



*International Journal Of*  
**Recent Scientific  
Research**

ISSN: 0976-3031  
Volume: 7(3) March -2016

**COMPARATIVE *IN SILICO* GENOMIC ANALYSIS OF *MYXOCOCCUS* — AN  
ENIGMATIC EUBACTERIAL GENUS**

Ayon Pal



THE OFFICIAL PUBLICATION OF  
INTERNATIONAL JOURNAL OF RECENT SCIENTIFIC RESEARCH (IJRSR)  
<http://www.recentscientific.com/> [recentscientific@gmail.com](mailto:recentscientific@gmail.com)

**RESEARCH ARTICLE****COMPARATIVE *IN SILICO* GENOMIC ANALYSIS OF *MYXOCOCCUS* — AN ENIGMATIC EUBACTERIAL GENUS****Ayon Pal**

Department of Botany, Raiganj University, Raiganj-733134, West Bengal, INDIA

---

**ARTICLE INFO****Article History:**Received 06<sup>th</sup> December, 2015Received in revised form 14<sup>th</sup>

January, 2016

Accepted 23<sup>rd</sup> February, 2016Published online 28<sup>th</sup>

March, 2016

**Keywords:**Gliding bacteria; fruit-body; *Myxococcus*; myxococci; codon usage bias; Nc; GC3; correlation analysis; GRAVY

---

**ABSTRACT**

The large amount of genomic information present in the global databases may be analyzed from a wide variety of perspectives to obtain novel information regarding the functional relation between different organisms as well as their comparative physiology, and lifestyle. In this study, an in-depth comparative codon utilization analysis has been employed to study the genome design of an interesting bacterial species *Myxococcus*, utilizing the whole genome of four myxococci species. *Myxococcus* represents the myxobacteria characterized by gliding motility, fruit-body formation and the production of lytic principles capable of attacking different bacterial and fungal cells on which the formation of fruiting bodies readily takes place. Among the four species analyzed in this study, *M. fulvus* was found to display the most codon biased genome with ninety percent of its protein coding genes having Nc values below 40. Correlation between the different codon usage parameters showed strong negative correlation between Nc and GC3; the GRAVY score, which is a measure of hydropathicity of a protein was also found to be influenced by the genic GC content. An interesting observation of this comparative intra-specific codon usage utilization study was the relatively weak positive correlation between GC3 and GC content in the protein coding genes of the four *Myxococcus* species.

**Copyright © Ayon Pal., 2016**, this is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

**INTRODUCTION**

The advent of modern sequencing methods (Mardis 2008, Morozova and Marra 2008, Metzker 2010), and the creation of state of the art biological databases (Baxevanis 2001, Birney and Clamp 2004, Benson, Cavanaugh *et al.* 2013), puts forward the attractive proposition of comparative genomics. At present, there are more than forty thousand fully sequenced bacterial genomic datasets in the genome databases worldwide (Markowitz, Chen *et al.* 2012, Benson, Cavanaugh *et al.* 2013). The large amount of bacterial genomic information present in the global databases may be analyzed from a wide variety of perspectives to obtain novel information regarding the functional relation between the different organisms, as well as their comparative physiology, and lifestyle. In this study, codon utilization analysis has been employed to study the genome design of an interesting bacterial species *Myxococcus*. This organism belongs to the order Myxococcales of the class – Proteobacteria, and represents Gram-negative, spore-forming, chemoorganotrophic, obligate aerobes (Goldman, Bhat *et al.* 2007). Myxococci are nonflagellated elongated rods with rounded or tapered ends. *Myxococcus* represents the myxobacteria characterized by gliding motility, fruit-body

formation and the production of lytic principles capable of attacking different bacterial and fungal cells on which the formation of fruiting bodies readily takes place (Singh 1947, Nolte 1957, Mason and Powelson 1958, Berg 1975); three peculiar features possessed together by any known bacteria till date. Important species of *Myxococcus* include *M. fulvus*, *M. stipitatus*, *M. virescens*, *M. xanthus*, *M. coralloides*, *M. flavaescens*, and *M. macrosporus*.

