

RESEARCH ARTICLE

Open Access

# Clustering of classical swine fever virus isolates by codon pair bias

Immanuel Leifer<sup>1\*</sup>, Dirk Hoepfer<sup>2</sup>, Sandra Blome<sup>2</sup>, Martin Beer<sup>2</sup> and Nicolas Ruggli<sup>1</sup>

## Abstract

**Background:** The genetic code consists of non-random usage of synonymous codons for the same amino acids, termed codon bias or codon usage. Codon juxtaposition is also non-random, referred to as codon context bias or codon pair bias. The codon and codon pair bias vary among different organisms, as well as with viruses. Reasons for these differences are not completely understood. For classical swine fever virus (CSFV), it was suggested that the synonymous codon usage does not significantly influence virulence, but the relationship between variations in codon pair usage and CSFV virulence is unknown. Virulence can be related to the fitness of a virus: Differences in codon pair usage influence genome translation efficiency, which may in turn relate to the fitness of a virus. Accordingly, the potential of the codon pair bias for clustering CSFV isolates into classes of different virulence was investigated.

**Results:** The complete genomic sequences encoding the viral polyprotein of 52 different CSFV isolates were analyzed. This included 49 sequences from the GenBank database (NCBI) and three newly sequenced genomes. The codon usage did not differ among isolates of different virulence or genotype. In contrast, a clustering of isolates based on their codon pair bias was observed, clearly discriminating highly virulent isolates and vaccine strains on one side from moderately virulent strains on the other side. However, phylogenetic trees based on the codon pair bias and on the primary nucleotide sequence resulted in a very similar genotype distribution.

**Conclusion:** Clustering of CSFV genomes based on their codon pair bias correlate with the genotype rather than with the virulence of the isolates.

## Background

Classical swine fever (CSF) is a serious and highly contagious disease of pigs that can cause important economical losses in the pig industries [1,2]. The disease is caused by the classical swine fever virus (CSFV), currently endemic in wild boars and in part also in domestic pigs in Asia, South America, and parts of Central and Eastern Europe [1,3,4]. Depending on the isolate, the disease can vary from an acute hemorrhagic fever to a chronic or unapparent infection. An acute infection with a highly virulent strain manifests with high fever, respiratory and gastrointestinal symptoms, multiple haemorrhages, neurological disorders, and a high mortality rate [5]. Chronic infections may not be immediately recognized due to the mild symptoms. Infections with

low virulent isolates can remain unapparent. Thus CSFV viruses are divided into strains of highly, moderately, and low to avirulent strains (mainly vaccine strains) [6,7], see also Table 1 with the references therein. A number of live attenuated vaccines are available. These vaccines are mostly based on the Chinese vaccine strain (C-strain) and are completely avirulent [8-10].

CSFV is classified within the genus *Pestivirus* of the family *Flaviviridae* together with Border disease virus (BDV) and Bovine viral diarrhoea virus (BVDV) [33]. Pestiviruses possess a single-stranded positive-sense RNA genome of approximately 12300 nucleotides, with 5'-terminal and 3'-terminal non-translated regions (5'-NTR, 3'-NTR) [34]. The genome encodes one polyprotein that is co- and post-translationally processed by the viral proteases N<sup>P<sup>ro</sup></sup>, NS2, NS3, and by cellular proteases [34]. The polyprotein is cleaved in the four structural proteins C, E<sup>ns</sup>, E1, E2, and in the eight non-structural

\* Correspondence: immanuel.leifer@ivi.admin.ch

<sup>1</sup>Institute of Virology and Immunoprophylaxis (IVI), Sensemattstrasse 293, CH-3147 Mittelhäusern, Switzerland

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

**Table 1 Overview of the CSFV strains used for this study.**

Isolate	Genotype	Virulence status	GenBank (NCBI) entry	Reference
ALD D49532 hv	1.1	hv	D49532	[11]
Alfort187 × 87939 hv	1.1	hv	X87939	[12]
AlfortA19 U90951 hv	1.1	hv	U90951	[13]
Brescia AF091661 hv	1.2	hv	AF091661	[14]
Brescia M31768 lv <sup>1</sup>	1.2	hv	M31768	[14]
BRESCIAX AY578687 hv	1.2	hv	AY578687	[13]
CAP X96550 lv <sup>2</sup>	1.1	Low virulent	X96550	[15]
cF114 AF333000 hv	1.1	hv	AF333000	[16]
Eystrup AF326963 hv	1.1	hv	AF326963	[17]
Eystrup NC_002657 hv	1.1	hv	NC_002657	[17]
Glentorf U45478 lv <sup>3</sup>	1.1	Low virulent	U45478	[18]
JL106 EU497410 hv	1.1	hv	EU497410	[19]
Koslov HM237795 hv	1.1	hv	HM237795	[19]
Shimen AF092448 hv	1.1	hv	AF092448	unpublished
Shimen-HVRI AY775178 hv	1.1	hv	AY775178	[19]
SWH DQ127910 hv	1.1	hv	DQ127910	[6]
C_strain AY259122 va	1.1	va	AY259122	[17]
C_strain AY382481 va	1.1	va	AY382481	unpublished
C_strain AY663656 va	1.1	va	AY663656	unpublished
C_strain C-ZJ-2008 va	1.1	va	HM175885	unpublished
C_strain HCLV AF531433 va	1.1	va	AF531433	unpublished
C_strain HVRI AY805221 va	1.1	va	AY805221	unpublished
C_strain U45477 va	1.1	va	U45477	unpublished
C_strain Z46258 va	1.1	va	Z46258	[20]
flc-LOM EU915211 va	1.1	va	EU915211	
GPE- D49533 va	1.1	va	D49533	[11]
HCLV AF091507 va	1.1	va	AF091507	[21]
India vaccine EU857642 va	1.1	va	EU857642	unpublished
LOM EU789580 va	1.1	va	EU789580	[22]
LPC AF352565 va	1.1	va	AF352565	[23]
RUCSFPLUM AY578688 va	1.2	va	AY578688	[13]
Thiverval EU490425 va	1.1	va	EU490425	[24]
944IL94TWN AY646427 mv	3.4	mv	AY646427	[25]
Alfort-Tuebingen J04358 mv	2.3	mv	J04358	[26]
Borken GU233731 mv	2.3	mv	GU233731	[3]
CSF 39 AF407339 mv <sup>4</sup>	recombinant	mv-hv	AF407339	[27]
Euskirchen GU233732 mv	2.3	mv	GU233732	[3]
GXW_Z02 AY367767 mv	2.1	mv	AY367767	[27]
Hennef GU233733 mv	2.3	mv	GU233733	[3]
Jambul CSF0864 mv	2.3	mv	HQ148062	[28]
Novska CSF0821 mv	2.3	mv	HQ148061	[28]
Paderborn GQ902941 mv	2.1	mv	GQ902941	[29,30]
Penevezys CSF1048 mv	2.1	mv	HQ148063	[28]
Roesrath GU233734mv	2.3	mv	GU233734	[3]
Sp01 FJ265020 mv	2.3	mv	FJ265020	unpublished
Uelzen GU324242 mv	2.3	mv	GU324242	[3]
96TD AY554397 uk	2.1	uk	AY554397	unpublished

**Table 1 Overview of the CSFV strains used for this study. (Continued)**

0406CH01TWN AY568569 uk	2.1	uk	AY568569	unpublished
HEBZ GU592790 uk	2.1	uk	GU592790	unpublished
SXCDK GQ923951 uk	2.1	uk	GQ923951	unpublished
SXYL2006 GQ122383 uk	2.1	uk	GQ122383	unpublished
ZJ0801 FJ529205 uk	2.1	uk	FJ529205	unpublished

Virulence status: highly virulent (hv), moderately virulent (mv), low virulent (lv) or vaccine strains (va), and unknown virulence (uk) is indicated according to the information available. If available, the references to the sequences are indicated.

