

Analysis of Codon Usage Bias in Interferon Alpha Gene of the Giant Panda (*Ailuropoda Melanoleuca*)

Yue Yi¹, Chengdong Wang^{1,2}, Zhiwen Xu¹, Desheng Li^{1,2}, Ling Zhu¹, Fengqin Li¹, Xinqiao Qi¹, and Wanzhu Guo¹

¹ Animal Biotechnology Center
College of Veterinary Medicine of Sichuan Agricultural University, Ya'an, Sichuan,
625014, China

² China Conservation and Research Center for the Grint Panda (CCRCGP), Wolong Nature Reserve, Sichuan, 623006, China

Abstract. The analysis on codon usage bias of IFN- α gene of giant panda (*Ailuropoda melanoleuca*) may provide a basis for understanding the evolution relationship of giant panda and for selecting appropriate host expression systems to improve the expression of target genes. In this paper, the codon usage bias in the mature IFN- α sequence of giant panda and 15 reference species have been analyzed. The results showed that the synonymous codons with G and C at the third codon position were widely used and the ENC-GC3S plot revealed that the genetic heterogeneity in IFN- α gene was main constrained by mutational bias. Contrastive analysis revealed that there were 40 codons showing distinct usage differences between GpIFN- α and *Escherichia coli*, 38 codons between GpIFN- α and yeast. and only 30 between GpIFN- α and *Homo sapiens*. Therefore the *Homo* expression system may be more suitable for the expression of GpIFN- α genes.

Keywords: *Ailuropoda melanoleuca*, interferon alpha, codon usage bias, analysis.

1 Introduction

The genetic code uses 61 nucleotide triplets (codons) to encode 20 amino acids and three to terminate translation[1]. So the genetic code is necessarily redundant which are more than amino acids. Each amino acid is encoded by between one (Met and Trp) and six (Arg, Leu and Ser) synonymous codons[2]. Different codons that encode the same amino acid are known as synonymous codons. Changes in the DNA sequence of a protein between two synonymous codons are often assumed to have no effect and are thus called synonymous changes or even silent changes[2]. In a wide variety of organisms such as bacteria, fungi, and insects (Sharp et al. 1992), codons within synonymous groups are used with different frequencies[3-6] the phenomenon called codon usage bias.

There is continuing speculation regarding the reasons what have made these differences in codon preferences[7]. the most commonly accepted hypothesis for the unequal usage of synonymous codons stated that it was the result of mutational biases

and natural selection acting at the level of translation. However, several other causes have also been reported, such as GC compositions, gene length, gene expression level, gene translation initiation signal, protein amino acid composition, protein structure, mutation frequency and patterns, and tRNA abundance [8-12]. In some other researches, it was also found that codon usage was related to gene function[13, 14] and cellular location of gene products [15-17]. Whatever the reasons for codon bias, it has become increasingly clear that codon biases can have profound impacts on the expression of heterologous proteins.

In this study, characterization of codon usage bias on interferon alpha gene without signal peptide of the giant panda had been evaluated. Giant panda is a much loved animal all over the world and is considered a symbol of China. However, it is also one of the world's most endangered species. In addition, giant pandas are considered to be particularly vulnerable to infectious disease and parasites[18]. Therefore, analysis and study on cytokine genes of the giant panda are important to facilitate the use of these cytokines in the immunotherapy of infectious diseases of the giant panda[19].

The interferons are a group of natural antiviral substances first discovered in 1957[20]. Many studies have performed on the function of IFN- α . It influences many biological processes including broad spectrum antiviral effects, inhibition of tumor cell proliferation and enhancement of immune functions. And it is considered to be a potential agents for therapy of neoplastic diseases and cancers[21].

So far, research on the giant panda genome is still rare. only a few genes with immunity function have been cloned. Recently, the interferon alpha gene of giant panda had been cloned in our laboratory called GpIFN- α . Here, we first time analyzed the codon usage data of GpIFN- α and those of other 15 species' IFN- α . Moreover, the codon usage bias in the GpIFN- α gene was also compared with *Escherichia coli*, yeast, and *Homo sapiens*. In this research, codon usage data of GpIFN- α and the comparison results might give some clues to the features of GpIFN- α gene and some evolutionary information of IFN- α gene. And also provide insights into the expression and possible function of GpIFN- α .

