BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis
Volume 8(19)

Niche specific amino acid features within the core genes of the genus *Shewanella*

Rachana Banerjee* & Subhasis Mukhopadhyay

Department of Biophysics, Molecular Biology and Bioinformatics and Distributed Information Centre for Bioinformatics, University of Calcutta, 92, A.P. C. Road Calcutta-700009; Rachana Banerjee – Email: rachanabanerjee23@yahoo.co.in; Phone: +91-9830136799; *Corresponding author

Received September 15, 2012; Accepted September 19, 2012; Published October 01, 2012

Abstract:
Shewanella species are found to dwell in various ecological niches. The widespread habitation where they live requires specific adaptations. Recent advances in genomic approaches, such as in sequencing technologies, generate huge amount of genomic data that lend support towards understanding the microbial evolution and diversity through comparative study. In this manuscript, we discuss a comparative analysis of core genes of phylogenetically related twelve members from the genus *Shewanella*. Phylogenetic analysis based on the core genes, differentiated two subgroups of the genus, one group comprises of species characterized as high-pressure cold-adapted while the other group is characterized as mesophilic pressure-sensitive species. By analyzing the differences of amino acid composition of these two groups, we have identified the specific trend of amino acid usage that has been adopted by the psychro-peizo-tolerant *Shewanella* species. The functional categories have also been recognized which are responsible for rendering the particular amino acid compositional pattern in psychropeizophilic *Shewanella* species facilitating their niche adaptation.

Keywords: *Shewanella sp*, core gene, amino acid composition, phylogenetic tree

Background:

The increasing number of complete genomic sequences has unwrapped loads of ways to interpret the adaptation of genomes in their respective niche in relation to their structure, function etc. **[1]**. An estimation of amino acid preferences of different organisms is now possible through comparative genomic study **[2]**. Comparative genomics helps to determine the genes that are present across bacterial genomes of the same species (or genus) known as conserved (core) genes and disparity within these core genes is an indication that can be used to unravel highly variable and conserved genes within the genomes of interest. The conserved genes usually evolve more slowly; therefore, they will play a role to infer molecular evolution of the genomes **[3]**. Moreover, molecular phylogenetic trees built using the sequences of core genes can also determine the relationship among the genomes.

Shewanella is a genus from which adequate number of species had already been sequenced facilitating detection of core genes. Different species belonging to the genus *Shewanella* are capable of inhabiting many aquatic (from fresh water to deep sea) and sedimentary ecosystems under aerobic as well as anaerobic conditions **[4]**. Some of the *Shewanella* species, for example *Shewanella violacea,* had been isolated from cold environments, such as seawater in Antarctica or in the North Sea, implying that they are not only peizophilic (can breed better under high hydrostatic pressure conditions), but also psychrophilic (needs low temperatures, ranging from −15°C to +10°C for growth and breeding). They can be defined as psychro-peizo-tolerant *Shewanella* species.

It is known that pressure influences protein structure and folding **[5].** Functional properties of proteins might get affected by pressure range that is experienced by organism **[6]**. In this study, we discuss the comparative analysis of the core genes of phylogenetically related twelve members from the genus *Shewanella* and focus on the recognition of a common trend towards a certain pattern of amino acid usage of the psychropeizo-tolerant *Shewanella* species.

Methodology:

We have downloaded the fully annotated complete nucleotide sequences of all the twelve *Shewanella* genomes, considered in the present study, from NCBI FTP site (www.ncbi.nlm.nih.gov/Ftp/). Details of the 12 bacteria are listed in **Table 2 (see supplementary material).**

Identification of core genes

Highly similar paralogous sequences have been removed from each of the twelve set of genome sequences, based on a comparison of the gene list against itself (with identity >=90%) using Blastclust program (www.ncbi.nlm.nih.gov/Web/Newsltr/Spring04/blastlab.html). From the rest of the genes, core genes were detected through putative one-to-one orthologous gene identification, using BLASTP (>= 85% identity and 0% gap). This search identified a total of 121 orthologous groups that had all twelve species represented, which were used for further analysis.

