



Published in final edited form as:

Curr Microbiol. 2016 May ; 72(5): 503–510. doi:10.1007/s00284-015-0964-x.

Conservation of the essential genome among *Caulobacter* and *Brevundimonas* species

Derrick Scott* and Bert Ely

Department of Biological Sciences, University of South Carolina, Columbia SC 29208, USA

Abstract

When the genomes of *Caulobacter* isolates NA1000 and K31 were compared, numerous genome rearrangements were observed. In contrast, similar comparisons of closely related species of other bacterial genera revealed nominal rearrangements. A phylogenetic analysis of the 16S rRNA indicated that K31 is more closely related to *Caulobacter henricii* CB4 than to other known *Caulobacters*. Therefore, we sequenced the CB4 genome and compared it to all of the available *Caulobacter* genomes to study genome rearrangements, discern the conservation of the NA1000 essential genome, and address concerns about using 16S rRNA to group *Caulobacter* species. We also sequenced the novel bacteria, *Brevundimonas* DS20, a representative of the genus most closely related to *Caulobacter* and used it as part of an outgroup for phylogenetic comparisons. We expected to find that there would be fewer rearrangements when comparing more closely related *Caulobacters*. However we found that relatedness was not correlated with the amount of observed “genome scrambling”. We also discovered that nearly all of the essential genes previously identified for *C. crescentus* are present in the other *Caulobacter* genomes and in the *Brevundimonas* genomes as well. However, a few of these essential genes were only found in NA1000, and some were missing in a combination of one or more species, while other proteins were 100% identical across species. Also, phylogenetic comparisons of highly conserved genomic regions revealed clades similar to those identified by 16S rRNA-based phylogenies, verifying that 16S rRNA sequence comparisons are a valid method for grouping *Caulobacters*.

INTRODUCTION

Alphaproteobacteria comprise a large and metabolically diverse group of bacteria that includes the genus *Caulobacter*. *Caulobacters* are found in essentially all habitats ranging from fresh and salt water, soil, root systems, and water treatment plants. They thrive in low nutrient conditions and exhibit a rare dimorphic phenotype consisting of a stalked non-motile cell and a motile swarmer cell produced at cell division. The motile cell is immature and must first shed its flagellum and differentiate into the stalked form before it replicates its chromosome and divides asymmetrically to regenerate itself and produce a flagellated daughter cell, thus continuing its life cycle. The ability to synchronize this cell cycle has allowed great advancements towards comprehending the genetic regulatory network and signal transduction pathway controlling the *C. crescentus* cell cycle [12, 29].

* corresponding author: scottdc@mailbox.sc.edu.

Compared to the wealth of information available to support cell cycle research, the amount of research dedicated to understanding the environmental and evolutionary biology of Caulobacters is minimal. However, as the genomic sequences of more Caulobacters are becoming available, a significant opportunity has arisen to add to this literature. Ribosomal RNA analyses show that bacteria previously defined as *Caulobacter* are actually grouped into two separate branches consisting of freshwater and marine species, *Caulobacter* and *Maricaulis*, respectively [1, 27]. Further 16S rDNA comparisons by Abraham *et al.* [1] revealed that the freshwater branch is clearly defined into two genera, *Caulobacter* and *Brevundimonas*. Thus *Brevundimonas* genomes are ideal for use as an outgroup for the analysis of *Caulobacter* genomes. The genus *Caulobacter* can be divided into two branches as well, based on their 16S rRNA gene sequences [1]. One branch contains *C. crescentus* and *C. segnis* while the other contains *C. henricii* and *Caulobacter sp.* K31 (Fig. 1). This separation influenced us to compare these genomes to the essential genome that has been experimentally defined for the *C. crescentus* strain NA1000 [6]. The genomic DNA sequences of both *C. crescentus* strain CB15 and its derivative NA1000 [21] and *C. segnis* strain TK0059 have been published [5; 23]. In addition, *Caulobacter* strain K31, a groundwater isolate of particular interest for its ability to tolerate and degrade chlorophenols [18], has also had its sequence elucidated [2]. To provide a fourth strain for this genome comparison, the genome nucleotide sequence of *C. henricii* strain CB4 [24] was determined as part of this study. Although many other *Caulobacter* isolates are listed in the IMG genome database [13], no other *Caulobacter* genome sequences have been fully assembled. Similarly, *Brevundimonas subvibrioides* strain CB81 is the only *Brevundimonas* with an available genome sequence [5]. Therefore, we have determined the nucleotide sequence of the *Brevundimonas* DS20 genome to provide a second *Brevundimonas* genome. Comparing these six genomes, we found that the experimentally determined *C. crescentus* essential genome [6] was conserved in all six species with minor exceptions. However, we found extensive genome rearrangements among these six genomes.

