

## RESEARCH ARTICLE

# Interactions between staphylococcal enterotoxins A and D and superantigen-like proteins 1 and 5 for predicting methicillin and multidrug resistance profiles among *Staphylococcus aureus* ocular isolates

Min Lu<sup>1</sup>, Jean-Marie Parel<sup>2</sup>, Darlene Miller<sup>3</sup>\*✉

**1** Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL, United States of America, **2** Department of Ophthalmology, Ophthalmic Biophysics Center, Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, United States of America, **3** Department of Ophthalmology, Ocular Microbiology Laboratory, Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, United States of America

✉ Current address: McKnight Vision Research Ctr, Miami, FL, United States of America

\* [dmiller@med.miami.edu](mailto:dmiller@med.miami.edu)



## OPEN ACCESS

**Citation:** Lu M, Parel J-M, Miller D (2021) Interactions between staphylococcal enterotoxins A and D and superantigen-like proteins 1 and 5 for predicting methicillin and multidrug resistance profiles among *Staphylococcus aureus* ocular isolates. PLoS ONE 16(7): e0254519. <https://doi.org/10.1371/journal.pone.0254519>

**Editor:** Abdelazeem Mohamed Algammal, Suez Canal University, EGYPT

**Received:** March 31, 2021

**Accepted:** June 29, 2021

**Published:** July 28, 2021

**Copyright:** © 2021 Lu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data are publicly available on GitHub: [https://github.com/luminwin/mecA\\_SESSL](https://github.com/luminwin/mecA_SESSL).

**Funding:** Funding in part by a NIH Center Core Grant P30EY014801 (institutional) and a Research to Prevent Blindness Unrestricted Grant (institutional).

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

### Background

Methicillin-resistant *Staphylococcus aureus* (MRSA) and multidrug-resistant (MDR) *S. aureus* strains are well recognized as posing substantial problems in treating ocular infections. *S. aureus* has a vast array of virulence factors, including superantigens and enterotoxins. Their interactions and ability to signal antibiotics resistance have not been explored.

### Objectives

To predict the relationship between superantigens and methicillin and multidrug resistance among *S. aureus* ocular isolates.

### Methods

We used a DNA microarray to characterize the enterotoxin and superantigen gene profiles of 98 *S. aureus* isolates collected from common ocular sources. The outcomes contained phenotypic and genotypic expressions of MRSA. We also included the MDR status as an outcome, categorized as resistance to three or more drugs, including oxacillin, penicillin, erythromycin, clindamycin, moxifloxacin, tetracycline, trimethoprim-sulfamethoxazole and gentamicin. We identified gene profiles that predicted each outcome through a classification analysis utilizing Random Forest machine learning techniques.

### Findings

Our machine learning models predicted the outcomes accurately utilizing 67 enterotoxin and superantigen genes. Strong correlates predicting the genotypic expression of MRSA

were enterotoxins A, D, J and R and superantigen-like proteins 1, 3, 7 and 10. Among these virulence factors, enterotoxin D and superantigen-like proteins 1, 5 and 10 were also significantly informative for predicting both MDR and MRSA in terms of phenotypic expression. Strong interactions were identified including enterotoxins A (*entA*) interacting with superantigen-like protein 1 (*set6-var1\_11*), and enterotoxin D (*entD*) interacting with superantigen-like protein 5 (*ssl05/set3\_probe 1*): MRSA and MDR *S. aureus* are associated with the presence of both *entA* and *set6-var1\_11*, or both *entD* and *ssl05/set3\_probe 1*, while the absence of these genes in pairs indicates non-multidrug-resistant and methicillin-susceptible *S. aureus*.

## Conclusions

MRSA and MDR *S. aureus* show a different spectrum of ocular pathology than their non-resistant counterparts. When assessing the role of enterotoxins in predicting antibiotics resistance, it is critical to consider both main effects and interactions.

## Introduction

Multidrug resistance has been increased globally that is considered a public health threat. Several previous studies revealed the emergence of multidrug-resistant (MDR) bacterial pathogens from different origins, especially birds, animals, and food chains which may be transmitted to human consumers resulting in severe illness [1–8]. As a ubiquitous Gram-positive bacterium, *Staphylococcus aureus* is a leading cause of ocular morbidity and blindness worldwide [9]. Infections include blepharitis (lids), conjunctivitis (conjunctiva), keratitis (cornea), endophthalmitis (intraocular fluids), and dacryocystitis (lacrimal system). Methicillin-resistant *S. aureus* (MRSA) is well recognized as getting impassable toward numerous antibacterial specialists from the efficacy of infection treatments in the crisis center [10–13]. According to the infectious keratitis survey from the American Society of Cataract and Refractive Surgery in 2008, MRSA is the most common pathogen causing infections after laser-assisted in situ keratomileusis [14]. Hence, detecting strong correlates and interactions for predicting methicillin and multidrug resistant status could help us understand the prevalence of *S. aureus*, give essential insights into ocular pathology, and provide information for drug development against *S. aureus* ocular infection.

Recognizing the virulence factors of methicillin and multidrug resistant *S. aureus* are essential in developing preventive measures for ocular infectious diseases. The gene encoding the penicillin-binding protein 2a or 2' (PBP2a or PBP2') (*mecA*) was found to be integrated into the chromosomal element (SCC*mec*) of MRSA [15]. *S. aureus* isolates with the *mecA* gene are more likely to be MDR and difficult to treat. Researchers have studied the association between MRSA and cytolytins ( $\alpha$ -toxin,  $\gamma$ -toxin, Panton-Valentine leukocidin (PVL) [16–19]. The associations between *mecA* and staphylococcal superantigens and enterotoxins are still ambiguous. MRSA could be associated with superantigens such as staphylococcal enterotoxin [SE] B or C [20], and staphylococcal enterotoxin-like [SE-I] Q [19]. Enterotoxin gene cluster types *egc1*, *egc2* and *egc3* could also play a role in this association since they are correlated with enterotoxin and enterotoxin-like gene profiles in human nasal carriage and animal isolates of *Staphylococcus aureus* [21]. However, few studies focused on these virulence factors for *Staphylococcus aureus* ocular infection. Moreover, statistically the role of each virulence factor was

often assessed in a univariate fashion using separate Chi-square test or Fisher's exact test for each factor [22–25], without adjustment of other factors or consideration of interactions between these virulence factors. Given the high correlations between these virulence factors, it is critical to include as many staphylococcal superantigens and enterotoxins as possible in a unified model to examine the most significant ones.

One statistical challenge for including all the virulence factors in a classification model is the curse of dimensionality. For example, staphylococcal enterotoxins are members of a protein family of more than 20 different staphylococcal exotoxins sharing several biological activities and structural features [9, 26, 27]. We want to examine both the main and the interaction effects for many virulence genes. Moreover, we wonder if we can predict MRSA accurately in a classification model after adding as many staphylococcal superantigen and enterotoxin genes as possible as predictors since high prediction accuracy indicates systematic differences in MRSA and methicillin-susceptible *S. aureus* (MSSA) ocular pathologies. We also wonder if the predictors are also informative for distinguishing MDR status. Clinically this is important because it will allow us to determine the superantigen profile of persistent strains. In addition, high accuracy of the prediction models would potentially enable the routine application of antimicrobial susceptibility testing to prevent the emergence of antibiotic-resistant strains of potential public health concern, and informative virulence genes might be chosen for tracing the sources of infection in ocular infection. For this purpose, we adopted an ensemble machine learning classifier, Random Forest [28], to predict MRSA and MDR *S. aureus*, which over the last twenty years has become widely used for predictions in cardiovascular disease, cancer, HIV, and other non-medical endeavors [29–33]. The objective of this study is to identify virulence factors of staphylococcal superantigens and enterotoxins that are most important for predicting the risk of methicillin and multidrug resistance, and to demonstrate how these informative virulence factors differentiate the predicted probability or risk score of MRSA and MDR *S. aureus* ocular infection.

## Methods

This was a retrospective study of the molecular characteristics and interactions of *S. aureus* isolates recovered from ocular samples collected between January 2014–December 2019. Institutional review board approval was obtained from the University of Miami Miller School of Medicine Sciences Subcommittee for the Protection of Human Subjects and the research followed the tenets of the Declaration of Helsinki (IRB Protocol Study ID #20070960). All patients were informed and consented in writing. The Ocular Microbiology Department's Isolate database was searched to identify, non-consecutive, de-identified (no patient information) *S. aureus* isolates recovered from patients presenting with culture proven *S. aureus* ocular infections. No patient data was included or available for the study.

We used a DNA microarray to characterize the enterotoxin and superantigen gene profiles [34] of 98 *S. aureus* isolates collected from common ocular sources. Seventy-six of these were from the cornea. There is a combination of 64 MRSA and 34 MSSA, and 69 MDR and 29 non-multidrug resistant (NMDR) isolates in terms of phenotypic characterization. MRSA status was determined through the E tests, cefoxitin screen and automated system. MDR status was defined as resistance to three or more drugs, including oxacillin, penicillin, erythromycin, clindamycin, moxifloxacin, tetracycline, trimethoprim-sulfamethoxazole (SXT), and gentamicin. The genotypic characterization of MRSA produced 56 *mecA*-positive and 42 *mecA*-negative isolates. The *mecA* gene was not detected in eight isolates that were categorized as MRSA in phenotypic assays. The difference is due to varied test methods, and microarray was done after the phenotypic assays—hence freezing may impact the stability of *mecA*. We include 21

enterotoxin genes, 45 staphylococcal superantigens-like genes, and one variable indicating isolate sources to predict MRSA and MDR statuses. In total, we have 67 predictors.

