environmental
microbiology reports

Environmental Microbiology Reports (2015) doi:10.1111/1758-2229.12297

Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator–prey genotypes

Naseer Sangwan,[1,2] Carey Lambert,[3] Anukriti Sharma,[1] Vipin Gupta,[1] Paramjit Khurana,[4] Jitendra P. Khurana,[4] R. Elizabeth Sockett,[3] Jack A. Gilbert[2,5,6] and Rup Lal[1]*

[1]*Department of Zoology, University of Delhi, Delhi 110007, India.*
[2]*Biosciences Division (BIO), Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA.*
[3]*Institute of Genetics, School of Life Sciences, Nottingham University, Queen's Medical Centre, Nottingham, UK.*
[4]*Interdisciplinary Centre for Plant Genomics & Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110021, India.*
[5]*Department of Ecology and Evolution, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA.*
[6]*College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China.*

Summary

***Bdellovibrio bacteriovorus* are small *Deltaproteobacteria* that invade, kill and assimilate their prey. Metagenomic assembly analysis of the microbial mats of an arsenic rich, hot spring was performed to describe the genotypes of the predator *Bdellovibrio* and the ecogenetically adapted taxa *Enterobacter*. The microbial mats were enriched with *Bdellovibrio* (1.3%) and several Gram-negative bacteria including *Bordetella* (16%), *Enterobacter* (6.8%), *Burkholderia* (4.8%), *Acinetobacter* (2.3%) and *Yersinia* (1%). A high-quality (47 contigs, 25X coverage; 3.5 Mbp) draft genome of *Bdellovibrio* (strain ArHS; Arsenic Hot Spring) was reassembled, which lacked the marker gene *Bd0108* associated with the usual method of prey interaction and invasion for this genus, while maintaining genes coding for the hydrolytic enzymes necessary for prey assimilation. By filtering microbial mat samples (< 0.45 μm) to enrich for small predatory cell sizes, we observed *Bdellovibrio*-like cells attached side-on to *E. coli* through electron microscopy. Furthermore, a draft pan-genome of the dominant potential host taxon, *Enterobacter cloacae* ArHS (4.8 Mb), along with three of its viral genotypes ($n = 3$; 42 kb, 49 kb and 50 kb), was assembled. These data were further used to analyse the population level evolutionary dynamics (taxonomical and functional) of reconstructed genotypes.**

Received 12 July, 2014; accepted 13 April, 2015. *For correspondence. E-mail ruplal@gmail.com; Tel. +91-011-27666254; Fax +91-011-27667443.

Introduction

Bdellovibrio bacteriovorus is an obligate aerobe and actively motile *Deltaproteobacterium* that preys upon a wide range of Gram-negative bacteria. *Bdellovibrio* predation generally involves attachment to the prey or host cell, invasion of the prey periplasm, attachment to the inner-membrane and eventually formation of a structure called 'bdelloplast'. In the bdelloplast, the prey's cellular macromolecules are assimilated for growth and division using an arsenal of hydrolytic enzymes (Rendulic *et al.*, 2004; Hobley *et al.*, 2012). Other species, e.g. *Bdellovibrio exovorus* JSS (Koval *et al.*, 2013; Pasternak *et al.*, 2014) attaches to and kills Gram-negative prey without entering the cell. *Bdellovibrio bacteriovorus* was first isolated in 1963 (Stolp and Starr, 1963), yet the mechanisms of predation in natural environments remain poorly characterized, especially across extreme environments with very low genetic diversity. Two complete genome sequences (HD100 (Rendulic *et al.*, 2004), Tiberius (Hobley *et al.*, 2012)) and a comparative genome analysis with *B. exovorus* (Pasternak *et al.*, 2013) has been published for this genus. These reference genomes have partly resolved the genetics of the 'intra-periplasmic' and 'epibiotic' mode of predation by *B. bacteriovorus* and *B. exovorus* species respectively.

Here, we report the first characterization of an environmental *Bdellovibrio* using metagenomics. Samples of microbial mats (surface temperature of 57°C ± 2) were isolated from a Himalayan hot spring located in the Parvati river valley (Manikaran, India), which harbours a unique and extremophillic microbiota (Dwivedi *et al.*, 2012). Extreme heat (surface water temperature of > 95°C), high concentrations of arsenic (140 ppb), heavy metals and dissolved CO_2 (14.7 ± 0.1 cm^3 STP/L), combined with low O_2 potential (4.8 ± 0.2 cm^3 STP/L) provide intense selective pressure. Using comparative genomics, we identified the differences in 'predation-specific' gene content (Pasternak *et al*., 2013; 2014) between the two reference genomes (HD100 and Tiberius) and the reconstructed *Bdellovibrio* genome (strain ArHS). We also reconstructed a draft (4.8 Mb) pan-genome of the eco-genetically adapted taxa, i.e. *Enterobacter* ArHS and its phage genotypes and report microscopic evidence for the presence of a *Bdellovibrio*-like bacterium in this environment.

Results

Sample collection and geochemical analysis

Two microbial mat samples were collected from Manikaran hot springs located in the geothermal field of the Parvati River which is a 45 Km long stretch with multiple thermal discharges (Fig. S1). The hot water springs in this valley range in temperature from 32°C to 96°C (Geological survey of India, 2011) with the hottest surface temperatures found at Manikaran (96°C; 32°02'N, 72°21'E). At the sampling site (32°01′34.8″N, 077°20′50.3″E) the average ambient temperature of the hot spring water and the two microbial mat samples (MB-A and MB-B) were ~91°C ± 3 and ~57°C ± 2 respectively. Both the water and microbial mats had high concentrations of As (140ppb ± 20 and 80 $\mu g\ g^{-1}$ respectively), Mn (136 ppb ± 30 and 300 $\mu g\ g^{-1}$ respectively) and Fe (123 $\mu g\ g^{-1}$ ± 32 and 500 $\mu g\ g^{-1}$ respectively; Table S1). However, Ba (121 $\mu g\ g^{-1}$ ± 32) and Sr (184 $\mu g\ g^{-1}$) concentrations were observed only using X-ray fluorescence analysis of microbial mat samples [not with inductively coupled plasma mass spectrometry (ICPMS)]. While physiochemical analysis of water and microbial mat samples validated previous observations (Chandrasekharam and Antu, 1995), including pH (7.1 in water and 7.4 in mats), H_2S (0.5 ppm ± 0.01) and HCO_3 (140 mg l^{-1} ± 23) concentrations in water, the exceptionally high concentrations of As, Mn and Fe were never reported before.

Microbial taxonomic composition

The metagenomes of the two microbial mats (MB-A: 4.6 Gb and MB-B: 3.9 Gb; Fig. S2A) had distinct *k*-mer patterns, different from hot spring microbiome samples from the Yellowstone River National Park (USA), including the Mushroom, Octopus and Bison hot springs (Bhaya *et al*., 2007; Schoenfeld *et al*., 2008). Individual read-based *k*-mer patterns showed that Manikaran samples had greatest correlation to each other ($R^2 = 0.90$), while the ‘Mushroom’ data were significantly more correlated to ‘Manikaran’ (MB-A: $R^2 = 0.78$ and MB-B: $R^2 = 0.71$; Fig. S2A) than the Octopus or Bison data (maximum R^2 value = 0.55). These differences are expected as there are several geochemical differences between Manikaran and Yellowstone National Park hot springs. The taxonomic structure of the two Manikaran metagenomes (MB-A and MB-B) was highly similar based on MetaPhlAn (Segata *et al*., 2012) and relative abundance patterns from 16S rRNA gene mapping analyses ($R^2 = 0.906$, $P < 1e^{-15}$; Fig. S3). *Proteobacteria* (MB-A: 50.3% and MB-B: 38.6%), *Firmicutes* (MB-A: 35.6% and MB-B: 37.6%), *Bacteroidetes* (MB-A: 2.7% and MB-B: 3.1%), and *Cyanobacteria* (MB-A: 3.8% and MB-B: 2.6%) dominated, while the major difference was in the abundance of *Planctomycetes*, with an 11.6% difference between MB-A and MB-B samples (Fig. S3). Further analysis revealed that the assemblages were mostly dominated by the *Bacilli* (MB-A: 22.4% and MB-B: 28.4%), *Gammaproteobacteria* (MB-A: 28.4% and MB-B: 22.6%), *Clostridia* (MB-A: 14.5% and MB-B: 15.02%), *Deltaproteobacteria* (MB-A: 7.3% and MB-B: 7.5%) and *Alphaproteobacteria* (MB-A: 6.7% and MB-B: 7.1%); and within these, a predominance of *Enterobacteria, Bdellovibrio, Clostridium, Citrobacter and Achromobacter* was clearly evident (Fig. S3).

