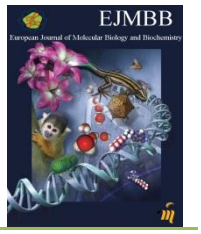
e - ISSN - 2348-2206
Print ISSN - 2348-2192

European Journal of Molecular Biology and Biochemistry

Journal homepage: www.mcmed.us/journal/ejmbb



ANALYSIS OF CODON USAGE PATTERN AND PREDICTION OF POTENTIALLY HIGHLY EXPRESSED GENES IN *DROSOPHILA*

[1]Poonam Chilana, [1]Jyotika Bhati, [2]Anu Sharma*, [2]D.C. Mishra, [3]Anil Rai

[1]Research Associate, [2]Scientist, [3]Head and Principal Scientist, Centre for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Pusa, New Delhi – 110 012, India.

Article Info

Received 02/08/2014
Revised 16/08/2014
Accepted 23/08/2014

Key words:-
Insecticide resistance, CYPs, GSTs, COEs, synonymous codon usage, Correspondence analysis.

ABSTRACT

Insecticide resistance development is one of the most serious concerns worldwide. The mechanism of this resistance has been attributed to evolutionary changes in pest/insect genomes such as alteration of drug target sites, up-regulation of degrading enzymes, and enhancement of drug excretion. In most of insects, this resistance is associated with higher amounts of insecticide detoxification enzymes as cytochrome P450s (CYPs), glutathione-s-transferase (GSTs) and carboxylesterases (COEs). The main objective of this study is to apply synonymous codon usage bias to understand the expressivity level of different categories of resistant genes in model insect, *Drosophila melanogaster*. Codon usage of different detoxifying gene families was analyzed and a comparative analysis of codon usage and RSCU values was calculated. The high level of homogeneity is seen in different gene families of *D. melanogaster*. Codon usage of these gene families is mainly determined by compositional constraints however, translational selection is also operating in shaping the codon usage variation among the genes of different insects. The study reveals that G/C-ending codons are preferred over A/T-ending codons in highly expressed genes. This study also helps to develop new strategies to mitigate the insecticide resistance development in pest insects which is serious concern for agriculture and human health.

INTRODUCTION

Insecticide resistance is a phenomenon of reduction of sensitivity of insecticide on insects. Long-term intensive use of chemical insecticides to control insect pests and disease vectors is often cited as the reason behind the development of insecticide resistance in insect populations. The resistant genes are now common in insect genomes due to the frequent use of chemicals and insecticides. In fact it is an on-going challenge to agricultural production managers. Approximately 600 species of insects are identified as resistant to at least one insecticide (www.pesticideresistance.org). In insects, three families of protein play crucial roles in detoxification of insecticides/xenobiotics *i.e.* the cytochrome P450s (CYPs), glutathione-s-transferase (GSTs) and carboxylesterases (COEs). In this new genomic era it is very much pertinent to identify and understand gene expression levels of various genes involved in this process of insecticide resistance at the molecular level. Extensive studies have also shown that synonymous codon usage biases are associated with gene expression level, gene length, gene translation initiation signal, protein amino acid composition, protein structure, tRNA abundance, mutation frequency and patterns, and GC compositions. Quantification of codon usage bias, especially at genomic scale, also helps to understand evolution of living organisms [1]. *Drosophila* use the genetic and biochemical mechanisms to develop insecticide resistance including single-site changes in target molecules resulting from point mutations and up regulation of degradative enzymes, particularly CYP and GST enzymes. In this specific area of insecticide metabolism and resistance, the *Drosophila*

