Prokaryotic functional gene diversity in the sunlit ocean: stumbling in the dark

Isabel Ferrera*[†], Marta Sebastian[†], Silvia G. Acinas, Josep M. Gasol

Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar, CSIC, Barcelona, Catalunya, Spain

[†]Equal contributors to this article

*Corresponding author:
Dpt. Biologia Marina i Oceanografia
Institut de Ciències del Mar, CSIC
Pg Marítim de la Barceloneta 37-49
E-08003 Barcelona, Catalunya, Spain
Phone: +34 93 230 95 00
Fax: + 34 93 230 95 55
email: iferrera@icm.csic.es

27   **Abstract**

28   Prokaryotes are extremely abundant in the ocean where they drive

29   biogeochemical cycles. The recent development and application of –*omics*

30   techniques has provided an astonishing amount of information revealing the

31   existence of a vast diversity of functional genes and a large heterogeneity within

32   each gene. The big challenge for microbial ecologists is now to understand the

33   ecological relevance of this variability for ecosystem functioning, a question that

34   remains largely understudied. This brief review highlights some of the latest

35   advances in the study of the diversity of biogeochemically relevant functional

36   genes in the sunlit ocean.

37   **Introduction**

38   The ocean is the largest ecosystem on Earth. Prokaryotes are the most abundant

39   cells in the ocean (ca. ~10^{29}, [1]) where they are the engines driving nutrient

40   cycles and energy flow. In the last two decades, molecular studies based on

41   rRNA genes have unveiled that the phylogenetic diversity of marine

42   microorganisms is immensely larger than the few thousand species formally

43   described [2]. After the initial surveys of ribosomal genes, the study of protein-

44   coding genes followed by the development of *meta'omics* approaches has

45   opened a window to the exploration of marine microbial diversity at the functional

46   level. In fact, the first large-scale ocean metagenomic survey identified novel

47   protein families and added a tremendous diversity to known protein families [3].

48   Hence, the study of functional diversity has improved our view of the roles that

49   marine microbes play in global biogeochemical cycles.

50   Some clear examples are the discovery of novel metabolisms, such as

51   proteorhodopsin-driven photoheterotrophy [4], or the role that urea plays in

52   nitrification by marine archaea [5]. These studies have shown that prokaryotes

53   play more functions than previously thought and that there is large sequence

54   diversity for each gene [6-9]. But, what are the ecological implications of this

55   variability? Addressing this important question is a major challenge in marine

56   microbial ecology.

57   The tremendous functional diversity of marine prokaryotes occurs at different

58   levels of complexity, such as variability in gene content, differences in operon

59   structure, heterogeneity in gene sequence, or existence of different proteins

60   which drive redundant functions. Furthermore, the presence of functional

61   metabolic genes in viral genomes, such as those reported to be involved in

62   photosynthesis, carbon or phosphorus metabolism, have added another

63   dimension to this complexity [10]. Most of the studies conducted so far have

64   focused on genes that confer an ecological advantage to prokaryotic populations,

65   such as nitrogen fixation or photoheterotrophy. In this review, we summarize the

66   current knowledge in prokaryotic (mostly bacterial) functional gene diversity in

67   relation to the four major marine biogeochemical cycles (carbon, phosphorus,

68   sulfur and nitrogen) (Figure 1). We focus on the sunlit ocean where most of the

69   studies have been conducted, and discuss the potential relevance of genetic

70   diversity for ecosystem functioning.

Biogeochemically relevant genes in the sunlit ocean

Carbon cycle. The paradigm of the ocean carbon flow depicting phototrophs as the only autochtonous producers of organic carbon (C) and heterotrophic bacteria as the consumers fell apart in the last decade. Nowadays we know that mixotrophy, the ability of an organism to use both light and organic matter to obtain energy, seems to be the rule rather than the exception in sunlit marine microbial communities. Since the discovery of a proteorhodopsin (PR) gene in a metagenomic fragment almost 15 years ago [4], we now know that up to 80% of bacteria inhabiting surface waters can harbor this gene [11]. Proteorhodopsins display large genetic diversity [12] that translates for example in different spectral tuning, likely as an adaptation to absorb the prevalent light wavelengths (blue or green) found in different marine waters. While this proton pump has been demonstrated to enhance growth under light in certain marine Bacteroidetes [13], no direct effect on growth was observed in *Pelagibacter,* in which, however, PR increased long-time survival during starvation [14], as was also observed in a *Vibrio* isolate [15]. Recently, new types of evolutionarily distinct rhodopsins have been uncovered which translocate either Na^+ or Cl^- [16, 17]. Interestingly, multiple rhodopsin types, each with different ion specificities, can be found within a single strain of marine Bacteroidetes [17]. However, the ecological implications of containing as many as three functionally different rhodopsins are unknown. Rhodopsin is indeed a neat example in which gene diversity likely results in diverse physiological and ecological functions. Yet, we appear to know only a small fraction of the functional significance of the diversity of this gene.

