ECRONICON OPEN ACCESS

EC PROTEOMICS AND BIOINFORMATICS
Research Article

Structure, Function and Sub-Cellular Localization of Hypothetical Proteins: An *In Silico* Study in *Candidatus Riesia pediculicola* Plasmid

Hitesh S Thakare[1,2]*, Satish Sawale[1], Dilip B Meshram[3], Kunal Roychoudhary[2] and Arun B Ingle[2]

[1]*National Environmental Engineering Research Institute, Nagpur, Maharashtra, India*

[2]*Department of Microbiology, Seth Kesarimal Porwal College, Kamptee, Maharashtra, India*

[3]*Department of Zoology, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur, Maharashtra, India*

***Corresponding Author**: Hitesh S Thakare, National Environmental Engineering Research Institute, Nagpur and Department of Microbiology, Seth Kesarimal Porwal College, Kamptee, Maharashtra, India.

Received: September 01, 2017; **Published:** September 19, 2017

Abstract

In the present *in silico* study, based on the structure, function and sub-cellular localization of hypothetical proteins in *Candidatus Riesia pediculicola* plasmid were explained. For the determination of functional annotations or functionality the CDD-BLAST, INTERPROSCAN and PFAM were used. The PS2 Server-Protein Structure Prediction server and Cello were used for understanding and identified the presence of templates for conserved domains and sub- cellular localizations within the cell. Also, for the determination of 3-D structures, E-value and aligned percentage of the predicted hypothetical proteins PS2 Server was used. There are total 8 genes were screened for understanding the structures, functions and sub-cellular localization of proteins in *Candidatus Riesia pediculicola* plasmid out of these 4 genes were unknown for their structures, functions and sub-cellular localization which was predicted for hypothetical proteins. This study may be useful for understanding the role of bacterium life cycle by characterizing structure and functionality as well as genetics and metabolic pathways at the molecular level.

Keywords: *Endosymbionts; Pediculus; PS2 Server; Initiator Protein; Plasmid*

Introduction

Over 500 species of Sucking lice were described well in the world. The Sucking lice are ectoparasites of mammals and they take mammalian blood as a nutrient diet [1]. There are two closely related species of human lice viz. body louse (*Pediculus humanus*) and the head louse (*Pediculus capitis*). These are morphologically and genetically similar. The *P. humanus* lives in clothes and feeds from the body and *P. capitis* lives in the hair and feeds the scalp [2].

A small number of primary endosymbionts (P-endosymbionts) of about 14,000 predictable species of hematophagous insects have been described. It is observed that nutrient-poor diet of P-endosymbionts is blood in which all sucking lice to be expected that have obligated primary endosymbionts [3].

The primary endosymbionts (P-endosymbionts) were first seen microscopically in the human head and body louse over 340 years ago [4]. Most of the primary endosymbionts drift to the ovaries after leaving their mycetomes so that they could be united into developing eggs (transovarial transmission). Due to this, they can be inherent in the host generation [5,6], for long-term. During transovarial transmission, the coevolutionary history between the insects and their symbionts may be pooled [3].

The *Candidatus Riesia pediculicola* is an endosymbiont of requisite louse having short, linear chromosome and a circular plasmid encodes less than 600 genes. The plasmid harbors is an essential and a sole of genes for the synthesis of pantothenate, an essential vitamin in

the louse diet deficient [7]. *Candidatus Riesia pediculicola* is the P-endosymbiont present in the mycetomes of *P. humanus* [2,6]. As per the coevolutionary history, it is divided into three parts. The P-endosymbionts are parasitic lice found in primates like humans, chimpanzees and gorillas having genus *Candidatus Riesia* [2,3,7]. In recent literature, the P-endosymbionts of chimpanzee lice (*P. schaeffi*) and human pubic lice (*Pthirus pubis*) were characterized at the molecular level [2,3].

Methodology

Sequence Retrieval

The KEGG database was used for retrieval of *Candidatus Riesia pediculicola* plasmid whole genome sequences (http://www.genome.jp/kegg/).

