bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Woeseiales transcriptional response in Arctic fjord surface sediment

Joy Buongiorno[1#*], Katie Sipes[1], Kenneth Wasmund[2,3], Alexander Loy[2,3], Karen G. Lloyd[1]

[1]Department of Microbiology, University of Tennessee, Knoxville, USA

[2]Division of Microbial Ecology, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

[3]Austrian Polar Research Institute, Vienna, Austria

[*]Corresponding author: JBuongiorno@carnegiescience.edu

[#]Current affiliation: Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA

Running title: Responses to burial in Woeseiales

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Distinct lineages of Gammaproteobacteria clade Woeseiales are globally-distributed in marine sediments, based on metagenomic and 16S rRNA gene analysis. Yet little is known about why they are dominant or their ecological role in Arctic fjord sediments, where glacial retreat is rapidly imposing change. This study combined 16S rRNA gene analysis, metagenome-assembled genomes (MAGs), and genome-resolved metatranscriptomics uncovered the *in situ* abundance and transcriptional activity of Woeseiales with burial in four shallow sediment sites of Kongsfjorden and Van Keulenfjorden of Svalbard (79°N). We present five novel Woeseiales MAGs and show transcriptional evidence for metabolic plasticity during burial, including sulfur oxidation with reverse dissimilatory sulfite reductase (*dsrAB*) down to 4 cm depth and nitrite reduction down to 6 cm depth. A single stress protein, spore protein SP21 (*hsp*A), had a tenfold higher mRNA abundance than any other transcript, and was a hundredfold higher on average than other transcripts. At three out of the four sites, SP21 transcript abundance increased with depth, while total mRNA abundance and richness decreased, indicating a shift in investment from metabolism and other cellular processes to build-up of spore protein SP21. The SP21 gene in MAGs was often flanked by genes involved in membrane-associated stress response. The ability of Woeseiales to shift from sulfur oxidation to nitrite reduction with burial into marine sediments with decreasing access to overlying oxic bottom waters, as well as enter into a dormant state dominated by SP21, may account for its ubiquity and high abundance in marine sediments worldwide, including those of the rapidly shifting Arctic.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Seafloor sediments host a significant portion of Earth's biomass [1]. Microorganisms must have physiological adaptations to be active in marine sediments. Steep gradients in oxygen, redox active compounds, organic matter compositions, metal distributions, and metabolite pools are some of the many characteristics to which a microorganism must be able to respond [2-4]. Alternatively, increasing energy limitation with depth in the subseafloor environment may promote a strategy of metabolic standby until conditions again become optimal [5, 6], perhaps through exhumation through bioturbation, sediment slumping, or storm events in shallow waters. The particular strategy used by sedimentary microbes will likely depend on the sedimentological regime, degree of bioturbation, and depositional history.

One group that may have adaptive strategies for dealing with diverse sediment environments is the order Woeseiales (formerly 'JTB255') [7]. The Woeseiales are a globally-distributed group of Gammaproteobacteria that have been shown through 16S rRNA gene amplicon sequence surveys, as well as catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH), to be highly abundant in both coastal and deep-sea surface sediments [7-11]. Metagenomic studies [7] and 16S rRNA gene oligotyping studies [12] have revealed that this clade is composed of distinct lineages that inhabit either coastal or deep-sea sediments, suggesting that subclades within Woeseiales may interact with their sediment environment differently. Single-cell amplified genomes (SAGs), metagenome-assembled genomes (MAGs), transcriptomics, and incorporation of ^{13}C-labeled inorganic carbon showed the potential for chemolithoautotrophy via the oxidation of reduced sulfur compounds and the Calvin Benson cycle in coastal subclades [13]. A stronger potential for heterotrophy via

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

71 oxidation of carbohydrates, peptides, aromatics [7] and molecular hydrogen [10] has been
72 suggested for deep-sea subclades [12]. Interestingly, the only cultured representative, *Woeseia*
73 *oceani* XK5, from a coastal sediment sample, appears incapable of autotrophic growth [14]. This
74 supports the idea that different populations of Woeseiales may have fundamentally different
75 physiologies. Even single representatives of Woesiales have been hypothesized to be metabolic
76 generalists, with evidence for a high degree of metabolic flexibility [15].
77
78 The physiological responses of individual populations to burial in marine sediments, such
79 as transcriptional regulation and dormancy strategies, remain largely unexamined and may be
80 key to understanding how Woeseiales taxa modulate their behavior in different sediment
81 environments. We hypothesize that a high depth-resolution metatranscriptomic analysis will
82 demonstrate metabolic plasticity as well as strategies for long-term survival and subsistence in
83 the sediment environment. Altogether, this work will shed light on the ability to extrapolate the
84 functions made previously from genomic inferences of Woeseiales.
85
86 Woeseiales have also been found in Arctic fjord sediments in Svalbard, Norway [11, 13].
87 The Arctic is currently experiencing climatic changes from increased temperatures [16],
88 frequency and severity of storm events [17], and glacial retreat [18]. The ability to predict
89 whether members of a globally important microbial taxon such as Woeseiales will be resilient to
90 this change requires understanding the mechanisms by which it responds to the stresses of burial
91 in marine sediments. We therefore examined changes in gene transcript recruitment with depth to
92 MAGs of uncultured Woeseiales over the upper 5 to 6 cm of sediments from two Svalbard
93 fjords: Kongsfjorden and Van Keulenfjorden. Here, glaciers terminate in marine-influenced fjord
94 systems, where high rates of sediment loading (5-10 cm/year [19, 20]) derive directly from

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

glacial run-off. However, as glaciers retreat, this sediment loading may decrease [20, 21]. We examined two sites from Kongsfjorden (sites F and P) and one site from Van Keulenfjorden (site AC) with heavy loading of glacially-derived sediment, as well as another site from Van Keulenfjorden (site AB) with minimal input of glacially-derived sediments. Five novel medium to high quality Woeseiales MAGs were reconstructed from metagenomic sequences from these four sites. Mapping of transcripts to these MAGs demonstrated broad decreases in transcription with depth for the majority of genes, while one gene, *hsp*A (encoding spore protein SP21), comparatively increased transcription with depth by multiple orders of magnitude.

Results

Sites within Van Keulenfjorden (sites AB and AC) and Kongsfjorden (sites P and F), Svalbard, were investigated for their geochemical properties and relative abundance of Woeseiales with depth (Fig 1 a-c). Previous geochemical analysis on porewaters at these sites showed higher relative concentrations of Fe (up to ~300 nM) at site AB compared to <50 nM at site AC in the top 5 cm of sediment [22]. Within shallow sediments at site AB, Mn concentrations never exceed 65 nM, while at site AC, Mn concentrations exceed 250 nM [22]. Similarly high Fe concentrations (~300 nM) have been observed in shallow sediments at site P in Kongsfjorden, with Mn concentrations not exceeding 20 nM [23].

Figure 1: Physical and geochemical setting. Overview map of Spitzbergen (a), containing red box indicating Kongsfjorden (b) and green box indicating Van Keulenfjorden (c). Images taken from USGS Land Look web server (https://landlook.usgs.gov/viewer.html). The relative abundance of Woeseiales by 16S rRNA gene amplicon analysis for sites AB (d) and AC (e) in

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Van Keulenfjorden are shown along with measurements of hydrogen (f), total organic carbon (g), carbon to nitrogen ratios (h), and stable isotopes of organic carbon (i). Outliers were determined with Cook's distance.

The relative abundance of Woeseiales 16S rRNA gene amplicons remained steady with depth in Van Keulenfjorden sediment, between 1-3 % of 16S rRNA gene amplicon libraries down to ~20 cm at sites AB and AC in Van Keulenfjorden (Fig 1d-e). 16S rRNA genes did not amplify for sites P and F in Kongsfjorden, likely due to the high concentrations of Fe, which can interfere with DNA extraction and amplification [24].

In Kongsfjorden, hydrogen, total organic carbon (TOC), C/N ratios, and stable isotopes of organic carbon were variable in the upper five cm (Fig 1f-i). This is similar to sites AB and AC in Van Keulenfjorden [22, 25]. High C/N ratios and low TOC values at sites AC, F, and P indicate the deposition of refractory terrestrial organic material relative to site AB where corresponding values were lower [22, 25].

Recruitment of metagenomic and metatranscriptomic reads to Woeseiales MAGs

Metagenomic sequence libraries were generated from the depth-integrated upper five cm at three of the four sites (AB, AC, and F), with read numbers ranging from 30,615,734 to 83,705,875 per library. The metagenomic sequencing failed for site P, perhaps due to the same coextraction of inhibitors that probably prevented amplification of 16S rRNA genes. Five Woeseiales MAGs were reconstructed from the metagenomic data (Fig S1) and found to be

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

141 closely related to the Woeseiales Lineage 11 representative JSS Woes1 [12]. Amino acid identity
142 analysis shows that the majority of proteins in our MAGs match unclassified
143 Gammaproteobacteria and/or Chromatiales (Fig S2). Less than half of the proteins in the
144 Woeseiales MAGs have average amino acid identity to the proteins from the isolate *Woeseia*
145 *oceani*, pointing to proteomic novelty in our environmental MAGs compared to the isolate.
146
147 MAG completeness ranged from 80 to 94 %, with 2 to 9% contamination (Table S1).
148 Woeseiales collectively recruited from 24% of 6,145,182 reads mapped to the metagenome at
149 site F to 33% of 15,328,660 reads mapped to the metagenome at site AC, highlighting the
150 importance of this clade to the sediment microbial community. Across all transcriptomes (which
151 ranged from 272,522 reads to 1,307,568 reads after removal of rRNA), Woeseiales persistently
152 recruited from 0.5 to 3% of mapped reads. A high similarity in predicted gene functions with
153 KEGG-decoder [26] between the five Woeseiales MAGs permitted a combined analysis for
154 transcripts (Fig S3).
155
156 Most genes that recruited transcripts (54% of total transcribed genes) were recruited only
157 from one of the four sites (Figs S4, S5). However, the majority of these transcripts that only
158 appeared at one site (93%) occurred only in a single depth. We therefore restricted analyses to
159 transcripts occurring at more than one depth, to filter out potential noise. This showed transcripts
160 unique to a single site included electron transport proteins, proteins involved in cellular
161 signaling, and chaperones (Fig S5). Specifically at site AB, we identified transcripts for the
162 oxidoreductase complex protein subunits RnfE (Em = -500 mV [27]) and RsxB, which are
163 involved in electron transfer to regulatory proteins that respond to oxidative stress [28]. At site
164 AC, we identified exclusive transcripts for bifunctional (p)ppGpp synthase/hydrolase SpoT. This

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

enzyme catalyzes the synthesis and degradation of guanosine pentaphosphate ((p)ppGpp), an alarmone involved in signaling and bacterial stringent response to nutritional starvation and the transition to a state of dormancy under stress [29]. At site F, we detected unique transcripts for chaperone ClpB. This chaperone is involved in recovery after heat-induced damage and works in concert with the DnaK chaperone system (transcribed at F, AB, and P) to refold denatured protein aggregates [30]. Finally, at site P, we identified unique transcripts for RNA polymerase sigma factor SigV, an extracytoplasmic function sigma factor involved in transcription under cell envelope stress [31].

