The Use of Machine Learning Methodologies to Analyse Antibiotic and Biocide Susceptibility in *Staphylococcus aureus*

Joana Rosado Coelho¹*, João André Carriço^{1,2}, Daniel Knight^{3,4}, Jose-Luis Martínez⁵, Ian Morrissey^{3,6}, Marco Rinaldo Oggioni⁷, Ana Teresa Freitas¹

1 INESC-ID/IST, Technical University of Lisbon, Lisbon, Portugal, 2 Instituto de Microbiologia/Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal, 3 Quotient Bioresearch, Fordham, United Kingdom, 4 University of Western Australia, Queen Elizabeth II Medical Centre, Nedlands, Australia, 5 Departamento de Biotecnologia Microbiana, Centro Nacional de Biotecnologia-CSIC, Madrid, Spain, 6 IHMA Europe Sàrl, Epalinges, Switzerland, 7 Dipartimento Biotecnologie, Università di Siena, Siena, Italy

Abstract

Background: The rise of antibiotic resistance in pathogenic bacteria is a significant problem for the treatment of infectious diseases. Resistance is usually selected by the antibiotic itself; however, biocides might also co-select for resistance to antibiotics. Although resistance to biocides is poorly defined, different *in vitro* studies have shown that mutants presenting low susceptibility to biocides also have reduced susceptibility to antibiotics. However, studies with natural bacterial isolates are more limited and there are no clear conclusions as to whether the use of biocides results in the development of multidrug resistant bacteria.

Methods: The main goal is to perform an unbiased blind-based evaluation of the relationship between antibiotic and biocide reduced susceptibility in natural isolates of *Staphylococcus aureus*. One of the largest data sets ever studied comprising 1632 human clinical isolates of *S. aureus* originated worldwide was analysed. The phenotypic characterization of 13 antibiotics and 4 biocides was performed for all the strains. Complex links between reduced susceptibility to biocides and antibiotics are difficult to elucidate using the standard statistical approaches in phenotypic data. Therefore, machine learning techniques were applied to explore the data.

Results: In this pioneer study, we demonstrated that reduced susceptibility to two common biocides, chlorhexidine and benzalkonium chloride, which belong to different structural families, is associated to multidrug resistance. We have consistently found that a minimum inhibitory concentration greater than 2 mg/L for both biocides is related to antibiotic non-susceptibility in *S. aureus*.

Conclusions: Two important results emerged from our work, one methodological and one other with relevance in the field of antibiotic resistance. We could not conclude on whether the use of antibiotics selects for biocide resistance or *vice versa*. However, the observation of association between multiple resistance and two biocides commonly used may be of concern for the treatment of infectious diseases in the future.

Citation: Coelho JR, Carriço JA, Knight D, Martínez J-L, Morrissey I, et al. (2013) The Use of Machine Learning Methodologies to Analyse Antibiotic and Biocide Susceptibility in *Staphylococcus aureus*. PLoS ONE 8(2): e55582. doi:10.1371/journal.pone.0055582

Editor: John Conly, University of Calgary, Canada

Received June 4, 2012; Accepted December 27, 2012; Published February 19, 2013

Copyright: © 2013 Coelho et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by FCT (INESC-ID multiannual funding) through the PIDDAC Program funds and under project PEst-OE/EEI/LA0021/2011 and the FP7 Cooperation Work Programme: Food, Agriculture and Fisheries, and Biotechnologies, KBBE-227258 (BIOHYPO project). Quotient Bioresearch received part-funding from the European Union in the scope of BIOHYPO project. The work of J. R. Coelho was supported by the PhD grant SFRH/BD/33719/2009 from Fundação para a Ciência e a Tecnologia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: During the elaboration of this manuscript, lan Morrissey and Daniel Knight were employed by Quotient Bioresearch and belonged to the BIOHYPO European project. However, currently these two authors are no longer employed by Quotient Bioresearch. Ian Morrissey is an employee of IHMA Europe. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* E-mail: jrcoelho@kdbio.inesc-id.pt

Introduction

Antibiotics are among the most successful therapeutic agents so far developed by humankind. They have been used for decades for the treatment of bacterial infections in humans and animals. However, the widespread use of these compounds has led to the selection of resistant microorganisms, and consequently, to a loss of efficacy for this family of drugs [1,2]. This problem is further increased by the circulation of microorganisms among humans, animals and agricultural hosts creating the opportunity for the exchange of resistance genes/mechanisms directing the spread of resistance [3]. This is a public health problem on a global scale since antimicrobial resistance can result in increased disease burden, morbidity and mortality [4,5]. The problem is compounded by the fact that the pace of introduction of new chemotherapeutic drugs in the market and in clinical practice has slowed in the last few years [6]. The consequence of this slowdown is that many antibiotics lose their treatment effect against organisms that have developed antibiotic resistance and the use of other drugs, which are often reserved for last line treatment, is required. In some cases, therapeutic options are limited to just one antibiotic.

Besides antibiotics, another group of antimicrobials that might be relevant in selecting for antibiotic resistant bacteria are biocides. Biocides, used in antiseptics, disinfectants and preservatives, have been widely employed to control the growth of microorganisms such as bacteria, fungi, protozoa or viruses [7]. Their use has increased in a wide range of applications, namely in the domiciliary and in the healthcare environments, food industry and industrial settings, and in veterinary environments [8].

In recent years, there has been considerable discussion and controversy on whether biocide and antibiotic decreased susceptibility are connected [6,9,10]. This was highlighted some years ago, with the remarkable discovery of a shared bacterial target (FabI) between a biocide (triclosan) and an antibiotic (isoniazid) [11]. In the same study, it was described that triclosan exposure was followed by mutations in this bacterial target, which led to antibiotic resistance. Furthermore, even for those biocides which do not share a target with any antibiotic, reduced susceptibility can be achieved either when two elements, one conferring antibiotic resistance and another conferring reduced susceptibility to biocides are present in the same mobile element (co-resistance), or when a single resistance determinant, such as a multidrug (MDR) efflux pump, can provide resistance to both antibiotics and biocides (cross-resistance).

In this regard, some authors have suggested that there is an association between biocide-resistance and enrichment of antibiotic-resistance genes [12–14] and others have described increased bacterial resistance to biocides in areas where they have been used [15–17].

