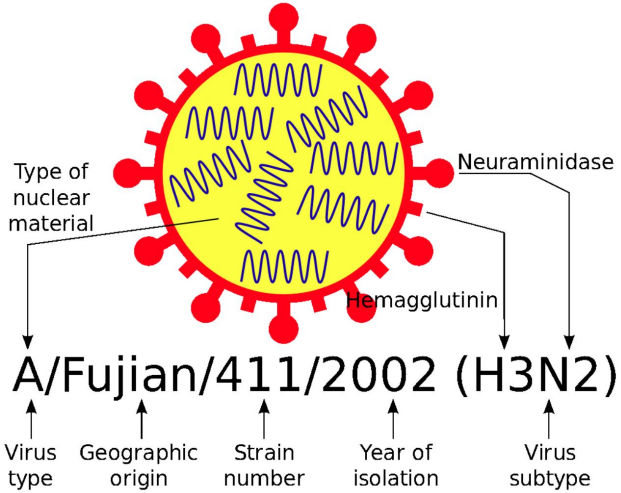
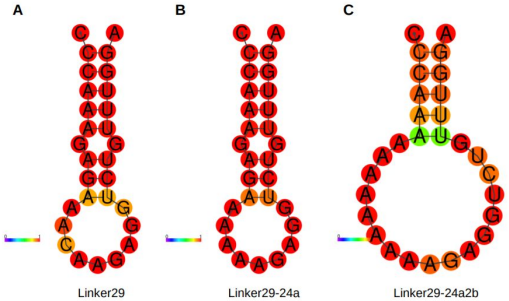
Structural predisposition to acquire a Polybasic Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin and Coronavirus.

RNA viruses exist as genetically diverse populations. It is thought that diversity and genetic structure of viral populations determine the rapid adaptation observed in RNA viruses and hence their pathogenesis. This study shows the potential role of the viral RNA secondary structure for nucleotide insertions and demonstrates a key mechanism explaining why the acquisition of the polybasic HA cleavage site is restricted to particular HA subtypes in nature. Aquatic birds, especially migratory ducks are reservoir of IAVs in nature.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5312086/#B26

INTRODUCTION

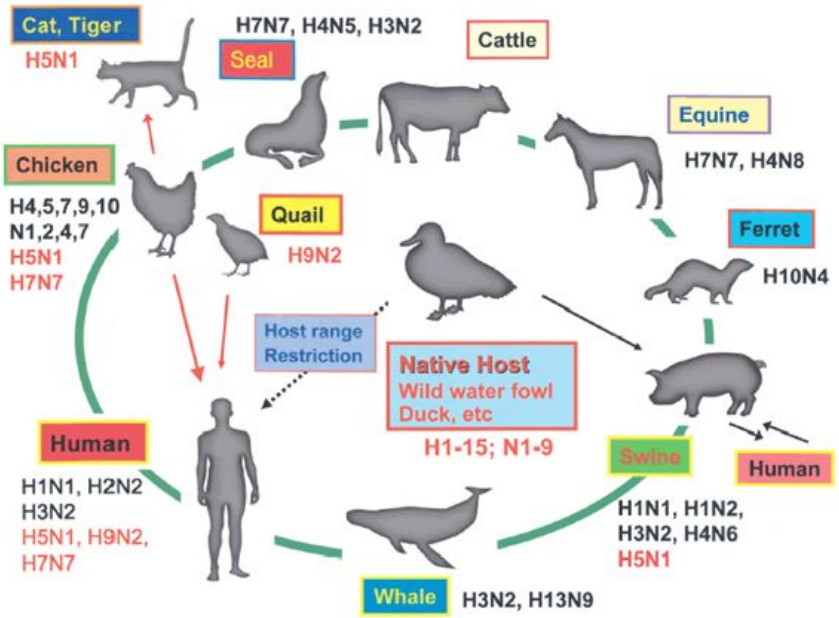
Influenza A viruses are divided into subtypes based on the antigenicity of the viral surface glycoproteins hemagglutinin (HA) and neuraminidase. Of the 16 HA subtypes (H1 to -16) maintained in waterfowl reservoirs of influenza A viruses, H5 and H7 viruses often become highly pathogenic through the acquisition of multiple basic amino acid residues at the HA cleavage site. Although this mechanism has been observed to occur naturally only in these HA subtypes, little is known about the genetic basis for the acquisition of the polybasic HA cleavage site.





Here we show that consecutive adenine residues and a stem-loop structure, which are frequently found in the viral RNA region encoding amino acids around the cleavage site of low-pathogenic H5 and H7 viruses isolated from waterfowl reservoirs, are important for nucleotide insertions into this RNA region.