Functional Annotation and Categorization

The bioinformatics web tools were used to screened and analyzed the presence of functional domains and 3-D structures of the hypothetical proteins in *Candidatus Riesia pediculicola plasmid*. The CDD-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/; [8-11]), INTERPROSCAN (http://www.abi.ac.uk/interpro; [12]), Pfam (http://www.pfam.sanger.ac.uk/; [13]) and Cello (http://cello.life.nctu.edu.tw/) were used as online bioinformatics web servers. For the presence of conserved domains and functionality in the genome sequences were identified as per the information existing in CDD-Blast, Interproscan, and Pfam databases. Determination of the sub-cellular localization of proteins or enzymes present in the cell was identified by using the Cello server v.2.5.

Protein Structure Prediction

The 3-D structure of unknown proteins predicted by using PS2: Protein Structure Prediction Server (http://www.ps2.life.nctu.edu.tw/; [11,14,15]). The prediction of protein 3D structures was created by running the FASTA format of protein sequence via online web server also the detection of templates having the structural model of the protein is based on the functional annotations.

Results and Discussion

For characterization of 8 hypothetical proteins from the complete genome sequences of *Candidatus Riesia pediculicola* plasmid, the computational studies were carried out. The predictions of structure and function characteristics of total 4 hypothetical proteins were carried out by using CDD- Blast, Interproscan, Pfam, Cello, and PS2server. Also, the Sub-cellular localization of all the hypothetical proteins in *Candidatus Riesia pediculicola* plasmid was characterized successfully. The PS2 structure template for 3-D structures is depicted in the order as Template ID, E-value and aligned percentage in the following table.

We have successfully characterized probable functions of gene products by using CDD-Blast, Interproscan, and Pfam which was found to be 1, 2 and 1 respectively. Out of 4 screened hypothetical proteins, only single 3D structure prediction template was successfully characterized using PS2 online bioinformatics web server.

After the screening of the NCBI gene ID 502799949, it was found that transmembrane protein phobius with a membrane-bound protein predicted to be embedded in the membrane but there is no template formation. Another NCBI gene ID 908689879 explains the predicted 3D structure (Figure 1) and the presence of Replication initiator protein A; Members of this family of bacterial proteins are single-stranded DNA binding proteins that are involved in DNA replication, repair, and recombination (Table 1). The predicted 3D structure is asymmetric and stoichiometrically monomeric. It has one and two unique protein and nucleic acid chains respectively (http//www.rcsb.org/pdb/explore/).

The remaining two NCBI genes IDs 502799947 and 502799952 have the sub-cellular localization within the cell but it does not explain any structural and functional characterization (Table 1).

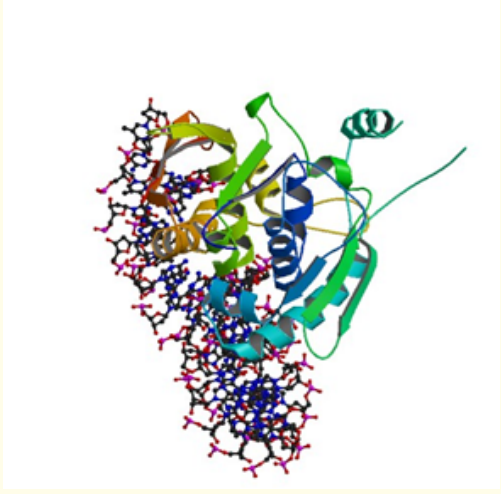

Figure 1: *3D structure of NCBI gene ID No. 908689879.*

NCBI Gene ID	CDD BLAST	Interproscan	Pfam	Cello	PS2: Structure Prediction Server		
					Template	E-Value	Aligned Percentage
502799947	NA	NA	NA	Cytoplasmic 2.307	NA	NA	NA
502799949	NA	Phobius Transmembrane Region of a membrane-bound protein predicted to be embedded in the membrane. Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	NA	Cytoplasmic 2.457	NA	NA	NA
502799952	NA	NA	NA	Cytoplasmic 2.241	NA	NA	NA
908689879	Replication initiator protein A; Members of this family of bacterial proteins are single-stranded DNA binding proteins that are involved in DNA replication, repair and recombination.	Replication initiator protein A	Replication initiator protein A	Cytoplasmic 3.608	2nraC	0.16	72.79

Table 1: *Structural, functional and sub-cellular localizations of hypothetical proteins in Candidatus Riesia Pediculicola plasmid*