Despite the fact that the term resistance to biocides has been used in some papers [18,19], resistance to these compounds is still to be clearly defined, unlike the case of antibiotics. Chapman suggests that the definition of resistance to biocides should proceed in a manner analogous to the definition for antibiotic resistance [18]. For antibiotics, there are two definitions of resistance, either based on the clinical breakpoints or on the epidemiological cut-off values (ECOFFs) of the minimal inhibitory concentration (MIC) whose values may differ. The setting of clinical breakpoints is a widely debated topic among the scientific, regulatory, and industrial community. Clinical breakpoints are discriminatory antimicrobial concentrations used to classify isolates as resistant, intermediate or susceptible, on the basis of the relevance of these values for treatment of infections [20]. Breakpoints are defined based on many factors, namely pharmacokinetic, dose-effect relationships from *in vitro* studies and animal studies, available formulations, clinical indications and practices, and target organisms [21]. Moreover, Monte Carlo simulations have been applied to help in establishing clinical breakpoints [22]. On the other hand, ECOFFs are used to define the intrinsic susceptibility to antibiotics of a given microbial species based on the MICs of wild type populations [23]. The ECOFF is independent of clinical treatment and will not be altered by changing circumstances [21]. Under this definition, any isolate presenting a MIC above the ECOFF value of the agent in question is considered as being non-susceptible. In the case of biocides, to date, none of these definitions have been established and we still do not have any regulatory or industrial organization involved in the normalization of classical breakpoints or of ECOFF values, which is important for studying biocide reduced susceptibility and changes therein.

There are many studies on the evaluation and evolution of antibiotic resistance [24-27] and some on the efficacy of biocides [27-29]. However, studies comparing biocide and antibiotic resistance are less common and the data sets used in these studies are typically small. Some studies follow the approach of simple statistical analysis using bivariate correlations and tests for independence [30–33]. Another study [34] uses a simple statistical analysis approach and hierarchical clustering on a small data set to identify statistical correlations between bacterial patterns of reduced susceptibility for disinfectants and antibiotics. Another paper [35] describes the use of classical association tests as ANOVA, Spearman-Rho correlations and principal component analysis to investigate the existence of correlations between the biocides and the antibiotics tested. The authors do not conclude whether there is an association or not, but with the data and tools used in the analysis it is hard to conclude if increased biocide resistance is a cause of increased antibiotic resistance. All of these studies deal with small data sets and the results present no evidence of statistical correlation between those antimicrobials, either antibiotics or biocides.

In this paper, we propose a new framework to analyse and cope with the scale and complexity of biocide susceptibility data, antibiotic susceptibility data and patient demographic data. In one of the largest phenotype studies ever performed on biocide and antibiotic reduced susceptibility, a total of 1632 world-wide clinical strains of Staphylococcus aureus (S. aureus) strains were analysed. S. aureus is a major human pathogen, a major cause of nosocomial infections, and also a significant cause of food borne infection. We combined different machine learning methodologies, namely decision trees and clustering, to explore the data in order to find biological and statistical significant results. Machine learning methods can be very powerful in finding unexpected patterns and hidden knowledge as well as establishing new rules from large data sets [36,37]. Furthermore, using this approach we established a framework of analysis that may help determine values for the ECOFF of two biocides commonly used in our daily lives.

Materials and Methods

Strain Collection

A total of 1632 *S. aureus* strains were analysed in this study, taken from the collection of Quotient Bioresearch (Fordham, UK) [15]. The strains were collected between 2002 and 2003, from different geographical origins (including Europe, East Asia, Oceania and both North and South Americas) and represented both hospital and community acquired infections.

Antibiotic and Biocide Susceptibility Testing

The study was conducted using the Clinical and Laboratory Standards Institute (CLSI) standardised broth MIC and MBC (minimum bactericidal concentration) methodology [38,39]. Application of CLSI procedures to the study ensured that MIC data was generated using internationally recognized and standardised methods. Biocide and antibiotic susceptibility was determined by measuring both bacteriostatic activity (MIC) and bactericidal activity (MBC). The MIC was determined as the lowest concentration of compound required to completely inhibit the growth of an organism. The MBC was determined as the lowest concentration of a compound required to produce a 1×10^3 reduction in colony forming units [38,39].

Bivariate Correlations

A bivariate correlation analysis was performed between biocides and antibiotics' variables. Spearman's correlation coefficient was computed for each bivariate combination of these variables in order to find non-linear associations between the biological variables. Bivariate correlations were calculated using Matlab[®] 7.10.0.499 (R2010a). For each computed Spearman's correlation coefficient, a hypothesis test was performed in order to test for statistical association between each pair of variables and a *p-value* calculated, assuming statistical significance at a level of 5%. When performing a set of multiple statistical inferences it is likely that a multiple testing problem of incorrectly rejecting the null hypothesis will occur [40]. To control the family-wise error rate, the classical conservative approach of Bonferroni correction was applied [41] leading to a critical value at level

$$\alpha_{pt} \approx \frac{\alpha_{pf}}{n_t} = \frac{0.05}{12 \times 13} \approx 3.2 \times 10^{-4}$$

where α_{pt} is the probability of Type I error (false positive) *per test*; α_{pf} is the probability of Type I error *per family* of tests, also called, *familywise* or the *experimentwise* alpha; n_t is the number of pairwise tests performed.

Decision Trees with Permutation Tests validation and Visualisation

Within Machine Learning, Decision Tree algorithms are some of the most widely used for classification [37]. Learned trees have an intuitive form and represent a set of if-then rules. This method is generally very well accepted by domain experts since it is very simple to understand and interpret. For this particular case, decision trees were built using all the data as a training set. A decision tree inductor algorithm was applied in such a way that whenever it encountered a set of items (training set), it identified the attribute that discriminated the various instances most clearly. Features that lead to the highest information gain were successively chosen to build the decision tree. Trees complexity was controlled by a pruning stopping criteria. In the context of this analysis, the objective was to apply a pruning that could generate less complex decision trees: on the one hand it would have better readability, and on the other, it would include enough features to explain the data. Leaves were allowed to have a minimum number of objects of 1% of the data set dimension.

