

Ultra-Deep Pyrosequencing of Partial Surface Protein Genes from Infectious Salmon Anaemia Virus (ISAV) Suggest Novel Mechanisms Involved in Transition to Virulence

Turhan Markussen^{1*}, Hilde Sindre¹, Christine Monceyron Jonassen^{1^{na}}, Torstein Tengs¹, Anja B. Kristoffersen², Jon Ramsell^{1^{nb}}, Sanela Numanovic¹, Monika J. Hjortaa¹, Debes H. Christiansen³, Ole Bendik Dale¹, Knut Falk²

1 Department of Laboratory Services, Norwegian Veterinary Institute, Oslo, Norway, **2** Department of Health Surveillance, Norwegian Veterinary Institute, Oslo, Norway, **3** National Reference Laboratory for Fish Diseases, Food and Veterinary Authority, Torshavn, Faroe Islands

Abstract

Uncultivable HPR0 strains of infectious salmon anaemia viruses (ISAVs) infecting gills are non-virulent putative precursors of virulent ISAVs (vISAVs) causing systemic disease in farmed Atlantic salmon (*Salmo salar*). The transition to virulence involves two molecular events, a deletion in the highly polymorphic region (HPR) of the hemagglutinin-esterase (HE) gene and a Q₂₆₆→L₂₆₆ substitution or insertion next to the putative cleavage site (R₂₆₇) in the fusion protein (F). We have performed ultra-deep pyrosequencing (UDPS) of these gene regions from healthy fish positive for HPR0 virus carrying full-length HPR sampled in a screening program, and a vISAV strain from an ISA outbreak at the same farming site three weeks later, and compared the mutant spectra. As the UDPS data shows the presence of both HE genotypes at both sampling times, and the outbreak strain was unlikely to be directly related to the HPR0 strain, this is the first report of a double infection with HPR0s and vISAVs. For F amplicon reads, mutation frequencies generating L₂₆₆ codons in screening samples and Q₂₆₆ codons in outbreak samples were not higher than at any random site. We suggest quasispecies heterogeneity as well as RNA structural properties are linked to transition to virulence. More specifically, a mechanism where selected single point mutations in the full-length HPR alter the RNA structure facilitating single- or sequential deletions in this region is proposed. The data provides stronger support for the deletion hypothesis, as opposed to recombination, as the responsible mechanism for generating the sequence deletions in HE.

Citation: Markussen T, Sindre H, Jonassen CM, Tengs T, Kristoffersen AB, et al. (2013) Ultra-Deep Pyrosequencing of Partial Surface Protein Genes from Infectious Salmon Anaemia Virus (ISAV) Suggest Novel Mechanisms Involved in Transition to Virulence. PLoS ONE 8(11): e81571. doi:10.1371/journal.pone.0081571

Editor: Maureen J. Donlin, Saint Louis University, United States of America

Received: July 13, 2013; **Accepted:** October 14, 2013; **Published:** November 26, 2013

Copyright: © 2013 Markussen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by The Norwegian Research Council (project no. 207024/E40). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: turhan.markussen@vetinst.no

^{na} Current address: Department of Laboratory Medicine, Østfold Hospital Trust, Fredrikstad, Norway

^{nb} Current address: Department of Medical Sciences, Uppsala University, Uppsala, Sweden

Introduction

Infectious salmon anaemia virus (ISAV) is an orthomyxovirus that has caused systemic infection and disease in farmed Atlantic salmon (*Salmo salar*) in Norway, Canada, Scotland, Shetland Islands, the Faroe Islands, USA and Chile [1-8]. The virus' two major surface glycoproteins are the hemagglutinin-esterase (HE), responsible for receptor binding and release (RDE), and the fusion protein (F), responsible for virus uptake through the fusion of viral and cellular membranes [9-12]. The

HEs from different virulent ISAV strains (vISAVs) often vary in length as determined by the size of their highly polymorphic region (HPR) located in the stalk immediately upstream of the transmembrane domain of the protein. The shortening of the stalk is thought to arise from differential deletions from a full-length precursor HE (HPR0) [13]. Recombination through template switching during replication has been proposed as an alternative mechanism for generating the deletions [14]. The deletions in the HE stalk region could be analogues to the varying lengths in the influenza A virus neuraminidase stalk,

which has been associated with host switching [15-18]. A shortening of the HE stalk could also affect the functional balance between the HE receptor-binding and -destroying activities similar to that found between the hemagglutinin and neuraminidase of influenza A viruses through a number of studies [19-23].

ISAV HPR0 genotypes containing full-length HEs are non-virulent, non-cultivable and are primarily found in gills [24-26]. They have been detected in apparently healthy wild and farmed Atlantic salmon in most regions with Atlantic salmon farming [24-28]. Frequent findings of HPR0s in fish farms in Norway and the Faroe Islands, the former with few disease outbreaks and the latter with no ISA outbreaks since 2005, suggests the transition to virulence is an infrequent occurrence [27,29]. In addition to a full-length HE, all HPR0s also contain a glutamine in position 266 (Q₂₆₆) next to the one of the two putative cleavage sites in the F protein (R₂₆₇) [30]. In contrast, all vISAVs carry a leucine in this position (L₂₆₆), except for a few strains which have small sequence insertions close to this site [30-32]. F gene insertions through template switching, as well as reassortment of gene segments, have both been linked to virulence [30,31]. For the F protein, the single amino acid substitution or insertion close to R₂₆₇ may be analogous to that in highly pathogenic avian influenza A virus subtypes H5 and H7, where pathogenicity is acquired through the mutational change of cleavage specificity resulting in altered tissue tropism [33,34]. These differences between HPR0s and vISAVs may represent viral adaptation leading to disease in densely populated industrial farming operations.

RNA viruses are known to exist in the host as a swarm of closely related mutant genomes known as quasispecies [35-37]. The molecular basis for this genomic heterogeneity is that replication of RNA viruses is highly error prone since viral RNA-dependent RNA polymerases do not possess proofreading-repair activity [38-40]. Such a mutant spectrum within an individual host, combined with the ability to recombine and reassort their gene segments, can enable these viruses to rapidly adapt in response to selective pressures [36,37]. This genetic diversity can also be viewed as a measure of the fitness of the virus and be linked directly to pathogenesis [41-43]. Ultra-deep pyrosequencing (UDPS) allows for the simultaneous analysis of thousands of clonally amplified PCR amplicons, enabling the detection of the within-host minority variants [44-49]. The technology is now well established in many fields within virology, including viral evolution and pathogenesis, detection of antiviral resistance markers and diagnostics. For ISAV, aside from a recent publication describing the presence of quasispecies in the non-coding regions of the viral genome, based on traditional cloning and Sanger sequencing [50], the within-host genomic heterogeneity for this virus has not been investigated.

An ISA epidemic from 2007 to 2010 in Astafjord of Norway was monitored by both disease diagnostics and screening of apparently healthy fish. Sequence-based analysis of partial HE- and F genes combined with epidemiological information concluded that a single virulent ISAV strain had transmitted horizontally between proximal fish farms and caused all 17 ISA outbreaks in the region from 2007 to 2009 (i.e. Astafjord strain)

[28,51]. Here we present results from UDPS of amplicons from gene regions spanning the HPR in HE and the putative R₂₆₇ cleavage site in the F protein. From a site in Astafjord 2010, mutant spectra originating from gill tissue samples found screening positive for a non-virulent HPR0 are compared with that of a vISAV from an ISA outbreak at the same farming site three weeks later. We report, for the first time, a double infection involving a HPR0 and a vISAV. F reads containing L₂₆₆- and Q₂₆₆ codons in screening and outbreak samples, respectively, were present at mutational frequencies comparable to that found at any random site in the amplicons. Similarly, HE reads from screening samples containing deletions within the full-length HPR, differing from the HPR deletion pattern of the outbreak strain (delHPR, present at very low levels) were not observed at all. The results from UDPS data analyses, together with RNA structure predictions, provide stronger support for the HE deletion hypothesis over recombination in creating the sequence deletions in vISAVs, and we propose a mechanism in which selected single point mutations in the full-length HPR alter the RNA structure in such a way that it facilitates single- or sequential deletions in this region of the HE gene.

