African Journal of Agricultural Research Vol. 7(21), pp. 3255-3260, 5 June, 2012
Available online at http://www.academicjournals.org/AJAR
DOI: 10.5897/AJAR12.043
ISSN 1991-637X © 2012 Academic Journals

Full Length Research Paper

Analysis of the codon use frequency of peroxisome proliferator-activated receptors (PPAR) family genes from different species

Wenzong Lu*, Guangfeng Jia, Xiangyan Meng, Chen Zhao, Liang Zhang, Haixian Pan, Yumiao Ren and Yuan Ni

Department of Biomedical Engineering, Xi'an Technological University, Xi'an, Shaanxi Province, 710032, People's Republic of China.

Accepted 10 April, 2012

In mammals, peroxisome proliferator-activated receptors (PPAR) play essential roles in the regulation of cellular differentiation, development, and metabolism (carbohydrate, lipid, protein), and tumorigenesis of higher organisms. In this paper, the relative synonymous codon use frequency of 32 PPAR family genes from seven mammal species (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus*, *Macaca mulatta*, *Oryctolagus cuniculus*, *Sus scrofa*) were analyzed by correspondence analysis and hierarchical cluster method. The results indicated that the gene function is the dominant factor that determines codon usage bias in PPAR family genes, while species is a minor factor that determines further difference in codon usage bias for genes with similar functions.

Key words: Peroxisome proliferator-activated receptors, synonymous codon use, correspondence analysis, codon usage bias.

INTRODUCTION

Codon usage bias refers to differences in the frequency of occurrence of synonymous codons in coding DNA. It is generally acknowledged that codon preferences reflect a balance between mutational biases and natural selection for translational optimization (Angellotti et al., 2007). Different factors have been proposed to be related to codon usage bias, including gene expression level (reflecting selection for optimizing translation process by tRNA abundance), GC composition (reflecting horizontal gene transfer or mutational bias), GC skew (reflecting strand-specific mutational bias), amino acid conservation, protein hydropathy, transcriptional selection, RNA stability, optimal growth temperature and hypersaline adaptation (Ermolaeva, 2001; Lynn et al., 2002; Paul et al., 2008). The most obvious factor that determines codon usage is mutational bias that shapes genome GC composition. This factor is most significant in genomes with extreme base composition: species with high GC content use more G and C-ending codons than species with low GC content. Mutational bias is responsible not only for intergenetic difference in codon usage but also for codon usage bias within the same genome (Ermolaeva, 2001). On the contrary, some reports showed that translational selection served as the most important role in shaping codon usage in the genome (Horn, 2008; Iida and Akashi, 2000).

The peroxisome proliferator-activated receptors (PPAR) are a group of nuclear receptor proteins that function as transcription factors regulating the expression of genes (Michalik et al., 2006). PPAR plays essential roles in the regulation of cellular differentiation, development, and metabolism (carbohydrate, lipid, protein), and tumorigenesis of higher organisms (Belfiore et al., 2009; Berger and Moller, 2002; Feige et al., 2006). PPAR not only plays a crucial role in intracellular lipid metabolism, but is also an important regulator of extracellular lipid metabolism (Schoonjans et al., 1996). PPAR has three

*Corresponding author. E-mail: wenzonglu@126.com. Tel: +86-29-86173097.

Abbreviations: PPAR, Peroxisome proliferator-activated receptors; **RSCU,** relative synonymous codon usage; **ENC,** effective number of codons; **COA,** correspondence analysis.

subtypes (PPAR-α, β/δ, and γ) showing different expression patterns in vertebrates. PPAR-α regulates fatty acid homeostasis via transcriptional activation of genes encoding key enzymes in fatty acid metabolism. PPAR-β/δ is almost ubiquitously expressed and transcriptionally regulates fatty acid oxidation. PPAR-γ promotes adipocyte differentiation and thought to play a vital role in regulating fat storage, and this PPAR through alternative splicing is expressed in three forms (γ-1, γ-2, γ-3) (Finck, 2007).

