Clinical Microbiology: Open Access

Koskulu et al., Clin Microbiol 2017, 6:1
DOI: 10.4172/2327-5073.1000275

Research Article Open Access

Identification and Characterization of Replication Origins of Multiple Chromosomes in *Rhodobacter sphaeroides*

Berra Koskulu[1], **Abha Choudhary**[1], **Hannah Johnson**[1], **Hyuk Cho**[2] **and Madhusudan Choudhary**[1*]

[1]*Department of Biological Sciences, Sam Houston State University, Huntsville, TX 77341, USA*

[2]*Department of Computer Science, Sam Houston State University, Huntsville, TX 77341, USA*

[*]**Corresponding author:** Madhusudan Choudhary, Associate Professor, Department of Biological Sciences Sam Houston State University, Huntsville, Texas 77341, USA, Tel: 936-294-4850; Fax: 935-294-1540; E-mail: mchoudhary@shsu.edu

Received date: December 22, 2016; **Accepted date:** January 31, 2017; **Published date:** February 10, 2017

Abstract

DNA replication has been extensively studied in a number of bacterial species, which possess a unipartite genome consisting of a single circular chromosome. Approximately 10% of sequenced bacterial species have a multipartite genome structure, which is comprised of multiple chromosomes. However, the coordination and regulation of multi-chromosomal replication in bacteria remains poorly understood. *Rhodobacter sphaeroides* possesses a multipartite genome consisting of two chromosomes, the primary chromosome (CI) of approximately 3Mb and the secondary chromosome (CII) of 0.9 Mb. Z-curve and GC skew analyses revealed that CI and CII of *R. sphaeroides* exhibit three and five putative chromosomal origin regions, respectively. Then, the flanking regions of these putative regions were further analyzed in terms of gene conservation, gene density, and gene ratios between the corresponding forward and complement strands, previously identified near the biologically confirmed replicative origins of bacterial species that were closely related to *R. sphaeroides*. Subsequently, all the putative replicator regions were cloned into a pLO1 plasmid, a suicide vector in *R. sphaeroides*. The autonomous replication of these recombinant plasmids in *R. sphaeroides* was further examined using conjugation and molecular methods. Results demonstrated that CI and CII of *R. sphaeroides* have a single replication origin on its chromosomes, respectively, and this will provide the basis of future work on coordination and control of replication and segregation of multiple chromosomes in bacteria.

Keywords: Origin of Replication; Multipartite genome; *Rhodobacter sphaeroides*; Z-curve analysis

Abbreviations: Chromosome I (CI); Chromosome II (CII); Replication origin of chromosome (oriC); Sistrom (SIS); Luria-Broth (LB).

Introduction

The continuity of bacterial species requires replication of their chromosomes, partitioning of the chromosomes into two haploid nucleoids, and formation of the septum, which together assure each of the newly divided cells with a full set of genetic information. The bacterial cell cycle consists of three stages, B period, C period, and D period, which correspond broadly to G1, S, and G2/M phases of eukaryotic cell cycle, respectively [1]. These three stages are sequentially coordinated and genetically regulated. The bacterial cell cycle begins with a newly divided cell, and when it reaches to an optimal size, the chromosomal replication initiates and continues till the genome gets fully replicated. The genome replication is followed by segregation of the replicated chromosomes and mid-cell septum formation. The chromosomal segregation mechanism of unipartite genome is more studied than that of a multipartite bacterial genome. In the multipartite genome, multiple chromosomes are of different sizes, and therefore synchronization of DNA replication is necessary for the proper chromosomal segregation and cytokinesis.

Chromosomal replication in bacteria initiates at the replicator (*oriC*) of its chromosome. The replicator varies in length across bacterial species; however, it contains highly conserved consensus sequences, where DnaA (replication initiator protein) binds to promote denaturation of the double-stranded DNA locally. The resulting single stranded DNA then acts as templates on which their corresponding complementary DNA strands are synthesized. After the chromosome is fully replicated, two sister chromosomes are separated at the terminus sequences located approximately 180°, the opposite site of the replication origin. Chromosomal replication occurs once per cell cycle, and the control of this vital process is achieved by the accessibility of DnaA to the *oriC*. In prokaryotes, a single replication origin exists on the chromosome; however, multiple origins of replication have recently been identified in archaebacterial species, *Sulfolobus acidocaldarius* and *Sulfolobolus solfataricus* [2,3].

The genome of *R. sphaeroides* was discovered to be multipartite [4], consisting of two circular chromosomes, the large primary chromosome (CI), and the small secondary chromosome (CII), both of which have been completely sequenced and fully annotated [5,6]. Since 10% of the sequenced bacterial genomes have been shown to possess a multipartite genome structure [7], the multipartite genome structure has become an accepted genome organization in bacteria. Both the chromosomes contain house-keeping genes, and many of these genes are involved in essential molecular processes, including replication, transcription and translation. The replication and segregation of a multipartite genome of *Vibrio cholerae* was previously studied and its genome consists the primary and the secondary chromosomes which are 2.9 Mbps and 1 Mbps, respectively [8]. The primary chromosome encodes essential proteins such as DnaA, DNA polymerases, DNA gyrase for chromosomal replication; while the secondary chromosome

does not contain any such important gene functions, but it contains genes such as *repABC*, *parA*, and *parB*, which are essential for plasmid replication and segregation. It has been previously demonstrated that the replication initiation in *V. cholerae* in the primary and the secondary chromosomes is replicated and regulated by DnaA and RctB, respectively [9]. An origin of replication on the primary chromosome interacts with DnaA protein, while the replication origin of the secondary chromosome interacts with another replication initiator protein, called RctB. Both the chromosomal origins are AT rich regions and they contain DnaA binding sequences. RctB plays an important role in synchronizing the replication of the two chromosomes by its auto-regulatory control which represses its own transcription [10].

Several species with multiple chromosomes, for example, *V. cholera* [11], *Brucella melitensis* [12], and *Agrobacterium tumifaciens* [13], are pathogenic to humans, other animals, or plants. Therefore, it is cumbersome to study as animal or plant cell-culture systems for the investigation are required. While, *R. sphaeroides* as a free-living model bacterium would offer similar understanding of replication mechanism of the multipartite genomes as it grows in a wide variety of growth conditions, including aerobic, anaerobic, or photosynthetic conditions. Furthermore, the genome of *R. sphaeroides* is completely sequenced and fully annotated [5,6]. The present study carries out an investigation for the identification and biological characterization of the replication origin of two chromosomes (OriCI and OriCII) in *R. sphaeroides* by employing genomic, genetic, and molecular analyses. The findings of this study will provide further insights for the study of the synchronization and coordination of chromosomal replication and segregation of multipartite genomes.

Materials and Methods

Bacterial strains, plasmids, and growth media

Bacterial strains and plasmids, and their corresponding genotypes and/or characteristics are listed in Table 1. *R. sphaeroides* cells were grown aerobically in Sistrom (SIS) minimal medium at 30°C, while *E. coli* cells were grown in Luria-Broth (LB) liquid medium or solid plates at 37°C with or without appropriate antibiotics as needed. For *E. coli* and *R. sphaeroides*, 50 µg/mL and 25 µg/mL of Kanamycin, respectively, were used. In addition, the genome sequences of the bacterial species are downloaded from the NCBI in a local database for genome analysis.

