ACTA SCIENTIFIC MEDICAL SCIENCES

Volume 3 Issue 7 July 2019

Research Article

Random Sequences May Have Putative Beta-Lactamase Properties

Brenda ZN Kwek[1,2], Keerthana D Ardhanari-Shanmugam[1,2], Jun Hong Woo[1,2], Sharlene Usman[1,2], Jing Wen Chua[1,2], Vicnesh B[1,2], Khadija Shahrukh[1,2], Chakrit Thong-Ek[1,2], Maurice HT Ling[1,2,3]*

[1]*Department of Applied Sciences, Northumbria University, United Kingdom*

[2]*School of Life Sciences, Management Development Institute of Singapore, Singapore*

[3]*HOHY PTE LTD, Singapore*

***Corresponding Author:** Maurice HT Ling, School of Life Sciences, Management Development Institute of Singapore, HOHY PTE LTD, Singapore.

Received: June 04, 2019; **Published:** June 13, 2019

Abstract

Beta-lactamases, which confer resistance to beta-lactam antibiotics, is of medical and healthcare concerns globally. Studies had placed the emergence of beta-lactamases to more than 2 billion years ago. However, it is not known where the first beta-lactamase originate. In this study, we examine the probability of de novo emergence of putative beta-lactamase from random sequences. A set of 10 thousand randomly generated sequences were aligned using Smith-Waterman algorithm and Needleman-Wunsch algorithm to a set of known class D beta-lactamases isolated from GenBank to determine the probability of each randomly generated sequence as putative beta-lactamases. Our results suggest that substantial proportion of randomly generated sequences may be putative beta-lactamases, with 4% of the randomly generated sequences showing 99% probability as putative beta-lactamases. To test whether a putative beta-lactamase can evolve over generations to have more characteristics of known beta-lactamases, *in silico* evolution was carried out using DOSE, an evolution simulation software. Our simulation results also suggest that a putative beta-lactamase may rapidly evolve into a more functional beta-lactamase under selection. Hence, *de novo* origination of beta-lactamase from random sequences is plausible.

Keywords: *de novo*; Beta-Lactamase; *in silico*

Introduction

Beta-lactamases are enzymes which inactivate beta-lactam antibiotics [1], such as penicillin, by hydrolysing the amide bond [2] in the beta-lactam ring. Enzymatically, beta-lactamases can be classified as serine beta-lactamases or metallo-beta-lactamases. Serine beta-lactamases catalyses the hydrolysis via a serine-bound intermediate [2] while metallo-beta-lactamase require zinc as intermediate [3]. The first beta-lactamase isolated is penicillinase from *Escherichia coli* in 1940 [4], before the beginning of antibiotic era in 1942 [5]. However, phylogenetic analysis suggests serine beta-lactamases to be more than 2 billion years old [6], strongly suggesting that beta-lactamases originated before antibiotic era, with the main function to aid in cell survival by defending early microorganisms from naturally occurring beta-lactams produced by other competitors in the environment [7]. Clinically, carbapenem-hydrolysing class D beta-lactamases (CHDLs) are important as they confer resistance against carbapenems, which are antibiotics of last resort [8,9]. A study [10] has shown that OXA-2 beta-lactamases are CHDLs, even though currently they are classified as narrow-spectrum. However, it is not known where the first beta-lactamase originated.

Studies has shown that some functional sequences can originate randomly. In vitro selection experiments demonstrate that functional RNAs and proteins can be obtained from random sequence libraries [11]. Keefe and Szostak [12] have produced four new functional ATP-binding proteins from random-sequence libraries. Many random sequences can be bioactive peptide or RNA [13]. Moreover, recent studies demonstrate the possibility of putatively functional genes [14] and peptides [15] from random DNA sequences and amino acid sequences respectively. Hence, is it possible for functional beta-lactamases to originate from random sequences?

This study examines the possibility of functional Oxa-2 beta-lactamases originating from random sequences by evaluating ran-

domly generated sequences against known Oxa-2 beta-lactamases. Our results suggest that substantial proportion of randomly generated sequences may be putative beta-lactamases, with 4% of the randomly generated sequences showing 99% probability as putative beta-lactamases.

Methods

Sequence Data Sets

A set of Oxa-2 beta-lactamases sequences (known as baseline sequences) were retrieved from GenBank using *Salmonella enterica* subsp. enterica serovar Typhimurium R46 blaOXA gene for oxacillin-hydrolyzing class D beta-lactamase OXA-2 (Accession number NG_049496.1) as the query on BLASTN with default parameters and an E-value threshold of 1e-9. A set of 10,000 random sequences between 74 and 1,029 nucleotides in length and 2,654 adenine, 2,296 cytosine, 2,745 guanine and 2,304 thymine per 10,000 bases; without start and stop codons; were generated using RANDOMSEQ [16]. Length of 74 to 1,029 represented 20th to 100th percentile of the lengths of baseline sequences.

