

Genome dynamics in three different geographical isolates of white spot syndrome virus (WSSV)

Gaurav Sablok · Arturo Sánchez-Paz ·
XianMing Wu · Jayant Ranjan ·
Jimmy Kuo · Ingo Bulla

Received: 23 February 2012 / Accepted: 21 May 2012 / Published online: 27 July 2012
© Springer-Verlag 2012

Abstract White spot syndrome virus (WSSV), the sole member of the monotypic family *Nimaviridae*, is considered an extremely lethal shrimp pathogen. Despite its impact, some essential biological characteristics related to WSSV genome dynamics, such as the synonymous codon usage pattern and selection pressure in genes, remain to be elucidated. The results show that compositional limitations and mutational pressure determine the codon usage bias

and base composition in WSSV. Furthermore, different forces of selective pressure are acting across various regions of the WSSV genome. Finally, this study points out the possible occurrence of two major recombination events.

Keywords Codon usage bias · Correspondence analysis · Gene expression level · White spot syndrome virus

G. Sablok, A. Sánchez-Paz and X. Wu contributed equally to work.

G. Sablok (✉)
Fondazione Edmund Mach, Centro Ricerca ed Innovazione,
Via Mach 2, 38010 San Michele all'Adige, TN, Italy
e-mail: sablokg@gmail.com

A. Sánchez-Paz (✉)
Laboratorio de Sanidad Acuicola, Centro de Investigaciones
Biológicas del Noroeste (CIBNOR), S.C. (Campus Hermosillo),
Calle Hermosa No. 101. Col. Los Angeles, 83106 Hermosillo,
Sonora, Méjico
e-mail: asanchez04@cibnor.mx

X. Wu
School of Biological Science and Technology, ShenYang
Agricultural University, ShenYang 110866, LiaoNing,
People's Republic of China

J. Ranjan
Fondazione Edmund Mach, Centro Ricerca ed Innovazione,
Via Mach 2, 38010 San Michele all'Adige, TN, Italy

J. Kuo
Department of Planning and Research, National Museum
of Marine Biology and Aquarium, Pingtung 944, Taiwan

I. Bulla
Universität Greifswald, Institut für Mathematik und Informatik,
Walther-Rathenau-Straße 47, 17487 Greifswald, Germany

Introduction

Viral diseases have become a limiting factor to the sustainable growth of the global shrimp culture industry [1]. White spot syndrome virus (WSSV), the sole member of the monotypic family *Nimaviridae*, genus *Whispovirus* [2], is an extremely lethal pathogen that can cause cumulative mortality of up 100 % within 2–10 days after the onset of clinical signs [3, 4]. Furthermore, it has been reported that WSSV is capable of infecting most of the commercially cultivated shrimp species and has consistently emerged as a highly prevalent and widespread virus. The WSSV virion consists of an enveloped nucleocapsid containing a circular double-stranded DNA genome that shows size variations among different geographic isolates (307,287, 305,107, and 292,967 base pairs [bp] for the Taiwan, China, and Thailand isolates, respectively) [5–7].

It has long been known that all organisms have a specific codon usage signature, and the degeneracy of the genetic code implies that multiple codons specify the same amino acid. Several factors have been proposed to explain the deviations in codon patterns, and such uneven usage is not selectively neutral as previously proposed, but related to compositional constraints and translational selection, gene expression [8], protein structure formation

[9], viral packaging [10], GC-biased gene conversion [11] and even tRNA base modifications [12]. It has been proposed that new viruses and strains may emerge as a result of selection pressures from environmental fluctuations or by host shifts [13].

Viral recombination, here defined as the exchange of genetic material between at least two viral genomes [14], is a process that influences viral fitness at many different levels [15, 16]. It has been suggested that viral recombination may be a way in which viruses adapt quickly to changing environmental conditions, new hosts and ecological niches [17, 18], as recombination can enable access to evolutionary innovations that would otherwise be inaccessible by mutation alone [19].