Bacterial strains or plasmids	Genotype and/or characteristics	Reference or source
Escherichia coli DH5αphe	F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1	Eraso and Kaplan [34]
Escherichia coli S17-1	pro, res2 hsdR17 (rK2 mK+) recA2 with an integrated RP4-2-Tc::Mu-Km::Tn7, Tp	Simon et al. [35]
Rhodobacter sphaeroides 2.4.1	wild-type	Suwanto and Kaplan [4]
pRK415	pRK404 derivative, Tcr OriT oriV lacZ	Keen et al. [36]
pLO1	Km^r, sacB RP4-oriT ColE1-oriV	Lenz O et al. [37]
pLO1CI1	Km^r, sacB RP4-oriT ColE1-oriV, CI-1 putative origin of replication cloned into Xba1 site	This study
pLO1CI3	Km^r, sacB RP4-oriT ColE1-oriV, CI-3 putative origin of replication cloned into Xba1 sit	This study
pLO1CII1	Km^r, sacB RP4-oriT ColE1-oriV, CII-1 putative origin of replication cloned into Xba1 site	This study
pLO1CII2	Km^r, sacB RP4-oriT ColE1-oriV, CII-2 putative origin of replication cloned into Xba1 site	This study
pLO1CII3	Km^r, sacB RP4-oriT ColE1-oriV, CII-3 putative origin of replication cloned into Xba1 site	This study
pLO1CII4	Km^r, sacB RP4-oriT ColE1-oriV, CII-4 putative origin of replication cloned into Xba1 site	This study
pLO1CII5	Km^r, sacB RP4-oriT ColE1-oriV, CII-5 putative origin of replication cloned into Xba1 site	This study

Table 1: Bacterial strains and plasmids and their corresponding characteristics.

Asymmetry of nucleotide composition

CI- and CII-specific DNA sequences of *R. sphaeroides* were downloaded from the NCBI genome database. The putative origins of replication were identified using OriFinder [14]. GC-Skew analysis was performed through a sliding window to compute GC=(G-C)/(G+C), to predict the *OriC* regions on the chromosome based on the GC-skew in leading and lagging strand of DNA. The nucleotide asymmetry and the location of *dnaA* were analyzed using Z-curve method [15]. DNA composition near the origin of replication displays various nucleotide disparities, which include AT disparity (AT skew), GC disparity (GC skew), amino/keto disparity (MK disparity), and purine/pyrimidine disparity (RY disparity). The disparities were plotted as $RY=x_n$,

MK=y_n, AT=$(x_n+y_n)/2$, and GC=$(x_n+y_n)/2$ line patterns along the DNA sequence of each chromosome. In addition to these lines, the graph also includes DnaA box clusters, indicator genes, and putative replication origin sites, as previously studied [16].

Gene distribution analysis

Two different approaches were employed to analyze the gene distribution of essential genes on the forward and complement strands. Essential gene-functions included genes involved in information processing, cellular structures, and ribosomal RNA genes. First, each 20 kb DNA fragments containing different putative origin sites of CI and CII were independently selected to calculate cumulative gene numbers, and then increasingly 20 kb region of DNA sequences were added (e.g., bin number=20 Kb, 20+20 kb, 40+20 kb, 60+20 kb, and so on), and the corresponding gene numbers were computed within those respective DNA fragments up to the end of the chromosome. The ratio of the number of genes located in the forward to complement strands was calculated for all corresponding bin numbers. The second analysis was performed on three distinct chromosomal fragments, one-third replication origin, another one-third middle region, and the last one-third terminus region. A total number of genes, essential genes, and ribosomal RNA genes were calculated for the forward and complement strands in each of these three chromosomal fragments. The analysis was performed on each 20 kb fragment containing one of the putative replication origin sites previously identified on CI and CII of *R. sphaeroides*. For each putative replication origin, all genes located on the forward and the complement strands were counted, and the ratios of the gene numbers of the forward to the complement strands were determined. In addition, specific 20 kb DNA fragments containing biologically known origin of replication sites of five bacterial species, *Bacillus subtilis*, *Sinorhizobium meliloti*, *Agrobacterium tumefaciens*, *Caulobacter cresentus*, and *Pseudomonas putida* were also analyzed in the same manner.

DNA isolation, polymerase chain reaction, and molecular cloning

Genomic DNA was purified using the Wizard® Genomic DNA Purification Kit. The fragments containing putative replication origins were amplified using genomic DNA of *R. sphaeroides*. Both the forward and the reverse primers were designed using OligoPerfect™ (http://tools.thermofisher.com) with XbaI linkers. The putative replication origins, sequences of the forward and the reverse primers, and the expected sizes of PCR products are described in Table 2. The polymerase chain reaction consisted of 30 cycles with denaturation at 92°C for one minute, annealing at 62°C for one minute, and extension at 72°C for one minute. PCR products were purified using the PCR Clean-up system (Wizard®, A1460). The pL01 plasmid vector was digested with XbaI, and the linearized vectors were ligated with the DNA fragments of each putative replication origin. *E. coli* DH5α cells were transformed with the recombinant pLO1 plasmids, and then plated on LB plus kanamycin plates. The resulting Kn^r transformants were then selected for further molecular analysis. Recombinant plasmids were isolated and purified using the Wizard® Plus SV Minipreps DNA purification system and transformed into *E. coli* S17-1 cells, which were then used as donor cells in the conjugation experiment.

DNA Regions	Chromosome coordinates (length)	Forward primer	Reverse primer
OriCI1	1700795-1701533 (738)	GCTCTAGAGGAGTGACTGAATGAAAGGCA	GCTCTAGATCACCACGAAAATGGGGG
OriCI2	2379797-2380253 (456)	GCTCTAGACGAGCTCTCCAGCGCAAG	GCTCTAGAGAGGGCGAGGTTCTTCCG
OriCI3	3004720-3005267 (581)	GCTCTAGAAGGCCAGCACCGAAATCA	GCTCTAGACCTCTTCGTCGATGCGGG
OriCII1	205633-205962 (308)	GCTCTAGACCGCGAAGCAGGAAGGGG	GCTCTAGACAGCTTCTGCATCTGCTGGCTG
OriCII2	365498-365929 (431)	GCTCTAGACCTGCGCCACGGTCTTCC	GCTCTAGAGGCGAGGCTGCCAAATCG
OriCII3	583596-584030 (434)	GCTCTAGAGCGCGGACCGACAAATGG	GCTCTAGAGCGAGACCTTGATGCGCG
OriCII4	942524-000508 (1003)	GCTCTAGATTCGAGGGAATCGCGGCCT	GCTCTAGAGTGCGCTTACGCCCAGTAATTCC
OriCII5	737949-738451 (502)	GCTCTAGATGCGGAGGGTGTCAGCCA	GCTCTAGAGAGAAGGAAGGCTGCTGGTCA
pLO11F/1R	NA	GCCGGATCCTTAATTAAG	GCAGGTTTAAACAGTCGA

Table 2: PCR primers used to amplify the putative origins of replication sequences and their corresponding coordinates and lengths (bp) of PCR products. PCR was performed for 30 cycles with denaturation at 92°C for one minute, annealing at 62°C for one minute, and extension at 72°C for one minute.

Conjugation experiment and molecular analysis

E. coli and *R. sphaeroides* liquid cultures were grown until mid-log phase, (3 hours for *E. coli* and 24 hours for *R. sphaeroides*). Having the optical density (OD=0.6) approximately the same, *E. coli* cells were washed several times with LB media to remove any traces of the Kanamycin, and then cells of *R. sphaeroides* and *E. coli* were mixed in 1:10 ratio. The mating mixture was spotted on LB plates and incubated overnight at 30°C. The following day, cells were collected and washed three times with SIS medium, and then the washed mating mixtures were plated out on the SIS-agar plates plus kanamycin for screening the *R. sphaeroides*' exconjugants. *R. sphaeroides* exconjugants were streaked out for several generations until pure colonies were obtained. The purified *R. sphaeroides* colonies were screened for the presence of autonomously replicating pLO1 recombinant plasmids in the cell. Plasmid DNA was isolated from these purified exconjugants for further plasmid analysis performing PCR on these plasmids and analyzing the PCR products by gel electrophoresis.

