Zetaproteobacteria Pan-Genome Reveals Candidate Gene Cluster for Twisted Stalk Biosynthesis and Export

Elif Koeksoy[1,2], Oliver M. Bezuidt[1], Timm Bayer[3], Clara S. Chan[4,5] and David Emerson[1]*

[1] Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States, [2] Leibniz Institute DSMZ (German Collection of Microorganisms and Cell Cultures), Braunschweig, Germany, [3] Geomicrobiology Group, Center for Applied Geoscience, University of Tübingen, Tübingen, Germany, [4] Department of Earth Sciences, University of Delaware, Newark, DE, United States, [5] School of Marine Sciences and Policy, University of Delaware, Newark, DE, United States

Twisted stalks are morphologically unique bacterial extracellular organo-metallic structures containing Fe(III) oxyhydroxides that are produced by microaerophilic Fe(II)-oxidizers belonging to the Betaproteobacteria and Zetaproteobacteria. Understanding the underlying genetic and physiological mechanisms of stalk formation is of great interest based on their potential as novel biogenic nanomaterials and their relevance as putative biomarkers for microbial Fe(II) oxidation on ancient Earth. Despite the recognition of these special biominerals for over 150 years, the genetic foundation for the stalk phenotype has remained unresolved. Here we present a candidate gene cluster for the biosynthesis and secretion of the stalk organic matrix that we identified with a trait-based analyses of a pan-genome comprising 16 Zetaproteobacteria isolate genomes. The "**s**talk **f**ormation in **Z**etaproteobacteria" (sfz) cluster comprises six genes (sfz1-sfz6), of which sfz1 and sfz2 were predicted with functions in exopolysaccharide synthesis, regulation, and export, sfz4 and sfz6 with functions in cell wall synthesis manipulation and carbohydrate hydrolysis, and sfz3 and sfz5 with unknown functions. The stalk-forming Betaproteobacteria *Ferriphaselus* R-1 and OYT-1, as well as dread-forming Zetaproteobacteria *Mariprofundus aestuarium* CP-5 and *Mariprofundus ferrinatatus* CP-8 contain distant sfz gene homologs, whereas stalk-less Zetaproteobacteria and Betaproteobacteria lack the entire gene cluster. Our pan-genome analysis further revealed a significant enrichment of clusters of orthologous groups (COGs) across all Zetaproteobacteria isolate genomes that are associated with the regulation of a switch between sessile and motile growth controlled by the intracellular signaling molecule c-di-GMP. Potential interactions between stalk-former unique transcription factor genes, sfz genes, and c-di-GMP point toward a c-di-GMP regulated surface attachment function of stalks during sessile growth.

Keywords: zetaproteobacteria, Fe-biosignatures, biomineralization, twisted stalks, pan-genome, microaerophilic Fe(II) oxidation

INTRODUCTION

Twisted stalks are organo-mineral composites that were first recognized in the early nineteenth century for their distinctive ribbon-shaped morphology (Ehrenberg, 1836). They are formed by neutrophilic microaerophilic chemolithoautotrophs that obtain energy from the enzymatic oxidation of Fe(II) to Fe(III) with oxygen (O_2) as the terminal electron acceptor (Emerson and Revsbech, 1994; Emerson and Moyer, 1997), and that classify into the Betaproteobacteria (Kato et al., 2013, 2015) and Zetaproteobacteria (Emerson et al., 2007). Only four isolates of the Gallionellaceae family within the ecologically and physiologically diverse Betaproteobacteria have been characterized as freshwater Fe(II)-oxidizers to date, with two of them lacking the stalk phenotype (Emerson et al., 2013; Kato et al., 2015). Contrastingly, all previously reported cultured Zetaproteobacteria are characterized as marine Fe(II)-oxidizers (Emerson et al., 2007), with two isolates reported as being additionally capable of H_2 oxidation (Mori et al., 2017). The full metabolic potential of the Zetaproteobacteria is yet to be uncovered (Field et al., 2015; Brooks and Field, 2020) and so are the evolutionary grounds for the rise of the stalk formation phenotype in two distinct Proteobacteria classes.

Their requirement for O_2 as terminal electron acceptor limits microaerophilic Fe(II)-oxidizers to low-O_2 environmental niches (typically $< 50\ \mu M$) where they can outcompete the kinetics of chemical Fe(II) oxidation with ambient O_2 (Rentz et al., 2007; Druschel et al., 2008; McAllister et al., 2019). Contrary to their specialized metabolic demands, both culture-dependent and culture-independent approaches revealed that Zetaproteobacteria occupy a diverse range of habitats, including marine hydrothermal vents and submarine volcanoes (Scott et al., 2015; Emerson et al., 2017; Vander Roost et al., 2017), subtropical coastal catchments (Lin et al., 2012), continental margins (Rubin-Blum et al., 2014), coastal bays (Mumford et al., 2016), estuarine water columns (Chiu et al., 2017; Garrison et al., 2019), marine sediments (Laufer et al., 2017; Beam et al., 2018), and worm burrow openings in coastal sediments (McAllister et al., 2015; Beam et al., 2020). The source environments of neutrophilic Fe(II)-oxidizing Betaproteobacteria were primarily reported as ferruginous groundwater discharge points (Emerson et al., 2013; Kato et al., 2015), while metagenomic data also indicates their presence in marine environments such as the Arctic Mid-Ocean Ridge, where they co-occur with Zetaproteobacteria (Vander Roost et al., 2017).

Stalk formation has been recognized as a characteristic morphological trait of Fe(II)-oxidizing bacteria since the 1800s (Ehrenberg, 1836). It has only been relatively recently that successful isolation of Fe(II)-oxidizing Zeta- and Betaproteobacteria has enabled detailed morphological and chemical characterization of their twisted stalk products. For instance, time lapse microscopy of the Zetaproteobacterium *Mariprofundus ferrooxydans* PV-1 revealed single cells excreting multiple nanometer-thin fibers at the cell concavity that elongate at a rate of 2.2 $\mu m\ h^{-1}$ as cells grow and oxidize Fe(II). A contemporaneous rotation of the cell was suggested to cause the helical morphology of stalks (Chan et al., 2011).

A combination of spectroscopy and microscopy approaches applied on stalks of isolate cultures and in environmental Fe-mat samples resolved the stalk nanofibrils to consist of acidic polysaccharides with carboxylic functional groups (Chan et al., 2009, 2011). The organic matrix is proposed to adsorb Fe(III) originating from biotic Fe(II) oxidation (Chan et al., 2004), resulting in its encrustation with amorphous Fe(III) oxyhydroxide (FeOOH) during initial stalk growth. As the stalks age, amorphous FeOOH becomes metastable ferrihydrite which is overprinted with the FeOOH lepidocrocite (Chan et al., 2011). Based on these observations, stalks have been suggested to serve as biomarkers for primordial microaerophilic Fe(II) oxidation in the geologic record (Krepski et al., 2013; Picard et al., 2015; Chan et al., 2016; Dodd et al., 2017; Little et al., 2021), that may be of astrobiological interest as well.

Despite substantial progress on the chemical composition and nano-scale morphology of stalks, the underlying genetic machinery of this unique bacterial product remains unresolved. Certain Fe(II)-oxidizing Beta- and Zetaproteobacteria, i.e., *Sideroxydans lithotrophicus* ES-1, *Gallionella capsiferriformans* ES-2, and *Ghiorsea bivora* TAG-1 and SV-108, were reported not to form twisted stalks but to produce amorphous Fe(III) oxyhydroxides instead (Emerson et al., 2013; Mori et al., 2017). Other Zetaproteobacteria isolates, i.e., *Mariprofundus aestuarium* CP-5 and *Mariprofundus ferrinatatus* CP-8, were demonstrated to excrete shorter and thicker exopolymers encrusted in Fe(III) oxyhydroxides that differ morphologically from stalks, known as "dread"-biominerals due to their unique morphology (Chiu et al., 2017). Collectively, these observations imply that stalks are not essential for the survival of Fe(II)-oxidizing bacteria. However, since stalk-biosynthesis must incur some bioenergetic cost, it is logical that they confer physiological and/or ecological advantages over Fe(II)-oxidizers that form amorphous Fe(III) oxides. Proposed functions include anchoring of the cell in environments with favorable gradients of Fe(II) and O_2, and prevention of cell encrustation in Fe(III) through templating its sorption to an organic matrix (Chan et al., 2011).

Identifying the genes that encode stalk formation is a vital step in deciphering the evolutionary and physiological relevance of this unique bacterial trait. The organic nature of the stalk fibrils together with their secretion from the cell surface suggest an interaction of multiple genes with different functions, including carbohydrate synthesis, secretory pathways, and cellular export. Furthermore, we hypothesize that there may be a reciprocal regulation with Fe(II) oxidation pathways which requires special attention in the search for candidate genes involved in stalk biomineralization.

Kato et al. (2015) previously reported a cluster of four genes with putative involvement in stalk formation that are shared between stalk-forming freshwater Fe(II)-oxidizers and the Zetaproteobacteria *M. ferrooxydans* PV-1, M34, and *Mariprofundus* sp. EKF-M39. Of these, three genes were identified to be similar to the exopolysaccharide synthesis encoding *xagBCD* genes of plant-pathogenic *Xanthomonas* species that attach to their targets with a xanthan-based biofilm

(Tao et al., 2010; Zhang et al., 2013). The fourth gene was annotated as a BcsB-like cellulose synthase regulator protein (Kato et al., 2015). Due to the lack of a systematic comparison of all currently available Zetaproteobacteria isolate genomes, we chose to search for genes connected to stalk formation with a pan-genome approach and subsequently compared our results to those of Kato et al. (2015). Here, we describe the size and trajectory of a pan-genome that comprises 16 genomes of cultured Zetaproteobacteria isolates. Our data reveals substantial genetic diversity within the Zetaproteobacteria, and provides insight into gene clusters that are likely involved in stalk generation. These results on stalk formation align with, and significantly extend those from Kato et al. (2015), while also pointing toward a potential signaling cascade that may regulate a switch between sessile and motile growth in Zetaproteobacteria.

MATERIALS AND METHODS

Functional Annotation of Zetaproteobacteria Genomes

All 16 sequenced genomes of cultured Zetaproteobacteria isolate strains (**Table 1**) were retrieved from the Joint Genome Institute's Genome Portal[1]. Automated functional annotation of all genomes was conducted using Prokka (Seemann, 2014), which predicted coding sequences using Prodigal (Hyatt et al., 2010), ribosomal RNA genes using RNAmmer (Lagesen et al., 2007), transfer RNA genes using Aragorn (Laslett and Canback, 2004), signal leader peptides using SignalP (Petersen et al., 2011), and non-coding RNA regions using Infernal (Nawrocki et al., 2009). Using BLAST+ (Camacho et al., 2009) and

HMMER (Eddy, 2011), Prokka hierarchically annotates against the UniProt (Apweiler et al., 2004), RefSeq (O'Leary et al., 2016), Pfam (Punta et al., 2012), and TIGRFAMs (Haft et al., 2013) databases to predict gene functions for previously identified coding sequences.