2 Materials and Methods

2.1 Sequence Data

The mature sequence of GpIFN- α which cloned in our laboratory had been submitted to GenBank and the accession number is HQ378189. The overall IFN- α reference sequences were download from NCBI (<http://www.ncbi.nlm.nih.gov/>). The signal peptide of every reference sequence had been removed to kept the statistical significance of codon usage bias. The informations such as GenBank accession No., Natural host and Length of the IFN- α gene of overall 16 reference sequence are shown in Table 1.

2.2 Analysis on Codon Usage in GpIFN- α Genes and 15 Reference Sequences

Codon bias was calculated according to methods described elsewhere. Relative synonymous codon usage (RSCU) is the observed frequency of a codon divided by the frequency expected if all synonyms for that amino acid were used equally. It is

Table 1. The Information about IFN-A Gene Used in Our Research

Species	GenBank Accession No.	Animals' IFN- α	Natural animal	Length (bp)
Mammal	AY323972	Bubalus bubalis IFN- α	Bubalus bubalis	498
	DQ520882	Canis familiaris IFN- α	Canis familiaris	495
	FJ959074	Capra hircus IFN- α	Capra hircus	501
	NM_001099441	Equus caballus IFN- α	Equus caballus	486
	NM_001009851	Felis catus IFN- α	Felis catus	516
	NM_001194384	Macaca mulatta IFN- α	Macaca mulatta	501
	EU863618	Mustela putorius furo IFN- α	Mustela putorius furo	495
	HM636502	Pteropus vampyrus IFN- α	Pteropus vampyrus	540
	AY331298	Sus scrofa IFN- α	Sus scrofa	501
	EF990625	Vulpes vulpes IFN- α	Vulpes vulpes	540
	BC074928	Homo sapiens IFN- α	Homo sapiens	501
	HQ378189	grint panda IFN- α	grint panda	495
Birds	EU029159	Anser anser rubrirostris IFN- α	Anser anser rubrirostris	492
	X84764	duck IFN- α	duck	492
	AB021154	Gallus gallus IFN- α	Gallus gallus	492
	EU022750	Shitou Goose IFN- α	Shitou Goose	492

independent of amino acid composition and is thus useful for comparing different sets of genes[22]. The effective number of codons (ENC) is a measure analogous to the effective number of alleles. Based on simulation studies, ENC appears to be low sensitive to gene length[23]. The ENC value of the IFN-a gene of each reference was computed with the EMBOSS CHIPS online service program. The degree of synonymous-codon usage bias was also measured by the codon adaptation index (CAI) which estimates the extent of bias toward codons that are known to be favored in highly expressed genes. which was calculated with the EMBOSS CAI program [24]. The peculiarity in codon usage frequency ,the G+C content of the gene sequences and the G+C content at the synonymous third position of codons (GC3S) were also calculated with the EMBOSS CUSP program[25].

2.3 Phylogenetic Analysis of the GpIFN-a Gene

The mature nucleotide sequences of the GpIFN-a gene and 15 reference IFN-a were translated into amino acid sequences by using DNASTAR software. Then, multiple sequence alignment and phylogenetic analysis by the bootstrap option were performed by using ClustalW and TREEVIEW software.

2.4 Comparison of Codon Preferences of GpIFN-a Gene with Those of E. coli, Yeast, Adenovirus, and Human

A comparison analysis was performed between the GpIFN-a, Escherichia coli, yeast, and Homo sapiens. The database of the codon usage in E. coli, yeast, and Homo sapiens was available at (<http://www.kazusa.or.jp/codon/>).

3 Results

3.1 Characterization of Synonymous Codon Usage in the IFN-a Gene of Giant Panda

While the relative synonymous codon usage values (RSCU) and the related measures indicate the overall GpIFN-a codon bias, it is also important to look more closely at the patterns of codon bias. Table 2 shows the details of the codon usage patterns and codon usage data of giant panda IFN- α gene (excluding Met, Trp, and the termination codons). A high level of diversity in codon usage bias existed for coding the Ala, Leu and Ser amino acids (emphasized with underline). From the table 2 we can see that the frequencies of codons usage in coding the same amino acids are different. There are 26 codons, such as CTG(Leu), AGG (Arg), ATC(Ile), CCC(Pro), TGG(Trp), TAC(Tyr) and so on, were more frequently used. Among the preferred codons nineteen synonymous codons strong bias toward C-ended and nine toward G-ended at the third codon position. It is evident that codon usage bias in the mature polypeptide is strong bias towards the synonymous codons with G and C in the wobble.