Among the genes that were transcribed at all sites (8% of total transcribed genes; Fig S4), about 20% encoded proteins involved in energy production and conversion (Table S2). This included transcripts for the central metabolism genes encoding citrate synthase, acetyl-coenzyme A synthase, and isocitrate lyase. In addition, several genes involved in sulfide oxidation metabolism were found at all sites. This included the cytochrome subunit of sulfide dehydrogenase (*fcc*A, present in all MAGs) and the beta subunit of dissimilatory sulfite reductase (*dsr*B). Sequence analysis of the alpha and beta subunits (*dsr*AB) within two Woeseiales MAGs (Woeseia sp. stnF and Woeseia sp. stnAC-2) confirms that members of this clade encode a reverse dissimilatory sulfite reductase involved in oxidation of sulfur (Fig 2) [7]. In Woeseiales, this may facilitate thiosulfate oxidation through a truncated SOX pathway involving SoxXA, SoxYZ, SoxCD, SoxB and a c-type cytochrome (c-555) (Fig 3a). The oxidation of reduced sulfur and sulfur intermediates is potentially linked to carbon fixation through the Calvin-Benson Cycle using an encoded Type II RubisCO (Figs 3a S6; Supplementary File S1); however, we were not able to detect transcripts for *sox* genes. Transcription of *dsr* subunits decreased with depth (Fig 3b).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 2: Phylogeny of DsrA encoded on Woeseiales MAGs. The CIPRES Science Gateway v. 3.3 online server was used for RAxML and the tree was visualized with the Interactive Tree of Life (iToL) visualization tool. Nodes with bootstrap values equal to or greater than 70% are indicated with a black circle. Reductive-type archaeal and bacterial sequences are represented in pink and orange fields, respectively, with solid tree branches. Oxidative-type bacterial sequences are represented in the blue field with dashed tree branches. Woeseiales DsrA sequences are highlighted in red text.

Figure 3: Sulfur oxidation and denitrification in Woeseiales MAGs. Schematic representation of metabolic pathways reconstructed in Woeseiales MAGs (a). The gray/black bar at the top corresponds to the redox potential for the key cytochromes depicted in the figure (sulfide quinone reductase, cbb_3-type c oxidase, cyt bc1 complex, NADPH oxidoreductase, and the Rnf complex). Colors in circles next to each protein represent the presence in the MAGs. Depth trends in transcriptional coverage (reported in TPM) for *nir*S and the alpha and/or beta subunits of *dsr* are reported (b).

Another gene that was transcribed at all sites was *nir*S. This gene encodes nitrite reductase, a key enzyme in denitrification. Transcripts for *nir*S were detected at every sediment depth (Fig 3b). Similar to r*dsrAB,* transcription of *nir*S decreased with depth. Transcripts of *nir*S were detected in deeper sediment depths compared to transcripts of *dsrAB*, which did not penetrate past 4-5 cmbsf (Fig 3b). While transcription levels of both genes decreased overall, transcription of *nirS* remained elevated below 2 cm, after which *dsr* transcripts were nearly zero

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

at most sites. This may indicate a transcriptional response to decreased availability of respective substrates, with oxygen and intermediates of sulfur becoming depleted at shallower depths relative to nitrite [32-34].

Transcriptional trends of hspA (SP21)

The gene *hsp*A, encoding SP21, also called heat-shock protein A, recruited at least ten times more transcripts in our Woeseiales MAGs than the next most highly transcribed gene all sites (chaperone protein DnaK, 1,916 TPM; Fig S7). SP21 is a small, membrane-associated stress protein related to heat shock proteins [35, 36] (Fig S8). This gene had very high transcriptional coverage, with values an order of magnitude greater on average (TPM = 2,770) compared to all other genes (average TPM = 531). This gene also had the largest transcription observed for any gene at any site, with 309,186 TPM at Site P, 5-6 cmbsf (Fig S7).

Depth analysis of SP21 coverage with post-hoc Tukey testing showed statistically-significant increases with depth at sites AB and P (ANOVA $p < 0.05$; Fig 4). At sites AC and F, transcriptional coverage did not show statistically-supported trends with depth (ANOVA $p > 0.05$). Instead, at these sites, transcription of spore protein SP21 either remains high at all depths (site F) or in the low-to-middle-range of coverage at all depths (site AC).

Figure 4: Transcriptional coverage of SP21 with depth at each site. Each point represents a copy of the gene encoding SP21 (*hsp*A) in a Woeseiales MAG. Some MAGs have more than one copy of the gene (see Fig S9). The upper and lower hinges correspond to the 75th and 25th percentiles, respectively, and the median is represented by a horizontal bar. Statistical differences

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in mean values across depths for a single site was calculated with Tukey means testing. Depths that do not share a letter/color have statistically significant different mean TPM values across the different copies of the gene.

In addition to SP21, transcripts for other stress proteins, transcriptional regulators, and starvation genes were detected (Fig S9). However, transcript recruitment to these genes was much lower than that observed for SP21, at 89 TPM on average. In addition, none of these genes showed increases with depth, even those co-localized on the same contig as the spore protein SP21 gene *hsp*A (Fig S10). In fact, transcripts for the co-localized chaperonins encoded by *gro*S and *gro*L decreased with depth (Fig. S9), while other chaperones either showed no trend (co-localized chaperon DnaK) or were only detected at the surface (chaperone SurA and chaperone HtpG; Fig S9).

Other genes flanking the SP21 gene *hsp*A included those encoding the protein-methionine-sulfoxide reductase MsrPQ system (*msr*P and *msr*Q; Fig S10). This membrane-associated system works with the periplasmic chaperone SurA to repair oxidatively damaged periplasmic proteins containing methionine sulfoxide residues, thereby maintaining cell envelope integrity [37]. Transcripts for chaperone SurA were identified in surface sediments of sites AB and AC (Fig S9). We identified several additional nearby genes involved in membrane-associated processes. These included genes that participate in toxin removal/efflux (*tol*C transcribed at AB, P, and F; *xcp*T and *eph*A) [38-40], protein repair (*deg*Q transcribed at AC, F, and P; *pcm*) [41, 42], and cell wall reorganization (*upp*P and *glu*P) [43].

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Transcriptional coverage and richness with depth

To get a broader understanding of transcriptional responses to burial, we classified Woeseiales transcripts into Clusters of Orthologous Groups (COGs) (Fig 5). We took this approach because low overall transcriptional coverage and inconsistent detection of transcripts for an individual gene across several depths precluded analysis at the gene-level. Spearman's rho was calculated to assess the degree of association between depth and TPM for each COG category. For this analysis, SP21 was removed to allow equitable comparison.

Figure 5: Coverage trends and transcriptional richness. Changes in transcription for each of the three main COGs (information storage and processing, cellular processing and signaling, and metabolism) at each site (a). For equitable comparison, SP21 transcripts were removed (because of its anomalously-high coverage values) as well as the two libraries representing the bottom 10% of library sizes (2-3 cm and 4-5 cm at site F). Metagenome coverage is not reported for site P, as we did not sequence a metagenome for this site. Changes in transcriptional richness, or the number of transcribed genes, with depth (b). Lines correspond to a subCOG category within a main category, each rendered to share the same color palette: khaki lines represent information storage and processing transcripts, pink lines represent cellular processes and signaling transcripts, and green lines represent metabolism transcripts. Spearman's rho was calculated for both COG-specific and site-wide trends in transcriptional richness. Only site-wide trends had significant p-values at alpha = 0.05.

Information storage and processing genes (mostly involved in transcription/translation) had the highest average transcription amongst the three main COG categories (average TPM =

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

281 91) compared to cellular processes and signaling (average TPM = 65) and metabolism (average
282 TPM = 60). With the exception of site P, all sites had significant decreases ($p < 0.05$) in
283 transcripts for all three of these main COG categories (Fig 5a). Metabolism genes had the
284 greatest decrease (Spearman's rho from -0.5 at site AC to -0.6 at site F), followed by cellular
285 processes and signaling (Spearman's rho from -0.24 at site AB to -0.58 at site F) and information
286 storage and processing (Spearman's rho from -0.25 at site AB to -0.47 at site F).
287
288 The COG with the largest number of genes that were transcribed was energy production
289 and conversion, followed at most sites by translation, ribosomal structure, and biogenesis (Fig
290 5b). We tested how richness within these COGs and others changed along a depth gradient
291 through Spearman analysis. This analysis revealed that the number of transcribed genes at site
292 AB decreased overall (Spearman's rho = -0.26, p value = 0.018), while at sites AC and P, there
293 was a slight increase (Spearman's rho = 0.30 and 0.41, respectively, $p \leq 0.005$). By contrast, site
294 F had no statistically-supported depth trend in transcriptional richness, either when evaluating
295 the entire data set or sub-setting by COG category (Spearman's rho = 0.11, p value = 0.31).
296
297 Combined, these analyses suggest that overall transcription generally decreased, while
298 trends in transcript richness were not uniform across sites. Instead, trends in transcriptional
299 richness broadly intersected with increased SP21 transcription (Figs 4 and 5). For example, at
300 site AB, overall decreases in transcriptional richness coincided with increased SP21 transcription
301 from 0-1 cm to 2-3 cm. At site P, increased richness from 0-1 cm to 2-3 cm was followed by a
302 dramatic decrease from 2-3 cm to 3-4 cm (Fig 5b). This decrease coincided with statistically
303 elevated amounts of SP21 transcription (Fig 4). At site AC, SP21 transcripts remained low at all
304 depths and were accompanied by an increase in total transcript richness. Interestingly, at site F,

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

305 we observed no change in either transcriptional richness or in SP21 transcription. Altogether, this
306 suggests that mRNA generation may be attenuated once a threshold of SP21 transcription is
307 reached.
308

309 # Discussion

310 Woeseiales populations made up 1-3% of the microbial community by 16S rRNA gene
311 amplicon analysis, and had high metagenomic read recruitment and transcriptional activity in the
312 sediments of Kongsfjord and Van Keulenfjord, Svalbard. This is in line with previous studies
313 that showed Woeseiales is an abundant lineage in marine sediments worldwide [7, 12, 13],
314 although the properties that determine such ecological success are poorly understood. Previous
315 work has suggested metabolic plasticity to be at the root of the global distribution and high
316 abundance continuously observed for Woeseiales worldwide [15]. Our high-resolution
317 transcriptional evidence supports this idea and suggests that the Woeseiales in these Arctic fjords
318 are able to sense and adapt to their environment upon burial, most likely in response to changes
319 in redox chemistry. In addition to metabolic plasticity, transcriptional evidence also supports an
320 additional mechanism that is potentially driving their ecological success: the transcription of a
321 small, membrane-associated stress protein, SP21 (*hspA*) [35, 36].
322