Results and Discussion

Nucleotide asymmetry patterns in multiple chromosomes in bacteria

Z-curve analysis has been used to predict the origin of replication in the bacterial chromosomes [16,17]. Results of the Z-curve analysis are shown in Figures 1, 2 and 3. Figure 1 shows the Z-curve analysis of the chromosome of the unipartite genomes of *E. coli*, *B. subtilis*, and *C. crescentus*. Figure 2 depicts the Z-curve analysis of the primary and secondary chromosomes of the multipartite genomes of *V. cholera*, and *B. melitensis* and the origins of chromosomal replication in the above organisms have been bioligically confirmed. Figure 3 shows the Z-curve analysis of the primary and secondary chromosomes of two closely related species, *R. sphaeroides* and *S. meliloti*.

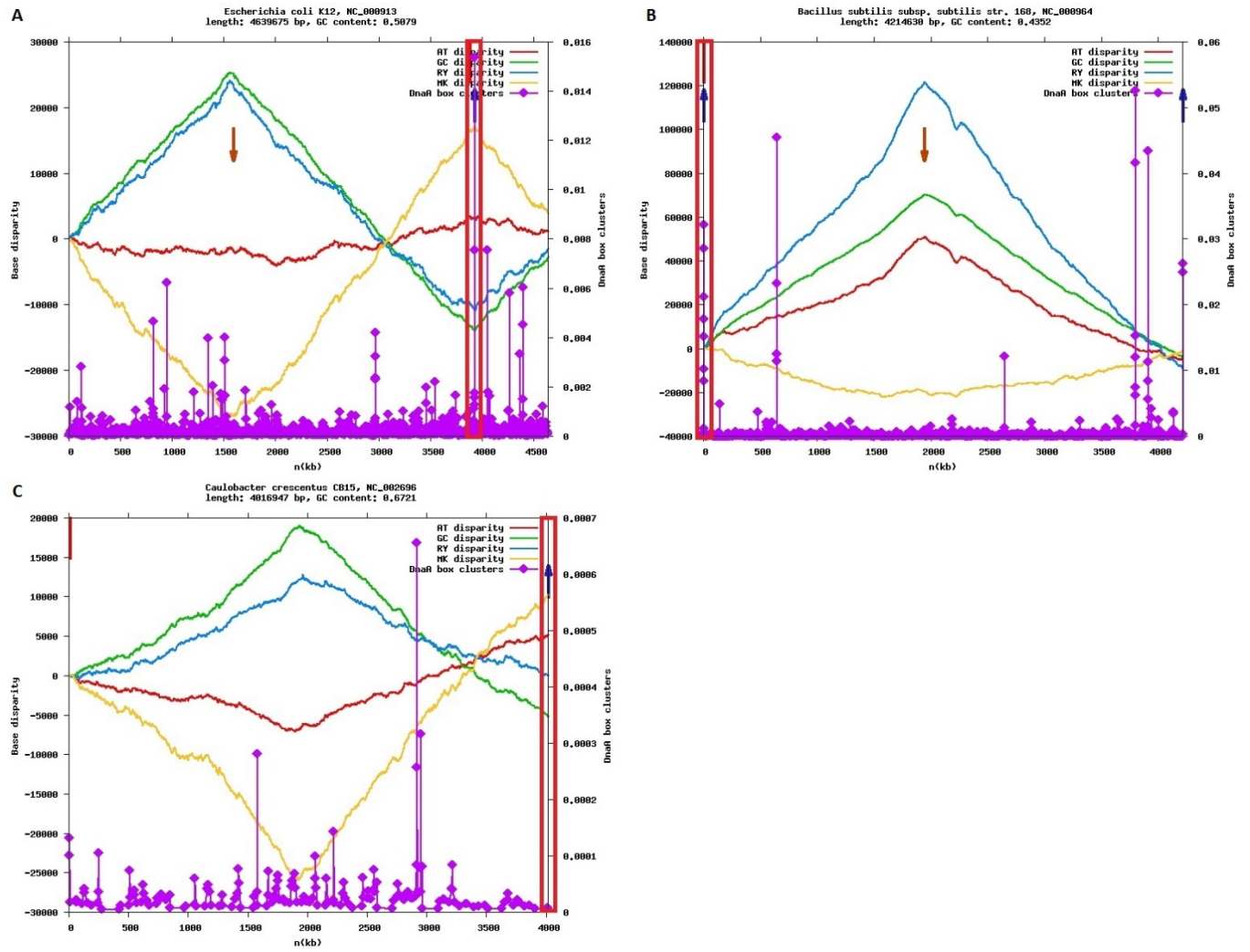

Figure 1: Bacterial chromosomes and their corresponding *oriC* regions (red box) predicted by Z-curve analysis. Z-curve data are shown for the primary chromosome of *Escherichia coli* (A), *Bacillus subtilis*, (B), and *Caulobacter crescentus* (C). The graph shows the original DNA sequence with four disparity curves (GC in green, AT in red, MK in yellow, and RY in blue), distribution of DnaA boxes (purple peaks with diamonds) and locations of putative indicator genes (vertical red lines), the predicted *oriC* region (vertical up arrow), and the predicted terminus region (vertical down arrow).

Results shown in Figure 1 reveal that *E. coli*, *B. subtilis*, and *C. crescentus* have a single origin of replication on their respective chromosomes, and these replication origins have been biologically confirmed [18-20]. The Z-curve analyses of the two chromosomes of *V. cholera* and *B. melitensis* are shown in Figure 2, which reveals a similar pattern of nucleotide diaparities in their corresponding chromosomes. However, *S. meliloti* and *R. sphaeroides* show a similar pattern as shown in Figure 3, but it is different from that of *V. cholera* and *B. melitensis*. While the primary chromosomes of all the four species (*V. Cholera*, *B. Melitensis*, *S. Meliloti*, and *R. sphaeroides*) exhibit a similar pattern of Z-cureve analyses as it was seen for the chromosome of *E. coli*, the secondary chromosomes show different patterns of nucleotide disparity. The secondary chromosomes of *V. cholera* and *B. melitensis* are similar to the Z-curve pattern observed for primary chromosomes, which implies that the secondary chromosomes of these two species possibly originated from their respective primary chromosomes.

However, the secondary chromosome of *R. sphaeroides* reflected a different nucleotide asymmetry, and showed a very complex pattern of nucleotide disparity similar to the one shown in the secondary chromosome of *S. meliloti*, a closely related species to *R. sphaeroides*. This suggests that the secondary chromosomes of *R. sphaeroides* and *S. meliloti* may have originated from their native plasmids by acquiring essential genes from their respective primary chromosomes. Based on these analyses, a total of eight putative replication origins (three on CI and five on CII) of *R. sphaeroides* were identified. They were designated as OriCI1, OriCI2, OriCI3, OriCII1, OriCII2, OriCII3, OriCII4, and OriCII5 and listed in Table 2.