Thirteen different decision trees were obtained, one for each different antibiotic. The antibiotic classification (Susceptible) or Non-Susceptible) was selected as the dependent variable whereas the independent variables included MIC and MBC values for the four biocides', the year and region of collection, the age and gender of the host, the source and type of infection as well as the status of the host: hospital inpatient or outpatient. Decision trees were produced using Weka (Waikato Environment for Knowledge Analysis) Version 3.6.4.

rmine whether the observed decision trees is due to the randomness introduced in the sampling or not [42]. These are nonparametric statistical tests for determining statistical significance by rearranging the labels of a data set. Permutation tests can help reducing the multiple testing burden [43] and can be used to compare statistical tests [44].

ngement of the labels is performed n times and the values of the statistic in use are calculated giving rise to a distribution of permutation values. On the other hand, a test statistic is computed from the experimental data set and then compared with the distribution of permutation values previously obtained. The main drawback of this method is that it needs a very large number of permutations to accurately estimate small p-values, a procedure that is computationally expensive [45].

A hypothesis test was performed where the null hypothesis, H_{0} , represents the absence of the effect given by the decision tree profile obtained. Small p-values are evidence against the null hypothesis and in favour of a real effect in the population. The statistic used is based on the conditional probability of the Fisher's Exact Test [46].

In order to visualise the distribution of the antibiotic resistance patterns found in the dataset and its relationships with the biocide MICs, graphic matroids were created. Graphic matroids are representations of unrooted trees similar to a minimum spanning tree but using a set of rules instead of distances. For each isolate, a pattern of susceptibility was obtained by binary coding the antibiotic susceptibility of each of the 13 antibiotics. Each node on the resulting tree represents a unique antibiotic susceptibility pattern found in the dataset. The graphic matroid was created using the goeBURST algorithm rules [47] with the different susceptibility patterns as an input. The size of each node is a function of the number of isolates found in the dataset with that pattern.PHYLOViZ software [48] was used to create the graphic matroid and the visualisation of the decision tree rules.

Clustering

The goal of clustering is the separation of objects into groups or clusters using only the input vectors. This separation is based on a measure of similarity between objects [49,50].

Hierarchical clustering was used to produce dendograms that could demonstrate co-resistance of antibiotics among the isolates. The measure of dissimilarity was based on the Euclidean distance metric and the hierarchical clustering method with average linkage was chosen. Clustering was obtained with Cluster 3.0 using the C Clustering Library version 1.50 and the visualisation was performed with Java TreeView 1.1.6r2.

Results

In this paper, we first performed the standard approach described in the literature [30–33] for epidemiological data analysis by calculating the bivariate correlations between biocide and antibiotic phenotypes. This approach failed to find any associations between the variables in analysis (data not shown). We then followed a different approach represented in Figure 1 by applying machine learning techniques, namely decision trees validated by permutation tests and clustering analysis of the data set followed by a visualisation step. We propose this new framework, based on machine learning methodologies, to analyse the scale and complexity of molecular epidemiology data and its application to study the potential linkage between antibiotic resistance and biocide reduced susceptibility.

Generation of the Data Set

A collection of 1632 *S. aureus* strains isolated between 2002 and 2003, from different geographical origins, and representing both hospital and community acquired infections was analysed [15]. For all isolates, the MIC values of 13 different antibiotics, namely amoxicillin/clavulanate, cefuroxime, cefaclor, cefpodoxime, clindamycin, erythromycin, clarithromycin, azithromycin, telithromycin, ciprofloxacin, levofloxacin, gatifloxacin and moxifloxacin, were determined in addition to the susceptibility data for methicillin. A bacterial isolate was considered non-susceptible to an antibiotic agent when it tested resistant or intermediate when using clinical breakpoints as interpretative criteria according to the Clinical and Laboratory Standards Institute (CLSI) [51]. In addition, for all isolates, susceptibility data for four different biocides (MIC and MBC), namely chlorhexidine, benzalkonium



Figure 1. New framework of analysis. In this paper, we propose the use of machine learning techniques to analyse epidemiological data sets, namely decision trees validated by permutation tests and clustering analysis of the data set followed by a visualisation step of the results. This methodology is successful in dealing with the complexity of molecular epidemiology data and its application in studying the potential linkage between antibiotic resistance and biocide reduced susceptibility.

chloride, sodium hypochlorite and triclosan, were obtained. Since there are no established breakpoints for biocides the information on the susceptibility to these compounds was introduced in the data set as actual MIC and MBC values and not as categories (susceptible or non-susceptible) like in the case of antibiotics. According to the classification of Magiorakos et al [52], strains were considered as multidrug resistant (MDR) when they were resistant to at least three different families of antibiotics. All the MIC and MBC values are in a log₂ scale due to the methodology used to determine the phenotypic data.

A data set was generated which included, for each *S. aureus* isolate: the MIC value and the susceptible/non-susceptible classification for each of the 13 antibiotics; the MIC and MBC values for each biocide; the year (2002 or 2003) and region of collection (Europe, East Asia, Oceania and both North or South Americas); the age (ranging from 0 to 97 years old) and gender of the host; the source of the sample and infection type; and if the host was an inpatient or outpatient of the hospital were collection occurred.

Bivariate Correlations

This is the standard approach reported in the literature to analyse data of epidemiological data sets. However, using this methodology there was no clear correlation between the biocide and antibiotic phenotype variables. Indeed, the data analysed showed weak bivariate correlations. This is a result that matches with previous studies of smaller data sets. After performing the correction for family-wise error rate, we found some pairs of variables (represented in bold in Figure S1, in supporting material) with a statistically significant Spearman's correlation coefficient. In spite of having a statistically significant correlation coefficient, the pair of variables in bold have a considerably low absolute value for the coefficient so one cannot assume a strong correlation between those variables. The ratio between MBC and MIC (Fold Change) of each biocide was also analysed as this may indicate a difference in biocide reduced susceptibility. Kendall's correlation coefficient was also computed for the same pairwise test variables and the results were similar (data not shown).

