Journal of Applied Biosciences (2008), Vol. 10: 471 - 482.
ISSN 1997 – 5902: www.biosciences.elewa.org

Predicted role of small non-coding Ribonucleic acids (ncRNAs) and Ribonucleoproteins (RNPs) in Influenza genetic shifts and drifts

[redacted]

[redacted]

ABSTRACT
Objective: Influenza disease has been observed to be severe in animal hosts in associations with co-infection with H.*Influenza* species. This phenomenon is widely related to H.*Influenza* associated inflammation, but may as well be due to emergence of new virus strains. If the later is true, then H. *Influenza* may be releasing modulators of Influenza genomic variation aside of innate virus' error prone polymerase. Group II Retroposon elements (RTE) are a mobile class of small noncoding RNAs (snRNA) with both RNA and DNA catalytic potential found distributed across organella genomes, eubacteria and archeabacteria. To aimed to investigate if *H. Influenza* RTE derived snRNA may be exogenous mediators of Influenza genomic splicing
Methodology and Results: we employed Insilco palindromics to search for potential cleavage sites in 8 Influenza genomic RNA segments using 14 H. Influenza derived REases; AND a search for ORF encoding RT, X and En domains in 16 *H. influenza* species genome databases by BLASTP with query H.s.I1. Palindromes were found distributed in order of polymerase acidic protein gene PA 14 (100%), S2 13 (93%); HA;7 (50%), NA;7 (50%), PB1;8 (57%), PB2;7 (50%), M1&2;10 (71%), S1;8 (57%), S3;9 (64%), S4;7 (50%), S5;7 (50%), S6;7 (50%), S7;6 (43%), S8;6 (43%), NP;4 (29%), and the NS1&2;3 (21%). 15 low score blast hits were found, 7 of which were MTase- subunits of putative type I R-M systems, but no ortholog of the H.s.I1 ORF except in the query source genome H.*somnus*.
Conclusions: While the distribution of Palindromes supports the existence of HNHc/HHVR cleavage sites in Influenza RNA, the role of Group II RTE snRNA in drifts and shifts is limited by lack of conclusive homologs of RT ORF.

Key words: Group II Retroelements; Haemophilus Influenza species; Influenza antigenic drifts and reassortments; Restriction Endonuclease (REases).

Abbreviations: DNA-Deoxyribonucleic acid; RNA-Ribonucleic acids; NA-Neuraminidase; HA-Haemagglutinin; PA-Polymerase acidic protein; PB1&2-Polymerase basic proteins 1&2; M1,2-Matrix proteins 1,2; S1-8-genomic RNA segments 1-8; NP-Nucleoprotein; NS1&2-Non-structural proteins 1&2, RM –Restriction modification:

This paper has supplementary files that can be accessed on a separate link provided next to the paper title at the journal website.

[redacted]

INTRODUCTION

Influenza viruses belong to the family Orthomyxoviridae and are causes of significant morbidity in the general population and immunocompromised patients (Vilchez *et al*., 2002; Treanor, 2004). Influenza is classified into three distinct subtypes based on antigenic differences: influenza A, influenza B and influenza C (Nicholson, 1998; Treanor, 2004). Structurally, Influenza A viruses are enveloped, single-stranded negative sense Ribonucleic acids (RNA) viruses with a segmented genome consisting of eight gene segments. Two of the most important gene products are the surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). HA is produced by RNA segment 4 and is responsible for host-cell membrane attachment and membrane fusion. The NA is produced by RNA segment 5 and cleaves sialic acid from the cell surface, thereby allowing for cleavage of viral progeny from infected cell surfaces. There are at least 15 antigenically distinct HA types that have been described in influenza A (H1–H15) and at least 9 NA (N1–N9) types (Nicholson, 1998).

Influenza viruses undergo antigenic changes at a high frequency. The variability in antigens generally involves changes in the external glycoproteins HA and NA. Minor variability is referred to as antigenic drift, while larger changes are called shifts (CDC, 2007; WHO, 2007). This 'antigenic shift' is responsible for pandemic influenza and render vaccine efficacy obsolete in a year or so (Belshe, 2005; Russell & Webster, 2005). Until now, no single exogenous ecological factor has been described as a likely facilitator of these antigenic drifts and shifts, with much of the occurrences being attributed to the fact that viruses replicating by RNA polymerase generally exhibit higher mutations than their DNA polymerase counterparts. Generally, many viruses with RNA genomes have genetically diverse populations called quasispecies. The representation of any particular sequence within this quasispecies is a result of interactions between the host and environmental factors affecting the replication of the virus. Important biological properties are a direct result of the levels of diversity in the quasispecies 'cloud size', including adaptability and host range. RNA viruses have become the model system for the analysis of viral evolution due to the inherent error-prone nature of their genome-replicating enzymes that lack a proof-reading function (Gould, 2004). This view is debatable, given that closely contested mutations have been demonstrated in small DNA viruses (Laura *et al*., 2005), a finding that emphasizes the need to investigate if there are other exogenous biotic factors contributing to this high level of reassortment (Horimoto & Kawaoka, 2005; Tumpey *et al*., 2005; Monto 2005; Taubenberger *et al*., 2005; Kaye & Pringle, 2005).