Although DNA sequence of PPAR family has been published and many studies have reported it in different species in recent years (Ershov and Bazan, 2000; Son et al., 2010; Winegar et al., 2001; Zimin et al., 2009), no in depth analyses have so far been made on codon usage, which may provide more information on the features of PPAR family genes. In this paper, codon usage of 32 PPAR family genes from *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus*, *Macaca mulatta*, *Oryctolagus cuniculus*, *Sus scrofa* were analyzed using correspondence analysis (COA) and hierarchical cluster method and our attention is focused on genes with similar functions while from different species.

MATERIALS AND METHODS

Data sets

PPAR family genes in *Sus scrofa* used in this paper were from Genbank (http://www.ncbi.nlm.nih.gov). Orthologous genes from *Sus scrofa* were considered as predicted orthologous genes from *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Macaca mulatta*, *Oryctolagus cuniculus*, *Bos taurus* by using Inparanoid to study whether gene function determines codon usage bias in PPAR family genes. The genes are greater than 500 bp to minimize the sampling errors. We have eliminated the genes, which have internal termination codons. 32 genes were selected for analysis and number, access number, definition and protein ID of each gene were showed in Table 1.

Statistical analysis

Comparison of the variation of codon usage between different genes was done by one tailed t-test. Cluster analysis was performed using Hierarchical cluster method and the distances between selected sequences were computed by the Euclidean distance method. It can be calculated as:

$$d_{ik}=\sqrt{\sum_{j=1}^{59}(\mathrm{RSCU}_{ij}-\mathrm{RSCU}_{kj})^2}$$

Where $RSCU_{ij}$ is the relative synonymous codon-use frequency of codon *j* in sequence *i*. Similarly, $RSCU_{kj}$ is the relative synonymous codon-use frequency of codon *j* in sequence *k*.

Measures of synonymous codon usage bias

Most amino acids can be coded by more than one triplet of nucleotides (codon). Relative synonymous codon usage (RSCU) is defined as the ratio of the observed frequency of codons to the expected frequency given that all the synonymous codons for the same amino acids are used equally. This codon usage measure corresponds to the ratio between the observed numbers of a codon over its expected value under the hypothesis of a random distribution of all the synonymous codons encoding for a given amino acid. RSCU values have no connection with the amino acids usage and the abundance ratio of synonymous codons, which can directly reflect the bias of synonymous codon usage (Sharp et al., 1986). Effective number of codons (ENC) is a measure to study the state of codon usage biases in genes and genomes. The way that ENC is computed has obvious similarities to the computation of effective population size in population genetics. When ENC value approaches 20, only one codon is used with extreme bias for one amino acid and, if the value is up to 61, the anonymous codons are used equally with no bias (Wright, 1990).

Correspondence analysis (COA)

Correspondence analysis is an ordination technique that identifies the major trends in the variation of the data and distributes genes along continuous axes in accordance with these trends. It is conceptually similar to principal component analysis, but applies to categorical rather than continuous data. COA was performed by the values of RSCU in each gene, and was plotted in a 59-dimensional hyperspace according to their usage of the 59 sense codons (excluding Met, Trp, and termination codons). Major variation trends can be determined using these RSCU values and genes ordered according to their positions along the major axis, which can also be used to distinguish the major factors influencing the codon usage of a gene. Generally, the major trend influences codon usage variation among genes and occurs when the variability is more 10% (Greenacre, 1984).

Analysis tools

The RSCU, ENC and COA were calculated using the online analysis tools (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=codonw). Cluster analysis was carried out by using the multi analysis software SPSS version 13.0.