<sup>1</sup> Brescia M31768 lv<sup>1</sup> is representing the sequence of strain Brescia C1.1.1 which is a low virulent strain obtained after the 30<sup>th</sup> passage of strain Brescia on PK-15 cells [31].

<sup>2</sup> CAP X96550 lv<sup>2</sup> is described as highly virulent strain in some publications, but was originally described as cell culture adapted strain of low virulence [15].

<sup>3</sup> Glentorf U45478 lv<sup>3</sup> is described as low virulent or as highly virulent strain, depending on the report. In this study it is considered to be low virulent according to the publication of Handel et al. and Ahrens et al. [18,32]

<sup>4</sup> CSF 39 AF407339 mv<sup>4</sup> is a recombinant CSFV from China [27]. The virulence of this strain cannot be related to a particular genotype because the 5'NTR and the 3'NTR as well as the NS5A/B genes are homologous to genotype 1.1 strains, while the structural genes are homologous to genotype 2.1 strains. Furthermore the sequence of the original isolate is not known since the 32<sup>nd</sup> cell culture passage was used for sequence analysis.

proteins N<sup>pro</sup>, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B [35-37].

By phylogenetic classification, the CSFV strains are divided into the three different genotypes 1, 2, and 3, every genotype consisting of three to four subgenotypes [38]. The most recent CSFV outbreaks in the European Community are associated essentially with isolates that cluster in the genotype 2. In general, the genotype 2 isolates are of moderate or low virulence, as are the isolates of genotype 3 [3,4,39]. To the best of our knowledge, the CSFV strains with the highest virulence identified so far all belong to genotype 1. However, there is no absolute relationship between genotype and virulence, as there are also low virulent field isolates (e.g. "Glentorf") of genotype 1, and isolates within the genotype 2 group (e.g. "Uelzen") for which infection of piglets results in higher mortality than with most genotype 2 isolates. In addition, all vaccine strains belong also to genotype 1, as they were derived from genotype 1 strains by attenuation through multiple passages in non natural hosts, typically rabbits and guinea pigs, or in cell cultures derived from them.

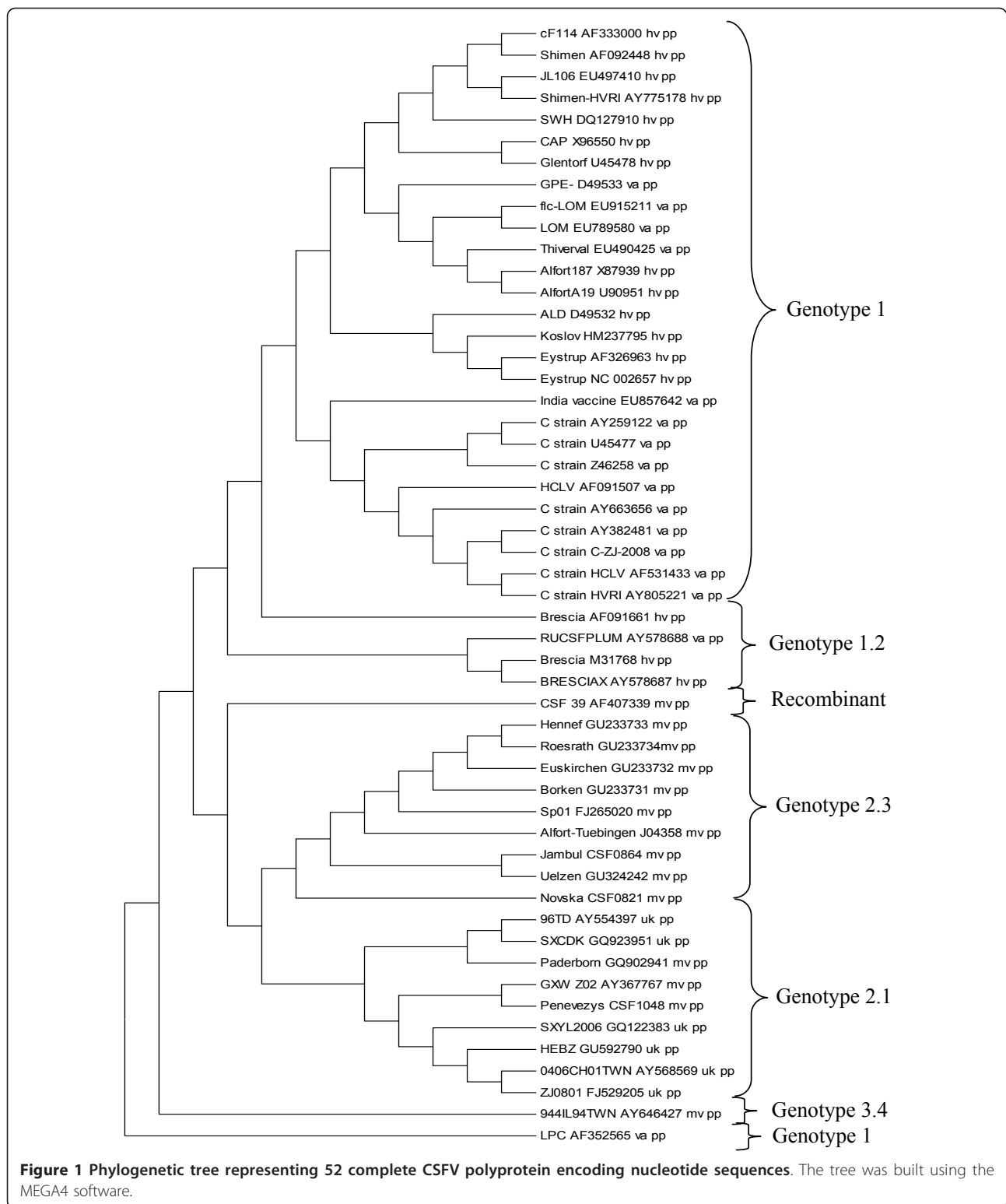
Various experimental approaches were implemented with the aim of identifying the virulence determinants related to a particular CSFV isolate. Numerous mutants with deletions, insertions, peptide or amino acid exchanges were analyzed and described in detail [40-46]. All mutants described so far were attenuated, leading to the conclusions that the modified positions may be relevant for the virulence of a specific strain or of CSFV in general. Certainly, strain-specific virulence factors determine whether an infection results in acute hemorrhagic fever, chronic disease or subclinical infection. Whether virulence determinants can be associated with particular amino acid positions remains unanswered. From a general point of view however, one may speculate that virulence depends mostly on the speed and level of virus replication. For poliovirus and influenza virus it was

shown that the codon pair bias can influence fitness and virulence [47,48]. The codon pair bias refers to the non random juxtaposition of codons, while the non random usage of synonymous codons for the same amino acids is referred to as codon bias. Previous studies showed that differences in synonymous codon usage did not relate to the virulence of CSFV isolates [7]. There is no analysis of codon pair usage of CSFV available. Therefore, the aim of this study was to investigate whether the codon pair usage of CSFV may relate to virulence or simply cluster the isolates into their genotype.

