



OPEN

Genomic, morphological, and biochemical analyses of a multi-metal resistant but multi-drug susceptible strain of *Bordetella petrii* from hospital soil

Urmi Halder¹, Raju Biswas¹, Ashutosh Kabiraj¹, Rajendar Deora², Moitri Let¹, Rajendra Kr Roy¹, Annapurna Chitikineni^{3,4}, Krishnendu Majhi¹, Shrabana Sarkar¹, Bhramar Dutta¹, Anubhab Laha¹, Arunava Datta¹, Dibyendu Khan¹, Rajeev K. Varshney^{3,4}, Dipnarayan Saha⁵, Saswati Chattopadhyay⁶ & Rajib Bandopadhyay¹✉

Contamination of soil by antibiotics and heavy metals originating from hospital facilities has emerged as a major cause for the development of resistant microbes. We collected soil samples surrounding a hospital effluent and measured the resistance of bacterial isolates against multiple antibiotics and heavy metals. One strain BMCSI 3 was found to be sensitive to all tested antibiotics. However, it was resistant to many heavy metals and metalloids like cadmium, chromium, copper, mercury, arsenic, and others. This strain was motile and potentially spore-forming. Whole-genome shotgun assembly of BMCSI 3 produced 4.95 Mb genome with 4,638 protein-coding genes. The taxonomic and phylogenetic analysis revealed it, to be a *Bordetella petrii* strain. Multiple genomic islands carrying mobile genetic elements; coding for heavy metal resistant genes, response regulators or transcription factors, transporters, and multi-drug efflux pumps were identified from the genome. A comparative genomic analysis of BMCSI 3 with annotated genomes of other free-living *B. petrii* revealed the presence of multiple transposable elements and several genes involved in stress response and metabolism. This study provides insights into how genomic reorganization and plasticity results in evolution of heavy metals resistance by acquiring genes from its natural environment.

Development of drug and heavy metal resistance in bacteria, particularly resulting in reduction of antibiotic efficacy has intensified in last few decades. Hospital activities like byproducts discharges are generally a big source of antibiotics and heavy metals^{1–4}. Health care facilities from different geographical locations are getting contaminated with ~ 1.5–310 g/day due to the uncontrolled and extensive use of drugs. Over the past few years, in different countries hospital effluents have resulted in the discharge of multiple antibiotics; erythromycin (13–7545 ng/L), azithromycin (113–7351 ng/L), clarithromycin (10–14,000 ng/L), spiramycin (40–2200 ng/L), josamycin (12–15 ng/L), roxithromycin (140–2189 ng/L), clindamycin (31–1465 ng/L), lincomycin (7–48,400 ng/L), ofloxacin (662–37,000 ng/L), ciprofloxacin (11–101,000 ng/L), lomefloxacin (313–1162 ng/L), levofloxacin (150–750 ng/L), enoxacin (450–480 ng/L), norfloxacin (100–44,000 ng/L), sulfadiazine (19.2–6640 ng/L), and

¹Microbiology Section, Department of Botany, The University of Burdwan, Burdwan, West Bengal 713104, India. ²Department of Microbial Infection and Immunity, The Ohio State University, Columbus, OH 43210, USA. ³Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India. ⁴State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Murdoch University, Murdoch 6500, Australia. ⁵Biotechnology Unit, Division of Crop Improvement, ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, West Bengal 700121, India. ⁶Department of Microbiology, Burdwan Medical College, Burdwan, West Bengal 713104, India. ✉email: rajibindia@gmail.com

sulfamethoxazole (16.1–37,300 ng/L)¹. Furthermore, hospitals are also a major source of toxic metal and metalloids contamination like, cyanide, gadolinium (2–320 µg/L), platinum (0.02–300 µg/L), mercury (0.05–6 µg/L), silver (120–500 µg/L), copper (60–250 µg/L), nickel (8–80 µg/L), lead (4–20 µg/L), zinc (80–700 µg/L) and arsenic (0.7–12 µg/L)¹.

Due to these exposures in the natural environment, clinically-relevant bacteria evolve complex protective metabolic activities that help them to withstand the deleterious effects of antibiotics, resulting in a negative impact on public health^{5–8}. Similarly, exposure to toxic metals like cadmium, arsenic, chromium, lead, mercury, etc. leads to acquisition of metal resistant genes and efflux pumps to minimize toxicity. For example, CadA and ArsAB proteins are involved in ATP-dependent efflux of cadmium and arsenite ions, respectively from bacterial cells⁹. In most cases, cytosolic, membrane bound or periplasmic enzymes like arsenate reductase, chromate reductase, mercuric reductase, arsenite oxidase, etc. are involved in metal detoxification^{10–12}. In the presence of arsenite methyltransferase (ArsM), arsenite is methylated into different forms allowing cells to survive under arsenic like toxic material¹³. A primary focus of ongoing research is to determine if any connection exists between antibiotics and heavy metal resistance in microbes. Many studies suggest that there is a link between multidrug and heavy metal resistance in bacteria^{14,15}. Mutations within the genome or the existence of resistant genes within the same mobile genetic element (MGE) are known to mediate cross-resistance to both antibiotics and heavy metals¹⁶. However, natural selection leads to the evolution in acquisition of resistance development in a random way that both antibiotic resistance and heavy metal tolerance are sometimes not linked¹⁷. One such interesting phenomenon has been highlighted in this study, where we isolated a bacterium *B. petrii* BMCSI 3 from effluent adjacent soil sample of the Medical College and Hospital located in Burdwan, West Bengal, India. Although, this strain was tolerant to multiple heavy metals, it was susceptible to several groups of antibiotics.

Bordetella petrii is a free-living, widely distributed environmental species. Earlier, a *B. petrii* strain was isolated from a cystic fibrosis patient, shown to be resistant to a variety of antibiotics¹⁸. In this investigation, we have sequenced the genome, characterized different genomic regions, and identified morphological features. Furthermore, we have pointed out the major genomic features of this strain with other *Bordetella* species and strains of *B. petrii*.

Results and discussion

Isolation, multi-drug and multi-metal tolerance of strain BMCSI 3 from soil. Four individual bacterial colonies grew on Luria–Bertani Agar plate of which one was designated as BMCSI 3. First, we tested its sensitivity against multiple antibiotics. It was found to be sensitive to all the tested antibiotics including Amino benzyl penicillin, Penicillin, Penicillin beta-lactam, 3rd and 4th generation Cephalosporins, Amino Glycosides, Tetracyclines, Fluoroquinolone, Macrolides, Quinolone, and Sulphonamide group of antibiotics (Fig. 1a). Using X-ray fluorescence spectroscopy (XRF), we determined that the soil was contaminated with multiple toxic metals like Titanium (Ti), Iron (Fe), Copper (Cu), Zinc (Zn), Molybdenum (Mo), Cadmium (Cd), and Lead (Pb). The presence of Potassium (K) and Calcium (Ca) were also detected from XRF spectroscopy. Therefore, we tested the growth of BMCSI 3 in the presence of toxic heavy metals. BMCSI 3 grew in the presence of very high concentrations of salts of Lead, Molybdenum, Manganese (~3000 mg/L), Copper (~1000 mg/L), and Iron (~750 mg/L) (Fig. 1b). In comparison to these heavy metals, BMCSI 3 tolerated lower concentrations (~10 mg/L) of Chromium, Cobalt, Arsenic (III and V) salts and did not grow in the presence of Zinc, Cadmium and Mercury (Fig. 1c). However, by analyzing these results, in one way, BMCSI 3 is susceptible to antibiotics, other way, it is highly resistant to multiple metals.

Whole-genome sequences and phylogenomic inference. The draft genome of BMCSI 3 was assembled using the three genome assemblers, SPAdes, Velvet, and ABySS; the latter two were based on the best suited K-mer statistics. For instance, the best assembly statistics were obtained with K-mer length of 99 in Velvet and k-mer length of 96 in ABySS. The contigs produced by the SPAdes, Velvet, and ABySS programs are 66, 14, and 114, respectively. The comparative assembly statistics using the QUAST software revealed that the assemblies consisted of approximately 4.9 Mb genome with the highest N50 length of 3.8 Mb was found with Velvet assembly, which was finally processed for further analysis. The finalized draft assemblies of BMCSI 3 were 99.53% complete without any contaminations (Supp. file S1).

Whole-genome sequence of BMCSI 3 was analyzed to predict its taxonomic status through different web tools. As per the predictions made by MiGA, StrainSeeker, and TYGS webservers, the closest relative determined was *B. petrii* DSM 12804^T. The tree inferred with FastME 2.1.6.1¹⁹ from GBDP (Genome BLAST Distance Phylogeny) distances calculated from genome sequences. The branch lengths were scaled in terms of GBDP distance formula d_5 . The numbers above branches are GBDP pseudo-bootstrap support values > 60% from 100 replications, with average branch support of 97.9%. The tree was rooted at the midpoint²⁰ (Fig. 2a). The taxonomic classification was thus resolved to *B. petrii* under the family Alcaligenaceae, order Burkholderiales, and phylum Proteobacteria. Different features of *B. petrii* BMCSI 3 genome were analyzed including the coding sequence of genes and GC content distributed across the scaffolds. From outermost to centre, distribution of scaffolds (Ring-1); protein-coding genes (CDS) in forward and reverse strand (Ring-2 and 3); Blastp hit with a reference *B. petrii* DSM 12804^T genome (Ring-4); GC skew plot with a value on above average and below average (Ring-5); and plots of GC content (Ring-6) (Fig. 2b).

