Research paper

Codon usage in *Alphabaculovirus* and *Betabaculovirus* hosted by the same insect species is weak, selection dominated and exhibits no more similar patterns than expected

Sheng-Lin Shi *, Yi-Ren Jiang, Rui-Sheng Yang, Yong Wang, Li Qin *

Insect Resource Engineering Research Center of Liaoning Province, College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang 110866, China

A B S T R A C T

Mutations shape synonymous codon usage bias in certain organism genomes, while selection shapes it in others. Lepidopteran-specific *Alphabaculovirus* and *Betabaculovirus* are two large genera in the family of *Baculoviridae*. In this study, we analyzed the codon usage patterns in 17 baculoviruses, including 10 alphabaculoviruses and 7 betabaculoviruses, which were isolated from seven insect species, and we characterized the codon usage patterns between *Alphabaculovirus* and *Betabaculovirus*. Our results show that all the baculoviruses possessed a general weak trend of codon bias. The differences of ENc (effective number of codons) values, nucleotide contents and the impacts of nucleotide content on ENc value within alpha-/betabaculovirus pairs were independent of whether the host species are the same or different. Furthermore, the majority of amino acid sequences adopted codons unequally in all viruses, but the numbers of common preferred codons between alpha- and betabaculoviruses hosted by the same insect species were not significantly different from the differences observed between alpha- and betabaculoviruses hosted by different insect species. In addition, the amino acids that adopt the same synonymous codon composition between alpha- and betabaculoviruses hosted by the same insect species were statistically as few as those between alpha- and betabaculoviruses hosted by different insect species. Correspondence analysis revealed that no major factors resulted in the codon bias in these baculoviruses, implying multiple minor influential factors exist. Neutrality plot analysis indicated that selection pressure dominated mutations in shaping the codon usage. However, the levels of selection pressure were not significantly different among viruses hosted by the same insect species. We expect that evolution would cause the alpha- and betabaculoviruses hosted by the same insect species to share more patterns, but this effect was not observed.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Synonymous codon usage bias (codon usage bias or codon bias) refers to the preference for a particular codon in a given organism. Codon bias in a genome or between different genomes is usually caused by adaptive changes (Goodarzi et al., 2008; Ingvarsson, 2008) and is widely present in biological systems from viruses to mammals (Eyre-Walker, 1991; Mirsafian et al., 2014; Shi et al., 2013). Synonymous codon usage affects protein biogenesis, such as translation efficiency and gene function, beyond specifying the amino acid sequence of a protein (Chaney and Clark, 2015; Supek, 2016). Hence, understanding codon bias is central to fields from molecular evolution to biotechnology (Isaacs et al., 2011; Plotkin and Kudla, 2011). Selection and/or mutation

bias are two general explanations for the existence of codon bias (Chen et al., 2014a; Chithambaram et al., 2014). Selectional explanation assumes that codon bias contributes to the efficiency and/or the accuracy of protein expression and is thus generated and maintained by selection, while the mutational or neutral explanation posits that codon bias exists because of nonrandomness in the mutational patterns (Hershberg and Petrov, 2008). At present, there is no common rule about the influential factors on codon bias. It is hard to predict whether selection or mutation dominates the codon usage in a given organism. As commonly known, mutational bias is the main force in shaping codon usage in certain viruses (Cristina et al., 2015; Zhou et al., 2015), whereas natural selection pressure plays important roles in others (Chen et al., 2014b; Shi et al., 2013).

Baculoviridae is a family of enveloped viruses with double-stranded DNA genomes ranging from 80 to 180 kb. According to the ninth virus taxonomy report (King et al., 2012), this family includes four genera: *Alphabaculovirus*, *Betabaculovirus*, *Gammabaculovirus* and *Deltabaculovirus*. *Alphabaculovirus* includes lepidopteran-specific nucleopolyhedroviruses, and *Betabaculovirus* comprises lepidopteran-

* Corresponding authors at: College of Bioscience and Biotechnology, Shenyang Agricultural University, No.120 Dongling Road, Shenyang 110866, China.
E-mail addresses: shishenglin@126.com (S.-L. Shi), jiangyiren56@126.com (Y.-R. Jiang), khankhan2000@163.com (R.-S. Yang), yongwang216@163.com (Y. Wang), qinli1963@163.com (L. Qin).

specific granuloviruses. Several studies focused on baculovirus phylogeny and evolution and are useful for improving the application of baculoviruses as pesticides, foreign gene expression and display systems, and mammalian cells transducing vectors (Herniou and Jehle, 2007; Herniou et al., 2003; Volkman, 2015). Presently, we know little about the synonymous codon usage pattern in baculoviruses. A previous study demonstrated that notable codon usage differences exist in *Autographa californica* multiple nucleopolyhedrovirus (Ranjan and Hasnain, 1995). Later, it was determined that codon bias in nucleopolyhedroviruses correlates with GC content but not with the gene length and gene expression level (Levin and Whittome, 2000). A recent study revealed that 40 of the 42 baculovirus genomes lack strong codon bias (Jiang et al., 2008). However, much remains unknown about codon usage in baculoviruses.