Phylogenetic analysis

All core genes from 12 *Shewanella* genomes were used for generating a neighbor joining tree. For each twelve species, the amino acid sequences of 121 core gene sets were concatenated in order to produce twelve continuous sequences of average length 34,746 amino acids. Multiple Alignments of these 12 sets of concatenated core amino acid sequences were performed using CLUSTALW (http://www.ebi.ac.uk/clustalw/). A Neighbor-joining Phylogenetic tree based on these concatenated core amino acid sequences from the *Shewanella* species of interest, was generated using MEGA 5 **[7]**.The bootstrap values are shown in the Phylogenetic tree.

Estimation of amino acid composition

We have computed the frequencies of amino acid residues in the protein sequences of two groups of *Shewanella* sp., as depicted in the Phylogenetic tree. To compare the means of two groups of data, t-test is essentially a good tool. The t-values are calculated as follows:

$$\mathbf{t = (F_{Group1} - F_{Group2}) / \sqrt{[(Var_{Group1}/ n_{Group1}) + (Var_{Group2}/ n_{Group2})]}}$$

where, Var_{Group1} and Var_{Group2} are the variance of amino acid residues. F_{Group1} and F_{Group2} are mean frequencies of Group1 and Group2 proteomes respectively. The n_{Group1} and n_{Group2} are the total number of Group1 and Group2 proteins investigated in this study. Based on student's t-distribution table of significance, critical values for such t-test at various probabilities are as follows:

Probability	10%	5%	2.5%	2%	1%	0.5%
Critical Value	1.372	1.812	2.228	2.359	2.764	3.169

If t-value is positive and greater than critical value at 10% probability (1.372), then the mean frequency (F_{Group1}) of Group1 core proteins are significantly greater than that of the Group2 core proteins (F_{Group2}) at 90% or higher confidence level. If the frequency of residue or property group t-value is negative and less than -1.372, then the mean frequency of Group1 core proteins are (F_{Group1}) is significantly less than that of Group2 core proteins (F_{Group2}) at 90% or higher confidence level **[8-10]**.

Functional classification

The core proteins were functionally classified according to Clusters of Orthologous Groups (COGs) of proteins categories. BLAST homology study had been carried out against the COG database. Proteins that were classified in two COG categories were registered in both categories.

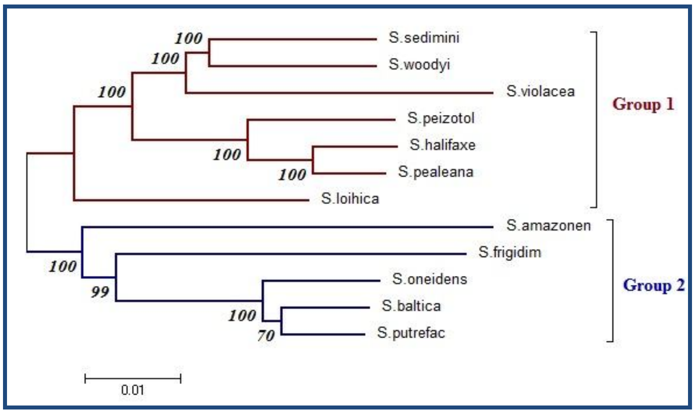

Figure 1: Phylogenetic tree constructed using core genes of 12 Shewanella species.Group1 consists of psychro-peizo-tolerant *Shewanella* species and Group2 contains mesophilic pressure-sensitive *Shewanella* species. Bootstrap values for all the branches are mentioned in the figure.

Results and Discussion:

A final list containing a total of 121 core gene sets from all the twelve *Shewanella* species, were used for further analyses. Alignment of these 12 sets of concatenated core amino acid sequences using CLUSTALW (http://www.ebi.ac.uk/clustalw/) reveals 17% non-conserved amino acid sites. Neighbor-joining phylogenetic tree based on the concatenated core amino acid sequences from the *Shewanella* species of interest, was constructed using MEGA 5, which separates 12 *Shewanella* species into two groups i.e., Group 1 and Group 2 **(Figure 1)**. Literature search describes Group 1 members of *Shewanella* as cold adapted species that grow at high pressure, while Group2 members are mesophilic, pressure sensitive species **[11]**.