Calculation of Average Amino Acid Identities Among Zetaproteobacteria

We identified the degree of genomic similarity between all Zetaproteobacteria and Betaproteobacteria genomes included in our study by calculating their average amino acid identity (AAI) using CompareM[2] with default parameters of 30% sequence identity and 70% sequence alignment length. We computed Bray-Curtis dissimilarity and average linkage hierarchical clustering using the R package vegan (Oksanen et al., 2016), and visualized our results with the R libraries gplots (Warnes et al., 2016), Heatplus (Ploner, 2019), and RColorBrewer (Neuwirth, 2014).

Pan-Genome Computation

We used functionally annotated genome files (see section "Functional Annotation of Zetaproteobacteria Genomes") of 16 Zetaproteobacteria isolates (**Table 1**) for computing the size and trajectory of 7 pan-genomes that differed in minimum blastp percentage identities between 30 and 90% (Altschul et al., 1990) using the Roary software (v. 3.11.2; Page et al., 2015). By comparing multiple pan-genomes computed with a range of different blastp cutoffs, we aimed at a subsequent unbiased screening for potential stalk formation genes (see section "Identification of Stalk-Associated Genes"), which involved the consideration that stalk formation genes may or may not be highly homologous. Our strategy also

---

[1]https://genome.jgi.doe.gov/portal/

[2]https://github.com/dparks1134/CompareM

TABLE 1 | Source environments, biomineral phenotypes, and general genome features of cultured Zetaproteobacteria isolate genomes included in this pan-genome analyses.

	Source	ZOTU	Strain	Genome size [Mbp]	GC content [%]	Contigs	References
Twisted stalks	Deep sea	11	*M. ferrooxydans* PV-1	2.87 (PD)	54.01	32	Singer et al., 2011
		11	*M. ferrooxydans* JV-1	2.85 (D)	53.96	39	Fullerton et al., 2015
		11	*M. ferrooxydans* M34	2.74 (PD)	53.87	36	Field et al., 2015
		36	*Mariprofundus* sp. EKF-M39	2.72 (PD)	51.86	45	Field et al., 2015
	Coastal bay	18	*Mariprofundus* sp. ET2	2.5 (D)	48.76	57	Makita et al., 2017
		18	*Mariprofunuds* sp. DIS-1	2.95 (D)	48.58	57	Mumford et al., 2016
	FeS Coupon	14	*Zetaproteobacteria* sp. EBB1	2.88 (D)	46.65	63	Lopez et al., 2019
	Worm burrow	14	*Zetaproteobacteria* sp. CSS1	2.99 (D)	46.89	87	Beam et al., 2018
		3	*Zetaproteobacteria* sp. SR1	2.42 (D)	50.78	25	Beam et al., 2018
	Steel coupon	n.a.	*Mariprofundus erugo* P3	2.89 (D)	54.7	59	Garrison et al., 2019
		n.a.	*Mariprofunuds erugo* P7	2.98 (D)	54.7	115	Garrison et al., 2019
Dreads	Estuarine bay	18	*Mariprofundus aestuarium* CP-5	2.54 (D)	51.49	1	Chiu et al., 2017
		37	*Mariprofundus ferrinatatus* CP-8	2.3 (D)	53.66	1	Chiu et al., 2017
Amorphous oxides	Worm burrow	1	*Zetaproteobacteria* sp. ECHO-1	1.94 (D)	50.62	8	Beam et al., 2018
	Deep sea	9	*Ghiorsea bivora* SV108	2.14 (D)	43.02	54	Mori et al., 2017
		9	*Ghiorsea bivora* TAG-1	2.16 (PD)	42.68	13	Mori et al., 2017

ZOTU, Zetaproteobacteria operational taxonomic unit (McAllister et al., 2018); PD, permanent draft; D, Draft; n.a., not available.

allowed us to search for potential stalk formation gene clusters composed of multiple genes with different conservation levels. Briefly, Roary converted coding regions identified by Prokka (Seemann, 2014) into protein sequences, removed partial sequences, and iteratively pre-clustered coding regions with CD-HIT (Fu et al., 2012). After pairwise sequence alignment with blastp (Altschul et al., 1990), Roary clustered sequences with the Markov cluster algorithm (MCL) (Enright et al., 2002) with an MCL inflation value of 1.5 and subsequently merged pre-clustering results from CD-HIT with those from MCL (Page et al., 2015). We visualized the calculated pan-genome size and trajectory with R using the ggplot2 package (Wickham et al., 2016).

Functional Analyses of Pan-Genome Subdivisions and Individual Genomes

We analyzed the pan-genomes in three categories as defined by Koonin and Wolf (2008), i.e. (1) core (genes present in all 16 genomes), (2) accessory (genes present in 2–15 genomes), and (3) strain-specific genes (genes present in only 1 genome). After extracting core, shell, and unique genes from the pan-genome computed at 50% minimal blastp identity using the "query_pan_genome -a *.gff" script of Roary (Page et al., 2015), we analyzed the distribution and enrichment of clusters of orthologous groups (COGs) in each pan-genome subdivision and in individual genomes by assigning COG categories to query protein sequences using cdd2cog (Leimbach, 2016).

Identification of Stalk-Associated Genes

Using the "query_pan_genome –a difference" function in Roary (Page et al., 2015), we extracted genes uniquely shared among all 11 stalk-former genomes that at the same time were absent in the genomes of stalk-less isolates from previously computed pan-genomes at blastp cutoffs between 30 and 90% (**Table 1**). We manually checked the query coverage of stalk-former unique genes using blastp (>70% amino acid identity). Protein coding sequences that were initially annotated as hypothetical proteins by Prokka were further analyzed for functional information using EggNOG mapper (Huerta-Cepas et al., 2019) and InterProScan (Quevillon et al., 2005). For stalk-former unique genes annotated as hypothetical proteins, we aligned the homologous amino acid sequences from all 11 stalk-forming Zetaproteobacteria using Muscle (Edgar, 2004) and visualized our results with Jalview2 (Waterhouse et al., 2009) to double check the query coverage and to search for highly conserved regions and identical amino acid residues. Finally, we analyzed whether stalk-former unique genes clustered within same gene cassettes according to continuous gene ID's that Roary automatically assigned during pan-genome computation. Using Gene Graphics (Harrison et al., 2018), we confirmed gene clustering and generated synteny plots of putative stalk formation genes in the genomes they were found in, given that they were available on the National Center for Biotechnology Information (NCBI) database.

SEM Sample Preparation and Imaging of Cell-Stalk-Aggregates

We obtained an active culture of *Mariprofundus micogutta* ET2 from the Japan Collection of Microbes (Ibaraki, Japan) that we cultivated on zero-valent iron plates with artificial seawater medium at 5% headspace oxygen after Laufer et al. (2016) for 5–8 days prior to analyses with an Olympus BX60 epifluorescence microscope (Olympus, Japan) and a Qicam 1394 (Qimaging, Canada). Similarly, we cultivated *Mariprofundus* sp. DIS-1 on zero-valent iron plates with artificial seawater medium after Laufer et al. (2016) for 5 days prior to harvesting 1 mL culture for sample preparation.

To ensure good preservation of biological and mineralogical structures, samples were split up for different treatments: either glutaraldehyde fixation followed by stepwise dehydration or air-drying. Biological samples were fixed in 2.5% glutaraldehyde on ice for 3 h. Samples were then washed three times with ultrapure water, pelleted at 2,800 g for 1 min, and mounted on poly-L-lysine (0.1% w/v aqueous solution) coated glass slides. Samples were then dehydrated sequentially with 30, 50, 70, and 95% ethanol for 5 min each, followed by two times 100% ethanol for 30 min each. Afterwards, samples were treated twice with hexamethyldisilazan for 30 s each. After final air drying, samples were stored in a dry chamber at room temperature.

Non-fixed samples for preservation of mineral structures were washed three times with ultrapure water, concentrated by centrifugation at 2,800 g for 1 min, mounted on poly-L-lysine glass slides, and air-dried afterward. All sample-slides were applied onto carbon-taped aluminum stubs and sputter-coated for 120 s at 4×10^{-2} bar to achieve a 12 nm platinum layer using a Bal-Tec SCD005 sputter coater (Baltic praeparation, Germany). Imaging was performed with a JSM-6500F field emission scanning electron microscope (JEOL, Germany) with a Schottky field emitter at 5kV acceleration voltage and a working distance of approximately 10 mm at the center for Light-Matter Interaction, Sensors and Analytics (LISA+, University of Tuebingen, Germany).

Data Availability

The Genbank accession ID's for *sfz1-sfz6* in *M. ferrooxydans* PV-1 are EUA54616-EUA54621, which we updated with the product annotations "stalk formation protein in Zetaproteobacteria 1–6," respectively. Amino acid sequences of $Sfz1_{PV-1}$-$Sfz6_{PV-1}$ and Genbank accession ID's are available in the SI.

RESULTS AND DISCUSSION

The diverse range of geographically distinct Fe(II)-rich environments inhabited by Zetaproteobacteria, including terrestrial, freshwater, coastal, and deep-sea environments (**Table 1**), hints toward their adaptive nature. Although the number of isolates is relatively limited across all these environments, there was no obvious differentiation of morphological type to a specific habitat-type (**Table 1**). We refer to the dread and amorphous oxide phenotypes collectively as "stalk-less" hereafter.

Pan-Genome Reveals High Genetic Diversity Among Zetaproteobacteria

The pan-genome is defined as the entire repertoire of genes accessible to a bacterial clade, subdivided into conserved "core" genes (shared between all n genomes), variable "accessory" genes (shared between two and n-1 genomes), and "strain-specific" genes (present in only one genome) (Tettelin et al., 2005; Koonin and Wolf, 2008; Vernikos et al., 2015). The relative proportions of these three subdivisions characterize a clade's genetic and functional richness and its capacity to acquire exogenous DNA, and define the trajectory of its pan-genome (Rouli et al., 2015). Roary regression curve models predicted the Zetaproteobacteria pan-genome to be open with infinite growth in total gene numbers and decreasing core gene numbers as new genomes are added (**Figure 1**). Accessory and strain-specific genes made up the majority of the Zetaproteobacteria pan-genome, whereas the core genome was relatively small with only 0.04–10.3% of the total gene count (90–30% blastp identity, respectively, **Supplementary Table 1**).

Open pan-genomes generally reflect sympatric lifestyles, i.e., life in mixed communities with high rates of horizontal gene transfer (Rouli et al., 2015; Vernikos et al., 2015), and are associated with niche versatility (Tettelin et al., 2005). Our results are consistent with the previously reported high genomic diversity across Zetaproteobacteria based on metagenomic data (Fullerton et al., 2017; McAllister et al., 2019) and indicate that Zetaproteobacteria may undergo considerable genetic exchange with other community members. We could not identify a correlation between source environment and the proportion of strain-specific and shared genes in isolate genomes (**Supplementary Table 2**) and so cannot conclude that a specific habitat may favor higher gene acquisition rates in Zetaproteobacteria. The high genomic diversity among currently available Zetaproteobacteria isolates is further reflected in amino acid identities (**Supplementary Figure 1**) that classify only deep sea members *M. ferrooxydans* PV-1, JV-1, and M34 (98.1 ± 5.4% to 99.9 ± 1.0%), and coastal estuarine isolates *M. erugo* P3 and P7 (98.9 ± 5.0%; **Supplementary Figure 1**) at the species rank (AAI of > 95%; Medlar et al., 2018) as previously reported by Garrison et al. (2019) and McAllister et al. (2019, 2020).