Determining open reading frames from random sequences

An open reading frame (ORF) can be defined as a sequence of codons, which begins and ends with a start and stop codon respectively [17]. A set of 10 sequences of 10 kilobases with uniform nucleotide distribution was generated using RANDOMSEQ [16] and ORFs of at least 33 nucleotides, corresponding to one of the shortest gene known [18], were identified.

Determining putative beta-lactamases from random sequences

Identification of putative beta-lactamases were carried out using a previously described method [14]. Two sets of pairwise sequence alignments were performed using both Smith-Waterman algorithm [19], also known as local alignment, and Needleman-Wunsch algorithm [20], also known as global alignment, via Bactome (https://github.com/mauriceling/bactome). In the first series, each baseline sequence was pairwise aligned to every other baseline sequence and the distribution of scores were used as measure for putative beta-lactamase sequences. In the second series, each of the 10,000 random sequence was pairwise aligned to every baseline sequence. A minimum and average alignment score were generated for each random sequence. Based on bootstrap statistics [21], the probability of each random sequence being a putative beta-lactamase sequence was determined by the proportion of baseline alignment scores below the minimum and average alignment score of the random sequence for stringent and relaxed criteria respectively.

In Silico evolution of putative beta-lactamase

To test whether a putative beta-lactamase can evolve over generations to have more characteristics of known beta-lactamases, *in silico* evolution was carried out using DOSE [22,23], using previously described methods [24,25]. Briefly, a single population of 100 digital organisms was created, deployed in the same ecological cell and simulated for 500 generations. A random sequence with minimum alignment score just above that of the baseline sequences was used as genome for the ancestral organism, which would be cloned into the initial population of 100 organisms. 10% point mutation rate [26,27] will be used. Organism fitness was calculated as average pairwise alignment of its genome to a random selection of 250 baseline sequences (known beta-lactamases). The lowest decile of the organisms by fitness were removed. However, in event where more than 50% of the population were removed, a random selection of 10 organisms were removed instead. A random selection of remaining organisms after removal were replicated to top up the population to 100 organisms for the next generation. The simulation was repeated 30 times.

Results and Discussion

Characterization of Oxa-2 Beta-lactamases

BLASTN of using *Salmonella enterica* subsp. enterica serovar Typhimurium R46 blaOXA gene for oxacillin-hydrolyzing class D beta-lactamase OXA-2 (Accession number NG_049496.1) yielded 6,555 hits within the E-value threshold, forming the set of baseline sequences. The distribution of nucleobases of the 6,555 baseline sequences is 26.54% adenine, 22.96% cytosine, 27.45% guanine, and 23.03% thymine. The minimum and maximum nucleotide length for baseline sequences are 24 and 1029 respectively (Figure 1), with an average and standard deviation of 94.9 and 106.55 bases respectively. The baseline sequences were pairwise aligned, and yield a total of 21,480,735 alignments (Figure 2). The mean alignment score is 57.1 (standard deviation of 18.46), with 11 and 1,028 as the minimum and maximum scores respectively. Alignment results from local [19] and global [20] were identical; hence, local alignment was used for subsequent analysis.

184 ORFs per 10 kilobases found.

We identified an average of 184 ORFs per random sequence of 10 kilobases (Figure 3). The shortest ORF found consists of 36 nucleotides while the longest ORF consists of 498 nucleotides. Our result suggests that ORFs, potential protein-coding region of a gene [28], can exist randomly and is consistent with a recent study showing 196 ORFs per 10 kb of random sequences [14]. Cardoso-Moreira and Long [29] had presented a model of *de novo* origins

of ORFs through mutations. However, our results suggest another possible route for *de novo* ORF – emerging from random sequences without the need for mutations, which does not contradict the model proposed by Cardoso-Moreira and Long [29].

