Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Gene-centric metegenome analysis reveals diversity of *Pseudomonas aeruginosa* biofilm gene orthologs in fresh water ecosystem

Rani Anupama[a], Amitava Mukherjee[b], Subramanian Babu[a,*]

[a] School of Bio Sciences and Technology, VIT University, Vellore 632014, India
[b] Centre for Nano Biotechnology, VIT University, Vellore 632014, India

ARTICLE INFO

Keywords:
Biofilm
Freshwater
Metagenome
Ortholog PCR
Pseudomonas aeruginosa

ABSTRACT

Metagenomic analysis of biofilm forming bacteria in environmental samples remains challenging due to the non-availability of gene sequences of most of the uncultivable bacteria. Sequences of *Pseudomonas aeruginosa* PAO1-UW genes involved either directly or indirectly in biofilm formation were analyzed using BLASTn to obtain matching sequences from different strain, species and genus. Conserved regions in the functional domain of the amino acid sequences were used to design common primers for direct PCR analysis of freshwater metagenomes. Seven key genes such as *ace*A, *clp*P, *typ*A, *cbr*A, *pho*R, *rpo*S and *gac*A involved in biofilm formation were validated. The ortholog genes belonged to wide range of *Pseudomonas* sp. indicating the diversity of biofilm genes and the conservation of protein functional domains. The approach would also help in analyzing the expression of biofilm genes in different bacteria of freshwater systems for monitoring toxic contaminations such as organic or inorganic pollutants.

1. Introduction

Most of the bacteria in the environment have the inherent capability to exist as biofilm. Microbial biofilms dominate biogeochemical processes in many sedimentary environments such as stream and riverbeds, lake sediments or ground water [1,2]. In bacteria, biofilms are involved in antimicrobial resistance, virulence and adaptation to environmental toxicity. The microbial diversity of natural biofilm has important role for the functioning of aquatic environments [3]. Studies on biofilms help to understand the bacterial ecology and community interactions in various environments [4]. In addition to the fundamental knowledge on bacterial systems gained through these studies, biofilms have wide application in commercial industries [5]. Plethora of work has already been done on *Pseudomonas aeruginosa* biofilms and myriads of data are available including whole genome sequence in Pseudomonas Genome Database (PGD); (earlier referred to as the *Pseudomonas aeruginosa* Genome Database http://www.pseudomonas.com/) [6]. Such available information has made *P. aeruginosa* PAO1-UW a reference strain for biofilm research in other bacteria.

In a recent study by Kumari et al. [7], effect of low concentrations of titanium dioxide on the biofilm related genes of *P. aeruginosa*, *Bacillus subtilis* and *B. altitudinus* was reported. The study describes expression of biofilm related genes of *P. aeruginosa* and *B. subtilis* using gene specific primers in PCR which resulted in background amplification of unknown genes in *B. altitudinus* [7]. *B. altitudinus* biofilm genes are yet to be characterized. The aforesaid observation indicated the presence of gene orthologs in bacteria towards common functionality (biofilm formation).

Metagenomic study of microbiomes of various environments such as marine water [8], freshwater [9], drinking water [10], sludge waste [11], soil [12], guts [13], acid mine drainage [14], oral cavity [15], ice shelves [16], glaciers [17] has already been reported. The common approach that has been used for culture-independent study of microbes from different environments mainly depend on high-throughput shotgun sequencing, followed by assembling large sequence data and software-based generation of contigs for the taxonomic classification of microbial metapopulation. However, most of the earlier works on targeting genes in the metagenome were based on construction of clonal library of the selected amplicons [18,19]. With the advent of metagenomics, where research is independent of conventional microbial culturing, there is a need for the development of gene specific probes/primers from reference strains that can be used to detect similar orthologs in other cultivable as well as uncultivable species in the environment.

To study the gene orthologs, Polymerase Chain Reaction (PCR) primer design based on conserved sequences has been usually common but the present study has incorporated a protein functional domain approach to use corresponding conserved region in genes. Although

* Corresponding author at: School of Biosciences and Technology, VIT University, Vellore 632014, Tamil Nadu, India.
E-mail address: babu.s@vit.ac.in (S. Babu).

http://dx.doi.org/10.1016/j.ygeno.2017.08.010
Received 19 June 2017; Received in revised form 14 August 2017; Accepted 30 August 2017

Table 1
P. aeruginosa PAO1- PAO-UW biofilm genes, ortholog based PCR primers, corresponding proteins and functional domains.

Genes	Primer sequences (5′ → 3′)	GC percentage	Annealing temp (°C)	Expected product size (bp)	Protein	Protein size (aa)	Domain range	Domain functions	Database used
Gene category I									
*clp*P	(F) TCTATTCGCGCCTGCTGAAGG	(F) 57	58	294	ATP-dependent Clp protease	213	35–205	Serine protease	NCBI
	(R) AATGCGGCAGGCAGTAGCGC	(R) 65							
*rpo*S	(F) CTGCTCGACCTGATCGA	(F) 59	52	551	RNA polymerase sigma factor RpoS	334	61–94	Sigma-70 factor	Uniprot
	(R) CGGGTCAGGCCGATTTCCT	(R) 63					99–169	″	
							179–254	″	
							267–320	″	
*rpo*N	(F) GCCTGGGAAGACATCTACCAG	(F) 57	55	1081	RNA polymerase factor sigma-54	497	3–51	Sigma-54 factor, Activator interacting domain	NCBI
	(R) CCAGTAAACCAGCGATCTTGCTG	(R) 52					123–324	Sigma-54 factor, core binding domain	
							336–495	Sigma-54, DNA binding domain	
Gene category II									
*gac*A	(F) CCAGATCGCCCTGATGATCGC	(F) 62	55	108	Response regulator GacA	214	1–119	Response regulatory	Uniprot
	(R) TTCTCGAAGATGCGGTAGCG	(R) 55					143–208	HTH-Lux R type	
*alg*D	(F) ACCGCGATCAAGGACTAC	(F) 56	52	210	GDP-mannose 6-dehydrogenase	436	1–190	UDPG_MGDP_dh family, NAD binding domain	NCBI
	(R) CTTGGCGATGTTGCCGATCT	(R) 55					203–300	UDPG_MGDP_dh central domain	
							317–424	UDP binding domain	
*cbr*A	(F) CCGAGGACATCCACTTCATCG	(F) 57	51	1034	Histidine kinase	983	1–451	Sodium solute symporter	Prosite
	(R) TTCCACGATCGAATAGACCAG	(R) 47					630–674	PAS domain	
							763–977	Histidine kinase	
*opg*H	(F) ATCGTCATGCCGATCTGCAAC	(F) 52	56	671	Glucans biosynthesis glucosyltransferase H	861	247–533	Nucleotide-diphospho-sugar transferases	Interpro
	(R) AGGTCGTAGGCGATCCACA	(R) 58					249–432	Glucan biosynthesis glucosyltransferase H	
*opg*G	(F) CTGCAGATCCATGCCGGCAACG	(F) 64	56	627	Glucans biosynthesis protein G	525	35–522	Galactose mutarotase-like domain	Interpro
	(R) CGGCTTCTTCGGATCCTTGAC	(R) 57					411–524	Immunoglobulin *E*-set	
							413–524	Immunoglobulin-like fold	
*typ*A	(F) CTGGACTACAACAGCTTCCT	(F) 50	52	300	Regulatory protein TypA	605	205–290	BipA_TypA_II: domain II of BipA	NCBI
	(R) GAAGGTCATGCTCACGGTCGG	(R) 62					398–476	BipA_TypA_C: a C-terminal portion of BipA or TypA	
Ppx	(F) GACATCGGCGGCGGCAGCAC	(F) 75	64	735	Exopolyphosphatase	506	17–315	Ppx/GppA phosphatase	Interpro
	(R) CAGGTAGGCGCCGTGCTTGTGG	(R) 68							
*pho*R	(F) GCGCACGCCGCTGACGGTGATC	(F) 73	56	549	Phosphate regulon sensor protein PhoR	443	93–182	PAS domain	Uniprot
	(R) TGCTTGACGATGGCCAGG	(R) 61					216–433	Histidine kinase	
Gene category III									
*che*Y	(F) GTCACCGACTGGAACATGCC	(F) 60	57	153	Chemotaxis protein CheY	124	1–119	Response regulatory	Uniprot
	(R) GCCGTTGACCCCGGCCTGGG	(R) 80							
*psl*A	(F) TGGTCGGCCCGCGCCC	(F) 88	63	124	PslA	478	150–221	NADP binding domain	Pfam
	(R) TTGATCTGCGCCCAGCCGGT	(R) 65					284–476	Bacterial sugar transferase	
Gene category IV									
*ace*A	(F) CAAGAAGCACCTGAAGACCACC	(F) 55	56	812	Isocitrate lyase AceA	531	11–278	Isocitrate lyase/phosphoenolpyruvate mutase enzyme family	Pfam, NCBI
	(R) GCGGCTTCTCGGTTTCGATCCA	(R) 59					315–408	″	
							413–530	″	
*bfi*S	(F) ACCGAGAACACCACCGACCTG	(F) 62	60	803	Histidine kinase	758	266–336	PAS domain S-box	Prosite
	(R) GGCTGGTTGATCTCGTGGGCGA	(R) 64					389–426	″	
							339–392	″	
							533–753	Histidine kinase	
*mif*R	(F) GCATGCCCGGCATGGACGG	(F) 74	59	481	MifR	447	5–115	Response regulator receiver	Pfam,
	(R) CCGAACAGCTCGCTCTCGAAC	(R) 62					144–311	Sigma-54 interaction domain	Interpro
							355–436	Homeodomain, DNA binding HTH domain	