METHODS

Media and Growth conditions

The *Caulobacter henricii* strain CB4 (ATCC 15253) was obtained from the American Type Culture Collection. It was grown at 30°C for 48 hours in PYE medium [14] that contains 2 g Bacto Peptone, 1 g Yeast Extract, 0.5M MgSO₄, and 0.5M CaCl₂ per L. In addition, we isolated a *Caulobacter*-like bacterium from a contaminated culture of *Caulobacter* FWC20 [16]. It was grown at 30°C under the same conditions as CB4. Based on genome comparisons, we determined this bacterium to be a novel member of the genus *Brevundimonas*. We named this isolate *Brevundimonas* DS20.

Genome sequence determination and annotation

Genomic DNA from *Caulobacter henricii* CB4 and *Brevundimonas* DS20 was isolated from a saturated PYE culture using the Qiagen DNeasy Tissue Kit following the manufacturer's protocol. Primers 16S_533F (GTGCCAGCMGCCGCGGTAA and 16S_U1492R (GGTTACCTTGTTACGACTT) were used to amplify the 16S rRNA region of the genome and the amplified DNA was sequenced using Sanger Sequencing on an ABI 3730 sequencer.

Genomic DNA library construction and nucleotide sequencing were carried out by the University of Washington Pacbio Sequencing Services using the Pacific Biosciences RSII sequencing system. The library prep template for the 10 kb protocol was used but the DNA was sheared for 20 kb fragments using a Covaris tube and a final 0.4x bead wash for a finished library. The collection protocols for the P4-C2 chemistry were:

Protocol: MagBead

Standard Seq v2

Movie Time: 120 min

Insert Size (bp): 20000

Stage Start: True

Control: DNA Control 3kb-10 kb.

The average read length for the CB4 DNA was 4289 bp with approximately 55X coverage [26]. The nucleotide sequence reads were assembled using HGAP2.0 as previously described [26]. For both strains the nucleotide sequence data were assembled into a single chromosomal contig. In addition, a 97894 bp circular plasmid was predicted to be present in the CB4 strain. To verify the chromosome assemblies, genomic DNA cut with the *PmeI*, *SnaBI*, or *SwaI* restriction enzymes was separated by Pulsed Field Gel Electrophoresis (PFGE) using the protocols described by [9]. For both genomes, the observed bands exactly matched the bands predicted from the assembled chromosomal sequence. For the CB4 genome, a *SspI* digest contained an extra band that was approximately 100 kb. This band confirmed the existence of the CB4 plasmid since it corresponded to the predicted 98 kb size of the plasmid. Annotations were performed using The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST) [22].

The resulting genome sequences and annotation are available in the NCBI database: *Caulobacter henricii* CB4 (accession numbers CP013002 and CP013003) and *Brevundimonas* DS20 (accession number CP012897).

Genome comparisons

Whole genome comparisons were performed using Progressive *MAUVE* Multiple Genome Alignment [8]. A BLAST comparison of the 480 experimentally-identified *C. crescentus* essential genes [6] to the predicted genes in the other five genomes was performed using the BlastStation version 1.3 software [31] to determine if homologous genes were present in the genomes of the *Caulobacter* strains CB4, TK0059, and K31 and the *Brevundimonas subvibrioides* CB81 and *Brevundimonas* DS20. BLAST matches with an e-value that was less than e^{-5} were considered significant. Phylogenetic and molecular evolutionary analyses were conducted using *MEGA* version 5.1 [28]. Phylogenetic trees were constructed using the maximum likelihood method [10]. All branches were recovered in both neighbor-joining and maximum-parsimony trees [11, 25].