94   Similarly, aerobic anoxygenic phototrophs (AAPs) can derive a portion of their

95   energy requirements harvesting light using bacteriochlorophyll *a* (BChl *a*). Marine

96   AAPs contain diverse *pufM* gene sequences (the phylogenetic marker for AAPs)

97   as well as *puf* operon structures that delimit several phylogroups distributed

98   across the *Alpha-* and *Gammaproteobacteria*. Recent experimental work with

99   isolates has shown that while light is the main factor controlling the regulation of

100   Bchl *a* in representatives of the *Alphaproteobacteria,* this is not the case in the

101   *Gammaproteobacteria* [18]. Likewise, evidences that distinct phylogroups have

102   preferences to live under different environmental conditions [8,19,20] suggest

103   that different AAPs posses diverse strategies to adapt to a changing environment

104   and may have different roles in the functioning of the ecosystem. However, a

105   direct linkage between *puf* diversity and its relevance in ecosystem function has

106   not yet been established.

107   CO lithoheterotrophy has also received attention as a potential form of obtaining

108   energy for bacterial metabolism [21]. CO is formed through the photochemical

109   degradation of organic matter in sunlit waters and can be oxidized by means of

110   the carbon monoxide dehydrogenase encoded by the *cox* genes [22]. These

111   genes are structured in operons that form two phylogenetically distinct groups, in

112   which the gene order also diverges. Some bacterial species contain one operon

113   from each phylogenetic group, whereas others can contain multiple operons of

114   the same group. The active site configuration of the two phylogenetic groups also

115   differs. Although the significance of these distinctions has not yet been resolved,

116   it is believed that they may affect substrate specificity and activity [22].

117 Nevertheless, although CO oxidation mediated by bacterioplankton is

118 biogeochemically significant, the energy obtained through this process seems to

119 have a negligible effect on bacterial metabolism [23].

120 Besides their ability to obtain energy from light, some heterotrophic bacteria may

121 obtain additional carbon by fixing CO_2 through RuBisCO [24], or anaplerotic

122 pathways [25]. Nevertheless, mixotrophy is not restricted to heterotrophic

123 prokaryotes. It was recently shown that *Prochlorococcus* cells use a sugar

124 transporter gene to take up glucose to tune their carbon metabolism [26] and

125 other evidences of organic matter incorporation by cyanobacteria exist [e.g., 27].

126 Yet, a major challenge to our comprehension of the global carbon cycle is to

127 understand bacterial processing of dissolved organic carbon (DOC). DOC is

128 composed by thousands of compounds, many of them uncharacterized, and

129 utilized by a diverse community of heterotrophic bacteria with varying enzymatic

130 capabilities and ecological strategies for carbon metabolism [28]. The number of

131 studies dealing with specific genes encoding for these enzymatic activities is

132 notably limited, with the exception of the genes involved in chitin degradation

133 [29]. Currently, *meta'omics* allows depicting the diverse enzymes involved in

134 DOC utilization [30]. However, the functional annotation of the genes encoding

135 these enzymes is frequently based on distant homology to proteins characterized

136 in non-marine isolates, which may translate into a misinterpretation of their

137 function. Besides the utilization of complex DOC compounds, genes encoding

138 pathways for demethylation and C1 oxidation have been identified in SAR11 and

139 other marine bacteria like *Roseobacter* or methylotrophs, supporting the

140   hypothesis that C1 oxidation might be a mechanism by which dissolved organic

141   carbon is significantly recycled to CO_2 in the upper ocean [31].

142   **Phosphorus cycle**. Phosphorus (P) availability is one of the dominant selective

143   forces driving niche partitioning and ecotype divergence in oceanic prokaryote

144   populations [32]. Inorganic phosphate (P_i) is the preferred source of P for

145   prokaryotes, and they have evolved mechanisms for monitoring and responding

146   to fluctuations of this essential nutrient. The Pho regulon integrates the sensing

147   of P_i availability with co-regulation of genes involved in P_i scavenging (like *pstS*)

148   or the use of alternative sources of P, such as phosphoesters (through *phoA*,

149   *phoX*, *phoD* and other phosphatase genes), phosphonates (*phn* genes) and

150   phosphite (*ptx* genes) [33,34]. Recent studies show that bacteria inhabiting

151   permanent P-depleted areas of the ocean contain a higher number and diversity

152   of Pho regulon genes, thereby increasing their ability to obtain P [32]. A recent

153   study provides a remarkable example among the ubiquitous SAR11, which

154   despite their streamlined genomes display striking differences in their P-related

155   gene content. An isolate from the P-rich northeast Pacific Ocean is only able to

156   grow on phosphate, whereas a strain isolated from the P-deprived Sargasso Sea

157   can utilize a broad range of alternative compounds for P nutrition [35]. Similar

158   results have been observed in *Prochlorococcus* strains pointing out how the

159   environment shapes the genetic diversity of different bacterial populations

160   [32,36].