An in-depth comparative codon usage analysis of the whole genome of four *Myxococcus* species was carried out in this study. These four genomes includes *M. fulvus* HW-1 (Li, Li *et al.* 2011), *M. stipitatus* DSM 14675 (Huntley, Kneip *et al.* 2013), *M. virescens* DSM 2260 and *M. xanthus* DK 1622 (Goldman, Nierman *et al.* 2006). *Myxococcus fulvus* is a halotolerant marine myxobacterium which is reported to exhibit complex social behaviors in the presence of low concentrations of seawater but adopts an asocial living pattern under oceanic conditions (Li, Li *et al.* 2011). *M. stipitatus* is characterized by the production of morphologically distinct fruiting bodies and secondary metabolites with cytotoxic or antibiotic actions (Huntley, Kneip *et al.* 2013). In comparison to the other members of *Myxococcus*, *M. stipitatus* generates a more complex fruit body structure where a mass of myxospores

---

\*Corresponding author: Ayon Pal

Department of Botany, Raiganj University, Raiganj-733134, West Bengal, INDIA

is placed on top of a cell-free stalk (Dawid 2000). *M. virescens* is known to produce a family of at least 12 closely related antibiotics, called the myxovirescins which are effective against Gram negative as well as Gram positive bacteria (Gerth, Irschik *et al.* 1982). *M. xanthus* is perhaps the most prolific of all the myxococci. It is a soil bacterium which commonly grows in damp top soil rich in organic matter and displays different forms of self-organizing behavior as a response to environmental cues. During normal conditions with abundant food, it exists as a predatory, saprophytic single-species biofilm called a swarm, whereas, during starvation period, it undergoes a multicellular development cycle (Kroos, Kuspa *et al.* 1986). So, *Myxococcus* represents a genus quite enigmatic in comparison to other known bacteria, and different members of this genus display unique features.

## MATERIALS AND METHODS

The whole genome sequences of the four *Myxococcus* species namely, *M. fulvus* HW-1 (Li, Li *et al.* 2011), *M. stipitatus* DSM 14675 (Huntley, Kneip *et al.* 2013), *M. virescens* DSM 2260 and *M. xanthus* DK 1622 (Goldman, Nierman *et al.* 2006) were obtained from the IMG database (Markowitz, Chen *et al.* 2012). The different codon usage parameters like effective codon number (Nc), GC content, GC content at the third position of the codon (GC3) and grand average of hydropathy (GRAVY) were calculated using the software CodonW (Peden 1999). Nc quantifies the "effective" number of codons that are used in a gene (Wright 1990). For the nuclear universal genetic code, the value of Nc ranges from 20 (*i.e.*, the codon bias is maximum) to 61 (*i.e.*, no codon bias). It has been shown that this is one of the best measures to show codon usage bias or CUB (Comeron and Aguade 1998). An Nc value of less than forty (Nc<40) is considered as the hallmark of major CUB (Lü, Zhao *et al.* 2005, Lahr, Nguyen *et al.* 2011, Liu, Zhang *et al.* 2011, Belalov and Lukashev 2013, Butt, Nasrullah *et al.* 2014), suggesting a significant CUB at the whole genome level. Information regarding intra-species and inter-species synonymous codon usage variation can be accounted for by studying the variation in GC content in the third position of a codon. GC3 represents the guanine and cytosine content at the third position of a codon (Wright 1990) and have been found to play a vital role in cell function (Epstein, Lin *et al.* 2000, Smith and Eyre-Walker 2001). It is one of the major driving force of CUB (Muto and Osawa 1987, Wan, Xu *et al.* 2004). GC3 is defined as the proportion of GC content in the third codon position, excluding methionine and tryptophan (nuclear universal genetic code) (Wright 1990).

A plot of Nc versus GC3s provides a useful visual display of the main features of codon usage patterns for a number of genes. Such a plot is referred to as the Nc-plot. Modified Nc-plot depicting the correlation between Nc and GC3 was used to explore the variation in inter-specific synonymous codon usage patterns within the four species of *Myxococcus* included in this study.

The grand average of hydropathicity or GRAVY (Kyte and Doolittle 1982) of the linear polypeptide sequence was calculated as the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence. Increasing

positive score indicates greater hydrophobicity. The calculation is based on the Kyte-Doolittle scale (Kyte and Doolittle 1982). It is a simple method for displaying the hydropathic character of a protein. The GRAVY scores of all the protein coding genes in the four myxococci was calculated in course of this study.