RESULTS AND DISCUSSION

Growth characteristics of CB4 and DS20

Caulobacter henricii CB4 [24] forms yellow colonies and can grow on minimal media glucose plates [14] in the presence of vitamin B₁₂. *Brevundimonas* DS20 forms yellow mucoidal colonies which are round, smooth, slightly raised, and glistening. It is able to grow on minimal media and the cells are rod shaped and lack the curved phenotype found in many *Caulobacters*. At 30°C, CB4 had a doubling time of 190 minutes when grown in PYE. In addition it had a doubling time of two weeks at 10°C. DS20 grew faster with a doubling time of 120 minutes at 30°C and 12 days at 10°C. When grown at 10°C light microscopy of both strains indicated that they appeared healthy with highly motile swarmer cells.

Genome overview

The assembly of the *Caulobacter henricii* CB4 genome resulted in a 3,864,204 bp chromosome and a 93,0084 bp plasmid (Table 1). The CB4 chromosome contains 3751 genes and has a GC content of 66.4%. As such, the codons in the protein coding regions should have a high G+C content, especially in the third codon position (GC3). Indeed, 29 of the 30 most used codons contain either a G or a C in the third position. The overall GC3 percentage for CB4 is 83.2%. The plasmid has a GC content of 65.4 % and a GC3 content of 80.8%. It contains one integrase gene but no transposases among its 97 genes. Since the K31 genome contains two megaplasmids [2], we compared the predicted amino acid sequences of the CB4 plasmid genes to those of the K31 plasmids and found that only four of the plasmid genes are homologous to any of the genes in either of the two K31 plasmids. However, the plasmid contains 11 heavy metal resistance genes (*czc*) in addition to 15 *czc* genes on the chromosome. Despite the extra *czc* genes, CB4 is resistant to the same concentrations of cadmium and zinc as NA1000 which has only 13 *czc* genes. Nearly all of the other genes that code for proteins with predicted functions code for other types of metal resistance or for plasmid functions. No other plasmids have been reported in any other *Caulobacter* genome.

The *Brevundimonas* DS20 genome consists of 3,457,610 bp and does not include a plasmid (Table 1). It has a GC content of 67% and contains 3411 genes. As in CB4, 29 of the 30 most used codons contain either a G or a C in the third position with an overall GC3 percentage is 86.3%. The pattern of codon usage is similar to that of CB4 as well. The crescentin gene responsible for the crescent cell shape in most *Caulobacter* species [20] was not found in the DS20 genome.

Genome Rearrangements

The gene order of closely related species is usually very similar and often the assembled genome of one species can be used as the template for the alignment of closely related species. However, a previous comparison of NA1000 and K31 revealed rearrangements an order of magnitude greater than previously described in other bacteria [2]. When the K31 chromosome was aligned to that of NA1000, more than 60 inversions and 45 large translocations were readily observed. Since this level of genome scrambling makes it difficult to identify the endpoints of individual inversion events, we hypothesized that the

level of observed genome scrambling would decrease in comparisons of more closely related genomes based on phylogenetic comparisons (Fig. 1).

When the *C. segnis* TK0059 genome was compared to the NA1000 genome, only 35 inversions and 11 translocations were estimated from the Mauve comparison (Fig. 2). Since TK0059 is more closely related to NA1000 than to K31 [1], this reduced level of genome scrambling was consistent with the hypothesis that two closely related strains might have a small enough number of gene inversions that individual events could be accurately identified. However, when we compared the *Caulobacter henricii* CB4 genome to the closely related K31 genome, we observed more than 75 inversions and over 45 translocation events (Fig. 3). Most of these translocations were small with only five being over 100,000 bp. The rearrangements were also mostly organized around the origin of replication as shown previously for the NA1000 and K31 comparison [2]. Intriguingly, when compared to the more distantly-related TK0059 genome, the CB4 genome had only three large translocations and fewer inversions than we observed in the CB4/K31 comparison (data not shown). Thus the number of inversions and translocations appears to be unrelated to genetic distance. The two *Brevundimonas* also exhibit these high levels of genome rearrangement (Fig. 4).

