# Genetic exchange shapes ultra-small Patescibacteria metabolic capacities in the terrestrial subsurface

Authors: Emilie Gios[1], Olivia E. Mosley[1], Nobuto Takeuchi[1], Kim M. Handley[1,*]

[1]School of Biological Sciences, The University of Auckland, Auckland, New Zealand

*Corresponding author: K. M. Handley kim.handley@auckland.ac.nz

**ABSTRACT**

Bacterial genomes are highly dynamic entities, mostly due to the extent of horizontal gene transfer (HGT) occurring in these organisms. HGT is thought to be the main driver of genetic variation and adaptation to local environment in bacteria. However, little is known about the modalities of HGT within natural microbial communities, especially the implications of genetic exchange for streamlined microorganisms such as Patescibacteria (Candidate Phyla Radiation). We searched for evidence of genetic exchange in 125 Patescibacteria genomes recovered from aquifer environments and detected the presence of hundreds of genomic islands, individually transferred genes and prophage combined, with up to 29% of genome length attributed to HGT. Results show that most individual gene transfer events occurred between Patescibacteria, but donors were also phylogenetically diverse groundwater microorganisms. Using gene donor-recipient information, we identified one potential host (Omnitrophota) of the ultra-small bacteria, and confirmed this by matching relative abundance patterns across 16 groundwater samples. A wide variety of metabolic functions were introduced in Patescibacteria genomes by HGT including transcription, translation and DNA replication, recombination and repair. This study illustrates the evolutionarily dynamic nature of Patescibacteria genomes despite the constraints of streamlining, and that HGT in these organisms is also mediated via viral infection.

## INTRODUCTION

The horizontal transfer of genetic material (or horizontal gene transfer; HGT), involving acquisition of exogenous DNA, is thought to be a key factor in the evolution of prokaryotes, especially in bacteria (Ochman et al., 2000). HGT accounts for a significant fraction of gene diversity and genomic fluidity in these organisms, leading to a variety of genome structures and physiological diversity (Soucy et al., 2015). This evolutionary process is important for niche adaptation, and is often driven by mobile genetic elements (MGEs) such as genomic islands (GIs), transposons and phage. Many routes exist for gene transfer: uptake of environmental DNA molecules (transformation), exchange between two cellular organisms via a physical bond (conjugation), or delivery through phage infection (transduction).

Recent metagenomic studies of subsurface ecosystems reveal a large group of unusual bacteria referred to as the Patescibacteria phylum (equating to the Candidate Phyla Radiation; CPR), which has been suggested to comprise up to 26% of all currently known bacterial diversity (Parks et al., 2017; Schulz et al., 2017). The group constitutes taxa consistently harbouring small cell and genome sizes (Brown et al., 2015). Implications of reduced size include limited metabolic functions, such as the lack of known amino acid and nucleotide biosynthesis pathways (Castelle et al., 2018). These observations led to the prediction that most ultra-small Patescibacteria have a symbiotic lifestyle (Brown et al., 2015), and some were confirmed to be episymbionts based on cultivation and microscopy studies (He et al., 2021). Close physical associations with other organisms due to metabolic dependencies, along with the long residence time of groundwater (Griebler & Lueders, 2009), and the importance of the biofilms in aquifers (Flynn et al., 2008), could favour genetic exchange in subsurface ultra-small microbial communities.

52 Despite proximity to other groundwater bacteria, extensive gene loss in
53 Patescibacteria evolution could result in selective pressures against the acquisition
54 and retention of exogenous DNA that are stronger than in non-streamlined genomes.
55 Another potential barrier to HGT lies in the fact that genome reduction in obligate
56 intracellular symbionts has often been associated with loss of functions involved in
57 DNA recombination and repair (Moran & Bennett, 2014), which could limit integration
58 of foreign DNA. Previous studies have, however, shown not only that a significant
59 fraction of Patescibacteria organisms still possesses genes required for homologous
60 recombination (Castelle et al., 2018), but that they are naturally competent for uptake
61 of eDNA molecules, mediated by DNA intake pumps (via comEC and pili complexes;
62 Castelle et al., 2017, 2018; Nelson & Stegen, 2015).
63 Investigating the mobility of adaptive genes via HGT could provide important
64 information about the evolutionary strategy of these ultra-small organisms, including
65 aspects governing putative host-symbiont interactions. Implications of HGT include
66 altered interaction modalities with host populations, as described in *Vibrio fischeri*.
67 Adding a single gene to the *V. fischeri* genome is sufficient to alter its host range in
68 squids (Mandel et al., 2009). Additionally, acquisition of MGEs has been shown to
69 constitute evidence of symbiotic relationships between HGT recipient and donor, such
70 as extensive gene transfer from *Wolbachia* endosymbionts to their insect hosts during
71 genome reduction (Dunning Hotopp et al., 2007), and putative acquisition of host-
72 derived ankyrin repeat proteins by bacterial sponge symbionts to inhibit phagocytosis
73 (Nguyen et al., 2014). How potentially adaptive genes aid niche adaptation in ultra-
74 small prokaryotes and fine-tune metabolic interactions with their predicted 'basibiont'
75 prokaryotic host cells remains largely unknown.

To understand the role of HGT in the evolution of complex, natural groundwater communities, and particularly streamlined Patescibacteria, we analyzed 396 metagenome-assembled genomes (MAGs), including 132 Patescibacteria genomes. We investigated genomic evidence for, and the frequency of, individual horizontal gene transfer events (defined here as HT genes) in the MAGs, and compared these to acquisitions of larger MGEs such as GIs and prophage. Results provide evidence for many HGT events among Patescibacteria, and exchanges with other prokaryotes in groundwater. We report that metabolic functions horizontally acquired by Patescibacteria taxa appear to be biased towards features distinct from the general groundwater communities.

## MATERIALS AND METHODS

### Sample collection

Groundwater samples were collected from 8 wells from 2 sand/gravel aquifers in Canterbury, New Zealand, as described by (Mosley et al., 2022). Briefly, wells were purged (×3–5 bore volumes), then 3–90 L of groundwater was filtered per sample. Biomass was captured on 0.22 µm mixed cellulose ester (MCE) filters, after passing through a 1.2 µm MCE pre-filter, using a 142 mm filter holder (Merck Millipore Ltd., Cork, Ireland). Samples were preserved in RNAlater (ThermoFisher Scientific, Waltham, MA, USA), transported on dry ice, and stored at -80°C.

After collecting a groundwater sample (as above), a second sample per well was collected following low frequency sonication (2.43 kW for 2 minutes) to induce biofilm (and sediment) detachment from the surrounding aquifer, using a custom sonication device (Close et al., 2020). 0.5–15 L of sediment (and hence biomass) enriched groundwater was filtered as above.

101

102 **Nucleic acid extraction**

103 To remove RNAlater, filters were centrifuged (2500 g for 5 min) and washed with
104 nuclease-free phosphate-buffered saline (pH 7.4, ThermoFisher Scientific, Waltham,
105 MA, USA). DNA was extracted using the DNeasy PowerSoil Pro Kit (Qiagen, Valencia,
106 CA, USA) and 0.25-0.9 g of whole filter per reaction. Samples were extracted in
107 replicate (1–47 reactions/sample). Replicates were pooled and concentrated using a
108 sodium acetate/ethanol precipitation with glycogen (0.1 µg/µL final concentration,
109 Roche, Basel, Switzerland). High molecular weight DNA was confirmed via gel
110 electrophoresis. DNA was quantified using Qubit fluorometry (ThermoFisher
111 Scientific), and quality was checked using a NanoPhotometer (Implen, Munich,
112 Germany). Extractions yielded 70–864 ng of DNA (8.7 ng for gwj02) for
113 metagenomics.

114

115 **Metagenome sequencing**

116 Whole genome shotgun sequencing was undertaken on the 16 Canterbury samples
117 (gwj01-16). DNA libraries were prepared using the Illumina TruSeq Nano DNA Kit with
118 ~550 bp inserts, by Otago Genomics (University of Otago, Dunedin, NZ). The low-
119 yield gwj02 sample was prepared with the ThruPLEX DNA-Seq Kit (Takara Bio USA,
120 Inc., Mountain View, CA, USA). Sequences (2×250 bp) were generated using the
121 Illumina HiSeq 2500 V4 platform. Adapter sequences were removed using Cutadapt
122 v2.10 (Martin, 2011), and reads were quality trimmed using sickle v01.33 (parameters
123 -q 30 -l 80). Read quality was checked using FastQC v0.11.7
124 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

125

126 **Metagenome assembly and binning**

127 The 16 metagenomes were individually assembled using SPAdes v3.11.1 (Bankevich
128 et al., 2012) (parameters: --meta, -k 43,55,77,99,121). Metagenomes from the same
129 well also were co-assembled (same parameters). A dereplicated set of metagenome-
130 assembled genomes (MAGs) (99% average nucleotide identity, ANI, dereplication
131 threshold) was generated as described previously (Tee et al., 2020), and refined
132 according to bin_detangling script (https://github.com/dwwaite/bin_detangling).

133 For ultra-small prokaryotes, genome completeness and contamination were
134 assessed based on the presence of 51 bacterial single-copy genes (SCGs) for
135 Dependentiae, 43 SCGs for Patescibacteria and 38 archaeal SCGs for DPANN
136 (Anantharaman et al., 2016). For other prokaryotes, completeness was estimated
137 using CheckM v1.0.13 (Parks et al., 2015). To determine coverage, reads were
138 mapped onto the de-replicated MAGs using Bowtie (Langmead et al., 2009). Sample-
139 specific genome relative abundance was calculated by normalizing based on the
140 highest read count between samples (Probst et al., 2018).

