

Genome-wide codon usage bias analysis in *Beauveria bassiana*

Prajna Muthabathula^{1,*}, Saragadam Suneetha¹, J. Ratna Grace²

¹Department of Botany, Andhra University, Visakhapatnam, Andhra Pradesh, India; ²Department of Botany, Dr. V. S. Krishna Government Degree College, Visakhapatnam, Andhra Pradesh, India; Prajna Muthabathula - Email: prajna.muthabathula@gmail.com; Phone: +919618426005; *Corresponding author

Received October 12, 2018; Revised October 27, 2018; Accepted October 27, 2018; Published December 28, 2018
doi: 10.6026/97320630014580

Abstract:

Codon usage bias analysis allows in identifying the factors that are influencing and contributing to shape the evolution of the organisms. Therefore, it is of interest to analyze 10363 gene sequences from *Beauveria bassiana*. The GC content with 51.50% is higher than the AT content (48.50%) in *B. bassiana*. The fungal nuclear genes tend to be GC rich and predominantly G/C ending. Codon usage bias exhibited by *B. bassiana* is based on the Relative Synonymous Codon Usage (RSCU) values of 61 sense codons, of which 28 codons are with RSCU value larger than 1. Other factors like Nucleotide composition, mutational pressure and selection also has a role in shaping the codon usage bias. We identified 24 optimal codons that end with G or C. Correlation analysis suggests existence of translational efficiency of amino acids. Based on the GC3s distribution evolution of the *B. bassiana* genes is by the contribution of mutation pressure. ENC may be the major factor in shaping the codon usage bias. This study provides insights into the compositional selection pressure of the genes in *B. bassiana*.

Keywords: *Beauveria bassiana*, codon usage, translational efficiency, mutational pressure

Background:

The probability of the codon used for an amino acid over a different may act resulting in different clusters of synonymous codon usage codon which codes for the same amino acid is regarded as codon bias. patterns among genes within the genome [3].

Different codons that encode the same amino acid are known as synonymous codons. Even though synonymous codons encode the same amino acid it has been shown that for a wide variety of their consequences and causes and helps in identifying the selective organisms different synonymous codons are used with different forces that are involved in shaping the evolution of the codon usage frequencies. This phenomenon is termed as codon bias [1]. It is found patterns, which help in understanding the perspectives of genome in all eukaryotic and prokaryotic genomes. Codons used more often biology [4]. Codon usage bias of several organisms have been are referred to as optimized codons or preferred codons. Synonymous analyzed however, very little is known about codon usage bias in codon usage identity may be varying or similar in the genome or *B. bassiana* entomopathogen, belonging to Hypocrealean fungi among different genes within the genome. Several factors that (cordycepsitaceae, Ascomycota) that is used as a potential biopesticide. influence the variations in the codon usage patterns which include It is an environmental friendly mycoinsecticide which is commercially genetic drift, mutational pressure and natural selection [2] and these available whose genome was sequenced and light has shed on its factors are highly responsible for differences in codon usage differential gene expression and adaptability to different niches [5]. It variations among different organisms. Multiple forms of selection has diversifying roles apart from biopesticidal activity, also found as an endophyte both naturally and from inoculated samples and had a

role in suppressing plant pathogens [6, 7] which makes it more interesting to make further investigations to go through the details of genetic content. The accuracy and efficiency of protein production can be modulated with differences in codon usage while maintaining the same protein sequence [8]. Synonymous codon usage patterns identification proves useful in identifying the genes likely under translational selection [9]. In this study we analyzed the codon usage bias of *B. bassiana*. The objectives of the present study are to investigate the presence of codon bias and to identify the preferred codons in the *B. bassiana* genome and to examine the contribution of influencing factors on the usage of synonymous codons.

Methodology:

Sequence data:

The 10363 CDS (Coding domain sequences) dataset of *B. bassiana* (ASM28067v1) from the whole genome sequence were downloaded from National Centre for Biotechnology Information (NCBI) in FASTA (fasta and fna) format (<http://www.ncbi.nlm.nih.gov/genome/>).

