

# Relative influence of natural selection and mutational pressure on the genes of primary and secondary chromosome of multichromosomal bacteria

## Abstract

Bacterial genome is divided into two chromosome, those bacteria's are known as multi chromosomal bacteria. In one chromosome the number of genes are higher and most essential genes are present called primary chromosome. Genes on secondary chromosome has weaker codon usage bias than those on primary chromosome. The relative influence of mutational pressure and various forms of natural selection on codon bias are different. Our study reveal that whether there is any compositional constraint bias and relative Influence of mutational forces shaping codon usage pattern between primary and secondary chromosome from three bacterial genera, *Leptospira Interrogans*, *Rhodobacter sphaeroides* and *Vibrio Cholerae* which are phylogenetically independent, and the genomic GC content of these bacteria's are 36.68%, 69.15% and 47.7% respectively. Our investigation focused on if there is any difference in the influences of natural selection and mutational pressure on the genes of different chromosome of Multichromosomal bacteria with different genomic GC content.

**Keywords:** GC content, gravity, aroma, primary chromosome, secondary chromosome, isochore, p value

Volume 6 Issue 3 - 2017

**Sushanta Deb, Surajit Basak**

Department of Molecular Biology &amp; Bioinformatics, Tripura University, India

**Correspondence:** Surajit Basak, Department of Molecular Biology & Bioinformatics, Tripura University, Suryamaninagar – 799022, Tripura, India, Email Sushantadeb26@gmail.com

**Received:** August 04, 2017 | **Published:** November 06, 2017

## Introduction

*Rhodobacter sphaeroides* is the first reported multiple chromosomal bacteria, a rod-shaped, Gram-negative bacterium.<sup>1</sup> The researchers provide a complete physical map of the *Rhodobacter sphaeroides* genome by obtaining chromosomal DNA fragments through restriction digestion with endonuclease from the genomic DNA, which support in establishing the existence of more than one chromosome.<sup>1</sup> Chromosomes may produce by three different mechanisms: by the separation of a single chromosome, by chromosome copying, or by integration of a large plasmid with essential genes.<sup>2</sup> Of these processes, the theory of plasmid derived chromosome is more widely acceptable and has the greatest support because origins of replication of some secondary chromosomes have similarity with the plasmids oriC.<sup>2</sup> The emerging picture is that Multichromosomal bacteria have one primary chromosome (chromosome I) carrying most housekeeping genes, and generally one or two secondary chromosomes having some plasmid features which is recognisable and also contains some essential genes.<sup>3</sup> Under a certain environmental (external) conditions specified amplification of one chromosome can alter gene pool and thereby the levels of gene expression.<sup>3</sup>

Two circular chromosomes are present in the *Leptospira genome* with a total of 4,627,366 base pairs (bp), chromosome I with 4,277,185 bp and chromosome II with 350,181 bp.<sup>4</sup> *Leptospira Interrogans* genome divided into different chromosome and it is a Multichromosomal bacterium, increasing number of Multichromosomal bacteria, such as *Vibrio cholerae* and *Ralstonia solanacearum*.<sup>4</sup> In the genome sequence of multichromosomal Bacteria the genes encoding enzymes for metabolic pathways, such as Citric acid cycle and the glycolysis, as well as the genes of the enzymes for amino acid biosynthesis pathways are also distributed between the two chromosomes. The

Whole genome of *Vibrio cholerae* has 4,033,460 base pairs (bp) with two circular chromosomes of 2,961,146 bp and 1,072,314 bp.<sup>5</sup> The share part of recognizable genes which are involved in essential cell functions (such as transcription, DNA replication, cell-wall biosynthesis and translation) and Pathogenicity (for example, surface antigens, adhesions, and toxins) are located on the large chromosome (Primary chromosome).<sup>5</sup> On the other hand, a larger fraction (59%) of hypothetical genes are resides in the small chromosome (secondary Chromosome) compared with the large chromosome (42%).<sup>5</sup>

## Methodology

Retrieved the Whole genome of three multi chromosomal bacteria from NCBI ftp site <ftp.ncbi.nlm.nih.gov/genbank/genomes>. Coding sequences are shorted according to chromosome I and chromosome II. The value of different gene parameter such as GC1, GC2, GC3, gene length and ENC value of each of the gene on chromosome I and chromosome II are obtained by using a online server CAIcal server ([genome.erv.es/CAIcal/](http://genome.erv.es/CAIcal/)).<sup>1</sup> Obtained values are further analysed through several statistical measure to find their correlation and their distribution among the genome of chromosome I and chromosome II. The GRAVY and AROMA values of genes are obtained by using codon w softwear (written by John Peden and taken from (<http://molbiol.ox.ac.uk/cu/codonW.tar.Z/>)). Statistical measurement is done using [www.statsoft.com/Products/STATISTICA/Data-Miner](http://www.statsoft.com/Products/STATISTICA/Data-Miner) to get the values of statistical measurement.