Corresponding Author

Anu Sharma
Email:- anu@iasri.res.in

genome sequence provided a first complete picture of the detoxifying genes (CYPs) diversity which is available in abundance in insects with about 85 active CYP genes [2] while other insect genomes harbour more or fewer CYP genes [3]. Resistance to insecticides in pest/insects is also attributed to increased levels of GST activity. GSTs can protect the insect either by increasing the rate of detoxification of insecticide into non-toxic product e.g. the dehydrochlorination of the organochlorine, DDT [1,1,1-trichloro-2,2-bis-(p-chlorophenyl)ethane] [4] and the Odealkylation or O-dearylation of organophosphate insecticides [5] or by relieving the damage from oxidative stress induced by insecticide exposure [6]. COEs is a multigene family and occurs in animals, plants, insects, and microbes. COEs are mainly attributed to β-esterases, which are essentially irreversibly inhibited by organophosphate insecticides (OPs). CYP enzymes are involved in the detoxification of many toxic and xenobiotic chemicals. Insects showing over expression of one or more P450 genes are resistant to various insecticides. Studies on pest insects and also on resistant *Drosophila* strains have implicated that cytochrome P450 monooxygenase mediated detoxification is a major mechanism of resistance to insecticides. Recent advances in genomics have resulted in the availability of microarray slides containing all 90 *Drosophila* cytochrome P450 genes, facilitating an analysis of this complex enzyme system [7]. Guzov *et al*. [8] reported that cytochrome P450 6D1 (CYP6D1) is responsible for monooxygenase-mediated resistance to different group of insecticides like pyrethroid in the Learn PyR (LPR) strain of house fly. Although there are number of studies but still there is a long way for completely understanding the molecular basis of this resistance. Analysis of codon usage pattern has been used to identify highly expressed genes in many species of microbes. A typical codon usage has been used to infer that genes have been acquired by horizontal transfer [9-11]. Therefore, in this article, we discussed about synonymous codon usage patterns of insecticide detoxification genes (CYPs, GSTs and COEs) in model insect *Drosophila melanogaster*. The expression levels or substrate specificities of CYPs, GSTs and COEs have been linked to many cases of resistance with the responsible enzyme shown to utilize the insecticide as a substrate. This study will help identification and understating of various genes involved in insecticide resistance development pathways of insects/pests.

MATERIALS AND METHODS

The coding sequences of insecticide resistant genes (88 CYPs, 67 GSTs and 58 COEs) of model insect *D. melanogaster* have been downloaded from the NCBI. For minimization of sampling errors, sequence length less than 300bp, redundant data and sequences with intermediate termination codons were excluded for this data analysis.

Codon Usage Indices

Codon usage indices are used for tabulation and investigation of codon usage variation among genes. Some of the important indices available in the literature for reduction of information are, Effective Number of Codons (ENc), G+C content at the third codon position (GC_{3s}), Relative Synonymous Codon Usage (RSCU), Frequency of optimal codons (Fop) and Codon Adaptation Index (CAI). ENc estimates the degree of codon usage bias, but does not provide information about the types of preferred codons. Thus two genes can exhibit same ENc values but may prefer totally different codons [12-15]. ENc values ranges from 20 to 61 depending upon whether one codon or all synonyms in equal frequencies were used per amino acid, respectively [15]. EN_c values for highly expressed genes are normally less than thirty, whereas, for poorly expressed genes is greater than fifty five [16-17]. GC_{3s} estimates codon usage bias towards either G/C or A/T ending codons. RSCU is defined as the ratio of the observed frequency of codons to the expected frequency, if all the synonymous codons for those amino acids are used equally. RSCU values greater than 1.0 indicate that the corresponding codons are used more frequently than the expected frequency, whereas, the reverse is true for RSCU value less than 1.0 [18]. Fop measures the frequency of optimum codons in a gene [19] used the CAI as a universal measure of dominating codon usage bias.

In this study for each gene sequence, GC_{3s}, A_3, T_3, G_3, C_3 (frequency of A, T, G, C at third positions), ENc, RSCU, Fop and CAI were calculated using CodonW 1.4.2 and General Codon Usage Analysis (GCUA) version 1.2 software.