Candidate Gene Cluster for Stalk Synthesis and Export Is Shared With Stalk-Forming Betaproteobacteria
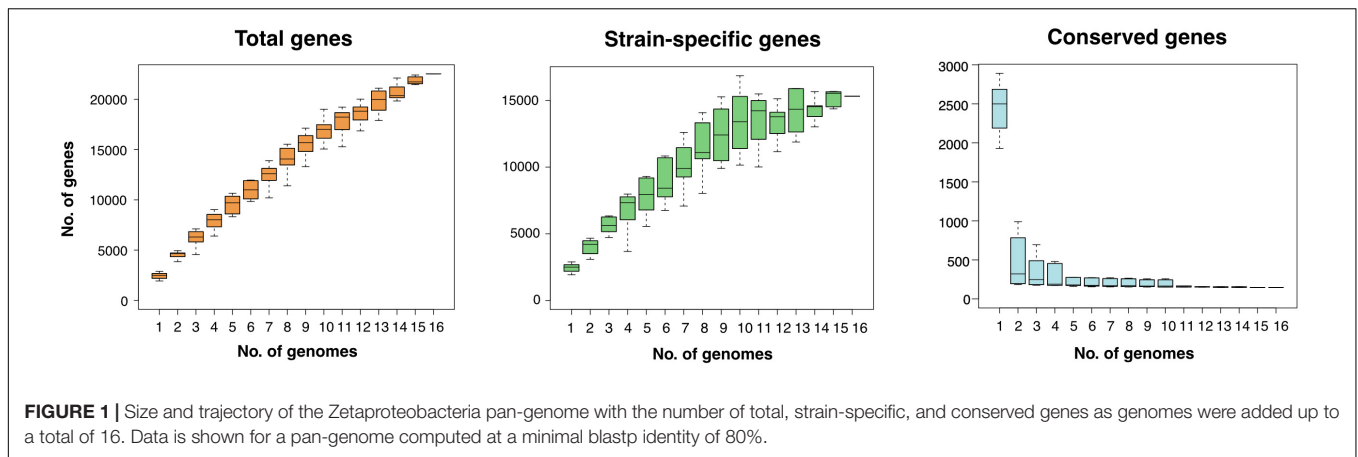
We used comparative genomics as an approach to identify the genetic basis for stalk formation where each Zetaproteobacteria isolate genome was assigned to either the stalk-former, dread-former, or amorphous-oxide-former trait groups (**Table 1**). While the assignment was clear for most isolate genomes from the documentation in the literature (Chan et al., 2011; Field et al., 2015; Mumford et al., 2016; Chiu et al., 2017; Mori et al., 2017; Beam et al., 2018) or from microscopic observations in our laboratory (**Figures 2A–F**), the trait was more ambiguous for strain *Mariprofundus micogutta* ET2 (JCM 30585^T), a member of the Zetaproteobacteria operational taxonomy unit 18 (ZOTU 18). Makita et al. (2017) reported strain ET2 to form thin Fe- and

C-rich extracellular filaments that differ morphologically from stalks. Using light microscopy, we identified stalk production in an active culture of *M. micogutta* ET2 (**Figures 2G,H**), although the majority of cells did not appear associated with obvious stalk structures. Hence, we assigned ET2 to the "stalk-forming" trait group that overall comprised 11 genomes (PV-1, JV-1, M34, ET2, DIS-1, EKF-M39, EBB1, CSS1, SR1, P3, and P7), while we assigned 5 genomes to the "stalk-less" (CP5, CP8, TAG-1, SV-108, and ECHO1) trait group during our analyses.

We extracted all genes from the Zetaproteobacteria pan-genome that are shared among stalk-former genomes and lacking in stalk-less Zetaproteobacteria for our analysis. Our search revealed 82 genes unique to stalk-forming Zetaproteobacteria at minimal blastp identities between 30 and 80% (**Supplementary Table 3**). We excluded genes that were identified above 90% minimal blastp identity from our analysis as they were only shared between 8 or less stalk-formers. Prokka annotated 34 stalk-former unique genes with functions in flagellar basal-body biosynthesis, motility, chemotaxis, and terminal electron transfer with cbb3-type cytochromes (**Supplementary Table 3**; see sections "Stalk-Former Unique Genes Emphasize a Functional Connection to c-di-GMP Signaling" and "Stalk-Former Unique Cytochromes Specific to Stalk-Forming Zetaproteobacteria" for further details). The other 48 genes were predicted as hypothetical proteins among which we found 6 conserved genes that cluster together with a shared synteny in all stalk-former genomes.

The high level of conservation above 50% amino acid identity (**Table 2**) and its lack in stalk-less Zetaproteobacteria emphasize the potential for this gene cluster as a candidate for the biosynthesis and export of the stalk organic matrix, which we therefore refer to as the **s**talk **f**ormation in **Z**etaproteobacteria (*sfz*) cluster with genes *sfz1-sfz6* hereafter. Protein motif analysis with InterProScan (Quevillon et al., 2005) predicted domains with functions in exopolysaccharide synthesis and export, cell wall synthesis inhibition, and carbohydrate hydrolysis in Sfz1, Sfz2, and Sfz4, respectively, that we describe in detail below (see sections "Sfz1 Is a Putative c-di-GMP-Dependent Regulator of Exopolysaccharide Synthesis" and "Sfz2 Putatively Catalyzes Exopolysaccharide Synthesis and Export"). We could not identify known domains and putative functions for Sfz3, Sfz5, and Sfz6 (see section "Sfz4 and Sfz6 Contain Domains for Carbohydrate Cleavage and Cell Wall Synthesis Inhibition").

The appearance of twisted stalks in Zeta- and Betaproteobacteria indicates that this trait evolved from shared ancestral genes encoding stalk biosynthesis and export. We elucidated the presence and degree of conservation of the *sfz* cluster in stalk-forming Betaproteobacteria using blastp (Altschul et al., 1990), and found homologs of *sfz1-sfz4* with a shared synteny in the stalk-formers *Ferriphaselus amnicola* OYT-1 and *Ferriphaselus* sp. R-1 (**Figure 3**; blastp identities below 40%). We refer to *sfz* homologs in Betaproteobacteria as *sfb1-sfb4* (**s**talk **f**ormation in **B**etaproteobacteria) hereafter. As with the *sfz* cluster in Zetaproteobacteria, we found no *sfb* genes in the stalk-less Fe(II)-oxidizing Betaproteobacteria *Gallionella capsiferriformans* ES-2 and *Sideroxydans lithotrophicus* ES-1. Our blastp search further revealed the dread-forming Zetaproteobacteria *M. aestuarium* CP-5 and *M. ferrinatatus*

FIGURE 1 | Size and trajectory of the Zetaproteobacteria pan-genome with the number of total, strain-specific, and conserved genes as genomes were added up to a total of 16. Data is shown for a pan-genome computed at a minimal blastp identity of 80%.

CP-8 to comprise less conserved homologs of *sfz1-sfz6* (blastp identities below 40%, **Figure 3**), which may point toward two distinct phenotypes among Zetaproteobacteria that are rooted in the same gene cluster.

FIGURE 2 | Scanning electron microscopy **(A–F)** and light microscopy images **(G,H)** of cell-stalks aggregates in cultures of *Mariprofundus* sp. DIS-1 **(A–F)** and *M. micogutta* ET2 **(G,H)**. White arrows in G-H indicate stalks.

Our findings align with a previous report of the *sfb* and *sfz* genes by Kato et al. (2015) as the "*xag*-like" gene cassette (CDS9-CDS12), that they found to be shared between freshwater stalk-formers and marine *M. ferrooxydans* PV-1, M34, and *Mariprofundus* sp. EKF-M39. Based on this limited genomic data, the authors surmised that respective genes might be involved in stalk formation. Our results here significantly expand the number of genomes, especially among the Zetaproteobacteria, and provide additional comparative analysis between stalk-forming and stalk-less strains. Thus, our present data support the proposal of Kato et al. and underline the strong potential of this gene cassette as a candidate gene cluster for stalk exopolysaccharide synthesis. Interestingly, Betaproteobacteria lack homologs for *sfz5* and *sfz6*, indicating that only *sfz1-sfz4* and *sfb1-sfb4* may be essential for the stalk phenotype, whereas the roles for *sfz5* and *sfz6* in stalk-formation may be more ambiguous.

While other stalk-former unique genes with conservation levels above 80% blastp identity such as HP_80_1 to HP_80_3 (see **Supplementary Table 3**) generally seem possible candidates for stalk formation, we found no homologs of respective genes in stalk-forming Betaproteobacteria. Moreover, our analysis revealed said genes to cluster within gene cassettes of distinct functions such as flagellar or cytochrome biosynthesis, pointing rather toward related functions instead of stalk formation. In general, exopolysaccharide synthesis strategies can vary significantly among bacteria, and are encoded by different sets of genes that can be more or less conserved in sequence and operon structure across taxa (for a review see Schmid et al., 2015). Thus, it is possible that additional accessory genes play a structural role in stalk-formation in the strains they are found in, but that other strains may use different structural polysaccharides in stalk-formation.

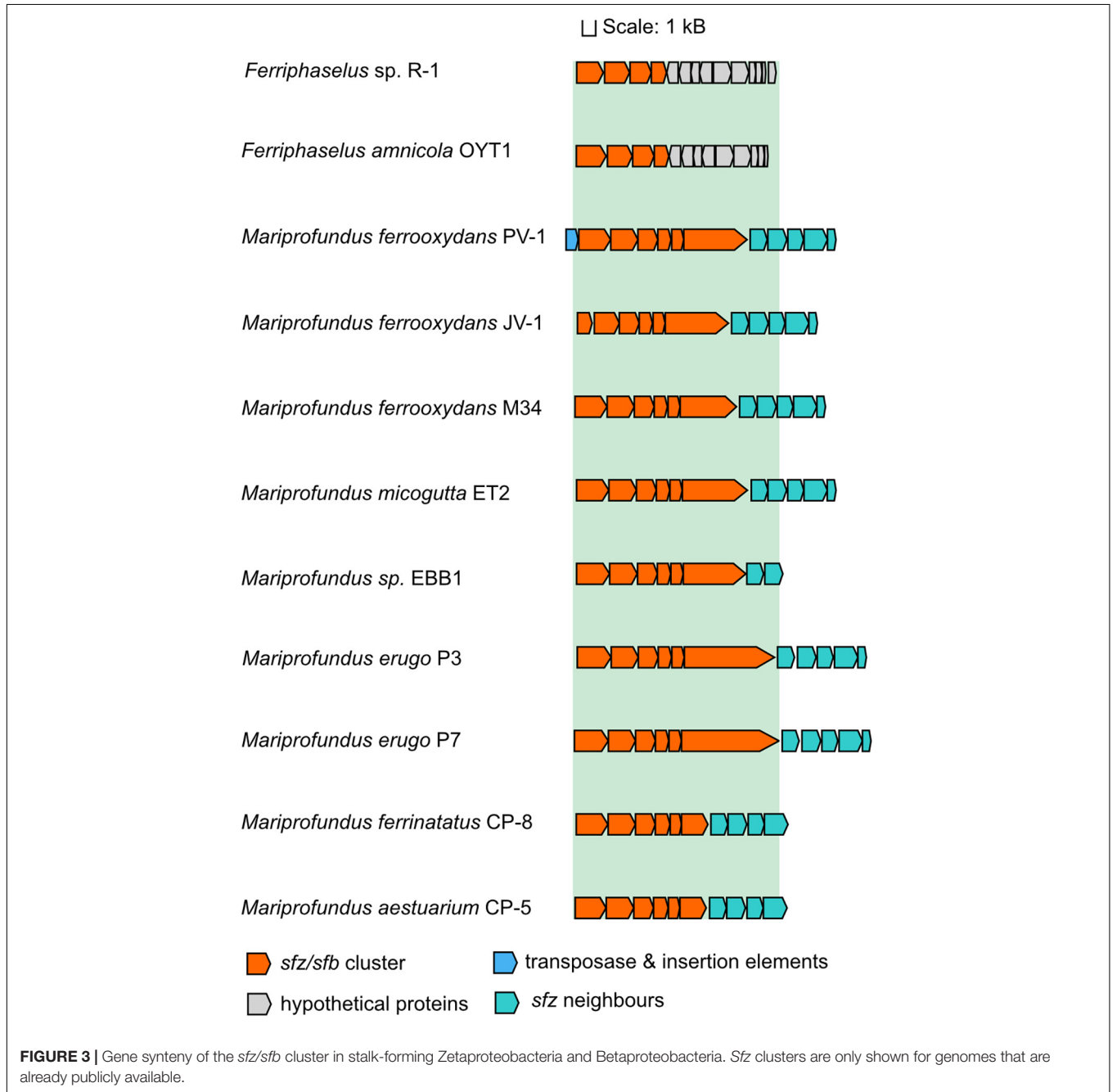## Sfz1 Is a Putative c-di-GMP-Dependent Regulator of Exopolysaccharide Synthesis

