

RESEARCH

Open Access

# Analysis of the use of codon pairs in the HE gene of the ISA virus shows a correlation between bias in HPR codon-pair use and mortality rates caused by the virus

Mario Tello\*, José Miguel Saavedra and Eugenio Spencer

## Abstract

**Background:** Segment 6 of the ISA virus codes for hemoagglutinin-esterase (HE). This segment is highly variable, with more than 26 variants identified. The major variation is observed in what is called the high polymorphism region (HPR). The role of the different HPR zones in the viral cycle or evolution remains unknown. However viruses that present the HPR0 are avirulent, while viruses with important deletions in this region have been responsible for outbreaks with high mortality rates. In this work, using bioinformatic tools, we examined the influence of different HPRs on the adaptation of HE genes to the host translational machinery and the relationship to observed virulence.

**Methods:** Translational efficiency of HE genes and their HPR were estimated analyzing codon-pair bias (CPB), adaptation to host codon use (codon adaptation index - CAI) and the adaptation to available tRNAs (tAI). These values were correlated with reported mortality for the respective ISA virus and the  $\Delta G$  of RNA folding. tRNA abundance was inferred from tRNA gene numbers identified in the *Salmo salar* genome using tRNAScan-SE. Statistical correlation between data was performed using a non-parametric test.

**Results:** We found that HPR0 contains zones with codon pairs of low frequency and low availability of tRNA with respect to salmon codon-pair usage, suggesting that HPR modifies HE translational efficiency. Although calculating tAI was impossible because one third of tRNAs (~60.000) were tRNA-ala, translational efficiency measured by CPB shows that as HPR size increases, the CPB value of the HE gene decreases ( $P = 2 \times 10^{-7}$ ,  $\rho = -0.675$ ,  $n = 63$ ) and that these values correlate positively with the mortality rates caused by the virus ( $\rho = 0.829$ ,  $P = 2 \times 10^{-7}$ ,  $n = 11$ ). The mortality associated with different virus isolates or their corresponding HPR sizes were not related with the  $\Delta G$  of HPR RNA folding, suggesting that the secondary structure of HPR RNA does not modify virulence.

**Conclusions:** Our results suggest that HPR size affects the efficiency of gene translation, which modulates the virulence of the virus by a mechanism similar to that observed in production of live attenuated vaccines through deoptimization of codon-pair usage.

## Background

The infectious salmon anemia virus (ISAV) is an orthomyxovirus, composed of eight segments. [1]. Two phylogenetic groups have been identified, the European and the North American, which probably diverged at the beginning of the last century [2,3]. Within each group the viruses present minimal genetic differences with the

exception of certain regions of high variability localized in genes encoding for glycoprotein F [4] and HE. The region of hyper-variability in HE genes (HPR) is located between the transmembrane region and the globular domain with hemagglutinin-esterase activity [5,6]. The function of HPR within the viral cycle is unknown. However, the type of HPR in the virus has been associated with the degree of observed virulence. For example, the HPR0 is found in avirulent ISAV strains, while HPR7b is associated with ISAV outbreaks in Chile with high mortality rates [7-9]. The molecular ratios that explain the

\* Correspondence: mario.tello@usach.cl  
Centro de Biotecnología Acuicola, Facultad de Química y Biología,  
Universidad de Santiago de Chile, Alameda 3363, Santiago, Chile

relationship between the HPR and observed virulence have not yet been identified. However, the HPR does not modify esterase or hemagglutinin activity [10].

Far from being silent, synonymous mutations affect diverse cellular processes, such as translation kinetics, mRNA structure and protein folding [11]. These mutations are distributed differently in the genes according to their levels of expression. Highly expressed genes tend to use preferentially the most abundant synonymous codons [12,13]. The relationship between bias in synonymous codon usage and gene expression seems to be the consequence of the synchrony observed between the frequency of codon usage and the abundance of tRNA that recognizes it [12,14-19].

As well as the bias in codon usage, there is a bias in codon-pair usage, which is characteristic for each organism and independent of the frequency of codon usage [20]. As with the synonymous codons, the codon pairs are distributed in the genes differently according to their level of gene expression [21].

The ribosome traffic jam in the mRNA has been suggested as a common mechanism that explains the reduction in translation efficiency by introducing codons with low tRNA availability, or tRNAs with little compatibility among themselves, as well as the negative relationship between the formation of stable structures in the mRNA and the level of protein expression [22-25].

There are few examples among viruses of the influence of codon use on the viral cycle. Studies about the adaptation to host codon usage indicate that viral genes that codify for critical proteins tend to use the synonymous codons most represented in the host genome [26]. In Influenza A, the structure acquired by the 5' end of the codifying zone is fundamental for viral packaging [27,28]. In the hepatitis A virus, the presence of underrepresented codons in the host genome seems to be a strategy to avoid competition with cellular mRNAs and to increase adequate folding of the capsid protein [29,30]. In the influenza A and polio viruses, the introduction of synonymous codon pairs in the viral genome with a negative bias produces viable attenuated viral particles with reduced rates of protein synthesis and replication [31-33].