Materials and Methods

Field Samples

Four HPR0-positive gill samples from healthy Atlantic salmon taken from a commercial farming site (hereafter named Site 10) in Astafjord (Norway) during screening in May 2010 (screening samples 1-4), and four gill samples taken three weeks later during the early phases of an ISA outbreak at the same site (outbreak samples 1-4), were chosen for UDPS. Sampling from both screening and disease outbreak, which were from different fish and different locations at the site, were performed on order by the Norwegian Food Safety Authority. Approval from Institutional Animal care and Use Committee (IACUC) or ethics was not required. No experiments that involved fish were performed. The screening samples were among 8 of 23 gill samples ISAV positive by real-time RT-PCR (Ct's 28-38, not shown). Traditional (Sanger) sequencing of HE- and F genes, covering the HPR and the putative encoded R₂₆₇ cleavage site, respectively, showed all to be of one non-virulent HPR0 genotype [29]. The ISA outbreak fish had pathology consistent with ISA, and the systemic infection was verified by real-time RT-PCR and Sanger sequencing of the HE gene from kidney samples of ten fish tested (Ct's 14-18, not shown). Real-time RT-PCR to determine viral loads in gill samples selected for UDPS was performed as previously described [28].

Preparation of PCR-generated Amplicons

Gill tissue was homogenized and total RNA extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany). cDNA synthesis was performed separately for each sample using random hexamers together with Superscript III reverse transcriptase, following the protocol recommended by the manufacturer (Invitrogen, Carlsbad, USA). Following first-strand synthesis, cDNA samples were used as templates in two separate PCR reactions to generate gene segment 5 (F gene) and segment 6

(HE gene) amplicons. Primer sequences are shown in Table S1. PCR was performed using high fidelity AccuPrime Pfx DNA polymerase (Invitrogen, Carlsbad, USA) with 2 µl cDNA in each reaction. PCR cycling conditions were 95°C/2 min, followed by 40 cycles of 95°C/45 sec, 55°C/45 sec and 68°C/2 min and a final extension step at 72°C/7 min. After PCR, a small volume from each reaction was run on 1.5 % agarose gel electrophoresis and the products visualized by ethidium bromide staining. In some cases the 2100 Bioanalyzer (Agilent Technologies) was also used to evaluate the quality of PCR products. For most samples, single PCRs produced sufficient amount of product, except for screening samples 2, 3 and 4 where mixing products from several parallel PCRs was necessary in order obtain the recommended amount of each product for pyrosequencing. The same cDNA sample was used in each PCR parallel. Samples from PCR were treated with ExoSAP-IT to remove unincorporated dNTPs and primers, following manufacturer instructions (Affymetrix). Purification was performed using the QIAquick PCR Purification kit (Qiagen, Hilden, Germany). Quantification was performed using Nanodrop 2000 (Thermo Scientific). In addition to pyrosequencing each amplicon in both directions, the eight purified HE- and F gene products were adjusted to equimolar amounts for multiplexing. Finally, two single samples (screening sample 1 and outbreak sample 1) and two pooled samples (screening samples 2-4 and outbreak samples 2-4) were prepared and distributed over four regions on a Genome Sequencer FLX (GS-FLX) Titanium platform (Roche) at the Norwegian Sequencing Centre, University of Oslo.

Data Analysis

The CLC Genomics Workbench program (CLC bio) was used for assembly of sequences. Prior to in-depth analysis of the pyrosequencing data, the reads were split only accepting reads containing 0 errors in the multiplex identifiers (MIDs). Sequences were then de-multiplexed, and adaptor sequences, keys, MIDs and primer ends trimmed away. This would, when compared to the Sanger sequence lengths (reference), correspond to amplicon sizes of 250 nucleotides (nts) (HPR HE), 187 nts (delHPR HE) and 194 nts (F). The comparative aspect between the screening and outbreak samples was the main focus of the present work. In order to avoid low-frequency mutational artifacts from the plasmid RT-step and risk of cross contamination during UDPS setup, which could mask true quasispecies present at low frequencies, *in vitro* transcribed plasmid cloned RT-PCR-amplified templates were not included as controls. The “RNA environments” as represented by short identical RNAs (from controls) and total RNA (from samples) could also introduce differential mutational artifacts following RT-PCR amplification. Therefore, a site by site manual comparison of observed frequency of nucleotides differing from the Sanger sequences was chosen as an alternative for all reads from screening and outbreak samples. Based on the results from these analyses a mutational cutoff value of 0.05% was chosen in order to identify mutations occurring only at particular high frequencies. Potential sequencing errors related to homopolymers was addressed by excluding variable sites present as stretches of four or more identical nucleotides. In

addition, mutation frequencies above cutoff, as well as reads containing sequence deletions and/or insertions (see below), would have to be present in both forward and reverse directions in order to be considered. When an estimation of the number of reads containing a particular deletion was to be determined, the minimum length fraction setting was always set to 1.0 while the minimum similarity fraction was set to 0.99. This would allow no flexibility with regards to length of the reads, but would allow for 1 to 3 mismatches, thus providing a necessary flexibility with regards to single nucleotide differences in the reads when determining the number of reads that contained the particular deletion type. CLC Main Workbench (CLC bio) and Align X (Vector NTI Advance™ 11 Package, Invitrogen) were used as additional tools for comparisons between individual sequence reads. RNA secondary structure predictions (minimal free-energy models) of viral (+)/(-)RNAs were performed using the Mfold program (version 2.3) [52]. Default parameters were used in the predictions except for temperature, which was set to 15°C, considered the optimal temperature for ISAV replication [53].

Real-Time PCR on Amplicons

As there was a theoretical possibility that the detection of delHPR reads in screening samples and full-length HPR reads in outbreak samples could have originated from cross-contamination during UDPS setup, TaqMAN real-time PCR was performed on all HE amplicon samples (except screening sample 2, no more sample). The assays were designed specifically to distinguish between full-length HPRs and delHPRs by placing the two reverse primers in this region. First, 1 µl of the amplicon samples used for UDPS was run on a 2% agarose gel for separation of the two size variants. Bands, both visible and estimated, corresponding to full-length or delHPR, were excised and purified using QIAquick Gel Extraction Kit following manufacturers protocol (Qiagen, Hilden, Germany). Real-time PCR was performed using 5 µl of sample and TaqMAN Universal PCR Master Mix (Invitrogen, Carlsbad, USA). Primer- and probes used are shown in Table S1. Real-time PCR was run on a Stratagene Mx3005P (Agilent Technologies) with cycling conditions 95°C/10 min followed by 40 cycles of 95°C/15 sec, 55°C or 60°C/1 min and 72°C/45 sec. Real-time PCR products were run on 2 % agarose gel, bands excised and purified as above, and sequenced using primer s6 HPR_F (Table S1) on a ABI 3130 Genetic Analyser (Invitrogen).

Statistical Approach

To evaluate whether events like mutations or deletions appear more frequently in one region compared to another, a chi-squared test was developed. Here, the actual number of events observed in each region, and the possible numbers of sites represented by the amplicon sizes of each of the regions, were compared. The null hypothesis tested each time was whether the frequencies of events were similar in the two amplicons.

Table 1. Raw data from UDPS of partial ISAV HE- and F gene amplicons.

ISAV Sample(s)	Sequence direction	No. of trimmed reads	Percent of reads identical to Sanger sequence
SCN1 ^a	HE gene forward	50552	82.4
	HE gene reverse	69684	91.4
	F gene forward	44498	91.9
	F gene reverse	49118	89.2
SCN pool (SCN2, SCN3, SCN4)	HE gene forward	25695	78.3
	HE gene reverse	68481	92.2
	F gene forward	31230	92.1
OBK1 ^b	HE gene reverse	35799	89.5
	HE gene forward	52038	92.4
	HE gene reverse	88172	93.0
	F gene forward	31546	91.4
OBK pool (OBK2, OBK3, OBK4)	F gene reverse	46835	89.7
	HE gene forward	40458	92.1
	HE gene reverse	51772	92.8
	F gene forward	45909	91.9
	F gene reverse	44498	91.9

^a. SCN=screening sample

^b. OBK=outbreak sample

doi: 10.1371/journal.pone.0081571.t001

Accession Numbers

The Sanger nucleotide sequences obtained and used in this study have GenBank accession numbers KC907278-KC907293. The UDPS dataset has been deposited into BioProject database (NCBI) with the accession number PRJNA221196.

