

RESEARCH

Open Access



# Analysis of the codon usage pattern of the RdRP gene of mycovirus infecting *Aspergillus* spp.

Mikyung Je<sup>1,2</sup>, Hayeon Kim<sup>3</sup> and Hyeon S. Son<sup>1,2\*</sup> 

## Abstract

**Background:** Mycoviruses that infect fungi generally do not have a significant effect on the host and, instead, reduce the toxicity of the fungi. However, recent studies have shown that polmycovirus-1, a mycovirus that infects *Aspergillus* species known to cause disease in humans, is related to increased virulence of the fungus.

**Methods:** Comparative analysis was performed of RdRP gene codon usage patterns of *Aspergillus fumigatus* polmycovirus-1 (AfuPmV-1) and other mycoviruses known to infect *Aspergillus* spp. to examine the genetic characteristics of AfuPmV-1. In addition, codon usage analysis was performed to determine whether the nucleotide composition and codon usage characteristics of AfuPmV-1 were also present in other polmycoviruses and hypervirulence-related mycoviruses. Phylogenetic analysis was also performed to investigate their evolutionary relationship.

**Results:** Analysis of nucleotide composition indicated that AfuPmV-1 had the highest GC content among analyzed mycoviruses and relative synonymous codon usage analysis indicated that all of the codons preferred by AfuPmV-1 ended with C or G, while codons ending with A or U were not observed. Moreover, the effective number of codons, the codon adaptation index, and correspondence analysis showed that AfuPmV-1 had greater codon preference compared with other mycoviruses and that AfuPmV-1 had relatively high adaptability to humans and fungi. These results were generally similar among polmycoviruses.

**Conclusions:** The codon usage pattern of AfuPmV-1 differs from other mycoviruses that infect *Aspergillus* spp. This difference may be related to the hypervirulence effect of AfuPmV-1. Analysis of AfuPmV-1 codon usage patterns could contribute to the identification and prediction of virulence effects of mycoviruses with similar genetic characteristics.

**Keywords:** Mycovirus, Polmycovirus, AfuPmV-1, Codon usage pattern, RSCU

## Background

Mycoviruses are viruses that infect fungi and are known to be infectious to most fungal species. While mycoviruses have little or no influence on the fungal host in most cases, some have been shown to control the pathogenicity of the host by increasing or decreasing the virulence [1, 2]. To date, research on mycoviruses has mainly focused on plant pathogenic fungi. The hypovirulence

effect of mycoviruses in fungi has led to the development of biological control agents to reduce the virulence of fungi on important crops and plant resources [2, 3]. There have been relatively few studies on the hypervirulence effect of mycoviruses, although some results have been reported recently. In 2015, the A78 virus, which infects the human pathogenic fungus, *Aspergillus fumigatus*, was reported to have significant mild hypervirulence effects in the moth species, *Galleria mellonella* [4], and similar virulence effects have been reported in *Aspergillus fumigatus* tetramycovirus-1 (AfuTmV-1) found in the same fungus [5]. AfuTmV-1 was renamed AfuPmV-1, and its virus family name was changed from *Tetramycoviridae* to *Polmycoviridae* as more viruses showing similar characteristics to AfuTmV-1 were discovered, including BbPmV-1,

\* Correspondence: [hss2003@snu.ac.kr](mailto:hss2003@snu.ac.kr)

<sup>1</sup>SNU Bioinformatics Institute, Interdisciplinary Graduate Program in Bioinformatics, College of Natural Science, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea

<sup>2</sup>Laboratory of Computational Biology & Bioinformatics, Institute of Public Health and Environment, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea

Full list of author information is available at the end of the article



**Table 1** The genome of polymycovirus-1 (AfuTmV-1)

Host	Segment (bp)	ORF		Product	
		Gene	Length	Protein	Length (aa)
<i>Aspergillus fumigatus</i>	dsRNA1 (2403)	RdRP ORF	2292 nt	RNA dependent RNA polymerase	763
	dsRNA2 (2233)	unknown_ORF	2091 nt	hypothetical protein	696
	dsRNA3 (1970)	Methyl transferase ORF	1845 nt	methyl transferase	614
	dsRNA4 (1131)	PAS-rp ORF	840 nt	proline/alanine/serine (PAS)-rich protein	279

which infects the insect pathogenic fungus, *Beauveria bassiana* [6] (Table 1).

*Aspergillus* is a fungus that belongs to the phylum *Ascomycota* and is related to various human diseases through infection, including aspergillosis on infection of the lungs, asthma, and allergies. Moreover, it is one of the fatal human pathogenic fungi, with a high mortality rate in patients with low immunity by opportunistic infection [7, 8]. Although studies have focused on polymycovirus-1 with regard to the increase in fungal pathogenicity in *A. fumigatus*, it is also important to examine other mycoviruses, such as partitivirus, chrysovirus, and victorivirus, which are infectious to other species in the genus *Aspergillus*, including *A. fumigatus*, *A. foetidus*, *A. ochraceus*, and *A. niger*. This study was performed to explore other mycoviruses that may increase the pathogenicity of the genus *Aspergillus* by comparison with polymycovirus-1, and to investigate the genetic characteristics of mycoviruses that enhance the toxicity of human pathogenic fungi.

## Methods

### Sequence data collection

The AfuPmV-1 genome consists of four capped double-stranded RNAs (dsRNAs), of which the largest dsRNA1 segment encodes the RNA-dependent RNA polymerase (RdRP) [5, 6]. The data were downloaded from the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genbank>) and the complete nucleotide sequences of the RdRP coding region of mycoviruses that infect *Aspergillus* species were used. Eight sets of data published between 2006 and 2015 were studied in the analysis. Other polymycoviruses and hypervirulence-related mycoviruses data were downloaded from NCBI for comparison (Table 2).

### Codon usage analysis

The programs CodonW (<https://sourceforge.net/projects/codonw/>) and CALcal (<http://genomes.urv.es/CALcal/>) were used to conduct analyses of nucleotide composition, overall/local G + C content, relative synonymous codon