141

142 **Gene prediction and functional annotation**

143 Protein coding sequences were predicted using Prodigal v2.6.3 (Hyatt et al., 2010).
144 Genes were annotated using USEARCH (Edgar, 2010) against UniRef100 (Suzek et
145 al., 2015), Uniprot (The UniProt Consortium, 2017) and KEGG (Kanehisa et al., 2016)
146 databases (e-value ≤0.001 and identity ≥50%), and HMMs against PFAM
147 (Sonnhammer et al., 1998) and TIGRFAM (Haft et al., 2003) databases using HMMER
148 v3.1b2 (Potter et al., 2018) (e-value ≤0.001). Additionally, functional annotation was

149 carried out using eggNOG-mapper v2 (Huerta-Cepas et al., 2017) against the
150 eggNOG v5.0 database (Huerta-Cepas et al., 2019), using default parameters.
151
152 **Genome taxonomy**

153 Taxonomic classification of MAGs was performed using the Genome Taxonomy
154 Database Toolkit (GTDB-Tk, v0.2.1) (Chaumeil et al., 2020) with database release 89
155 (Parks et al., 2020). Maximum likelihood trees were constructed using 120 bacterial
156 and 122 archaeal concatenated marker genes with IQ-TREE v1.6.12 (L.-T. Nguyen et
157 al., 2015) using 100 bootstraps, and ModelFinder (Kalyaanamoorthy et al., 2017) best-
158 fit model LG+F+G4 for bacteria and LG+F+I+G4 for archaea. Trees were visualized
159 and annotated with iTOL (Letunic & Bork, 2007). MAGs (≥80% complete) were
160 compared to GTDB representative genomes (downloaded 20-Jan-2020) by
161 calculating Average Amino-acid Identities (AAI) for blastp matches sharing ≥30%
162 identity over ≥70% of alignment length.

163

164 **Identification of horizontal gene transfer**

165 Individual HT events among MAGs were identified using MetaCHIP v1.10.0 (Song et
166 al., 2019) at phylum, class, order, family and genus levels. False-positive HT genes
167 (potentially introduced via assembly contamination) were filtered out by removing HT
168 genes marked by MetaCHIP as “full-length match” (where the match represents a
169 large proportion of contig) or “end match” (HT genes located at the end of contig) as
170 described by Dong et al. (2021). Information about HGT events involving members of
171 the Patescibacteria phylum were retrieved and visualised using the Circlize package
172 in R (Gu et al., 2014).

173 For genomic island prediction, contigs were concatenated into contiguous
174 sequences and uploaded to IslandViewer4 web server (Bertelli et al., 2017) for
175 genomic island prediction. GIs that were predicted by at least one method, and that
176 did not overlap two concatenated contigs, were considered for the analysis. Genes
177 located on GIs were compared to community gene content using blastn.
178 Prophage integrated in the Patescibacteria genomes were identified using
179 VIBRANT 1.2.1 (Kieft et al., 2020). Whenever a predicted viral sequence was
180 overlapping a predicted GI, CheckV quality assessment information was used and only
181 "High-quality" genomes were retained as prophage (Nayfach et al., 2021).
182

183 **Phylogenetic distance**

184 Average Amino-acid Identities (AAI) between MAGs (≥80% complete) involved in HGT
185 events were determined via blastp matches sharing ≥30% identity over ≥50% of query
186 and database alignment length.

187

188 **Orthology analysis**

189 To investigate the Patescibacteria pangenome, we performed an orthologous
190 clustering of 118,635 protein-coding sequences encoded by the recovered MAGs.
191 Pairwise all-vs-all blastp searches were conducted using the following cut-offs: ≥30%
192 nucleotide identity over ≥50% of query sequence length, e-value ≤0.001. BLAST table
193 outputs were used as input for orthogroup clustering using Orthofinder v2.3.1 (Emms
194 & Kelly, 2019) with -b option and default parameters.
195

196 **RESULTS AND DISCUSSION**

197 **Prevalence of HGT events in Patescibacteria genomes**

198 Using the community-level HGT detection tool MetaCHIP (Song et al., 2019),
199 we investigated the potential horizontal transfer of individual genes between aquifer
200 microorganisms. A total of 1,487 transfers were identified among the 396 groundwater
201 MAGs (Table S1), including 1,407 unique HT genes. Taxonomic breakdown of
202 detected HT events revealed that over half of the Patescibacteria genomes
203 experienced HT with another genome in the dataset, regardless of their taxonomy or
204 estimated genome size (Fig. 1a,b). One hundred and twenty-three (9% of) horizontally
205 acquired genes were identified in 68 Patescibacteria organisms, representing 1.0 ±
206 1.2 HT (average ± standard deviation) genes received per genome and 1.1 ± 1.3 HT
207 genes per Mbp (Fig. 1b). In comparison, genomes of other members of the community
208 contained on average 3.5 ± 4.8 horizontally transferred genes per genome, but a
209 comparable 1.4 ± 2.1 per Mbp. The comparison of HT events per Mbp indicates that
210 the rate of exchange of individual genes is mostly independent of genome size (i.e.
211 larger genomes do not receive proportionally more genes). This is supported by
212 significant positive correlations between HT events per genome and estimated
213 genome size (both donor and recipient HT events, Fig. 2), and is consistent with
214 previous observations of HT acquired genes in prokaryotic genomes (Garcia-Vallvé et
215 al., 2000), and of transfers occurring between distantly related organisms (i.e. cross-
216 order transfers, >85% of the total HT events detected here) (Cordero & Hogeweg,
217 2009).

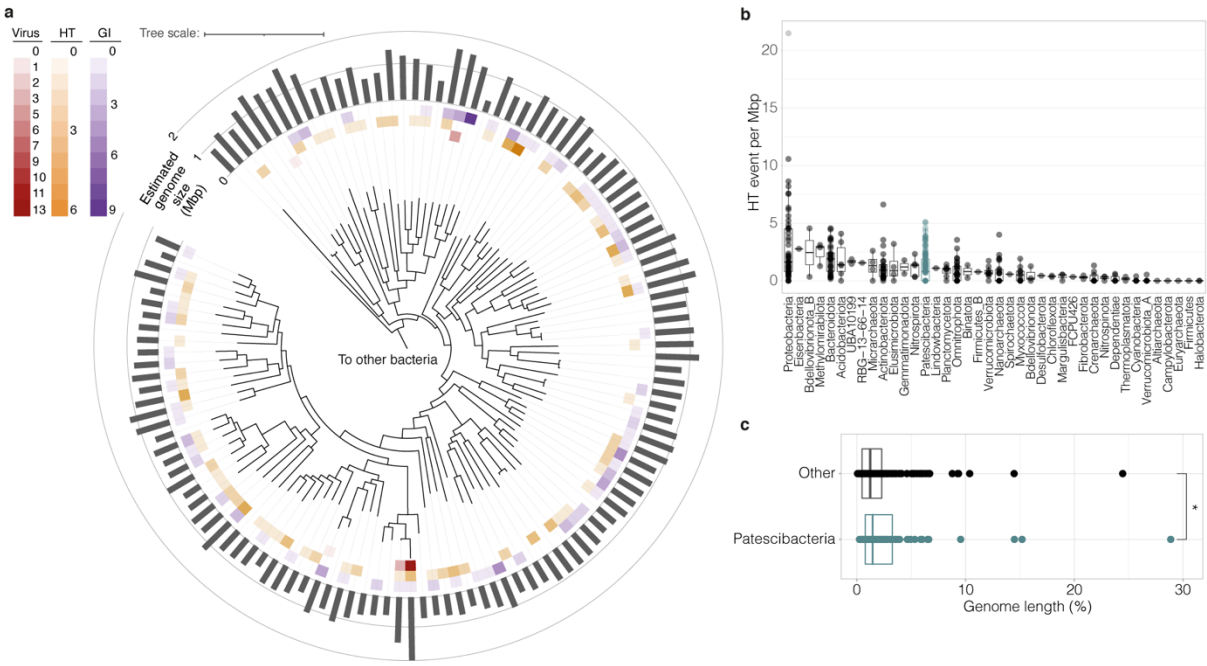
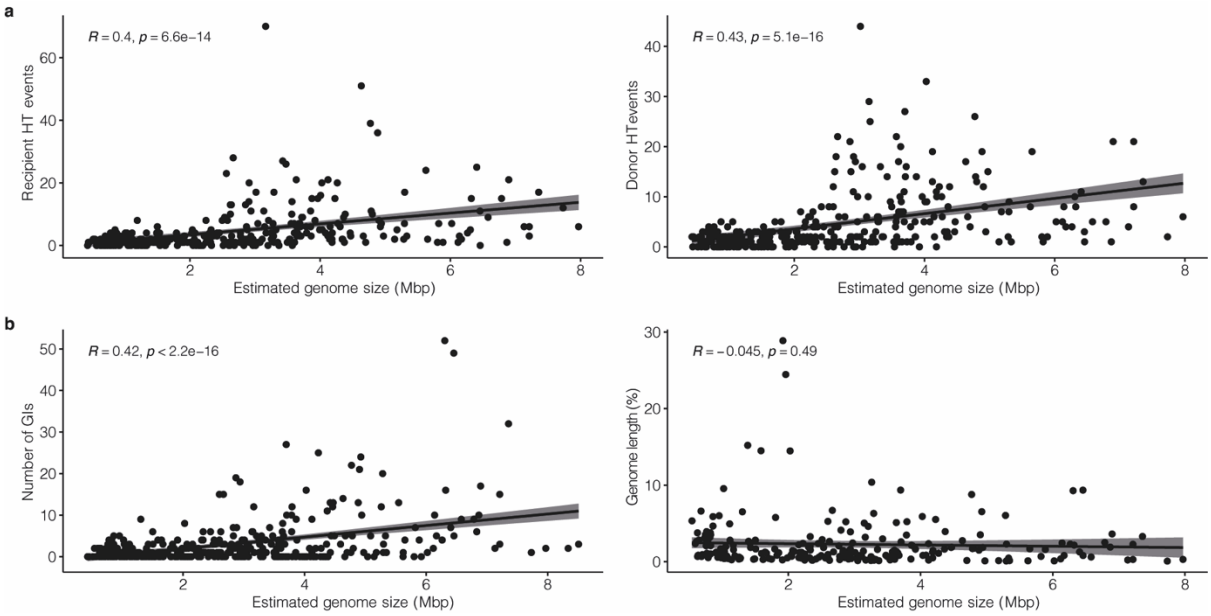



**Figure 1. Prevalence of HGT in Patescibacteria and the wider groundwater community.** (a) Maximum likelihood phylogenetic trees of 132 Patescibacteria MAGs recovered in this study. Trees are based on 120 concatenated bacterial marker gene alignment from GTDB-Tk. Rings from the center: number of prophage; number of HT events; number of GIs; estimated genome size (Mbp). The scale bar indicates the number of substitutions per site. (b) Number of recipient HT events between all groundwater MAGs per phylum normalized by genome size. Patescibacteria phylum is indicated in teal. The center line of each boxplot represents the median; the top and bottom lines are the first and third quartiles, respectively; and the whiskers show 1.5× the interquartile range. (c) Proportion of each genome occupied by acquired GIs and prophage sequences, comparing Patescibacteria (teal) and the rest of the community (black). Significant differences were assessed for each group using Wilcoxon Signed Rank (*: $p < 0.05$).