GRAVY and AROMA are the two parameter which measures the hydrophobicity and aromaticity of a single protein (gene product) encoded by a gene. The values of these two parameter can be obtained using Codon W program. It has been established that these two factors are the result of translational selection and according to the

natural selection.<sup>6</sup> A GRAVY or AROMA score are proportional to the hydrophobic or aromatic amino acid product in a protein encoding gene.

## Result and discussion

### Influence of mutational bias on chromosome I and chromosome II codon usage bias

Three different multi chromosomal bacteria with high genomic GC level 69.15 % of *Rhodobacter sphaeroides*, intermediate GC level of 47.7% in *Vibrio Cholerae* and low GC level of 36.68% in *Leptospira Interrogans* shows the different trends of compositional distribution In chromosome I and chromosome II. In case of low genomic GC content multi chromosomal bacteria *Leptospira Interrogans* shows positive correlation between GC3 and GC12, in both of the chromosome I and

chromosome II (Table 1 & 2).

Here we measure the similarity or dissimilarity of compositional distribution between the genome of chromosome I and Chromosome II. From this study it is found that all the bacterial species are not showing the same trends of compositional correlation, the *Rhodobacter sphaeroides* species shows a highly different distribution of composition, it shows negative correlation of GC3 with GC2. Other hand rest of two bacterial species shows same type of correlation, i.e they show positive correlation. If we consider about *Vibrio Cholerae* there is no correlation of GC3, GC2 and GC1 of chromosome I genes, chromosome II shows the higher correlation level between GC3, GC2 and GC1. This is different in case of *Rhodobacter sphaeroides* which shows chromosome I have higher level of compositional correlation but chromosome II shows no correlation.

**Table 1** The correlation between the codon compositions

Bacteria	Chromosome I			Chromosome 2		
	GC3 vs GC2	GC3 vs GC1	GC1 vs GC2	GC3 vs GC2	GC3 vs GC1	GC1 vs GC2
<i>Vibrio cholerae</i>	r=0.1422	r=0.1710	r=0.1533	r=0.2321	r=0.3535	r=0.1440
	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
<i>Leptospira interrogans</i>	r=0.1688	r=0.2708	r=0.2715	r=0.1356	R=0.3401	R=0.3087
	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
<i>Rhodobacter sphaeroides</i>	r=-0.1339	r=-0.1465	r=0.5499	r=-0.0704	r=-0.0334	r=0.5360
	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01

**Table 2** The correlation between the codon compositions the GRAVY value and the AROMA value of the chromosome I and chromosome II genes

Bacteria	Chromosome	Gravy	Aroma
<i>Leptospira interrogans</i>	Chromosome I	-0.0051	-0.2326
	Chromosome II	-0.0627	0.0239
<i>vibrio cholerae</i>	Chromosome I	0.0415	-0.1005
	Chromosome II	0.1333	-0.0941
<i>Rhodobacter sphaeroides</i>	Chromosome I	0.1292	0.1133
	Chromosome II	0.3072	0.023
		P<0.01	

Mutational forces shaping the compositional distribution and codon usage pattern has higher influence on genes of chromosome II rather than chromosome I in case of lower and moderate genomic GC content bacteria.<sup>7-9</sup> Where in higher GC content bacteria the codon usage pattern of genes of chromosome I is more influenced by mutational forces.<sup>9</sup> The cytosine and thiamine contents were higher than that of guanosine and adenosine on the third position of codons in the genes of chromosome II, it has been found that in high genomic GC content bacteria genes of chromosome II shows greater number of codon ending with T indicating mutation of cytosine to thymine through deamination is predominant in this chromosome, *Leptospira Interrogans* also shows the same trend of T ending codon distribution. Somewhat different trend have been noticed in genes of *Vibrio Cholerae*, same number of T ending codons are distributed between

