Rajshahi University Journal of Science & Engineering **ISSN 2309-0952**
Vol. 44: 85-93, 2016

Molecular Evolutionary Analysis of α-Defensin Peptides in Vertebrates

Arafat Rahman*, M Sahidul Islam, Otun Saha and Titon Chandra Saha

Department of Microbiology, Noakhali Science and Technology University, Noakhali, Bangladesh.

*Corresponding author: *ac.arafat@gmail.com*

Abstract

α-Defensin is a group of polypeptides with antimicrobial activity found in the host-defense system and it is widely distributed in, but not limited to mammalian epithelial cells and phagocytes. These molecules protect the organism from a diverse spectrum of bacteria, viruses, fungi, and protozoan parasites. Different studies have revealed wide sequence variation within α-defensin sequences, but the underlying evolutionary cause is not well-studied. In this study, the α-defensin gene from 25 vertebrate species has been comprehensively collected and computationally analyzed. NCBI gene and nucleotide databases were accessed to extract meta-information about α-defensin gene's defensin domain and leader propeptide sequences. Full coding sequences downloaded from nucleotide database by splitting out intron sequence. MEGA software used to construct phylogenetic tree using Neighbor-Joining method, which indicates that α-defensin gene evolution does not matches with species evolution. Selection analysis was carried out using Data Monkey web-server's FEL, SLAC, IFEL, MEME, TOGGLE and REL program on both propeptide and defensin super-family codon-aligned sequences to test different hypothesises. Positively selected sites were found on both propeptide and defensin domain, but the effect of negative selection pressure was higher on leader sequences. It was found that mutations in the positively selected sites of defensin domain had stabilizing effect on protein. Phyre2 web-server was used for homology modeling of selected α-defensin genes. Structural variation is observed on α-defensinproteins which may indicate heterogeneous structure-function relationship between species that reflects its interaction with diverse pathogens. This study provides a new perspective on the relationships among α-defensin gene repertoires which will help to infer its evolution.

Keywords: α-defensin, antimicrobial peptides, vertebrates, evolutionary analysis, selection pressure, protein stability

INTRODUCTION

Antimicrobial peptides, which are polypeptides of fewer than 100 amino acids, commonly found in animal defense systems and defensins are antimicrobial peptides of innate immunity [1]. In mammalian animals, defensins and cathelicidins are two main peptide families, but defensins are particularly prominent in human [2]. Structurally, defensins are a family of evolutionary related vertebrate antimicrobial peptides with a characteristic β-sheet-rich fold and a framework of six disulphide-linked cysteines that can be divided into two major subfamilies, namely α- and β-defensins, which differ in the length of peptide segments between the six cysteines and the pairing of the cysteines that are connected by disulphide bonds [1]. Defensins form pores in the cell membrane of bacteria by carpet-wormhole model of action using its amphipathic properties and protect the host [1].

Defensins are a much diverged group of molecules. What drives the evolution of defensin is a fundamental question and there are several studies to explore the nature and effect of

selection pressure on β-defensins [3][4]. But there is no such study on α-defensins. To inspect this research gap, this study was conducted by a selection analysis approach. The ground of selection analysis is based on the occurrences of random mutations which can be fixed in a populationeventually. In general, advantageous mutations are extremely rare because existing proteins already function quite well. Therefore the chance is very low that any change to them is an improvement. Similarly, a deleterious mutation will have a little chance because they will be selected against and will not rise in frequency. Based on that, it can be inferred that most mutations are neutral in nature and will not change an amino acid [5]. This relationship can be expressed using Ka/Ks ratio, where Ka/Ks = non-synonymous mutation rate per non-synonymous site / synonymous mutation rate per synonymous site [6][7]. This metric enables to measure the effect of evolutionary pressure on sequence level. Recently, many statistical methods have been implemented in Data Monkey web server (www.datamonkey.org) that are based on the calculation of Ka/Ks ratio and further optimized for different scenario [8]. For example, fixed effect likelihood (FEL) is an overall method in terms of the tradeoff between statistical performance and computational expense [9]. IFEL (Internal FEL) used to test for site-wise selection on internal branches of the tree [10]. Meanwhile, single likelihood ancestor counting (SLAC) is used as the most conservative method to detect selection pressure [9]. Mixed effect fixed evolution (MEME) is an extension of FEL which combines fixed effects at the level of a site with random probability and used to detect episodic diversifying selection affecting individual codon sites [11]. On the other hand, TOGGLE analysis can identify sites which toggle between a wild-type and escaped amino acid state [12]. Finally, the method REL (Random Effect Likelihood) can be used because it allows synonymous rate variation [9]. Inferred positively selected sites can be further analyzed for their effect on protein stability.The stability (ΔG) of a protein is defined by the free energy (quantified by kcal/mol) [13]. A protein is stable when free energy is low. A mutation that brings energy (ΔG>0 kcal/mol) will destabilize the structure, while a mutation that remove energy (ΔG<0 kcal/mol) will stabilize the structure. A threshold to detect a significant mutation is that if ΔG is >1 kcal/mol, which roughly corresponds to a single hydrogen bond. Molecular dynamics can be used to detect free energy change but it can be very time-consuming. FoldX uses an empirical method to estimate the stability effect of a mutation which provides good approximation with experimental studies [14].