Table 2. Codon Usage Data of Grint Panda Interferon Alpha Gene Analyzed by Cusps Program

Amino acid	codon ^a	Fract ^b	Frequency 1000 ^c	Number ^d	RSCU ^e
<u>A(Ala)</u>	GCA	0.077	6.061	1	0.30
A	GCC ^f	0.462	36.364	6	1.84
A	GCG	0.154	12.121	2	0.61
A	GCT	0.308	24.242	4	1.23
C(Cys)	TGC	0.500	18.182	3	1.00
C	TGT	0.500	18.182	3	1.00
D(Asp)	GAC ^f	0.833	30.303	5	1.67
D	GAT	0.167	6.061	1	0.33
E(Glu)	GAA	0.333	24.242	4	0.67
E	GAG ^f	0.667	48.485	8	1.33
F(Phe)	TTC ^f	0.625	30.303	5	1.25
F	TTT	0.375	18.182	3	0.75
G(Gly)	GGA	0.571	24.242	4	1.14

Table 2. (continued)

G	GGC ^f	0.286	12.121	2	1.14
G	GGG	0.143	6.061	1	0.57
G	GGT	0.000	0.000	0	0.00
H(His)	CAC	0.250	6.061	1	0.50
H	CAT	0.750	18.182	3	1.50
I(Ile)	ATA	0.000	0.000	0	0.00
I	ATC ^f	1.000	30.303	5	2.00
I	ATT	0.000	0.000	0	0.00
K(Lys)	AAA	0.143	6.061	1	0.29
K	AAG ^f	0.857	36.364	6	1.71
<u>L(Leu)</u>	CTA	0.111	12.121	2	0.67
L	CTC ^f	0.222	24.242	4	1.33
L	CTG ^f	0.389	42.424	7	2.33
L	CTT	0.056	6.061	1	0.33
L	TTA	0.000	0.000	0	0.00
L	ATG ^f	0.222	24.242	4	1.33
N(Asn)	AAC ^f	0.750	18.182	3	1.50
N	AAT	0.250	6.061	1	0.50
P(Pro)	CCC ^f	0.667	24.242	4	2.67
P	CCG	0.000	0.000	0	0.00
P	CCA	0.000	0.000	0	0.00
P	CCT	0.333	12.121	2	1.33
Q(Gln)	CAA	0.500	36.364	6	1.00
Q	CAG	0.500	36.364	6	1.00
R(Arg)	AGA	0.444	24.242	4	2.67
R	AGG ^f	0.556	30.303	5	3.33
R	CGA	0.000	0.000	0	0.00
R	CGC	0.000	0.000	0	0.00
R	CGG	0.000	0.000	0	0.00
R	CGT	0.000	0.000	0	0.00
<u>S(Ser)</u>	AGC ^f	0.294	30.303	5	1.76
S	AGT	0.000	0.000	0	0.00
S	TCA	0.176	18.182	3	1.06

Table 2. (continued)

S	TCG	0.118	12.121	2	0.71
S	TCT	0.235	24.242	4	1.41
T(Thr)	ACA	0.250	12.121	2	1.00
T	ACC	0.125	6.061	1	0.50
T	ACGf	0.375	18.182	3	1.50
T	ACT	0.250	12.121	2	1.00
V(Val)	GTA	0.000	0.000	0	0.00
V	GTCf	0.455	30.303	5	1.82
V	GTGf	0.455	30.303	5	1.82
V	GTT	0.091	6.061	1	0.36
Y(Tyr)	TACf	1.000	24.242	4	2.00
Y	TAT	0.000	0.000	0	0.00

^aThe preferentially used codons for each amino acid are displayed in bold; ^bThe ‘‘Fract’’ column shows the proportion of usage of a given codon in its redundant set (i.e., the set of codons that code for the same amino acid); ^cThe ‘‘Frequency’’ column lists the number of codons present per 1000 bases in the input sequence(s); ^dThe ‘‘Number’’ column lists the number of codons; ^eThe ‘‘RSCU’’ column shows the proportion of relative synonymous codon usage; ^fShows a strong bias towards the codons with A and T at the third codon position.