The *Caulobacter* Essential Genome

Previous experimental work had identified 480 protein coding regions that are essential for the growth of *C. crescentus* strain NA1000 in a nutrient rich medium [6]. We hypothesized that genes that were essential for growth of NA1000 in PYE would also be essential and highly conserved for the other five bacterial strains in this study as well. We used Blaststation software to BLAST the 480 essential genes against the CDS regions of the other genomes and determined that 94% of the genes coded for proteins that had homologs in the other five species. In fact, the SSU ribosomal protein S10P had 100% amino acid identity in all four *Caulobacter* species, another ribosomal protein and IF-1 had 100% identity in three species, and 17 more were 100% identical in two *Caulobacter* species. Most of these highly conserved genes code for ribosomal proteins where amino acid sequence conservation is expected due to the fact that these proteins bind to an rRNA and to each other to form a very precise protein manufacturing machine.

There were nine NA1000 essential genes that were absent in all five of the other genomes (Table 2). Four of these essential genes, CCNA 761, 1304, 2841, and 3307, have an unknown function, CCNA 2844 codes for an antitoxin protein, and CCNA 465, 466, 467, and 469 code for proteins involved in cell wall synthesis. In addition, 8 other *C. crescentus* essential genes are present in at least one other species, but are missing in at least one other species. Three of the 8 are present in all four of the *Caulobacter* genomes but not in the two *Brevundimonas* genomes. Two other genes that are present in only some of the *Caulobacter* genomes also code for antitoxin proteins. It is unsurprising that an antitoxin protein would be essential. These genes code for proteins that neutralize a specific toxin. The absence of these antitoxin genes paired with the presence of the corresponding toxin gene would prove fatal for the organism, but the gene is not needed if the toxin gene is not present. The four essential genes involved in cell wall synthesis (CCNA 465, 466, 467, and 469) are contained in a region annotated as a prophage element (coordinates 473044 to 499074) in the NA1000

genome. It is unlikely that genes gained from a prophage would become essential unless they are needed to protect against lethality due to some other part of the prophage. In fact numerous antisense transcripts have been detected in this region (Schrader et al. 2014) so complex regulatory circuits may be present. This phenomenon appears to have occurred in *C. segnis* TK0059 with the NA1000 essential gene IF-2 that is found in every genome that we compared except for the TK0059 genome. However, the TK0059 genome contains an alternate gene Cseg_3298 that codes for a protein that was predicted to function as IF-2 even though it has an unrelated amino acid sequence. In addition, the gene has a GC content of only 55% suggesting that it was recently obtained by horizontal gene transfer. However, more testing needs to be done to verify that this TK0059 protein can actually function as an IF-2 translation initiation factor.

Phylogenetic relationships

Comparative sequence analysis of the 16S ribosomal RNA genes is currently the most widely used approach for the reconstruction of microbial phylogeny since the rRNA operon size, nucleotide sequence, and secondary structures of the three rRNAs (16S, 23S, 5S) are highly conserved within a bacterial species [17]. The 16S rRNA is the most conserved of these subunits and has been used widely as a sort of “evolutionary clock” [30]. A comparison of the 16S rRNA nucleotide sequences shows within genus differences ranging up to 3% and between genus differences in the range of 5-7% (Table 3). However, a single gene cannot be used to assess genome divergence since different parts of the genome diverge at different rates.

Similarly whole genome comparisons only provide an average rate of divergence. Therefore, we decided to perform a phylogenetic analyses of two large clusters of genes that span thousands of base pairs: the divisional cell wall (dcw) cluster containing 26 genes (CCNA_2622-2647) involved in cell division and cell wall synthesis [3] and a ribosomal protein region containing 28 genes (CCNA_1304-1332). These conserved regions are large enough for robust statistical comparisons and provide the opportunity to study the divergence of individual functional units. The nucleotide sequences of the *Caulobacter* dcw operon differ by as much 19% in pairwise comparisons of the *Caulobacter* species and by 27% when the *Caulobacter* operons were compared with those of the two Brevundimonads (Table 4). However, the phylogenetic tree of the dcw gene cluster was essentially the same as the 16S tree.