323 ## Metabolic response to burial

324 In these surface sediments, where redox cycling of carbon, nitrogen, and sulfur
325 compounds provide key feedbacks to the overlying water [22, 23, 25], we found metabolic
326 potential for chemolithoautotrophy with sulfur oxidation paired to carbon fixation (Fig 3a). This
327 agrees with previous work that hypothesized Woeseiales as members of the sulfur-oxidizing

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

328 community in marine sediments [7, 12, 13]. We also found strong transcriptional evidence that
329 Woeseiales contribute to denitrification by performing the NO_2 reduction step in the examined
330 sediment surface layers down to 5-6 cm (Fig 3b; Table S3). Transcriptional evidence suggests
331 that these metabolisms are occurring *in situ* and suggests sulfur oxidation is down-regulated
332 quickly with depth in response to changing redox conditions (Fig 3b). The electron sources
333 driving nitrite reduction need to be examined in the future. Metabolic plasticity within the
334 Woeseiales was first proposed by Dyksma et al. (2016). Using combined molecular and isotopic
335 approaches, they found support for Woeseiales to switch between sulfur and hydrogen oxidation.
336 They hypothesized that the global distribution and ubiquity of the Woeseiales clade may in part
337 be attributable to their metabolic flexibility. Likewise, analysis of MAGs from permeable marine
338 sediments in Australia suggest Woeseiales encode genes for aerobic heterotrophy, denitrification,
339 sulfur oxidation, and hydrogenotrophic sulfur reduction [15]. Further, meta-analysis of deep-sea
340 versus coastal Woeseiales taxa found chemoheterotrophic metabolism is a deep-sea trait [12].
341 Our analysis extends these findings of metabolic flexibility to include Arctic fjord sediments,
342 suggesting that metabolic flexibility likely drives the ecology and distribution of Woeseiales in
343 the marine sediment realm.
344

345 ## Dormancy and quiescence strategies

346 An additional mechanism that may facilitate the worldwide distribution of Woeseiales is
347 its transcriptional response to burial, suggesting cellular quiescence (Fig 4). The extremely high
348 number of transcripts of the single gene encoding SP21, or *hspA* (6 - 161 times higher than the
349 next most abundant transcript; Fig S7) that increases with depth (1.5 to 9 times increase over the
350 upper 5 cm at three sites) introduces a potentially important and overlooked strategy available to

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Woeseiales. SP21 is transcribed alongside other chaperones and genes for repair, perhaps as a general response to stressful conditions that occur as Woeseiales is buried. While the transcription of chaperones, repair proteins, and other genes related to stress was observed, transcription did not increase with depth and remained several orders of magnitude lower than transcription for spore protein SP21 on average (Fig S9). Therefore, the transcription of spore protein SP21 may indicate a quiescence strategy related to stress-induced dormancy [44].

In *Stigmatella aurantiaca*, the expression of membrane-localized clusters of SP21 is inducible under stress conditions like heat shock, anoxia, and nutrient deprivation [35, 45]. When expressed under heat shock, antigens for this stress protein have area densities ranging from ~400 to >800 conjugates/μm^2 within the cell wall [35], suggesting that the protein is in extremely high concentrations in the cell. Unlike the sulfate-reducing Firmicutes whose spores are inducible under high temperatures in Svalbard fjord sediment [46], our transcriptional evidence for putative spore-like dormancy in Woeseiales is limited to just one protein located in the cell wall. Thus, it is very unlikely that Woeseiales can form spores *sensu stricto* as is observed in the Firmicutes, such as *Bacillus subtilis*. This is supported by the inability of the isolate *Woeseia oceani* XK5 to form spores [14]. However, the extremely high transcript abundance of SP21 relative to all other genes, and its sharp increase with depth suggest that it might also have a high protein abundance, as it does in *S. aurantiaca*, where it provides a spore-like quiescence.

Indeed, decreases in abundance and diversity of transcripts for metabolism, information storage and processing, and cellular processes may represent a stress response pattern indicating

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

374 entry into a stage of quiescence. The transcriptional behavior observed for the Woeseiales MAGs
375 is similar to quiescence strategies noted for other organisms, such as *B. subtilis* and
376 *Mycobacterium tuberculosis* [44]. For example, the overall high transcription of a single protein
377 coupled to much lower overall mRNA diversity and abundance agrees with a recent RNA-seq
378 study on *B. subtilis* spores conducted by Korza et al. (2019). *B. subtilis* spores had overall low
379 transcriptional diversity and abundance, with the exception of a few high-abundance transcripts
380 that were ~4,000-fold more abundant than the least abundant mRNAs [47]. These high-
381 abundance transcripts encoded proteins involved in stringent response and spore coat formation
382 are thought to remain abundant post-sporulation because of mRNA stability conferred by
383 dehydration upon spore maturation [47, 48]. Stabilization of mRNA is also observed for
384 sporulating Actinobacteria [49, 50] and Mycobacteria entering into quiescence [51]. For most
385 sites, mRNA diversity and abundance were highest in uppermost sediment layers prior to build
386 up of SP21 transcripts. This is consistent with increased macromolecular synthesis observed to
387 occur during entry and exit from quiescence in *E. coli* and *M. tuberculosis* [54] in order to
388 facilitate cellular remodeling between growth states [55]. After quiescence is established,
389 transcription has been shown to decrease 3- to 5- fold in *Saccharomyces cerevisiae* [56]. And so,
390 while Woeseiales may not have the ability to form true spores, it may have a strategy that is
391 similar to cellular quiescence that has been observed for both pathogenic bacteria [52] and
392 environmental bacteria [53].
393
394 Altogether, high transcription of SP21, coupled to the transcription of chaperones,
395 transcriptional regulators, membrane-associated stress genes, spore coat protein A (Fig S9), and
396 peptidases related to sporulation factors (M50B), suggests that Woeseiales has a strategy for
397 cellular quiescence for surviving periods of stress. Such metabolic standby of microbial

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

398 populations, wherein growth and metabolism are extremely depressed or put on hold entirely
399 until conditions again become optimal, has been proposed as a key strategy of life in marine
400 sediments [6, 57]. These dormant populations act as seed banks that can potentially be
401 reactivated under different conditions [58], such as those resulting from resuspension events or
402 bioturbation [59]. Such a subsistence strategy may confer resilience to Woeseiales populations,
403 allowing their persistence in the environment and contributing, in part, to their global distribution
404 and abundance.
405

406 # Conclusions

407 Our results suggest that the Woeseiales MAGs reconstructed from Arctic fjord sediments
408 responded physiologically to burial over the upper 5 cm of sediments. Transcriptional evidence
409 for metabolic plasticity was found with shallow termination of transcription of rDSR likely
410 involved in sulfur oxidation tied to carbon fixation and deeper transcription of the nirS gene for
411 denitrification. A single membrane-associated stress protein, SP21, was highly transcribed at all
412 depths and at all sites. Increases in transcripts for SP21 at three sites coincided with decreased
413 investment in mRNA for metabolism, cellular processes and signaling, and information storage
414 and processing genes with depth. The site where SP21 did not increase with depth did not show
415 decreases in the abundance and diversity of cellular transcripts. Altogether, this suggests that in
416 addition to metabolic flexibility, the success of Woeseiales as a ubiquitous and abundant marine
417 sediment clade may be attributable at least in part to a novel subsistence strategy that involves
418 SP21. With multiple subsistence strategies (metabolic flexibility and cellular quiescence)
419 available to Woeseiales both within and outside the reach of large glacial deposits, they may be

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

420 well-prepared to withstand shifts in the sub-benthic conditions of Arctic fjords due to climate
421 change.
422

423 Materials and methods

424 Sediment collection

425 Sediment was collected in the summer of 2016 from four different sites within Svalbard
426 fjords (79°N) with a HAPs corer [60] (Fig. 1a). Sediments included in this study are from sites
427 AB (77°35.249' N, 15°05.121'E) and AC (77°32.260' N, 15°39.434' E) in Van Keuelenfjorden
428 (outer and middle sites, respectively; Fig. 1b) and sites F (78°55.075' N, 12°15.929' E) and P
429 (78°57.915'N, 12°15.600'E) in Kongsfjorden (both located at areas of glacier outflow) (Fig. 1c).
430

431 Geochemical measurements

432 Acrylic liners were used for collecting cores from each site only cm apart within each
433 larger HAPs corer. These cores were kept at ~5°C until they were brought back to land and
434 processed 1-3 days later. Cores were subsectioned every 3 cm down to a depth of ~25 cm.
435 Samples for hydrogen analysis followed a previously published headspace equilibration method
436 [61]. Briefly,1 mL of sediment was placed into a dark glass serum vial which was then crimp
437 sealed and gassed with N_2 for 15 min prior to storage at 4°C. Headspace was measured with
438 glass syringes on a Peak Performer GC with a mercuric chloride detector (San Jose, California)
439 at the University of Tennessee, Knoxville (UTK) after 2 days. Sediment for analysis of organic
440 matter was freeze-dried after thawing from -80°C and subjected to acid fumigation overnight
441 before analysis [62]. Total organic carbon (TOC) and isotope composition of carbon ($\delta^{13}C_{org}$)

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

from bulk organic matter was measured using a Thermo-Finnigan Delta XL mass spectrometer coupled to an elemental analyzer at UTK. Carbon to nitrogen (C/N) ratios were calculated by dividing percent C by percent N. Isotopic values were calibrated against the USGS40 and USGS41 international standards. In-house standard sets were run every 12 samples. Across multiple runs, one standard deviation was 0.1-0.2 ‰ for $\delta^{13}C_{org}$, 1.1-1.8 % for mgN, and 1.0-2.2 % for mgC. Values for Van Keulenfjorden are published elsewhere [22].