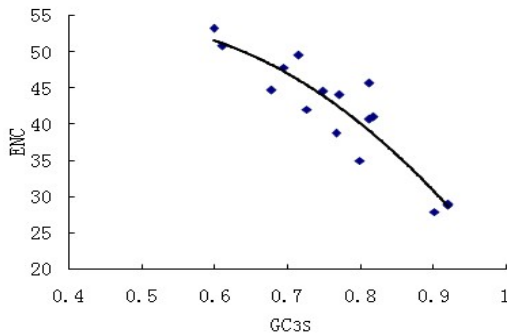
3.2 Codon Usage Analysis of the IFN- α Gene of Grint Panda and Reference Species

The value of CAI, ENC and GC and GC3S content analysis of grint panda IFN- α and other kinds of IFN- α genes were obtained by EMBOSS. The results are shown in Table 3. the ENC values of different IFN- α genes varied from 27.971 to 53.276, with a mean value of 41.50 and standard deviation (S.D.) of 7.86. The GC3S contents of each IFN- α gene ranged from 36.78 to 92.07% with a mean of 58.66% and S.D. of 0.097. The CAI values of different IFN- α genes varied from 0.759 to 0.85 with a mean value of 0.808 and standard deviation (S.D.) of 0.022.

It has been reported that a plot of ENC against GC_{3S} can be effectively used to explore the heterogeneity of codon usage among genes [23]. Figure 1 shows the distribution plot of the ENC and GC3S values for the IFN- α gene in the reference species. The solid line represented the curve if codon usage was only determined by GC content on the third codon position[26]. It is evident that there is only a small number of genes lying on the expect plot curve, most of the plot lay near to the solid line of this distribution (shown in Figure. 1).

Table 3. Summary Analysis of IFN- α Gene in Different Species

Species	Animals' IFN- α	CAI	ENC	Coding GC(%)	GC3S (%)
	<i>Equus caballus</i> IFN- α	0.809	45.690	60.20%	81.21%
	<i>Macaca mulatta</i> IFN- α	0.782	53.276	48.30%	59.88%
	<i>Mustela putorius furo</i> IFN- α	0.791	49.523	56.16%	71.52%
Mammal	<i>Pteropus vampyrus</i> IFN- α	0.797	47.855	54.49%	69.46%
	<i>Vulpes vulpes</i> IFN- α	0.813	41.025	60.61%	81.82%
	<i>Homo sapiens</i> IFN- α	0.791	50.840	48.50%	61.08%
	<i>Capra hircus</i> IFN- α	0.797	44.528	59.48%	74.85%
	<i>Bubalus bubalis</i> IFN- α	0.824	44.146	59.44%	77.11%
	<i>Felis catus</i> IFN- α	0.803	42.099	56.01%	72.67%
	<i>Canis familiaris</i> IFN- α	0.820	40.714	60.20%	81.21%
Birds	<i>grint panda</i> IFN- α	0.759	44.692	53.94%	67.88%
	<i>Sus scrofa</i> IFN- α	0.850	38.773	59.08%	76.65%
	<i>Anser anser rubrirostris</i> IFN- α	0.831	29.000	68.09%	92.07%
	<i>Duck</i> IFN- α	0.815	27.971	66.87%	90.24%
	<i>Gallus gallus</i> IFN- α	0.807	35.008	59.35%	79.88%
	<i>Shitou Goose</i> IFN- α	0.831	28.807	67.89%	92.07%

**Fig. 1.** The relationship between the effective number of codons (ENC) and the GC content of the third codon position (GC3) of the IFN- α gene of grint panda and 15 reference IFN- α genes

3.3 Phylogenetic Analysis on the IFN- α Gene of Grint Panda and Reference Species

A phylogenetic tree based on the amino acid sequences of the IFN- α gene of the 16 reference species was shown in Fig. 2. It shows that there are mainly two branches for the IFN- α family. GpIFN- α has been clustered in the mammal branch. And we can also see that the grint panda and *Mustela putorius furo* clustered in a monophyletic clade, and have a closer relationship with *Canis familiaris* and *Vulpes vulpes*. Multiple sequence alignment of the amino acid sequences testify to the conclusion that grint panda is evolutionarily closer to the animal mentioned above. The figure also shows that the IFN- α gene of grint panda is alienated to those whose natural host is avian on the evolution relationship.