Codon usage indices:

Codon usage indices such as GC, GC1, GC2, GC3s, A3s, T3s, C3, and G3s were calculated using CALcal tool [10]. Relative Synonymous Codon Usage (RSCU) values and Codon Adaptation Index (CAI) values were also calculated [10]. General average hydropathicity (GRAVY) and Aromo values (frequency of aromatic amino acids) in the hypothetical translated gene product were also calculated [8].

Effective Number of Codons (ENC) and ENC plot:

ENC is assessment of non-uniformity of usage within synonymous groups of codons [11]. ENC values vary from 20 (extreme bias i.e., only one codon is used for one amino acid) and 61 (random bias i.e., codons used randomly). ENC values were plotted against GC3s values to find out the codon usage bias influencing factor [11].

Relative Synonymous Codon Usage (RSCU):

RSCU is defined as the ratio of observed frequency of codons to the expected frequency. If the RSCU value is equal to 1 the codon is not biased and if RSCU value is >1 codon is frequently used.

Codon Adaptation Index (CAI):

CAI is a measurement of the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes. CAI values range from 0-1. The higher values indicate a higher level gene expression as well as codon bias [12].

Neutrality plot:

The GC content is calculated according to the first, second and third codon positions (GC1, GC2 and GC3 respectively). GC12 is the average of GC1 and GC2 used for the analysis of neutrality plot (GC12 against GC3). Neutrality plot is used to analyze the relationship between GC12 and GC3, and the factors influencing the codon usage bias [13, 15].

Software and statistical analysis:

RSCU, ENC, total G+C genomic content, as well as COA, were calculated by codonW 1.4 version (<http://codonw.sourceforge.net/>). Values of CAI, GC1, GC2 and GC3 were calculated by CALcal server (<http://genomes.urv.cat/CALcal/>). Statistical analysis was done using R software 3.4.1 version (www.r-project.org) and GraphPad Prism 7.03. Graphs were constructed using GraphPad Prism 7.03 (GraphPad Software Inc., La Jolla, CA, USA).

Results:

Base composition:

The GC content of 10363 genes distributed between 23.66% to 72.70%, GC12 being distributed between 40.00% to 60.00% (Figure 1). There is a great difference in the GC content of GC2 and GC3, 45.56±5.58 and 66.96±10.44 respectively (Table 1).

Neutrality plot:

To characterize the correlation among three positions of GC the neutrality plot is drawn. The relationship between GC12 and GC3 was revealed with neutrality plot (Figure 1). The neutrality plot reveals that the genes of *B. bassiana* exhibit a wide range of GC3 values, ranging from 20.16% to 95.78%. If a gene is located on the diagonal line with a significant correlation between GC12 and GC3, it indicates that the gene is under neutral selection pressure. The points (genes) were located above the regression curve (bold line) with a slope less than 1, indicating that the natural selection pressure is dominating the composition of coding codons in *B. bassiana*. GC12 and GC3s showed a significant positive correlation ($r = 0.3348, p < 0.001$). The slope of regression line for all genes was 0.1196 which indicates that the effect of mutation pressure is 11.96% and the influence from other factors is around 88.04%.

Effective Number of codons (ENC) and GC3s association

The ENC of *B. bassiana* ranges from 24.68 to 61.00 with an average of 48.02. Among 10363 genes 808 genes exhibited high codon bias (ENC < 35), indicating that *B. bassiana* genes, in general exhibit random codon usage without strong codon bias.

Table 1: Base composition, ENC, GRAVY and AROMO of codons of *Beauveria bassiana*

Class	Genes	Codons	GC%	GC3s	GC1 (%)	GC2 (%)	GC3 (%)		
Total	10363	5182650	57.24±4.97	65.79±10.88	59.19±4.86	45.56±5.58	66.96±10.44		
Class	T3s (%)	C3s (%)	A3s (%)	G3s (%)	Gravy	Aro	ENC	CAI	
Total	19.37±5.95	38.55±9.30	13.66±6.12	28.40±6.44	0.32±0.37	0.07±0.02	48.06±8.44	0.62±0.62	