### Statistical analysis

Random Forest [28] model was applied for the prediction of outcome, a modern machine learning technique that permits exploration of nonlinear, complex interrelationships. It has been utilized to explore a large number of predictors and identify replicable sets of risk factors [35–38]. The Random Forest technology is related to recursive partitioning and classification tree analyses, wherein the variables that are most related to an outcome of interest are first optionally split to improve prediction, followed by more and more splits to create a tree. A single tree is inherently unstable with large prediction variance. To overcome this instability, a forest of trees is “grown” from bootstrap samples of the original dataset (sampling with replacement until a dataset of equal size is generated; an average of 37% of the data will not be sampled, which is referred as out-of-bag data), permitting an ensemble average to be calculated across the individual trees [28]. Because this method is completely nonparametric without any restrictive underlying model assumptions, complex relationship and interactions among variables can be robustly accounted for.

**Classification analyses.** Classification analysis was applied to two statuses, MRSA versus MSSA, and NMDR versus MDR, using the Random Forest classification model. Estimated variable importance (VIMP) of each predictor [28, 39] is adopted, which utilizes a prediction-based approach by estimating prediction error attributable to the predictor, evaluated as the cross-validated misclassification error estimated via the out-of-bag data. The VIMP can be interpreted as the increase in the misclassification error when the corresponding predictor is randomly permuted into a noise variable. Positive VIMP values identify variables that are predictive after adjusting for all the other variables. For example, a VIMP of 4.29% indicates that a variable improves by 4.29% the ability of the model to classify the status of the outcome. Negative VIMP values indicate “noisy” variables that degrade model performance. Standard errors and *P* values are generated by a delete-*d*-jackknife procedure [39]. Strong correlates were based on an  $\alpha = 0.05$  level confidence for VIMP. Variables with positive VIMP estimates whose *P* values less than 0.05 are selected as informative ones to predict the outcome.

Odds ratio is used to inform the direction of the associations between the outcome and the selected informative predictors, calculated for the negative and positive categories. The ambiguous category is not considered. Odds ratio  $> 1$  means that positive isolate for the corresponding virulence gene is more likely to be MRSA or MDR, while Odds ratio  $< 1$  indicates that positive isolate for the corresponding virulence gene is associated with MSSA or NMDR status. *P* values are calculated using Fisher’s exact test.

Random Forest model and odds ratio calculation were implemented in the open-source R software using the `randomForestSRC` [40] and `epitools` [41] R packages correspondingly. From the `randomForestSRC` R package, the function `rfsrc` was used for building the prediction model under default setting with 1000 trees; then inferences of VIMP were estimated using the function `subsample`. The default settings for this function were adopted with 1000 subsamples, except the ratio for subsampling approach which was increased from 10% (the default setting is the square root of the sample size) to 90% because of small sample size. The function `oddsratio.fisher` from `epitools` package was used to calculate odds ratios, confidence intervals and *P* values.

**Partial plot.** For the Random Forest classification analysis, partial dependence plots were used to visualize the variables’ impact on the predicted probabilities of the statuses of the outcome through mapping the marginal effect of the selected variable to uncovers the relationship

[42–44], where predicted probability is defined as out-of-bag predicted probability adjusted by integrating out all variables other than the targeted variable of interest. The integration is approximated utilizing the data by averaging over variables, implemented through setting the partial parameter in the `plot.variable` function from the `randomForestSRC` R package.

## Results

### Summary of methicillin and multidrug resistant isolates

We summarized all the predictors stratified by MRSA and MDR statuses in [Table 1](#). Pearson's Chi-squared test is applied and implemented in the open-source R software for testing for independence. The percentages of positive isolates are listed in parentheses in [Table 1](#). Most virulence genes display different percentages of positives between different MRSA and MDR statuses. For MRSA, such a difference is more likely to be significant for staphylococcal superantigen/enterotoxin-like genes than enterotoxin genes. For MDR *S. aureus*, enterotoxins C, D, and L levels are significantly different between NMDR and MDR isolates. As to isolate sources, fewer MRSA isolates were collected from the cornea (66.07% as 37 out of 56) than MSSA (92.86% as 39 out of 42,  $p = 0.021$ ) in the microarray test of *mecA*. Such a difference was not significant for the phenotypic expression of MDR or MRSA.

Two enterotoxins, *entA* (320E) and *entE*, contain no positive isolates, and therefore they are unable to predict any outcome. There are virulence genes with identical detection results, for example, enterotoxins M, N, and O. In order to reduce the dimension and collinearity of the classification model, when the values of a group of variables are the same, we only add one of them in the classification model whose result will represent the entire group. Variables with duplicated results are listed in [S2 Table](#). There are 46 variables in each prediction model.

The Random Forest classification model predicts the outcomes accurately: the out-of-bag misclassification error is 23.47% for predicting the genotypic expression of MRSA, 17.35% for the phenotypic expression of MRSA and 14.29% for MDR *S. aureus*. The model predicting MDR *S. aureus* has the lowest misclassification error, indicating that superantigens and enterotoxins are more associated with MDR *S. aureus* than MRSA. As to the phenotypic expressions, we found that MRSA isolates are highly likely to be MDR with an infinite odds ratio (see [S3 Table](#) for the contingency table of the three outcomes). The genotypic expression of MRSA is also significantly associated with MDR status (odds ratio = 18.4,  $p < .001$ ).

### Relationship of outcomes to important variables

**Genotypic expression of MRSA.** Variables of importance for predicting each outcome were identified by machine learning using Random Forest classification model. For the genotypic expression of MRSA, 16 variables are significant from its classification analysis, shown in [Table 2](#). Strong correlates predicting MRSA include enterotoxin A, P, D, J, and R and staphylococcal superantigen-like proteins 1, 3, 5, 7, and 10. The full list of VIMP and odds ratios for the 67 predictors can be found in [S1 Table](#). Isolate sources are not informative (see the bottom of [S1 Table](#)), which means that after adjusting for differences in virulence genes, isolate sources are balanced between MSSA and MRSA isolates. Univariately, none of the informative enterotoxins are significantly associated with MRSA. However, after adjusting for other variables, enterotoxin P positive indicates a higher probability of MRSA, as shown in the partial plot of *entA* (N315) / *entP* in subfigure A of [S1 Fig](#). All the selected staphylococcal superantigen-like proteins are significantly associated with MRSA from Fisher's exact test (see [Table 2](#)), except *ssl05/set3* (RF122, probe-611) and *ssl07/set1* (MRSA252). After adjusting for other variables, *ssl05/set3* (RF122, probe-611)-positive is associated with MRSA (subfigure B of [S1 Fig](#)) and