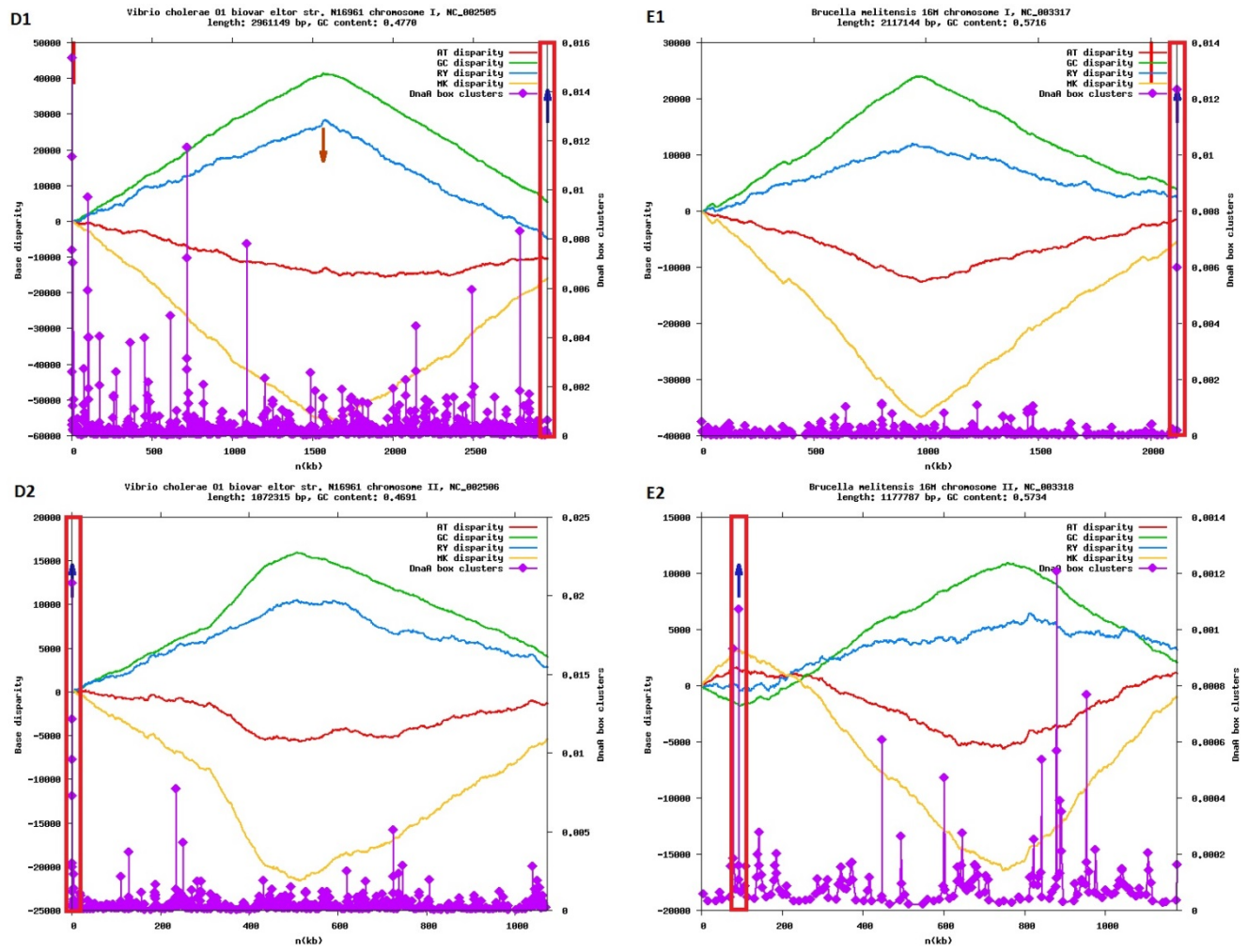

Figure 2: Bacterial chromosomes and the corresponding *oriC* regions (red box) predicted by Z-curve analysis. Z-curve data are shown for the primary chromosomes of *Vibrio cholera* (D1) and *Brucella melitensis* (E1), and for the secondary chromosomes of *V. cholera* (D2) and *B. melitensis* (E2). The graph shows the original DNA sequence with four disparity curves (GC in green, AT in red, MK in yellow, and RY in blue), distribution of DnaA boxes (purple peaks with diamonds) and locations of putative indicator genes (vertical red lines), the predicted *oriC* region (vertical up arrow), and the predicted terminus region (vertical down arrow).

Gene conservation patterns in the replication origin regions

As genes located in the *oriC* region are conserved across closely related bacterial species [21], the conserve gene-cluster within the known replication-origin regions can be utilized to predict putative origins of replication in related species, as shown in Figure 4 and Figure 5. For instance, comparison of the genome sequences of *C. crescentus* and *R. prowazekii* reveals that both species share a conserved gene-cluster near the established replication origins in *C. crescentus* and *R. prowazekii* [22-24]. Gene-conservation in the replication-origin regions of the primary chromosmes among bacterial species is shown in Figure 4. Of the three putative replication origins of CI of *R. sphaeroides*, OriCI3 region shares the common features of the true-origin of replicator region. The OriCI3 region conserves several genes (*parA*, *parB*, *gidA*, *gidB*, *trmE*, *maf*, and *rho*) as seen within the known *oriC* regions of closely related species such as *S. meliloti*, *A. tumefaciens*, and *C. crescensus*. In adiition, the replication initiator gene (*dnaA*) is located close to OriCI3, and it contains five DnaA binding sites. Also, the AT content of the OriCI3 region was higher (44%) compared to the genome average AT content (30%). In contrast, the genes located in the replication origin regions of the secondary chromosomes of different species as shown in Figure 5 are poorly

conserved. Of the five putative origins of replication predicted on CII, OriCII4 seems to be the most formidable replicator candidate. The AT content within OriCII4 was 42%, significantly higher than that of the genome average of 30% AT content. Also, the OriCII4 region contains *rep*, *parA* and *parB* genes, which have been shown to be conserved in known origins of replication of the secondary chromosomes and plasmids of other bacterial species.