Amino acid composition preferences

We have analyzed the amino acid composition of 121 core gene sequence sets for each of the twelve *Shewanella* species. The frequencies of individual amino acids were further analyzed using student t-test, which shows that a few of the amino acid differed significantly in the core gene sets of *Shewanella* species present in Group1 when compared to the core gene sets of *Shewanella* species present in Group 2 **Table 1 (see supplementary material)**. As specified by the t-value, amino acid residues A, D, S, N and C, are significantly preferred (marked in bold red in **Table 1**), while G, T, V, M, I are moderately preferred and residues R, Y, E, P and L are significantly avoided (marked in blue in **Table 1**) by Group 1 *Shewanella* species compared to Group 2. Residues S and D are helix breakers and residue E favors the formation of helical structure. On the other hand, presence of residue L stabilizes the helical conformations **[12]**. It is known that amino acid substitutions that diminish protein flexibility and compressibility results in an increase of stability of the protein at high pressure **[13]**. In addition, helix destabilizing beta branch residues I, T and V are preferred by peizophilic proteins **[14]**. All the above results signify that increased structural homogeneity is perhaps favored by the high-pressure environments of the deep-sea, which is attained by favoring a decreased number of helix breaking and helix destabilizing residues. Considering the amino acid properties, Group1 *Shewanella* members most significantly, favor the amino acid residues S, D, and N, all of which are strongly polar as well as have very low molecular weight. Interestingly, the hydrostatic pressure asymmetry index is positively correlated with the polarity of amino acids and inversely correlated with molecular weight of amino acids **[15]**.

Thus, the amino acid composition of the core genes of the twelve species of *Shewanella* considered in our study shows a particular trend, which points towards a strong favor of polar and small amino acids with adequate propensity of breaking and destabilizing the helical structure, for the Group 1 members of *Shewanella* **(Figure 1)** sustaining their psychropeizophilic adaptation for residing in deep sea environment, while an opposite amino acid compositional trend is featured by the Group 2 members of *Shewanella* **(Figure 1)** supporting their mesophilic, pressure sensitive characteristics.