Because of the rising prevalence of WSSV virus and the global implications of this virus in the aquaculture trade, it is imperative to understand its genome dynamics, which may facilitate its evolution to create novel variants that can adapt to a changing environment and host genotype [20]. We studied all three of the different geographical isolates sequenced so far to illustrate the genome dynamics of WSSV. The findings of the present study established the presence of compositional constraints in the WSSV genome. Interestingly, we found that most of the genes that are under the influence of positive selection are associated with the control of virus replication, which confers some characteristics to the viral genome that may ensure its efficient replication and may consequently provide increased fitness. The presence of recombination hotspots in the WSSV genome was also evaluated, and some factors that might influence the recombination rate are proposed.

Materials and methods

Genome sequence data and multivariate analysis

The W-70, W-93 and W-29 genome sequences were retrieved from the GenBank database (accession numbers AF440570, AF332093, and AF369029). A threshold of 100 codons was applied to sequence filtering, and finally, 260, 252, and 146 complete coding sequences (CDS) from W-70, W-93 and W-29, respectively, were extracted directly to avoid sampling bias in calculations of codon usage [21]. The G + C frequency distribution was calculated as described earlier [22]. The effective number of codons (N_c) was calculated as described previously [21]. The relative synonymous codon usage (RSCU) values were calculated to normalize and identify the intra-genomic variations with differing amino acid compositions [23]. Correspondence analysis (COA) for RSCU was implemented using Codon W (<http://codonw.sourceforge.net>) [24].

The codon adaptation index—A measure of gene expression and evolution

The codon adaptation index (CAI), determines the bias of codon usage in highly expressed genes. For the calculation of CAI, a set of 16 highly expressed genes was selected for each genome. These included wsv151 (latency related), wsv427 (latency related), wsv366 (latency related), wsv230 (ICP11), wsv360 (VP664), wsv421 (VP28), wsv069 (iE1), wsv254 (VP37), wsv514 (DNA polymerase), wsv129 (VP357), wsv214 (VP15), wsv311 (VP26), wsv414 (VP19), wsv002 (VP24), wsv386 (VP68), wsv001 (Collagen-like) as previously suggested [25]. All correlations were based on the nonparametric Spearman's rank correlation (ρ) analysis method using R (<http://www.r-project.org/>). In order to compute orthologs, the best reciprocal BLAST hit approach (RBH) approach was used to find the best bidirectional hits. To calculate dN/dS ratios, amino acid sequences were first aligned using ClustalW1.83 with default parameters [26], and the corresponding codon alignment and dN/dS ratio were then calculated using an in-house Perl script.

Characterization of potential recombination events

Detection of potential recombinant sequences, identification of potential parental sequences, and localization of possible recombination breakpoints was done using the GENECON, BOOTSCAN, MaxChi, CHIMAERA, SISCAN and 3SEQ methods embedded in the RDP3 software package [27]. A multiple-comparison-corrected P-value cutoff of 0.01 was used throughout the study.

Results and discussion

Codon usage pattern in the WSSV genome and highly expressed genes

It has been shown that strong codon bias is common in highly expressed genes compared to those that are not highly expressed within the same genome [23, 28]. The overall RSCU values of the 59 sense codons in the whole genome of the WSSV isolates and for 16 highly expressed genes are shown in Table 1. Codon usage in the WSSV isolates is preponderantly A- or T-ended (W-29, 67 % T-ended, 22 % A-ended, and 11 % G- or C-ended; W-70 and W-93, 56 % T-ended, 39 % A-ended, and 5 % G-ended), which correlates with the low overall GC₃ content in all three of the isolates (~39.0 %). It was further observed that the average GC content of the three isolates at the first position was higher than at the second and third codon positions, which clearly demonstrates the

Table 1 Summary of the average relative synonymous codon usage (RSCU) of the 59 degenerate codons used in the three WSSV genomes for whole-genome analysis (all) and for 16 HE genes (high)