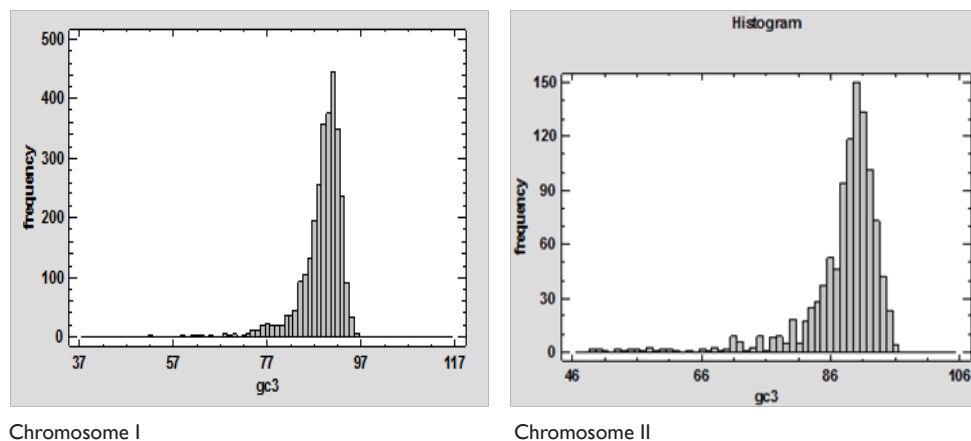
chromosome I and Chromosome II genes suggesting cytosine to thymine substitution is not influence the genes of chromosome II, it most probably due to the wide range of GC3 distribution in chromosome II genes. With C and T used more frequently than G and A in the third position of codon in chromosome II genes. This result indicates that nature selection affects the codon usage of the genes of chromosome II, but mutational bias has the major influence of codon usage.

### Distribution of GC3 values among genes of chromosome I and chromosome II

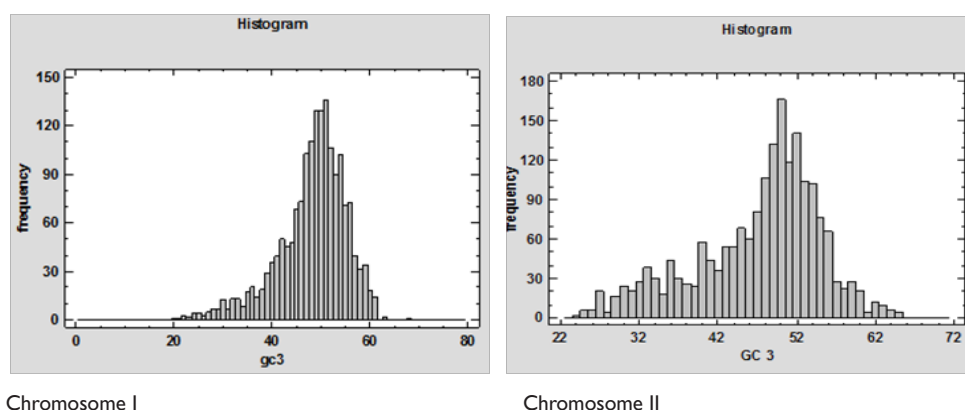
The range of distribution of GC3 values are very similar between chromosome I and chromosome II of moderate and low genomic GC content bacteria but it show a little bit of alteration in case of high GC content where genes of chromosome II shows Wide range of GC3 distribution. Similar GC3 distribution on the genes of both the chromosome indicates that gene density is same in the GC rich Isochores of both chromosome I and Chromosome II genome (Figures 1-3).<sup>10</sup>

### Variation of optimal codons on the genes of chromosome I and chromosome II

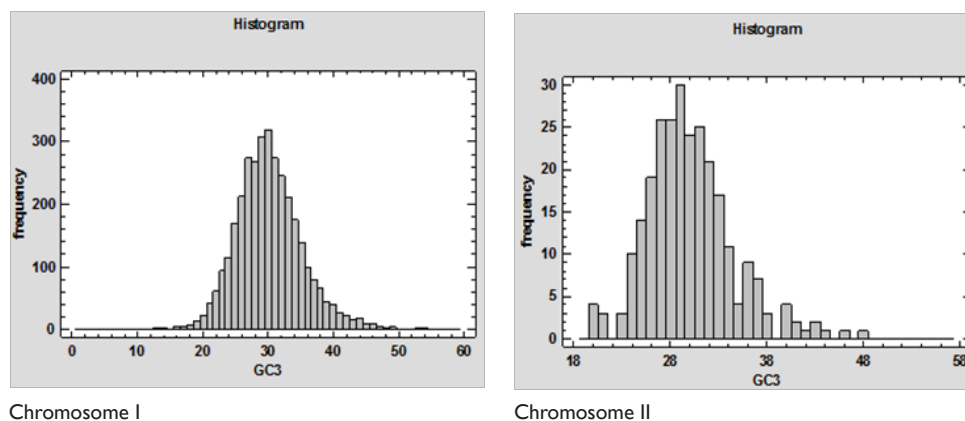
Optimal codons of chromosome I and chromosome II genes of three multichromosomal bacteria are identified and it reveals that number of optimal codons are approximately uniform on the genes of chromosome I and chromosome II of *Leptospira interrogans*, *Vibrio cholera* and *Rhodobacter sphaeroides*.<sup>4</sup> Uniform distribution of optimal codons on the genes of both chromosome indicate that there is rarely any variation of translational selection pressure on the genes of chromosome I and chromosome II of multichromosomal bacteria. Optimal codons are identified as described by Michele Stenico et al.<sup>4</sup>



