Phylogenetic analysis based evolutionary study of 16S rRNA in known *Pseudomonas sp*

Arindam Adhikari[1,2] , Suvodip Nandi[2], Indrabrata Bhattacharya[3], Mithu De Roy[4], Tanusri Mandal[1] & Subrata Dutta[3]*

[1]Department of Biotechnology, O.I.S.T. Vidyasagar University, Paschim Medinipur, India; [2]AICRP on Vegetables crops, Directorate of research, BCKV, Kalyani, India; [3]Department of Plant Pathology, BCKV, Mohanpur, Nadia, India; [4]Soil Testing Laboratory, PORS, Berhampore, Murshidabad, India; Subrata Dutta – Email: subrata_mithu@yahoo.co.in; *Corresponding author

Abstract:
Molecular evolution analysis of 16S rRNA sequences of native *Pseudomonas* strains and different fluorescent pseudomonads were conducted on the basis of Molecular Evolutionary Genetics Analysis version 5.2 (MEGA5.2). Topological evaluations show common origin for native strains with other known strains with available sequences at GenBank database. Phylogenetic affiliation of different *Pseudomonas sp* based on 16S rRNA gene shows that molecular divergence contributes to the genetic diversity of *Pseudomonas sp*. Result indicate direct dynamic interactions with the rhizospheric pathogenic microbial community. The selection pressure acting on 16S rRNA gene was related to the nucleotide diversity of *Pseudomonas sp* in soil rhizosphere community among different agricultural crops. Besides, nucleotide diversity among the whole population was very low and tajima test statistic value (D) was also slightly positive (Tajima's test statistics D value 0.351). This data indicated increasing trends of infection of soil-borne pathogens under gangetic-alluvial regions of West Bengal due to high degree of nucleotide diversity with decreased population of plant growth promoting rhizobacteria like fluorescent Pseudomonads in soil.

Key Words: Molecular diversity, Positive selection, 16S rRNA, MEGA5.2, *Pseudomonas sp.*

Background:
To determine the distribution of genotypes among the known species of fluorescent pseudomonads, the phylogenetic analysis was initially run with the sequences of 16S rDNA. Understanding of molecular evolution was needed to analyze phylogenetics based on 16S rDNA analysis. Several workers had been worked on phylognetic study of *Pseudomonas sp.* [1-3]. In the phylogenetic analysis research reported so far, identification of the primary phylogenetic groups based on their 16S rDNA sequences, and to compare the relative diversity were documented [4, 5]. There are now a large number of phylotypes that can be distinguished based on their 16S rDNA sequences but detection of purification or diversifying selection of nucleotide of 16S rDNA sequences were not yet studied in detail. Two types selection pressure was occurring in molecular evolution. Negative selection in 16S rDNA sequences might have decreased the population of rhizobacteria, while there had been particular interest in identifying positive selection because it provides evidence for adaptive changes in function. The statistics used to identify positive selection in these studies is the nonsynonymous to synonymous substitution rate ratio, (dN/dS). An excess of nonsynonymous substitutions (dN) over synonymous substitutions (dS) is an unambiguous indicator of positive natural selection at the molecular level [6, 7]. The novel objective of this study was to characterize effect of positive or diversifying selection on 16S rRNA gene of different soil rhizospher.

Methodology:
DNA sequencing of 16S rDNA region of Pseudomonas sp and sequence analysis
The sequence reaction was conducted using the Sanger dideoxy sequencing Sequencing kit (Applied Biosystems, Foster City) following the manufacturer's instructions. The trimmed nucleotide sequences of different *Pseudomonas sp* isolates were first locally aligned using Basic Local Alignment Search Tool (BLAST) algorithm provided by National Centre for Biotechnology Information (NCBI). 16S rDNA region of *Pseudomonas sp* isolates were deposited to the NCBI nucleotide