Results

This paper describes the results obtained from UDPS of the variable regions in ISAV HE- and F genes from non-virulent HPR0-positive fish collected from screening, and a virulent strain sampled from an ISA outbreak at the same farming site three weeks later. From UDPS, mutant spectra were compared and sequence data linked to predicted RNA structural properties of the two genes. Table 1 shows the total number of reads (25695 to 88172) obtained from bidirectional UDPS of HE- and F amplicons following trimming and de-multiplexing. On average, the number of reads identical to Sanger sequence was 90% for both HE and F. In the remaining 10%, mutations, deletions and insertions were observed.

HE Gene Amplicons

The majority of high frequency mutations in HE reads were found within the HPR0 full-length HPR. Generally, the prevalence (i.e. # of different mutational sites compared to Sanger sequence) and frequency (i.e. # of reads containing a particular mutation differing from Sanger sequence) of

mutations above 0.05 % in reads from HE amplicons (i.e. high frequency mutations) seemed higher for screening- compared to outbreak samples, as illustrated in Figure 1a and Table S2. There were 22 sites in total with mutational frequency above 0.05%, 18 in screening- and 5 in outbreak samples (Table S2). A chi-squared test correcting for HPRO HE and delHPR HE amplicon sizes (250 and 187 nts respectively) gave $p = 0.075$, supporting the tendency for a higher number of mutational sites occurring in HE amplicons from screening- compared to those from outbreak. There were a total of 4209 reads containing mutations in HE amplicons with frequency above 0.05%, of which 3319 were from screening samples (Table S2). Comparing the total number of mutations, correcting for differences in amplicon size, showed that HE amplicons from screening samples had a higher frequency of mutational sites compared to those from outbreak ($p < 0.001$, chi-squared test). The majority of high frequency mutational sites were found within the HPR0 full-length HPR which constitutes 63/250 nts of the amplicon. In the single sample from screening, nine sites were in the full-length HPR, six in the rest of the amplicon ($p = 0.008$, chi-squared test). In the pooled sample the corresponding number was five of eight sites in total ($p = 0.051$, chi-squared test) (Figure 1a, Table S2).

Half of the deletion types in HE reads are located at or in close proximity to the T₁₀₄₃-C₁₀₄₄ site. Insertions were not observed in any HE reads with the exception of the low-frequency findings of the non-deleted full-length HPR in outbreak samples (the opposite, the outbreak delHPR pattern in screening samples was also observed, see separate section). On the other hand, a number of different types of deletions were observed in HE reads, especially from outbreak samples. Although present in low frequencies (4-45 reads) they appeared mostly to have a non-random distribution. A total of fourteen different deletion types were observed for HE from outbreak samples (Figure 1a, 2b, Table S2). Two of these are further deletions from T₁₀₄₃-C₁₀₄₄ (positions relative to start of open reading frame). This is the site in the HE gene where a putative precursor full-length HE is hypothesized to have undergone a deletion generating the outbreak delHPR pattern (i.e. the site where the Sanger sequences of full-length HPR and delHPR differ in a multiple sequence alignment). Furthermore, reads containing three other deletion types occurring immediately-, four- and fourteen nts upstream of this site, and two deletion types starting further upstream on the amplicon and ending immediately downstream of the T₁₀₄₃-C₁₀₄₄ site, were observed as well (Figures 1a, 2b, Table S2). The latter deletion type was present in both the single sample from outbreak and the pooled sample. Together, these seven deletion types ("7-dels") all occur within fourteen nts of the T₁₀₄₃-C₁₀₄₄ site, constituting half of all deletion types observed in HE reads from outbreak samples ($p < 0.001$, chi-squared test). Although in some cases the differences were only a few nucleotides, none of the above deletion types produced delHPR patterns identical to that observed in outbreak strains. HE reads containing deletions starting downstream of T₁₀₄₃-C₁₀₄₄ was not observed. The remaining deletion types, both from screening- and outbreak samples, are shown in Figures 1a, 2a, 2b and Table S2.

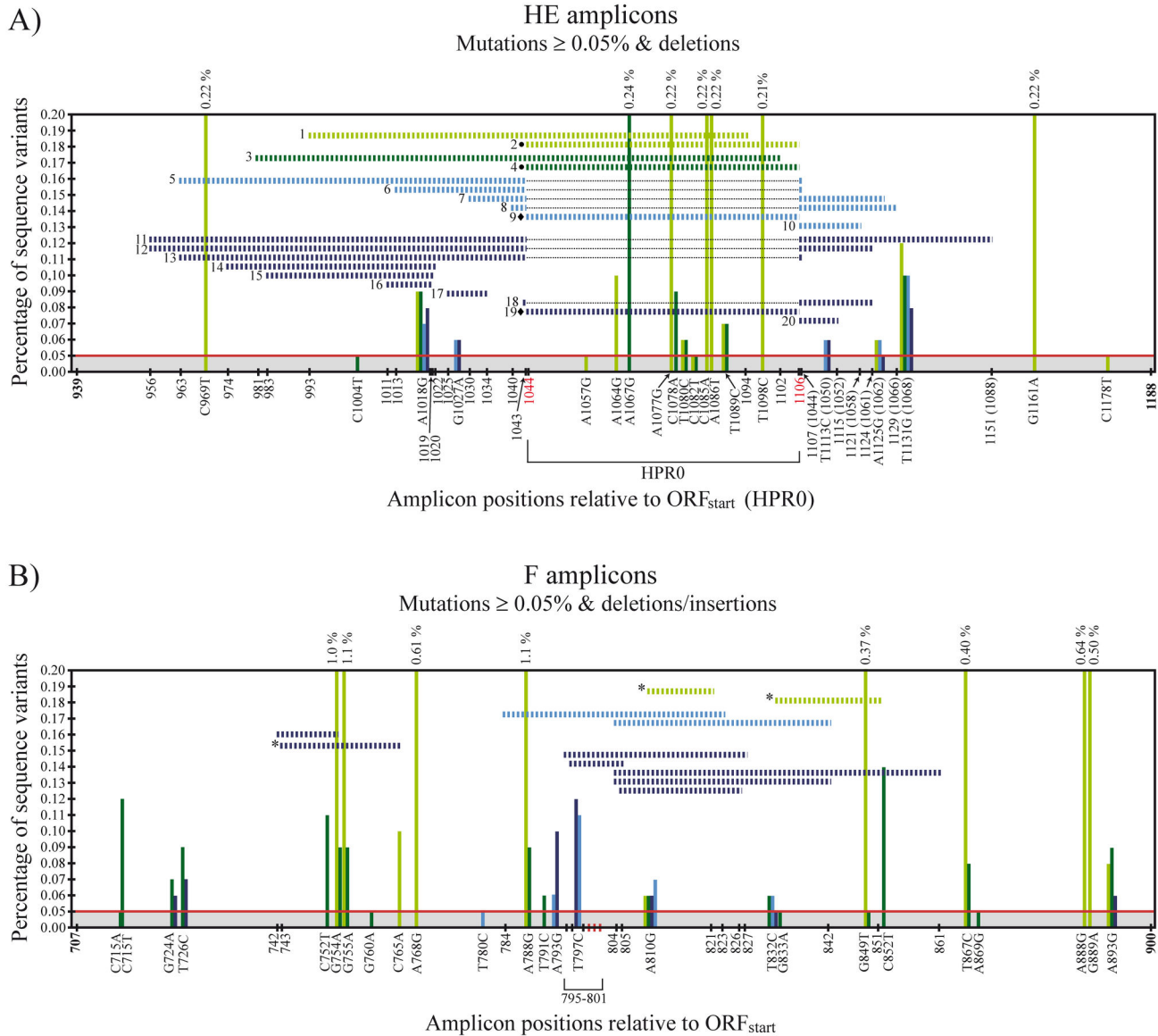


Figure 1. Position, percentage and type of single nucleotide variants (vertical lines), and deletion-insertions and deletions (horizontal lines) detected in reads from UDPS of HE gene (A) and F gene (B) amplicons from a non-virulent HPR0- and a virulent ISAV strain. Grey shading and the red horizontal line illustrate that only mutations present in frequencies $\geq 0.05\%$ are included. Color coding; light green = single screening sample, dark green = pooled screening sample, light blue = single outbreak sample, dark blue = pooled outbreak sample. The three red vertical lines in B) is the L_{266}/Q_{266} codon in the F gene. Numbering of deletion events in A) correspond to numbering in Figures 2a, b. * = deletion-insertion types observed in F reads (regions deleted are shown, see Table S3 for details), • = outbreak delHPRs in screening samples, and ♦ = full-length HPRs in outbreak samples. Numbers in bold indicate ORF_{start} and ORF_{end} positions in HPR0 HE- and F genes not containing insertions. Red numbers indicate start and end of the full-length HPR, and numbers in parenthesis downstream of this region are the corresponding positions relative ORF_{start} in the outbreak delHPR. Dotted lines illustrate the full-length HPR portion not present in the delHPR.