**Figure 1** Distribution of GC3 content in *Rhodobacter sphaeroides* genome.



**Figure 2** Distribution of GC3 content in *Vibrio cholerae* genome.



**Figure 3** Distribution of GC3 content in *Leptospira interrogans* Genome.

### The influence of natural selection on chromosome I and Chromosome II genes codon usage bias

To investigate the influence of natural selection pressure on the genes of both chromosomes codon usage bias, we perform a correlation analysis using the Gravy and Aroma values and the codon compositions. The result reveal that gravy value is not correlated with the GC3 only correlation found in case of chromosome II genes of *Rhodobacter sphaeroides* and that the aroma value is rarely correlated with GC3 except the chromosome I genes of *Leptospira interrogans*. This confirmed that natural selection has no influence on the genes of both the chromosome of multichromosomal bacteria.<sup>11–14</sup>

### Conclusion

Our study reveals that *Vibrio cholera* and *Leptospira interrogans* with genomic GC content 36.68% and 47.7% shows relatively higher influence of mutational pressure on the genes of chromosome II. Where in other case *Rhodobacter sphaeroides* with genomic GC content 69.15%, chromosome I genes are mostly influenced by mutational pressure than chromosome II. Influence of Natural selection pressure is similar on the codon usage bias of the genes of chromosome I and chromosome II of multichromosomal bacteria. Considering the multichromosomal bacteria's with high genomic GC content, their primary chromosomes are greatly influenced by

mutational pressure in compare to secondary chromosome. From the analysis of these three category of multichromosomal genome in context of their GC content (High, Intermediate, Low) it is found that mutational pressure on different chromosomes are varied but natural selection on the genes of both the chromosome are same throughout genomes of multichromosomal bacteria.

## Acknowledgements

None.

## Conflict of interest

The author declares no conflict of interest.

## References

1. Puigbò P, Bravo IG, García-Vallvé S. CAIcal: A combined set of tools to assess codon usage adaptation. *Biol Direct*. 2008;3:38.
2. Chen, Youhua. A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: quantifying the relative importance of mutational pressure and natural selection. *BioMed research international*. 2013;2013:406342.
3. Jha JK, Baek JH, Venkova-Canova T, et al. Chromosome dynamics in multichromosome bacteria. *Biochim Biophys Acta*. 2012;1819(7):826–829.
4. Stenico M, Lloyd AT, Sharp PM. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res*. 1994;22(13):2437–2446.
5. Heidelberg JF, Eisen JA, Nelson WC, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature*. 2000;406(6795):477–483.
6. Lobry JR, Gautier C. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res*. 1994;22(15):3174–3180.
7. Liu Q. Analysis of codon usage pattern in the radioresistant bacterium *Deinococcus radiodurans*. *Biosystems*. 2006;85(2):99–106.
8. Yang X, Luo X, Cai X. Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset. *Parasit Vectors*. 2014;7:527.
9. Chen Y. A Comparison of Synonymous Codon Usage Bias Patterns in DNA and RNA Virus Genomes: Quantifying the Relative Importance of Mutational Pressure and Natural Selection. *Biomed Res Int*. 2013;2013:406342.
10. Joy F, Basak S, Gupta SK, et al. Compositional Correlations in Canine Genome Reflects Similarity with Human Genes. *Journal of Biochemistry and Molecular Biology*. 2006;39(3):240–246.
11. Salanoubat M, Genin S, Artiguenave F, et al. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*. 2002;415(6871):497–502.
12. Suwanto A, Kaplan S. Physical and Genetic Mapping of the *Rhodobacter sphaeroides* 2.4.1 Genome: Presence of Two Unique Circular Chromosomes. *J Bacteriol*. 1989;171(11):5850–5859.
13. Cooper VS, Vohr SH, Wrocklage SC, et al. Why Genes Evolve Faster on Secondary Chromosomes in Bacteria. *PLoS Comput Biol*. 2010;6(4):e1000732.
14. Nascimento AL, Ko AI, Martins EA, et al. Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J Bacteriol*. 2004;186(7):2164–2172.