AA	Codon	RSCU						AA	Codon	RSCU					
		(W-29)		(W-70)		(W-93)				(W-29)		(W-70)		(W-93)	
		All	HE	All	HE	All	HE			All	HE	All	HE	All	HE
Asn	AAU	1.16	0.88	1.18	0.89	1.17	0.89	Pro	CCU	1.72	1.24	1.52	1.23	1.52	1.23
	AAC	0.83	0.99	0.78	0.99	0.78	0.99		CCC	0.88	1.31	0.84	1.32	0.85	1.32
Asp	GAU	1.23	1.15	1.17	1.15	1.17	1.15		CCA	1.14	1.36	1.25	1.36	1.25	1.36
	GAC	0.77	0.72	0.76	0.72	0.77	0.72		CCG	0.27	0.09	0.36	0.09	0.35	0.09
Cys	UGU	1.22	1.04	1.18	1.04	1.17	1.04	Thr	ACU	1.46	1.11	1.29	1.11	1.29	1.11
	UGC	0.66	0.58	0.65	0.58	0.64	0.58		ACC	0.77	1.29	0.72	1.30	0.72	1.30
His	CAU	1.01	1.10	0.99	1.11	0.98	1.11		ACA	1.36	1.16	1.45	1.16	1.46	1.16
	CAC	0.90	0.65	0.84	0.64	0.83	0.64		ACG	0.41	0.44	0.50	0.44	0.49	0.44
Lys	AAA	1.09	0.85	1.16	0.85	1.16	0.85	Val	GUU	1.23	1.50	1.21	1.50	1.19	1.49
	AAG	0.91	1.15	0.83	1.15	0.83	1.15		GUC	0.67	0.95	0.71	0.96	0.71	0.96
Phe	UUU	1.19	0.99	1.14	0.98	1.14	0.98		GUA	1.07	0.63	1.06	0.62	1.06	0.62
	UUC	0.81	0.89	0.83	0.89	0.83	0.89		GUG	1.02	0.91	1.03	0.93	1.04	0.93
Tyr	UAU	1.01	0.82	0.99	0.82	0.98	0.82	Arg	CGU	0.80	1.01	0.75	1.01	0.76	1.01
	UAC	0.95	1.18	0.88	1.18	0.88	1.18		CGC	0.53	0.64	0.51	0.64	0.53	0.63
Gln	CAA	1.25	1.09	1.07	1.09	1.06	1.09		CGA	0.47	0.45	0.54	0.45	0.53	0.45
	CAG	0.72	0.67	0.69	0.67	0.69	0.67		CGG	0.30	0.19	0.35	0.19	0.35	0.19
Glu	GAA	1.42	1.20	1.38	1.20	1.38	1.20		AGA	2.28	2.28	2.25	2.28	2.22	2.23
	GAG	0.57	0.80	0.61	0.80	0.61	0.80		AGG	1.61	1.43	1.58	1.44	1.59	1.49
Ile	AUU	1.26	1.36	1.16	1.36	1.18	1.36	Leu	UUA	1.17	1.07	1.22	1.08	1.21	1.08
	AUC	0.74	1.18	0.74	1.19	0.74	1.19		UUG	1.31	1.38	1.38	1.38	1.38	1.38
Ala	AUA	1.00	0.45	1.07	0.45	1.05	0.45		CUU	1.11	1.81	1.06	1.82	1.06	1.82
	GCU	1.74	1.07	1.19	1.08	1.18	1.08		CUC	0.85	0.76	0.86	0.75	0.85	0.75
Gly	GCC	0.62	1.46	0.93	1.46	0.94	1.46	Ser	CUA	0.82	0.52	0.74	0.51	0.74	0.51
	GCA	0.64	1.09	1.44	1.08	1.44	1.08		CUG	0.75	0.46	0.75	0.47	0.75	0.46
	GCG	1.37	0.38	0.41	0.37	0.41	0.38		UCU	2.00	1.54	1.87	1.54	1.88	1.54
	GGU	0.91	1.10	0.99	1.10	1.01	1.10		UCC	0.90	1.53	0.97	1.53	0.98	1.53
	GGC	1.34	0.80	0.65	0.80	0.64	0.80		UCA	1.21	0.78	1.28	0.78	1.28	0.78
	GGA	0.35	1.60	1.44	1.60	1.44	1.60		UCG	0.40	0.24	0.50	0.24	0.50	0.24
	GGG	1.07	0.50	0.83	0.49	0.82	0.49	AGU	1.01	0.97	0.86	0.97	0.87	0.98	
								AGC	0.48	0.94	0.49	0.94	0.48	0.94	

The highest RSCU of codons for each amino acid is in boldface. Isolate names: W-29, Thailand; W-70, Taiwan; and W-93, China
AA represents amino acid

GC compositional pressure on the biased codon usage. We further evaluated the relationship between nucleotide content and codon usage using an effective number of codons (Nc) - plot. The Nc- GC_{3s} plot has been widely used to study codon usage variation among different genes, as it has been shown that this index has a relationship to GC₃. The Nc value showed a wide variation, ranging from 23 to 61 in W-93. This shows that some of the genes with low Nc values have a stronger bias in comparison to genes with higher Nc values, which supports the presence of compositional constraints and a bias gradient in WSSV genomes.