**Table 1. Distribution of virulence factors among methicillin and multidrug resistance profiles.**

Virulence factor	Virulence gene	Negatives/Positives (percentage of positives)						Sig†
		Genotype		Phenotype		Phenotype		
		MSSA (n = 42)	MRSA (n = 56)	MSSA (n = 34)	MRSA (n = 64)	NMDR (n = 29)	MDR (n = 69)	
Enterotoxin A	<i>entA</i>	40/2 (5)	48/8 (14)	32/2 (6)	56/8 (12)	28/1 (3)	60/9 (13)	
Enterotoxin A, 320E	<i>entA (320E)</i>	42 (0)	56 (0)	34 (0)	64 (0)	29 (0)	69 (0)	
Enterotoxin A, N315	<i>entA (N315) / entP</i>	38/4 (10)	52/4 (7)	30/4 (12)	60/4 (6)	26/3 (10)	64/5 (7)	
Enterotoxin B	<i>entB</i>	37/5 (12)	49/7 (12)	29/5 (15)	57/7 (11)	26/3 (10)	60/9 (13)	
Enterotoxin C	<i>entC</i>	39/3 (7)	55/1 (2)	31/3 (9)	63/1 (2)	25/4 (14)	69/0 (0)	‡‡
Enterotoxin D	<i>entD</i>	40/2 (5)	46/10 (18)	33/1 (3)	53/11 (17)	29/0 (0)	57/12 (17)	‡
Enterotoxin E	<i>entE</i>	42 (0)	56 (0)	34 (0)	64 (0)	29 (0)	69 (0)	
Enterotoxin G	<i>entG</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
Enterotoxin H	<i>entH</i>	40/2 (5)	56/0 (0)	32/2 (6)	64/0 (0)	28/1 (3)	68/1 (1)	
Enterotoxin I	<i>entI</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
Enterotoxin J	<i>entJ</i>	37/5 (12)	46/10 (18)	30/4 (12)	53/11 (17)	26/3 (10)	57/12 (17)	
Enterotoxin K	<i>entK</i>	36/6 (14)	40/16 (29)	28/6 (18)	48/16 (25)	26/3 (10)	50/19 (28)	
Enterotoxin L	<i>entL</i>	39/3 (7)	55/1 (2)	31/3 (9)	63/1 (2)	25/4 (14)	69/0 (0)	‡‡
Enterotoxin M	<i>entM</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
Enterotoxin N	<i>entN (cons)</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
	<i>entN (other than RF122)</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
Enterotoxin O	<i>entO</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
<i>egc</i> cluster	<i>egc (total)</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
Enterotoxin Q	<i>entQ</i>	36/6 (14)	40/16 (29)	28/6 (18)	48/16 (25)	26/3 (10)	50/19 (28)	
Enterotoxin R	<i>entR</i>	38/4 (10)	46/10 (18)	30/4 (12)	54/10 (16)	26/3 (10)	58/11 (16)	
Enterotoxin U and/or Y	<i>entU</i>	19/23 (55)	21/35 (62)	16/18 (53)	24/40 (62)	12/17 (59)	28/41 (59)	
SSL protein 1	<i>set6-var1_11</i>	16/26 (62)	0/56 (100)	16/18 (53)	0/64 (100)	15/14 (48)	1/68 (99)	***++ +‡‡‡
	<i>set6-var2_11</i>	27/14 (33)	54/0 (0)	19/14 (41)	62/0 (0)	15/13 (45)	66/1 (1)	***++ +‡‡‡
	<i>set6-var1_12</i>	12/24 (57)	3/48 (86)	9/19 (56)	6/53 (83)	8/14 (48)	7/58 (84)	**+‡‡
	<i>set6-var2_12</i>	37/4 (10)	55/0 (0)	29/4 (12)	63/0 (0)	25/3 (10)	67/1 (1)	+
	<i>set6-var4_11</i>	6/36 (86)	1/55 (98)	6/28 (82)	1/63 (98)	6/23 (79)	1/68 (99)	*+‡‡
	<i>ssl01-RF122</i>	39/3 (7)	56/0 (0)	31/3 (9)	64/0 (0)	26/3 (10)	69/0 (0)	‡
	<i>ssl01/set6 (COL)</i>	32/10 (24)	35/21 (38)	27/7 (21)	40/24 (38)	25/4 (14)	42/27 (39)	‡
	<i>ssl01/set6 (Mu50+N315)</i>	17/15 (36)	1/34 (61)	17/10 (29)	1/39 (61)	17/8 (28)	1/41 (59)	***++ +‡‡‡
	<i>ssl01/set6 (MW2+MSSA476)</i>	38/4 (10)	56/0 (0)	30/4 (12)	64/0 (0)	26/3 (10)	68/1 (1)	+
	<i>ssl01/set6 (MRSA252)</i>	38/4 (10)	56/0 (0)	30/4 (12)	64/0 (0)	25/4 (14)	69/0 (0)	+‡‡
	<i>ssl01/set6 (RF122)</i>	41/1 (2)	56/0 (0)	33/1 (3)	64/0 (0)	28/1 (3)	69/0 (0)	
<i>ssl01/set6 (other alleles)</i>	34/8 (19)	55/1 (2)	26/8 (24)	63/1 (2)	20/9 (31)	69/0 (0)	*+‡‡‡	
SSL protein 2	<i>ssl02/set7</i>	5/34 (81)	0/56 (100)	5/26 (76)	0/64 (100)	5/21 (72)	0/69 (100)	***+‡‡‡‡
	<i>ssl02/set7 (MRSA252)</i>	11/7 (17)	15/0 (0)	7/7 (21)	19/0 (0)	7/7 (24)	19/0 (0)	**+‡‡‡‡
SSL protein 3	<i>ssl03/set8_probe 1</i>	7/35 (83)	0/56 (100)	7/27 (79)	0/64 (100)	7/22 (76)	0/69 (100)	**+‡‡‡‡
	<i>ssl03/set8_probe 2</i>	7/35 (83)	0/56 (100)	7/27 (79)	0/64 (100)	7/22 (76)	0/69 (100)	**+‡‡‡‡
	<i>ssl03/set8 (MRSA252, SAR0424)</i>	40/2 (5)	56/0 (0)	32/2 (6)	64/0 (0)	27/2 (7)	69/0 (0)	
SSL protein 4	<i>ssl04/set9</i>	8/34 (81)	0/56 (100)	8/26 (76)	0/64 (100)	8/21 (72)	0/69 (100)	**+‡‡‡‡
	<i>ssl04/set9 (MRSA252, SAR0425)</i>	34/7 (17)	53/0 (0)	26/7 (21)	61/0 (0)	21/7 (24)	66/0 (0)	**+‡‡‡‡

(Continued)

Table 1. (Continued)

Virulence factor	Virulence gene	Negatives/Positives (percentage of positives)						Sig†
		Genotype		Phenotype		Phenotype		
		MSSA (n = 42)	MRSA (n = 56)	MSSA (n = 34)	MRSA (n = 64)	NMDR (n = 29)	MDR (n = 69)	
SSL protein 5	<i>ssl05/set3_probe 1</i>	9/33 (79)	0/56 (100)	9/25 (74)	0/64 (100)	9/20 (69)	0/69 (100)	**++++‡‡‡
	<i>ssl05/set3 (RF122, probe-611)</i>	9/8 (19)	3/0 (0)	7/8 (24)	5/0 (0)	7/8 (28)	5/0 (0)	***++ +‡‡‡
	<i>ssl05/set3_probe 2 (612)</i>	7/27 (64)	0/56 (100)	7/19 (56)	0/64 (100)	7/14 (48)	0/69 (100)	***++ +‡‡‡
	<i>ssl05/set3 (MRSA252)</i>	35/7 (17)	56/0 (0)	27/7 (21)	64/0 (0)	22/7 (24)	69/0 (0)	**++++‡‡‡
SSL protein 6	<i>ssl06/set21</i>	24/18 (43)	36/20 (36)	19/15 (44)	41/23 (36)	18/11 (38)	42/27 (39)	
	<i>ssl06 (NCTC8325+MW2)</i>	17/20 (48)	30/21 (38)	12/17 (50)	35/24 (38)	11/13 (45)	36/28 (41)	
SSL protein 7	<i>ssl07/set1</i>	6/35 (83)	0/56 (100)	6/27 (79)	0/64 (100)	6/22 (76)	0/69 (100)	**++++‡‡‡
	<i>ssl07/set1 (MRSA252)</i>	5/2 (5)	0/0 (0)	5/2 (6)	0/0 (0)	5/2 (7)	0/0 (0)	**++++‡‡‡
	<i>ssl07/set1 (AF188836)</i>	24/5 (12)	41/0 (0)	17/5 (15)	48/0 (0)	14/5 (17)	51/0 (0)	*++‡‡‡
SSL protein 8	<i>ssl08/set12_probe 1</i>	7/35 (83)	0/56 (100)	7/27 (79)	0/64 (100)	7/22 (76)	0/69 (100)	**++++‡‡‡
	<i>ssl08/set12_probe 2</i>	7/35 (83)	0/56 (100)	7/27 (79)	0/64 (100)	7/22 (76)	0/69 (100)	**++++‡‡‡
SSL protein 9	<i>ssl09/set5_probe 1</i>	7/35 (83)	3/53 (95)	7/27 (79)	3/61 (95)	7/22 (76)	3/66 (96)	+‡‡
	<i>ssl09/set5_probe 2</i>	7/35 (83)	3/53 (95)	7/27 (79)	3/61 (95)	7/22 (76)	3/66 (96)	+‡‡
	<i>ssl09/set5 (MRSA252)</i>	35/7 (17)	56/0 (0)	27/7 (21)	64/0 (0)	22/7 (24)	69/0 (0)	**++++‡‡‡
SSL protein 10	<i>ssl10/set4</i>	11/31 (74)	0/56 (100)	11/23 (68)	0/64 (100)	11/18 (62)	0/69 (100)	***++ +‡‡‡
	<i>ssl10 (RF122)</i>	33/4 (10)	49/0 (0)	25/4 (12)	57/0 (0)	19/4 (14)	63/0 (0)	+‡‡
	<i>ssl10/set4 (MRSA252)</i>	3/6 (14)	1/0 (0)	1/6 (18)	3/0 (0)	1/6 (21)	3/0 (0)	**++++‡‡‡
SSL protein 11	<i>ssl11/set2 (COL)</i>	32/10 (24)	35/21 (38)	27/7 (21)	40/24 (38)	25/4 (14)	42/27 (39)	‡
	<i>ssl11+set2(Mu50+N315)</i>	28/14 (33)	22/34 (61)	25/9 (26)	25/39 (61)	22/7 (24)	28/41 (59)	*++‡‡
	<i>ssl11+set2(MW2+MSSA476)</i>	39/3 (7)	56/0 (0)	31/3 (9)	64/0 (0)	27/2 (7)	68/1 (1)	
	<i>ssl11/set2 (MRSA252)</i>	36/6 (14)	56/0 (0)	28/6 (18)	64/0 (0)	23/6 (21)	69/0 (0)	*++‡‡‡
	<i>setB3</i>	7/35 (83)	0/56 (100)	7/27 (79)	0/64 (100)	7/22 (76)	0/69 (100)	**++++‡‡‡
SEL protein, second locus	<i>setB3 (MRSA252)</i>	35/7 (17)	56/0 (0)	27/7 (21)	64/0 (0)	22/7 (24)	69/0 (0)	**++++‡‡‡
	<i>setB2</i>	7/35 (83)	0/56 (100)	7/27 (79)	0/64 (100)	7/22 (76)	0/69 (100)	**++++‡‡‡
	<i>setB2 (MRSA252)</i>	35/7 (17)	56/0 (0)	27/7 (21)	64/0 (0)	22/7 (24)	69/0 (0)	**++++‡‡‡
	<i>setB1</i>	2/38 (90)	0/56 (100)	2/30 (88)	0/64 (100)	2/25 (86)	0/69 (100)	+‡‡
	Sources	<i>sources</i>	2/1/39/	16/0/36/	2/1/31/	16/0/44/	2/1/26/	16/0/49/
		0/0/0 (93)	1/1/2 (64)	0/0/0 (91)	1/1/2 (69)	0/0/0 (90)	1/1/2 (71)	

SSL indicates staphylococcal superantigen-like; SEL indicates Staphylococcal exotoxin-like; MSSA is short for methicillin-susceptible *S. aureus*, MRSA for methicillin-resistant *S. aureus*, NMDR for non-multidrug-resistant *S. aureus*, MDR for multidrug-resistant *S. aureus*. For isolate sources listed in the bottom, numbers of isolates are listed in the order of conjunctiva, conjunctiva/lids, cornea (percentage is listed in the parenthesis), lids, socket, and suture.