Community metabolic potential and genetic adaptation against chemical stress

Individual metagenome read-based abundance of functional genes (KEGG categories) was highly correlated between the two metagenomes ($R^2 = 0.98$; Fig. S2B). Discounting core metabolism, several ‘pathogenesis and predation’-related genetic traits were observed, including antibiotic production (MB-A = 3.1% and MB-B = 2.9%), bacterial chemotaxis (MB-A = 1.37% and MB-B = 1.32%) and flagellar assembly (MB-A = 1.49% and MB-B = 1.37%). A complete pathway for bacterial chemotaxis (26 genes, KEGG ID: k02030, representing > 5 taxa) and flagellar assembly (41 genes, KEGG ID: k02040) was reconstructed in both the datasets (Fig. S2B).

Using total metagenome assembly, 27 contigs were found corresponding to the bacterial arsenic resistance operon (BLASTX, E-value = 10^{-5}; Silver and Phung, 2005), including *ArsA, B, C, R and H* (Table S2). These genes were assigned to *Staphylococcus*, *Klebsiella*, *Salmonella*, *Cronobacter*, *Escherichia* and *Achromobacter*. Genes associated with the respiratory arsenic reductase (*Arr*) operon were not observed, but an arsenic oxidase large subunit (*AsoA*) gene was assembled, and was closely related to the *AsoA* gene of *Chloroflexus aurantiacus* J-10-fl (nucleotide identity = 80% and query coverage = 99%). Manganese-dependent superoxide dismutase (Mn-*SodA*) gene homologues were observed on 14 contigs (Table S2) and were assigned to *Enterobacter, Xanthomonas, Pseudomonas, Delftia,* *Achromobacter, Cirobacter, Klebsiella and Shigella*. Anoxygenic photosynthesis potential was observed, with 11 *pufM* and 19 *pufL* gene homologues identified in the metagenome assembly with homology (BLASTX, E-value = 10^{-5}) to *Methylobacterium, Rhodospirillium, Halochromatium, Methyloversatillus, Rhodobacter* and *Chloroflexus*.

Phylogenetic characterization of assembled metagenomic contigs

To further resolve the microbial diversity of the Manikaran microbial mats, the metagenomes were combined and assembled ($n = 83$, 656, > 500 bp contig^{-1}). Phylogeny was assigned to majority of the contigs ($n = 60$, 906; 75%) with 54321 (89% of total assigned) assigned to Bacteria (Fig. S4; average contig size = 3.5 kb, largest contig = 123 kb), 6104 assigned to Eukaryota (average contig size = 750 bp), 409 assigned to viruses (average contig size = 750 bp) and 72 assigned to Archaea (average contig size = 1.5 kb). Taxonomic analysis of the assembled contigs revealed 298 bacterial genera, with *Proteobacteria* comprising mainly of Gram-negative genera i.e. *Bordetella* (16%), *Burkholderia* (4.8%)*, Yersinia* (1%)*, Acinetobacter* (2.3%)*, Salmonella* (10%) *and Xanthomonas* (6.8 %) (Fig. S4). The genus *Bdellovibrio* comprised 0.45% (437 contigs) of the assembled fragments, but with an average contig length (40 kb) significantly longer than either the other genera (2 kb) or the whole metagenome assembly (750 bp).

Population genomes reconstruction

(i) The Bacterial Predator: *Bdellovibrio* ArHS

Genetic abundance of the *Bdellovibrio* genotypes was validated ($P < 0.0001$; Fisher's Exact *t*-test) across the microbial mat samples using three different microbial diversity typing methods i.e. 16S rRNA (MB-A: 0.8% and MB-B: 0.9%), MetaPhlAn (MB-A: 1.3% and MB-B: 0.91%) and BLAST2LCA (437 contigs). Tetranucleotide frequency correlation, %G + C and coverage-based clustering was performed on longer (> 10 kb) contigs ($n = 23$ 906) and an individual graph was constructed for each cluster. Clustering analysis revealed that larger contigs were making a separate cluster ($n = 154$, largest contig = 123 kb and average contig length 25 kb). BLAST2LCA (BLASTN-based) analysis of this cluster against National Center for Biotechnology Information (NCBI)-nt database (ftp.ncbi.nlm.nih.gov/blast/db/FASTA/) suggested the contigs were most similar to *B. bacteriovorus*. Reciprocal smallest distance (RSD) analysis (Wall and Deluca, 2007) revealed that the assembly (3.5 Mb) had 2270 orthologous genes with HD100 (3.78 Mb: 3,587 genes), 2292 with Tiberius (3.98 Mb: 3738 genes) and 1771 with *B. exovorus* JSS (2.65 Mb: 2,618 genes). The assembly was refined using PRICE (Ruby *et al.*, 2013) to 47 contigs, which was named '*Bdellovibrio* ArHS', and comprised a near-complete (compared with *B. bacteriovorus*) draft genome size of 3.5 Mb, with 30 tRNA and 3468 protein coding genes and an average of 47% G + C content. Tetranucleotide profile correlation, single-copy marker gene (19/31), bacteria-specific conserved gene (83/107) (Dupont *et al.*, 2012) profile comparison (Fig. S5) and average nucleotide identity analysis (ANI) (Konstantinidis and Tiedje, 2005) of the reconstructed '*Bdellovibrio* ArHS' and three reference genomes (HD100, Tiberius and *B. exovorus* JSS) revealed that the reconstructed *'*ArHS*'* genotype was a genetically distinct *Bdellovibrio* species (Fig. 1A).

Additionally, OSfinder analysis (Hachiya *et al.*, 2009) predicted six genomic regions (total 1.9 Mb) as orthologous segments inherited in the three *Bdellovibrio* genomes. Enzyme level differentiation was typed (BLASTP at E-value cut-off 10^{-5}) across these genomes. Pairwise dN/dS analysis of orthologous genes predicted from RSD analysis highlighted (Fig. 2A) the strength of negative selection (dN/dS < 1) across this thermophillic *B. bacteriovorus* strain. Comparative genomic analysis of *Bdellovibrio* ArHS, HD100 and Tiberius revealed (Fig. S5C) novel genetic elements in ArHS, including *bcpB* (bacterioperoxidin), cyanophycinase, *radC*, four HNH-endonuclease, three esterases (*lpqC*, *lpqP* and *lptB*), one IS element (IS*6521*), a superoxide dismutase gene and an aero tolerance operon. Additionally, this comparison identified the presence of two 'hit locus' regions, i.e. the genomic regions of predatory *Bdellovibrio* taxa associated with predation versus host/prey independent growth (Cotter and Thomashow, 1992; Roschanski *et al.*, 2011). The hit locus regions of HD100, Tiberius and ArHS showed synteny except for the genes *Bd0105* and *Bd0108* which were congenitally deleted in *Bdellovibrio* ArHS (Fig. 1). Since previous studies have established *Bd0108* as a marker gene for the invasion of prey cells by *B. bacteriovorus* species (Cotter and Thomashow, 1992; Evans *et al.*, 2007; Roschanski *et al.*, 2011), the absence of this marker gene in ArHS was considered to indicate the existence of a hitherto concealed predatory mechanism, or that ArHS might be a host independent strain. However, ArHS maintained the majority of the genetic infrastructure required for prey/host attachment and assimilation including 86% of the 369 hydrolytic enzyme coding sequences and the complete profile ($n = 25$) of 'flagella and type IV pilus genes' as found in strain HD100 (Rendulic *et al.*, 2004). Detailed comparison of predation-specific gene content of the ArHS and *B. exovorus* JSS clearly revealed the species-specific demarcations, i.e. 39 hydrolytic