161   **Sulfur cycle**. The marine sulfur cycle has been focus of intense research in the

162   last years due to the potential role of dimethylsulphide (DMS) in regulating

163 climate. Dimethylsulphoniopropionate (DMSP) is released by some

164 phytoplankton cells and is either demethylated to methylmercaptopropionate

165 (MMPA), or cleaved to DMS by DMSP lyases [37,38]. DMS is a volatile

166 compound that is later photo-oxidized to sulfate aerosols that form cloud

167 condensation nuclei, initiating cloud cover over the oceans which leads to

168 increases in albedo. As for the other essential nutrients, the combination of

169 classical techniques and *-omics* has been crucial for the advancement on the

170 knowledge of sulfur cycling in the ocean. Metagenomic data revealed that the key

171 gene in the demethylation process (*dmdA*) is diverse among different taxa and

172 can be up to two orders of magnitude more abundant that DMSP lyases [38].

173 Indeed, the demethylation pathway provides reduced carbon for energy or

174 biomass and reduced sulfur that can be incorporated into S-containing

175 aminoacids or oxidized for energy. Hence, this pathway seems ecologically more

176 advantageous in oligotrophic environments than the cleavage pathway.

177 Nevertheless the DMSP lyases are a wide group of enzymes that seem to

178 catalyze the same reaction, whereas *dmdA* is the only known demethylase [37].

179 **Nitrogen cycle**. Traditionally, *Trichodesmium* and other heterocyst-forming

180 cyanobacteria were thought to be the predominant N_2-fixing microorganisms in

181 the ocean [39]. However we currently know that unicellular cyanobacteria such

182 as *Croscosphaera watsonii* [40] or the prymnesiophyte symbiont UCYN-A [41],

183 together with heterotrophic bacteria [6] dominate the diazotrophic community in

184 ocean surface waters. Furthermore, ammonia-oxidizing archaea were discovered

185 a decade ago [12] and are now known to be responsible for most of the

186 nitrification in the upper ocean [42]. These findings have arisen interest towards

187 genes involved in N cycling with several studies that have explored the diversity

188 of *nifH* (involved in nitrogen fixation) and *amoA* (involved in ammonia oxidation)

189 among other genes. Some of these studies have allowed linking gene diversity

190 with functional significance. For example, a recent study showed that two distinct

191 types of archaeal *amoA* genes have different biogeographies, one operating at

192 medium, and one at low ammonia concentrations [43]. In addition, recent

193 metagenomic analyses have shown that Thaumarchaeota contain genes

194 involved in urea transport and degradation (*ureA*, *ureB*, *ureC* genes) and that

195 these organisms may fuel nitrification in oceanic polar regions [5].

196 **Towards understanding functional diversity**

197 Rapid technological development has enabled the description of the functional

198 diversity of ocean microbes. Yet, the astonishing amount of genetic information

199 generated is not accompanied by a significant improvement in our understanding

200 of its functional significance and therefore, many relevant ecological questions

201 remain unsolved. Here, we discuss some of these questions.

202 *What is the significance of the genetic variability for well-studied genes?*
203 Molecular techniques have unveiled the diversity of the key players in marine

204 biogeochemical cycles and how genetic diversity is structured in relation to

205 environmental variables. For example, differential effects of abiotic factors on

206 AAP bacteria harboring different *pufM* sequences suggest that not all

207 phylogroups are functionally and ecologically equal [8]. Likewise, experimental

208 data have shown that the abundance of different sequence clusters of the

209 demethylation gene *dmdA* linked to specific marine taxa including *Roseobacter*-

210 and SAR11-like bacteria, correlate with certain environmental characteristics [38].

211 A global study of *nifH*-harboring bacterioplankton has reveled specific

212 biogeography for different taxa [6]. Additionally, studies on diversity of enzymes

213 degrading organic compounds show that there is a whole suit of different

214 enzymes carrying out the same function but with very different kinetics [44],

215 probably reflecting the heterogeneity of substrate concentration at the microscale

216 [45]. Microbes may also overcome limitation of enzyme cofactors by swapping

217 out one metal for another in the same enzyme [46] or synthesizing different

218 metalloenzymes for the same purpose. For example, calcium based

219 phosphatases like PhoX and PhoD seem to be prevalent in oligotrophic waters,

220 as opposed to the classical zinc-binding form of the enzyme (PhoA) [33,47].

221 Nevertheless, perhaps the best-known example of the significance of genetic

222 variability is the case of the rhodopsins, which as seen above, display large

223 genetic diversity that translates into various physiological and ecological

224 functions.

225 *What is the function of the astounding number of hypothetical proteins? –Omics*

226 datasets contain thousands of sequences encoding hypothetical proteins whose

227 characterization requires tedious experimental work. While the genetic

228 engineering approach is rarely conducted in marine studies, the results obtained

229 so far stress its value. Heterologous expression was used to prove that

230 proteorhodopsins are light-induced proton pumps [4] and through mutagenesis

231 we learned more about their differential spectral tuning [48]. Functional screening

232 of fosmid libraries allowed the identification of a phosphonatase pathway as well

233 as a novel pair of genes that allow utilization of 2-aminoethylphosphonate,

234 alternative sources of P for marine microorganisms [49]. These approaches have

235 led to major advances in our understanding of functional diversity and thus, more

236 effort should be put in applying these techniques to advance in our knowledge of

237 the functional capabilities within microbial populations.