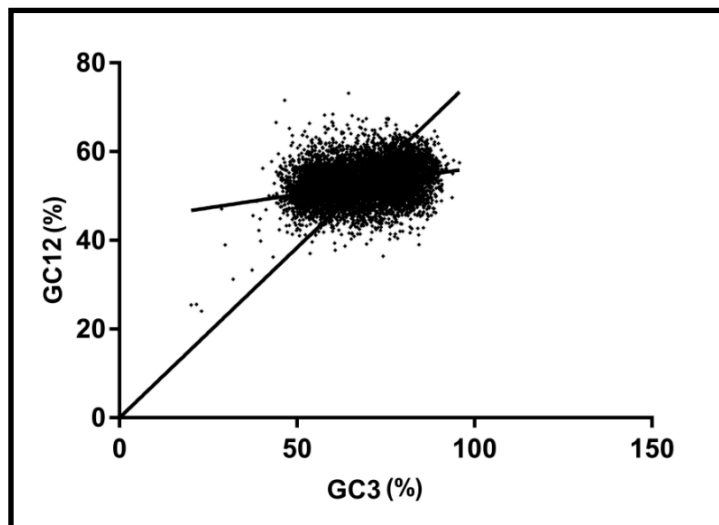


Figure 1: Neutrality plots (GC12 vs. GC3). GC12 is the average value of GC content in the first and second position of the codons. GC3 is the GC content at third position. The solid line is the linear regression of GC12 against GC3, $R^2=0.1105$, $p<0.001$

An ENC plot was generated to explore the influence of GC3s on codon bias in *B.bassiana*. If a gene is located on the expected curve, the codons of that gene are no bias. The GC3s distribution was in between 0.4 and 0.99, indicating that *B.bassiana* mainly evolved by mutation pressure (Figure 2). The distribution of ENC versus GC3s reveals, most of the points with low ENC values lay below the expected curve. This indicates that the mutational pressure and other factors are likely to be involved in determining the selective contribution on codon bias.

Table 2: Correlation coefficients between the codon usage indices along the first two major axes and the position of genes

	Length	GC	GC1	GC2	GC3	GC3s	A3s	T3s	C3s	G3s	GRAVY	AROMO	ENC	CAI	AXIS1
GC	-0.153**														
GC1	0.001	0.409**													
GC2	-0.044**	0.299**	0.155**												
GC3	-0.190**	0.595**	0.387**	0.129**											
GC3s	-0.192**	0.598**	0.398**	0.135*	0.999**										
A3s	0.135**	-0.509**	-0.270**	-0.088**	-0.879**	-0.879**									
T3s	0.213**	-0.525**	-0.416**	-0.147**	-0.850**	-0.848**	0.514**								
C3s	-0.140**	0.473**	0.182**	0.124**	0.810**	0.814**	-0.858**	-0.534**							
G3s	-0.074**	0.288**	0.388**	0.026**	0.457**	0.448**	-0.180**	-0.095**	-0.628**						
GRAVY	-0.030**	0.040**	-0.090**	-0.025*	0.096**	0.091**	-0.158**	-0.011	0.154**	-0.067**					
AROMO	0.001	-0.213**	-0.240**	-0.160**	-0.004	-0.016	-0.052**	0.062**	0.047**	-0.078**	0.362**				
ENC	0.089**	-0.774**	-0.254**	-0.090**	-0.596**	-0.598**	0.585**	0.450**	-0.556**	-0.148**	-0.136**	0.010			
CAI	0.087**	-0.920**	-0.364**	-0.197**	-0.608**	-0.608**	0.474**	0.583**	-0.427**	-0.385**	-0.058**	0.103**	0.717**		
AXIS1	-0.078**	0.825**	0.268**	0.091**	0.637**	0.639**	-0.612**	-0.490**	0.593**	0.179**	0.154**	0.016	-0.934**	-0.793**	
AXIS2	0.023**	-0.328**	-0.177**	-0.089**	-0.194**	-0.190**	0.017**	0.332**	0.061**	-0.458**	0.006	0.033**	-0.005	0.530**	-0.043**

Correlation between codon usage bias, gene length, Hydrophobicity and Aromaticity in *B.bassiana*

Correlation between the codon usage indices such as gene length, codon usage bias and hydrophobicity and aromaticity was determined using Spearman correlation analysis (Table 2). The values showed that the gene length was positively correlated with ENC ($r=0.089$, $p<0.001$), suggesting the contribution of gene length to codon usage bias. ENC was negatively correlated with first and second axes ($r= -0.934$, $p<0.001$; $r= -0.005$, $p<0.05$) and also with GC3 ($r= -0.598$, $p<0.001$). GRAVY (General Average Hydrophobicity) is negatively correlated with ENC ($r= -0.136$, $p<0.001$).