In this study, these methods were employed to explore selection pressure on α-defensin sequences, phylogenetic and structural analysis were conducted to infer about the effect of selection process, andmutations on the positively selected sites were analyzed for their effect on protein stability.

MATERIALS AND METHODS

A. Dataset formation

Nucleotide sequences of α-defensin were retrieved in the NCBI Gene database. Each hit was accessed;intron part of the sequences has beencleaved out by using meta-data of Gen Bank format file, subsequences combined to form coding sequences (CDS) and downloaded. PROSITE and SMART were used for checking the presence and verification of defensin domain and signal sequence [15][16]. Two datasets in fasta format were prepared - one for signal propeptide and another for defensin domain sequences. 85 sequences from 25 mammalian species were used to prepare propeptide

dataset; while 101 sequences of 25 mammalian species were used to form defensin domain dataset (Table I).

B. Selection Analysis

Both datasets were aligned by codon using ClustalW algorithm in MEGA5.2. Datamonkey.org web server was used for evolutionary selection analysis by using following six models: FEL, SLAC, MEME, REL, TOGGLE and IFEL [17]. Before analysis, model selection was carried out. For each analysis, default parameter setting was kept and p = 0.1 was used as threshold.

C. Phylogenetic analysis

Both datasets were used to reconstruct phylogenetics using MEGA5.2. Neighbor-Joining method was used with bootstrap test of 1000 replication. Kimura 2-parametric model was used as substitution model with Gamma distributed rates among sites in both case. Both trees were visualized in Fig Tree software [18].

D. Sequence logo visualization

Sequence logo of defensin domain dataset was visualized using WebLogo web server (http://weblogo.berkeley.edu/logo.cgi) after translating in amino acid sequences to demonstrate the consensus amino acid at various positions of the sequence [20].

TABLE I. DESCRIPTION OF DATASET USED IN THIS STUDY

List of Species			
Propeptide dataset		**Defensin domain dataset**	
Aotusnancymaae	*Macacanemestrina*	*Aotusnancymaae*	*Musmusculus*
Chrysochlorisasiatica	*Microtusochrogaster*	*Cricetulusgriseus*	*Microtusochrogaster*
Colobusangolensis	*Macacafascicularis*	*Cercocebusatys*	*Microcebusmurinus*
Callithrixjacchus	*Microcebusmurinus*	*Colobusangolensis*	*Nomascusleucogenys*
Chlorocebussabaeus	*Musmusculus*	*Chrysochlorisasiatica*	*Pan troglodytes*
Cricetulusgriseus	*Nomascusleucogenys*	*Chlorocebussabaeus*	*Papioanubis*
Cercocebusatys	*Pan paniscus*	*Callithrixjacchus*	*Pongoabelii*
Dipodomysordii	*Pan troglodytes*	*Dipodomysordii*	*Pan paniscus*
Equuscaballus	*Papio Anubis*	*Equuscaballus*	*Peromyscusmaniculatus*
Gorilla gorilla	*Pongoabelii*	*Gorilla gorilla*	*Rattusnorvegicus*
Homo sapiens	*Peromyscusmaniculatus*	*Homo sapiens*	
Jaculusjaculus	*Rattusnorvegicus*	*Jaculusjaculus*	
Macacamulatta		*Macacamulatta*	
		Macacafascicularis	