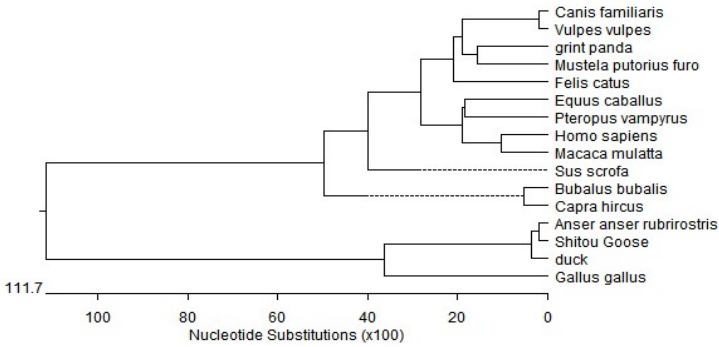


Fig. 2. Phylogenetic tree based on the amino acid sequences of the IFN- α gene of the 16 reference species (Table 1), and constructed by CLUSTAL-W and TREEVIEW software

3.4 Comparison of Codon Usage between GpIFN- α and *E. coli*, Yeast and *H. Sapiens*

The codon usage bias in genes remains at a certain level across species. We compared GpIFN- α gene with those of *E.coli*, yeast and *H. sapiens* to see which will be the suitable host for the optimal expression of GpIFN- α gene. Contrastive analysis showed that there are 40 codons showing distinct usage differences between GpIFN- α and *E.coli*, 38 codons between GpIFN- α and yeast, but only 30 between GpIFN- α and *H. sapiens*, suggesting that codon usage of GpIFN- α genes more closely resembles that of *H. sapiens* genes.

Table 4. Comparison of codon preferences between the GpIFN- α gene and *E. coli*, yeast and human

Amino acid	Codon	1/1000				DuIFN - α /		
		<i>IFN-α</i>	<i>E. coli</i>	Yeast	Human	<i>E. coli</i>	Yeast	Human
A(Ala)	GCA	6.061	20.6	16.1	16.1	0.29	0.38	0.38
A	GCC	36.364	25.5	12.5	28.4	1.43	2.91	1.28
A	GCG	12.121	31.7	6.1	7.5	0.38	1.99	1.62
A	GCT	24.242	15.6	21.1	18.6	1.55	1.15	1.30
C(Cys)	TGC	18.182	6.9	4.7	12.2	2.64	3.87	1.49
C	TGT	18.182	5.5	8	10	3.31	2.27	1.82
D(Asp)	GAC	30.303	18.6	20.2	25.6	1.63	1.50	1.18
D	GAT	6.061	32.1	37.8	21.9	0.19	0.16	0.28
E(Glu)	GAA	24.242	38.2	48.5	29	0.63	0.50	0.84
E	GAG	48.485	17.7	19.1	39.9	2.74	2.54	1.22
F(Phe)	TTC	30.303	16.9	18.2	20.6	1.79	1.67	1.47
F	TTT	18.182	23.2	26.1	17.1	0.78	0.70	1.06
G(Gly)	GGA	24.242	9	10.9	16.4	2.69	2.22	1.48
G	GGC	12.121	27.9	9.7	22.5	0.43	1.25	0.54
G	GGG	6.061	11.3	6	16.3	0.54	1.01	0.37

Table 4. (continued)