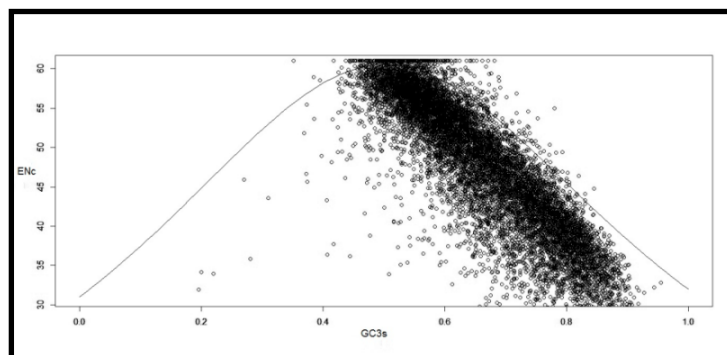


Figure 2: ENC plot between Effective Number of Codons and GC3s. Standard curve represents the respected ENC to GC3s. The codon usage pattern is affected by other factors besides nucleotide composition as most genes are far away from the standard curve.

Table 3: Codon usage of *Beauveria bassiana*

Amino Acid	Codon	Total count	RSCU	Amino Acid	Codon	Total count	RSCU
Phe	UUU	97,124	<u>1.04</u>	Ser	UCU	63,684	0.92
	UUC	90,092	0.96		UCC	81,713	<u>1.18</u>
	UUA	15,257	0.2		UCA	44,120	0.64
Leu	UUG	61,046	0.8	UCG	84,448	<u>1.22</u>	
	CUU	69,405	0.91	AGU	34,859	0.5	
	CUC	1,50,168	<u>1.96</u>	AGC	1,06,765	<u>1.54</u>	
	CUA	36,366	0.47	CCU	64,901	0.86	
Ile	CUG	1,27,510	<u>1.66</u>	CCC	98,366	<u>1.3</u>	
	AUU	1,00,384	<u>1.25</u>	CCA	56,636	0.75	
	AUC	1,12,518	<u>1.41</u>	CCG	82,011	<u>1.09</u>	
	AUA	27,324	0.34	ACU	56,368	0.74	
Met	AUG	1,15,524	<u>1</u>	ACC	99,999	<u>1.31</u>	
	GUU	63,597	0.8	ACA	58,385	0.77	
Val	GUC	1,43,919	<u>1.8</u>	ACG	89,814	<u>1.18</u>	
	GUA	28,295	0.35	GCU	1,01,796	0.82	
	GUG	83,697	<u>1.05</u>	GCC	2,02,360	<u>1.63</u>	
Tyr	UAU	44,250	0.64	GCA	76,891	0.62	
	UAC	93,320	<u>1.36</u>	GCG	1,16,536	0.94	
Cys	UGU	17,074	0.51	UGG	74,428	<u>1</u>	
	UGC	49,552	<u>1.49</u>	CGU	44,390	0.82	
His	CAU	45,803	0.73	CGC	1,25,865	<u>2.33</u>	
	CAC	80,397	<u>1.27</u>	CGA	47,855	0.89	
Gln	CAA	77,403	0.73	CGG	40,937	0.76	
	CAG	1,33,706	<u>1.27</u>	AGA	36,237	0.67	
Asn	AAU	66,104	0.73	AGG	29,101	0.54	
	AAC	1,15,356	<u>1.27</u>	GGU	72,400	0.81	
Lys	AAA	68,568	0.57	GGC	1,92,920	<u>2.16</u>	
	AAG	1,73,994	<u>1.43</u>	GGA	52,839	0.59	
Asp	GAU	1,12,138	0.74	GGG	39,116	0.44	
	GAC	1,92,144	<u>1.26</u>	UGA	3,421	0.99	
Glu	GAA	1,10,930	0.73	UAA	3,595	<u>1.04</u>	
	GAG	1,93,581	<u>1.27</u>	UAG	3,348	0.97	

RSCU is the relative synonymous codon usage, preferential codons are underlined.