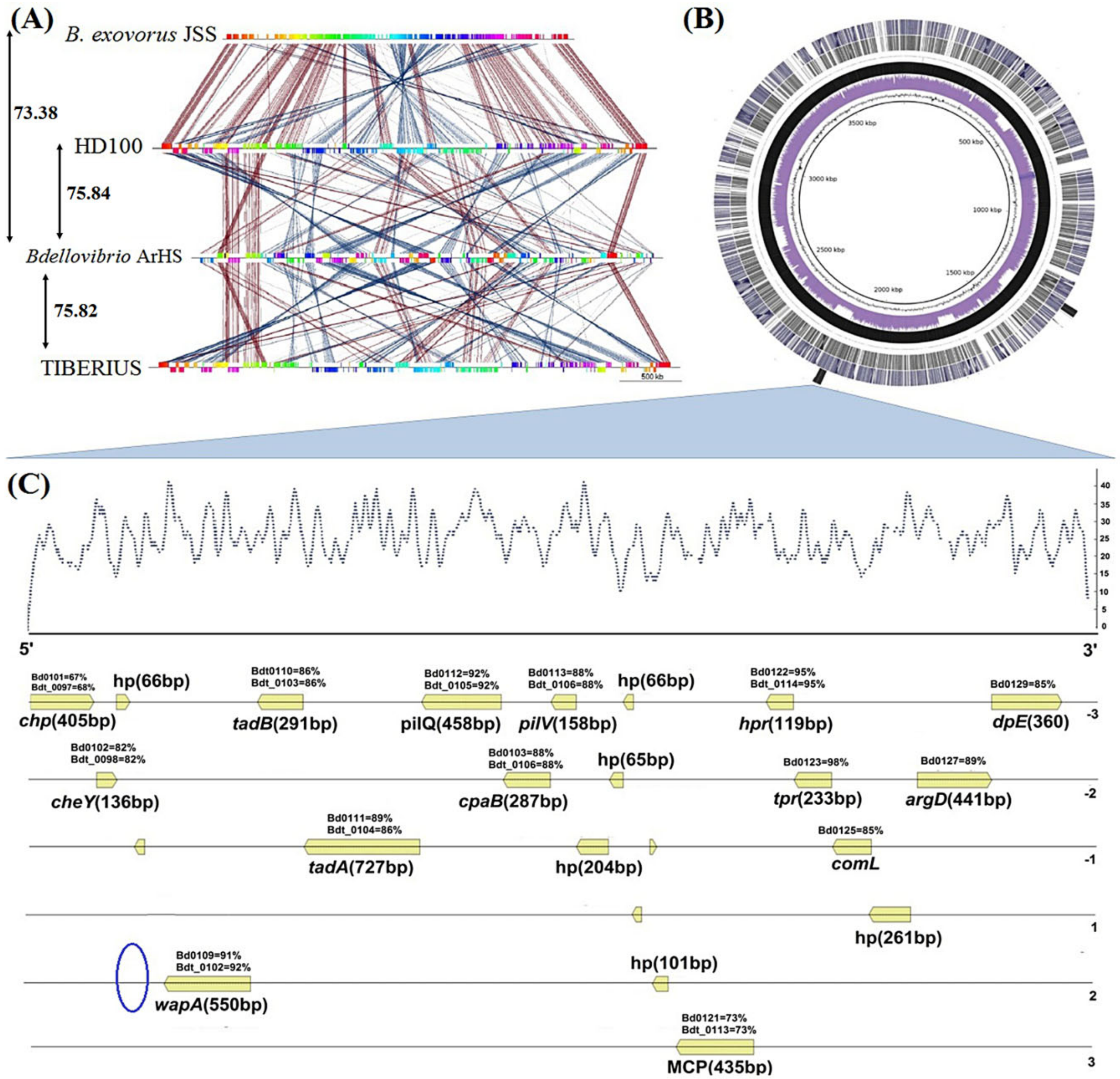
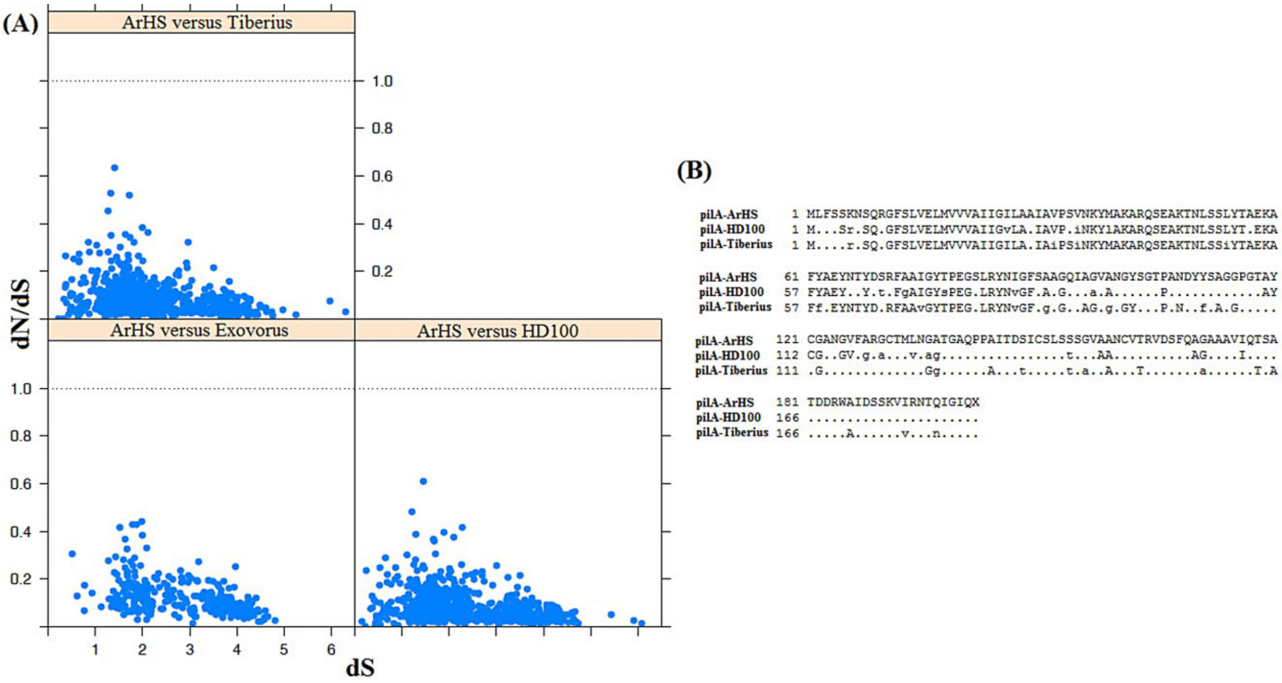


Fig. 1. Metagenomic reconstruction of the predator genotype **(A)** Whole genome based synteny comparisons of *B. bacteriovorus* HD100, Tiberius, ArHS and *B. exovorus* JSS (synteny bloc size: 500 kb) **(B)** Circular representation of the *Bdellovibrio* ArHS and its BLASTN based comparison with reference *B. bacteriovorus* strains; black color intensity represents the percentage identity. From outside towards the center: outermost circle, hit region (*Bd0101-Bd0129*) homologues of *B. bacteriovorus* HD100; circle 2, complete genome sequence of *B. bacteriovorus* Tiberius; circle 3, complete genome sequence of *B. bacteriovorus* HD100; circle 4, Single nucleotide polymorphisms (SNPs) predicted in the genome of *B. bacteriovorus*HD100; circle 5, *Bdellovibrio* ArHS contigs (ordered against HD100 genome); circle 6, read depth across reconstructed contigs (each coordinate represents 10bp region and minimum coverage cut-off = 15X); circle 7, Cumulative GC skew diagram for the *B. bacteriovorus* HD100 genome. **(C)** Synteny comparisons of hit locus region across of *Bdellovibrio* HD100 and ArHS. Upper panel represents the cumulative metagenome coverage and lower panel represents protein coding genes (all six frames) of *Bdellovibrio* ArHS hit locus and comparison (BLASTP) with their homologues in HD100 (locus tag = Bd) and Tiberius genomes (locus tag = Bdt).