† The association between the outcome and the virulence gene in category negative, positive, and ambiguous is tested through Pearson’s Chi-squared test. Sig indicates significant level according to *P* values: when the outcome is genotypic expression of MRSA, \* for  $p \leq 0.05$ , \*\* for  $p \leq 0.01$ , \*\*\* for  $p \leq 0.001$ ; when the outcome is phenotypic expression of MRSA, + for  $p \leq 0.05$ , ++ for  $p \leq 0.01$ , and +++ for  $p \leq 0.001$ ; when the outcome is MDR, ‡ for  $p \leq 0.05$ , ‡‡ for  $p \leq 0.01$ , and ‡‡‡ for  $p \leq 0.001$ .

\* indicates the genotypic expression between MSSA and MRSA; + for the phenotypic expression between MSSA and MRSA; ‡ for the phenotypic expression between NMDR and MDR.

<https://doi.org/10.1371/journal.pone.0254519.t001>

**Table 2. Informative virulence genes for predicting methicillin and multidrug resistance profiles.**

Virulence factor	Virulence gene	Fisher's exact test		Random Forest				
		Genotype		Genotype			Phenotype	
		MRSA		MRSA		MRSA†	MDR†	
		OR	95% CI	Est	SE	Sig	Sig	Sig
Enterotoxin A	<i>entA</i>	3.3	0.61–33.61	0.89	0.49	*		
Enterotoxin A, allele from strain N315 = enterotoxin P	<i>entA (N315) / entP</i>	0.73	0.13–4.2	1.09	0.52	*	+	
Enterotoxin D	<i>entD</i>	4.29	0.84–42.53	3.1	0.98	***	+++	‡‡‡
Enterotoxin J	<i>entJ</i>	1.6	0.45–6.51	0.77	0.32	**		‡
Enterotoxin R	<i>entR</i>	2.05	0.54–9.68	1.78	0.63	**		‡
Staphylococcal superantigen-like protein 1	<i>set6-var1_11</i>	Inf	7.39–Inf	1.8	0.41	***	+++	‡‡
	<i>set6-var2_11</i>	0	0–0.17	1.07	0.32	***	+++	
	<i>ssl01/set6 (MW2+MSSA476)</i>	0	0–1.1	0.14	0.06	**	++	
Staphylococcal superantigen-like protein 3	<i>ssl03/set8_probe 1</i>	Inf	2.12–Inf	0.07	0.04	*	+	
Staphylococcal superantigen-like protein 5	<i>ssl05/set3_probe 1</i>	Inf	3.06–Inf	0.35	0.16	*	++	‡‡‡
	<i>ssl05/set3 (RF122, probe-611)</i>	0	0–3.54	1.71	0.46	***	++	‡‡
	<i>ssl05/set3_probe 2 (612)</i>	Inf	2.71–Inf	2.64	0.55	***	+++	‡‡‡
	<i>ssl05/set3 (MRSA252)</i>	0	0–0.47	0.08	0.05	*	+	
Staphylococcal superantigen-like protein 7	<i>ssl07/set1 (MRSA252)</i>	0	0–Inf	0.18	0.1	*		
	<i>ssl07/set1 (AF188836)</i>	0	0–0.71	2.13	0.84	**		
Staphylococcal superantigen-like protein 10	<i>ssl10/set4</i>	Inf	4.11–Inf	0.41	0.15	**	++	‡‡

OR indicates odds ratio, which is used to indicate the direction of the correlation: if the value is larger than one, the virulence gene is more likely to be detected (as more positive isolates) in MRSA isolates; if the value is smaller than one, the virulence gene is less likely to be detected (as more negative isolates) in MRSA isolates. Inf indicates infinity value since some entry of the contingency table is zero.

Est indicates estimated VIMP from Random Forest model in terms of contribution to classification accuracy in percentage; SE indicates standard error; Sig indicates significant level according to *P* values: \* for  $p \leq 0.05$ ; \*\* for  $p \leq 0.01$ ; \*\*\* for  $p \leq 0.001$ .

† Two Random Forest models were fitted to classify phenotypic expressions of MRSA and MDR statuses, and the significant level of estimated VIMP was used to mark the informative virulence genes: when the outcome is MRSA, + for  $p \leq 0.05$ , ++ for  $p \leq 0.01$ , and +++ for  $p \leq 0.001$  (see S4 Table for full list of informative predictors); when the outcome is MDR *S. aureus*, ‡ for  $p \leq 0.05$ , ‡‡ for  $p \leq 0.01$ , and ‡‡‡ for  $p \leq 0.001$  (see S5 Table for full list of informative predictors).

<https://doi.org/10.1371/journal.pone.0254519.t002>

*ssl07/set1 (MRSA252)* negative is associated with MRSA (subfigure C of S1 Fig). As to enterotoxins, the most informative virulence gene is *entD*, contributing 3.1% prediction accuracy (standard error (SE) = 0.98,  $p < .001$ ). In other words, without taking into consideration *entD*, the prediction error or misclassification rate of the Random Forest model will increase from 23.47% to 26.57%. As to staphylococcal superantigen-like proteins, *ssl05/set3\_probe 2 (612)* is the most informative (VIMP = 2.64, SE = 0.55,  $p < .001$ ).

**Phenotypic expression of MRSA.** As mentioned before, eight isolates were categorized as MRSA in phenotypic assays but classified as *mecA*-negative in the microarray test. These eight isolates cause differences in the classification results between the genotypic and phenotypic expressions of MRSA. However, the informative virulence genes with large estimates of VIMP are still very similar. For example, *entD* and *ssl05/set3\_probe 2 (612)* are the most informative genes for the genotypic expression of MRSA. For the phenotypic expression of MRSA, *entD* (VIMP = 2.82, SE = 0.83,  $p < .001$ ) and *ssl05/set3\_probe 2 (612)* (VIMP = 2.33, SE = 0.56,  $p < .001$ ) are also the most informative genes. As shown in S4 Table, strong correlates predicting the phenotypic expression of MRSA include enterotoxins P and D and staphylococcal superantigen-like proteins 1, 3, 4, 5, 7, and 10. For staphylococcal superantigen-like proteins 3, 5 and 10, the informative genes are identical for predicting both the genotypic and phenotypic



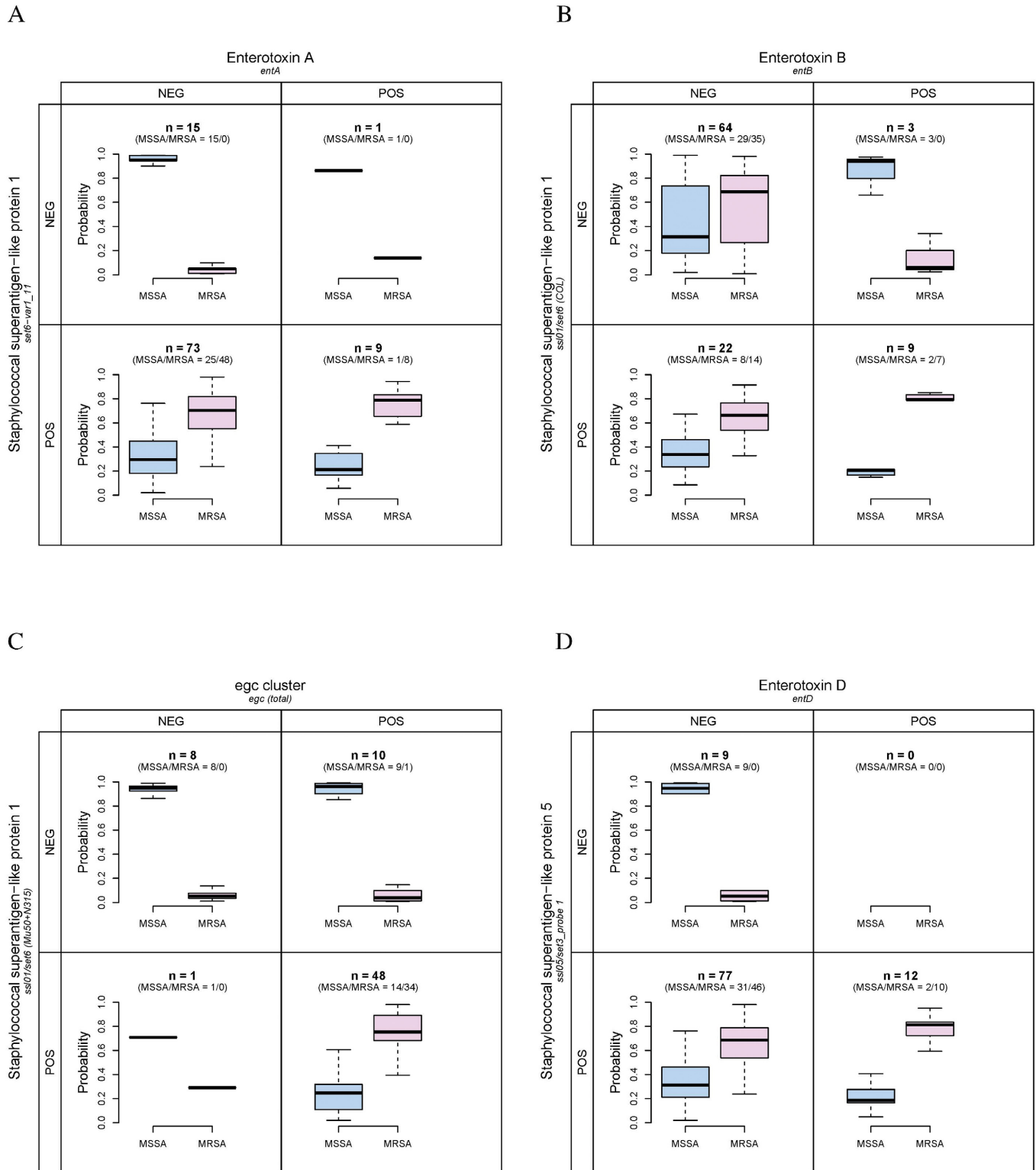
expressions of MRSA. Three virulence genes, *set6-var4\_11*, *ssl01/set6* (*Mu50+N315*) and *ssl04/set9*, are significantly informative for predicting the phenotypic expression of MRSA, but their effects for predicting the genotypic expression of MRSA are not significant. Four virulence genes, *entA*, *entJ*, *entR* and *ssl07/set1* (*AF188836*), are significantly informative for predicting the genotypic expression of MRSA, but their effects for predicting the phenotypic expression of MRSA are not significant. The contribution of these seven virulence genes on the prediction accuracy is limited, since their VIMP estimates are small.