**Table 2** Summary of datasets

No.	Accession no.		Fungal host	Mycovirus	Abbr.	Length		Ref.
	Nucleotide	Protein				Nucleotide (nt)	Protein (aa)	
1	FN376847	CAY25801	<i>Aspergillus fumigatus</i>	partitivirus-1	AfuPV-1	1629	542	[22]
2	FN178512	CAX48749	<i>Aspergillus fumigatus</i>	chrysovirus	AfuCV	3345	1114	[23]
3	HG975302	CDP74618	<i>Aspergillus fumigatus</i>	polymycovirus-1 <sup>b</sup>	AfuPmV-1	2292	763	[5]
4	HE588144	CCD33020	<i>Aspergillus foetidus</i>	alternavirus <sup>c</sup>	AfAltV	3375	1124	[24]
5	HE588147	CCD33024	<i>Aspergillus foetidus</i>	victorivirus <sup>d</sup>	AfW	2520	839	[25]
6	HE588148	CCD33025	<i>Aspergillus foetidus</i>	exartavirus <sup>e</sup>	AfExV	2889	962	[26]
7	DQ270031	ABC86749	<i>Aspergillus ochraceus</i>	partitivirus	AoPV	1620	539	[27]
8	EU289897 <sup>a</sup>	ABX79997	<i>Aspergillus niger</i>	alternavirus	AniAltV	3375	1124	[28]
9	MF444217	AWY10945	<i>Sclerotinia sclerotiorum</i>	tetramycovirus-1	SstRV1	2403	800	[29]
10	MF317878	AVZ65983	<i>Penicillium digitatum</i>	polymycoviruses 1	PdPmV1	2283	760	[30]
11	LN896307	CUS18595	<i>Beauveria bassiana</i>	polymycovirus 1	BbPmV-1	2328	775	[6]
12	LN896311	CUS18599	<i>Beauveria bassiana</i>	polymycovirus 2	BbPmV-2	2304	767	[6]
13	LC350277	BBC27878	<i>Alternaria alternata</i>	chrysovirus 1	AaCV1	3354	1174	[31]
14	KM235317	AKF14154	<i>Penicillium marneffeii</i>	partitivirus-1	TmPV1	1638	545	[32]

No. 1~8: mycoviruses infecting *Aspergillus* (Group 1); No. 9~12: other polymycoviruses (Group 2); No. 13~14: hypervirulence-related mycoviruses (Group 3)

<sup>a</sup>putative RdRP gene

<sup>b</sup>originally named *Aspergillus fumigatus* tetramycovirus-1

<sup>c</sup>proposed name of *Aspergillus foetidus* fast virus

<sup>d</sup>proposed name of *Aspergillus foetidus* slow virus 1

<sup>e</sup>proposed name of *Aspergillus foetidus* slow virus 2

usage (RSCU), effective number of codons (ENC), codon adaptation index (CAI), and correspondence analysis (COA) for each of the selected genetic data [9, 10]. Nucleotide composition analysis was conducted based on the overall frequency of nucleic acid occurrence (A%, C%, U%, and G%), total AU%/GC%, frequency of the third nucleic acid in synonymous codons (A3s, C3s, U3s, and G3s), and the GC3s values. The RSCU is a value represented as the ratio between predicted and observed usage rates of specific codons, assuming that all synonymous codons were used equally for an amino acid. An RSCU value of 1.0 indicates that all codons were used randomly or equally. The value would be greater (less) than 1.0 if certain codons were used more (less) frequently [11, 12]. In particular, codons with RSCU values  $\geq 1.6$  are considered over-represented codons, while those with RSCU values  $\leq 0.6$  are considered under-represented codons [13, 14]. The ENC is a simple and absolute method to measure codon usage bias in genes and genomes. The GC3s values are used in the calculation, and the ENC value ranges from 20 to 61. The ENC is inversely proportional to codon usage bias, and lower ENC values indicate stronger codon usage bias. A gene that does not show any codon usage bias would have an ENC value of 61. On the other hand, a gene that shows a high degree of codon usage bias would have an ENC value of 20, indicating that only one codon was used for an amino acid. Generally, ENC values  $< 35$  are interpreted as showing strong codon usage bias [15, 16]. The CAI is the geometric mean of relative adaptation level and is a quantitative method to calculate the differences in codon usage bias against highly expressed known reference data. The CAI value ranges between 0 and 1, with values closer to 1 indicating a high degree of similarity between the reference data and the

codon usage pattern and expression level, whereas smaller values indicate lower similarity in the codon usage pattern and expression level [17]. The reference data of codon usage levels in humans (*Homo sapiens*) and fungus (*Aspergillus fumigatus*) were acquired from the codon usage database (<http://www.kazusa.or.jp/codon/>) for comparison [18]. Furthermore, the RSCU values were examined by the COA method with the XLSTAT 2016 program for visualization of the CodonW results. The COA method is the preferred method in the area of multivariate analysis and represents the data as vectors consisting of rows and columns [19]. Each individual data of the RdRP codon region was represented as a vector with 59 dimensions covering 59 codons except methionine (AUG) and tryptophan (UGG), which lack synonymous codons, as well as termination codons. A comparative analysis was performed including three groups: mycoviruses infecting *Aspergillus* (Group 1), polycycoviruses with a fungal host other than *Aspergillus* (Group 2), and newly reported mycoviruses that enhance host virulence (Group 3).

**Phylogenetic analysis**

The phylogenetic relationships among the mycoviruses used for the analysis were inspected and phylogenetic analysis was conducted using the MEGA7 program (<http://www.megasoftware.net>) to infer the influence of evolutionary processes on codon usage patterns. RdRP coding sequences were aligned with the MUSCLE algorithm, and the phylogenetic tree was constructed by applying the Maximum Likelihood method and the Kimura 2-parameter substitution model. The robustness of the tree was verified with the bootstrap value set to 1000 [20].

**Table 3** Nucleotide compositions of the RdRP genes of *Aspergillus*-infecting mycoviruses

No.	A%	C%	U%	G%	AU%	GC%	A3s	C3s	U3s	G3s	GC3s	CAI <sup>H</sup>	CAI <sup>A</sup>	ENC
1	24.37	23.70	25.11	26.83	49.48	50.52	25.97	31.76	35.53	32.69	50.30	0.729	0.741	56.58
2	27.53	20.00	24.30	28.16	51.84	48.16	29.39	32.19	29.57	38.74	53.50	0.740	0.709	56.73
3	18.63	35.99	18.63	26.75	37.26	62.74	9.29	61.22	16.18	32.01	78.30	0.762	0.843	40.67
4	19.53	23.56	26.31	30.61	45.84	54.16	16.59	27.75	38.85	38.50	53.10	0.704	0.744	52.69
5	21.15	29.25	22.38	27.22	43.53	56.47	17.26	43.42	25.10	33.17	63.60	0.715	0.760	54.29
6	29.15	22.53	30.53	17.79	59.67	40.33	34.51	24.81	48.25	16.90	32.80	0.714	0.676	51.02
7	25.37	22.10	27.41	25.12	52.78	47.22	29.33	28.84	41.61	26.55	43.00	0.699	0.722	54.79
8	19.53	23.56	26.37	30.55	45.90	54.10	16.35	27.78	39.00	38.62	53.20	0.705	0.745	52.40
9	19.10	30.50	20.93	29.46	40.03	59.97	24.74	43.63	16.34	33.90	64.90	0.746	0.799	49.02
10	19.45	29.00	20.37	31.19	39.82	60.18	10.88	45.16	20.78	43.85	73.20	0.733	0.804	49.17
11	19.16	26.76	20.45	33.63	39.60	60.40	22.78	34.56	13.11	51.31	69.60	0.710	0.770	49.68
12	16.84	33.12	21.18	28.86	38.02	61.98	25.53	49.25	6.72	35.04	71.60	0.749	0.832	47.34
13	21.65	23.55	22.72	32.08	44.36	55.64	20.83	29.46	27.99	44.99	59.80	0.696	0.736	55.19
14	24.42	22.41	25.70	27.47	50.12	49.88	27.69	29.63	36.81	30.67	47.60	0.691	0.724	58.02