Decision Trees

Since standard statistical tests did not enable us to obtain a clear picture of the relationships between antibiotic and biocide reduced susceptibilities, we decided to apply a different approach. The approach is shown in Figure 1 and is based on the application of machine learning methodologies that allow extraction of hidden knowledge and unexpected patterns from data as well as establishing new rules from large data sets [36].

After applying the tree induction algorithm to each of the 13 antibiotics under analysis, a common pattern emerged from the decision trees. The structure of these trees shows interesting relationships. Among all independent variables, the decision trees modelled the dependent variable based on two specific predictor variables as shown in Figure 2. Whereas most independent variables, including region and year of collection, were not the ones that better discriminate antibiotic susceptibility, the MICs of two common biocides chlorhexidine (CLX) and benzalkonium chloride (BZC) are the best predictors of susceptibility and non-susceptibility to each of the 13 antibiotics, with a *p-value* < 1×10^{-4} , validated with permutation tests. This is a common profile to all 13 antibiotics and basically it can be summarized in two simple rules:

- 1. MIC(CLX) ≤2 mg/L: strains are susceptible to the antibiotic in test
- 2. **MIC(CLX)** >2 mg/L \land **MIC(BZC)** >2 mg/L: strains are non-susceptible to the antibiotic in test

The leaf **MIC(CLX)** >2 mg/L \land **MIC(BZC)** \leq 2 mg/L involves a small number of strains, a total of 13 strains for each of the 13 antibiotics, representing less than 0.8% of the data set dimension.

The decision tree leaves have an average misclassification rate of 16% and 23% for the first and second rule, respectively. Addressing this misclassification rate, we have further analysed these isolates. In fact, the leaf of the first rule has an average of 84% true positives, which represent 74% of the total number of isolates of the data set. On the other hand, the second rule has an average of 77% correctly classified isolates representing 8% of the total number of isolates of the data set. Furthermore, the population of non-susceptible isolates is, on average, divided in a proportion of 2:1 between the first and second rules.

In spite of not having strong correlations between any biocide or antibiotic variables, we have found interesting relationships between two specific biocides and all the 13 antibiotics analysed. This methodology allows finding out relationships other than nonlinear monotonic ones. The CLX and BZC MICs that separate antibiotic susceptible from non-susceptible are very consistent among all the 13 antibiotics, 2 mg/L being the borderline for two biocides. These two biocides commonly used in our daily life belong to different structural families: Chlorhexidine a bisbiguanide and Benzalkonium Chloride a quaternary ammonium compound. Furthermore, using this approach we can suggest values for the ECOFF of Chlorhexidine and Benzalkonium Chloride, indicating a potential ECOFF of MIC <2 mg/L for



Figure 2. Decision tree profile for each of the 13 antibiotics. The level of the MIC of two common biocides is the best feature to predict the susceptibility and non susceptibility to each of the 13 antibiotics with ap-value $< 1 \times 10^{-4}$. doi:10.1371/journal.pone.0055582.g002

the susceptible population and of MIC >2 mg/L for the non-susceptible strains for both biocides.

Visualisation

Figure 3 shows the distribution of the susceptibility patterns. Each link represents a unique change in the overall pattern. Individual clusters are produced that are different among themselves by two or more differences (variations) in the pattern. A Blue to Red colour map was created to help the visualisation of the overall antimicrobial resistance dispersal on the tree: the isolates in dark blue are susceptible to all of the antibiotics, while isolates in red share a greater non-susceptibility to antibiotics. The isolates in blue are clearly concentrated on the right-hand side of the graph and the isolates in red, on the left-hand side, represent a gradient of multidrug resistance in our dataset.

Based on this distribution of susceptible complexes we can also visualise each of the decision tree rules previously described. Since all the antibiotics have the same two rules obtained using the decision trees approach, we can colournap each of these rules in order to visualise them in the graphic matroid. Figure 4 represents the isolates classified as susceptible to a given antibiotic in *green* based on rule **1**. They are predominantly present on the right-hand side of the graph, which corresponds to the region of antibiotic susceptible strains. There are some false positive strains represented in green on the left-hand side of the graph. Moreover, Figure 5 represents in red the isolates classified as non-susceptible based on rule **2**. These isolates are largely represented on the left-hand side of the graph; however, some isolates in red are also present on the right-hand side, corresponding to the susceptible antibiotic region of the tree.

Clustering

We found two groups of antibiotic resistant strains: those that are misclassified as susceptible according with rule **1** and those that are correctly classified as non-susceptible according to rule **2**. To address whether these subpopulations represent specific subsets of non-susceptible strains they were further analysed. For this purpose, a hierarchical clustering analysis was performed on the two subpopulations of non-susceptible strains.

Of note here is that even for each single antibiotic whose number of misclassified strains is low according to rule 1 (in the range of 16%), the final percentage of isolates presenting resistance to at least one antibiotic is higher, about 34%. In agreement with

this issue, the clustering of misclassified isolates of rule 1, which are in fact non-susceptible isolates, shows that few of them have cumulative resistance to different antibiotic families (Figure 6) sharing only 17.5% of co-resistance. This result suggests that, for this group of non-susceptible strains, acquisition of resistance to one antibiotic is independent of the acquisition of resistance to another and is independent as well of the development of reduced susceptibility to the tested biocides. A different situation can be observed for the non-susceptible isolates, classified based on rule 2(Figure 7). This population presents a high prevalence of strains, about 62%, displaying a phenotype of multiple resistance to the different families of antibiotics tested and presents as well reduced susceptibility to both biocides analysed.

Discussion

The possibility that the use of biocides might be a risk for the selection of antibiotic resistant bacteria is a controversial issue still under discussion. Two fundamental concerns are raised: 1) there are some elements such as multidrug efflux pumps that can confer simultaneously resistance to antibiotics and biocides (cross-resistance) when they are over expressed, and 2) antibiotic resistance determinants transferred by horizontal gene transfer are usually associated to other determinants that are transferred together on the same genetic element, some of which are involved in resistance to biocides (co-resistance). In both situations, the presence of the biocide will select for antibiotic resistance and *vice versa*.