DNA extraction and 16S rRNA gene amplicon libraries

Cores were collected with acrylic liners a few cm apart from those taken for geochemical analysis and were subsectioned at 1 cm depth intervals in the Ny-Ålesund Marine Lab down to ~20 cm depth. Sediment was frozen immediately on dry ice and remained frozen during transport using a dry shipper. Sediment was stored at -80°C until processing. Nucleic acids were extracted by both the Lloyd (sites AB, AC, P, F) using the Qiagen RNeasy Powersoil® Kit (Hilden, Germany) for RNA with the DNA extraction accessory according to manufacturer's instructions. DNA from site AC at 18 cm depth was extracted by the Loy lab using the DNeasy PowerSoil Kit (Qiagen), following manufacturer's instructions. 16S rRNA gene amplification from sites P and F in Kongsfjorden failed. The Phusion Master Mix (Thermo Fisher) was used with the primer set 515F/806R [63] at the Center for Environmental Biology at UTK for amplification. Reads were sequenced with Illumina MiSeq and trimmed for quality with Trimmomatic [64] using a window 10 base pairs wide and a minimum phred score of 28. Trimmed reads were then processed in mothur 1.35.1[65] using the computational cluster at the Bioinformatics Resource Facility (BRC) at UTK. Operational taxonomic units (OTUs) were

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

464 clustered with mothur as previously described [22]. All 16S rRNA gene amplicon library
465 sequences can be accessed at NCBI under Bioproject PRJNA493859.
466

467 Metatranscriptomic and metagenomic sequencing

468 RNA extracts were treated with DNAse I at UTK (Qiagen) and further DNase treatment
469 was performed at MRDNA (Shallowater, TX), followed by sequencing of metatranscriptomic
470 libraries with Illumina HiSeq, PE 2x250 bp. Individual 1 cm-depth resolved metatranscriptomic
471 libraries were generated with RNA extracts from the first 5 to 6 cm of sediment from sites AB,
472 AC, F, and P, for a total of 20 metatranscriptomes. Metagenomic libraries were generated from
473 the combined extracts from the first 5 cm (spanning 0 to 5 cm downcore) in sites AB and F in the
474 Center for Environmental Biotechnology, Knoxville, using Illumina HiSeq, PE 2x250 bp. The
475 decision to make high depth resolution metatranscriptomes, and only a single metagenome for
476 each site was based on financial constraints. The Loy Lab sequenced DNA extracts from Site AC
477 at 18 cm depth using Illumina HiSeq 3000, PE 2x150 by the Biomedical Sequencing Facility, at
478 The University of Vienna.
479

480 Prior to metagenomic assembly, raw reads were trimmed for quality and adapters were
481 removed using in-house scripts in the Loy lab (removing the leading eight 5' bases, bases with
482 QC < 15 and reads below 50 bp in length) and Illumina adapters were removed using
483 AdapterRemoval using default settings [66]. The Lloyd Lab used Trimmomatic [64] for
484 trimming both metagenomes and metatranscriptomes, with a sliding window of 10 and a phred
485 cut off score of 28 for all reads above 90 bp. The quality of trimmed reads was assessed with

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Quast 4.5 [67]. Metagenomic sequences can be accessed in the NCBI SRA under BioSamples SAMN10372305, SAMN10372307, and SAMN10372306.

Metagenomic assembly: MetaSPAdes

Metagenomic assembly was accomplished using metaSPAdes [68] both in command line for Linux and in KBase, a browser interface with bioinformatics modules and applications [69]. Site F was the only metagenome that was assembled via metaSPAdes version 3.11 in the command line, with kmer size set to 21, 33, 55, 77, 99, and 127 to find the best assembly. All other metaSPAdes assemblies were completed on KBase with the default parameters (1000 bp length and kmer sizes of 21, 33, and 55). Assembled contigs were then filtered to contain only contigs with more than 5x coverage and 1000 bp length using in-house scripts.

Metagenomic assembly: IDBA and Megahit

To reduce RAM utilization and wall clock time, larger sequence datasets were normalized to 100X with BBN prior to assembly with IDBA version 1.1.3 [70] and Megahit version 1.0.5 [71]. Assemblies were completed in either the command line or on KBase [69]. All assemblies were generated with 1000 minimum contig length, except IDBA assembly in the Loy lab, which used a 500 bp cutoff.

Taxonomic binning of contigs into metagenome assembled genome (MAGs)

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Contig binning was carried out with MaxBin2 version 2.2.3 [72], CONCOCT (version 0.4.1) [73] and MetaBAT (version 2.12.1) [74] in command line. Each of these binning tools utilizes genomic signatures within contigs, such as coverage and tetranucleotide frequency, to identify discrete clusters of contigs that likely represent a population's genome. MaxBin2 and CONCOCT binning were performed with default parameters and 1000 bp minimum contig length. MetaBat binning was conducted with an interactive pipeline that applies decreasing levels of stringency within each successive iteration, collecting the best bins and their contigs each time. The remaining contigs are then placed into the next step with a lower stringency.

DASTool (version 1.1.0) [75] was applied to each set of binned contigs to create medium to high quality MAGs. This tool takes several different MAGs as input and identifies a consensus, non-redundant genome for each MAG, leading to higher quality genomes. Quality was determined for our MAGs according to Bowers et al. (2017). Medium quality drafts have completeness and contamination values of ≥ 50% and <10%, respectively whereas high quality MAGs have completeness and contamination values >90% and contamination <5%. Completeness and contamination were determined with CheckM [77]. MAG taxonomy was determined through phylogenetic analysis using a concatenated alignment of proteins from single copy marker genes included in the CheckM suite and a tree was built with FastTree [78].

Transcript mapping and annotation

Ribosomal sequences were identified and removed using SortmeRNA version 2.1 [79]. Filtered transcripts were mapped to MAGs using Bowtie2 version 2.3.4 [80] with sensitive local mapping, which requires neither end-to-end alignment nor mismatches when evaluating seed

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

substrings 20bp in length. Genomes were annotated with Prokka version 1.12 [81]. Mapping files, Prokka gene calls, and fasta files for each MAG were then used in the metagenomics pipeline of Anvi'o version 5.1 [82] for COG identification, visualization, and transcript abundance estimates. Transcript abundance is the coverage of the gene normalized by gene length. Transcripts per million was calculated by dividing transcript abundance by the number of mapped reads for each library and multiplying by 1,000,000.

COG annotations were computed using eggNOG-mapper [83] based on eggNOG 4.5 ontology data [84]. Spearman correlation coefficient analysis of changes in transcriptional richness associated with depth was performed in R [85]. Intersection analysis was performed with the UpsetR package in R [86]. TPM values for each *hspA* gene copy were log-transformed and normality was tested with a shapiro-wilks test in R. ANVOA (alpha = 0.05) was performed in R to facilitate post-hoc Tukey testing with the agricolae package (alpha = 0.05). Co-localization analysis was performed in CLC Genomics Workbench.

Amino acid identity analysis for each Woeseiales MAG was conducted with AAI-profiler [86]. This tool uses SANSparallel [88] to find homology matches within the Uniprot database (update Apr2020).

Data Access

Sequences for 16S rRNA gene libraries, transcriptomes, and metagenomes are deposited at NCBI under BioProject PRJNA493859. MAGs can be accessed from IMG with Gold Analysis Project IDs Ga0315927, Ga0315928, Ga031529, Ga0315930, and Ga0315931. R Markdown

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

553 files for Spearman calculations and depth-resolved transcript coverage plots for all individual
554 genes are deposited at GitHub (https://github.com/JBuongio/Woes).
555

556 RubisCO, rDSR, and hspA sequence analyses

557 Publicly available full length (> 480 aa) sequences for large and small chains of ribulose
558 bisphosphate carboxylase (RubisCO) were compared with MAG RubisCO sequences. Mega
559 version 7 [89] was used for Clustal alignments and neighbor-joining tree construction.
560

561 The alpha subunits of Woeseiales rDSR sequences were aligned in the *dsr*AB ARB
562 database [90]. RAxML was used within the CIPRES Science Gateway v. 3.3 server [91] and the
563 Interactive Tree of Life (iToL) server was used for tree visualization [92]. For spore protein
564 SP21, publicly available amino acid sequences for *hsp*A were acquired from environmental
565 samples and isolates and compared with Woeseiales sequences. RAxML was used within the
566 CIPRES Science Gateway v. 3.3 server [91] and the Interactive Tree of Life (iToL) server was
567 used for tree visualization [92].
568

569 Acknowledgements

570 This work was primarily supported by grants from the Simons Foundation (404586 to K.G.L.)
571 and the Austrian Science Fund (P29426-B29 to KW; P25111-B22 to AL), with additional
572 funding from a student research grant from the Explorer's Club (to J.B.) and the U.S.
573 Department of Energy, Office of Science, Office of Biological and Environmental Research,
574 Genomic Science Program under Award Number DE-SC0020369. Computing resources were
575 provided by the Center for Dark Energy Biosphere Investigations (C-DEBI, publication number

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

576 XXX). We thank Dr. B. Tully for assistance with computer programming. We also thank Captain
577 Stig Henningsen and first mate Reidar Sorensen of MS Farm. We thank the Alfred Wegener
578 Institute—Institute Paul Emile Victor (AWIPEV) station and staff for housing and excellent
579 logistics support. We thank all the participants of the 2016 Svalbard KOP 56/RiS 10528
580 expedition for help with sample collection.

581

582 References

583 1. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. Global distribution of
584 microbial abundance and biomass in subseafloor sediment. Proceedings of the National
585 Academies of Sciences. 2012;109(40):16213-6.
586 2. Cai W-J, Sayles FL. Oxygen penetration depths and fluxes in marine sediments. Marine
587 Chemistry. 1996;52(2):123-31.
588 3. Thamdrup B, Fossing H, Jørgensen BB. Manganese, iron and sulfur cycling in a coastal
589 marine sediment, Aarhus Bay, Denmark. Geochimica et Cosmochimica Acta. 1994;58(23):5115-
590 29.
591 4. Leloup J, Fossing H, Kohls K, Holmkvist L, Borowski C, Jørgensen BB.
592 Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity
593 related to geochemical zonation. Environ Microbiol. 2009;11(5):1278-91.
594 5. Hoehler TM, Jørgensen BB. Microbial life under extreme energy limitation. Nat Rev
595 Micro. 2013;11(2):83-94.
596 6. Jørgensen BB. Deep subseafloor microbial cells on physiological standby. Proceedings of
597 the National Academies of Sciences. 2011;108(45):18193-4.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

598 7. Mußmann M, Pjevac P, Krüger K, Dyksma S. Genomic repertoire of the
599 Woeseiaceae/JTB255, cosmopolitan and abundant core members of microbial communities in
600 marine sediments. The ISME Journal. 2017;11(5):1276.
601 8. Probandt D, Eickhorst T, Ellrott A, Amann R, Knittel K. Microbial life on a sand grain:
602 from bulk sediment to single grains. The ISME Journal. 2017;12:623.
603 9. Bowman JP, McCuaig RD. Biodiversity, Community Structural Shifts, and
604 Biogeography of Prokaryotes within Antarctic Continental Shelf Sediment. Applied and
605 Environmental Microbiology. 2003;69(5):2463-83.
606 10. Kessler AJ, Chen Y-J, Waite DW, Hutchinson T, Koh S, Popa ME, et al. Bacterial
607 fermentation and respiration processes are uncoupled in anoxic permeable sediments. Nature
608 Microbiology. 2019:1.
609 11. Fang X-M, Zhang T, Li J, Wang N-F, Wang Z, Yu L-Y. Bacterial community pattern
610 along the sediment seafloor of the Arctic fjorden (Kongsfjorden, Svalbard). Antonie van
611 Leeuwenhoek. 2019;112(8):1121-36.
612 12. Hoffmann K, Bienhold C, Buttigieg PL, Knittel K, Laso-Pérez R, Rapp JZ, et al.
613 Diversity and metabolism of Woeseiales bacteria, global members of marine sediment
614 communities. The ISME Journal. 2020;14(4):1042-56.
615 13. Dyksma S, Bischof K, Fuchs BM, Hoffmann K, Meier D, Meyerdierks A, et al.
616 Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. The ISME
617 Journal. 2016;10(8):1939.
618 14. Du Z-J, Wang Z-J, Zhao J-X, Chen G-J. *Woeseia oceani* gen. nov., sp. nov., a
619 chemoheterotrophic member of the order Chromatiales, and proposal of Woeseiaceae fam. nov.
620 International Journal of Systematic and Evolutionary Microbiology. 2016;66(1):107-12.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