When searching for large MGEs, we found 120 GIs in 63 Patescibacteria MAGs, ranging from 2.7 to 45 kbp long, with 40 GIs over 10 kb long (average length 17.4 ± 0.7 kb) (Table S2). Additionally, 24 putative prophage genomes (30% of those overall) were found to be concentrated in 5 Patescibacteria MAGs – with the proportion of prophage in Patescibacteria being comparable to the proportion of Patescibacteria MAGs in the community (33%). This frequency of prophage in ultra-small bacterial genomes is surprising, as prophage frequency has been shown to increase with genome size (Touchon et al., 2016). Accordingly, our analysis shows that, as also

241 observed for HT genes, the frequency of acquisition of longer complex DNA
242 sequences, constituting GIs, increased with the genome size of groundwater-derived
243 MAGs (Fig. 2b).
244



245

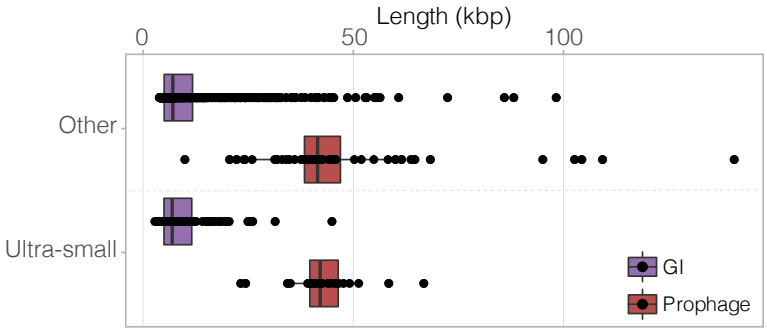
246 **Figure 2.** Correlations between the estimated genome size of groundwater-derived
247 MAGs and HGT. (a) Estimated genome size compared with recipient HT events (left)
248 and donor HT events (right). (b) Estimated genome size compared with the number of
249 GIs detected (left), and the proportion of genome length occupied by GI and prophage
250 sequences (right).
251

252 When considering the groundwater community overall, the proportion of
253 genome length occupied by GIs and prophages, combined, remained constant
254 regardless of recipient genome size (Fig. 2b), and despite GI and prophage sequences
255 being similar in length between ultra-small and other microorganisms (Fig. 3, Table
256 S2). This is consistent with observations, by Ochman et al. (2000), that the proportion
257 of horizontally transferred genes is independent of the amount of protein-coding
258 sequences in complete prokaryotic genomes. However, when comparing the two
259 community fractions, results here show the proportion of genome length occupied by
260 putative GIs and prophage was slightly, but significantly higher in Patescibacteria than

261 the rest of the communities – on average 3.0 ± 4.4% versus 2.0 ± 2.8%, respectively
262 (Fig. 1c).

263 The extent of gene transfer events detected here is comparable to those
264 detected in a variety of other phylogenetically diverse microbial communities, e.g.
265 90.4% of groundwater MAGs containing detectable HGT events, compared to 78.8%
266 in cheese-associated communities, and 89.5% in prokaryotic isolate genomes
267 (Bonham et al., 2017; Ochman et al., 2000). However, the extent of genetic exchange
268 observed in the groundwater microbial community is likely underestimated. Ancient
269 transfer events will escape detection due to robust integration within the recipient
270 genome via modification of GC content and tetranucleotide frequencies (Lawrence &
271 Ochman, 1997). Moreover, while MetaCHIP has proven efficient in predicting HGT
272 donors/recipients relationships within microbial communities, donor taxa might not
273 have been captured in the MAG dataset.

274




275

276 **Figure 3.** Length distribution of GIs and prophage in ultra-small MAGs and other
277 groundwater microbial MAGs. The center line of each boxplot represents the median;
278 the top and bottom lines are the first and third quartiles, respectively; and the whiskers
279 show 1.5× the interquartile range. No significant differences were detected in GI and
280 prophage sequence length between ultra-small and other groundwater
281 microorganisms (Wilcoxon Signed Rank, p>0.05).

282

283 **HGT origins in Patescibacteria taxa**

Most HT events detected occurred between members of the Patescibacteria (Fig. 4a), with 104 of the 123 HT genes received by Patescibacteria originating from counterparts. This observation is consistent with research, which demonstrates that HGT is most prevalent among closely related organisms (Bonham et al., 2017), and that intra-species HGT occurs at far greater frequency (Majewski et al., 2000). HGT was likely also facilitated by the spatial proximity of microorganisms (Gogarten et al., 2002), in particular by microbial cohorts in groundwater communities (Hug et al., 2015). This includes cohorts formed by ultra-small prokaryotes, where individuals spatially co-occur as described previously in subsurface and soil ecosystems (Geesink et al., 2020; Herrmann et al., 2019; Nicolas et al., 2021).

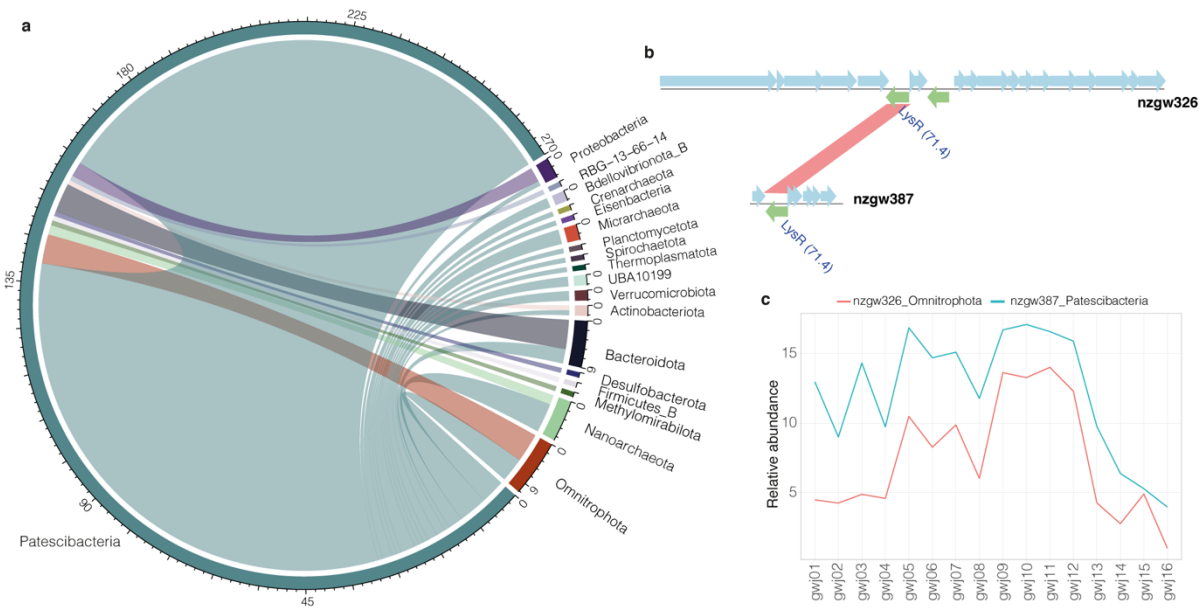



**Figure 4. Origins of HT genes in Patescibacteria taxa.** (a) Gene flow among phyla donating to Patescibacteria taxa. Scale represents number of HT events and inner bands show donor/recipient relationship (coloured according to donor organism). (b) Example of HT gene transferred from Omnitrophota nzgw326 (upper plot) to Patescibacteria nzgw387 (lower plot). Genes encoded on the forward strand are displayed in blue, and genes coded on the reverse strand are displayed in green. (c) Genome relative abundance profiles of these 2 organisms across 16 groundwater samples.

305 Although a smaller number of MAGs were recovered from Omnitrophota and
306 Bacteroidota, organisms in these phyla shared genes more frequently with
307 Patescibacteria (3.5x and 3x more, respectively) than, for example, Proteobacteria,
308 which was the second most populous phylum (Tables 1, S3). Frequent sharing
309 suggests potential close relationships between these organisms and ultra-small
310 prokaryotes (i.e. host/symbiont). The former two bacterial groups have been identified
311 previously as putative HGT donors to Patescibacteria in seafloor environments (Dong
312 et al., 2021). Moreover, Omnitrophota and Bacteroidota organisms were found to
313 consistently co-occur with the ultra-small bacteria in aquifers (Chaudhari et al., 2021).
314

315 **Table 1.** Frequency of HT events to Patescibacteria at phylum level. Note the ratio of
316 donor HGT events to number of MAGs is biases in donor phyla with small MAG sample
317 sizes (Methylomirabilota, Firmicutes_B, Desulfobacterota, RBG-13-66-14).

| Donor phylum | Donor HGT events | Reconstructed donor MAGs | Ratio donor HGT events / number of MAGs |
|---|---|---|---|
| Patescibacteria | 109 | 132 | 0.83 |
| Omnitrophota | 6 | 28 | 0.21 |
| Bacteroidota | 6 | 33 | 0.18 |
| Nanoarchaeota | 2 | 23 | 0.09 |
| Proteobacteria | 3 | 51 | 0.06 |
| Actinobacteriota | 1 | 24 | 0.04 |
| Methylomirabilota | 1 | 3 | 0.33 |
| Firmicutes_B | 1 | 1 | 1.00 |
| Desulfobacterota | 1 | 1 | 1.00 |
| RBG-13-66-14 | 1 | 1 | 1.00 |

318

319 We observed further evidence for recent putative horizontal transfers to
320 groundwater Patescibacteria from patescibacterial counterparts and other taxa based
321 on positive correlations between the relative abundances of HGT pairs across sites
322 (29 out of 131 pairs involving Patescibacteria as donor or recipient with Pearson's
323 correlation coefficient >0.5, $p < 0.05$, Table S4). For example, Patescibacteria nzgw387

324 received a copy of the *lysR* gene from Omnitrophota nzgw326 (Fig. 4b), a highly
325 conserved transcriptional regulator involved in numerous cellular functions (Maddocks
326 & Oyston, 2008), and is shown to co-occur with its donor, Omnitrophota nzgw326, in
327 groundwater (Fig. 4c).