238 *What genes are actually expressed and what environmental variables regulate*

239 *gene expression?* Metatranscriptomic studies provide insights into which genes

240 are transcribed, and therefore likely expressed, in the environment. Comparative

241 expression studies have evidenced for example that coexisting heterotrophic

242 bacterial populations exhibit diel oscillations in their expression patterns,

243 resembling those shown by their photosynthetic counterparts [50], or that

244 bacterioplankton increase the transcription of genes involved in cell aggregation

245 when accompanying a phytoplankton bloom [51]. Studies combining both

246 metagenomics and metatranscriptomics are essential to better understand what

247 genes are expressed under which circumstances, but these studies are only

248 starting to appear [52]. Furthermore, current metatranscriptomic studies recover

249 only a very small percentage of the total pool of transcripts (~0.000001%, [52]).

250 Significant gains in the analysis of biogeochemically informative gene expression

251 patterns require a greater sequencing investment. Initiatives like the global

252 circumnavigation expedition TARA Oceans [53] are presently applying massively

253 high-throughput sequencing to describe the functional profiles of the world's

254 oceans and promise to greatly advance our understanding of ecosystem

255 functioning. Yet, to understand what factors drive the observed trends, a precise

256 characterization of the environment is necessary, something that is not always

257 conducted.

258 *What mechanisms drive the observed functional genetic diversity?* Comparative

259 gene and genome analyses have revealed the ecological and evolutionary forces

260 that influence genome content. Genome streamlining seems to be a successful

261 adaptation to live under permanent low nutrient concentration, as seen in SAR11,

262 *Prochlorococcus* [54] and in wide array of marine single amplified genomes [55].

263 However, even those reduced genomes display a significant variability in gene

264 content among closely related populations. As mentioned above, one of the

265 dominant selective forces driving niche partitioning and ecotype divergence

266 within these populations is nutrient availability [32]. Indeed, many of the genes

267 involved in the efficient acquisition of nutrients are horizontally transferred and

268 often found in genomic islands [32]. For example, certain microdiverse lineages

269 of uncultured *Prochlorococcus* harbor a genomic island that contains the genes

270 necessary to assimilate nitrite and nitrate, as opposed to their cultured

271 representatives that rely on ammonia [56]. Likewise, alkaline phosphatases

272 (*phoA*, *phoX* and *phoD*) are not found in SAR11 isolates, but are present in

273 uncultured representatives of this ubiquitous clade [47,57]. Phosphatases are

274 needed for the utilization of a major component of the dissolved organic P pool,

275 and have broad substrate specificity. Therefore, the acquisition of just one

276 phosphatase gene enables the cell to access a wide spectrum of P compounds,

277 which can be very advantageous in oligotrophic environments. Other examples of

278 horizontally transferred genes are those involved in the adaptation to different

279 light regimes or in obtaining additional energy from light [59]. Moreover, genomic

280 islands are enriched in genes encoding hypothetical proteins which could play a

281 role in increasing bacterial fitness under changing environmental conditions [60],

282 but determining their function is necessary to confirm this hypothesis. Thus,

283 despite not being the only mechanism (e.g. mutation), horizontal gene transfer is

284 largely responsible for the observed diversity.

285 In summary, although several key questions remain to a large extend

286 unanswered, the *–omics* approach provide a framework for evaluating functional

287 gene diversity in an environmental context while at the same time is a powerful

288 tool to formulate hypotheses on microbially mediated processes. What the field is

289 currently lacking is taking a step further to experimentally test *–omics* emerged

290 hypotheses. Function-driven metagenomics and combination of *–omics* with

291 classical microbiology techniques are necessary to obtain a more comprehensive

292 view of the functioning of the very complex marine microbial communities, a

293 relevant matter for marine ecology.

294 **Acknowledgments**

300
301   **REFERENCES**
302
303   1.   Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: the unseen**
304   **majority.** *Proc Natl Acad Sci USA* 1998, **95:**6578-6583.

305   2.   Pedrós-Alió C: **Marine microbial diversity: can it be determined?**
306   Trends in microbiology. *Trends Microbiol* 2006, **14:**257-263.

307   3.   Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington
308   K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L *et al.*: **The**
309   **Sorcerer II Global Ocean Sampling expedition: expanding the universe of**
310   **protein families.** *PLoS Biol* 2007, 5:e16

311   4.   Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP,
312   Jovanovich SB, Gates C, Feldman RA, Spudich JL *et al.*: **Bacterial rhodopsin:**
313   **evidence for a new type of phototrophy in the sea.** *Science* 2000, 289:1902-
314   1906.

315   5.   Alonso-Sáez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL,
316   Lovejoy C *et al.*: **Role for urea in nitrification by polar marine Archaea.** *Proc*
317   *Natl Acad Sci USA* 2012,**109**:17989-17994.

318   6.   Farnelid H, Andersson AF, Bertilsson S, Al-Soud WA, Hansen LH,
319   Sørensen S, Steward GF, Hagström Å, Riemann L: **Nitrogenase gene**
320   **amplicons from global marine surface waters are dominated by genes of**
321   **non-cyanobacteria.** *PLoS One* 2011, **6**:e19223.

