamicin resistance is mediated by the drug-modifying enzyme *SaurJH1_2806(aacA-aphD)*. Applying our methodology we identified the gene encoding it as likely to be associated with drug resistance (maximal support). Moreover, we identified also the gene *SaurJH1_2805* as putatively associated with gentamicin resistance. The close proximity of these two genes in the genomes and their highly correlated gene gain/loss profiles suggest co-evolution. We hypothesize that the gene *SaurJH1_2805* plays some role in drug resistance for gentamicin.

Conclusion

In this work we present a novel approach to associate genes and mutations with drug resistance phenotypes by comparative analysis of fully sequenced bacterial strains (within the same species).

In order to apply our approach we collected genotype and phenotype data. Genome sequences and annotations were downloaded for 100 fully sequenced *S. aureus* strains. A challenge was to collect drug resistance information, which is spread throughout the literature. We retrieved the data from 71 publications. The collected dataset is available in the supplementary material (additional file 1).

In our method we consider two types of genetic differences as potentially associated with drug resistance: nonsynonymous point mutations and gene acquisition, represented by their mutation and gene gain/loss profiles, respectively. Then, the approach is based on the newly introduced concept of *support*, which is a score

assigned to all mutation and gene gain/loss profiles. Intuitively, the higher support of a mutation profile, the better chance of the mutation to be associated with the drug-resistant phenotype. We also generalize the concept of support into *weighted support*, which incorporates phylogenetic information.

Applying our approach, we were able to successfully re-identify most of the known drug resistance determinants. Here, on average, *weighted support* outperforms *support* and *odds ratio*.

Moreover, applying our methodology, we identified some putative novel resistance-associated genes and mutations. We expect that these associations and drug resistance predictions will attract the experimental research community to verify their role in drug resistance mechanisms.

Finally, although the presented approach shows promise, it has some obvious limitations. Firstly, it is not clear what threshold for the *weighted support* is appropriate. Secondly, in this approach we only consider genome variations, whereas some drug resistance mechanisms may be related to changes in protein production rates (such as efflux pumps). Thirdly, the current approach ignores the role of non-coding RNA in drug resistance mechanisms. That, would not be detected by our approach. We plan to address these mentioned problems in some future work.

Additional material

Additional file 1: Collected phenotype and genotype data. Collected dataset of phenotypes (results of drug susceptibility tests). Columns represent drugs, rows represent *S. aureus* strains included in the study, put in the order corresponding to the reconstructed phylogenetical tree of strains. Green, yellow and red color cells represent collected information: susceptibility, intermediate resistance and resistance of isolates, respectively.

Additional file 4: Excel table with detailed results for point mutation profiles. Excel table providing results of our approach applied to point mutation profiles for ten drugs: penicillin, methicillin, oxacillin, tetracycline, clindamycin, erythromycin, gentamicin, ciprofloxacin, rifampicin and vancomycin

Additional file 2: Summary table for the top scored gene gain/loss profiles (same thresholds as for Table 2 are applied). The columns refer to: gene identifier of the corresponding gene family; normalized *weighted support* (NWS); p-value and the drug resistance profiles put together with gene gain/loss profiles. Each cell in the gene gain/loss profiles corresponds to one strain, ordered according to the order in Figure 2. Cells corresponding to drug-resistant and drug-susceptible strains are colored red and green, respectively. Strains without drug resistance information are left white. For each gene gain/loss profile p and its corresponding row, if a cell in this row corresponds to strain *i*, such that *rv(p) = p(i)*, then it is colored blue, otherwise it is colored pink.

Additional file 3: Excel table with detailed results for gene gain/loss profiles. Excel table providing results of our approach applied to gene gain/loss profiles for ten drugs: penicillin, methicillin, oxacillin, tetracycline, clindamycin, erythromycin, gentamicin, ciprofloxacin, rifampicin and vancomycin.

Additional file 5: Table with point mutation profiles for top scored mutations profiles. Summary table for the top scored gene point

mutation profiles (same thresholds as for Table 3 are applied). The columns refer to: gene identifier of the corresponding gene family; *normalized weighted support* (NWS); p-value and the drug resistance profiles put together with point mutation profiles. Each cell in the point profiles corresponds to one strain, ordered according to the order in Figure 2. Cells corresponding to drug-resistant and drug-susceptible strains are colored red and green, respectively. Strains without drug resistance information are left white. For each point mutation profile p and its corresponding row, if a cell in this row corresponds to strain i (assuming the corresponding gene is present in the strain sequence), such that $r_v(p) = p(i)$, then it is colored blue, otherwise it is colored pink. Cells corresponding to strains without the corresponding gene are left white.