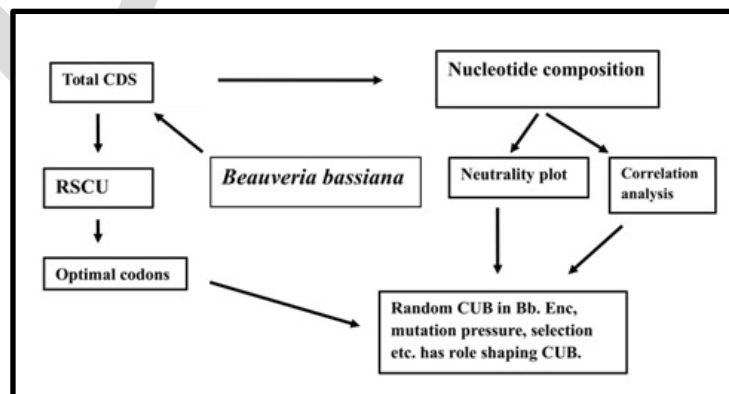


Figure 3:

CAI (Codon Adaptation Index) was negatively correlated with GC1, GC2, G3s and Axis1 ($r = -0.364, p < 0.001$; $r = -0.197, p < 0.001$; $r = -0.385, p < 0.001$ and $r = -0.934, p < 0.001$ respectively) and gene length and ENC were positively correlated with CAI ($r = 0.087, p < 0.001$; $r = 0.717, p < 0.001$) indicating that Nc and gene expression levels contribute to codon usage in a major way. This suggests that ENC may be the major factor in shaping the codon usage in *B. bassiana*.

GRAVY and CAI values showed positive correlation with Aromo ($r = 0.362, p < 0.001$; $r = 0.103, p < 0.001$) indicating that the hydrophobicity, CAI and aromaticity are the most important factors in amino acid usage. This provides a strong evidence for the existence of the selection for translational efficiency of amino acids in *B. bassiana*.

Table 4: Optimal codons of *Beauveria bassiana*

Amino acid	codon	High RSCU	N	Low RSCU	N
Phe	UUU	0.99	3493	1.02	4006
	UUC	1.01	3581	0.98	3865
	UUA	0.02	52	0.52	1782
	UUG	0.21	546	1.14	3947
Leu	CUU	0.37	956	1.25	4325
	CUC*	3.53	9236	1.14	3937
	CUA	0.13	351	0.78	2684
	CUG*	1.74	4564	1.17	4053
Ile	AUU	1.09	3160	1.19	4431
	AUC*	1.85	5373	1.11	4157
	AUA	0.05	159	0.7	2611
Met	AUG	1	3892	1	4860
	GUU	0.33	1072	1.11	3542
Val	GUC*	2.83	9154	1.21	3887
	GUA	0.08	245	0.66	2128
	GUG	0.76	2474	1.02	3260
Tyr	UAU	0.2	531	1.02	2903
	UAC*	1.8	4668	0.98	2809
His	CAU	0.22	434	1.09	3329
	CAC*	1.78	3436	0.91	2780
Gln	CAA	0.31	931	1.01	5088
	CAG*	1.69	4985	0.99	4946
Asn	AAU	0.27	815	1.04	4228
	AAC*	1.73	5282	0.96	3879
Lys	AAA	0.2	885	0.93	4872
	AAC*	1.8	8076	1.07	5585
Asp	GAU	0.29	1419	1.01	6628
	GAC*	1.71	8446	0.99	6509
Glu	GAA	0.33	1498	0.97	6665
	GAG*	1.67	7699	1.03	7095
	UCU	0.47	971	1.16	3849
	UCC*	2.01	4120	0.83	2740
Ser	UCA	0.16	324	1.06	3500
	UCG*	1.52	3103	0.92	3056
	AGU	0.12	243	0.78	2572
	AGC*	1.72	3523	1.26	4164
Pro	CCU	0.41	939	1.04	3239
	CCC*	2.41	5557	0.74	2302
	CCA	0.15	337	1.32	4115
	CCG*	1.04	2409	0.9	2816
Thr	ACU	0.33	862	1.03	3441
	ACC*	2.22	5828	0.86	2873
	ACA	0.21	549	1.23	4104
	ACG*	1.24	3259	0.88	2945
Ala	GCU	0.43	2078	1.07	5240
	GCC*	2.69	13066	0.9	74782
	GCA	0.13	616	1.15	5659
	GCG	0.76	3677	0.81	3992
Cys	UGU	0.13	135	0.86	1491
	UGC*	1.87	1916	1.14	1971
Trp	UGG	1	2681	1	3263
Arg	CGU	0.53	858	0.77	1920
	CGC*	4.5	7293	1.11	2778
	CGA	0.18	288	1.35	3373
	CGG	0.47	756	0.76	1916