## RESULTS AND DISCUSSION

CUB is one of the major forces responsible for genome evolution and is a result of the culmination of a variety of factors (Pal, Banerjee *et al.* 2015). The codon usage pattern of 30,229 protein coding genes belonging to the four species of *Myxococcus* were thoroughly analyzed during this study. Figure 1 depicts the number of protein coding genes present in the four myxococci species along with their genome size. From Figure 1, it is evident that *M. stipitatus* has the largest genome among the four myxococci species, together with the largest number of protein coding genes (8043 genes). *M. fulvus* was found to display the lowest number of protein coding genes (7284 genes) followed by *M. xanthus* (7369 genes) and *M. virescens* (7533 genes). The coding efficiency, which is calculated as the total percentage of genomic DNA bases involved in coding the different genes, was found to be the lowest in *M. fulvus*. While all the other three species uses more than ninety percent of their genome size to code for different proteins and RNA, *M. fulvus* demonstrated a coding efficiency of about 88% only.

The Nc score, which is a satisfactory reflection of codon bias (Wright 1990), of the 30,229 protein coding genes belonging to the four species of *Myxococcus* was estimated and histograms were constructed (Figure 2). From figure 2, it is clearly evident that the myxococci have a codon biased genome, and the genome of *M. fulvus* is quite unique from the rest of the myxococci considered in this study. Out of the 7270 protein coding genes in *M. fulvus*, about 55% have biased codon usage pattern and 40% have strongly biased codon usage pattern. In comparison, the other genomes have about 13% to 18% of the protein coding genes demonstrated strongly biased codon usage pattern.

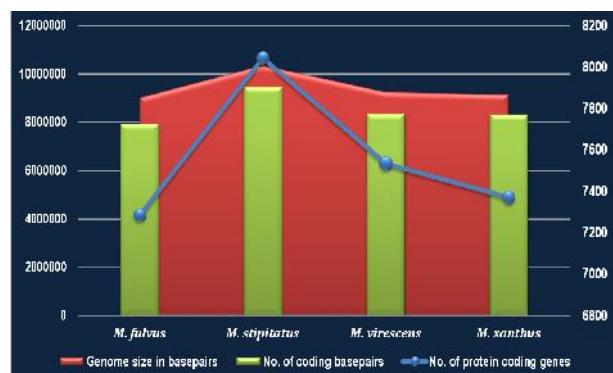
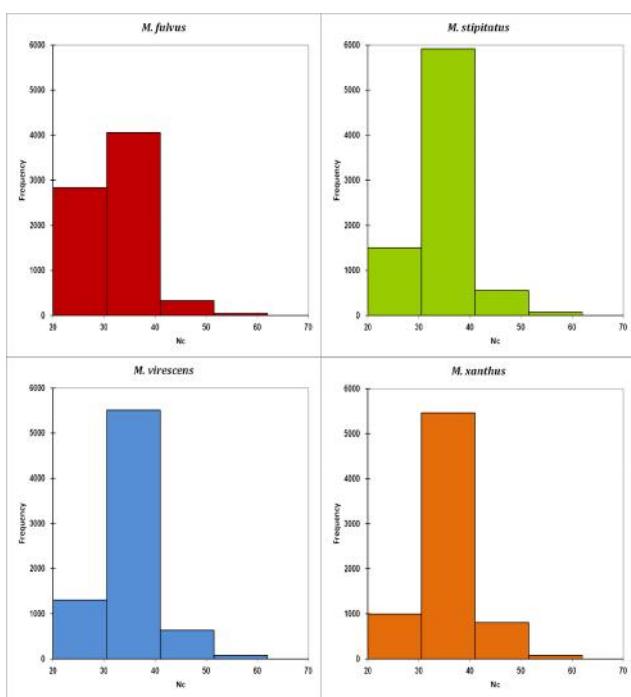
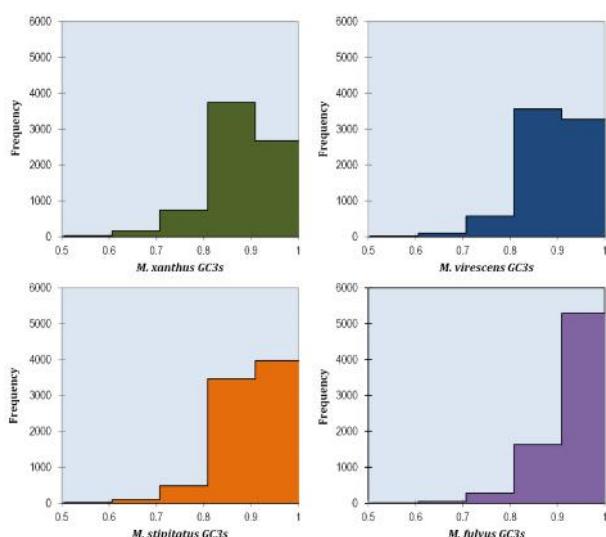