322   7.   Fernàndez-Guerra A, Casamayor EO: **Habitat-associated phylogenetic**
323   **community patterns of microbial ammonia oxidizers.** *PLoS* One 2012,
324   **7**:e47330.

325   8.   Ferrera I, Borrego CM, Salazar G, Gasol JM: **Marked seasonality of**
326   **aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean**
327   **Sea as revealed by cell abundance, pigment concentration and**
328   **pyrosequencing of** *puf*M **gene.** *Environ Microbiol* 2014, **16**:2953-2965. (*)

329   This study applied for the first time high throughput sequencing to investigate
330   the diversity of the functional gene *puf*M and showed that it is larger than
331   previously reported. The results show a clear seasonality of different
332   phylogroups suggesting that AAP subpopulations replace each other over time.
333   Furthermore, differential effects of abiotic factors on different groups were
334   determined likely indicating that not all AAP phylogroups are functionally and
335   ecologically equal.

336   9.   Varaljay VA, Gifford SM, Wilson ST, Sharma S, Karl DM, Moran MA:
337 **Bacterial dimethylsulfoniopropionate degradation genes in the oligotrophic**
338 **North Pacific Subtropical Gyre.** *Appl Environ Microbiol* 2014, **78**: 2775-2782.
339 (*)

340   An exhaustive study of the diversity of the dimethylsulfoniopropionate (DMSP)
341   demethylase gene, *dmdA*. The authors designed PCR primers targeting 10
342   environmental clades and subclades of the DmdA protein, and showed that
343   *dmdA* richness in free-living versus particle-associated bacteria differs. The
344   paper also highlights how the availability of metagenomic sequence data has
345   significantly enhanced the design of PCR primer pairs for targeting key
346   functional genes.

347   10.   Breitbart. **Marine viruses: Truth or dare**. Ann Rev Mar Sci 2012, **4**:425-
348   448.

349   11.   DeLong EF, Béjà O: **The light-driven proton pump proteorhodopsin**
350 **enhances bacterial survival during tough times.** *PLoS Biol* 2010, **8**:e1000359.

351   12.   Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA,
352 Wu D, Paulsen I, Nelson KE, Nelson W *et al.*: **Environmental genome shotgun**
353 **sequencing of the Sargasso Sea.** *Science* 2004, **304**:66-74.

354   13.   Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T,
355 Neutze R, Pedrós-Alió C, Pinhassi J: **Light stimulates growth of**
356 **proteorhodopsin-containing marine Flavobacteria.** *Nature* 2007, 445:210-
357 213.

358   14.   Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ**: Energy**
359 **starved Candidatus *Pelagibacter ubique* substitutes light-mediated ATP**
360 **production for endogenous carbon respiration.** *PLoS One* 2011, **6**:e19725.

361   15.   Gómez-Consarnau L, Akram N, Lindell K, Pedersen A, Neutze R, Milton
362 DL, González JM, Pinhassi J: **Proteorhodopsin phototrophy promotes**
363 **survival of marine bacteria during starvation.** *PLoS Biol* 2010, **8**:e1000358.

364   16.   Kwon SK, Kim BK, Song JY, Kwak MJ, Lee CH, Yoon JH, Oh TK, Kim JF:
365 **Genomic makeup of the marine flavobacterium *Nonlabens* (*Donghaeana*)**
366 ***dokdonensis* and identification of a novel class of rhodopsins.** *Genome Biol*
367 *Evol* 2013, **5**:187-199. (**)

368   This study uncovered a gene encoding an unexpected type of microbial
369   rhodopsin containing a characteristic NQ motif, which forms a phylogenetically
370   distinct group. Genomic and metagenomic surveys demonstrated the
371   prevalence of this rhodopsin among microbial communities inhabiting
372   hypersaline environments, suggesting its involvement in sodium metabolism
373   and the sodium-adapted lifestyle. This paper is a good example of how gene
374   diversity may result in diverse physiological and ecological functions.

375 17. Yoshizawa S, Kumagai Y, Kim H, Ogura Y, Hayashi T, Iwasaki W, *et al.*:
376 **Functional characterization of flavobacteria rhodopsins reveals a unique**
377 **class of light-driven chloride pump in bacteria.** *Proc Natl Acad Sci USA* 2014,
378 **111**:6732-6737. (**)

379 A recent report on a new light-driven chloride pump found in a marine
380 bacterium, the flavobacterium *Nonlabens marinus*. Interestingly, this bacterium
381 also possesses a proteorhododpsin and a light-driven sodium pump, being the
382 first reported microbe to possess three functionally different types of
383 rhodopsins.
384
385 18. Spring S, Riedel T: **Mixotrophic growth of bacteriochlorophyll *a*-**
386 **containing members of the OM60/NOR5 clade of marine**
387 **gammaproteobacteria is carbon-starvation independent and correlates with**
388 **the type of carbon source and oxygen availability.** *BMC Microbiol* 2013,
389 13:117.

390 19. Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB, Béjà O:
391 **Assessing diversity and biogeography of aerobic anoxygenic phototrophic**
392 **bacteria in surface waters of the Atlantic and Pacific Oceans using the**
393 **Global Ocean Sampling expedition metagenomes.** *Environ Microbiol* 2007,
394 **9**:1464-1475.