Our protein motif analyses predicted the Sfz1 amino acid sequence (60% in stalk-formers) to contain two non-cytoplasmic BcsB-like domains flanked by an N-terminal transmembrane signal peptide domain and a C-terminal transmembrane domain.
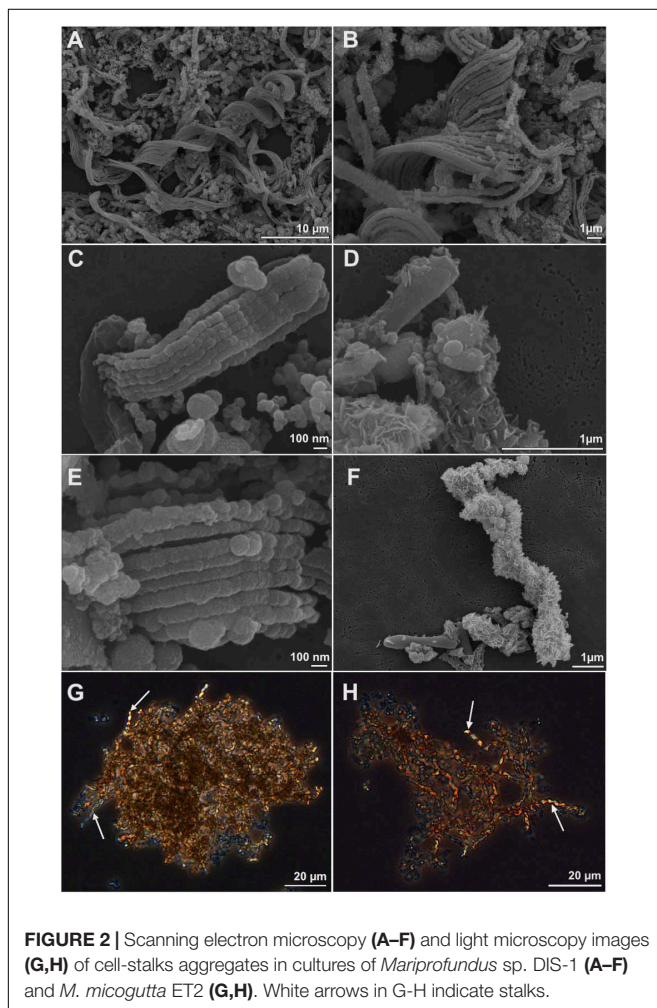
TABLE 2 | Sfz genes with minimal and maximal basepair length in different stalk-former genomes at minimal blastp identities between 50 and 70%.

Gene	Annotation	Putative function	Min. length	Max. length	Min. blastp identity
Sfz1	Hypothetical protein	Regulator of EPS synthase	1,067	2,576	60
Sfz2	Hypothetical protein	EPS synthesis and export	2,051	2,084	70
sfz3	Hypothetical protein	Unknown	1,499	1,535	60
sfz4	Hypothetical protein	carbohydrate/cell wall hydrolysis	971	1,037	50
sfz5	Hypothetical protein	Unknown	929	983	50
sfz6	Hypothetical protein	Unknown	3,311	7,352	50

EPS, exopolysaccharide.

FIGURE 3 | Gene synteny of the *sfz/sfb* cluster in stalk-forming Zetaproteobacteria and Betaproteobacteria. *Sfz* clusters are only shown for genomes that are already publicly available.

We found the N-terminal part of Sfz1 to be less conserved than the C-terminal fraction that contains 76 identical residues in all stalk-formers (**Supplementary Figure 2**). *M. ferrooxydans* JV-1 only contains the second part of *sfz1*, which likely is an artifact as the gene is split over two contigs in its genome.

BcsB is the periplasmic membrane-anchored regulatory subunit of the dimeric cellulose synthase enzyme in *Acetobacter xylinum* (Wong et al., 1990) that regulates the polymerization of uridine 5′-disphosphate glucose (UDP-glucose) to cellulose, an exopolysaccharide that builds the structural basis of a variety of biofilms. The reaction is catalyzed by the family II glycosyltransferase BcsA upon activation by the BcsB subunit, which depends on the binding of the positive effector cyclic dimeric guanylate monophosphate (c-di-GMP). High intracellular c-di-GMP levels were shown to stimulate exopolysaccharide synthesis and biofilm formation as part of surface attachment and sessile growth in different bacteria, including *Pseudomonas aeruginosa, Escherichia coli, Salmonella enterica serovar Typhimurium,* and *Shewanella oneidensis* MR-1. In contrast, low c-di-GMP levels induce cell detachment and motility (Simm et al., 2004; Merritt et al., 2007; Kuchma et al., 2015).

Generally, c-di-GMP is known to be a central signaling molecule for a variety of signaling cascades, and it particularly controls the switch between motile and sessile growth in bacteria, where biofilm formation is up- or downregulated for surface attachment or detachment (for a review see D'Argenio and Miller, 2004; Valentini and Filloux, 2016). An interaction between Sfz1 and c-di-GMP could point toward a surface holdfast function of stalks as previously proposed by Singer et al. (2011), and our pan-genome data provides additional hints for such a functional relationship (see section "Stalk-Former Unique Genes Emphasize a Functional Connection to c-di-GMP Signaling"). Our detailed analysis of Sfz1 indicates it could encode the conserved c-di-GMP binding motifs RxxD (glutamine-x-x-aspartate; **Supplementary Figure 2**, residues 295–298 and 830–833; Chou and Galperin, 2016) and RxxxR (glutamine-x-x-x-glutamine; **Supplementary Figure 2**, residues 227–231 and 326–330, Morgan et al., 2014). Sfb1 similarly contained c-di-GMP binding sites with an additional RxxD motif compared to Sfz1 (**Supplementary Figure 2**, residues 270–273), whereas it lacked the C-terminal RxxD motif in Sfz1 (residues 830–833) and contained aspartic acid instead of arginine at the RxxxR site at residues 326–330.

Functional exopolysaccharide synthesis would require at least a second gene that encodes a glycosyltransferase function for polysaccharide polymerization similar to BcsA, which we identified to be the case for *sfz2* (see section "Putatively Catalyzes Exopolysaccharide Synthesis and Export"). Interestingly, the BcsA and BcsB couple was previously proposed to carry out both the synthesis and translocation of exopolysaccharides, as BcsA can, in addition to its glycosyltransferase activity, also form a channel across the cell membrane that allows for the export of the synthesized polysaccharide chain (Morgan et al., 2013; Omadjela et al., 2013). Whether Sfz1

and Sfz2 can carry out a similar function in stalk-formers remains to be explored.

Sfz2 Putatively Catalyzes Exopolysaccharide Synthesis and Export

Our analyses predicted the highly conserved Sfz2 sequence (70% in stalk-formers; **Supplementary Figure 3**) to contain an N-terminal EpsE-like type II secretion system (T2SS) protein domain (IPR007831) and a C-terminal family II glycosyltransferase domain (IPR001173). The presence of the conserved glycosyltransferase motifs QxxRW (glutamine-x-x-arginine-tryptophane; residues 523–527), DxD (aspartate-x-aspartate; residues 488–490), and TED (threonine-glutamate-aspartate; residues 486–488; **Supplementary Figure 3**; Morgan et al., 2016) supports a glycosyltransferase function at the C-terminal end of Sfz2. We found the QxxRW and DxD motifs in Sfb2 as well, whereas it lacked the TED site.

The N-terminal part of Sfz2 may involve a translocation function, as EpsE is a cytosolic, hexameric ATPase that causes asymmetric conformational rearrangements in the T2SS inner membrane platform through ATP hydrolysis, promoting the extension of secreted exopolymers (Korotkov et al., 2012). Guttenplan et al. (2010) proposed EpsE in *B. subtilis* to be a bifunctional enzyme where it acts either as a family II glycosyltransferase (Garinot-Schneider et al., 2000), or as a molecular clutch that controls the flagellar rotor and inhibits cellular motility (Blair et al., 2008). Due to its gene architecture Sfz2 may encode a similar function, although a translocation function seems questionable. Functional exopolymer secretion via T2SS was demonstrated to require at least 12 different proteins that are generally encoded on a single operon (Sandkvist, 2001) and no other *eps* homologs were identified in the *sfz* cluster or its gene neighborhood (see section "Sfz Gene Neighborhood").

Importantly, *eps* gene expression was demonstrated to be upregulated by c-di-GMP in *Vibrio cholerae* (Sloup et al., 2017). As in Sfz1, we found the c-di-GMP binding motifs RxxxR (**Supplementary Figure 2**, residues 522–526, Morgan et al., 2014) and RxxD (**Supplementary Figure 2**, residues 97–100 and 325–328) in Sfz2 and two additional RxxD motifs in Sfb2 (residues 282–285 and 297–300), which further emphasizes a possible connection between the *sfz and sfb* gene clusters and a c-di-GMP regulated switch into sessile or motile growth in stalk-formers (see section "Stalk-Former Unique Genes Emphasize a Functional Connection to c-di-GMP Signaling" for further details). Together, the conservation within Sfz2 with these different motifs involved in exopolysaccharide production and environmental response is circumstantial evidence for similar functions in stalk formation; however experimental work will be necessary to confirm these functions.