G	GGT	0.000	24.4	24	10.8	0.00	0.00	0.00
H(His)	CAC	6.061	9.8	7.7	15	0.62	0.79	0.40
H	CAT	18.182	13.6	13.7	10.5	1.34	1.33	1.73
I(Ile)	ATA	0.000	5.4	17.8	7.7	0.00	0.00	0.00
I	ATC	30.303	24.2	17	21.6	1.25	1.78	1.40
I	ATT	0.000	29.8	30.4	16.1	0.00	0.00	0.00
K(Lys)	AAA	6.061	33.2	42.2	24.1	0.18	0.14	0.25
K	AAG	36.364	10.7	30.7	32.2	3.40	1.18	1.13
L(Leu)	CTA	12.121	4	13.3	7.8	3.03	0.91	1.55
L	CTC	24.242	11	5.4	19.8	2.20	4.49	1.22
L	CTG	42.424	50.9	10.4	39.8	0.83	4.08	1.07
L	CTT	6.061	11.7	12.1	13	0.52	0.50	0.47
L	TTA	0.000	13.9	26.7	7.5	0.00	0.00	0.00
L	TTG	24.242	14	27	12.6	1.73	0.90	1.92
M(Met)	ATG	24.242	27	20.9	22.2	0.90	1.16	1.09
N(Asn)	AAC	18.182	21.4	24.9	19.5	0.85	0.73	0.93
N	AAT	6.061	18.6	36.3	16.7	0.33	0.17	0.36
P(Pro)	CCA	0.000	8.5	18.2	16.7	0.00	0.00	0.00
P	CCC	24.242	5.8	6.8	20.1	4.18	3.57	1.21
P	CCG	0.000	21.8	5.3	6.9	0.00	0.00	0.00
P	CCT	12.121	7.3	13.6	17.3	1.66	0.89	0.70
Q(Gln)	CAA	36.364	15	27.5	12	2.42	1.32	3.03
Q	CAG	36.364	29.5	12.1	34.1	1.23	3.01	1.07
R(Arg)	AGA	24.242	2.9	21.3	11.5	8.36	1.14	2.11
R	AGG	30.303	1.9	9.2	11.4	15.95	3.29	2.66
R	CGA	0.000	3.9	3	6.3	0.00	0.00	0.00
R	CGC	0.000	21	2.6	10.7	0.00	0.00	0.00
R	CGG	0.000	6.3	1.7	11.6	0.00	0.00	0.00
R	CGT	0.000	20.3	6.5	4.6	0.00	0.00	0.00
S(Ser)	AGC	30.303	16	9.7	19.3	1.89	3.12	1.57
S	AGT	0.000	9.5	14.2	11.9	0.00	0.00	0.00
S	TCA	18.182	7.8	18.8	12	2.33	0.97	1.52
S	TCC	18.182	8.9	14.2	11.9	2.04	1.28	1.53
S	TCG	12.121	8.7	8.5	4.4	1.39	1.43	2.75
S	TCT	24.242	8.7	23.5	14.7	2.79	1.03	1.65
T(Thr)	ACA	12.121	8.2	17.8	15.1	1.48	0.68	0.80
T	ACC	6.061	22.8	12.6	19.4	0.27	0.48	0.31
T	ACG	18.182	14.8	7.9	6.1	1.23	2.30	2.98
T	ACT	12.121	9.1	20.3	13	1.33	0.60	0.93
V(Val)	GTA	0.000	11.1	11.8	7.2	0.00	0.00	0.00
V	GTC	30.303	15.1	11.6	14.6	2.01	2.61	2.08

Table 4. (continued)

V	GTG	30.303	25.5	10.6	28.4	1.19	2.86	1.07
V	GTT	6.061	18.5	22	11	0.33	0.28	0.55
W(Trp)	TGG	18.182	15.2	10.3	12.7	1.20	1.77	1.43
Y(Tyr)	TAC	24.242	12.1	14.6	15.5	2.00	1.66	1.56
Y	TAT	0.000	16.5	18.9	12.1	0.00	0.00	0.00
*	TAA	0.000	2	1	0.7	0.00	0.00	0.00
*	TAG	0.000	0.3	0.5	0.6	0.00	0.00	0.00
*	TGA	6.061	1.1	0.7	1.5	5.51	8.66	4.04

4 Discussion

Codon bias refers to the nonrandom usage of synonymous codons for encoding amino acids in organisms. As it is related to the carrier molecular of genetic information(DNA) and functional molecular (protein) of life, this phenomenon implicates important biological sense[27]. Extensive studies had shown that synonymous codon usage in various organisms. Although the codon usage pattern among different species is a complex phenomenon, it plays an important role in illuminating the underlying mechanisms of codon usage patterns in order to understand the evolution of the species.

Signal peptides consist of short stretches of amino acids which, after protein delivery to the correct subcellular compartment, are frequently removed by specialized signal peptidases[28]. Only after removal of the signal peptide sequence, the precursor protein may be allowed to entry into the secretory pathway and become a normal function of the mature protein [29]. We intend to express the GpIFN- α gene in prokaryotic expression system and test its antiviral effects in future. In order to keep the higher biological activity, we cloned the gene fragment coding for mature protein of GpIFN- α . In this study, we analyzed the codon usage data of GpIFN- α without signal peptide and those of other 15 species of IFN- α . Additionally, we investigated the correlation of codon usage between GpIFN- α gene with *Escherichia coli*, yeast, and *Homo sapiens* to find the suitable expression system.