CategoryI: Protein domains information available for all the three selected orthologs.
CategoryII: Protein domains information available in different species and genus selected orthologs but not in same subgroup.
CategoryIII: Protein domains information available only till same species level but was not found in different genus selected orthologs.
CategoryIV: Protein domain information available only for *P.aeruginosa* PAO1 but not for any of the selected orthologs.
F – forward primer; R – reverse primer; bp – base pairs.

protein domain-based PCR primer design has been reported earlier by Jin et al. [11] using CODEHOPs (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) tool to analyze carboxypeptidase gene in metagenome of activated sludge waste water, the study focused on a single gene target. Consequently, for designing substantial number of primers based on conserved gene sequence among orthologs, MetCap pipeline was also created for largescale targeted metagenomics [20]. We targeted multiple genes of *P. aeruginosa* that are involved in a complex process of biofilm formation. We have used a fixed criterion of choosing possible ortholog genes in i) different strain of the same species ii) distinct species of the same genus (*Pseudomonas*) and iii) different genera.

2. Results

2.1. Genome wide computational analysis of biofilm genes

Details of genes retrieved from *P. aeruginosa* PAO1-UW genome are listed in Supplementary Table 1. The table in addition represents orthologs in other bacteria, functions and the sequence coverage homology with orthologs. Total number of genes involved in biofilm formation was found to be 71 using the keyword “biofilm” in search option in the whole genome annotation file. Among these, seven genes were reported to be hypothetical, whereas other genes were found functionally annotated with cellular processes such as translation, transcription, signal transduction, transport, motility, chemotaxis and metabolism. However, most of the genes were found to contribute in swarming motility and transcription regulation and exopolysaccharide synthesis. Size of the genes ranged from 237 bp to 3582 bp. The smallest gene was found to be *clp*S (PA2621) with a size of 237 bp and the largest gene was *pel*B (PA3063) with a size of 3582 bp.

Different strains with maximum sequence identity and query coverage were extracted using nucleotide BLAST as described. PA0705 (*mig*A) was the only gene for which no sequence similarity was found with other strains, species and genus. Gene sequence of asterisked strains of Supplementary Table 1 were reverse complemented before performing ClustalW alignment to get matching score in accordance with the BLAST result. For genes *alg*Q, *wap*B, *rsa*L, *psl*D, *psl*O, *clp*S, *pel*D, *pel*B and *exo*S, sequence similarity was confined to same species but different substrains. Therefore, for such genes, alignment was obtained only for two sequences. For 29 genes (*spr*P, *lop*A, *psl*F, *psl*G, *psl*H, *psl*I, *psl*J, *psl*K, *psl*L, *czc*R, *pel*E, *pel*C, *pel*A, *hpt*B, *lec*B, *amr*Z, *alg44*, *alg*K, *alg*X, *ppr*B, *mva*T, PA4354, PA4819, *est*A, *alg*P, *alg*R, *rnk*, *sad*B, *ton*B1) sequence alignment was found with different species of *Pseudomonas* also. Thus, for these genes, alignment of 3 sequences was obtained. Besides this, 32 genes (*pfp*I, *fle*R, *bdl*A, *che*Y, *clp*P, *psl*A, *psl*B, *psl*C, *psl*E, *psl*M, *psl*N, *gac*A, *ace*A, *pel*G, *pel*F, *alg*D, *rpo*S, *bfm*R, PA4108, *amp*R, *bfi*S, *rpo*N, *cbr*A, PA4781, *opg*H, *opg*G, *typ*A, *ppx*, *pho*R, *alg*B, *mif*R) were noticed to display alignments even among different genus as well. Four sequence alignments were performed for these 32 genes.

2.2. Gene selection for design of universal primers

In the ClustalW sequence alignments, although 32 genes exhibited homology even beyond species level, only 18 genes showed significant matching of sequences. Out of these, *bfm*R and *alg*B were not used for primer designing since the conserved nucleotide regions were found much closer, resulting in a too smaller expected PCR product (73 bp and 63 bp respectively). After this shortlisting, the genes were grouped into four categories. Genes *clp*P, *rpo*S, *rpo*N (Category I) had a explicit domain information of all the three orthologs available in the protein databases, whereas no domain information was available on any of the gene orthologs of *bfi*S, *mif*R, *ace*A in other bacteria, except for *P. aeruginosa* PAO1-UW sequence (Category IV). For category I, four protein sequence alignment displayed high degree of homology even across genus. The nucleotide sequence corresponding to the conserved protein domain was calculated and from that conserved nucleotide sequences, two oligonucleotides were selected for primer designing. For genes *gac*A, *alg*D, *cbr*A, *opg*G, *opg*H, *typ*A, *ppx* and *pho*R, protein domain information was available for different species and different genera but was not found in the selected subgroup (Category II). Hence, three sequence alignments were performed for them and oligonucleotides corresponding to the conserved domain were used for primers. Alignment of functional domain of protein coded by *psl*A indicated conserved regions across subspecies and distinct species of *Pseudomonas*. (Category III). Therefore, two protein sequence alignments were used in this category. Likewise, after all the domain screening, above mentioned 16 genes were finally selected for analysis (listed in Table 1). Primers contained a minimum of 17 nucleotides from the conserved domain after assuring the suitability for use in PCR (Table 1). Protein sequence alignments displayed more homology than nucleotide sequence alignments for all the shortlisted genes. The hypothetical interaction among these genes based on the literature search and their relative participation in biofilm formation are represented in Fig. 1 using Chilibot tool [21] which highlights the direct link of gene *cbr*A, *ace*A and *rpo*S with biofilm (green) and inhibitory connections between *rpo*N, *bfi*S and *gac*A with biofilms.