Analysis of Codon Usage

Various statistical techniques are applied on codon usage indices data for further analysis to capture the information of codon usage variation among genes. Correspondence analysis (CA) is a data dimension reduction technique tailored for categorical data [20]. It provides a means of displaying or summarising a set of data in two-dimensional graphical form. In the present study, CA was employed to study the codon usage variation among the genes in which all the genes are plotted in a 59-dimensional hyperspace according to their usage of the 59 informative codons (excluding Met, Trp, and stop codons). A matrix is created in which the rows correspond to the genes in the genome and the columns to the 59 codons, such that each row has the codon usage information for a specific gene. In order to reduce the dimensionality of visualization space of genes for its proper interpretation and drawing valid biological conclusions, CA were applied on RSCU values of 59 codons. The most prominent axes, which contribute maximum to the codon usage variation among the genes were then determined. The major sources of variations in genes have been studied through correlation analysis of these axes. This correlation analysis was used for

explanation of variation and association of gene feature values with axes scores. Furthermore, to investigate the difference between high and low expressed genes, the codon usage variation between 10% of the genes located at the extreme right of major axis and 10% of the genes located towards the extreme left identified by CA using RSCU were compared. Chi-squared contingency test of the two groups were performed to estimate the optimal codons i.e. synonymous codons frequently used in highly expressed genes. Codons whose frequency of usage were significantly higher ($P<0.01$) in highly expressed genes than the genes with low level of expression were determined as optimal codons. Codon W 1.4.2 and MS-Excel were used for carrying out this statistical analysis.

RESULTS AND DISCUSSION

Codon bias: ENc, CAI and GC_{3s} content variation in *Drosophila melanogaster*

The codon usage data of detoxifying genes (CYPs, GSTs and COEs) suggested that the codons ending with G and/or C in all synonymous codon family are predominately used as compared to codons ending with A and/or T. Moreover, it was found that G ending codons are preferred over C ending codons in GST and COE family whereas, *vice versa* with CYP family genes. The codon bias data were first analyzed in terms of ENc value. The ENc value of CYP, GST and COE genes ranged in between 33.49-59.39, 28.94-60.66 and 37.84-59.46 respectively (Fig. 1). This result indicates that some of the genes are highly expressed. Table 1 showed that the average GC3s values for different detoxifying genes in *D. melanogaster* varied from 60.8 to 70.41. There is a good deal of variation of GC_{3s} values rather than CAI and Fop values, which seems to be almost similar among different categories of detoxifying genes. These results suggest that GC3 is a major source of variation in codon usage bias. GST genes have higher CAI and Fop values than the COE and CYP genes. This result indicates that codon biasness is playing a major role among GST genes in comparison to other gene families.

Nc Plots

A plot of ENc versus GC3 (Nc plot) has been widely used to study the codon usage variation among genes in different organisms [21]. Here, actual distribution of genes is compared with the distribution expected under no selection (if codon usage is determined only by GC content). This may indicate that whether the sequences of genes are submitted to some constraints different from just the composition or not. In this study, Nc vs. GC_{3s} plots are used to explore the codon usage variations among various detoxifying genes (Fig. 2A, B, C). The Nc plot of the detoxifying genes suggests that in case of COE genes, maximum points lie on the expected curve with almost towards GC rich region. This result demonstrates that compositional-constraints are playing a role in codon usage bias. Points demonstrating CYP genes are clustered towards GC region ranging between 0.2-0.4 except only one gene lying above the curve. Result suggests that major contributor in codon usage bias is compositional constraints. In case of GST genes maximum points lie below the expected curve towards GC rich region except one or two lying on or above the curve. It is also the consequence of compositional constraints. Points demonstrating CYP, GST and COE genes from *D. melanogaster* in majority lie towards GC rich region.

Differential base usage at third codon position

The correlation of the frequencies of A, T, G and C at the third position against N_c values of all genes were estimated as shown in Table 2. In all the three families of detoxifying genes Nc value is negatively correlated with G_{3s} and C_{3s} and positively correlated with A_3 and T_3. Hence, it can be assumed that the influence of mutational bias of these genes is reflected in the choice of bases at the third position. Due to lack of information on gene expression level of all these gene families so far, we may consider the genes having low Nc values as highly expressed and vice versa [22]. However, strong influence of compositional constraints on codon usages bias in all the gene families analyzed could also be understood from the presence of significant negative correlation between GC_{3s} and Nc ($r=-0.85$, -0.85 and -0.89 at $p<0.001$ respectively with CYP, GST and COE) as shown in Table 2. There seems to be a considerable homogeneity in compositional bias and codon usage patterns in the different detoxifying gene families. It is also studied that C-ending codons are preferred over G-ending codons in highly expressed genes. Preference of C-ending codons in the highly expressed genes might be related to the translational efficiency of the genes as it has been reported that RNY (R-Purine, N-any nucleotide base, and Y-pyrimidine) codons are more advantageous for translation [23]. Thus, compositional mutation bias possibly plays an important role in shaping the genome of these genes.