Sfz4 and Sfz6 Contain Domains for Carbohydrate Cleavage and Cell Wall Synthesis Inhibition

The 323–345 aa long Sfz4 sequence (50% minimal blastp identity) is predicted with a function as a family 10 glycosyl hydrolase. Glycosyl hydrolases cleave glycosidic bonds

between carbohydrates, and family 10 enzymes have been specifically reported with exopolysaccharide-degrading activities as for instance in xylanase, endo-1,3-beta-xylanase, and cellobiohydrolase enzymes (Henrissat, 1991; Do et al., 2013). Cell wall degrading glycosyl hydrolases were demonstrated with roles in the assembly of supramolecular membrane-spanning structures such as secretion systems, and phenotypes of cells with non-functional cell wall hydrolases included impaired biofilm formation, surface attachment, or chemotaxis (Nambu et al., 1999; Mercier et al., 2002; Vollmer et al., 2008; Vermassen et al., 2019). A possible role of Sfz4 in stalk formation could be in hydrolyzing the stalk exopolysaccharide for cell detachment as part of a transition from sessile to motile growth, given that stalks function as surface holdfasts.

Sfz6 forms an exception to the other *sfz* genes in that it is distantly located from the *sfz* cluster in Zetaproteobacteria sp. SR-1 and EKF-M39. Since stalk-forming Betaproteobacteria lack a homolog for *sfz6*, it is presumably not essential for stalk formation but rather functions as an accessory gene complementing essential stalk formation genes. The 1,415–2,270 aa long Sfz6 protein sequence (50% blastp identity) contains an N-terminal SH3 (sarcoma homology-3) domain and a C-terminal CotH domain. Among a variety of different cellular functions, SH3 domains were suggested to function as cell wall binding sites during cell wall hydrolysis (Vermassen et al., 2019). In contrast, the function of CotH is presumably determined by additional adjacent protein domains. For instance, cellulose-binding domains may indicate a role of CotH in the formation of the cellulose-degrading multi-enzyme cellulosome complex or generic interactions with the cell wall (Nguyen et al., 2016). Generally, a domain architecture with a combination of SH3 and CotH seems unique and due to the wide variety of functions encoded by SH3 domains, we cannot confer a potential function for *sfz6*.

Sfz Gene Neighborhood

A cassette of five genes that is directly adjacent to the *sfz* cluster is conserved across stalk-forming Zetaproteobacteria. Their annotated functions are an ATP-ADP antiporter, a phosphomannose-isomerase, a N-acetylmuramoyl-L-alanine amidase, a MsbA-like ABC-type protein exporter, and a protein of unknown function (**Figure 3**), respectively. Phosphomannose-isomerases interconvert mannose-6-phosphate and fructose-6-phosphate, and were demonstrated to be involved in the synthesis of the bacterial exopolysaccharides xanthane and alginate (Shinabarger et al., 1991; Köplin et al., 1992), whereas N-acetylmuramoyl-L-alanine amidases cleave amide bonds in the net-like peptidoglycan structure and degrade it (Wilmes et al., 2017; Vermassen et al., 2019). None of the five genes are unique to stalk-formers, and Zetaproteobacteria sp. CSS1 and EBB1 contain only the phosphomannose-isomerase and the ATP-ADP antiporter genes in their *sfz* gene cluster neighborhood. Dread-formers contain four of the five conserved genes in the *sfz* gene neighborhood as well, whereas stalk-forming Betaproteobacteria lack the entire cassette. Altogether, their functional connection to carbohydrate metabolism and close association with the *sfz* cluster point

toward a potential involvement in stalk formation, possibly as complementary genes.
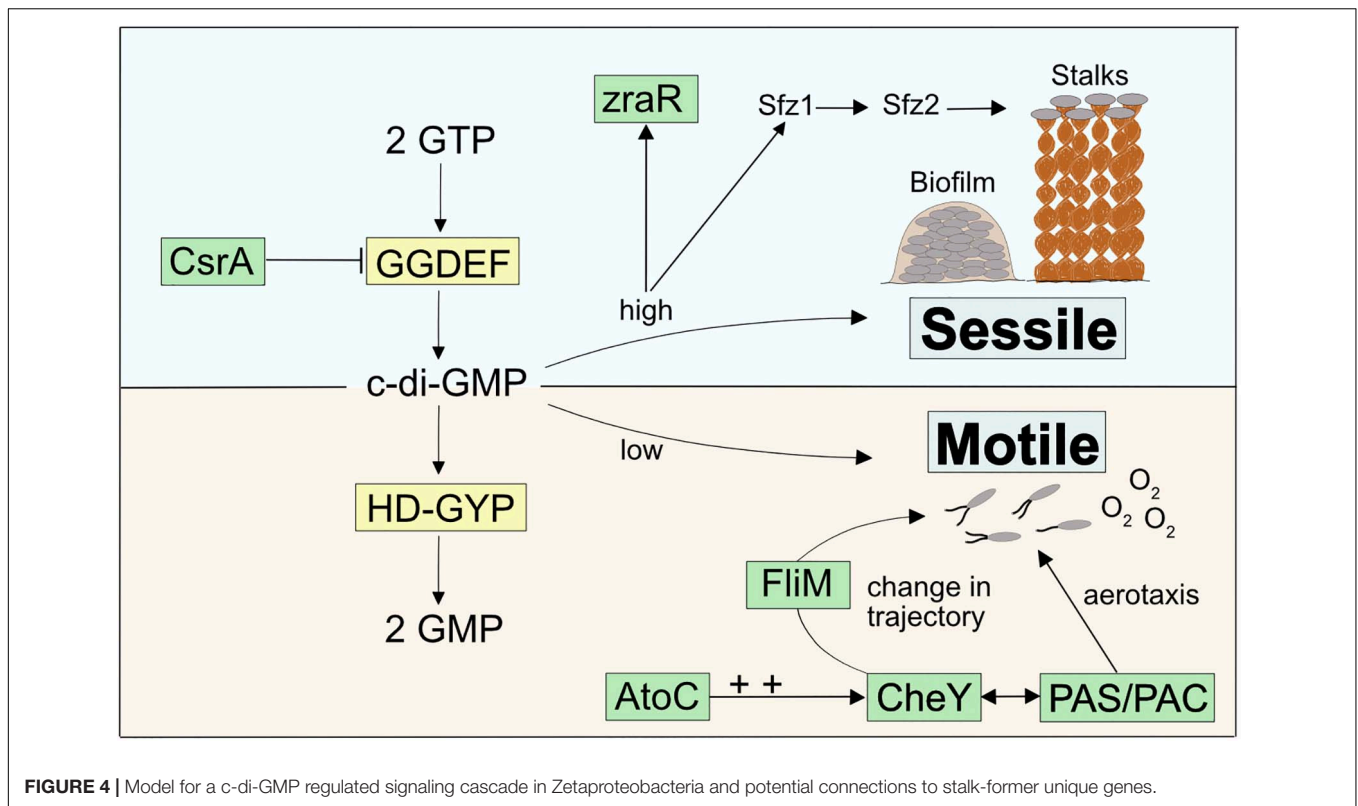
Generally, our *sfz/sfb* gene neighborhood analysis revealed these gene cassettes could confer functions in carbohydrate metabolism, chemotaxis, extracellular export, electron transfer, and c-di-GMP related signaling. There is variation in the gene architecture of these cassettes across all stalk-formers (**Supplementary Figure 4**). It is possible that this diversity could reflect species-specific adjustments to stalk formation and functional roles that are niche-driven; however, the functional assignments are deduced from homologous genes in other bacteria, and the true phenotypic ramifications for different stalk-formers will require further elucidation.

Stalk-Former Unique Genes Emphasize a Functional Connection to c-di-GMP Signaling

Zetaproteobacteria are under constant pressure to compete against abiotic Fe(II) oxidation by oxygen (Rentz et al., 2007; Druschel et al., 2008). Therefore, sensing and quickly responding to geochemical changes as well as sustaining stable growth in environments with optimal Fe(II) and O_2 gradients are essential attributes for their survival. Our pan-genome analysis revealed a general enrichment of COG categories associated with c-di-GMP signaling (CheY-like response regulator domains, GGDEF, AAA-type ATPase, HD-GYP, and PAS/PAC domains, **Supplementary Table 4**) across stalk-forming and stalk-less Zetaproteobacteria. The signaling molecule c-di-GMP upregulates biofilm formation and sessile growth at high intracellular concentrations, whereas declining c-di-GMP concentrations result in increasing surface detachment and motile growth (**Figure 4**). While these results point toward a pronounced importance of switching between motile and sessile growth in stalk-forming and stalk-less Zetaproteobacteria (**Figure 4**), our domain analysis of *sfz1-6* and additional stalk-former unique genes highlights a particular emphasis of such a switching response in stalk-formers.

For instance, we found two highly conserved transcriptional regulators unique to stalk-formers that both neighbor additional stalk-former unique genes (**Supplementary Figure 4**) and that favor cellular motility, i.e., *atoC* and *csrA* (**Supplementary Table 3**; NCBI accession EAU54699.1 and EAU54987.1 in *M. ferrooxydans* PV-1, respectively). Both genes are located in the close neighborhood of c-di-GMP signaling genes with GGDEF- and HD-GYP domains that are known to antagonistically regulate intracellular levels of c-di-GMP (**Figure 4**). GGDEF-containing diguanylate cyclases catalyze c-di-GMP synthesis and hence an increase in its concentrations (Tal et al., 1998), while HD-GYP-containing phosphodiesterase enzymes catalyze c-di-GMP degradation (Ryan et al., 2009).

CsrA was reported to negatively regulate the expression of GGDEF-containing proteins that upregulate intracellular c-di-GMP levels (Jonas et al., 2008), thereby inhibiting the expression of the exopolysaccharide gene *pgaA* in *Campylobacter jejuni* (Huertas-Rosales et al., 2017). *CsrA*-mutant cells were demonstrated to show poor swarming ability and lowered resistance to oxidative stress (Fields and Thompson, 2008). AtoC

FIGURE 4 | Model for a c-di-GMP regulated signaling cascade in Zetaproteobacteria and potential connections to stalk-former unique genes.

is a transcriptional regulator that was demonstrated to induce the expression of the flagellar *flhDC* and *fliAZY* operons in *E. coli.* Cells deficient in *atoC* were reported to be unresponsive to chemoattractants and showed a general lack of cellular motility. AtoC-dependent regulation of motility and chemotaxis also required an upregulation of CheY (Theodorou et al., 2012), a cytoplasmic chemotaxis response regulator that receives environmental signals and controls cellular swarming trajectories upon phosphorylation. By binding to the flagellar motor protein FliM (Clegg and Koshland, 1984; Bray and Bourret, 1995), phosphorylated CheY induces a clockwise rotation of the flagellum and causes a decrease in torque and swarming speed. In contrast, de-phosphorylated CheY does not bind FliM, which maintains the cellular swarming direction and speed with a counterclockwise flagellar rotation (**Figure 4**; Scharf et al., 1998; Yuan et al., 2010; Nesper et al., 2017). We found that the stalk-former unique genes comprise the flagellar motor protein gene *fliM* (NCBI accession EAU54562.1 for *M. ferrooxydans* PV-1) within a gene cluster containing other stalk-former unique genes annotated with functions in flagellar motility (**Supplementary Table 3** and **Supplementary Figure 4**), which may be related to a specialized switching response of the flagellum in stalk-formers, and could indicate a functional connection between stalk formation and chemotaxis regulation.