Coevolution of baculoviruses with their insect hosts is well established (Herniou et al., 2004; Cory and Myers, 2003). We assumed that same insect species shape the codon usage of baculoviruses that they host similarly and that different host species shape the codon usage of their hosted baculoviruses differently. Under this assumption, the codon usage between alpha- and betabaculovirus hosted by the same insect species should be more similar than that between those hosted by different insect species. If we select alpha- and betabaculoviruses hosted by the same insect species to compare their codon usage, then the host's influence is controlled, and the differences obtained can represent the differences between the two virus genera, not only between the two virus species. In the present study, we used 17 baculoviruses that share common insect host species to differentiate the codon usage patterns between *Alphabaculovirus* and *Betabaculovirus*. Our results show that selection pressure dominated mutations in shaping the weak codon usage in these baculoviruses. The alpha- and betabaculovirus hosted by the same insect species shared no more similar codon usage patterns than expected.

2. Materials and methods

2.1. Genome sequences

In NCBI (National Center for Biotechnology Information) genome database, 69 baculovirus genome sequences had been documented by May 26, 2015. We selected 17 of those genomes that represent alpha- and betabaculoviruses isolated from seven insect species [Table S1]. In each genome, we deleted the codons containing ambiguous bases, and omitted the open reading frames (ORFs) of <300 bp.

2.2. Effective number of codons and nucleotide composition

The effective number of codons (ENc value, ranging from 20 to 61) is defined as the number of codons that would yield the observed level of codon usage if all codons were equally frequent and is widely used to measure the degree of synonymous codon usage bias (Fuglsang, 2006; Wright, 1990). If only one codon is used for each amino acid, the ENc value would be 20 (extreme codon bias), and if all codons are used equally, the value would be 61 (no codon bias). In a genome, the majority of genes possess ENc values lower than 61, and only small number of genes possess ENc values equal to 61. As a result, the mean ENc value of a genome is always lower than 61. In general, ENc values less than or equal to 35 are described as strong, >45 are described as weak, and values in between these values are described as moderate (Chen et al., 2014b; Roychoudhury et al., 2011). ENc values correlated with total GC content (GC%) and GC content at the third synonymous codon position (GC3s%). ENc values were plotted against GC3s contents (ENc-plot analysis) to examine the influence of nucleotide composition on codon usage. The expected ENc value assumes equal use of G and C (A and T) in degenerate codon groups and can be calculated from GC3s content according to the equation provided by Wright (Wright, 1990). In the plot, a gene or genome spot whose codon choice is subject to mutation

pressure will lie on or just below the curve of the predicted values. In our plot, the spots represent individual genes. We used software CodonW (http://codonw.sourceforge.net/) to calculate ENc values and nucleotide composition. We employed one-way analysis of variance to compare ENc values, GC contents and GC3s contents, and employed Games-Howell's method to perform post hoc comparisons. We employed non-parametric Spearman rho to indicate correlation, and used Benjamini-Hotchberg correction (Benjamini and Hochberg, 1995) with false discovery rate $q^* = 0.05$ to set the significance level. The one-way analysis of variance and the correlation analysis were performed using software SPSS 16.0 (SPSS Inc., Chicago, IL, USA).

2.3. Relative synonymous codon usage

Relative synonymous codon usage (RSCU) is defined as the observed frequency of a codon usage divided by the frequency of that expected if all synonymous codons for that amino acid were used equally (Sharp and Li, 1986). A codon with RSCU value greater than one (>1) represents a positive codon usage bias, while a value less than one (<1) represents a negative codon usage bias. We identified preferred codons by comparing their RSCU values. For each amino acid, we selected the first two highest frequently used codons to perform a chi-square goodness-of-fit test to examine whether their frequencies are significantly different. When the *p* value was <0.05, we referred to the highest frequently used codon as the preferred codon. When the *p* value was >0.05, we reselected the first three highest frequently used codons to do a second round of chi-square goodness-of-fit test. Again, if the *p* value was <0.05, we referred the two highest frequently used codons both as the preferred codons. If the *p* value was >0.05, we tried another round of testing for the four highest frequently used codons. In any round of the chi-square goodness-of-fit test, the RSCU value of the codon selected should be greater than one (>1); in other words, the preferred codon selected must be positively biased. We used CodonW to calculate RSCU value, and used SPSS 16.0 to perform the chi-square goodness-of-fit test.