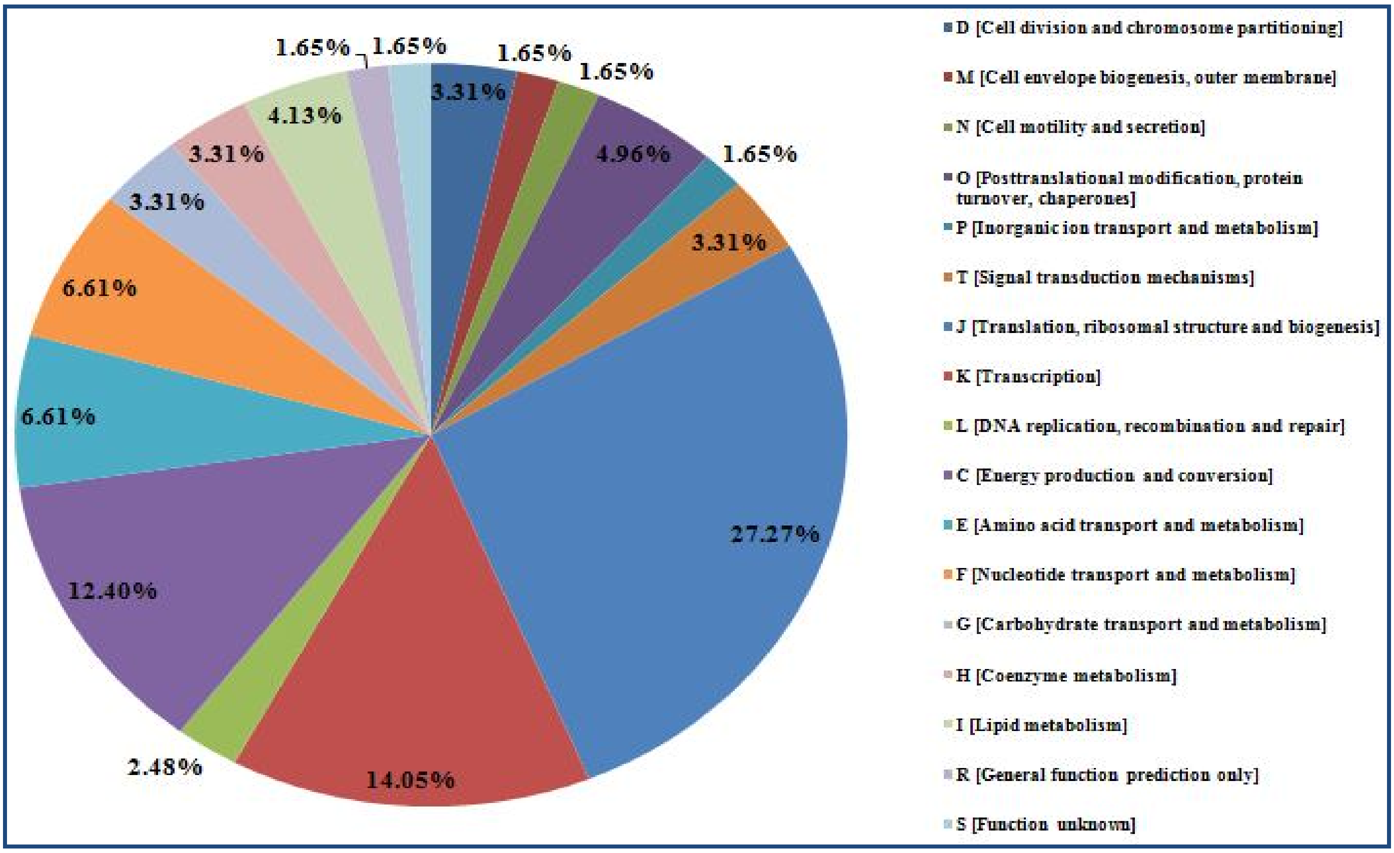

Figure 2: Distribution of COG classification of 121 core genes in twelve *Shewanella* species considered in this study**.**

Functional Classification

Core genes considered in our study can be divided into two categories depending on the average percent identity of the blast score. (a) Core gene showing low variation (54% of the total core genes with identity >95%); (b) Core gene that are highly variable (46% of the total core genes with identity <95%). Functional profiles of the core genes of the twelve *Shewanella* species considered in our study have been determined based on COG categories. It has been found that higher proportion of the core genes (44%) are involved in information storage and processing, 62% of which are involved in translation, ribosomal structure and biogenesis **(Figure 2).**

Variable core genes as well as conserved core genes are classified according to COG functional categories and compared. The comparison points to the fact that functional categories like “Translation, ribosomal structure and biogenesis”, “Carbohydrate transport and metabolism” and “Amino acid transport and metabolism” are more common in conserved core genes **(Figure 3)**. On the other hand, functional categories like “Energy production and conversion”, “Cell envelope biogenesis” and “Cell motility and secretion” are more common in variable core genes.

Core genes with low variation experience selection against mutations that leads to amino acid changes. But in case of highly variable core genes, positive selection for amino acid changes takes place **[16]**. Consequently, three functional categories which are more common in variable core genes are accountable for the specific amino acid compositional trend in psychropeizophilic *Shewanella* species (members of Group 1), favoring polar and small amino acids with helix breaking and destabilizing property.