enzymes, highly divergent hit locus, *pilA* (Fig. 2B) and *fliC* genes (Fig. S6). Recently, similar genetic differences (species specific) were also observed in a comparative genomic analysis targeted to resolve the epibiotic (*B. exovorus* JSS*)* and intra-periplasmic (*B. bacteriovorus* HD100) mode of predation across *Bdellovibrio* genotypes (Pasternak *et al*., 2013). Interestingly, the alignment of the PilA protein demonstrated its divergence in ArHS (Fig. 2B), which may reflect prey surface diversity or thermal tolerance of the pilus fiber; however, this still



```
pilA-ArHS       1 MLFSSKNSQRGFSLVELMVVVAIIGILAAIAVPSVNKYMAKARQSEAKTNLSSLYTAEKA
pilA-HD100      1 M...Sr.SQ.GFSLVELMVVVAIIGvLA.IAVP.iNKYlAKARQSEAKTNLSSLYT.EKA
pilA-Tiberius   1 M....r.SQ.GFSLVELMVVVAIIGILA.IAiPSiNKYMAKARQSEAKTNLSSiYTAEKA

pilA-ArHS      61 FYAEYNTYDSRFAAIGYTPEGSLRYNIGFSAAGQIAGVANGYSGTPANDYYSAGGPGTAY
pilA-HD100     57 FYAEY..Y.t.FgAIGYsPEG.LRYNvGF.A.G...a.A......P............AY
pilA-Tiberius  57 Ff.EYNTYD.RFAAvGYTPEG.LRYNvGF.g.G..AG.g.GY...P.N..f.A.G.....

pilA-ArHS     121 CGANGVFARGCTMLNGATGAQPPAITDSICSLSSSGVAANCVTRVDSFQAGAAAVIQTSA
pilA-HD100    112 CG..GV.g.a...v.ag.................t...AA..........AG....I....
pilA-Tiberius 111 .G.............Gg......A...t.....t.a..A...T.......a......T.A

pilA-ArHS     181 TDDRWAIDSSKVIRNTQIGIQX
pilA-HD100    166 ......................
pilA-Tiberius 166 .....A......v...n.....
```

Fig. 2. (A) Codon substitution patterns between orthologous genes of *B. bacteriovorus* HD100 and *B. bacteriovorus* ArhS. The average dN/dS ratio (y axis) is plotted against the synonymous substitution rates (dS) (x axis). (B) Multiple sequence alignment of pila protein from *B. Bacteriovorus* strains (ArhS, HD100 and Tiberius) and *B. exovorus* JSS. Dissimilarity based matrix was used to plot the alignments.

needs to be biochemically validated. All six *fliC* gene homologues (70% to 95% identity) were also observed in ArHS (Fig. S6), but phylogenetic variation (amino acid variation in multiple alignment) in these genes suggested a possible aberration of this phenotype (flagella) in the ArHS strain.

(ii) Eco-genetically adapted genotype: *Enterobacter cloacae* ArHS

Individual metagenome read mapping (global alignment at sequence identity cut-off = 85%) against NCBI-nt database clearly highlighted the genetic predominance of *Enterobacter cloacae* genotypes, with 14% (*n* = 11 234 530) of the total assembled data (10.6 Gb) (*n* = 78 862 135) aligning to five *E. cloacae* genomes. This finding was also supported by MetaPhlAn (MB-A 22% and MB-B 25%), 16S rRNA gene mapping [MB-A 3% and MB-B 2.4%, (Fig. S3)] and BLAST2LCA analysis (Fig. S4). Pan-genomic level (tetranuleotide correlation: 0.9 and %G + C: mean ± 1) clustering of metagenomic contigs enabled us to reconstruct a draft (4.78 Mb; Fig. S7A) *Enterobacter cloacae* pan-genome (*E. cloacae* ArHS) representing five *E. cloacae* strains; *E. cloacae* ATCC 13407, *E. cloacae* subsp. Dissolvens SDM, *E. cloacae* subsp. cloacae ENHKU01, *E. cloacae* subsp. cloacae NCTC 9394 and *E. cloacae* ECWSUI, which represent 44%, 8%, 7.6%, 33% and 5.6% relative abundance within the pan-genome. *Enterobacter cloacae* ArHS was assembled in 1040 contigs, with 37 tRNA and 4285 protein coding genes with an average of 55%G + C (Fig. S7B). This draft pan-genome was further analysed using single-copy gene profiling, average nucleotide identity and BLAST2LCA analysis. Reciprocal smallest distance analysis [E-value (10^{-5}) and divergence cut-off (0.15)] showed that the *E. cloacae* ArHS and *E. cloacae* sub-sp ATCC (pairwise ANI value = 95%) shared only 49 orthologous protein coding genes. *Enterobacter cloacae* ArHS was predicted to be a facultative aerobe, with the presence of *fnr* (fumarate nitrate reduction regulator) and selenate reductase genes, as well as a complete pathway for menaquinone-8 biosynthesis, suggesting at its competence to survive in anaerobic niches, as shown previously (Ma *et al*., 2009).

(iii) Viral Predators in the Mat Environment: Phages

The majority (85%) of the viral contigs (*n* = 409; average size = 1.5 kb) were identified as 'uncultured viruses', with *Caudovirales* dominating (70%) those that could be annotated. Tetranucleotide frequency correlation (minimum R^2 value = 0.9), %G + C and coverage-based clustering of viral contigs enabled the reconstruction of a 200 kb viral genotype (*n* = 4, max contig size = 101 kb, Fig. S8A), sharing only three homologous genes with the GenBank-nr database. A 49 kb reconstructed viral genotype (contigs = 2) was assigned (coverage = 99% and minimum identity = 97% at 44.5 kb region) to

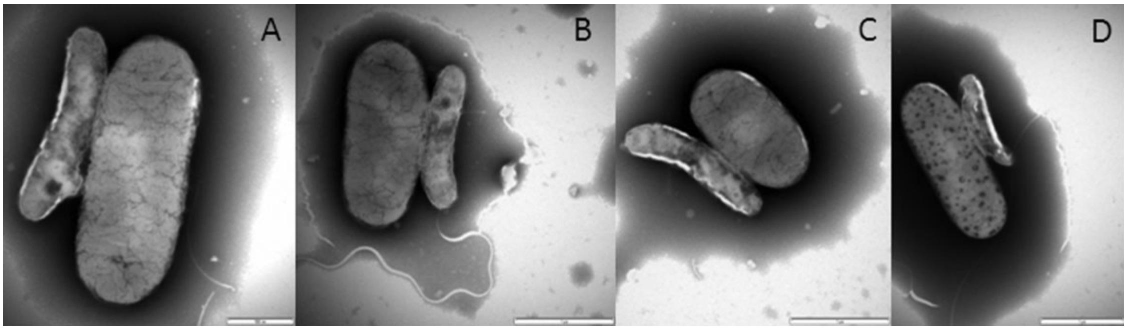


Fig. 3. Electron micrographs showing side on attachment of *Bdellovibrio*-like cells to *E. coli* prey cells. Cells were stained with 2% phosphotungstic acid (PTA). Scale bar is 500 nm in A, 1 μm in B-D.

Enterobacteria phage mEp237 (Fig. S8B). Another 42 kb contig (Fig. S8C) was assigned (coverage = 94% and minimum identity = 93% at 38 kb region) as *Caudovirales*, P2-like virus (*Enterobacter* phage PsP3) and downstream analysis (BLASTN E-value 10^{-15}) of this contig revealed close homology (coverage = 91% and maximum identity = 100%, at 38 kb region) with a prophage-like region of *E. cloacae* subsp. dissolvens SDM (Xu *et al.*, 2012) and also with *E. cloacae* ArHS (coverage = 55% and maximum identity = 100% 38 kb). Another 48 kb viral contig was assigned to the *Caudovirales*: Lambda-like viruses again showed homology (coverage = 55% and maximum identity = 100%) with the prophage-like region of *Enterobacter cloacae* subsp. cloacae ATCC 13047 (Ren *et al.*, 2010) and also with *E. cloacae* ArHS (coverage = 35% and maximum identity = 100%). Finally, a 38 kb (Fig. S8D) contig was assigned to *Caudovirales* specifically, *Caulobacter* phage Cd1 (Query coverage = 31% and maximum identity = 100%, 9.3 kb).