## Results

### Sequencing of complete genomes of recent CSFV isolates

In order to include some of the latest European CSFV isolates in the codon pair bias analysis, the genome of three recent field isolates were sequenced. The complete nucleotide sequences of the isolates CSFV/2.3/dp/CSF0821/2002/HR/Novska, CSFV/2.3/dp/CSF864/2007/BG/Jambul, and CSFV/2.1/dp/CSF1048/2009/LT/Penevezys were deposited to the NCBI GenBank nucleotide database [GenBank: HQ148061-HQ148063]. The genomes of the newly sequenced isolates encode a polyprotein of 3898 amino acids. The 5'NTRs of the three isolates are 373 nucleotides long. The 3'NTR is composed of 225 nucleotides for the "Novska" isolate and of 226 nucleotides for the two other isolates. These three sequences were included in a phylogenetic tree together with 49 complete CSFV genome sequences obtained from GenBank (Figure 1). The three newly sequenced isolates belong to the genotype 2, with the isolates "Novska" and "Jambul" clustering with the subgenotype 2.3 strains and the isolate "Penevezys" belonging to subgenotype 2.1. With experimental infections of pigs, the isolate "Penevezys" was classified as low to moderately virulent whereas the two other isolates were moderately virulent. Detailed information on the genotype, virulence, and origin of the 52 CSFV isolates analysed are provided in Table 1.



**The relative synonymous codon usage (RSCU) does not vary among different CSFV isolates**

In order to determine the variations in RSCU between CSFV isolates of different genotypes and virulence, the

frequency of each codon was determined for the 52 complete genome sequences available. As an example, the codon usage of three prototype isolates of different virulence, the low virulent "Glentorf" strain, the highly

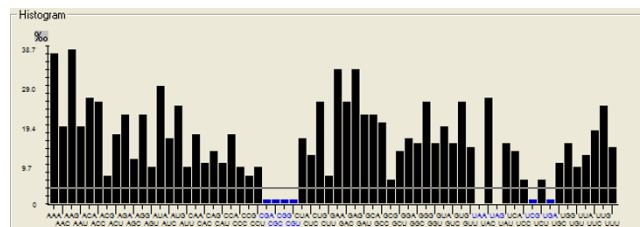
virulent “Koslov” strain, and the moderately virulent “Euskirchen” isolate is shown in Figure 2A-C . All three virus isolates have a very similar RSCU pattern. The two codons encoding the amino acid lysine (AAA and AAG) are the most frequent codons appearing in the CSFV genomes. The AAG triplet is slightly preferred. AAA is found in average 142,4 times/polyprotein with a standard deviation of 3.8 whereas AAG is found in average 151,6 times/polyprotein with a standard deviation of 3.6, independently of genotype and virulence. For the amino acid arginine there is a total of six different codons possible: CGA, CGC, CGG, CGU, AGA, and AGG. The four codons CGA, CGC, CGG, CGU are amongst the rarest codons used in all isolates. Thus, arginine is encoded almost exclusively by AGA and AGG, but here again, no major differences between strains of different virulence can be observed. Overall, no significant differences were observed between the different isolates confirming

earlier results showing that the RSCU does not vary between strains of different virulence [7].

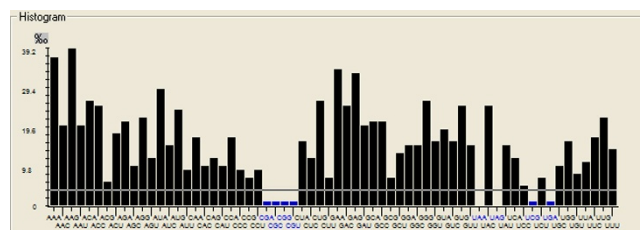
### The codon pair bias clusters CSFV into groups of different genotypes

Since the analysis of RSCU did not reveal any obvious differences among isolates of different virulence, it was of interest to determine whether the codon pair usage differs between CSFV isolates. To this end, the ANACONDA 2.0 software was applied to analyse the codon pair bias of the polyprotein encoding sequences of the 52 CSFV isolates. As opposed to the RSCU, clear differences were observed between different isolates (Figure 3). The codon pair analysis clustered the isolates in two groups, one representing the avirulent and the highly virulent strains, and the other the moderately virulent strains. The codon pairs CAA-AGA and GCA-GGG for instance are preferred by moderately virulent strains, but strongly rejected among vaccine viruses and highly

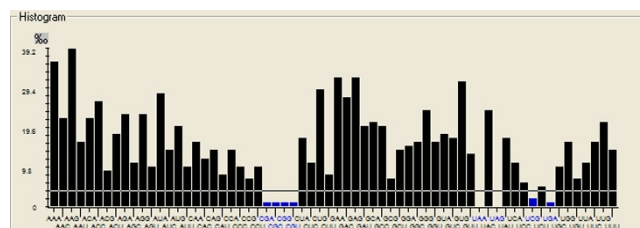
#### A Glentorf [GenBank:U45478]



