Prediction of Potential Drug Targets and Vaccine Candidates Against Antibiotic-Resistant *Pseudomonas aeruginosa*

Fatemeh Sadat Shamsinejad1 · Zahra Zafari1

Accepted: 20 September 2022 / Published online: 7 November 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

*Pseudomonas aeruginosa* is one of the leading causes of nosocomial infections, characterized by increasing antibiotic resistance, severity and mortality. Therefore, numerous efforts have been made nowadays to identify new therapeutic targets. This study aimed to find potential drug targets and vaccine candidates in drug-resistant strains of *P. aeruginosa*. Extensive antibiotic-resistant and carbapenem-resistant strains of *P. aeruginosa* with complete genome were selected and ten common hypothetical proteins (HPs) containing more than 200 amino acids were obtained. The structural, functional and immunological predictions of these HPs were performed with the utility of bioinformatics approaches. Two common HPs (Gene ID: 2877781645 and 2877781936) among other investigated proteins were revealed as potential candidates for pharmaceutical and vaccine purposes based on structural and physicochemical properties, functional domains, subcellular localizations, signal peptides, toxicity, virulence factor, antigenicity, allergenicity and immunoinformatic predictions. The consequence of this predictive study will assist in novel drug and vaccine design through experimental investigations.

Keywords *Pseudomonas aeruginosa* · Hypothetical proteins · Drug resistance · Vaccine

Introduction

*Pseudomonas aeruginosa* is a gram-negative bacterium, a gamma proteobacter, belonging to the Pseudomonadaceae family with an excellent adaptability and survival in a variety of different environments and ecological niches (Klockgether and Tümmler 2017; Pang et al. 2019). This bacterium is an opportunistic pathogen and one of the most frequent and severe cause of nosocomial infections (NIs) leading to a wide range of infections, from acute to chronic life-threatening, in humans (Liu et al. 2015; Lund-Palau et al. 2016). The mortality rate due to the infections caused by this bacterium is very high, especially in people with compromised immunity and people with cystic fibrosis (CF) (López-Causapé et al. 2018). *Pseudomonas aeruginosa* is inherently resistant to a wide range of antibiotics and belongs to the group of ESKAPE pathogens that have potential mechanisms of drug resistance (Santajit and Indrawattana 2016; van Duin and Paterson 2016). The emergence of MDR (multi-drug resistant), XDR (extensively drug-resistant), and PDR (pan-drug resistant) strains of *P. aeruginosasa* and the increasing number of the resistant strains of this bacterium are considered as a major crisis and paints a scary future for current antibiotics and challenges the effectiveness of these antibiotics (Bassetti et al. 2013; Ventola 2015). Despite the efforts made, there is still no vaccine available for this bacterium and the last line of treatment is carbapenem group of antibiotics (meropenem, doripenem, ertapenem and imipenem) but the growing number of carbapenem-resistant strains has made them critical priorities with an urgent need for developing new treatment strategies by the World Health Organization (WHO) (Priebe and Goldberg 2014; Haenni et al. 2017; World Health Organization 2019).

The genome of *P. aeruginosa* is large (5.5–7 million base pairs) and contains many open reading frames predicted as hypothetical proteins (Klockgether et al. 2011; Newman et al. 2017). Hypothetical proteins (HPs) are predicted proteins with no experimental evidences for their structures and functions but they may be associated with human diseases and may play an important role in understanding biological and functional pathways, finding new structures, functions, domains, motifs and markers, fighting pathogens and early detection...
or treatment of infectious diseases, and developing new vaccines and drug candidates (Naqvi et al. 2015; Varma et al. 2015; Ijaq et al. 2019). Functional annotation and curation of HPs in several human pathogens have been also reported (Rabbi et al. 2021).

In the present study hypervirulent strains of \textit{P. aeruginosa} were isolated, and then structural, functional, and immuninformatics analysis of common hypothetical proteins were performed with the utility of in silico methods to identify potential drug and vaccine candidates against the resistant strains of \textit{P. aeruginosa}.

