The ISME Journal (2008) 2, 790–793

www.nature.com/ismej

SHORT COMMUNICATION

A rare SAR11 fosmid clone confirming genetic variability in the '*Candidatus* Pelagibacter ubique' genome

Jack A Gilbert, Martin Mühling and Ian Joint
Plymouth Marine Laboratory, Prospect Place, Plymouth, UK

A sequence analysis is described of a fosmid clone from a coastal marine metagenomic library that contains a 16S rRNA gene with high sequence similarity to that of the SAR11 bacterium '*Candidatus* Pelagibacter ubique' HTCC1062. The sequence of the fosmid clone was 32 086 bp in length and contained 23 187 bp of the 48-kb hyper-variable region 2 (HVR2) present in the genome of '*Cand.* P. ubique'. However, half of the sequences within the HVR2 region of the fosmid clone show little sequence similarity to or have no representative homologues in the genome sequence of '*Cand.* P. ubique' HTCC1062. Given their putative functions, the acquisition of these genes suggests that SAR11 could harbour more diverse phenotypes than represented by the 16S rRNA taxonomy. Variation in SAR11 genomes from different locations might explain why SAR11 is abundant in so many diverse marine provinces.
The ISME Journal (2008) **2,** 790–793; doi:10.1038/ismej.2008.49; published online 22 May 2008
Subject Category: integrated genomic and post-genomic approaches in microbial ecology
Keywords: SAR11; pelagibacter; fosmid; genome plasticity; horizontal gene transfer; bacteria

Recent analysis of the Sargasso Sea and Global Ocean Survey metagenomes has provided evidence of regions of hyper variability in the genomes of SAR11 (Rusch *et al.*, 2007; Wilhelm *et al.*, 2007). This was achieved using fragment recruitment (Wilhelm *et al.*, 2007), whereby environmental sequence data were compared against the genome of '*Candidatus* Pelagibacter ubique' HTCC1062 and then assembled (Rusch *et al.*, 2007).

We have confirmed a region of hyper variability in an SAR11 genome by screening a metagenomic fosmid library for SAR11-like 16S rRNA gene-containing clones. SAR11-specific 16S rRNA gene primers (Béjà *et al.*, 2000) were used to identify clones that cover the 48-kb hyper-variable region 2 (HVR2) of '*Cand.* P. ubique' HTCC1062. A clone was identified that contained an SAR11 16S rRNA gene sequence from a library of 10 000 clones, constructed with surface seawater from the Western Channel Observatory (WCO: http://www.westernchannelobservatory.org.uk/, 50.25°N: 04.212°W) (see Supplementary Information for description of methods).

PCR screening of the fosmid library yielded only one positive clone (clone 01–003783: GenBank accession number EU410957; G/C content 43.2%), which contained a 16S rRNA gene with 98.8% sequence identity to that of '*Cand.* P. ubique' HTCC1062. Phylogenetic analysis of the 16S rRNA gene sequence confirmed that it belongs to the SAR11 clade of bacteria (Supplementary Figure S1). The small number of SAR11 fosmid clones produced is consistent with the findings of others (Suzuki *et al.*, 2004; DeLong *et al.*, 2006) and is puzzling given the high abundance and ubiquitous distribution of this clade (Morris *et al.*, 2002). It has been suggested that underrepresentation might be due to the presence of genes, in proximity to the ribosomal RNA operon, which express products toxic to the host cell (Béjà *et al.*, 2000). To our knowledge, there is no evidence of such genes from published genomes.

Table 1 summarizes the identity and location of each coding sequence (CDS) of fosmid clone 01–003783. Apart from the ribosomal RNA operon, a total of 25 complete and one partial CDS were predicted. Of these, only 16 showed high sequence similarity to genes from '*Cand.* P. ubique' HTCC1062 (Table 1). To exclude the possibility that inclusion of these CDSs in the fosmids resulted from a chimeric cloning event, PCR primers were designed to amplify fragments between each of the CDSs from 13 to the rRNA operon (Figure 1). All of these regions were detected in the environmental DNA used to produce the fosmid library. As all CDSs were complete, a chimera could only have formed

Correspondence: JA Gilbert, Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, UK.
E-mail: jagi@pml.ac.uk
Received 6 March 2008; revised 22 April 2008; accepted 24 April 2008; published online 22 May 2008

Table 1 Details of coding sequences (CDSs) and ribosomal RNA sequences from the fosmid clone (EU410957)