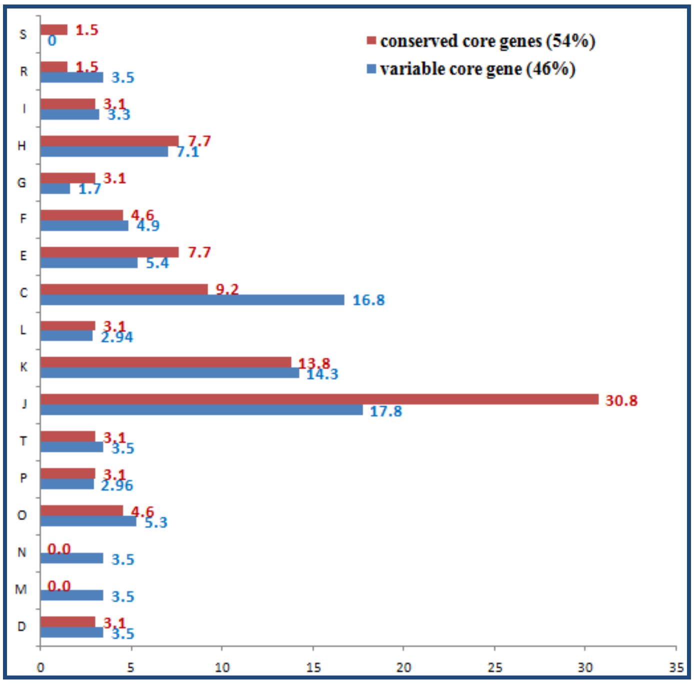

Figure 3: COG classification terms for conserved core genes (red bars) and highly variable core genes (blue bars). Functional classes like Translation, ribosomal structure and biogenesis (J), Carbohydrate transport and metabolism (G), Amino acid transport and metabolism (E), are more common in conserved core genes. Variable core genes have a preference over functional categories like Energy production and conversion (C), Cell envelope biogenesis (M) and Cell motility and secretion (N).

Conclusion:

Phylogenetic study based on the concatenated core amino acid sequences of twelve *Shewanella* species separated two distinct groups of *Shewanella*. Group 1 comprises of psychropeizophilic *Shewanella* species, whereas Group2 members are mesophilic, pressure sensitive species. Our studies on the composition of individual amino acid residues within these two groups of bacteria revealed that the psychropeizophilic *Shewanella* species show a specific trend of amino acid usage that favors the increase in frequency of strongly polar, small and tiny amino acids having the potential of avoiding helical structures. Amino acid residues S, D and N are mostly preferred by them. Functional profiles of the core genes of the twelve *Shewanella* species show that information storage and processing represents higher proportion of the core genes, with 62% of them involved in translation, ribosomal structure and biogenesis, signifying the importance of these two functional categories in maintaining the most important cellular processes of the genus *Shewanella*. We have also divided the core genes into two types: conserved core genes and variable core genes, which are evaluated on the basis of COG functional classification. Three functional categories (Cell envelope biogenesis, Cell motility and secretion, Energy production and conversion) which are quite more common in variable core genes are responsible for displaying the specific amino acid compositional trend in psychropeizophilic *Shewanella* species (members of Group 1). Thus, they seem to have played an important role for niche adaptation of the twelve *Shewanella* species considered in our study.

Acknowledgement:

We wish to thank Department of Biotechnology, Government of India, for providing financial support through two project grants, (BT/BI/04/026/93 and BT/BI/010/019/99).

References:

[1] Medigue C *et al. Genome Res.* 2005 **15**: 1325 [PMID: 16169927]
[2] Saunders NF *et al. Genome Res.* 2003 **13**: 1580 [PMID: 12805271]
[3] Urwin R *et al. Trends Microbiol.* 2003 **11**: 479 [PMID: 14557031]
[4] Venkateswaran K *et al. Int J Syst Bacteriol.* 1999 **49**: 705 [PMID: 10319494]
[5] Gross M & Jaenicke R, *Eur J Biochem*. 1999 **221**: 617 [PMID: 8174542]
[6] Li H *et al. J Agric Food Chem*. 2011 **59**: 12028 [PMID: 21981038]
[7] Tamura K *et al. Mol Biol Evol.* 2011 **28**: 2731 [PMID: 21546353]
[8] Jahandideh S *et al. J Theor Biol*. 2007 **246**: 159 [PMID: 17275036]
[9] Pack SP & Yoo YJ, *J Biotechnol.* 2004 **111**: 269 [PMID: 15246663]
[10] Vogt G *et al. J Mol Biol*. 1995 **249**: 816 [PMID: 7602593]
[11] Kato C & Nogi Y, *FEMS Microbiol Ecol.* 2001 **35**: 223 [PMID: 11311432]
[12] Lewis PN *et al. Proc Natl Acad Sci U S A*. 1970 **65**: 810 [PMID: 5266152]
[13] Royer CA *et al. Biochemistry.* 1993 **32**: 5222 [PMID: 8494899]
[14] Kumar S *et al. Protein Eng.* 2000 **13**: 179 [PMID: 10775659]
[15] Simonato F *et al. J Biotechnol.* 2006 **126**: 11 [PMID: 16780980]
[16] Leekitcharoenphon P *et al. BMC Genomics.* 2012 **13**: 88 [PMID: 22409488]