2.4. Correspondence analysis

Correspondence analysis is widely used to evaluate the major variation trend in codon usage among genes (Fellenberg et al., 2001; Perriere and Thioulouse, 2002). In correspondence analysis, each gene is represented as a 59-dimensional vector (because there are 59 synonymous codons), and each dimension corresponds to the RSCU value of one sense codon (excluding AUG, UGG and three stop codons). Correspondence analysis partitions the variation along 59 orthogonal axes and the first two axes often explain the largest fraction of variation in data (Andrea et al., 2011; Suzuki et al., 2008). We used CodonW to perform the correspondence analysis and focused on the first two axes to interpret the variation.

2.5. Neutrality plot analysis

Neutrality plot estimates the extent of neutrality of directional mutation pressure against selection and regarded the regression coefficient (slope) as the mutation-selection equilibrium coefficient (Sueoka, 1988). In general, the third codon position of synonymous codons includes an equal number A/T and G/C nucleotide pairs for most amino acids except for tryptophan (TGG), methionine (ATG) and one (ATA) of the three isoleucine codons. In the analysis, P_1, P_2, and P_3 are the observed GC contents of the first, second, and third codon positions of an individual gene; P_{12} is the average of P_1 and P_2. In the calculation, six codons (ATG, TGG, ATA, TAA, TAG, or TGA) were excluded (Sueoka, 1999a, 1999b). The removal of these six codons from the analysis eliminates odd-numbered synonymous codon sets and therefore avoids an extra cause of potential bias from Parity Rule 2, an intrastrand rule where A = T and G = C are expected if there is no bias in mutation and selection between the two complementary strands of DNA (Sueoka, 1999a,

1999b). Difference existed between P_1 and P_2 in their regressions against P_3 due to directional mutation pressure. Consequently, researchers usually use P_{12} instead of using P_1 and P_2 separately in the regression analysis. We used Bioperl (Stajich et al., 2002) script to calculate P_1, P_2, P_{12}, and P_3 values. The perl script is freely available on request from the author. We used SPSS 16.0 to plot P_{12} against P_3, to perform the simple linear regression analysis, and to compare the regression slopes by covariate analysis. Benjamini-Hotchberg correction (Benjamini and Hochberg, 1995) with false discovery rate $q^* = 0.05$ was used to set the significance level for multiple comparisons.

3. Results

3.1. ENc value and nucleotide composition

The mean ENc values of the 17 baculoviruses ranged from 47.30 (ChocNPV) to 53.65 (AgseGV) [Table 1]. The mean GC contents ranged from 33.34% (ChocGV) to 51.69% (ChocNPV), and the mean GC3s contents ranged from 30.06% (ChocGV) to 63.22% (ChocNPV) [Table 1].

In the correlation analysis (Table 2), GC contents positively correlated with GC3s contents in all viruses ($p < 0.05$, adjusted cutoff). ENc values positively correlated with GC contents and GC3s contents in AdorGV, AdorNPV and AgseGV, and negatively correlated with GC contents and GC3s contents in AgseNPV, AgseNPVB, ChocNPV, SfGV and SpltNPVII ($p < 0.029$, adjusted cutoff). ENc values positively correlated only with GC contents in HearMNPV and SfMNPV, and positively correlated only with GC3s contents in ChocGV and SpliGV ($p < 0.029$, adjusted cutoff). ENc values had no correlation with both GC contents and GC3s contents in HearGV, HearNPVNNg1, PlxyGV, PlxyNPV and SpltNPV ($p > 0.029$, adjusted cutoff). Altogether, 3 of the 10 alpha-/betabaculovirus pairs hosted by the same insect species and 9 of the 60 alpha-/betabaculovirus pairs hosted by different insect species showed similar correlation between ENc value and nucleotide composition. However, these two sets of data are not significantly different (Fisher's Exact Test, $x^2 = 1.358$, $p = 0.359$).

Comparisons of ENc values, GC contents and GC3s contents showed that the differences within alpha-/betabaculovirus pairs hosted by the same insect species were not significantly different from the differences observed within those hosted by different insect species. Among the 10 alpha-/betabaculovirus pairs hosted by the same insect species, ENc values within 2 pairs, GC contents and GC3s contents within 7 pairs were significantly different ($p < 0.05$) [Table S2]. In contrast, ENc values, GC contents and GC3s contents within 23, 47 and 50 of the 60 alpha-/betabaculovirus pairs hosted by different insect species, respectively, were significantly different ($p < 0.05$) [Table S2]. Between alpha-/

Table 2
Correlations between ENc (effective number of codons) values and nucleotide contents within alpha-/betabaculovirus pairs are independent of whether the host species are same or different.