CDS[a]	*Nucleotide range*	*Direction of transcription*	*% GC content*	*Predicted function*	*BLASTp analysis*			*Cand. P. ubique HTCC1062 gene locus*[d]
					Most similar homologue[b]	*% Amino-acid identity*	*% Coverage*[c]	
1	2–502	+	55.25	Thioesterase (*paaI*)	ZP_01011710	70	93	None
2	3505–1803	+	52.38	DNA-damage inducible protein F	ZP_00958559	59	99.5	SAR11_1228
3	1816–2871	–	52.5	Dihydroorotate dehydrogenase	ZP_01055946	67	100	SAR11_0209
4	2871–3206	–	51.62	DUF952	YP_168113	59	99.5	SAR11_0785
5	3440–4896	+	50.32	5′-nucleotidase	YP_684099	70	99.8	None
6	4982–6274	–	52.55	Serine hydroxymethyl- transferase	YP_684127	87	100	None
7	6444–7202	+	53.02	COG0061 sugar kinase	YP_168147	69	99.6	SAR11_1132
8	7205–8638	–	55.46	Amidase	ZP_01057112	61	99.6	SAR11_0263
9	8764–10650	–	54.87	Propionate-CoA ligase	YP_684123	78	100	SAR11_1132
10	10777–12141	–	50.88	Diguanylate cyclase	ZP_01747305	36	100	None
11	12233–12508	+	50.9	Hypothetical protein	ZP_00960013	49	96.7	None
12	12568–13257	+	47.04	HAD hydrolase	ZP_01034650	52	98.6	SAR11_0464
13	13372–14928	+	53.97	Trimethylamine methyltransferase	YP_682774	76	96.5	None
14	15030–15644	+	46.6	Hypothetical protein	YP_167955	39	94.2	None
15	15586–16590	+	51.79	Diguanylate cyclase	ZP_01015615	38	94	None
16	17056–17598	+	52.01	Putative adenine deaminase	ZP_01750103	74	54	None
17	17451–17963	–	34.11	Pyridoxal phosphate-dependant aminotransferase	NP_893375	60	87.1	SAR11_1317
18	17967–18941	–	24.03	Epimerase	ZP_01385431	30	100	SAR11_0078
19	18954–20006	–	28.83	Acetoin dehydrogenase β-chain	YP_487412	44	98.8	None
20	20011–20967	–	25.5	Acetoin dehydrogenase α-chain	YP_487413	45	95.6	None
21	20976–21632	–	26.7	SAM binding protein	NP_949261	51	100	SAR11_0528
22	21685–23187	+	25.7	Cytidyltransferase	NP_772062	30	100	SAR11_0539
23	23364–26241		44.9	23S rRNA gene	CP000084	np	np	SAR11_23S_rRNA
24	26248–26323		59.2	TRNA-Ala	CP000084	np	np	SAR11_tRNA_f_5
25	26336–26412		57.1	TRNA-Ile	CP000084	np	np	SAR11_tRNA_f_4
26	26513–27985		48.7	16S rRNA gene	CP000084	np	np	SAR11_16S_rRNA
27	28323–29612	+	27.6	M37 peptidase	AAR37629	67	93.5	SAR11_0521
28	29625–30032	–	27.7	Hypothetical protein	ZP_01264309	67	69.1	SAR11_0520
29	30049–30888	–	23.6	HemK methylase	AAR37631	58	100	SAR11_0519
30	30896–31963	–	30.5	Protein chain release factor A	ZP_01264311	84	100	SAR11_0518

Abbreviation: np, not performed.
[a]CDS number from 5′ to 3′ on cloned insert.
[b]As compared to the non-redundant protein sequence database in GenBank on 15 April 2008.
[c]Percentage of CDS covered by the most similar homologue from the BLASTp analysis.
[d]'*Candidatus* Pelagibacter ubique' homologues were inferred by the presence of same or closely related gene nomenclature in the HTCC1062 genome annotation available in GenBank.

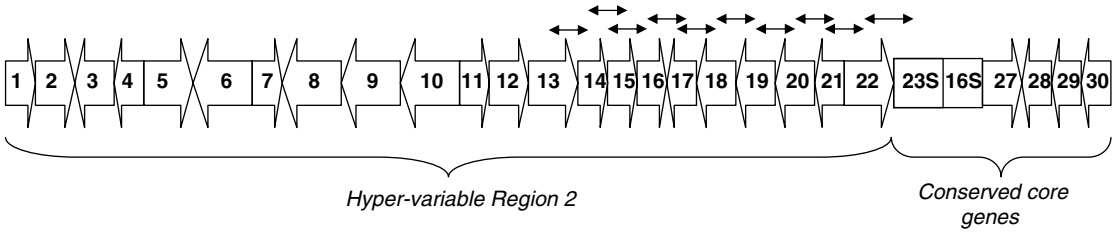

Figure 1 Diagram of the fosmid clone sequence. Large arrows represent relative size and orientation of open reading frames (ORFs). 16S, 23S rRNA and the intergenic transcribed spacer region (ITS) are represented by two boxes titled 23S and 16S. Small solid arrows represent regions amplified by PCR from environmental DNA to validate clone structure. Core conserved genes represent area with synteny to '*Candidatus* Pelagibacter ubique'.

between the CDSs. By amplifying and sequencing these regions from environmental DNA, we have demonstrated that this pattern of genes existed prior to library construction and hence cannot be a chimera.

