Identification of Quorum Sensing Peptides using Random Forest and Instance-based Classifier

Akanksha Rajput, Manoj Kumar*

Bioinformatics Centre, Institute of Microbial Technology, Sector 39A, Chandigarh-160036, India.
Email: akanksha@imtech.res.in; manojk@imtech.res.in

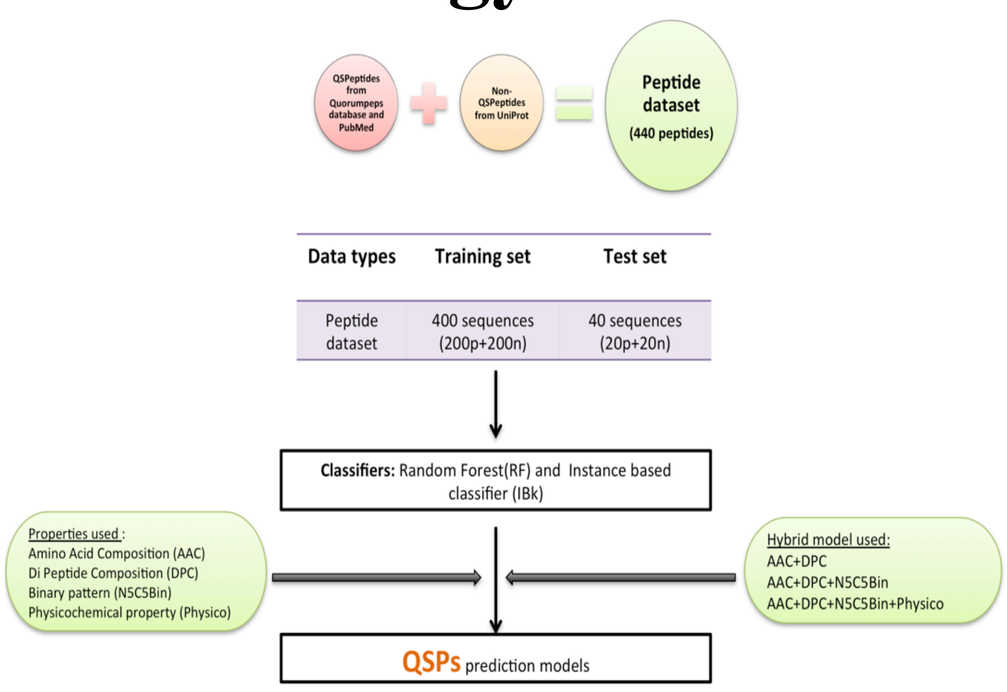
Introduction

- Quorum sensing is mechanism of communication among bacteria. Quorum sensing is exhibited by signaling molecules (Miller and Bassler 2001).
- Quorum sensing peptides (QSPs) are signaling molecules in Gram-positive bacteria. QSPs help bacteria in various functions like biofilm formation, virulence, etc (Schauder and Bassler 2001).
- Therefore, identification of QSPs are important

Objective

- To identify QSPs, we used Random Forest (RF) and Instance-based Classifier (IBk) from Weka (Waikato Environment for Knowledge Analysis) package (Frank, Hall et al. 2004)

Methodology

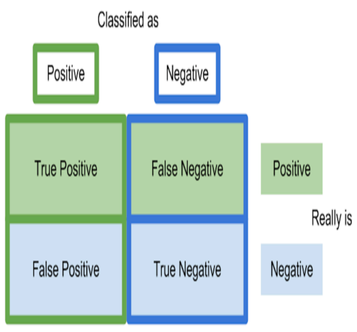


Classifiers used

- **Random forests** are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees (Breiman 2001).

- **IBk** in Weka implements K nearest neighbors algorithm which is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions) (Altman 1992).

Parameters used for measuring performance of various classifiers



$$Sensitivity = \frac{TP}{TP+FN} \times 100$$

$$Specificity = \frac{TN}{TN+FP} \times 100$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \times 100$$