Virus	ENc & GC		ENc & GC3s		GC& GC3s	
	ρ	P value	ρ	p	ρ	p
AdorGV	0.459	<0.001	0.495	<0.001	0.683	<0.001
AdorNPV	0.228	0.023	0.349	<0.001	0.691	<0.001
AgseGV	0.315	0.001	0.519	<0.001	0.638	<0.001
AgseNPV	−0.412	<0.001	−0.650	<0.001	0.758	<0.001
AgseNPVB	−0.482	<0.001	−0.652	<0.001	0.762	<0.001
ChocGV	0.100	0.338	0.259	0.012	0.519	<0.001
ChocNPV	−0.529	<0.001	−0.739	<0.001	0.731	<0.001
HearGV	0.109	0.193	0.062	0.459	0.739	<0.001
HearMNPV	0.259	0.002	0.130	0.123	0.736	<0.001
HearNPVNNg1	0.100	0.295	0.086	0.370	0.669	<0.001
PlxyGV	−0.139	0.175	−0.118	0.249	0.849	<0.001
PlxyNPV	−0.042	0.645	−0.095	0.300	0.817	<0.001
SfGV	−0.375	<0.001	−0.497	<0.001	0.718	<0.001
SfMNPV	0.246	0.006	0.161	0.075	0.650	<0.001
SpliGV	0.134	0.166	0.227	0.017	0.722	<0.001
SpltNPV	−0.111	0.233	−0.123	0.185	0.734	<0.001
SpltNPVII	−0.441	<0.001	−0.625	<0.001	0.753	<0.001

Note: Benjamini-Hotchberg correction with false discovery rate $q^* = 0.05$ was used to set the significance level. The cutoff for ENc & GC was $\alpha = 0.029$, for ENc & GC3s was $\alpha = 0.029$, and for GC& GC3s was $\alpha = 0.05$.

betabaculovirus pairs hosted by the same insect species and those hosted by different insect species, the ratios of virus/virus pairs that were significantly different to that of those not significantly different were not significantly different (Fisher's Exact Test, ENc value [$x^2 = 1.255$, $p = 0.314$], GC content [$x^2 = 0.338$, $p = 0.685$], GC3s content [$x^2 = 1.008$, $p = 0.380$]).

3.2. Synonymous codon usage

We used chi-square goodness-of-fit test to check whether the synonymous codons encoded each of the 18 amino acids (exclude methionine and tryptophan) randomly in each virus (Table S3). AgseGV, AgseNPV, HearNPVNNg1, PlxyMNPV, SfGV, SfMNPV, and SpltNPVII each contained one amino acid that was encoded randomly ($p > 0.05$). In contrast, AdorGV, PlxyGV, and SpltNPV each contained two amino acids that were encoded randomly ($p > 0.05$). HearMNPV and SpliGV each contained three amino acids that were encoded randomly ($p > 0.05$). The other five viruses had no amino acids that were encoded randomly ($p < 0.05$). These results suggest that although all viruses

Table 1
ENc (effective number of codons) values, GC contents and GC3s contents in the 17 baculovirus genomes.

Virus	ENc			GC%			GC3s%		
	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
AdorGV	50.11	4.62	49.18–51.04	35.05	3.50	34.34–35.75	35.57	5.34	34.50–36.65
AdorNPV	51.93	4.06	51.12–52.74	35.54	3.89	34.77–36.31	35.78	5.85	34.62–36.94
AgseGV	53.65	4.07	52.90–54.41	38.00	3.33	37.38–38.62	36.40	4.85	35.50–37.31
AgseNPV	50.16	4.46	49.38–50.94	47.69	4.04	46.99–48.40	59.22	7.19	57.96–60.47
AgseNPVB	52.27	5.47	51.33–53.21	47.47	4.05	46.78–48.17	57.45	7.71	56.13–58.78
ChocGV	48.57	5.31	47.48–49.66	33.34	4.29	32.46–34.21	30.06	5.70	28.89–31.23
ChocNPV	47.30	4.98	46.40–48.20	51.69	5.02	50.78–52.59	63.22	8.59	61.67–64.77
HearGV	53.45	4.38	52.73–54.17	42.32	3.81	41.69–42.94	46.01	7.15	44.84–47.19
HearMNPV	53.05	4.26	52.35–53.76	41.52	3.85	40.88–42.16	43.92	6.84	42.79–45.06
HearNPVNNg1	52.00	4.38	51.18–52.83	40.34	4.34	39.52–41.16	41.64	6.09	40.50–42.79
PlxyGV	51.98	5.47	50.88–53.08	42.35	6.01	41.14–43.57	51.06	10.43	48.96–53.16
PlxyNPV	51.59	4.45	50.79–52.39	41.55	4.90	40.67–42.44	47.21	8.73	45.63–48.78
SfGV	50.87	5.14	49.96–51.78	47.50	4.29	46.74–48.26	58.14	8.81	56.58–59.70
SfMNPV	52.38	4.60	51.56–53.20	40.71	4.00	40.00–41.43	47.16	6.77	45.95–48.37
SpliGV	53.32	4.48	52.47–54.17	40.00	4.14	39.21–40.78	43.47	7.04	42.13–44.81
SpltNPV	52.88	3.56	52.23–53.53	44.17	4.39	43.36–44.97	50.68	7.39	49.33–52.04
SpltNPVII	50.55	4.70	49.70–51.40	47.21	4.25	46.44–47.98	59.73	8.50	58.19–61.26

lacked strong codon bias, almost all amino acids adopted codons unequally.