data base through GenBank Bankit submission tools **Table 1 (see supplementary material).**
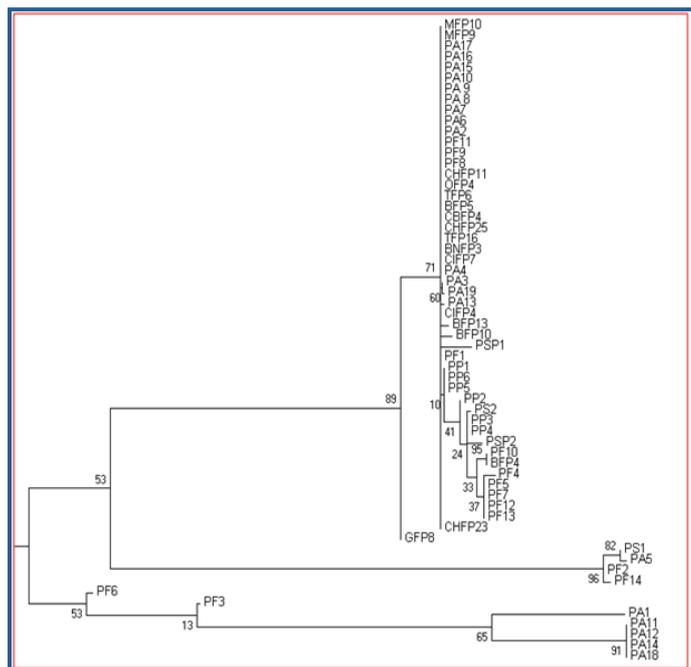

Figure 1: PSP: *Pseudomonas sp*, PA: *Pseudomonas aeruginosa*, PF: *Pseudomonas fluorescens*, PP: *Pseudomonas putida*, PS: Pseudomonas strutzeri; Maximum Likelihood relationship among different species of *Pseudomonas* in Worldwide collection.

Sequence retrieval
To compare the similarity or diversity, the nucleotide sequences need to be downloaded. 16S rDNA partial sequence of 60 selected *Pseudomonas* species namely *P. aeruginosa, P. fluorescence, P. putida, P. Strutzeri* were retrieved from NCBI: GenBank (http://www.ncbi.nlm.nih.gov/) with different geographical regions of world. The locations of sample's accession number depicted in **Table 2 (see supplementary material).**

Multiple sequence alignment and model selection for maximum likelihood analysis
The multiple sequence alignment was performed using CLUSTALW 1.6 integrated in software MEGA5.05 **[8]**. To identify the most appropriate nucleotide substitution model for the 16S rDNA sequences, an algorithm was used which is implemented in MEGA 5.05 **[8]**. The nucleotide substitution matrix and transition-transversion bias (R) were estimated for best model of maximum likelihood method. The pattern of nucleotide substitution and transition-transversion bias was also estimated in Maximum Composite Likelihood algorithm with Tamura-Nei model **[9]**.

Molecular evolutionary relationships analysis
The evolutionary history was inferred using the Maximum likelihood method (ML) with most suitable model. The bootstrap consensus trees inferred from 1000 replicates were taken to represent the evolutionary history of the taxa analyzed. The ML trees were obtained using the Nearest-Neighbor-Interchange heuristic algorithm. All positions containing gaps and missing data were eliminated from the
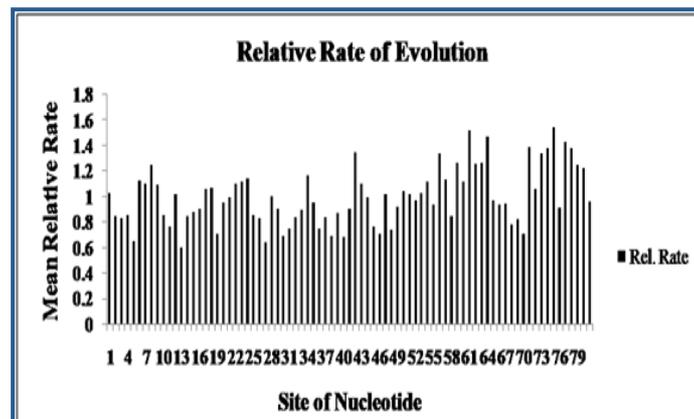
dataset (Complete Deletion option). Phylogenetic analyses were conducted in MEGA5.05 **[10]**.

Figure 2: The mean (relative) evolutionary rates (calculated in K2 + I parameter model) were scaled such that the average evolutionary rate across all sites was 1. This means that sites showing a rate < 1 are evolving slower than average and those with a rate > 1 are evolving faster than average.