doi: 10.1371/journal.pone.0081571.g001

Regions flanking the delHPR T_{1043} - C_{1044} site but not the full-length HPR are predicted to fold into a RNA hairpin motif. The lowest free energy predicted (+)RNA secondary structures for the Sanger sequences of the HE amplicons from screening and outbreak suggests that the full-length HPR in the

former folds into a structure with less basepairing (i.e. "HPR0 RNA motif") compared to a single hairpin motif predicted to be formed by sequences flanking the T_{1043} - C_{1044} site in the latter (Figures 2a, b). These two motifs, displayed only when the predictions were run in the sense (+) orientation, were

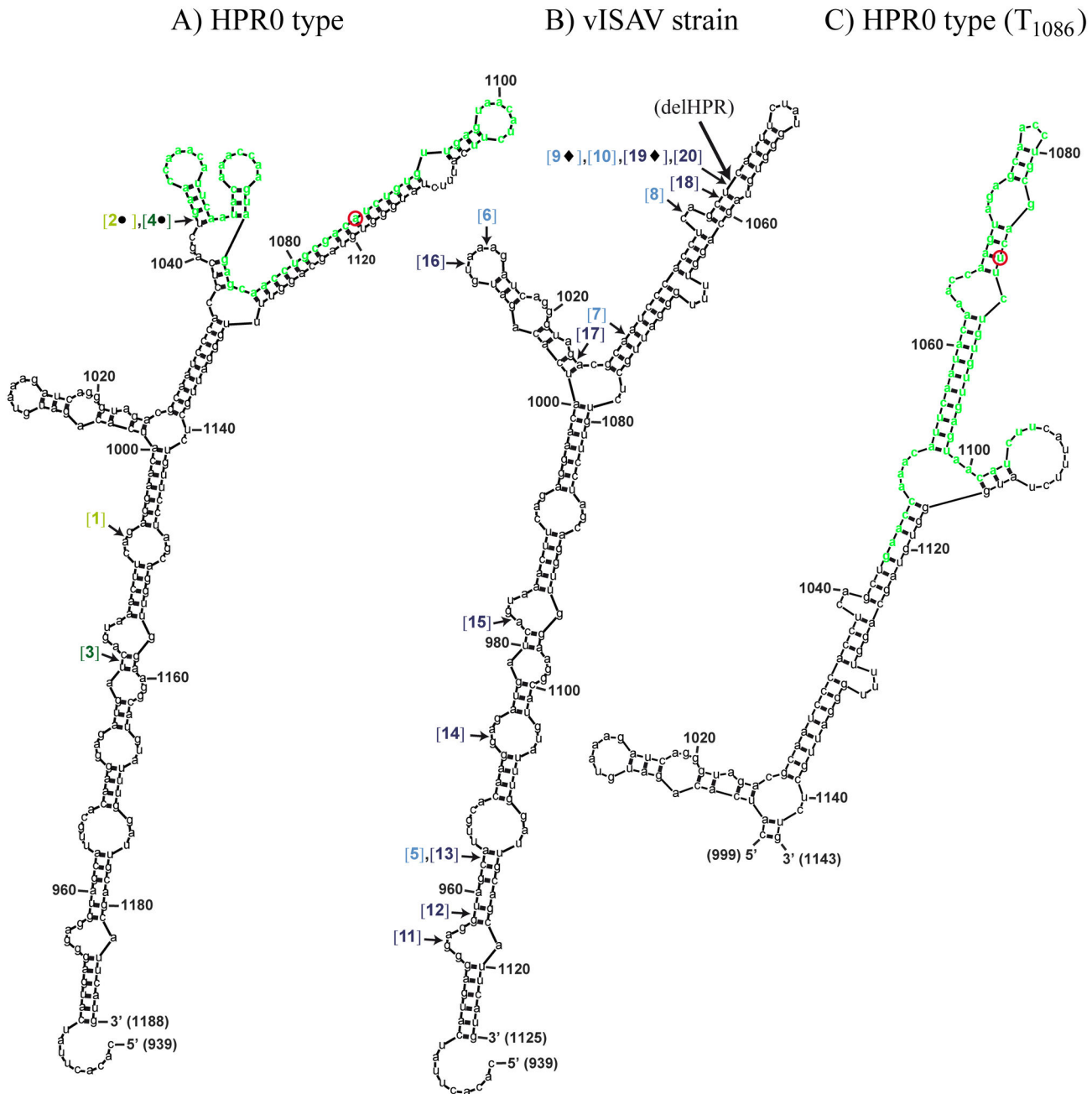


Figure 2. RNA secondary structure predictions of HE amplicons using mFold. The lowest free energy predicted structures of the RNAs in the sense (+) orientations are shown. A) Sanger sequence of non-virulent HPR0 strain, B) Sanger sequence of the virulent ISAV strain, and C) partial Sanger sequence from screening with a single $A_{1086} \rightarrow T_{1086}$ mutation (position indicated by red circle in A) and C)). Default settings were used in the predictions, except for temperature which was set to 15°C. Sequence numbering is according to ORF_{start}. The nucleotides constituting the HPR0 full-length HPR in A) and C) are shown in green color, and the site where this region is missing in the sequence from the virulent strain (between T_{1043} - C_{1044}) in B) is indicated by a large arrow and “delHPR”. Numbers in brackets indicate the 5'-flanking positions of deletion sites. Here, the numbering, symbols and color coding correspond to that used in Figure 1a.

doi: 10.1371/journal.pone.0081571.g002

consistently present with varying input sequence lengths up to 800 nts (the largest input length allowed in the program) (not

shown). Comparing UDPS data for HE from outbreak samples with predicted RNA structure shows that the 7-dels all start or

end within the predicted hairpin motif, with four starting at or close to the T₁₀₄₃-C₁₀₄₄ site in the top half of this structure (Figure 2b). As a higher proportion of high frequency mutation sites (11) were present within the full-length HPR from screening samples, we also checked whether any of these single nucleotide variants would alter the predicted RNA structure of the Sanger sequence in this region. Tested separately, only the A₁₀₈₆→T₁₀₈₆ mutation predicted a significant structural change in (+)RNA, transforming most of the predicted HPR0 RNA motif into a single hairpin structure similar to that predicted for the delHPR (Figures 2a, c). In the pooled samples, not a single read containing the A₁₀₈₆→T₁₀₈₆ mutation was present.

F Gene Amplicons

F reads from the single screening sample contain sites with high mutation frequencies. As observed for HE reads, the prevalence and frequency of mutations in F reads above cutoff ($\geq 0.05\%$) were higher for screening samples compared to outbreak samples (Figure 1b, Table S3). There were 25 sites in total with mutational frequency above 0.05%, 22 in screening- and 7 in outbreak samples (Table S3) ($p = 0.013$, chi-squared test) supporting also for F the tendency for a higher number of mutational sites occurring in amplicon samples from screening- compared to those from outbreak. There were a total of 7584 reads containing mutations in F amplicons with frequency above 0.05%, of which 6799 were from screening samples (Table S3). Similar to HE, F amplicons from screening samples has a higher frequency of mutational sites compared to those from outbreak ($p < 0.001$, chi-squared test). Opposed to all other samples though, the single sample from screening contained several sites displaying exceptionally high variability ($\geq 1\%$) (Figure 1b, Table S3). These sites were also present in high frequency in the pooled sample, although with a prevalence approximately 10x lower.

Frequencies of L₂₆₆ codon in screening samples and Q₂₆₆ codon in outbreak samples were not higher than at any other random site. As the codon immediately upstream of the putative encoded cleavage sites R₂₆₇ in the F gene marks a crucial difference between HPR0s and vISAVs it was of interest to see whether the codon for this site displayed higher variability. We found that the mutation frequency in this codon, CA₇₉₇G↔CT₇₉₇G (Q₂₆₆↔L₂₆₆) was not higher than the mutation frequency at any random site in the two amplicon types ($<<0.05\%$). Also, the A₇₉₇ vs. T₇₉₇ does not introduce changes in predicted (+)RNA or (-)RNA structures. In contrast, the only other site that differs between the F gene Sanger sequences from screening and outbreak, G₇₅₅ vs. A₇₅₅, does introduce significant changes in predicted RNA structure in the immediate region in both (+)- and (-) RNA (not shown). Hence, the contribution of this nt difference to the differing mutant spectra profiles between the two types of F amplicons, cannot be excluded.