#### B Koslov [GenBank:HM237795]

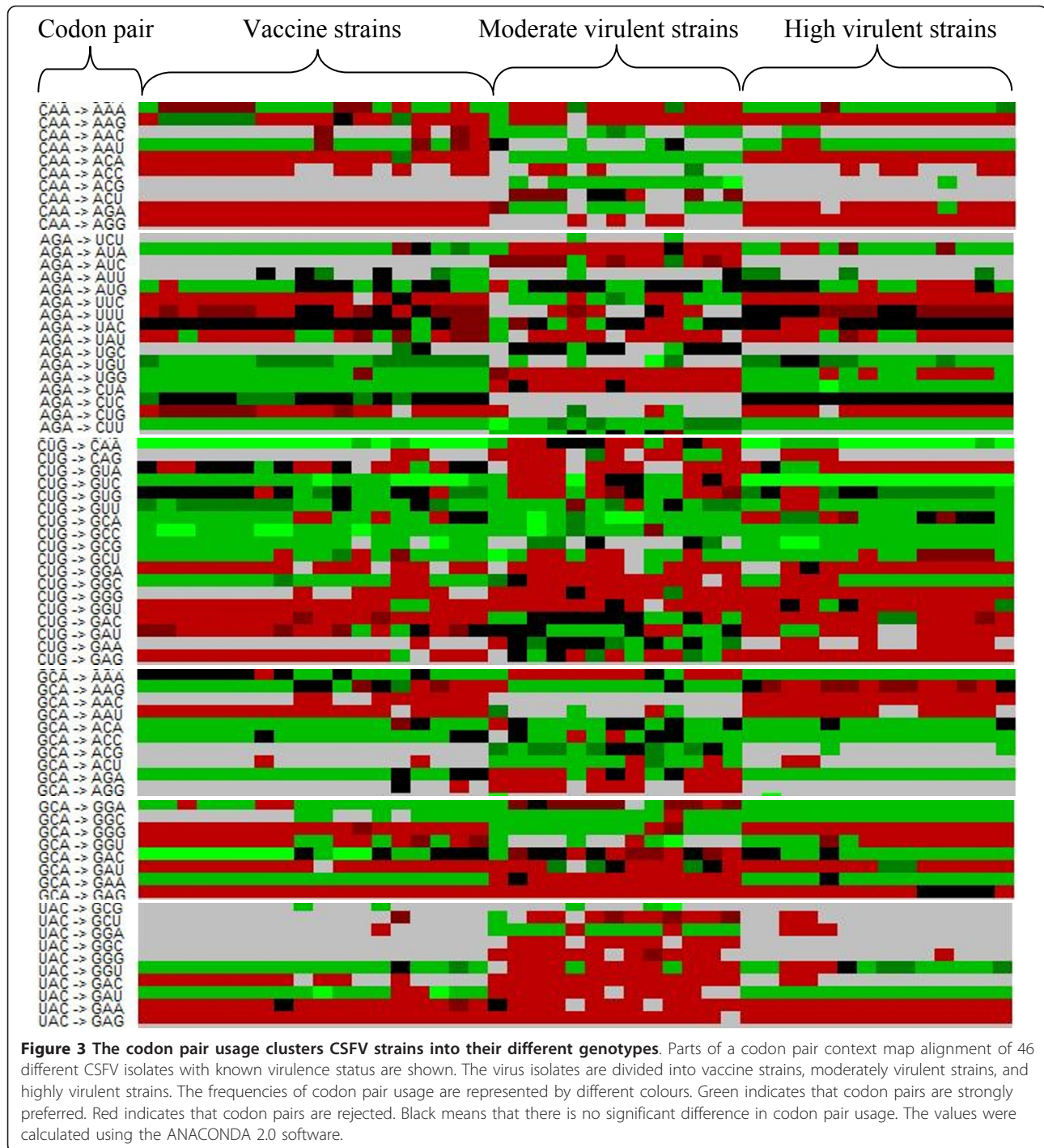


#### C Euskirchen [GenBank:GU233732]



**Figure 2** Relative synonymous codon usage exemplified with three prototype CSFV isolates. The histograms show the frequencies of synonymous codon usage in per mille for the strains “Glentorf” [GenBank:U45478], “Koslov” [GenBank:HM237795], and “Euskirchen” [GenBank:GU233732]. The values were calculated using the ANACONDA 2.0 software.

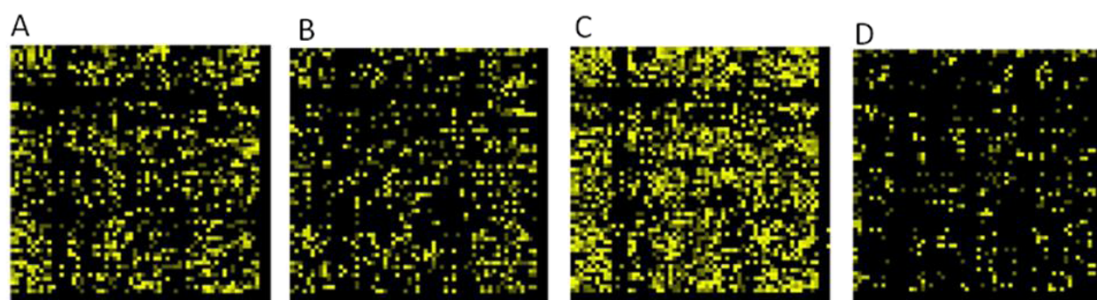




virulent strains. On the other hand, the pairs UAC-GGU, UAC-GAU, AGA-CUA, and GCA-GAA are preferred by vaccine and highly virulent viruses and are rejected by moderately virulent strains. Remarkably, the codon pairs with the pattern UAC-GGN and UAC-GAN strongly rejected among moderately virulent strains. Pairwise comparison of the complete codon pair context maps of vaccine strains among each other, and

those of vaccine and highly virulent strains revealed similar degrees of variability (exemplified in Figure 4A-D). Interestingly, the diversity within the group of vaccine strains is in some cases higher compared to the diversity between vaccine strains and highly virulent strains (Figure 4A-D).

It was also hypothesised that the codon pair bias may affect specifically the genome replication efficiency. In



**Figure 4** Codon pair bias overlays of vaccine and highly virulent CSFV strains using the differential display codon pair context tool of ANACONDA 2.0;  $61 \times 64$  codon pair bias matrices are shown. Yellow spots indicate differences in the corresponding codon pair usage whereas black colour means that codon pairs are used with similar residual values. Shown are codon pair overlays of the "GPE" vaccine strain and the parental highly virulent "ALD" strain (A), of the highly virulent "ALD" and "Koslov" strains (B), of the "GPE" and "C-strain Riems" vaccine viruses (C), and of the "C-strain Riems" and "HCLV" vaccine viruses (D).

order to determine whether the codon pair bias differs between the replicase and the structural proteins, which would suggest a potential effect of the codon pair bias on replication efficiency and virulence, artificial open reading frames ORFs were constructed covering the structural proteins and the NS5B protein of each CSFV strain. These ORFs were compared with respect to codon pair usage. No obvious differences in codon pair usage between structural and replicase genes were found, irrespectively of genotype and virulence (data not shown). Therefore, analysis of the individual genes did not allow discrimination between virulence either.