**Materials and Methods**

**Sequence Retrieval**

\textit{Pseudomonas aeruginosa} strains with completed genome sequencing were collected through Integrated Microbial Genome (IMG) site (https://img.jgi.doe.gov/) (Chen et al. 2021). MDR, XDR and strains resistant to carbapenem antibiotics were selected by literature review. The Pseudomonas database (https://www.pseudomonas.com/) was also used to evaluate the resistance of the strains (Winsor et al. 2016). The most resistant strain was selected as a query and placed in the homologous part in IMG site to find the common hypothetical proteins. In the non-homologous part, \textit{Homo sapiens} and two human microbiome bacteria (\textit{Escherichia coli} K-12 MG1655 and \textit{E. coli} K12-W3110) were placed. Subsequently, 10 common hypothetical proteins with a length of > 200 amino acids were selected for further study (Fig. 1). The sequences of these final proteins were submitted to several servers and predictive databases to characterize the hypothetical proteins.

**Primary Structural Analysis of Hypothetical Proteins**

The physicochemical properties of the selected hypothetical proteins were predicted by ProtParam tool (https://web.expasy.org/protparam/). The physicochemical properties of a protein provide basic information about its properties (Wilkins et al. 1999).

The subcellular localization of the hypothetical proteins was predicted using 4 servers: PSORTb(V.3) (https://www.psort.org/psortb/) (Yu et al. 2010), CELLO II (http://cello.life.nctu.edu.tw/) (Yu et al. 2006), PSLpred (https://webs.iiitd.edu.in/raghuva/pslpred/index.html) (Bhasin et al. 2005) and Gneg-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/Gneg-multi/) (Shen and Chou 2010). The CELLO II server predicts the position of proteins based on the support vector machine (SVM) method (Yu et al. 2006). The PSLpred server was developed to predict the location of gram-negative bacterial proteins (Bhasin et al. 2005). The Gneg-mPLoc server uses a combination of gene ontology (GO) information and domain function to predict the position of proteins (Shen and Chou 2010). According to various reports, bioinformatics servers cannot accurately detect the location of some proteins to predict the subcellular localization alone. In this study, 4 different servers were used to identify the location of proteins to cover each other’s weaknesses (Gardy and Brinkman 2006; Brown et al. 2012).

We used TMHMM, HMMTOP, SignalP and SecretomeP for more precise analysis of HPs. We used TMHMM(V.2) (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al. 2001) and HMMTOP (http://www.enzim.hu/hmmtop/) (Tusnády and Simon 2001) to predict the transmembrane helices for the hypothetical proteins. Both servers are based on hidden Markov model (HMM) (Krogh et al. 2001; Tusnády and Simon 2001).

The SignalP(V.5) server (http://www.cbs.dtu.dk/services/SignalP/) was used to investigate the presence of signal peptides in the hypothetical proteins. This server is based on the deep neural network (Almagro Armenteros et al. 2019). The existence of non-classical secretory pathway in hypothetical proteins was predicted by SecretomeP(V.2) server (http://www.cbs.dtu.dk/services/SecretomeP/) (Bendtsen et al. 2005).

**Secondary Structural Analysis of Hypothetical Proteins**

Two servers, SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=NPSA/npsa_sopma.html) and PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/), were used to predict the secondary structure of hypothetical proteins. The PSIPRED server uses a two-stage neural network to predict the second structure (Geourjon and Deléage 1995; Buchan and Jones 2019).

**Three-Dimensional Structure Analysis of Hypothetical Proteins**

The biological functions of a protein are determined by its third structure. The I-TASSER server (https://zhanggroup.org/I-TASSER/) was used to predict the tertiary structure of the putative proteins. The I-TASSER server is an integrated platform for automatically predicting protein structure and its performance is based on sequence-to-structure-to-function pattern (Roy et al. 2010).

**Evaluate of the Quality of Three-Dimensional Structures**

The quality evaluation of three-dimensional structures of hypothetical proteins was performed using ERRAT tool (https://saves.mbi.ucla.edu/). Types of different atoms due
to Energetic and Geometric effects on each other non-randomly proteins are distributed. Errors in model construction lead to a more random distribution of different types of atoms. Statistical methods can be used to distinguish them from correct distributions. Subject that the basis of development is ERRAT (Colovos and Yeates 1993).