328 A BLAST analysis of GI genes against the gene content of whole groundwater
329 communities showed that most of the genes located on GIs could not be matched to
330 any organisms within the communities (91.9% of GI genes; Table S5). Nonetheless,
331 we found that 7% of GI genes were related to those in Patescibacteria other than the
332 one carrying the GI. Interestingly, the content of one 4,250 bp long GI in
333 Patescibacteria nzgw468 (74% genome completeness, no contamination) was highly
334 similar to genes in the archaeon Halobacterota nzgw24 (99% completeness, 1.3%
335 contaminated), indicating a potential recent transfer event between the two distantly
336 related organisms. None of the 11 genes located on the GI could be functionally
337 characterized (as for a large proportion of genes in Patescibacteria genomes), but
338 were between 85 and 100% identical in amino acid composition to the ones of nzgw24.
339 While inter-domain transfers do occur between Archaea and Bacteria, they are less
340 frequent than bacterial-domain confined HGT events (Nelson et al., 1999, López-
341 García et al., 2015). Gene exchange between Patescibacteria and their archaeal ultra-
342 small counterparts (DPANN) has been reported (León-Zayas et al., 2017), particularly
343 in the context of membrane-associated protein encoding genes potentially involved in
344 host-symbiont interactions (Jaffe et al., 2016). Accordingly, we detected 9 HT events
345 from Patescibacteria to DPANN archaea (Micrarchaeota and Nanoarchaeota; function
346 of transferred genes discussed below). A further, 2 genes (one ABC-transporter and
347 one gene related to the BRO family, of which the function remains unknown) were
348 transferred from Nanoarchaeota to Patescibacteria (Table S6).

349

350 **Metabolic functions encoded in horizontally acquired regions**

351 Patescibacteria, or CPR bacteria, have reduced genomes and limited biosynthetic
352 capacities (Castelle et al., 2018). The metabolic function of horizontally transferred
353 genes that are retained in Patescibacteria genomes could provide clues about the
354 evolution of dependencies in these ultra-small prokaryotes regarding their symbiotic
355 interactions with hosts. Although a substantial portion of genes located in horizontally
356 acquired genomic regions could not be assigned any function (61.6% of all genes for
357 Patescibacteria, 38.9% for other taxa), genes located in those regions in
358 Patescibacteria genomes that were able to be annotated are involved in a wide range
359 of metabolic pathways (Fig. 5), and include a large number of genes involved in
360 translation and transcription. Horizontally acquired translation and transcription genes
361 predominantly comprised ribosomal proteins (small and large subunits), elongation
362 factor *tuf*, which plays a central role in protein synthesis, and chaperonin *groS* (Table
363 S6). These results reflect trends identified when considering the most frequently
364 transferred genes to other groundwater prokaryotes, which notably included the *rpsL*
365 gene encoding ribosomal protein S12 and transcription initiation factor *infA*, alongside
366 translation elongation factor *tuf*. Kanhere and colleagues described similar findings
367 (Kanhere & Vingron, 2009). By inferring HGT events based on evolutionary distances
368 within orthologous protein families (using the COG database), authors found that
369 genes exchanged among bacteria were primarily involved in translation (Kanhere &
370 Vingron, 2009).

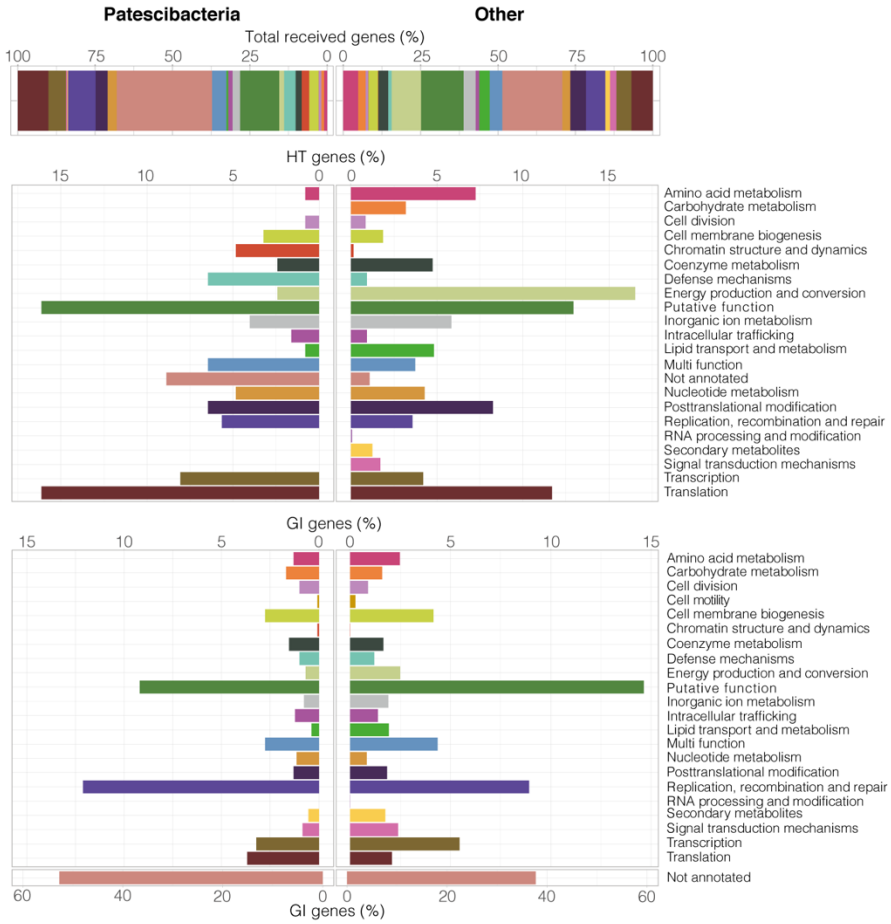



**Figure 5. Metabolic functions of genes received through horizontal transfer.** Metabolic functions acquired through HGT (HT and GI genes) by Patescibacteria (124 HT genes + 120 GIs received) and other groundwater microorganisms (1,283 HT genes + 926 GIs received) from any other member of the groundwater community (top). Breakdown of metabolic functions of HT genes (middle) and genes located on GIs (bottom).

HT genes in Patescibacteria were found to be highly enriched, compared to other phyla, in genes assigned to the EggNOG "Defense mechanisms" category (Fig. 5). These included genes (e.g. *macB* and *msbA*) encoding ABC-type systems for multidrug and antimicrobial peptide transport. Membrane-located multidrug transporters have been identified previously in relatives of ultra-small bacterium *Babela massiliensis* (Dependentiae phylum, previously Candidate division TM6). Their specialization is thought to be important at the host-symbiont interface (Jaffe et al., 2016). There is growing evidence that drug transporters are required for survival of

387 intracellular pathogenic bacteria in host cells (Du et al., 2015). Genes assigned to
388 chromatin structure were also more prevalent in the HT gene pool in Patescibacteria.
389 Among these genes were six SWIB/MDM2 domain containing proteins, which are
390 chromatin-remodelling proteins involved in transcriptional activation (Tang et al.,
391 2010), and were almost exclusively transferred between Patescibacteria (Table S6).
392 DNA packaging in the form of chromatin is well documented in Eukaryotes and
393 Archaea, but remains poorly understood in Bacteria (Shen & Landick, 2019). Whether
394 ultra-small bacteria use chromatin-remodelling proteins to manipulate their own
395 genomic DNA structure, or for the modification of eukaryotic/archaeal symbiotic host
396 DNA is unclear.

397 In contrast, other groundwater taxa more frequently acquired genes involved in
398 energy production (respiratory chain) and amino acid biosynthesis, which
399 Patescibacteria are known to consistently lack (Castelle et al., 2018). This could
400 further indicate that ultra-small bacteria must scavenge resources such as protons and
401 amino acids from the host or surrounding environment. A higher proportion of genes
402 involved in secondary metabolite pathways and signal transduction mechanisms was
403 also detected in other taxa (426 HT and GI genes in 105 'other' MAGs versus 18 genes
404 in 15 ultra-small MAGs) (Table S6). Secondary metabolites are considered luxury
405 metabolism (Demain & Fang, 2000) and can constitute public goods (e.g.
406 siderophores) (Weigert & Kümmerli, 2017). As such, biosynthetic gene clusters
407 encoding for these metabolites are uncommon in streamlined genomes
408 (Konstantinidis & Tiedje, 2004; Sharrar et al., 2020), which have reduced auxiliary
409 functions (Giovannoni et al., 2014), and are more likely to depend on the metabolites
410 produced by other organisms (Morris et al., 2012). Similarly, loss of signal transduction
411 mechanisms, which facilitate environmental monitoring, is a characteristic feature of

412 streamlined bacterial genomes, including those of bacteria with non-intracellular
413 lifestyles (e.g. bacterioplankton) (Gabaldón & Huynen, 2007; Swan et al., 2013).