Multivariate analysis of codon usage

Correspondence analysis on RSCU (COA) for isolates W-29, W-70 and W-93 revealed that axis 1 accounted for ~12.79, ~9.06, and ~9.09 %, respectively, of the total variation of the 59-dimensional space. Interestingly, the RSCU values observed in the W-29 isolate were higher than those observed in the W-70 and W-93 isolates. These results may indicate that the reduction of the genome size of W-29 has favored the usage of T-ending codons, while isolates W-70 and W-93 (both of which have larger genome sizes), may seem to have more varied options for

codon usage, and taken together, this may reflect that compositional limitations played a central role in shaping the codon usage pattern of WSSV. It seems probable that the reduction of the genome size of WSSV over time has selectively driven the eradication of background nucleotide content, specifically affecting the number of A-ending codons, but also favoring the appearance of C-ending codons. This can be interpreted as an adaptive strategy that may offer an advantage to W-29 by allowing unrestricted access to the full pool of tRNAs of the host to exploit its translation machinery in order to replicate unrestrictedly. Aragonès et al. [29] found that poliovirus shows a highly optimized codon usage that conforms to that of the host cell, confirming that viral replication reaches its maximum level when the correspondence between codon usage (demand) and tRNA availability (supply) is optimal.

Gene expression in WSSV

To confirm the assumption that highly expressed genes are clustered along the first major axis, the codon adaptation index (CAI) was calculated for all of the genes identified in the WSSV genome. CAI was calculated taking highly expressed genes as a reference (see “[Materials and methods](#)”). A weak but significantly positive correlation (W-29; $r = 0.065$, W-70; $r = 0.010$, and W-93; $r = 0.002$, $P < 0.001$) was observed between the positions of the genes along the first major axis and their corresponding CAI values in all isolates. The CAI value, which ranges between 0 and 1, indicates that genes with a CAI value close to 1 are composed of very frequently occurring codons. In this case, all WSSV isolates showed CAI values that range from ~ 0.5 to ~ 0.85 , but most of the WSSV genes showed CAI values between 0.7 and 0.8. It is worth noting that the CAI values for W-29 span a narrower range than those of W-70 and W-93, which may indicate that the WSSV genes avoid the use of rare codons, resulting in a codon usage bias. It is known that the introduction of rare codons, or pairs of rare codons, into an ORF reduces viral translation efficiency [30, 31].

Substitution rate and evolutionary constraints

Differences in the synonymous and non-synonymous nucleotide substitution ratio (K_a/K_s , termed as the “acceptance rate”) between WSSV isolates were also investigated. The acceptance rate has been widely used as an estimator of the stringency of the purifying selection or the strength of adaptive evolution. Among the 132 orthologous genes, a total of 51 genes appear to be under positive selection. Furthermore, it was found that the average synonymous (K_s) rate for the orthologs under positive selection

is 0.111 ± 0.0023 , and the average non-synonymous (K_a) substitution rate is 0.040 ± 0.0036 . Interestingly, only five of the WSSV orthologs under positive selection showed a relatively high ratio of synonymous substitutions over non-synonymous substitutions (ORF134, $K_s = 4/1218$, $K_a = 1/1218$; ORF42, $K_s = 3/1279$, $K_a = 2/1279$), while most of these orthologs showed high ratios of non-synonymous to synonymous substitutions (ORF14, $K_s = 39/301$, $K_a = 5/301$; ORF30, $K_s = 190/1683$, $K_a = 37/1683$; ORF183, $K_s = 147/496$, $K_a = 31/496$; ORF61, $K_s = 68/579$, $K_a = 16/579$; ORF40, $K_s = 73/1534$, $K_a = 19/1534$). According to van Hulten et al. [6], ORF30 encodes a collagen-like protein, and ORF61 encodes a putative serine/threonine protein kinase.