2.3. Analyses of water samples

Pond water samples collected from different locations of Vellore were analyzed for their physical properties. Supplementary Table 2 shows the average value of pH, temperature, conductivity, Total Dissolved Solids (TDS) and salinity of all samples which varied among locations. The pH of the water samples was mostly alkaline with maximum pH of 9.62 for VFP03 but RAN10 was found to be slightly acidic with pH 6.26. Temperature of the samples was recorded in the range of 29.5 °C (ALV06) to 33.5 °C (MAR09). Conductivity was found maximum in RAN10 with 5730 µs followed by MAR09 with 4230 µs and minimum of 357.5 µs in OTT077 with the majority lying in the range of around 1100 µs. TDS was relatively very high for RAN10 as compared to the other samples with the concentration of 4.07 µg/ml. MAR09 and VFP03 were next to RAN10 with 2.97 and 2.32 µg/ml of TDS, respectively while the normal TDS was found in the range of 0.27–1.1 µg/ml. Salinity was observed maximum for RAN10 and VFP03 with 2187.5, 1710 and 1560 ppm, though the remaining were below 800 ppm. Dissolved Oxygen (DO) and Biochemical Oxygen Demand (BOD) were found in the range of 4–5.7 mg/l and 12–114 mg/l respectively, with the highest value of both belonging to RAN10. The physical parameters along with DO and BOD indicate the level of overall pollution. All these analyses represent different pollution level in waters samples which is the clear indication of microbial population level in them [22]. Based on all these physical parameters sample RAN10 was found to be the most contaminated and sample OTT07 was the least contaminated followed by PEN04 which in turn denote lesser microbial diversity in them.

2.4. Metagenomic PCR analyses

The PCR performed using freshwater metagenomes as template showed that the amplification of selected genes was as per the expected product size as shown in Fig. 2. Although for every gene, there was variation observed in the gel band pattern for each sample, nevertheless, the expected amplicon bands were commonly present in all the samples except a few. This common PCR amplification pattern though with different band intensities confirms the direct amplification of genes involved in biofilm formation in the pond fresh water metagenome. Multiple bands indicate the possible non-specific amplification due to the unknown complexity of metagenome. However, *ace*A (Fig. 2a) and *cbr*A (Fig. 2d) gene did not show multiple bands for all the samples which might be due to the uniqueness of primer targeted region of gene sequence among freshwater bacterial community. Sample

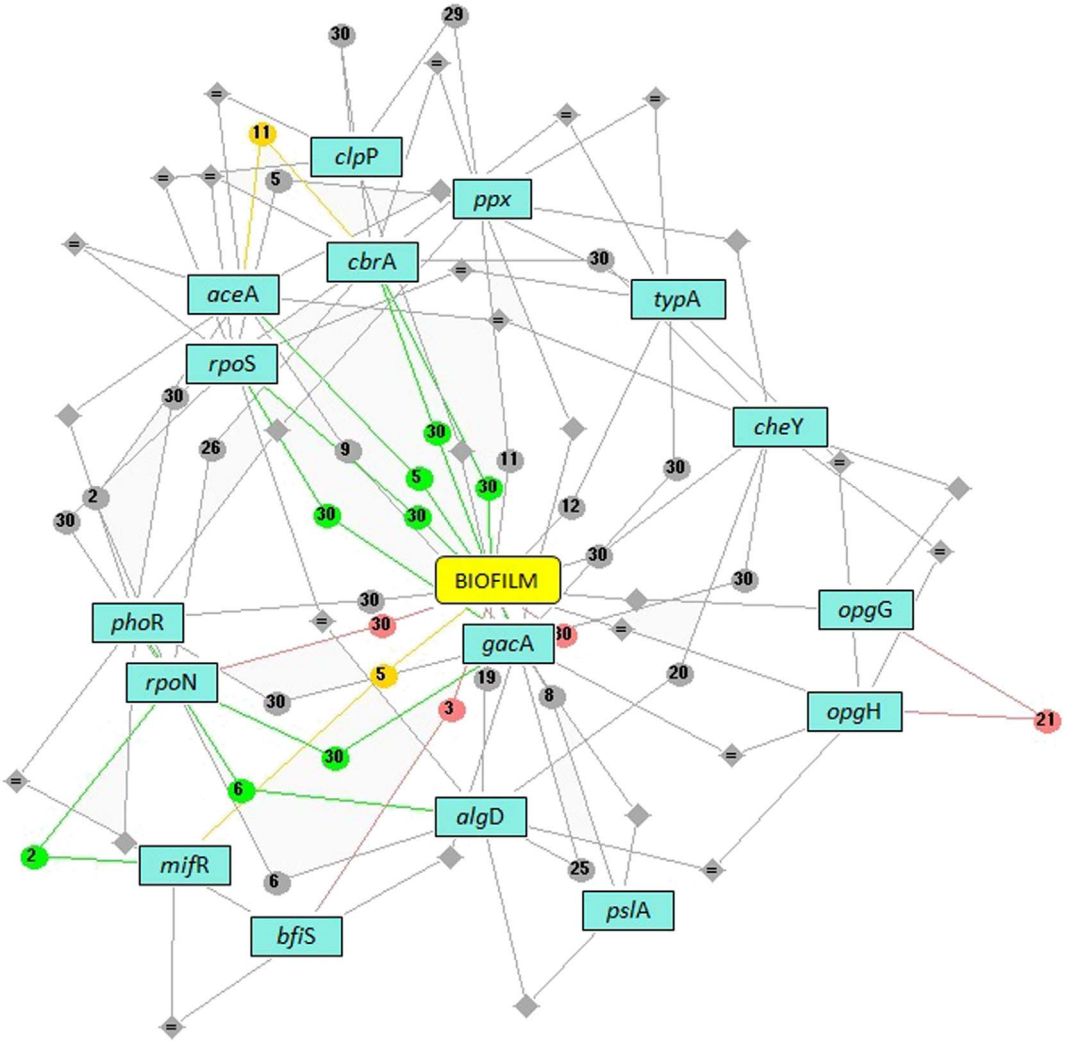

Fig. 1. Network map of biofilm related genes constructed using Chilibot.
Lines and nodes indicate the interactions between genes and relation of genes with biofilm. Colour represent stimulatory relationship (green), inhibitory (red), stimulatory/inhibitory (yellow), queried terms (cyan blue), neutral interaction (grey). The numbers denote the size of the abstract searched in literature. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

STH05 and THR02 displayed single expected band for all the genes. Unlike other samples, no amplification was observed for any gene in sample OTT07 (Fig. 2). No amplification was observed for *cbr*A gene in samples VIT01 and THR02 (Fig. 2d). OTT07 was the sample with least TDS, salinity, conductivity and BOD values that indicate less suitability of this water sample for biofilm bacteria. The genes that are validated by this approach and the amplified gene and their sources are given in Table 2.

2.5. Sequence analysis and phylogeny

Sequence analysis using BLAST search confirmed that the sequences coded for the target genes on which the primers were basically designed. The first hit for each gene sequence search with maximum sequence identity (> 98%) was *P. balearica* DSM 6083 (*ace*A), *P. alcaligenes* NEB585 (*clp*P), *P. stutzeri* CCUG 29243 (*typ*A), *P. putida* W619 (*cbr*A), *P. mendocina* S5.2 (*pho*R), *P. pseudoalcaligenes* (*rpo*S) and *P. resinovorans* NBRC 106553 (*gac*A). Although the obtained sequences were found to be conserved not only in wide range of *Pseudomonas* sp. but even in unrelated genus as well. To see the significant relatedness of the gene and its evolutionary trend among different bacterial genus and species, phylogenetic analysis was performed for the top hundred BLAST hits with distinct species of *Pseudomonas* as well as different genus of bacteria having the same gene. Bootstrap consensus trees were constructed using maximum likelihood method with bootstrap value of 1000 for high robustness is shown in Fig. 3. Gene *ace*A evolved in wide range of different *Pseudomonas* species as well as bacteria of different genus *Azotobacter chroococcum*. This bacterial species of different genus along with three cross genus bacteria *Thioalkalivibrio virsutus*, *Hafnia alvei* and *Aeromonas schubertii* were also found to carry *rpo*S gene. *Azotobacter vinelandii*, *Halotalea alkalilenta* and *Marinobacter similis* of outside genus were found in the phylogeny of *gac*A, *clp*P and *typ*A respectively. On the contrary, *cbr*A and *pho*R were seen only within the *Pseudomonas* genus. Among all the orthologs, *P. alcaligenes* and *P. stutzeri* were two bacteria which could be seen in all the cladograms. *P. aeruginosa*, *P. balearica* and *P. citronellolis* were found common among all the trees except for *cbr*A. Similarly, *P. mendocina* also existed in all except *rpo*S. On the other hand the uniquely spotted orthologs were *P. oryzihabitan* for *clp*P; *P. koreensis*, *P. agrici*, *P. plecoglossida*, *P.fragi* for *cbr*A; *P. proteogens* and *P. cichorii* for *pho*R.