The *Caulobacter* ribosomal protein operons differed by as much 10% in pairwise comparisons of the nucleotide sequences and 21% when compared with those of the Brevundimonads (Table 5). Thus this region is more highly conserved than the dcw operon probably because the amino acid sequences of ribosomal proteins are more constrained since they are involved in complex intermolecular interactions. In both comparisons, the nucleotide sequences were much more diverse than the 16S rDNA sequences, but the phylogenetic trees were essentially the same

We also compared a conserved prophage region that spans approximately 20 genes (CCNA_2861-2880) and found that the nucleotide sequence differs by as much 17% in pairwise comparisons among the *Caulobacters* (Table 6). Thus the divergence of the

Caulobacter prophage gene nucleotide sequences appears to be slightly less than that of the *dcw* operon even though a selective constraint is unknown. However, none of the *Caulobacter* prophage nucleotide sequences had significant identity ($< e10^{-5}$) to the corresponding *Brevundimonas* sequences.

Upon closer inspection, we found that there was significant amino acid identity among the genes in this region in all six genomes. Part of the explanation for the disparity between the nucleotide and amino acid sequence identities may be that the *Caulobacter* phage regions display a codon usage bias for CTG (Leucine), GGG (Glycine), GCG (Alanine), and CGG (Arginine) in contrast to the *Brevundimonas* phage regions which display a bias towards CTC (Leucine), GGC (Glycine) CGC (Arginine), and GCC (Alanine). Interestingly, this difference in codon usage bias was not observed when the genomic codon usage was compared. This difference in codon usage would facilitate diversity at the nucleotide level while conserving the amino acid sequence. There may be some environmental or evolutionary pressure that is influencing the conservation of these genes, but it is not obvious. Also, even though the *Brevundimonas subvibrioides* genome contained all genes of the conserved phage operon, we detected two translocations of genes to locations away from this region. Three genes were grouped together in what appears to be a four gene operon along with a recombinase gene that is absent in four of the strains in this study but is found in K31. A fourth gene was found in a different four gene operon along with three other genes not found in any of the other bacteria in our study. Since these translocated genes were found 21,084 and 25,619 base pairs before the start of the prophage region, it is also possible that there was a single translocation event followed by an insertion between the first and second genes. In either case, we can conclude that the prophage region was present in the common ancestor of *Caulobacter* and *Brevundimonas* and has remained intact until recently despite the high level of other genome rearrangements observed in these species. Furthermore, when phylogenetic trees were produced using this set of 20 prophage genes, the resulting trees were similar to the ribosomal RNA gene trees except that the K31 and CB4 genes did not form a monophyletic group.

Conclusions

In summary, we found that despite the extensive scrambling of the *Caulobacter* and *Brevundimonas* genomes, most genes shown previously to be essential for *C. crescentus* [6] are highly conserved in other species of *Caulobacter* and *Brevundimonas*. Also, a prophage present in the genomes of both genera exhibits conserved amino acid sequences in the protein coding regions, but altered codon use between the two genera even though the overall codon usage patterns are very similar. Thus, the prophage genes may be subject to some unrecognized evolutionary pressure.

ACKNOWLEDGEMENTS

This work was funded in part by a fellowship from The Southern Region Educational Board (SREB) to DS and by NIH grant GM076277 and NSF grant EF-0826792 to BE.