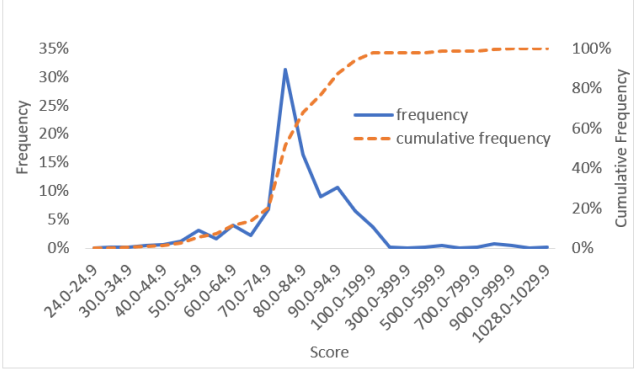

Figure 1: Distribution of baseline sequence lengths.

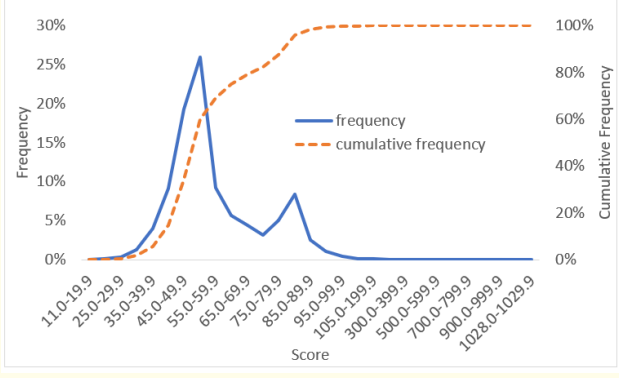

Figure 2: Beta-lactamase local pairwise alignment score.

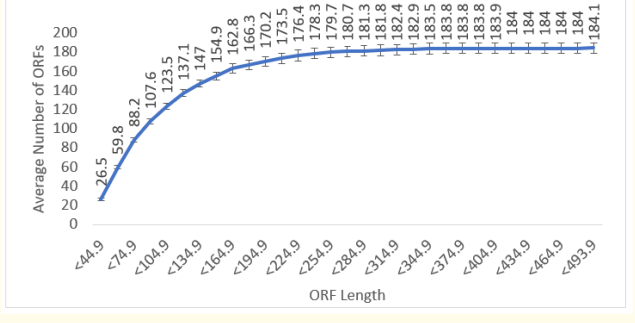

Figure 3: Total number of ORFs identified per 10 kilobases of randomly generated sequences.

4% Random sequences with more than 99% probability as putative beta-lactamases

Our results show that all 10,000 randomly generated sequences have minimum pairwise alignment score equal or higher than the minimum local pairwise alignment score of 6,555 from baseline sequences. As the range of pairwise alignment scores among baseline sequences represents the sequence diversity of beta-lactamases; therefore, if a random sequence is not likely a putative beta-lactamases, then its minimum pairwise alignment score with known beta-lactamases (baseline sequences) should be lower than the minimum pairwise alignment score among known beta-lactamases. Our results show that all 10,000 randomly generated sequences have 34% probability of being putative beta-lactamases (Table 1), based on the probability that the average pairwise alignment scores of all randomly generated sequences are above 34% of the 21,480,735 baseline pairwise scores. A possibility can be a left-skew of baseline alignment scores due to short sequences in the baseline set. Despite so and using the same argument, 4% of the 10,000 randomly generated sequences have 99% probability of being putative beta-lactamases as their pairwise alignment scores are at or above 99th percentile of the baseline pairwise alignment scores.

Alignment Score	Minimum Score	Average Score	Probability of Beta-lactamase Function
>10.9	100%	100%	0%
>19.9	100%	100%	0.2%
>24.9	0%	100%	1%
>29.9	0%	100%	2%
>34.9	0%	100%	6%
>39.9	0%	100%	15%
>44.9	0%	100%	34%
>49.9	0%	99%	60%
>54.9	0%	97%	69%
>59.9	0%	95%	75%
>64.9	0%	93%	79%
>69.9	0%	90%	82%
>74.9	0%	86%	87%
>79.9	0%	80%	96%
>84.9	0%	67%	98%
>89.9	0%	4%	99%
>94.9	0%	0%	100%

Table 1: Prediction of random sequences probability of functionality. Random sequences minimum and average and pairwise alignment scores are projected against baseline sequences alignment scores.

This is consistent with several studies. Yona., *et al*. [30] report about 60% of random sequences have wild-type promoter efficiency with only 1 base mutation out of 103 bases. Ling [15] suggests that nearly 27% of randomly generated amino acid chains may contain putative protein domains. Although Ling [15] uses amino acid chains, Neme., *et al*. [13] report that 25% of randomly generated 150 nucleotide sequences exhibit beneficial effect on *E. coli* growth rate when expressed as RNA or peptide. This is also supported by Thong-Ek [14] suggesting the possibility of de novo emergence of putative archaebacterial genes. Zhang., *et al*. [31] also suggest an average emergence of 51.5 *de novo* genes per million years in *Oryza* by studying the genomes of 13 closely related *Oryza species*.