3. Discussion

The protein products of the genes analyzed in the study can be categorized into signal transduction by transcriptional regulators, those involved in transcription, adaptation or protection, motility,

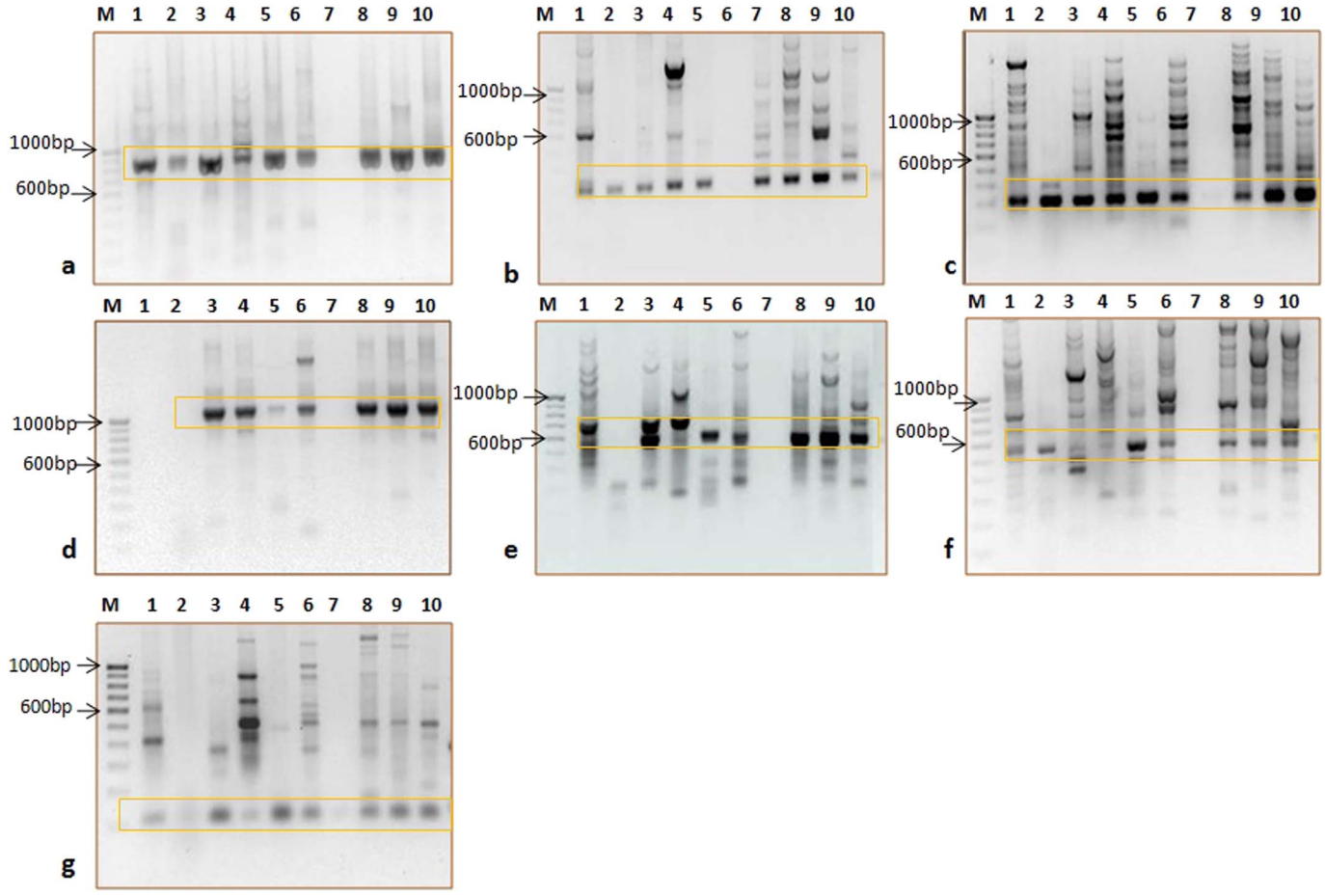

Fig. 2. PCR analysis of freshwater metagenome using universal primers for biofilm related genes.
a. *ace*A b. *clp*P c. *typ*A d. *cbr*A e. *pho*R f. *rpo*S g. *gac*A, M - 100 bp DNA ladder, 1-VIT01, 2- THR02, 3- VFP03, 4- PEN04, 5- STH05, 6- ALV06, 7- OTT07, 8- VIR08, 9- MAR09, 10- RAN10.

chemotaxis, transport and metabolism, energy production and conservation and cell wall/membrane/cell envelope biogenesis. Out of the 16 genes for which PCR primers were designed based on protein functional domain in ortholog genes of related and unrelated species and genera to *P. aeruginosa*, we could validate the primers for seven genes (*ace*A, *clp*P, *typ*A, *cbr*A, *pho*R, *rpo*S, *gac*A). The proteins of the genes *che*Y and *gac*A have been known to be involved in phosphorelay response [23,24]. CheY is known to be involved in phosphate metabolism and chemotaxis [23] whereas, phoR in phosphate regulation [25]. The reading frame of *che*Y resides at the downstream of *fli*A gene which codes for protein known to be involved in flagellin synthesis. Gene *gac*A is a global activator, highly conserved gene and synthesizes

Table 2
List of biofilm genes validated by metagenome analysis.

Gene	Ortholog sources used for primer design	Amplified gene	Source microbe in the metagenome	GenBank Acession ID
*ace*A	*Pseudomonas aeruginosa* PAO1-VE13 *Pseudomonas knackmussii* B13 *Azotobacter vinelandii* CA6	*aceA*	*P. balearica* DSM 6083	KX756976.1
*clp*P	*Pseudomonas aeruginosa* PAO1-VE13 *Pseudomonas denitrificans* ATCC 13867 *Marichromatium purpuratum* 984	*clpP*	*P. alcaligenes* NEB585	KX772807.1
*typ*A	**Pseudomonas aeruginosa* PAO1-VE13 **Pseudomonas denitrificans* ATCC 13867 **Azotobacter vinelandii* CA6	*typA*	*P. stutzeri* CCUG 29243	KX779040.1
*cbr*A	*Pseudomonas aeruginosa* PAO1-VE13 *Pseudomonas denitrificans* ATCC 13867 *Azotobacter vinelandii* CA6	*cbrA*	*P. putida* W619	KX772806.1
*pho*R	*Pseudomonas aeruginosa* PAO1-VE13 **Pseudomonas chlororaphis* strain PA23 **Thioalkalivibrio sulfidophilus* HL-EbGr7	*phoR*	*P. mendocina* S5.2	KX772808.1
*rpo*S	**Pseudomonas aeruginosa* PAO1-VE13 **Pseudomonas resinovorans* NBRC 106553 DNA **Halomonas elongata* DSM 2581	*rpoS*	*P. pseudoalcaligenes*	KX779039.1
*gac*A	**Pseudomonas aeruginosa* PAO1-VE13 **Pseudomonas knackmussii* B13 *Azotobacter vinelandii* CA6	*gacA*	*P. resinovorans* NBRC 106553	KX831930.1