Fig. 1 Schematic workflow of the common hypothetical proteins in silico. Common hypothetical proteins of resistant strains were obtained and then structural, functional and immunological predictions were made for these proteins. Finally, two proteins were introduced as vaccine and drug candidates.

**Functional Analysis of Hypothetical Proteins**

Interpretation of different domains of a protein is a very important topic in protein-related studies. SMART(V.9) (http://smart.embl-heidelberg.de/) (Letunic et al. 2021) and Pfam(V.34) (http://pfam.xfam.org/) databases (Mistry et al.
were used to predict the functional domains in hypothetical proteins. The SMART database is a web resource for identifying and interpreting functional domains (Letunic et al. 2021). The Pfam database is widely used to classify protein sequences in protein families and domains (Mistry et al. 2021).

The CDD database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was used to examine the conserved domains of hypothetical proteins. The CDD database is part of the Entrez NCBI search and retrieval system (Marchler-Bauer et al. 2013).

BLASTP. BLASTP tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) was used to search for similar sequences of hypothetical proteins. Investigating the similarity of the protein sequences studied in the study with other proteins in other organisms can provide useful information about finding different domains. This tool compares the query protein sequences with the protein sequences in the protein database (Altschul et al. 1997).

The toxicity of hypothetical proteins was investigated using the BTXpred server (https://webs.iiitd.edu.in/raghava/btxpred/index.html). This server classifies toxins into two groups: (I) exotoxins and (II) endotoxins (Saha and Raghava 2007).

Function prediction for hypothetical proteins was performed by the VICMPred server (https://webs.iiitd.edu.in/raghava/vicmpred/index.html). This server classifies pathogenic microbial proteins into 4 functional classes: (I) virulence factor, (II) information molecule, (III) cellular process, and (IV) metabolism (Saha and Raghava 2006a, b).

**Immunoinformatics Analysis**

The antigenicity of hypothetical proteins was predicted with VaxiJen(V.2) server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). Predicting conserved antigens is crucial for the development of vaccine candidates. This server predicts conserved antigens using an alignment-independent approach (Doytchinova and Flower 2007). Hypothetical proteins with a threshold above 0.5 were considered as antigens.

AllerTOP(V.2) (https://www.ddg-pharmfac.net/AllerTOP/) (Dimitrov et al. 2014) server was used to predict the allergenicity of hypothetical proteins.

Major histocompatibility complex (MHC) is a basic cell surface protein of the vertebrate immune system (Kaufman 2018). MHC molecules that present antigens are highly polymorphic and are classified into two main classes, MHCI and MHCII (Reche et al. 2004). MHCI is involved in the determination of epitopes for cytotoxic T cells (TCDB8+) and MHCII is involved in the determination of epitopes for helper T cells (TCDA4+) (Reche et al. 2004). The IEDB database (https://www.iedb.org/) was used to predict MHC-binding epitopes in hypothetical proteins (Vita et al. 2019). For this evaluation, the list of alleles in the default site was used.

Epitopes of hypothetical proteins detected by B lymphocytes were predicted using two servers, ABCpred (https://webs.iiitd.edu.in/raghava/abcpred/index.html) (Saha and Raghava 2006a, b) and Ellipro (http://tools.iedb.org/ellipro/) (Ponomarenko et al. 2008). ABCpred server was used to predict continuous B lymphocyte epitopes with the default threshold. The approach of this server is to use recurrent neural network (Saha and Raghava 2006a, b). The Ellipro server was used to predict discontinuous B lymphocyte epitopes with the default threshold. This server makes predictions based on the geometric properties of the protein structure. Based on the geometrical properties of the protein structure, this server predicts discontinuous epitopes of B lymphocytes (Ponomarenko et al. 2008).

**Results**

**Sequence Retrieval**

Thirty five resistant strains with finished genome sequences were detected (Table 1). The 10 common hypothetical proteins are shown in Table 2.