In addition, we identified preferred codons for amino acids that were encoded by biased codons in each virus according to the synonymous codon frequencies (Table S4) and counted the common preferred codons between viruses (Table S5). The numbers of common preferred codons between alpha- and betabaculoviruses hosted by the same insect species were not significantly different from those between alpha- and betabaculoviruses hosted by different insect species ($p = 0.717$). Furthermore, we compared synonymous codon composition for each amino acid between viruses and found that the amino acids adopted the same synonymous codon composition between alpha- and betabaculoviruses hosted by the same insect species were statistically as few as those between alpha- and betabaculoviruses hosted by different insect species ($p = 0.748$).

3.3. ENc-plot analysis

ENc-plot analysis is generally used to investigate synonymous codon usage patterns. For each virus, the majority of the spots representing individual genes lay below the expected curve (Fig. 1), suggesting that apart from compositional constraints, other factors might have dominated influences on the codon usage variation. The plots of all viruses scattered in the same way, far below the curve of expected ENc values, implying no difference existed as to the influential factors between viruses hosted by the same insect species and viruses hosted by different insect species.

3.4. Correspondence analysis

The first two axes of correspondence analysis usually account for the major variation of codon usage bias. We also employed the correspondence analysis but found the values of the first two axes were not sufficiently large to account for the major variation in all viruses (Table S6). The first axis of PlxyGV possessed the largest value of 0.1529, accounting for 15.29% of the total variation, while the first axis of AgseGV possessed the smallest value of 0.0808, accounting for 8.08% of the total variation. The relatively smaller values of the first two axes suggest that no single major factor could account for the codon usage patterns in these viruses.

3.5. Neutrality plot analysis

Neutrality plot analysis (Fig. 2, Table 3) showed that the regressions of P_{12} to P_3 in AgseGV and SfGV were not statistically significant ($p > 0.044$, adjusted cutoff), while those of the other 15 viruses were statistically significant ($p < 0.044$, adjusted cutoff). The regression coefficients of P_{12} to P_3 in the 15 viruses ranged from 0.094 to 0.316 (Table 3), indicating a relatively lower and diverse neutrality. These results suggested that selection pressure dominated mutations in shaping codon usage in these baculoviruses. Slope comparison showed that no significant difference existed among regression coefficients of viruses hosted by the same insect species ($p > 0$, adjusted cutoff) [Table 3], indicating the extents of selection pressure on codon bias are not significantly different.

4. Discussion

Synonymous codon usage bias varies widely within genomic sequences of different organisms. Understanding the extent and causes of synonymous codon usage bias is essential to research focused on viral evolution and transmission and is particularly important for interpreting the interplay between viruses and their hosts (Shackelton et al., 2006). In the present study, we compared the synonymous codon usage between alpha- and betabaculovirus hosted by the same insect species to infer viral evolution at codon usage level.

A virus would be regarded as weakly biased when the mean ENc value is higher than 45; thus, all 17 baculoviruses we examined lack strong codon bias, which is consistent with a previous report (Jiang et al., 2008). A possible explanation for weak codon bias in baculoviruses is that it is essential for efficient replication, re-adaption and survival in host cells with potentially distinct codon preferences (Cristina et al., 2015; Shi et al., 2013; Zhou et al., 2015). Natural selection and mutation pressure are two main factors that account for codon usage variation in different organisms. We found that selection pressure dominates over mutation in determining the codon usage bias in the 17 baculoviruses, which is similar to the results we obtained in family *Parvoviridae* (Shi et al., 2013), but the selection pressure exert on viruses hosted by the same insect species are similar or not significantly different.
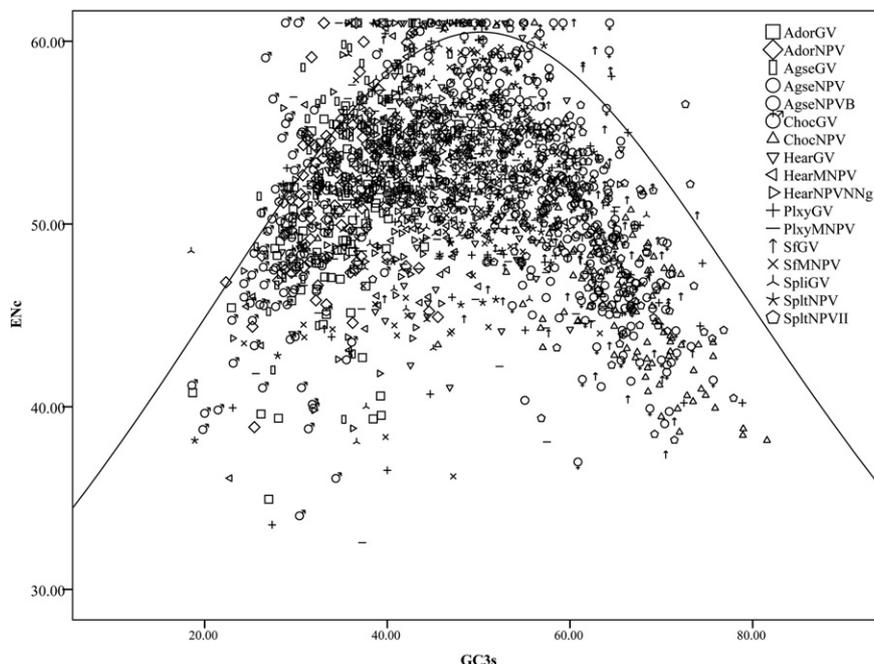