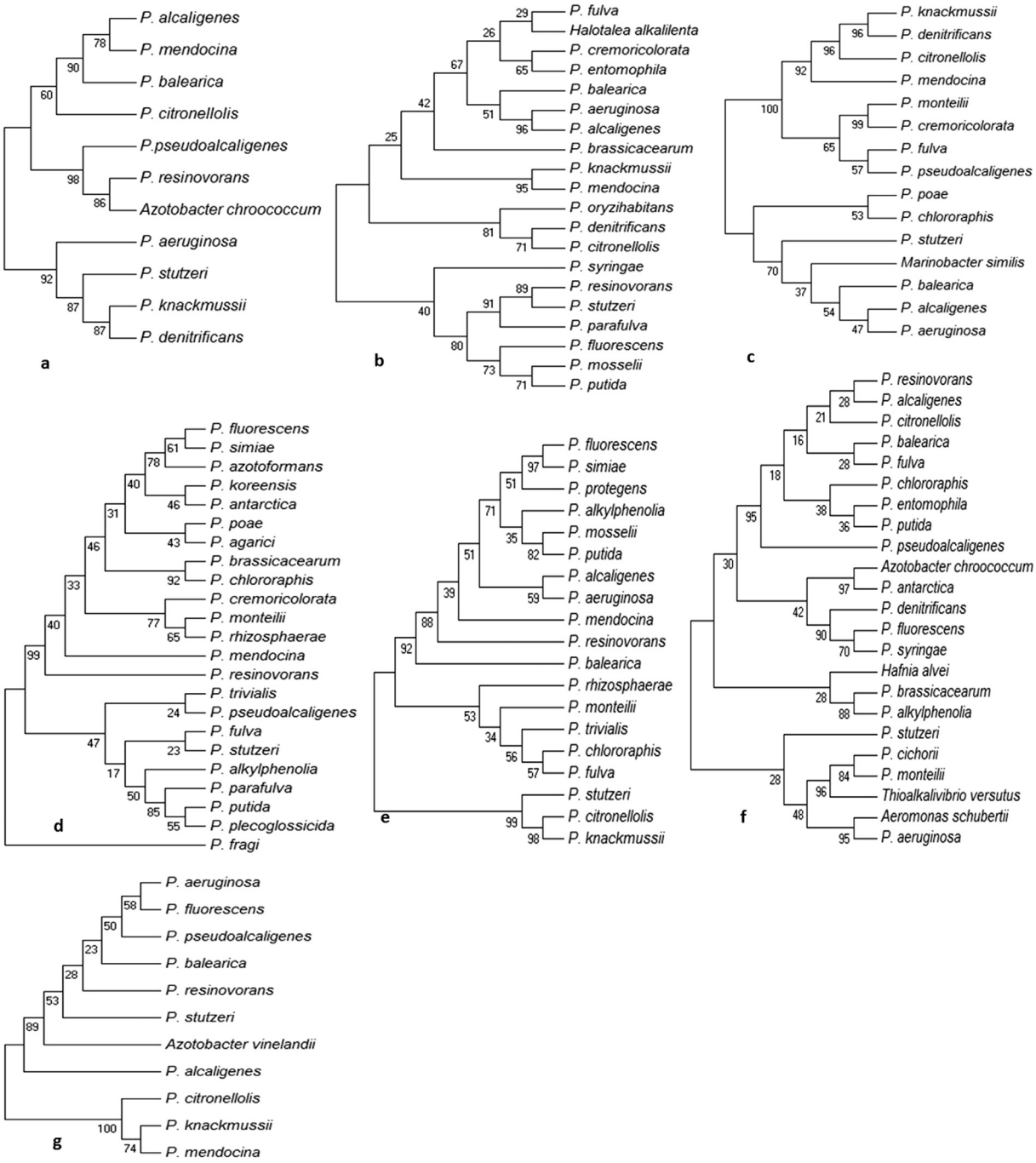

Fig. 3. Bootstrap consensus maximum likelihood tree representing evolutionary phylogenetic relationship of biofilm forming genes among different *Pseudomonas* species and other genera.
a. *ace*A b. *clp*P *c. typA* d. *cbr*A e. *pho*R f. *rpo*S g. *gac*A.

exoenzyme along with secondary metabolites [24].

Secondary metabolite synthesis is also negatively regulated by rpoS protein, well known as sigma factor, mediating transcription of other stress related genes. Suh et al. [26] demonstrated its regulatory activity in various kinds of stress conditions through *rpo*S mutant study in *P. aeruginosa*. Similarly, *rpo*N produces sigma 54 and known to be involved in transcriptional regulatory activity [27]. Besides, the other transcription regulators are *cbr*A, *alg*B and *bfi*S. The role of *clp*P is in ATP dependent protease system, which couples with *clp*A and the product formed is a cylindrical core of serine protease. ClpP constitutes the subunit of ATP dependent ClpP protease, exhibiting peptidase activity [28] in addition to other functions such as post translational modification, intra cellular trafficking, attachment, motility and antibiotic resistance. In our search, complete protein domain information of *rpo*N, *rpo*S and *clp*P was found available in databases in all the selected orthologs. Polysaccharide synthesis related gene *psl*A encodes for sugar transferase required for lipopolysaccharide synthesis during biofilm formation in non-mucoid strains without relying on alginate, which is a major component of biofilm. Alginate based exoplolysaccharide synthesis in *P. aeruginosa* is controlled by protein products of *alg*B and *alg*D genes; the latter produces GDP-mannose dehydrogenase [29].

Among the 16 genes analyzed in the present study, *ace*A was known to be involved in energy production and conservation by encoding isocitrate lyase (ICL), an enzyme of TCA cycle. This gene is

monocistronic having an independent existence *i.e.* it does not rely on flanking genes for its function [30]. In *P. aeruginosa*, three main genes (*bfm*R, *bfi*S and *mif*R) have been characterized for similar functions, in the absence of which fabrication of biofilm tends to collapse as explicitly manifested by Petrova and Sauer [31]. For the genes *ace*A, *bfi*S and *mif*R, information about protein functional domain is not available in our chosen orthologs. Therefore, primers were spotted in the conserved region which was corresponding to the protein domains of the reference strain *P. aeruginosa*. Osmoregulated periplasmic glucans (OPGs), a membrane glycosyltransferase, has been reported to be the result of *opg*G and *opg*H gene transcription where, the protein opgG also reconciles inorganic ion transport and metabolism [32]. Another key enzyme for the transport of ions and nucleotides is ppx encoding exophosphatases causing hydrolysis of polyphosphate [33]. One of the genes analyzed in this study was *typ*A which has a role in virulence of *P. aeruginosa* [34]. This is also a highly-conserved gene present in other pathogenic bacteria such as *Vibrio cholerae*, *Yersinia pestis*, *Mycobacterium tuberculosis* and *Escherichia coli* [35–37]. These genes that are validated in the study since are known to play key roles in bacterial biofilm formation, they could be considered as signature genes for bacterial biofilms. In this context, the simple approach described for analyzing these genes could be widely used for metagenomic studies in other environments.

Metagenomic analysis to understand microbial diversity in environmental samples has been shown usually through 16S rRNA clone libraries. Next Generation Sequencing (NGS) and tag pyrosequencing helps in deeper analysis of microbial diversity [38,39]. Chao et al. [10] were able to generate and identify occurrences and abundances of specific genes including those involved in glutathione metabolism, SoxRS system, OxyR system, RpoS regulated genes and production of extracellular polymeric substances. Based on specific pattern frequencies and tetranucleotides, it was possible to construct near complete draft genomes. Binning to cluster nucleotide sequences of NGS analysis based on base composition to identify functional roles of microorganisms is yet another approach described [40]. Recently, gene-centric metagenome analysis to identify specific functionality based microorganisms in a community gained importance in parallel. Bonilla-Rosso et al. [41] designed and evaluated primers that targeted genes encoding nitrate reductases. However, the design of PCR primers was based on known *nir*S and *nir*R genes of denitrifying microorganisms. Similarly, PCR amplification of *nif*H gene was used to develop and analyze a metagenomic *nif*H gene clone library from soil samples [42]. Muller et al. [43] used gene specific primers to amplify non ribosomal peptide synthetases to investigate bioactive substance related genes in old vegetation forms. Huson et al. [44] performed protein-alignment-guided assembly of ortholog gene families from microbiome sequencing reads. This was done in the short reads obtained from microbiome projects. Using this approach, a gene family can be selected based on classification such as KEGG and the reads binned to the gene family can be assembled.