Author details

[1]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland. [2]School of Computing, National University of Singapore, Singapore.

Authors' contributions

All authors contributed to the design of the method, analysis of results and writing of the manuscript. MW wrote software and performed experiments. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Published: 13 December 2012

References

1. Multidrug and extensively drug-resistant Tuberculosis: 2010 global report on surveillance and response. *WHO* 2010.
2. Antimicrobial resistance surveillance in Europe. *ECDC* 2009.
3. Shlaes DM, Projan SJ, Edwards JE: Antibiotic discovery: state of the state. *ASM News* 2004, **70(6)**:275-281.
4. Chong CR, Sullivan DJ: New uses for old drugs. *Nature* 2007, **448(7154)**:645-646.
5. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE: Drug Discovery Using Chemical Systems Biology: Repositioning the Safe Medicine Comtan to Treat Multi-Drug and Extensively Drug Resistant Tuberculosis. *PLoS Comput Biol* 2009, **5(7)**:e1000423.
6. Wright GD: Molecular mechanisms of antibiotic resistance. *Chem Commun* 2011, **47**:4055-4061.
7. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS: DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Research* 2011, **39(Suppl 1)**:D1035-D1041.
8. Liu B, Pop M: ARDB-Antibiotic Resistance Genes Database. *Nucleic Acids Research* 2009, **37(Suppl 1)**:D443-D447.
9. Giannouli S, Labrou M, Kyritsis A, Ikonomidis A, Pournaras S, Stathopoulos C, Tsakris A: Detection of mutations in the FemXAB protein family in oxacillin-susceptible mecA-positive Staphylococcus aureus clinical isolates. *Journal of Antimicrobial Chemotherapy* 2010, **65(4)**:626-633.
10. Ikonomidis A, Michail G, Vasdeki A, Labrou M, Karavasilis V, Stathopoulos C, Maniatis AN, Pournaras S: In vitro and in vivo evaluations of oxacillin efficiency against mecA-positive oxacillin-susceptible Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy* 2008, **52(11)**:3905-3908.
11. Welch WD, Southern PM: Unusual susceptibility of methicillin-resistant Staphylococcus aureus to erythromycin, clindamycin, gentamicin, and tetracycline at 30 degrees C but not at 35 degrees C. *Journal of Clinical Microbiology* 1984, **19(6)**:831-833.
12. Johnson A, O'Donnell C: An open access database of genome-wide association results. *BMC Medical Genetics* 2009, **10**:6.
13. Manolio TA: Genomewide association studies and assessment of the risk of disease. *New England Journal of Medicine* 2010, **363(2)**:166-176.
14. Levin BR, Perrot V, Walker N: Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* 2000, **154(3)**:985-997.
15. Handel A, Regoes RR, Antia R: The role of compensatory mutations in the emergence of drug resistance. *PLoS Computational Biology* 2007, **3(2)**:e46.
16. Comas I, Borrell S, Roetzer A, Rose G, Malla B, KatoMaeda M, Galagan J, Niemann S, Gagneux S: Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics* 2012, **44**:106-110 [http://www.nature.com/ng/journal/v44/n1/full/ng.1038.html].
17. Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, Bruce D, Rubin E, Myers E, Siggia ED, Tomasz A: Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. *Proceedings of the National Academy of Sciences* 2007, **104(22)**:9451-9456.
18. Ioerger TR, Koo S, No E, Chen X, Larsen MH, Jacobs WR Jr, Pillay M, Sturm AW, Sacchettini JC: Genome analysis of multi- and extensively-drug-resistant Tuberculosis from KwaZulu-Natal, South Africa. *PLoS ONE* 2009, **4(11)**:e7778.
19. Fournier P, Vallenet D, Barbe V, Audic S, Ogata H, Poirel L, Richet H, Robert C, Mangenot S, Abergel C, Nordmann P, Weissenbach J, Raoult D, Claverie J: Comparative genomics of multidrug resistance in Acinetobacter baumannii. *PLoS Genetics* 2006, **2**:e7.
20. Habib F, Johnson AD, Bundschuh R, Janies D: Large scale genotype-phenotype correlation analysis based on phylogenetic trees. *Bioinformatics* 2007, **23(7)**:785-788.
21. Peleg AY, Miyakis S, Ward DV, Earl AM, Rubio A, Cameron DR, Pillai S, Moellering RC, Eliopoulos GM: Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of Staphylococcus aureus. *PLoS ONE* 2012, **7**:e28316.
22. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: GenBank. *Nucleic acids research* 2008, **36(Database issue)** [http://nar.oxfordjournals.org/content/36/suppl_1/D25.full].
23. Snyder EE, Kampanya N, Lu J, Nordberg EK, Karur HR, Shukla M, Soneja J, Tian Y, Xue T, Yoo H, Zhang F, Dharmanolla C, Dongre NV, Gillespie JJ, Hamelius J, Hance M, Huntington KI, Jukneliene D, Koziski J, Mackasmiel L, Mane SP, Nguyen V, Purkayastha A, Shallom J, Yu G, Guo Y, Gabbard J, Hix D, Azad AF, Baker SC, Boyle SM, Khudyakov Y, Meng XJ, Rupprecht C, Vinje J, Crasta OR, Czar MJ, Dickerman A, Eckart JD, Kenyon R, Will R, Setubal JC, Sobral BWS: PATRIC: The VBI PathoSystems Resource Integration Center. *Nucleic Acids Research* 2007, **35(Database issue)**:D401-D406[http://nar.oxfordjournals.org/content/35/suppl_1/D401.long].
24. Wozniak M, Wong L, Tiuryn J: CAMBer: an approach to support comparative analysis of multiple bacterial strains. *BMC Genomics* 2011, **12(Suppl 2)**:S6.
25. Rost B: Twilight zone of protein sequence alignments. *Protein Engineering* 1999, **12(2)**:85-94.
26. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004, **32(5)**:1792-1797.
27. Felsenstein J: PHYLIP (Phylogeny Inference Package) version 3.6 Seattle: Department of Genome Sciences, University of Washington; 2005.
28. Highlander SK, Hulten KG, Qin X, Jiang H, Yerrapragada S, Mason EO, Shang Y, Williams TM, Fortunov RM, Liu Y, Igboeli O, Petrosino J, Tirumalai M, Uzman A, Fox GE, Cardenas AM, Muzny DM, Hemphill L, Ding Y, Dugan S, Blyth PR, Buhay CJ, Dinh HH, Hawes AC, Holder M, Kovar CL, Lee SL, Liu W, Nazareth LV, Wang Q, Zhou J, Kaplan SL, Weinstock GM: Subtle genetic changes enhance virulence of methicillin resistant and sensitive Staphylococcus aureus. *BMC Microbiology* 2007, **7**:99.
29. Holden MTG, Feil EJ, Lindsay JA, Peacock SJ, Day NPJ, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R, Barron A, Bason N, Bentley SD, Chillingworth C, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James KD, Lennard N, Line A, Mayes R, Moule S, Mungall K, Ormond D, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S, Barrell BG, Spratt BG, Parkhill J: Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101(26)**:9786-9791.