E. Protein stability

Experimentally resolved crystal structure of human alpha defensin-6 (1ZMQ) was downloaded in pdb format from RCSB protein data bank (http://www.rcsb.org/pdb/home/home.do) which had 2.1 Å resolutions. This model was

used as reference structure on which different mutations were simulated and the stability effect of a mutation was empirically estimated using FoldX (version 4) [21]. Problem like steric clashes are frequently present in pdb structures. FoldX was used to fix those problems by lowering the global energy (ΔG). After repairing pdb, mutant defensin data-set was prepared by aligning and comparing defensin domain data-set sequences with 1ZMQ and registering change in amino acid residue. These mutations were performed in the A chain of 1ZMQ and free-energy change was simulated in FoldX. Also single mutation in defensin domain sequences was selected to observe their individual effect on free-energy change. In total 56 mutant sequences and 203 individual mutations were analyzed by FoldX. Histograms of free-energy change were generated and analyzed using R (version 3.0)[22].

F. **Homology modelling**

To predict and analyze structure and characteristic protein folding of α-defensin peptides Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) was used in intensive modeling mode with the default parameter of the web server, homology model of α-defensin peptides weregeneratedand visualized using PyMol software [19].

RESULTS

Selection Analysis: After model selection, general reversible model (REV) and Hasegawa, Kishino and Yano (HKY85) were found as the best substitution model for defensin domain dataset and leader propeptide dataset respectively. For defensin domain, FEL found 10 positively and 9 negatively selected codon. For propeptide domain, FEL found 5 positively and 13 negatively selected codon. For defensin domain, IFEL found 6 positively and 7 negatively selected codon. For propeptide domain, it found 5 positively and 12 negatively selected codon. For defensin domain, SLAC found 5 positively and 9 negatively selected codon. For propeptide domain, this method found 5 positively and 7 negatively selected codon. MEME found 10 sites with evidence of episodic diversifying selection on both datasets. In TOGGLE analysis, it was observed that toggling occured more on defensin domain comparatively leader propeptide domain. These results are summarized in table II.

Phylogenetic Analysis: Phylogenetic analysis reveals that in general, most species are clustered in the same clad. But in some cases, the occurrences of distinct species in multiple clad were observed like *Macacamulatta* and *Homo sapiens* (tree of propeptide not shown). This occurrence is particularly higher in defensin domain sequence set (Figure 1). In both cases, the α-defensin gene propeptide and defensin domain sequences did not maintain species-evolution pattern.

TABLE II. SELECTION ANALYSIS RESULTS ON DEFENSIN DOMAIN AND PROPEPTIDE DATASET

Method	Defensin Domain		Propeptide	
	+ve Selection	-ve Selection	+ve Selection	-ve Selection
FEL	7, 8, 9, 11, 12, 13, 17, 24, 26, 28	1, 4, 5, 10, 14, 18, 20, 33, 34	19, 33, 44, 48, 55	3, 4, 7, 11, 13, 16, 20, 23, 27, 38, 39, 54
IFEL	9, 12, 13, 17, 24, 28	1, 4, 5, 10, 14, 20, 34	19, 24, 33, 37, 55	3, 9, 11, 13, 16, 23, 25, 27, 31, 38, 39, 43
SLAC	7, 9, 11, 13, 19	1, 4, 5, 10, 14, 18, 20, 33, 34	11, 31, 42, 48, 55	3, 4, 7, 13, 16, 20, 24
MEME	**Evidence of Episodic Diversifying Selection**			
	7, 9, 11, 12, 13, 17, 19, 21, 24, 26		4, 19, 32, 33, 37, 42, 44, 55, 56, 61.	

Sequence Logo: In the defensin domain amino acid sequence logo, six cysteine amino acid residues were consensus at position 1, 4, 10, 20, 33, 34 which indicates the characteristics conserve amino acid sequence of α-defensin (Figure 2(a)). Besides, there are major consensus of glutamic acid and glycine at 14 and 18 positions respectively.