395 20. Waidner LA, Kirchman DL: **Diversity and distribution of ecotypes of the**
396 **aerobic anoxygenic phototrophy gene *pufM* in the Delaware estuary.** *Appl*
397 *Environ Microbiol* 2008, 74:4012-4021.

398 21. Moran MA, Miller WL: **Resourceful heterotrophs make the most of light**
399 **in the coastal ocean.** *Nat Rev Microbiol* 2007, **5**:792-800.

400 22.   King GM, Weber CF: **Distribution, diversity and ecology of aerobic co-**
401 **oxidizing bacteria.** *Nat Rev Microbiol* 2007, **5**: 107-118.

402 23. Cunliffe M: **Physiological and metabolic effects of carbon monoxide**
403 **oxidation in the model marine bacterioplankton ruegeria pomeroyi DSS-3.**
404 *Appl Environ Microbiol* 2013, 79:738-740.

405 24. Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, *et al.*:
406 **Genome characteristics of a generalist marine bacterial lineage.** *ISME J*
407 2010, 4:784-798.

408 25. Palovaara J, Akram N, Baltar F, Bunse C, Forsberg J, Pedrós-Alió C,
409 González JM, Pinhassi J: **Stimulation of growth by proteorhodopsin**
410 **phototrophy involves regulation of central metabolic pathways in marine**
411 **planktonic bacteria.** *Proc Natl Acad Sci USA* 2014, **111**:3650-3658. (*)

412 This study explores the metabolic pathways mediating the light responses of
413 proteorhodopsin-containing bacteria. Changes in gene expression between light
414 and dark conditions revealed a tightly regulated shift in central metabolic

pathways. Carbon acquisition pathways seem to shift towards anaplerotic CO_2 fixation in the light. Their results show that regulatory interactions between dissolved organic carbon quality and central metabolic pathways critically determine the fitness of surface ocean bacteria engaging in proteorhodopsin phototrophy.

26.   Muñoz-Marín MC, Luque I, Zubkov MV, Hill PG, Diez J, García-Fernández JM: **Prochlorococcus can use the Pro1404 transporter to take up glucose at nanomolar concentrations in the Atlantic Ocean**. *Proc Natl Acad Sci USA* 2013, **110**:8597-8602. (**)

It was known that autotrophic cyanobacteria could incorporate organic compounds containing limiting nutrients, and this paper experimentally shows that *Prochlorococcus* can take up glucose to tune their metabolism to energetically benefit from changing environmental conditions. To prove this hypothesis, these authors introduced the gene that confers the capability for glucose uptake in a cyanobacterium unable to take up glucose. Furthermore, they were able to observe glucose uptake by *Prochlorococcus* in the environment.

27.   Gómez-Pereira PR, Hartmann M, Grob C, Tarran GA, Martin AP, Fuchs BM, Scanlan DJ, Zubkov MV: **Comparable light stimulation of organic nutrient uptake by SAR11 and *Prochlorococcus* in the North Atlantic subtropical gyre.** ISME J 2013, 7:603-614.

28.   Kujawinski EB: **The impact of microbial metabolism on marine dissolved organic matter.** *Ann Rev Mar Sci* 2011, **3**:567-599.

29.   Beier S, Bertilsson S: **Bacterial chitin degradation-mechanisms and ecophysiological strategies.** *Front Microbiol* 2013, **4**:149. (*)

A comprehensive review on the different mechanisms and the principal steps involved in chitin degradation at the molecular level while discussing the coupling of bacterial community composition to the measured chitin hydrolysis activities and substrate uptake rates.

30.   Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J: **Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom.** *Science* 2012, **336**:608-611.

31.   Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, Landry ZC, Giovannoni SJ: **One carbon metabolism in SAR11 pelagic marine bacteria.** *PLoS One* 2011, **6**:e23973.

32.   Coleman ML, Chisholm SW: **Ecosystem-specific selection pressures revealed through comparative population genomics.** *Proc Natl Acad Sci USA* 2010, **107**, 18634-18639.

454   33.   Sebastian M, Ammerman JW: **The alkaline phosphatase PhoX is more**
455   **widely distributed in marine bacteria than the classical PhoA.** *ISME J* 2009,
456   **3:**563-572.

457   34.   Martínez A, Osburne MS, Sharma AK, DeLong EF, Chisholm SW:
458   **Phosphite utilization by the marine picocyanobacterium *Prochlorococcus***
459   **MIT9301.** *Environ Microbiol* 2012, **14:**1363-1367.

460   35.   Carini P, White AE, Campbell EO, Giovannoni SJ: **Methane production**
461   **by phosphate-starved SAR11 chemoheterotrophic marine bacteria.** *Nat*
462   *Commun* 2014, 5:4346 (**)

463   This study shows that despite having streamlined genomes, SAR11 strains
464   display striking differences in their phosphorus (P)-related gene content
465   depending on the phosphate concentration of the waters where they were
466   isolated from. Due to differences in gene content, the isolate from P-rich
467   Northeast Pacific only grows on phosphate, whereas the isolate from the P-
468   deprived Sargasso Sea is able to grow on different sources of P. Interestingly,
469   the Northeast Pacific isolate initiates a general stress response upon P-
470   depletion because is not able to adapt to fluctuations of this essential nutrient.