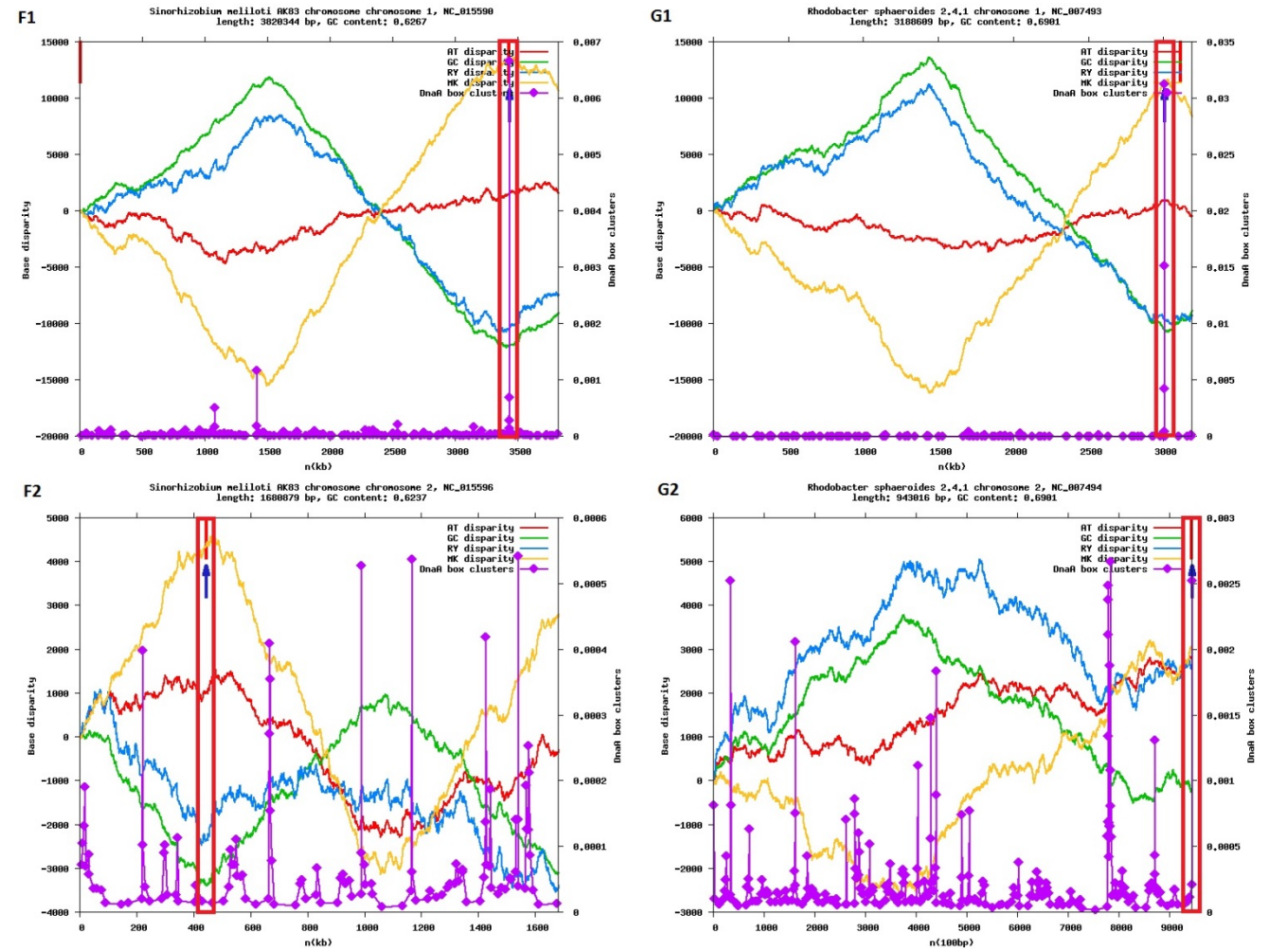

Figure 3: Bacterial chromosomes and their corresponding *oriC* region (red box) predicted by Z-curve analysis. Z-curve data are shown for the primary chromosomes of *Sinorhizobium meliloti* (F1) and *Rhodobacter sphaeroides* (G1), and for the secondary chromosomes of *S. meliloti* (F2) and *R. sphaeroides* (G2). Figure shows the original DNA sequence with four disparity curves (GC in green, AT in red, MK in yellow, and RY in blue), distribution of DnaA boxes (purple peaks with diamonds) and locations of putative indicator genes (vertical red lines), the predicted *oriC* region (vertical up arrow), and the predicted terminus region (vertical down arrow).

Gene distribution and cumulative gene density ratios of the leading and lagging strands

The analysis of the cumulative ratios of the number of all genes in the forward and the complement strands of CI and CII sequences revealed similar patterns as predicted by the disparity of genes located on the forward and reverse strands of bacterial chromosomes [25,26]. Figure 6 shows the cumulative ratios of CI and CII of *R. sphaeroides*, plotted from the start of the 20 kb region. Two distinct switch points were observed along the CI sequence, where one was located around the OriCI3 region and the other distantly located from the origin, close to the replication-terminus region. In contrast to CI, multiple switch points were observed along the CII sequence. This result suggests that the primary chromosomes in bacteria have conserved essential genes near the replicator [27,28], which implies the common origin of replicator sequence on the primary or large chromosome; while the secondary chromosomes may have originated from different plasmids which acquired genes from other species, resulting in the discontinuity of ratios along the CII sequence.

Furthermore, the number of essential genes is computed in three equally divided different chromosomal regions (the origin, the middle, and the terminus), considering each putative replicator regions as possible origin of replication in CI and CII of *R. sphaeroides*. Both OriCI3 and OriCII4 origin-regions revealed the highest frequency of essential genes including ribosomal RNA genes as shown in Figure 7, which is consistent with the previous findings [26].

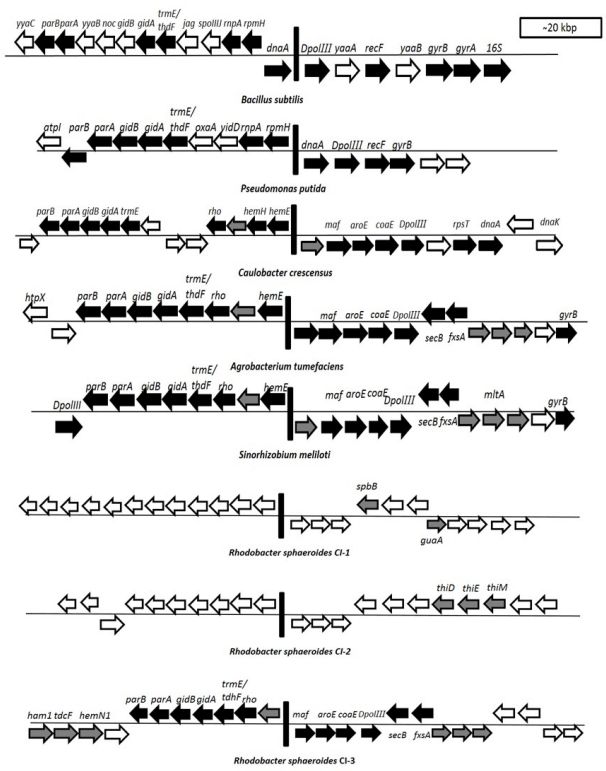

Figure 4: Conservation of gene clusters in the flanking regions of the confirmed chromosomal replication origin (*oriC*) of five bacterial species, as well as three putative origins of the primary chromosome of *R. sphaeroides*. Three 20 kb regions containing a separate putative replication origins of the CI are located at the coordinates in Table 2. Black, gray, and white arrows represent conserved genes with known biological functions, non-conserved genes with known biological functions, and non-conserved genes encoding hypothetical proteins, respectively. Each bold-faced vertical black line is the location of the replication origin. Each horizontal line is the center of the two DNA strands. Some genes are coded on one strand, while others are on the other strand of the DNA. Each arrow direction displays the orientation of gene transcription. The genes and their functions are as follows: *dnaA*, DnaA chromosomal replication initiator protein; *dnaPolIII*, DNA Polymerase III subunit; *parA*, ATPase segregation protein; *parB*, kinetochore binding protein; *gidA* glucose-inhibited division protein A; *gidB*, 16S rRNA m(7)G-527 methyltransferase; *thdF*, tRNA modification GTPase *trmE*; tRNA modification GTPase; rho, transcription termination factor; *maf*, putative Maf/YceF/YhdE family protein; *gyrB*, DNA gyrase or topoisomerase II (subunit B); *hemE*, uroporphyrinogen decarboxylase; *hemH*, ferrochelatase; and *rnpA*, Ribonuclease P.

Table 3 lists the ratios of the number of genes on the forward and the complement strands located in the 20 kb region of all the predicted putative origin sites of CI and CII in *R. sphaeroides*. Of the total eight putative sites, only OriCI3 of CI and OriCII4 of CII revealed gene ratios of the corresponding forward and the complement strands close to 1, while other putative origin regions have gene ratios either higher or lesser than 1. Thus, the results described here further aid to the credence that OriCI3 and OriCII4 regions are the functional replicators for the large and small chromosome in *R. Sphaeroides*, respectively. This result was further validated by similar analysis of the biologically confirmed replication origins of other five bacterial species listed also in Table 3.

Thus, all three approaches yielded similar observations supporting OriCI3 and OriCII4 regions as functional origins of replication of primary and secondary chromosomes of *R. sphaeroides*, respectively.