Genome features. BMCSI 3 contains a total of 4,708 genes of which 4,638 were protein-coding genes (CDS) and 41 were pseudogenes. The number of genes that encode tRNAs, rRNAs, and ncRNAs were found to be 54, 12, and 4, respectively. Annotation by RAST resulted in 29% (1337) of the protein-coding genes being covered under the 324 subsystems. The metabolism of amino acids and derivatives occupied the largest number

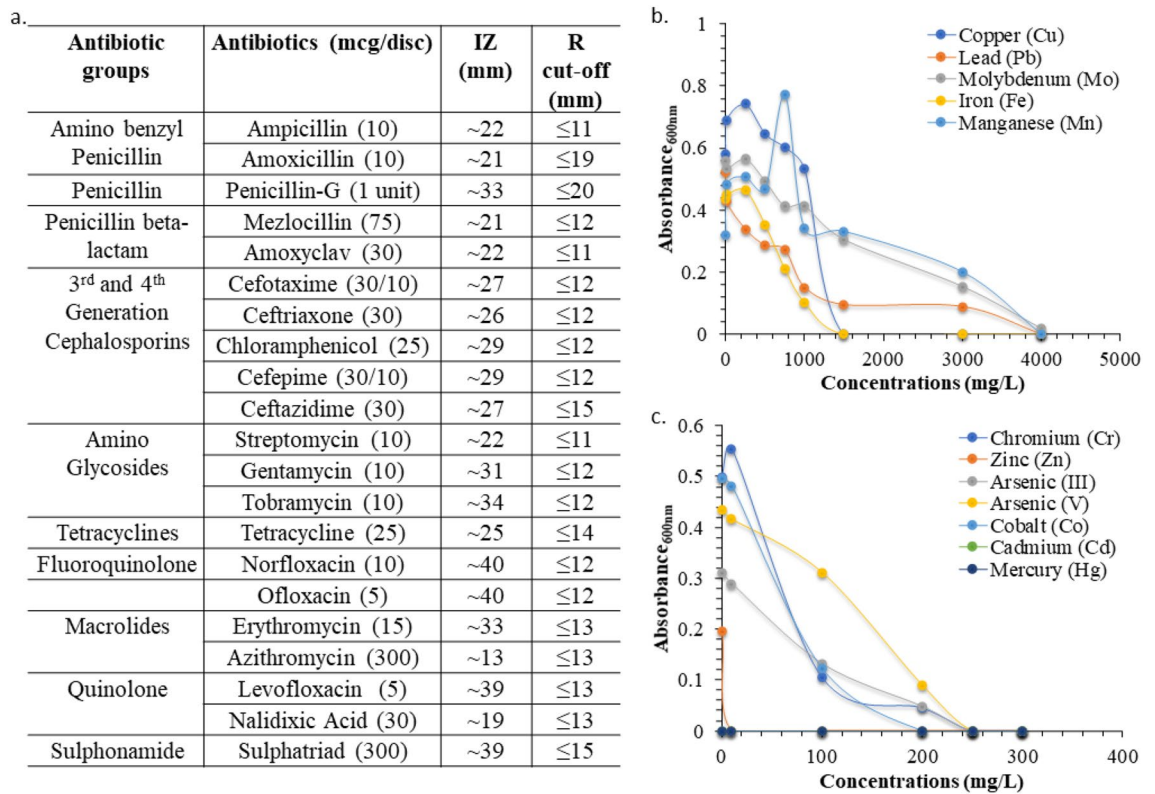


Figure 1. Multi-drug and multi-metal tolerance capacity of BMCSI 3. Antibiotic sensitivity against different antimicrobial agents (IZ inhibition zone; R resistance) (a); metal tolerance capacity against different heavy metals (b,c).

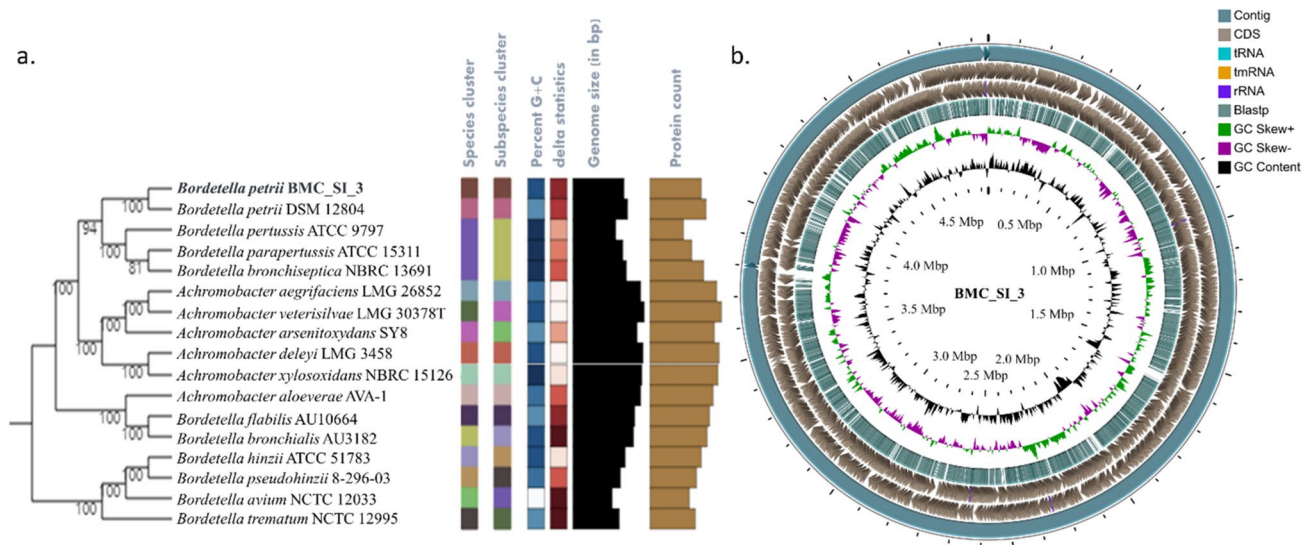


Figure 2. TYGS result for the BMCSI 3 genome (a). Circular representation of BMCSI 3 genome features as per the CG viewer server (b).

of genes (434) followed by carbohydrate metabolism (226), and protein metabolism (206). A significant number of genes (54) were related to virulence, disease, and defense.

Consistent with its tolerance to multiple metals, BMCSI 3 harbored a large number of genes with potential functions in multidrug efflux and heavy metal resistance like MATE family efflux transporter; multidrug efflux SMR transporter; chromate efflux transporter; multidrug efflux RND transporter permease subunit and periplasmic adaptor subunit; outer membrane efflux transporter subunit; fluoride efflux transporter CrcB; Bcr/CflA family efflux MFS transporter; DHA2 family efflux MFS transporter permease subunit; nickel/cobalt efflux

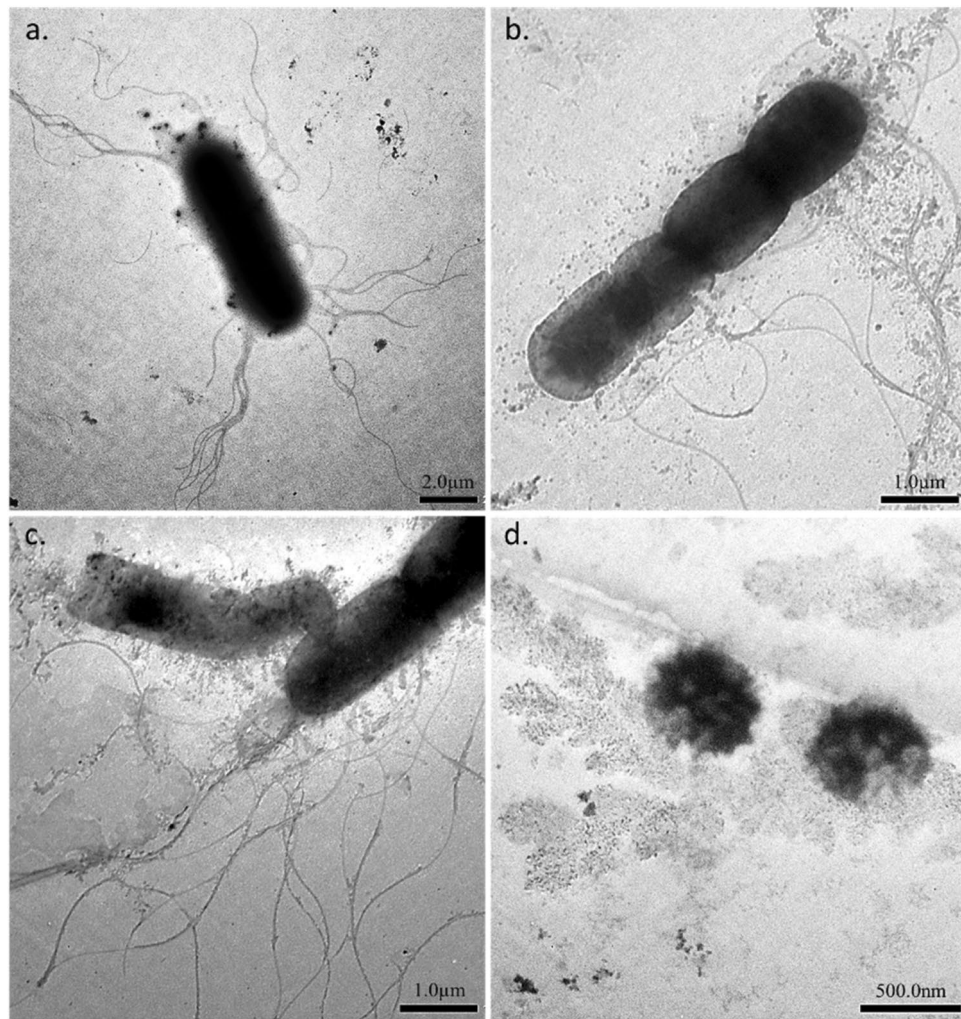


Figure 3. Transmission Electron micrographs showing mature vegetative cells with peritrichous flagellation (a); different stages of cell division (b,c); and free endospores (d) of BMCSI 3.