It has been suggested that the different HPR are probably the result of in-frame deletions from sectors of HPR0 [6,34]. These deletions could modify translation efficiency by elimination of underrepresented codons in a similar but opposite mechanism to what is observed in the generation in vitro of live attenuated vaccines by deoptimization of their codon-pair usage. In this work we analyze this hypothesis of the effect of HPR on the adaptation to host codon usage, tRNA availability, host codon-pair usage bias and the folding energy of the HE gene.

## Results

### The effect of the HPR on translation efficiency

Translation efficiency is governed by secondary structure of the mRNA [35] and by its adaptation to the tRNA pool [12]. To evaluate if the HPR affects virulence through the possible formation of a stable secondary structure that reduces the advance of ribosome, thus causing a decrease in the HE translation rate, we determined if there is a relationship between the HPR folding energy and observed virulence. Our results indicate the absence of a statistically significant relationship between folding  $\Delta G$  and the size of the HPR region or associated mortality. However, the HPR0 region showed the most favorable energy ( $\Delta G = -24.2$  Kcal/mol) (Additional file 1: Figure S1).

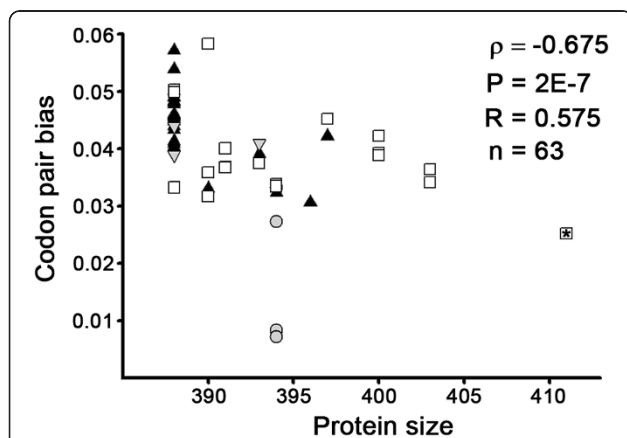
The effect of the presence of the different HPRs on translation efficiency can also be evaluated by analyzing the degree of adaptation to the tRNA pool through the use of the tAI value, which corresponds to the geometric mean of the relative abundance of tRNA in the cells that recognize the codons present in the gene. The relative abundance of the tRNA can be calculated based on the number of copies of each tRNA present in the genome [12]. Using the tRNAScan-SE program, we identified 63857 genes in the *Salmo salar* genome that codify for tRNAs. Approximately 21085 of these genes codify for pseudogenes and another 20546 codify for alanine tRNA. Of the rest of the tRNAs, their abundances range from 2724 for methionine tRNA to 236 for tryptophan tRNA. As well, 353 untyped tRNA were identified and 20 genes for tRNA suppressors (Additional file 2: Table S1). As in the calculation of tAI, the relative abundance of the tRNAs that recognize the codons in the gene was normalized with respect to the codon with greater availability of tRNAs. The extreme disproportion between Ala-tRNA copies and the other tRNA for the other amino acids makes interpretation of the data problematic and could lead to erroneous conclusions. To avoid this problem, we continue the study using adaptation to host codon use and bias in codon-pair usage.

### Adaptation to host codon usage and host codon pair bias of the ISAV hemagglutinin-esterase gene

The codon adaptation index (CAI) is a good indicator of the adaptation to the host codon usage and a good predictor of translation efficiency [13]. Previous studies have indicated that ISAV is the least adapted orthomyxovirus to its host's genome, sharing CAI values with less than 3% of the genes of *Salmo salar* (Tello, unpublished). Among these, the North American virus presents the lowest level of adaptation. If the HPR modifies translation efficiency through the accumulation of high or low frequency codons, it can be expected that the CAI values of hemagglutinin genes are proportional to the size of the HPR and the protein.

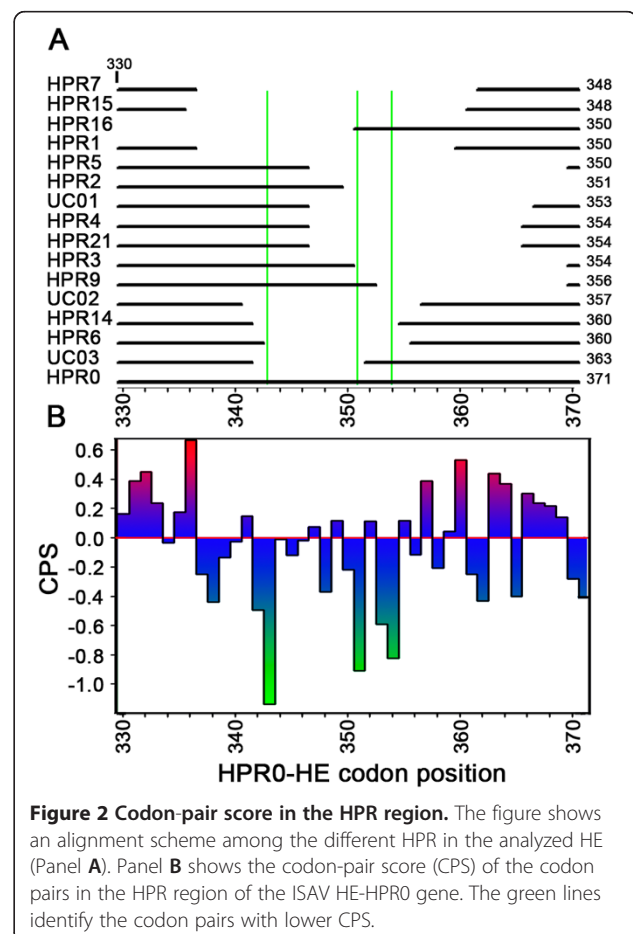
Among the 63 analyzed sequences, no statistically significant relationships ( $P < 0.05$ ) were found between the CAI value of the HE gene and the size of the proteins or observed virulence, but a relationship was observed between the CAI value of the HPR and the size of this region ( $\rho=0.393$ ,  $P=0.002$ ,  $n=63$ ). As the translation efficiency is affected by the presence of underrepresented codon pair usage in the genome, we analyzed bias in the codon pair usage of the HE genes (Figure 1). The results show that HE shares CPB values with 20% of the salmon genes (Additional file 1: Figures S2 and S3). A statistically significant relationship was identified ( $\rho=-0.675$ ,  $P=2 \times 10^{-7}$ ,  $n=63$ ) between CPB value and HE size. Of the 63 analyzed sequences, 3 were genes of Canadian viruses, which presented low CPB values. Among the viruses classified as European, the gene containing HPR0 had the lowest CPB value (Figure 1). This correlation increases when analyzing the relationship between the CPB value of the HPR region of the European isolates and the size of HPR region ( $\rho=-0.757$ ,  $P=2 \times 10^{-7}$ ,  $n=60$ ).