CAI<sup>H</sup>: compare with *Homo sapiens* as reference set

CAI<sup>A</sup>: compare with *Aspergillus fumigatus* as reference set

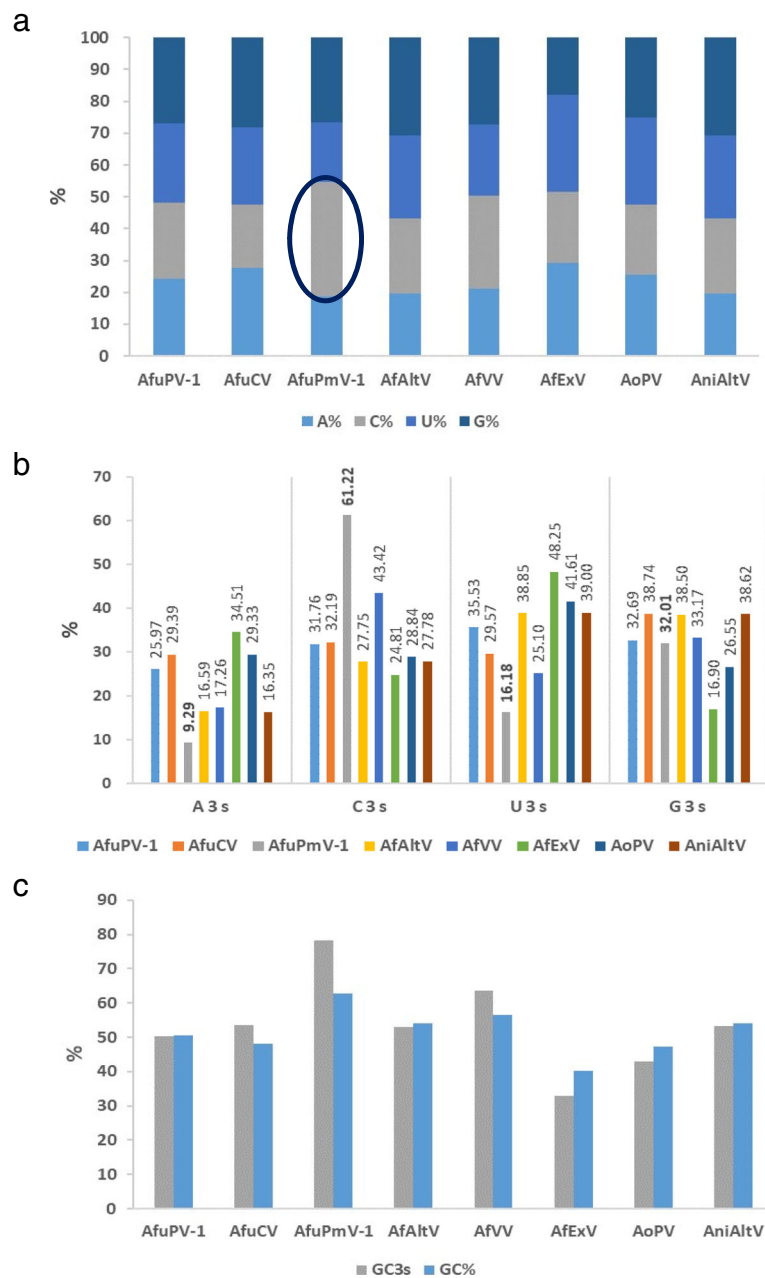
The dotted line shows the result of AfuPmV-1 analysis, and the solid line shows the result of analysis of polycycoviruses

**Results**

**Nucleotide composition features**

Basic nucleotide composition analysis was conducted for the RdRP gene of mycoviruses infecting *Aspergillus* spp. (Table 3, Fig. 1). The mean and standard deviation (SD) of A%, C%, U%, and G% were  $23.16 \pm 4.00$ ,  $25.09 \pm 5.13$ ,  $25.13 \pm 3.54$ , and  $26.63 \pm 4.04$ , respectively. The mean and SD of total AU% and GC% were  $48.29 \pm 6.76$  and  $51.71 \pm 6.76$ , respectively. The mean and SD of A3s, C3s, U3s, G3s, and GC3s were  $22.34 \pm 8.66$ ,  $34.72 \pm 12.09$ ,

$34.26 \pm 10.20$ ,  $32.15 \pm 7.47$ , and  $53.48 \pm 13.45$ , respectively. AfuPmV-1 showed the lowest frequencies of nucleotides A and U and the highest frequency of nucleotide C among the *Aspergillus*-infecting mycoviruses included in the analysis. Similarly, the frequencies of the four types of nucleotides in the third position of synonymous codons were the lowest with A3s and U3s and highest with C3s. As nucleotide C occurred relatively abundantly in AfuPmV-1, the GC content and GC3s indices were highest in this virus with values of 62.74 and 78.30,



**Fig. 1** Nucleotide composition features. **a** Nucleotide distribution of A, C, U, and G in the RNA-dependent RNA polymerase (RdRp) gene. **b** Nucleotide distribution frequency calculated only for the third codon base. **c** Analysis of GC content at all codon positions (GC%) and at the third position (GC3s)

respectively (Fig. 1). The GC content of AfuPmV-1 was reported to be approximately 63% in a previous study of the entire or partial genome sequences of *Polymyoviridae* viruses [6]. As a result of the comparison, Group 2 showed a similar pattern to that of AfuPmV-1. The GC content was between 59.97 and 61.98%, and the frequencies of nucleotides at the third position in the synonymous codons were higher in C3s and G3s than in A3s and U3s, respectively. Among these, BbPmV-1 is an experimentally reported virus that may be related to mild hypervirulence. The Group 3 results differed from those of Groups 1 and 2 (Table 3).