Edited by P Kangueane
Citation: Banerjee & Mukhopadhyay, Bioinformation 8(19): 938-942 (2012)

Supplementary material:

Table 1: The composition of individual amino acids in protein sequences of core genes of Group1 and Group2 *Shewanella* sp. Groups are referred as depicted by the phylogenetic tree in Figure 1. Significantly preferred and avoided amino acids by Group1 members of *Shewanella*, as indicated by the t-test parameters are marked in bold red and bold blue respectively.

	Group 1		**Group 2**		
Amio acids	**Average**	**Variance**	**Average**	**Variance**	**t-test**
PHE [F]	3.4782	0.0246	3.4842	0.0246	-0.296
LEU [L]	8.4929	0.0851	8.5791	0.0861	**-2.290**
ILE [I]	6.4327	0.0343	6.4160	0.0361	0.691
MET [M]	3.1395	0.0190	3.1251	0.0217	0.785
VAL [V]	8.4373	0.0523	8.4300	0.0531	0.250
SER [S]	5.3233	0.0336	5.1677	0.0321	**6.676**
PRO [P]	3.6378	0.0173	3.6707	0.0171	**-1.951**
THR [T]	5.4971	0.0294	5.4715	0.0268	1.186
ALA [A]	9.2356	0.0719	9.1881	0.0692	**1.392**
TYR [Y]	2.2622	0.0179	2.3004	0.0176	**-2.229**
HIS [H]	2.0344	0.0152	2.0407	0.0149	-0.397
GLN [Q]	3.4405	0.0223	3.4582	0.0241	-0.904
ASN [N]	3.6560	0.0201	3.6169	0.0196	**2.158**
LYS [K]	6.6604	0.0942	6.6697	0.0942	-0.237
ASP [D]	5.1047	0.0412	5.0069	0.0404	**3.766**
GLU [E]	7.1301	0.0757	7.2304	0.0778	**-2.817**
CYS [C]	0.8945	0.0063	0.8728	0.0064	**2.131**
TRP [W]	0.5511	0.0042	0.5519	0.0043	-0.092
ARG [R]	6.5165	0.1267	6.4226	0.1277	**2.048**
GLY [G]	8.2383	0.0681	8.2164	0.0703	0.648

Table 2: Information about *Shewanella* strains used in this study

Genome Name	**NCBI Taxon ID**	**RefSeq Project ID**	**GenBank Project ID**	**Genome Size**	**CDS Count**
Shewanella amazonensis SB2B	326297	58257	13385	4306142	3645
Shewanella baltica OS223	407976	58775	17985	5358884	4482
Shewanella frigidimarina NCIMB 400	318167	58265	13391	4845257	4066
Shewanella halifaxensis HAW-EB4	458817	59007	20241	5226917	4306
Shewanella loihica PV-4	323850	58349	13906	4602594	3870
Shewanella oneidensis MR-1	211586	57949	335	5131408	4502
Shewanella pealeana ANG-SQ1, ATCC 700345	398579	58705	17415	5174581	4254
Shewanella piezotolerans WP3	225849	58745	17675	5396476	4933
Shewanella putrefaciens CN-32	319224	58267	13393	4659220	4005
Shewanella sediminis HAW-EB3	425104	58835	18789	5517674	4501
Shewanella violacea DSS12	637905	47085	34739	4962103	4346
Shewanella woodyi MS32, ATCC 51908	392500	58721	17455	5935403	4923