transporter RcnA; ACR3 family arsenite efflux transporter; arsenical efflux pump membrane protein ArsB; $\text{Co}^{2+}/\text{Mg}^{2+}$ efflux protein ApaG; HlyD family efflux transporter periplasmic adaptor subunit; quaternary ammonium compound efflux SMR transporter SugE and QacE; CusA/CzcA family heavy metal efflux RND transporter; MdtB/MuxB and MdtA/MuxA family multidrug efflux RND transporter periplasmic adaptor subunit; AdeC/AdeK/OprM family multidrug efflux complex outer membrane factor basically involve in effluxing, detoxifying or scavenging excess metal ions²¹ was identified in the genome of metal-loving BMCSI 3. Presence of proteins like heavy metal sensor, response regulator and efflux system could also influence heavy metal resistant^{22,23}. The presence of two-component response regulators and multiple genes with potential function in heavy metal resistance resulted in successful adaptation and survival of this strain in an environment that is heavily contaminated with toxic metals²⁴.

Morphology and biochemical features. BMCSI 3 is a Gram-negative, rod-shaped bacterium. This strain tested positive for catalase, amylase, lipase and utilized citrate as carbon source. A total of 14 protein-coding genes related to flagella under the category of motility and chemotaxis were identified including flagellar biosynthesis proteins like FlhA, FlhB, FliR, FlhF; motor switch protein FliM, FliN; motor rotation protein MotA, MotB; basal-body rod modification protein FlgD; L-ring protein FlgH. Other chemotaxis related proteins, like CheR, CheW, CheB, CheA, response regulator CheY including a type IV pili methyl-accepting chemotaxis transducer N-terminal domain-containing protein that usually mediates chemotactic response were also found to be present²⁵. Generally, *Bordetella* species sense environmental cues and control transcription of virulence genes by the activity of BvgAS (*Bordetella* virulence gene) system. When BvgAS is inactive, *Bordetella* produce functional flagella and thus are motile. The genome of BMCSI 3 lacks genes homologous to *bvgA* and *bvgS*²⁶. Based on these findings, we hypothesized that BMCSI 3 will be flagellated. Consistent with this hypothesis, transmission electron micrographs revealed peritrichous flagellation and ~8–10 μm elongated mature vegetative cells (Fig. 3a–c).

A forespore-like formation and cell outburst was observed within the mature vegetative cell of BMCSI 3 (Fig. 3c). Free endospore-like structures (Fig. 3d) were also observed. To date, the phenotype of sporulation has

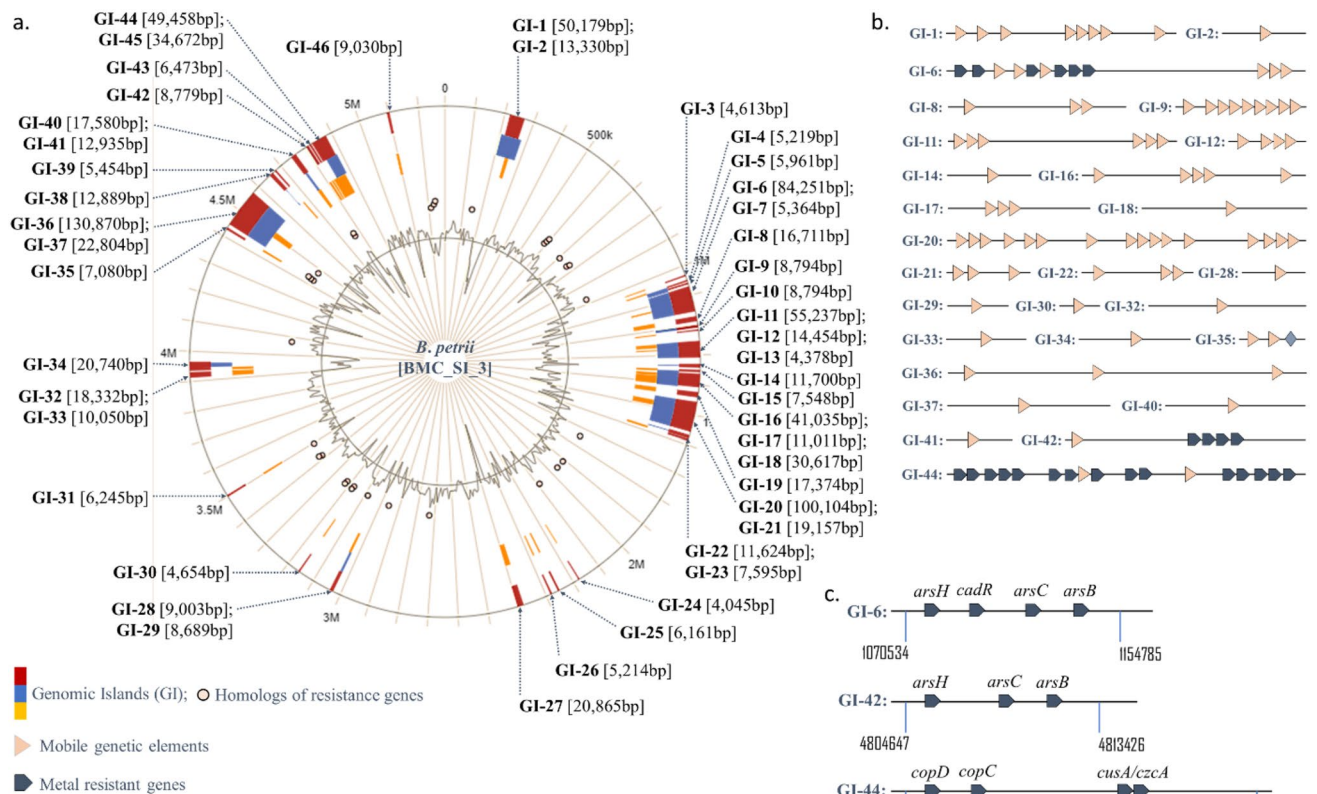


Figure 4. Genomic Islands (GIs) of BMCSI 3 genome (a); presence of mobile genetic elements (b) and metal resistant genes (c) were marked on the GIs (Developed from Supp. file S2).

not been reported in *Bordetella*. Additionally, micrographs clarifying the cell morphology are not available for *B. petrii*²⁷. In the draft genome of BMCSI 3, *spoVC* that encodes peptidyl-tRNA hydrolase and is essential for vegetative growth was present along with spore maturation proteins²⁸. However, genes homologous to sporulation-specific sigma factors of *Bacillus* species were absent^{29–31}.

Genomic islands. A total of 46 Genomic Islands (GIs) of BMCSI 3 were identified which covered about 20% of the genome. Approximately, 10% (88,743 bp) of the genomic islands were comprised of mobile genetic elements in the form of integrative and conjugative elements like the most abundant IS3 family including ISL3, IS5, IS21, IS66, IS91, IS110, IS360, IS1066, and IS1663 families (Fig. 4). Consistent with previously published report, the repetitive element IS481, which is a frequent target of diagnosis of other *Bordetella* species was not found in BMCSI 3^{27,32}. Most of the putative laterally acquired GIs found in BMCSI 3 harbored heavy metal resistant genes, multiple response regulators and other transcription factors, transporters, and multi-drug efflux pumps (Supp. file S2). GI-1 contains a TniQ family protein having a role in transposition of the mercury-resistance transposon Tn5053³³. GI-1 and 2 harbours genes having homology to the ZorAB system, which is involved in the opening of the proton pump leading to the membrane depolarization upon phage infection and ensuring abortive transduction³⁴. This may lower the chances of acquiring new genes, especially antibiotic resistance genes into the bacterial genome. Consistent with this, BMCSI 3, has genes that are usually associated with antibiotic resistance, such as small multidrug resistance (SMR) and resistance-nodulation-division (RND) antibiotic efflux pump³⁵. Interestingly, no such cross-resistance link between heavy metals resistance and antibiotic resistance was found. Eventually, almost all the SMR and RND groups of efflux pump systems are likely to involve metals resistance, instead of antibiotic resistance during different environmental stresses (Fig. 1). HlyD, an important component of CusCFA efflux membrane fusion protein is present in multiple GIs. This type of resistance-nodulation-division group (RND) has evolved to secrete toxic metallic ions outside the periplasmic space³⁶. MFS transporters involved in both efflux and influx of various substances including drug efflux from the cell are also present in the genome³⁷. Another major uncharacterized transporter belonging to the DMT family of transporters is also integrated into the genome. GI-6 and GI-42 harboured arsenical resistance protein-coding gene *arsH*, Cd(II)/Pb(II)-responsive transcriptional regulator protein-coding gene *cadR*, arsenate reductase protein-coding gene *arsC*, and ACR3 family arsenite efflux transporter related gene *arsB*. GI-44 harbors multicopper oxidase domain-containing protein, copper-translocating P-type ATPase, efflux RND transporter periplasmic adaptor and permease subunit, copper-binding protein, metal-binding protein, copper homeostasis membrane and periplasmic binding protein, copper resistance protein B, copper resistance system multicopper oxidase, heavy metal sensor histidine kinase, heavy metal response regulator transcription factor, CusA/CzcA family heavy metal efflux RND