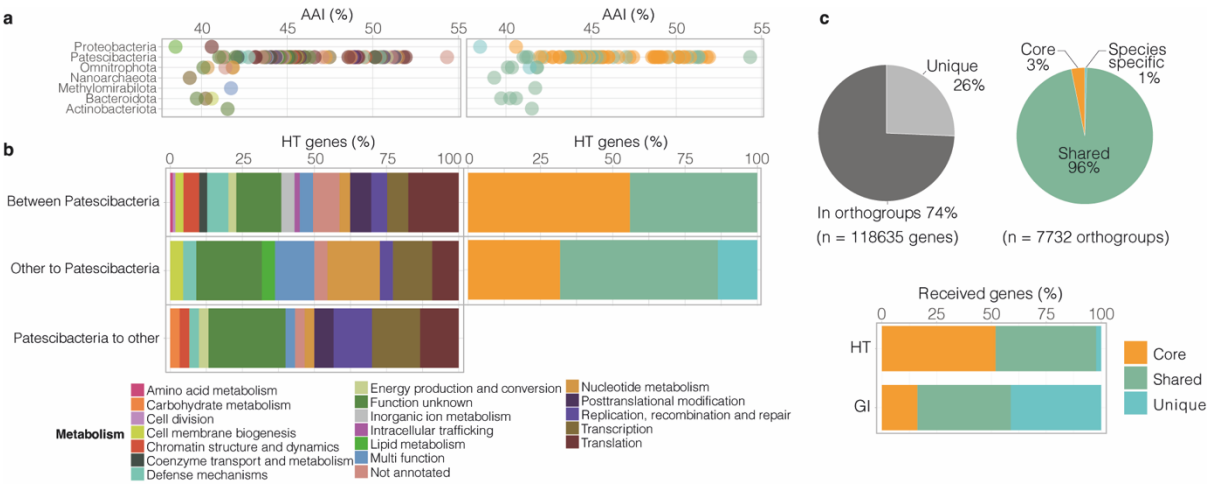
414 Although it has been shown that gene function impacts HGT overall, studies
415 assessing functional categories of horizontally transferred genes have reported
416 diverging results. The 'complexity hypothesis' suggests that genes involved in cellular
417 housekeeping are more frequently subjected to horizontal transfer, while informational
418 genes (involved in transcription, translation, and related processes are more resistant
419 to transfers (Jain et al., 1999). Nakamura et al. (2004) found that horizontally
420 transferred genes are biased towards cell surface, DNA binding and pathogenicity
421 related functions. Conversely, other studies have found that the functional category of
422 a gene family is an insignificant factor in determining HGT (Choi & Kim, 2007; Cohen
423 et al., 2011). While we found that the genes localized in GIs harbour similar metabolic
424 functions overall between Patescibacteria and the rest of the groundwater community,
425 functions of individual genes acquired by the ultra-small bacteria through horizontal
426 transfer tend to differ (Fig. 5). Results obtained here therefore suggest lifestyle of the
427 recipient impacts the traits of genes transferred and also retained.

428 Biases towards horizontally acquired features distinct from the general
429 groundwater communities by Patescibacteria may be explained by constraints on
430 streamlined genomes. Selection of distinct features could occur by two mechanisms:
431 (i) retention of horizontally transferred genes and concentration due to selective
432 deletion of other less desirable features, and (ii) selective retention of acquired genes
433 with desirable features. The caveat being that the distribution of functions among the
434 hypothetical coding DNA sequence (CDS) fraction in Patescibacteria is unknown.

435 We could question whether the length of transferred DNA is an important factor
436 affecting integration and retainment of transferred genes and related metabolic

437 functions in streamlined ultra-small and other groundwater prokaryotes. The length of
438 transferred DNA was 10,324 ± 9,397 bp on average for the 1,046 detected GIs (10.8
439 Mbp total in all groundwater taxa) and 899 ± 562 bp for the 1,501 HT genes (1.3 Mbp
440 overall) (Table S2). Therefore, while a third less GIs than individually transferred genes
441 were detected in terms of number, the total length of sequences transferred was
442 substantially greater with GIs, and broadly equivalent between Patescibacteria (9,689
443 bp per MAG) and other groundwater microorganisms (10,406 bp per MAG). Besides
444 sequence length, it has been shown that recombination efficiency decreases
445 exponentially with sequence divergence (Majewski et al., 2000; Majewski & Cohan,
446 1999). However, the presence of flanking regions of identity in exchangeable regions
447 can remove most barriers to recombination. The shortest length of sequence
448 homology necessary for efficient recombination can vary greatly depending on the
449 organism, the recombination pathway used, and other factors (Thomas & Nielsen,
450 2005). In *Escherichia coli*, efficient recombination has been observed with as little as
451 23 bp of sequence homology (Shen & Huang, 1986).
452
453 **Impact of phylogenetic distance and direction of transfer**
454 We assessed how phylogenetic distance between donor-recipient pairs associates
455 with the metabolic function of genes acquired by Patescibacteria. Genetic exchange
456 occurred over a wide range of phylogenetic distances based on average amino acid
457 identities between genomes of groundwater microorganisms (as low as 38.5% AAI,
458 Fig. 6a). Within the Patescibacteria phylum, genes of diverse metabolic functions were
459 transferred regardless of phylogenetic distance (41.0-54.3% AAI). However, direction
460 of HT events (between members of the Patescibacteria phylum, other groundwater
461 microorganisms to Patescibacteria, and Patescibacteria to other groundwater

462 microorganisms) appeared to have an impact on the metabolic function of acquired
463 genes (Fig. 6b). A number of metabolic functions were found to be exclusively
464 transferred between Patescibacteria (encoded by 12 out of 104 HT genes shared
465 between Patescibacteria), and were involved in processes such as amino acid
466 metabolism (serine hydroxymethyltransferase, GlyA), cell division (actin homolog,
467 MreB), intracellular trafficking, co-enzyme metabolism (riboflavin biosynthesis) and
468 inorganic ion metabolism (cation transport) (Fig. 6b, Table S6).
469



470
471 **Figure 6. Pangenome of Patescibacteria and effect of phylogenetic distance and**
472 **direction of transfer on acquired genes.** (a) Impact of phylogenetic distance of HGT
473 donor to Patescibacteria (calculated as AAI) on metabolic function of horizontally
474 acquired genes (left) and pangenome category (right). (b) Metabolic function of HT
475 genes according to direction of transfer. (c) Proportion of Patescibacteria total gene
476 pool contained in orthogroups vs. unique, and breakdown of proportion of orthogroups
477 in 'core', 'shared' and 'species-specific' pangenome categories (top). Proportion of
478 horizontally acquired genes (HT genes and GI genes) by Patescibacteria taxa in each
479 pangenome category (bottom).
480

481 Others were exclusively shared from Patescibacteria to other phyla (out of n=29
482 transfers overall, Fig. 6b), namely two genes associated with chromatin structure and
483 dynamics (SWIB/MDM2 domain-containing protein) and carbohydrate metabolism
484 (mannose-6-phosphate isomerase, COG0662). Moreover, Patescibacteria donated a
485 greater fraction of replication, recombination and repair genes (GIY-YIG catalytic

486 domain-containing protein, UvrD subfamily helicase, bacterial nucleoid DNA-binding
487 protein COG0776, DNA-3-methyladenine glycosylase) to other bacteria. Among the
488 nine HT events from Patescibacteria to DPANN archaea reported above, three were
489 related to a glyoxalase (Table S6), which is key metalloenzyme in the glycolytic
490 pathway involved in the detoxification of reactive methylglyoxal into D-lactate. It is a
491 common feature across domains of life, including in Archaea (Rawat & Maupin-Furlow,
492 2020), and was previously reported in CPR single amplified genomes (León-Zayas et
493 al., 2017). The reason ultra-small archaea exhibit an apparent preference for acquiring
494 bacterial homologs is unclear.

495 A very small number of HGT events (two) also occurred exclusively from other
496 prokaryotes to Patescibacteria (out of n=20 transfers overall, Fig 6b). This includes
497 the *acpP* gene encoding an acyl carrier protein, and a gene encoding a putative
498 lysophospholipase (COG2267, alpha beta hydrolase superfamily).
499 Lysophospholipases catalyse phospholipid hydrolysis and generate fatty acids. A
500 recent study coupling lipidomics and metagenomics approaches suggested that
501 members of the Patescibacteria phylum, that lack capacities for fatty acid biosynthesis
502 (Castelle et al., 2018), are able to recycle lipids from other bacteria for cell membrane
503 biogenesis (Probst et al., 2020). This suggests that HGT could facilitate the sourcing
504 of fatty acids in Patescibacteria. Phospholipase genes have also been reported in the
505 genomes of other ultra-small CPR prokaryotes (for example, in amoeba endosymbiont
506 *Vampirococcus lugosii*; Moreira et al., 2021). Patescibacteria were also more likely to
507 receive genes from other groundwater microorganisms associated with nucleotide
508 metabolism (n=2), including one thymidylate synthase genes *thyA.* The enzyme
509 encoded catalyzes the *de novo* synthesis of dTMP (deoxythymidine monophosphate),
510 an essential precursor for DNA biosynthesis (Voeller et al., 1995), which is a capacity

511 already encoded by the minimal biosynthetic gene repertoire of groundwater
512 Patescibacteria. Our results indicate that this essential function can also be
513 horizontally acquired from other members of groundwater microbial communities.
514 While the number of transfer events recorded from general groundwater
515 microorganisms to Patescibacteria and vice versa is low, limiting robust conclusions,
516 these numbers are consistent with known associations between HGT and
517 phylogenetic distance among prokaryotes more generally (Majewski et al., 2000).
518 Overall, metabolic functions gained via HGT in ultra-small prokaryotes remain poorly
519 studied, regardless of phylogenetic origin. Events identified here tend to confer
520 Patescibacteria with critical functions, such as the ability to degrade lipids, and carry
521 out the biosynthesis of nucleotides.
522
523 **Patescibacteria pangenome and HGT**
524 The evolutionary fate of a HGT event is determined by fitness conferred to the recipient
525 genome and cell, i.e., whether the acquisition of exogenous DNA is beneficial, neutral
526 or deleterious. In order to evaluate the importance of the metabolic functions acquired
527 through HGT in Patescibacteria, we performed a pangenome analysis of the 125
528 reconstructed MAGs. The pangenome of all the examined Patescibacteria contained
529 7,732 orthogroups, comprising 88,187 genes (74.3%) out of the total 118,635
530 predicted genes. Genes in the pan-genome were classified into 3 groups based on
531 their occurrence: (1) the “core” genes, which are present in at least 2/3 of all 132
532 genomes, (2) the “shared” genes are found in more than one genome, but are not core
533 genes, and (3) the “unique” genes, comprising species-specific genes (i.e. genes
534 present in one genome only) and genes not assigned to an orthogroup. Both shared
535 and unique genes comprise the accessory genome. Results indicated the vast majority

536 of orthogroups (96%) were shared between the groundwater Patescibacteria, and only
537 3% of orthogroups (2,634 genes) comprised the core genome (Fig. 6c). This
538 represents the lower end of core genome size among bacterial pangenomes (Bosi et
539 al., 2017; Shin et al., 2016) and could be partially due to MAG incompleteness, but is
540 consistent with the variation found in the core protein families by Méheust and
541 colleagues across members of the CPR radiation (Méheust et al., 2019).