30. Nubel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK, Zemlickova H, Leblois R, Wirth T, Jombart T, Balloux F, Witte W: **A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant Staphylococcus aureus.** *PLoS Pathog* 2010, **6**(4):e1000855.

31. Li Y, Cao B, Zhang Y, Zhou J, Yang B, Wang L: **Complete genome sequence of Staphylococcus aureus T0131, an ST239-MRSA-SCCmec type III clone isolated in China.** *Journal of Bacteriology* 2011, **193**(13):3411-3412.

32. Smith TC, Male MJ, Harper AL, Kroeger JS, Tinkler GP, Moritz ED, Capuano AW, Herwaldt LA, Diekema DJ: **Methicillin-resistant Staphylococcus aureus (MRSA) strain ST398 is present in midwestern U. S. swine and swine workers.** *PLoS ONE* 2009, **4**:e4258.

33. Gill SR, Fouts DE, Archer GL, Mongodin EF, DeBoy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM: **Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain.** *Journal of Bacteriology* 2005, **187**(7):2426-2438.

34. Howden BP, Seemann T, Harrison PF, McEvoy CR, Stanton JL, Rand CJ, Mason CW, Jensen SO, Firth N, Davies JK, Johnson PDR, Stinear TP: **Complete cenome sequence of Staphylococcus aureus strain JKD6008, an ST239 clone of methicillin-resistant Staphylococcus aureus with intermediate-level vancomycin resistance.** *Journal of Bacteriology* 2010, **192**(21):5848-5849.