621 15. Chen Y-J, Leung PM, Bay SK, Hugenholtz P, Kessler AJ, Shelley G, et al. Metabolic
622 flexibility allows generalist bacteria to become dominant in a frequently disturbed ecosystem.
623 BioRxiv [Preprint]. 2020. bioRxiv 945220 [posted Feb 13; cited June 1]. Available from:
624 https://doi.org/10.1101/2020.02.12.945220
625 16. Hop H, Pearson T, Hegseth EN, Kovacs KM, Wiencke C, Kwasniewski S, et al. The
626 marine ecosystem of Kongsfjorden, Svalbard. Polar Research. 2002;21(1):167-208.
627 17. Rinke A, Maturilli M, Graham RM, Matthes H, Handorf D, Cohen L, et al. Extreme
628 cyclone events in the Arctic: Wintertime variability and trends. Environmental Research Letters.
629 2017;12(9):094006.
630 18. Svendsen H, Beszczynska-Møller A, Hagen JO, Lefauconnier B, Tverberg V, Gerland S,
631 et al. The physical environment of Kongsfjorden–Krossfjorden, an Arctic fjord system in
632 Svalbard. Polar Research. 2002;21(1):133-66.
633 19. Elverhøi A, Lønne Ø, Seland R. Glaciomarine sedimentation in a modern fjord
634 environment, Spitsbergen. Polar Research. 1983;1(2):127-50.
635 20. Powell RD. A model for sedimentation by tidewater glaciers. Annals of Glaciology.
636 1981;2:129-34.
637 21. Kempf P, Forwick M, Laberg JS, Vorren TO. Late Weichselian and Holocene
638 sedimentary palaeoenvironment and glacial activity in the high-arctic Van Keulenfjorden,
639 Spitsbergen. The Holocene. 2013;23(11):1607-18.
640 22. Buongiorno J, Herbert L, Wehrmann L, Michaud A, Laufer K, Røy H, et al. Complex
641 microbial communities drive iron and sulfur cycling in Arctic fjord sediments. Applied and
642 Environmental Microbiology. 2019;85(14):00949-19.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

643 23. Wehrmann LM, Riedinger N, Brunner B, Kamyshny Jr A, Hubert CR, Herbert LC, et al.
644 Iron-controlled oxidative sulfur cycling recorded in the distribution and isotopic composition of
645 sulfur species in glacially influenced fjord sediments of west Svalbard. Chemical Geology.
646 2017;466:678-95.
647 24. He J, Xu Z, Hughes J. Pre-lysis washing improves DNA extraction from a forest soil.
648 Soil Biology and Biochemistry. 2005;37(12):2337-41.
649 25. Wehrmann LM, Formolo MJ, Owens JD, Raiswell R, Ferdelman TG, Riedinger N, et al.
650 Iron and manganese speciation and cycling in glacially influenced high-latitude fjord sediments
651 (West Spitsbergen, Svalbard): Evidence for a benthic recycling-transport mechanism. Geochim
652 Cosmochim Acta. 2014;141:628-55.
653 26. Graham E, Heidelberg J, Tully B. Potential for primary productivity in a globally-
654 distributed bacterial phototroph. The ISME Journal. 2018;12(7):1861-1866.
655 27. Biegel E, Schmidt S, González JM, Müller V. Biochemistry, evolution and physiological
656 function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes.
657 Cellular and Molecular Life Sciences. 2011;68(4):613-34.
658 28. Koo MS, Lee JH, Rah SY, Yeo WS, Lee JW, Lee KL, et al. A reducing system of the
659 superoxide sensor SoxR in Escherichia coli. The EMBO journal. 2003;22(11):2614-22.
660 29. Korch SB, Henderson TA, Hill TM. Characterization of the hipA7 allele of Escherichia
661 coli and evidence that high persistence is governed by (p)ppGpp synthesis. Molecular
662 Microbiology. 2003;50(4):1199-213.
663 30. Mogk A, Schlieker C, Strub C, Rist W, Weibezahn J, Bukau B. Roles of individual
664 domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP
665 hydrolysis, and chaperone activity. Journal of Biological Chemistry. 2003;278(20):17615-24.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

666 31. Guariglia-Oropeza V, Helmann JD. Bacillus subtilis σV confers lysozyme resistance by
667 activation of two cell wall modification pathways, peptidoglycan O-acetylation and D-
668 alanylation of teichoic acids. Journal of Bacteriology. 2011;193(22):6223-32.
669 32. Jørgensen BB, Glud RN, Holby O. Oxygen distribution and bioirrigation in Arctic fjord
670 sediments (Svalbard, Barents Sea). Marine Ecology Progress Series. 2005;292:85-95.
671 33. Gihring TM, Lavik G, Kuypers MMM, Kostka JE. Direct determination of nitrogen
672 cycling rates and pathways in Arctic fjord sediments (Svalbard, Norway). Limnology and
673 Oceanography. 2010;55(2):740-52.
674 34. Kostka JE, Thamdrup B, Glud RN, Canfield DE. Rates and pathways of carbon oxidation
675 in permanently cold Arctic sediments. Marine Ecology Progress Series. 1999:7-21.
676 35. Lünsdorf H, Schairer HU, Heidelbach M. Localization of the stress protein SP21 in
677 indole-induced spores, fruiting bodies, and heat-shocked cells of *Stigmatella aurantiaca*. Journal
678 of Bacteriology. 1995;177(24):7092-9.
679 36. Heidelbach M, Skladny H, Schairer HU. Purification and characterization of SP21, a
680 development-specific protein of the myxobacterium *Stigmatella aurantiaca*. Journal of
681 Bacteriology. 1993;175(3):905-8.
682 37. Gennaris A, Ezraty B, Henry C, Agrebi R, Vergnes A, Oheix E, et al. Repairing oxidized
683 proteins in the bacterial envelope using respiratory chain electrons. Nature. 2015;528(7582):409.
684 38. Stone BJ, Miller VL. Salmonella enteritidis has a homologue of *tolC* that is required for
685 virulence in BALB/c mice. Molecular Microbiology. 1995;17(4):701-12.
686 39. Durand É, Bernadac A, Ball G, Lazdunski A, Sturgis JN, Filloux A. Type II protein
687 secretion in *Pseudomonas aeruginosa*: the pseudopilus is a multifibrillar and adhesive structure.
688 Journal of Bacteriology. 2003;185(9):2749-58.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

689 40. Lisabeth EM, Falivelli G, Pasquale EB. Eph Receptor Signaling and Ephrins. Cold
690 Spring Harbor Perspectives in Biology. 2013;5(9).
691 41. Waller P, Sauer RT. Characterization of degQ and degS, Escherichia coli genes encoding
692 homologs of the DegP protease. Journal of Bacteriology. 1996;178(4):1146-53.
693 42. Chatterjee T, Mukherjee D, Banerjee M, Chatterjee BK, Chakrabarti P. Crystal structure
694 and activity of protein L-isoaspartyl-O-methyltransferase from Vibrio cholerae, and the effect of
695 AdoHcy binding. Archives of Biochemistry and Biophysics. 2015;583:140-9.
696 43. Mesak LR, Mesak FM, Dahl MK. Expression of a novel gene, gluP, is essential for
697 normal Bacillus subtilis cell division and contributes to glucose export. BMC microbiology.
698 2004;4(1):13.
699 44. Rittershaus ES, Baek S-H, Sassetti CM. The normalcy of dormancy: common themes in
700 microbial quiescence. Cell Host & Microbe. 2013;13(6):643-51.
701 45. Heidelbach M, Skladny H, Schairer HU. Heat shock and development induce synthesis of
702 a low-molecular-weight stress-responsive protein in the myxobacterium Stigmatella aurantiaca.
703 Journal of Bacteriology. 1993;175(22):7479-82.
704 46. Hubert C, Arnosti C, Brüchert V, Loy A, Vandieken V, Jørgensen BB. Thermophilic
705 anaerobes in Arctic marine sediments induced to mineralize complex organic matter at high
706 temperature. Environmental Microbiology. 2010;12(4):1089-104.
707 47. Korza G, Camilleri E, Green J, Robinson J, Nagler K, Moeller R, et al. Analysis of the
708 mRNAs in Spores of Bacillus subtilis. Journal of Bacteriology. 2019;201(9):e00007-19.
709 48. Cowan AE, Koppel DE, Setlow B, Setlow P. A soluble protein is immobile in dormant
710 spores of Bacillus subtilis but is mobile in germinated spores: implications for spore dormancy.
711 Proceedings of the National Academies of Sciences. 2003;100(7):4209-14.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

712 49. Bobek J, Šmídová K, Čihák M. A waking review: old and novel insights into the spore
713 germination in Streptomyces. Frontiers in Microbiology. 2017;8:2205.
714 50. Mikulík K, Palečková P, Felsberg J, Bobek J, Zídková J, Halada P. SsrA genes of
715 streptomycetes and association of proteins to the tmRNA during development and cellular
716 differentiation. Proteomics. 2008;8(7):1429-41.
717 51. Rustad TR, Minch KJ, Brabant W, Winkler JK, Reiss DJ, Baliga NS, et al. Global
718 analysis of mRNA stability in Mycobacterium tuberculosis. Nucleic acids research.
719 2013;41(1):509-17.
720 52. Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K. Evaluation of a nutrient
721 starvation model of Mycobacterium tuberculosis persistence by gene and protein expression
722 profiling. Molecular Microbiology. 2002;43(3):717-31.
723 53. Lewis D, Gattie D. The ecology of quiescent microbes. ASM American Society for
724 Microbiology News. 1991;57(1):27-32.
725 54. Shaikh AS, Tang YJ, Mukhopadhyay A, Martín HG, Gin J, Benke PI, et al. Study of
726 stationary phase metabolism via isotopomer analysis of amino acids from an isolated protein.
727 Biotechnology Progress. 2010;26(1):52-6.
728 55. Sherrid AM, Rustad TR, Cangelosi GA, Sherman DR. Characterization of a Clp protease
729 gene regulator and the reaeration response in Mycobacterium tuberculosis. PLoS One. 2010;5(7).
730 56. Choder M. A general topoisomerase I-dependent transcriptional repression in the
731 stationary phase in yeast. Genes & Development. 1991;5(12a):2315-26.
732 57. Bradley JA, Amend JP, LaRowe DE. Survival of the fewest: Microbial dormancy and
733 maintenance in marine sediments through deep time. Geobiology. 2019;17(1):43-59.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