Several articles have shown the existence of the aforementioned linkage between antibiotic resistance and biocide reduced susceptibility [19,53–59], however, most of these studies have been performed *in vitro* using model strains and the evidence of such association by means of epidemiological studies is much weaker. In fact, some papers that describe the use of classical association tests such as ANOVA, Spearman-Rho correlations and principal component analysis, arrive at the conclusion that "it is very difficult to support a hypothesis that increased biocide resistance is a cause of increased antibiotic resistance" [35]. The authors do not conclude whether there is or there is not an association, but that with the data and tools used in the analysis "it is very difficult" to reach such a conclusion.

In the case of the potential associations between biocides and antibiotic resistance, there are mainly two reasons for this lack of potential association. First, the number of isolates analysed in



Figure 3. Graphic matroid of *5. aureus* **antibiotic susceptibility.** Graphic matroid of *5. aureus* antibiotic susceptibility with a red-blue colour map. Isolates in *blue* are susceptible to a greater number of antibiotics and isolates in *red* share a greater non susceptibility to antibiotics. S-Susceptible. NS-Non-susceptible. ATB-Antibiotics. doi:10.1371/journal.pone.0055582.g003



Figure 4. Graphic matroid of *S. aureus* isolates colour mapped based on the first decision tree rule. Isolates with CLX MIC \leq 2 (classified as susceptible to each antibiotic, based on the first rule) are coloured in *green*. In *grey* are represented the remaining isolates. doi:10.1371/journal.pone.0055582.g004



Figure 5. Graphic matroid of *S. aureus* isolates colour mapped based on the second decision tree rule. Isolates with CLX MIC >2ABZC MIC >2 (classified as non-susceptible to each antibiotic, based on the second rule) are coloured in *red*. In *grey* are represented the remaining isolates. doi:10.1371/journal.pone.0055582.g005



Figure 6. Clustering of the non-susceptible population for rule 1. The different clusters for the non-susceptible population of rule 1 are shown in the dendogram with the respective global percentage of shared resistance of each set of antibiotics. Among the 487 isolates eligible for the clustering analysis, only 17.5% isolates are non-susceptible to all the 13 antibiotics. We have identified 4 clusters of antibiotics which may be related with the potential mechanisms of resistance already reported on the literature [60]. doi:10.1371/journal.pone.0055582.g006



Figure 7. Clustering of the non-susceptible population for rule 2. The different clusters for the non-susceptible population of rule 2 are shown in the dendogram with the respective global percentage of shared resistance of each set of antibiotics. Among the 155 isolates eligible for the clustering analysis, 61.9% isolates are non-susceptible to all the 13 antibiotics. We have identified 4 clusters of antibiotics which may be related with the potential mechanisms of resistance already reported on the literature [60]. doi:10.1371/journal.pone.0055582.q007

epidemiological studies with antibiotic and biocide susceptibility data is not usually high enough to get statistically robust conclusions. Secondly, performing these epidemiological studies highlights several difficulties due to the complexity of this type of data, mainly because of the large number of variables from different sources and the typical non-linearities of biological signals. To improve our current analysis, we have used decision trees validated with permutation tests for analysing biological data sets and in particular to unveil hidden associations between antibiotic resistance and biocide reduced susceptibility in S. aureus. These methodologies were successful in finding rules within the data set where the standard and most used statistical approaches found no associations. We studied the largest collection of isolates so far analysed for this purpose (1632 isolates). The main conclusion emerging from these unbiased analyses is that reduced susceptibility to two biocides, CLX and BZC, which belong to different structural families, is associated with resistance to several antibiotics, whereas resistance to just one antibiotic is not rare even in biocide-susceptible populations. A recent paper has proposed a classification of strains presenting resistance in three categories: multidrug resistant (MDR), when they present resistance to antibiotics belonging to three different families; extensively drug resistant (XDR), when they remain susceptible just to one or two families of antibiotics; and pan-drug resistant (PDR) when they are resistant to all known antibiotics [52]. Following these rules, 73.5% of isolates presenting low susceptibility to biocides are MDR S.

aureus (Figure 7). The fact that neither the time nor the place of isolation were relevant predictors of resistance indicates that there is not a bias in the type of isolates used in our study.

As stated above, simultaneous reduced susceptibility to biocides can be achieved either by the acquisition of elements containing an array of genes, each one encoding resistance to a different drug, or by the overproduction of efflux pumps capable of extruding both biocides and antibiotics. Although we have not studied here the molecular basis of resistance of our isolates, our results strongly suggest that reduced susceptibility to CLX and BZC is associated to multidrug resistance in S. aureus. Whether this association is the consequence of biocide usage or it is a side effect of the selection of resistance by antibiotics remains to be established. It is relevant to note that using machine learning technologies we have been able to define two subpopulations with very different patterns of antibiotic non-susceptibility and the phenotype of two biocides, CLX and BZC, is enough to discriminate them. These populations were not apparent using more classical statistical tests. We have consistently found that a CLX MIC greater than 2 mg/L and a BZC MIC greater than 2 mg/L is related to antibiotic nonsusceptibility in S. aureus. Given that those values distinguished between two populations presenting different phenotypes, it might be possible that these values represent cut-off values for biocide non-susceptibility. In this regard, our study may help establishing the ECOFF value of two biocides, CLX and BZC, commonly used in our daily lives. Nevertheless, the ECOFF values need to be confirmed with other types of analyses [15]. These studies are now in progress in our laboratories. Two important results, one methodological and another with relevance in the field of antimicrobial resistance emerge from our work. Machine learning methods are useful for the extraction of hidden knowledge and unexpected patterns from large datasets, such as those generated in our work, in situations in which more classical statistical tests may fail to find associations. Multidrug resistance is associated with reduced susceptibility to benzalkonium chloride and chlorhexidine in S. aureus, whereas this association is not observed at the level of resistance to a single antibiotic. From our results, we cannot conclude on whether the use of antibiotics selects for biocide resistance or vice versa. However, the observation of association between multiple resistance and two biocides commonly used may be of concern for the treatment of infectious diseases, in the future.