RESULTS AND DISCUSSION

To study the codon usage variation among different PPAR family genes, ENC values of different PPAR family genes were calculated (Figure 2). ENC values of different PPAR family genes vary from 48.81 to 54.42 with a mean value of 52.81 and the SD values varying from 1.22 to 3.92. All the ENC values of these genes are more than 45. The ENC was often used to measure the magnitude of codon bias for an individual gene, yielding values ranging from 20, for a gene with extreme bias using only one codon per amino acid, to 61 for a gene with no bias using synonymous codons equally (Wright, 1990). The data suggests the homogeneity of synonymous codon usage among PPAR family genes examined. Our results showed that the GC contend of different PPAR family genes change to 60.74 from 46.93, and one interesting thing were obviously found; the ENC value of PPAR-β/δ was the least in the genes but the GC contend of PPAR-β/δ was the richest (Figure 2). One research on coding sequences of RNA viruses and their genome polarity

Table 1. PPAR family genes under our research.

S/N	Access no.	Protein ID	Definition
1	NM_001127330	NP_001120802.1	*Mus musculus* peroxisome proliferator activated receptor gamma (Pparg), transcript variant 1, mRNA.
2	MMU09138	AAA62277.1	*Mus musculus* peroxisome proliferator activated protein-gamma-2, mRNA, complete cds.
3	EF062476	ABK39948.1	*Mus musculus* peroxisome proliferator activated receptor gamma 2 (Pparg) mRNA, complete cds.
4	NM_011145	NP_035275.1	*Mus musculus* peroxisome proliferator activator receptor delta (Ppard), mRNA.
5	NM_011144	NP_035274.2	*Mus musculus* peroxisome proliferator activated receptor alpha (Ppara), transcript variant 1, mRNA.
6	NM_138711	NP_619725.2	*Homo sapiens* peroxisome proliferator-activated receptor gamma (PPARG), transcript variant 3, mRNA.
7	NM_015869	NP_056953.2	*Homo sapiens* peroxisome proliferator-activated receptor gamma (PPARG), transcript variant 2, mRNA.
8	NM_006238	NP_006229.1	*Homo sapiens* peroxisome proliferator-activated receptor delta (PPARD), transcript variant 1, mRNA.
9	HSU63415	AAB04028.1	Human peroxisome proliferator activated receptor gamma 2 mRNA, complete cds.
10	AB307690	BAH02281.1	*Homo sapiens* NR1C1 mRNA for peroxisome proliferator activated receptor alpha, complete cds.
11	AY179866	AAN75018.1	*Bos taurus* peroxisome proliferator activated receptor gamma-2 mRNA, complete cds.
12	Y12419	CAA73032.1	*B. taurus* mRNA for peroxisome proliferator activated receptor gamma 1.
13	NM_001034036	NP_001029208.1	*Bos taurus* peroxisome proliferator-activated receptor alpha (PPARA), mRNA.
14	BC134636	AAI34637.1	*Bos taurus* peroxisome proliferator-activated receptor delta, mRNA (cDNA clone MGC:151611 IMAGE:8315367), complete cds.
15	AY188501	AAO34393.1	*Sus scrofa* peroxisome proliferator-activated receptor delta (PPARd) mRNA, complete cds.
16	DQ437884	ABE01102.1	*Sus scrofa* peroxisome proliferator-activated receptor gamma 1 mRNA, complete cds.
17	AF059245	AAC14348.1	*Sus scrofa* peroxisome proliferator-activated receptor gamma 2 (PPARG) mRNA, complete cds.
18	AF103946	AAD19577.1	*Sus scrofa* peroxisome proliferator activated receptor gamma 2 (PPAR gamma 2) mRNA, complete cds.
19	NM_001044526	NP_001037991.1	*Sus scrofa* peroxisome proliferator-activated receptor alpha (PPARA), mRNA.
20	AJ006756	CAA07224.1	*Sus scrofa* mRNA for peroxisome proliferator-activated receptor gamma 1.
21	NM_013141	NP_037273.2	*Rattus norvegicus* peroxisome proliferator-activated receptor delta (Ppard), mRNA.
22	NM_013196	NP_037328.1	*Rattus norvegicus* peroxisome proliferator activated receptor alpha (Ppara), mRNA.
23	AF156666	AAD40119.1	*Rattus norvegicus* peroxisome proliferator-activated receptor gamma 2 (PPARgamma2) mRNA, complete cds.
24	AF156665	AAD40118.1	*Rattus norvegicus* peroxisome proliferator-activated receptor gamma 1 (PPARgamma1) mRNA, complete cds.
25	Y12882	CAA73382.2	*Rattus norvegicus* mRNA for peroxisome proliferator activated receptor gamma 2.
26	AY048696	AAL05263.1	*Macaca fascicularis* peroxisome proliferator-activated receptor gamma 3 (PPARgamma3) mRNA, complete cds.
27	AY048694	AAL05261.1	*Macaca fascicularis* peroxisome proliferator-activated receptor gamma 1 (PPARgamma1) mRNA, complete cds.
28	DQ062812	AAY64435.1	*Macaca mulatta* peroxisome proliferator activated receptor (PPARA) mRNA, complete cds.
29	AF033103	AAB87480.1	*Macaca mulatta* peroxisome proliferator-activated receptor gamma 2 (PPARg2) mRNA, complete cds.
30	OCU84893	AAB96380.1	*Oryctolagus cuniculus* peroxisome proliferator activated receptor gamma-1 mRNA, complete cds.
31	XM_002723354	XP_002723400.1	PREDICTED: *Oryctolagus cuniculus* peroxisome proliferative activated receptor, alpha (LOC100356422), mRNA.
32	NM_001082148	NP_001075617.1	*Oryctolagus cuniculus* PPAR gamma 3 (LOC100008892), mRNA.