Putative beta-lactamase can evolve under selective pressure

Next, we examine whether a putative beta-lactamase can evolve under selective pressure by selecting a random sequence with the lowest average pairwise alignment score with the baseline sequences and perform *in silico* evolution. The selected sequence, Test_2716, has the sequence length of 103 nucleobases and has the lowest average score of 53.05. Our simulation results (Figure 4) show a possibility for Test_2716 to mutate into a functional beta-lactamase as its average maximum score cross the baseline average score of 57.1 at the 4th simulated generation. This is consistent with previous studies [24,25] showing a rapid increase in fitness under selective pressure. Using the stricter criteria of grand mean (mean of means), the fitness of Test_2716 increases but did not surpass the baseline average of known beta-lactamases. This suggests that Test_2716 may not reach the average functionality of a putative beta-lactamase. However, the grand mean plateau between 56.3 and 56.5, which is more than 69% probability of being putative beta-lactamases (Table 1), from 20th to 500th generation. This suggests that a putative beta-lactamase may rapidly evolve into a more functional beta-lactamase under selection.

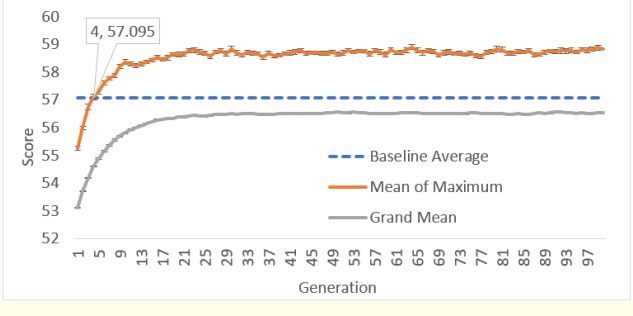

Figure 4: Simulation result of random sequence Test_2716. Error bars represent standard errors.

Carvunis., *et al*. [32] defined "proto-gene" as a gene born from non-genic sequence by random processes without selection, and must fulfil 3 criteria; namely, the DNA sequence must be transcribed and translated, and the protein product must be beneficial to the organism. A proto-gene is the first stage of a *de novo* gene origin with a beneficial and selectable phenotype that can be selected [11]. As Yona., *et al*. [30] demonstrate that substantial proportion of random sequences can be promoters, it is plausible to consider that the requirement for transcription has substantial chance of being randomly fulfilled.

Conclusion

Our results demonstrate the feasibility of *de novo* origination of ORFs and putative beta-lactamases from random sequences; hence, the requirement for translation is fulfilled. This is supported by Zhang., *et al*. [31] reporting that 56.6% of the *de novo* genes identified are translated. As antibiotic resistance genes such as beta-lactamase can improve survivability in stressful ecological conditions [33-35], the requirement for beneficial function is also fulfilled. Baym., *et al*. [36] have demonstrated the emergence of high-resistant mutants from susceptible strains within 2 weeks under a strong selective pressure. Therefore, it is plausible to consider that beta-lactamase may originate *de novo* from random sequences and selected for survival benefits.

Bibliography

1. AA Shah., *et al.* "Extended-spectrum beta-lactamases (ESbLs): characterization, epidemiology and detection." *Critical Reviews in Microbiology* 30.1 (2004): 25-32.

2. T Palzkill. "Metallo-β-lactamase structure and function." *Annals of the New York Academy of Sciences* 1277 (2013): 91-104.

3. R Fernandes., et al. "β-Lactams: chemical structure, mode of action and mechanisms of resistance." *Reviews in Medical Microbiology* 24.1 (2013): 7-17.

4. E P Abraham and E Chain. "An enzyme from bacteria able to destroy penicillin." *Nature* 146.3713 (1940): 837.

5. W A Adedeji. "The treasure called antibiotics." *Annals of Ibadan Postgraduate Medicine* 14.2 (2016): 56-57.

6. G Hall and M Barlow. "Evolution of the serine beta-lactamases: past, present and future." *Drug Resistance Updates* 7.2 (2004): 111-123.

7. K. Bush. "Past and present perspectives on β-lactamases." *Antimicrobial Agents and Chemotherapy* 62.10 (2018): e01076.

8. F S Codjoe and E S Donkor. "Carbapenem resistance: A review." *Medical Sciences (Basel)* 6.1 (2017): 1.

9. C C Sheu., *et al.* "Infections caused by carbapenem-resistant enterobacteriaceae: an update on therapeutic options." *Frontiers in Microbiology* 10 (2019): 80.