Gly	AGA	0.13	215	1.21	3024
	AGG	0.19	312	0.81	2022
	GGU	0.47	1675	0.88	3074
	GGC*	3.22	11372	1.3	94816
	GGA	0.14	486	1.06	3678
	GGG	0.17	614	0.67	2332
TER	UAA	1.73	299	1.01	175
	UAG	0.76	132	1.01	175
	UGA	0.5	87	0.97	168

Th

The number of codons in the high bias dataset was 177993 and number of codons in low bias dataset was 226356. The low and high indicate the top and bottom of the dataset ordered by ENC ratio value respectively. *Optimal codons.

Optimal codons in *Beauveria bassiana*

Based on the RSCU values of 61 codons, codon bias exhibited by *B. bassiana* is weak. Twenty eight codons were frequently used which showed the high RSCU values such as CGC (RSCU=2.33), GGC (RSCU= 2.16) encoding Arg and Gly respectively. Most frequent codons ended with C or G, such as CUC (RSCU=1.96), GUC (RSCU=1.80), CUG (RSCU=1.66) and GCC (RSCU=1.63) (Table 3).

Each amino acid has the synonymous codons, the putative optimal codons of *B. bassiana* are given in (Table 4). There is a difference in number of synonymous codons for each amino acid. There were 25 optimal codons that end with G or C (G= 10/25, C= 15/25), which suggests the third position in the preferred codons may be related to the GC content. There are two or three optimal codons for each amino acid indicating that the codons were significantly correlated with translation levels.

Discussion:

Codon usage bias is an essential feature of all genomes [14]. The GC rich genome of *B. bassiana* can result in the dominance of the G/C ended codons, where AT rich genome of bacteria show the A/T ended optimized codons [15]. GC content close to 50% indicates little overall mutational bias in *Aspergillus nidulans*, an ascomycete [16], here *B. bassiana* also has a GC rich genome which too indicates a chance for mutational bias. One of the hypothesis proposed to explain variation of GC content in genome evolution is the "mutational biases hypothesis" is that GC content is driven by heterogenous mutational biases along genomes [17]. During the evolution of genomic structures G+C content could be the most important factors [18]. Neutrality plot results showed a significant positive correlation in *B. bassiana* indicating that the effect on the GC contents by the intragenomic GC mutation bias was similar at all three codon positions [19].

To investigate the synonymous codon usage, plotting ENC versus GC3s is an effective strategy [11]. ENC may play a role in shaping the codon usage in *B. bassiana*. Around 7.79% of genes only exhibited codon bias which indicates that there is random codon usage

without strong codon bias in *B.bassiana* [8]. Random codon usage bias in *B.bassiana* may result due to translational selection as it is responsible for the unequal codon usage of synonymous codons in protein coding genes in a wide variety of organisms [20]. The protein produced may not be affected by the synonymous codon chosen but it may relate to the expression of gene [21].

In particular, for species of fungi codon usage bias was driven by selection [22, 23, 24] and partly genetic interference in the model organism *Neurospora crassa* [25]. Codon usage bias is recognized as a critical factor contributing to gene expression and cellular function with its effects on processes like RNA processing to translation and protein folding [26]. Optimal codons were identified by comparing the low and high bias datasets, these codons if significantly correlate with translational levels [19], they would be helpful in designing degenerate primers in order to investigate evolutionary aspects of *B. bassiana*. *B.bassiana* exhibit no strong codon usage bias, there is a random codon usage bias. There is a strong evidence for selection of translational efficiency of amino acids and also there is the contribution of mutational pressure and other factors to codon usage bias. The codon usage in *B.bassiana* is dominated by the natural selection pressure.