Host transmission.

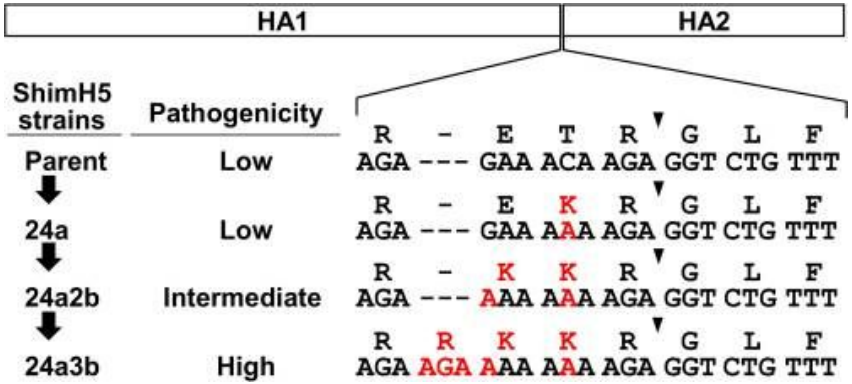


Upon transmission to other host animals, IAVs evolve rapidly due to the high error rate of the viral RNA (vRNA) polymerase and strong immune-driven natural selection. It is also well known that human pandemic viruses emerge through reassortment among RNA segments between avian and human viruses. Highly pathogenic avian influenza (HPAI) viruses with the H5 and H7 HA subtypes produce high mortality in poultry. In addition to infection of avian species, it has been reported that H5N1 HPAI viruses are occasionally transmitted to humans and cause severe pneumonia with high case fatality rates. It is known that HPAI viruses evolve from low-pathogenic H5 and H7 viruses. The key determinant for the different pathogenicities is the proteolytic cleavage of HA. Low-pathogenic IAVs contain a single arginine(R) residue at the cleavage site of HA, which is cleaved only by trypsin-like proteases and therefore produces localized infection of the respiratory and/or intestinal tracts, causing asymptomatic or mild infection. After introduction into domestic poultry, low-pathogenic viruses often acquire multiple basic amino acids at the HA cleavage site, which is recognized by ubiquitous cellular proteases such as furin and PC6. The polybasic HA cleavage site is known to be generated by multiple nucleotide insertions/substitutions to create codons for basic amino acids or by recombination with cellular or viral RNAs and is considered to be the primary virulence marker of HPAI viruses. Most outbreaks of HPAI are caused only by IAVs with the H5 and H7 subtypes. Other types of virus also acquired high pathogenicity for chickens via the introduction of a pair of dibasic amino acid residues at the HA cleavage site and subsequent passages in chickens.

Research

Whistling swan

In this study A/whistling swan/Shimane/499/83 (H5N3) (ShimH5) strain (GenBank: GU052809.1), which was originally isolated as a low-pathogenic strain and shown to become highly pathogenic after passaging through experimentally infected chickens. During serial passages through chickens, the ShimH5 strain first underwent two point mutations at nucleotide positions 1050 (C to A [ShimH5 24a]) and 1046 (G to A [ShimH5 24a2b]) in the RNA sequence encoding the HA cleavage site motif (R-E/K-T/K-R) and then acquired 5 consecutive basic amino acids, R-R-K-K-R, via the insertion of a codon for an arginine residue at the cleavage site (ShimH5 24a3b).