A third stalk-former unique gene located in the *atoC* gene neighborhood (**Supplementary Figure 4**), i.e., *zraR* (**Supplementary Table 3**; NCBI accession EAU54701.1 in *M. ferrooxydans* PV-1), encodes a σ^{54} transcriptional regulator for metal tolerance and resistance (Sallai and Tucker, 2005).

Its gene expression was shown to be induced by high intracellular levels of c-di-GMP, although the effect of higher *zraR* transcription has not been resolved yet (Méndez-Ortiz et al., 2006).

Stalk-Former Unique Cytochromes Specific to Stalk-Forming Zetaproteobacteria

In addition to genes with potential functional links to c-di-GMP regulation, we found six stalk-former unique genes with annotations as cbb3-type cytochrome subunits and as *fixGH* genes (**Supplementary Table 3**; NCBI accession EAU53376— EAU53382 in *M. ferrooxydans* PV-1, respectively). The latter are putatively required for the assembly of cbb3-type cytochromes (Preisig et al., 1996) and cluster together in stalk-formers (**Supplementary Figure 4**).

Cbb3-type cytochromes are terminal oxidases for aerobic respiration with a high affinity for oxygen (Pitcher and Watmough, 2004) that were reported of being primarily expressed under O_2-limited conditions (Mandon et al., 1994). They were found among the most highly expressed proteins in the proteome of *M. ferrooxydans* PV-1 (Barco et al., 2015), which is reasonable with respect to the microaerophilic metabolism of Zetaproteobacteria (Field et al., 2015). All Zetaproteobacteria contain cbb3-type cytochrome c genes in their genomes, and the stalk-former unique subunits may be an adaptation required for sensing optimal oxygen levels as part of microaerobic growth during stalk formation.

Generally, Per-Arnt-Sim motifs (PAS) and PAS-associated C-terminal motifs (PAC) were prevalent domains among the strain-specific genes of stalk-forming and stalk-less Zetaproteobacteria (**Supplementary Table 4**). PAS/PAC domains are known to respond to environmental signals such as oxygen concentrations, light, and redox potential (Taylor and Zhulin, 1999). The PAC motif was demonstrated to control flagellar rotation during aerotaxis of *E. coli* upon sensing an emitted signal by the PAS domain depending on the surrounding redox conditions (Rebbapragada et al., 1997) with an involvement of CheY (Bibikov et al., 2000). Given their abundance, PAS/PAC domains may determine the swarming trajectories of Zetaproteobacteria during aero- and/or chemotaxis by a similar mechanism (**Figure 4**).

CONCLUSION

Our pan-genome analysis revealed a high genomic diversity among currently available Zetaproteobacteria isolates, and a highly conserved gene cluster in stalk-forming Zetaproteobacteria that is shared at lower conservation levels with stalk-forming Betaproteobacteria and dread-forming Zetaproteobacteria. The *sfz* and *sfb* clusters are absent in stalk-less Zetaproteobacteria and Betaproteobacteria, and as previously postulated by Kato et al. (2015), they are likely candidates for encoding the twisted stalk trait. The presence of conserved protein domains for the binding of the signaling messenger c-di-GMP in *sfz* and *sfb* genes together with stalk-former unique genes that putatively encode transcription factors regulating a c-di-GMP dependent signaling cascade for a switch between sessile and motile growth point toward a functional connection between stalk formation and surface attachment. Singer et al. (2011) previously proposed a model according to which stalks serve as holdfasts to anchor cells to surfaces during sessile growth phases. In such a scenario, stalks could provide an advantage over amorphous biofilms as cells can grow within gradients while remaining surface-attached (Chan et al., 2011, 2016). By comparison, stalk-less cells forming amorphous biofilms would have to frequently detach to swarm toward more favorable conditions. Generally, the fact that stalk-formers bear highly conserved unique genes that putatively control cellular motility and surface attachment is supportive of a holdfast function of stalks, though the actual involvement of respective genes in such processes and the extent to which they are linked to stalk formation requires detailed investigation. Our observations reported here require further investigation with lab-based experiments involving gene deletion/mutagenesis, differential

transcriptomics and gene expression analyses to uncover whether the *sfz* and *sfb* gene clusters indeed encode stalk formation in Zetaproteobacteria and Betaproteobacteria, and to identify whether stalks are functionally linked to surface attachment. In particular, the potential for stalk-less Zetaproteobacteria to harbor alternative stalk formation genes that may be non-operative requires investigation with laboratory experiments.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

EK, DE, OB, and CC contributed conception and design of the study. EK generated the data, performed data analyses, and wrote the first draft of the manuscript. TB conducted SEM sample analyses and supported EK in SEM sample preparation. OB supported EK in workflow design, script writing, and computational data analyses. All authors contributed to manuscript revision, and read and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb. 2021.679409/full#supplementary-material

REFERENCES

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

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32(Suppl. 1), D115–D119.

Barco, R. A., Emerson, D., Sylvan, J. B., Orcutt, B. N., Meyers, M. E. J., Ramírez, G. A., et al. (2015). New insight into microbial iron oxidation as revealed by

the proteomic profile of an obligate iron-oxidizing chemolithoautotroph. *Appl. Environ. Microbiol.* 81, 5927–5937. doi: 10.1128/AEM.01374-15

Beam, J. P., George, S., Record, N. R., Countway, P. D., Johnston, D. T., Girguis, P. R., et al. (2020). Mud, microbes, and macrofauna: seasonal dynamics of the iron biogeochemical cycle in an intertidal mudflat. *Front. Mar. Sci.* 7:562617. doi: 10.3389/fmars.2020.562617

Beam, J. P., Scott, J. J., McAllister, S. M., Chan, C. S., McManus, J., Meysman, F. J., et al. (2018). Biological rejuvenation of iron oxides in bioturbated marine sediments. *ISME J.* 12, 1389–1394. doi: 10.1038/s41396-017-0032-6

Bibikov, S. I., Barnes, L. A., Gitin, Y., and Parkinson, J. S. (2000). Domain organization and flavin adenine dinucleotide-binding determinants in the aerotaxis signal transducer aer in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5830–5835. doi: 10.1073/pnas.100118697

Blair, K. M., Turner, L., Winkelman, J. T., Berg, H. C., and Kearns, D. B. (2008). A molecular clutch disables flagella in the *Bacillus subtilis* biofilm. *Science* 320, 1636–1638. doi: 10.1126/science.1157877

Bray, D., and Bourret, R. B. (1995). Computer analysis of the binding reactions leading to a transmembrane receptor-linked multiprotein complex involved in bacterial chemotaxis. *Mol. Biol. Cell* 6, 1367–1380. doi: 10.1091/mbc.6.10.1367

Brooks, C. N., and Field, E. K. (2020). Iron flocs and the three domains: microbial interactions in freshwater iron mats. *mBio* 11:e02720-20.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Chan, C. S., De Stasio, G., Welch, S. A., Girasole, M., Frazer, B. N., Nesterova, M. V., et al. (2004). Microbial polysaccharides template assembly of nanocrystal fibers. *Science* 303, 1656–1658. doi: 10.1126/science.1092098

Chan, C. S., Fakra, S. C., Edwards, D. C., Emerson, D., and Banfield, J. F. (2009). Iron oxyhydroxide mineralization on microbial extracellular polysaccharides. *Geochim. Cosmochim. Acta* 73, 3807–3818. doi: 10.1016/j.gca.2009.02.036

Chan, C. S., Fakra, S. C., Emerson, D., Fleming, E. J., and Edwards, K. J. (2011). Lithotrophic iron-oxidizing bacteria produce organic stalks to control mineral growth: implications for biosignature formation. *ISME J.* 5, 717–727. doi: 10.1038/ismej.2010.173

Chan, C. S., McAllister, S. M., Leavitt, A. H., Glazer, B. T., Krepski, S. T., and Emerson, D. (2016). The architecture of iron microbial mats reflects the adaptation of chemolithotrophic iron oxidation in freshwater and marine environments. *Front. Microbiol.* 7:796. doi: 10.3389/fmicb.2016.00796

Chiu, B. K., Kato, S., McAllister, S. M., Field, E. K., and Chan, C. S. (2017). Novel pelagic iron-oxidizing Zeta*proteobacteria* from the chesapeake bay oxic-anoxic transition zone. *Front. Microbiol.* 8:1280. doi: 10.3389/fmicb.2017.01280

Chou, S. H., and Galperin, M. Y. (2016). Diversity of cyclic di-GMP-binding proteins and mechanisms. *J. Bacteriol.* 198, 32–46. doi: 10.1128/jb.00333-15

Clegg, D. O., and Koshland, D. E. (1984). The role of a signaling protein in bacterial sensing: behavioral effects of increased gene expression. *Proc. Natl. Acad. Sci.U.S.A.* 81, 5056–5060. doi: 10.1073/pnas.81.16.5056

D'Argenio, D. A., and Miller, S. I. (2004). Cyclic Di-GMP as a bacterial second messenger. *Microbiology* 150, 2497–2502.

Do, T. T., Quyen, D. T., and Nguyen, T. N. (2013). Molecular characterization of a glycosyl hydrolase family 10 xylanase from *Aspergillus niger*. *Protein Exp. Purif.* 92, 196–202. doi: 10.1016/j.pep.2013.09.011

Dodd, M. S., Papineau, D., Grenne, T., Slack, J. F., Rittner, M., Pirajno, F., et al. (2017). Evidence for early life in earth's oldest hydrothermal vent precipitates. *Nature* 543, 60–64. doi: 10.1038/nature21377

Druschel, G. K., Emerson, D., Sutka, R., Suchecki, P., and Luther, G. W. (2008). Low-oxygen and chemical kinetic constraints on the geochemical niche of neutrophilic iron(ii) oxidizing microorganisms. *Geochim. Cosmochim. Acta* 72, 3358–3370. doi: 10.1016/j.gca.2008.04.035

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Ehrenberg, C. G. (1836). Vorläufige Mitteilungen Über Das Wirkliche Vorkommen Fossiler Infusorien Und Ihre Große Verbreitung. *Ann. Phys.* 114, 213–227. doi: 10.1002/andp.18361140520

Emerson, D., and Moyer, C. L. (1997). Isolation and Characterization Of Novel Iron-Oxidizing Bacteria That Grow At Circumneutral pH. *Appl. Environ. Microbiol.* 63, 4784–4792. doi: 10.1128/aem.63.12.4784-4792.1997

Emerson, D., and Revsbech, N. P. (1994). Investigation of an iron-oxidizing microbial mat community located near Aarhus, Denmark: laboratory studies. *Appl. Environ. Microbiol.* 60, 4032–4038. doi: 10.1128/aem.60.11.4032-4038.1994

Emerson, D., Field, E. K., Chertkov, O., Davenport, K. W., Goodwin, L., Munk, C., et al. (2013). Comparative genomics of freshwater fe-oxidizing bacteria: implications for physiology, ecology, and systematics. *Front. Microbiol.* 4:254. doi: 10.3389/fmicb.2013.00254

Emerson, D., Rentz, J. A., Lilburn, T. G., Davis, R. E., Aldrich, H., Chan, C. S., et al. (2007). A novel lineage of *proteobacteria* involved in formation of marine fe-oxidizing microbial mat communities. *PLoS One* 2:e667.