Estimation of different evolutionary parameters among different species of Pseudomonas
The sequences of same species of *Pseudomonas* were grouped separately for easy estimation of diversity among different species. The mean diversity within subpopulations, diversity within a species population, inter-population diversity and coefficient of differentiation among the species groups were also estimated under kimura-2-parameter model **[11]** with 1000 bootstrap variance estimate method. The codon positions included were 1st+2nd+3rd+Noncoding for 16S rDNA gene. All the positions containing gaps and missing data were eliminated from the dataset (Complete deletion option).

Estimate of relative evolutionary rate and Tajima's Neutrality test
The mean (relative) evolutionary rate for each site of nucleotide was estimated under Kimura-2-Parameter model **[11]** with invariant site (+I) as this model has third lower Bayesian Information Criterion (BIC) value just higher than Kimura 2-parameter model **[11]** and Jukes-Cantor **[12]** model with uniform rate. This final data set was subjected to Tajima's Neutrality Test **[13]** to evaluate the neutral theory of molecular evolution and to identify the population structure.

Results & Discussion:
The sequences of the 16S rDNA region of some isolates have been submitted in NCBI database **Table 1 (See supplementary material).** All native *Pseudomonas sp* strains used in the study exhibited 96 to 99% sequence similarity to the either *Pseudomonas aeruginosa* or *Pseudomonas sp* available in NCBI database with lowest E-value and maximum query coverage and maximum identity. Similar works also reported by several scientists **[14, 3, 15, 1, 16, 2].** Models with the lowest BIC scores (Bayesian Information Criterion) were considered to describe the substitution pattern the best. In this context K2 model (Kimura's two parameters) displayed lowest BIC scores with the value of 2732.4 **Table 3 (see supplementary material)** used to make consensus ML tree was drawn from these aligned sequences **(Figure 1)**. The multiple alignment file was then used to create phylogram using MEGA5 software using the K2

model based on bootstrap analysis of 1000 replicates was performed to estimate the confidence of tree topologies. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model [8]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 60 nucleotide sequences. All positions containing gaps and missing data were eliminated. The Phylogenetic tree revealed that all the native strain fit in to an evolutionary cluster comprising members of *Pseudomonas aeruginosa* and other *Pseudomonas sp* grouped into separate cluster according to their relatedness **(Figure 1)**. Molecular evolution of West Bengal *Pseudomonas sp* population was much closely related to Indian *Pseudomonas* strains which strongly supported the topology of native strain. On the basis of 16S rDNA sequence homology, different evolutionary parameters and topology of ML, it can be clearly depicted that West Bengal *Pseudomonas sp* population is much more closely related with *Pseudomonas aeruginosa*. The null hypothesis of equal evolutionary rate throughout the ML tree topology was rejected at a 5% significance level **(Figure 1)**. So, the alternative hypothesis i.e., nucleotides showed different evolutionary rate was accepted. Using 16S rRNA sequences, **[17]** showed the phylogenetic affiliation of most *Pseudomonas sp*. According to relatedness of 16S rRNA sequences monophyletic clade was formed with genera *Escherichia* **[18]**. The closest phylogenetic relative was *Escherichia coli* with 99% 16S rDNA gene sequence similarity. Researcher also compared the environmental conditions of strains to reveal whether the phylogentic groups were correlated with their original geographical sites based on 16S rDNA sequences **[19]**.

The mean (relative) evolutionary rates (calculated in K2 parameter model) were scaled such that the average evolutionary rate across all sites was 1 **[11]**. This means that sites showing a rate < 1 are evolving slower than average and those with a rate > 1 are evolving faster than average **(Figure 2)**. These relative rates were estimated under the model **[20]** (+I). A discrete Gamma (+G) distribution was used to model evolutionary rate differences among sites (1 category). Some sites were allowed to stay invariant (+I). It was clearly depicted from **Figure 2**. The maximum sites showed a mean (relative) evolutionary rate < 1and thus it can be concluded that maximum sites showed evolutionary rate slower than the average rate.