RSCU, CAI, ENC values, GC and GC3s content of the IFN- α gene were predicted by using the EMBOSS CUSP program and CHIPS program. RSCU with direct assessment codon usage bias of different codons in each gene sample. When its value greater than 1.0 indicated that the corresponding codon was used more frequently than expected. Table 2 showed that there were 26 preferred codons, most of them were strong bias towards the synonymous codons with G and C in the wobble, coincides with high G+C content at the third codon position (the mean value was 76.85%). CAI was used as a measure to predict the level of gene expression. The value was between 0 and 1[30], a higher value means a stronger codon usage bias and a higher expression level. the CAI values for IFN- α genes were slight variation among different species and with a mean value of 0.808, so we concluded IFN- α gene was likely to be moderately expressed in giant panda genome. The ENC is a measure analogous to the effective number of alleles. It ranges from 20, where only one codon is used for each amino acid to 61 if all codons are used equally[31]. In general, if the ENC value of a gene is 35 or less, that gene is thought to possess strong codon bias[32]. Analyzing the

ENC values of all the IFN- α genes, the results showed the ENC values of mammal's IFN- α genes (mean value of 45.26) were higher than birds' IFN- α gene (mean value of 30.20), So the codon usage bias was slightly low in mammal but strong in bird.

The ENC-plot (ENC plotted against GC) is used as part of a general strategy to investigate patterns of synonymous codon usage. In fig.1, the solid line represents the curve if codon usage is only determined by GC content on the third codon position [33], There is only a small number of genes lying on the expect plot curve, most of the plot lay near to the solid line. In principle, proprietary proportion of points lay near to the solid line on this distribution. It suggested that mutational bias was the main factor determining the codon usage variation among these IFN- α genes.

Comparative analysis of IFN- α genes in giant panda and the reference species indicated that synonymous codon usage in mammal's IFN- α genes are phylogenetically conserved. Data in table 3 show that the IFN- α genes of mammal have similar CAI, ENC value and GC content. But a slight difference with birds'. Therefore we can conclude species with near homology relationship on evolution share the near codon usage frequency and preference index or the similar index. The phylogenetic tree analysis based on the IFN- α genes testify to the conclusion. So species has a certain influence to the preference of codon usage.

The expression of functional proteins in heterologous hosts is a cornerstone of modern biotechnology. In order to find the suitable host for the optimal expression of GpIFN- α genes. we compared it with those of *E. coli*, yeast, and *H. sapiens*. The ratio higher than 2 or lower than 0.050 indicated the codon usage preference differed (shown in table 4.). The results revealed that *Homo sapiens* may more efficiently to express the GpIFN- α genes. If we want to express this gene in prokaryotic expression system, maybe some strategies should be taken in the target gene so that they more closely reflect the codon usage of the host.

In summary, codon usage patterns and the phylogenetic results we proposed here were useful to understand the characteres of GpIFN- α gene and a better understanding of both the evolutionary history of these genes. And also might help in increasing the efficiency of gene delivery/expression systems.

Acknowledgment. The authors are grateful to Ya'an Research Base of Giant Panda Breeding for providing blood samples of the giant panda. Chengdong Wang contributed equally to this work and should be considered as first authors. Zhi-wen Xu, Ling Zhu are the corresponding authors. Key Laboratory of Animal Disease and Human Health of Sichuan Province & Animal Biotechnology Center, College of Veterinary Medicine of Sichuan Agricultural University, 46# Xinkang Road, Yucheng district, Ya'an 625014, Sichuan province of China. Tel.: +86 835 2885477; fax: +86 835 2885477; E-mail address: abtcxzw@126.com(Z.Xu).

References

1. Gustafsson, C., Govindarajan, S., Minshull, J.: Codon bias and heterologous protein expression. *Trends in Biotechnology* 22(7), 346–353 (2004)
2. Hershberg, R., Petrov, D.: Selection on codon bias. *Annual Review of Genetics* 42, 287–299 (2008)