**Phenotypic expression of MDR *S. aureus*.** For MDR *S. aureus*, enterotoxin D (*entD*, VIMP = 3.63, SE = 1.01,  $p < .001$ ) and staphylococcal superantigen-like protein 1 (*ssl01/set6* (*Mu50+N315*), VIMP = 2.27, SE = 0.48,  $p < .001$ ) are the most informative factors (see [S5 Table](#)). Eight virulence genes, *entD*, *entJ*, *entR*, *set6-var1\_11*, *ssl05/set3\_probe 1*, *ssl05/set3* (*RF122, probe-611*), *ssl05/set3\_probe 2* (612) and *ssl10/set4*, are significantly informative for predicting both MDR *S. aureus* and the genotypic expression of MRSA. Nine virulence genes, *entD*, *set6-var1\_11*, *set6-var4\_11*, *ssl01/set6* (*Mu50+N315*), *ssl04/set9*, *ssl05/set3\_probe 1*, *ssl05/set3* (*RF122, probe-611*), *ssl05/set3\_probe 2* (612) and *ssl10/set4*, are significantly informative for predicting the phenotypic expressions of both MDR and MRSA. Unlike MRSA, MDR status is associated with enterotoxins in a directly proportional relationship (see [Table 1](#) for *entC*, *entD*, and *entL*). In [Table 2](#), predictors that are also significantly informative for the phenotypic expressions of MDR and MRSA are marked in the last two columns: *entD*, *set6-var1\_11*, *ssl05/set3\_probe 1*, *ssl05/set3* (*RF122, probe-611*), *ssl05/set3\_probe 2* (612), and *ssl10/set4* are significantly informative for predicting all the outcomes. The outcomes are highly correlated. Therefore, the classification results are similar in terms of the sign of the VIMP estimates. However, the inferences could vary since the predictors correlate with each other, especially among virulence genes under the same category of virulence factor. The model predicting MDR *S. aureus* has the highest prediction accuracy: utilizing the combination of these 46 genes encoding staphylococcal enterotoxins and superantigen-like proteins, we can predict 85.71% *S. aureus* isolates correctly in terms of NMDR versus MDR status.

## Interactions between virulence factors

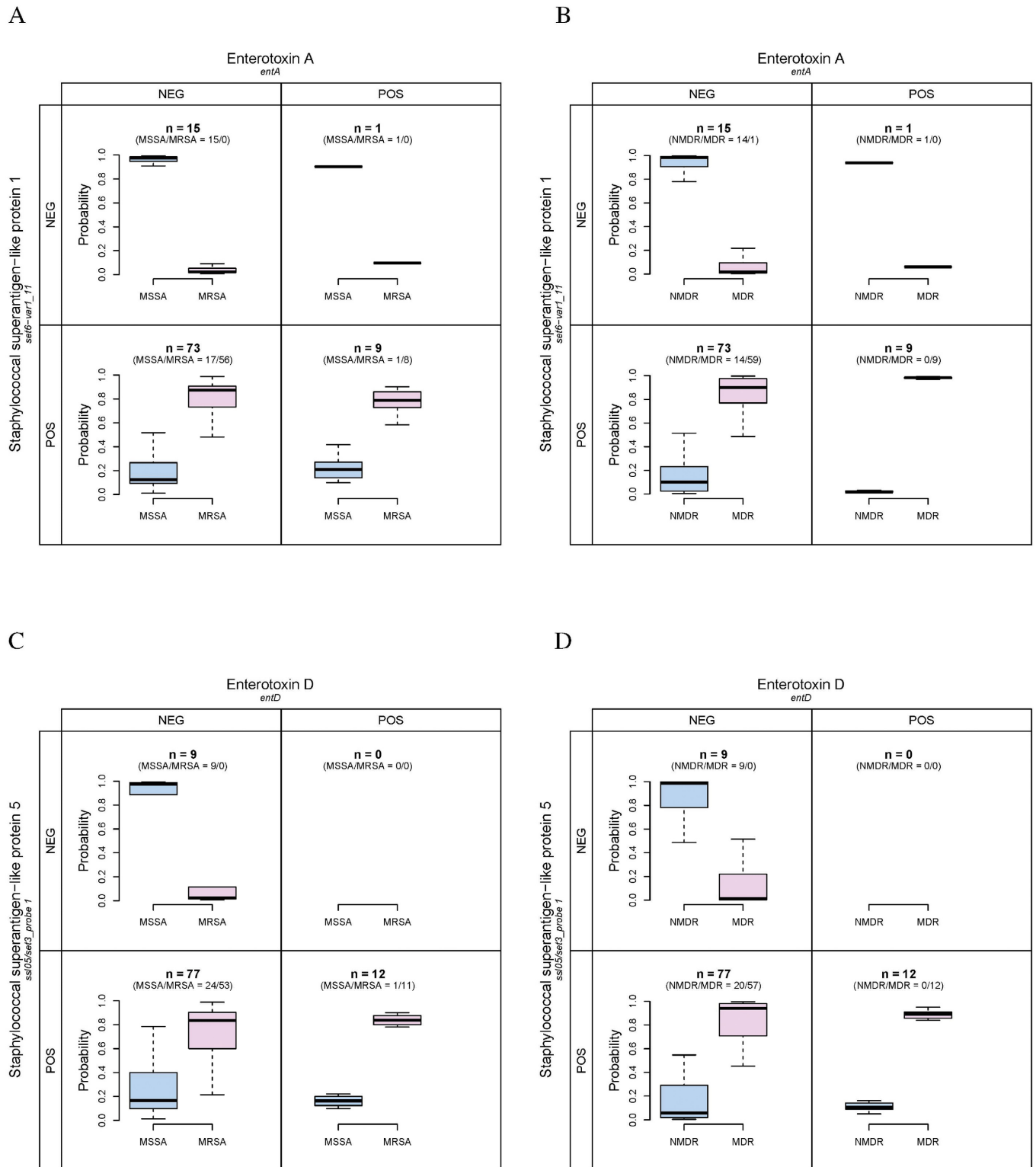
The interaction detection results for all three outcomes are very similar. Enterotoxins strongly interact with staphylococcal superantigen-like proteins. Enterotoxin A, B, and *egc* cluster interact with staphylococcal superantigen-like protein 1. As shown in [Fig 1A](#) for the genotypic expression of MRSA, when both *entA* and *set6-var1\_11* are negative, the predicted probability of MRSA is much higher than MSSA (range 90.13%–99.06% for MSSA and 0.94%–9.87% for MRSA,  $n = 15$ ). When both are positive, the predicted probability of MSSA is much lower than MRSA (range 5.69%–41.25% for MSSA and 58.75%–94.31% for MRSA,  $n = 9$ ). For the phenotypic expressions of MRSA and MDR *S. aureus*, these results are similar. As shown in [Fig 2A](#), when both *entA* and *set6-var1\_11* are negative, the predicted probability of MSSA is much higher than MRSA (range 86.06%–99.33% for MSSA and 0.67%–13.94% for MRSA,  $n = 15$ ). When both are positive, the predicted probability of MSSA is much lower than MRSA (range 9.86%–41.63% for MSSA and 58.37%–90.14% for MRSA,  $n = 9$ ). The result for MDR *S. aureus* is shown in [Fig 2B](#): when both *entA* and *set6-var1\_11* are negative, the predicted probability of NMDR is much higher than MDR (range 45.92%–99.73% for NMDR and 0.27%–54.08% for MDR,  $n = 15$ ). When both are positive, the predicted probability of NMDR is much lower than MDR (range 0.98%–29.14% for NMDR and 70.86%–99.02% for MDR,  $n = 9$ ).

As shown in [Table 1](#), the percentage of enterotoxin B positive is similar for different MRSA and MDR statuses. Enterotoxin B possibly interacts with staphylococcal superantigen-like



**Fig 1. Random Forest estimated probabilities of MSSA and MRSA in terms of genotypic expression plotted against selected interactions between staphylococcal superantigens and enterotoxins.** The numbers of detected MSSA and MRSA are listed in the parentheses. A: The interaction between *entA* and *set6-var1\_11*. B: The interaction between *entB* and *ssl01/set6 (COL)*. C: The interaction between *egc (total)* and *ssl01/set6 (Mu50+N315)*. D: The interaction between *entD* and *ssl05/set3\_probe 1*.