542 GIs brought many unique, and hence divergent, genes to the Patescibacteria
543 pangenome (41.2% of GI genes, Fig. 6c). These observations are consistent with
544 previous work suggesting that prokaryotes contain a higher proportion of novel genes
545 in GIs compared with the rest of their genome (Hsiao et al., 2005). HT genes acquired
546 by Patescibacteria species were spread across the core and the shared gene fractions
547 of these organisms (45.8% shared and 51.9% core, Fig. 6c), and not the unique
548 fraction. A high proportion of core or shared genes would be expected given predicted
549 transfers occurred mostly between Patescibacteria species (83.2% of HT events, Fig.
550 4a). However, the lack of any unique genes reflects limitations in the detection of HT
551 events by the MetaCHIP tool, which considers self matches to be false positive (i.e. a
552 donor is required) (Song et al., 2019).

553 Streamlined genomes are characterised by low numbers of paralogs
554 (Giovannoni et al., 2014). Accordingly, among all genes in Patescibacteria, we
555 detected a low number of paralogous orthogroups based on the pangenome analysis
556 results (89 genes in 36 species-specific orthogroups – i.e. 89 paralogs out of 118,635
557 genes). We found that no genes located on GI regions nor HT genes were paralogs
558 of other genes in the Patescibacteria pangenome. Gene duplicates are generally
559 under purifying selection (Lynch & Conery, 2000), and have been shown to evolve
560 faster than non-duplicated genes with a similar level of divergence (Kondrashov et al.,

561 2002), introducing new proteins. Nonetheless, gene duplication is usually widespread
562 in bacterial genomes, with an estimated 7–41% of bacterial proteins encoded by
563 paralogs (Gevers et al., 2004), and is a common strategy for creating novel gene
564 functions for niche expansion in prokaryotes (Hooper & Berg, 2003). However, it has
565 been quantitatively demonstrated that gene gain via HGT rather than gene duplication
566 is the principal factor of innovation in the evolution of prokaryotes (Treangen & Rocha,
567 2011). Based on a comparably high proportion of HGT into Patescibacteria as for other
568 phyla, results here demonstrate HGT, and unique genes conferred by GIs, are almost
569 exclusively used over duplication by Patescibacteria to expand niche ranges.
570

571 **GI in Patescibacteria nzgw456 encoding a full type I restriction-modification**
572 **system**

573 Streamlined bacteria (and archaea) have been hypothesized to be largely lacking
574 classical defense mechanisms, such as CRISPR-Cas (Clustered Regularly
575 Interspaced Short Palindromic Repeats – CRISPR associated), restriction-
576 modification (R-M) and toxin-antitoxin systems (Koonin et al., 2017). Similarly, the
577 analysis of hundreds of CPR/Patescibacteria genomes recovered from diverse
578 environments revealed that most are missing CRISPR-Cas adaptive immune systems,
579 lost in the process of genome streamlining, and instead rely nonspecific defence, such
580 as R-M systems (Castelle et al., 2018). Although first believed to constitute a barrier
581 to genetic exchange between organisms (Vasu & Nagaraja, 2013), it has been recently
582 suggested that these systems play an important role in HGT rates by increasing the
583 frequency of exchange in organisms possessing numerous R-M systems (Oliveira et
584 al., 2016).
585

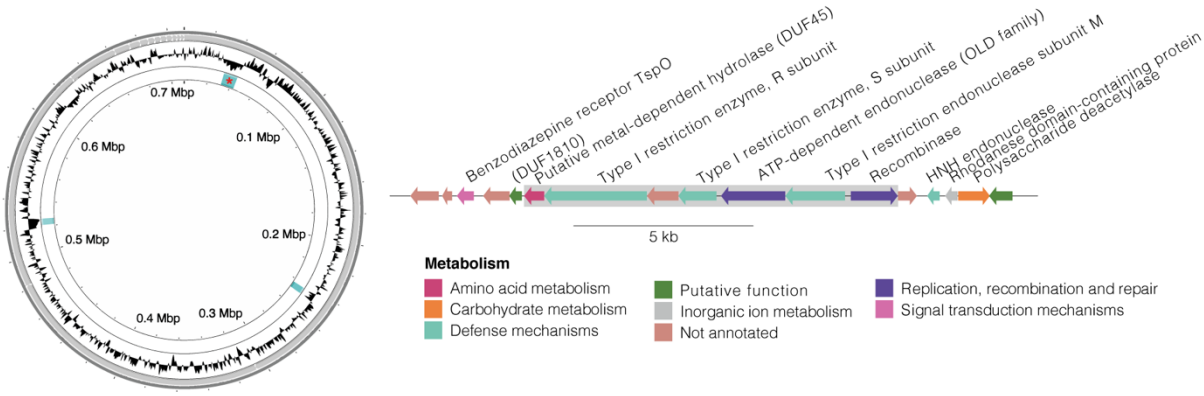



**Figure 7. Genomic island encoding a full type I restriction-modification system in Patescibacteria nzgw456.** Left: circular genome plot with following rings (inner to outer): GIs (in blue; star indicates location of GI depicted in the right panel); GC content; contig boundaries. Right: Linear map of GI (indicated by grey box) and protein-coding genes. Manual examination of GI was undertaken to further confirm boundaries using the DarkHorse prediction tool (Podell et al., 2008).

Out of the 120 GIs detected in Patescibacteria we identified four encoding restriction modification systems or endonucleases in patescibacterial genomes. All four MAGs contained a high number of GIs compared to other patescibacterial counterparts (average 5 ± 2.8 SD vs 1.7 ± 1.0), and three acquired more HT per Mbp than the average for Patescibacteria overall (average 2.5 ± 1.3 SD vs 1.2 ± 1.5, Table S2). One GI in Patescibacteria nzgw456 (Paceibacteria class) encodes a full type I R-M system (Fig. 7). Type I R-M systems, which are the most complex of all four known R-M systems (Murray, 2000), comprise three subunits, namely HsdR and HsdM (required for methyltransferase activity), and HsdS (required for restriction). Co-located with genes encoding these subunits on the GI was one putative ATP-dependent endonuclease. Type I R-M related genes were also found in a GI present in Patescibacteria nzgw350, and other restriction endonucleases in nzgw410 and nzgw504 (Table S7). Acquisition of type I R-M genes via HGT was recently described in CPR bacterium *V. lugosii* (Moreira et al., 2021), however not as syntenic genes. It has been suggested that R-M systems could be used by putative epibionts as a way

609 to source of nucleotides by degrading exogenous DNA, including that derived from
610 viruses (Burstein et al., 2016), in addition to protecting against phage infection (Stern
611 & Sorek, 2011).
612

613 **CONCLUSION**

614 Recent phylogenetic analyses have shed light on the massive bacterial group that is
615 the Patescibacteria phylum/CPR (Hug et al., 2016). Due to the reductive evolution of
616 Patescibacteria, it is essential to consider the processes of gene gain and loss in
617 respect of the predicted symbiotic lifestyles of these ultra-small microorganisms.
618 Results here indicate that Patescibacteria are involved in dynamic genetic exchange
619 with their ultra-small counterparts and other members of the microbial community. We
620 found a total 124 HT events involving Patescibacteria, 24 prophage and over 100 GIs
621 in Patescibacteria genomes. While most transfers occurred among Patescibacteria,
622 HGT origins were diverse and widespread across both bacterial and archaeal
623 domains, in particular with members of the Omnitrophota and Bacteroidota phyla. A
624 putative host symbiont pair was identified based on HGT evidence and co-occurrence
625 across groundwater samples. Functions encoded by horizontally acquired regions
626 comprised diverse pathways, but mainly transcription, translation and DNA replication,
627 recombination and repair, such as ribosomal proteins and transcription factors.
628 Findings show that acquired genes were distributed across core and accessory
629 portions of Patescibacteria genomes, and that GIs generally introduced
630 phylogenetically distant sequences. Some of the metabolic functions horizontally
631 acquired by Patescibacteria are distinct from those acquired by the general
632 groundwater communities, suggesting unique pressures on gene gain/loss dynamics
633 occurring in ultra-small prokaryotes.

## ACKNOWLEDGMENTS

Funding was provided by a MBIE Smart Ideas grant awarded to K.M.H. (project UOAX1720). We thank D. Waite and J.S. Boey for help with bioinformatics. Computational resources were provided by New Zealand eScience Infrastructure.

The authors declare no conflict of interest.