10. N T Antunes., *et al*. "Class D β-lactamases: are they all carbapenemases?." *Antimicrobial Agents and Chemotherapy* 58.4 (2014): 2119-2125.

11. M Weisman and S R Eddy. "Gene evolution: getting something from nothing." *Current Biology* 27.13 (2017): R661-R663.

12. D Keefe and J W Szostak. "Functional proteins from a random-sequence library." *Nature* 410.6829 (2001): 715-718.

13. R Neme., *et al*. "Random sequences are an abundant source of bioactive RNAs or peptides." *Nature Ecology & Evolution* 1.6 (2017): 0217.

14. Thong-Ek., *et al.* "Potential de novo origins of archaebacterial glycerol-1-phosphate dehydrogenase (G1PDH)." *Acta Scientific Microbiology* 2.6 (2019): 106-110.

15. M H Ling. "De novo putative protein domains from random peptides." *Acta Scientific Microbiology* 2.4 (2019): 109-112.

16. M H Ling. "RANDOMSEQ: Python command–line random sequence generator." *MOJ Proteomics & Bioinformatics* 7.4 (2018): 206-208.

17. P Sieber., *et al.* "The Definition of open reading frame revisited." *Trends Genet* 34.3 (2018): 167-170.

18. J I Pueyo and J P Couso. "The 11-aminoacid long Tarsal-less peptides trigger a cell signal in Drosophila leg development." *Developmental Biology* 324.2 (2008): 192-201.

19. T F Smith and M S Waterman. "Identification of common molecular subsequences." *Journal of Molecular Biology* 147.1 (1981): 195-197.

20. S B Needleman and C D Wunsch. "A general method applicable to the search for similarities in the amino acid sequence of two proteins." *Journal of Molecular Biology* 48.3 (1970): 443-453.

21. D Boos. "Introduction to the Bootstrap World." *Statistical Science* 18.2 (2003): 168-174.

22. C F Castillo and M H Ling. "Digital Organism Simulation Environment (DOSE): a library for ecologically-based in silico experimental evolution." *Advances in Computer Science: an International Journal* 3.1 (2014): 44-50.

23. J Z Lim., *et al.* "A genetic algorithm framework grounded in biology." *The Python Papers Source Codes* 2 (2010): 6.

24. C F Castillo and M H Ling. "Resistant traits in digital organisms do not revert preselection status despite extended deselection: implications to microbial antibiotics resistance." *BioMed Research International* (2014): 648389.

25. C F Castillo., *et al.* "Resistance maintained in digital organisms despite guanine/cytosine-based fitness cost and extended de-selection: implications to microbial antibiotics resistance." *MOJ Proteomics & Bioinformatics* 2.2 (2015): 00039.

26. J Rattray and J N Strathern. "Error-prone DNA polymerases: when making a mistake is the only way to get ahead." *Annual Review of Genetics* 37 (2003): 31-66.

27. D F Lee., *et al*. "Mapping DNA polymerase errors by single-molecule sequencing." *Nucleic Acids Research* 44.13 (2016): e118.

28. B J Woodcroft., *et al*. "OrfM: a fast open reading frame predictor for metagenomic data." *Bioinformatics* 32.17 (2016): 2702-2703.

29. M Cardoso-Moreira and M Long. "The origin and evolution of new genes." *Methods in Molecular Biology* 856 (2012): 161-186.

30. H Yona., *et al.* "Random sequences rapidly evolve into de novo promoters." *Nature Communications* 9.1 (2018): 1530.

31. L Zhang., *et al*. "Rapid evolution of protein diversity by de novo origination in Oryza." *Nature Ecology & Evolution* 3.4 (2019): 679-690.

32. A R Carvunis., *et al.* "Proto-genes and de novo gene birth." *Nature* 487.7407 (2012): 370-374.

33. N. Citri. "Beta-lactamase: lessons in the art of survival." *Journal of Chemotherapy* 3.2 (1991): 75-78.

34. T Duangurai., *et al.* "Burkholderia pseudomallei adaptation for survival in stressful conditions." *BioMed Research International* (2018): 3039106.

35. J M Munita and C A Arias. "Mechanisms of Antibiotic Resistance." *Microbiology Spectrum* 4 (2016): 2.

36. M Baym., *et al*. "Spatiotemporal microbial evolution on antibiotic landscapes." *Science* 353.6304 (2016): 1147.

Volume 3 Issue 7 July 2019