<https://doi.org/10.1371/journal.pone.0254519.g001>



**Fig 2. Interactions between staphylococcal enterotoxins A and D and superantigen-like proteins 1 and 5 for predicting the phenotypic expressions of MRSA and MDR *S. aureus*.** The numbers of detected MSSA, MRSA, NMDR and MDR isolates are listed in the parentheses. A: The interaction between *entA* and *set6-var1\_11* for MRSA. B: The interaction between *entA* and *set6-var1\_11* for MDR *S. aureus*. C: The interaction between *entD* and *ssl05/set3\_probe 1* for MRSA. D: The interaction between *entD* and *ssl05/set3\_probe 1* for MDR *S. aureus*.

<https://doi.org/10.1371/journal.pone.0254519.g002>

protein 1. The presence of enterotoxin B predicts higher probability of MDR or MRSA only when *ssl01/set6* (*COL*) is positive. If *ssl01/set6* (*COL*) is negative, *entB* positive is associated with higher probability of NMDR or MSSA. The result for the genotypic expression of MRSA is shown in Fig 1B. A similar interaction exists for *egc* (*total*) and *ssl01/set6* (*Mu50+N315*). As shown in Fig 1C, when both of them are negative, the predicted probability of MSSA is much higher than MRSA (range 86.26%–98.98% for MSSA and 1.02%–13.74% for MRSA,  $n = 8$ ). When both are positive, the predicted probability of MSSA is much lower than MRSA (range 1.98%–76.21% for MSSA and 23.79%–98.02% for MRSA,  $n = 48$ ). When the two genes are both positive or negative for these two interactions, the range for the probability of MDR *S. aureus* or the genotypic expression of MRSA is much overlap with its non-resistant counterpart, indicating that the interaction effects are weaker than the pair of *entA* and *set6-var1\_11*. However, for the phenotypic expression of MRSA, the interaction between *egc* (*total*) and *ssl01/set6* (*Mu50+N315*) is strong: when both of them are negative, the predicted probability of MSSA is much higher than MRSA (range 86.06%–98.65% for MSSA and 1.35%–13.94% for MRSA,  $n = 8$ ). When both are positive, the predicted probability of MSSA is much lower than MRSA (range 1.19%–55.54% for MSSA and 44.46%–98.81% for MRSA,  $n = 48$ ).

An interaction between enterotoxin D and staphylococcal superantigen-like protein 5 exists. For the prediction of the genotypic expression of MRSA, the result is shown in Fig 1D: when both *entD* and *ssl05/set3\_probe 1* are negative, the predicted probability of MSSA is much higher than MRSA (range 65.98%–99.06% for MSSA and 0.94%–34.02% for MRSA,  $n = 9$ ). When both are positive, the predicted probability of MSSA is much lower than MRSA (range 4.85%–46.4% for MSSA and 53.6%–95.15% for MRSA,  $n = 12$ ). For the phenotypic expressions of MDR and MRSA, these results are similar. As shown in Fig 2C, when both *entD* and *ssl05/set3\_probe 1* are negative, the predicted probability of MSSA is much higher than MRSA (range 51.85%–99.33% for MSSA and 0.67%–48.15% for MRSA,  $n = 9$ ). When both are positive, the predicted probability of MSSA is much lower than MRSA (range 9.88%–47.17% for MSSA and 52.83%–90.12% for MRSA,  $n = 12$ ). The result for MDR *S. aureus* is shown in Fig 2D: when both *entD* and *ssl05/set3\_probe 1* are negative, the predicted probability of NMDR is much higher than MDR (range 48.54%–99.73% for NMDR and 0.27%–51.46% for MDR,  $n = 9$ ). When both are positive, the predicted probability of NMDR is much lower than MDR (range 4.97%–30.95% for NMDR and 69.05%–95.03% for MDR,  $n = 12$ ).

The interactions between enterotoxins are weaker. We found two interactions: enterotoxins A and Q, and enterotoxins D and R. When these genes present in pairs, the probability of MDR and MRSA is high. Take the genotypic expression of MRSA for example. As shown in S2 Fig, the predicted probability of MRSA is quite high when both enterotoxins A and Q are positive (range 58.75%–94.31%,  $n = 7$ ). Similarly, when enterotoxins D and R are positive, the predicted probability of MRSA is high, ranging from 53.6%–95.15% in 11 isolates. However, the absence of enterotoxins A and Q or enterotoxins D and R in pairs does not indicate that the probability of MDR or MRSA is lower than its non-resistant counterpart, which means that these enterotoxins alone could not distinguish any outcome deterministically.

## Discussion

The recovery of MDR strains of *S. aureus* gave a warning to the usage of antibiotics. Methicillin and multidrug resistant *S. aureus* has high clinical significance and poses a potential public health hazard. We analyzed staphylococcal superantigen and enterotoxin genes to examine their potential for bacterial pathogenicity and probe their potential mechanisms of resistance to antibiotics. We found that enterotoxin D and staphylococcal superantigen-like proteins 5 are the most predominant virulence factors associated with MRSA, while enterotoxin D and

staphylococcal superantigen-like proteins 1 are the most predominant virulence factors associated with MDR *S. aureus*. The presence of *entA* and *set6-var1\_11*, as well as *entD* and *ssl05/set3\_probe 1* in pairs, predicts MDR and MRSA. This, along with our findings of enterotoxins A, J, R and staphylococcal superantigen-like proteins 1, 3, 7, and 10 could evoke new research about their roles in the initiation and pathogenesis of the disease. Although staphylococcal enterotoxins were found to be highly correlated with MRSA, we found that they are more predictive for MDR status in ocular isolates. Since the patterns of staphylococcal enterotoxin genes are probably source-associated, identifying these genes is a potential method to trace the sources of infection in ocular infection. Clinically these findings are important because they allow us to determine the superantigen profile of persistent strains and may bring inspiration into the routine application of antimicrobial susceptibility testing to prevent the emergence of antibiotic-resistant strains of potential public health concern. Therefore, our results provide important antimicrobial resistance and hygienic information on the control and screening of *S. aureus* in ocular medicine.

The high accuracy of our machine learning Random Forest model indicates that there are significant differences in the spectrum of ocular pathology between different methicillin and multidrug resistant statuses. Since *S. aureus* isolates with the *mecA* gene are more likely to be MDR and difficult to treat, our model is valuable for facilitating optimal treatments of *S. aureus* ocular infections. Overall, staphylococcal superantigen-like proteins correlate with MRSA stronger than enterotoxins. This finding is similar to a previous study of livestock-associated MRSA isolates from farms and farmers with hospital-acquired MRSA [45]; the sources of our isolates are different, and we included more virulence factors to choose the most informative ones more selectively through the Random Forest model. We found that the information from staphylococcal superantigen-like proteins is redundant because 30 variables showed significant association with the genotypic expression of MRSA from 45 Pearson's Chi-squared tests, but we only discovered 11 significant variables from our classification model. These 11 variables, including *set6-var1\_11*, *set6-var2\_11*, *ssl01/set6 (MW2+MSSA476)*, *ssl03/set8\_probe 1*, *ssl05/set3\_probe 1*, *ssl05/set3 (RF122, probe-611)*, *ssl05/set3\_probe 2 (612)*, *ssl05/set3 (MRSA252)*, *ssl07/set1 (MRSA252)*, *ssl07/set1 (AF188836)*, and *ssl10/set4*, cover staphylococcal superantigen-like proteins 1, 3, 5, 7, and 10, which indicates that staphylococcal superantigen-like proteins 2, 4, 6, 8, 9, 11 and staphylococcal exotoxin-like protein second locus could be highly correlated with these 11 virulence genes. Staphylococcal superantigen-like protein 6 seems not associated with MDR or MRSA.

Enterotoxin B was found to be more prevalent in MSSA than MRSA; whether such association is significant [23, 24] or not [25] is unclear. We found that this association was stronger when *ssl01/set6 (COL)* was negative. Our findings for enterotoxins A to D are similar to the results of isolates from urinary tract infections in Baba-Moussa et al. study [25] that enterotoxins A and D are more informative than enterotoxins B and C. After adjusting for other virulence factors, our model suggested that *entA (N315) / entP* positive indicated a higher probability of MRSA; the main effects of other enterotoxins were not as strong as *entD*. In other studies, enterotoxin genes demonstrated heterogeneous main effects [24, 46–48]. We believe that the complex interactions between enterotoxin genes and staphylococcal superantigen-like proteins contribute to these heterogeneous findings and potentially explain some temporal or geographic variation in the MDR and MRSA epidemic. Moreover, treatment with staphylococcal enterotoxin B has been used to suppress immune rejection during corneal transplantation in mice potentially due to its effects on T-cell depletion and acquiring donor-specific immunosuppression [49]. Our study inspires future research on staphylococcal enterotoxin B treatment for preventing resistance and maintaining the effectiveness of antibiotics.

Enterotoxin genes are more informative to predict MDR *S. aureus* compared with MRSA. The interactions between enterotoxin genes exist, but enterotoxins alone could not distinguish MRSA versus MSSA or NMDR versus MDR deterministically. For example, the predicted probability of MDR and MRSA is quite high when both enterotoxins A and Q or both enterotoxins D and R are positive. However, if these virulence factors were not produced, the probability of MDR or MRSA was not significantly lower. These results indicate an association between enterotoxins and *S. aureus* strains, regardless of the methicillin and multidrug resistance phenotype. However, when interacting with staphylococcal superantigen-like proteins, the effect of enterotoxins is deterministic. We found that when both *entA* and *set6-var1\_11*, or both *egc (total)* and *ssl01/set6 (Mu50+N315)*, or both *entD* and *ssl05/set3\_probe 1* are negative in pairs, the probability of MDR and MRSA is much lower than NMDR and MSSA, while the presence of these genes in pairs is associated MDR and MRSA. These genes may act in pairs or groups to exert profound toxic effects upon the immune system. Although it is unclear whether these combinations link to aggravation of the corneal damage or exacerbation of the ocular surface inflammatory response, these results provide valuable insight to effective treatments of *S. aureus* ocular infections and strict hygiene as well as preventative measures. Further research is needed to determine why the interaction between staphylococcal superantigens and enterotoxins is so strong.