Figure 1 A comparative depiction of the genome size, number of coding base pairs and the frequency of protein coding genes in the four *Myxococcus* species considered in this study. The red area depicts the genome size (in base pairs), the green bars represent the number of coding pairs and the blue line denotes the number of protein coding genes in the four *Myxococcus* species (plotted on the secondary axis on the right side).



**Figure 2** Histograms showing the frequency of the genes in the different Nc range for the four species of *Myxococcus* considered in this study. The frequency or the number of genes is plotted on the y-axis whereas, the x-axis denotes the Nc values.

The GC3 (Wright 1990) distribution of the four myxococci species were also calculated and the data was used to construct the histograms shown in Figure 3.



**Figure 3** Histograms showing the frequency of genes in the different GC3 range for the four species of *Myxococcus* considered in this study. The frequency or the number of genes is plotted on the y-axis whereas, the x-axis denotes the GC3 values.

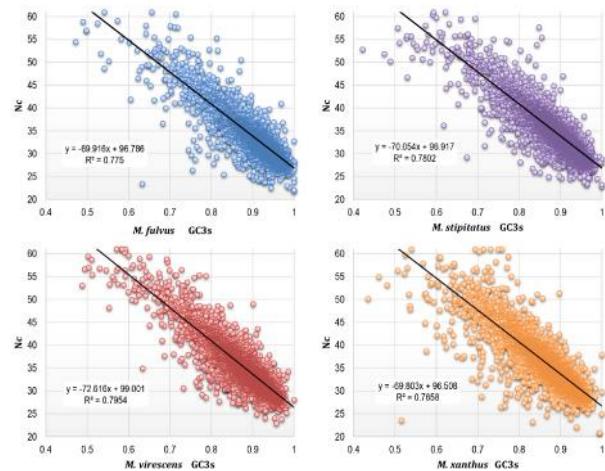
From this figure it is clearly evident that the four species included in this analysis, have unique genomic GC3 signature. A common attribute of the GC3 signature was that, more than eighty percent of the protein coding genes in all the species had GC3 scores greater than 80% and in the case of *M. fulvus* ninety percent (i.e., 6924 protein coding genes) of the genes were found to have GC3 scores above 80%. Table 1 depicts the range in GC3 fluctuation along with the mean and standard

deviation found within the four myxococci analyzed in this study.

**Table 1** A comparative account of the GC3 values in the four myxococci species analyzed in this study.

Organism	No. of protein coding genes	Minimu m GC3	Maximu m GC3	Mean GC3	Std. deviation
<i>M. xanthus</i>	7349	0.435	1.000	0.877	0.064
<i>M. virescens</i>	7529	0.488	1.000	0.887	0.060
<i>M. stipitatus</i>	8042	0.422	1.000	0.895	0.059
<i>M. fulvus</i>	7270	0.472	1.000	0.922	0.055

Modified Nc-plots were constructed to gain a better idea about the inter-species codon usage pattern of *Myxococcus* (Figure 4), and it was observed that *M. fulvus* have the most number of scattered or isolated genes on the Nc-plot compared to the other three species.



**Figure 4** A modified Nc-plot showing the Nc values on the y-axis plotted against the GC3 values on the x-axis. A black continuous line depicting the linear trendline fitting the data along with the equation and  $R^2$  value is also shown separately for the four *Myxococcus* species considered in this study.