471   36.   Martiny AC, Coleman ML, Chisholm SW: **Phosphate acquisition genes**
472   **in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation.** *Proc*
473   *Natl Acad Sci USA* 2006, **103**:12552-12557.

474   37.   Curson AR, Todd JD, Sullivan MJ, Johnston AW: **Catabolism of**
475   **dimethylsulphoniopropionate: microorganisms, enzymes and genes.** *Nat*
476   *Rev Microbiol* 2011, **9**:849-859.

477   38.   Moran MA, Reisch CR, Kiene RP, Whitman WB: **Genomic insights into**
478   **bacterial DMSP transformations.** *Ann Rev Mar Sci* 2012, **4**:523-442.

479   39.   Bergman B, Sandh G, Lin S, Larsson J, Carpenter EJ: ***Trichodesmium*-a**
480   **widespread marine cyanobacterium with unusual nitrogen fixation**
481   **properties.** *FEMS Microbiol Rev* 2013, **37:**286-302.

482   40.   Moisander PH, Beinart RA, Hewson I, White AE, Johnson KS, Carlson
483   CA, Montoya JP, Zehr JP: **Unicellular cyanobacterial distributions broaden**
484   **the oceanic N_2 fixation domain.** *Science* 2010, **327**:1512-1514.

485   41.   Zehr JP, Kudela RM: **Nitrogen cycle of the open ocean: from genes to**
486   **ecosystems.** *Ann Rev Mar Sci* 2011, **3**:197-225.

487   42.   Newell SE, Facwett SE, Ward BB. **Depth distribution of ammonia**
488   **oxidation rates and ammonia-oxidizer community composition in the**
489   **Sargasso Sea.** *Limnol Oceanogr* 2013, **58**:1491-1500.

43.   Sintes E, Bergauer K, De Corte D, Yokokawa T, Herndl GJ: **Archaeal** *amoA* **gene diversity points to distinct biogeography of ammonia‑oxidizing Crenarchaeota in the ocean.** *Environ Microbiol* 2013, 15:1647-1658 (*)

An example on how genetic diversity may translate into functional significance. Using two different *amoA* primer sets, these authors show that distinct ecotypes of marine Crenarchaeota Group I containing diverse *amo*A genes have differential affinities for ammonia, one operating at medium and the other at low concentrations.

44.   Arrieta JM, Herndl GJ: **Changes in bacterial β-glucosidase diversity during a coastal phytoplankton bloom.** *Limnol Oceanogr* 2002, 47:594-599.

45.   Dittmar T, Stubbins A: **Dissolved organic matter in aquatic systems**. In: Birrer B, Falkowski P, Freeman K (eds) Treatise of Geochemistry. 2nd edition, Elsevier, 2014, 12: 125-156.

46.   Dupont CL, Neupane K, Shearer J, Palenik B: **Diversity, function and evolution of genes coding for putative Ni-containing superoxide dismutases.** *Environ Microbiol* 2008, **10**: 1831-1843.

47.   Kathuria S, Martiny AC: **Prevalence of a calcium-based alkaline phosphatase associated with the marine cyanobacterium P*rochlorococcus* and other ocean bacteria.** *Environ Microbiol* 2011, **13**:74-78.

48.   Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R, Spudich EN, Spudich JL, Béja O: **Diversification and spectral tuning in marine proteorhodopsins.** *EMBO J* 2003, **22**:1725-1731.

49.   Martinez A, Tyson GW, Delong EF: **Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses.** *Environ Microbiol* 2010, 12: 222-223.

50.   Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, Schuster SC, Scholin CA, DeLong EF: **Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages**. *Science* 2014, 345:207-212. (**)

Ottesen *et al.* apply metatranscriptomics to samples collected during three days to analyze the *in situ* expression profiles of naturally occurring oceanic bacterial populations. They find that unexpectedly, several different heterotrophic bacterioplankton groups display diel cycling in many of their gene transcripts. These waves of diel gene transcription may influence both the tempo and the mode of matter and energy transformation in the sea.

51.   **Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA: Bacterial community transc*ription patterns d*uring a marine phytoplankton bloom.** *Environ Microbiol* 2012, **14:**228–239.

527   52.   Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M,
528   Meng J, Sun S, Medeiros PM, Paul JH *et al.*: **Microspatial gene expression**
529   **patterns in the Amazon River Plume.** *Proc Natl Acad Sci USA* 2014,
530   111:11085-11090. (**)

531   This article performs the first fully quantitative inventory of genes and transcripts
532   by combining metagenomics and metatranscriptomics, to study the presence
533   and expression of genes mediating elemental cycling at the microspatial scale
534   in the ocean's largest river plume, the Amazon River Plume. These authors
535   hypothesize that in response to patchiness in carbon, nutrients, and light at the
536   micrometer scale, Amazon Plume microbes regulate the expression of genes
537   relevant to biogeochemical processes at the ecosystem scale.