**Primary Structural Analysis of Hypothetical Proteins**

Physicochemical properties of 10 hypothetical proteins including molecular weight, isoelectric point, instability index, aliphatic index, and hydrophobicity (GRAVY) were evaluated. HP2, HP3, HP4, HP5, and HP6 proteins are stable and others are unstable. Only HP2 protein has pI > 7 and others have pI < 7. The aliphatic index of HP6 protein is the highest. The GRAVY index of all hypothetical proteins is negative, so they are all hydrophilic (Table 3).

Results of prediction of subcellular localization, transmembrane helixes, and prediction of SignalP and SecretomeP are shown in Table 4.

**Secondary Structural Analyzes of Hypothetical Proteins**

In general, the most common secondary structure in hypothetical proteins was the coil structure. The results of SOPMA and PSIPRED servers for predicting the secondary structure of 10 hypothetical common proteins are given in Supplementary File 1.
The results of I-TASSER server for predicting the tertiary structure of 10 hypothetical proteins are given in Fig. 2.

**Evaluate the Quality of Three-Dimensional Structures of Hypothetical Proteins**

The quality of all three-dimensional structure models predicted by the I-TASSER server was evaluated with the ERRAT tool. For the ERRAT tool, a quality tertiary structure model must score above 50. All models obtained from the I-TASSER server had a score above 50.

**Functional Analysis of Hypothetical Proteins**

Functional and conserved domains of hypothetical proteins were predicted using SMART, Pfam, and CDD databases. No functional and conserved domains were predicted for HP6. The results of three databases are as follows: HP1, HP5, HP7, and HP10 proteins have the DUF domain, HP2, and HP8 proteins have the OprD domain, HP4 protein has the MoaF and MoaF-C domains, HP9 protein has the FeCR domain in SMART and Pfam database and has the COG4254 domain in CDD database, HP3 protein has the M60-like domain in SMART database and has the peptidase_M60 like super family, IMPa_N_2, and IMPa_helical domains.

BLASTP was used for the prediction of similarity between 10 hypothetical proteins with other proteins in other organisms. In general, the hypothetical proteins were most similar to organisms of *Acinetobacter*. 

### Table 1: The discovered resistant strains of *P. aeruginosa*

| Strain  | Reference                        | Strain  | Reference                        |
|---------|----------------------------------|---------|----------------------------------|
| AR_0446 | Pseudomonas database             | AR444   | Pseudomonas database             |
| AR_455  | Pseudomonas database             | AR_0443 | Pseudomonas database             |
| AR441   | Pseudomonas database             | PA298   | Turano et al. (2017)             |
| AR442   | Pseudomonas database             | PA1088  | Turano et al. (2017)             |
| AR_0360 | Pseudomonas database             | LESB58  | Fournier et al. (2013), Kos et al. (2015) |
| AR_0110 | Pseudomonas database             | ST277   | de Oliveira Santos et al. (2019), Nascimento et al. (2019), do Nascimento et al. (2020) |
| AR_0111 | Pseudomonas database             | WPB099  | Ng et al. (2018)                 |
| AR_0230 | Pseudomonas database             | WPB100  | Ng et al. (2018)                 |
| AR_0356 | Pseudomonas database             | WPB101  | Ng et al. (2018)                 |
| AR439   | Pseudomonas database             | PA58    | Espinosa-Camacho et al. (2017)   |
| AR_0353 | Pseudomonas database             | PA124   | Espinosa-Camacho et al. (2017)   |
| AR_0354 | Pseudomonas database             | PA127   | Espinosa-Camacho et al. (2017)   |
| AR445   | Pseudomonas database             | K34-7   | Taiaroa et al. (2018)            |
| NCGM1900| Xiong et al. (2017)              | PB367   | Huang et al. (2020)              |
| NCGM1984| Tada et al. (2016)               | PA-VAP-1| Abdellatif et al. (2016), Borgatta et al. (2017) |
| NCGM2.S1| Miyoshi-Akiyama et al. (2011), Miyoshi-Akiyama et al. (2017), Xiong et al. (2017) | PA-VAP-3| Abdellatif et al. (2016), Borgatta et al. (2017), Persyn et al. (2019) |
| PB350   | Huang et al. (2020)              | PASGNDM345| Ding et al. (2018)            |
| PASGNND699| Ding et al. (2018)             |         |                                  |