Enrichment of Bdellovibrio*-like cell with a novel predatory interaction*

The *hit* locus for the ArHS strain was remarkably different from other sequenced strains of predatory *B. bacteriovorus* species, and it is well known that *hit* locus genes control the switching between predatory and non-predatory lifestyle of *Bdellovibrio*. Thus, it was possible that the *Bdellovibrio* ArHS strain was capable of growing independent of any host/prey species or had evolved a different mechanism of prey interaction. Strain Tiberius revealed a difference of three amino acids in the *Bd0108* gene product of the *hit* locus relative to strain HD100 which is already established as an intra-periplasmic predator, but with some cells simultaneously growing host independently (Hobley *et al.*, 2012). To identify the ArHS strain's growth mode, filtered (< 0.45 μm) community extract was incubated with *E. coli* prey cells in Ca/HEPES buffer at 37°C and 45°C (the prey cells were expected to be alive in the former and dead in the latter). Electron microscopy of these samples revealed many *Bdellovibrio*-like cells of approximate dimensions 1.5 × 0.5 μm often with a single polar flagellum. Many of these cells were seen to be attached *side on* to the *E. coli* prey cells (Fig. 3), suggesting that in case if this observed cell was the ArHS strain, it might have a novel mechanism of interaction with prey cells in these conditions. Unfortunately, despite many attempts, these *Bdellovibrio*-like cells could not be successfully purified and cultured from these samples, even using elevated temperatures and other environmentally isolated potential prey bacteria in place of *E. coli*. Further, these enrichments were rapidly overgrown by other bacteria from the environmental sample which had passed through the filter along with the *Bdellovibrio*-like cells and so deoxyribonucleic acid (DNA) extracted from these samples were dominated by these bacteria. It is likely that *Bdellovibrio* from this hot spring sample require specific growth conditions that would take extensive investigation to replicate in laboratory conditions which remains out of the scope of this study. Using *Bdellovibrio*-specific forward primer, we were however able to show that community DNA from this sample, where the *Bdellovibrio*-like taxa were visibly present yielded a 628 bp *Bdellovibrio* 16S rRNA sequence identical (100%) to that of *B. bacteriovorus* HD100 and *Bdellovibrio* ArHS (538 bp contig).

Discussion

The Himalayan hot springs located in the valley of the Parvati River at Manikaran, India are extreme environments with low prokaryotic diversity. Here, we have leveraged this ecosystem to describe the genotype of a novel *Bdellovibrio* taxon, a potential prey taxon (*E. cloacae*), and three prey-related phages. We have also demonstrated a potential *Bdellovibrio* cell living in this ecosystem with a probable unique predatory phenotype.

The majority of the community in the two microbial mats was dominated by Gram-negative taxa (mostly known pathogens), and *Bdellovibrio bacteriovorus* which is known to kill, digest and assimilate mostly Gram-negative pathogenic bacterial strains (Dashiff *et al*., 2011); therefore, it is reasonable to assume that these enriched pathogenic bacterial strains could represent potential prey. The most abundant of these potential preys was *E. cloacae*, which comprised 11% of the total metagenome data (10.6 Gbp). The only abundant *Bdellovibrio* genotype that could be reconstructed from these mats (*Bdellovibrio* ArHS) lacked the key genes assumed necessary for host selection and penetration, i.e. *Bd0105* and *Bd0108*. However, the ArHS genotype still maintained the predation-specific gene content of *B. bacteriovorus* HD100, required for the production of hydrolytic enzymes, flagella and pili, suggesting that it maintained the predatory genetic infrastructure required for prey attachment and assimilation of the cellular contents. Electron microscopy of the filtered community revealed many *Bdellovibrio*-like cells apparently attached side-on to prey *E. coli* cells. While we could not culture these organisms, possibly because we failed to replicate the hot spring *in situ* conditions, this does represent a unique opportunity for individual cell capture and single-cell genomic sequencing, but this was beyond the remit of the current study. However, this evidence does suggest the existence of a new predatory *Bdellovibrio* species with an 'epibiotic' mode of predation, instead of the classical intra-periplasmic 'nose up' mode of attachment and invasion into the prey periplasm.

The predominant genotype, *E. cloacae* ArHS, comprised of at least five different strains that were enriched in these samples. The reconstructed *Enterobacter* genotype (*E. cloacae* ArHS) marked by two important genetic characteristics supporting the hypothesis that it can survive the aerobic and anaerobic niches in this ecosystem using *fnr* (Constantinidou *et al*., 2006) and selenate reductase genes *luxS* gene (Rezzonico *et al*., 2012) and Menaquinone biosynthesis operon (*menFDHBCE*)) (Table S3). Interestingly, this FNR protein is also known to activate the selenate reductase gene in *E. cloacae* sub. spp., which further reduces the selenate [Se(VI), SeO_4^{-2}; water soluble and toxic] into elemental selenium [Se(0)] using menaquinones as anaerobic electron carriers (Yee *et al*., 2007). X-ray diffraction (XRD) analysis revealed the high concentration ($1.5\ \mu g\ g^{-1}$) of elemental selenium throughout these microbial mats, which supports the potential activity of this pathway.

According to the Lotka–Voltera model ('kill the winner' hypothesis), viruses (mostly phages) are the apex species-specific predators in every microbial ecosystem (Suttle, 2005). Prophages are known to generate strain-specific differences within species and thus can change the genetic repertoire of the host (Canchaya *et al*., 2003). Interestingly, five 'near complete' viral genomes were assembled including four *Caudovirales* and one novel virus genotype (200 kb). Three of the *Caudovirales* genotypes were assigned as *P2 (Enterobacter* phage PSP3), *Enterobacteria* phage mEp237 and *Lambda* (prophage-like region of *E. cloacae sub. sp. cloacae* ATCC 13047)-like viruses. Since, the viruses are known to be species-specific microbial predators and two of the reconstructed phage genomes shared syntenous genetic regions (including replication protein GpA) with the reconstructed *E. cloacae* ArHS pan-genome (Supporting Information Fig. S7B), we assume that these genotypes are the natural phages of the *in situ E. cloacae* strains. These reconstructed temperate phages possess lysogenic conversion genes (LCG) (Canchaya *et al*., 2003), which would change the phenotype of their *in situ* host (*Enterobacter*) following infection. The LCG genes include periplasmic protein YqjC, amino acid transmembrane gene (COG1126) and DNA polymerase (Fig. S8). These results suggest that the phage can introduce and maintain strain-specific variations across these eco-genetically adapted host species.

Conclusions

We have used shotgun metagenomics to reveal the population-level dynamics of prokaryotic genotypes maintained across hot-spring associated microbial mats. Metagenomic diversity analysis highlighted the genetic predominance of *B. bacteriovorus* genotypes as the primary bacterial predator and assigned *E. cloacae* as the eco-genetically adapted taxon. Owing to the low phylogenetic diversity and evenness, we were able to reconstruct the still-uncultivated genotypes of these prey and predator taxa. Downstream analysis of the reconstructed *Bdellovibrio* ArHS genome revealed a significant difference in the predatosomes of ArHS and reference predatory strains (i.e. HD100 and Tiberius). Despite the presence of complete hydrolytic arsenal, the invasion-specific marker gene *Bd0108* was congenitally deleted in two 'hit-locus' regions of ArHS. Phenotypic aberration was also predicted for the *fliC* genes (flagella). *In vitro* predation experiments revealed the presence of *Bdellovibrio*-like cells involved in novel membrane interaction that did not involve pili or cell invasion. Interestingly, detailed analysis of reconstructed *E. cloacae* ArHS pan-genome enabled us to highlight that this taxon is capable of living across the aerobic and anaerobic regions of the mats using the *luxS* and auto-inducer-based quorum sensing pathway and FNR, menaquinone and selenate reductase-based anaerobic metabolism of selenium [Se(0)] respectively. Although we have provided some detailed insights into novel *Bdellovibrio* taxa, some questions remained unanswered. For example, do bacteriophages control the population of *Bdellovibrio?* Also, is mutational resistance to *Bdellovibrio* predation possible for *in situ* prey genotypes? *In vitro* predation assays can now be targeted to resolve the *in situ* prey–predator dynamics.

Experimental procedures

Site selection and sampling

We obtained the microbial mat samples from a hot spring opening located atop the Himalayan ranges in north-eastern India (Dwivedi *et al*., 2012). This site (32°01′34.8″N, 077°20′50.3″E) represents one of the closely located natural hot spring openings with highest temperature (> 90°C) reported for any hot spring across the country (Cinti *et al*., 2009). Sampling was performed on 12th June of 2012. On site, two samples were obtained from different layers of a microbial mat (sedimented over stones that faces continuous water flow) and stored into a sterile microcentrifuge tube and immediately transferred on dry ice. The ambient temperature of water and microbial mat was 93°C and 52°C respectively. Samples were transported back (on ice) to the laboratory (University of Delhi) and stored at −80°C till processed for further analysis.

Physiochemical analysis

Physical, chemical and mineralogical properties of the microbial mat and water samples were analysed using XRD and ICP-MS analysis respectively. Detail explanation of the methods is described in (Dahl *et al*., 2013).