Both deletions and insertions are found in F reads, the majority located in proximity to the R₂₆₇ codon. ISAV strains with small insertions (originating from other parts of the ISAV genome) close to the putative encoded cleavage site R₂₆₇ in the F gene have been documented [30-32]. In the present

study, reads containing sequence insertions and/or deletions were observed in low frequency from both screening- and outbreak samples (4-30 reads). These are all shown in Figure 1b and Table S3. Of a total of three deletion-insertion events, two are from the single screening sample where one, a 23 nt insertion, is located only 6 positions downstream of where insertions have been found in some ISAV outbreak strains. In contrast, the number of reads containing deletions of varying lengths was higher. Five of a total of eight deletion types ("5-dels"), all from outbreak samples, start at the same position (3), immediately upstream (1) or downstream (1) of where insertions have been observed in outbreak strains (at two sites separated by 11 nts). These 5-dels occur within a 12 nt region, which is a higher occurrence compared to the rest of the amplicon region ($p < 0.001$, chi-squared test). RNA secondary structure predictions on the Sanger sequences using mFold suggest that the 5-dels, and the 23 nt insertion type from the screening sample may, similar to that seen for the HE from outbreak (see above), occur within hairpin structures (not shown). For amplicons from screening samples, the hairpin motif was most prominent when the prediction was run on (-)RNA. Of the 5-dels from outbreak, three were located at the distal tip of a predicted hairpin, the structural feature displayed in all lowest energy predicted (+/-)RNA structures (up to the maximum allowed input sequence length of 800 nts).

Double Infection with HPR0 and vISAV

DelHPR's in screening samples and vice versa. Sequencing (Sanger) of upstream regions of the HE genes from one screening sample (screening sample 3) and one outbreak sample (outbreak sample 3) verified that the outbreak strain from Site 10 was identical to the outbreak strain from two years earlier at the same site. Together with information previously obtained from the Astafjord region [28], the sequencing showed that the HPR0 genotype at this site was not the likely precursor of the outbreak strain that emerged three weeks later (not shown). From UDPS, both the single- and pooled sample from screening, low frequency reads (21-30) containing deletion pattern identical to the Site 10 delHPR Sanger sequence, were observed (Figure 1a, Table S2). The opposite was also found, reads (22,23) containing full-length HPR in outbreak samples. To confirm their presence and rule out the possibility of sample cross-contamination, two real-time PCR assays using reverse primers differentiating between the full-length HPR and the delHPR were run on all samples prior to pooling and shipping for UDPS (Table S1). The results, including sequencing of the real-time PCR products, verified the presence of these low frequency reads (not shown).

Discussion

UDPS of non-virulent ISAV HPR0- and virulent ISAV strains is presented for the first time. Gill tissues were sampled from healthy fish positive for the HPR0 strain from a screening regime, and a virulent (v) ISAV strain from a subsequent ISA disease outbreak at the same fish farming site three weeks later. HE- and F gene regions including the HPR and the codon

for the putative encoded R₂₆₇ cleavage site were amplified by PCR, pyrosequenced and detailed analysis of the sequence data was performed. The raw UDPS data revealed that roughly 10% of the reads in each sample differed from the Sanger sequences. Mutant spectra from the two ISAV strains were compared and linked to predicted viral RNA structures.

Different mutant spectra patterns were observed for both HE- and F amplicons between the HPRO- and vISAV strains, with a higher number of high frequency mutation sites in screening than outbreak samples (see below). Moreover, HE reads from screening samples revealed a strong preference for high frequency mutations within the full-length HPR, suggesting this region to be more prone to mutations compared to the flanking regions of the amplicon. For many RNA viruses, studies have shown that nucleotide sites not forming molecular base pairs tend to show higher variability compared to those who do [54]. It should therefore not be excluded that the predicted RNA structure characteristics of the loop-containing HPRO motif contributes to the higher frequency of mutations observed in this region. In F reads from single screening sample, the exceptionally high mutational frequency ($\geq 1.0\%$) observed at several positions may correlate with the fact that this sample initially contained a viral load 10-100x higher, as estimated by real-time RT-PCR (not shown), compared to the individual samples included in the pooled screening sample. On the other hand, the CT₇₉₇G vs. CA₇₉₇G codons (i.e. the L₂₆₆ vs. Q₂₆₆ virulence marker) were not found in higher frequencies in screening vs. outbreak samples, respectively, than the mutational frequency at any random position.

Reads containing deletions were observed in both HE- and F reads while insertions were only observed in the latter. There was a higher number of different deletion types in HE reads than in F reads. For HE, half of the deletion types from outbreak samples (7 of 14) start or end within a fourteen nt region, where two are further deletions from T₁₀₄₃-C₁₀₄₄, the site where the original deletion from a precursor HPRO genotype can be hypothesized to have occurred. Similarly, of a total of ten different insertion-deletion- or deletion types found in F reads, three were found to start at the same position were most F gene insertions have been found in outbreak strains [30-32]. It cannot be excluded that several of these events have been artificially generated during the PCR amplification step or during UDPS [55,56]. However, studies have indicated that the *in vitro* recombination rate in UDPS is low [56]. For HE, this suggests a tendency towards further deletions from T₁₀₄₃-C₁₀₄₄ following a primary deletion event, possibly linked to increased viral fitness associated with larger deletions. In fact, closely related vISAVs isolated in the course of the same outbreak at the same fish farming site have been found to differ only by the size of their deletions in the HPR, where one isolate is likely to have originated from a further deletion of the other [57].

The results obtained from UDPS suggest a link between mutant spectra profiles and the predicted hairpin motifs observed with Sanger sequences of HE (+)RNA and F (+/-)RNAs from outbreak, and F (-)RNA from screening. Such a RNA structure may facilitate mutational events such as insertions (F) and deletions (HE) through mechanisms such as RNA polymerase jumping- and/or template switching during

replication. For influenza A viruses, jumping of the RNA polymerase has been suggested as a mechanism for generating deletions and insertions in hemagglutinin and neuraminidase genes [58]. The involvement of RNA structure in the generation of the polybasic cleavage site in highly pathogenic avian influenza A strains, by the polymerase slipping during strand synthesis upon arrival at a region of higher stability, has also been suggested [59]. For F genes, a sequence-based non-homologous replicase-driven template switch, as observed in positive-stranded RNA viruses, has been proposed as the most likely mechanism behind the sequence insertions observed in some vISAVs [30,60]. The inserted sequences in F genes were all found to originate from other parts of the ISAV genome, and not the host, we believe most likely because these sequences are in close proximity to one another, associated with replicase complex, during ISAV replication in the nucleolus [61].

From the present UDPS data, only one deletion-insertion event (F reads, pooled outbreak sample) and one deletion event (HPRO HE reads, single screening sample) seem to have been generated through the involvement of extensive sequence homology (not shown), although whether these were generated artificially or are true quasispecies, is not known. In general though, recombination is known to be rare in negative strand- compared to positive-stranded RNA viruses like *Picornaviridae* and *Coronaviridae* [60,62]. Our results from UDPS data- and RNA secondary structure prediction analysis suggests that RNA structural properties may be involved in generating the F gene insertions, supported also by the fact that non-homologous recombination is very rare in orthomyxoviruses. For ISAV HE, based on the results from both UDPS and RNA structure predictions, a plausible mechanism for creating the deletions in HE is likely to involve the replicase complex bypassing a sequence stretch on the same template or jumping to a homologous template reinitiating strand synthesis further downstream. This event is hypothesized to be driven through alterations in RNA structure (i.e. hairpin structure) following point mutations in the full-length HPR. The distributions of deletions and the total absence of insertions in HE reads together with current epidemiological information on vISAVs supports this hypothesis. Compared to previous publications on this issue, the present results strongly suggest that the recombination mechanisms creating the deletions do not involve insertions nor recombination between HPR regions between different virulent strains [14].