Conclusion:

The present study brings out the codon usage details of entomopathogenic fungus *Beauveria bassiana*. We found no strong codon bias in *B.bassiana*. Reason for random or selective contribution of codon bias is mutational pressure and other factors like natural selection. There is also influence of translational efficiency of amino acids in shaping codon usage bias. Our analysis forms the footwork of genetic evolutionary aspects of *B.bassiana*. Further studies may reveal more details relating to the evolution and other molecular aspects of these fungi.

Acknowledgements:

I'm obliged and grateful to my guide Prof. K Uma Devi for all her support and blessings.

References:

- [1] Hershberg R & Petrov DA. *Annu. Rev. Genet.* 2008, **42**, 287-299. [PMID: 18983258]
- [2] Sharp PM & Li WH. *J. Mol. Evol.* 1986, **24**, 28-38. [PMID: 3104616]
- [3] Kober KM & Pogson GH. *G3: Genes, Genomes, Genetics.* 2013, **3**, 1069-1083. [PMID: 23637123]
- [4] Behura SK & Severson DW. *PloS one.* 2012, **7**, e43111. [PMID: 22912801]
- [5] Xiao G *et al. Sci. Rep.* 2012, **2**, 483. [PMID: 22761991]
- [6] Ownley BH *et al. J. Invertebr. Pathol.* 2008, **98**, 267-270. [PMID: 18442830]
- [7] Vega FE. *J. Invertebr. Pathol.* 2008, **98**, 277-279. [PMID: 18406422]
- [8] Li X *et al. Int. J. Mol. Sci.* 2016, **17**, 1138. [PMID: 27428961]
- [9] Sharp PM. *Nucleic Acids Res.* 1986, **14**, 5125-5143. [PMID: 3526280]
- [10] Puigbo P., Bravo IG & Garcia-Vallve S. *Bio. Direct.* 2008, **3.1**, 38. [PMID: 18796141]
- [11] Wright F. *Gene.* 1990, **87**, 23-29. [PMID: 2110097]
- [12] Sharp PM & Li WH. *Nucleic Acids Res.* 1987, **15**, 1281-1295. [PMID: 3547335]
- [13] Sueoka N. *Proc. Natl. Acad. Sci.* 1988, **85**, 2653-2657. [PMID: 3357886]
- [14] Zhou Z *et al. Proc. Natl. Acad. Sci.* 2016, **113**, E6117-E6125. [PMID: 27671647]
- [15] Liu J *et al. Int. J. Mol. Sci.* 2016, **17**, 1304. [PMID: 27517915]
- [16] Lloyd AT & Sharp PM. *Mol. Gen. Genet. MGG.* 1991, **230**, 288-294. [PMID: 1745237]
- [17] Niu Z., Xue Q., Wang H., Xie X., Zhu S., Liu W., & Ding X. 2017. *Int. J. Mol. Sci.* **18**(11), 2307. [PMID: 29099062]
- [18] Bellgard M *et al. J. Mol. Evol.* 2001, **53**, 465-468. [PMID: 11675606]
- [19] Zhang WJ *et al. J. Integr. Plant Biol.* 2007, **49**, 246-254.
- [20] Reis MD *et al. Nucleic Acids Res.* 2004, **32**, 5036-5044. [PMID: 15448185]
- [21] Grantham R *et al. Nucleic Acids Res.* 1980, **8**, 197-197. [PMID: 6986610]
- [22] Bennetzen JL & Hall BD. *J. Bio. Chem.* 1982, **257**, 3026-3031. [PMID: 7037777]
- [23] Ikemura T. *Mol. Biol. Evol.* 1985, **2**, 13-34. [PMID: 3916708]
- [24] Novembre JA. *Mol. Biol. Evol.* 2002, **19**, 1390-1394. [PMID: 12140252]
- [25] Whittle CA., Sun Y & Johannesson H. *Mol. Biol. Evol.* 2012, **29**, 1975-1986. [PMID: 22334579]
- [26] Plotkin JB & Kudla G. *Nat. Rev. Genet.* 2011, **12**, 32-42. [PMID: 21102527]

Edited by P. Kanguane

Muthabathula *et al.* *Bioinformatics* 14(9): 580-586 (2018)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.

BIOINFORMATION

Discovery at the interface of physical and biological sciences



Biomedical Informatics Society

Agro Informatics Society



Journal