Two groups of intra-genomic and in-viscera derived RNA-protein complexes (Ribonucleoproteins-RNP) are hypothetically likely to splice viral RNA. These are (1) the mobile class of Group II introns found to be interspersed within organelle genomes, Eubacteria, archeabacteria plus eukaryotes, and (2) Ribonucleases (RNases) that are part of various taxa digestive (and other tissue)secretions. The major emphasis of this study was placed on the small non coding RNAs of the highly mobile class of group II Introns. While a section of these systems are Apurinic apyraminic Endonucleases (APE-) like, majority are Restriction endonuclease (REase)-like, coding a REase motif within the Reverse transcriptase (RT) moiety of the REase-like retroposons. These have a HNHc and HHVR specificity amino acid motif within the C-terminal Zn^{2+} REase whose function is consistent with several bacterial derived Deoxyribonucleases (DNases) by way of recognizing palindromic sequences and cleaving within or near them. REase-like ncRNAs have the same Deoxyribonuclease (DNase) ability, and most bacteria group II introns are actually a complexion of Maturases (X), Reverse trancripatse (RT) and Dnases (REases) (Michel & Ferat, 1995; Zimmerly *et al*., 1995[a]). They function by recognizing a 4-8 bp sequence within target DNA/RNA and cleave within, or close to it. Despite the absence of an Insilco model for assessing the

RNomics of small non-coding RNAs (ncRNA) on RNA or DNA splicing, a model based on searching for palindromic sequences recognizable by REases may suffice (; Nelson *et al*., 1972; Kessler & Manta, 1990; Roberts & Macelis, 1991; Janulaitis *et al*., 1992; Radasci & Bickle, 1991; Barcus and Murray,1995; Michel & Ferat, 1995; Zimmerly *et al*., 1995[a]; Murray, 2000). Such a model may be a vital source of information regarding the likely effect of REase ncRNAs on RNA genomes.

This study aimed to use the REase palindrome recognition model to search for catalytic RNA sites of the Group II RTE within Influenza genomes. In addition, a search was carried out for ORF of Group II intron type within 16 related H. Influenza genomes using the Reverse transcriptase domain of Haemophilus somnus Group II RTE H.s.I1 which codes 2 other peptides: Maturase(X), REase(En).

METHOD AND MATERIALS

Method I: *Design:* Insilico Palindromics

Materials and selection: 16/18 Influenza genes and genomic RNA segments [6/7 core genes of Influenza virus: HA, NA, PA, PB1, PB2; 8 genomic RNA segments S1-8; and 2/3 others: M1/2, NP, NS1/NS2,] derived from the NCBI Influenza Virus Resource Database (Suppl file A1-3) (CDC *et al*., 2007b), Webcutter version2 [http://rna.lundberg.gu.se/cgi-bin/cutter2/cutter], 14 Haemophilus *spp* derived restriction enzymes, and 1 control from the Chloroflex bacteria *Herpesiphon aphrophilus* (Table 1).

Interventions: Complete Influenza gene and genomic RNA segment-sequences were fed into Webcutter version 2 present with the above 14 *Haemophilus* spp and H. *aphrophilus* derived restriction enzymes to analyze for site cleavage and mapping of linear nucleotide sequences by recognizing 6 or more base pair palindromes.

Method II: *Design*: Insilco comparative genomics.

Materials: Haemophilus *somnus* retron RT protein ID ZP_00131874.1, H. *somnus* complete genome and 15 other Haemophilus *influenza* genomes (6 of which were complete (Table 5), NCBI Genomic BLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi).

Interventions: The Haemophilus *somnus* retron RT protein ID ZP_00131874.1 was used to search for ORF of Group II retroelements within the 16 genomes.

.