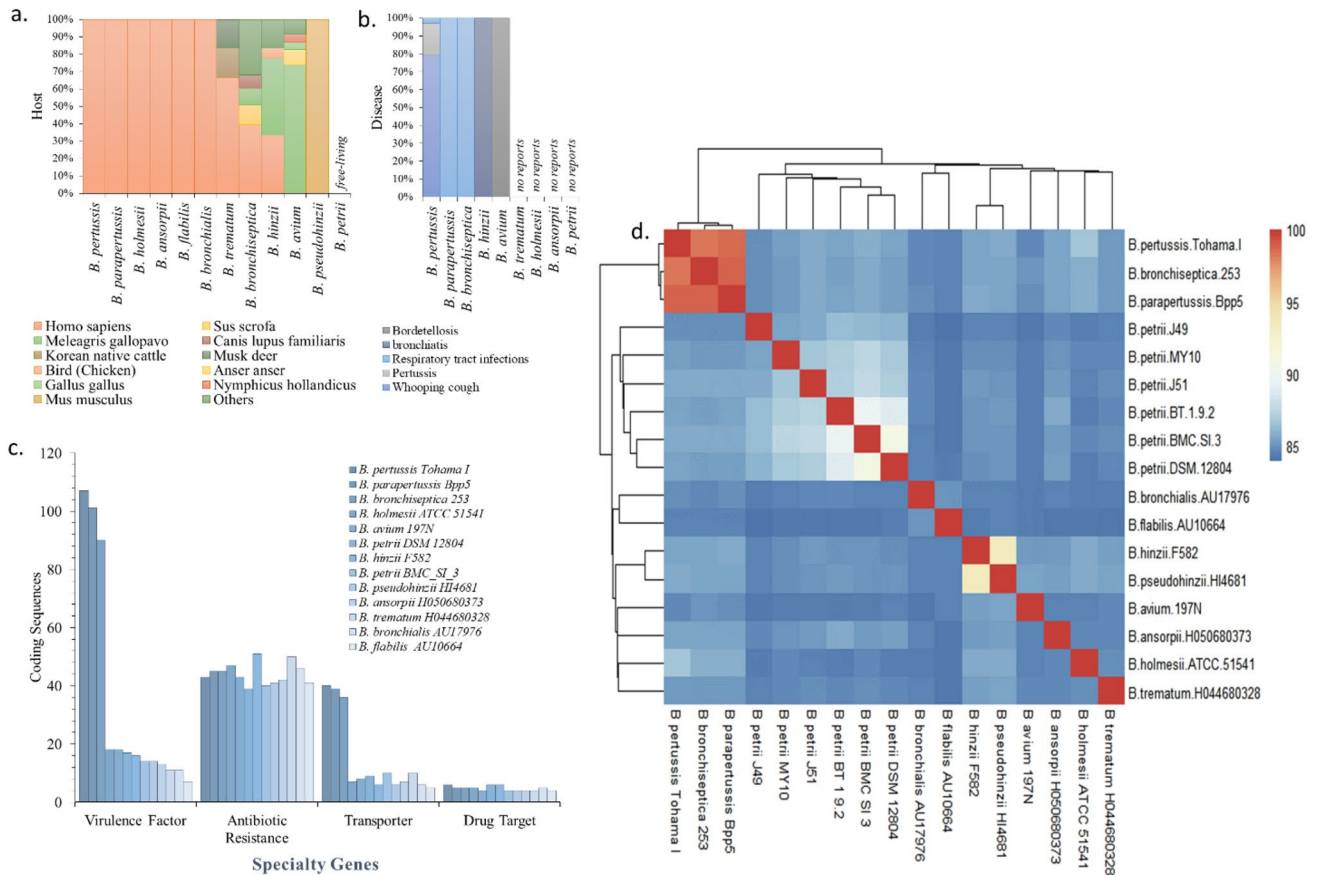


Figure 5. Genome information of the genus *Bordetella*. Host specification (a); disease causing abilities (b); specialty genes (c) (developed from Supp. file S3), and ANI heat map cluster of all representative species including strains of *B. petrii* (d).

transporter, and periplasmic adaptor subunit. GI-33 acquired a universal stress protein that may regulate a broad range of cellular responses against biotic and abiotic stress³⁸. Additionally, GI-44 harboured the gene encoding for the copper-translocating P-type ATPase whose primary function is to transport copper across the biological membrane³⁹. Copper homeostasis membrane protein CopD, periplasmic protein CopC, copper resistant protein B, multicopper oxidase are also present in the same genomic island which may promote tolerance in the presence of copper. A MATE family efflux transporter CusA/CzcA family heavy metal efflux is present in GI-44. Additionally, genes encoding heavy metal sensor histidine kinase and heavy metal response regulator transcription factor are integrated within the GI-44 at multiple sites.

Genome comparison. Bacteria belonging to the genus *Bordetella* are gram-negative coccobacilli of the phylum proteobacteria. Most species in this genus are capable of causing a wide spectrum of pulmonary and bronchial diseases, in humans, animals, and birds^{40,41}. To date, of the described nine species, *B. pertussis* and *B. parapertussis* cause whooping cough in humans⁴², while *B. bronchiseptica* infects primarily animals, and rarely humans⁴³. Despite the fact that, evolutionary trend analysis demonstrated that the ancestors of this genus were of environmental origin, a significant loss of metabolic genes and acquisition or retention of virulence factors within the genome has driven this particular genus to emerge as an opportunistic pathogen^{44,45}. Comparative analysis within the representative strains from *Bordetella* spp. revealed the abundance of transposable elements and presence of stress response genes within genomes of free-living *B. petrii*. Virulence factor and transporter-related protein numbers were significantly high in *B. pertussis*, *B. parapertussis*, *B. bronchiseptica* species. In comparison, *B. petrii* contains a relatively lower number of virulence genes; antibiotic resistance and drug target-related protein numbers were almost equal among all the selected species. Genome similarities were represented and clusters were determined within the species of *Bordetella* based on Average Nucleotide Identity (ANI) analysis, where 95–96% was the species distinction cut-off value. Likewise, species of *B. pertussis*, *B. parapertussis*, *B. bronchiseptica*, strains of *B. petrii* were not correctly grouped with that threshold value. ANI value <90% is usually evidenced by taxonomic differences⁴⁶ (Fig. 5).

All the sixteen genomes of different species of *Bordetella* were compared along with the available strains of *B. petrii* using *B. petrii* DSM 12804 as a reference. From outermost to centre (Fig. 6), *B. pseudohinzii* (Ring-1), *B. hincii* (Ring-2), *B. avium* (Ring-3), *B. trematum* (Ring-4), *B. holmesii* (Ring-5), *B. parapertussis* (Ring-6), *B. bronchiseptica* (Ring-7), *B. pertussis* (Ring-8), *B. flabilis* (Ring-9), *B. bronchialis* (Ring-10), *B. ansorpii* (Ring-11), *B. petrii* J49 (Ring-12), *B. petrii* J51 (Ring-13), *B. petrii* BMCSI 3 (Ring-14), *B. petrii* BT 1 9.2 (Ring-15), *B. petrii*