Emerson, D., Scott, J. J., Leavitt, A., Fleming, E., and Moyer, C. L. (2017). In situ estimates of iron-oxidation and accretion rates for iron-oxidizing bacterial mats at Lō'ihi seamount. *Deep Sea Res. I Oceanogr. Res. Pap.* 126, 31–39. doi: 10.1016/j.dsr.2017.05.011

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575

Field, E. K., Sczyrba, A., Lyman, A. E., Harris, C. C., Woyke, T., Stepanauskas, R., et al. (2015). Genomic insights into the uncultivated marine Zeta*proteobacteria* at loihi seamount. *ISME J.* 9, 857–870. doi: 10.1038/ismej.2014.183

Fields, J. A., and Thompson, S. A. (2008). Campylobacter jejuni CsrA mediates oxidative stress responses, biofilm formation, and host cell invasion. *J. Bacteriol.* 190, 3411–3416. doi: 10.1128/jb.01928-07

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Fullerton, H., Hager, K. W., and Moyer, C. L. (2015). Draft genome sequence of Mariprofundus ferrooxydans strain JV-1, isolated from Loihi Seamount, Hawaii. *Genome Announc.* 3, e01118–15.

Fullerton, H., Hager, K. W., McAllister, S. M., and Moyer, C. L. (2017). Hidden diversity revealed by genome-resolved metagenomics of iron-oxidizing microbial mats from lo'ihi seamount, Hawai'i. *ISME J.* 11, 1900–1914. doi: 10.1038/ismej.2017.40

Garinot-Schneider, C., Lellouch, A. C., and Geremia, R. A. (2000). Identification of essential amino acid residues in the sinorhizobium meliloti glucosyltransferase ExoM. *J. Biol. Chem.* 275, 31407–31413. doi: 10.1074/jbc.m004524200

Garrison, C. A., Price, K. A., and Field, E. K. (2019). Environmental evidence for and genomic insight into the preference of iron-oxidizing bacteria for more-corrosion-resistant stainless steel at higher salinities. *Appl. Environ. Microbiol.* 85, e00483–19.

Guttenplan, S. B., Blair, K. M., and Kearns, D. B. (2010). The EpsE flagellar clutch is bifunctional and synergizes with EPS biosynthesis to promote Bacillus subtilis biofilm formation. *PLoS Genet.* 6:e1001243. doi: 10.1371/journal.pgen.1001243

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, D387–D395.

Harrison, K. J., Crécy-Lagard, V. D., and Zallot, R. (2018). Gene graphics: a genomic neighborhood data visualization web application. *Bioinformatics* 34, 1406–1408. doi: 10.1093/bioinformatics/btx793

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309–316. doi: 10.1042/bj2800309

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314.

Huertas-Rosales, Ó., Romero, M., Heeb, S., Espinosa-Urgel, M., Cámara, M., and Ramos-González, M. I. (2017). The *Pseudomonas* Putida CsrA/RsmA homologues negatively affect c-Di-GMP pools and biofilm formation through the GGDEF/EAL response regulator CfcR. *Environ. Microbiol.* 19, 3551–3566. doi: 10.1111/1462-2920.13848

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

Jonas, K., Edwards, A. N., Simm, R., Romeo, T., Römling, U., and Melefors, Ö (2008). The RNA binding protein CsrA controls cyclic Di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Mol. Microbiol.* 70, 236–257. doi: 10.1111/j.1365-2958.2008.06411.x

Kato, S., Chan, C., Itoh, T., and Ohkuma, M. (2013). Functional Gene analysis of freshwater iron-rich flocs at circumneutral Ph and isolation of a stalk-forming microaerophilic iron-oxidizing bacterium. *Appl. Environ. Microbiol.* 79, 5283–5290. doi: 10.1128/aem.03840-12

Kato, S., Ohkuma, M., Powell, D. H., Krepski, S. T., Oshima, K., Hattori, M., et al. (2015). Comparative genomic insights into ecophysiology of neutrophilic, microaerophilic iron oxidizing bacteria. *Front. Microbiol.* 6:1265. doi: 10.3389/fmicb.2015.01265

Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719. doi: 10.1093/nar/gkn668

Köplin, R., Arnold, W., Hötte, B., Simon, R., Wang, G., and Pühler, A. (1992). Genetics of xanthan production in Xanthomonas campestris: the xanA and

xanB genes are involved in UDP-glucose and GDP-mannose biosynthesis. *J. Bacteriol.* 174, 191–199. doi: 10.1128/jb.174.1.191-199.1992

Korotkov, K. V., Sandkvist, M., and Hol, W. G. (2012). The type II secretion system: biogenesis, molecular architecture and mechanism. *Nat. Rev. Microbiol.* 10, 336–351. doi: 10.1038/nrmicro2762

Krepski, S. T., Emerson, D., Hredzak-Showalter, P. L., Luther, G. W. III , and Chan, C. S. (2013). Morphology of biogenic iron oxides record microbial physiology and environmental conditions: toward interpreting iron microfossils. *Geobiology* 11, 457–471. doi: 10.1111/gbi.12043

Kuchma, S. L., Delalez, N. J., Filkins, L. M., Snavely, E. A., Armitage, J. P., and O'Toole, G. A. (2015). Cyclic Di-GMP-mediated repression of swarming motility by *Pseudomonas aeruginosa* PA14 requires the MotAB stator. *J. Bacteriol.* 197, 420–430. doi: 10.1128/jb.02130-14

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152

Laufer, K., Kappler, A., Nordhoff, M., Jørgensen, B. B., Røy, H., Schmidt, C., et al. (2016). Coexistence of microaerophilic, nitrate-reducing, and phototrophic Fe(II) oxidizers and Fe(III) reducers in coastal marine sediment. *Appl. Environ. Microbiol.* 82, 1433–1447. doi: 10.1128/aem.03527-15

Laufer, K., Nordhoff, M., Halama, M., Martinez, R. E., Obst, M., Nowak, M., et al. (2017). Microaerophilic Fe(II)-Oxidizing Zeta*proteobacteria* isolated from Low-Fe marine coastal sediments?: physiology and composition of their twisted stalks. *Appl. Environ. Microbiol.* 83, e03118–16.

Leimbach, A. (2016). *Bac-Genomics-Scripts: Bovine E. coli Mastitis Comparative Genomics Edition. Zenodo.* Available online at: http://dx.doi.org/10.5281/zenodo.215824 (accessed December 21, 2016).

Lin, C., Larsen, E. I., Nothdurft, L. D., and Smith, J. J. (2012). Neutrophilic, microaerophilic Fe(II)-oxidizing bacteria are ubiquitous in aquatic habitats of a subtropical australian coastal catchment (Ubiquitous FeOB in Catchment Aquatic Habitats). *Geomicrobiol. J.* 29, 76–87. doi: 10.1080/01490451.2010.523446

Little, C. T. S., Johannessen, K. C., Bengtson, S., Chan, C. S., Ivarsson, M., Slack, J. F., et al. (2021). A late Paleoproterozoic (1.74 Ga) deep−sea, low−temperature, iron−oxidizing microbial hydrothermal vent community from Arizona, USA. *Geobiology* 19, 228–249. doi: 10.1111/gbi.12434

Lopez, A., Albino, D., Beraki, S., Broomell, S., Canela, R., Dingmon, T., et al. (2019). Genome sequence of mariprofundus sp. Strain EBB-1, a novel marine autotroph isolated from an iron-sulfur mineral. *Microbiol. Resour. Announc.* 8:e00995-19.

Makita, H., Tanaka, E., Mitsunobu, S., Miyazaki, M., Nunoura, T., Uematsu, K., et al. (2017). Mariprofundus micogutta sp. nov., a novel iron-oxidizing zetaproteobacterium isolated from a deep-sea hydrothermal field at the Bayonnaise knoll of the Izu-Ogasawara arc, and a description of Mariprofundales ord. nov. and Zeta*proteobacteria* classis nov. *Arch. Microbiol.* 199, 335–346. doi: 10.1007/s00203-016-1307-4

Mandon, K., Kaminski, P. A., and Elmerich, C. (1994). Functional analysis of the fixNOQP region of *Azorhizobium caulinodans*. *J. Bacteriol.* 176, 2560–2568. doi: 10.1128/jb.176.9.2560-2568.1994

McAllister, S. M., Barnett, J. M., Heiss, J. W., Findlay, A. J., MacDonald, D. J., Dow, C. L., et al. (2015). Dynamic hydrologic and biogeochemical processes drive microbially enhanced iron and sulfur cycling within the intertidal mixing zone of a beach aquifer. *Limnol. Oceanogr.* 60, 329–345. doi: 10.1002/lno.10029

McAllister, S. M., Moore, R. M., and Chan, C. S. (2018). ZetaHunter, a reproducible taxonomic classification tool for tracking the ecology of the Zeta*proteobacteria* and other poorly resolved taxa. *Microbiolol. Resour. Announc.* 7, e00932–18.

McAllister, S. M., Moore, R. M., Gartman, A., Luther, G. W. III, Emerson, D., and Chan, C. S. (2019). The Fe (II)-oxidizing Zeta*proteobacteria*: historical, ecological and genomic perspectives. *FEMS Microbiol. Ecol.* 95:fiz015.

McAllister, S. M., Polson, S. W., Butterfield, D. A., Glazer, B. T., Sylvan, J. B., and Chan, C. S. (2020). Validating the Cyc2 neutrophilic iron oxidation pathway using meta-omics of Zeta*proteobacteria* iron mats at marine hydrothermal vents. *Msystems* 5, e00553–19.

Medlar, A. J., Törönen, P., and Holm, L. (2018). AAI-Profiler: fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination. *Nucleic Acids Res.* 46, W479–W485.