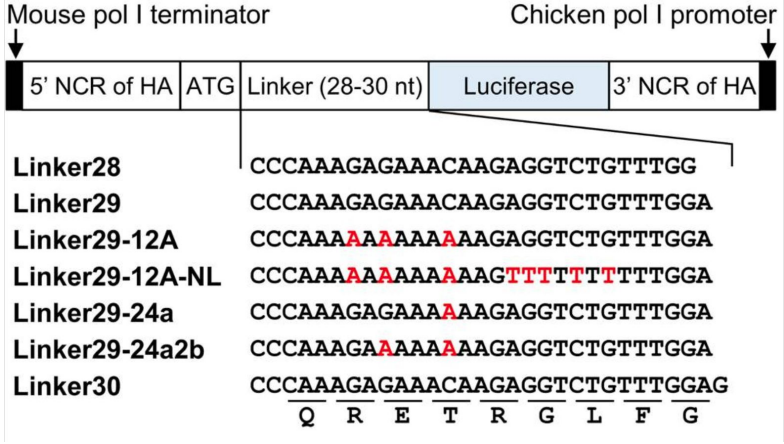


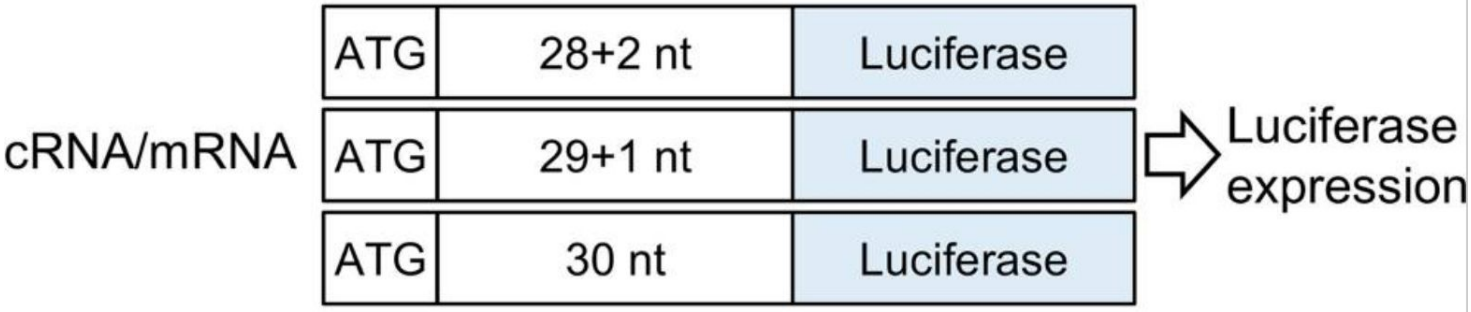
Interestingly, consecutive adenine residues and large stem-loop structures in this RNA region were frequently found only in particular HA subtypes (e.g., H5). The hypothesis about the direct link between the frequency of nucleotide insertions and predicted stem-loop structures of the viral RNA was tested and confirmed. RNA sequence determining the HA cleavage site amino acid motif has a key role in inducing viral polymerase slippage, which increases the frequency of nucleotide insertions, and that this mechanism contributes to the acquisition of additional codons for basic amino acids to create the polybasic HA cleavage site.

RESULTS

Reporter gene expression resulting from nontemplated nucleotide insertions into the ShimH5 HA sequence.

Reporter assay was established to detect nucleotide insertions into the RNA sequence encoding amino acids across the HA cleavage site (29 nucleotides, CCCAAAGAGAAACAAGAGGTCTGTTTGGA [designated RNAseqHAclv]) of the strain ShimH5. In the reporter plasmid, the firefly luciferase gene lacking its start codon was inserted downstream of the start codon/linker region (e.g., 29 polynucleotides corresponding to RNAseqHAclv [designated Linker29]). Since the firefly luciferase gene that followed the 28- or 29-nucleotide linkers was out of frame, the luciferase was expected to be expressed only when nucleotides were inserted into the linker region of mRNA, cRNA, and/or viral RNA (vRNA) to make the linker sequence in frame with the open reading frame (ORF) of the reporter gene.

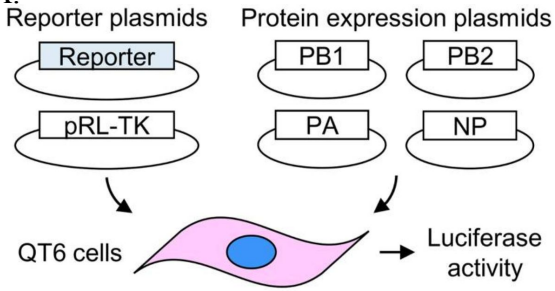


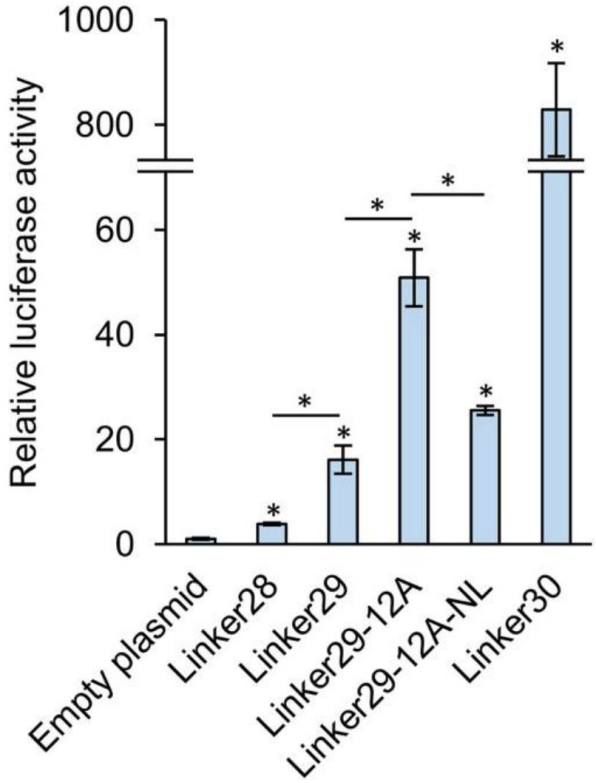


The nucleotide positions different from those of the parental ShimH5 sequence are indicated in red.