Supporting Information

Figure S1 Matrix with Spearman's correlation coefficients between biocides phenotypes (MIC, MBC or fold

References

- Falagas ME, Bliziotis IA (2007) Pandrug-resistant Gram-negative bacteria: the dawn of the post-antibiotic era? International Journal of Antimicrobial Agents 29: 630–636. Available: http://www.ncbi.nlm.nih.gov/pubmed/17306965. Accessed 2012 Sep 6.
- Jansen WTM, Van Der Bruggen JT, Verhoef J, Fluit AC (2006) Bacterial resistance: a sensitive issue complexity of the challenge and containment strategy in Europe. Drug resistance updates reviews and commentaries in antimicrobial and anticancer chemotherapy 9: 123–133. Available: http://www.ncbi.nlm.nih. gov/pubmed/16807066. Accessed 2012 Sep 6.
- WHO (2007) The world health report-A safer future: global public health security in the 21st century. Available: http://www.who.int/whr/2007/en/ index.html. Accessed 2012 Sep 10.
- EARSS (2005) Annual Report European Antimicrobial Resistance Surveillance System. Available: http://www.ecdc.europa.eu/en/activities/ surveillance/EARS-Net/Documents/2005_EARSS_Annual_Report.pdf. Accessed 2012 Sep 10.
- Harbarth S, Samore MH (2005) Antimicrobial resistance determinants and future control. Emerging Infectious Diseases 11: 794–801. Available: http:// www.ncbi.nlm.nih.gov/pubmed/15963271. Accessed 2012 Sep 7.
- Russell AD (2002) Antibiotic and biocide resistance in bacteria: Introduction. Symposium series (Society for Applied Microbiology): 1S–3S. Available: http:// www.ncbi.nlm.nih.gov/pubmed/12481823. Accessed 2011 May 30.
- SCCS (2010) Opinion on Triclosan-Antimicrobial Resistance. European Commission: Scientific Committee on Consumer Safety. Available: http://ec. europa.eu/health/scientific_committees/consumer_safety/docs/sccs_0_023. pdf. Accessed 2012 Sep 7.
- Maillard J-Y, Denyer SP (2009) Emerging bacterial resistance following biocide exposure: should we be concerned? Chimica oggi 27: 26–28. Accessed 2012 Sept 7.
- Aiello AE, Larson EL, Levy SB (2007) Consumer antibacterial soaps: effective or just risky? Clinical infectious diseases: an official publication of the Infectious Diseases Society of America 45 Suppl 2: S137–47. Available: http://www.ncbi. nlm.nih.gov/pubmed/17683018. Accessed 2012 Apr 12.
- Levy SB (2001) Antibacterial household products: cause for concern. Emerging Infectious Diseases 7: 512–515. Available: http://www.pubmedcentral.nih.gov/ articlerender.fcgi?artid = 2631814&tool = pmcentrez&rendertype = abstract. Accessed 2012 May 30.
- Parikh SL, Xiao G, Tonge PJ (2000) Inhibition of InhA, the enoyl reductase from *Mycobacterium tuberculosis*, by triclosan and isoniazid. Biochemistry 39: 7645– 7650. Available: http://pubs.acs.org/doi/abs/10.1021/bi0008940. Accessed 2012 Sep 7.
- Lucey B, Crowley D, Moloney P, Cryan B, Daly M, et al. (2000) Integronlike structures in *Campylobacter spp.* of human and animal origin. Emerging Infectious Diseases 6: 50–55. Available: http://www.pubmedcentral.nih.gov/articlerender. fcgi?artid = 2627975&tool = pmcentrez&rendertype = abstract. Accessed 2011 Jun 14.
- Lyon BR, Skurray R (1987) Antimicrobial resistance of *Staphylococcus aureus*: genetic basis. Microbiology and Molecular Biology Reviews: MMBR 51: 88– 134. Available: http://mmbr.asm.org. Accessed 2012 Sep 7.
- 14. Paulsen IT, Littlejohn TG, Rådström P, Sundström L, Sköld O, et al. (1993) The 3' conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. Antimicrobial Agents and Chemotherapy 37: 761–768. Available: http://www.pubmedcentral.nih.gov/

change) and antibiotics MIC values. *P*-values were also computed in order to find statistical significant associations. In blue and bold, we have the statistical significant associations, i.e., the correlation coefficients whose *p*-value is lower than the critical value $\alpha_{pt} \approx 3.2 \times 10^{-4}$.

(TIF)

Acknowledgments

Authors are thankful to all the participants of the BIOHYPO European project (http://kdbio.inesc-id.pt/biohypo).

Author Contributions

Conceived and designed the experiments: JRC ATF. Performed the experiments: JRC DK. Analyzed the data: JRC ATF. Contributed reagents/materials/analysis tools: JRC DK IM. Wrote the paper: JRC JAC DK JLM IM MRO ATF.

articlerender.fcgi?artid = 187754&tool = pmcentrez&rendertype = abstract. Accessed 2011 Jun 14.