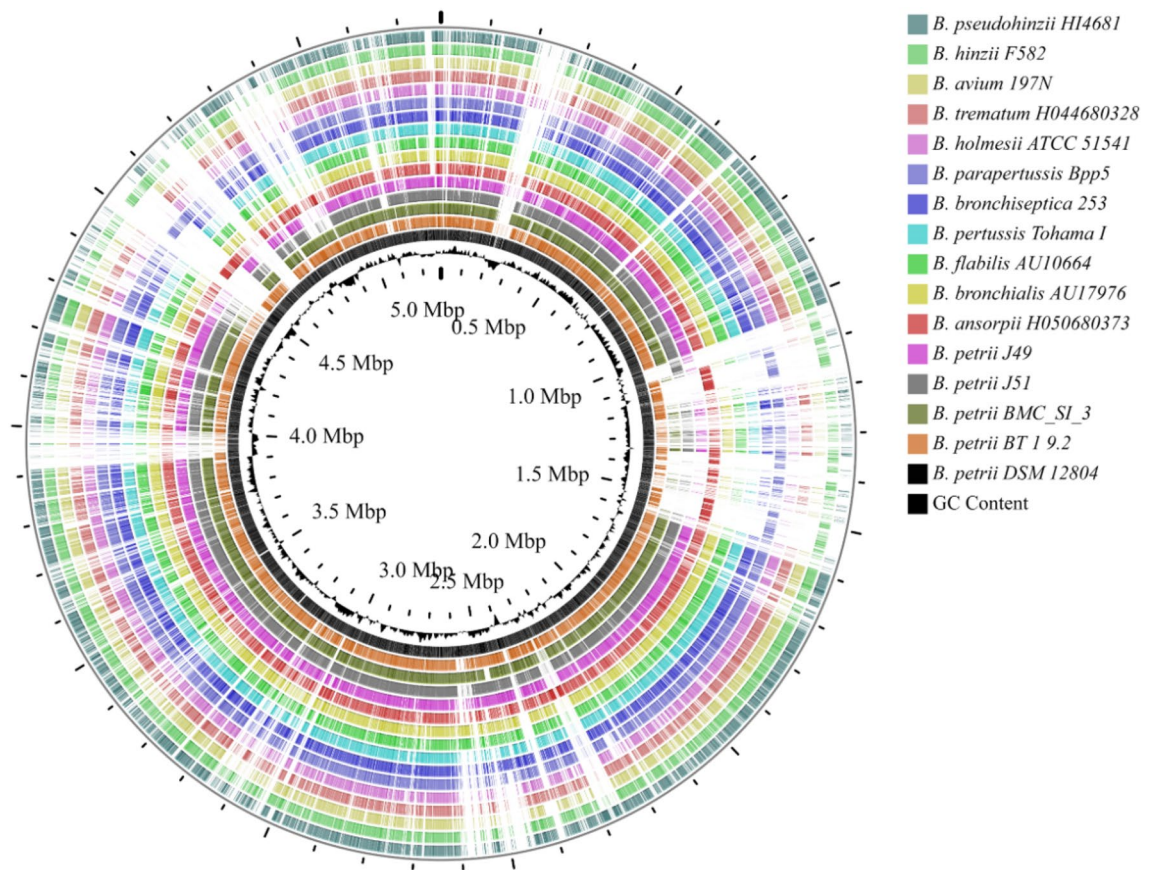


Figure 6. Blast analysis and circular genome comparison of the species of *Bordetella*.

DSM 12804 (Ring-16) and plots of GC content (Ring-17). Transposable elements were most abundant within the genomes of free-living *B. petrii*. A lot of metabolism related genes like putative monooxygenase coding gene, *alsT*, *yhiB*, *menH* (2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase)⁴⁷, *arcD* (arginine/ornithine antiporter), *areA* (Nitrogen regulatory proteins which are GATA type transcription factors)⁴⁸, *hdfR* (encodes a LysR family protein)⁴⁹, *cysIJ*, *hemN* (oxygen-independent coproporphyrinogen-III oxidases)⁵⁰, *nrdD* (Anaerobic ribonucleoside-triphosphate reductase activating protein)⁵¹, *yhbU*, *ptlB*, *nimT* (2-nitroimidazole transporter and similar proteins of the Major Facilitator Superfamily of transporters)⁵², *tmoT*, NADH dehydrogenase like protein coding gene, *sasA* (Adaptive-response sensory-kinase)⁵³, *yheG* putative multidrug efflux transporter coding gene, *norB*, *qoxC* (subunit III of the aa3-type quinone oxidase)⁵⁴, *nirBQST*, *mftC*, *eysG*, *ccmH* (Cytochrome C biogenesis protein)⁵⁵, *zupT* (Zinc transporter)⁵⁶, *cusAB*, *dsbD*, *cnrA* (membrane-bound protein complex catalyzing an energy-dependent efflux of Ni²⁺ and Co²⁺)⁵⁷, *hcaR*, *gcd*, *iolG*, *wbpl*, *mshA*, *pglK* (six-transmembrane helical domain of the ABC transporter)⁵⁸ were only found within the genome of *B. petrii*. Interestingly, virulence regulon transcriptional activator related gene *virB* was only found within the genomes of *B. petrii* (Supp. file S4).

Next, we carried out a comparative genome analysis of *B. petrii*. To date, six soil-born strains of *B. petrii* whose genome has been sequenced (Table 1) of which only the Type Strain *B. petrii* DSM 12804 with a complete genome was reported from a river sediment enriched dechlorinating bioreactor²⁷. Other strains viz., *B. petrii* J49 and *B. petrii* J51 were reported from aquatic soil and strain BT 1 9.2 was reported from contaminated soil. Another strain MY 10 was also reported from soil⁵⁹. Genome sizes of these other strains ranged from 4.21 to 6.10 Mb, with an average GC% ranging from 65.4 to 67.3. The 4.95 Mb genome of BMCSI 3 was smaller than three other *B. petrii* strains and the overall GC content of 67.3% was the second highest among the six sequenced strains. The different categories of the genes present in the five genomes are shown in Table 1. Both BMCSI 3 and BT 1 9.2 isolated from contaminated soil harboured a significantly large number of heavy metal response and efflux transporter related proteins, but a smaller number of ABC transporter, virulence, motility and chemotaxis related proteins compare to the other four strains. Only the strain BT 1 9.2 is devoid of any type IV pili methyl-accepting chemotaxis transducer protein related to chemotaxis. Transmembrane signaling receptor PAS domain-containing methyl-accepting chemotaxis protein was found within strain DSM 12804, J49 and BMCSI 3. Two SOS response-associated peptidase family proteins were found within the chromosome of strain DSM 12804 but as a sign of rapid genome adaptation, this protein was acquired within the island of strain BMCSI 3 genome. Only three strains (J49, J51, and BT 1 9.2) of *B. petrii* harbour genes involved in sulfur metabolism and only two strains BMCSI 3 and DSM 12804 harbored the Type IV secretion system, multi-subunit secretion apparatus involved in secretion of macromolecules across the membranes. The highest numbers of tRNA were found within BMCSI 3 followed by DSM 12804. The decoding of mRNA into protein is governed by tRNA. Since

	<i>B. petrii</i> DSM 12804 ^T	<i>B. petrii</i> BMCSI 3	<i>B. petrii</i> BT 19.2	<i>B. petrii</i> MY 10	<i>B. petrii</i> J49	<i>B. petrii</i> J51
Isolation source	River sediment	Hospital soil	Contaminated soil	Soil	Aquatic soil	Aquatic soil
Assembly	GCA_000067205.1	GCA_017356245.1	GCA_017745595.1	GCA_020991325.1	GCA_000518845.1	GCA_000518965.1
Size	5.29	4.95	6.10	4.95	4.21	5.04
GC%	65.50	67.30	65.90	65.70	65.40	68.50
Genes (total)	5105	4708	5686	5055	3990	4870
CDSs (total)	5041	4638	5630	4991	3899	4779
rRNAs	9	12	5	11	5	8
tRNAs	51	54	47	49	47	47
ncRNAs	4	4	4	4	4	4
Pseudo genes (total)	132	41	58	77	35	32
Motility and chemotaxis related protein	17	9	8	13	11	13
Methyl-accepting chemotaxis proteins	19	13	12	17	13	16
Heavy metal related proteins	18	23	26	16	10	18
Transporter	656	671	760	767	597	771
ABC transporter	315	225	294	340	319	386
Efflux transporter related proteins	42	53	51	36	29	41
Virulence related proteins	2	1	1	2	2	3
Stress-related proteins	40	38	43	35	30	43

Table 1. NCBI PGAP annotated genome features comparison between all available strains of *B. petrii*.

the structural diversity of tRNA most likely co-evolved with their processing RNA splicing endonuclease⁶⁰, it may be significant in the evolution of BMCSI 3.

Pangenome analysis of six genomes of *B. petrii* represents 52,87,950 bp to be pangenome size including 10,69,661 bp core genes, 12,34,707 bp accessory genes and 29,83,582 bp strain-specific genes shared by at least 2 strains. So, the genome diversity of *B. petrii* represents an open pangenome model. However, functional analysis of strain-specific accessory and core genome suggests a usual core metabolism by those strains (Fig. 7).

Conclusion

Bacteria belonging to the genus *Bordetella* can cause a wide range of pulmonary and bronchial infections in humans, animals, and birds. However, *B. petrii*, a member of this genus has a wide spectrum of occurrence within the natural habitat. In this study, we report that a soil-inhabiting *B. petrii* strain isolated from soil sample adjacent to a hospital's effluent can tolerate toxic metalloids by acquiring different metal tolerant genes. Surprisingly, despite being of hospital origin, this strain was susceptible to multiple antibiotics. BMCSI 3 featured putative multidrug efflux transporters like MATE, MSF, RND, DMT, as well as ABC family proteins that are possibly involved in the effluxing of excess metal ions but not the drug molecules. The genomic islands represented a significant number of integrative and conjugative transposable elements. Moreover, genomic islands carried genes that encode metal efflux transporter (ACR3), copper resistant operon (Cop), and mercury-resistance transporter (mer) proteins that may result in metal resistance. Collectively, our results suggest that resistance to antibiotics and tolerance to metals are not always linked. A detailed phenotypic investigation of *B. petrii* is needed to understand the mechanisms through which microbes become resistant to heavy metals.