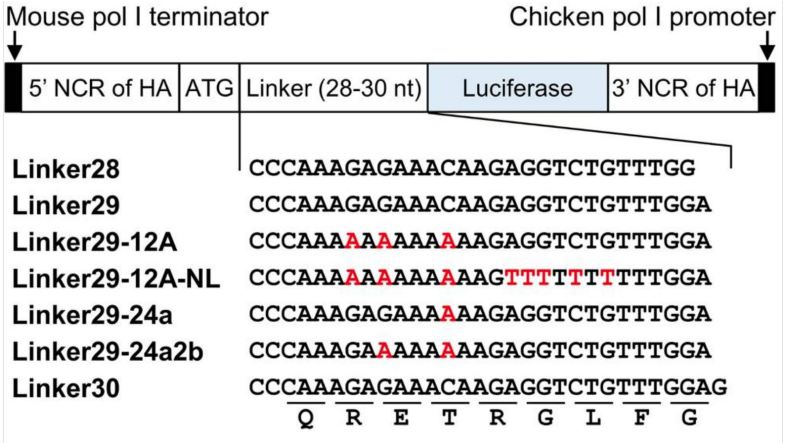
Evaluation

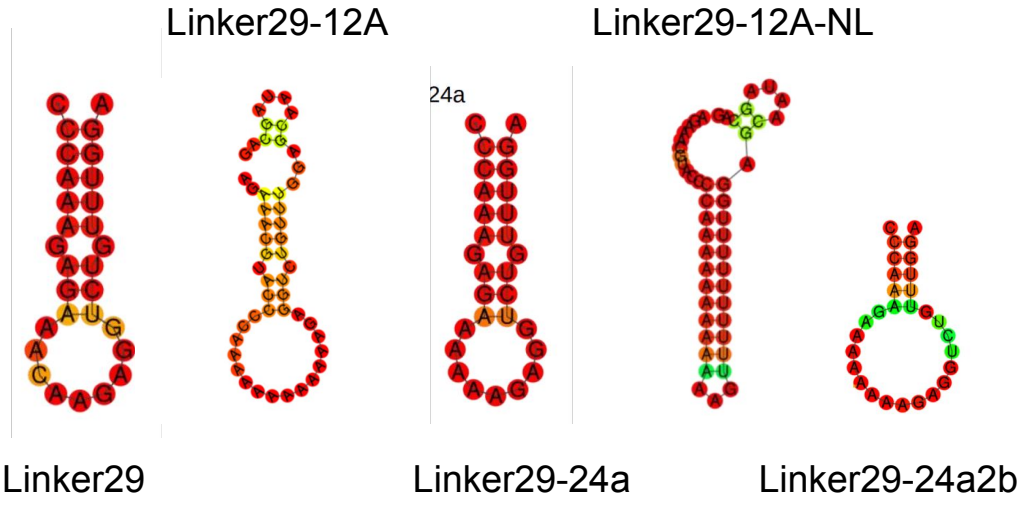
To evaluate this system, the plasmids containing the linkers (Linker28, Linker29, and Linker30) was tested, all of which had the sequence derived from the parent ShimH5 strain. QT6 cells were transfected with these plasmids, and luciferase activities of cell lysates were measured. As expected, a high level of luciferase activity was detected in the cells transfected with the construct with Linker30, which had the sequence in frame with the ORF of the reporter gene. Interestingly, we found that the plasmids containing Linker28 and Linker29 expressed slightly but significantly higher levels of luciferase than the empty plasmid. Significantly higher luciferase expression was observed in cells transfected with the plasmid containing Linker29 than with Linker28. These results suggested that nontemplated nucleotide insertions into these linker regions occurred during the synthesis of mRNA, cRNA, and/or vRNA.





Comparison of secondary structures of the linker sequences.