734 58. Lennon JT, Jones SE. Microbial seed banks: the ecological and evolutionary implications
735 of dormancy. Nature Reviews Microbiology. 2011;9(2):119-30.
736 59. Chen X, Andersen TJ, Morono Y, Inagaki F, Jørgensen BB, Lever MA. Bioturbation as a
737 key driver behind the dominance of Bacteria over Archaea in near-surface sediment. Scientific
738 Reports. 2017;7(1):1-14.
739 60. Kanneworff E, Nicolaisen W. The "Haps" a frame-supported bottom corer. Ophelia.
740 1972;10(2):119-28.
741 61. Hoehler TM, Alperin MJ, Albert DB, Martens CS. Field and laboratory studies of
742 methane oxidation in an anoxic marine sediment: Evidence for a methanogen-sulfate reducer
743 consortium. Global Biogeochemical Cycles. 1994;8(4):451-63.
744 62. Harris D, Horwáth WR, Van Kessel C. Acid fumigation of soils to remove carbonates
745 prior to total organic carbon or carbon-13 isotopic analysis. Soil Science Society of America
746 Journal. 2001;65(6):1853-6.
747 63. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et
748 al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.
749 Proceedings of the National Academies of Sciences. 2011;108:4516-22.
750 64. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
751 data. Bioinformatics. 2014;30(15):2114-20.
752 65. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al.
753 Introducing mothur: open-source, platform-independent, community-supported software for
754 describing and comparing microbial communities. Applied and Environmental Microbiology.
755 2009;75(23):7537-41.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

66. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC research notes. 2016;9(1):88.

67. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.

68. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Research. 2017;27(5):824-34.

69. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nature Biotechnology. 2018;36:566.

70. Peng Y, Leung HC, Yiu S-M, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28(11):1420-8.

71. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6.

72. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2015;32(4):605-7.

73. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nature Methods. 2014;11(11):1144.

74. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

777 75. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of
778 genomes from metagenomes via a dereplication, aggregation, and scoring strategy. Nature
779 Microbiology. 2018; 3:836-843.
780 76. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy T, et al.
781 Minimum information about a single amplified genome (MISAG) and a metagenome-assembled
782 genome (MIMAG) of bacteria and archaea. Nature Biotechnology. 2017;35(8):725.
783 77. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the
784 quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome
785 Research. 2015;25(7):1043-55.
786 78. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for
787 large alignments. PloS One. 2010;5(3):e9490.
788 79. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal
789 RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7.
790 80. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods.
791 2012;9(4):357-9.
792 81. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014:btu153.
793 82. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an
794 advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.
795 83. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al.
796 Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper.
797 Molecular Biology and Evolution. 2017;34(8):2115-22.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

798 84. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al.
799 eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for
800 eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research. 2015;44(D1):D286-D93.
801 85. R CoreTeam. R: A Language and Environment for Statistical Computing. R Foundation
802 for Statistical Computing, Vienna, Austria. 2013.
803 86. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of
804 intersecting sets and their properties. Bioinformatics. 2017;33(18):2938-40.
805 87. Medlar AJ, Törönen P, Holm L. AAI-profiler: fast proteome-wide exploratory analysis
806 reveals taxonomic identity, misclassification and contamination. Nucleic Acids Research.
807 2018;46(W1):W479-W85.
808 88. Somervuo P, Holm L. SANSparallel: interactive homology search against Uniprot.
809 Nucleic Acids Research. 2015;43(W1):W24-W9.
810 89. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis
811 version 7.0 for bigger datasets. Molecular Biology and Evolution. 2016;33(7):1870-1874.
812 90. Müller AL, Kjeldsen KU, Rattei T, Pester M, Loy A. Phylogenetic and environmental
813 diversity of DsrAB-type dissimilatory (bi) sulfite reductases. The ISME Journal. 2015;9(5):1152.
814 91. Miller MA, Pfeiffer W, Schwartz T, editors. Creating the CIPRES Science Gateway for
815 inference of large phylogenetic trees. 2010 gateway computing environments workshop (GCE);
816 2010: 1-8
817 92. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
818 developments. Nucleic Acids Research. 2019;47(W1)W256-W259.
819

820 # Supporting information

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

821 **Figure S1: Phylogenetic tree for Woeseiales MAGs placement.** Maximum likelihood was
822 calculated in Mega v. 7 with 1000 bootstraps on a concatenated alignment of 49 ribosomal
823 proteins. Only nodes with >75% support are shown. MAGs from this study are indicated with
824 blue circles. Tree was rooted using Roseobacter species as the outgroup.
825
826 **Figure S2: Comparison of genomic contents across the 5 Woeseiales MAGs to show**
827 **similarity**. Figure generated with KEGG-decoder [26].
828
829 **Figure S3: Amino acid identity analysis.** Scatterplots, Krona diagrams, and histograms relating
830 amino acid identity to the cultured isolate, *Woeseia oceani*, are presented for the MAGs Woeseia
831 sp. stnAB (a, b, c), Woeseia sp. stnAC (d, e, f), Woeseia sp. stnAC-2 (g, h, i), Woeseia sp. stnF
832 (j, k, l), and Woeseia sp. stnF-2 (m, n, o).
833
834 **Figure S4: Intersection analysis of all transcripts.** Set intersections of transcripts visualized
835 with UpsetR [86]. The total number transcribed genes ("set size") is indicated as horizontal bars.
836 Vertical bars represent the intersection size, or number of transcribed genes that fall into the 15
837 different sets of site combinations. Depth analysis of transcript abundance for unique sets of
838 transcripts are depicted in Fig S5.
839
840 **Figure S5: Recruitment of unique transcripts.** Transcriptional coverage (transcript abundance
841 normalized to gene length) is reported for the unique transcripts detected at each site for the
842 COGs Cellular Processing and Signaling, Information Storage and Processing, Metabolism, and
843 Function Unknown. Those with predicted function only were removed from this analysis. Dot

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

844 size scales with metagenome coverage and color saturation scales from lighter to darker blue for
845 increased transcriptional coverage. Site P did not have a metagenome.
846
847 **Figure S6: Neighbor-joining tree of RubisCO sequences.** Only nodes with >80% support after
848 1000 bootstraps are shown.
849
850 **Figure S7: Coverage of all Woeseiales MAG transcripts.** The transcripts per million (TPM)
851 value is reported for each transcribed gene (along x-axis). Names of genes are not included to
852 ease illustration and interpretation. COG assignment of genes is represented by color. The gene
853 encoding for spore protein SP21 is indicated with arrows.
854
855 **Figure S8: Maximum likelihood tree of *hsp*A genes encoding for spore protein SP21**.
856 Bootstrap support after 1000 bootstraps is reported for nodes with > 70% support. Woeseiales
857 sequences are indicated in red, cultured microorganisms in purple, and experimentally-verified
858 spore protein SP21 indicated in green [36, 45]. Archaeal sequences of spore protein SP21 were
859 used as the outgroup. The scale bar represents the number of substitutions per site.
860
861 **Figure S9: Coverage with depth for transcripts of stress proteins, chaperones, and**
862 **transcriptional regulators.** TPM values reported for genes associated with starvation response,
863 stress mitigation, protein repair/folding, and transcriptional regulation. Yellow border labels
864 indicate genes that co-occur on the same contigs as *hsp*A (encoding spore protein SP21; Figure
865 S9).
866

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

867 **Figure S10: Co-localization of genes on the same contig as the gene for spore protein SP21**
868 **(*hsp*A), highlighted with orange rectangle**. Visualizations from CLC Genomics Workbench.

869
870 **Table S1: Genome statistics for the Woeseiales MAGs in this study.** Completeness and
871 contamination were determined for each genome with CheckM.

872

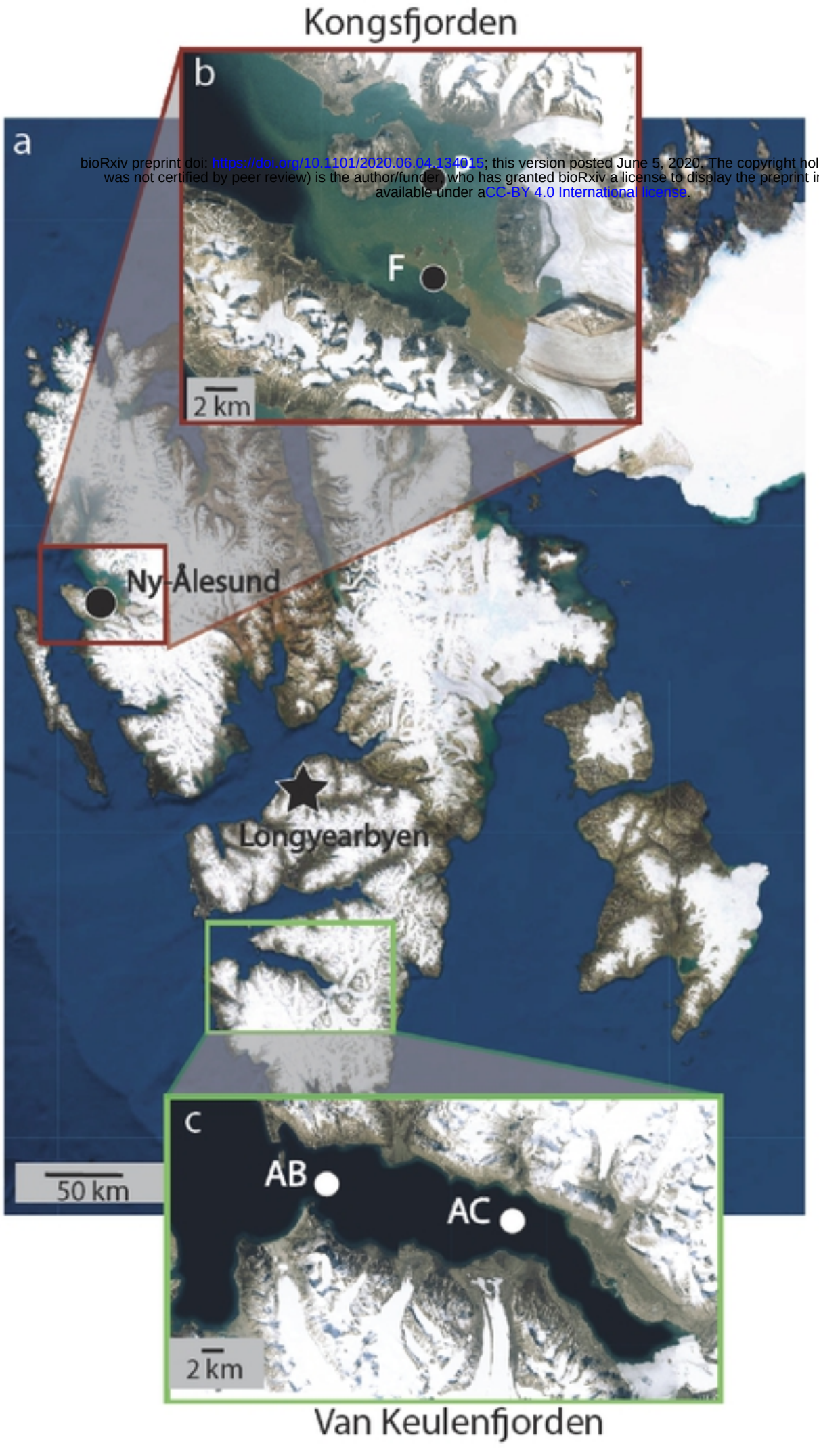
873 **Table S2: Genes with transcripts detected at all sites. C** = Energy production and conversion;
874 **D** = Cell cycle control, cell division, chromosome partitioning; **E** = Amino acid transport and
875 metabolism; **F** = Nucleotide transport and metabolism; **G** = Carbohydrate transport and
876 metabolism; **H** = Coenzyme transport and metabolism; **I** = Lipid transport and metabolism; **J** =
877 Translation, ribosomal structure and biogenesis; **K** = Transcription; **L** = Replication,
878 recombination and repair; **N** = Cell motility; **O** = Post-translational modification, protein
879 turnover and chaperones; **P** = Inorganic ion transport and metabolism; **Q** = Secondary
880 metabolites biosynthesis, transport and catabolism; **S** = Function unknown; **T** = Signal
881 transduction mechanisms; **U** = Intracellular trafficking, secretion and vesicular transport; **V** =
882 Defense mechanisms.