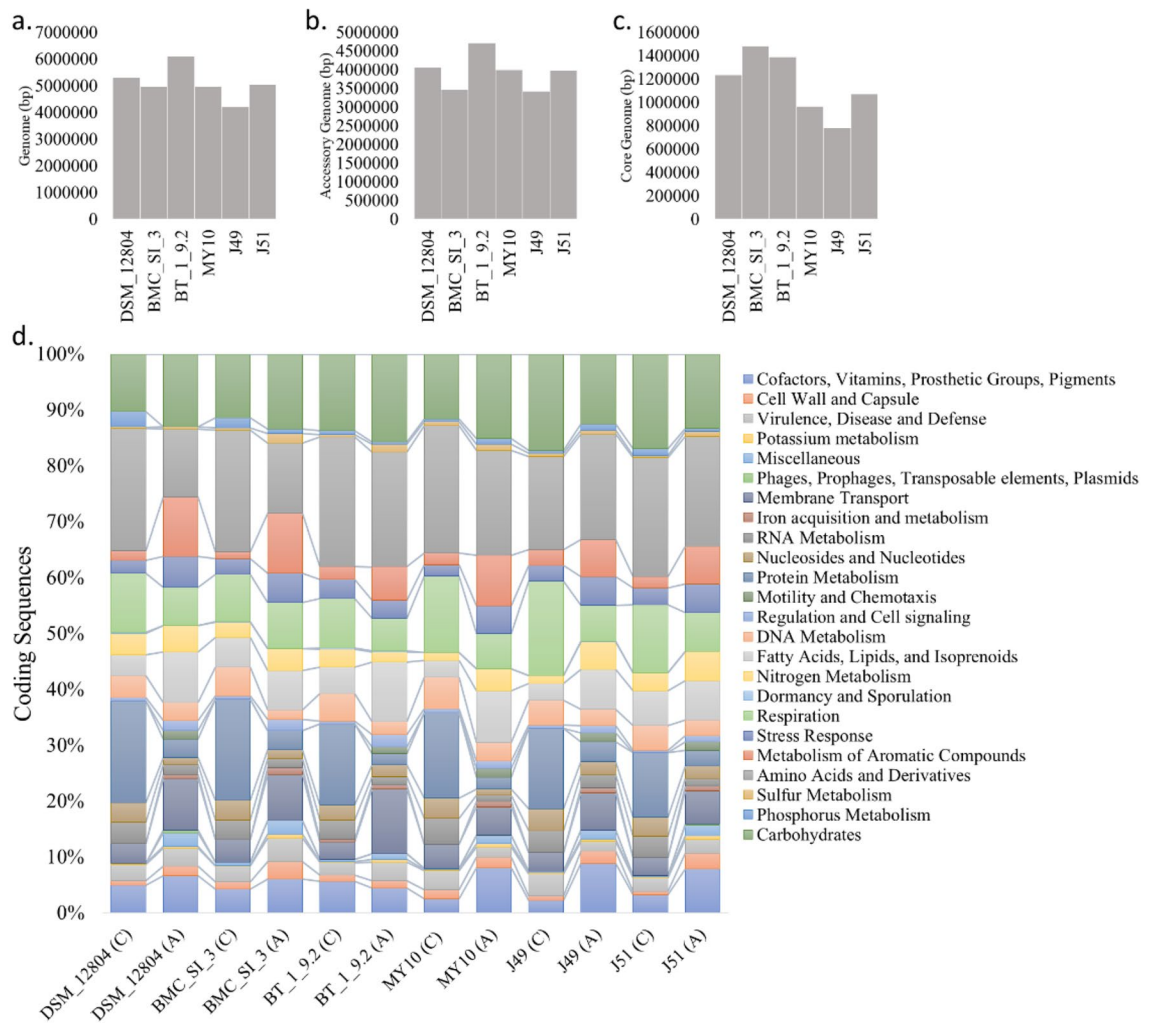


Figure 7. Pangenome analysis strains representing strain-specific genome size (a), accessory genome size (b), core genome size (c), and functional annotation of *Bordetella petrii*.

Methodology

Sampling and isolation. Soil samples were aseptically collected from effluent adjacent dump backyard area of the Emergency Unit, Burdwan Medical College and Hospital, Burdwan, West Bengal, India (23.2489° N, 87.8536° E). Qualitative elemental analysis of the soil sample was analyzed using X-ray fluorescence spectroscopy (Artax, Bruker). For isolation of bacterial species, 1.0 g of soil sample was diluted gradually up to 10^{-4} in sterile saline Milli-Q (0.1% NaCl, HiMedia) and 100 μ L of each dilution was plated onto Luria Bertani Agar medium (HiMedia) and incubated at 30 °C for 24 h⁶¹.

Screening of multi-drug and multi-metal tolerance. All individual bacterial colony forming units were subjected to the preliminary antibiotic sensitivity assay against different groups of antibiotics, viz., ampicillin (10 mcg), amoxicillin (10 mcg), penicillin G (10 IU), mezlocillin (75 mcg), amoxiclav (30 mcg), cefotaxime (30/10 mcg), ceftriaxone (390 mcg), chloramphenicol (30 mcg), cefepime (30/10 mcg), ceftazidime (30 mcg), streptomycin (10 mcg), gentamicin (10 mcg), tobramycin (10 mcg), tetracycline (30 mcg), norfloxacin (10 mcg), ofloxacin (5 mcg), erythromycin (15 mcg), azithromycin (15 mcg), levofloxacin (5 mcg), nalidixic acid (10 mcg), sulphatriad (300 mcg) (HiMedia). Freshly prepared bacterial culture was spread onto the culture plates containing Mueller Hinton Agar media (HiMedia) and was allowed to incubate overnight at 30 °C⁶¹. This experiment was performed in triplicate and inhibition zones (IZ) were measured. The antibiotic concentrations were selected and the sensitivity was determined on the basis of clinical breakpoints provided by the Clinical and Laboratory Standards Institute and European Committee on Antimicrobial Susceptibility Testing^{62,63}. Multi-metal resistant capability was also tested in triplicate against 10 different heavy metal salts (Fe, Cu, Cr, Pb, Cd, Hg, Mn, Mo, Co, Zn) and 1 metalloid (As)⁶⁴. Nutrient Broth (HiMedia) media were prepared and supplemented with 200 μ g/L and (10, 100, 200, 250, 300, 500, 750, 1000, 1500, 3000, and 4000) mg/L concentrations of FeCl₃·6H₂O; CuSO₄·5H₂O; CrO₃; PbCl₂; CdCl₂; HgCl₂; MnCl₂·4H₂O; Na₂MoO₄; CoCl₂·6H₂O; ZnSO₄·7H₂O along with Na₂AsH₄O₄·7H₂O and NaAsO₂ (HiMedia; Merck). A freshly prepared 10 μ L of bacterial culture was inoculated on each

prepared broth and were allowed to incubate overnight at 30 °C, 130 rpm. Cell density was measured by the absorbance value taken at 600 nm using a Spectrophotometer (Lasany LI-721). Based on the multi-drug and multi-metal tolerance capability, strain BMCSI 3 was selected for this study.

Morphology and physicochemical study. Cell morphology of BMCSI 3 was studied by performing Transmission electron microscopy. Cell culture was drop cast onto a carbon-coated Cu grid (Sigma Aldrich), air-dried, and images were acquired under JEM 1400 plus, JEOL Transmission Electron Microscope (120 keV).

Different physicochemical properties of strain BMCSI 3 like gram characteristics, extracellular enzyme activities (catalase, amylase, and protease), substrate hydrolysis (tributarin, gelatin, starch), and substrate utilization (lysine, citrate) were also performed⁶⁵.

Genome sequencing. Cells of BMCSI 3 were cultured on Luria Bertani Agar media (HiMedia) at 30 °C, overnight and genomic DNA was extracted using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, USA). Final DNA concentration and purity were obtained using NanoDrop 1000 (Thermo Scientific, USA) and the genomic DNA integrity was checked on 1.5% agarose gel electrophoresis⁶⁶. Approximately, 1 µg of genomic DNA was used to construct sequencing libraries by using the TruSeq™ DNA PCR-Free library preparation kit (Illumina, Inc., USA). Before library preparation, the genomic DNA was fragmented using Covaris followed by end repair and adapter ligation. Finally, the sequencing was performed on Illumina MiSeq (Illumina, USA) platform.

Assembly and annotation. The raw paired-end fastq reads (2 × 301 bp) were quality checked using FastQC v.0.11.5⁶⁷ followed by trimming of low-quality bases in a sliding window approach using Trimmomatic v.0.39⁶⁸. The cleaned reads were assembled separately using the SPAdes v.3.13.0⁶⁹, VelvetOptimiser v.2.2.4⁷⁰, and ABySS v.1.9.0⁷¹. The comparative evaluation of the assemblies generated from the three assemblers was carried out using the Quality Assessment Tool (QUAST v.5.0.2)⁷². The best assembly (with the highest N50 length) found with the QUAST analysis proceeded further for genome assembly-assisted bacterial strain identification using the MiGA webserver⁷³, StrainSeeker⁷⁴, and TYGS server⁷⁵. The contigs from the Velvet assembly were processed for scaffolding and gap filling using the programs SSPACE v3.0⁷⁶ and GapFiller v.1.10⁷⁷. Finally, the contigs were ordered using the Progressive Mauve⁷⁸ and a similar reference genome of *Bordetella petrii* DSM 12804 (NC_010170.1) to obtain the draft assembly of BMCSI 3. The assembly quality and genomic contaminations, if any, were evaluated using the CheckM—v.1.0.18⁷⁹ in KBase⁸⁰. Further genomic analysis, annotation, and other comparative genomics studies were carried out using this BMCSI 3 draft assembly.