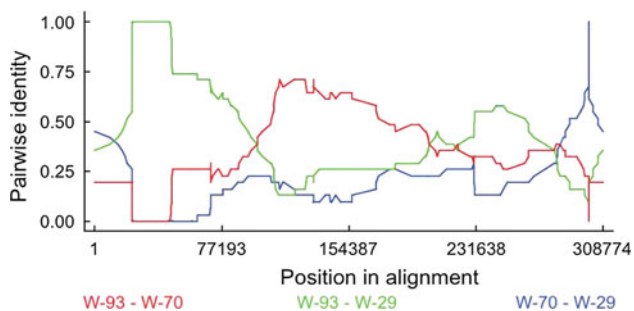
Furthermore, nine out of the 51 genes ($\sim 18\%$) showing the most evidence for positive selection have inferred putative functions. Most of these orthologs appear to be associated with the replication of WSSV, collagen-like protein, serine/threonine protein kinase, class I cytokine receptor, DNA metabolism, and transcription. For example, ORF27 encodes a DNA polymerase, ORF92 and ORF98 encode the large and small subunits of the ribonucleotide reductase, respectively, ORF171 encodes a chimeric thymidine kinase-thymidylate kinase, and ORF 149 putatively encodes a TATA box binding protein). Thus, WSSV seems to be under the influence of a balance of different selective forces at different regions and sites that display different functional constraints. Similar results have been described previously for other viruses. In a recent study, it was found that one gene (*tat*) of simian immunodeficiency virus exhibits positive selection, while the overlapping gene (*vpr*) shows signs of strong purifying selection [32].

Characterization of potential recombination events

A recombination detection analysis using RDP3 identified three potential events. The genome segment affected by the first putative recombination event starts at positions 23,227 and ends at position 44,587. It has been reported previously that this part of the genome includes both a highly variable region, which is located at position 22,961–23,619 in the W-29 isolate, and a genomic deletion when compared with the W-70 and W-93 isolates [33]. In addition, it has been suggested that the chimeric thymidine kinase (TK) and thymidylate kinase (TMK) genes were incorporated into the WSSV genome via homologous recombination [34]. Moreover, the vast majority of the informative sites on which the respective recombination signals were based lie in these variable regions. Since RDP assumes that differences between sequences arise from independent point mutations, the first recombination event was discarded as unrealistic. As the remaining two recombination events were located in regions of low variability and few deletions,

Table 2 The average p-values of three recombination events occurring in WSSV calculated by six different recombination detection methods

Event	RDP	GENOCOV	BOOTSCAN	MaxChi	CHIMAERA	3Seq
I	1.7×10^{-89}	3.2×10^{-77}	3.4×10^{-64}	1.1×10^{-41}	1.3×10^{-18}	1.4×10^{-91}
II	2.9×10^{-11}	-	2.6×10^{-8}	-	9.9×10^{-6}	1.4×10^{-9}
III	1.2×10^{-34}	6.2×10^{-43}	1.2×10^{-34}	5.8×10^{-19}	4.8×10^{-14}	4.2×10^{-40}

**Fig. 1** Identification of recombination events II and III occurring in the WSSV genome, using the results of bootscan analysis for the recombination origin on the basis of pairwise distance, modeled with a window size of 200, a step size of 20 and 100 bootstrap replicates

and as low p-values were achieved during the analysis (Table 2), they were considered accurate and trustworthy.

Based on the findings of RDP and the distribution of sequence similarity (Fig. 1), it was concluded that the following recombination events (event II and event III) are the most plausible: W-93 resulted from a recombination of W-29, W-70 and an unknown WSSV sequence (probably an ancestral WSSV variant) in which the segments (measured in reference to the alignment used) started at position 1 to 96,803 and 201,825–280,000 stem from W-29, while the segment 96,803–201,825 was derived from W-70, and the segment 280,000–end originated from an unknown sequence. Furthermore, according to the results obtained in this study, it is proposed that W-29 resulted from a recombination of W-70 and an unknown WSSV isolate in which the segment starting at position 1–285,484 stems from an unknown sequence and the segment 285,484–end originated from WSSV-70. These results contrast with those reported recently, in which a model of gradual WSSV genome shrinkage has been proposed [35].