- 15. Ciusa ML, Furi L, Knight D, Decorosi F, Fondi M, et al. (2012) A novel resistance mechanism to triclosan that suggests horizontal gene transfer and demonstrates a potential selective pressure for reduced biocide susceptibility in clinical strains of *Staphylococcus aureus*. International Journal of Antimicrobial Agents 40: 210–220. Available: http://linkinghub.elsevier.com/retrieve/pii/ S0924857912002051?showall = true. Accessed 2012 Sep 7.
- Ferber D (2000) Superbugs on the Hoof? Science 288: 792–794. Available: http://www.sciencemag.org/content/288/5467/792.short. Accessed 2012 Sep 7.
- Mølbak K, Baggesen DL, Aarestrup FM, Ebbesen JM, Engberg J, et al. (1999) An Outbreak of Multidrug-resistant *Salmonella enterica* serotype Typhimurium DT104. The New England Journal of Medicine 341: 1420–1425. Available: http://www.ncbi.nlm.nih.gov/pubmed/10547404. Accessed 2012 Sep 7.
- Chapman J (1998) Characterizing bacterial resistance to preservatives and disinfectants. International Biodeterioration & Biodegradation 41: 241–245. Available: http://linkinghub.elsevier.com/retrieve/pii/S0964830598000250. Accessed 2011 May 30.
- McMurry LM, Oethinger M, Levy SB (1998) Overexpression of marA, soxS, or acrAB produces resistance to triclosan in laboratory and clinical strains of *Escherichia coli*. FEMS Microbiology Letters 166: 305–309. Available: http:// www.ncbi.nlm.nih.gov/pubmed/9770288. Accessed 2011 Jun 3.
- MacGowan AP, Wise R (2001) Establishing MIC breakpoints and the interpretation of *in vitro* susceptibility tests. The Journal of Antimicrobial Chemotherapy 48 Suppl 1: 17–28. Available: http://www.ncbi.nlm.nih.gov/ pubmed/11420334. Accessed 2012 Apr 16.
- EUCAST (2012) European Committee on Antimicrobial Susceptibility Testing-Clinical Breakpoints. Available: http://www.eucast.org/clinical-breakpoints/. Accessed 2012 Sep 10.
- Ambrose PG (2006) Monte Carlo Simulation in the Evaluation of Susceptibility Breakpoints: Predicting the Future: Insights from the Society of Infectious Diseases Pharmacists. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy 26: 129–134. Available: http://dx.doi.org/10.1592/phco. 2006.26.1.129. Accessed 2012 Sep 7.
- Turnidge J, Kahlmeter G, Kronvall G (2006) Statistical characterisation of bacterial wild-type MIC value distributions and the determination of epidemiological cut-off values. Clinical Microbiology and Infection 12: 418– 425. Available: http://www.ncbi.nlm.nih.gov/pubmed/16643517. Accessed 2012 Apr 16.
- Thabaut A, Durosoir JL, Meyran M (1983) La sensibilite de Staphylococcus aureus aux antibiotiques en milieu hospitalier evolution et etat actuel. Médecine et Maladies Infectieuses 13: 128–132. Available: http://www.sciencedirect.com/ science/article/pii/S0399077X83801103. Accessed 2012 Sep 7.
- Tankovic J, Aubry-Damon H, Leclercq R (1997) Résistance aux antibiotiques autres que les bêta-lactamines chez *Staphylococcus aureus*. Médecine et Maladies Infectieuses 27, Supple: 207–216. Available: http://www.sciencedirect.com/ science/article/pii/S0399077X97800224. Accessed 2012 Sep 7.
- 26. Hugard L, Ndoye B, Saccharin C (1994) Evolution des résistances aux antibiotiques. Comparaison des principaux germes isolés de pus en 1989 et 1992 à l'Hôpital Principal de Dakar. Médecine d'Afrique Noire 41: 31–36. Available: http://cat.inist.fr/?aModele = afficheN&cpsidt = 3969009. Accessed 2012 Sep 7.

- Espigares E, Bueno A, Espigares M, Gálvez R (2006) Isolation of Salmonella serotypes in wastewater and effluent: Effect of treatment and potential risk. International Journal of Hygiene and Environmental Health 209: 103–107. Available: http://www.ncbi.nlm.nih.gov/pubmed/16373208. Accessed 2012 Sep 6.
- Rouillon S, Ourdanabia S, Jamart S, Hernandez C, Meunier O (2006) Étude de l'efficacité d'un produit détergent désinfectant pour sols et surfaces sur les souches bactériennes isolées à partir de l'environnement hospitalier. Pathologie Biologie 54: 325–330. doi:10.1016/j.patbio.2006.01.011. Accessed 2012 Feb 29.
- Rutala WA, Weber DJ (2001) Surface disinfection: should we do it? The Journal of Hospital Infection 48: 64–68. Accessed 7 March 2012.
- Aase B, Sundheim G, Langsrud S, Rørvik LM (2000) Occurrence of and a possible mechanism for resistance to a quaternary ammonium compound in *Listeria monocytogenes*. International Journal of Food Microbiology 62: 57–63. Available: http://www.ncbi.nlm.nih.gov/pubmed/11139022. Accessed 2012 March 2.
- Beier RC, Bischoff KM, Ziprin RL, Poole TL, Nisbet DJ (2005) Chlorhexidine susceptibility, virulence factors, and antibiotic resistance of beta-hemolytic *Escherichia coli* isolated from neonatal swine with diarrhea. Bulletin of Environmental Contamination and Toxicology 75: 835–844. Available: http://www.ncbi.nlm.nih.gov/pubmed/16400568. Accessed 2012 Mar 7.
- Soumet C, Ragimbeau C, Maris P (2005) Screening of benzalkonium chloride resistance in *Listeria monocytogenes* strains isolated during cold smoked fish production. Letters in Applied Microbiology 41: 291–296. Available: http:// www.ncbi.nlm.nih.gov/pubmed/16108923. Accessed 2012 Mar 2.
- 33. Copitch JL, Whitehead RN, Webber MA (2010) Prevalence of decreased susceptibility to triclosan in *Salmonella enterica* isolates from animals and humans and association with multiple drug resistance. International Journal of Antimicrobial Agents 36: 247–251. Available: http://www.ncbi.nlm.nih.gov/ pubmed/20541914. Accessed 2012 Mar 2.
- 34. Beier RC, Anderson PN, Hume ME, Poole TL, Duke SE, et al. (2011) Prevalence of decreased susceptibility to triclosan in *Salmonella enterica* isolates from animals and humans and association with multiple drug resistance. Foodborne Pathogens and Disease 8: 593–600. Available: http://www.ncbi.nlm. nih.gov/pubmed/21235389. Accessed 2012 Mar 7.
- Lambert RJW (2004) Comparative analysis of antibiotic and antimicrobial biocide susceptibility data in clinical isolates of methicillin-sensitive Staphylacoccus aureus, methicillin-resistant Staphylacoccus aureus and Pseudomonas aeruginosa between 1989 and 2000. Journal of Applied Microbiology 97: 699–711. Available: http://dx.doi.org/10.1111/j.13652672.2004.02345.x. Accessed 2012 Sep 17.
- Adriaans P, Zantinge D (1997) Data Mining. Addison-Wesley Longman Publishing Co., Inc. p.
- 37. Mitchell T (1997) Machine Learning. McGraw-Hill Education. p.
- CLSI (2009) Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that grow aerobically; Approved Standard - Eighth Edition. CLSI document M07-A8. Available: http://www.clsi.org/source/orders/free/m07a8.pdf. Accessed 2011 Jun 8.
- NCCLS (1999) Methods for Determining Bactericidal Activity of Antimicrobial Agents; Approved Guideline. NCCLS document M26-A. Available:http:// www.clsi.org/source/orders/free/m26-a.pdf. Accessed 2012 Sep 12.
- Hochberg Y, Tamhane A (1987) Multiple Comparison Procedures. John Wiley & Sons. p.
- Abdi H (2007) The Bonferonni and Šidák Corrections for Multiple Comparisons. In: Salkind NJ, editor. Encyclopedia of Measurement and Statistics. Thousand Oaks, Vol. 1. 103–107.
- Hesterberg T, Monaghan S, Moore D, Clipson A, Epstein R (2003) Bootstrap Methods and Permutation Tests. The Practice of Business Statistics. W. H. Freeman, Vol. 47. 1–70.
- Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. Nature Genetics 37: 413–417. Available: http://www.ncbi.nlm.nih.gov/pubmed/15793588. Accessed 2011 May 4.
- Li M, Wang K, Grant SFA, Hakonarson H, Li C (2009) ATOM: a powerful gene-based association test by combining optimally weighted markers. Bioinformatics 25: 497–503. Available: http://www.ncbi.nlm.nih.gov/pubmed/ 19074959. Accessed 2011 May 4.
- Anunciação O, Gomes B, Vinga S, Gaspar J, Oliveira A, et al. (2010) A Data Mining Approach for the Detection of High-Risk Breast Cancer Groups. In:

Rocha M, Riverola F, Shatkay H, Corchado J, editors. Advances in Bioinformatics. Springer Berlin Heidelberg, Vol. 74. 43–51. Available: http:// www.springerlink.com/content/155k63p457721384/. Accessed 2012 Sep 12.

- Weisstein EW (2012) Fisher's Exact Test. From MathWorld-A Wolfram Web Resource. Available: http://mathworld.wolfram.com/FishersExactTest.html. Accessed 2012 Sep 12.
- 47. Francisco AP, Bugalho M, Ramirez M, Carriço JA (2009) Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics 10: 152. Available: http://www.pubmedcentral.nih.gov/ articlerender.fcgi?artid = 2705362&tool = pmcentrez&rendertype = abstract. Accessed 2012 Sep 12.
- Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, et al. (2012) PHYLOViZ: Phylogenetic Inference and Data Visualization for Sequence Based Typing Methods. BMC Bioinformatics 13: 87. Available: http://www. biomedcentral.com/1471-2105/13/87. Accessed 2011 May 9.
- Tarca AL, Carey VJ, Chen X, Romero R, Dräghici S (2007) Machine Learning and Its Applications to Biology. PLoS Computational Biology 3: e116. Available: http://www.ncbi.nlm.nih.gov/pubmed/17604446. Accessed 2012 Sep 12.
- Kraskov A, St H, Andrzejak RG, Grassberger P (2008) Hierarchical Clustering Based on Mutual Information. Europhysics Letters 70. Available: http:// iopscience.iop.org/0295-5075/70/2/278. Accessed 2012 Mar 12.
- CLSI (2010) Performance Standards for Antimicrobial Susceptibility Testing?; Twentieth Informational Supplement. CLSI document M100–S20.
- 52. Magiorakos A-P, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, et al. (2012) Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clinical Microbiology and Infection 18: 268–281. Available: http:// dx.doi.org/10.1111/j.1469-0691.2011.03570.x. Accessed 2012 Sep 17.
- Russell AD (2004) Whither triclosan? The Journal of Antimicrobial Chemotherapy 53: 693–695. Available: http://www.ncbi.nlm.nih.gov/pubmed/ 15073159. Accessed 14 August 2010.
- 54. Bailey AM, Constantinidou C, Ivens A, Garvey MI, Webber M a, et al. (2009) Exposure of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium to triclosan induces a species-specific response, including drug detoxification. The Journal of Antimicrobial Chemotherapy 64: 973–985. Available: http://www. ncbi.nlm.nih.gov/pubmed/19759044. Accessed 2012 Mar 5.
- Birosová L, Mikulásová M (2009) Development of triclosan and antibiotic resistance in Salmonella enterica serovar Typhimurium. Journal of Medical Microbiology 58: 436–441. Available: http://www.ncbi.nlm.nih.gov/pubmed/ 19273638. Accessed 2012 May 30.
- 56. Chuanchuen R, Beinlich K, Hoang TT, Becher A, Karkhoff-Schweizer RR, et al. (2001) Cross-resistance between triclosan and antibiotics in *Pseudomonas aeruginosa* is mediated by multidrug efflux pumps: exposure of a susceptible mutant strain to triclosan selects *nfxB* mutants overexpressing MexCD-OprJ. Antimicrobial Agents and Chemotherapy 45: 428–432. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11158736. Acccessed 2012 Sep 12.
- 57. Chuanchuen R, Narasaki CT, Schweizer HP (2002) The MexJK efflux pump of *Pseudomonas aeruginosa* requires OprM for antibiotic efflux but not for efflux of triclosan. Journal of Bacteriology 184: 5036–5044. Available: http://www. p u b m e d c e n t r a l . n i h . g o v / a r t i c l e r e n d e r . fcgi?artid = 135324&tool = pmcentrez&rendertype = abstract. Accessed 2012 Sep 12.
- 58. Hernández A, Ruiz FM, Romero A, Martínez JL (2011) The Binding of Triclosan to SmeT, the Repressor of the Multidrug Efflux Pump SmeDEF, Induces Antibiotic Resistance in *Stenotrophomonas mallophilia*. PLoS Pathogens 7: 12. Available: http://www.pubmedcentral.nih.gov/articlerender. fcgi?artid = 3128119&tool = pmcentrez&rendertype = abstract. Accessed 2012 Sep 12.
- Sanchez P, Moreno E, Martinez JL (2005) The biocide triclosan selects Stenotrophomonas maltophilia mutants that overproduce the SmeDEF multidrug efflux pump. Antimicrobial Agents and Chemotherapy 49: 781–782. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC547238/. Accessed 2012 Sep 12.
- Katzung BG, Masters SB, Trevor AJ (2009) Basic and Clinical Pharmacology. 11th ed. McGraw-Hill Medical (Lange Basic Science). p.