



# Effects of nucleotide usage on the synonymous codon usage patterns of biofilm-associated genes in *Haemophilus parasuis*

L.Y. Wang, L.N. Ma and Y.S. Liu

State Key Laboratory of Veterinary Etiological Biology,  
Lanzhou Veterinary Research Institute,  
Chinese Academy of Agricultural Sciences, Lanzhou, Gansu, China

Corresponding author: Y.S. Liu  
E-mail: liuyongsheng@caas.cn

Genet. Mol. Res. 15 (2): gmr.15027226  
Received July 16, 2015  
Accepted September 25, 2015  
Published June 10, 2016  
DOI <http://dx.doi.org/10.4238/gmr.15027226>

**ABSTRACT.** To provide a new perspective on the evolutionary characteristics shaping the genetic diversity of *Haemophilus parasuis* biofilms, the relative synonymous codon usage values, codon usage bias values, effective number of codons (ENC) values, codon adaptation index (CAI) values, and the base components were calculated. Our objective was to implement a comparative analysis to evaluate the dynamic evolution of biofilm-associated genes in *H. parasuis*. The analysis of genetic diversity provides evidence that some biofilm-associated genes have similar genetic features. However, other genes show some variations in genetic direction. Furthermore, preferential selection of the synonymous codons and amino acids is apparent in biofilm-associated genes. Additionally, the ENC and CAI data from this study all strongly suggested that biofilm-associated genes may depend on deoptimization to adapt to environmental changes, and the mutation effect of biofilm-associated genes in *H. parasuis* plays an important role

in shaping the genetic features. Our results reveal that the mutations of biofilm-associated genes form a set of sophisticated strategies for combating the environmental changes arising from the host cell in the evolution of *H. parasuis*.

**Key words:** *Haemophilus parasuis*; Synonymous codon usage; Codon usage bias; Dynamic evolution

## INTRODUCTION

*Haemophilus parasuis* is a small, polymorphous, gram-negative bacillus that belongs to the Pasteurellaceae family. It is a commensal organism found in the upper respiratory tract of swine (Bello-Ortí et al., 2014). According to the differences in virulence of *H. parasuis*, at least 15 serovars have been described (Tang et al., 2010). It is the etiological agent of Glässer's disease, which is characterized by polyarthritis, fibrinous polyserositis, and meningitis. Although *H. parasuis* only infects pigs, it has caused significantly increased mortality and morbidity rates, leading to major economic losses in the pig-rearing industry (Zou et al., 2013). Because of its ability to form biofilms, *H. parasuis* colonies are resilient and exhibit high survival rates (Ishak et al., 2014). Bacterial cells enclosed in the self-produced polymer matrix form the biofilm; the polymer matrix is composed of proteins, nucleic acids, and exopolysaccharides. Biofilms can adhere to solid surfaces and protect bacteria from adverse environmental conditions, thereby resisting attack from the host's immune defenses. Biofilms are one of the main causes of chronic *H. parasuis* infection (Bello-Ortí et al., 2014).

In the wake of developments in sequencing technology for microorganisms, many complete genomes or single genes have been described, and some of these genes are related to biofilm formation. Analyzing these genes would improve our understanding of the function of biofilms in the etiopathogenesis of *H. parasuis*. Apart from analyzing the amino acid and nucleotide sequences, the synonymous codon use pattern could provide more useful genetic information about the mechanism underlying biofilm formation and the role of biofilms in the etiopathogenesis of *H. parasuis*. It is accepted that information on protein sequences and the formation of codon usage bias (CUB) arises from the non-randomness of nucleotide compositions (Lloyd and Sharp, 1992; Xie and Ding, 1998). In general, translation selection and compositional constraints under mutational pressure are thought to be major factors accounting for codon usage variation among genomes in microorganisms (Karin and Mrázek, 1996; Lesnik et al., 2000; Zhou et al., 2006; Zhou et al., 2010). In some microorganisms, compared with translation selection in nature, mutation pressure plays an important role in the synonymous codon usage pattern (Zhou et al., 2014).

The codon usage pattern of the biofilm-associated genes may reveal the molecular mechanism and evolutionary process of *H. parasuis* free from the host cell response. To our knowledge, this is the first study to systemically analyze the synonymous codon usage pattern and evolutionary dynamics driven by mutation pressure or natural selection in *H. parasuis*. In this investigation, synonymous codon usage analyses were used to determine the evolutionary characteristics of the biofilm-associated genes in *H. parasuis*. The key evolutionary determinants of codon usage bias in *H. parasuis* were evaluated.

## MATERIAL AND METHODS

### Sequence data of biofilm-associated genes in *H. parasuis*

Fifteen genes are related to biofilm formation in *H. parasuis*: *aidA*, *capD*, *galE*, *galU*, *iscR*, *lsgB*, *luxS*, *ompP2*, *ompP5*, *opsX*, *pilA*, *rfaF*, *wza*, *wzz*, and *yadA* (Li et al., 2013; Zou et al., 2013; Zhang et al., 2014). The 391 complete coding sequences of the 15 biofilm-associated genes in *H. parasuis* were downloaded from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/Genbank/>), and detailed information about the microorganisms is listed in [S1 Table](#). The nucleotide content of the open reading frames (ORFs) of each biofilm-associated gene was analyzed using biosoftware DNASTar 7.0 (Madison, WI, USA) for windows.

### Calculation of the relative synonymous codon usage (RSCU)

To investigate the characteristics of the synonymous codon usage of the 15 biofilm-associated genes in *H. parasuis* without the confounding influence of amino acid composition among different sequences, RSCU values among the different codons in each ORF were calculated. The RSCU value of the  $i^{\text{th}}$  codon for the  $j^{\text{th}}$  amino acid was calculated according to the published equation (Sharp et al., 1986). RSCU:

$$RSCU = \frac{g_{ij}}{\sum_j g_{ij}} \cdot n_i \quad (\text{Equation 1})$$

where  $g_{ij}$  is the observed number of the  $i^{\text{th}}$  codon for the  $j^{\text{th}}$  amino acid, which has  $n_i$  type synonymous codons. Codons with RSCU values more than 1.0 have positive CUB, while those with RSCU values less than 1.0 have relatively negative CUB. When the RSCU value is equal to 1.0, the codon is chosen equally and randomly.

### Analysis of codon usage bias

The codon adaptation index (CAI) is a measure that was used to account for the unequal usage of synonymous codons in the 15 biofilm-associated genes in *H. Parasuis* (Sharp and Li, 1987). CAI was originally proposed to provide a normalized estimate that could be used across the whole coding sequence, and the threshold ranges from 0 to 1, i.e., a higher value indicates a stronger synonymous codon bias. The CAI value of a gene correlates with its expression level (Sharp and Li, 1986). The “effective number of codons” (ENC) is another means of quantifying the codon usage bias of the whole coding sequence, and is a useful estimator of absolute codon usage bias; the ENC value ranges from 20 (when only one synonymous codon is chosen by the corresponding amino acid) to 61 (when all synonymous codons are used equally) (Wright, 1990).

Additionally, there is a simple method that supposes that statistically equal and random usage of all available synonymous codons is the “neutral point” ( $RSCU_0 = 1.00$ ) for the development of serotype-specific codon usage (Zhou et al., 2010). CUB:

$$CUB = \sum_{i=1}^n (RSCU_{ij} - RSCU_0) / n \quad (\text{Equation 2})$$

More simply, *CUB* is the average value of difference between  $RSCU_{ij}$  and  $RSCU_0$  at each position of the target region;  $n$  represents all codons appearing in this position. When all *RSCU* values according to a particular position in the target region are  $RSCU_0$ , *CUB* is equal to zero. This means that few preferential or non-preferential codons exist at this position. In contrast, when the *CUB* value shows much greater deviation than  $RSCU_0$ , codons with *CUB* are preferentially chosen at a particular position.

### Principal component analysis

Principal component analysis (PCA), a commonly used multivariate statistical method, was carried out to analyze the major trend in codon usage patterns among the different biofilm-associated genes. PCA involves a mathematical procedure that transforms some correlated variable (here the *RSCU* value) into a smaller number of uncorrelated variables called principal components. Each gene is represented as a 59-dimensional vector, and each dimension corresponds to the *RSCU* value of each sense codon, which only includes several synonymous codons for a particular amino acid, excluding the codons AUG, UGG, and three stop codons.

The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. We created a two-dimensional coordinate system, which comprised the first principal component (Axis 1) and the second principal component (Axis 2). In this study, this two-dimensional coordinate system reported the genetic relationships among all strains and the differences in codon usage patterns among the different biofilm-associated genes in *H. parasuis*.

### Correlation analysis

Correlation analysis of the biofilm-associated genes was used to identify the relationship between nucleotide composition and synonymous codon usage patterns (Ewens and Grant, 2005). This analysis was implemented based on the Pearson's rank correlation analysis method. All statistical processes were carried out using statistical software SPSS11.5 for windows.

## RESULTS

### Synonymous codon usage in biofilm-associated genes

The nucleotide contents in each biofilm-associated gene were accumulated and the comparison among the values of A, T, C, G,  $A_3$ ,  $T_3$ ,  $C_3$ , and  $G_3$ % indicated that  $A_3$  and  $T_3$ % were higher than  $G_3$  and  $C_3$ %, based on A and T% being higher than C and G% ([S2 Table](#)), suggesting that nucleotide usages probably shaped the synonymous codon usage pattern. The relationship

between  $(A_3+T_3)$  and  $(A+T)\%$  revealed some special features, because it can reflect the real variation of  $A_3$  or  $T_3\%$ . Thus, the variation of  $(A_3+T_3)\%$  probably had a slight correlation with the synonymous codon usage pattern in the biofilm. The overall RSCU values of 59 sense codons of the biofilm-associated genes are listed in Table 1. It is interesting that Ser, Gly, Arg, Cys, Ile, Lys, Gln, His, Ala, and Leu were associated with optimal A-ended or U-ended codons, while Pro and Thr were associated with C-ended or G-ended optimal codons (Table 1). Also noteworthy is the phenomenon whereby although the amino acids (Leu, Pro) were associated with optimal T- or A-ended codons, they were associated with some preferential C- or G-ended codons (Table 1). This implies that the pattern of synonymous codon usage might not be directly and simply correlated to nucleotide compositional limitation, but to mutation pressure from environmental changes. These results suggest that the compositional limitation of the bacterium itself and the mutation effects played important roles in the evolution of the biofilm-associated genes in *H. parasuis*.

**Table 1.** Synonymous codon usage patterns of biofilm-associated genes in *Haemophilus parasuis*.

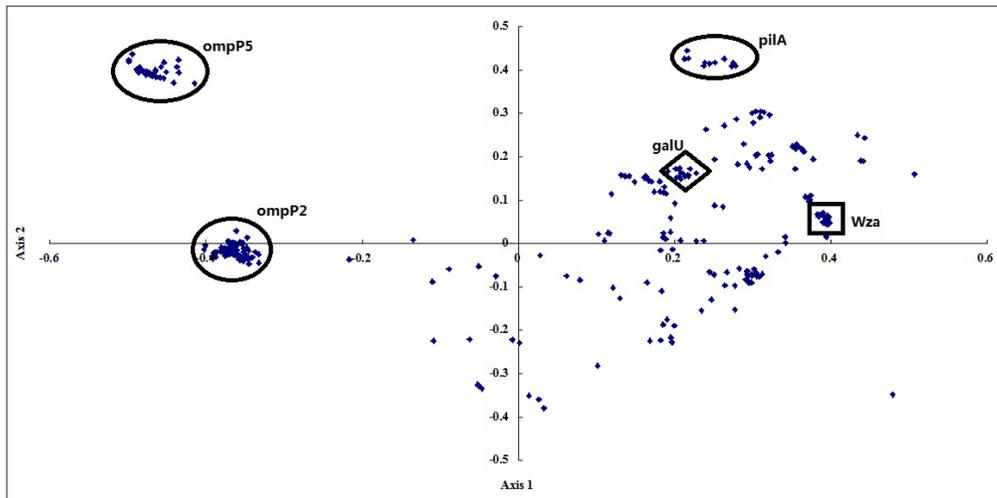
Codon (amino acid)	RSCU	Codon (amino acid)	RSCU
UUU (F)	1.34	ACU (T)	1.08
UUC (F)	0.66	ACC (T)	1.02
UUA (L)	3.39	ACA (T)	1.53
UUG (L)	0.75	ACG (T)	0.37
CUU (L)	1.26	GCU (A)	1.28
CUC (L)	0.40	GCC (A)	0.61
CUA (L)	0.15	GCA (A)	1.66
CUG (L)	0.05	GCG (A)	0.45
AUU (I)	1.82	UAU (Y)	1.09
AUC (I)	1.10	UAC (Y)	0.91
AUA (I)	0.08	CAU (H)	1.68
GUU (V)	1.54	CAC (H)	0.32
GUC (V)	0.40	CAA (Q)	1.68
GUA (V)	1.15	CAG (Q)	0.32
GUG (V)	0.91	AAU (N)	1.21
UCU (S)	2.34	AAC (N)	0.79
UCC (S)	0.74	AAA (K)	1.74
UCA (S)	0.78	AAG (K)	0.26
UCG (S)	0.34	GAU (D)	1.52
AGU (S)	1.24	GAC (D)	0.48
AGC (S)	0.56	GAA (E)	1.16
CCU (P)	0.99	GAG (E)	0.84
CCC (P)	0.00	UGU (C)	2.00
CCA (P)	1.00	UGC (C)	0.00
CCG (P)	2.01	CGU (R)	2.12
GGU (G)	2.21	CGC (R)	0.39
GGC (G)	0.95	CGA (R)	1.56
GGA (G)	0.33	CGG (R)	0.11
GGG (G)	0.51	AGA (R)	1.03
		AGG (R)	0.79

RSCU = relative synonymous codon usage.

## Genetic relationship based on synonymous codon usage in biofilm-associated genes

PCA was performed for the examined ORFs of all isolations. The method discovered one major trend in the first axis, which accounted for 30% of the total synonymous codon usage variation, and another major trend in the second axis, which accounted for 14% of the total variation. The results appeared to be a little complex with some overlapping plots representing different epidemic areas (Figure 1). Compared with other biofilm-associated

genes, the plots for *ompP5*, *ompP2*, *pilA*, *galU*, and *wza*, showed high aggregation, implying that different biological functions played a role in influencing the synonymous codon usage pattern to some degree.



**Figure 1.** Correspondence analysis of biofilm-associated genes in *Haemophilus parasuis*.

### Compositional properties of biofilm-associated genes

To identify the formation of synonymous codon usage regulated by natural selection or mutation pressure, A, U, C, and G% were compared with  $A_3$ ,  $U_3$ ,  $C_3$ , and  $G_3$ %, respectively. There were three different correlation degrees among the nucleotides: positive correlation, negative correlation, and non-correlation. As shown in Table 2, positive correlation between A, U, C, G, and  $A_3$ ,  $U_3$ ,  $C_3$ , and  $G_3$ %, and negative correlation among different nucleotide compositions reflected the fact that compositional constraint played a role in the formation of codon usage patterns, while positive correlation between G and  $A_3$ % probably implies that the formation of a codon usage pattern is regulated by other factors. In addition, the synonymous codon usage indices (Axis 1 and Axis 2) were compared with the nucleotide composition of each sample (Table 3). The synonymous codon usage indices appeared to have significant correlation with all examined nucleotide compositions. Taken together, these analyses indicate

**Table 2.** Summary of the correlation analysis between A, U, C, and G contents and  $A_3$ ,  $U_3$ ,  $C_3$ , and  $G_3$  contents in all selected samples.

	$A_3$ %	$U_3$ %	$C_3$ %	$G_3$ %
A%	$r = 0.733^{**}$	$r = 0.039^{NS}$	$r = 0.016^{NS}$	$r = -0.719^{**}$
U%	$r = 0.243^{**}$	$r = 0.823^{**}$	$r = -0.694^{**}$	$r = -0.448^{**}$
C%	$r = -0.677^{**}$	$r = -0.941^{**}$	$r = 0.958^{**}$	$r = 0.496^{**}$
G%	$r = 0.131^{**}$	$r = 0.674^{**}$	$r = -0.820^{**}$	$r = 0.222^{**}$

The r values in this table were calculated in each correlation analysis. NS means non-significant ( $P > 0.05$ ). \* $0.01 < P < 0.05$ . \*\* $P < 0.01$ .

that nucleotide compositions play a role in synonymous codon usage patterns to some degree. Furthermore, mutation effects caused by environmental changes were the main factors responsible for the variation of synonymous codon usage among biofilm-associated genes.

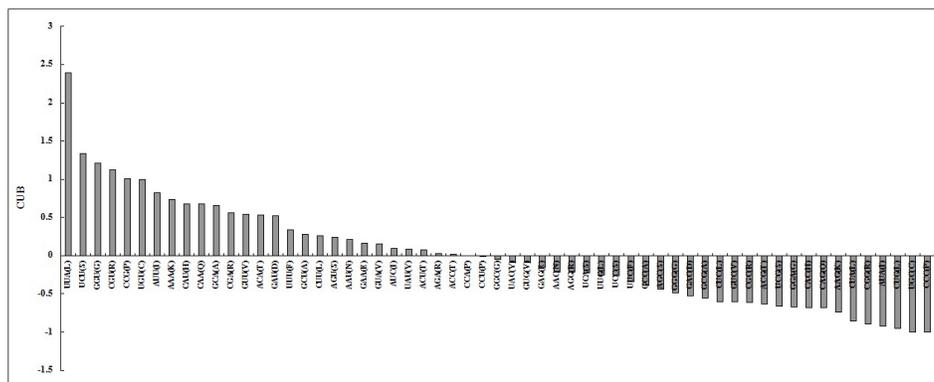
**Table 3.** Summary of the correlation analysis between the first two axes in principle and nucleotide contents in samples.

Base compositions	Axis 1	Axis 2
A <sub>3</sub> %	r = -0.573**	r = -0.248**
U <sub>3</sub> %	r = -0.529**	r = -0.242**
C <sub>3</sub> %	r = 0.518**	r = 0.235**
G <sub>3</sub> %	r = 0.454*	r = 0.241**

\*0.01 < P < 0.05. \*\*P < 0.01.

### Relationship between amino acids and codon usage pattern of biofilm-associated genes in *H. parasuis*

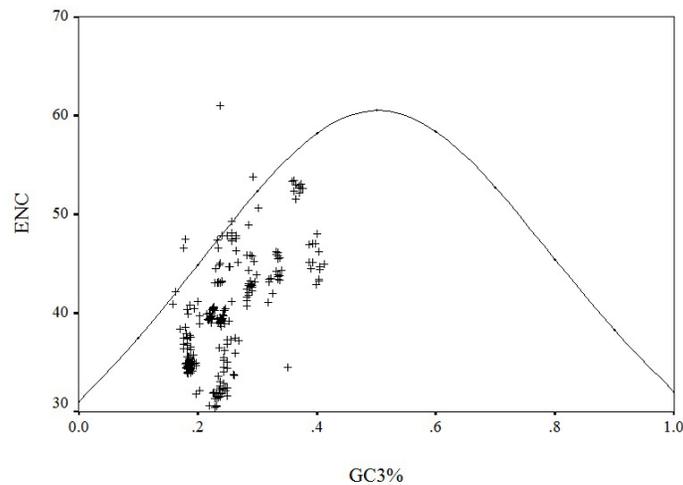
To determine whether the evolution of CUB in biofilm-associated genes in *H. parasuis* was controlled by mutation effects or by translational selection in nature, the CUB values were calculated based on the data listed in Table 1. This table displays a numerical representation of the translational machinery. The distribution of CUB values is illustrated in Figure 2. The transition from maximum-negative to maximum-positive values was smooth and there was no obvious or unambiguous border between the so-called dominant and prohibited codons, i.e., all possible codons were used. The result indicated that translational selection in nature had no effect on the pattern of synonymous codon usage and the evolutionary pattern of biofilm-associated genes in *H. parasuis*.



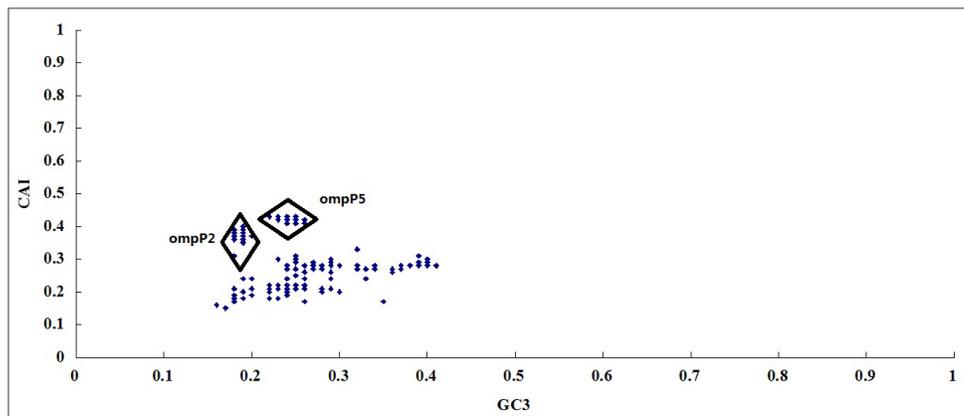
**Figure 2.** Distribution of the codon usage bias (CUB) of codons for each amino acid. CUB was taken from Table 1 and sorted in ascending order.

To analyze the effects of environmental changes on the evolution of biofilm-associated genes, the values of ENC among the biofilm-associated genes were examined. They were highly variable, and ranged from 30.45 to 61, suggesting the effects of synonymous codon usages on the genetic trend of biofilm-associated genes in *H. parasuis*. At the same time, we created a plot that reflected the relationship between GC<sub>3</sub>% and ENC values in the biofilm-

associated genes (Figure 3), and found that most of the data points of the biofilm-associated genes fell under the expected curve, which may reflect a direct relationship between the ENC values and GC<sub>3</sub>% (Wright, 1990). This suggests that the codon usage pattern of the host cells did not regulate the evolution of the codon usage of the biofilm-associated genes. Additionally, to detect the expression level of biofilm-associated genes in the host cells, the CAI data were calculated; the data values were low and ranged from 0.15 to 0.43. Moreover, the CAI data revealed that the expression levels of *ompP2* and *ompP5* genes were higher than those of the other genes (Figure 4). There was a significantly negative correlation between the CAI data and GC<sub>3</sub>% (Pearson  $r = -0.259$ ,  $P < 0.001$ ) (Figure 4), implying that biofilm-associated genes in *H. parasuis* might have low expression levels in host cells because several optimal codons of biofilm-associated genes correspond to rare codons in the host cells.



**Figure 3.** Relationship between GC<sub>3</sub>% and effective number of codons (ENC) values of biofilm-associated genes in *Haemophilus parasuis*.



**Figure 4.** Correlation between GC<sub>3</sub>% and codon adaptation index (CAI) values of biofilm-associated genes in *Haemophilus parasuis*.

## DISCUSSION

Previous reports have generally indicated that organisms, including *Sulfolobus acidocaldarius*, *Taenia pisiformis*, nematodes, and spirochaetes, preferentially use C-/G-ended or A-/T-ended codons (Lafay et al., 1999; Cutter et al., 2006; Chen et al., 2013; Nayak, 2013). However, a few reports have described a synonymous codon usage pattern that influences the evolution of biofilm-associated genes in *H. parasuis*. There is no doubt that during evolution, biofilm-associated genes shape the synonymous codon usage pattern to adapt to the environmental changes resulting from the host. In this study, it was noteworthy that AT<sub>3</sub>% was higher than GC<sub>3</sub>%, because ordinarily AU% is higher than GC% in biofilm-associated genes in *H. parasuis*; amino acids that are associated with two synonymous codons show at least some diversity, although some amino acids are coded by the optimal A- or U-ended codons. These results suggest that the optimal codons usage pattern does not simply reflect the predominant nucleotide in the composition, while compositional limitation and mutation effects play a coefficient role in the genetic variation of synonymous codon usage. The general view is that optimal codons usage increases the efficiency of translation (Ehrenberg and Kurland, 1984; Andersson and Kurland, 1990). Ribosomes affect about two-thirds of the protein content of an *Escherichia coli* cell when growing rapidly (Pedersen et al., 1978), and the ribosome population is a major factor in the control of the growth rate of some organisms (Sharp et al., 2010).

Although the biofilm-associated genes in *H. parasuis* examined in this study do not reflect serotypes depending on the RSCU values of the whole coding sequences, the variation of RSCU values reflects the genetic diversity of biofilm-associated genes with specific biological functions (Costerton et al., 1999; Hall-Stoodley et al., 2004). Bacterial biofilms are structural clusters of bacterial cells enclosed in a self-produced polymer matrix and are attached to a solid surface. Bacteria within a biofilm can resist attack from the host immune response and are less sensitive than planktonic cells to desiccation and to the action of antibiotics. The synonymous codon usage patterns of biofilm-associated genes in *H. parasuis* can be changed to meet the requirements of environmental changes; thus, the specific biological functions of these genes can be separated from those of other similar genes at the evolutionary progress.

The traditional view is that two main forces, namely natural selection and mutational pressure, drive evolutionary progress. With regards to biofilm-associated genes, the role of evolutionary forces in genetic diversity is not clear. When we attempted to determine the main force driving codon usage variation in biofilm-associated genes in *H. parasuis*, we found that CUB was strongly correlated with the nucleotide usages of target genes, implying that compositional constraints under mutational pressure, rather than natural selection for specific coding triplets, was responsible. Furthermore, in Figure 3, the relationship between the ENC data and CG<sub>3</sub>% indicated that mutation effects were probably the main dynamic force driving the evolution of biofilm-associated genes in *H. parasuis* and their adaptation to environmental changes. In this study, the identity of each base at the third position in the synonymous codon was always correlated to the synonymous codon usage variation (Table 3), suggesting that the mutational pressure derived from the environmental changes affecting the evolution of these genes. A general mutational pressure that affects the whole genome would certainly account for the majority of the codon usage patterns that have been reported among microorganisms (Zhou, et al., 2014). Since the mutation rates of the biofilm-associated genes were high, it was understandable that mutation pressure played a key role in shaping the synonymous

codon usage pattern in the biofilm-associated genes in *H. parasuis* examined in this study. In addition, the general association between synonymous codon usage indices and compositional constraints showed that mutational pressure played an important role in determining codon usage variation in the biofilm-associated genes, which was supported by the highly significant correlation between the codon usage indices (Axis 1) and A<sub>3</sub>, U<sub>3</sub>, G<sub>3</sub>, C<sub>3</sub>, and GC<sub>3</sub>%. Moreover, the codon usage indices (Axis 1) were strongly correlated with compositional constraint (Table 3). GC<sub>3</sub>% had a stronger correlation with the Axis 1 values than with the Axis 2 values, implying that compositional limitation played a role in the formation of the overall codon usage pattern of the biofilm-associated genes in *H. parasuis*. In addition, the effect of natural translation selection in shaping synonymous codon usage was not observed in this study, which was consistent with previous reports (Levin and Whittome, 2000). To the best of our knowledge, this is the first report of codon usage analysis of biofilm-associated genes in *H. parasuis*, and it has provided a basic understanding of the mechanisms underlying codon usage bias and the process governing the evolution of biofilm-associated genes in *H. parasuis*.

### Conflicts of interest

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

Research supported in part by grants from the International Science & Technology Cooperation Program of China (#2010DFA32640), and from the National Natural Science Foundation of China (#31302100 and #31172335). Research also supported by the Innovative Engineering of Bacterial Disease in Grazing Animals Team. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### REFERENCES

- Andersson SG and Kurland CG (1990). Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54: 198-210.
- Bello-Orti B, Deslandes V, Tremblay YD, Labrie J, et al. (2014). Biofilm formation by virulent and non-virulent strains of *Haemophilus parasuis*. *Vet. Res. (Faisalabad)* 45: 104-117. <http://dx.doi.org/10.1186/s13567-014-0104-9>
- Chen L, Liu T, Yang D, Nong X, et al. (2013). Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochem. Biophys. Res. Commun.* 430: 1344-1348. <http://dx.doi.org/10.1016/j.bbrc.2012.12.078>
- Costerton JW, Stewart PS and Greenberg EP (1999). Bacterial biofilms: a common cause of persistent infections. *Science* 284: 1318-1322. <http://dx.doi.org/10.1126/science.284.5418.1318>
- Cutter AD, Wasmuth JD and Blaxter ML (2006). The evolution of biased codon and amino acid usage in nematode genomes. *Mol. Biol. Evol.* 23: 2303-2315. <http://dx.doi.org/10.1093/molbev/msl097>
- Ehrenberg M and Kurland CG (1984). Costs of accuracy determined by a maximal growth rate constraint. *Q. Rev. Biophys.* 17: 45-82. <http://dx.doi.org/10.1017/S0033583500005254>
- Ewens WJ and Grant GR (2005). *Statistical methods in bioinformatics*. Springer-Verlag, New York.
- Ishak N, Tikhomirova A, Bent SJ, Ehrlich GD, et al. (2014). There is a specific response to pH by isolates of *Haemophilus influenzae* and this has a direct influence on biofilm formation. *BMC Microbiol.* 14: 47-56. <http://dx.doi.org/10.1186/1471-2180-14-47>
- Hall-Stoodley L, Costerton JW and Stoodley P (2004). Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol.* 2: 95-108. <http://dx.doi.org/10.1038/nrmicro821>
- Karlin S and Mrázek J (1996). What drives codon choices in human genes? *J. Mol. Biol.* 262: 459-472. <http://dx.doi.org/10.1006/jmbi.1996.0528>
- Lafay B, Lloyd AT, McLean MJ, Devine KM, et al. (1999). Proteome composition and codon usage in spirochaetes:

- species-specific and DNA strand-specific mutational biases. *Nucleic Acids Res.* 27: 1642-1649. <http://dx.doi.org/10.1093/nar/27.7.1642>
- Lesnik T, Solomovici J, Deana A, Ehrlich R, et al. (2000). Ribosome traffic in *E. coli* and regulation of gene expression. *J. Theor. Biol.* 202: 175-185. <http://dx.doi.org/10.1006/jtbi.1999.1047>
- Levin DB and Whittome B (2000). Codon usage in nucleopolyhedroviruses. *J. Gen. Virol.* 81: 2313-2325. <http://dx.doi.org/10.1099/0022-1317-81-9-2313>
- Li Y, Kwok AH, Jiang J, Zou Y, et al. (2013). Complete genome analysis of a *Haemophilus parasuis* serovar 12 strain from China. *PLoS One* 8: e68350. <http://dx.doi.org/10.1371/journal.pone.0068350>
- Lloyd AT and Sharp PM (1992). Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 20: 5289-5295. <http://dx.doi.org/10.1093/nar/20.20.5289>
- Nayak KC (2013). Comparative genome sequence analysis of *Sulfolobus acidocaldarius* and 9 other isolates of its genus for factors influencing codon and amino acid usage. *Gene* 513: 163-173. <http://dx.doi.org/10.1016/j.gene.2012.10.024>
- Pedersen EH, Mikkelsen JV, Vaaben J and Taulbjerg K (1978). Interference effect in resonant double-charge transfer. *Phys. Rev. Lett.* 41: 1541-1544. <http://dx.doi.org/10.1103/PhysRevLett.41.1541>
- Sharp PM and Li WH (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24: 28-38. <http://dx.doi.org/10.1007/BF02099948>
- Sharp PM and 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. <http://dx.doi.org/10.1093/nar/15.3.1281>
- Sharp PM, Tuohy TM and Mosurski KR (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14: 5125-5143. <http://dx.doi.org/10.1093/nar/14.13.5125>
- Sharp PM, Emery LR and Zeng K (2010). Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365: 1203-1212. <http://dx.doi.org/10.1098/rstb.2009.0305>
- Tang C, Zhang B, Yue H, Yang F, et al. (2010). Characteristics of the molecular diversity of the outer membrane protein A gene of *Haemophilus parasuis*. *Can. J. Vet. Res.* 74: 233-236.
- Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29. [http://dx.doi.org/10.1016/0378-1119\(90\)90491-9](http://dx.doi.org/10.1016/0378-1119(90)90491-9)
- Xie T and Ding D (1998). The relationship between synonymous codon usage and protein structure. *FEBS Lett.* 434: 93-96. [http://dx.doi.org/10.1016/S0014-5793\(98\)00955-7](http://dx.doi.org/10.1016/S0014-5793(98)00955-7)
- Zhang L, Wen Y, Li Y, Wei X, et al. (2014). Comparative proteomic analysis of the membrane proteins of two *Haemophilus parasuis* strains to identify proteins that may help in habitat adaptation and pathogenesis. *Proteome Sci.* 12: 38-48. <http://dx.doi.org/10.1186/1477-5956-12-38>
- Zhou JH, Zhang J, Chen HT, Ma LN, et al. (2010). Analysis of synonymous codon usage in foot-and-mouth disease virus. *Vet. Res. Commun.* 34: 393-404. <http://dx.doi.org/10.1007/s11259-010-9359-4>
- Zhou JH, Ding YZ, He Y, Chu YF, et al. (2014). The effect of multiple evolutionary selections on synonymous codon usage of genes in the *Mycoplasma bovis* genome. *PLoS One* 9: e108949. <http://dx.doi.org/10.1371/journal.pone.0108949>
- Zhou T, Sun X and Lu Z (2006). Synonymous codon usage in environmental chlamydia UWE25 reflects an evolutionary divergence from pathogenic chlamydiae. *Gene* 368: 117-125. <http://dx.doi.org/10.1016/j.gene.2005.10.035>
- Zou Y, Feng S, Xu C, Zhang B, et al. (2013). The role of galU and galE of *Haemophilus parasuis* SC096 in serum resistance and biofilm formation. *Vet. Microbiol.* 162: 278-284. <http://dx.doi.org/10.1016/j.vetmic.2012.08.006>

## Supplementary material

[S1 Table](#). Detailed information of the biofilm-associated genes in *Haemophilus parasuis*.

[S2 Table](#). Average content of each nucleotide in biofilm-associated genes in *Haemophilus parasuis*.