Tajima's neutrality test **[13]** clearly indicated that the nucleotide diversity (π) among the population was 0.230 and Tajima test statistic (D) was 0.351 **Table 4 (see supplementary material).** In this study, comparative sequence analysis was carried out to determine molecular evolutionary divergence and classify the sequences into different species of *Pseudomonas* like *Pseudomonas aeruginosa, Pseudomonas fluorescens, Pseudomonas putida* etc. In this study, comparative sequence analysis was carried out to determine molecular evolutionary divergence and classify the sequences into different species of *Pseudomonas* like *Pseudomonas aeruginosa, Pseudomonas fluorescens, Pseudomonas putida* etc. In addition, nucleotide substitution was evaluated in respect of different species of *Pseudomonas* to understand nucleotide divergences like mean evolutionary

diversity within entire population and Subpopulations; mean inter population diversity, coefficient differentiation over species pairs but it was remarkable that the species of *Pseudomonas* from West Bengal attributed the low degree of diversity (0.010) **Table 5 (see supplementary material).** Several works had been demonstrated for considering the possible source of diversifying selection of different protein coding gene of bacteria like *Pseudomonas*. The different gene responsible for production of different protein showed diversifying selection **[21-23]**, but this study first time demonstrated diversifying selection based molecular evolution at 16S rRNA gene. Molecular evolution of West Bengal *Pseudomonas sp* population was much closely related to Indian strains which strongly supported the topology of native strain. In this study Selection pressure supportive molecular evolution was hypothesized. The current analysis clearly depicted that the maximum sites in 16S rDNA region evolved in much slower rate than average. Besides, nucleotide diversity among the whole population was very low and tajima test statistic value (D) was also slightly positive. From the consequences of above result two possibilities can be assumed that either the population may have suffered a recent bottleneck (or be decreasing) or may have evidence for over dominant selection at this locus.

Conclusion:
Distinct selection pressure was first time detected to estimate the mechanism underlying the evolution of the nucleotide sequences and population structure of *Pseudomonas sp* in soil community. In the present study, the determination of molecular evolution and population structure of native fluorescent Pseudomonads using bioinformatics tools, it can be concluded that it would create a new threat of pathogen infection in crop plants under gangetic alluvial zone of West Bengal, India due to bottleneck population structure of *Pseudomonas sp* in soil. If the *Pseudomonas sp* population has recently undergone a bottleneck, the nucleotide diversity of *Pseudomonas sp* will little be affected in soil community gradually resulting fluorescent Pseudomonads strains would not be able to control the crop disease in eco-friendly manner by reducing the use of harmful chemicals, along with enhancing the productivity of different crops. In this concern, this is the right time to focus our attention for altering the soil environment through organic and bio inputs for enhancing the population of native rhizobacteria and thereby converting the soil suppressive against pathogens. However, it is becoming important to map these pseudomonad populations in order to select ''consortia'' of different rhizospheric plant growth promoting bacteria for biological control for different host-pathogenic system instead of using any one species/strain of Fluorescent Pseudomonad.

References:
[1] Elomari M *et al*. *Int J Syst Bacteriol*. 1997 **47**: 846 [PMID: 9226917]
[2] Dabboussi F *et al*. *Res Microbiol*. 1999 **150**: 303 [PMID: 10422691]
[3] Elomari M *et al*. *Int J Syst Bacteriol*. 1996 **46**: 1138 [PMID: 8863448]
[4] Godfrey SAC *et al*. *J Appl Microbiol*. 2001 **91:** 412 [PMID: 11556905]
[5] Behrendt U *et al*. *Int J Syst Evol Microbiol*. 2007 **57**: 979 [PMID: 17473245]
[6] Hughes AL & Nei M, *Nature*. 1988 **335:** 167 [PMID: 3412472]

[7]   Yang Z & Bielawski JP. *Trends Ecol Evol.* 2000 **15:** 496 [PMID: 11114436]

[8]   Tamura K *et al*. *Mol Biol Evol.* 2011 **28**: 2731 [PMID: 21546353]

[9]   Tamura K & Nei M. *Mol Biol Evol.* 1993 **10**: 512   [PMID: 8336541]

[10] Tamura K *et al*. *Mol Biol Evol.* 2007 **24**: 1596 [PMID: 8336541]

[11] Kimura M. *J Mol Evol.* 1980 **16**: 111 [PMID: 7463489]

[12] Jukes TH & Cantor CR. In: Munro HN (ed), *Mammalian Protein Metabolism*, Academic Press, New York 1969 pp.121-132.