When viruses were compared according to their country of isolation, there were significant differences ( $t$ -test,  $P = 0.008$ ) in the mean CPB values of the HE gene of the Norwegian ( $CPB = 0.0394 \pm 0.007$ ) and Chilean viruses ( $CPB = 0.0446 \pm 0.006$ ), with the Chilean isolates presenting the higher CPB values (Additional file 1: Figure S4). When the CPB values of the HE gene are classified according to their HPR type, we note that the lowest CPB values are found in the HE genes containing HPR21 and HPR0, while the highest values are found in the HE genes containing HPR1, HPR15 and HPR7 (Additional file 1: Figure S5).



**Figure 1 Relationship between CPB and HE size.** The figure shows the CPB values of 63 proteins with complete sequences in the encoding region of the HE gene. The CPB values are shown for HE genes in the Norwegian virus (white squares), North American virus (gray circles), the Chilean virus (black triangles) and the British virus (inverted gray triangles). The square with a star indicates the HE gene of the HPR0 genotype.

To assess the impact of the HPR region on the total CPB value of HE, we determined the QCPB value. The results of this analysis show that the HPR0, HPR9, HPR3, HPR16, HPR2, HPR5, HPR4, UC010, UC011 and HPR6 regions negatively affect the CPB value of the gene that carries them, while HPR7, HPR1, HPR14, HPR15, UC04 and HPR21 increase total CPB value (Additional file 1: Figure S6). With the exception of HPR9 and HPR3, the HPR0 segment has the lowest CPB value, while HPR15, HPR7 and HPR1 present the highest CPB values (Additional file 1: Figure S7). The HPR0 and HPR9 regions had the most negative effect on total CPB value. Because the HPR0s had one of the lowest CPB values and most negatively affected total CPB value, we analyzed this region to find zones rich in codon pairs that are underrepresented in *Salmo salar*. This analysis shows that there are three points in the 343–344, 351–352 and 354–355 residues in the HPR0 region that present CPS values  $< 0.8$ , that is, that these codon pairs are underrepresented by at least two times. Most of the points with low CPB values are absent in the rest of the HPR variants (Figure 2). Comparison among the CPS profiles of the different genes shows that the most



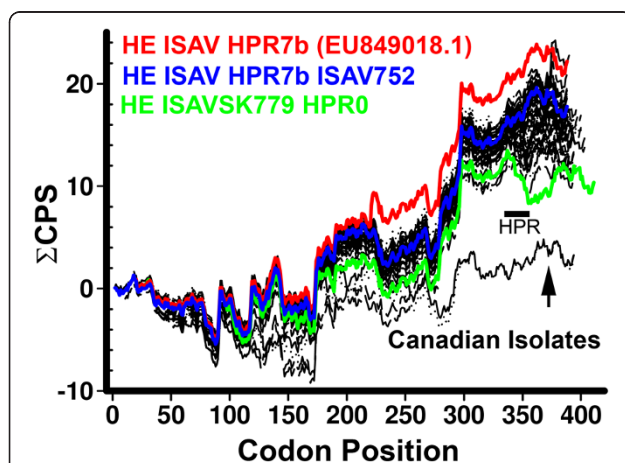
**Figure 2 Codon-pair score in the HPR region.** The figure shows an alignment scheme among the different HPR in the analyzed HE (Panel A). Panel B shows the codon-pair score (CPS) of the codon pairs in the HPR region of the ISAV HE-HPR0 gene. The green lines identify the codon pairs with lower CPS.

significant difference in the carboxyl terminal profiles are found in the HPR region. The viral isolates in Chile with the HPR7b show profiles with greater magnitude, in contrast to what is observed in the North American viruses and ISAV SK779 (HPR0). In the gene with HPR0, the HPR zone presents a descending profile that breaks an ascending ramp of CPB values present in the HPR variant with important deletions in the HPR zone (Figure 3).