538   53.   Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas, C, Raes J, Sullivan
539   M, Arendt D, Benzoni F, Claverie JM *et al.*: **A holistic approach to marine eco-**
540   **systems biology.** *PLoS Biology* 2011, 9:e1001177

541   54.   Giovannoni SJ, Thrash JC, Temperton B: **Implications of streamlining**
542   **theory for microbial ecology.** *ISME J* 2014, **8:**1553-1565.

543   A review addressing the implications of genome streamlining in both symbiotic
544   and commensal bacteria, as well as in cosmopolitan free-living bacteria having
545   small genomes. The causes of genome reduction in these two groups of
546   organisms are different but they imply loss of function. The consequences of
547   reductive evolution in streamlined organisms are discussed.

548   55.   Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González
549   JM, Luo H, Wright JJ, Landryh ZC, Hansoni NW *et al.*: **Prevalent genome**
550   **streamlining and latitudinal divergence of planktonic bacteria in the surface**
551   **ocean.** *Proc Natl Acad Sci USA* 2013, **110**:11463-11468. (**)

552   Using large-scale single cell genomics, this paper explores the genome content
553   of many bacterial lineages inhabiting the surface ocean. By comparison with
554   existing cultures, they found that natural bacterioplankton have smaller
555   genomes, fewer gene duplications, and are depleted in guanine and cytosine,
556   noncoding nucleotides, and genes encoding transcription, signal transduction,
557   and non-cytoplasmic proteins. These findings provide strong evidence that
558   genome streamlining and oligotrophy are prevalent features among diverse
559   free-living bacterioplankton.

560   56.   Martiny AC, Kathuria S, Berube PM: **Widespread metabolic potential for**
561   **nitrite and nitrate assimilation among *Prochlorococcus* ecotypes.** *Proc Natl*
562   *Acad Sci* USA 2009, **106**:10787-10792.

563   57.   Luo H, Moran MA: **Assembly-free metagenomic analysis reveals new**
564   **metabolic capabilities in surface ocean bacterioplankton.** *Environ Microbiol*
565   *Rep* 2013, **5**:686-696

566   58.   Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF,
567 Chisholm SW: **Genomic islands and the ecology and evolution of**
568 ***Prochlorococcus.*** *Science* 2006, **311**:1768-1770.

569   59.   Cho JC, Stapels MD, Morris RM, Vergin KL, Schwalbach MS, Givan SA,
570 Barofsky DF, Giovannoni SJ: **Polyphyletic photosynthetic reaction centre**
571 **genes in oligotrophic marine *Gammaproteobacteria*.** *Environ Microbiol* 2007,
572 **9**:1456-1463.

573   60.   Fernández-Gómez B, Fernàndez-Guerra A, Casamayor EO, González JM,
574 Pedrós-Alió C, Acinas SG: **Patterns and architecture of genomic islands in**
575 **marine bacteria.** *BMC Genomics* 2012, **13**:347.

576

Figure legends

Figure 1. Simplified view of major biogeochemical cycles in the sunlit ocean. Solid lines indicate prokaryotic mediated processes and the key functional genes involved in the processes are shown in boxes. Asterisks denote groups of genes with a related function. Briefly, phytoplankton fix carbon through photosynthesis, incorporate inorganic nutrients and release dissolved organic matter, which includes dissolved organic carbon (DOC), dissolved organic phosphorus (DOP), and dissolved organic nitrogen (DON). Certain phytoplankton species also release dimethylsulphoniopropionate (DMSP). Zooplankton also excretes dissolved organic matter, including urea. Prokaryotes, especially heterotrophic bacteria but also certain cyanobacteria, use the organic matter released by phyto- and zooplankton, and compete with phytoplankton for the inorganic nutrients. DMSP can be either cleaved by different DMSP lyases (encoded by *ddd* genes) to DMS and eventually released to the atmosphere, or demethylated by *dmdA* and used by prokaryotes as a reduced sulfur source. DOC is metabolized by a diverse number of enzymes, such as glucosidases, sulfatases or chitinases, but only the chitinases genes are well characterised. DOP is mainly composed of phosphoesters and phosphonates, which to be used require the action of alkaline phosphatases (*phoX*, *phoA*, *phoD*) and phosphonates genes (*phn* genes). The key enzymes for the utilization of DON are aminopeptidases, but the molecular basis of these enzymes is still largely unknown. Some heterotrophic bacteria are able to obtain additional carbon by fixing CO_2 through RuBisCO (*rbcL*) or alternative pathways (see main text for details). Other heterotrophic bacterial cells can obtain additional energy from sunlight by means of the proteorhodopsin or the *puf* operon, or from the oxidation of carbon monoxide (CO). Prokaryotes can obtain nitrogen from ammonia (NH_4^+) or nitrate (NO_3^-), and some cyanobacteria and heterotrophic bacteria are capable of fixing N_2 by means of nitrogenases (*nif*). The oxidation of ammonia in the sunlit ocean is mainly performed by archaea by means of the *amoA* gene, and the degradation of urea (*ure* genes) plays a major role in fuelling this process. Viruses lyse prokaryotic and eukaryotic cells, releasing DOM, particulate organic matter and nutrients. Viruses also play a crucial role in the exchange of genetic information among marine prokaryotes.