REFERENCES

1. Abraham WR, Strompl C, Meyer H. Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis gen. nov.* with *Maricaulis maris* (Poindexter) *comb. nov.* as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*. *Int J Syst Bacteriol.* 1999; 49(Pt 3):1053–1073. other authors. [PubMed: 10425763]
2. Ash K, Brown T, Watford T, Scott LE, Stephens C, Ely B. A comparison of the *Caulobacter* NA1000 and K31 genomes reveals extensive genome rearrangements and differences in metabolic potential. *Open Biology.* 2014; 4:140128. [PubMed: 25274120]
3. Ayala, JA.; Garrido, T.; de Pedro, MA.; Vincente, M. Molecular biology of bacterial septation. *Bacterial cell wall.* In: Ghuysen, JM.; Hakenbeck, R., editors. Elsevier; Amsterdam: 1994. 1994. p. 73-101. p. 73-101. Molecular biology of bacterial septation. *Bacterial cell wall*
4. Beare PA, Unsworth N, Andoh M. Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus *Coxiella*. *Infect Immun.* 2009; 77:642–656. other authors. [PubMed: 19047403]
5. Brown PJ, Kysela DT, Buechlein A, Hemmerich C, Brun YV. Genome sequences of eight morphologically diverse Alphaproteobacteria. *J Bacteriol.* 2011; 193:4567–4568. [PubMed: 21705585]
6. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Collier JA, Fero MJ, McAdams HH, Shapiro L. The essential genome of a bacterium. *Mol Syst Biol.* 2011; 7:528. [PubMed: 21878915]
7. Darling AE, Miklos I, Ragan MA. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet.* 2008; 4:e1000128. [PubMed: 18650965]
8. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 2010; 5:e11147. [PubMed: 20593022]
9. Ely B, Gerardot CJ. Use of pulsed-field-gradient gel electrophoresis to construct a physical map of the *Caulobacter crescentus* genome. *Gene.* 1988; 68:323–333. [PubMed: 2851498]
10. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* 1981; 17:368–376. [PubMed: 7288891]
11. Fitch WM. Towards defining the course of evolution: minimum change for a specific tree topology. *Syst Zool.* 1971; 20:406–416.
12. Holtzendorff J, Hung D, Brende P, Reisenauer A, Viollier PH, McAdams HH, Shapiro L. Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science.* 2004; 304:983–987. [PubMed: 15087506]
13. img.jgi.doe.gov img.jgi.doe.gov
14. Johnson RC, Ely B. Isolation of spontaneously derived mutants of *Caulobacter crescentus*. *Genetics.* 1977; 86:25–32. [PubMed: 407126]
15. Konstantinidis KT, Ramette A, Tiedje JM. The bacterial species definition in the genomic era. *Philos Trans R Soc Lond B Biol Sci.* 2006; 361:1929–1940. [PubMed: 17062412]
16. MaCrae JD, Smit J. Characterization of *Caulobacters* Isolated from Wastewater Treatment Systems. *Applied and Environmental Microbiology.* 1991; 57:751–758.
17. Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR. The RDP (Ribosomal Database Project). *Nucleic Acids Res.* 1997; 25:109–111. [PubMed: 9016515]
18. Mannisto MK, Tirola MA, Salkinoja-Salonen MS, Kulomaa MS, Puhakka JA. Diversity of chlorophenol-degrading bacteria isolated from contaminated boreal groundwater. *Arch Microbiol.* 1999; 171:189–197. [PubMed: 10201097]
19. McDonald IR, Kampfer P, Topp E. *Aminobacter ciceronei* sp. nov. and *Aminobacter lissarensis* sp. nov., isolated from various terrestrial environments. *Int J Syst Evol Microbiol.* 2005; 55:1827–1832. other authors. [PubMed: 16166673]
20. Møller-Jensen J, Löwe J. Increasing complexity of the bacterial cytoskeleton. *Curr. Opin. Cell Biol.* 2005; 17(1):75–81. [PubMed: 15661522]
21. Nierman WC, Feldblyum TV, Laub MT. Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci U S A.* 2001; 98:4136–4141. other authors. [PubMed: 11259647]

22. Overbeek R, Olson R, Pusch GD. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014; 42:D206–214. other authors. [PubMed: 24293654]
23. Patel S, Fletcher B, Scott DC, Ely B. Genome sequence and phenotypic characterization of *Caulobacter segnis*. *Current Microbiology.* 2015; 70:355–363. [PubMed: 25398322]
24. Poindexter JS. Biological Properties and Classification of the *Caulobacter* Group. *Bacteriol Rev.* 1964; 28:231–295. [PubMed: 14220656]
25. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4:406–425. [PubMed: 3447015]
26. Scott D, Ely B. Comparison of genome sequencing technology and assembly methods for the analysis of a GC-rich bacterial genome. *Current Microbiology.* 2015; 70:338–344. [PubMed: 25377284]
27. Stahl DA, Key R, Flesher B, Smit J. The phylogeny of marine and freshwater caulobacters reflects their habitat. *J Bacteriol.* 1992; 174:2193–2198. [PubMed: 1551840]
28. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
29. Wang SP, Sharma PL, Schoenlein PV, Ely B. A histidine protein kinase is involved in polar organelle development in *Caulobacter crescentus*. *Proc Natl Acad Sci U S A.* 1993; 90:630–634. [PubMed: 8421698]
30. Woese CR. Bacterial evolution. *Microbiol Rev.* 1987; 51:221–271. [PubMed: 2439888]
31. www.blaststation.comwww.blaststation.com
32. Schrader JM, Zhou B, Li G-W, Lasker K, Childers WS, et al. The Coding and Noncoding Architecture of the *Caulobacter crescentus* Genome. *PLoS Genet.* 2014; 10(7):e1004463. doi: 10.1371/journal.pgen.1004463. [PubMed: 25078267]