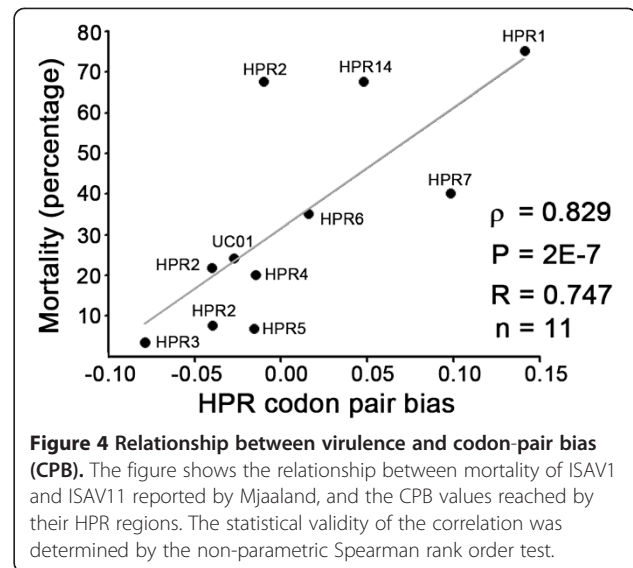
#### Relationship among codon-pair bias, HPR type and virulence

Our analysis consistently shows that the genes containing the HPR0 and HPR7 regions have extreme CPB values. The two HPRs are associated with opposite phenotypes: the HPR0 region is found in avirulent strains of ISAV, while the HPR7 region has been systematically found in Chilean outbreaks that have been responsible for eliminating 90% of salmon in production centers. To determine if there is a relationship among virulence, HPR type and codon-pair bias, we correlated our CPB values to the mortality rates reported by Mjeeland [36].

These mortality rates constitute valuable information given that the virulence of 11 isolates of the ISA virus with different HPRs were analyzed and compared. The results of this correlation show that the CPB values of the HE gene correlate positively with mortality by direct injection ( $\rho = 0.597$ ,  $P = 0.0467$ ,  $n = 11$ ) and with mortality rates with fish-to-fish transmission (TG2,  $\rho = 0.633$ ,  $P=0.032$ ,  $n=11$ ). When we use only the CPB value of the HPR region, the correlation with mortality of the cohabitants (fish inoculated with the virus by injection) increases ( $\rho = 0.829$ ,  $P = 2 \times 10^{-7}$ ,  $n = 11$ ) (Figure 4).



**Figure 3 Profile of CPS throughout the HE gene.** The figure shows the profile of accumulated CPS values for the 63 HE genes. We highlight the Chilean isolates with HPR 7b, the Canadian isolates and the HPR0 region.



**Figure 4 Relationship between virulence and codon-pair bias (CPB).** The figure shows the relationship between mortality of ISAV1 and ISAV11 reported by Mjaaland, and the CPB values reached by their HPR regions. The statistical validity of the correlation was determined by the non-parametric Spearman rank order test.

Extrapolation of mortality based on the CPB value of the HPR shows that the Norwegian viruses with HPR15 (AF378180.2 and AF302799.1) and the Chilean viruses with HPR1C (GU830908.1 and FJ594285.1) have an estimated mortality  $> 80\%$ , which is higher than values observed in ISAV2 with HPR01. When we analyzed the CPB of the complete HE gene, we observed that most of the Chilean viruses have values between those of ISAV1 (HPR14) and ISAV2 (HPR01). The CPB value of the HE gene from Chilean isolates with HPR7b has a value of 0.122, which by extrapolation becomes a virus with similar characteristics to HPR1 (CPB = 0.146) in the isolates of ISAV2. ISAV2 has the highest mortality rate by ISAV injection (75%) and the second highest rate by cohabitation (28.9%), followed by the HPR14 in the ISAV1 isolate (46.1%) (Additional file 3: Table S2).

#### Discussion

The results presented in this work indicate the existence of a relationship between bias in the use of codon pairs and the size of the HPR region or the HE gene. In particular, our analysis shows that the different types of HPR tend to eliminate regions that present underrepresented codon pairs (negative CPS index). In Influenza A and polio, the introduction of codon pairs with low CPS values is an effective way to reduce translation of the viral gene, reducing the production of viral particles and virulence. In contrast, the introduction of codon pairs with positive CPS increases protein expression [31,32]. The molecular explanation for this relationship is not clear, but it is probably because not all tRNA are compatible among themselves upon simultaneously occupying the A and P ribosome sites, as multivariate analysis suggests [37].

A universal codon distribution pattern has recently been described. In general the genes tend to accumulate

codons poorly adapted to their tRNA pool among the first 50 codons, while the more adapted codons are found among the last 150 to 300 nucleotides [22]. It has been suggested that the presence of this adaptation ramp to the tRNA pool serves to avoid the ribosome traffic jam in the mRNA, the translation efficiency being inverse to the quantity of ribosomes in the mRNA. Thus, the presence of underrepresented codons could reduce translational efficiency and the advance and turn-over of ribosomes [22-24].