## REFERENCES

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., Thomas, B. C., Singh, A., Wilkins, M. J., Karaoz, U., Brodie, E. L., Williams, K. H., Hubbard, S. S., & Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications*, *7*, 1–11. https://doi.org/10.1038/ncomms13219

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V, Sirotkin, A. V, Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., & Brinkman, F. S. L. (2017). IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research*, *45*(W1). https://doi.org/10.1093/nar/gkx343

Bonham, K. S., Wolfe, B. E., & Dutton, R. J. (2017). Extensive horizontal gene transfer in cheese-associated bacteria. *ELife*, *6*.

https://doi.org/10.7554/eLife.22144

Bosi, E., Fondi, M., Orlandini, V., Perrin, E., Maida, I., de Pascale, D., Tutino, M. L., Parrilli, E., Lo Giudice, A., Filloux, A., & Fani, R. (2017). The pangenome of (Antarctic) Pseudoalteromonas bacteria: Evolutionary and functional insights. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-016-3382-y

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., Wilkins, M. J., Wrighton, K. C., Williams, K. H., & Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*, *523*(7559), 208–211. https://doi.org/10.1038/nature14486

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., Thomas, B. C., & Banfield, J. F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms10613

Castelle, C. J., Brown, C. T., Anantharaman, K., Probst, A. J., Huang, R. H., & Banfield, J. F. (2018). Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nature Reviews Microbiology*, *16*(10), 629–645. https://doi.org/10.1038/s41579-018-0076-2

Castelle, C. J., Brown, C. T., Thomas, B. C., Williams, K. H., & Banfield, J. F. (2017). Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. *Scientific Reports*, *7*(October 2016), 1–12. https://doi.org/10.1038/srep40101

Chaudhari, N. M., Overholt, W. A., Figueroa-Gonzalez, P. A., Taubert, M., Bornemann, T. L. V., Probst, A. J., Hölzer, M., Marz, M., & Küsel, K. (2021). The economical lifestyle of CPR bacteria in groundwater allows little preference for environmental drivers. *BioRxiv*, 2021.07.28.454184.

https://www.biorxiv.org/content/10.1101/2021.07.28.454184v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.07.28.454184v1.abstract

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Choi, I. G., & Kim, S. H. (2007). Global extent of horizontal gene transfer. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(11), 4489–4494. https://doi.org/10.1073/pnas.0611557104

Close, M., Abraham, P., Webber, J., Cowey, E., Humphries, B., Fenwick, G., Howard, S., Huynh, K., Grace, T., Dupont, P. Y., & Weaver, L. (2020). Use of Sonication for Enhanced Sampling of Attached Microbes from Groundwater Systems. *Groundwater*, *58*, 901–915. https://doi.org/10.1111/gwat.12984

Cohen, O., Gophna, U., & Pupko, T. (2011). The complexity hypothesis revisited: Connectivity Rather Than function constitutes a barrier to horizontal gene transfer. *Molecular Biology and Evolution*, *28*(4). https://doi.org/10.1093/molbev/msq333

Cordero, O. X., & Hogeweg, P. (2009). The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21748–21753. https://doi.org/10.1073/pnas.0907584106

Demain, A. L., & Fang, A. (2000). The natural functions of secondary metabolites. In *Advances in biochemical engineering/biotechnology* (Vol. 69). https://doi.org/10.1007/3-540-44964-7_1

Dong, X., Zhang, C., Li, W., Weng, S., Song, W., Li, J., & Wang, Y. (2021). Functional diversity of microbial communities in inactive seafloor sulfide

deposits. *FEMS Microbiology Ecology*, *97*(8), 1–10. https://doi.org/10.1093/femsec/fiab108

Du, D., van Veen, H. W., Murakami, S., Pos, K. M., & Luisi, B. F. (2015). Structure, mechanism and cooperation of bacterial multidrug transporters. In *Current Opinion in Structural Biology* (Vol. 33). https://doi.org/10.1016/j.sbi.2015.07.015

Dunning Hotopp, J. C., Clark, M. E., Oliveira, D. C. S. G., Foster, J. M., Fischer, P., Muñoz Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards, S., Spiro, D. J., Ghedin, E., Slatko, B. E., Tettelin, H., & Werren, J. H. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science*, *317*(5845). https://doi.org/10.1126/science.1142490

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1832-y

Flynn, T. M., Sanford, R. A., & Bethke, C. M. (2008). Attached and suspended microbial communities in a pristine confined aquifer. *Water Resources Research*, *44*(7), 1–7. https://doi.org/10.1029/2007WR006633

Gabaldón, T., & Huynen, M. A. (2007). From endosymbiont to host-controlled organelle: The hijacking of mitochondrial protein synthesis and metabolism. *PLoS Computational Biology*, *3*(11). https://doi.org/10.1371/journal.pcbi.0030219

Garcia-Vallvé, S., Romeu, A., & Palau, J. (2000). Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Research*, *10*(11).

https://doi.org/10.1101/gr.130000

Geesink, P., Wegner, C.-E., Probst, A. J., Herrmann, M., Dam, H. T., Kaster, A.-K., & Küsel, K. (2020). Genome-inferred spatio-temporal resolution of an uncultivated Roizmanbacterium reveals its ecological preferences in groundwater. *Environmental Microbiology*, *22*(2), 726–737. https://doi.org/https://doi.org/10.1111/1462-2920.14865

Gevers, D., Vandepoele, K., Simillion, C., & Van De Peer, Y. (2004). Gene duplication and biased functional retention of paralogs in bacterial genomes. In *Trends in Microbiology* (Vol. 12, Issue 4). https://doi.org/10.1016/j.tim.2004.02.007

Giovannoni, S. J., Cameron Thrash, J., & Temperton, B. (2014). Implications of streamlining theory for microbial ecology. In *ISME Journal* (Vol. 8, Issue 8). https://doi.org/10.1038/ismej.2014.60

Gogarten, J. P., Doolittle, W. F., & Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution*, *19*(12). https://doi.org/10.1093/oxfordjournals.molbev.a004046

Griebler, C., & Lueders, T. (2009). Microbial biodiversity in groundwater ecosystems. *Freshwater Biology*, *54*(4), 649–677. https://doi.org/10.1111/j.1365-2427.2008.02013.x

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*(19). https://doi.org/10.1093/bioinformatics/btu393

Haft, D. H., Selengut, J. D., & White, O. (2003). The TIGRFAMs database of protein families. In *Nucleic Acids Research* (Vol. 31, Issue 1). https://doi.org/10.1093/nar/gkg128

He, C., Keren, R., Whittaker, M., Farag, I. F., Doudna, J., Cate, J. H. D., & Banfield, J. (2021). Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nature Microbiology*, *6*, 354–365. https://doi.org/10.1101/2020.05.14.094862

Herrmann, M., Wegner, C.-E., Taubert, M., Geesink, P., Lehmann, K., Yan, L., Lehmann, R., Totsche, K. U., & Küsel, K. (2019). Predominance of Cand. Patescibacteria in Groundwater Is Caused by Their Preferential Mobilization From Soils and Flourishing Under Oligotrophic Conditions. *Frontiers in Microbiology*, *10*(June), 1–15. https://doi.org/10.3389/fmicb.2019.01407

Hooper, S. D., & Berg, O. G. (2003). Duplication is more common among laterally transferred genes than among indigenous genes. *Genome Biology*, *4*(8). https://doi.org/10.1186/gb-2003-4-8-r48

Hsiao, W. W. L., Ung, K., Aeschliman, D., Bryan, J., Brett Finlay, B., & Brinkman, F. S. L. (2005). Evidence of a large novel gene pool associated with prokaryotic genomic Islands. *PLoS Genetics*, *1*(5). https://doi.org/10.1371 /journal.pgen.0010062

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314.

https://doi.org/10.1093/nar/gky1085

Hug, L. A., Thomas, B. C., Brown, C. T., Frischkorn, K. R., Williams, K. H., Tringe, S. G., & Banfield, J. F. (2015). Aquifer environment selects for microbial species cohorts in sediment and groundwater. *ISME Journal*, *9*(8), 1846–1856. https://doi.org/10.1038/ismej.2015.2

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*, 119. https://doi.org/10.1186/1471-2105-11-119

Jaffe, A. L., Corel, E., Pathmanathan, J. S., Lopez, P., & Bapteste, E. (2016). Bipartite graph analyses reveal interdomain LGT involving ultrasmall prokaryotes and their divergent, membrane-related proteins. *Environmental Microbiology*, *18*(12), 5072–5081. https://doi.org/10.1111/1462-2920.13477

Jain, R., Rivera, M. C., & Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(7). https://doi.org/10.1073/pnas.96.7.3801

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070

Kanhere, A., & Vingron, M. (2009). Horizontal gene transfers in prokaryotes show differential preferences for metabolic and translational genes. *BMC Evolutionary*

809 *Biology*, *9*(1), 1–13. https://doi.org/10.1186/1471-2148-9-9
810 Kieft, K., Zhou, Z., & Anantharaman, K. (2020). VIBRANT: automated recovery,
811 annotation and curation of microbial viruses, and evaluation of viral community
812 function from genomic sequences. *Microbiome*, *8*(1), 90.
813 https://doi.org/10.1186/s40168-020-00867-0
814 Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Selection in the
815 evolution of gene duplications. *Genome Biology*, *3*(2).
816 https://doi.org/10.1186/gb-2002-3-2-research0008
817 Konstantinidis, K. T., & Tiedje, J. M. (2004). Trends between gene content and
818 genome size in prokaryotic species with larger genomes. *Proceedings of the*
819 *National Academy of Sciences*, *101*(9), 3160–3165.
820 https://doi.org/10.1073/pnas.0308653100
821 Koonin, E. V., Makarova, K. S., & Wolf, Y. I. (2017). Evolutionary Genomics of
822 Defense Systems in Archaea and Bacteria. *Annual Review of Microbiology*, *71*.
823 https://doi.org/10.1146/annurev-micro-090816-093830
824 Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and
825 memory-efficient alignment of short DNA sequences to the human genome.
826 *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25
827 Lawrence, J. G., & Ochman, H. (1997). Amelioration of bacterial genomes: Rates of
828 change and exchange. *Journal of Molecular Evolution*, *44*(4).
829 https://doi.org/10.1007/PL00006158
830 León-Zayas, R., Peoples, L., Biddle, J. F., Podell, S., Novotny, M., Cameron, J.,
831 Lasken, R. S., & Bartlett, D. H. (2017). The metabolic potential of the single cell
832 genomes obtained from the Challenger Deep, Mariana Trench within the
833 candidate superphylum Parcubacteria (OD1). *Environmental Microbiology*,

*19*(7), 2769–2784. https://doi.org/10.1111/1462-2920.13789

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)*, *23*(1), 127–128. https://doi.org/10.1093/bioinformatics/btl529

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, *290*(5494). https://doi.org/10.1126/science.290.5494.1151

Maddocks, S. E., & Oyston, P. C. F. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology*, *154*(12), 3609–3623. https://doi.org/10.1099/mic.0.2008/022772-0

Majewski, J., & Cohan, F. M. (1999). DNA sequence similarity requirements for interspecific recombination in Bacillus. *Genetics*, *153*(4). https://doi.org/10.1093/genetics/153.4.1525

Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F. M., & Dowson, C. G. (2000). Barriers to genetic exchange between bacterial species: Streptococcus pneumoniae transformation. *Journal of Bacteriology*, *182*(4). https://doi.org/10.1128/JB.182.4.1016-1023.2000