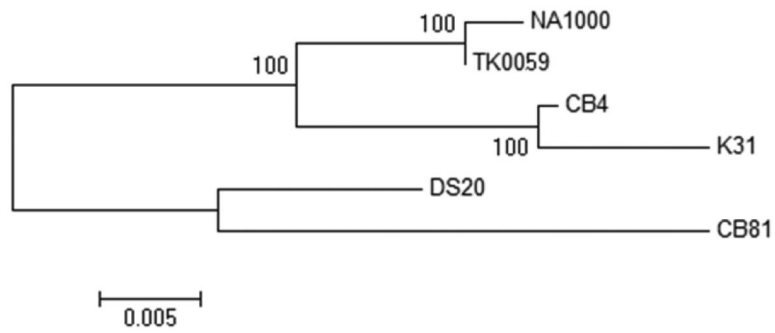


Fig. 1. Phylogenetic tree indicating the relationship between *C. crescentus* NA1000, *C. segnis* TK0059, *Caulobacter* K31, and *C. henricii* CB4 based on 16S rRNA gene sequence analysis. The tree was constructed using the maximum likelihood method [10]. *Brevundimonas subvibrioides* CB81 and *Brevundimonas* DS20 were used as outgroup taxa. All branches were recovered in both neighbor-joining and maximum-parsimony trees [11], [25]. Bootstrap values greater than 50% are given at branch points. Bar, 0.005 substitutions per nucleotide position.

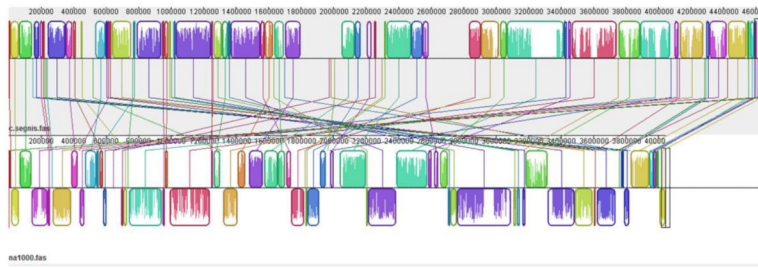


Fig. 2. *MAUVE* comparison of *C. segnis* TK0059 (top) with NA1000 (bottom).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

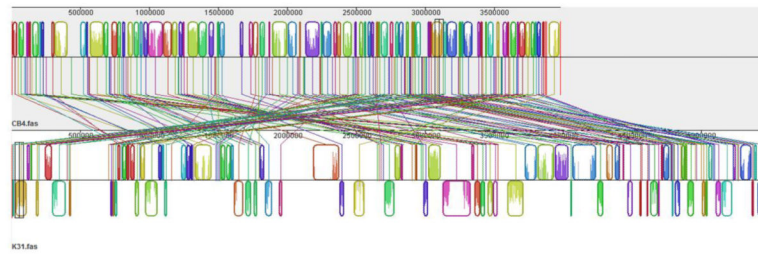


Fig. 3.
MAUVE alignment of CB4 (top) and K31 (bottom).