35. Yamamoto T, Takano T, Higuchi W, Iwao Y, Singur O, Reva I, Otsuka Y, Nakayashiki T, Mori H, Reva G, Kuznetsov V, Potapov V: **Comparative genomics and drug resistance of a geographic variant of ST239 methicillinresistant Staphylococcus aureus emerged in Russia.** *PLoS ONE* 2012, **7**:e29187.

36. Edgeworth JD, Yadegarfar G, Pathak S, Batra R, Cock-field JD, Wyncoll D, Beale R, Lindsay JA: **An outbreak in an intensive care unit of a strain of methicillinresistant Staphylococcus aureus sequence type 239 associated with an increased rate of vascular access device-related bacteremia.** *Clinical Infectious Diseases* 2007, **44**(4):493-501.

37. Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K: **Genome sequence of Staphylococcus aureus strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands.** *Journal of Bacteriology* 2008, **190**:300-310.

38. Nair D, Memmi G, Hernandez D, Bard J, Beaume M, Gill S, Francois P, Cheung AL: **Whole-genome sequencing of Staphylococcus aureus strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain.** *Journal of Bacteriology* 2011, **193**(9):2332-2335.

39. Le Marechal C, Hernandez D, Schrenzel J, Even S, Berkova N, Thiery R, Vautor E, Fitzgerald JR, Fran?s P, Le Loir Y: **Genome sequences of two Staphylococcus aureus ovine strains that induce severe (strain O11) and mild (strain O46) mastitis.** *Journal of Bacteriology* 2011, **193**(9):2353-2354.

40. Herron-Olson L, Fitzgerald JR, Musser JM, Kapur V: **Molecular correlates of host specialization in Staphylococcus aureus.** *PLoS ONE* 2007, **2**(10):e1120.

41. Neoh Hm, Cui L, Yuzawa H, Takeuchi F, Matsuo M, Hiramatsu K: **Mutated response regulator graR is responsible for phenotypic conversion of Staphylococcus aureus from heterogeneous vancomycin-intermediate resistance to vancomycin-intermediate resistance.** *Antimicrobial Agents and Chemotherapy* 2008, **52**:45-53.

42. Rolain J, Francois P, Hernandez D, Bittar F, Richet H, Fournous G, Mattenberger Y, Bosdure E, Stremler N, Dubus J, Sarles J, Reynaud-Gaubert M, Boniface S, Schrenzel J, Raoult D: **Genomic analysis of an emerging multiresistant Staphylococcus aureus strain rapidly spreading in cystic fibrosis patients revealed the presence of an antibiotic inducible bacteriophage.** *Biology Direct* 2009, **4**:1.

43. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki Ki, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue Ri, 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: **Whole genome sequencing of meticillin-resistant Staphylococcus aureus.** *The Lancet* 2001, **357**(9264):1225-1240.

44. Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki Ki, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, Kuroda H, Cui L, Yamamoto K, Hiramatsu K: **Genome and virulence determinants of high virulence community-acquired MRSA.** *Lancet* 2002, **359**(9320):1819-1827.

45. Holt DC, Holden MTG, Tong SYC, Castillo-Ramirez S, Clarke L, Quail MA, Currie BJ, Parkhill J, Bentley SD, Feil EJ, Giffard PM: **A very early-branching Staphylococcus aureus lineage lacking the carotenoid pigment staphyloxanthin.** *Genome Biology and Evolution* 2011, **3**:881-895.

46. Lindqvist M, Isaksson B, Grub C, Jonassen T, Hällgren A: **Detection and characterisation of SCCmec remnants in multiresistant methicillin-susceptible Staphylococcus aureus causing a clonal outbreak in a Swedish county.** *European Journal of Clinical Microbiology & Infectious Diseases* 2012, **31**(2):141-147.

47. Lowder BV, Guinane CM, Ben Zakour NL, Weinert LA, Conway-Morris A, Cartwright RA, Simpson AJ, Rambaut A, Nubel U, Fitzgerald JR: **Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus.** *Proceedings of the National Academy of Sciences* 2009 [http://www.pnas.org/content/106/46/19545.long].

48. Chua K, Seemann T, Harrison PF, Davies JK, Coutts SJ, Chen H, Haring V, Moore R, Howden BP, Stinear TP: **Complete genome sequence of Staphylococcus aureus strain JKD6159, a unique Australian clone of ST93-IV community methicillin-resistant Staphylococcus aureus.** *Journal of Bacteriology* 2010, **192**(20):5556-5557.