### Table 2: The Gene IDs of 10 common hypothetical proteins among resistant strains of *P. aeruginosa*

| Name of hypothetical proteins | Amino acid sequence length of hypothetical proteins | Gene ID       |
|-------------------------------|----------------------------------------------------|---------------|
| HP1                           | 370                                                | 2877781414    |
| HP2                           | 460                                                | 2877781645    |
| HP3                           | 908                                                | 2877781936    |
| HP4                           | 274                                                | 2877782301    |
| HP5                           | 641                                                | 2877782543    |
| HP6                           | 201                                                | 2877781738    |
| HP7                           | 353                                                | 2877781975    |
| HP8                           | 418                                                | 2877782246    |
| HP9                           | 333                                                | 2877782354    |
| HP10                          | 238                                                | 287782514     |

### Three-Dimensional Structure Analysis of Hypothetical Proteins

The results of I-TASSER server for predicting the tertiary structure of 10 hypothetical proteins are given in Fig. 2.
baumannii, Klebsiella pneumonia, Streptococcus dysgalactiae, and Enterobacter cloacae. BLASTP results are given in Supplementary File 2.

The BTXpred server was used to predict the toxicity of 10 hypothetical proteins. According to the results of this server, HP2, HP3, HP4, HP5, HP7, HP8, and HP9 are exotoxins and other proteins are not toxins.

The VICMpred server was used for function prediction. Based on the result of this server, HP1, HP4, and HP10 proteins are involved in metabolism, HP2, HP3, and HP9 proteins are involved in virulence factor, and other proteins are involved in cellular process.

**Immunoinformatics Analysis**

HP2, HP3, HP4, HP5, HP7, HP8, and HP9 are antigens according to the VaxiJen server, while HP1, HP6, and HP10 proteins are not. HP6 protein is an allergen based on AllerTOP server results.

Prediction of T cell epitopes and epitopes identified by B lymphocytes for 10 hypothetical proteins are given in Supplementary File 3 and Supplementary File 4 respectively.
Discussion

*Pseudomonas aeruginosa* is one of the most important health challenges in the medical community due to its antibiotic resistance. Until now, many efforts have been made to prevent and treat infections caused by this bacterium. In this study, the structure and function of 10 common hypothetical proteins from resistant strains of *P. aeruginosa*
were investigated. The goal of analyzing common hypothetical proteins in resistant strains was to find potential candidates for vaccine and drug design. The structural and functional characterization of hypothetical bacterial proteins can answer many of our questions about bacterial physiology, but experimental approaches are time-consuming and cost a lot of money. In silico techniques along with laboratory techniques can overcome these problems and more accurate results can be achieved (Omeershoffudden and Kumar 2019, Uddin et al. 2019).

Immunoinformatics offers new approaches to the design of new vaccines, diagnostic targets, and the study of the pathology of infectious diseases (He et al. 2010). Studies have shown that the best candidates for vaccine design should be able to stimulate both humoral and cellular immunity (Kozakiewicz et al. 2013). Moreover, they should have antigenicity and should be extracellular proteins, not homologous to human proteins and microbiome bacteria and not allergenic to humans (Barat et al. 2012; Sudha et al. 2019). Virulence factors expressed by bacteria are indispensable for the survival and growth of pathogenic bacteria and thus can be a valuable drug targets (Sudha et al. 2019). Vaccines are a safe and cost-effective solution to fight infectious diseases (Doro et al. 2009). Peptide vaccines are suitable, safe, and contain immunogenic epitopes (Rashid et al. 2017). Immunogenic potential is primarily dependent on the affinity of MHC binding (Rashid et al. 2017). Therefore, predicting epitopes with higher binding potential for MHC is essential for designing a peptide vaccine (Rashid et al. 2017).