RESULTS

Distribution of H. *influenza* REase palindromes within influenza genomic RNA: All 14 H. *influenza* REases (100%) had cleavage palindromes in the polymerase acidic protein gene (PA), with the genomic RNA segment S2 being second populated; 13 (93%). Other genes/genomic RNA segments were palindrome populated in the ascending order of: NS1&2; 3 (21%), NP;4 (29%), S7;6 (43%), S8;6 (43%), HA;7 (50%), NA;7 (50%), PB2;7 (50%), S4;7 (50%), S5;7 (50%), S6;7 (50%), PB1;8 (57%),S1;8 (57%), S3;9 (64%), and the M1&2;10 (71%), respectively. A correlation was noted between results with S4 and S5 with those of HA and NA, as S4 and S5 encode HA and NA, respectively. The Haemophilus *influenza* derived HinfI demonstrated the highest recognizable palindromes in all influenza genes except segment 6, while H. *aegypticus* derived HaeIII had cleavage sites present in all but 5 genes/genomic RNA segments. A *Herpesiphon aphrophilus* derived HgiEI DNase used as a control had at least one cutting cleavage palindromic site in all but segment 7 of genomic RNA (Table 2, 3 and 4). For details of influenza gene/genomic RNA segment used (see suppl File A1-3) and cleavage Sites (see Supp Files B). Site mapping of palindromic sequences is demonstrated in Supplementary files C1; C2 and C3.

Table 1: Recognition sites of the 14 Haemophilus spp derived Restriction enzymes, and the *Herpetosiphon auranticus* derived HgiEI.

Restriction enzyme	Recognition site	Source
HaeIII	5'...GG▼CC...3' 3'...CC▲GG...5'	*Haemophilus aegypticus*
HapII	5'-C^C G G-3' 3'-G G C^C-5'	*Haemophilus aphrophilus*
HgiEI	G^GWCC CCWG^G	*Herpetosiphon aurantiacus* *(Bacteria; Chloroflexi)*
HhaI	5'...GCG▼C...3' 3'...C▲GCG...5'	*Haemophilus haemolyticus*
Hin6I	Asa for HhaI, but Unlike HhaI, Hin6I produces DNA fragments with a 2-base 5'-extension	*Haemophilus influenzae* RFL6
HinP1I.	5'...G▼CGC...3' 3'...CGC▲G...5'	*Haemophilus influenzae* P_1
HindIII.	5'...A▼AGCTT...3' 3'...TTCGA▲A...5'	*Haemophilus influenzae* Rd
HinfI	5'...G▼ANTC...3' 3'...CTNA▲G...5'	*Haemophilus influenzae* Rf
HpaII	5'...C▼CGG...3' 3'...GGC▲C...5'	*Haemophilus parainfluenzae*
HaeII,	5'...RGCGC▼Y...3' 3'...Y▲CGCGR...5'	*Haemophilus aegypticus*
HgaI,	5'...GACGC$(N)_5$▼...3' 3'...CTGCG$(N)_{10}$▲...5'	*Haemophilus gallinarum*
Hin1I,	5'-G Pu^C G Py C-3' 3'-C Py G C^Pu G-5'	*Haemophilus influenza* RFL 1
HincII,	5'...GTY▼RAC...3' 3'...CAR▲YTG...5'	*Haemophilus influenzae* Rc
HindII	As for HincII	*Haemophilus influenzae* Rd
HpaI	5'...GTT▼AAC...3' 3'...CAA▲TTG...5'	*Haemophilus parainfluenzae*

Distribution of group II intron ORF in Haemophilus *influenza* genomes: Of all 16 genomes searched (genomic protein databases of 44,923 sequences; 12,652,512 total letters), only 16 hits to database(DB) were obtained, of which, the only orthologous hit found was within the H.*somnus* 2336 Cluster of Orthologous Group 3344; the Retron-type reverse transcriptase H.s.I1 Open reading flame of Length = 575 used in the search. Other low score Blast hits were found within other Haemohipilus *species* protein genomic databases corresponding to magnesium/nickel/cobalt transporter CorA; putative type I restriction-modification; COG0286: Type I restriction-modification; lic-1 operon protein and phosphorylcholine transferase. Although the low alignment [29-41] scores relative for a query of 575 (<10%) pre ludes these to labeled homologs (removing the need for a further evolutionary classification into either orthologs OR parolog); the 7 hits obtained with methyltranferases (MTases) offer evidence for a likely paralogous evolutionary relationship to putative type I restriction modification systems. Further details of blast hits and full results of BLASTP searches done with H.s.I1 are presented (Table 6; Figure 1; Suppl file D).

Table 2: Distribution of palindrome sequences (thus cleavage sites) recognizable by Haemophilus derived restriction enzymes in HA, NA, PA, PB1, PB2 and M1&2 influenza genes.