The HPR region is located at approximately 80 codons from the carboxyl terminal with respect to the gene with HPR0. According to the ramp observed in eukaryotic [22], this region should contain codons more adapted to the tRNA pool. The presence of underrepresented codon pairs in the HPR zone could decrease the velocity of the advance of ribosomes by the incompatibility among tRNA, being functionally equivalent to the use of codons with low levels of available tRNA. On the other hand, the elimination of these codon pairs could increase translation efficiency by decreasing the ribosome traffic jam in this region. The Chilean isolates with HPR7b present a  $\Sigma$ CPS profile with high CPS values and a continuous ramp toward the carboxyl terminal (Figure 3), which is interrupted in the virus with HPR0. Consistent with this, viruses containing HPR0 are associated with avirulent ISAV strains. On the other hand, HPR regions where codon pairs with negative CPS have been eliminated are associated with outbreaks with high mortality rates, such as HPR1 or HPR7b found in 80% of the pathogenic outbreaks in Chile [38,39]. The virulence values extrapolated for the Chilean isolates are consistent with the effect of these viruses on the Chilean salmon industry.

Current efforts to cultivate the virus in vitro with the HPR0 genotype have not been fruitful, which is why it remains unknown if these viruses have lower HE protein synthesis rates than viruses with the HPR7b genotype. Nevertheless, the correlations between the CPB values in the HE gene with the size of the region and the degree of pathogenicity of the ISA virus are consistent with the effect of HPR on translation efficiency.

Both immune escape due to antigenic variation produced by different HPRs [40] and modification of hemagglutinin-esterase activity can act as selection forces for the different HPR described. Recent experiments have determined that HPR does not modify hemagglutinin-esterase activity [10], suggesting that the biological activity of the protein does not explain the difference in observed virulence among ISAV isolates. There are variations in the protein sequence of the glycoproteins of influenza A that give rise to antigenic drift [41]. However, in influenza A, the variations occur by specific mutations that produce variations in the

epitopes, in contrast to what occurs in ISAV, where the greatest variation observed in the HPR occurs by partial deletions in the HPR0 [34]. In ISAV, lesser variations have been observed on the surface of the HE protein [39]. However, it is not known if these mutations have an important effect on antigen-antibody recognition. Although antigenic variation in the ISAV hemagglutinin-esterase could act as a selective force, in our opinion this concept makes more sense in populations whose individuals are long-lived and acquire immunity after surviving a viral infection or vaccination. Unlike humans, salmon are short-lived (four years) and suffer outbreaks of ISAV with high levels of mortality. Moreover, variants in the HPR region appeared prior to the development of vaccines. Determining the degree of adaptation to the tRNA pool could help in confirming that the codons in the HPR region reduce translation efficiency. However, the estimation of the number of tRNA copies in *Salmo salar*, using the tRNAScan-SE program (which recognizes close to 100% of the tRNA in a genome), showed that *Salmo salar* has a large number of tRNA genes (>65000), which is much higher than the approximately 500 copies in humans and is on the order of the number found in other teleosts, like zebra fish (~45,000) [42]. Interestingly, close to a third of the identified tRNA is tRNA-Ala and another third is pseudogenes. This disproportionately high level of tRNA suggests the presence of a repeated SINE-type element, which presents homologous regions to those of the tRNA [43]. Repeated SINE-type elements have been described in salmonoids and other teleosts, but with leucine and lysine tRNA [44-47]. This suggests the presence of a new SINE-type in *Salmo salar*, although further studies will be necessary once the *S. salar* genome has been completely annotated to characterize this type of element. Despite the significant disproportion of Ala-tRNA compared to the other tRNAs, the W values of the codons in the HPR0 region show three areas, close to the area with negative CPS, where codons with low tRNA availability accumulate (Additional file 4: Table S3, Additional file 5: Table S4 and Additional file 1: Figure S8). This reinforces the idea that HPR0 is a zone of reduced translation efficiency.

In summary, our analysis shows a correlation between the CPB value of the HE gene or the HPR region and the virulence of the ISA virus. To date, we have not determined if the CPB value in the ISAV is related to gene translation. However, in vitro data from influenza A and the poliovirus strongly suggest the existence of this relationship. Although it is an over-simplification of the pathogenesis to explain virulence based solely on the adaptation of viral genes to the host's codon pair usage, to our knowledge the correlations described in this work constitute the only natural example of de-optimization of codon pairs as the element correlating to the level of

pathogenicity of a virus and constitute a reliable explanation of the role of the HPR region in the ISA virus as a modulating factor of virulence. Without doubt, the development of a reverse genetic technique could verify this relationship, but until this technique becomes available, sequence analysis is a useful tool.

## Materials and methods

### Calculation of tAI

We calculated tAI in accordance with what was established by dos Reis [12] (Equation 1 and 2), estimating the relative abundance of the tRNAs in function of the number of copies of each tRNA in the *Salmo salar* genome. The number of copies of each tRNA was determined with the tRNASCAN-SE program [48] in Linux, which was used to analyze the 555960 contigs in the GenBank database (Project ID: 72713) corresponding to the recently released salmon genome.

$$W_i = \sum_{j=1}^{n_i} (1 - S_{ij}) tGNC_{ij} \quad (1)$$

where  $W_i$  is the relative adaptiveness of the codon  $i^{\text{th}}$  to the tRNA pool,  $n_i$  is the number of tRNA isoacceptors that recognize the  $i^{\text{th}}$  codon,  $tGNC_{ij}$  is the number of copies of the  $j^{\text{th}}$  tRNA gene that recognize the  $i^{\text{th}}$  codon, and  $S_{ij}$  is the selective constraint in the efficiency of codon-anticodon pairing.

$$tAI_g = \left( \prod_{k=1}^{l_g} w_{ikg} \right)^{1/l_g} \quad (2)$$

The adaptation of a gene to the tRNA pool is calculated according to equation 2, in which  $w_i$  is defined as the quotient between  $W_i$  and  $W_{\max}$  ( $W_i/W_{\max}$ ),  $ikg$  is the codon defined by the  $k^{\text{th}}$  triplet of gene  $g$ , and  $l_g$  is the length of gene  $g$  in codifying codons.