Correspondence analysis using RSCU values

Correspondence analysis was carried out on RSCU values of all detoxifying genes in data. The position of all CYP, GST and COE genes of the first two major axis shows that almost all genes are clustered. As shown in Fig. 3A, B, C, only the distributions of the genes along the first two major axes were considered as these accounted for maximum variation (32.84 and 30.42% respectively). Among these axes, axis 1 is a major explanatory axis as it accounted the maximum codon usage variation among the different gene families.

The first major axis is positively correlated with A_{3s} and T_{3s} while it is correlated negatively with G_{3s} and C_{3s} (Table 3). Also, strong positive correlation exists between positions of all the genes along the first axis with Nc (Fig. 4A, B, C) while high degree of negative correlation is seen with GC_{3s} (Fig. 5A, B, C). These findings suggest that highly biased genes, those with G and

C ending codons, are clustered on the negative side, whereas the codons ending in A and T predominate on the positive side of the first major axis.

As there is significant negative correlation is observed with Nc against GC_{3s} and GC, thus it shows that highly expressed genes tend to use "C" or "G" at the synonymous position as compared to lowly expressed genes.

Translational optimal codons

RSCU values for each codon for all the three detoxifying gene categories are shown in cluster bar chart in Fig. 6. In case of CYPs, 24 codons were determined using χ^2 test at $P<0.01$ as the 'optimal codons' which were significantly more frequent among the highly expressed genes. Out of 24 codons, there is 17 C-ending viz. UUC, UCC, CCC, CUC, AUC, ACC, GUC, GCC, UAC, UGC, CAC, CGC, AAC, AGC, GAC, GGC, GAG and 7-G ending codons viz. UCG,CUG, CCG, GUG, CAG, CGG, AAG. We determined 18 optimal codons in GST family of CAG of which 13 C-ending codons viz., UUC, UCC, UAC, UGC, CCC, CAC, CGC, AUC, ACC, AAC, GCC, GAC, GGC and 5 G-ending codons viz., GUG, GAG, AAG, CAG, CUG. Similarly, in case of COE family, there are 13 codons found as optimal out of which 7 C-ending are UUC, UAC, CGC, AUC, GCC, GAC, GGC and 6 are as G-ending *viz*., UCG, CUG, CCG, CAG, ACG, GCG. All of these optimal codons have C and/or G at the third position. We found that the 10 codons i.e. UAC, UCG, UGC, GCC, GGC, AUC, CGC, CUG, GAC, and CAG are common in all the different detoxifying gene families. However, AGC, CGG, CUC, GUC is the unique codons for CYP genes and ACG is for COE, whereas GST shares the codon from both the CYP and COE detoxifying gene families.

Gene expressivity level

Codon heterogeneity is usually associated with gene expression level. Since, high codon bias promotes high expression level of genes and thereby assists the cells to fight insecticide/toxin and indicate high resistance in insects. Thus, highly expressed genes contain a higher frequency of codons that are considered translationally optimal. Many studies have also demonstrated a positive correlation between degree of codon bias and level of gene expression [18, 24, 25]. The scores of axis 1 of correspondence analysis on RSCU of codons, which accounted the major codon usage variation among the genes, were classified into three categories on the basis of mean and standard deviation values to indicate the three differential levels of gene expression.

Three different gene expression levels with corresponding cut-off values, viz., high, moderate, and low expressed genes are shown in Fig. 7. The analysis depicted that in all the three detoxifying gene families, GST contains the highest percent of gene under the category of high. Genes of high expression of different detoxifying gene families are listed in Table 4.