49. Garcia-Alvarez L, Holden MT, Lindsay H, Webb CR, Brown DF, Curran MD, Walpole E, Brooks K, Pickard DJ, Teale C, Parkhill J, Bentley SD, Edwards GF, Girvan EK, Kearns AM, Pichon B, Hill RL, Larsen AR, Skov RL, Peacock SJ, Maskell DJ, Holmes MA: **Meticillin-resistant Staphylococcus aureus with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study.** *The Lancet Infectious Diseases* 2011, **11**(8):595-603.

50. Guinane CM, Ben Zakour NL, Tormo-Mas MA, Weinert LA, Lowder BV, Cartwright RA, Smyth DS, Smyth CJ, Lindsay JA, Gould KA, Witney A, Hinds J, Bollback JP, Rambaut A, Penades JR, Fitzgerald JR: **Evolutionary genomics of Staphylococcus aureus reveals insights into the origin and molecular basis of ruminant host adaptation.** *Genome Biology and Evolution* 2010, **2**:454-466.

51. Tenover FC, Lancaster MV, Hill BC, Steward CD, Stocker SA, Hancock GA, O'Hara CM, Clark NC, Hiramatsu K: **Characterization of Staphylococci with reduced susceptibilities to vancomycin and other glycopeptides.** *Journal of Clinical Microbiology* 1998, **36**(4):1020-1027.

52. Bikels-Goshen T, Landau E, Saguy S, Shapira R: **Staphylococcal strains adapted to epigallocathechin gallate (EGCG) show reduced susceptibility to vancomycin, oxacillin and ampicillin, increased heat tolerance, and altered cell morphology.** *International Journal of Food Microbiology* 2010, **138**(1-2):26-31.

53. Dickgiesser N, Wallach U: **Toxic shock syndrome toxin1 (TSST-1): influence of its production by subinhibitory antibiotic concentrations.** *Infection* 1987, **15**(5):351-353.

54. Pucci MJ, Dougherty TJ: **Direct quantitation of the numbers of individual penicillin-binding proteins per cell in Staphylococcus aureus.** *Journal of Bacteriology* 2002, **184**(2):588-591.

55. Kuroda M, Kuwahara-Arai K, Hiramatsu K: **Identification of the up- and down-regulated genes in vancomycinresistant Staphylococcus aureus strains Mu3 and Mu50 by cDNA differential hybridization method.** *Biochemical and Biophysical Research Communications* 2000, **269**(2):485-490.

56. Eyyunni L, Pettyjohn R, Sahm DF: *Network on antimicrobial resistance in Staphylococcus aureus* 2011 [http://www.narsa.net].

57. Friedman L, Alder JD, Silverman JA: **Genetic changes that correlate with reduced susceptibility to daptomycin in Staphylococcus aureus.** *Antimicrobial Agents and Chemotherapy* 2006, **50**(6):2137-2145.

58. WYKE AW, WARD JB, HAYES MV: **Synthesis of peptidoglycan in vivo in methicillin-resistant Staphylococcus aureus.** *European Journal of Biochemistry* 1982, **127**(3):553-558.

59. Holden MTG, Lindsay JA, Corton C, Quail MA, Cock-field JD, Pathak S, Batra R, Parkhill J, Bentley SD, Edgeworth JD: **Genome gequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant Staphylococcus aureus, sequence type 239 (TW).** *Journal of Bacteriology* 2010, **192**(3):888-892.

60. Desbois AP, Coote PJ: **Wax moth larva (Galleria mellonella): an in vivo model for assessing the efficacy of antistaphylococcal agents.** *Journal of Antimicrobial Chemotherapy* 2011.

61. Hershberger E, Aeschlimann JR, Moldovan T, Rybak MJ: **Evaluation of bactericidal activities of LY333328, vancomycin, teicoplanin, ampicillin-sulbactam, trovaoxacin, and RP59500 alone or in combination with rifampin or gentamicin against different strains of vancomycin-intermediate Staphylococcus aureus by time-kill curve methods.** *Antimicrobial Agents and Chemotherapy* 1999, **43**(3):717-721.