Protein Stability Simulation: Distribution of free-energy change in mutants indicates that overall free-energy change is positive with central tendency was around 9 kcal/mol (data not shown). It's to note that there were at least four mutations in those mutants. When the free-energy was normalized, the central tendency was around to 0.5 kcal/mol (fig 2(b)). Effect of individual mutation on defensin's stability has central tendency around 0 kcal/mol and ranges in both negative and positive direction (Figure 2(c)). Free energy change due to individual mutation in the positively selected sites was marked by vertical dashed-lines on the distribution of free energy change of single mutations (Figure 2(c)). This shows although some mutations are destabilizing (i.e.> 0 kcal/mol), most mutations in the positively selected sites are stabilizing (i.e.< 0 kcal/mol). There are 12 and 42 mutations in the five positively selected sites which are destabilizing and stabilizing respectively in the defensin domain data-set.

Homology Modeling: Thirty computational models of α-defensin structure were constructed. In the α-defensin peptides, the overall structure composed of 2-3 β-strands, where anti parallel β-sheet is dominant. But 3 predicted (*Macaca mulatta*-574196; *Mus musculus*-68009; *Equus caballus*-100307027) models did not contain any β-strands, although they contained α-helix coil. These models suggest that there are differences in the arrangement of disulfide bond and folding.

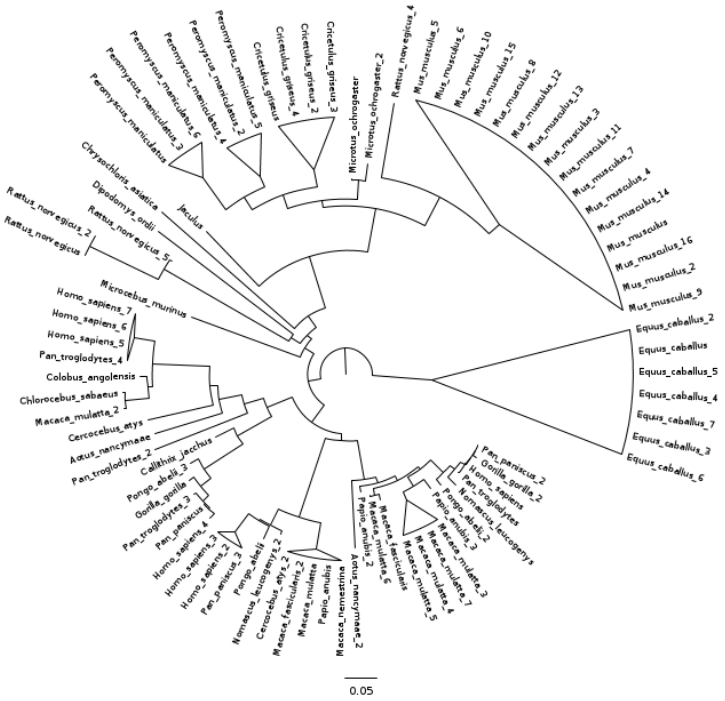
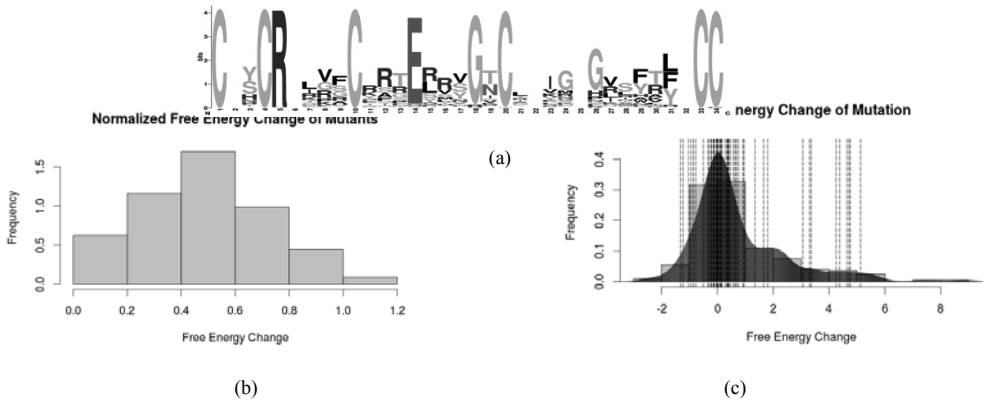


Fig. 1. Phylogenetic reconsturction of α –defensin defensin domain sequences.