In our study, freshwater metagenome was targeted to validate a simplified strategy of gene analysis based on known orthologs without the need for metagenome/microbiome sequencing. The primers designed based on the conserved protein domains in orthologs of other strain, species and genus were able to amplify ortholog genes in entirely different species of *Pseudomonas* which were not used for primer design (Table 2). Primers of all the seven genes could amplify orthologs from wide range of species of *Pseudomonas* including *P. balenrica* (*ace*A), *P. alcaligenes* (*clp*P), *P. stutzeri* (*typ*A), *P. putida* (*cbr*A), *P. mendocina* (*pho*R), *P. pseudoalcaligenes* (*rpo*S) and *P. resinovorans* (*gac*A). This indicates i) presence of these *Pseudomonas* species in the pond water samples studied ii) distribution of biofilm genes across various species of *Pseudomonas* iii) conservation of functional domains in proteins coded by these biofilm genes iv) ability of the our PCR primes to amplify ortholog genes in other bacteria which are not used for designing PCR primers.

Molecular fingerprinting of bacterial communities in fresh water biofilms can serve as water quality monitoring tool as well [45] as the community may differ according to the geographical location, seasonal environmental changes and contaminants. The sample sources used in the present study are mainly used for irrigation purpose and also a source of drinking water for farm animals. In addition, the rural poor in India use this water for house hold purposes except drinking. The bacterial biofilms in fresh water systems could serve as reservoirs for pathogenic microbes and cause water borne diseases [38] in human and cattle. Hence, it becomes indispensible to study the bacterial communities in pond water for monitoring the quality as well as to take up sanitation measures for preventing any water borne diseases. We strongly believe that the molecular approach described here using a simplified PCR amplification strategy could serve as a step forward towards understanding bacterial biofilms in fresh water ecosystems.

4. Materials and methods

4.1. Mining of biofilm related genes in P. aeruginosa PAO1-UW genome

Genes that are known to be involved in biofilm formation in *P. aeruginosa* PAO1-UW strain were filtered out from PGD [6] using 'biofilm' in text search. All information related to the biofilm genes *viz*. nomenclature, locus, size and function were collected.

4.2. Ortholog search using P. aeruginosa biofilm gene sequences

Sequences of biofilm related genes were used for similarity search using nucleotide BLAST. Among the BLAST search results, gene sequences of bacterial strains one each from different stain, different species and different genus to *P. aeruginosa* PAO-UW were chosen. If there were more than one bacterium in each category, the bacterium with highest query coverage and highest homology to the test sequences was selected. Matching gene sequences of the corresponding strains were collected from GenBank database and converted to properly aligned clusters of sequence in FASTA format.

4.3. Multiple sequence alignment using ClustalW

For each biofilm related gene of *P. aeruginosa*, four nucleotide sequences (gene sequences of PAO1-UW strain, a different strain of *P. aeruginosa*, a different species of *Pseudomonas* and a genus different from *Pseudomonas*) as selected from BLAST results were used for alignment in ClustalW software. Wherever required, sequences were converted to reverse complemented sequences since some of the BLAST matches were on negative strand of the bacterial genomes. For the gene sequences, which did not show BLAST matches with other genus, three sequences were used for alignment (PAO1-UW strain, different strain of *P. aeruginosa* and a different species of *Pseudomonas*). Similarly, for the gene sequences which did not show BLAST matches with other genus and other species, only two sequences were aligned (PAO1-UW strain and a different strain of *P. aeruginosa*).

4.4. Protein domain based comparison

Uniprot, Prosite, Pfam, Interpro and NCBI are the databases used to obtain information on the functional domain of the proteins encoded by biofilm related genes under study. Protein functional domains of the orthologs were also obtained from these databases. Thus, the amino acid sequences of each biofilm related protein of *P. aeruginosa* PAO1-UW was aligned with the protein sequences of other bacteria (other strain, species and genus) as obtained from nucleotide BLAST results. The nucleotide sequence corresponding to the protein domain was computed manually for each selected gene. From this nucleotide sequence, two similar conserved regions were selected visually from the ClustalW alignment for the design of PCR primers (17 nucleotides or

more contiguous sequences). The flanked region was fixed within the range 100–2000 bp nucleotides for easy analysis in PCR.

4.5. Sample collection

Freshwater samples (2.5 l) from two diagonally opposite sites of ten different freshwater ponds of Vellore (12°54′40″N 79°8′10″E) district, located at 220 m above the mean sea level (http://www.vellore.tn.nic.in/) in Tamil Nadu, India, were collected in sterile container during morning hours, during the month of September 2016 (monsoon weather) and pooled to 5 l of each, respectively. The water samples were checked for physical parameters that included pH, temperature, colour, electrical conductivity, total dissolved solids and salinity using portable multi-parameter (Oaklon Instrument, PCSTestr™ 35). All the samples were stored at 4 °C immediately to avoid further microbial growth and allotted unique identifier based of the sequential collection and location name (Supplementary Table 2). Fresh samples were filtered by Whatman filter paper (grade 1) used for DO and BOD analysis which was performed within a week using standard Winkler's method.

4.6. Isolation of metagenomic DNA

Metagenomic DNA from all the water samples was extracted according to the method described by Ranjan et al. [46] with few modifications. Sample volume of 2 l was centrifuged (Eppendorf 5804 R, Germany) at 8000*g* for 10 min to harvest total cells and suspended in 8 ml of extraction buffer (100 mM Tris-HC1, pH 8.0, 100 mM EDTA, 100 mM sodium phosphate, 1.5 M NaCl, 1% CTAB), followed by incubation at 37 °C for 15 min. Lysozyme (5 mg/ml) was added to the mixture and again incubated for 1 h at 37 °C with gentle shaking. Proteinase K (2.5 mg) and 1% SDS were added and incubated at 65 °C for 1 h. The lysate was centrifuged at 8000*g* for 10 min and to the supernatant, equal volume of phenol-chloroform (1:1) was added and aqueous phase was separated in a fresh vial. DNA of the aqueous phase was precipitated using 0.6 volume of isopropanol (kept undisturbed for 1 h) and centrifuged at 10,000*g* for 20 min at 4 °C. DNA pellet was washed using 70% ethanol, air dried and dissolved in TE buffer (10 mM Tris and 1 mM EDTA buffer, pH 8.0). DNA bands were checked in 0.7% agarose gel under UV transillumination (AlphaImager HP) and quantified using Nanodrop 2000 (ThermoScientific) facility.