[13] Tajima F. *Genetics.* 1989 **123**: 585 [PMID: 2513255]

[14] Kersters K *et al*. *Syst Appl Microbiol.* 1996 **19**: 465

[15] Coroler L *et al*. *Syst Appl Microbiol.* 1996 **19**: 600

[16] Dabboussi F *et al*. *Int J Syst Bacteriol.* 1999 **49**: 1091[PMID: 10425766]

[17] Anzai Y *et al*. *Int J Syst Evol Microbiol.* 2000 **50**: 1563 [PMID: 10939664]

[18] Siraj Ud Din MM *et al*. *Bioinformation.* 2011 **6**: 370 [PMID: 21904423]

[19] Peng H *et al*. *Biochem Mol Biol.* 2006 **39**: 178 [PMID: 16584633]

[20] Tamura K. *Mol Biol Evol.* 1992 **9**: 678 [PMID: 1630306]

[21] Smith EE *et al*. *J Bacteriol.* 2005 **187**: 2138 [PMID: 15743962]

[22] Polley SD & Conway DJ, *Genetics.* 2001 **158**: 1505 [PMID: 11514442]

[23] Boyd EF & Hart DL, *J Mol Evol.* 1998 **47**: 258 [PMID: 9732452]

Supplementary material:

Table 1: Result of NCBI BLAST of 16S rDNA region of different native isolates

SL.No.	Strains	Maximum identity with BLAST Seq ID.	Gen Bank Accession no
1.	BFP4	JN975155	KC414324
2.	MFP9	KC212085	KC414325
3.	CHFP11	KM280025	KC414326
4.	MFP10	JX109938	KC414330
5.	CHFP23	KF782794	KC414329
6.	OFP4	KM280025	KC414327
7.	CIFP4	KC959478	KF430021
8.	BNFP3	CP010555	KF430023
9.	CIFP7	JQ660570	KF430022
10.	PFP3	KM280025	KC414328
11.	TFP6	JQ917006	KC414331
12.	GFP8	HE862283	KC414332
13.	BFP5	JX441328	KC414333
14.	CBFP4	EF071945	KC414334
15.	BFP10	JF758866	KC414335
16.	CHFP25	KJ958500	KC414336
17.	TFP16	KM982694	KC414337
18.	BFP13	FJ194518	KF430024

Table 2: List of the strain name with accession no and location

Strain name	Gen Bank accession no	Location
PSP 1	AY275488	Canada
PSP2	AY275487	Canada
PS1	HF675140	India
PS2	KF532951	Belgium
PF 1	EF418612	India
PF 2	JX905208	Canada
PF 3	DQ310732	India
PF 4	U63902	Japan
PF 5	U63901	Japan
PF 6	EF159157	India
PF 7	PFU63901	Japan
PF 8	JQ660571	India
PF 9	HQ907732	India
PF 10	DQ073039	Spain
PF 11	JQ801441	India
PF 12	EF413073	India
PF 13	FJ639838	Egypt
PF14	HQ696487	India
PA 1	AJ496326	UK
PA 2	DQ211696	Pakistan
PA 3	AY337580	Taiwan
PA4	AY074896	Egypt
PA 5	EU167929	India
PA 6	AY999032	India
PA 7	AY956935	India
PA 8	AM293357	UK
PA 9	AM263499	UK
PA 10	AM263498	UK
PA 11	AM263497	UK
PA 12	AM263496	UK
PA 13	DQ310732	India
PA 14	AM263495	India
PA 15	AM263494	India
PA 16	AM263493	India
PA 17	AM263490	India
PA 18	AM263488	India

PA 19	AM263486	India
CHFP23	KC414329	India
MFP10	KC414330	India
BFP4	KC414324	India
MFP9	KC414325	India
CHFP11	KC414326	India
OFP4	KC414327	India
TFP6	KC414331	India
GFP8	KC414332	India
BFP5	KC414333	India
CBFP4	KC414334	India
BFP10	KC414335	India
CHFP25	KC414336	India
TFP16	KC414337	India
BNFP3	KF430023	India
BFP13	KF430024	India
CIFP7	KF430022	India
CIFP4	KF430021	India
PP1	AJ310536	Vietnam
PP2	AB060973	Japan
PP 3	HQ238582	China
PP 4	HG313638	Switzerland
PP 5	AB362881	Japan
PP 6	FR877564	Romania

PSP: Pseudomonas sp, PA: Pseudomonas aeruginosa, PF: Pseudomonas fluorescens, PP: Pseudomonas putida, PS: Pseudomonas strutzeri