Fig. 1. ENc (effective number of codons) values of the majority of the genes in each virus are lower than the expected values. The curve indicates the expected ENc (effective number of codons) values in the case that GC compositional constraints alone account for codon usage bias.
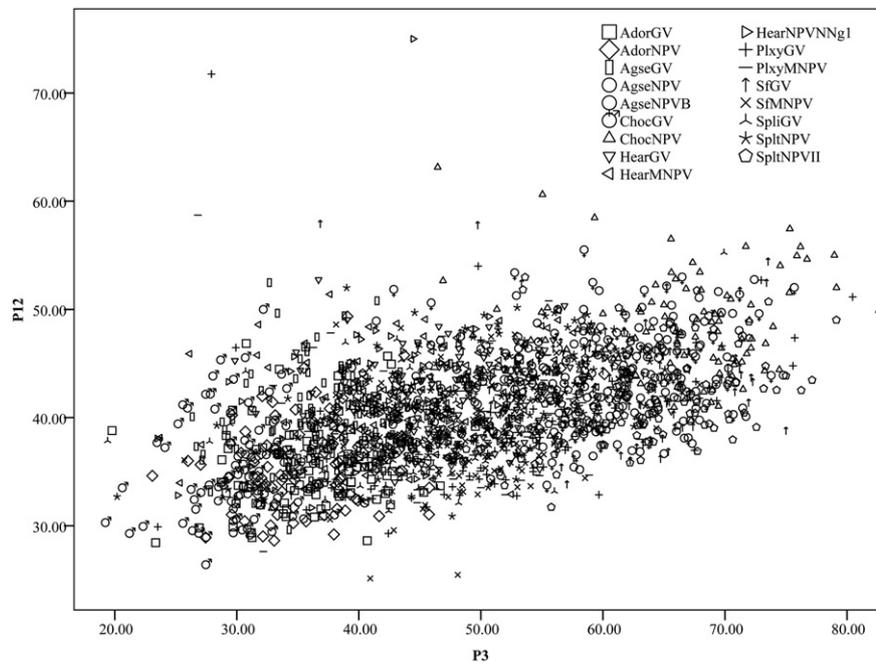
Fig. 2. Plot of P_{12} (average of P_1 and P_2) against P_3 for each virus.

As to the impact of a host on the codon usage of viruses, the early study demonstrated that the codon usage in severe acute respiratory syndrome coronavirus is not host specific (Gu et al., 2004). However, more studies have indicated: a) codon usage bias of viruses is related with their hosts during the adaptation process (Bahir et al., 2009; Chantawannakul and Cutler, 2008; Ma et al., 2015; Nasrullah et al., 2015), and b) viruses coincide their codon usage with hosts (Cheng et al., 2012; Kattoor et al., 2015). In the present research, we aimed to differentiate the codon usage patterns between *Alphabaculovirus* and *Betabaculovirus* by comparing their ENc values, nucleotide contents, preferred codons, synonymous codon compositions for each amino acid and the influential factors of codon usage. We found that the differences of these indices within alpha-/betabaculovirus pairs are independent of whether the host species are the same or different. This result disagrees with our assumption that the viruses hosted by the same insect species should share more common codon usage patterns because the hosts' impacts are similar.

5. Conclusions

Comparative analysis of codon usage patterns between *Alphabaculovirus* and *Betabaculovirus* genera members that were isolated from the same insect species has provided a basic understanding of the evolutionary characteristics of codon usage in these viruses. Our results demonstrate that codon biases in these 17 baculoviruses are weak. The differences of ENc values, GC contents and the impacts of GC content on ENc value within alpha-/betabaculovirus pairs hosted by the same insect species are not significantly different from that of those hosted by different insect species. The numbers of common preferred codons and amino acids that have the same synonymous codon composition between viruses hosted by the same insect species are not significantly different from that of those between viruses hosted by different insect species. Though no major factors could solely account for the codon usage pattern, we find selection pressure dominates the codon usage over mutation. Additionally, the

Table 3
Slopes of P_{12} (average of P_1 and P_2) against P_3 are small and have no significant difference among viruses hosted by the same insect species.