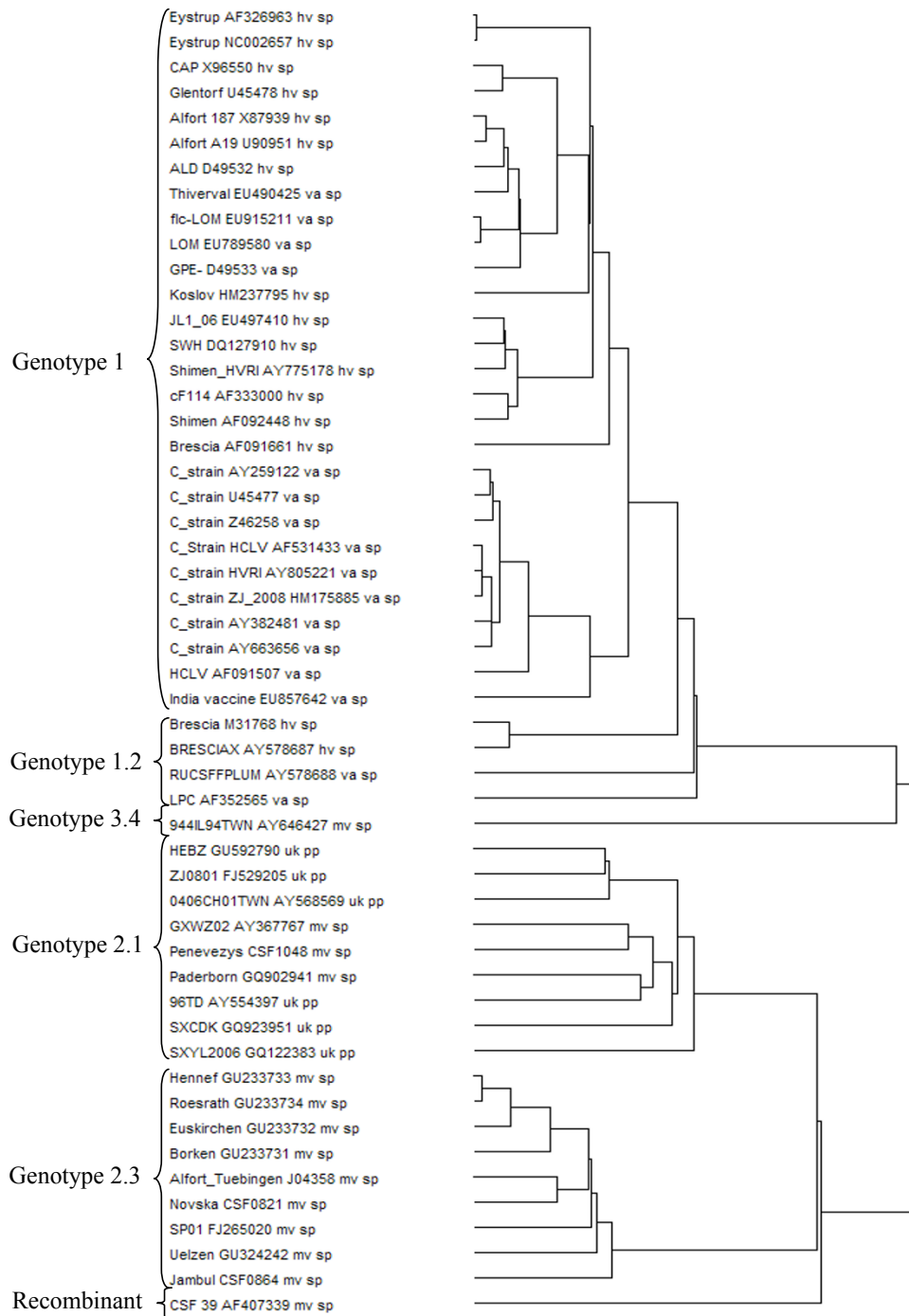
Finally, a phylogenetic tree based on the codon pair usage of the complete polyprotein encoding nucleotide sequences of the 52 isolates was constructed using the ANACONDA 2.0 software (Figure 5). The codon pair usage clusters the isolates in genotypes 1, 2, 3, and sub-genotypes 2.1 and 2.3, similarly to the phylogenetic tree based on the primary nucleotide sequence (compare Figures 1 and 5). Interestingly, some vaccine strains are grouped with the highly virulent strains, e. g. the strains "Alfort" and "Thiverval". According to these data, the codon pair bias clusters the CSFV isolates by genotype rather than by virulence.

## Discussion

Despite numerous efforts, CSFV virulence could not be linked to any particular genome sequence signature so far. Most if not all highly virulent CSFV strains belong to genotype 1 as do the vaccine strains (Table 1 and the references listed therein). Moderately virulent strains belong essentially to genotypes 2 and 3. The genetic variability within the genotype 1 is lower compared to strains of genotype 2 and 3 [49,50]. This lets hypothesize that sequence signatures of virulence may be found, especially with full sequence data of vaccine strains and parental highly virulent strains [51].

From the functional point of view, virulence may depend on viral replication efficiency, which can be influenced by differences in protein expression. Codon and codon pair bias can have an impact on translation efficiency and protein expression as it was shown for bacteria and yeast [52,53]. For poliovirus and influenza A virus, the artificial use of rare codons and of underrepresented codon pairs reduced viral protein translation and viral fitness, resulting in virus attenuation in vivo [47,48]. Consequently, a potential influence of codon usage and in particular of codon pair usage on CSFV virulence was considered. The analysis of the RSCU of 52 virus isolates covering the whole spectrum of virulence did not reveal any relationship with virulence. This confirmed earlier results obtained with the complete genome sequences of 35 isolates [7]. Thus codon usage between CSFV isolates is very similar, which is in agreement with the findings that RNA viruses of the same host category have the same codon usage preferences [54]. For the human immunodeficiency virus type-1, the RSCU is different from that of the human host. Adaptation towards human RSCU was attributed to the homogenization of the codon usage by mutation pressure rather than host adaptation [55].

Analysis of the codon pair bias of the complete coding sequence of the 52 isolates revealed a clear clustering (Figure 3). Vaccine strains and highly virulent strains showed mostly the same pattern, differing from the codon pair usage of moderately virulent strains. Because highly virulent and vaccine strains belong to genotype 1 and moderately virulent strains belong essentially to genotype 2, similarities in codon pair usage within a genotype might be due to the high proportion of sequence identity. Indeed, the genotype clustering obtained with phylogenetic analysis based on the codon pair usage and on the primary nucleotide sequence was nearly identical (compare Figures 5 and 1). Nevertheless, this does not exclude a possible relationship of the



**Figure 5** Phylogenetic tree of the CSFV polyprotein sequence alignment of 52 different isolates based on their codon pair bias, for analysis the ANACONDA 2.0 software was used.

codon pair usage with the virulence phenotype. The codon pair UAC-GNN for instance is less preferred by CSFV strains of genotype 2. Cytosin-phosphatidyl-Guanin (CpG) dinucleotides are signals for DNA methylation in

eukaryotes and regulate gene expression [56,57]. A reduction of UAC-GNN codon pair usage could reflect a host-specific adaptation, as it might influence the host anti-viral response as described for other viruses [58,59]. For CSFV



it is unknown whether adaptation to the host is linked to a gain of viral fitness. One could hypothesize that highly virulent CSFV strains would emerge through increased viral replication in the host. However, adaptation to the host is likely to result in optimized rather than in enhanced replication since occurrence of higher virulent strains has not been observed in CSFV field isolates during the last years [3,4,39,60]. From the evolutionary point of view, natural selection or adaptation towards a moderately virulent strain makes sense, because the mortality of the host is lower [39,60,61]. In addition, failure in early diagnosis due to mild clinical symptoms contributes to the dissemination and survival of the virus [59]. Thus, a moderately or low virulent virus has a greater chance of circulating in a pig or wild boar population without being detected [62-64]. Hence the reduced virulence observed with the CSFV isolates from the more recent outbreaks in Europe could result from several driving forces representing advantages for the virus. In fact, during the last decades CSFV outbreaks in Europe and Asia were increasingly caused by genotype 2 and 3 isolates, while the older CSFV field isolates belong to genotype 1. This suggests that evolution of CSFV is directed towards genotype 2 and 3. However, it is unknown if this is applicable to South American isolates since sequence information is missing. The development of live attenuated CSFV vaccine strains was based on isolates belonging to the genotype 1, which explains the close phylogenetic relationship between highly virulent and vaccine strains within the same genotype. Interestingly, there are nevertheless obvious differences in codon pair usage among strains of genotype 1 as seen from the overlays of codon pair matrices. These differences are the most prominent between the two unrelated "GPE" and "C-strain" vaccine strains attenuated in guinea pigs and rabbits, respectively. It is likely that these differences are in part caused by the propagation of the viruses in different hosts.