Community DNA extraction, sequencing and quality filtering

Total community DNA was extracted from the homogenized biofilm samples (5 g) using PowerMax(R) DNA isolation kit (MoBio Inc., USA) according to the manufacturer's instructions. Following extraction, DNA concentration and integrity was measured using NanoDrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA) and gel electrophoresis respectively. Samples of DNA were later pre-processed for sequencing using DNA sample preparation kit protocol (Illumina Inc., San Diego, CA, USA). Sequencing was performed as per the manufacturer's (Illumina) protocol. Briefly, for both the samples, Illumina GAII technology was used to generate 2 bp × 90 bp paired-end reads with an average insert size of 170 bp. Base calling was performed using Illumina Genome Analyzer software version 1.5.1. Individual metagenome reads were trimmed using following parameters; Q_{20} quality cut-off, a minimum read length of 85 bp and allowing no ambiguous nucleotides.

Metagenome assembly and annotation

Quality filtered sequence reads (78 891 278) were assembled by Velvet (Zerbino and Birney, 2008) set at *k*-mer length of 51, insertion length of 170 bp, an expected coverage of 15 and coverage cut-off of 2. Coverage was calculated by mapping raw sequences back to the contigs using bwa-0.5.9 (Li and Durbin, 2009) at default parameters. Approximately 36% of the total reads were assembled into contigs (total length = 455 Mbp).

To predict and compare the community metabolism across these microbial mat samples, individual metagenome reads were compared (BLASTX E-value = 10^{-5}) against KEGG (Kanehisa *et al*., 2004) database. BLASTX-based analysis was used to calculate the relative abundance and coverage using HUMANN (Abubucker *et al*., 2012). To predict the overall community metabolism, metagenome assembly (contigs > 500 bp) was used. Gene calling was performed on the selected contigs using FragGeneScan (Rho *et al*., 2010) at the parameters -genome -complete = 0 -train = sanger_5. Predicted open reading frames were annotated against NCBI-nr database (downloaded on September 2012), KEGG (Kanehisa *et al*., 2004) and COGG (Tatusov *et al*., 2003) using BLASTP (Altschul *et al*., 1990) at an E-value cut-off of 1×10^{-5}.

Assigning phylogenetic status to the community gene content

To identify the relative abundance of sequenced genotypes in our datasets, we mapped raw sequence reads against the RefSeq database (Pruitt *et al*., 2007) using global alignment short sequence search tool (GASSST) (Rizk and Lavenier, 2010) at the parameters p = 85, r = 1 and h = 1. Relative abundance per reference sequence was calculated according to the 'paired-end' criterion. Briefly, a 'hit' was counted against any reference sequence if both the paired-end reads were mapped over reference sequence (minimum sequence length = 800 bp) at an intra-read distance of estimated insert size (170 ± 20 bp). Metagenomic rRNA gene survey was performed against GREENGENES database (DeSantis *et al*., 2006) (16S) and SILVA database (Pruesse *et al*., 2007) (18S) using BLASTN at an E-value cut-off of 1×10^{-15}. We also used the MetaPhlAn (Segata *et al*., 2012) classifier to estimate the microbial diversity across these biofilm samples. To reveal the overall microbial diversity across this extreme environment contigs, above 1000bp length were compared against NCBI-nt and nr database using BLASTN and BLASTX (Altschul *et al*., 1990) set at an E-value cut-off of 1×10^{-10} and 10^{-5} respectively. BLAST output was afterwards processed using BLAST2LCA algorithm (https://github.com/emepyc/Blast2lca), and phylogenetic identity (up to genus level) was assigned to the individual contigs.

Genome and genotype reconstruction

For the genotype reconstruction, metagenome contigs with length > 10 kb (*n* = 23, 906) were clustered based on their tetranucleotide frequency correlations (minimum $R^2 = 0.9$), read coverage and %G + C criterion (Mackelprang *et al*., 2011), and a separate graph was constructed for each cluster. Subsequently, we recruited the smaller contigs (1–10 kb) to individual clusters using similar criterions. Preliminary analysis of BLAST2LCA (https://github.com/emepyc/Blast2lca) results assigned majority of the 'larger contigs' (> 10 kb) to the *Bdellovibrionales* and *Enterobacteriales*. Therefore, we targeted our further assembly efforts to reconstruct the genotypes from these taxa. Largest cluster (minimum R^2 value = 0.9) was manually checked for coverage parameter and the outliers (contigs above the standard deviation cut-off = ± 1) were removed. Subsequently, remaining contigs were also recruited on cluster using similar criterions. Tetranucleotide frequencies were calculated for larger contigs using TETRA (Teeling *et al*., 2004) and a ‘contigs versus tetra-mer (total 256)’ matrix was constructed and later hierarchical clustering (squared Pearson correlation, average linkage) was performed on this matrix using MEV (Saeed, 2003). Sequence length of ‘targeted clusters’ (contigs with similar coverage and tetranucleotide frequency correlation value above 0.9) was extended and contigs were reassembled using PRICE (Ruby *et al*., 2013).

Clusters representing draft *Bdellovibrio* ArHS ($n = 53$; total length = 3.89 Mbp) and *E. cloacae* ArHS ($n = 1097$; total length = 4.78 Mbp) bacterial genotypes were submitted to RAST server (Aziz *et al*., 2008) for gene calling and annotations. Viral genotypes were annotated using tBLASTX against NCBI-nr database and PHAGE RAST (http://www.phantome.org/PhageSeed/Phage.cgi?page=phast) pipeline. For the single-copy gene analysis, 15 protein coding genes (*dnaG, frr, infC, nusA, pgk, pyrG, rplA, rplK, rplL, rplM, rplS, rplT, rpmA, rpsB, tsf*) were selected, and their protein sequences were extracted from the strain ArHS and also downloaded from the five closest phylogenetic neighbours (NC_005363, NC_019567, NC_020813, NC_016620 and NC_007503) predicted by the RAST server (Aziz *et al*., 2008) and whole genome-based ANI analysis (Konstantinidis and Tiedje, 2005). Protein sequences of these single-copy genes were concatenated and aligned using MUSCLE (Edgar, 2004) and TRIMAL v1.3 (Capella-Gutierrez *et al*., 2009) was used to trim the low confidence parts of the alignment. Trimmed alignment was fed to RAxML 7.0.4 (Stamatakis, 2006) for constructing a phylogenetic tree based upon PROTGAMMAWAG model. The phylogenetic analysis was conducted with 100 bootstrap replicates.

Comparative genomics and metagenomics

Metagenome sequences were compared against previously published hot spring community genomics datasets (Bhaya *et al*., 2007; Schoenfeld *et al*., 2008) using *k*-mer-based microbiome similarity analysis using RTG Investigator. A correlation network diagram was generated (minimum correlation = 0.9) from the tetranucleotide frequency matrix using ‘QGRAPH’ package in R (R Development Core Team, 2011: http://www.R-project.org/). For viral contigs, tetranucleotide frequency matrix (contigs versus tetramer type) was generated from custom sequence database generated using NCBI virus database (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/), virus contigs from this study ($n = 409$, taxonomically assigned as ‘virus’ by BLAST2LCA analysis) and from previously published hot spring viromes (Schoenfeld *et al*., 2008). Correlation matrix was processed using hierarchical clustering (bootstrap = 1000, squared Pearson correlation, average linkage) using PV CLUST package (Suzuki *et al*., 2006) implemented in R (R Development Core Team, 2011: http://www.R-project.org/).

Orthologous gene identification was performed between *B. bacteriovorus* ArHS, *B. bacteriovorus* HD100 (Rendulic *et al*., 2004), *B. bacteriovorus* Tiberius (Hobley *et al*., 2012) and *B. exovorus* JSS (Pasternak *et al*., 2013) genomes using RSD algorithm (Wall and Deluca, 2007) with the following parameters, E-value = 10^{-15} and the divergence cut-off = 0.2. Reciprocal smallest distance combines reciprocal best BLAST hit approach and likelihood evolutionary distance for accurate orthologous gene identification. The ancestral genotype (orthologous segments) inherited in these *Bdellovibrio* genomes was reconstructed as explained earlier (Sangwan *et al*., 2014). Comparative functional analysis was also performed for three *Bdellovibrio* genomes using automated annotation web service of xBASE2 (Chaudhuri *et al*., 2008). The nucleotide sequences of orthologous proteins was aligned by ‘pal2nal’ script, (Suyama *et al*., 2006) using protein alignment as guide, and dN/dS ratios were calculated for the alignments using yn00 module of PAML package (Yang, 2007). To reveal the time-independent effect of ‘periodic selection’ dN/dS values were plotted (y axis) against dS values (x axis) as explained earlier. Reads contributed in the reconstructed genome of *Bdellovibrio* ArHS were mapped across *B. bacteriovorus* HD100, and *B. bacteriovorus* Tiberius genomes and mapping results (SAM files) were processed for the single nucleotide polymorphism (SNP) detection using VARSCAN (Koboldt *et al*., 2009). For the comparative genomics of *E. cloacae* ArHS and its reference genomes, pair-wise ANI comparisons, orthologous gene prediction, dN/dS calculations and SNP detection was performed using similar bioinformatics tools and parameters as explained above for *Bdellovibrio* genomes.