A thorough Spearman's rank correlation analysis (with a significance level alpha = 0.01) of the different codon usage parameters such as Nc, GC3, GC, GRAVY and the occurrence of the four bases A, T, G and C at the third position of a codon (A3, T3, G3 and C3 respectively) was worked out to have an idea about the influence of the different factors in shaping codon usage pattern, and genome design of the four species of *Myxococcus*. It was observed that, the occurrence of T and A at the third position of a codon in a protein coding gene, was significantly correlated with the Nc in all the four *Myxococcus* species. This suggests that in the myxococci, codon bias is significantly reduced by the occurrence of T and A residues at the third position of a codon. The codon bias was also found to be strongly anti-correlated with the GC3 value ( $r_s > -0.8$ ,  $p < 0.01$ ) as is expected for a highly codon biased genome like the ones depicted by the myxococci considered in this study. An interesting feature that emerged from the correlation analysis was the significantly feeble possible correlation between the GC3 and GC values of the protein coding genes in the four myxococci with correlation coefficient running as low as  $r_s = -0.237$  ( $p < 0.01$ ). This is an aberration since high GC containing genomes are supposed to depict strong positive correlation between GC and GC3 content. Weak positive correlation was also observed between the GC content of the protein coding gene and its GRAVY score. This is indicative of

the fact that, in the four *Myxococcus* species included in this study, the hydropathicity of the proteins are to some extent, determined by the genic GC content.

## CONCLUSIONS

The *Myxococcus* represents a unique genus bestowed with the ability of gliding motion and fruit body formation. Among the four species analyzed in this study, *M. fulvus* was found to display the most codon biased genome with ninety percent of its protein coding genes having Nc values below 40. *M. fulvus* was also found to have the most number of scattered or isolated genes on the Nc-plot compared to the other three species considered in this study. Correlation between the different codon usage parameters in the myxococci showed strong negative correlation between Nc and GC3, and the GRAVY score, which is a measure of hydropathicity of a protein was found to be influenced by the genic GC content. Another interesting observation of this comparative intra-specific codon usage utilization study was the relatively weak positive correlation of GC3 with GC values in the protein coding genes of the four *Myxococcus* species analyzed in course of this study.