62. Burnside K, Lembo A, de los Reyes M, Iliuk A, BinhTran N, Connelly JE, Lin W, Schmidt BZ, Richardson AR, Fang FC, Tao WA, Rajagopal L: **Regulation of hemolysin expression and virulence of Staphylococcus aureus by a serine/threonine kinase and phosphatase.** *PLoS ONE* 2010, **5**(6):e11071.

63. Suzuki M, Yamada K, Nagao M, Aoki E, Matsumoto M, Hirayama T, Yamamoto H, Hiramatsu R, Ichiyama S, Iinuma Y: **Antimicrobial ointments and methicillin-resistant Staphylococcus aureus USA300.** *Emerging Infectious Diseases* 2011, **17**(10):1917-1920.

64. Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, Davidson MG, Lin F, Lin J, Carleton HA, Mongodin EF, Sensabaugh GF, Perdreau-Remington F: **Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus.** *The Lancet* 2006, **367**(9512):731-739.

65. Ferrero L, Cameron B, Crouzet J: **Analysis of gyrA and grlA mutations in stepwise-selected ciprofloxacinresistant mutants of Staphylococcus aureus.** *Antimicrobial Agents and Chemotherapy* 1995, **39**(7):1554-1558.

66. Pietiainen M, Francois P, Hyyrylainen H, Tangomo M, Sass V, Sahl H, Schrenzel J, Kontinen VP: **Transcriptome analysis of the responses of Staphylococcus aureus to antimicrobial peptides and characterization of the roles of vraDE and vraSR in antimicrobial resistance.** *BMC Genomics* 2009, **10**:429.

67. Crossley KB, Jefferson KK, Archer GL, Fowler VG: *Staphylococci in human disease* John Wiley & Sons; 2009.

68. Brackman G, Cos P, Maes L, Nelis HJ, Coenye T: **Quorum Sensing Inhibitors Increase the susceptibility of bacterial biofilms to antibiotics in vitro and in vivo.** *Antimicrobial Agents and Chemotherapy* 2011, **55**(6):2655-2661.

69. Sá-Leao R, Sanches IS, Couto I, Alves CR, de Lencastre H: **Low prevalence of methicillin-resistant strains among Staphylococcus aureus colonizing young and healthy members of the community in Portugal.** *Microbial Drug Resistance (Larchmont, N.Y.)* 2001, **7**(3):237-245.

70. Molestina RE: **BEI Resources: a biological resource center for parasitologists.** *Trends in parasitology* 2010, **26**(12):559-560.

71. Howden BP, Stinear TP, Allen DL, Johnson PDR, Ward PB, Davies JK: **Genomic analysis reveals a point mutation in the two-component sensor gene graS that leads to intermediate vancomycin resistance in clinical Staphylococcus aureus.** *Antimicrobial Agents and Chemotherapy* 2008, **52**(10):3755-3762.

72. O'Neill AJ, Chopra I: **Insertional inactivation of mutS in Staphylococcus aureus reveals potential for elevated mutation frequencies, although the prevalence of mutators in clinical isolates is low.** *Journal of Antimicrobial Chemotherapy* 2002, **50**(2):161-169.

73. Kuroda M, Kuroda H, Oshima T, Takeuchi F, Mori H, Hiramatsu K: **Two-component system VraSR positively modulates the regulation of cell-wall biosynthesis pathway in Staphylococcus aureus.** *Molecular Microbiology* 2003, **49**(3):807-821.

74. LaPlante KL, Leonard SN, Andes DR, Craig WA, Rybak MJ: **Activities of clindamycin, daptomycin, doxycycline, linezolid, trimethoprim-sulfamethoxazole, and vancomycin against community-associated methicillin-resistant Staphylococcus aureus with inducible clindamycin resistance in murine thigh infection and in vitro pharmacodynamic models.** *Antimicrobial Agents and Chemotherapy* 2008, **52**(6):2156-2162.

75. Lee MJ: *Phenotypic and genetic characterization of borderline oxacillin-resistant Staphylococcus aureus* University of Toronto; 1999.

76. Beltramini AM, Mukhopadhyay CD, Pancholi V: **Modulation of cell wall structure and antimicrobial susceptibility by a Staphylococcus aureus eukaryote-like serine/threonine kinase and phosphatase.** *Infection and Immunity* 2009, **77**(4):1406-1416.

77. Oliva B, O'Neill AJ, Miller K, Stubbings W, Chopra I: **Anti-staphylococcal activity and mode of action of clofazimine.** *Journal of Antimicrobial Chemotherapy* 2004, **53**(3):435-440.