3. Ikemura, T.: Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution* 2(1), 13 (1985)
4. Sharp, P.: Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Research* 14(19), 7737 (1986)
5. Sharp, P., Li, W.: An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of Molecular Evolution* 24(1), 28–38 (1986)
6. Shields, D., Sharp, P., Higgins, D., Wright, F.: " Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Molecular Biology and Evolution* 5(6), 704 (1988)
7. Smith, N., Eyre-Walker, A.: Synonymous codon bias is not caused by mutation bias in G+C-rich genes in humans. *Molecular Biology and Evolution* 18(6), 982 (2001)
8. Bains, W.: Codon distribution in vertebrate genes be used to predict gene length. *Journal of Molecular Biology* 197(3), 379–388 (1987)
9. D'Onofrio, G., Ghosh, T., Bernardi, G.: The base composition of the genes is correlated with the secondary structures of the encoded proteins. *Gene* 300(1-2), 179–187 (2002)
10. Bernardi, G.: Compositional constraints and genome evolution. *Journal of Molecular Evolution* 24(1), 1–11 (1986)
11. Gouy, M., Gautier, C.: Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research* 10(22), 7055 (1982)
12. Tao, X., Dafu, D.: The relationship between synonymous codon usage and protein structure. *FEBS Letters* 434(1-2), 93–96 (1998)
13. Angellotti, M., Bhuiyan, S., Chen, G., Wan, X.: CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Research* 35(Web Server issue), W132 (2007)
14. Epstein, R., Lin, K., Tan, T.: A functional significance for codon third bases. *Gene* 245(2), 291–298 (2000)
15. Ma, J., Zhou, T., Gu, W., Sun, X., Lu, Z.: Cluster analysis of the codon use frequency of MHC genes from different species. *Biosystems* 65(2-3), 199–207 (2002)
16. Coghlan, A., Wolfe, K.: Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast-Chichester* 16(12), 1131–1146 (2000)
17. Marais, G., Duret, L.: Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. *Journal of Molecular Evolution* 52(3), 275–280 (2001)
18. Tao, Y., Zeng, B., Xu, L., Yue, B., Yang, D., Zou, F.: Interferon- γ of the Giant Panda (*Ailuropoda melanoleuca*): Complementary DNA Cloning, Expression, and Phylogenetic Analysis. *DNA and Cell Biology* 29(1), 41–45
19. Xu, L., Zeng, B., Peng, R., Zou, F., Yue, B.: Molecular cloning and sequence analysis of the gene encoding interleukin-6 of the giant panda (*Ailuropoda melanoleuca*). *Journal of Natural History* 42(39), 2585–2591 (2008)
20. Isaacs, A., Lindenmann, J.: "Virus interference: I". The interferon. *CA: A Cancer Journal for Clinicians* 38(5), 280 (1988)
21. Bin, F., Bu-feng, L., Guang-yuan, H.: Synonymous codon usage bias and overexpression of a synthetic gene encoding Interferon $\alpha 2b$ in yeast. *Virologica Sinica* 22(3), 226–232 (2007)
22. Duret, L., Mouchiroud, D.: Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 96(8), 4482 (1999)
23. Wright, F.: The effective number of codons' used in a gene. *Gene* 87(1), 23–29 (1990)
24. Sakai, H., Washio, T., Saito, R., Shinagawa, A., Itoh, M., Shibata, K., et al.: Correlation between sequence conservation of the 5' untranslated region and codon usage bias in *Mus musculus* genes. *Gene* 276(1-2), 101–105 (2001)

25. Lu, H., Zhao, W., Zheng, Y., Wang, H., Qi, M., Yu, X.: Analysis of synonymous codon usage bias in Chlamydia. *Acta Biochimica et Biophysica Sinica* 37(1), 1 (2005)
26. Gu, W., Zhou, T., Ma, J., Sun, X., Lu, Z.: Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus Research* 101(2), 155–161 (2004)
27. Wu, X., Wu, S., Ren, D., Zhu, Y., He, F.: The analysis method and progress in the study of codon bias. *Yi Chuan= Hereditas/Zhongguo Yi Chuan Xue Hui Bian Ji* 29(4), 420 (2007)
28. Tjalsma, H., Bolhuis, A., Jongbloed, J., Bron, S., Van Dijk: Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiology and Molecular Biology Reviews* 64(3), 515 (2000)
29. von Heijne, G.: The signal peptide. *Journal of Membrane Biology* 115(3), 195–201 (1990)
30. Zhang, S., Cheng, A., Wang, M.: Characterization of Codon Usage Bias in the Newly Identified DEV UL53 Gene, pp. 1–9. IEEE
31. Shen, F., Cheng, A., Wang, M., Huang, H., Li, C., Jiang, J., et al.: Factors affecting synonymous codon usage bias in the gC gene of DPV CHv strain. *African Journal of Microbiology Research* (5), 343–353
32. Xiang, J., Cheng, A., Wang, M., Zhang, S., Chang, H., Zhu, D., et al.: Analysis of Synonymous Codon Usage in the Capsid Gene UL38 of Duck Enteritis Virus, pp. 1–8. IEEE
33. Jia, R., Cheng, A., Wang, M., Xin, H., Guo, Y., Zhu, D., et al.: Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. *Virus Genes* 38(1), 96–103 (2009)