## Conclusions

The present results describe the first extensive codon pair bias analysis of a representative number of CSFV isolates covering the complete spectrum of virulence. Overall, the CSFV strains can be grouped in two main clusters according to the codon pair usage. Thus codon pair bias analysis can support CSFV phylogeny. However, based on the data presented here, a direct link between the codon pair usage and CSFV virulence cannot be established.

## Methods

### Sequencing of complete CSFV genomes

Nucleotide sequence analysis of complete CSFV genome was performed by pyrosequencing with a FLX Genome Sequencer (Roche Diagnostics, Mannheim, Germany) as

described previously [3]. Briefly, full CSFV genome DNA fragments (obtained by long-range RT-PCR) were separated by agarose gel electrophoresis and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corporation, Orange, CA, USA) prior to analysis with the FLX Genome Sequencer. The 5'NTR and 3'NTR were sequenced using commercial kits for RACE RT-PCR (5'RACE System and 3'RACE System, Invitrogen, Carlsbad, CA, USA) according to the manufacturers recommendations. Minor modifications were performed as described previously [3]. The raw sequence data were assembled using the GS assembler software newbler (v. 2.0.00.22; Roche, Mannheim). The nucleotide sequence information was deposited to the NCBI GenBank nucleotide database [65].

### Sequence data source and additional sequence information

Complete genome sequences of 49 different CSFV isolates were obtained from the NCBI GenBank nucleotide database. Detailed information on the virus isolates is provided in Table 1. Virus isolates were grouped in highly virulent (hv), moderately virulent (mv), and low virulent (lv) or vaccine strains (va). According to the information available, 46 virus isolates were subdivided into these three groups composed of 16 highly virulent, 14 moderately virulent, and 16 vaccine strains (Table 1). For the remaining virus isolates virulence could not be determined.

### Analysis of RSCU and codon pair usage

The relative synonymous codon usage is expressed as RSCU value of a codon [53]. The RSCU value expresses the relationship between the observed and the expected codon frequency and was calculated with the ANACONDA 2.0 software (Universidade de Aveiro, Portugal) [66]. The codon context bias of the complete polyprotein encoding nucleotide sequence of 52 different CSFV isolates was investigated using the software package ANACONDA 2.0 as described [67-69]. In addition, different regions of the genomes were analysed separately. To this end, artificial ORFs for the NS5B and the structural protein genes were constructed by adding a start and stop codon to the corresponding coding regions. Codon pair biases were analysed according to their relative occurrence. Statistical calculation of the codon pairs is given in relation to its real occurrence and the expected incidence independently of their distribution. The ANACONDA 2.0 software displays a codon pair context map for each viral ORF. This context map consists of 3904 possible codon pairs given in a vertical row with one coloured square for each codon pair. The colours represent the frequency of occurrence: red coloured squares indicate codon pairs that are strongly

rejected, whereas preferred codon pairs are represented in green colour. Codon pairs represented by black squares are statistically not significant.

Phylogenetic trees based on codon pair bias were created with the ANACONDA 2.0 software. Neighbour-joining trees with the maximum composite likelihood method using complete polyprotein encoding nucleotide sequences were constructed with the MEGA4 software (Molecular Evolutionary Genetics Analysis, Center for Evolutionary Medicine and Informatics, Tempe, USA) software [69].

#### Acknowledgements

This study was supported by the Swiss Federal Veterinary Office (BVET) grant # 1.10.13. We thank Moctezuma Reimann for excellent technical assistance and the CSFV Community Reference Laboratory (Hannover, Germany) for providing the CSFV isolates.

#### Author details

<sup>1</sup>Institute of Virology and Immunoprophylaxis (IVI), Sensemattstrasse 293, CH-3147 Mittelhäusern, Switzerland. <sup>2</sup>Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Südufer 10, 17493 Greifswald-Insel Riems, Germany.

#### Authors' contributions

IL has done the statistical analyses and written the manuscript. NR is responsible for the design and supervision of the project and for the manuscript. DH has done most parts of the sequencing and MB has been involved in carefully revising the manuscript and giving substantial input during interpretation of the data. SB has determined the virulence phenotype of selected CSFV isolates in pigs. All authors have read and approved the final manuscript.

#### Conflict of interests

The authors declare that they have no competing interests.