showed that positive-stranded RNA viruses have significantly higher GC contents than negative-stranded RNA viruses. Coding sequences of all negative-stranded RNA viruses are biased toward high A in coding strands (high T in genomes) (Auewarakul, 2005). A research revealed that

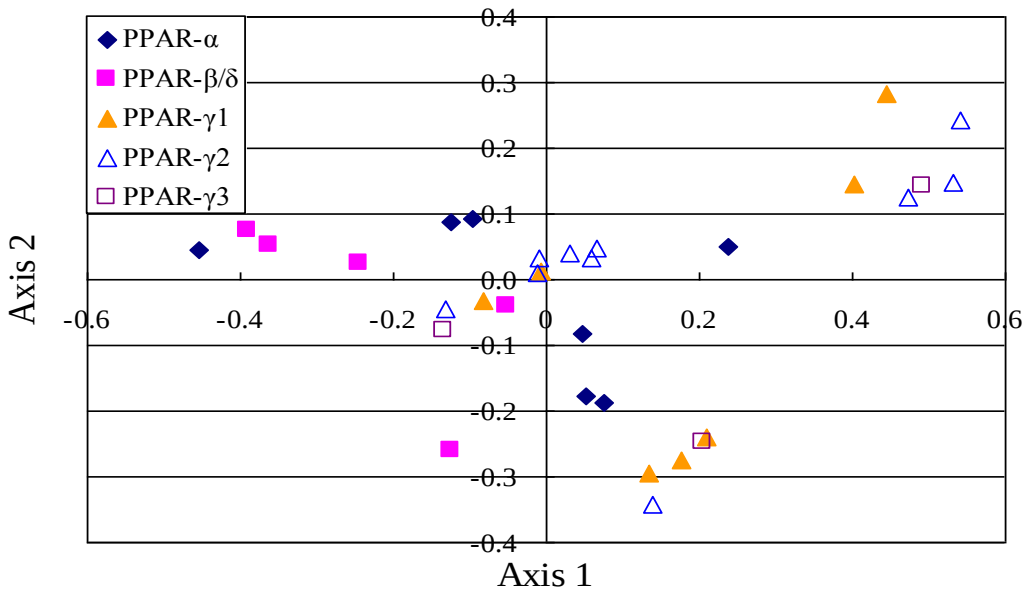

Figure 1. A plot of the values of the first axis and the second axis of each gene under study in COA.