Table 3 : Maximum likelihood fits of 24 different nucleotide substitution models

Model	#Param	BIC	AICc	lnL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G
K2	118	**2732.44**	1977.4	-867.62	n/a	n/a	1.17956	0.25	0.25	0.25	0.25
K2+G	119	2734.72	1973.34	-864.538	n/a	5.591788	1.26137	0.25	0.25	0.25	0.25
JC	117	2739.27	1990.58	-875.262	n/a	n/a	0.5	0.25	0.25	0.25	0.25
T92	119	2740.06	1978.67	-867.205	n/a	n/a	1.18492	0.25903	0.25903	0.24097	0.24097
K2+I	119	2740.53	1979.15	-867.441	0.0094781	n/a	1.18978	0.25	0.25	0.25	0.25
T92+G	120	2742.47	1974.74	-864.184	n/a	5.705324	1.26731	0.25903	0.25903	0.24097	0.24097
K2+G+I	120	2743.17	1975.44	-864.536	0.0008917	5.591788	1.26249	0.25	0.25	0.25	0.25
JC+G	118	2743.98	1988.94	-873.39	n/a	7.045977	0.5	0.25	0.25	0.25	0.25
JC+I	118	2747.19	1992.15	-874.995	0.0112392	n/a	0.5	0.25	0.25	0.25	0.25
T92+I	120	2747.93	1980.2	-866.917	0.0099158	n/a	1.19684	0.25903	0.25903	0.24097	0.24097
T92+G+I	121	2750.88	1976.8	-864.163	0.0018308	5.64827	1.27054	0.25903	0.25903	0.24097	0.24097
JC+G+I	119	2752.45	1991.06	-873.399	0.0038026	7.261655	0.5	0.25	0.25	0.25	0.25
HKY	121	2763.36	1989.29	-870.405	n/a	n/a	1.24007	0.23253	0.28532	0.18273	0.29921
TN93	122	2770.88	1990.46	-869.937	n/a	n/a	1.21292	0.23253	0.28532	0.18273	0.29921
HKY+I	122	2770.88	1990.47	-869.939	0.0121188	n/a	1.26256	0.23253	0.28532	0.18273	0.29921
TN93+I	123	2777.94	1991.18	-869.244	0.0139844	n/a	1.23753	0.23253	0.28532	0.18273	0.29921
HKY+G	122	2778.73	1998.31	-873.863	n/a	5.425706	1.35536	0.23253	0.28532	0.18273	0.29921
TN93+G	123	2785.93	1999.18	-873.24	n/a	4.956476	1.2925	0.23253	0.28532	0.18273	0.29921
HKY+G+I	123	2787.12	2000.36	-873.831	0.0056768	5.53587	1.36616	0.23253	0.28532	0.18273	0.29921
GTR	125	2793.91	1994.47	-868.777	n/a	n/a	1.19645	0.23253	0.28532	0.18273	0.29921
TN93+G+I	124	2794	2000.9	-873.048	0.009718	5.057112	1.31211	0.23253	0.28532	0.18273	0.29921
GTR+G	126	2796.12	1990.34	-865.655	n/a	5.480511	1.2537	0.23253	0.28532	0.18273	0.29921
GTR+I	126	2800.93	1995.16	-868.065	0.0141917	n/a	1.22258	0.23253	0.28532	0.18273	0.29921
GTR+G+I	127	2804.26	1992.15	-865.502	0.0093857	5.64827	1.27085	0.23253	0.28532	0.18273	0.29921

Different nucleotide substitution models GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor

Table 4: Tajima's neutrality test

m	*S*	*p*$_s$	*Θ*	*π*	*D*
60	79	0.975309	0.209150	0.230488	0.351733

m = number of sequences, S = Number of segregating sites, p_s = S/m, $Θ$ = p_s/a_1, $π$ = nucleotide diversity, and D is the Tajima test statistic.

Table 5: Estimation of different molecular evolutionary factors

Strains	Mean Evolutionary Diversity for the Entire Population	Mean Diversity within sub population	Mean Inter-population Diversity over sp. pairs	Co-efficient of differentiation over sp. pairs
Pseudomonas sp West Bengal	0.010			
P. fluorescence	0.492	0.299	0.007	0.024
P. aeruginosa	0.529	0.317	0.020	0.060
P. putida	0.021	0.031	0.030	0.093
P. strutzeri	2.273	n/c	n/c	n/c