4.7. PCR analysis

PCR reactions for amplification of biofilm genes that were filtered during computational analysis were optimized to give the specific expected band size using *P. aeruginosa* VITLWS3 DNA as template, which was isolated from VIT freshwater lake and has been reported to form biofilm mass of 0.48 at OD 590 nm [7]. For DNA isolation from *P. aeruginosa* VITLWS3, a loop full of colony from culture plate was inoculated in 50 ml sterile nutrient broth in flask and left overnight for incubation at 37 °C in shaker incubator. When OD_{600} exceeded 0.8, 1.5 ml of liquid culture was used to isolate genomic DNA using standard DNA isolation protocol [47]. The optimized PCR conditions (annealing temperature, template concentration, primer concentration and cycles) which were different for different genes were then used for metagenomic PCR analysis. PCR amplifications were performed in reaction volume of 20 µl containing template DNA in the range of 1–2 µl (50–100 ng), 1–2 µl (0.5–1 µM) forward and reverse primer, 10 µl PCR master mix [(2 ×)(Taq polymerase 2 units), Amplicon] and remaining volume of sterile water. Amplifications were carried out in thermal cycler (Eppendorf, Germany) under following conditions: 30 cycles of initial denaturation at 94 °C for 2 min, denaturation at 94 °C for 1 min, annealing at 51 °C for *rpo*N, 58 °C for *clp*P, 55 °C for *pho*R, 55 °C for *che*Y for 1 min, extention at 72 °C for 1 min and final extension at 72 °C for 7 min. Touchdown PCR was used for amplification of the genes *typ*A, *ace*A, *mif*R, *psl*A, *gac*A, *opg*H, *rpo*S, *cbr*A and *bfi*S with similar conditions except first annealing of 10 cycles at 50, 55, 50, 50, 47, 54, 49, 49 and 50 °C respectively followed by 20 cycles at 52, 58, 59, 59, 52, 57, 52, 52 and 58 °C. For all the genes, PCR was run with ten samples containing metagenomic DNA from ten water samples as template. The amplified products were analyzed in 1.5% agarose gel containing ethidium bromide (0.33 mg/l) for amplicon < 500 bp and 1.2% for 500 bp and visualized under short-wavelength UV light.

4.8. Sequencing and phylogenetic analysis

Amplified DNA bands of expected product size were excised from agarose gel using sterile scalpel. DNA was eluted from the gel band piece using SureTrap® Gel Extraction Kit. Randomly, two samples from each gene sample set were used for sequencing (Amnion Biosciences Pvt. Ltd., Bangalore). Raw sequence reads exported in FASTA format were fed in BLASTn and BLAST hits were scrutinized to confirm the gene function and validation. Top hundred hits belonging to different genus and different species of *Pseudomonas* in BLAST result were used for phylogenetic study of the gene among the orthologs. Molecular evolutionary genetics analysis software MEGA (6.06) [48] was used to construct phylogenetic tree using maximum likelihood method keeping bootstrap value of 1000 for higher accuracy.

The partial gene sequences were submitted in GenBank database using Bankit submission tool and accession number KX756976, KX772806, KX772807, KX772808, KX779039, KX779040 and KX831930 for gene *ace*A, *cbr*A, *clp*P, *pho*R, *rpo*S, *typ*A, and *gac*A respectively were obtained. For all the gene sequences the translated protein amino acid sequences were also added in the features with partial 5′ and 3′ coding end during submission.

5. Data Access

The sequences of the bacterial biofilm genes that were amplified from fresh water genome are available in the GenBank database with accession numbers KX756976, KX772806, KX772807, KX772808, KX779039, KX779040, KX831930 for genes *ace*A, *cbr*A, *clp*P, *pho*R, *rpo*S, *typ*A, and *gac*A respectively.

Acknowledgement

The authors gratefully acknowledge the support rendered by the management of VIT University, Vellore, India in performing this research work.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial or non-for-profit sectors.

Disclosure of statement

The authors declare that there are no conflict of interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2017.08.010.

References

[1] T.J. Battin, L.A. Kaplan, S. Findlay, C.S. Hopkinson, E. Marti, A.I. Packman, D. Newbold, F. Sabater, Biophysical controls on organic carbon fluxes in fluvial networks, Nat. Geosci. 1 (2008) 95–100.
[2] C. Griebler, F. Malard, T. Lefebure, Current developments in ground water ecology – from biodiversity to ecosystem function and services, Curr. Opin. Biotechnol. 27 (2014) 159–167.
[3] K. Besemer, Biodiversity, community structure and function of biofilms in stream