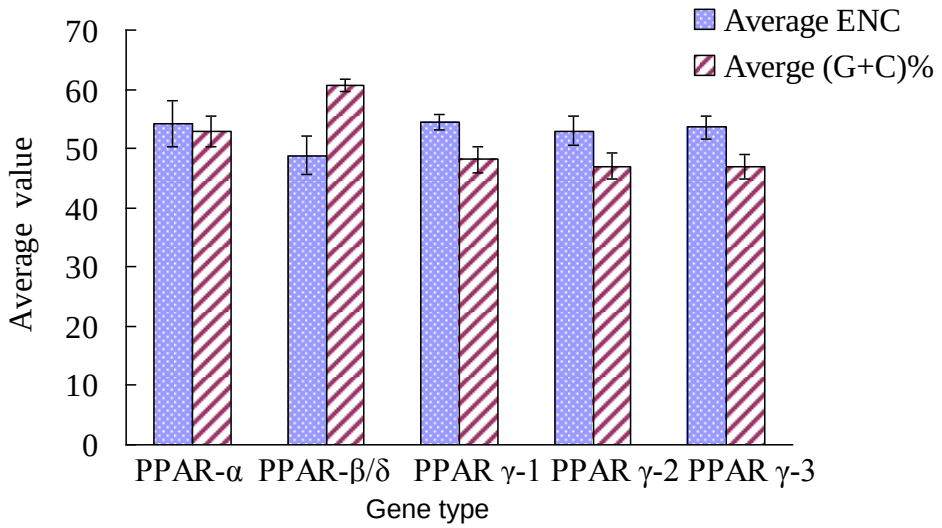

Figure 2. Average ENC value and GC content of each group according to the gene type among different species in study.

patterns of codon usage bias are strongly correlated with overall genomic GC content, suggesting that genome-wide mutational pressure, rather than natural selection for specific coding triplets, is the main determinant of codon usage (Shackelton et al., 2006). But in *Chlamydomonas reinhardtii* genome which had high GC contents, there was no evidence that the genome composition shaped the codon usages of genes (Naya et al., 2001). Correspondence analysis was implemented on these 32 PPAR family genes examined as a single dataset based on the RSCU value of each gene to investigate the variation of RSCU values among the genes. The axis of a correspondence analysis identifies the source of the variation among a set of multivariate data point. The four largest trends in codon usage among these genes were observed: the first axis accounts for 23.09% of all variation among genomes, whereas the next three axes accounts for 10.12, 3.87 and 0.14%, respectively. A plot of the first axis and the second axis of each gene has been shown in Figure 1. It is clear in Figure 1 that the functionally homologous genes in different species tend to have close value of the first axis and second axis in COA. Because the closeness of any two genes on this value reflects the similarities of their codon usages,

Table 2. Clustering result of PPAR family genes under study.

Lineage number	Number
I	11, 13, 15, 16, 19, 21, 25, 26, 32
II	2, 3, 4, 9, 12, 18, 22, 23, 28, 30
III	8, 10, 14, 20, 27, 31
IV	1, 5, 6, 7, 17, 24, 27