Virus	Slope (95% CI)	R Square	Significance of the regression	Significance of slope comparison
AdorGV	0.166 (0.018–0.315)	0.049	0.029	0.348
AdorNPV	0.260 (0.129–0.390)	0.138	<0.001	
AgseGV	0.148 (−0.007–0.303)	0.031	0.060	
AgseNPV	0.163 (0.067–0.259)	0.083	0.001	0.607
AgseNPVB	0.098 (0.005–0.191)	0.032	0.039	
ChocGV	0.316 (0.147–0.485)	0.131	<0.001	0.279
ChocNPV	0.207 (0.105–0.310)	0.120	<0.001	
HearGV	0.130 (0.045–0.215)	0.061	0.003	
HearMNPV	0.123 (0.026–0.219)	0.043	0.013	0.420
HearNPVNNg1	0.223 (0.073–0.373)	0.074	0.004	
PlxyGV	0.207 (0.099–0.314)	0.133	<0.001	0.758
PlxyNPV	0.185 (0.091–0.279)	0.113	<0.001	
SfGV	0.062 (−0.027–0.152)	0.015	0.171	0.273
SfMNPV	0.143 (0.028–0.258)	0.048	0.015	
SpliGV	0.204 (0.097–0.312)	0.117	<0.001	
SpltNPV	0.213 (0.106–0.319)	0.120	<0.001	0.147
SpltNPVII	0.094 (0.008–0.181)	0.038	0.033	

Note: Benjamini-Hotchberg correction with false discovery rate $q^* = 0.05$ was used to set the significance level. The cutoff for the regression was $\alpha = 0.44$, for slope comparison was $\alpha = 0$.

levels of selection pressure are not significantly different between viruses hosted by the same insect species.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2016.07.042.

Acknowledgments

References

Andrea, L.D., Pintó, R.M., Bosch, A., Musto, H., Cristin, J., 2011. A detailed comparative analysis on the overall codon usage patterns in hepatitis A virus. Virus Res. 157, 19–24.

Bahir, I., Fromer, M., Prat, Y., Linial, M., 2009. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol. Syst. Biol. 5, 311.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

Chaney, J.L., Clark, P.L., 2015. Roles for synonymous codon usage in protein biogenesis. Annu. Rev. Biophys. 44, 143–166.

Chantawannakul, P., Cutler, R.W., 2008. Convergent host-parasite codon usage between honeybee and bee associated viral genomes. J. Invertebr. Pathol. 98, 206–210.

Chen, H., Sun, S., Norenburg, J.L., Sundberg, P., 2014a. Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). PLoS One 9, e85631.

Chen, Y., Shi, Y., Deng, H., Gu, T., Xu, J., Ou, J., Jiang, Z., Jiao, Y., Zou, T., Wang, C., 2014b. Characterization of the porcine epidemic diarrhea virus codon usage bias. Infect. Genet. Evol. 28, 95–100.

Cheng, X.F., Wu, X.Y., Wang, H.Z., Sun, Y.Q., Qian, Y.S., Luo, L., 2012. High codon adaptation in citrus tristeza virus to its citrus host. Virol. J. 9, 113.

Chithambaram, S., Prabhakaran, R., Xia, X., 2014. The effect of mutation and selection on codon adaptation in Escherichia coli bacteriophage. Genetics 197, 301–315.

Cory, J.S., Myers, J.H., 2003. The ecology and evolution of insect baculoviruses. Annu. Rev. Evol. Syst. 34, 239–272.

Cristina, J., Moreno, P., Moratorio, G., Musto, H., 2015. Genome-wide analysis of codon usage bias in Ebolavirus. Virus Res. 196, 87–93.

Eyre-Walker, A.C., 1991. An analysis of codon usage in mammals: selection or mutation bias? J. Mol. Evol. 33, 442–449.

Fellenberg, K., Hauser, N.C., Brors, B., Neutzner, A., Hoheisel, J.D., Vingron, M., 2001. Correspondence analysis applied to microarray data. Proc. Natl. Acad. Sci. U. S. A. 98, 10781–10786.

Fuglsang, A., 2006. Estimating the "effective number of codons": the Wright way of determining codon homozygosity leads to superior estimates. Genetics 172, 1301–1307.

Goodarzi, H., Torabi, N., Najafabadi, H.S., Archetti, M., 2008. Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons. Gene 407, 30–41.

Gu, W., Zhou, T., Ma, J., Sun, X., Lu, Z., 2004. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. Virus Res. 101, 155–161.