Méndez-Ortiz, M. M., Hyodo, M., Hayakawa, Y., and Membrillo-Hernández, J. (2006). Genome-wide transcriptional profile of *Escherichia coli* in response to high levels of the second messenger 3′, 5′-cyclic diguanylic acid. *J. Biol. Chem.* 281, 8090–8099. doi: 10.1074/jbc.m510701200

Mercier, C., Durrieu, C., Briandet, R., Domakova, E., Tremblay, J., Buist, G., et al. (2002). Positive role of peptidoglycan breaks in lactococcal biofilm formation. *Mol. Microbiol.* 46, 235–243. doi: 10.1046/j.1365-2958.2002.03160.x

Merritt, J. H., Brothers, K. M., Kuchma, S. L., and O'Toole, G. A. (2007). SadC reciprocally influences biofilm formation and swarming motility via modulation of exopolysaccharide production and flagellar function. *J. Bacteriol.* 189, 8154–8164. doi: 10.1128/jb.00585-07

Morgan, J. L. W., McNamara, J. T., and Zimmer, J. (2014). Mechanism of activation of bacterial cellulose synthase by Cyclic Di-GMP. *Nat. Struct. Mol. Biol.* 21, 489–496. doi: 10.1038/nsmb.2803

Morgan, J. L., McNamara, J. T., Fischer, M., Rich, J., Chen, H. M., Withers, S. G., et al. (2016). Observing cellulose biosynthesis and membrane translocation in crystallo. *Nature* 531, 329–334. doi: 10.1038/nature16966

Morgan, J. L., Strumillo, J., and Zimmer, J. (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* 493, 181–186. doi: 10.1038/nature11744

Mori, J. F., Scott, J. J., Hager, K. W., Moyer, C. L., Küsel, K., and Emerson, D. (2017). Physiological and ecological implications of an iron- or hydrogen-oxidizing member of the Zeta*proteobacteria*, ghiorsea bivora, gen. Nov., Sp. Nov. *ISME J.* 11, 2624–2636. doi: 10.1038/ismej.2017.132

Mumford, A. C., Adaktylou, I. J., and Emerson, D. (2016). Peeking under the iron curtain: development of a microcosm for imaging the colonization of steel surfaces by mariprofundus Sp. DIS-1, an oxygen-tolerant fe-oxidizing bacterium. *Appl. Environ. Microbiol.* 82, 6799–6807. doi: 10.1128/aem.01990-16

Nambu, T., Minamino, T., Macnab, R. M., and Kutsukake, K. (1999). Peptidoglycan-hydrolyzing activity of the FlgJ protein, essential for flagellar rod formation in *Salmonella typhimurium*. *J. Bacteriol.* 181, 1555–1561. doi: 10.1128/jb.181.5.1555-1561.1999

Nawrocki, E. P., Kolbe, D. L., and Eddy, S. R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25, 1335–1337. doi: 10.1093/bioinformatics/btp157

Nesper, J., Hug, I., Kato, S., Hee, C. S., Habazettl, J. M., Manfredi, P., et al. (2017). Cyclic Di-GMP differentially tunes a bacterial flagellar motor through a novel class of CheY-like regulators. *ELife* 6:e28842.

Neuwirth, E. (2014). *Package 'RColorBrewer'. ColorBrewer Palettes.* Vienna: R Foundation for Statistical Computing.

Nguyen, K. B., Sreelatha, A., Durrant, E. S., Lopez-Garrido, J., Muszewska, A., Dudkiewicz, M., et al. (2016). Phosphorylation of spore coat proteins by a family of atypical protein kinases. *Proc. Natl. Acad. Sci.U.S.A.* 113, E3482–E3491.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2016). *Vegan: Community Ecology Package. R Package Version 2.4-3.* Vienna: R Foundation for Statistical Computing.

Omadjela, O., Narahari, A., Strumillo, J., Mélida, H., Mazur, O., Bulone, V., et al. (2013). BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17856–17861. doi: 10.1073/pnas.1314063110

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Picard, A., Kappler, A., Schmid, G., Quaroni, L., and Obst, M. (2015). Experimental diagenesis of organo-mineral structures formed by microaerophilic Fe(II)-oxidizing bacteria. *Nat. Commun.* 6:6277.

Pitcher, R. S., and Watmough, N. J. (2004). The bacterial cytochrome cbb3 oxidases. *Biochim. Biophys. Acta* 1655, 388–399. doi: 10.1016/j.bbabio.2003.09.017

Ploner, A. (2019). *Heatplus: Heatmaps with Row and/or Column Covariates and Colored Clusters, v2. 28.0.*

Preisig, O., Zufferey, R., and Hennecke, H. (1996). The *Bradyrhizobium japonicum* fixGHIS genes are required for the formation of the high-affinity cbb3-type cytochrome oxidase. *Arch. Microbiol.* 165, 297–305. doi: 10.1007/s002030050330

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33(Suppl. 2), W116–W120.

Rebbapragada, A., Johnson, M. S., Harding, G. P., Zuccarelli, A. J., Fletcher, H. M., Zhulin, I. B., et al. (1997). The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for *Escherichia coli* behavior. *Proc. Natl. Acad. Sci.U.S.A.* 94, 10541–10546. doi: 10.1073/pnas.94.20.10541

Rentz, J. A., Kraiya, C., Luther, G. W., and Emerson, D. (2007). Control of ferrous iron oxidation within circumneutral microbial iron mats by cellular activity and autocatalysis. *Environ. Sci. Technol.* 41, 6084–6089. doi: 10.1021/es062203e

Rouli, L., Merhej, V., Fournier, P. E., and Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 7, 72–85. doi: 10.1016/j.nmni.2015.06.005

Rubin-Blum, M., Antler, G., Tsadok, R., Shemesh, E., Austin, J. A. Jr., Coleman, D. F., et al. (2014). First evidence for the presence of iron oxidizing *Zetaproteobacteria* at the levantine continental margins. *PLoS One* 9:e91456. doi: 10.1371/journal.pone.0091456

Ryan, R. P., Lucey, J., O'Donovan, K., McCarthy, Y., Yang, L., Tolker-Nielsen, T., et al. (2009). HD-GYP Domain proteins regulate biofilm formation and virulence in *Pseudomonas Aeruginosa*. *Environ. Microbiol.* 11, 1126–1136. doi: 10.1111/j.1462-2920.2008.01842.x

Sallai, L., and Tucker, P. A. (2005). Crystal structure of the central and C-terminal domain of the σ54-activator ZraR. *J. Struct. Biol.* 151, 160–170. doi: 10.1016/j.jsb.2005.05.006

Sandkvist, M. (2001). Biology of type II secretion. *Mol. Microbiol.* 40, 271–283. doi: 10.1046/j.1365-2958.2001.02403.x

Scharf, B. E., Fahrner, K. A., Turner, L., and Berg, H. C. (1998). Control of direction of flagellar rotation in bacterial?chemotaxis. *Proc. Natl. Acad. Sci.U.S.A.* 95, 201–206. doi: 10.1073/pnas.95.1.201

Schmid, J., Sieber, V., and Rehm, B. (2015). Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. *Front. Microbiol.* 6:496. doi: 10.3389/fmicb.2015.00496

Scott, J. J., Breier, J. A., Luther, G. W., and Emerson, D. (2015). Microbial iron mats at the mid-atlantic ridge and evidence that Zetaproteobacteria may be restricted to iron-oxidizing marine systems. *PLoS One* 10:e0119284. doi: 10.1371/journal.pone.0119284

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shinabarger, D., Berry, A., May, T. B., Rothmel, R., Fialho, A., and Chakrabarty, A. M. (1991). Purification and characterization of phosphomannose isomerase-guanosine diphospho-D-mannose pyrophosphorylase. a bifunctional enzyme in the alginate biosynthetic pathway of *Pseudomonas aeruginosa*. *J. Biol. Chem.* 266, 2080–2088. doi: 10.1016/s0021-9258(18)52212-2

Simm, R., Morr, M., Kader, A., Nimtz, M., and Römling, U. (2004). GGDEF and EAL domains inversely regulate cyclic Di-GMP levels and transition from sessility to motility. *Mol. Microbiol.* 53, 1123–1134. doi: 10.1111/j.1365-2958.2004.04206.x

Singer, E., Emerson, D., Webb, E. A., Barco, R. A., Kuenen, J. G., Nelson, W. C., et al. (2011). *Mariprofundus Ferrooxydans* PV-1 the first genome of a marine Fe(II) oxidizing zetaproteobacterium. *PLoS One* 6:e25386. doi: 10.1371/journal.pone.0025386

Sloup, R. E., Konal, A. E., Severin, G. B., Korir, M. L., Bagdasarian, M. M., Bagdasarian, M., et al. (2017). Cyclic di-GMP and VpsR induce the expression of type II secretion in *Vibrio cholerae*. *J. Bacteriol.* 199, e00106–17.

Tal, R., Wong, H. C., Calhoon, R., Gelfand, D., Fear, A. L., Volman, G., et al. (1998). Three cdg operons control cellular turnover of cyclic Di-GMP in *Acetobacter Xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol.* 180, 4416–4425. doi: 10.1128/jb.180.17.4416-4425.1998

Tao, F., Swarup, S., and Zhang, L. H. (2010). Quorum sensing modulation of a putative glycosyltransferase gene cluster essential for Xanthomonas campestris biofilm formation. *Environ. Microbiol.* 12, 3159–3170. doi: 10.1111/j.1462-2920.2010.02288.x

Taylor, B. L., and Zhulin, I. B. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* 63, 479–506. doi: 10.1128/mmbr.63.2.479-506.1999

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus Agalactiae*: implications for the microbial 'Pan-Genome'. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955.

Theodorou, M. C., Theodorou, E. C., and Kyriakidis, D. A. (2012). Involvement of AtoSC two-component system in *Escherichia coli* flagellar regulon. *Amino acids* 43, 833–844. doi: 10.1007/s00726-011-1140-7

Valentini, M., and Filloux, A. (2016). Biofilms and cyclic Di-GMP (c-Di-GMP) signaling: lessons from *Pseudomonas Aeruginosa* and other bacteria. *J. Biol. Chem.* 291, 12547–12555. doi: 10.1074/jbc.r115.711507

Vander Roost, J., Thorseth, I. H., and Dahle, H. (2017). Microbial analysis of zetaproteobacteria and co-colonizers of iron mats in the troll wall vent field, arctic mid-ocean ridge. *PLoS One* 12:e0185008. doi: 10.1371/journal.pone.0185008

Vermassen, A., Leroy, S., Talon, R., Provot, C., Popowska, M., and Desvaux, M. (2019). Cell wall hydrolases in bacteria: insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan. *Front. Microbiol.* 10:331. doi: 10.3389/fmicb.2019.00331

Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. (2015). Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* 23, 148–154. doi: 10.1016/j.mib.2014.11.016

Vollmer, W., Joris, B., Charlier, P., and Foster, S. (2008). Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol. Rev.* 32, 259–286. doi: 10.1111/j.1574-6976.2007.00099.x

Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., and Huber, W. (2016). *Package 'gplots'. Various R Programming Tools for Plotting Data*. Available online at: https://cran.r-project.org/web/packages/gplots/index.html

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Wickham, H., Chang, W., and Wickham, M. H. (2016). *Package 'Ggplot2.' Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2*:1–189.

Wilmes, M., Meier, K., Schiefer, A., Josten, M., Otten, C. F., Klöckner, A., et al. (2017). AmiD is a novel peptidoglycan amidase in Wolbachia endosymbionts of *Drosophila melanogaster*. *Front. Cell. Infect. Microbiol.* 7:353. doi: 10.3389/fcimb.2017.00353

Wong, H. C., Fear, A. L., Calhoon, R. D., Eichinger, G. H., Mayer, R., Amikam, D., et al. (1990). Genetic organization of the cellulose synthase operon in *Acetobacter xylinum*. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8130–8134. doi: 10.1073/pnas.87.20.8130

Yuan, J., Fahrner, K. A., Turner, L., and Berg, H. C. (2010). Asymmetry in the clockwise and counterclockwise rotation of the bacterial flagellar motor. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12846–12849. doi: 10.1073/pnas.1007333107

Zhang, Y., Wei, C., Jiang, W., Wang, L., Li, C., Wang, Y., et al. (2013). The HD-GYP domain protein RpfG of Xanthomonas oryzae pv. oryzicola regulates synthesis of extracellular polysaccharides that contribute to biofilm formation and virulence on rice. *PLoS One* 8:e59428. doi: 10.1371/journal.pone.0059428