It was observed that in case of CYPs maximum genes of high category belongs to CYP4 class and are presented in table in bold format. Bassett *et al*., [26] observed High Cyp18 expression in body wall and gut while negligible expression was observed in salivary glands and body fat. Tijet *et al*. [2] studied that 83 putative CYP genes of *D. melanogaster* are classified either as CYP4 or CYP6 genes, none of which belongs to the CYP1 family. Insect CYP4 class genes are induced by phenobarbital (PB) and alkaloids, are involved in some forms of insecticide resistance [27].Therefore, it has been suggested that the CYP4 enzymes of insects are involved in toxin metabolism, a function that is distinct from its primary role in fatty acid and eicosanoid metabolism in vertebrates [28, 29]. Some of these CYP enzymes are believed to be involved in the metabolism of steroids and xenobiotics [30-35]. The CYP4 family presents a high degree of structural diversity and comprises numerous members among invertebrates, however, for most of these enzymes a well-described function has not yet been assigned [32]. However, the mechanism of over expression of these detoxifying genes in insect is not well-understood. These results will be useful for research on CYP, GST and COEs mediated insecticide resistance in agriculturally important insects. DDT resistance in mosquitoes has been associated with up regulation of glutathione-S-transferases [36], such as GSTE2 [37]. In *Drosophila melanogaster*, ten members of GST genes belonging to Epsilon class were clustered on chromosome 2R division 55C [38].

Table 1. Mean values of GC, GC3, CAI and Fop for three detoxifying gene families of *D. melanogaster*

Gene Family	GC%	GC3%	CAI	Fop
CYPs	50.8±0.05	60.8±0.10	0.23±0.03	0.45±0.04
GSTs	54.56±0.04	70.41±0.10	0.27±0.03	0.47±0.04
COEs	54.2±0.035	62.9±0.08	0.24±0.03	0.46±0.04

Table 2. Correlation coefficient values of A, T, G, C at third position with N_c in detoxifying genes

Gene Families	GC_{3s}	G_{3s}	C_{3s}	T_{3s}	A_{3s}
CYP	-0.85	-0.85	-0.81	0.81	0.89
GST	-0.85	-0.76	-0.84	0.83	0.81
COE	-0.89	-0.60	-0.88	0.79	0.74

Table 3. Correlation analysis of Nc values and A, T, G, C at third position with Axis1

Gene	Nc	C_{3s}	T_{3s}	A_{3s}	G_{3s}	$_{GC3s}$
CYP	0.85	-0.92	0.93	0.96	-0.89	-0.99
GST	0.87	-0.93	0.91	0.85	-0.59	-0.97
COE	0.92	-0.85	0.88	0.92	-0.88	-0.94

Table 4. List of highly expressed genes in different families

CYP	GST	COE
Cyp18a1	GstD1	Gliotactin (Gli)
Cyp4d1	Ef1gamma	Alpha-Esterase-10 (alpha-Est10)
Cyp6a2	Mgstl (microsomal glutathione S-transferase)	CG10175
Cyp4g15	Fax (failed axon connections)	CG31146
Cyp4d1	GstS1	
Cyp6v1	CG1702	
Cyp6g1	CG30185	
Cyp6g2		

Furthermore, in *Anophelesgambiae*, a cluster of eight Epsilon GST genes was located in chromosome 3R division 33B [39]. These observations in dipteran insects suggest that GST genes of the Epsilon class occurred via local duplication that led to independent expansions of the gene class [40].

Considering P450s are important enzymes involved in insecticide resistance, many of these newly identified genes as GSTs and COEs are potential candidates for inhibiting the pathway of insecticide metabolism and for targets of lepidopteran-selective insecticides. Molecular biologists can find some more clues in modifying these highly expressed genes by reading this paper and our work in this paper is only a complementary to the experts' efforts. This can also help to identify the metabolic targets for new synergists to block insecticide resistance and to simple molecular tools to detect insecticide resistant alleles in field populations, an important requirement for effective insecticide resistance management strategies.