### Determination of bias in codon-pair usage

Bias in codon-pair usage was determined according to what was described by Coleman [31]. To do this, we recalculated the codon-usage tables in Kazusa (<http://www.kazusa.or.jp/codon/>), using the information in the codifying regions of the *Salmo salar* genes available in Genebank. The sequence of the codifying regions of these genes was downloaded from the RefSeq database and subsequently filtered to eliminate redundant genes or genes with ambiguous nucleotides or with incomplete codifying regions. A total of 3485 sequences passed through the filters, which represented the entrance of 1168431 codons. Based on this set of sequences (Additional file 6: Table S5), we calculated the codon and codon-pair frequency, using scripts in the language Python. The observed frequency of amino acid pairs was

deduced based on the sum of all the codon pairs that codify for the same combination of amino acids. Subsequently, we used these frequencies to calculate the CPS index (codon-pair score) using the equation described by Mueller et al. 2006 [33] (Equation 3), where  $f(AB)$  corresponds to the frequency of the appearance of the AB codon pair expressed in parts-per-thousands with respect to the total number of codon pairs in the genome.  $f(A)$  and  $f(B)$  correspond to the abundance of each of the individual codons expressed in parts-per-thousands with respect to the total number of codons in the genome.  $f(X)$  and  $f(Y)$  corresponds to the amino acids codified by the A and B codons respectively and  $f(XY)$  is the abundance of the amino acid pairs in all the proteins of the organism under study. The CPS values of the 3271 codon pairs in *Salmo salar* can be found in the supplementary material (Additional file 7: Table S6). We calculated the bias in codon usage as the arithmetic mean of the codon-pair score of a gene (CPS) described by Coleman et al. (2006) Equation 4, in which  $i$  is the  $i^{\text{th}}$  codon pair of a gene and  $l$  is the total length of the gene expressed in the quantity of codified amino acids.

$$CPS = Ln \left( \frac{f(AB)}{f(A) \times f(B)} \times \frac{f(X) \times f(Y)}{f(XY)} \right) \quad (3)$$

$$CPB = \frac{\sum_i^l CPS_i}{l-1} \quad (4)$$

### Calculation of the CAI

The codon adaptation index was calculated with the Emboss program [49] based on the frequency of codon usage of *Salmo salar*.

### Prediction of free folding energy

Free folding energy was calculated with the Mfold program through the web server (<http://mfold.rna.albany.edu/?q=mfold>) using the standard parameters. The energy was calculated for the region between residuals 330 and 371 with respect to HE SK779 HPR0.

### Virulence data

The virulence of 11 ISA viruses with known genomes was obtained from data published by Mjaaland et al. in 2005, in which they reported mortality rates 60 days after infection. This work analyzed the mortality rates of two fish populations from the same parents and the mortality rates of fish with genetic heterogeneity. The fish with the same parents were infected by cohabitation (TGI and TGII) with genetically heterogeneous fish that were previously infected by intraperitoneal injection (COH) [36].

## HE genes of ISAV

The complete sequence of the codifying regions of the ISAV HE genes was downloaded from Genbank. All the genes that did not begin with ATG were filtered out, ending with a stop codon and ambiguous nucleotides. A total of 63 sequences passed through the filter. The GenBank codes of these sequences can be found in the supplementary material.

## Graphs and statistical methods

The graphs and statistical tests were made using the program SigmaPlot 11. The significance of the differences or correlations among the data groups obtained was evaluated with non-parametric tests (Rank Sum Test for comparing two groups or Spearman Rank Order for correlations) using a value of  $p < 0.05$  as a cutoff.

## Additional files

**Additional file 1:** This is a Microsoft Word document containing supplementary figures about Gibb's free energy of folding acquired by different HPRs, Pattern of codon-pair usage in *Salmo salar*, Pattern of distribution of the bias of codon-pair use in *S. Salar*, CPB values of HE genes, CPB values according to the HPR type, Effect of the HPR region on bias in the use of codon pairs, and CPB values of the HPR regions, and W values of codons present in the HPRO region of HE gene.

**Additional file 2:** This is a Microsoft Excel document containing a supplementary table of *Salmo salar* tRNA copy numbers.

**Additional file 3:** This is a Microsoft Excel document of a supplementary table containing HE gene ID, HE Protein ID, HPR or group, Country of isolation, HE Protein size, CPB value of HE, CPB of HPR, Name of ISAV strain and mortality rates of different ISAV isolates.

**Additional file 4:** This is a Microsoft Excel document containing a supplementary table with tRNA availability (W values) for each coding codon of *Salmo salar*.

**Additional file 5:** This is a Microsoft Excel document containing a supplementary table with tRNA availability (W values) for each codon present in HPRO.