Accordingly, the WSSV genome has been shrinking by removing some variable regions while at the same time its virulence has increased due to a faster replication of a smaller genome. Thus, the suggested expansion of the WSSV genome is not paradoxical. A genome is a collection of genes that controls and coordinates the essential functions of an organism through a dynamic interaction of its elements [36]. Thus, it is clear that a virus containing a small genome will depend extensively on the host cell as a provider of the elements needed for its replication. It is also clear that a reduction in the genome size may confer some

evolutionary advantages to the virus. However, a virus genome reduction is not necessarily a straightforward process. According to the results obtained in the present study, a reduction in the WSSV genome indeed occurred early during its evolution; however, successive recombination events have caused an increase in the genome size. Similar findings involving recombinational events during viral genome size increase have been reported previously, suggesting that some genome components of the geminiviruses may have experienced homologous and non-homologous recombination events that finally caused a size increase [37]. If this hypothesis is proven correct, it may be clear that the WSSV genome is non-static, and recombination is certainly an important force in its evolution, conferring an outstanding ability to adapt to any given environment.

Acknowledgments XMW thanks ShenYang Agricultural University for computational facilities. Part of this work was funded by the Consejo Nacional de Ciencia y Tecnología (CONACyT), México, for grant 102744 (to ASP). Thanks are also due to the supportive staff of the Laboratorio de Sanidad Acuicola (CIBNOR, Hermosillo), particularly to MVZ Fernando Mendoza, Daniel Coronado Molina and to Dr. Adriana Muhlia for careful reading and critical review of this manuscript.

References

- De La Peña LD, Lavilla-Pitogo CR, Villar CBR, Paner MG, Sombito CD, Capulos GC (2007) Prevalence of white spot syndrome virus (WSSV) in wild shrimp *Penaeus monodon* in the Philippines. *Dis Aquat Organ* 77:175–179
- Vlak JM, Bonami JR, Flegel TW, Kou GH, Lightner DV, Lo CF, Loh PC, Walker PW (2004) Nimaviridae. In: VIIIth report of the International Committee on Taxonomy of Viruses. Elsevier, Amsterdam, The Netherlands, pp 187–192
- Chou HY, Huang CY, Lo CF, Kou GH (1998) Studies on transmission of white spot syndrome associated baculovirus (WSBV) in *Penaeus monodon* and *P. japonicus* via waterborne contact and oral ingestion. *Aquaculture* 164:263–276
- Marks H, van Duijse JJA, Zuidema D, van Hulten MCW, Vlak JM (2005) Fitness and virulence of an ancestral white spot syndrome virus isolate from shrimp. *Virus Res* 110:9–20
- Wang CH, Lo CF, Leu JH, Chou CM, Yeh PY, Chou HY, Tung MC, Chang CF, Su MS, Kou GH (1995) Purification and genomic analysis of baculovirus associated with white spot syndrome (WSBV) of *Penaeus monodon*. *Dis Aquat Org* 23:239–242
- van Hulten MCW, Witteveldt J, Peters S, Kloosterboer N, Turchini R, Fiers M, Sandbrink H, Lankhorst RK, Vlak JM (2001) The white spot syndrome virus DNA genome sequence. *Virology* 286:7–22