Only the four CDSs transcribed in the 3′-direction after the rRNA operon shared synteny with the '*Cand.* P. ubique' HTCC1062 and HTCC1002 genomes (Table 1). The 11 CDSs with no homologue in the genome of HTCC1062 represent potential niche-specific acquisitions that may provide a competitive advantage (Wilhelm *et al.*, 2007). For example, acquisition of a gene involved in 5′-nucleotidase and uridine diphosphate (UDP)-sugar hydrolysis (CDS 5) might benefit utilization of extracellular nucleotides as a phosphate source under conditions of phosphate starvation (Rittmann *et al.*, 2005). The missing gene products required for these processes to be fully functional may be encoded within the 25 kb of the 48-kb large HVR2 that were not part of the sequence of fosmid clone 01–003783. However, we were not successful in identifying a contiguous clone within the fosmid clone library. Their absence suggests that these genes are either non-functioning or involved in other processes. The former is unlikely due to the genomic streamlining of SAR11 organisms (Giovannoni *et al.*, 2005), which would result in the loss of non-functioning elements.

It is interesting that 22 of the CDSs of fosmid clone 01–003783 show higher sequence similarity and have a closer phylogenetic relationship to the respective sequences of bacteria not belonging to the SAR11 clade (Table 1; Supplementary Figure S2). This is in contrast to the phylogenetic inference based on the 16S rRNA gene sequence within the same fosmid clone (Supplementary Figure S1). A possible explanation for this may be that CDSs 1–22 are a result of horizontal gene transfer. Although there is no direct evidence for phage-mediated horizontal gene transfer in fosmid clone 01–003783, the possibility is supported by the low G/C content (potentially indicative of phage genes) of six of the CDSs in the fosmid (Table 1). Also, phage integrase genes have been documented in the other 49% of HVR2 that was not in this fosmid clone (Rusch *et al.*, 2007). The ecological ramifications of horizontal gene transfer for SAR11 bacteria have been discussed previously (for example, Rusch *et al.*, 2007; Wilhelm *et al.*, 2007).

In conclusion, the presence of a region of extreme variability within bacterial genomes (Rusch *et al.*, 2007, Wilhelm *et al.*, 2007) has been confirmed in fosmid clone 01–003783; this is a rare example of a fosmid that contains the SAR11 16S rRNA gene. The variable region HVR2 provides evidence in support of the hypothesis of bacterial promiscuity (Coleman *et al.*, 2006; Fraser *et al.*, 2007, Vergin *et al.*, 2007). In particular, this study demonstrates the presence of genes within the HVR2 region, which are not present in the genome of '*Cand.* P. ubique' HTCC1062 or HTCC1002. Therefore, fosmid clone 01–003783 provides further evidence for the genomic plasticity of this bacterial lineage and adaptability to specific environmental conditions, which may have been acquired through horizontal gene transfer.

Acknowledgements

This work is supported by a grant from the Natural Environment Research Council (NE/C507902/1) and is part of the core research programme of the Plymouth Marine Laboratory, a collaborative centre of NERC.

References

Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP *et al.* (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**: 516–529.

Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.

Fraser C, Hanage WP, Spratt BG. (2007). Recombination and the nature of bacterial speciation. *Science* **315**: 476–480.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al.* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.

Rittmann D, Sorger-Herrmann U, Wendisch VF. (2005). Phosphate starvation-inducible gene *ushA* encodes a 5′ nucleotidase required for growth of *Corynebacterium glutamicum* on media with nucleotides as the phosphorus source. *Appl Environ Microbiol* **71**: 4339–4344.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical pacific. *PLoS Biol* **5**: 398–431.

Suzuki MT, Preston CM, Beja O, de la Torre JR, Steward GF, DeLong EF. (2004). Phylogenetic screening of ribosomal RNA gene-containing clones in Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey Bay. *Microb Ecol* **48**: 473–488.

Vergin KL, Tripp HJ, Wilhelm LJ, Denver DR, Rappe MS, Giovannoni SJ. (2007). High intraspecific recombination rate in a native population of *Candidatus* Pelagibacter ubique (SAR11). *Environ Microbiol* **9**: 2430–2440.

Wilhelm LJ, Tripp HJ, Givan SA, Smith DP, Giovannoni SJ. (2007). Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* **2**: 27.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)