78. Pillai SK, Wennersten C, Venkataraman L, Eliopoulos GM, Moellering RC, Karchmer AW: **Development of reduced vancomycin Susceptibility in methicillin-susceptible Staphylococcus aureus.** *Clinical Infectious Diseases* 2009, **49**(8):1169-1174.

79. Lin Y, Anderson MJ, Kohler PL, Strandberg KL, Olson ME, Horswill AR, Schlievert PM, Peterson ML: **Proinflammatory exoprotein characterization of toxic shock syndrome Staphylococcus aureus.** *Biochemistry* 2011, **50**(33):7157-7167.

80. Sieradzki K, Leski T, Dick J, Borio L, Tomasz A: **Evolution of a vancomycin-intermediate Staphylococcus aureus strain In vivo: multiple changes in the antibiotic resistance phenotypes of a single lineage of methicillin-resistant S. aureus under the impact of antibiotics administered for chemotherapy.** *Journal of Clinical Microbiology* 2003, **41**(4):1687-1693.

81. Ekkelenkamp MB, Sekkat M, Carpaij N, Troelstra A, Bonten MJM: **Endocarditis due to meticillin-resistant Staphylococcus aureus originating from pigs.** *Nederlands Tijdschrift Voor Geneeskunde* 2006, **150**(44):2442-2447.

82. Schijffelen MJ, Boel CE, van Strijp JA, Fluit AC: **Whole genome analysis of a livestock-associated methicillin-resistant Staphylococcus aureus ST398 isolate from a case of human endocarditis.** *BMC Genomics* 2010, **11**:376.

83. Robinson DA, Kearns AM, Holmes A, Morrison D, Grundmann H, Edwards G, O'Brien FG, Tenover FC, McDougal LK, Monk AB, Enright MC: **Re-emergence of early pandemic Staphylococcus aureus as a community-acquired meticillin-resistant clone.** *Lancet* 2005, **365**(9466):1256-1258.

84. Schaaff F, Reipert A, Bierbaum G: **An elevated mutation frequency favors development of vancomycin resistance in Staphylococcus aureus.** *Antimicrobial Agents and Chemotherapy* 2002, **46**(11):3540-3548.

85. Meehl M, Herbert S, Götz F, Cheung A: **Interaction of the GraRS two-component system with the VraFG ABC transporter to support vancomycin-intermediate resistance in Staphylococcus aureus.** *Antimicrobial Agents and Chemotherapy* 2007, **51**(8):2679-2689.

86. Howden BP, Johnson PDR, Ward PB, Stinear TP, Davies JK: **Isolates with low-level vancomycin resistance associated with persistent methicillin-resistant Staphylococcus aureus bacteremia.** *Antimicrobial Agents and Chemotherapy* 2006, **50**(9):3039-3047.

87. Hanaki H, Kuwahara-Arai K, Boyle-Vavra S, Daum RS, Labischinski H, Hiramatsu K: **Activated cell-wall synthesis is associated with vancomycin resistance in methicillin-resistant Staphylococcus aureus clinical strains Mu3 and Mu50.** *Journal of Antimicrobial Chemotherapy* 1998, **42**(2):199-209.

88. Coleman DC, Pomeroy H, Estridge JK, Keane CT, Cafferkey MT, Hone R, Foster TJ: **Susceptibility to antimicrobial agents and analysis of plasmids in gentamicinand methicillin-resistant Staphylococcus aureus from Dublin hospitals.** *Journal of Medical Microbiology* 1985, **20**(2):157-167.

89. Cui L, Neoh Hm, Shoji M, Hiramatsu K: **Contribution of vraSR and graSR point mutations to vancomycin resistance in vancomycin-intermediate Staphylococcus aureus.** *Antimicrobial Agents and Chemotherapy* 2009, **53**(3):1231-1234.

90. Stegger M, Price LB, Larsen AR, Gillece JD, Waters AE, Skov R, Andersen PS: **Genome sequence of Staphylococcus aureus strain 11819-97, an ST80-IV European community-acquired methicillin-resistant isolate.** *Journal of Bacteriology* 2012, **194**(6):1625-1626.

91. Kennedy AD, Porcella SF, Martens C, Whitney AR, Braughton KR, Chen L, Craig CT, Tenover FC, Kreiswirth BN, Musser JM, DeLeo FR: **Complete nucleotide sequence analysis of plasmids in strains of Staphylococcus aureus clone USA300 reveals a high level of identity among isolates with closely related core genome sequences.** *Journal of Clinical Microbiology* 2010, **48**(12):4504-4511.