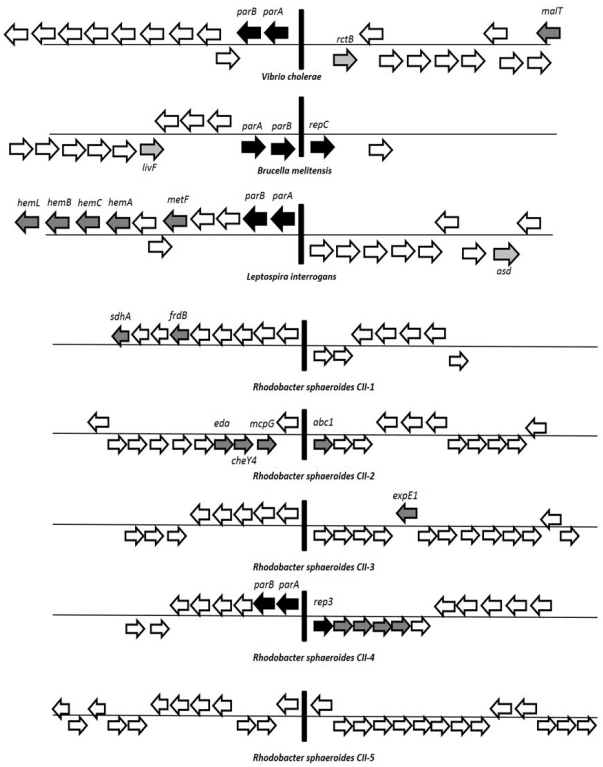

Figure 5: Conservation of gene clusters in the flanking regions of the confirmed chromosomal replication origin (*oriC*) of three bacterial species, as well as five putative origins of the secondary chromosome of *R. sphaeroides*. Five 20 kb regions containing a separate putative replication origin of the CII are located at the coordinates in Table 2. Black, gray, and white arrows represent conserved genes with known biological functions, non-conserved genes with known biological functions, and non-conserved genes encoding hypothetical proteins, respectively. Each bold-faced vertical black line is the location of the replication origin. Each horizontal line is the center of the two DNA strands. Some genes are coded on one strand, while others are on the other strand of the DNA. Each arrow direction displays the orientation of gene transcription. Some known genes and their funcations are as follows: *rctB*, Replication Initiation Protein; *repC* (*rep3*), Plasmid-like Replication Initiation Protein; *parA*, ATPase segregation protein; and *parB*, kinetochore binding protein.

Species[1]	Coordinate of Origin	Forward (F) strand	Complement (C) strand	F/C ratio
C. crescentus **CB15**	**4016703–0000234**	**11 (52%)**	**10 (48%)**	**1.1**
A. tumefaciens **C58**	**2841174–2841581**	**11 (50%)**	**11 (50%)**	**1**
S. meliloti **AK83**	**3425340–3425744**	**12 (55%)**	**10 (45%)**	**1.2**
B. subtilis **168**	**1751–1938**	**9 (50%)**	**9 (50%)**	**1**
P. putida **KT2440**	**8947–9541**	**8 (47%)**	**9 (53%)**	**0.889**
R. sphaeroides CI-1	1700795–1701533	9 (40%)	13 (60%)	0.69
R. sphaeroides CI-2	2379797–2380253	4 (19%)	17 (81%)	0.23
R. sphaeroides **CI-3**	**3004720–3005267**	**13 (54%)**	**11 (46%)**	**1.188**
R. sphaeroides CII-1	205632–205962	3 (17%)	14 (83%)	0.21
R. sphaeroides CII-2	365498–365516	14 (58%)	8 (42%)	1.75
R. sphaeroides CII-3	583596–584030	13 (65%)	7 (35%)	1.85
R. sphaeroides **CII-4**	**942503–000516**	**11 (47%)**	**12 (53%)**	**0.916**
R. sphaeroides CII-5	737949–738451	17 (56%)	13 (44%)	1.3

[1]Species names in bold represent those species where origin of replication has been previously characterized.

R. sphaeroides' replication origins (in bold) are identified in this study.

Table 3: Gene distribution in leading and lagging strands in the replication origin region.

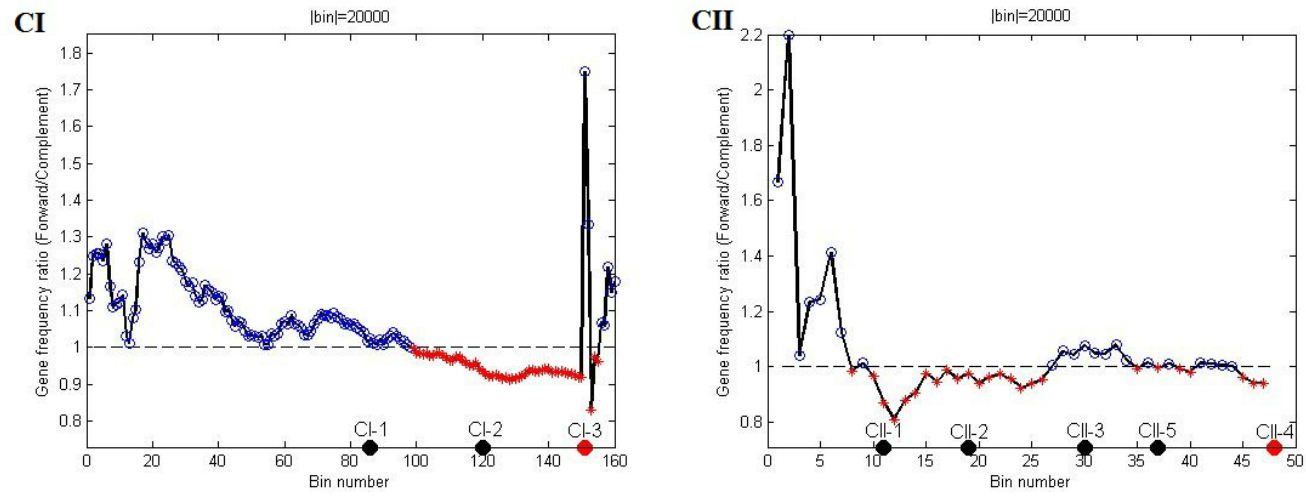

Figure 6: Gene distributions along with the chromosomal sequences of *R. sphaeroides*. Each marker depicts the cumulative gene frequency ratio (defined as the ratio of the number of genes on the forward strand to the complement strand) for subsequent cumulative bins (each bin size=20kb) along the chromosomal sequence. Blue circle markers and red star markers depict the ratio >1 and <1, respectively. Left and right panels are for CI and CII. The potential replicator regions, OriCI3 and OriCII4, are denoted as CI-3 and CII-4 by red filled circles, and the other putative replicators are denoted by black filled circles.

Biological confirmation of origin of replication

In this study, a positional cloning approach was employed to biologically characterize all eight putative replication origins of *R. sphaeroides*. Of the eight putative replicators, seven were successfully cloned in the pLO1 suicidal plasmid vector as shown in Figure 8. Only OriCI3 putative replicator of CI could not be cloned in *E. coli*, even though the given genomic fragment was successfully amplified using the standard polymerase chain reaction. Over several repeated transformations, no single transformant was obtained with the pLO1CI2 recombinant plasmid. It has been previously reported that

some gene sequences are unstable due to lack of supercoiling and secondary structures or encode regulatory proteins toxic to the *E. coli* cells [29]. It is suspected that the OriCI2 insert of the pLO1CI2 recombinant plasmid may contain some sequence features mentioned above, which may have induced plasmid instability in *E. coli* cells.