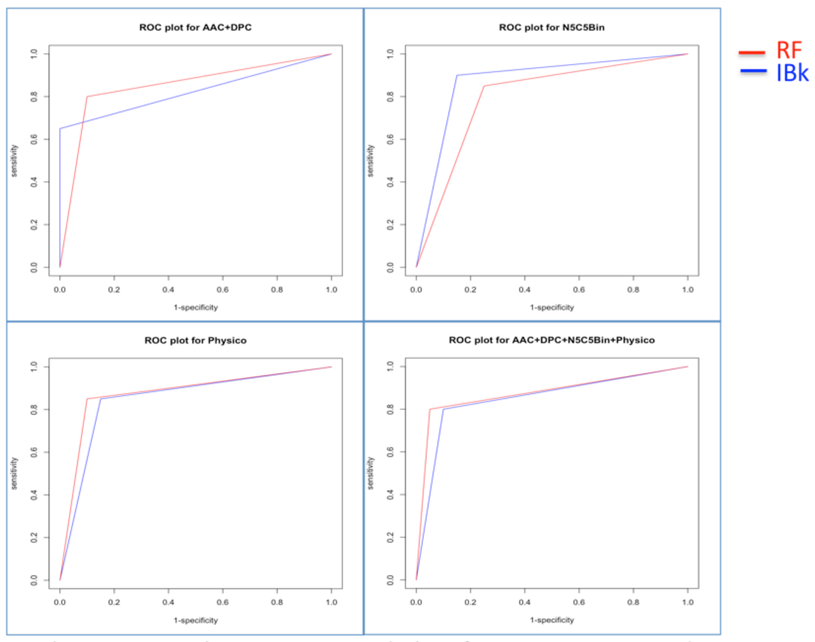
Results

Performance of Random Forest (RF) and Instance-based Classifier (IBk) by employing distinct peptide properties during 10- fold cross validation

Properties	Techniques	Sen	Spec	Acc	MCC	AUC
AAC	RF	85.00	90.00	87.50	0.75	0.92
	IBk	75.00	90.00	82.50	0.66	0.83
DPC	RF	75.00	100.00	87.50	0.77	0.90
	IBk	60.00	95.00	77.50	0.59	0.78
AAC+DPC	RF	80.00	90.00	85.00	0.70	0.96
	IBk	65.00	100.00	82.50	0.69	0.83
N5Bin	RF	80.00	75.00	77.50	0.55	0.87
	IBk	85.00	85.00	85.00	0.70	0.91
C5Bin	RF	85.00	85.00	85.00	0.70	0.90
	IBk	80.00	95.00	87.50	0.76	0.93
N5C5Bin	RF	85.00	75.00	80.00	0.60	0.87
	IBk	90.00	85.00	87.50	0.75	0.91
Physico	RF	85.00	90.00	87.50	0.75	0.94
	IBk	85.00	85.00	85.00	0.70	0.85
AAC+DPC+ N5C5Bin	RF	85.00	80.00	82.50	0.65	0.94
	IBk	80.00	90.00	85.00	0.70	0.85
AAC+DPC+N5C5Bin+Physico	RF	80.00	95.00	87.50	0.76	0.97
	IBk	80.00	90.00	85.00	0.70	0.85

AAC, Amino Acid Composition; DPC, Di Peptide Composition; N5AAC, Amino Acid Composition of 5 N-terminal residues; C5AAC, Amino Acid Composition of 5 C-terminal residues; N5Bin, Binary pattern of 5 N-terminal residues; C5Bin, Binary pattern of 5 C-terminal residues; N5C5Bin, Binary pattern of 5 N and 5 C terminal residues; Physico, top 10 physicochemical properties; Sen, Sensitivity; Spec, specificity; Acc, Accuracy; MCC, Mathew's correlation coefficient; AUC, Area Under the curve

ROC of various hybrid models



RF,Random Forest; IBk, Instance Based Classifier; AAC, Amino acid composition; DPC, Dipeptide composition; N5C5Bin, Binary pattern of 5 N and 5 C terminal residues; Physico, top 10 physico;

Conclusion

- In this study, we concluded that classifiers *viz.* RF and IBk are good classifiers to identify QSPs

References

1. Miller, M. B. and B. L. Bassler (2001). "Quorum sensing in bacteria." Annu Rev Microbiol **55**: 165-199.
2. Schauder, S. and B. L. Bassler (2001). "The languages of bacteria." Genes Dev **15**(12): 1468-1480.
3. Breiman, L. (2001). "Random Forests." Machine Learning **45**(1): 5-32.
4. Frank, E., M. Hall, et al. (2004). "Data mining in bioinformatics using Weka." Bioinformatics **20**(15): 2479-2481
5. Altman, N. S. (1992). "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression." The American Statistician **46**(3): 175-185

Acknowledgements

- Council of Scientific and Industrial Research (CSIR)
- Department of Biotechnology(DBT), Govt. of India