CONCLUSION

In the present study, codon usages bias of detoxifying gene of three different families CYP, GST and COE were studied in the model organism Drosophila melanogaster. Comparative analysis of codon usage and RSCU values was calculated. The high level of homogeneity is seen in different gene families of D. melanogaster. Codon usage of these gene families is mainly determined by compositional constraints among the genes of different insects. G/C-ending codons are preferred over A/T-ending codons in highly expressed genes. Optimal codon were also analysed, 10 codons i.e. UAC, UCG, UGC, GCC, GGC, AUC, CGC, CUG, GAC, and CAG are found common in all the different detoxifying gene families. Considering CYP, GST and COE as an important detoxifying gene families involved in insecticide resistance, many of other potential candidates for inhibiting the pathway of insecticide resistance for targets of lepidopteran - selective insecticides can be identified. Molecular biologists can find some more clues in identifying individual detoxifying gene by reading this paper and our work in this paper is only a complementary to the experts' efforts.

ACKNOWLEDGEMENT

We acknowledge the support from National Agricultural Bioinformatics Grid (NABG), National Agricultural Innovation Project (NAIP) and Indian Council for Agricultural Research (ICAR).

REFERENCES

1. Angellotti MC, Bhuiyan SB, Chen G *et al*. (2007). CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res*, 35, 132–136.
2. Tijet N, Helvig C, Feyereisen R. (2001). The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intro-exon organization and phylogeny. *Gene*, 262, 189–198.
3. Feyereisen R. (2005). Insect cytochrome P450. In Comprehensive Molecular Insect Science. Gilbert LI, Iatrou K, Gill SS ed, 1–77. Elsevier Oxford.
4. Clark AG, Shamaan NA. (1984). Evidence that DDT-dehydrochlorinase from the housefly is a glutathione S-transferase. *Pestic Biochem Physiol*, 22, 249-261.
5. Chiang FM, Sun CN. (1993). Glutathione transferase isozymes of diamondback moth larvae and their role in the degradation of some organophosphorus insecticides. *Pestic Biochem Physiol*, 45, 7-14.
6. Vontas JG, Small GJ, Hemingway J. (2001). Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*. *Biochem J*, 357, 65–72.