Received: 6 October 2011 Accepted: 29 November 2011

Published: 29 November 2011

#### References

1. Edwards S, Fukusho A, Lefevre P, Lipowski A, Pejsak Z, Roehe P, Westergaard J: Classical swine fever: the global situation. *Vet Microbiol* 2000, **73**:103-119.
2. Vandeputte J, Chappuis G: Classical swine fever: the European experience and a guide for infected areas [In Process Citation]. *Rev Sci Tech* 1999, **18**:638-647.
3. Leifer I, Hoffmann B, Hoper D, Rasmussen TB, Blome S, Strebelow G, Horeth-Bontgen D, Staubach C, Beer M: Molecular epidemiology of current classical swine fever virus isolates of wild boar in Germany. *J Gen Virol* 2010, **91**:2687-2697.
4. Pol F, Rossi S, Mesplede A, Kuntz-Simon G, Le Potier MF: Two outbreaks of classical swine fever in wild boar in France. *Vet Rec* 2008, **162**:811-816.
5. Kaden V, Ziegler U, Lange E, Dedek J: Classical swine fever virus: clinical, virological, serological and hematological findings after infection of domestic pigs and wild boars with the field isolate "Spante" originating from wild boar. *Berl Munch Tierarztl Wochenschr* 2000, **113**:412-416.
6. Li X, Xu Z, He Y, Yao Q, Zhang K, Jin M, Chen H, Qian P: Genome comparison of a novel classical swine fever virus isolated in China in 2004 with other CSFV strains. *Virus Genes* 2006, **33**:133-142.
7. Tao P, Dai L, Luo M, Tang F, Tien P, Pan Z: Analysis of synonymous codon usage in classical swine fever virus. *Virus Genes* 2009, **38**:104-112.
8. Kaden V, Lange E, Fischer U, Strebelow G: Oral immunisation of wild boar against classical swine fever: evaluation of the first field study in Germany. *Vet Microbiol* 2000, **73**:239-252.
9. Kaden V, Lange B: Oral immunisation against classical swine fever (CSF): onset and duration of immunity. *Vet Microbiol* 2001, **82**:301-310.
10. Kaden V, Lange E, Riebe R, Lange B: Classical swine fever virus Strain 'C'. How long is it detectable after oral vaccination? *J Vet Med B Infect Dis Vet Public Health* 2004, **51**:260-262.
11. Sharp PM, Tuohy TM, Mosurski KR: Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 1986, **14**:5125-5143.
12. Su MW, Lin HM, Yuan HS, Chu WC: Categorizing host-dependent RNA viruses by principal component analysis of their codon usage preferences. *J Comput Biol* 2009, **16**:1539-1547.
13. Meintjes PL, Rodrigo AG: Evolution of relative synonymous codon usage in Human Immunodeficiency Virus type-1. *J Bioinform Comput Biol* 2005, **3**:157-168.
14. Chan SW, Henderson IR, Jacobsen SE: Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* 2005, **6**:351-360.
15. Robertson KD: DNA methylation and human disease. *Nat Rev Genet* 2005, **6**:597-610.
16. Karlin S, Doerfler W, Cardon LR: Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J Virol* 1994, **68**:2889-2897.
17. Sugiyama T, Gursel M, Takeshita F, Coban C, Conover J, Kaisho T, Akira S, Klinman DM, Ishii KJ: CpG RNA: identification of novel single-stranded RNA that stimulates human CD14 + CD11c + monocytes. *J Immunol* 2005, **174**:2273-2279.
18. Everett H, Salguero FJ, Graham SP, Haines F, Johns H, Clifford D, Nunez A, La Rocca SA, Parchariyanon S, Steinbach F, et al: Characterisation of experimental infections of domestic pigs with genotype 2.1 and 3.3 isolates of classical swine fever virus. *Vet Microbiol* 2010, **142**:26-33.
19. Rivera H, Angeles R, Sandoval N, Manchego A: Persistencia del virus de la Peste Porcina Clásica de baja virulencia en el sistema nervioso central de lechones de granjas tecnificadas. *Rev Investig Vet* 1999, **10**:1-10.
20. Artois M, Delahay R, Guberti V, Cheeseman C: Control of infectious diseases of wildlife in Europe. *Vet J* 2001, **162**:141-152.
21. Fritzemeier J, Teuffert J, Greiser-Wilke I, Staubach C, Schluter H, Moennig V: Epidemiology of classical swine fever in Germany in the 1990s. *Vet Microbiol* 2000, **77**:29-41.
22. Laddomada A: Incidence and control of CSF in wild boar in Europe. *Vet Microbiol* 2000, **73**:121-130.
23. NCBI GenBank nucleotide database. 2011 [http://www.ncbi.nlm.nih.gov/nucleotide].
24. ANACONDA 2.0 software. 2011 [http://bioinformatics.ua.pt].
25. Moura G, Pinheiro M, Silva R, Miranda I, Afreixo V, Dias G, Freitas A, Oliveira JL, Santos MA: Comparative context analysis of codon pairs on an ORFeome scale. *Genome Biol* 2005, **6**:R28.
26. Moura G, Pinheiro M, Freitas AV, Oliveira JL, Santos MA: Computational and statistical methodologies for ORFeome primary structure analysis. *Methods Mol Biol* 2007, **395**:449-462.
27. Pinheiro M, Afreixo V, Moura G, Freitas A, Santos MA, Oliveira JL: Statistical, computational and visualization methodologies to unveil gene primary structure features. *Methods Inf Med* 2006, **45**:163-168.
28. Kumar S, Nei M, Dudley J, Tamura K: MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 2008, **9**:299-306.
29. Fauquet CM, Fargette D: International Committee on Taxonomy of Viruses and the 3,142 unassigned species. *Virus* 2005, **2**:64.
30. Ruggli N, Tratschin JD, Mittelholzer C, Hofmann MA: Nucleotide sequence of classical swine fever virus strain Alfort/187 and transcription of infectious RNA from stably cloned full-length cDNA. *J Virol* 1996, **70**:3478-3487.
31. Hulst MM, Van Gennip HG, Moormann RJ: Passage of classical swine fever virus in cultured swine kidney cells selects virus variants that bind to heparan sulfate due to a single amino acid change in envelope protein E(rns). *J Virol* 2000, **74**:9553-9561.
32. Ahrens U, Kaden V, Drexler C, Visser N: Efficacy of the classical swine fever (CSF) marker vaccine Porcilis Pesti in pregnant sows. *Vet Microbiol* 2000, **77**:83-97.
33. Laude H: Isolation of a cytolytic strain of hog cholera virus from IB-RS2 cells (author's transl). *Ann Microbiol (Paris)* 1978, **129**:553-561.
34. Nie Y, Ke Y, Chen J, Ding M: [Construction of the full-length cDNA clone of Chinese virulent strain-F114]. *Wei Sheng Wu Xue Bao* 2001, **41**:452-456.
35. Risatti G, Holinka L, Lu Z, Kutish G, Callahan JD, Nelson WM, Brea TE, Borca MV: Diagnostic evaluation of a real-time reverse transcriptase PCR