The high frequency of mutations within the full-length HPR of screening samples prompted investigation of whether any of these mutations (11 in all) tested individually would change the predicted (+/-)RNA structure of the Sanger HPRO motif. Results from several studies suggest a direct link between the structure of viral RNA and the fidelity of the RNA polymerase [54]. Only one mutation, A₁₀₈₆ to T₁₀₈₆, found in reads from the single screening sample resulted in a change in the predicted structure, transforming most of the predicted HPRO motif into hairpin-like structure. Hence, single mutations in the full-length HPR may have the potential to transform this RNA motif into one resembling the regions flanking the delHPR site, potentially facilitating deletions in this region. In fact, for avian influenza A

viruses the generation of the polybasic cleavage site characteristic of high pathogenic strains has been suggested also to be mediated by a mechanism in which single nucleotide changes alter RNA secondary structure facilitating the introduction of small sequence insertions in this region [63]. Here, aside from the fact that the HPR from the single screening sample contained roughly twice the amount of high frequency mutation sites compared to the pooled sample, it is noteworthy that not a single read from the pooled sample contained the T₁₀₈₆ mutation. This might reflect a link between the severity of infection by a HPR0 strain, possibly influenced by a high-stress fish farming environment, and the composition of mutant spectra. Studies have shown that differences in HPR0 viral loads between individuals is linked to the point in time of infection and that the virus may be present, although transiently, in the population for some time, with no clinical signs of ISA [27]. The effect may thus be a broadening of the mutant spectra in HPR0 quasispecies in spite of low replication rate for these non-virulent strains. In contrast, upon mutation leading to vISAVs, one sequence that harbours the mutation will outcompete the other sequences and spread as a single strain displaying narrower mutant spectra, at least in the early stages of the disease outbreak. In the present study, sampling of the vISAV strain was performed in the early phases of the disease outbreak and displayed many magnitudes higher viral loads compared to the samples from screening.

We hypothesize that the mutant spectra displayed by HPR0 infections in a fish farming environment is directly linked to the potential for deletions in the full-length HPR, and that RNA structure plays a central role in this transition. Also, the observed differences in mutant spectra by the two ISAV strains (like the higher number of high frequency mutation sites in screening vs. outbreak samples) may not only be linked to differences in infection period but also to replication rates, as well as to possible variations in the viral RdRps that may affect both replication rate and replication fidelity, thus changing quasispecies populations and hence viral fitness [42,64-68]. Hence, the potential involvement of sequence differences in internal genes on the differential mutational frequencies observed between the HPR0- and vISAV strains cannot be excluded. Also, the putative gill tropism displayed by HPR0s in farmed and wild Atlantic salmon compared to the systemic nature of vISAV infections, possibly involving different cell types in the gills [69] and different parts of the immune system, may also be a contributing factor to the observed differences in the mutant spectra. As ISA has not been observed in wild Atlantic salmon, a future comparative study between mutant spectra from HPR0-positive wild salmon with that obtained in the present study could provide valuable insight into the mechanisms of ISAV transition into virulence. It should also be considered that Atlantic salmon may not be the natural host for HPR0s and the transition to vISAV could be viewed as an adaptation of the virus in this particular host.

This study represents the first documentation of a double infection involving non-virulent HPR0s and vISAVs. The low number of delHPR reads in screening samples corresponds well with the number of reads containing L₂₆₆ codons, as do the low number of full-length HPRs vs. Q₂₆₆ codons from outbreak

samples. Although statistically unlikely, it cannot be excluded that one or several of the deletions observed in HE- and F amplicons are linked to the low frequency presence of the other ISAV variant. For single site mutations though, the threshold level set eliminates this potential misinterpretation. From Sanger sequencing of larger portions of the HE genes it was deemed unlikely that the HPR0 strain from the screening samples was the direct precursor of the vISAV strain in this particular fish farm. Hence, many individuals showing no signs of ISA disease were already infected with a virulent strain at the time of screening, suggesting that infection with a HPR0 strain may not protect the fish from ISA disease. Also, vaccination against ISA has shown little or no protection against infection with HPR0s [27]. The opposite finding, i.e. the presence of full-length HPRs in outbreak samples, could either be due to the fact that sampling was performed in the early stages of the disease outbreak, or that an earlier infection with a HPR0 virus persists for some time at basal levels in ISA-diseased individuals caused by a virulent strain. Such double infections may actually play an important role in ISAV evolution.

UDPS technology has shown itself to be a valuable tool in detecting minority within-host mutant variants of viral genomes not detected by traditional means, such as low-level persistence of drug resistant mutations in individuals following prophylactic treatment, or double infections [44,70-73]. The latter is also the case here, where the low-frequency presence of delHPRs in screening samples were not detected through the standard diagnostic procedures involving RT-PCR and Sanger sequencing, thus demonstrating the future potential UDPS technology may have in ISAV screening regimes and diagnostics.

In conclusion, we have performed UDPS of HE- and F variable gene regions from samples containing a non-virulent HPR0 strain and a virulent ISAV strain. Detailed analysis and comparisons of mutant spectra from both individual and pooled samples revealed marked differences between the two virus strains. We propose a new hypothesis were the introduction of selected mutations in the full-length HPR can alter the RNA structure in such a way that facilitates deletion events in this gene region, making both quasispecies composition and RNA structure important factors involved in the two known molecular events leading to ISAV virulence. The results further strengthen the hypothesis that deletions, and not recombination, form the variability of delHPRs in virulent ISAVs and support observations from Norway and the Faroe Islands that this transition to virulence is an infrequent event.

Supporting Information

Table S1. PCR primers and probes.
(DOCX)

Table S2. Type, number and prevalence of mutations present in $\geq 0.05\%$ of reads and deletions as detected by UDPS of HE amplicons.
(DOCX)

Table S3. Type, number and prevalence of mutations present in $\geq 0.05\%$ of reads, deletions and insertions as detected by UDPS of F amplicons. (DOCX)