**Additional file 6:** This is a Microsoft Excel document containing a supplementary table with the Genebank access numbers for *Salmo salar* genes analyzed in this work.

**Additional file 7:** This is a Microsoft Excel document containing a supplementary table with CPS values for *Salmo salar*.

## Competing interests

The authors declare no competing interests.

## Authors' contributions

MT designed all *in silico* experiments. JMS performed all computational analysis and calculations. ES contributed to data analysis and preparing the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We would like to thank to the Vice Rectoría de Investigación y Desarrollo (VRID) of the Universidad de Santiago de Chile (USACH) for administrative support and travel assistance, Project Bicentenario PDA20 to MTR, CONICYT Fellowship to JMS, and CONICYT-COOP. INTERNACIONAL ACE-02 and INNOVA-CORFO 09MCCS-6698.

Received: 28 January 2013 Accepted: 28 May 2013

Published: 6 June 2013

## References

1. Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE: *Field's Virology*. 5th edn; 2007.
2. Krossoy B, Nilsen F, Falk K, Endresen C, Nylund A: Phylogenetic analysis of infectious salmon anaemia virus isolates from Norway, Canada and Scotland. *Dis Aquat Organ* 2001, **44**:1–6.
3. Blake S, Bouchard D, Keleher W, Opitz M, Nicholson BL: Genomic relationships of the North American isolate of infectious salmon anemia virus (ISAV) to the Norwegian strain of ISAV. *Dis Aquat Organ* 1999, **35**:139–144.
4. Devold M, Karlsen M, Nylund A: Sequence analysis of the fusion protein gene from infectious salmon anemia virus isolates: evidence of recombination and reassortment. *J Gen Virol* 2006, **87**:2031–2040.
5. Devold M, Falk K, Dale B, Krossoy B, Biering E, Aspehaug V, Nilsen F, Nylund A: Strain variation, based on the hemagglutinin gene, in Norwegian ISA virus isolates collected from 1987 to 2001: indications of recombination. *Dis Aquat Organ* 2001, **47**:119–128.
6. Mjaaland S, Hungnes O, Teig A, Dannevig BH, Thorud K, Rimstad E: Polymorphism in the infectious salmon anemia virus hemagglutinin gene: importance and possible implications for evolution and ecology of infectious salmon anemia disease. *Virology* 2002, **304**:379–391.
7. Godoy MG, Aedo A, Kibenge MJ, Groman DB, Yason CV, Grothusen H, Lisperguer A, Calbuco M, Avendano F, Imilan M, et al: First detection, isolation and molecular characterization of infectious salmon anaemia virus associated with clinical disease in farmed Atlantic salmon (*Salmo salar*) in Chile. *BMC Vet Res* 2008, **4**:28.
8. Christiansen DH, Ostergaard PS, Snow M, Dale OB, Falk K: A low-pathogenic variant of infectious salmon anemia virus (ISAV-HPRO) is highly prevalent and causes a non-clinical transient infection in farmed Atlantic salmon (*Salmo salar* L.) in the Faroe Islands. *J Gen Virol* 2011, **92**:909–918.
9. McBeath AJ, Bain N, Snow M: Surveillance for infectious salmon anaemia virus HPRO in marine Atlantic salmon farms across Scotland. *Dis Aquat Organ* 2009, **87**:161–169.
10. McBeath A, Fourrier M, Munro E, Falk K, Snow M: Presence of a full-length highly polymorphic region (HPR) in the ISAV haemagglutinin-esterase does not affect the primary functions of receptor binding and esterase activity. *Arch Virol* 2011, **156**:2285–2289.
11. Plotkin JB, Kudla G: Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 2011, **12**:32–42.
12. dos Reis M, Savva R, Wernisch L: Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* 2004, **32**:5036–5044.
13. Sharp PM, Li WH: The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987, **15**:1281–1295.
14. Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985, **2**:13–34.
15. Sorensen MA, Kurland CG, Pedersen S: Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol* 1989, **207**:365–377.
16. Moriyama EN, Powell JR: Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol* 1997, **45**:514–523.
17. Varenne S, Buc J, Lloubes R, Lazdunski C: Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* 1984, **180**:549–576.
18. Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T: Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol* 2001, **53**:290–298.
19. Gingold H, Pilpel Y: Determinants of translation efficiency and accuracy. *Mol Syst Biol* 2011, **7**:481.
20. Tats A, Tenson T, Remm M: Preferred and avoided codon pairs in three domains of life. *BMC Genomics* 2008, **9**:463.
21. Gutman GA, Hatfield GW: Nonrandom utilization of codon pairs in *Escherichia coli*. *Proc Natl Acad Sci U S A* 1989, **86**:3699–3703.
22. Tuller T, Carmi A, Vestigian K, Navon S, Dorfan Y, Zaboroske J, Pan T, Dahan O, Furman I, Pilpel Y: An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 2010, **141**:344–354.