Herniou, E.A., Jehle, J.A., 2007. Baculovirus phylogeny and evolution. Curr. Drug Targets 8, 1043–1050.

Herniou, E.A., Olszewski, J.A., Cory, J.S., O'Reilly, D.R., 2003. The genome sequence and evolution of baculoviruses. Annu. Rev. Entomol. 48, 211–234.

Herniou, E.A., Olszewski, J.A., O'Reilly, D.R., Cory, J.S., 2004. Ancient coevolution of baculoviruses and their insect hosts. J. Virol. 78, 3244–3251.

Hershberg, R., Petrov, D.A., 2008. Selection on codon bias. Annu. Rev. Genet. 42, 287–299.

Ingvarsson, P.K., 2008. Molecular evolution of synonymous codon usage in Populus. BMC Evol. Biol. 8, 307.

Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., Tolonen, A.C., Gianoulis, T.A., Goodman, D.B., Reppas, N.B., Emig, C.J., Bang, D., Hwang, S.J., Jewett, M.C., Jacobson, J.M., Church, G.M., 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333, 348–353.

Jiang, Y., Deng, F., Wang, H., Hu, Z., 2008. An extensive analysis on the global codon usage pattern of baculoviruses. Arch. Virol. 153, 2273–2282.

Kattoor, J.J., Malik, Y.S., Sasidharan, A., Rajan, V.M., Dhama, K., Ghosh, S., Banyai, K., Kobayashi, N., Singh, R.K., 2015. Analysis of codon usage pattern evolution in avian rotaviruses and their preferred host. Infect. Genet. Evol. 34, 17–25.

King, A.M., Adams, M.J., Carstens, E.B., Lefkowitz, E., 2012. Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses. 1 ed. Elsevier Academic Press, San Diego.

Levin, D.B., Whittome, B., 2000. Codon usage in nucleopolyhedroviruses. J. Gen. Virol. 81, 2313–2325.

Ma, Y.P., Liu, Z.X., Hao, L., Ma, J.Y., Liang, Z.L., Li, Y.G., Ke, H., 2015. Analysing codon usage bias of cyprinid herpesvirus 3 and adaptation of this virus to the hosts. J. Fish Dis. 38, 665–673.

Mirsafian, H., Mat Ripen, A., Singh, A., Teo, P.H., Merican, A.F., Mohamad, S.B., 2014. A comparative analysis of synonymous codon usage bias pattern in human albumin superfamily. TheScientificWorldJOURNAL 2014, 639682.

Nasrullah, I., Butt, A.M., Tahir, S., Idrees, M., Tong, Y., 2015. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. BMC Evol. Biol. 15, 174.

Perriere, G., Thioulouse, J., 2002. Use and misuse of correspondence analysis in codon usage studies. Nucleic Acids Res. 30, 4548–4555.

Plotkin, J.B., Kudla, G., 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12, 32–42.

Ranjan, A., Hasnain, S.E., 1995. Codon usage in the prototype baculovirus–Autographa californica nuclear polyhedrosis virus. Indian J. Biochem. Biophys. 32, 424–428.

Roychoudhury, S., Pan, A., Mukherjee, D., 2011. Genus specific evolution of codon usage and nucleotide compositional traits of poxviruses. Virus Genes 42, 189–199.

Shackelton, L.A., Parrish, C.R., Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol. 62, 551–563.

Sharp, P.M., Li, W.H., 1986. Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons. Nucleic Acids Res. 14, 7737–7749.

Shi, S.L., Jiang, Y.R., Liu, Y.Q., Xia, R.X., Qin, L., 2013. Selective pressure dominates the synonymous codon usage in parvoviridae. Virus Genes 46, 10–19.

Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G.R., Korf, I., Lapp, H., Lehväslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D., Birney, E., 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res. 12, 1611–1618.

Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. U. S. A. 85, 2653–2657.

Sueoka, N., 1999a. Translation-coupled violation of parity rule 2 in human genes is not the cause of heterogeneity of the DNA G + C content of third codon position. Gene 238, 53–58.

Sueoka, N., 1999b. Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. J. Mol. Evol. 49, 49–62.

Supek, F., 2016. The code of silence: widespread associations between synonymous codon biases and gene function. J. Mol. Evol. 82, 65–73.

Suzuki, H., Brown, C.J., Forney, L.J., Top, E.M., 2008. Comparison of correspondence analysis methods for synonymous codon usage in bacteria. DNA Res. 15, 357–365.

Volkman, L.E., 2015. Baculoviruses and nucleosome management. Virology 476, 257–263.

Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87, 23–29.

Zhou, H., Yan, B., Chen, S., Wang, M., Jia, R., Cheng, A., 2015. Evolutionary characterization of Tembusu virus infection through identification of codon usage patterns. Infect. Genet. Evol. 35, 27–33.