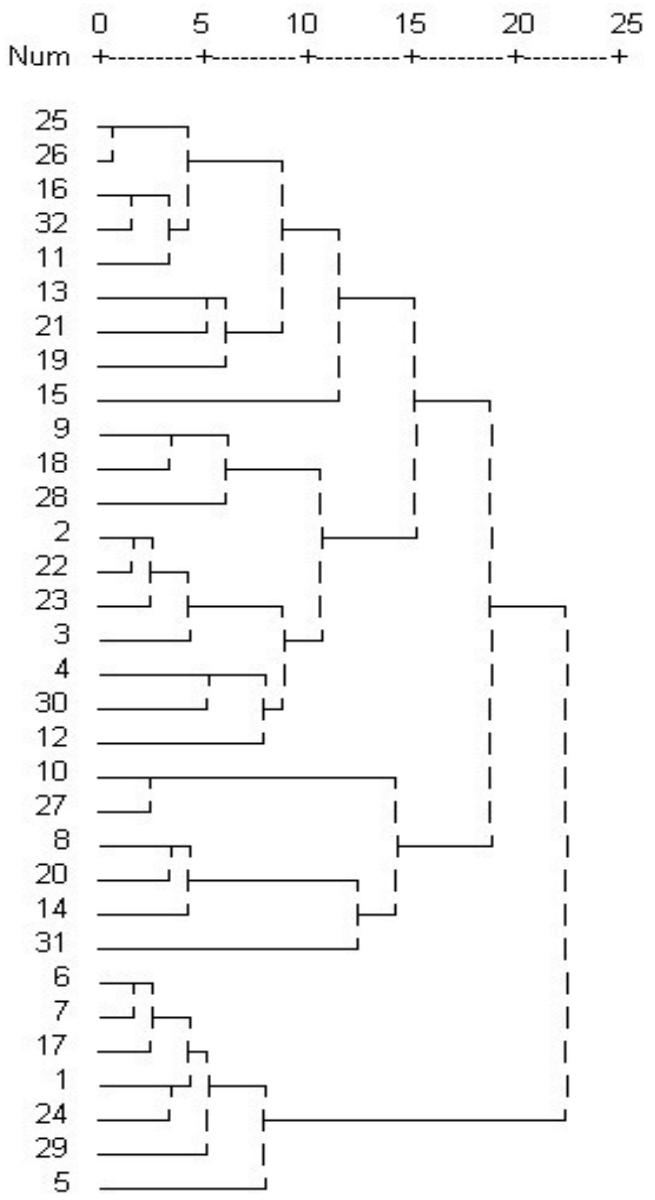

Figure 3. Dendroid chart of the cluster result of 32 PPAR family genes in study based on hierarchical cluster method.

synonymous codon usage bias appears to be conservative between genes that are functionally closely related. All genes were divided into several classes by gene function to display whether gene function was correlated with the found variation in codon bias. Five genes of PPAR family (PPAR-α, β/δ, γ-1, γ-2, and γ-3) were selected to study whether there is a correlation between gene function and codon usage. The value of ENC was computed in each group, and the SD values were small (Figure 2). One tailed t-test was then performed on ENC values and values of both axes 1 and axes 2 in COA of these genes with the hypothesis that there is no correlation between codon usage bias and gene function ($P<10^{-2}$). It suggests that the gene's function contributes to the codon usage in PPAR family genes.

Based on the RSUC variation of these 32 PPAR family genes examined, a cluster tree was generated by using Hierarchical cluster method. The cluster result of these genes is listed in Table 2 and the dendroid chart obtained using SPSS is shown in Figure 3. From Table 2 and Figure 3, it could be found that all 32 genes in Table 1 are classified into four main lineages (I, II, III and IV). All PPAR family genes from these species examined were grouped into lineage I. Lineage II was composed of all PPAR family genes except PPAR γ-3 from these species examined, and the percentage of PPAR γ-2 was 50% in lineage II genes. Lineage III included six genes of PPAR-α, PPAR-β/δ and PPAR γ-1. Two PPAR-α, and five PPAR-γ genes were classified into lineages IV. Distances between sample sequences are standardized in Figure 3. It could be found from Figure 3 that PPAR family genes form a few sub-lineages in lineage I, lineage II and lineage III. The genes of these sub-lineages were from different species (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Macaca mulatta*, *Oryctolagus cuniculus*, *Bos Taurus*). From the above, we could see that genes with same function yet from different species are classified into the same lineage or same sub-lineage. In Figure 3, for a certain lineage or sub-lineage, distances between genes from the close species are closer than distances between distant species. It could be concluded that the codon usage bias of PPAR family genes is related mainly to gene functions and species is also a minor factor that can affect codon usage. For instance, we could see five PPAR γ-2 genes mainly from the lineage II in which the degree of codon usage bias of species with close consanguinity is closer than that of species with distant consanguinity in Figure 1. In general, it showed that PPAR-α protein originates more early than the others and PPAR-γ protein originates later in different mammalian species (Table 1, Figure 3). From above, it could be inferred that the codon usage bias of PPAR family genes is determined mainly by gene functions; in the precondition of similar gene functions, species is also a factor that may cause codon usage bias difference.