(a)

(b)

(c)

Fig. 2. Amino acid sequence logo visualization of defensin domain sequences showing conserved and diverged base-positions (a), distribution of normalized free energy change of different mutant defensin sequences (b) and distribution of free energy change of various single mutations..

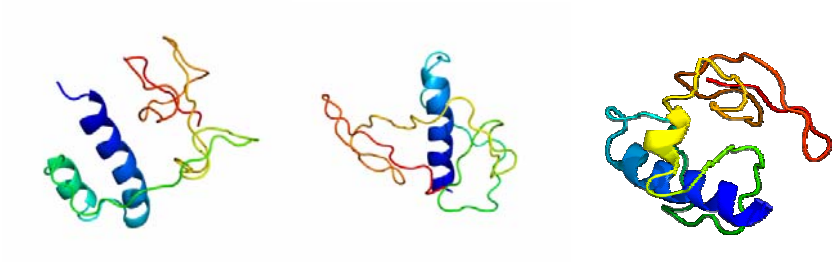


Fig. 3. Homology model of *Macaca mulatta*-574196; *Mus musculus*-68009; *Equus caballus*-100307027 has been showed respectively.

DISCUSSION

In general, positive selection on codon position 7, 9, 11, 13, 17 and negative selection on codon position 1, 4, 5, 10, 14, 18, 20, 34 was persistently reported by different methods on defensin domain dataset. On the other hand, positive selection found on codon 19, 33, 55 while negative selection on codon position 3, 4, 7, 13, 16, 20 was reported by these methods on propeptide domain dataset. Mixed effect selection was also reported on these positions. The implication of these results is that relatively positive selection pressure isgreater on defensin domain while negative selection is greater on propeptide sequences.

In the sequence logo of defensin dataset, the positively selected positions show diverged base-variance while negatively selected positions shows conserve nature.This phenomenon is also reflected on the reconstructed phylogenetic tree. In the tree based on propeptide sequences, same species are clusteredwithin a clad in general. But in the case of defensin domain dataset tree, interleaving nature of species between different clades was more frequent. In homology modeling study, structural variations of defensin protein were evident. In the distribution of free energy change due to single mutation, it was found that average mutation does not change stability of protein much. The distribution is positively skewed, that is some mutation has extreme effect on protein stability making the structure more non-stable. On the other hand, mutations causing negative free energy change stabilize the structure but they do not have extreme values. This suggests a random mutation itself hardly have enormous stabilizing effect and is a rare event. Most of the mutations (42 out of 54) on the positively selected sites were found stabilizing, which is expected. But some of the mutations have destabilizing effect on those sites, although this is not improbable. Evolutionary selection can positively select a codon position in a way such that the protein become more stable and its function get more efficient; but it's appears that some destabilizing mutations can happen in positively selected sites.

This study examined the effect of natural selection on α-defensin gene. The conclusion of the study is that although both positive and negative selection is acting on both propeptide and defensin domain of the gene, there is a dominant role of positive and negative selection on two distinct parts of gene – negative selection is dominant on propeptide and positive selection is dominant on defensin domain CDS. The explanation of this partial dominance can be hypothesized that function of propeptide is to carry premature

defensinpeptide to the specific location of a cell so that further maturation of it can take place, and this function, in general, did not change drastically throughout various species in different tissues. But defensin domain, which mainly takes part in host-pathogen interaction, has to face a wide range of circumstances in which pathogen is continuously changing and trying to evade the host immune system [23]. To cope with, defensin peptides need to change itself – which might be the cause behind the dominance of positive selection in this region. Further computational and experimental analysis is required to understand this observation.

References

[1] T. Ganz, "Defensins: antimicrobial peptides of innate immunity", Nat Rev Immunol, vol. 3, no. 9, pp. 710-720, 2003.

[2] T. Ganz, "Defensins and Other Antimicrobial Peptides: A Historical Perspective and an Update", Combinatorial Chemistry & High Throughput Screening, vol. 8, no. 3, pp. 209-217, 2005.