Enrichment of Bdellovibrio *and electron microscopy analysis*

Microbial mat sample (5 gm of MB-A) was suspended in Ca/HEPES buffer (100 ml of 25 mM HEPES, 2 mM $CaCl_2$, pH 7.6) and incubated at 45°C for 48 h. After incubation, samples were filtered three times using 0.45 µm pore size Millex filter (Millipore, Billerica, MA). Fifty microlitres of this sample was spotted onto yeast-extract-peptone-sulfate-cysteine (YPSC) double-layer agar plate containing *E. coli* S17-1 in the top layer and incubated at 37°C for 5 days. Endogenous Gram-negative bacteria were also tested but we were not able to demonstrate predation with these. Areas of clearing in the lawn of *E. coli* were picked and re-suspended in 2 ml Ca/HEPES buffer. These samples were enriched for *Bdellovibrio* by addition of 150 µl stationary phase *E. coli* S17-1 (grown for 16 h in YT broth at 37°C with shaking at 200 r.p.m.) and incubation at either 37°C or 45°C for 24 h. Fifteen microlitre samples were placed on carbon formvar grids (200 mesh) for 5 min, then stained with 15 µl 2% phosphotungstic acid pH 7.0 for 45 s before being imaged using a JEOL JEM 1010 transmission electron microscope.

Sequence availability

Metagenome sequence data has been deposited in DDBJ/EMBL/GenBank under the accession number of PRJEB4614 (http://www.ebi.ac.uk/ena/data/view/PRJEB4614).
Bdellovibrio ArHS contigs have been submitted in NCBI under the accession number of JTEV00000000.

Metagenome assembly contigs have been submitted in MGRAST with MG-RAST ID: 4600092.3.

Acknowledgements

We thank Scott Fendorf Department of Environmental Earth System Science, Stanford University, California for geochemical analysis of the samples. The work was supported by Grants from Government of India under the project, National Bureau of Agriculturally Important Microorganisms (NBAIM) AMASS/2006-07/NBAIM/CIR, the University of Delhi/ Department of Science and Technology Promotion of University Research and Scientific Excellence (PURSE)-DU DST – PURSE GRANT Department of Biotechnology (DBT). NS, VG and AS gratefully acknowledge the Council for Scientific and Industrial Research (CSIR) and the National Bureau of Agriculturally Important Microorganisms (NBAIM) for providing research fellowships. This paper was revised during a renewed visit under an Alexander von Humboldt Fellowship (at the University of Freiburg, Germany) awarded to RL.

References

Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., *et al*. (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* **8:** e1002358.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al*. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9:** 75.

Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., *et al*. (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analysis. *ISME J* **1:** 703–713.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brussow, H. (2003) Prophage genomics. *Microbiol Mol Biol Rev* **67:** 238–276.

Capella-Gutierrez, S., Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972–1973.

Chandrasekharam, D., and Antu, M.C. (1995) Geochemistry of Tattapani thermal springs, Madhya Pradesh, India – field and experimental investigations. *Geothermics* **24:** 553–559.

Chaudhuri, R.R., Loman, N.J., and Pallen, M.J. (2008) xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Res* **36:** D543–D546.

Cinti, D., Pizzino, L., Voltattorni, F., Quattrocchi, F., and Walia, V. (2009) Geochemistry of thermal waters along fault segments in the Beas and Parvati valleys (north-west Himalaya, Himachal Pradesh) and in the Sohna town (Haryana), India. *J Geochemical* **43:** 65–76.

Constantinidou, C., Hobman, J.L., Griffiths, L., Patel, M.D., Penn, C.W., Cole, J.A., and Overton, T.W. (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K-12 adapts from aerobic to anaerobic growth. *J Biol Chem* **281:** 4802–4815.

Cotter, T.W., and Thomashow, M.F. (1992) Identification of a *Bdellovibrio bacteriovorus* genetic locus, hit, associated with the host-independent phenotype. *J Bacteriol* **1174:** 6018–6024.

Dahl, T.W., Ruhl, M., Hammarlund, E.U., Canfield, D.U., Rosing, M.T., and Bjerrum, C.J. (2013) Tracing euxinia by molybdenum concentrations in sediments using handheld X-ray fluorescence spectroscopy (HHXRF). *Chem Geol* **360–361:** 241–251.

Dashiff, A., Keeling, T.G., and Kadouri, D.E. (2011) Inhibition of predation by *Bdellovibrio bacteriovorus* and *Micavibrio aeruginosavorus* via host cell metabolic activity in the presence of carbohydrates. *Appl Environ Microbiol* **77:** 2224–2231.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al*. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.J., Richter, R.A., Valas, R., *et al*. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* **6:** 1186–1199.

Dwivedi, V., Sangwan, N., Nigam, A., Garg, N., Niharika, N., Khurana, P., *et al*. (2012) Draft genome sequence of *Thermus* sp. strain RL, isolated from a hot water spring located atop the Himalayan ranges at Manikaran, India. *J Bacteriol* **194:** 3534.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Evans, K.J., Lambert, C., and Sockett, R.E. (2007) Predation by *Bdellovibrio bacteriovorus* HD100 requires type IV pili. *J Bacteriol* **189:** 4850–4859.

Hachiya, T., Osana, Y., Popendorf, K., and Sakakibara, Y. (2009) Accurate identification of orthologous segments among multiple genomes. *Bioinformatics* **25:** 853–860.

Hobley, L., Lerner, T.R., Williams, L.E., Lambert, C., Till, R., Milner, D.S., *et al*. (2012) Genomic analysis of a simultaneously predatory and prey-independent, novel *Bdellovibrio bacteriovorus* from the river Tiber, supports in silico predictions of both ancient and recent lateral gene transfer from diverse bacteria. *BMC Genomics* **13:** 670.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32:** 277–280.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., *et al*. (2009) Varscan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25:** 2283–2285.

Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102:** 2567–2572.

Koval, S.F., Hynes, S.H., Flannagan, R.S., Pasternak, Z., Davidov, Y., and Jurkevitch, E. (2013) *Bdellovibrio exovorus* sp. nov., a novel predator of *Caulobacter crescentus*. *Int J Syst Evol Microbio* **63:** 146–151.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **26:** 589–595.

Ma, J.C., Kobayashi, D.Y., and Yee, N. (2009) Role of menaquinone biosynthesis genes in selenate reduction by *Enterobacter cloacae* SLD1a-1 and *Escherichia coli* K12. *Environ Microbiol* **11:** 149–158.

Mackelprang, R., Waldrop, M.P., DeAngelis, K.M., David, M.M., Chavarria, K.L., Blazewicz, S.J., *et al.* (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480:** 368–371.

Pasternak, Z., Pietrokovski, S., Rotem, O., Gophna, U., Weinberger, M.N.L., and Jurkevitch, E. (2013) By their genes ye shall know them: genomic signatures of predatory bacteria. *ISME J* **7:** 756–769.

Pasternak, Z., Njagi, M., Shani, Y., Chanyi, R., Rotem, O., Lurie-Weinberge, M.N., *et al.* (2014) In and out: an analysis of epibiotic vs periplasmic bacterial predators. *ISME J* **8:** 625–635.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Pruitt, K.D., Tatusova, T., and Magcott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acid Res* **35:** 65–67.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. Vienna, Austria.: R Foundation for Statistical Computing. [WWW document]. URL http://www.R-project.org/.

Ren, Y., Ren, Y., and Wang, L. (2010) Complete genome sequence of *Enterobacter cloacae* subsp. *cloacae* Type Strain ATCC 13047. *J Bacteriol* **9:** 2463–2464.