## References

- Baxevanis, A. D. (2001). "The Molecular Biology Database Collection: an updated compilation of biological database resources." *Nucleic Acids Research* **29**(1): 1-10.
- Belalov, I. S. and A. N. Lukashev (2013). "Causes and Implications of Codon Usage Bias in RNA Viruses." *PLoS ONE* **8**(2): e56642.
- Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers (2013). "GenBank." *Nucleic Acids Res* **41**(Database issue): D36-42.
- Berg, H. C. (1975). "Bacterial behaviour." *Nature* **254**(5499): 389-392.
- Birney, E. and M. Clamp (2004). "Biological database design and implementation." *Briefings in bioinformatics* **5**(1): 31-38.
- Butt, A. M., I. Nasrullah and Y. Tong (2014). "Genome-Wide Analysis of Codon Usage and Influencing Factors in Chikungunya Viruses." *PLoS ONE* **9**(3): e90905.
- Comeron, J. M. and M. Aguade (1998). "An evaluation of measures of synonymous codon usage bias." *J Mol Evol* **47**(3): 268-274.
- Dawid, W. (2000). "Biology and global distribution of myxobacteria in soils." *FEMS Microbiol Rev* **24**(4): 403-427.
- Epstein, R. J., K. Lin and T. W. Tan (2000). "A functional significance for codon third bases." *Gene* **245**(2): 291-298.
- Gerth, K., H. Irschik, H. Reichenbach and W. Trowitzsch (1982). "The myxovirescins, a family of antibiotics from *Myxococcus virescens* (Myxobacterales)." *J Antibiot (Tokyo)* **35**(11): 1454-1459.
- Goldman, B., S. Bhat and L. J. Shimkets (2007). "Genome Evolution and the Emergence of Fruiting Body Development in <italic>Myxococcus xanthus</italic>." *PLoS ONE* **2**(12): e1329.
- Goldman, B. S., W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. A. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand and H. B. Kaplan (2006). "Evolution of sensory complexity recorded in a myxobacterial genome." *Proc Natl Acad Sci U S A* **103**(41): 15200-15205.
- Huntley, S., S. Kneip, A. Treuner-Lange and L. Søgaard-Andersen (2013). "Complete genome sequence of *Myxococcus stipitatus* strain DSM 14675, a fruiting myxobacterium." *Genome Announc* **1**(2): e0010013.
- Huntley, S., S. Kneip, A. Treuner-Lange and L. Søgaard-Andersen (2013). "Complete Genome Sequence of *Myxococcus stipitatus* Strain DSM 14675, a Fruiting Myxobacterium." *Genome Announcements* **1**(2): e00100-00113.
- Kroos, L., A. Kuspa and D. Kaiser (1986). "A global analysis of developmentally regulated genes in *Myxococcus xanthus*." *Dev Biol* **117**(1): 252-266.
- Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." *J Mol Biol* **157**(1): 105-132.
- Lahr, D. J., T. B. Nguyen, E. Barbero and L. A. Katz (2011). "Evolution of the actin gene family in testate lobose amoebae (Arcellinida) is characterized by two distinct clades of paralogs and recent independent expansions." *Mol Biol Evol* **28**(1): 223-236.
- Li, Z. F., X. Li, H. Liu, X. Liu, K. Han, Z. H. Wu, W. Hu, F. F. Li and Y. Z. Li (2011). "Genome sequence of the halotolerant marine bacterium *Myxococcus fulvus* HW-1." *J Bacteriol* **193**(18): 5015-5016.
- Liu, W.-q., J. Zhang, Y.-q. Zhang, J.-h. Zhou, H.-t. Chen, L.-n. Ma, Y.-z. Ding and Y. Liu (2011). "Compare the differences of synonymous codon usage between the two species within cardiovirus." *Virology Journal* **8**(1): 325.
- Lü, H., W.-M. Zhao, Y. Zheng, H. Wang, M. Qi and X.-P. Yu (2005). "Analysis of Synonymous Codon Usage Bias in Chlamydia." *Acta Biochim Biophys Sin (Shanghai)* **37**(1): 1-10.
- Mardis, E. R. (2008). "The impact of next-generation sequencing technology on genetics." *Trends in Genetics* **24**(3): 133-141.
- Markowitz, V. M., I.-M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N. N. Ivanova and N. C. Kyrpides (2012). "IMG: the integrated microbial genomes database and comparative analysis system." *Nucleic Acids Research* **40**(D1): D115-D122.
- Mason, D. J. and D. Powelson (1958). "The cell wall of *Myxococcus xanthus*." *Biochim Biophys Acta* **29**(1): 1-7.
- Metzker, M. L. (2010). "Sequencing technologies—the next generation." *Nature reviews genetics* **11**(1): 31-46.

- Morozova, O. and M. A. Marra (2008). "Applications of next-generation sequencing technologies in functional genomics." *Genomics* **92**(5): 255-264.
- Muto, A. and S. Osawa (1987). "The guanine and cytosine content of genomic DNA and bacterial evolution." *Proc Natl Acad Sci U S A* **84**(1): 166-169.
- Nolte, E. M. (1957). "[Studies on nutrition and fruit body formation in myxobacteria]." *Arch Mikrobiol* **28**(2): 191-218.
- Pal, A., R. Banerjee, U. K. Mondal, S. Mukhopadhyay and A. K. Bothra (2015). "Deconstruction of Archaeal Genome Depict Strategic Consensus in Core Pathways Coding Sequence Assembly." *PLoS ONE* **10**(2): e0118245.
- Peden, J. F. (1999). Analysis of codon usage, University of Nottingham.
- Singh, B. N. (1947). "Myxobacteria in soils and composts; their distribution, number and lytic action on bacteria." *J Gen Microbiol* **1**(1): 1-10.
- Smith, N. G. and A. Eyre-Walker (2001). "Why are translationally sub-optimal synonymous codons used in *Escherichia coli*?" *J Mol Evol* **53**(3): 225-236.
- Wan, X.-F., D. Xu, A. Kleinhofs and J. Zhou (2004). "Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes." *BMC Evol Biol* **4**(1): 19.
- Wright, F. (1990). "The 'effective number of codons' used in a gene." *Gene* **87**(1): 23-29.

\*\*\*\*\*

**How to cite this article:**

Ayon Pal.2016, Comparative *In Silico* Genomic Analysis of *Myxococcus* — An Enigmatic Eubacterial Genus. *Int J Recent Sci Res.* 7(3), pp. 9661-9665.

TSSN 0976-3031



9 770976 303009 >