We analyzed virulence factors in a more inclusive fashion to identify the complex interactions between virulence genes and find the ones that are the most informative to predict each outcome after adjusting for multiple predictors. It is possible that after adjustment for different virulence factors, the discoveries of informative virulence genes could be more consistent and reproducible. When the variables are multifactorial and interacted, a large sample size and flexible statistical assumption of the prediction model are often required. For this dataset, we tried classical logistic regression and lasso penalized logistic regression. None of the coefficients were significant from the classical logistic regression, and the lasso penalized logistic regression did not predict the outcome as accurately as the Random Forest model. Moreover, when the virulence genes are highly correlated and interacted, nonparametric variable important index [39, 50–52], instead of odds ratio or regression coefficient, may be more suitable for providing insights into ocular pathology. In looking back, we can now see that the success of the Random Forest model can be largely attributed to its ability to accommodate interactions. It delivered an ideal prediction accuracy and has a high potential for practical implementation. We demonstrated that combining the results of MDR and MRSA testing and staphylococcal enterotoxin and superantigen profiling is feasible in Random Forest models for the discrimination of the genetic diversity and drug resistance in *S. cereus*.

## Supporting information

**S1 Table. Variable importance results for predicting the genotypic expression of MRSA (full version of Table 2).**

(DOCX)

**S2 Table. Summary of duplicated results of S1 Table.**

(DOCX)

**S3 Table. Role of *mecA* genotype in the phenotypic expressions of methicillin and multi-drug resistance.**

(DOCX)

**S4 Table. Informative virulence genes for predicting the phenotypic expression of MRSA.**

(DOCX)

**S5 Table. Informative virulence genes for predicting the phenotypic expression of MDR *S. aureus*.**

(DOCX)

**S1 Fig. Adjusted probability of the genotypic expression of MRSA from Random Forest prediction plotted against candidate virulence genes.** A: Higher prevalence of enterotoxin P in MRSA isolates after adjusting for other virulence genes. In other words, *entA* (*N315*) / *entP* positive is associated with MRSA. B: After adjusting for other virulence genes, *ssl05/set3* (*RF122*, *probe-611*) positive is associated with MRSA. C: After adjusting for other virulence genes, *ssl07/set1* (*MRSA252*) negative is associated with MRSA.

(TIF)

**S2 Fig. Random Forest estimated probabilities of MSSA and MRSA in terms of genotypic expression plotted against selected interactions between enterotoxins.** A: The interaction between *entQ* and *entA*. B: The interaction between *entR* and *entD*.

(TIF)

## Author Contributions

**Conceptualization:** Darlene Miller.

**Formal analysis:** Min Lu.

**Funding acquisition:** Darlene Miller.

**Investigation:** Darlene Miller.

**Methodology:** Min Lu.

**Resources:** Darlene Miller.

**Supervision:** Darlene Miller.

**Validation:** Min Lu, Darlene Miller.

**Visualization:** Min Lu.

**Writing – original draft:** Min Lu.

**Writing – review & editing:** Min Lu, Jean-Marie Parel, Darlene Miller.

## References

1. Abolghait SK, Fathi AG, Youssef FM, Algammal AM. Methicillin-resistant *Staphylococcus aureus* (MRSA) isolated from chicken meat and giblets often produces staphylococcal enterotoxin B (SEB) in non-refrigerated raw chicken livers. *International Journal of Food Microbiology*. 2020; 328:108669. <https://doi.org/10.1016/j.ijfoodmicro.2020.108669> PMID: 32497922
2. Algammal AM, Enany ME, El-Tarabili RM, Ghobashy MO, Helmy YA. Prevalence, antimicrobial resistance profiles, virulence and enterotoxin-determinant genes of MRSA isolated from subclinical bovine mastitis samples in Egypt. *Pathogens*. 2020; 9(5):362. <https://doi.org/10.3390/pathogens9050362> PMID: 32397408
3. Enany ME, Algammal AM, Shagar GI, Hanora AM, Elfeil WK, Elshaffy NM. Molecular typing and evaluation of Sidr honey inhibitory effect on virulence genes of MRSA strains isolated from catfish in Egypt. *Pakistan journal of pharmaceutical sciences*. 2018; 31(5). PMID: 30150182
4. Algammal AM, El-Kholy AW, Riad EM, Mohamed HE, Elhaig MM, Yousef SAA, et al. Genes encoding the virulence and the antimicrobial resistance in enterotoxigenic and shiga-toxigenic *E. coli* isolated from diarrheic calves. *Toxins*. 2020; 12(6):383. <https://doi.org/10.3390/toxins12060383> PMID: 32532070

5. Algammal AM, Mohamed MF, Tawfik BA, Hozzein WN, El Kazzaz WM, Mabrok M. Molecular typing, antibiogram and PCR-RFLP based detection of *Aeromonas hydrophila* complex isolated from *Oreochromis niloticus*. *Pathogens*. 2020; 9(3):238. <https://doi.org/10.3390/pathogens9030238> PMID: 32235800
6. Algammal AM, El-Sayed ME, Youssef FM, Saad SA, Elhaig MM, Batiha GE, et al. Prevalence, the antibiogram and the frequency of virulence genes of the most predominant bacterial pathogens incriminated in calf pneumonia. *AMB Express*. 2020; 10(1):1–8. <https://doi.org/10.1186/s13568-020-01037-z> PMID: 32472209
7. Algammal AM, Mabrok M, Sivaramasamy E, Youssef FM, Atwa MH, El-Kholy AW, et al. Emerging MDR-*Pseudomonas aeruginosa* in fish commonly harbor opr L and tox A virulence genes and bla TEM, bla CTX-M, and tet A antibiotic-resistance genes. *Scientific Reports*. 2020; 10(1):1–12. <https://doi.org/10.1038/s41598-020-72264-4>
8. El-Sayed M, Algammal A, Abouel-Atta M, Mabrok M, Emam A. Pathogenicity, genetic typing, and antibiotic sensitivity of *Vibrio alginolyticus* isolated from *Oreochromis niloticus* and *Tilapia zillii*. *Rev Med Vet*. 2019; 170:80–86.
9. Rutar T, Zwick OM, Cockerham KP, Horton JC. Bilateral blindness from orbital cellulitis caused by community-acquired methicillin-resistant *Staphylococcus aureus*. *American journal of ophthalmology*. 2005; 140(4):740–742. <https://doi.org/10.1016/j.ajo.2005.03.076> PMID: 16226533
10. Shanmuganathan V, Armstrong M, Buller A, Tullo A. External ocular infections due to methicillin-resistant *Staphylococcus aureus* (MRSA). *Eye*. 2005; 19(3):284–291. <https://doi.org/10.1038/sj.eye.6701465> PMID: 15375372
11. Hsiao CH, Chuang CC, Tan HY, Ma DH, Lin KK, Chang CJ, et al. Methicillin-resistant *Staphylococcus aureus* ocular infection: a 10-year hospital-based study. *Ophthalmology*. 2012; 119(3):522–527. <https://doi.org/10.1016/j.ophtha.2011.08.038> PMID: 22176801
12. Sato Ki. External ocular infections due to methicillin-resistant *Staphylococcus aureus* and medical history. *Canadian Journal of Ophthalmology*. 2015; 50(5):e97–e99. <https://doi.org/10.1016/j.cjco.2015.05.006> PMID: 26455994
13. Freidlin J, Acharya N, Lietman TM, Cevallos V, Witcher JP, Margolis TP. Spectrum of eye disease caused by methicillin-resistant *Staphylococcus aureus*. *American journal of ophthalmology*. 2007; 144(2):313–315. <https://doi.org/10.1016/j.ajo.2007.03.032> PMID: 17659970
14. Solomon R, Donnenfeld ED, Holland EJ, Yoo SH, Daya S, Güell JL, et al. Microbial keratitis trends following refractive surgery: results of the ASCRS infectious keratitis survey and comparisons with prior ASCRS surveys of infectious keratitis following keratorefractive procedures. *Journal of Cataract & Refractive Surgery*. 2011; 37(7):1343–1350. <https://doi.org/10.1016/j.jcrs.2011.05.006> PMID: 21700112
15. Schulte RH, Munson E. *Staphylococcus aureus* Resistance Patterns in Wisconsin 2018 Surveillance of Wisconsin Organisms for Trends in Antimicrobial Resistance and Epidemiology (SWOTARE) Program Report. *Clinical medicine & research*. 2019; 17(3-4):72–81. <https://doi.org/10.3121/cm.2019.1503> PMID: 31582419
16. Bispo PJ, Ung L, Chodosh J, Gilmore MS. Hospital-Associated Multidrug-Resistant MRSA Lineages Are Trophic to the Ocular Surface and Cause Severe Microbial Keratitis. *Frontiers in Public Health*. 2020; 8:204. <https://doi.org/10.3389/fpubh.2020.00204> PMID: 32582610
17. Otto M. A MRSA-terious enemy among us: end of the PVL controversy? *Nature medicine*. 2011; 17(2):169–170. <https://doi.org/10.1038/nm0211-169> PMID: 21297612
18. Sola C, Paganini H, Egea AL, Moyano AJ, Garnera A, Kevric I, et al. Spread of epidemic MRSA-ST5-IV clone encoding PVL as a major cause of community onset staphylococcal infections in Argentinean children. *PLoS One*. 2012; 7(1):e30487. <https://doi.org/10.1371/journal.pone.0030487> PMID: 22291965
19. Schlievert PM. Cytolysins, superantigens, and pneumonia due to community-associated methicillin-resistant *Staphylococcus aureus*. *The Journal of infectious diseases*. 2009; 200(5):676–678. <https://doi.org/10.1086/605333> PMID: 19653828
20. Assimacopoulos AP, Strandberg KL, Rotschafer JH, Schlievert PM. Extreme pyrexia and rapid death due to *Staphylococcus aureus* infection: analysis of 2 cases. *Clinical infectious diseases*. 2009; 48(5):612–614. <https://doi.org/10.1086/597009> PMID: 19191649
21. Collery MM, Smyth DS, Tumilty JJ, Twohig JM, Smyth CJ. Associations between enterotoxin gene cluster types egc1, egc2 and egc3, agr types, enterotoxin and enterotoxin-like gene profiles, and molecular typing characteristics of human nasal carriage and animal isolates of *Staphylococcus aureus*. *Journal of medical microbiology*. 2009; 58(1):13–25. <https://doi.org/10.1099/jmm.0.005215-0> PMID: 19074649
22. Ronco T, Klaas IC, Stegger M, Svennensen L, Astrup LB, Farre M, et al. Genomic investigation of *Staphylococcus aureus* isolates from bulk tank milk and dairy cows with clinical mastitis. *Veterinary microbiology*. 2018; 215:35–42. <https://doi.org/10.1016/j.vetmic.2018.01.003> PMID: 29426404