Interestingly, all linker RNAs potentially formed stem-loop structures mainly consisting of adenine and guanine residues. These stem-loop structures were maintained even when longer sequences (i.e., 39 and 49 nucleotides, but not 69 nucleotides or more) including the RNAseqHAclv region in the middle were used for the analysis. It was noted that the loop magnitude was correlated with the efficiency of luciferase expression. To confirm the importance of the stem-loop structure, an additional reporter plasmid containing Linker29-12A-NL was tested, which was artificially designed to have 12 consecutive adenines but to minimize the loop structure. As expected, significantly lower luciferase expression was observed in the cells transfected with this plasmid than in those transfected with the plasmid containing Linker29-12A

Stem-loops comparison.

Loop of encoding the amino acid around the HA cleavage sites of all analysed viruses. Statistics. Mean

Influenza A type	1st loop nucls qnt	A%	G%	U%	C%
H1	8.3	31	18	26	25
H2	6.8	40	45	1	14
H3	7	29	31	17	22
H4	8.6	38	35	9	18
H5	12.3	49	33	1	16
H6	7	28	37	19	16
H7	7.4	36	44	4	15
H8	7	32	30	23	14
H9	7.1	31	40	16	13
H10	7.7	27	51	12	10
H11	10.6	26	28	29	17
H12	8.2	36	29	8	27
H13	6.8	38	35	15	12
H14	10	45	38	0	18
H15	5.7	18	65	0	18
H16	8.4	31	30	31	8

In terms of structure similarity, more similar are H5 and H14.

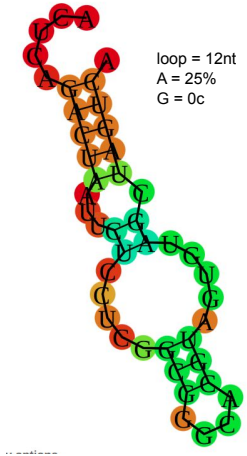
table generation:
https://github.com/cappelchi/Bioinformatics-for-Infectious-Diseases/blob/master/Pine_Bio_Infection_Dseases_presentation_part1.ipynb

RNAfold-predicted RNA structures of COVID19 sequence.

Polybasic cleavage site, sequence reference ID: NC_045512.2:

nucleotide: CAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATC

amino acids: QTQTNSPRRARSVASQ



similarity with H5 & H7 is not high.

Could COVID19 be more pathogenic?

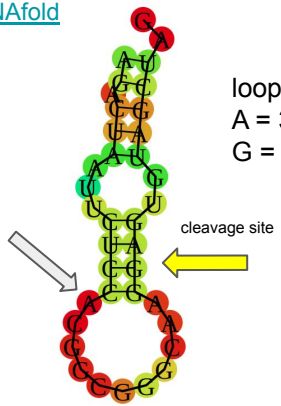
Let's look at some possible substitutions:

nucleotide modified: CAGACTCAGACTAATTCTCC**A**CGGCGGGCA**A**G**G**'GTGTAGCTAGTCAATC

amino acids modified: QTQTNS**P**RRA**R'**SVASQ (synonymous mutation)

in amino acids nothing change, what about structure?

RNAfold



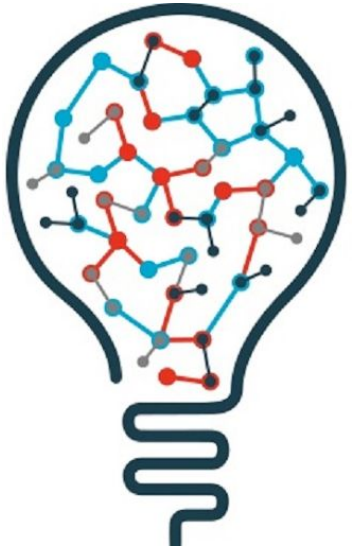
loop = 11 nt
A = 30%
G = 50%

Looks like predisposal for pathogenesis

High probability, that nontemplated nucleotide insertions will occur during the synthesis of mRNA, cRNA, and/or vRNA, that can lead to high pathogenicity.

colab notebook

Conclusion



1. Significance of virus RNA structure is underestimated.
2. The viruses that are identical in terms of amino acids can have different secondary structures (codon degeneracy and synonymous nucleotide substitutions).
3. Therefore, the same viruses can theoretically have different pathogenesis since the possible stem-loops structures are not part of the sequence.
4. Analysis of the RNA secondary structure can predict the possible evolution of pathogenicity and host transmission.

What's next.

The proteolytic cleavage or polybasic cleavage secondary structure can be studied programmatically (with python) to find possible high pathogenic mutations (insertions, deletions) for further experiments.

Mikhail Kosaretskiy

about me

Bioinformatician at Atlas Biomed

cappelchi@gmail.com

BIO.INFO

ROSALIND

- Passionate Data Scientist
- Live and work in Moscow, Russia