NA: Not Available

Conclusion

The present study revealed that total 4 functionally and structurally significant hypothetical proteins from *Candidatus Riesia pediculicola* plasmid has been sorted successfully. It is observed that *Candidatus Riesia* pediculicola plasmid has very few possible functional proteins. Out of 8 NCBI genes, a total of 4 hypothetical proteins are successfully characterized for understanding the structure as well as functions using CDD- Blast, Interproscan, Pfam, Cello and PS2 server. The life cycle of the bacterium can be supporting in establish their role by characterized predicted functions and three-dimensional structures. This study may be useful for understanding the metabolic pathways and genetics at the molecular level.

Acknowledgments

The authors are thankful to Dr. Swapnil Sanmukh, a postgraduate researcher at Sau Paulo state university, Brazil, for his grateful help and support during the present study. The authors thank Principal and Head, Department of Microbiology, S. K. Porwal College, Kamptee for valuable guidance. The continuous support and valuable guidance from Scientist and Head, Water Technology and Management Division, CSIR-NEERI, Nehru Marg, Nagpur are highly acknowledged.

Bibliography

1. Durden La and G Musser. "The Sucking Lice (Insecta, Anoplura) of the World - a Taxonomic Checklist with Records of Mammalian Hosts and Geographical Distributions". *Bulletin of the American Museum of Natural History* (1994).

2. Sasaki-Fukatsu Kayoko., *et al*. "Symbiotic Bacteria Associated with Stomach Discs of Human Lice". *Applied and Environmental Microbiology* 72.11 (2006): 7349-7352.

3. Allen Julie M., *et al*. "Evolutionary Relationships of 'Candidatus Riesia Spp.,' Endosymbiotic Enterobacteriaceae Living within Hematophagous Primate Lice". *Applied and Environmental Microbiology* 73.5 (2007): 1659-1664.

4. Hooke Robert. "Micrographia, Or, Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses: with Observations and Inquiries Thereupon /by R. Hooke". Council of the Royal Society of London for Improving Natural Knowledge London United Kingdom (1665).

5. Douglas AE. "Mycetocyte Symbiosis in Insects". *Biological Reviews of the Cambridge Philosophical Society* 64.4 (1989): 409-434.

6. Perotti M Alejandra., *et al*. "Rickettsia as Obligate and Mycetomic Bacteria". *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 20.13 (2006): 2372-2374.

7. Kirkness Ewen F., *et al*. "Genome Sequences of the Human Body Louse and Its Primary Endosymbiont Provide Insights into the Permanent Parasitic Lifestyle". *Proceedings of the National Academy of Sciences of the United States of America* 107.27 (2010): 12168-12173.

8. Schäffer Alejandro A., *et al*. "Improving the Accuracy of PSI-BLAST Protein Database Searches with Composition-Based Statistics and Other Refinements". *Nucleic Acids Research* 29.14 (2001): 2994-3005.

9. Altschul Sf., *et al*. "Gapped BLAST and PSI- BLAST: A New Generation of Protein Database Search Programs". *Nucleic Acids Research* 25.17 (1997): 3389-3402.

10. Marchler-Bauer Aron., *et al*. "CDD: A Conserved Domain Database for Interactive Domain Family Analysis". *Nucleic Acids Research* 35.1 (2007): D237-D240.

11. Notredame C., *et al*. "T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment". *Journal of Molecular Biology* 302.1 (2000): 205-217.

12. Zdobnov EM and R Apweiler. "InterProScan--an Integration Platform for the Signature-Recognition Methods in InterPro". *Bioinformatics* 17.9 (2001): 847-848.

13. Bateman Alex., *et al*. "@Pfam@The Pfam Protein Families Database". *Nucleic Acids Research* 32 (2004): D138-D141.

14. Aydin Zafer., *et al*. "Protein Secondary Structure Prediction for a Single-Sequence Using Hidden Semi-Markov Models". *BMC Bioinformatics* 7 (2006): 178.

15. Siddiqui Azeem., *et al*. "Genome Annotation and Structure Predictions for Hypothetical Proteins in Agrobacterium Fabrum Str. C58 Plasmid At". *International Journal of Computer Applications* 85.1 (2014): 22-24.

Volume 2 Issue 1 September 2017