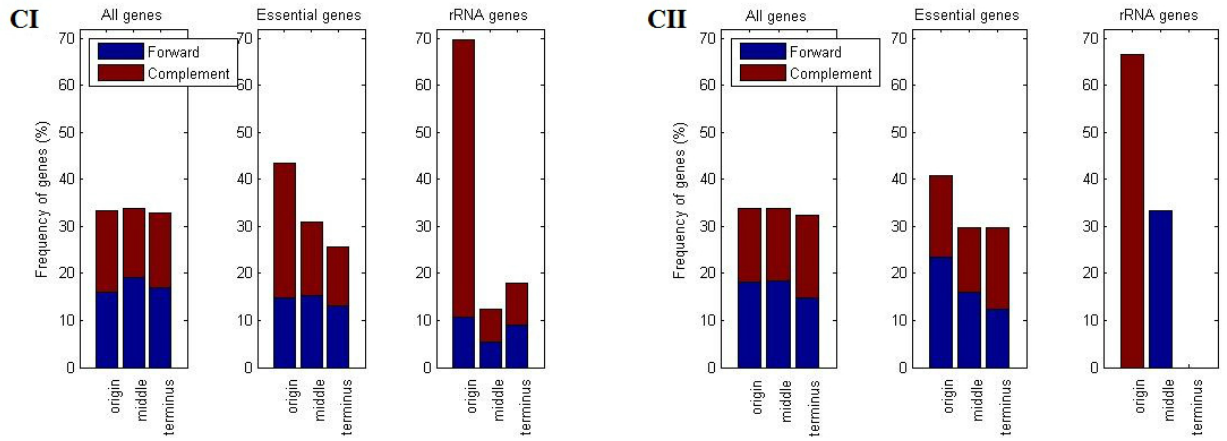

Figure 7: Gene distributions within the three distinct fragments (replication origin region, middle region, and terminus region) of *R. sphaeroides*. For each fragment, the proportion of genes in forward and complement strands is represented by red and blue bars, respectively. Also, the proportions for all genes, essential genes, and rRNA gene were separately computed. Left panel shows the result with OriCI3 and right panel with OriCII4 as functional replication origins.

The conjugation efficiency was estimated to be 5 X 10^6, which was comparable to the range of conjugation efficiencies previously estimated for *R. Sphaeroides* [30]. This observation confirmed that the conjugal plasmid transfer in the current study was efficient. The result of a single colony exconjugant from the pLO1 plasmid from *E. coli* to *R. sphaeroides* suggests that *R. sphaeroides* may survive on Kanamycin even without the plasmid, possibly due to a chromosomally encoded gene mutation [31].

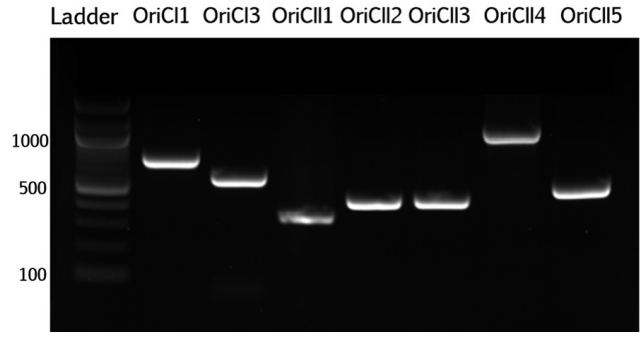

Figure 8: PCR products amplified from plasmids from *E. coli* transformants. Lane 1 is a 100 bps ladder, and Lanes 2-8 were of insert size of OriCI1 (738 bps), OriCI3 (581 bps), OriCII1 (308 bps), OriCII2 (431 bps), OriCII3 (434 bps), OriCII4 (1003 bps), and OriCII5 (502 bps), respectively.

The exconjugants were grown on SIS plus Kanamycin minimal plates on which they grew in different colony sizes. A majority of exconjugants were of small colony size and they were not recovered after passing though several generations of colony purification. We suspect a large number of Kn^r exconjugants are lacking the recombinant plasmids, and therefore these false positives are attributed to mutations of chromosomally encoded genes in *R. sphaeroides*. Kanamycin belongs to the group of aminoglycoside antibiotics, and their bactericidal activity is attributed to binding to the 30S ribosomes and inhibits protein synthesis, as well as interacting with other cellular structures and metabolic processes [32].

The antibiotic resistance proteins alter ribosome structure, decrease membrane permeability, and/or inactivate kanamycin through aminoglycoside modifying enzymes, such as acetyl transferases, nucleotidyl transferases, and phospho-transferases [33]. The high level of resistance can also result from mutations in chromosomal genes encoding ribosomal proteins, rpsL, rpsD, or rpsE, and such compensatory mutations in these genes generate a low level of kanamycin resistance by altering ribosomal structures, which may provide a false Kn^r phenotype [32].

Of the two recombinant plasmids (pLO1CI1 and pLO1CI3) that represented the cloned putative origins from the large chromosome, only the pLO1CI3 plasmid was maintained after getting transferred from *E. coli* to *R. sphaeroides*. Of the total 105 exconjugants screened, 25 exconjugants grew very slowly, and did not maintain the recombinant plasmid in *R. sphaeroides*.

Of the 80 normally grown exconjugants, most of them survived when the colonies were purified over five generations. However, all the five recombinant plasmids (pLO1CII1, pLO1CII2, pLO1CII3, pLO1CII4, and pLO1CII5) that contain each of the predicted replication origin regions of CII yielded 20-30 of *R. spaheroides* exconjugants.

None the less, following the colony purification over five generations, only pLO1CII4 plasmid was seen to be maintained in *R. sphaeroides*. Results shown in Figure 9 provided the further evidence that the expected inserts (581 and 1000 nucleotide DNA fragments)

were successfully amplified from these plasmids, when the internal primers were used to amplify the respective cloned sequences. Thus, these two plasmids containing respective replication origins from CI and CII are stably replicating in *R. sphaeroides*.

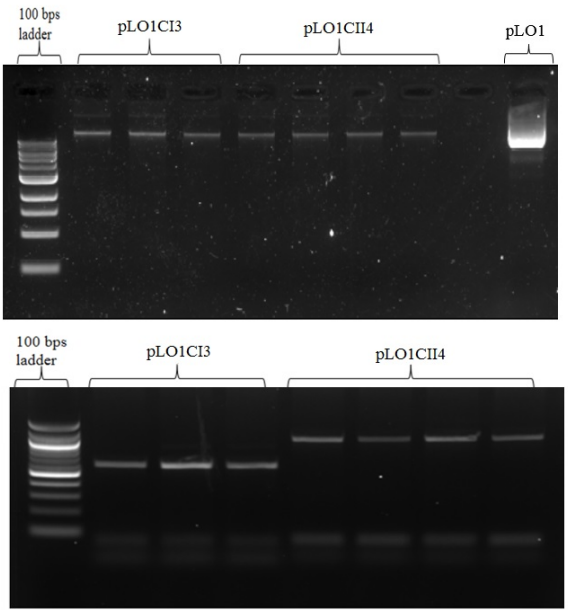

Figure 9: Plasmid and their corresponding PCR products from *R. sphaeroides* exconjugants. Top panel exhibits plasmid DNA (pLO1CI3 and pLO1CII4) isolated from *R. sphaeroides* exconjugants. Bottom panel shows the amplified inserts of the recombinant plasmids (pLO1CI3 and pLO1CII4). OriCI3 and OriCII4 regions were amplified using external pLO1F and pLO1R primers, which were 587 bps and 1004 bps, respectively.