883

884 File S1: **Statistics for genes included in Figure 3a.** Presence of genes and their transcripts were
885 obtained through manual inspection of Ghost Koala annotations and Prokka annotations. Black
886 boxes indicate no transcriptome is available, gray indicates no detection, and yellow indicates
887 detection.

888

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.04.134015; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



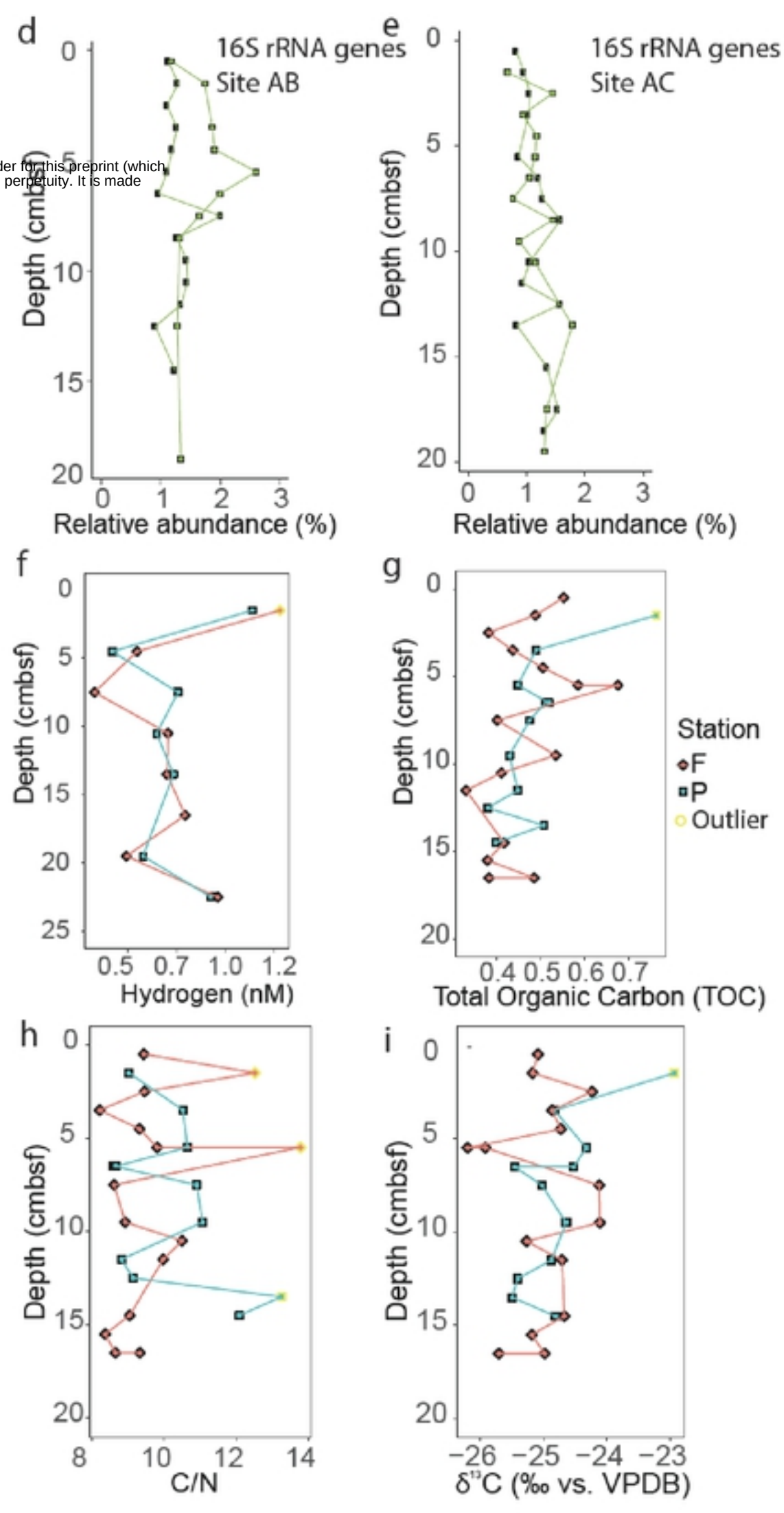


Figure 1

Uncultured bacterium ws101A12 JQ256780 JQ256780 Oxidative bacterial Marine
Rhodomicrobium vannielii RhmVann6 CP002292 Oxidative bacterial Unknown
uncultured marine FM246807 FM246807 Oxidative bacterial Marine
uncultured marine FM246801 FM246801 Oxidative bacterial Marine
uncultured marine FM246804 FM246804 Oxidative bacterial Marine
uncultured bacterium HQ191188 HQ191188 Oxidative bacterial Marine
uncultured bacterium HQ191181 HQ191181 Oxidative bacterial Marine
uncultured bacterium HQ191214 HQ191214 Oxidative bacterial Marine
uncultured sulfate-reducing GU181329 GU181329 Reductive bacterial Soil
uncultured sulfate-reducing HQ657192 HQ657192 Reductive bacterial Soil
Desulfovibrio cuneatus DslCune3 AB061537 Reductive bacterial Freshwater
uncultured Reductive bacterial Industrial (5)
Desulfovibrio vulgaris (15)
Desulfovibrio sp. (13)

Figure 2

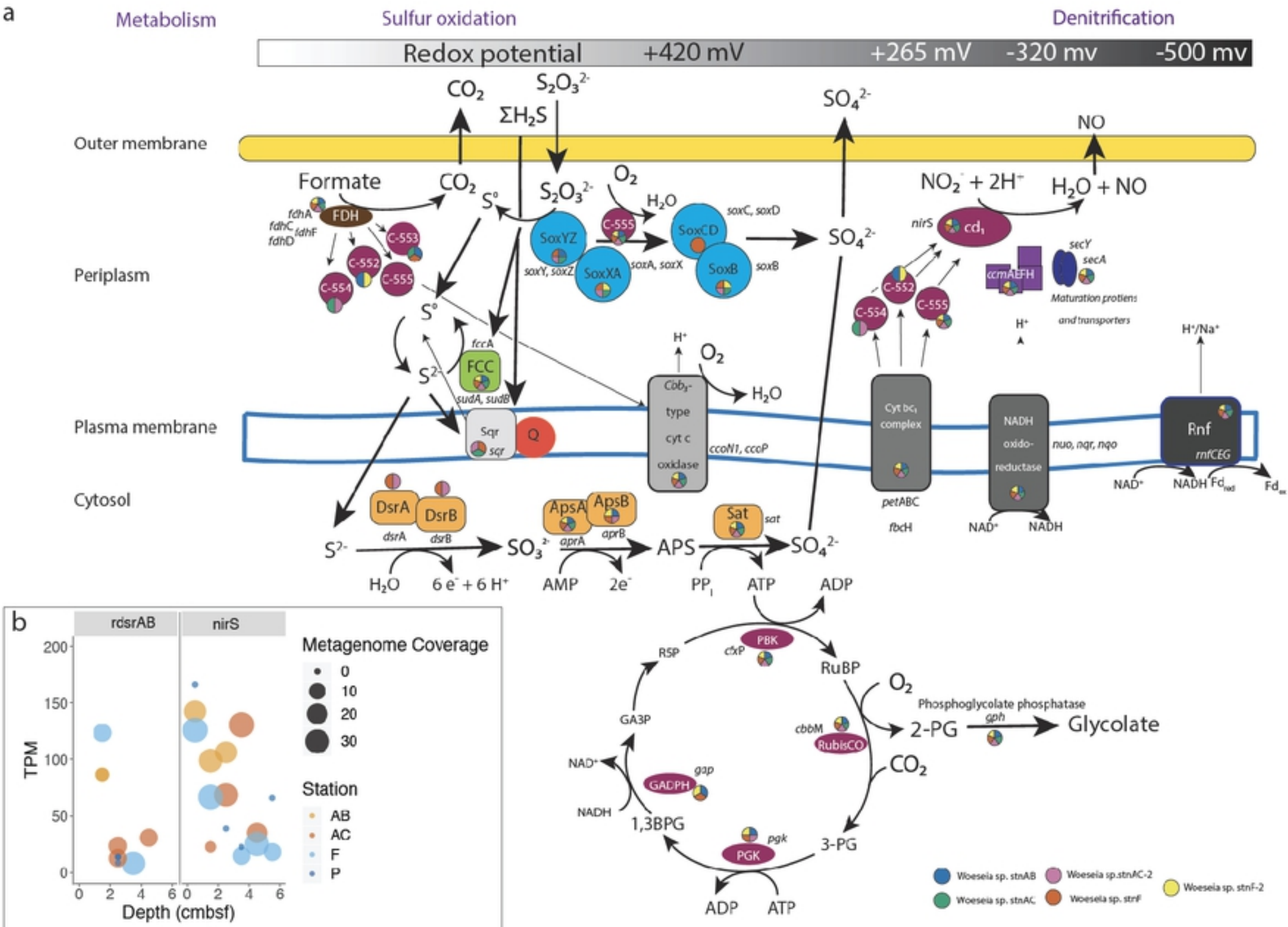

a
Metabolism
Sulfur oxidation
Denitrification
Redox potential
+420 mV
+265 mV
-320 mv
-500 mv
Outer membrane
Periplasm
Plasma membrane
Cytosol
Formate
FDH
fdhA
fdhC
fdhD
fdhF
C-553
C-552
C-554
C-555
SoxYZ
SoxXA
soxY, soxZ
soxA, soxX
SoxCD
SoxB
soxC, soxD
soxB
FCC
fccA
sudA, sudB
Sqr
sqr
Q
Cbb3-type cyt c oxidase
ccoN1, ccoP
Cyt bc1 complex
petABC
fbcH
NADH oxido-reductase
nuo, ngr, nqo
Rnf
rnfCEG
nirS
cd1
ccmAEFH
secY
secA
Maturation protiens
and transporters
DsrA
DsrB
dsrA
dsrB
ApsA
ApsB
aprA
aprB
Sat
sat
APS
AMP
2e-
PPi
ATP
ADP
6 e- + 6 H+
RuBP
PBK
cfxP
RSP
GA3P
NAD+
NADH
GAPDH
gap
1,3BPG
PGK
pgk
3-PG
RubisCO
cbbM
2-PG
Phosphoglycolate phosphatase
gph
Glycolate
Woeseia sp. stnAB
Woeseia sp. stnAC
Woeseia sp.stnAC-2
Woeseia sp. stnF
Woeseia sp. stnF-2
b
rdsrAB
nirS
TPM
Depth (cmbsf)
Metagenome Coverage
Station
AB
AC
F
P