- assay for detection of classical swine fever virus. *J Clin Microbiol* 2005, **43**:468-471.
36. Moormann RJ, Warmerdam PA, van der Meer B, Schaaper WM, Wensvoort G, Hulst MM: Molecular cloning and nucleotide sequence of hog cholera virus strain Brescia and mapping of the genomic region encoding envelope protein E1. *Virology* 1990, **177**:184-198.
  37. Mayer D, Thayer TM, Hofmann MA, Tratschin JD: Establishment and characterisation of two cDNA-derived strains of classical swine fever virus, one highly virulent and one avirulent. *Virus Res* 2003, **98**:105-116.
  38. Handel K, Kehler H, Hills K, Pasick J: Comparison of reverse transcriptase-polymerase chain reaction, virus isolation, and immunoperoxidase assays for detecting pigs infected with low, moderate, and high virulent strains of classical swine fever virus. *J Vet Diagn Invest* 2004, **16**:132-138.
  39. Zhao JJ, Cheng D, Li N, Sun Y, Shi Z, Zhu QH, Tu C, Tong GZ, Qiu HJ: Evaluation of a multiplex real-time RT-PCR for quantitative and differential detection of wild-type viruses and C-strain vaccine of Classical swine fever virus. *Vet Microbiol* 2008, **126**:1-10.
  40. Moormann RJ, Van Gennip HG, Miedema GK, Hulst MM, van Rijn PA: Infectious RNA transcribed from an engineered full-length cDNA template of the genome of a pestivirus. *J Virol* 1996, **70**:763-770.
  41. Wu HX, Wang JF, Zhang CY, Fu LZ, Pan ZS, Wang N, Zhang PW, Zhao WG: Attenuated lapinized chinese strain of classical swine fever virus: complete nucleotide sequence and character of 3'-noncoding region. *Virus Genes* 2001, **23**:69-76.
  42. Cho HS, Park SJ, Park NY: Development of a reverse-transcription polymerase chain reaction assay with fluorogenic probes to discriminate Korean wild-type and vaccine isolates of Classical swine fever virus. *Can J Vet Res* 2006, **70**:226-229.
  43. Pan CH, Jong MH, Huang YL, Huang TS, Chao PH, Lai SS: Rapid detection and differentiation of wild-type and three attenuated lapinized vaccine strains of Classical swine fever virus by reverse transcription polymerase chain reaction. *J Vet Diagn Invest* 2008, **20**:448-456.
  44. Fan Y, Zhao Q, Zhao Y, Wang Q, Ning Y, Zhang Z: Complete genome sequence of attenuated low-temperature Thiverval strain of classical swine fever virus. *Virus Genes* 2008, **36**:531-538.
  45. Lin YJ, Chien MS, Deng MC, Huang CC: Complete sequence of a subgroup 3.4 strain of classical swine fever virus from Taiwan. *Virus Genes* 2007, **35**:737-744.
  46. Meyers G, Rumenapf T, Thiel HJ: Molecular cloning and nucleotide sequence of the genome of hog cholera virus. *Virology* 1989, **171**:555-567.
  47. He CQ, Ding NZ, Chen JG, Li YL: Evidence of natural recombination in classical swine fever virus. *Virus Res* 2007, **126**:179-185.
  48. Blome S, Grotha I, Moennig V, Greiser-Wilke I: Classical swine fever virus in South-Eastern Europe—retrospective analysis of the disease situation and molecular epidemiology. *Vet Microbiol* 2010, **146**:276-284.
  49. Uttenthal A, Storgaard T, Oleksiewicz MB, de Stricker K: Experimental infection with the Paderborn isolate of classical swine fever virus in 10-week-old pigs: determination of viral replication kinetics by quantitative RT-PCR, virus isolation and antigen ELISA. *Vet Microbiol* 2003, **92**:197-212.
  50. Rasmussen TB, Reimann I, Uttenthal A, Leifer I, Depner K, Schirmer H, Beer M: Generation of recombinant pestiviruses using a full-genome amplification strategy. *Vet Microbiol* 2009.
  51. Meyers G, Thiel HJ: Molecular characterization of pestiviruses. *Adv Virus Res* 1996, **47**:53-118.
  52. Rumenapf T, Unger G, Strauss JH, Thiel HJ: Processing of the envelope glycoproteins of pestiviruses. *J Virol* 1993, **67**:3288-3294.
  53. Tautz N, Elbers K, Stoll D, Meyers G, Thiel HJ: Serine protease of pestiviruses: determination of cleavage sites. *J Virol* 1997, **71**:5415-5422.
  54. Thiel HJ, Stark R, Weiland E, Rumenapf T, Meyers G: Hog cholera virus: molecular composition of virions from a pestivirus. *J Virol* 1991, **65**:4705-4712.
  55. Greiser-Wilke I, Depner K, Fritzsche J, Haas L, Moennig V: Application of a computer program for genetic typing of classical swine fever virus isolates from Germany. *J Virol Methods* 1998, **75**:141-150.
  56. Kaden V, Lange E, Polster U, Klopffleisch R, Teifke JP: Studies on the virulence of two field isolates of the classical Swine Fever virus genotype 2.3 rostock in wild boars of different age groups. *J Vet Med B Infect Dis Vet Public Health* 2004, **51**:202-208.
  57. Risatti GR, Borca MV, Kutish GF, Lu Z, Holinka LG, French RA, Tulman ER, Rock DL: The E2 glycoprotein of classical swine fever virus is a virulence determinant in swine. *J Virol* 2005, **79**:3787-3796.
  58. Risatti GR, Holinka LG, Lu Z, Kutish GF, Tulman ER, French RA, Sur JH, Rock DL, Borca MV: Mutation of E1 glycoprotein of classical swine fever virus affects viral virulence in swine. *Virology* 2005, **343**:116-127.
  59. Risatti GR, Holinka LG, Carrillo C, Kutish GF, Lu Z, Tulman ER, Sainz IF, Borca MV: Identification of a novel virulence determinant within the E2 structural glycoprotein of classical swine fever virus. *Virology* 2006, **355**:94-101.
  60. Risatti GR, Holinka LG, Fernandez S, Carrillo C, Lu Z, Borca MV: N-linked glycosylation status of classical swine fever virus strain Brescia E2 glycoprotein influences virulence in swine. *J Virol* 2007, **81**:924-933.
  61. Risatti GR, Holinka LG, Fernandez S, Carrillo C, Kutish GF, Lu Z, Zhu J, Rock DL, Borca MV: Mutations in the carboxyl terminal region of E2 glycoprotein of classical swine fever virus are responsible for viral attenuation in swine. *Virology* 2007, **364**:371-382.
  62. Sainz IF, Holinka LG, Lu Z, Risatti GR, Borca MV: Removal of a N-linked glycosylation site of classical swine fever virus strain Brescia Erns glycoprotein affects virulence in swine. *Virology* 2008, **370**:122-129.
  63. Van Gennip HG, Vlot AC, Hulst MM, de Smit AJ, Moormann RJ: Determinants of virulence of classical swine fever virus strain Brescia. *J Virol* 2004, **78**:8812-8823.
  64. Coleman JR, Papamichail D, Skiena S, Fitcher B, Wimmer E, Mueller S: Virus attenuation by genome-scale changes in codon pair bias. *Sci* 2008, **320**:1784-1787.
  65. Mueller S, Coleman JR, Papamichail D, Ward CB, Nimmual A, Fitcher B, Skiena S, Wimmer E: Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotechnol* 2010, **28**:723-726.
  66. Greiser-Wilke I, Dreier S, Haas L, Zimmermann B: Genetic typing of classical swine fever virus isolates - current status. *Deutsche Tierärztliche Wochenschrift* 2006, **113**:134-138.
  67. Paton DJ, McGoldrick A, Greiser-Wilke I, Parchariyanon S, Song JY, Liou PP, Stadejek T, Lowings JP, Bjorklund H, Belak S: Genetic typing of classical swine fever virus. *Vet Microbiol* 2000, **73**:137-157.
  68. Ishikawa K, Nagai H, Katayama K, Tsutsui M, Tanabayashi K, Takeuchi K, Hishiyama M, Saitoh A, Takagi M, Gotoh K: Comparison of the entire nucleotide and deduced amino acid sequences of the attenuated hog cholera vaccine strain GPE- and the wild-type parental strain ALD. *Arch Virol* 1995, **140**:1385-1391.
  69. Gouy M, Gautier C: Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* 1982, **10**:7055-7074.

doi:10.1186/1756-0500-4-521

Cite this article as: Leifer et al: Clustering of classical swine fever virus isolates by codon pair bias. *BMC Research Notes* 2011 **4**:521.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