23. Tuller T: Codon bias, tRNA pools and horizontal gene transfer. *Mob Genet Elements* 2011, **1**:75–77.
24. Kudla G, Murray AW, Tollervey D, Plotkin JB: Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 2009, **324**:255–258.
25. Jia M, Li Y: The relationship among gene expression, folding free energy and codon usage bias in *Escherichia coli*. *FEBS Lett* 2005, **579**:5333–5337.
26. Bahir I, Fromer M, Prat Y, Linial M: Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol* 2009, **5**:311.
27. Gog JR, Afonso Edos S, Dalton RM, Leclercq I, Tiley L, Elton D, von Kirchbach JC, Naffakh N, Escriou N, Digard P: Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic Acids Res* 2007, **35**:1897–1907.
28. Marsh GA, Rabadan R, Levine AJ, Palese P: Highly conserved regions of influenza A virus polymerase gene segments are critical for efficient viral RNA packaging. *J Virol* 2008, **82**:2295–2304.
29. Pinto RM, Aragones L, Costafreda MI, Ribes E, Bosch A: Codon usage and replicative strategies of hepatitis A virus. *Virus Res* 2007, **127**:158–163.
30. Aragones L, Guix S, Ribes E, Bosch A, Pinto RM: Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog* 2010, **6**:e1000797.
31. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S: Virus attenuation by genome-scale changes in codon pair bias. *Science* 2008, **320**:1784–1787.
32. Mueller S, Coleman JR, Papamichail D, Ward CB, Nimnual A, Futcher B, Skiena S, Wimmer E: Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotechnol* 2010, **28**:723–726.
33. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E: Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol* 2006, **80**:9687–9696.
34. Nylund A, Plarre H, Karlsen M, Fridell F, Ottem KF, Bratland A, Saether PA: Transmission of infectious salmon anaemia virus (ISAV) in farmed populations of Atlantic salmon (*Salmo salar*). *Arch Virol* 2007, **152**:151–179.
35. Gray NK, Hentze MW: Regulation of protein synthesis by mRNA structure. *Mol Biol Rep* 1994, **19**:195–200.
36. Mjaaland S, Markussen T, Sindre H, Kjøglum S, Dannevig BH, Larsen S, Grimholt U: Susceptibility and immune responses following experimental infection of MHC compatible Atlantic salmon (*Salmo salar* L.) with different infectious salmon anaemia virus isolates. *Arch Virol* 2005, **150**:2195–2216.
37. Buchan JR, Aucott LS, Stansfield I: tRNA properties help shape codon pair preferences in open reading frames. *Nucleic Acids Res* 2006, **34**:1015–1027.
38. Kibenge FS, Godoy MG, Wang Y, Kibenge MJ, Gherardelli V, Mansilla S, Lisperger A, Jarpa M, Larroquete G, Avendano F, et al: Infectious salmon anaemia virus (ISAV) isolated from the ISA disease outbreaks in Chile diverged from ISAV isolates from Norway around 1996 and was disseminated around 2005, based on surface glycoprotein gene sequences. *Virol J* 2009, **6**:88.
39. Cottet L, Cortez-San Martin M, Tello M, Olivares E, Rivas-Aravena A, Vallejos E, Sandino AM, Spencer E: Bioinformatic analysis of the genome of infectious salmon anemia virus associated with outbreaks with high mortality in Chile. *J Virol* 2010, **84**:11916–11928.
40. Kibenge FS, Kibenge MJ, McKenna PK, Stothard P, Marshall R, Cusack RR, McGeachy S: Antigenic variation among isolates of infectious salmon anaemia virus correlates with genetic variation of the viral haemagglutinin gene. *J Gen Virol* 2001, **82**:2869–2879.
41. Webster RG, Laver WG, Air GM, Schild GC: Molecular mechanisms of variation in influenza viruses. *Nature* 1982, **296**:115–121.
42. Bermudez-Santana C, Attolini CS, Kirsten T, Engelhardt J, Prohaska SJ, Steigele S, Stadler PF: Genomic organization of eukaryotic tRNAs. *BMC Genomics* 2010, **11**:270.
43. Kramerov DA, Vassetzky NS: Origin and evolution of SINEs in eukaryotic genomes. *Heredity (Edinb)* 2011, **107**:487–495.
44. Matveev V, Nishihara H, Okada N: Novel SINE families from salmon validate Parahucho (*Salmonidae*) as a distinct genus and give evidence that SINEs can incorporate LINE-related 3'-tails of other SINEs. *Mol Biol Evol* 2007, **24**:1656–1666.
45. Morescalchi MA, Barucca M, Stingo V, Capriglione T: Polypteridae (*Actinopterygii*: *Cladistia*) and DANA-SINEs insertions. *Mar Genomics* 2010, **3**:79–84.
46. Kawagoe-Takaki H, Nameki N, Kajikawa M, Okada N: Probing the secondary structure of salmon Smal SINE RNA. *Gene* 2006, **365**:67–73.
47. Izsvak Z, Ivics Z, Garcia-Estefania D, Fahrenkrug SC, Hackett PB: DANA elements: a family of composite, tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish (*Danio rerio*). *Proc Natl Acad Sci U S A* 1996, **93**:1077–1081.
48. Lowe TM, Eddy SR: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997, **25**:955–964.
49. Rice P, Longden I, Bleasby A: EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000, **16**:276–277.

doi:10.1186/1743-422X-10-180

Cite this article as: Tello et al.: Analysis of the use of codon pairs in the HE gene of the ISA virus shows a correlation between bias in HPR codon-pair use and mortality rates caused by the virus. *Virology Journal* 2013 **10**:180.

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