The contig assembly of BMCSI 3 was annotated using the software Prokka v.1.13.7⁸¹, Rapid Annotation System Technology (RAST) Pipeline⁸², PATRIC Pipeline⁸³ and National Center for Biotechnology Information (NCBI) stand-alone Prokaryotic Genome Annotation Pipeline⁸⁴. The visualization of the genome and its typical features was carried out using the CGviewer server beta⁸⁵. Genomic Islands within the genome of BMCSI 3 were predicted using IslandViewer 4 server⁸⁶.

Comparative genomics. Genome sequences available in the public domain of eight representative strains of each *Bordetella* species i.e., *B. pertussis* Tohama I (NC_002929.2), *B. parapertussis* Bpp5 (NC_018828.1), *B. bronchiseptica* 253 (NC_019382.1), *B. trematum* H044680328 (NZ_LT546645.1), *B. holmesii* ATCC 51,541 (NZ_CP007494.1), *B. avium* 197 N (NC_010645.1), *B. hinzii* F582 (NZ_CP012076.1), *B. ansorpii* H050680373 (NZ_FKIF01000001.1), *B. flabilis* AU10664 (NZ_CP016172.1), *B. bronchialis* AU17976 (NZ_CP016171.1) and *B. pseudohinzii* HI4681 (NZ_CP016440.1) including all available strains of *B. petrii* were retrieved from the National Center for Biotechnology Information (NCBI) GenBank (<https://www.ncbi.nlm.nih.gov/genome>).

Genome-informed differential characteristics. Selected representative species of the genus *Bordetella* including all available strains of *B. petrii* were compared to identify the potential differences, viz, the presence or absence of specific genes related to virulence factors, antibiotic resistance, and transporter using PATRIC server⁸⁷. To study the relatedness of all the taxa, BLAST-based average nucleotide identity (ANIb) was investigated using JSpecies server⁸⁸ and heatmap was produced in R version 4.1.1 using "pheatmap" package. For agglomeration of dendrogram, the default method "complete" was used⁸⁹. Circular genome comparison was analyzed using CGView-Circular Genome Viewer⁹⁰. Pangenome were analyzed using Spine and AGent tool⁹¹ considering all six strains of *B. petrii*. Core and accessory genomes were further annotated using RAST Pipeline⁸².

Data availability

This whole-genome shotgun project was deposited at DDBJ/ENA/GenBank under the accession number JAFMZZ000000000, version JAFMZZ000000000.1 (BioProject PRJNA708297 and BioSample SAMN18228559).

Received: 28 March 2022; Accepted: 4 May 2022

Published online: 19 May 2022

References

1. Khan, N. A. *et al.* Occurrence, sources and conventional treatment techniques for various antibiotics present in hospital wastewaters: a critical review. *TrAC, Trends Anal. Chem.* **129**, 115921 (2020).
2. Almágor, J., Temkin, E., Benenson, I., Fallach, N. & Carmeli, Y. The impact of antibiotic use on transmission of resistant bacteria in hospitals: Insights from an agent-based model & DRIVE-AB consortium. *PLoS ONE* **13**, 5 (2018).

3. Ariffin, H., Navaratnam, P., Kee, T. K. & Balan, G. Antibiotic resistance patterns in nosocomial gram-negative bacterial infections in units with heavy antibiotic usage. *J. Trop. Pediatr.* **50**, 26–31 (2004).
4. Gould, I. M. & Jappy, B. Trends in hospital antibiotic prescribing after introduction of an antibiotic policy. *J. Antimicrob. Chemother.* **38**, 895–904 (1996).
5. Aminov, R. I. The role of antibiotics and antibiotic resistance in nature. *Environ. Microbiol.* **11**, 2970–2988 (2009).
6. Li, P. et al. Comparative genome analyses of *Serratia marcescens* FS14 reveals its high antagonistic potential. *PLoS One* **10**, (2015).
7. Bennett, J. W. & Bentley, R. Seeing red: The story of prodigiosin. (2000).
8. Magiorakos, A. P. et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* **18**, 268–281 (2012).
9. Das, S., Dash, H. R. & Chakraborty, J. Genetic basis and importance of metal resistant genes in bacteria for bioremediation of contaminated environments with toxic metal pollutants. *Appl. Microbiol. Biotechnol.* **100**, 2967–2984 (2016).
10. Jung, J., Jeong, H., Kim, H. J., Lee, D. W. & Lee, S. J. Complete genome sequence of *Bacillus oceanisediminis* 2691, a reservoir of heavy-metal resistance genes. *Mar Genomics* **30**, 73–76 (2016).
11. Shahpiri, A., Majnoun, Z., Kazemi-Nasab, A. & Zarei, M. Enhancement of chromate bioaccumulation by engineered *Escherichia coli* cells co-expressing chromate reductase (YieF) and a rice metallothionein isoform (OsMT1). *J. Chem. Technol. Biotechnol.* **96**, 1285–1291 (2021).
12. Zagui, G. S. et al. High occurrence of heavy metal tolerance genes in bacteria isolated from wastewater: A new concern? *Environ. Res.* **196**, (2021).
13. Miyatake, M. & Hayashi, S. Characteristics of Arsenic-methylation and Arsenic-removal by *Bordetella petrii* Strain KC42. *Resour. Process.* **63**, 18–23 (2016).
14. Komijani, M. et al. Heavy metal pollution promotes antibiotic resistance potential in the aquatic environment. *Environ. Polluti.* **274**, (2021).
15. Wang, X., Lan, B., Fei, H., Wang, S. & Zhu, G. Heavy metal could drive co-selection of antibiotic resistance in terrestrial subsurface soils. *J. Hazard. Material.* **411**, (2021).
16. Dickinson, A. W. et al. Heavy metal pollution and co-selection for antibiotic resistance: A microbial palaeontology approach. *Environ. Int.* **132**, 105117 (2019).
17. Biswas, R., Halder, U., Kabiraj, A., Mondal, A. & Bandopadhyay, R. Overview on the role of heavy metals tolerance on developing antibiotic resistance in both Gram-negative and Gram-positive bacteria. *Arch. Microbiol.* **203**, 2761–2770 (2021).
18. Carleton, A. et al. Clustered multidrug-resistant *Bordetella petrii* in adult cystic fibrosis patients in Ireland: case report and review of antimicrobial therapies. in (Society for General Microbiology, 2014).
19. Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).
20. Farris, J. S. Estimating phylogenetic trees from distance matrices. *Am. Nat.* **106**, 645–668 (1972).
21. Srivastava, P. & Kowshik, M. Mechanisms of bacterial heavy metal resistance and homeostasis. *Heavy Metals Environ. Microorgan. Bioremediat.* **15**, (2017).
22. Nies, D. H. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol. Rev.* **27**, 313–339 (2003).
23. Chandrangu, P., Rensing, C. & Helmann, J. D. Metal homeostasis and resistance in bacteria. *Nat. Rev. Microbiol.* **15**, 338–350 (2017).
24. Negrete, F. et al. Genome-wide survey of efflux pump-coding genes associated with *Cronobacter* survival, osmotic adaptation, and persistence. *Curr. Opin. Food Sci.* **30**, 32–42 (2019).
25. Yeh, J. I. et al. High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. *J. Mol. Biol.* **262**, 186–201 (1996).
26. Taylor-Mulneix, D. L. et al. *Bordetella bronchiseptica* exploits the complex life cycle of *Dictyostelium discoideum* as an amplifying transmission vector. *PLoS Biol.* **15**, e2000420 (2017).
27. Gross, R. et al. The missing link: *Bordetella petrii* is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic *Bordetellae*. *BMC Genom.* **9**, 1–14 (2008).
28. Menez, J., Buckingham, R. H., de Zamaroczy, M. & Campelli, C. K. Peptidyl-tRNA hydrolase in *Bacillus subtilis*, encoded by spoVC, is essential to vegetative growth, whereas the homologous enzyme in *Saccharomyces cerevisiae* is dispensable. *Mol. Microbiol.* **45**, 123–129 (2002).
29. Stragier, P. & Losick, R. Molecular genetics of sporulation in *Bacillus subtilis*. *Annu. Rev. Genet.* **30**, 297–341 (1996).
30. Cho, W.-I. & Chung, M.-S. *Bacillus* spores: A review of their properties and inactivation processing technologies. *Food Sci. Biotechnol.* **29**, 1447–1461 (2020).
31. Delbrück, A. I., Zhang, Y., Heydenreich, R. & Mathys, A. *Bacillus* spore germination at moderate high pressure: A review on underlying mechanisms, influencing factors, and its comparison with nutrient germination. *Comprehensive Rev. Food Sci. Food Saf.* **20**, 4159–4181 (2021).
32. Register, K. B. & Sanden, G. N. Prevalence and sequence variants of IS 481 in *Bordetella bronchiseptica*: implications for IS 481-based detection of *Bordetella pertussis*. *J. Clin. Microbiol.* **44**, 4577–4583 (2006).
33. Kholodii, G. Y. et al. Four genes, two ends, and a res region are involved in transposition of Tn5053: A paradigm for a novel family of transposons carrying either a mer operon or an integron. *Mol. Microbiol.* **17**, 1189–1200 (1995).
34. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (1979)* **359**, 6379 (2018).
35. Alcock, B. P. et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525 (2020).
36. Bilinovich, S. M., & Leeper, T. C. Silver binding to bacterial glutaredoxins observed by NMR. *Biophysica* **1**, 359–376 (2021).
37. Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
38. Chi, Y. H. et al. The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. *Front. Plant Sci.* **750**, (2019).
39. Axelsen, K. B. & Palmgren, M. G. Evolution of substrate specificities in the P-type ATPase superfamily. *J. Mol. Evol.* **46**, 84–101 (1998).
40. Basheer, S. M. et al. Structure activity characterization of *Bordetella petrii* lipid A, from environment to human isolates. *Biochimica* **120**, 87–95 (2016).
41. Hamidou Soumana, I., Linz, B. & Harvill, E. T. Environmental origin of the genus *Bordetella*. *Front. Microbiol.* **8**, 28 (2017).
42. Fullen, A. R., Yount, K. S., Dubey, P. & Deora, R. Whoop! There it is: The surprising resurgence of pertussis. *PLoS Pathog.* **16**, e1008625 (2020).
43. Kameyama, H. et al. Pathogenicity of *Bordetella bronchiseptica* isolated from apparently healthy rabbits in guinea pig, rat, and mouse. *J. Vet. Med. Sci.* 21–494 (2022).
44. Gross, R., Keidel, K. & Schmitt, K. Resemblance and divergence: the “new” members of the genus *Bordetella*. *Med. Microbiol. Immunol.* **199**, 155–163 (2010).
45. Wang, Z. et al. Comparative genomic analysis of *Bordetella bronchiseptica* isolates from the lungs of pigs with porcine respiratory disease complex (PRDC). *Infect. Genet. Evolut.* **81**, (2020).