Figure 3

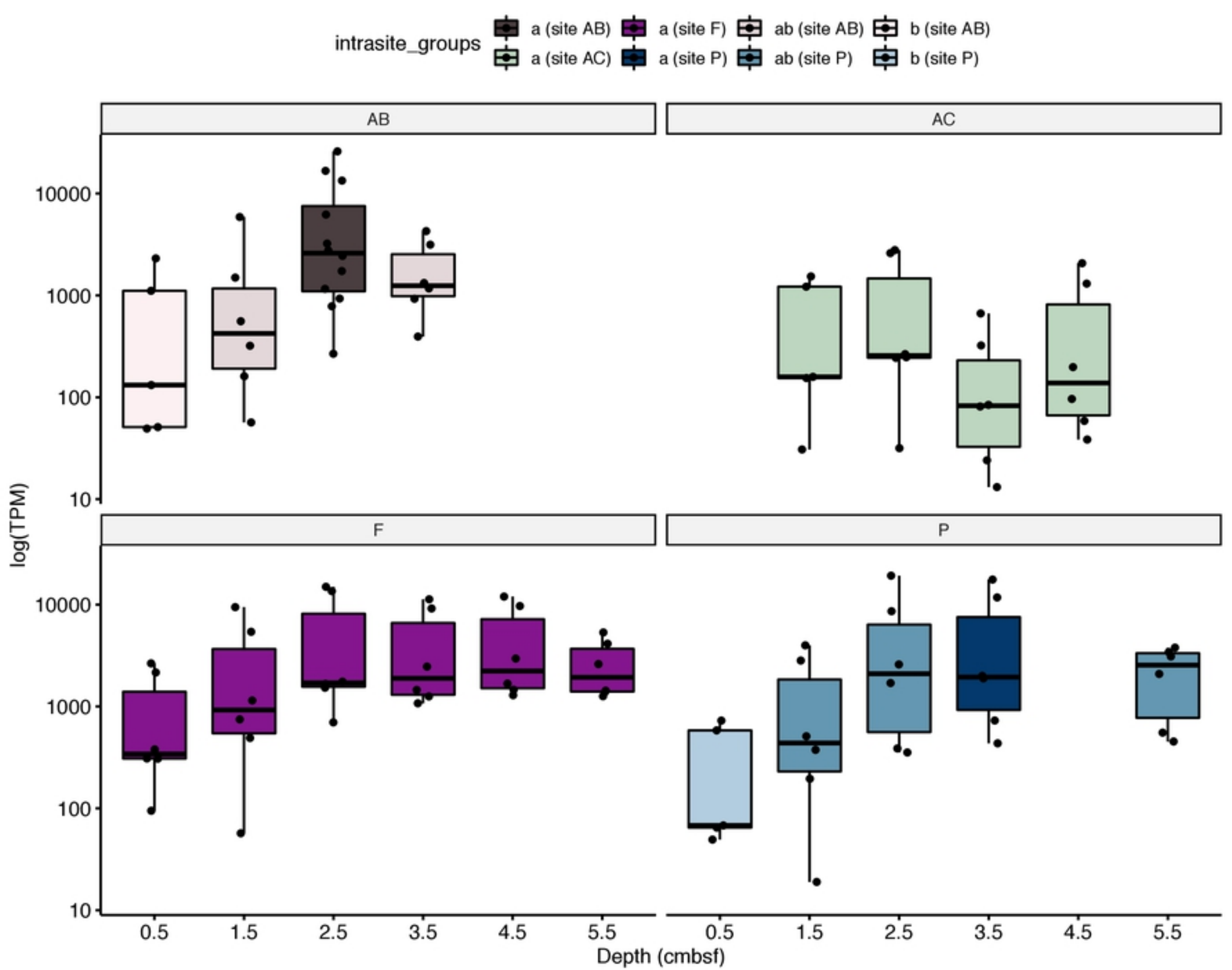

intrasite_groups
a (site AB)
a (site AC)
a (site F)
a (site P)
ab (site AB)
ab (site P)
b (site AB)
b (site P)
AB
AC
F
P
log(TPM)
10000
1000
100
10
0.5
1.5
2.5
3.5
4.5
5.5
Depth (cmbsf)

Figure 4

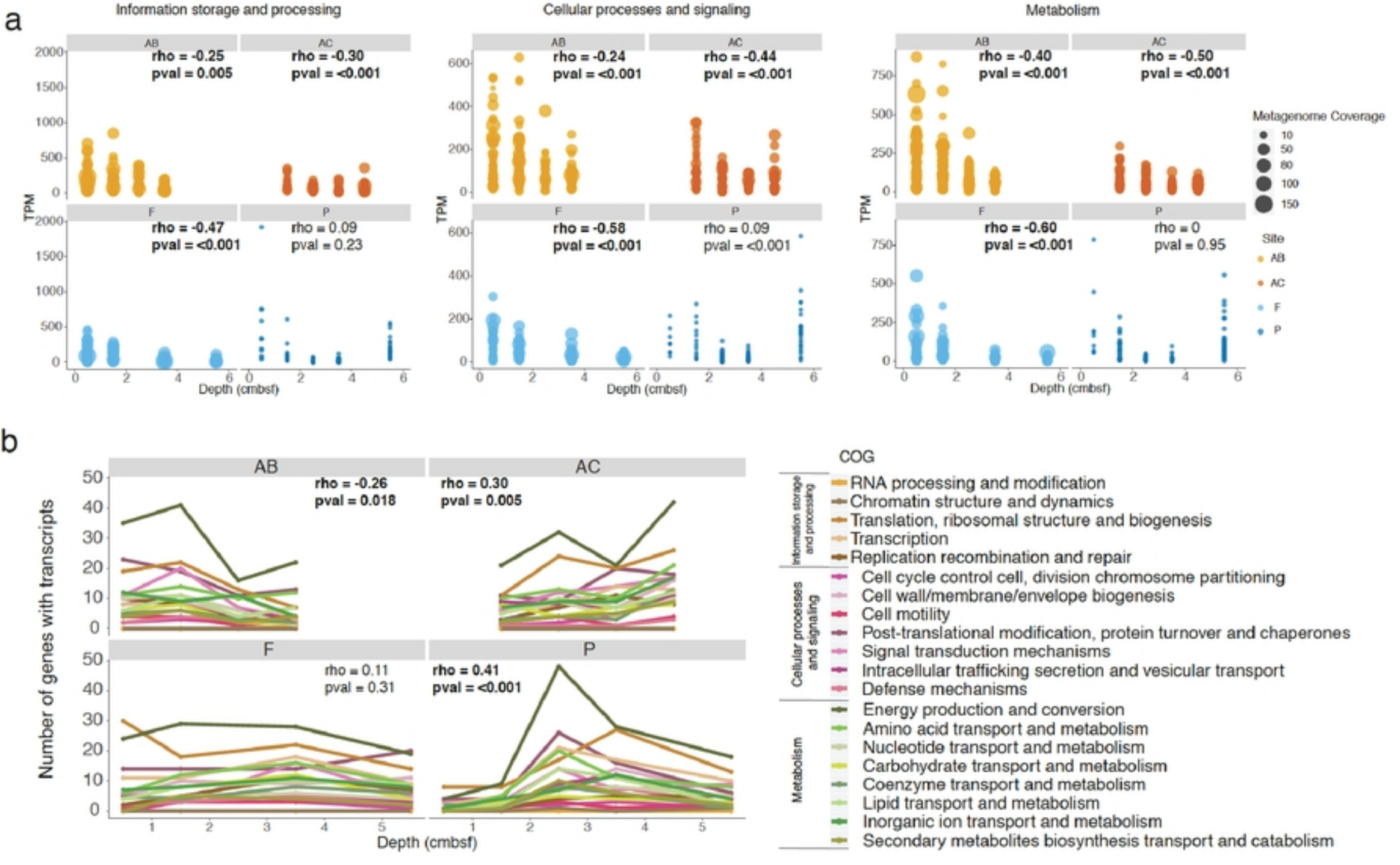

a
Information storage and processing
Cellular processes and signaling
Metabolism
AB
AC
F
P
rho = -0.25
pval = 0.005
rho = -0.30
pval = <0.001
rho = -0.24
pval = <0.001
rho = -0.44
pval = <0.001
rho = -0.40
pval = <0.001
rho = -0.50
pval = <0.001
rho = -0.47
pval = <0.001
rho = 0.09
pval = 0.23
rho = -0.58
pval = <0.001
rho = 0.09
pval = <0.001
rho = -0.60
pval = <0.001
rho = 0
pval = 0.95
TPM
Depth (cmbsf)
Metagenome Coverage
10
50
80
100
150
Site
AB
AC
F
P
b
AB
AC
F
P
rho = -0.26
pval = 0.018
rho = 0.30
pval = 0.005
rho = 0.11
pval = 0.31
rho = 0.41
pval = <0.001
Number of genes with transcripts
Depth (cmbsf)
COG
Information storage and processing
RNA processing and modification
Chromatin structure and dynamics
Translation, ribosomal structure and biogenesis
Transcription
Replication recombination and repair
Cellular processes and signaling
Cell cycle control cell, division chromosome partitioning
Cell wall/membrane/envelope biogenesis
Cell motility
Post-translational modification, protein turnover and chaperones
Signal transduction mechanisms
Intracellular trafficking secretion and vesicular transport
Defense mechanisms
Metabolism
Energy production and conversion
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis transport and catabolism

Figure 5