Conclusions

According to the present results, we could conclude that gene type and function is the main factor that is closely

related to synonymous codon usage bias in PPAR family genes.

ACKNOWLEDGMENT

This work was supported by the fund of Xi'an Technological University (XAGDXJJ1014).

REFERENCES

Angellotti MC, Bhuiyan SB, Chen G, Wan XF (2007). CodonO: codon usage bias analysis within and across genomes. Nucleic Acids Res., 35(Web Server issue): W132-136.

Auewarakul P (2005). Composition bias and genome polarity of RNA viruses. Virus Res., 109(1): 33-37.

Belfiore A, Genua M, Malaguarnera R (2009). PPAR-gamma agonists and their effects on IGF-I receptor signaling: implications for cancer. PPAR Res., 2009: 830501.

Berger J, Moller DE (2002). The mechanisms of action of PPARs. Annu. Rev. Med., 53: 409-435.

Ermolaeva MD (2001). Synonymous codon usage in bacteria. Curr. Issues Mol. Biol., 3(4): 91-97.

Ershov AV, Bazan NG (2000). Photoreceptor phagocytosis selectively activates PPAR gamma expression in retinal pigment epithelial cells. J. Neurosci. Res., 60(3): 328-337.

Feige JN, Gelman L, Michalik L, Desvergne B, Wahli W (2006). From molecular action to physiological outputs: peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. Prog. Lipid Res., 45(2): 120-159.

Finck BN (2007). The PPAR regulatory system in cardiac physiology and disease. Cardiovasc Res., 73(2): 269-277.

Greenacre MJ (1984). Theory and applications of correspondence analysis. Academic Press, London, pp. 19-20

Horn D (2008). Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. BMC Genomics, 9: 2.

Iida K, Akashi H (2000). A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. Gene, 261(1): 93-105.

Lynn DJ, Singer GA, Hickey DA (2002). Synonymous codon usage is subject to selection in thermophilic bacteria. Nucleic Acids Res., 30(19): 4272-4277.

Michalik L, Auwerx J, Berger JP, Chatterjee VK, Glass CK, Gonzalez FJ, Grimaldi PA, Kadowaki T, Lazar MA, O'Rahilly S, Palmer CN, Plutzky J, Reddy JK, Spiegelman BM, Staels B, Wahli W (2006). International Union of Pharmacology. LXI. Peroxisome proliferator-activated receptors. Pharmacol. Rev., 58(4): 726-741.

Naya H, Romero H, Carels N, Zavala A, Musto H (2001). Translational selection shapes codon usage in the GC-rich genome of Chlamydomonas reinhardtii. FEBS Lett., 501(2-3): 127-130.

Paul S, Bag SK, Das S, Harvill ET, Dutta C (2008). Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. Genome Boil., 9(4): R70.

Schoonjans K, Staels B, Auwerx J (1996). Role of the peroxisome proliferator-activated receptor (PPAR) in mediating the effects of fibrates and fatty acids on gene expression. J. lipid Res., 37(5): 907-925.

Shackelton LA, Parrish CR, Holmes EC (2006). Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol., 62(5): 551-563.

Sharp PM, Tuohy TM, Mosurski KR (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res., 14(13): 5125-5143.

Son NH, Yu S, Tuinei J, Arai K, Hamai H, Homma S, Shulman GI, Abel ED, Goldberg IJ (2010). PPARgamma-induced cardiolipotoxicity in mice is ameliorated by PPARalpha deficiency despite increases in fatty acid oxidation. J. Clin Invest., 120(10): 3443-3454.

Winegar DA, Brown PJ, Wilkison WO, Lewis MC, Ott RJ, Tong WQ, Brown HR, Lehmann JM, Kliewer SA, Plunket KD, Way JM, Bodkin NL, Hansen BC (2001). Effects of fenofibrate on lipid parameters in obese rhesus monkeys. J. Lipid Res., 42(10): 1543-1551.

Wright F (1990). The 'effective number of codons' used in a gene. Gene, 87(1): 23-29.

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marcais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Boil., 10(4): R42.