7. Yang F, He J, Lin X, Li Q, Pan D, Zhang X, Xu X (2001) Complete genome sequence of the shrimp white spot bacilliform virus. *J Virol* 75:11811–11820
8. Sharp PM, Emery LR, Zeng K (2010) Forces that influence the evolution of codon bias. *Philos Trans R Soc Lond* 365B: 1203–1212
9. Ding J, Doorbar J, Li B, Zhou F, Gu W, Zhao L, Saunders NA, Frazer IH, Zhao KN (2010) Expression of papillomavirus L1 proteins regulated by authentic gene codon usage is favoured in G2/M-like cells in differentiating keratinocytes. *Virology* 399: 46–58
10. Wong E, Smith D, Rabadan R, Peiris M, Poon L (2010) Codon usage bias and the evolution of influenza A viruses. *Codon usage biases of influenza virus*. *BMC Evol Biol* 10:253
11. Harrison RJ, Charlesworth B (2011) Biased gene conversion affects patterns of codon usage and amino acid usage in the *Saccharomyces sensu stricto* group of yeasts. *Mol Biol Evol* 28:117–129
12. Chiari Y, Dion K, Colborn J, Parmakelis A, Powell JR (2010) On the possible role of tRNA base modifications in the evolution of codon usage: queuosine and *Drosophila*. *J Mol Evol* 70:339–345
13. Liu X, Wu C, Chen AYH (2010) Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. *Arch Virol* 155:685–693
14. Froissart R, Roze D, Uzest M, Galibert L, Blanc S, Michalakakis Y (2005) Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biol* 3:e89
15. Martin DP, Van Der Walt E, Posada D, Rybicki EP (2005) The evolutionary value of recombination is constrained by genome modularity. *PLoS Genet* 1:e51
16. Domingo E (2010) Mechanisms of viral emergence. *Vet Res* 41:38
17. Worobey M, Holmes EC (1999) Evolutionary aspects of recombination in RNA viruses. *J Gen Virol* 80:2535–2543
18. Lin YCJ, Evans DH (2010) Vaccinia virus particles mix inefficiently, and in a way that would restrict viral recombination, in coinfecting cells. *J Virol* 84:2432–2443
19. Lefevre P, Lett JM, Varsani A, Martin DP (2009) Widely conserved recombination patterns among single-stranded DNA viruses. *J Virol* 83:2697–2707
20. Chen R, Holmes EC (2006) Avian influenza virus exhibits rapid evolutionary dynamics. *Mol Biol Evol* 23:2336–2341
21. Wright F (1990) The “effective number of codons” used in a gene. *Gene* 87:23–29
22. Sablok G, Nayak K, Vazquez F, Tatarinova TV (2011) Synonymous codon usage, GC₃ and Evolutionary patterns across plastomes of three pooid model species—Emerging grass genome models for monocots. *Mol. Biotechnol* 49:116–128
23. Sharp PM, Li WH (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for “rare” codons. *Nucleic Acids Res* 14:7737–7749
24. Grocock RJ, Sharp PM (2001) Synonymous codon usage in *Cryptosporidium parvum*: identification of two distinct trends among genes. *Int J Parasitol* 31:402–412
25. Sánchez-Paz A (2010) White spot syndrome virus: an overview on an emergent concern. *Vet Res* 41:43
26. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
27. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462
28. 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
29. Aragonès L, Guix S, Ribes E, Bosch A, Pintó RM (2010) Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS pathog* 6:e1000797
30. Irwin B, Heck JD, Hatfield G (1995) Codon pair utilization biases influence translational elongation step times. *J Biol Chem* 270:22801–22806
31. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol* 80:9687–9696
32. Hughes AL, Westover K, Da Silva J, O’Connor DH, Watkins DI (2001) Simultaneous positive and purifying selection on overlapping reading frames of the tat and vpr genes of simian immunodeficiency virus. *J Virol* 75:7966–7972
33. Marks H, Goldbach RW, Vlask JM, Van Hulten MCW (2004) Genetic variation among isolates of white spot syndrome virus. *Arch Virol* 149:673–697
34. Tsai MF, Yu HT, Tzeng HF, Leu JH, Chou CM, Huang CJ, Wang CH, Lin JY, Kou GH, Lo CF (2000) Identification and characterization of a shrimp white spot syndrome virus (WSSV) gene that encodes a novel chimeric polypeptide of cellular-type thymidine kinase and thymidylate kinase. *Virology* 277:100–110
35. Zwart MP, Dieu BTM, Hemerik L, Vlak JM (2010) Evolutionary trajectory of white spot syndrome virus (WSSV) genome shrinkage during spread in Asia. *PLoS One* 5:e13400
36. Rokyta D, Badgett MR, Molineux IJ, Bull JJ (2002) Experimental genomic evolution: extensive compensation for loss of DNA ligase activity in a virus. *Mol Biol Evol* 19:230–238
37. Gilbertson RL, Sudarshana M, Jiang H, Rojas MR, Lucas WJ (2003) Limitations on geminivirus genome size imposed by plasmodemesmata and virus-encoded movement protein: insights into DNA trafficking. *Plant Cell* 15:2578–2591