92. Storrs MJ, Courvalin P, Foster TJ: **Genetic analysis of gentamicin resistance in methicillin- and gentamicin-resistant strains of Staphylococcus aureus isolated in Dublin hospitals.** *Antimicrobial Agents and Chemotherapy* 1988, **32**(8):1174-1181.

93. Crisostomo MI, Westh H, Tomasz A, Chung M, Oliveira DC, de Lencastre H: **The evolution of methicillin resistance in Staphylococcus aureus: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(17):9865-9870.

94. O'Neill AJ, Huovinen T, Fishwick CWG, Chopra I: **Molecular genetic and structural modeling studies of Staphylococcus aureus RNA polymerase and the fitness of rifampin resistance genotypes in relation to clinical prevalence.** *Antimicrobial Agents and Chemotherapy* 2006, **50**:298-309.

95. Nicola FG, McDougal LK, Biddle JW, Tenover FC: **Characterization of erythromycin-resistant isolates of Staphylococcus aureus recovered in the United States from 1958 through 1969.** *Antimicrobial Agents and Chemotherapy* 1998, **42**(11):3024-3027.

96. Daurel C, Huet C, Dhalluin A, Bes M, Etienne J, Leclercq R: **Differences in potential for selection of clindamycinresistant mutants between inducible erm(A) and erm(C) Staphylococcus aureus genes.** *Journal of Clinical Microbiology* 2008, **46**(2):546-550.

97. Xue H, Lu H, Zhao X: **Sequence diversities of serineaspartate repeat genes among Staphylococcus aureus isolates from different hosts presumably by horizontal gene transfer.** *PLoS ONE* 2011, **6**(5):e20332.

98. Chua KYL, Seemann T, Harrison PF, Monagle S, Korman TM, Johnson PDR, Coombs GW, Howden BO, Davies JK, Howden BP, Stinear TP: **The dominant Australian community-acquired methicillin-resistant Staphylococcus aureus clone ST93-IV [2B] is highly virulent and genetically distinct.** *PLoS ONE* 2011, **6**(10):e25887.

99. Chopra I, Roberts M: **Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance.** *Microbiology and Molecular Biology Reviews: MMBR* 2001, **65**(2):232-260, second page, table of contents.

100. Connell SR, Tracz DM, Nierhaus KH, Taylor DE: **Ribosomal Potection Proteins and their mechanism of tetracycline resistance.** *Antimicrobial Agents and Chemotherapy* 2003, **47**(12):3675-3681.

101. McAleese FM, Foster TJ: **Analysis of mutations in the Staphylococcus aureus clfB promoter leading to increased expression.** *Microbiology* 2003, **149**:99-109.

102. Werckenthin C, Schwarz S, Roberts MC: **Integration of pT181-like tetracycline resistance plasmids into large staphylococcal plasmids involves IS257.** *Antimicrobial Agents and Chemotherapy* 1996, **40**(11):2542-2544.

103. Hu M, Nandi S, Davies C, Nicholas RA: **High-level chromosomally mediated tetracycline resistance in Neisseria gonorrhoeae results from a point mutation in the rpsJ gene encoding ribosomal protein S10 in combination with the mtrR and penB resistance determinants.** *Antimicrobial Agents and Chemotherapy* 2005, **49**(10):4327-4334.

104. Sabath LD: **Mechanisms of resistance to beta-lactam antibiotics in strains of Staphylococcus aureus.** *Annals of Internal Medicine* 1982, **97**(3):339-344.

105. Drawz SM, Bonomo RA: **Three decades of betalactamase inhibitors.** *Clinical Microbiology Reviews* 2010, **23**:160-201.

106. **Guidelines for ATC classification and DDD assignment.** 2010.

107. Schmitz F, Sadurski R, Kray A, Boos M, Geisel R, Kohrer K, Verhoef J, Fluit AC: **Prevalence of macrolide-resistance genes in Staphylococcus aureus and Enterococcus faecium isolates from 24 European University Hospitals.** *Journal of Antimicrobial Chemotherapy* 2000, **45**(6):891-894.

108. Shakil S, Khan R, Zarrilli R, Khan AU: **Aminoglycosides versus bacteria-a description of the action, resistance mechanism, and nosocomial battleground.** *Journal of Biomedical Science* 2008, **15**:5-14.