Considering all the features mentioned in this study, only two proteins, HP2 and HP3, among the 10 common hypothetical proteins are introduced as potential candidates for drugs and vaccines. According to the analysis performed, the location of these two proteins was predicted in an outer membrane. Moreover, for these two proteins, virulence factor function was predicted and they were exotoxins. They had antigenicity and they did not an allergen. Based on the functional and conserved domains analysis, it is predicted that the HP2 protein is a member of the OprD family. This family contains bacterial outer membrane porins with serine protease activity (Yoshihara et al. 1998). It was reported that the OprD2 protein of P. aeruginosa had protease activity (Yoshihara et al. 1998). The OprD protein in P. aeruginosa facilitates the passage of small amino acids and peptides (Ochs et al. 2000). Another role of this protein is to cross the antibiotic imipenem (Ochs et al. 2000). Long-term treatment of patients with imipenem results in the development of imipenem-resistant mutants due to mutations in the OprD gene (Ochs et al. 2000). These mutant strains have severely reduced expression levels of OprD (Ochs et al. 2000). Mutations can also lead to a lack of OprD expression in mutant strains (Ochs et al. 2000). For HP3 protein, M60-like, M60 peptidase family, IMPa helical, and IMPa-N-2 domains were predicted. The peptidase family contains a zinc metallopeptidase motif and has mucinase activity (Nakjang et al. 2012). The IMPa (immunomodulating metalloprotease of P. aeruginosa) domain is a host immune-modulatory metalloprotease in P. aeruginosa that protects the bacterium from neutrophil infestation (Bardoel et al. 2012). This is the domain belonging to the M60 peptidase family (Bardoel et al. 2012). Pseudomonas aeruginosa secretes various proteases that destroy proteins essential for host defense (Noach et al. 2017). The PA0572 protein of P. aeruginosa, which is a PSGL1 (P-selectin Glycoprotein Ligand1) inhibitor, and this secretory protease, called the P. aeruginosa immunomodulator metalloprotease or IMPa (Bardoel et al. 2012).

**Conclusion**

Antibiotic-resistant bacteria are becoming one of the biggest challenges in medicine and healthcare. Pseudomonas aeruginosa is one of the top priorities of WHO for the immediate development of treatment strategies. Identification of protein functions is very important for understanding the biological processes in bacteria. In this study, the structure and function of 10 common hypothetical proteins of resistant strains of P. aeruginosa were investigated. The results revealed that the two hypothetical proteins (Gene ID: 287781645 and 287781936) had the potential to be used in drug and vaccine design. However, in vitro and in vivo studies are required to determine the effectiveness of these candidates for drug and vaccine development.

**Acknowledgements** The authors wish to thank Department of biology, Shahed university for their support to conduct the present study.

**References**

Abdellatif S, Trifi A, Daly F, Majhoub K, Nasri R, Ben Lakhal S (2016) Efficacy and toxicity of aerosolised colistin in ventilator-associated pneumonia: a prospective, randomised trial. Ann Intensive Care 6(1):26

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37(4):420–423

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

Barat S, Willer Y, Rizos K, Claudi B, Schenmer AK, Kirchhoff D, Schmidt A, Burton N, Bumann D (2012) Immunity to intracellular Salmonella depends on surface-associated antigens. PLoS Pathog 8(10):e1002966

Bardoel BW, Hartsink D, Vughs MM, de Haas CJ, van Strijp JA, van Kessel KP (2012) Identification of an immunomodulating metalloprotease of Pseudomonas aeruginosa (IMPa). Cell Microbiol 14(6):902–913
a Canadian Community Hospital. Genome Announc. https://doi.org/10.1128/genomeA.00458-17
Yoshihara E, Yoneyama H, Ono T, Nakae T (1998) Identification of the catalytic triad of the protein D2 protease in Pseudomonas aerugi-
osa. Biochem Biophys Res Commun 247(1):142–145
Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins 64(3):643–651
Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahi-nalp SC, Ester M, Foster LJ, Brinkman FS (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26(13):1608–1615

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.