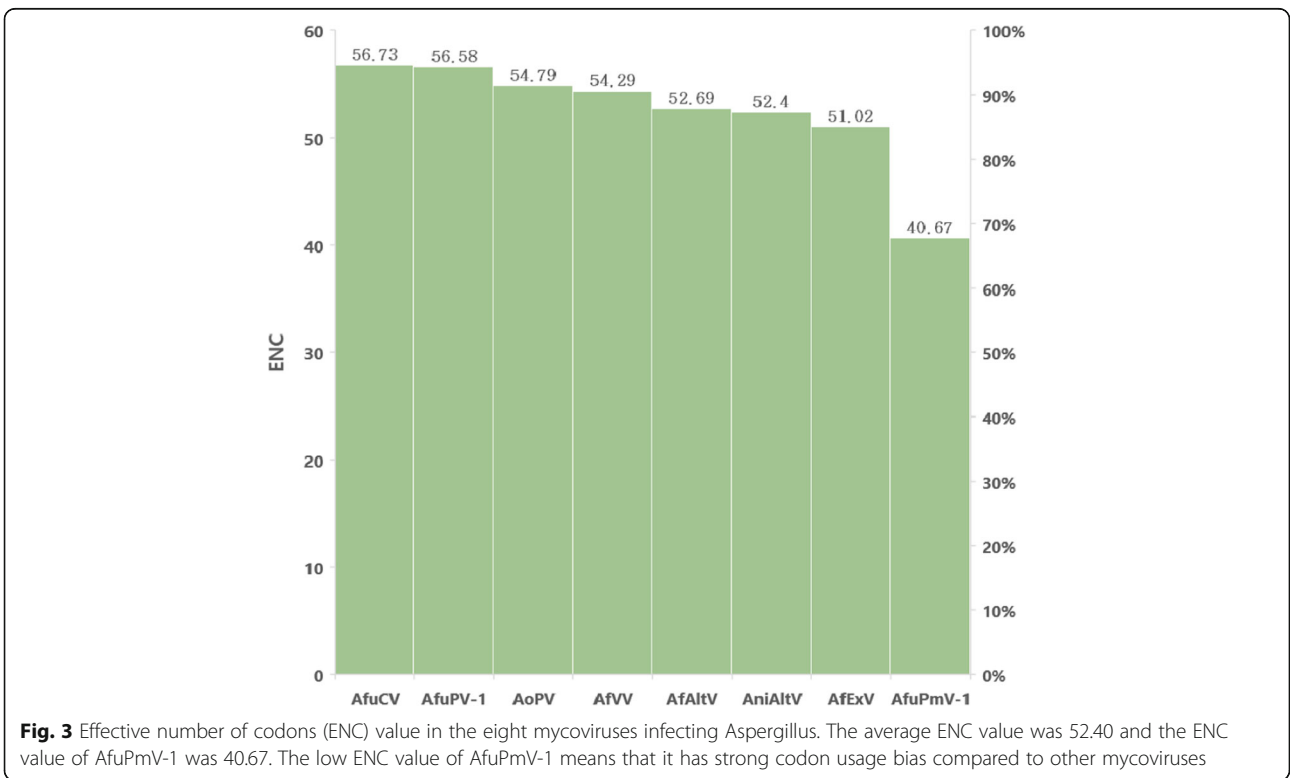
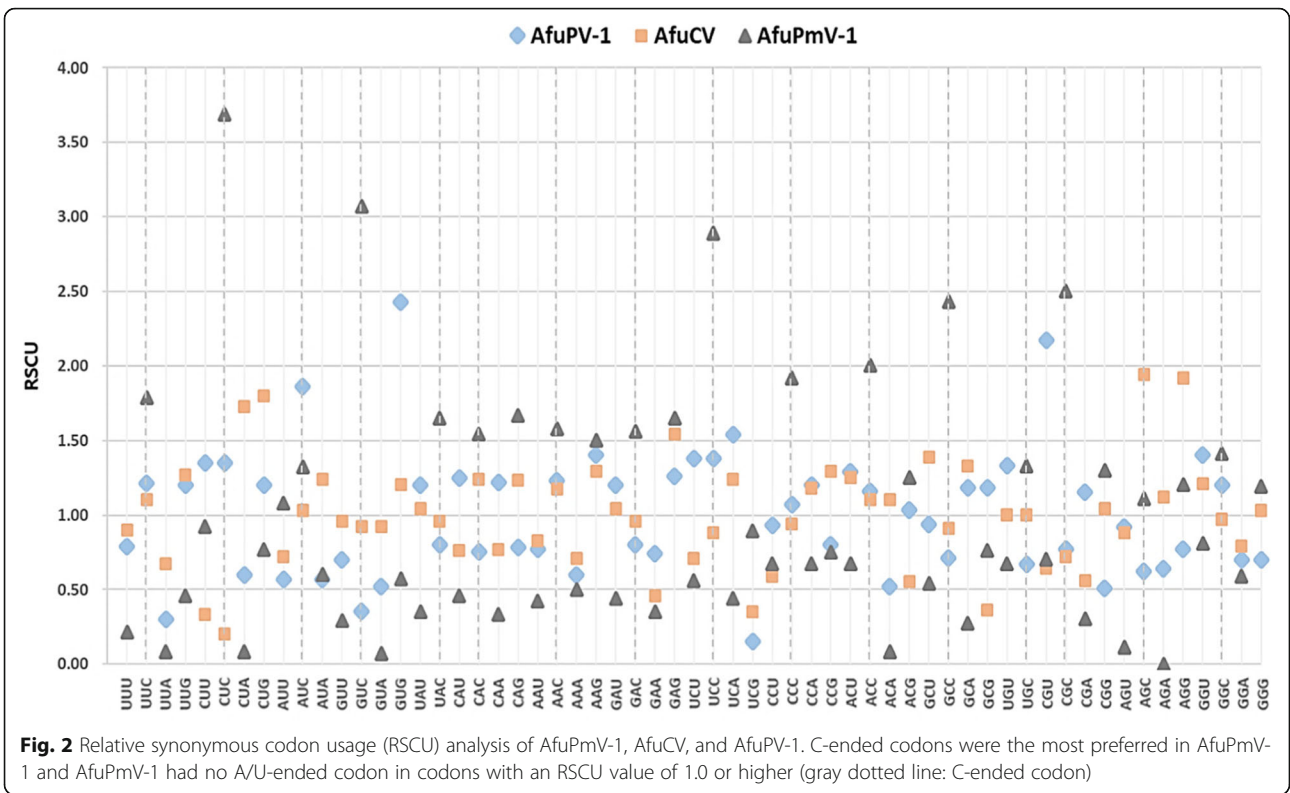
**RSCU value and codon usage preference**

The RSCU values of AfuPmV-1, AfuCV, and AfuPV-1 of Group1 were compared to inspect the codon usage bias according to virus species (Table 4, Fig. 2). Of codons related to the entire 18 amino acids, 18 codons were preferred in AfuPmV-1 of which 11 showed RSCU values  $\geq 1.6$ . AfuCV showed 17 preferred codons, three of which had RSCU values  $\geq 1.6$ . AfuPV had 20 preferred codons, three of which showed RSCU values  $\geq 1.6$ . AfuPmV-1 showed similarities to AfuCV in codons CAC (His), CAG (Gln) and to AfuPV-1 in codons CUC (Leu) and AUC (Ile). The codons preferred by each virus

**Table 4** RSCU analysis of AfuPmV-1, AfuCV, and AfuPV-1

AA	Codon	AfuPmV-1	AfuCV	AfuPV-1	AA	Codon	AfuPmV-1	AfuCV	AfuPV-1
Phe	UUU	0.21 <sup>a</sup>	0.90	0.79	Ser	UCU	0.56 <sup>a</sup>	0.71	1.38
	UUC	<b>1.79<sup>b</sup></b>	<b>1.10</b>	<b>1.21</b>		UCC	<b>2.89<sup>b</sup></b>	0.88	1.38
Leu	UUA	0.08 <sup>a</sup>	0.67	0.30	Pro	UCA	0.44 <sup>a</sup>	1.24	<b>1.54</b>
	UUG	0.46 <sup>a</sup>	1.27	1.20		UCG	0.89	0.35 <sup>a</sup>	0.15 <sup>a</sup>
	CUU	0.92	0.33 <sup>a</sup>	<b>1.35</b>		CCU	0.67	0.59 <sup>a</sup>	0.93
	CUC	<b>3.69<sup>b</sup></b>	0.20 <sup>a</sup>	<b>1.35</b>		CCC	<b>1.92<sup>b</sup></b>	0.94	1.07
	CUA	0.08 <sup>a</sup>	1.73	0.60 <sup>a</sup>		CCA	0.67	1.18	<b>1.20</b>
	CUG	0.77	<b>1.80<sup>b</sup></b>	1.20		CCG	0.75	<b>1.29</b>	0.80
Ile	AUU	1.08	0.72	0.57 <sup>a</sup>	Thr	ACU	0.67	<b>1.25</b>	<b>1.29</b>
	AUC	<b>1.32</b>	1.03	<b>1.86<sup>b</sup></b>		ACC	<b>2.00<sup>b</sup></b>	1.10	1.16
	AUA	0.60 <sup>a</sup>	<b>1.24</b>	0.57 <sup>a</sup>		ACA	0.08 <sup>a</sup>	1.10	0.52 <sup>a</sup>
Val	GUU	0.29 <sup>a</sup>	0.96	0.70	Ala	ACG	1.25	0.55 <sup>a</sup>	1.03
	GUC	<b>3.07<sup>b</sup></b>	0.92	0.35 <sup>a</sup>		GCU	0.54 <sup>a</sup>	<b>1.39</b>	0.94
	GUA	0.07 <sup>a</sup>	0.92	0.52 <sup>a</sup>		GCC	<b>2.43<sup>b</sup></b>	0.91	0.71
	GUG	0.57 <sup>a</sup>	<b>1.20</b>	<b>2.43<sup>b</sup></b>		GCA	0.27 <sup>a</sup>	1.33	<b>1.18</b>
Tyr	UAU	0.35 <sup>a</sup>	<b>1.04</b>	<b>1.20</b>	Cys	GCG	0.76	0.36 <sup>a</sup>	<b>1.18</b>
	UAC	<b>1.65<sup>b</sup></b>	0.96	0.80		UGU	0.67	1.00	<b>1.33</b>
His	CAU	0.46 <sup>a</sup>	0.76	<b>1.25</b>	Arg	UGC	<b>1.33</b>	1.00	0.67
	CAC	<b>1.54</b>	<b>1.24</b>	0.75		CGU	0.70	0.64	<b>2.17<sup>b</sup></b>
Gln	CAA	0.33 <sup>a</sup>	0.77	<b>1.22</b>		CGC	<b>2.50<sup>b</sup></b>	0.72	0.77
	CAG	<b>1.67<sup>b</sup></b>	<b>1.23</b>	0.78		CGA	0.30 <sup>a</sup>	0.56 <sup>a</sup>	1.15
Asn	AAU	0.42 <sup>a</sup>	0.83	0.77	Ser	CGG	1.30	1.04	0.51 <sup>a</sup>
	AAC	<b>1.58</b>	<b>1.17</b>	<b>1.23</b>		AGU	0.11 <sup>a</sup>	0.88	0.92
Lys	AAA	0.50 <sup>a</sup>	0.71	0.60 <sup>a</sup>	Arg	AGC	1.11	<b>1.94<sup>b</sup></b>	0.62
	AAG	<b>1.50</b>	<b>1.29</b>	<b>1.40</b>		AGA	0.00 <sup>a</sup>	1.12	0.64
Asp	GAU	0.44 <sup>a</sup>	<b>1.04</b>	<b>1.20</b>	Gly	AGG	1.20	<b>1.92<sup>b</sup></b>	0.77
	GAC	<b>1.56</b>	0.96	0.80		GGU	0.81	<b>1.21</b>	<b>1.40</b>
Glu	GAA	0.35 <sup>a</sup>	0.46 <sup>a</sup>	0.74		GGC	<b>1.41</b>	0.97	1.20
	GAG	<b>1.65<sup>b</sup></b>	<b>1.54</b>	<b>1.26</b>		GGA	0.59 <sup>a</sup>	0.79	0.70
							GGG	1.19	1.03

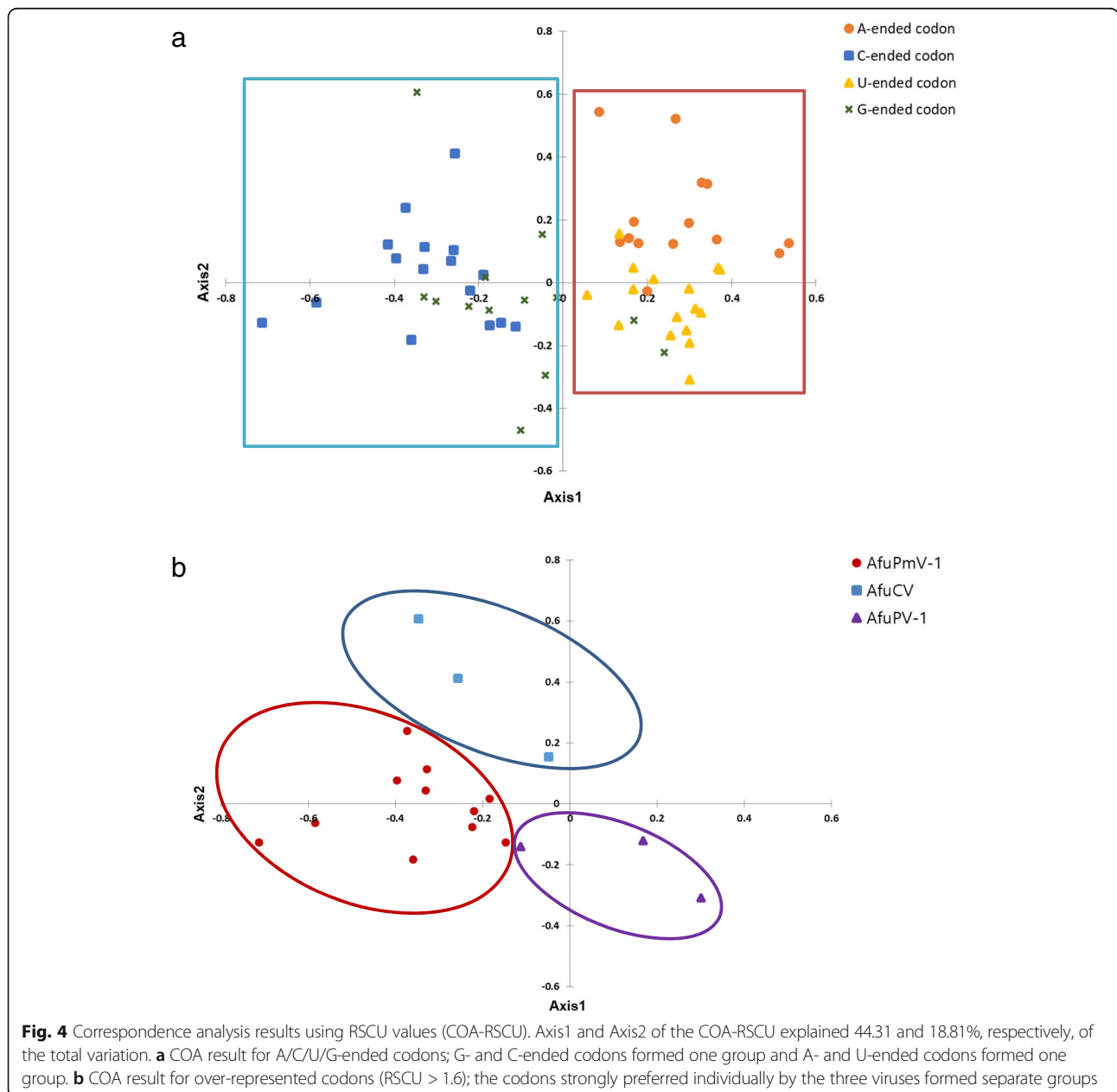
AA amino acid  
 Non-degenerate amino acids (Met, Trp) and stop codons are excluded  
 The boldface are the most preferred codons  
<sup>a</sup>RSCU value  $\leq 0.6$ ; under-represented codons  
<sup>b</sup>RSCU value  $\geq 1.6$ ; over-represented codons



individually were as follows. The codons solely preferred in AfuPmV-1 were GUC (Val), UAC (Tyr), GAC (Asp), UCC (Ser), CCC (Pro), ACC (Thr), GCC (Ala), UGC (Cys), CGC (Arg), and GGC (Gly) for 10 of the amino acids. Remarkably, all of these were C-ended codons (Fig. 2). Codons solely preferred in AfuCV were CUG (Leu), AUA (Ile), CCG (Pro), GCU (Ala), AGC (Ser), and AGG (Arg) for six of the amino acids, and in AfuPV-1, CUU (Leu), CAU (His), CAA (Gln), UCA (Ser), CCA (Pro), GCA (Ala), GCG (Ala), UGU (Cys), and CGU (Arg) for eight of the amino acids. The codons UUC (Phe), AAC (Asn), AAG (Lys), and GAG (Glu)

showed similar preferences in all three viruses. The RSCU values and end nucleotide composition indicated that, in the RdRP coding region, C-ended codons were strongly preferred in AfuPmV-1 (15 of 18), G-ended codons were preferred in AfuCV (7 of 17), and U-ended codons were preferred in AfuPV-1 (8 of 20). Interestingly, there were no A/U-ended codons among the preferred codons of AfuPmV-1, indicating that AfuPmV-1 has a codon bias toward C- and G-ended codons.

Group 2 showed a similar codon usage pattern to that of AfuPmV-1, which indicates a preference for the C/G-ended codon. Group 3 also preferred C- or G-ended



codons, but the four nucleic acids were distributed more evenly.

### General codon usage pattern

The ENC value was calculated to quantitatively measure the magnitude of RdRP gene codon usage bias of the eight mycoviruses infecting *Aspergillus* spp. An ENC value < 35 indicates a strong codon usage bias. The results showed that the ENC value was lowest for AfuPmV-1 (40.67) and the other viruses showed ENC values > 50 (Table 3, Fig. 3). Taking into account the results of previous studies on RNA viruses in which the ENC values of Zaire Ebola virus, Chikungunya virus, Hepatitis C virus, and West Nile virus were 57.23, 55.56, 52.62, and 53.81, respectively [21], AfuPmV-1 appeared to have stronger codon usage bias relative to other viruses. The CAI was calculated to compare the adaptability of synonymous genetic codon usage in mutually different individuals, and the codon usage patterns were considered similar to the reference individual with CAI values closer to 1. This study referred to the CAI values of *H. sapiens* and *A. fumigatus*, which have ranges of 0.699–0.762 and 0.676–0.843, respectively. The mean and SD were  $0.72 \pm 0.02$  and  $0.74 \pm 0.05$  for *H. sapiens* and *A. fumigatus*, respectively. Remarkably, AfuPmV-1 showed the highest values for both *H. sapiens* and *A. fumigatus* with values of 0.762 and 0.843, respectively. These results indicated that AfuPmV-1 has the greatest similarity to the reference data in codon usage pattern and expression level, and that it has higher adaptability to human and fungal hosts compared with other

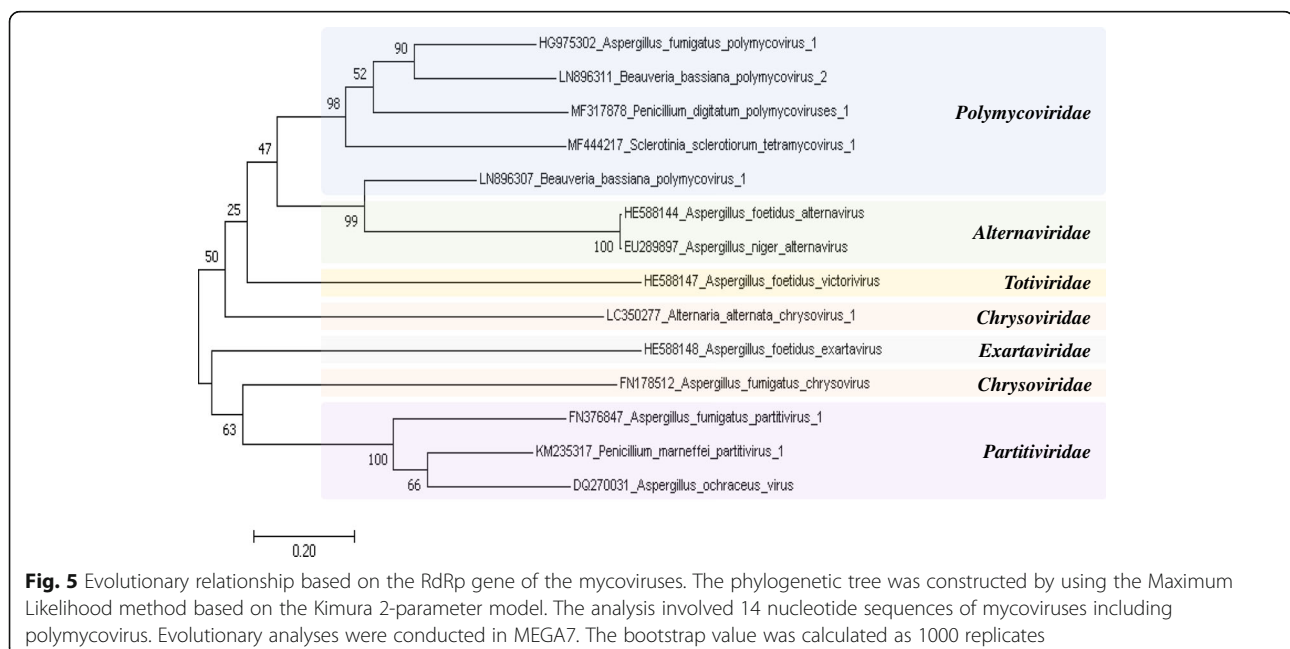
mycoviruses. In Group 2, CAI and ENC values were similar to those of AfuPmV-1 (Table 3).

### General trend of codon usage variation

To inspect the trends related to codon usage patterns of the *Aspergillus*-infecting viruses AfuPmV-1, AfuCV, and AfuPV-1, COA was performed with the RSCU values. Axis1, Axis2, Axis3, and Axis4 of the COA-RSCU explained 44.31, 18.81, 17.33, and 9.78% of the total variation, respectively. The results were based on 59 codons, excluding the three termination codons and methionine (AUG)/tryptophan (UGG) that do not have synonymous codons. Although there were exceptions in the COA results according to the third nucleotide in the codon, G- and C-ended codons formed one group and A- and U-ended codons formed another group (Fig. 4(a)). Moreover, the COA results for over-represented codons with RSCU values  $\geq 1.6$  showed that the codons strongly preferred individually by the three viruses formed separate groups (Fig. 4(b)). The observations verified that there were differences in the codon pattern preferences among the *Aspergillus*-infecting viruses.

### Evolutionary relationship between mycoviruses

A phylogenetic tree was constructed to examine the phylogenetic relationships among the 14 mycoviruses, including AfuPmV-1. Polymycoviruses were grouped with AfuPmV-1. Polymycoviruses showed a relatively close relationship with alternaviruses (Fig. 5). To provide more information, the RdRP family for each sequence was examined from Pfam. AfuPmV-1 was assigned to RdRP\_1 and all other polymycoviruses





**Table 5** RdRP family as assigned by Pfam

No.	Accession no.	Mycovirus	Family by Pfam	E-value
1	FN376847	<i>Aspergillus fumigatus</i> partitivirus-1	RdRP_1	1e-06
2	FN178512	<i>Aspergillus fumigatus</i> chrysovirus	RdRP_4	1.8e-73
3	HG975302	<i>Aspergillus fumigatus</i> polymycovirus-1	RdRP_1	3.9e-05
4	HE588144	<i>Aspergillus foetidus</i> alternavirus	RdRP_4	3.6e-07
5	HE588147	<i>Aspergillus foetidus</i> victorivirus	RdRP_4	1.2e-105
6	HE588148	<i>Aspergillus foetidus</i> exartavirus	RdRP_1	4.4e-10
7	DQ270031	<i>Aspergillus ochraceus</i> partitivirus	RdRP_1	7.7e-12
8	EU289897	<i>Aspergillus niger</i> alternavirus	RdRP_4	3.8e-07
9	MF444217	<i>Sclerotinia sclerotiorum</i> tetramycovirus-1	RdRP_1	0.00017
10	MF317878	<i>Penicillium digitatum</i> polymycoviruses 1	RdRP_1	4.4e-07
11	LN896307	<i>Beauveria bassiana</i> polymycovirus 1	RdRP_1	2.5e-07
12	LN896311	<i>Beauveria bassiana</i> polymycovirus 2	no match	–
13	LC350277	<i>Alternaria alternata</i> chrysovirus 1	RdRP_4	7.7e-44
14	KM235317	<i>Penicillium marneffeii</i> partitivirus-1	RdRP_1	4.1e-10

with similar codon patterns were assigned to the same RdRP\_1 (pfam00680). Other mycoviruses infecting *Aspergillus* were classified as RdRP\_1 or RdRP\_4, although they were found in the same fungus (Table 5). This result suggests that the hypervirulent effects of AfuPmV-1 may be more affected by viral genome characteristics than by the effect of *Aspergillus* as a host.

## Discussion

The mechanism underlying the hypervirulence effect of polymycoviruses has yet to be determined. However, experimental studies have demonstrated the existence of mycoviruses with mild hypervirulence effects, and other mycoviruses with similar sequences are continuously being discovered. The pathogenic effects of pathogenic fungi on the hosts may be increased by infection with mycoviruses that show hypervirulence effects. Therefore, it is necessary to determine the genetic characteristics of mycoviruses with hypervirulence effects. The results of the present study showed that AfuPmV-1 has a high GC content, and all of the strongly preferred codons (RSCU value  $\geq 1.6$ ) ended with either a C or G nucleotide. The distinctive codon usage pattern of AfuPmV-1 compared to other mycoviruses that infect *Aspergillus* spp. may be related to its hypervirulence effect. These characteristics did not appear in all mycoviruses with hypervirulent effects, but were shared by polymycoviruses. Nucleotide composition and codon usage patterns of polymycoviruses may be useful in predicting hypervirulent effects of unidentified mycoviruses.

## Conclusions

*Aspergillus* spp. are pathogenic fungi that cause various symptoms in humans. The hypervirulence effect of mycoviruses can increase the toxicity of *Aspergillus* spp. in human hosts, and thus increase the severity of symptoms. Here, AfuPmV-1 was shown to have distinct patterns in some codon usage indexes compared to other mycoviruses that infect *Aspergillus* spp. The distinctive codon usage pattern of AfuPmV-1 demonstrated in the present study indicated the need for monitoring of mycoviruses with similar characteristics. Research on mycoviruses has generally focused on their hypovirulence effects on fungi that infect plants. With the discovery of polymycoviruses, further research on the hypervirulence effects of mycoviruses is needed, particularly with regard to mechanisms of virulence control in mycoviruses, such as AfuPmV-1, which infects human pathogenic fungi.

## Abbreviations

AfuCV: *Aspergillus fumigatus* chrysovirus; AfuPmV-1: *Aspergillus fumigatus* polymycovirus-1; AfuPV-1: *Aspergillus fumigatus* partitivirus-1; AfuTmV-1: *Aspergillus fumigatus* tetramycovirus-1; BbPmV-1: *Beauveria bassiana* polymycovirus-1; CAI: Codon adaptation index; COA: Correspondence analysis; dsRNAs: Double-stranded RNAs; ENC: Effective number of codons; RdRP: RNA-dependent RNA polymerase; RSCU: Relative synonymous codon usage; SD: Standard deviation

## Acknowledgements

We'd like to thank those who made their invaluable data publicly available.

## Funding

This work was supported by a grant from the National Research Foundation of Korea funded by the Korea government (MSIP) (No. 2016R1C1B2015511) and the Ministry of Education (No. 2017R1D1A1B03033413).

## Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Authors' contributions**

HSS designed the study; MJ collected the data; MJ and HK analysed and interpreted the results; MJ and HK wrote the manuscript; All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>SNU Bioinformatics Institute, Interdisciplinary Graduate Program in Bioinformatics, College of Natural Science, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. <sup>2</sup>Laboratory of Computational Biology & Bioinformatics, Institute of Public Health and Environment, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. <sup>3</sup>Department of Biomedical Laboratory Science, Kyungdong University, 815 Gyeonhwon-ro, Munmak, Wonju, Gangwondo 24695, Korea.

Received: 22 August 2018 Accepted: 1 January 2019

Published online: 16 January 2019

**References**

- Ghabrial SA, Suzuki N. Viruses of plant pathogenic fungi. *Annu Rev Phytopathol.* 2009;47:353–84.
- Ghabrial SA, Caston JR, Jiang D, Nibert ML, Suzuki N. 50-plus years of fungal viruses. *Virology.* 2015;479:356–68.
- Pearson MN, Beever RE, Boine B, Arthur K. Mycoviruses of filamentous fungi and their relevance to plant pathology. *Mol Plant Pathol.* 2009;10:115–28.
- Ozkan S, Coutts Robert HA. *Aspergillus fumigatus* mycovirus causes mild hypervirulent effect on pathogenicity when tested on *Galleria mellonella*. *Fungal Genet Biol.* 2015;76:20–6.
- Kanhayuwa L, Kotta-Loizou I, Ozkan S, Gunning AP, Goutts Robert HA. A novel mycovirus from *Aspergillus fumigatus* contains four unique dsRNAs as its genome and is infectious as dsRNA. *Proc Natl Acad Sci.* 2015;112:9100–5.
- Kotta-Loizou I, Coutts Robert HA. Studies on the virome of the entomopathogenic fungus *Beauveria bassiana* reveal novel dsRNA elements and mild hypervirulence. *PLoS Pathog.* 2017. <https://doi.org/10.1371/journal.ppat.1006183>.
- Bennett JW. An overview of the genus *Aspergillus*. In: Machida M, Gomi K, editors. *Aspergillus: molecular biology and genomics*. Poole: horizon scientific press; 2010. p. 1–17.
- Kotta-Loizou I, Coutts Robert HA. Mycoviruses in *Aspergilli*: a comprehensive review. *Front Microbiol.* 2017;8:1–15.
- Peden J. Analysis of codon usage, PhD [dissertation]. In: Department of Genetics: University of Nottingham; 1999.
- Puigbo P, Bravo IG, Garcia-Vallve S. CAlcal: a combined set of tools to assess codon usage adaptation. *Biol Direct.* 2008;3–38.
- Sharp PM, Li WH. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res.* 1986;14:7737–49.
- Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol.* 1986b;24:28–38.
- Wong Emily HM, Smith DK, Rabadan R, Peiris M, Poon Leo LM. Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. *BMC Evol Biol.* 2010;10:253.
- Butt AM, Nasrullah I, Tong Y. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PLoS One.* 2014. <https://doi.org/10.1371/journal.pone.0090905>.
- Wright F. The 'effective number of codons' used in a gene. *Gene.* 1990;87:23–9.
- Comeron JM, Aguade M. An evaluation of measures of synonymous codon usage bias. *J Mol Evol.* 1998;47:268–74.
- Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 1987;15:1281–95.
- Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 2000;28:292.
- Greenacre MJ. Theory and applications of correspondence analysis. London: Academic Press; 1984.
- Kumar S, Stecher G, Tamra K. Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
- Cristina J, Fajardo A, Sonora M, Moratorio G, Musto H. A detailed comparative analysis of codon usage bias in Zika virus. *Virus Res.* 2016;223:147–52.
- Bhatti MF, Bignell EM, Coutts Robert HA. Complete nucleotide sequences of two dsRNAs associated with a new partitivirus infecting *Aspergillus fumigatus*. *Arch Virol.* 2011;156:1677–80.
- Jamal A, Bignell EM, Coutts Robert HA. Complete nucleotide sequences of four dsRNAs associated with a new chrysovirus infecting *Aspergillus fumigatus*. *Virus Res.* 2010;153:64–70.
- Kozlakidis Z, Herrero N, Ozkan S, Kanhayuwa L, Jamal A, Bhatti MF, Coutts Robert HA. Sequence determination of a quadripartite dsRNA virus isolated from *Aspergillus foetidus*. *Arch Virol.* 2013;158:267–72.
- Kozlakidis Z, Herrero N, Coutts Robert HA. The complete nucleotide sequence of a totivirus from *Aspergillus foetidus*. *Arch Virol.* 2013;158:263–6.
- Kozlakidis Z, Herrero N, Ozkan S, Bhatti MF, Coutts Robert HA. A novel dsRNA element isolated from the *Aspergillus foetidus* mycovirus complex. *Arch Virol.* 2013;158:2625–8.
- Kim JW, Choi EY, Kim YT. Intergeneric relationship between the *Aspergillus ochraceus* virus F and the *Penicillium stoloniferum* virus S. *Virus Res.* 2006; 120:212–5.
- Hammond TM, Andrews MD, Roossinck MJ, Keller NP. *Aspergillus* mycoviruses are targets and suppressors of RNA silencing. *Eukaryot Cell.* 2008;7:350–7.
- Mu F, Xie J, Cheng S, You MP, Barbetti MJ, Jia J, Wang Q, Cheng J, Fu Y, Chen T, Jiang D. Virome characterization of a collection of Sclerotinia sclerotiorum from Australia. *Front Microbiol.* 2018;8:2540.
- Niu Y, Yuan Y, Mao J, Yang Z, Cao Q, Zhang T, Wang S, Liu D. Characterization of two novel mycoviruses from *Penicillium digitatum* and the related fungicide resistance analysis. *Sci Rep.* 2018;8:5513.
- Okada R, Ichinose S, Takeshita K, Urayama S, Fukuhara T, Komatsu K, Arie T, Ishihara A, Egusa M, Kodama M, Moriyama H. Molecular characterization of a novel mycovirus in *Alternaria alternata* manifesting two-sided effects: Down-regulation of host growth and up-regulation of host plant pathogenicity. *Virology.* 2018;519:23–32.
- Lau Susanna KP, Lo George CS, Chow Franklin WN, Fan Rachel YY, Cai JJ, Yuen KY, Woo Patrick CY. Novel Partitivirus enhances virulence of and causes aberrant gene expression in *Talaromyces marneffei*. *MBio.* 2018. <https://doi.org/10.1128/mBio.00947-18>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