7. Richard HC, Daborn PJ, Goff GL. (2004). The genetics and genomics of insecticide resistance. Trends in Genet, 20, 163-170.
8. Guzov VM, Unnithan GC, Chernogolov AA, Feyereisen R. (1998). *CYP12A1*, a mitochondrial cytochrome P450 from the house fly. *Archives of Biochemistry and Biophysics*, 359, 231–340.
9. Delorme C, Godon JJ, Ehrlich SD Renault P. (1994). Mosaic structure of large regions of the *Lactococcuslactis* subsp cremoris chromosome. *Microbiology-UK*, 140, 3053-3060.
10. Groisman, E. A., Saier, M. H. J. and Ouchman, H. (1992) Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella*. *EMBO Journal*, 11, 1309-1316.
11. Medigue C, Rouxel T, Vigier P, Henaut A, Danchin A. (1991). Evidence for horizontal gene transfer in Escherichia coli speciation. *J Mol Biol*, 222, 851-856.
12. Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F. (1988). Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomycescerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res*, 16, 8207-8211.
13. Grocock RJ, Sharp PM. (2002). Synonymous codon usage in *Pseudomonas aeruginosa* PAO1. *Gene*, 289, 131–139.
14. Romero H, Zavala A, Mustom H, Bernardi G. (2003). The influence of translational selection on codon usage in fishes from the family Cyprinidae. *Gene*, 317, 141–147.
15. Wright F. (1990). The 'effective number of codons' used in a gene. *Gene*, 87, 23-29.
16. Sharp PM, Cowe E. (1991). Synonymous codon usage in *Saccharomyces cerevisiae*. *Yeast*, 7, 657-78.
17. Gupta SK, Ghosh TC. (2001). Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*. *Gene*, 273, 63-70.
18. Sharp PM, Li WH. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, 15, 1281-1295.
19. Carbone A, Zinovyev A, Kepes F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics*, 19, 2005-2015.
20. Abdi H, Williams LJ. (2010). Correspondence Analysis: In Encyclopedia of Research Design. Salkind NJ, ed. Thousand Oaks: Sage, 243-251.
21. Gupta SK, Bhattacharyya TK, Ghosh TC. (2004). Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J Biomol Structure Dynamics*, 21, 527-536.
22. Hassan S, Kumar V, Mahalingam V. (2009). Synonymous Codon Usage Analysis of Thirty Two *Mycobacteriophage* Genomes. *Advances in Bioinformatics*, 11.
23. Shepherd JC. (1981). Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. *Proc Natl Acad Sci USA*, 78, 1596-1600.
24. Gouy, M. and Gautier, C. (1982) Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res*, 10: 7055–7074.
25. Bennetzen JL, Hall BD. (1982). Codon selection in yeast. *J Biol Chem*, 257, 3026–3031.
26. Bassett MH, McCarthy JL, Waterman MR, Sliter TJ. (1997). Sequence and developmental expression of Cyp18, a member of a new cytochrome P450 family from *Drosophila*. *Mol Cell Endocrinol*, 131, 39-49.
27. Danielson PB, MacIntyre RJ, Fogleman JC. (1997). Molecular cloning of a family of xenobiotic-inducible drosophilid cytochrome P450s: evidence for involvement in host-plant allelochemical resistance. *Proc Natl Acad Sci USA*, 94, 10797–10802.
28. Dunkov B, Guzov VM, Mocelin G, Shotkoski FA, Brun A, Amichot M, Ffrench-Constant R, Feyereinsen R. (1997). The *Drosophila* cytochrome P450 gene Cyp6a2: structure, localization, heterologous expression, and induction by phenobarbital. *DNA Cell Biol*, 16, 1345–1356.
29. Scott JG. (1999). Cytochromes P450 and insecticide resistance. *Insect Biochem Mol Biol*, 29, 757–777.
30. Boyle SM, Popp MP, Smith WC, Greenberg RM, James MO. (1998). Expression of CYP2L1 in the yeast *Pichia pastoris*, and determination of catalytic activity with progesterone and testosterone. *Environ Res*, 46, 25– 28.
31. Brown DJ, Clark GC, Van Beneden RJ. (1998). A new cytochrome P450 (CYP30) family identified in the clam, *Mercenaria mercenaria*. *Comp Biochem Physiol*, 121, 351–360.
32. Dauphin-Villemant C, Bocking D, Tom M, Maibeche M, Lafont R. (1999). Cloning of a novel cytochrome p450 (CYPAC15) differently expressed in the steroidogenic glands of an arthropod. *Biochem Biophys Res Commun*, 264, 413–418.
33. Rewitz K, Styrishave B, Anderson O. (2003). CYP330A1 and CYP4C39 enzymes in the shore crab *Carcinus maenas*: sequence and expression regulation by ecdysteroids and xenobiotics. *Biochem Biophys Res Commun*, 310, 252–260.
34. Snyder MJ. (1998a). Cytochrome P450 enzymes belonging to the CYP4 family from marine invertebrates. *Biochem. Biophys Res Commun*, 249, 187– 190.
35. Snyder MJ. (1998b). Identification of a new cytochrome P450 family, CYP45, from the lobster, *Homarus americanus*, and expression following hormone and xenobiotic exposures. *Arch Biochem Biophys*, 358, 271– 276.

36. Enayati AA, Ranson H, Hemingway J. (2005). Insect glutathione transferases and insecticide resistance. *Insect Mol Biol*, 14, 3-8.
37. Lumjuan N, McCarroll L, Prapanthadara LA, Hemingway J, Ranson H. (2005). Elevated activity of an Epsilon class glutathione transferase confers DDT resistance in the dengue vector, *Aedes aegypti*. *Insect Biochem Mol Biol*, 35, 861-871.
38. Sawicki R, Singh SP, Mondal AK, Benes H, Zimniak P. (2003). Cloning, expression and biochemical characterization of one Epsilon-class (GST-3) and ten Delta-class (GST-1) glutathione S-transferases from *Drosophila melanogaster*, and identification of additional nine members of the Epsilon class. *Biochemical J*, 370, 661-669.
39. Ding Y, Ortelli F, Rossiter LC, Hemingway J, Ranson H. (2003). The *Anopheles gambiae* glutathione transferase supergene family: annotation, phylogeny and expression profiles. *BMC Genomics*, 4, 35.
40. Sonoda S, Ashfaq M, Tsumuki H. (2006). Genomic organization and developmental expression of glutathione S-transferase genes of the diamondback moth, *Plutellaxylostella*. *J Insect Sci*, 6, 1-9.