46. Moriuchi, R., Dohra, H., Kanesaki, Y. & Ogawa, N. Complete genome sequence of 3-chlorobenzoate-degrading bacterium *Cupriavidus necator* NH9 and reclassification of the strains of the genera *Cupriavidus* and *Ralstonia* based on phylogenetic and whole-genome sequence analyses. *Front. Microbiol.* **10**, 133 (2019).
47. Smith, H. B. *et al.* *Listeria monocytogenes* MenI encodes a DHNA-CoA thioesterase necessary for menaquinone biosynthesis, cytosolic survival, and virulence. *Infect. Immun.* **89**, (2021).
48. Lamb, H. K., Dodds, A. L., Swatman, D. R., Cairns, E. & Hawkins, A. R. Deletion of the N-terminal region of the AREA protein is correlated with a derepressed phenotype with respect to nitrogen metabolite repression. *J. Bacteriol.* **179**, 6649–6656 (1997).
49. Holden, E. R. *et al.* Massively parallel transposon mutagenesis identifies temporally essential genes for biofilm formation in *Escherichia coli*. *Microb. Genom.* **7**, (2021).
50. Dailey, H. A. Terminal steps of haem biosynthesis. *Biochem. Soc. Trans.* **30**, 590–595 (2002).
51. Padovani, D., Thomas, F., Trautwein, A. X., Mulliez, E. & Fontecave, M. Activation of Class III Ribonucleotide Reductase from *E. coli*. The Electron Transfer from the Iron–Sulfur Center to S-Adenosylmethionine. *Biochemistry* **40**, 6713–6719 (2001).
52. Ogasawara, H., Ohe, S. & Ishihama, A. Role of transcription factor NimR (YeaM) in sensitivity control of *Escherichia coli* to 2-nitroimidazole. *FEMS Microbiol. Lett.* **362**, 1–8 (2015).
53. Iwasaki, H. *et al.* A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. *Cell* **101**, 223–233 (2000).
54. Abramson, J. *et al.* The structure of the ubiquinol oxidase from *Escherichia coli* and its ubiquinone binding site. *Nat. Struct. Biol.* **7**, 910–917 (2000).
55. Fabianek, R. A., Hofer, T. & Thöny-Meyer, L. Characterization of the *Escherichia coli* CcmH protein reveals new insights into the redox pathway required for cytochrome c maturation. *Arch. Microbiol.* **171**, 92–100 (1999).
56. Grass, G., Wong, M. D., Rosen, B. P., Smith, R. L. & Rensing, C. ZupT is a Zn (II) uptake system in *Escherichia coli*. *J. Bacteriol.* **184**, 864–866 (2002).
57. Franke, S., Grass, G., Rensing, C. & Nies, D. H. Molecular analysis of the copper-transporting efflux system CusCFBA of *Escherichia coli*. *J. Bacteriol.* **185**, 3804–3812 (2003).
58. Locher, K. P. Structure and mechanism of ATP-binding cassette transporters. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 239–245 (2009).
59. <https://www.ncbi.nlm.nih.gov/genome>.
60. Fujishima, K. & Kanai, A. tRNA gene diversity in the three domains of life. *Front. Genet.* **5**, 142 (2014).
61. Halder, U. *et al.* Production of prodigiosin by a drug-resistant *Serratia rubidaea* HB01 isolated from sewage. *Environ. Sustain.* **3**, 279–287 (2020).
62. Valsesia, G., Roos, M., Böttger, E. C. & Hombach, M. A statistical approach for determination of disk diffusion-based cutoff values for systematic characterization of wild-type and non-wild-type bacterial populations in antimicrobial susceptibility testing. *J. Clin. Microbiol.* **53**, 1812–1822 (2015).
63. Sarker, M. R. *et al.* Studies of the impact of occupational exposure of pharmaceutical workers on the development of antimicrobial drug resistance. *J. Occup. Health* **12–14** (2014).
64. Saranya, K. *et al.* Biosorption of multi-heavy metals by coral associated phosphate solubilising bacteria *Cronobacter muytjensii* KSCAS2. *J. Environ. Manage* **222**, 396–401 (2018).
65. Technic., Committee on Bacteriological, Pelczar, M. J., Bard, R. C. & Burnett, G. W. *Manual of microbiological methods*. (McGraw-Hill Book Company, 1957).
66. Halder, U. *et al.* Draft genome report of *Bacillus altitudinis* SORB11, isolated from the Indian sector of the Southern Ocean. *Genome Announc.* **5**, (2017).
67. Andrews, S. *et al.* FastQC. A quality control tool for high throughput sequence data. **370**, (2010).
68. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
69. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
70. Gladman, S. & Seemann, T. Velvet optimiser. <http://bioinformatics.net.au/software/velvetoptimiser.shtml> (2008).
71. Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
72. quality assessment tool for genome assemblies. Gurevich, A., Saveliev, V. V. & N., T. G. QUAST. *Bioinformatics* **29**, 1072–1075 (2013).
73. Rodriguez-R, L. M. *et al.* The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* **46**, W282–W288 (2018).
74. Roosaare, M. *et al.* StrainSeeker: Fast identification of bacterial strains from raw sequencing reads using user-provided guide trees. *PeerJ* **5**, (2017).
75. Meier-Kolthoff, J. P. & Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **10**, 1–10 (2019).
76. Boetzer, M. & Pirovano, W. SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinf.* **15**, 1–9 (2014).
77. Nadalin, F., Vezzi, F. & Policriti, A. GapFiller: A de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics* **13**, 1–16 (2012).
78. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, 6 (2010).
79. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
80. Arkin, A. P. *et al.* KBase: The United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
81. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
82. Aziz, R. K. *et al.* The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 1–15 (2008).
83. Brettin, T. *et al.* RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 1–6 (2015).
84. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).
85. Grant, J. R. & Stothard, P. The CGView Server: A comparative genomics tool for circular genomes. *Nucleic Acids Res* **36**, W181–W184 (2008).
86. Bertelli, C., Laird, M. R. & Williams, K. P. Simon Fraser University Research Computing Group; Lau, BY; Hoard, G.; Winsor, GL; Brinkman, FSL IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* **45**(W1), W30–W35 (2017).
87. Davis, J. J. *et al.* The PATRIC bioinformatics resource center: expanding data and analysis capabilities. *Nucleic Acids Res* **48**(D1), D606–D612 (2020).
88. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* **106**, 19126–19131 (2009).
89. Team, R. C. R: A language and environment for statistical computing. (2013).
90. Stothard, P. & Wishart, D. S. Circular genome visualization and exploration using CGView. *Bioinformatics* **21**, 537–539 (2005).

91. Ozer, E. A., Allen, J. P. & Hauser, A. R. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics* **15**, 1–17 (2014).

Acknowledgements

We are thankful to the UGC-Centre for Advanced Study, Department of Botany, The University of Burdwan, and DST FIST II. We are thankful to the University Science Instrumentation Centre, The University of Burdwan for the TEM and XRF analysis. We are also thankful to ICRISAT, Hyderabad for providing us genome sequencing facilities and Director, ICAR-CRIJAF for helping with genome analysis facilities.

Author contributions

Sampling, R.B.1 U.H. and S.C.; conceptualization R.B. and methodology, U.H., R.B.1 and S.C.; experiments, U.H., A.K., M.L., S.S., B.D., A.L., K.M., R.K.R., A.D., D.K.; sequencing, A.C., R.K.V.; software and visualization, D.S. and U.H.; writing and editing, U.H., A.K., R.D., R.B., R.B.1, and D.S.

Funding

UH is thankful to SRF (State Fund) fellowship [Fc (Sc.)/RS/SF/ BOT./2017-18/22], AK is thankful to DHESTBT (WB-DBT), Memo No. [Fc (Sc.)/RS/SF/ BOT./2017-18/22], ML and RKR are thankful to UGC JRF for financial assistance to conduct this research work. AD is thankful to DS Kothari Fellowship.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12435-7>.

Correspondence and requests for materials should be addressed to R.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022