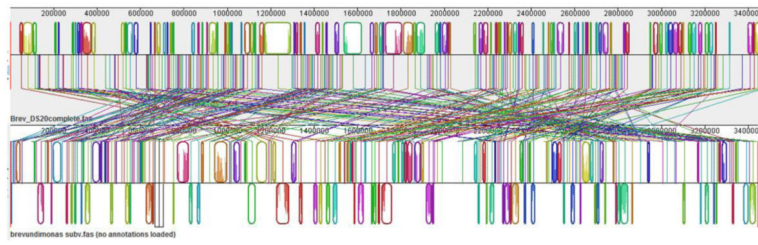


Fig. 4. MAUVE alignment of *Brevundimonas* DS20 (top) and *B. subvibrioides* CB81 (bottom).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1Features of the *Caulobacter* and *Brevundimonas* genomes

Genome Features	C. segnis TK0059	C. crescentus NA1000	Caulobacter sp. K31	C. henricii CB4	B. subvibriodes CB81	Brevundimonas sp. DS20
Base Pairs (Mb)	4.66	4.02	5.48	3.86	3.45	3.46
Plasmids	0	0	2	1	0	0
G/C content (%)	67.7	67.2	68.1	66.4	68.4	67.4
Protein-coding Genes	~4250	~3800	~5450	~3628	~3331	~3411

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Essential NA1000 genes not present in TK0059, CB4, K31, DS20, or CB81

NA1000 gene	Annotation	Distribution in other strains
CCNA_00465	UDP-galactopyranose mutase	none
CCNA_00466	glycosyltransferase	none
CCNA_00467	oligosaccharide translocase/flippase	none
CCNA_00469	glycosyltransferase	none
CCNA_00761	hypothetical protein	none
CCNA_01304	hypothetical protein	none
CCNA_02841	hypothetical protein	none
CCNA_02844	antitoxin protein parD-3	K31
CCNA_03307	hypothetical protein	none
CCNA_03630	socA antitoxin protein	none
CCNA_03474	SpoVT-AbrB transcription factor, phd antitoxin	CB4, K31
CCNA_00364	deoxyhypusine synthase	All but DS20
CCNA_01211	hypothetical protein	TK0059, CB4, K31
CCNA_01380	pole-organizing protein popZ	TK0059, CB4, K31
CCNA_02294	argininosuccinate lyase	All but DS20
CCNA_02644	cell division protein FtsL	TK0059, CB4, K31
CCNA_03213	polyhydroxyalkanoic acid system protein	TK0059, CB4, K31
CCNA_03277	glycosyltransferase	TK0059, CB4, K31
CCNA_03339	TolA protein	TK0059, CB4, K31
CCNA_03274	hypothetical protein	no not translated
CCNA_00684	anti-sigma factor chrR	TK0059, K31, DS20
CCNA_01864	transcriptional regulator, TetR family	TK0059, K31, DS20, CB81
CCNA_00041	bacterial translation initiation factor 2 IF-2	All but TK0059

Table 3

A comparison of 16S rRNA nucleotide sequences among the species included in this study (percent identity).

	NA1000	TK0059	CB4	K31	CB81	DS20
NA1000	100%	99%	98%	97%	93%	94%
TK0059	99%	100%	98%	97%	94%	95%
CB4	98%	98%	100%	99%	93%	94%
K31	97%	97%	99%	100%	93%	93%
CB81	93%	94%	93%	93%	100%	97%
DS20	94%	95%	94%	93%	97%	100%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

A comparison of dcw cluster nucleotide sequences among the species included in this study (percent identity).

	NA1000	TK0059	CB4	K31	CB81	DS20
NA1000	100%	88%	82%	83%	73%	75%
TK0059	88%	100%	82%	81%	75%	73%
CB4	82%	82%	100%	84%	76%	76%
K31	83%	81%	84%	100%	73%	75%
CB81	73%	75%	76%	73%	100%	79%
DS20	75%	73%	76%	75%	79%	100%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

A comparison of ribosomal protein operon nucleotide sequences among the species included in this study (percent identity)

	NA1000	TK0059	CB4	K31	CB81	DS20
NA1000	100%	96%	90%	90%	79%	80%
TK0059	96%	100%	91%	90%	79%	80%
CB4	90%	91%	100%	93%	79%	80%
K31	90%	90%	93%	100%	79%	80%
CB81	79%	79%	79%	80%	100%	86%
DS20	80%	80%	80%	80%	86%	100%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6

A comparison of conserved phage region nucleotide sequences among the species included in this study (percent identity).

	NA1000	TK0059	CB4	K31
NA1000	100%	83%	90%	87%
TK0059	83%	100%	84%	87%
CB4	90%	84%	100%	75%
K31	87%	87%	75%	100%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript