



Article

# Genome-Wide Analysis of the Synonymous Codon Usage Patterns in *Riemerella anatipestifer*

Jibin Liu <sup>1,2</sup>, Dekang Zhu <sup>2,†</sup>, Guangpeng Ma <sup>3,†</sup>, Mafeng Liu <sup>1,2</sup>, Mingshu Wang <sup>1,2,\*</sup>,  
Renyong Jia <sup>1,2</sup>, Shun Chen <sup>1,2</sup>, Kunfeng Sun <sup>1,2</sup>, Qiao Yang <sup>1,2</sup>, Ying Wu <sup>1,2</sup>, Xiaoyue Chen <sup>2</sup>  
and Anchun Cheng <sup>1,2,\*</sup>

<sup>1</sup> Institute of Preventive Veterinary Medicine, Sichuan Agricultural University, Wenjiang, Chengdu 611130, China; eaeas12@163.com (J.L.); liumafengra@163.com (M.L.); cqrc\_jry@163.com (R.J.); sophia\_cs@163.com (S.C.); 18981613263@163.com (K.S.); yangqiao721521@sina.com (Q.Y.); yingzi\_no1@126.com (Y.W.)

<sup>2</sup> Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Wenjiang, Chengdu 611130, China; zdk24@sicau.edu.cn (D.Z.); chenxy\_24@sina.cn (X.C.)

<sup>3</sup> China Rural Technology Development Center, Beijing 100045, China; maguangpeng1977@163.com

\* Correspondence: mshwang@163.com (M.W.); chenganchun@vip.163.com (A.C.);  
Tel.: +86-28-8629-1905 (M.W.)

† These authors contributed equally to this work.

Academic Editor: Patrick C. Y. Woo

Received: 25 June 2016; Accepted: 2 August 2016; Published: 10 August 2016

**Abstract:** *Riemerella anatipestifer* (RA) belongs to the *Flavobacteriaceae* family and can cause a septicemia disease in poultry. The synonymous codon usage patterns of bacteria reflect a series of evolutionary changes that enable bacteria to improve tolerance of the various environments. We detailed the codon usage patterns of RA isolates from the available 12 sequenced genomes by multiple codon and statistical analysis. Nucleotide compositions and relative synonymous codon usage (RSCU) analysis revealed that A or U ending codons are predominant in RA. Neutrality analysis found no significant correlation between GC<sub>12</sub> and GC<sub>3</sub> ( $p > 0.05$ ). Correspondence analysis and ENc-plot results showed that natural selection dominated over mutation in the codon usage bias. The tree of cluster analysis based on RSCU was concordant with dendrogram based on genomic BLAST by neighbor-joining method. By comparative analysis, about 50 highly expressed genes that were orthologs across all 12 strains were found in the top 5% of high CAI value. Based on these CAI values, we infer that RA contains a number of predicted highly expressed coding sequences, involved in transcriptional regulation and metabolism, reflecting their requirement for dealing with diverse environmental conditions. These results provide some useful information on the mechanisms that contribute to codon usage bias and evolution of RA.

**Keywords:** *Riemerella anatipestifer*; codon usage bias; natural selection; highly expressed gene

## 1. Introduction

*Riemerella anatipestifer* (RA), belonging to the *Flavobacteriaceae* family, is a non-spore-forming, rod-shaped, and atrichous Gram-negative bacterium [1]. It can cause a contagious disease in domestic ducks, geese, turkeys, and various other wild birds. To date, more than 21 serovars have been identified [2]. In addition, no cross-protection has been observed with inactivated bacterins made from different serotypes of RA [3]. Thus, RA can easily cause large economic losses in the duck industry over the world.

Codon usage bias (CUB) of genes generally exists in prokaryotes and eukaryotes. The genetic code in organisms is not strictly one-for-one code. Most amino acids, except Trp (UGG) and Met (ATG) can allow more than one codon (called synonymous codon). Synonymous codons usually

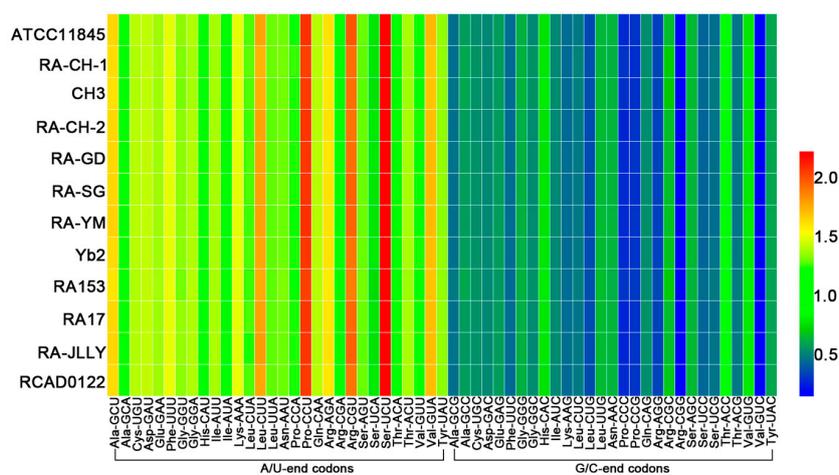
differ by one base in the third codon position (or for some amino acids, in the second position) [4,5]. Among prokaryotes, it is well known that CUB is mainly influenced by mutational bias and natural selection [6,7]. Mutational bias can drive the change in the G + C content of the whole genome. Examples of mutational bias affecting codon usage can be illustrated in many prokaryotes with extremely AT or GC-rich genome [8]. Moreover, CUB may be associated with some other factors, including gene expression level [9,10], gene length [11], amino acid conservation, protein structure [12], gene function [13], and isoaccepting tRNA [14]. There are some variations in codon usage among the genomes of bacteria, which suggests that these genomes bear different pressure in evolution process. CUB analysis has important significance in many aspects. It was proved useful in studying molecular genetic engineering for codon optimization and heterologous protein expression in some species [15]. CUB analysis at genomic scale can also reveal the genetic information about the molecular evolution of individual genes and help to understand evolution of living organisms [16]. Furthermore, CUB can enrich our understanding about the relationship between pathogens and their hosts by analyzing their codon usage patterns [17].

At present, CUB in RA has not been investigated in any detail, and is not clear which factors shape the codon usage pattern. In this study, we analyzed the genome-wide codon usage patterns of 12 RA species. Our results show that natural selection is the main driving factor for codon usage patterns of RA. Additionally, the evolutionary relationship of the species shown in our study is different from that of the traditional classification.

## 2. Results

### 2.1. The Codon Usage Pattern between *Riemerella anatipestifer* (RA)

To identify and understand codon usage patterns of RA, the values of relative synonymous codon usage (RSCU) were computed for every codon in each genome. A codon with an RSCU value of more than 1.0 has a positive codon usage bias, while a value of less than 1.0 has a negative codon usage bias. When the RSCU value is equal to 1.0, it means that this codon is chosen equally and randomly [18]. The results showed a general bias toward codons having nucleotides A or U in the third position while U was more frequently detected (Figure 1 and Table 1). There were 30 codons having the high RSCU values (RSCU > 1, Table 1; in bold), and optimal codons (shown in \*, Table 1) identified by  $\chi$ -squared test which were similarly biased. Among the RA strains, it can be clearly observed that the frequencies of UCU (Ser) and CCU (Pro) are considerably high.



**Figure 1.** Comparison of RSCU between 12 different species of RA. The heat-map was drawn by HemI using hierarchical clustering method. The higher RSCU value, suggesting more frequent codon usage, was represented with darker shades of red. In *Riemerella anatipestifer* (RA) genomes, codons ending in A or U have higher RSCU value than codons ending in G or C.

**Table 1.** The RSCU analysis of the preferred codons, the optimal codons and the rare codons for RA.

Amino Acids	Codon	RSCU <sup>1</sup>	Amino Acids	Codon	RSCU <sup>1</sup>	
Ala	GCG	0.49	Pro	CCC	0.31	
	GCC	0.61		CCG	0.33	
	GCU *	<b>1.66</b>		CCA *	<b>1.27</b>	
	GCA *	<b>1.24</b>		CCU *	<b>2.10</b>	
Cys	UGC	0.57	Arg	AGG	0.37	
	UGU *	<b>1.43</b>		CGC	0.71	
Asp	GAC	0.55	Arg	CGG	0.19	
	GAU *	<b>1.45</b>		AGA *	<b>1.63</b>	
Glu	GAG	0.60	Arg	CGA	<b>1.14</b>	
	GAA	<b>1.40</b>		CGU *	<b>1.96</b>	
Phe	UUC	0.47	Ser	AGC	0.68	
	UUU *	<b>1.53</b>		UCC	0.46	
Gly	GGG	0.66	Ser	UCG	0.49	
	GGC	0.56		AGU *	<b>1.32</b>	
	GGU *	<b>1.36</b>		UCA	0.80	
	GGA *	<b>1.42</b>		UCU *	<b>2.24</b>	
His	CAC	0.80	Thr	ACC	0.89	
	CAU *	<b>1.20</b>		ACG	0.50	
Ile	AUC	0.55	Thr	ACA *	<b>1.20</b>	
	AUU *	<b>1.41</b>		ACU *	<b>1.41</b>	
	AUA	<b>1.04</b>		GUG	0.82	
Lys	AAG	0.47	Val	GUC	0.18	
	AAA *	<b>1.53</b>		GUU *	<b>1.26</b>	
Leu	CUC	0.51	Try	GUA *	<b>1.74</b>	
	CUG	0.39		UAC	0.62	
	UUG	0.68		UAU *	<b>1.38</b>	
	CUA *	<b>1.30</b>		Gln	CAG	0.57
	CUU *	<b>1.81</b>			CAA *	<b>1.43</b>
Asn	UUA *	<b>1.32</b>	Stop	UGA	0.20	
	AAC	0.66		UAG	0.61	
	AAU *	<b>1.34</b>		UAA *	<b>2.19</b>	
Met	AUG	1.00				

<sup>1</sup> Average value of RSCU in 12 RA genomes; \* represents the optimal codons ( $p$ -value < 0.01). The preferred codons (RSCU > 1) are in bold.

## 2.2. The Codon Usage Bias of RA not Affected by Mutation Bias

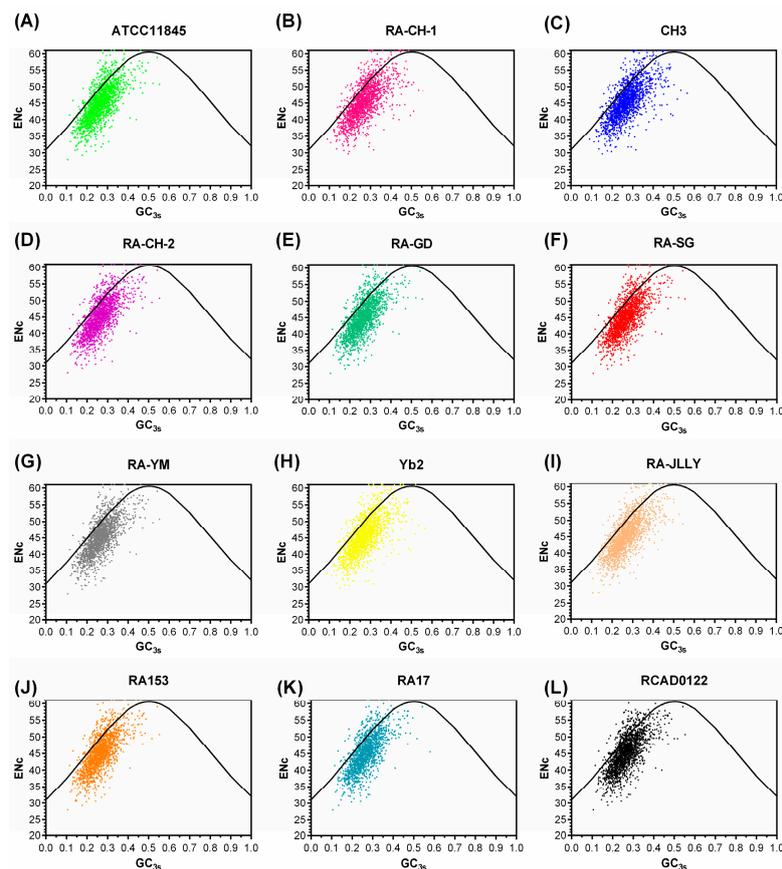
The preference of A or U in the third position of the codon in RA observed in the RSCU comparative analysis could be due to the overall GC bias within the genome. Differences in GC content were the greatest at the third codon position followed by the first and second positions [19]. The GC3s values of RA strains varied from 27.07% to 26.50% with a mean of 26.6% and standard deviation (SD) of 0.23. The effective number of codons (ENc) has been widely used to measure the codon bias level of individual genes. Among the 12 isolates, the values of ENc were higher than 40 (Table 2) ranging from 45.04 and 45.47. With the mean value of 45.20 and S.D. of 0.16 ( $p > 0.05$ ), this indicates that CUB has no bias in RA genomes.

Plotting ENc versus GC3s is an effective strategy to investigate patterns of synonymous codon usage [20]. The distribution plot of ENc and GC3s values for these genes have been presented in

Figure 2. The solid line represents the curve if codon usage is only determined by GC3s. The actual ENc values for some genes lay near to the solid line on the left region of this distribution, and a majority of the points with low ENc values lay below the expected curve. This implies that not only mutation but also other factors are likely to be involved in determining the selective constraints on codon bias in RA genomes.

**Table 2.** Characteristic in the indices of codon bias of RA genes.

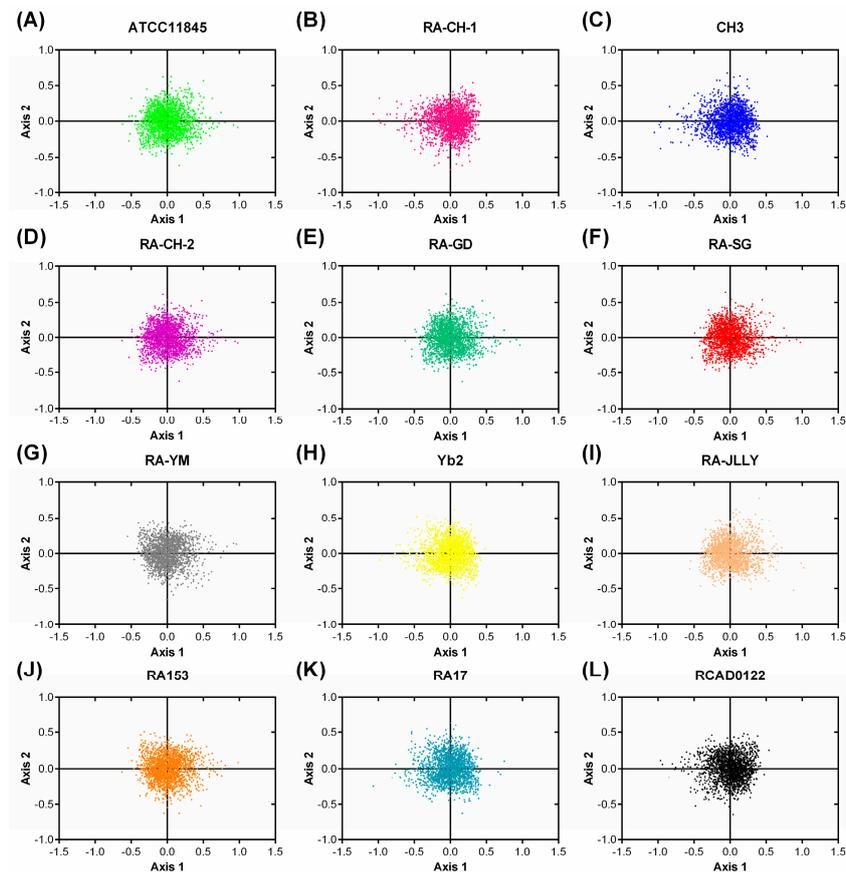
RA Strain	GC%	GC3s%	ENc	CAI
ATCC11845	35.42 ± 3.63	26.50 ± 5.75	45.04 ± 5.13	0.616 ± 0.062
RA-CH-1	35.51 ± 3.85	27.05 ± 6.28	45.47 ± 5.15	0.604 ± 0.064
CH3	35.59 ± 3.89	27.07 ± 6.30	45.46 ± 5.17	0.582 ± 0.070
RA-CH-2	35.39 ± 3.62	26.64 ± 5.98	45.19 ± 5.19	0.616 ± 0.064
RA-GD	35.33 ± 3.64	26.67 ± 5.95	45.16 ± 5.20	0.602 ± 0.069
RA-SG	35.40 ± 3.60	26.52 ± 5.17	45.10 ± 5.07	0.609 ± 0.063
RA-YM	35.50 ± 3.58	26.69 ± 5.80	45.15 ± 5.09	0.610 ± 0.063
Yb2	35.34 ± 3.64	26.50 ± 5.74	45.12 ± 5.14	0.613 ± 0.062
RA-JLLY	35.45 ± 3.89	26.91 ± 6.25	45.34 ± 5.21	0.605 ± 0.067
RA153	35.44 ± 3.61	26.57 ± 5.86	45.18 ± 5.18	0.604 ± 0.064
RA17	35.52 ± 3.59	26.55 ± 5.77	45.14 ± 5.22	0.690 ± 0.064
RCAD0122	35.40 ± 3.65	26.69 ± 5.86	45.21 ± 5.14	0.607 ± 0.064



**Figure 2.** The ENc vs. GC3s plots of RA genomes. (A) ATCC11845; (B) RA-CH-1; (C) CH3; (D) RA-CH-2; (E) RA-GD; (F) RA-SG; (G) RA-YM; (H) Yb2; (I) RA-JLLY; (J) RA153; (K) RA17; and (L) RCAD0122. The standard curve represents the expected ENc to GC3s. Most RA genes are far away from the standard curve, showing that their codon usage pattern might be affected by other factors besides nucleotide composition. Some genes with the ENc score of 61 display no bias and use all the 61 sense codons.

### 2.3. Correspondence Analysis (COA)

To investigate the synonymous codon usage variation among RA strains, COA was performed on the variation of RSCU value for this study. The coordinate of each coding sequence (CDS) on the two principal axes (Axes 1 and 2) is shown in Figure 3. The relative inertia explained by the first axis in RA contributes approximately 10% of the total variation. It must be remembered that although the first principal axis explains a substantial amount of variation of codon usage among the genes in RA, its value is not remarkably high for relative inertia explained by the first axis in other organisms studied earlier [11,21,22]. The low value might be due to the AT-rich genomic composition of this genome. As it can be seen, these strains of RA isolated from different places, even the same serotype, have the same trend in codon usage variation. The previous studies have shown that the codon usage variation among the genes in the extremely AT or GC rich organisms is only shaped by compositional bias, The third codon position in the preferred codons should also have the base composition of A or T [23,24]. The mutation bias toward a high G + C content seems to have resulted in a preponderance of GC-rich optimal codons [25]. As shown in Table 1, the third positions of optimal codons in RA were preferred in A or T, which suggests that the strongest influence on the choice of codon usage might not be mutation bias, but translation optimality in RA.

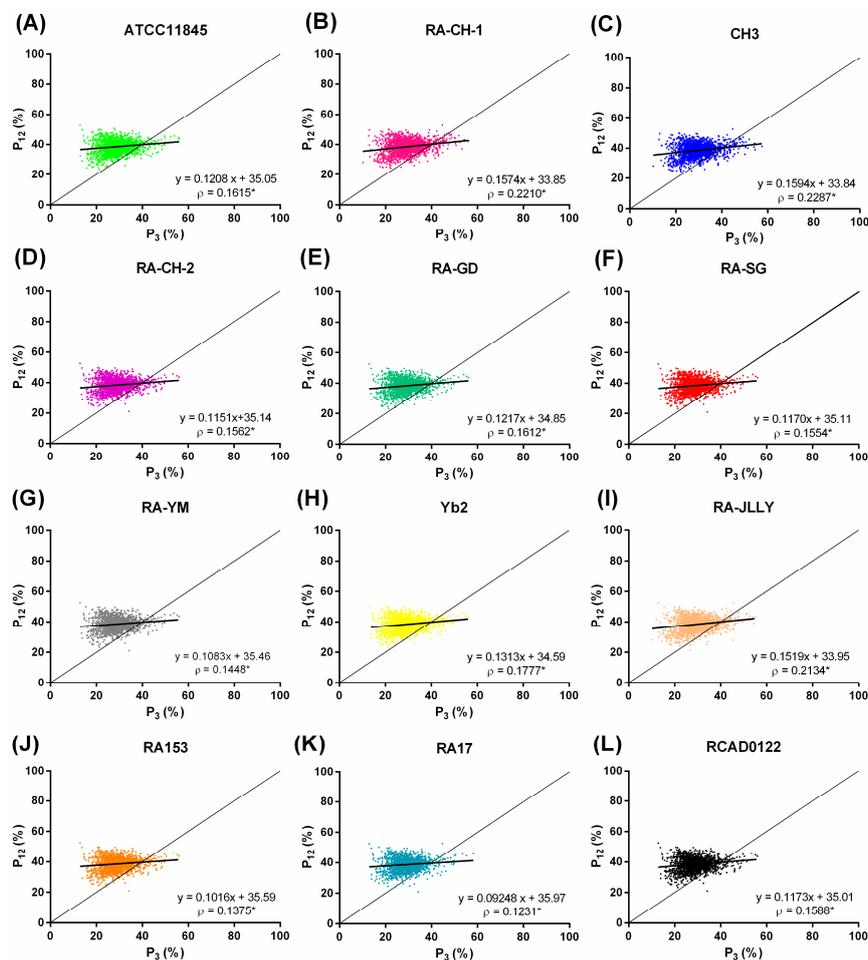


**Figure 3.** The correspondence analysis (COA) of the genes in RA genomes. (A) ATTCC11845; (B) RA-CH-1; (C) CH3; (D) RA-CH-2; (E) RA-GD; (F) RA-SG; (G) RA-YM; (H) Yb2; (I) RA-JLLY; (J) RA153; (K) RA17; and (L) RCAD0122. Each point represents a gene corresponding to the coordinates of the first and second axes of variation generated from the correspondence analysis.

### 2.4. Natural Selection Influences the Codon Bias of RA

The GC content is calculated according to the first, second, and third codon positions ( $P_1$ ,  $P_2$  and  $P_3$  respectively).  $P_{12}$  is the average of  $P_1$  and  $P_2$ , it is used for analysis of neutrality plot ( $P_{12}$  against  $P_3$ ).

The neutrality plot is drawn to characterize the correlation among the three codon positions, and then used to estimate the extent of directional mutation pressure against selection on CUB. In the neutrality plot, each point represents one gene (Figure 4). If a gene is under neutral selection pressure, a point should be located on diagonal line with a significant correlation between its  $P_{12}$  and  $P_3$ . If a gene is close to X-axis, below the diagonal line, meaning the gene is under mutational pressure. Thus, the slope less than 1 should indicate a whole genome trend of non-neutral mutational pressure [26,27]. In this study, all RA species had relative neutralities ranging from 9% to 15% (Figure 4). It means CUB was affected a little by neutral evolution since natural selection was more than 85%. The points in all RA species were located above the diagonal distribution and the regression curve (bold line) with a slope less than 1, indicated the whole genomes in RA species trend of non-neutral mutational pressure. The subsequent correlation analysis revealed little positive correlation between  $P_{12}$  and  $P_3$ . These results showed that natural selective pressure dominated over mutation shaping the composition of coding sequences.

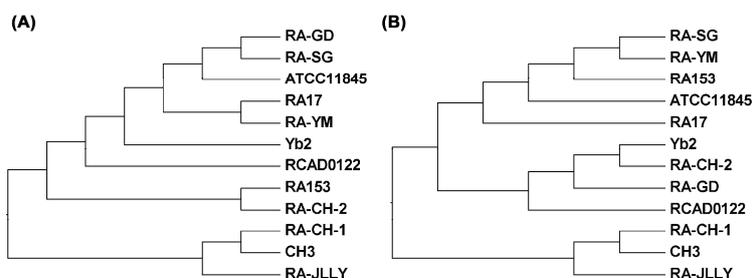


**Figure 4.** Neutrality plots of RA genomes. (A) ATCC11845; (B) RA-CH-1; (C) CH3; (D) RA-CH-2; (E) RA-GD; (F) RA-SG; (G) RA-YM; (H) Yb2; (I) RA-JLLY; (J) RA153; (K) RA17; and (L) RCAD0122. Individual genes are plotted based on the mean GC content in the first and second codon position ( $P_{12}$ ) versus the GC content of the third codon position ( $P_3$ ). Regression lines and Spearman's rank correlation coefficients ( $\rho$ ) are shown, with the asterisk (\*) denoting  $p$ -values < 0.01.

### 2.5. Cluster Analysis

To gain more insight into evolution of the RA, the RSCU values between 12 species were used in hierarchical clustering (Figure 5B). Cluster analysis for RA family yielded five major clusters, similar to

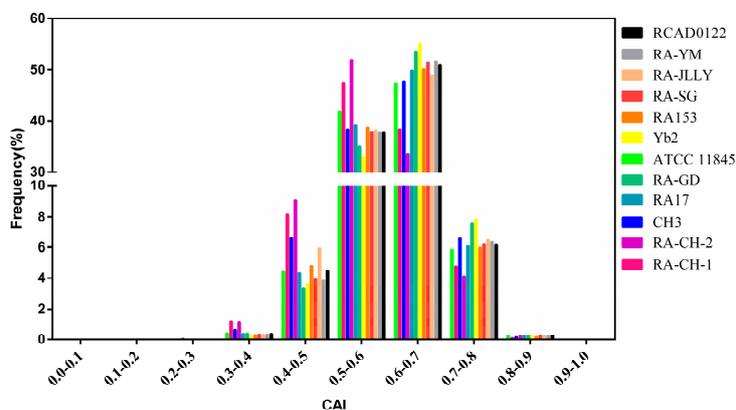
dendrogram based on genomic BLAST by neighbor-joining method (Figure 5A). Cluster I is composed of RA-CH-2, RA-GD, Yb2 and RCAD0122, meanwhile RA-CH-2 and Yb2 stay the closest and are isolated almost from the root. Cluster II contains ATCC11845, RA153, RA-SG and RA-YM. RA-SG and RA-YM appear closely related to RA153, but are on different branches compared to ATCC11845. RA17 has a close relationship with cluster II divided into cluster III. RA-CH-1 and CH3, belonging to the serotype 1, are clustered in cluster IV. The highly biased RA-JLLY is clustered alone as a minor cluster of cluster V and close to the branch of RA-CH-1 and CH3.



**Figure 5.** Comparison of phylogenetic tree with RSCU based clustering of RA strains. (A) The phylogenetic tree derived for the RA genome using genomic BLAST by neighbor-joining method; (B) Cluster analysis of the 12 species in RA based on RSCU value. The observed distances range from 1 to 25, the ratio of the rescaled distances within the dendrogram is the same as the ratio of the original Squared Euclidean distances.

## 2.6. Understanding Pathway Level Functions in RA through CUB

The codon adaptation index (CAI) for a gene is a measurement of its optimal codons usage, which is the codon commonly used by highly expressed proteins in a given genome [28]. CAI values of all CDS in RA genomes were calculated using the ribosomal protein codon usages as a reference set. As shown in Figure 6, the CAI values of all RA genes were distributed over a very wide range from 0.3 to 0.8 (the mean value of 0.6), but most of the genes had CAI values between 0.5 and 0.7. Only about 6% of the CAI values were greater than or equal to 0.7. No obvious correlation was observed between CAI values and the corresponding gene lengths ( $p > 0.05$ ). This implies that codon bias is not the primary mechanism determining the translational efficiency of long genes in RA. Within each RA strain, the top 5% of genes with the highest CAI value were predicted to be highly expressed genes. This is corresponded to CAI cut-off of 0.701 in ATCC11845, 0.698 in RA-CH-1, 0.708 in CH3, 0.691 in RA-CH-2, 0.709 in RA-GD, 0.706 in RA-SG, 0.706 in RA-YM, 0.715 in Yb2, 0.708 in RA-JLLY, 0.703 in RA153, 0.700 in RA17, and 0.706 in RCAD0122 (included about 100 genes for each RA strains), respectively.



**Figure 6.** Frequency distribution of codon adaptation index (CAI) values for coding sequence (CDS) in the genomes of RA strains.

**Table 3.** Orthologous high-level expression genes found in 12 RA strains.

Category	Gene	Proteins	Strains
Ribosome	<i>rplA</i>	Large subunit ribosomal protein L1	RA-CH-1, RA-GD, RA17
	<i>rplB</i>	Large subunit ribosomal protein L2	+
	<i>rplD</i>	Large subunit ribosomal protein L4	+
	<i>rplE</i>	Large subunit ribosomal protein L5	+
	<i>rplF</i>	Large subunit ribosomal protein L6	+
	<i>rplL</i>	Large subunit ribosomal protein L7/L12	+
	<i>rplI</i>	Large subunit ribosomal protein L9	+
	<i>rplJ</i>	Large subunit ribosomal protein L10	+
	<i>rplK</i>	Large subunit ribosomal protein L11	RA-CH-1
	<i>rplN</i>	Large subunit ribosomal protein L14	+
	<i>rplO</i>	Large subunit ribosomal protein L15	+
	<i>rplP</i>	Large subunit ribosomal protein L16	RA-CH-2, CH3, ATCC11845
	<i>rplQ</i>	Large subunit ribosomal protein L17	+
	<i>rplR</i>	Large subunit ribosomal protein L18	+
	<i>rplS</i>	Large subunit ribosomal protein L19	+
	<i>rplU</i>	Large subunit ribosomal protein L21	+
	<i>rplV</i>	Large subunit ribosomal protein L22	+
	<i>rplX</i>	Large subunit ribosomal protein L24	+
	<i>rpsA</i>	Small subunit ribosomal protein S1	+
	<i>rpsB</i>	Small subunit ribosomal protein S2	+
Ribosome	<i>rpsC</i>	Small subunit ribosomal protein S3	+
	<i>rpsD</i>	Small subunit ribosomal protein S4	+
	<i>rpsE</i>	Small subunit ribosomal protein S5	RA-CH-1, CH3
	<i>rpsG</i>	Small subunit ribosomal protein S7	+
	<i>rpsH</i>	Small subunit ribosomal protein S8	Except RA17, RA-GD
	<i>rpsI</i>	Small subunit ribosomal protein S9	+
	<i>rpsK</i>	Small subunit ribosomal protein S11	+
	<i>rpsO</i>	Small subunit ribosomal protein S15	+
	<i>rpsP</i>	Small subunit ribosomal protein S16	CH3
	<i>rpsR</i>	Small subunit ribosomal protein S18	+

Table 3. Cont.

Category	Gene	Proteins	Strains
Elongation factor	<i>tuf</i>	Elongation factor Tu	+
	<i>fusA</i>	Elongation factor G	+
	<i>tsf</i>	Elongation factor Ts	+
Chaperone	<i>dnaK</i>	Molecular chaperone DnaK	Except CH3
	<i>groEL</i>	Chaperonin GroEL	+
	<i>tig</i>	Trigger factor	+
Enzymes	<i>acnA</i>	Aconitate hydratase	+
	<i>sucC</i>	Succinyl-CoA synthetase $\beta$ subunit	+
	<i>sucD</i>	Succinyl-CoA synthetase $\alpha$ subunit	+
	<i>mdh</i>	Malate dehydrogenase	+
	<i>gapA</i>	Glyceraldehyde 3-phosphate dehydrogenase	+
	<i>ccoP</i>	Cytochrome c oxidase cbb3-type subunit III	+
	<i>ccp</i>	Cytochrome c peroxidase	+
	<i>atpA</i>	F-type H <sup>+</sup> -transporting ATPase subunit $\alpha$	RA-CH-1, RA-CH-2, RA17, ATCC11845
	<i>atpF</i>	F-type H <sup>+</sup> -transporting ATPase subunit b	Except CH3
	<i>pncA</i>	Nicotinamidase/pyrazinamidase	+
	<i>ndk</i>	Nucleoside-diphosphate kinase	+
	<i>tlpA</i>	Alkyl hydroperoxide reductase/thiol specific antioxidant/mal allergen	+
	<i>ppiA</i>	Peptidyl-prolyl isomerase	+
	<i>dsrO</i>	Molybdopterin-containing oxidoreductase	RA-CH-1
	<i>katE</i>	Catalase	RA153, ATCC11845, Yb2, RA-SG
	<i>ahpC</i>	Peroxiredoxin	Except RA153, RA17
<i>sdhB</i>	Succinate dehydrogenase/fumarate reductase	RA153, RA17, RA-SG, RA-YM	
<i>eno</i>	Enolase	+	

Table 3. Cont.

Category	Gene	Proteins	Strains
Enzymes	<i>nrfA</i>	Nitrite reductase	RA-CH-2, RA153, RA17, RA-GD
	<i>dam</i>	DNA adenine methylase	RA17, ATCC11845, RA-GD
	<i>tatD</i>	TatD DNase family protein	RA-CH-1
	<i>pabC</i>	4-Amino-4-deoxychorismate lyase	RA-JLLY
	<i>ald</i>	Alanine dehydrogenase	RA-JLLY
	<i>ribBA</i>	3,4-Dihydroxy 2-butanone 4-phosphate synthase	RA-JLLY
	-	Peptidase s8 and s53 subtilisin kexin sedolisin	+
	-	Peptidase s46	RA17
	-	Putative FAD dependent oxidoreductase	RA17
	-	Septum formation initiator	RA17
-	Serine protease	ATCC11845	
-	Nodulation protein X acyltransferase 3	ATCC11845	
Binding protein	-	Cyclic nucleotide-binding protein	Except RA17
Transport protein	<i>arac</i>	Transcriptional regulator	RA17
Apoptosis protein	<i>cys</i>	Cytochrome c	Except RA17, RA-JLLY
Structure protein	<i>gldI</i>	Gliding motility protein gldI	RA17, ATCC11845
	<i>ompH</i>	Outer membrane protein	+
	<i>ompa/motb</i>	ompa/motb domain-containing protein	+
	<i>ftnA</i>	Ferritin	RA153, ATCC11845, Yb2, RCAD0122
	<i>hinT</i>	Histidine triad (HIT) family protein	CH3
	-	Phosphate-selective porin o and p protein	RA-CH-1

+ represents the gene found in all RA strains.

To further analyze the highly expressed genes estimated by functional analysis, we used blastKOALA based on KEGG annotations [29]. As the limitation of gene annotation and functional studies, about forty-five orthologous high-level expression gene pairs from the all RA genomes were annotated as hypothetical proteins. In this way, their CAI values could rightfully indicate the gene expression level. These hypothetical proteins with the predicted high expression may become attractive candidates for experimental characterization, thus we assumed that they should have important functions in those organisms. Functional analysis showed that only half of genes in all 12 genomes were classified. The high-level expression genes were involved in genetic information processing, carbohydrate metabolism, energy metabolism, metabolism of cofactors and vitamins, nucleotide metabolism and cellular processes (Table 3). The high-level expression genes involved in genetic information processing were the largest functional group. An investigation of the functional categories to which the CAI reference genes (top 1% of genes) belong has revealed that RA contains a significant fraction of ribosomal proteins (large subunit ribosomal in 62.5% and small ribosomal subunits in 37.5%). This is in agreement with the ribosomal criterion defined by Carbone [30], which states that ribosomal proteins have significantly higher CAI value than other protein encoding genes in translationally biased organisms. The *rplL* encoding ribosomal protein L9 with the highest CAI value (0.834) was one of the most abundant proteins under the rapid growth conditions in RA while codon selection was expected to be effective. The second most high-level expression genes was for various enzymes including carbohydrate metabolism, metabolism of cofactors and vitamins, energy metabolism and nucleotide metabolism. As we know, *acnA*, *mdh*, *sucC* and *sucD* gene encoding aconitate hydratase, malate dehydrogenase and succinyl-CoA synthetase are participant in tricarboxylic acid cycle (TCA) pathway. Several genes encoding cytochrome, transferases, and ATP synthase were also found in the 12 RA strains. Enolase is involved in secondary metabolism. Apart from ribosomal proteins and enzymes, three genes encoding elongation factor Tu, G, Ts and two chaperone encoding GroEL and DnaK were observed as the high-level expression genes in RA genomes. In addition, the outer membrane protein was also found high in expression. This analysis has offered the prospective method to further carry out the characterization on those genes.

### 3. Discussion

To confirm the observed dominance of mutational bias, the RSCU patterns are conducted in these strains. As a general rule, AT-rich genome of bacteria can result in the dominance of the A/U-ended codons. RA has extremely AT-rich genome, which is the main reason why there are 31 optimized codons ending with A/U among 32 optimized codons. This predominance of A and T at the synonymous sites is better displayed in Table 2, which reveals that amino acid usage is strongly associated with AT content in AT-rich genome [31]. In bacteria with extreme genomic GC compositions, synonymous codon usage could be dominated by strong compositional bias [32]. What is more, the mutation is universally biased towards AT in bacteria [33,34]. Therefore, it likely can be concluded that the main force driving codon usage in RA is the strong compositional bias towards A and T. It is reasonable that compositional bias may be a potential bias in the evolution of the codons in RA.

The codon usage bias was conserved in RA strains. The RSCUs of each codon were very similar in 12 RA strains. Meanwhile, the distributions of the plot of Axes 1 and 2 in each CDS were almost in the same region. The plot of Axes 1 and 2 of each open reading frame (ORF) shows that there is a quite small amount of the codon usage variation in RA strains. In addition, the COA also has highly negative correlation with the GC3s value, which suggests codon usage variation is directly related to mutational bias. The ENc values of RA genome are all more than 45, which demonstrates that codon usage bias is low in RA strains. The ENc-plot suggests that not only mutational pressure but also other factors affect the codon bias among the genes. This conclusion is also supported by the highly significant correlation analysis. Comparisons of 12 RA species show a significant positive correlation between ENc and GC3s ( $p < 0.01$ ). Moreover, it is obviously that the codon usage bias has no significant

difference by comparing the ENc-plot of 12 RA strains. In summary, the data presented herein reveals that the differences of codon usage are small among different RA strains.

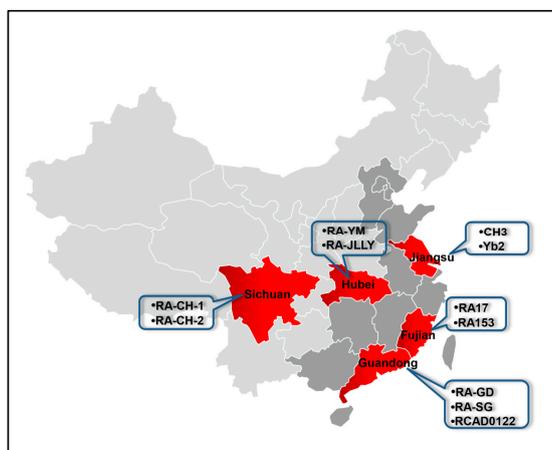
Most CAI values of RA genes are near to 0.6 that is lower than other bacteria, such as *E. coli*, *Nocardia farcinica*, and *Streptomyces coelicolor* [35–37]. The results provide evidence why RA strains need rich nutrition to grow but still slow and consequently have low environment adaptability. By correlation analysis between average RSCU values of RA ORFs and high/low ENc value groups, there are high correlations between RA ORFs and high/low ENc value groups. The codon usage patterns have no obvious difference between high and low ENc value genes. Hence, gene expression levels only have a weak influence on codon usage bias in RA.

Finally, the CAI values were set as the expression level indicator of genes in RA. The notion of gene expression by CAI values was proposed for a long time ago, however, in recent years, the methods have been widely used to qualitatively assess high-level expression genes in prokaryote and eukaryote [38–41]. Fast development of the whole-genome analysis technologies, especially whole genome sequencing as well as proteomics has made it possible to compare computational data of codon usage and expression ability with experimental evidence. In our research, the highly expressed genes can be considered as the strength of relative codon bias, most of the highly expressed genes are identified by ribosomal proteins genes. Moreover, the genes encoding elongation factor, chaperone proteins, enzymes of essential carbon metabolism pathways of TCA cycle, genes of ATP synthesis, nucleotide biosynthesis, outer membrane protein, transport and binding protein are identified as highly expressed genes in our approach. The study also proves our prediction, based on their codon usage, that some of hypothetical proteins would be highly expressed. Further research of hypothetical proteins by integrated computational and experimental data will enhance our knowledge of the metabolism in RA.

## 4. Materials and Methods

### 4.1. Sequences Data

A total of 12 RA genomes were used in this study. The coding sequences (CDS) datasets from the whole genome sequences were obtained from National Center of Biotechnology Information (NCBI). To minimize sampling bias in codon usage calculations only CDS of at least 100 codons in length with correct initiation were used in further analysis. Detailed information about these strains is listed in Table 4, and the distribution of these strains except ATCC11845 in the different provinces of China is shown in Figure 7.



**Figure 7.** Geographical location of the RA analyzed in this study. The main distribution of duck industries in China are shown in dark gray. The provinces (regions) of the RA strains in this study are indicated in red.

**Table 4.** RA strains used in this study.

Strain	Serotype	Geographic Location	Accession No.	CDS	CDS (>300 bp)	Reference
ATCC11845	6	USA	CP003388.	1941	1764	[42,43]
RA-CH-1	1	Sichuan	CP003787	2187	1953	
CH3	1	Jiangsu	CP006649	2181	1916	[44]
RA-YM	1	Hubei	AENH00000000	2010	1796	[45]
RA-CH-2	2	Sichuan	CP004020	2044	1844	[46]
RA-GD	2	Guangdong	CP002562	1985	1815	[47]
Yb2	2	Jiangsu	CP007204	2021	1877	[48]
RA153	2	Fujian	CP007504	1919	1730	
RA17	ND <sup>1</sup>	Fujian	CP007503	1656	1613	
RA-SG	ND <sup>1</sup>	Guangdong	ANGF00000000	2066	1838	[49]
RA-JLLY	ND <sup>1</sup>	Hubei	LAVB01000000	2089	1858	[50]
RCAD0122	ND <sup>1</sup>	Guangdong	LUDU00000000	2149	1892	[51]

<sup>1</sup> ND: Not determined.

#### 4.2. Measurement Indices of Codon Usage Bias

In order to normalize codon usage within datasets of different amino acid compositions, relative synonymous codon usage (RSCU) values were calculated by dividing the observed codon usage by the expected ones under the condition that all codons for the same amino acid are used equally. The RSCU was used to compute relative codon frequency. The codon adaptation index (CAI) has been proved to be the best gene expression value index and was extensively used as a measure of gene expression level. The CAI was generally calculated using the codon preference of genes for highly expressed proteins, such as ribosomal proteins and elongation factors. In this study, the values of CAI were calculated using a reference set of ribosomal proteins. Based on the calculated CAI value, 5% of the total genes with extremely high CAI values were regarded as the highly expressed datasets.

Effective number of codons (ENc) was often used to quantify the codon usage bias of a gene. The ENc value of a gene could range from 20 (extreme bias where one codon for each codon family was used) to 61 (all synonymous codons were used randomly). As in the previous report,  $P_1$ ,  $P_2$ , and  $P_3$  were calculated after excluding ATG, TGG, ATA, and the stop codons (TAA, TAG, or TGA) [52]. The value of GC3s was the frequency of G + C at the synonymous third position of sense codons and it was employed to better understand the codon usage variation and compositional constraints (i.e., excluding Met, Trp, and termination codons). The ENc value against GC3s was computed, which was assumed equal to the use of G and C (A and T) in degenerate codon groups. The expected ENc value under random codon usage was calculated for any value of GC3s as below:

$$ENc = 2 + s + 29[s^2 + (1 - s)^2]^{-1} \quad (1)$$

where  $s$  represents the given GC3s value. If the G + C content at the third position is the only determinant factor that shapes the codon usage, the point of ENc should fall on the standard curve described by Formula (1).

#### 4.3. Correspondence Analysis and Cluster Analysis

The correspondence analysis (COA) was used to investigate the major trend in codon usage variation among genes of 12 RA strains. The CDS of each gene was represented as a 59 dimensional vector (excluding ATG, TGG, and the stop codons), and each dimension corresponds to the RSCU value of one sense codon. Since the first two axes, compared to the other axes, would be enough to explain the higher fraction of the variance of the data, genes and codons were plotted on these two axes only [53,54]. In the cluster analysis, RA species were clustered according to their RSCU values by hierarchical methods through measurement of the Squared Euclidean distance.

#### 4.4. Software and Statistical Analysis

RSCU, ENc, total G + C genomic content, as well as COA, were calculated by CodonW 1.4 version [55]. The heat map was drawn with HemI (Huazhong University of Science and Technology, Wuhan, Hubei, China) [56] and clustered the RSCU values using an average linkage cluster algorithm. Values of CAI, P1, P2 and P3 were calculated by CAIcal Server [57]. The highly-expressed-gene datasets were interpretation of high-level functions by BlastKOALA [58]. Correlation analysis was performed using the statistical software SPSS 19.0 (IBM, Chicago, IL, USA). Graphs were generated with GraphPad Prism 6.0 (GraphPad Software Inc., La Jolla, CA, USA).

#### 5. Conclusions

To summarize, our study reveals that codon usage bias in RA is slightly biased, and there is no significant difference between the strains in codon usage. Natural selection is the main factor that affects codon usage variation in RA. Other factors, such as GC content and gene expression also have an influence on codon usage pattern. In addition, all RA strains have the common highly abundant proteins. To our knowledge, this research is the first work of its kind to report of codon usage analysis in RA, and it gives us a basic understanding of the mechanisms for codon usage bias and gene expression during the evolution of RA. Moreover, this study has provided a basis for further studies on the mechanisms of codon usage that affects the RA strains through evolution.

**Acknowledgments:** The research was supported by National Natural Science Foundation of China (31572521), Integration and Demonstration of Key Technologies for Duck Industrial in Sichuan Province (2014NZ0030), China Agricultural Research System (CARS-43-8) and Special Fund for Key Laboratory of Animal Disease and Human Health of Sichuan Province (2016JPT0004).

**Author Contributions:** Jibin Liu, Dekang Zhu, and Guangpeng Ma designed/performed the experiments and wrote the paper. Mafeng Liu, Shun Chen, Renyong Jia, Xiaoyue Chen, Kunfeng Sun, Qiao Yang and Ying Wu contributed to data collection and helped in data analysis. Mingshu Wang and Anchun Cheng contributed to figure modification.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Subramaniam, S.; Chua, K.L.; Tan, H.M.; Loh, H.; Kuhnert, P.; Frey, J. Phylogenetic position of *Riemerella anatipestifer* based on 16S rRNA gene sequences. *Int. J. Syst. Bacteriol.* **1997**, *47*, 562–565. [[CrossRef](#)] [[PubMed](#)]
2. Swayne, D.E.; Glisson, J.R.; McDougald, L.R. *Diseases of Poultry*; Wiley-Blackwell: Hoboken, NJ, USA, 2013; pp. 811–813.
3. Chang, C.F.; Lin, W.H.; Yeh, T.M.; Chiang, T.S.; Chang, Y.F. Antimicrobial susceptibility of *Riemerella anatipestifer* isolated from ducks and the efficacy of ceftiofur treatment. *J. Vet. Diagn. Investig.* **2003**, *15*, 26–29. [[CrossRef](#)]
4. Grantham, R.; Gautier, C.; Gouy, M.; Mercier, R.; Pave, A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* **1980**, *8*, r49–r62. [[CrossRef](#)] [[PubMed](#)]
5. Grantham, R.; Gautier, C.; Gouy, M.; Jacobzone, M.; Mercier, R. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* **1981**, *9*, r43–r74. [[CrossRef](#)] [[PubMed](#)]
6. Liu, G.; Wu, J.; Yang, H.; Bao, Q. Codon usage patterns in *Corynebacterium glutamicum*: Mutational bias, natural selection and amino acid conservation. *Comp. Funct. Genom.* **2010**, *2010*, 1–7.
7. Sharp, P.M.; Emery, L.R.; Zeng, K. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, *365*, 1203–1212. [[CrossRef](#)] [[PubMed](#)]
8. Venton, D. Highlight: Tiny bacterial genome opens a huge mystery: At mutational bias in *Hodgkinia*. *Genome Biol. Evol.* **2012**, *4*, 28–29. [[CrossRef](#)] [[PubMed](#)]
9. Jiang, P.; Sun, X.; Lu, Z. Analysis of synonymous codon usage in *Aeropyrum pernix* K1 and other *Crenarchaeota* microorganisms. *J. Genet. Genom.* **2007**, *34*, 275–284. [[CrossRef](#)]
10. Peixoto, L.; Zavala, A.; Romero, H.; Musto, H. The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium meliloti*. *Gene* **2003**, *320*, 109–116. [[CrossRef](#)]

11. Nayak, K.C. Comparative study on factors influencing the codon and amino acid usage in *Lactobacillus sakei* 23K and 13 other lactobacilli. *Mol. Biol. Rep.* **2012**, *39*, 535–545. [[CrossRef](#)] [[PubMed](#)]
12. Gu, W.; Tong, Z.; Ma, J.; Xiao, S.; Lu, Z. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *BioSystems* **2004**, *73*, 89–97. [[CrossRef](#)] [[PubMed](#)]
13. Ma, J.; Zhou, T.; Gu, W.; Sun, X.; Lu, Z. Cluster analysis of the codon use frequency of MHC genes from different species. *BioSystems* **2002**, *65*, 199–207. [[CrossRef](#)]
14. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **1985**, *2*, 13–34. [[PubMed](#)]
15. Cai, M.S.; Cheng, A.C.; Wang, M.S.; Zhao, L.C.; Zhu, D.K.; Luo, Q.H.; Liu, F.; Chen, X.Y. Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. *Intervirology* **2009**, *52*, 266–278. [[CrossRef](#)] [[PubMed](#)]
16. Sheng, Z.; Qin, Z.; Chen, Z.; Zhao, Y.; Zhong, J. The factors shaping synonymous codon usage in the genome of *Burkholderia mallei*. *J. Genet. Genom.* **2007**, *34*, 362–372.
17. Ma, X.X.; Feng, Y.P.; Bai, J.L.; Zhang, D.R.; Lin, X.S.; Ma, Z.R. Nucleotide composition bias and codon usage trends of gene populations in *Mycoplasma capricolum* subsp. *Capricolum* and *M. Agalactiae*. *J. Genet.* **2015**, *94*, 251–260. [[CrossRef](#)] [[PubMed](#)]
18. Sharp, P.M.; Li, W.H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [[CrossRef](#)] [[PubMed](#)]
19. Elhaik, E.; Landan, G.; Graur, D. Can GC content at third-codon positions be used as a proxy for isochore composition? *Mol. Biol. Evol.* **2009**, *26*, 1829–1833. [[CrossRef](#)] [[PubMed](#)]
20. Wright, F. The “effective number of codons” used in a gene. *Gene* **1990**, *87*, 23–29. [[CrossRef](#)]
21. Palacios, C.; Wernegreen, J.J. A strong effect of mutational bias on amino acid usage in *Buchnera* is mitigated at high-expression genes. *Mol. Biol. Evol.* **2002**, *19*, 1575–1584. [[CrossRef](#)] [[PubMed](#)]
22. Rispe, C.; Delmotte, F.; van Ham, R.C.; Moya, A. Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. *Genome Res.* **2004**, *14*, 44–53. [[CrossRef](#)] [[PubMed](#)]
23. Ohkubo, S.; Muto, A.; Kawauchi, Y.; Yamao, F.; Osawa, S. The ribosomal protein gene cluster of *Mycoplasma capricolum*. *Mol. Gen. Genet.* **1987**, *210*, 314–322. [[CrossRef](#)] [[PubMed](#)]
24. Wright, F.; Bibb, M.J. Codon usage in the G + C-rich *Streptomyces* genome. *Gene* **1992**, *113*, 55–65. [[CrossRef](#)]
25. Shields, D.C. Switches in species-specific codon preferences: The influence of mutation biases. *J. Mol. Evol.* **1990**, *31*, 71–80. [[CrossRef](#)] [[PubMed](#)]
26. Sueoka, N. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2653–2657. [[CrossRef](#)] [[PubMed](#)]
27. Sueoka, N. Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *J. Mol. Evol.* **1999**, *49*, 49–62. [[CrossRef](#)] [[PubMed](#)]
28. Sharp, P.M.; Li, W.H. The codon Adaptation Index—A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **1987**, *15*, 1281–1295. [[CrossRef](#)] [[PubMed](#)]
29. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [[CrossRef](#)] [[PubMed](#)]
30. Carbone, A.; Kepes, F.; Zinovyev, A. Codon bias signatures, organization of microorganisms in codon space, and lifestyle. *Mol. Biol. Evol.* **2005**, *22*, 547–561. [[CrossRef](#)] [[PubMed](#)]
31. Jon, B.; Ola, B.; Tammi, V.; Eystein, S.; Ussery, D.W. Amino acid usage is asymmetrically biased in AT- and GC-rich microbial genomes. *PLoS ONE* **2013**, *8*, e69878.
32. Muto, A.; Osawa, S. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 166–169. [[CrossRef](#)] [[PubMed](#)]
33. Hershberg, R.; Petrov, D.A. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.* **2010**, *6*, e1001115. [[CrossRef](#)] [[PubMed](#)]
34. Hildebrand, F.; Meyer, A.; Eyre-Walker, A. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet.* **2010**, *6*, e1001107. [[CrossRef](#)] [[PubMed](#)]
35. Wu, G. Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. *Microbiology* **2005**, *151*, 2175–2187. [[CrossRef](#)] [[PubMed](#)]
36. Wu, G.; Nie, L.; Zhang, W. Predicted highly expressed genes in *Nocardia farcinica* and the implication for its primary metabolism and nocardial virulence. *Antonie Van Leeuwenhoek* **2006**, *89*, 135–146. [[CrossRef](#)] [[PubMed](#)]

37. Willenbrock, H.; Friis, C.; Juncker, A.S.; Ussery, D.W. An environmental signature for 323 microbial genomes based on codon adaptation indices. *Genome Biol.* **2006**, *7*, 1–19. [[CrossRef](#)] [[PubMed](#)]
38. Cristina, J.; Moreno, P.; Moratorio, G.; Musto, H. Genome-wide analysis of codon usage bias in *Ebolavirus*. *Virus Res.* **2015**, *196*, 87–93. [[CrossRef](#)] [[PubMed](#)]
39. Subramanian, A.; Sarkar, R.R. Comparison of codon usage bias across *Leishmania* and *Trypanosomatids* to understand mRNA secondary structure, relative protein abundance and pathway functions. *Genomics* **2015**, *106*, 232–241. [[CrossRef](#)] [[PubMed](#)]
40. Ma, Q.P.; Li, C.; Wang, J.; Wang, Y.; Ding, Z.T. Analysis of synonymous codon usage in *FAD7* genes from different plant species. *Genet. Mol. Res.* **2015**, *14*, 1414–1422. [[CrossRef](#)] [[PubMed](#)]
41. Singha, H.S.; Chakraborty, S.; Deka, H. Stress induced *MAPK* genes show distinct pattern of codon usage in *Arabidopsis thaliana*, *Glycine max* and *Oryza sativa*. *Bioinformatics* **2014**, *10*, 436–442. [[CrossRef](#)] [[PubMed](#)]
42. Mavromatis, K.; Lu, M.; Misra, M.; Lapidus, A.; Nolan, M.; Lucas, S.; Hammon, N.; Deshpande, S.; Cheng, J.F.; Tapia, R.; et al. Complete genome sequence of *Riemerella anatipestifer* type strain (ATCC 11845). *Stand. Genom. Sci.* **2011**, *4*, 144–153. [[CrossRef](#)]
43. Wang, X.; Zhu, D.; Wang, M.; Cheng, A.; Jia, R.; Zhou, Y.; Chen, Z.; Luo, Q.; Liu, F.; Wang, Y.; et al. Complete genome sequence of *Riemerella anatipestifer* reference strain. *J. Bacteriol.* **2012**, *194*, 3270–3271. [[CrossRef](#)] [[PubMed](#)]
44. Wang, X.; Ding, C.; Han, X.; Wang, S.; Yue, J.; Hou, W.; Cao, S.; Zou, J.; Yu, S. Complete genome sequence of *Riemerella anatipestifer* serotype 1 strain CH3. *Genome Announc.* **2015**, *3*, e01594-14. [[CrossRef](#)] [[PubMed](#)]
45. Zhou, Z.; Peng, X.; Xiao, Y.; Wang, X.; Guo, Z.; Zhu, L.; Liu, M.; Jin, H.; Bi, D.; Li, Z.; et al. Genome sequence of poultry pathogen *Riemerella anatipestifer* strain RA-YM. *J. Bacteriol.* **2011**, *193*, 1284–1285. [[CrossRef](#)] [[PubMed](#)]
46. Wang, X.; Liu, W.; Zhu, D.; Yang, L.; Liu, M.; Yin, S.; Wang, M.; Jia, R.; Chen, S.; Sun, K.; et al. Comparative genomics of *Riemerella anatipestifer* reveals genetic diversity. *BMC Genom.* **2014**, *15*, 66–69. [[CrossRef](#)] [[PubMed](#)]
47. Yuan, J.; Liu, W.; Sun, M.; Song, S.; Cai, J.; Hu, S. Complete genome sequence of the pathogenic bacterium *Riemerella anatipestifer* strain RA-GD. *J. Bacteriol.* **2011**, *193*, 2896–2897. [[CrossRef](#)] [[PubMed](#)]
48. Wang, X.; Ding, C.; Wang, S.; Han, X.; Yu, S. Whole-Genome sequence analysis and Genome-Wide virulence gene identification of *Riemerella anatipestifer* strain Yb2. *Appl. Environ. Microbiol.* **2015**, *81*, 5093–5102. [[CrossRef](#)] [[PubMed](#)]
49. Yuan, J.; Li, L.; Sun, M.; Dong, J.; Hu, Q. Genome sequence of avirulent *Riemerella anatipestifer* strain RA-SG. *Genome Announc.* **2013**, *1*, e21812. [[CrossRef](#)] [[PubMed](#)]
50. Zhang, T.; Zhang, R.; Luo, Q.; Wen, G.; Ai, D.; Wang, H.; Luo, L.; Wang, H.; Shao, H. Genome sequence of avirulent *Riemerella anatipestifer* strain RA-JLLY. *Genome Announc.* **2015**, *3*. [[CrossRef](#)] [[PubMed](#)]
51. Song, X.H.; Zhou, W.S.; Wang, J.B.; Liu, M.F.; Wang, M.S.; Cheng, A.C.; Jia, R.Y.; Chen, S.; Sun, K.F.; Yang, Q.; et al. Genome sequence of *Riemerella anatipestifer* strain RCAD0122, a multidrug-resistant isolate from ducks. *Genome Announc.* **2016**, *4*. [[CrossRef](#)] [[PubMed](#)]
52. Sueoka, N. Translation-coupled violation of parity rule 2 in human genes is not the cause of heterogeneity of the DNA G + C content of third codon position. *Gene* **1999**, *238*, 53–58. [[CrossRef](#)]
53. Greenacre, M.J. *Theory and Applications of Correspondence Analysis*; Academic Press: Pittsburgh, PA, USA, 1984; pp. 326–339.
54. Lu, H.; Zhao, W.M.; Zheng, Y.; Wang, H.; Qi, M.; Yu, X.P. Analysis of synonymous codon usage bias in *Chlamydia*. *Acta Biochim. Biophys. Sin.* **2005**, *37*, 1–10. [[CrossRef](#)] [[PubMed](#)]
55. Codon, W. Available online: <http://codonw.sourceforge.net/> (accessed on 9 August 2016).
56. Deng, W.; Wang, Y.; Liu, Z.; Cheng, H.; Xue, Y. HemI: A toolkit for illustrating heatmaps. *PLoS ONE* **2014**, *9*, e111988. [[CrossRef](#)] [[PubMed](#)]
57. CAIcal Sever. Available online: <http://genomes.urv.es/CAIcal/intro.php> (accessed on 9 August 2016).
58. BlastKOALA. Available online: <http://www.kegg.jp/blastkoala/> (accessed on 9 August 2016).