23. Liu C, Chen Zj, Sun Z, Feng X, Zou M, Cao W, et al. Molecular characteristics and virulence factors in methicillin-susceptible, resistant, and heterogeneous vancomycin-intermediate *Staphylococcus aureus* from central-southern China. *Journal of Microbiology, Immunology and Infection*. 2015; 48(5):490–496. <https://doi.org/10.1016/j.jmii.2014.03.003> PMID: 24767415
24. Sabouni F, Mahmoudi S, Bahador A, Pourakbari B, Sadeghi RH, Ashtiani MTH, et al. Virulence factors of *Staphylococcus aureus* isolates in an Iranian referral children's hospital. *Osong public health and research perspectives*. 2014; 5(2):96–100. <https://doi.org/10.1016/j.phrp.2014.03.002> PMID: 24955319
25. Baba-Moussa L, Anani L, Scheftel J, Couturier M, Riegel P, Haikou N, et al. Virulence factors produced by strains of *Staphylococcus aureus* isolated from urinary tract infections. *Journal of hospital infection*. 2008; 68(1):32–38. <https://doi.org/10.1016/j.jhin.2007.10.010> PMID: 18069084
26. Argudín MÁ, Mendoza MC, Rodicio MR. Food poisoning and *Staphylococcus aureus* enterotoxins. *Toxins*. 2010; 2(7):1751–1773. <https://doi.org/10.3390/toxins2071751> PMID: 22069659
27. Fraser JD, Proft T. The bacterial superantigen and superantigen-like proteins. *Immunological reviews*. 2008; 225(1):226–243. <https://doi.org/10.1111/j.1600-065X.2008.00681.x> PMID: 18837785
28. Breiman L. Random forests. *Machine learning*. 2001; 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
29. Xu S, Huang X, Xu H, Zhang C. Improved prediction of coreceptor usage and phenotype of HIV-1 based on combined features of V3 loop sequence using random forest. *Journal of microbiology*. 2007; 45(5):441–446. PMID: 17978804
30. Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC bioinformatics*. 2008; 9(1):1–10. <https://doi.org/10.1186/1471-2105-9-319> PMID: 18647401
31. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications*. 2019; 134:93–101. <https://doi.org/10.1016/j.eswa.2019.05.028> PMID: 32968335
32. Rice TW, Rusch VW, Apperson-Hansen C, Allen MS, Chen LQ, Hunter JG, et al. Worldwide esophageal cancer collaboration. *Diseases of the Esophagus*. 2009; 22(1):1–8. <https://doi.org/10.1111/j.1442-2050.2008.00901.x> PMID: 19196264
33. Xu S, Zhang Z, Wang D, Hu J, Duan X, Zhu T. Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework. In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). IEEE; 2017. p. 228–232.
34. Peterson JC, Durkee H, Miller D, Maestre-Mesa J, Arboleda A, Aguilar MC, et al. Molecular epidemiology and resistance profiles among healthcare-and community-associated *Staphylococcus aureus* keratitis isolates. *Infection and drug resistance*. 2019; 12:831. <https://doi.org/10.2147/IDR.S190245> PMID: 31043797
35. Fang X, Liu W, Ai J, He M, Wu Y, Shi Y, et al. Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC infectious diseases*. 2020; 20(1):1–8. <https://doi.org/10.1186/s12879-020-4930-2> PMID: 32171261
36. Ong J, Liu X, Rajarethinam J, Kok SY, Liang S, Tang CS, et al. Mapping dengue risk in Singapore using Random Forest. *PLoS neglected tropical diseases*. 2018; 12(6):e0006587. <https://doi.org/10.1371/journal.pntd.0006587> PMID: 29912940
37. Bachert C, Zhang N, Holtappels G, De Lobel L, Van Cauwenberge P, Liu S, et al. Presence of IL-5 protein and IgE antibodies to staphylococcal enterotoxins in nasal polyps is associated with comorbid asthma. *Journal of allergy and clinical immunology*. 2010; 126(5):962–968. <https://doi.org/10.1016/j.jaci.2010.07.007> PMID: 20810157
38. Coates-Brown R, Moran JC, Pongchaikul P, Darby AC, Horsburgh MJ. Comparative Genomics of *Staphylococcus* reveals determinants of speciation and diversification of antimicrobial defense. *Frontiers in microbiology*. 2018; 9:2753. <https://doi.org/10.3389/fmicb.2018.02753> PMID: 30510546
39. Ishwaran H, Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in medicine*. 2019; 38(4):558–582. <https://doi.org/10.1002/sim.7803> PMID: 29869423
40. Ishwaran H, Kogalur UB. Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC); 2019. Available from: <https://cran.r-project.org/package=randomForestSRC>.
41. Aragon TJ. epitools: Epidemiology Tools; 2020. Available from: <https://CRAN.R-project.org/package=epitools>.
42. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*. 2015; 24(1):44–65. <https://doi.org/10.1080/10618600.2014.907095>

43. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media; 2009.
44. Molnar C. Interpretable machine learning. Lulu. com; 2020.
45. Mutters NT, Bieber CP, Hauck C, Reiner G, Malek V, Frank U. Comparison of livestock-associated and health care-associated MRSA—genes, virulence, and resistance. *Diagnostic microbiology and infectious disease*. 2016; 86(4):417–421. <https://doi.org/10.1016/j.diagmicrobio.2016.08.016> PMID: [27640079](https://pubmed.ncbi.nlm.nih.gov/27640079/)
46. van Trijp MJ, Melles DC, Snijders SV, Wertheim HF, Verbrugh HA, van Belkum A, et al. Genotypes, superantigen gene profiles, and presence of exfoliative toxin genes in clinical methicillin-susceptible *Staphylococcus aureus* isolates. *Diagnostic microbiology and infectious disease*. 2010; 66(2):222–224. <https://doi.org/10.1016/j.diagmicrobio.2009.08.021> PMID: [19828275](https://pubmed.ncbi.nlm.nih.gov/19828275/)
47. Sina H, Ahoyo TA, Moussaoui W, Keller D, Bankolé HS, Barogui Y, et al. Variability of antibiotic susceptibility and toxin production of *Staphylococcus aureus* strains isolated from skin, soft tissue, and bone related infections. *BMC microbiology*. 2013; 13(1):1–9. <https://doi.org/10.1186/1471-2180-13-188> PMID: [23924370](https://pubmed.ncbi.nlm.nih.gov/23924370/)
48. Kolawole DO, Adeyanju A, Schaumburg F, Akinyoola AL, Lawal OO, Amusa YB, et al. Characterization of colonizing *Staphylococcus aureus* isolated from surgical wards' patients in a Nigerian university hospital. *PLoS One*. 2013; 8(7):e68721. <https://doi.org/10.1371/journal.pone.0068721> PMID: [23935883](https://pubmed.ncbi.nlm.nih.gov/23935883/)
49. Zhang Y, Pan Z, Chen Y, Jie Y, He Y. Specific immunosuppression by mixed chimerism with bone marrow transplantation after *Staphylococcal Enterotoxin B* pretreatment could prolong corneal allograft survival in mice. *Molecular vision*. 2012; 18:974. PMID: [22550390](https://pubmed.ncbi.nlm.nih.gov/22550390/)
50. Lu M, Ishwaran H. A prediction-based alternative to P values in regression models. *The Journal of thoracic and cardiovascular surgery*. 2018; 155(3):1130. <https://doi.org/10.1016/j.jtcvs.2017.08.056> PMID: [29306487](https://pubmed.ncbi.nlm.nih.gov/29306487/)
51. Williamson BD, Gilbert PB, Carone M, Simon N. Nonparametric variable importance assessment using machine learning techniques. *Biometrics*. 2021; 77(1):9–22. <https://doi.org/10.1111/biom.13392> PMID: [33043428](https://pubmed.ncbi.nlm.nih.gov/33043428/)
52. Lu M, Ishwaran H. Discussion on “Nonparametric variable importance assessment using machine learning techniques” by Brian D. Williamson, Peter B. Gilbert, Marco Carone, and Noah Simon. *Biometrics*. 2021; 77(1):23–27. <https://doi.org/10.1111/biom.13391> PMID: [33290584](https://pubmed.ncbi.nlm.nih.gov/33290584/)