ecosystems, Res. Microbiol. 166 (2015) 774–781.
[4] A.M. Veach, J.C. Stegen, S.P. Brown, W.K. Dodds, A. Jumpponen, Spatial and successional dynamics of microbial biofilm communities in a grassland stream ecosystem, Mol. Ecol. 25 (2016) 4674–4688, http://dx.doi.org/10.1111/mec.13784.
[5] K. Velmourougane, R. Prasanna, A.K. Saxena, Agriculturally important microbial biofilms: present status and future prospects, J. Basic Microbiol. 57 (2017) 548–573.
[6] G.L. Winsor, D.K. Lam, L. Fleming, R. Lo, M.D. Whiteside, N.Y. Yu, R.E. Hancock, F.S. Brinkman, Pseudomonas genome database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes, Nucleic Acids Res. 39 (2011) D596–600.
[7] J. Kumari, D. Kumar, A. Mathur, A. Naseer, R.R. Kumar, P.T. Chandrasekaran, G. Chaudhuri, M. Pulimi, A.M. Raichur, S. Babu, N. Chandrasekaran, R. Nagarajan, A. Mukherjee, Cytotoxicity of TiO2 nanoparticles towards freshwater sediment microorganisms at low exposure concentrations, Environ. Res. 135 (2014) 333–345.
[8] J.L. Edwards, J.L. Smith, J. Connolly, J.E. McDonald, M.J. Cox, I. Joint, C. Edwards, A.J. McCarthy, Identification of carbohydrate metabolism genes in the metagenome of a marine biofilm community shown to be dominated by Gammaproteobacteria and Bacteroidetes, Genes 1 (2010) 371–384.
[9] B.L. Brown, R.V. LePrell, R.B. Franklin, R.V. Rivera, F.M. Cabral, H.L. Eaves, V. Gardiakos, K.P. Keegan, T.L. King, Metagenomic analysis of planktonic microbial consortia from a non-tidal urban impacted segment of James River, Stand. Genomic Sci. 10 (2015) 65, http://dx.doi.org/10.1186/s40793-015-0062-5.
[10] Y. Chao, Y. Mao, Z. Wang, T. Zhang, Diversity and functions of bacterial community in drinking water biofilms revealed by high-throughput sequencing, Sci Rep 5 (2015) 10044, http://dx.doi.org/10.1038/srep10044.
[11] H. Jin, B. Li, X. Peng, L. Chen, Metagenomic analyses reveal phylogenetic diversity of carboxypeptidase gene sequences in activated sludge of a wastewater treatment plant in Shanghai, China, Ann. Microbiol. 64 (2014) 689–697.
[12] L. Gutiérrez-Lucas, J. Montor-Antonio, N. Cortés-López, S. del Moral, Strategies for the extraction, purification and amplification of metagenomic DNA from soil growing sugarcane, Adv. Biol. Chem 4 (2014) 281–289.
[13] R.S. Mandal, S. Saha, S. Das, Metagenomic surveys of gut microbiota, Genomics Proteomics Bioinformatics 13 (2015) 148–158.
[14] G.J. Dick, A.F. Andersson, B.J. Baker, S.L. Simmons, B.C. Thomas, A.P. Yelton, J.F. Banfield, Community-wide analysis of microbial genome sequence signatures, Genome Biol. 10 (2009) R85, http://dx.doi.org/10.1186/gb-2009-10-8-r85.
[15] E.R. Hyde, F. Andrade, Z. Vaksman, K. Parthasarathy, H. Jiang, Metagenomic analysis of nitrate-reducing bacteria in the oral cavity: implications for nitric oxide homeostasis, PLoS One 9 (2014) 88645, http://dx.doi.org/10.1371/journal.pone.0088645.
[16] T. Varin, C. Lovejoy, A.D. Jungblut, W.F. Vincent, J. Corbeila, Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the high Arctic, Appl. Environ. Microbiol. 78 (2012) 549–559.
[17] C. Simon, A. Wiezer, A.W. Strittmatter, R. Daniel, Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome, Appl. Environ. Microbiol. 75 (2009) 7519–7526.
[18] S. Jia, Z. Wang, X. Zhang, B. Liu, W. Li, S. Cheng, Metagenomic analysis of cadmium and copper resistance genes in activated sludge of a tannery wastewater treatment plant, J. Environ. Biol. 34 (2013) 375–380.
[19] R.K. Kapardar, R. Ranjan, A. Grover, M. Puri, R. Sharma, Identification and characterization of genes conferring salt tolerance to *Escherichia coli* from pond water metagenome, Bioresour. Technol. 101 (2010) 3917–3924.
[20] S.K. Kushwaha, L. Manoharan, T. Meerupatil, K. Hedlund, D. Ahrén, MetCap: a bioinformatics probe design pipeline for large-scale targeted metagenomics, BMC Bioinf. 16 (2015) 1–11.
[21] H. Chen, B.M. Sharp, Content-rich biological network constructed by mining PubMed abstracts, BMC Bioinf. 5 (2004) 147.
[22] M.Y. Avasan, S.R. Ramakrishan, Effect of sugar mill effluent on organic resources of fish, Pollut. Res. 20 (2001) 167–171.
[23] M.N. Starnbach, S. Lory, The *fliA* (*rpoF*) gene of *Pseudomonas aeruginosa* encodes an alternative sigma factor required for flagellin synthesis, Mol. Microbiol. 6 (1992) 459–469.
[24] C. Reimmann, M. Beyeler, A. Latifi, H. Winteler, M. Foglino, A. Lazdunski, D. Haas, The global activator GacA of *Pseudomonas aeruginosa* PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase, Mol. Microbiol. 24 (1997) 309–319.
[25] M. Yamada, K. Makino, H. Shinagawa, A. Nakata, Regulation of the phosphate regulon of *Escherichia coli*: properties of *pho*R deletion mutants and subcellular localization of PhoR protein, Mol Gen Genet 220 (1990) 366–372.
[26] S.J. Suh, L. Silo-Suh, D.E. Woods, D.J. Hassett, S.E. West, D.E. Ohman, Effect of *rpoS* mutation on the stress response and expression of virulence factors in *Pseudomonas aeruginosa*, J. Bacteriol. 181 (1999) 3890–3897.
[27] S. Jin, K. Ishimoto, S. Lory, Nucleotide sequence of the *rpo*N gene and characterization of two downstream open reading frames in *Pseudomonas aeruginosa*, J. Bacteriol. 176 (1994) 1316–1322.
[28] A. Wawrzynow, D. Wojtkowiak, J. Marszalek, B. Baneckil, M. Jonsen, B. Graves, C. Georgopoulos, M. Zylicz, The ClpX heat-shock protein of *Escherichia coli*, the ATP-dependent substrate specificity component of the ClpP-ClpX protease, is a novel molecular chaperone, EMBO J. 14 (1995) 1867–1877.
[29] V. Deretic, J.F. Gill, A.M. Chakrabarty, Gene *algD* coding for GDPmannose dehydrogenase is transcriptionally activated in mucoid *Pseudomonas aeruginosa*, J. Bacteriol. 169 (1987) 351–358.
[30] A.L. Diaz-Perez, C. Roman-Doval, C. Diaz-Perez, C. Cervantes, C.R. Sosa-Aguirre, J.E. Lopez-Meza, J. Campos-Garcia, Identification of the *ace*A gene encoding isocitrate lyase required for the growth of *Pseudomonas aeruginosa* on acetate, acyclic terpenes and leucine, FEMS Microbiol. Lett. 269 (2007) 309–316.
[31] O.E. Petrova, K. Sauer, A novel signaling network essential for regulating *Pseudomonas aeruginosa* biofilm development, PLoS Pathog. 5 (2009) e1000668, http://dx.doi.org/10.1371/journal.ppat.1000668.
[32] Y. Lequette, E. Rollet, A. Delangle, E.P. Greenberg, J.P. Bohin, Linear osmoregulated periplasmic glucans are encoded by the *opg*GH locus of *Pseudomonas aeruginosa*, Microbiology 153 (2007) 3255–3263.
[33] A. Zago, S. Chugani, A.M. Chakrabarty, Cloning and characterization of polyphosphate kinase and exopolyphosphatase genes from *Pseudomonas aeruginosa* 8830, Appl. Environ. Microbiol. 65 (1999) 2065–2071.
[34] A. Neidig, A.T. Yeung, T. Rosay, B. Tettmann, N. Strempel, M. Rueger, O. Lesouhaitier, J. Overhage, TypA is involved in virulence, antimicrobial resistance and biofilm formation in *Pseudomonas aeruginosa*, BMC Microbiol. 13 (2013) 77, http://dx.doi.org/10.1186/1471-2180-13-77.
[35] M. Farris, A. Grant, T.B. Richardson, CD, BipA: a tyrosinephosphorylated GTPase that mediates interactions between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells, Mol. Microbiol. 28 (1998) 265–279.
[36] K. Scott, M.A. Diggle, S.C. Clarke, TypA is a virulence regulator and is present in many pathogenic bacteria, Br. J. Biomed. Sci. 60 (2003) 168–170.
[37] A.J. Grant, M. Farris, P. Alefounder, P.H. Williams, M.J. Woodward, C.D. O'Connor, Co-ordination of pathogenicity island expression by the BipA GTPase in enteropathogenic *Escherichia coli* (EPEC), Mol. Microbiol. 48 (2003) 507–521.
[38] M. Zhang, W. Liu, X. Nie, C. Li, J. Gu, C. Zhang, Molecular analysis of bacterial communities in biofilms of a drinking water clear well, Microbes Environ. 27 (2012) 443–448.
[39] N. Mukherjee, D. Bartelli, C. Patra, B.V. Chauhan, S.E. Dowd, P. Banerjee, Microbial diversity of source and point-of-use water in Rural Haiti – a pyrosequencing-based metagenomic survey, PLoS One 11 (2016) e0167353, , http://dx.doi.org/10.1371/journal.pone.0167353.
[40] I. Saeed, S.L. Tang, S.K. Halgamuge, Unsupervised discovery of microbial population structure within metagenomes using nucleotide base composition, Nucleic Acids Res. 40 (2012) e34, , http://dx.doi.org/10.1093/nar/gkr1204.
[41] G. Bonilla-Rosso, L. Wittorf, C.M. Jones, S. Hallin, Design and evaluation of primers targeting genes encoding NO-forming nitrite reductases: implications for ecological inference of identifying communities, Sci Rep 6 (2016) 39208, http://dx.doi.org/10.1038/srep39208.
[42] R. Soni, D.C. Suyal, S. Sai, R. Goel, Exploration of nifH gene through soil metagenomes of the western Indian Himalayas, 3 Biotech, 6 2016, p. 25, , http://dx.doi.org/10.1007/s13205-015-0324-3.
[43] C.A. Muller, L. Oberauner-Wappis, A. Peyman, G.C. Amos, E.M. Wellington, G. Berg, Mining of nonribosomal peptide synthetase and polyketide synthase genes revealed a high level of diversity in the sphagnum bog metagenome, Appl. Environ. Microbiol. 81 (2015) 5064–5072, http://dx.doi.org/10.1128/AEM.00631-15.
[44] D.H. Huson, R. Tappu, A.L. Bazinet, C. Xie, M.P. Cummings, K. Nieselt, R. Williams, Fast and simple protein-alignment-guided assembly of orthologous gene families from microbiome sequencing reads, Microbiome 5 (2015) 11, http://dx.doi.org/10.1186/s40168-017-0233-2.
[45] V. Loza, E. Perona, P. Mateo, Molecular fingerprinting of cyanobacteria from river biofilms as a water quality monitoring tool, Appl. Environ. Microbiol. 79 (2013) 1459–1472.
[46] R. Ranjan, A. Grover, R.K. Kapardar, R. Sharma, Isolation of novel lipolytic genes from uncultured bacteria of pond water, Biochem. Biophys. Res. Commun. 335 (2005) 57–65.
[47] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: a Laboratory Manual, 2nd ed., 3 Cold Springs Harbor Laboratory, Cold Springs Harbor, NY, 1989.
[48] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, Mol. Biol. Evol. 24 (2007) 1596–1599.