References

- Thorud KE, Djupvik HO (1988) Infectious anemia in Atlantic salmon (*Salmo salar* L.). Bull Eur Assoc Fish Pathol 8: 109-111
- Bouchard D, Brockway K, Giray C, Keleher W, Merrill PL (2001) First report of infectious salmon anaemia (ISA) in the United States. Bull Eur Assoc Fish Pathol 21: 86-88.
- Lovely JE, Dannevig BH, Falk K, Hutchin L, MacKinnon AM et al. (1999) First identification of infectious salmon anaemia virus in North America with haemorrhagic kidney syndrome. Dis Aquat Organ 35: 145-148. doi:10.3354/dao035145. PubMed: 10092978.
- Mullins J, Groman D, Wadowska D (1998) Infectious salmon anemia in salt water Atlantic salmon (*Salmo salar* L.) in New Brunswick, Canada. Bull Eur Assoc Fish Pathol 18: 110-114
- Rodger HD, Richards RH (1998) Haemorrhagic smolt syndrome: a severe anaemic condition in farmed salmon in Scotland. Vet Rec 142: 538-541. doi:10.1136/vr.142.20.538. PubMed: 9637379.
- Lyngrøy C (2003) Infectious salmon anaemia in Norway and the Faeroe Islands: an industrial approach. In: O MillerRC Cipriniano. International Response to Infectious Salmon Anaemia: Prevention, Control and Eradication. pp. 97-109.
- Rowley HM, Campbell SJ, Curran WL, Turnbull T, Bryson DG (1999) Isolation of infectious salmon anaemia virus (ISAV) from Scottish farmed Atlantic salmon, *Salmo salar* L. J Fish Dis 22: 483-487. doi:10.1046/j.1365-2761.1999.00190.x.
- Godoy MG, Aedo A, Kibenge MJ, Groman DB, Yason CV et al. (2008) 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 4: 28.
- Falk K, Aspehaug V, Vlasak R, Endresen C (2004) Identification and characterization of viral structural proteins of infectious salmon anemia virus. J Virol 78: 3063-3071. doi:10.1128/JVI.78.6.3063-3071.2004. PubMed: 14990725.
- Rimstad E, Mjaaland S, Snow M, Mikalsen AB, Cunningham CO (2001) Characterization of the infectious salmon anaemia virus genomic segment that encodes the putative hemagglutinin. J Virol 75: 5352-5356. doi:10.1128/JVI.75.11.5352-5356.2001. PubMed: 11333916.
- Krossøy B, Devold M, Sanders L, Knappskog PM, Aspehaug V et al. (2001) Cloning and identification of the infectious salmon anaemia virus haemagglutinin. J Gen Virol 82: 1757-1765. PubMed: 11413388.
- Aspehaug V, Mikalsen AB, Snow M, Biering E, Villøing S (2005) Characterization of the infectious salmon anemia virus fusion protein. J Virol 79: 12544-12553. doi:10.1128/JVI.79.19.12544-12553.2005. PubMed: 16160182.
- Mjaaland S, Hungnes O, Teig A, Dannevig BH, Thorud K et al. (2002) Polymorphism in the infectious salmon anemia virus hemagglutinin gene: importance and possible implications for evolution and ecology of infectious salmon anemia disease. Virology 304: 379-391. doi:10.1006/viro.2002.1658. PubMed: 12504577.
- Devold M, Falk K, Dale B, Krossøy B, Biering E, et al. (2001) Strain variation, based on the hemagglutinin gene, in Norwegian ISA virus isolates collected from 1987 to 2001: indications of recombination. Dis Aquat Organ 47: 119-128
- Castrucci MR, Kawaoka Y (1993) Biologic importance of neuraminidase stalk length in influenza A virus. J Virol 67: 759-764. PubMed: 8419645.
- Hossain MJ, Hickman D, Perez DR (2008) Evidence of expanded host range and mammalian-associated genetic changes in a duck H9N2 influenza virus following adaptation in quail and chickens. PLOS ONE 3: e3170. doi:10.1371/journal.pone.0003170. PubMed: 18779858.
- Li J, Cardona CJ (2010) Adaptation and transmission of a wild duck avian influenza isolate in chickens. Avian Dis 54: 586-590. doi:10.1637/8806-040109-ResNote.1. PubMed: 20521699.
- Sorrell EM, Song H, Pena L, Perez DR (2010) A 27-amino-acid deletion in the neuraminidase stalk supports replication of an avian H2N2 influenza A virus in the respiratory tract of chickens. J Virol 84: 11831-11840. doi:10.1128/JVI.01460-10. PubMed: 20826691.
- Xu R, Zhu X, McBride R, Nycholat CM, Yu W et al. (2012) Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. J Virol 86: 9221-9232. doi:10.1128/JVI.00697-12. PubMed: 22718832.
- de Vries E, de Vries RP, Wienholts MJ, Floris CE, Jacobs MS et al. (2012) Influenza A virus entry into cells lacking sialylated N-glycans. Proc Natl Acad Sci U S A 109: 7457-7462. doi:10.1073/pnas.1200987109. PubMed: 22529385.
- Lu B, Zhou H, Ye D, Kemble G, Jin H (2005) Improvement of influenza A/Fujian/411/02 (H3N2) virus growth in embryonated chicken eggs by balancing the hemagglutinin and neuraminidase activities, using reverse genetics. J Virol 79: 6763-6771. doi:10.1128/JVI.79.11.6763-6771.2005. PubMed: 15890915.
- Mitnaul LJ, Matrosovich MN, Castrucci MR, Tuzikov AB, Bovin NV et al. (2000) Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. J Virol 74: 6015-6020. doi:10.1128/JVI.74.13.6015-6020.2000. PubMed: 10846083.
- Shtyrya Y, Mochalova L, Voznova G, Rudneva I, Shilov A et al. (2009) Adjustment of receptor-binding and neuraminidase substrate specificities in avian-human reassortant influenza viruses. Glycoconj J 26: 99-109. doi:10.1007/s10719-008-9169-x. PubMed: 18661232.
- Cook-Versloot M, Griffiths S, Cusack R, McGeachy S, Richie R (2004) Identification and characterization of infectious salmon anaemia virus (ISAV) haemagglutinin gene highly polymorphic region (HPR) type 0 in North America. Bull Eur Assoc Fish Pathol 4: 203-208.
- Cunningham CO, Griffiths S, Black J, Simpson I, Raynard RS (2002) A novel variant of the infectious salmon anemia virus (ISAV) haemagglutinin gene suggests mechanisms for virus diversity. Bull Eur Assoc Fish Pathol 22: 366-374.
- McBeath AJ, Bain N, Snow M (2009) Surveillance for infectious salmon anaemia virus HPR0 in marine Atlantic salmon farms across Scotland. Dis Aquat Organ 87: 161-169. doi:10.3354/dao02128. PubMed: 20099410.
- Christiansen DH, Østergaard PS, Snow M, Dale OB, Falk K (2011) A low-pathogenic variant of infectious salmon anemia virus (ISAV-HPR0) is highly prevalent and causes a non-clinical transient infection in farmed Atlantic salmon (*Salmo salar* L.) in the Faroe Islands. J Gen Virol 92: 909-918. doi:10.1099/vir.0.027094-0. PubMed: 21148272.
- Lyngstad TM, Hjortaaas MJ, Kristoffersen AB, Markussen T, Karlens ET et al. (2011) Use of molecular epidemiology to trace transmission pathways for infectious salmon anaemia virus (ISAV) in Norwegian salmon farming. Epidemics 3: 1-11. doi:10.1016/j.epidem.2010.11.001. PubMed: 21420655.
- Lyngstad TM, Kristoffersen AB, Hjortaaas MJ, Devold M, Aspehaug V et al. (2012) Low virulent infectious salmon anaemia virus (ISAV-HPR0) is prevalent and geographically structured in Norwegian salmon farming. Dis Aquat Organ 101: 197-206. doi:10.3354/dao02520. PubMed: 23324416.
- Markussen T, Jonassen CM, Numanovic S, Braaen S, Hjortaaas M et al. (2008) Evolutionary mechanisms involved in the virulence of infectious salmon anaemia virus (ISAV), a piscine orthomyxovirus. Virology 374: 515-527. doi:10.1016/j.virol.2008.01.019. PubMed: 18280528.
- Devold M, Karlens M, Nylund A (2006) Sequence analysis of the fusion protein gene from infectious salmon anaemia virus isolates: evidence of recombination and reassortment. J Gen Virol 87: 2031-2040. doi:10.1099/vir.0.81687-0. PubMed: 16760406.
- Kibenge FS, Godoy MG, Wang Y, Kibenge MJ, Gherardelli V et al. (2009) 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 6: 88. doi:10.1186/1743-422X-6-88. PubMed: 19558648.
- Bosch FX, Garten W, Klenk HD, Rott R (1981) Proteolytic cleavage of influenza virus hemagglutinins: primary structure of the connecting peptide between HA1 and HA2 determines proteolytic cleavability and pathogenicity of Avian influenza viruses. Virology 113: 725-735. doi:10.1016/0042-6822(81)90201-4. PubMed: 7023022.

Author Contributions

Conceived and designed the experiments: TM HS TT MJH. Performed the experiments: TM SN MJH. Analyzed the data: TM CMJ ABK. Wrote the manuscript: TM HS CMJ TT ABK JR DHC OBD KF.