[3] D. Li, J. Tu, D. Li, Q. Li, L. Zhang, Q. Zhu, U. Gaur, X. Fan, H. Xu, Y. Yao, X. Zhao and M. Yang, "Molecular Evolutionary Analysis of β-Defensin Peptides in Vertebrates", Evolutionary Bioinformatics, p. 105, 2015.

[4] M. Boniotto, A. Tossi, M. DelPero, S. Sgubin, N. Antcheva, D. Santon, J. Masters and S. Crovella, "Evolution of the beta defensin 2 gene in primates", Genes and Immunity, vol. 4, no. 4, pp. 251-257, 2003.

[5] M. KIMURA, "Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution", Nature, vol. 267, no. 5608, pp. 275-276, 1977.

[6] M. KIMURA, "Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution", Nature, vol. 267, no. 5608, pp. 275-276, 1977.

[7] M. Kimura, "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences", J Mol Evol, vol. 16, no. 2, pp. 111-120, 1980.

[8] W. Delport, A. Poon, S. Frost and S. Kosakovsky Pond, "Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology", Bioinformatics, vol. 26, no. 19, pp. 2455-2457, 2010.

[9] S. Kosakovsky Pond, "Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection", Molecular Biology and Evolution, vol. 22, no. 5, pp. 1208-1222, 2005.

[10] S. Kosakovsky Pond, S. Frost, Z. Grossman, M. Gravenor, D. Richman and A. Brown, "Adaptation to Different Human Populations by HIV-1 Revealed by Codon-Based Analyses", PLoS Comp Biol, vol. 2, no. 6, p. e62, 2006.

[11] B. Murrell, J. Wertheim, S. Moola, T. Weighill, K. Scheffler and S. Kosakovsky Pond, "Detecting Individual Sites Subject to Episodic Diversifying Selection", PLoS Genetics, vol. 8, no. 7, p. e1002764, 2012.

[12] W. Delport, K. Scheffler and C. Seoighe, "Frequent Toggling between Alternative Amino Acids Is Driven by Selection in HIV-1", PLoS Pathog, vol. 4, no. 12, p. e1000242, 2008.

[13] Kortemme, T. and D. Baker. "A Simple Physical Model For Binding Energy Hot Spots In Protein-Protein Complexes". Proceedings of the National Academy of Sciences 99.22 (2002): 14116-14121.

[14] Christensen, Niels J. and Kasper P. Kepp. "Stability Mechanisms Of Laccase Isoforms Using A Modified Foldx Protocol Applicable To Widely Different Proteins". J. Chem. Theory Comput. 9.7 (2013): 3210-3223.

[15] C. Sigrist, E. de Castro, L. Cerutti, B. Cuche, N. Hulo, A. Bridge, L. Bougueleret and I. Xenarios, "New and continuing developments at PROSITE", Nucleic Acids Research, vol. 41, no. 1, pp. D344-D347, 2012.

[16] I. Letunic, T. Doerks and P. Bork, "SMART: recent updates, new developments and status in 2015", Nucleic Acids Research, vol. 43, no. 1, pp. D257-D260, 2014.

[17] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, "MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods", Molecular Biology and Evolution, vol. 28, no. 10, pp. 2731-2739, 2011.

[18] "FigTree", Tree.bio.ed.ac.uk, 2016. [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/. [Accessed: 01- Mar- 2016].

[19] L. Kelley, S. Mezulis, C. Yates, M. Wass and M. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis", Nat Protoc, vol. 10, no. 6, pp. 845-858, 2015.

[20] T. Schneider and R. Stephens, "Sequence logos: a new way to display consensus sequences", Nucl Acids Res, vol. 18, no. 20, pp. 6097-6100, 1990.

[21] Schymkowitz, J. W. H. et al. "Prediction Of Water And Metal Binding Sites And Their Affinities By Using The Fold-X Force Field". Proceedings of the National Academy of Sciences 102.29 (2005): 10147-10152. Web.

[22] Dobre, Gabriela-Roxana. "R Language: Statistical Computing And Graphics For Modeling Hydrologic Time Series". Mathematical Modelling in Civil Engineering 10.4 (2014): n. pag. web.

[23] P. Raj and A. Dentino, "Current status of defensins and their role in innate and adaptive immunity", FEMS Microbiology Letters, vol. 206, no. 1, pp. 9-18, 2002.