Conclusion

This study employed both bioinformatics and molecular methods to predict and biologically characterize the replication origins on CI and CII of *R. sphaeroides*. Z-curve and GC skew analyses indicated that CI and CII have three and five putative replication origins regions, respectively. They were named as OriCI1, OriCI2, OriCI3, OriCII1, OriCII2, OriCII3, OriCII4, and OriCII5. Analyses of gene conservation and gene density ratios distribution further imply that OriCI3 and OriCII4 are the formidable candidates of the true origins of CI and CII of *R. sphaeroides*.

The results of the conjugation experiment followed by molecular analysis of isolated plasmids (pLO1CI3) demonstrated that the 581 bps DNA segment (OriCI3) located at 3,004,720-3,005,267 bps region of CI is the functional origin of replication (OriCI) of the primary chromosome in *R. sphaeroides*. Similarly, the molecular analysis of pLO1CII4 plasmid demonstrated that a ~1000 bps DNA fragment (OriCII4) located at 942,503-000,516 bps region of CII is the functional origin of replication of the secondary chromosome in *R. sphaeroides*. Thus, the current study concludes the identification of a single origin of replication of CI and CII in *R. sphaeroides*.

References

1. Skarstad K, Steen HB, Boye E (1983) Cell cycle parameters of slowly growing Escherichia coli B/r studied by flow cytometry. J Bacteriol 154: 656-662.
2. Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R (2004) Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. Proc Nat Acad Sci USA 101: 7046-7051.
3. Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, et al. (2004) Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus. Cell 116: 25-38.
4. Suwanto A, Kaplan S (1989) Physical and genetic mapping of the Rhodobacter sphaeroides 2.4.1 genome: genome size, fragment identification, and gene localization. J Bacteriol 171: 5840-5849.
5. Mackenzie C, Choudhary M, Larimer FW, Predki PF, Stilwagen S, et al. (2001) The home stretch, a first analysis of the nearly completed genome of Rhodobacter sphaeroides 2.4.1. Photosynth Res 70: 19-41.
6. Kontur WS, Schackwitz WS, Ivanova N, Martin J, Labutti K, et al. (2012) Revised sequence and annotation of the Rhodobacter sphaeroides 2.4.1 genome. J Bacteriol 194: 7016-7017.
7. Choudhary M, Cho H, Bavishi A, Trahan C, Myagmarjav BE (2012) Evolution of multipartite genomes in prokaryotes. In: Pontarotti P, Editor, Springer-Verlag, Berlin, Heidelberg. Evolutionary Biology: Mechanisms and Trends 301-323.
8. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn M, et al. (2000) DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature 406: 477-483.
9. Egan ES, Waldor MK (2003) Distinct replication requirements for the two Vibrio cholerae chromosomes. Cell 114: 521-530.
10. Egan ES, Duigou S, Waldor MK (2006) Autorepression of RctB, an initiator of Vibrio cholerae chromosome II replication. J Bacteriol 188: 789-793.
11. DiRita VJ (1995) Vibrio cholerae and cholera: molecular to global perspectives. Trends in Microbiology 3: 79-80.
12. Smith LD, Ficht TA (1990) Pathogenesis of Brucella. Crit Rev Microbiol 17: 209-230.
13. Allardet-Servent A, Michaux-Charachon S, Jumas-Bilak E, Karayan L, Ramuz M (1993) Presence of one linear and one circular chromosome in the Agrobacterium tumefaciens C58 Genome. J Bacteriol 175: 7869-7874.
14. Gao F, Zhang CT (2008) Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. BMC Bioinformatics 9: 79.
15. Gao F (2014) Recent advances in the identification of replication origins based on the Z-curve method. Current Genomics 15: 104-112.
16. Worning P, Jensen LJ, Hallin PF, Stearfeldt HH, Ussery DW (2006) Origin of replication in circular prokaryotic chromosomes. Environ Microbiol 8: 353-361.
17. Zhang CT, Zhang R, Ou HY (2003) The Z curve database: A graphic representation of genome sequences. Bioinformatics 19: 593-599.
18. Asada K, Sugimoto K, Oka A, Takanami M, Hirota Y (1982) Structure of replication origin of the Escherichia coli K-12 chromosome: The presence of spacer sequences in the orl region carrying information for autonomous replication. Nucleic Acids Res 10: 3745-3754.
19. Ogasawara N, Seiki M, Yoshikawa H (1983) Replication origin region of Bacillus subtilis chromosome contains two rRNA operons. J Bacteriol 154: 50-57.
20. Marczynski GT, Dingwall A, Shapiro L (1990) Plasmid and chromosomal DNA replication and partitioning during the Caulobacter crescentus cell cycle. J Mol Biol 212: 709-722.
21. Wolanski M, Donczew R, Zawilak-Pawlik A, Zakrzewska-Czerwinska J (2015) oriC-encoded instructions for the initiation of bacterial chromosome replication. Front Microbiol 5: 735.
22. Brassinga AK, Marczynski GT (2001) Replication intermediate analysis confirms that chromosomal replication origin initiates from an unusual

intergenic region in Caulobacter crescentus. Nucleic Acids Res 29: 4441-4451.

23. Brassinga AK, Siam R, Marczynski GT (2001) Conserved gene cluster at replication origins of the alpha-proteobacteria Caulobacter crescentus and Rickettsia prowazekii. J Bacteriol 183: 1824-1829.
24. Brassinga AK, Siam R, McSween W, Winkler H, Wood D, et al. (2002) Conserved response regulator CtrA and IHF binding sites in the α-proteobacteria Caulobacter crescentus and Rickettsia prowazekii chromosomal replication origins. J Bacteriol 184: 5789-5799.
25. Zhao HL, Xia ZK, Zhang FZ, Ye YN, Guo FB (2015) Multiple factors drive replicating strand composition bias in bacterial genomes. Int J Mol Sci 16: 23111-23126.
26. Rocha EP (2004) The replication-related organization of bacterial Genomes. Microbiology 150: 1609-1627.
27. Rocha EP, Danchin A (2003) Gene essentiality determines chromosome organisation in bacteria. Nucleic Acids Res 31: 6570-6577.
28. Rocha EP, Danchin A (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat Genet 34: 377-378.
29. Godiska R, Mead D, Dhodda V, Wu C, Hochstein R, et al. (2010) Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. Nucleic Acids Res 38: e88.
30. Hunter CN, Turner G (1988) Transfer of genes coding for apoproteins of reaction centre and light-harvesting LH1 complexes to Rhodobacter sphaeroides. J General Microbiol 134: 1471-1480.
31. Munita JM, Arias CA (2016) Mechanisms of antibiotic resistance. Microbiol Spectr 4(2).
32. Vester B, Long K (2009) Antibiotic resistance in bacteria caused by modified nucleosides in 23S ribosomal RNA. In: Grosjean H. DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution, Austin, TX: Landes Bioscience 524-536.
33. Davis BD (1987) Mechanism of bactericidal action of aminoglycosides. Microbiol Rev 51: 341-350.
34. Eraso JM and Kaplan S (1994) prrA, a putative response regulator involved in oxygen regulation of photosynthesis gene expression in Rhodobacter sphaeroides. J Bacteriol 176: 32-43.
35. Simon R, Priefer U, and Pühler A (1983) A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. Nature Biotechnology 1: 784-791.
36. Keen NT, Tamaki S, Kobayashi D, Trollinger D (1988) Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria. Gene 70: 191-197.
37. Lenz O, Schwatz E, Dernedde J, Eitinger M, Friedrich B (1994) The Alicaligenes europhus H16 hoxX gene participates in hydrogenase regulation. J Bacteriol 176: 4385-4393.