34. Senne DA, Panigrahy B, Kawaoka Y, Pearson JE, Süß J et al. (1996) Survey of the hemagglutinin (HA) cleavage site sequence of H5 and H7 avian influenza viruses: amino acid sequence at the HA cleavage site as a marker of pathogenicity potential. *Avian Dis* 40: 425-437. doi: 10.2307/1592241. PubMed: 8790895.
35. Ojosnegros S, Perales C, Mas A, Domingo E (2011) Quasispecies as a matter of fact: viruses and beyond. *Virus Res* 162: 203-215. doi: 10.1016/j.virusres.2011.09.018. PubMed: 21945638.
36. Domingo E, Martin V Perales C, Grande-Perez A, Garcia-Arriaza J et al. (2006) Viruses as quasispecies: biological implications. *Curr Top Microbiol Immunol* 299: 51-82. doi:10.1007/3-540-26397-7_3. PubMed: 16568896.
37. Domingo E, Sheldon J, Perales C (2012) Viral quasispecies evolution. *Microbiol Mol Biol Rev* 76: 159-216. doi:10.1128/MMBR.05023-11. PubMed: 22688811.
38. Steinhauer DA, Domingo E, Holland JJ (1992) Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 122: 281-288. doi:10.1016/0378-1119(92)90216-C. PubMed: 1336756.
39. Ferrer-Orta C, Arias A, Pérez-Luque R, Escarmis C, Domingo E et al. (2007) Sequential structures provide insights into the fidelity of RNA replication. *Proc Natl Acad Sci U S A* 104: 9463-9468. doi:10.1073/pnas.0700518104. PubMed: 17517631.
40. Menéndez-Arias L (2002) Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases. *Prog Nucleic Acid Res Mol Biol* 71: 91-147. doi:10.1016/S0079-6603(02)71042-8. PubMed: 12102562.
41. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439: 344-348. doi:10.1038/nature04388. PubMed: 16327776.
42. Pfeiffer JK, Kirkegaard K (2005) Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* 1: e11. doi:10.1371/journal.ppat.0010011. PubMed: 16220146.
43. Jerzak GV, Bernard K, Kramer LD, Shi PY, Ebel GD (2007) The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology* 360: 469-476. doi:10.1016/j.virol.2006.10.029. PubMed: 17134731.
44. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ et al. (2007) DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res* 35: e91. doi:10.1093/nar/gkm435. PubMed: 17576693.
45. Mitsuya Y, Varghese V, Wang C, Liu TF, Holmes SP et al. (2008) Minority human immunodeficiency virus type 1 variants in antiretroviral-naive persons with reverse transcriptase codon 215 revertant mutations. *J Virol* 82: 10747-10755. doi:10.1128/JVI.01827-07. PubMed: 18715933.
46. Eriksson N, Pachter L, Mitsuya Y, Rhee SY, Wang C et al. (2008) Viral population estimation using pyrosequencing. *PLoS Comput Biol* 4: e1000074. PubMed: 18437230.
47. Nishijima N, Marusawa H, Ueda Y, Takahashi K, Nasu A et al. (2012) Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing. *PLOS ONE* 7: e35052. doi:10.1371/journal.pone.0035052. PubMed: 22523569.
48. Kuroda M, Katano H, Nakajima N, Tobiume M, Aina A et al. (2010) Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLOS ONE* 5: e10256. doi:10.1371/journal.pone.0010256. PubMed: 20428231.
49. Deng YM, Caldwell N, Hurt A, Shaw T, Kelso A et al. (2011) A comparison of pyrosequencing and neuraminidase inhibition assays for the detection of oseltamivir-resistant pandemic influenza A(H1N1) 2009 viruses. *Antiviral Res* 90: 87-91. doi:10.1016/j.antiviral.2011.02.014. PubMed: 21376084.
50. Kulshreshtha V, Kibenge M, Saloni K, Simard N, Riveroll A et al. (2010) Identification of the 3' and 5' terminal sequences of the 8 ma genome segments of European and North American genotypes of infectious salmon anemia virus (an orthomyxovirus) and evidence for quasispecies based on the non-coding sequences of transcripts. *Virology* 403: 338. doi:10.1016/j.virol.2010.07.038. PubMed: 21092282.
51. Bornø G, Sviland C, Jensen BB, Biering E, Johansen R, et al. (2010) Helsestatusjonen hos laksefisk 2010 (Health situation in farmed Atlantic salmon 2010; in Norwegian). In: *Fiskehelse rapporten 2010 (Fish Health Report 2010)*. Report from the Norwegian Veterinary Institute, Norway. pp. 1-36
52. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406-3415. doi: 10.1093/nar/gkg595. PubMed: 12824337.
53. Falk K, Namork E, Rimstad E, Mjåland S, Dannevig BH (1997) Characterization of infectious salmon anemia virus, an orthomyxo-like virus isolated from Atlantic salmon (*Salmo salar* L.). *J Virol* 71: 9016-9023
54. Sanjuán R, Bordería AV (2011) Interplay between RNA structure and protein evolution in HIV-1. *Mol Biol Evol* 28: 1333-1338. doi:10.1093/molbev/msq329. PubMed: 21135148.
55. Görzer I, Guelly C, Trajanoski S, Puchhammer-Stöckl E (2010) The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. *J Virol Methods* 169: 248-252. doi:10.1016/j.jviromet.2010.07.040. PubMed: 20691210.
56. Mild M, Hedskog C, Jernberg J, Albert J (2011) Performance of ultra-deep pyrosequencing in analysis of HIV-1 pol gene variation. *PLOS ONE* 6: e22741. doi:10.1371/journal.pone.0022741. PubMed: 21799940.
57. Lyngstad TM, Jansen PA, Sindre H, Jonassen CM, Hjortaa MJ et al. (2008) Epidemiological investigation of infectious salmon anaemia (ISA) outbreaks in Norway 2003-2005. *Prev Vet Med* 84: 213-227. doi: 10.1016/j.prevetmed.2007.12.008. PubMed: 18243376.
58. Fields S, Winter G (1982) Nucleotide sequences of influenza virus segments 1 and 3 reveal mosaic structure of a small viral RNA segment. *Cell* 28: 303-313. doi:10.1016/0092-8674(82)90348-8. PubMed: 7060132.
59. Perdue ML, García M, Senne D, Fraire M (1997) Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Res* 49: 173-186. doi:10.1016/S0168-1702(97)01468-8. PubMed: 9213392.
60. Nagy PD, Simon AE (1997) New insights into the mechanisms of RNA recombination. *Virology* 235: 1-9. doi:10.1006/viro.1997.8681. PubMed: 9300032.
61. Goić B, Bustamante J, Miquel A, Alvarez M, Vera MI et al. (2008) The nucleoprotein and the viral RNA of infectious salmon anemia virus (ISAV) are localized in the nucleolus of infected cells. *Virology* 379: 55-63. doi:10.1016/j.virol.2008.05.036. PubMed: 18632128.
62. Lai MM (1992) RNA recombination in animal and plant viruses. *Microbiol Rev* 56: 61-79. PubMed: 1579113.
63. García M, Crawford JM, Latimer JW, Rivera-Cruz E, Perdue ML (1996) Heterogeneity in the haemagglutinin gene and emergence of the highly pathogenic phenotype among recent H5N2 avian influenza viruses from Mexico. *J Gen Virol* 77: 1493-1504. doi: 10.1099/0022-1317-77-7-1493. PubMed: 8757992.
64. Arias A, Arnold JJ, Sierra M, Smidansky ED, Domingo E et al. (2008) Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. *J Virol* 82: 12346-12355. doi:10.1128/JVI.01297-08. PubMed: 18829745.
65. Coffey LL, Beeharry Y, Bordería AV, Blanc H, Vignuzzi M (2011) Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proc Natl Acad Sci U S A* 108: 16038-16043. doi:10.1073/pnas.1111650108. PubMed: 21896755.
66. Agudo R, Ferrer-Orta C, Arias A, de la Higuera I, Perales C et al. (2010) A multi-step process of viral adaptation to a mutagenic nucleoside analogue by modulation of transition types leads to extinction-escape. *PLOS Pathog* 6: e1001072.
67. Vignuzzi M, Wendt E, Andino R (2008) Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* 14: 154-161. doi: 10.1038/nm1726. PubMed: 18246077.
68. Andreoni M (2004) Viral phenotype and fitness. *New Microbiol* 27: 71-76. PubMed: 15646067.
69. Aamelfot M, Dale OB, Welis SC, Koppang EO, Falk K (2012) Expression of the infectious salmon anemia virus receptor on atlantic salmon endothelial cells correlates with the cell tropism of the virus. *J Virol* 86: 10571-10578. doi:10.1128/JVI.00047-12. PubMed: 22811536.
70. Deyde VM, Gubareva LV (2009) Influenza genome analysis using pyrosequencing method: current applications for a moving target. *Expert Rev Mol Diagn* 9: 493-509. doi:10.1586/erm.09.21. PubMed: 19580433.
71. Pinggen M, Nouwen JL, Dinant S, Albert J, Mild M et al. (2012) Therapy failure resulting from superinfection by a drug-resistant HIV variant. *Antivir Ther* 17: 1621-1625. doi:10.3851/IMP2267. PubMed: 22846173.
72. Ko SY, Oh HB, Park CW, Lee HC, Lee JE (2012) Analysis of hepatitis B virus drug-resistant mutant haplotypes by ultra-deep pyrosequencing. *Clin Microbiol Infect* 18: E404-E411. doi:10.1111/j.1469-0691.2012.03951.x. PubMed: 22757653.
73. Newman RM, Kuntzen T, Weiner B, Berical A, Charlebois P et al. (2013) Whole Genome Pyrosequencing of Rare Hepatitis C Virus Genotypes Enhances Subtype Classification and Identification of Naturally Occurring Drug Resistance Variants. *J Infect Dis* 208: 17-31. PubMed: 23136221.