Rendulic, S., Jagtap, P., Rosinus, A., Eppinger, M., Baar, C., Lanz, C., *et al.* (2004) A predator unmasked: life cycle of *Bdellovibrio bacteriovorus* from a genomic perspective. *Science* **303:** 689–692.

Rezzonico, F., Smitts Theo, H.M., and Duffy, B. (2012) Detection of AI-2 receptors in genomes of enterobacteriaceae suggests a role of type-2 quorum sensing in closed ecosystems. *Sensors* **12:** 6645–6665.

Rho, M., Tang, H., and Ye, Y. (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38:** e191.

Rizk, G., and Lavenier, D. (2010) GASSST: global alignment short sequence search tool. *Bioinformatics* **26:** 2534–2540.

Roschanski, N., Klages, S., Reinhardt, R., Linscheid, M., and Strauch, E. (2011) Identification of genes essential for prey-independent growth of *Bdellovibrio bacteriovorus* HD100. *J Bacteriol* **193:** 1745–1756.

Ruby, J.G., Bellare, P., and Derisi, J.L. (2013) PRICE: software for the targeted assembly of components of (meta) genomic sequence data. *Genes/Genomes/Genetics* **3:** 865–880.

Saeed, A.I. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **3:** 374–378.

Sangwan, N., Verma, H., Kumar, R., Negi, V., Lax, S., Khurana, P., *et al.* (2014) Reconstructing an ancestral genotype of two hexachlorocyclohexane-degrading Sphingobium species using metagenomic sequence data. *ISME J* **8:** 398–408.

Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M., and Mead, D. (2008) Assembly of viral metagenomes from Yellowstone Hot springs. *Appl Environ Microbiol* **74:** 4164–4174.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **8:** 811–814.

Silver, S., and Phung, L.T. (2005) Genes and enzyme involved in bacterial oxidation and reduction of inorganic arsenic. *Appl Environ Microbiol* **71:** 599–608.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stolp, H., and Starr, M.P. (1963) *Bdellovibrio bacteriovorus* gen. Et sp. N., a predatory, ectoparasitic, and bacteriolytic microorganism. *Antonie Van Leeuwenhoek* **29:** 217–248.

Suttle, C.A. (2005) Review viruses in the sea. *Nature* **437:** 356–361.

Suyama, M., Torrents, D., and Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34:** w609–w612.

Suzuki, M., Shigematsu, H., Iizasa, T., Hiroshima, K., Nakatani, Y., Minna, J.D., *et al.* (2006) Exclusive mutation in epidermal growth factor receptor gene, HER-2, and KRAS, and synchronous methylation of nonsmall cell lung cancer. *Cancer* **106:** 2200–2207.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41.

Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., and Glockner, F.O. (2004) TETRA: a web-service and a standalone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics* **5:** 163.

Wall, D.P., and Deluca, T. (2007) Ortholog detection using the reciprocal smallest distance algorithm. *Comparative genomics Methods in Molecular Biology* **396:** 95–110.

Xu, Y., Wang, A., Tao, F., Su, F., Tang, H., Ma, C., and Xu, P. (2012) Genome sequence of *Enterobacter cloacae* subsp. dissolvens SDM, an efficient biomass-utilizing producer of platform chemical 2,3-butanediol. *J Bacteriol* **194:** 897–898.

Yang, Z. (2007) PAML4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24:** 1586–1591.

Yee, N., Ma, J., Dalia, A., Boonfueng, T., and Kobayashi, D.Y. (2007) Se(VI) reduction and the precipitation of Se(0) by the facultative bacterium *Enterobacter cloacae* SLD1a-1 are regulated by FNR. *Appl Environ Microbiol* **73:** 1914–1920.

Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijin graphs. *Genome Res* **5:** 821–829.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. A geographical map of the study site showing microbial mats of Manikaran hotsprings (Himachal Pradesh, India).

Fig. S2. (a) Genetic correlation between hot-spring microbial mats: Inter sample correlation (based on arcsine square root transformed *K*-mer profile similarities) patterns were used for the clustering and network diagram construction (minimum correlation = 0.50 and cut-off = 0.91). **(b)** Community level functional typing of Manikaran microbial mat samples: Individual metagenome reads were compared (BLASTX, E-value = 10^{-5}) against KEGG database and pathway level inter sample variations (relative abundance) were analysed (Two sided Fisher's Exact t-test with Storey's FDR correction method).

Fig. S3. 16S rRNA (Left panel) and MetaPhlAn (Right panel) based taxonomical survey of Manikaran microbial mats. Taxonomical status was assigned at three main taxon levels: phylum (a, b), class (c, d) and genus (e, f).

Fig. S4. Microbial diversity across the Manikaran hot spring microbial mats: Each dot in figure represents a contig (>1 kbp) assigned to various phylogenetic taxa using BLAST2LCA analysis. Taxa clusters with white and blue rings represent bacterial genera for which genetic enrichment was confirmed by 16S rRNA gene typing and MetaPhlAn analysis, respectively.

Fig. S5. (a) Single copy gene (n = 15) based phylogeny of various *Bdellovibrio* taxon. Maximum likelihood based phylogeny was reconstructed using RAxML program (model = PROTGAMMAWAG and bootstrap = 1000) and *Carboxydothermus hydrogenoformans* Z-2901 was used as out-group. **(b)** Comparative genomics of various *Bdellovibrio* species: Tetranucleotide frequency based correlation network of metagenome contigs contributed in genome reconstructed *B. bacteriovorus* ArHS, contigs as nodes are marked with the numbers (1 to 53 = contigs of *B. bacteriovorus* ArHS, 54 = *B .bacteriovorus* HD100, 55 = *B. bacteriovorus* Tiberius and 56 = *B. exovorus* JSS) edges with Pearson's correlation coefficient (PCC) cut-off = 0.85 are indicated. **(c)** KEGG-Enzyme level comparisons between of HD100 and ArHS and Tiberius strains.

Fig. S6. Multiple sequence alignment of 6 *fli*C homologues genes between *B. bacteriovorus* strains; ArHS, HD100, Tiberius and *B. exovorus* JSS. Dissimilarity based matrix was used to plot the alignments.

Fig. S7. Metagenomic Reconstruction of the prey genotype *E. cloacae* ArHS: **(a)** Whole genome based synteny comparisons (synteny bloc size: 500 kb) of *E. cloacae* strains: ATCC 13047, SDM, ENHKU01, NCTC 9394, EcWSUI and SCFI strains **(b)** Circular representation of the *E. cloacae* ArHS and its BLASTN based comparison with reference *E. cloacae* strains; black color intensity represents the percentage identity. From outside toward the centre: outermost circle, reconstructed phage 1 (42 kb) homologues with *E. cloacae* ArHS and SDM strains; circle 2, reconstructed phage 2 (48 kb) homologues with *E. cloacae* ArHS and ATCC 13047 strains; circle 3, complete genome sequence of *E. cloacae* EcWSUI; circle 4, complete genome sequence of *E. cloacae* NCTC 9394; circle 5, complete genome sequence of *E. cloacae* ENHKU01; circle 6, complete genome sequence of *E. cloacae* SDM; circle 7, complete genome sequence of *E. cloacae* ATCC 13047. Outermost circle: GC plot of the reconstructed *E. cloacae* ArHS.

Fig. S8. Reconstructed *Caudovirales* viral genotypes; **(a)** 49kb viral genotype corresponding to *Enterobacteria* phage mEp237 **(b)** 48kb viral genotype corresponding to the 'prophage like region' of *Enterobacter cloacae* subsp. cloacae ATCC 13047. **(c)** 42kb viral genotype corresponding to the *Enterobacter* phage PsP3 and 'prophage like region' of *Enterobacter cloacae* subsp. dissolvens SDM **(d)** 38.1kb viral genotype corresponding to the '*Caulobacter phage* Cd1'. Phylogenetic status was assigned to the genotypes using BLASTN based comparison with reference genomes.

Table S1. Physicochemical analysis of microbial mat samples. ICPMS values are given in ppb (parts per billion) and soil XRF analysis values are given in relative percentage.

Table S2. Metagenomic recovery of manganese dismutase and arsenic resistant genes.

Table S3. Comparison of proteins for Men Operon, FNR, selenate reductase, quorum sensing between *E. cloacae* ArHS and *E. cloacae* 638, *E. cloacae* SDLa-1 and *E. Coli* K12.