Mandel, M. J., Wollenberg, M. S., Stabb, E. V., Visick, K. L., & Ruby, E. G. (2009). A single regulatory gene is sufficient to alter bacterial host range. *Nature*, *458*(7235). https://doi.org/10.1038/nature07660

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal; Vol 17, No 1: Next Generation Sequencing Data Analysis*. https://doi.org/10.14806/ej.17.1.200

Méheust, R., Burstein, D., Castelle, C. J., & Banfield, J. F. (2019). The distinction of CPR bacteria from other bacteria based on protein family content. *Nature*

*Communications*, *10*(1), 4173. https://doi.org/10.1038/s41467-019-12171-z

Moran, N. A., & Bennett, G. M. (2014). The Tiniest Tiny Genomes. *Annual Review of Microbiology*, *68*(1), 195–215. https://doi.org/10.1146/annurev-micro-091213-112901

Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The black queen hypothesis: Evolution of dependencies through adaptive gene loss. *MBio*, *3*(2), 1–7. https://doi.org/10.1128/mBio.00036-12

Mosley, O. E., Gios, E., Weaver, L., Close, M., Daughney, C., van der Raaij, R., Martindale, H., & Handley, K. M. (2022). Metabolic Diversity and Aero-Tolerance in Anammox Bacteria from Geochemically Distinct Aquifers. *MSystems*, *7*(1), 2021.09.16.460709. https://doi.org/10.1128/msystems.01255-21

Murray, N. E. (2000). Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews*, *64*(2). https://doi.org/10.1128/mmbr.64.2.412-434.2000

Nakamura, Y., Itoh, T., Matsuda, H., & Gojobori, T. (2004). Biased biological functions of horizontally-transferred genes in prokaryotic genomes. *Nature Genetics*, *36*(7), 760–766. https://doi.org/10.1038/ng1381

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, *39*(5). https://doi.org/10.1038/s41587-020-00774-7

Nelson, W. C., & Stegen, J. C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Frontiers in Microbiology*, *6*(JUN), 1–14. https://doi.org/10.3389/fmicb.2015.00713

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A

884 Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood
885 Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.
886 https://doi.org/10.1093/molbev/msu300
887 Nguyen, M. T. H. D., Liu, M., & Thomas, T. (2014). Ankyrin-repeat proteins from
888 sponge symbionts modulate amoebal phagocytosis. *Molecular Ecology*, *23*(6).
889 https://doi.org/10.1111/mec.12384
890 Nicolas, A. M., Jaffe, A. L., Nuccio, E. E., Taga, M. E., Firestone, M. K., & Banfield,
891 J. F. (2021). Soil Candidate Phyla Radiation Bacteria Encode Components of
892 Aerobic Metabolism and Co-occur with Nanoarchaea in the Rare Biosphere of
893 Rhizosphere Grassland Communities. *MSystems*, *6*(4).
894 https://doi.org/10.1128/msystems.01205-20
895 Ochman, Howard, Lawrence, Jeffrey G., & Groisman, Eduardo A. (2000). Lateral
896 gene transfer and the nature of bacterial innovation. *Nature*, *405*(6784), 299–
897 304. http://adsabs.harvard.edu/abs/2000Natur.405..299O
898 Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2016). Regulation of genetic flux
899 between bacteria by restriction-modification systems. *Proceedings of the*
900 *National Academy of Sciences of the United States of America*, *113*(20).
901 https://doi.org/10.1073/pnas.1603257113
902 Parks, D. H., Chuvochina, M., Chaumeil, P. A., Rinke, C., Mussig, A. J., &
903 Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria
904 and Archaea. *Nature Biotechnology*, *38*(September).
905 https://doi.org/10.1038/s41587-020-0501-8
906 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015).
907 CheckM: assessing the quality of microbial genomes recovered from isolates,
908 single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055.

https://doi.org/10.1101/gr.186072.114

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, *2*(11). https://doi.org/10.1038/s41564-017-0012-7

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, *46*(W1). https://doi.org/10.1093/nar/gky448

Probst, A. J., Elling, F. J., Castelle, C. J., Zhu, Q., Elvert, M., Birarda, G., Holman, H. Y. N., Lane, K. R., Ladd, B., Ryan, M. C., Woyke, T., Hinrichs, K. U., & Banfield, J. F. (2020). Lipid analysis of CO2-rich subsurface aquifers suggests an autotrophy-based deep biosphere with lysolipids enriched in CPR bacteria. *ISME Journal*, *14*(6). https://doi.org/10.1038/s41396-020-0624-4

Probst, A. J., Ladd, B., Jarett, J. K., Geller-McGrath, D. E., Sieber, C. M. K., Emerson, J. B., Anantharaman, K., Thomas, B. C., Malmstrom, R. R., Stieglmeier, M., Klingl, A., Woyke, T., Ryan, M. C., & Banfield, J. F. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature Microbiology*, *5*, 328–336. https://doi.org/10.1038/s41564-017-0098-y

Rawat, M., & Maupin-Furlow, J. A. (2020). Redox and thiols in archaea. In *Antioxidants* (Vol. 9, Issue 5). https://doi.org/10.3390/antiox9050381

Schulz, F., Eloe-Fadrosh, E. A., Bowers, R. M., Jarett, J., Nielsen, T., Ivanova, N. N., Kyrpides, N. C., & Woyke, T. (2017). Towards a balanced view of the bacterial tree of life. *Microbiome*, *5*(1). https://doi.org/10.1186/s40168-017-0360-9

Sharrar, A. M., Crits-Christoph, A., Méheust, R., Diamond, S., Starr, E. P., &

934 Banfield, J. F. (2020). Bacterial secondary metabolite biosynthetic potential in
935 soil varies with phylum, depth, and vegetation type. *MBio*, *11*(3).
936 https://doi.org/10.1128/MBIO.00416-20
937 Shen, B. A., & Landick, R. (2019). Transcription of Bacterial Chromatin. In *Journal of*
938 *Molecular Biology* (Vol. 431, Issue 20). https://doi.org/10.1016/j.jmb.2019.05.041
939 Shen, P., & Huang, H. V. (1986). Homologous recombination in Escherichia coli:
940 dependence on substrate length and homology. *Genetics*, *112*(3).
941 https://doi.org/10.1093/genetics/112.3.441
942 Shin, J., Song, Y., Jeong, Y., & Cho, B. K. (2016). Analysis of the core genome and
943 pan-genome of autotrophic acetogenic bacteria. In *Frontiers in Microbiology*
944 (Vol. 7, Issue SEP). https://doi.org/10.3389/fmicb.2016.01531
945 Song, W., Wemheuer, B., Zhang, S., Steensen, K., & Thomas, T. (2019). MetaCHIP:
946 Community-level horizontal gene transfer identification through the combination
947 of best-match and phylogenetic approaches. *Microbiome*, *7*(1).
948 https://doi.org/10.1186/s40168-019-0649-y
949 Sonnhammer, E. L. L., Eddy, S. R., Birney, E., Bateman, A., & Durbin, R. (1998).
950 Pfam: Multiple sequence alignments and HMM-profiles of protein domains.
951 *Nucleic Acids Research*, *26*(1). https://doi.org/10.1093/nar/26.1.320
952 Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building
953 the web of life. In *Nature Reviews Genetics* (Vol. 16, Issue 8).
954 https://doi.org/10.1038/nrg3962
955 Stern, A., & Sorek, R. (2011). The phage-host arms race: Shaping the evolution of
956 microbes. In *BioEssays* (Vol. 33, Issue 1).
957 https://doi.org/10.1002/bies.201000071
958 Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H., & Consortium, U.

(2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics (Oxford, England)*, *31*(6), 926–932. https://doi.org/10.1093/bioinformatics/btu739

Swan, B. K., Tupper, B., Sczyrba, A., Lauro, F. M., Martinez-Garcia, M., Gonźalez, J. M., Luo, H., Wright, J. J., Landry, Z. C., Hanson, N. W., Thompson, B. P., Poulton, N. J., Schwientek, P., Acinas, S. G., Giovannoni, S. J., Moran, M. A., Hallam, S. J., Cavicchioli, R., Woyke, T., & Stepanauskas, R. (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28). https://doi.org/10.1073/pnas.1304246110

Tang, L., Nogales, E., & Ciferri, C. (2010). Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. In *Progress in Biophysics and Molecular Biology* (Vol. 102, Issues 2–3). https://doi.org/10.1016/j.pbiomolbio.2010.05.001

Tee, H. S., Waite, D., Payne, L., Middleditch, M., Wood, S., & Handley, K. M. (2020). Tools for successful proliferation: diverse strategies of nutrient acquisition by a benthic cyanobacterium. *ISME Journal*, *14*(8), 2164–2178. https://doi.org/10.1038/s41396-020-0676-5

The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/nar/gkw1099

Touchon, M., Bernheim, A., & Rocha, E. P. C. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME Journal*, *10*(11). https://doi.org/10.1038/ismej.2016.47

Treangen, T. J., & Rocha, E. P. C. (2011). Horizontal transfer, not duplication, drives

984 the expansion of protein families in prokaryotes. *PLoS Genetics*, *7*(1).
985 https://doi.org/10.1371/journal.pgen.1001284
986 Vasu, K., & Nagaraja, V. (2013). Diverse Functions of Restriction-Modification
987 Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology*
988 *Reviews*, *77*(1). https://doi.org/10.1128/mmbr.00044-12
989 Voeller, D. M., Changchien, L. ming, Maley, G. F., Maley, F., Takechi, T., Turner, R.
990 E., Montfort, W. R., Allegra, C. J., & Chu, E. (1995). Characterization of a
991 specific interaction between Escherichia coli thymidylate synthase and
992 escherichia coli thymidylate synthase mRNA. *Nucleic Acids Research*, *23*(5).
993 https://doi.org/10.1093/nar/23.5.869
994 Weigert, M., & Kümmerli, R. (2017). The physical boundaries of public goods
995 cooperation between surface-attached bacterial cells. *Proceedings of the Royal*
996 *Society B: Biological Sciences*, *284*(1858).
997 https://doi.org/10.1098/rspb.2017.0631
998