Influenza virus gene/Genomic RNA segment	Freq of cuts; enzymes	Cumulative enzymes (%)	Enzymes cutting 3 or > times
Haemagglutinin gene, HA	1;4 2;1 4;1 5;2	7(50%)	4[HinfI], 5[HapII, HpaII]
Neuraminidase gene, NA	1;5 3;2	7(50%)	3[HaeIII, HinfI]
Polymerase acidic protein, PA	5;3 6;3 7;1 9;3 13,2 16;1 21;1	14(100%)	All 14 haemophilus spp restriction enzymes used
Polymerase basic protein B1	1;6 2;2 4;1 16;1	8(57%)	4 [HaeIII], 16[HinfI]
Polymerase basic protein B2	1;1 3;2 4;2 6;1 12;1	7(50%)	3[HpaII, Hap1I].,4[HindII HincII].,6[HaeIII].,12[HinfI]
Matrix Protein M1 and M2	1;4 2;4 4;1 5;1	10(71%)	4[HaeIII].,5[HinfI]

DISCUSSION

The finding from the results of the first methodology demonstrate the high prevalence of palindrome sequences within the Influenza genomic RNA that may be targeted by REase-like small ncRNAs, a mobile class of retroposons. While the activity of DNases (such as H. *influenza* derived ones used here) is mostly limited to DNA, a section of Intron derived small ncRNA have the potential under various complexions (Zimmerly *et al*., 1995[a]) to splice and ligate both RNA and DNA. They belong to a class of Group II introns, which are both catalytic RNAs and mobile genetic elements (Matsuura *et al*., 1997). The Group II RNAs can self splice, and can also carry out related transesterification reactions inclusive of reverse splicing, RNA and DNA ligation, and DNA cleavage (Mohr *et al*., 1993; Michel & Ferat, 1995; Zimmerly *et al*., 1995[a] ; Matsuura *et al*., 1997). These mobile GpII introns encode Reverse transcriptase open reading frames (ORF) moiety, which have 7 regions (structurally), with the most specialized being the Maturase(X), the Z thumb domain, and the C-terminal Zn^{2+} finger like region ZN; which contains amino acid sequences characteristic with class II Restriction endonucleases (REases)(Gorbalenya 1994; Shub *et al*., 1994). The relationship between the Group II intron Reverse transcriptase moiety and bacteria restriction endonulcease is demonstrated here by our findings from the results obtained with the second method, which show that despite the absence of orthologs to *H.somnus* intron H.s.I1 reverse transcriptase in the other 15 influenza genomes searched, several (7) hits were *H. influenza* derived methyltransferase subunit of putative type I restriction-modification system.

Table 3: Distribution of palindrome sequences (thus cleavage sites) recognizable by Haemophilus derived restriction enzymes in Influenza genomic RNA segments 1-8.

Influenza virus gene/genomic RNA segment	Freq of cuts; enzymes	Cumulative enzymes (%)	Enzymes cutting 3 or > times x[enzyme]
Segment 1	1;2 2;4 7;1 10;1	8(57%)	7[HaeIII].,10[HinfI]
Segment 2	1;2 2;6 3;1 5;1 7;2 14;1	13(93%)	3[HindIII].,5[HaeIII].,7[HapII ,HpaII].,14[HinfI]
Segment 3	1;4 2;3 8;1 13;1	9(64%)	8[HaeIII].,13[HinfI]
Segment 4	1;1 2;4 3;1 10;1	7(50%)	3[HaeIII].,10[HinfI]
Segment 5	1;2 2;3 3;1 5;1	7(50%)	3[HgaI].,5[HinfI]
Segment 6	1;5 2;1 3;1	7(50%)	3[HaeIII]
Segment 7	1;3 2;1 6;1 9;1	6(43%)	6[HinfI].,9[HaeIII]
Segment 8	1;4 2;1 5;1	6(43%)	5[HinfI]

Legend: Note that segments 4 and 5 code for HA and NA respectively

However, much of this relationship has already been previously documented in terms of functional domain, evolutionary origin and structural similarities (Shope, 1931a; Shope, 1931b; Andrewes *et al*., 1934; Burnet, 1935; Shope, 1937; Morl & Schmelzer, 1990; Gorbalenya 1994; Dalgaard *et al*., 1997; Zimmerly *et al*., 2001; Ostheimer *et al*., 2003). For instance, the domain HNHc (SMART ID: SM00507, SCOP nomenclature: HNH family) is associated with a range of DNA-binding proteins, performing a variety of binding and cutting functions (Gorbalenya, 1994; Shub *et al*., 1994). Several of the proteins are hypothetical or putative proteins of no well-defined function. The ones with known function are involved in a range of cellular processes including bacterial toxicity, homing functions in groups I and II introns and inteins, recombination, developmentally controlled DNA rearrangement, phage packaging, and restriction endonuclease activity(Zimmerly *et al*., 2001; Dalgaard *et al.,* 1997).

Table 4: Results of *Haemophilus* derived restriction enzyme activity on Nucleoprotein gene NA and Non-structural genes NS1 &NS2.

Influenza virus gene/genomic RNA segment	Freq of cuts; enzymes	Cumulative enzymes (%)	Enzymes cutting 3 or > times x[enzyme]
Nucleoprotein gene NP	2;2 3;1 8;1	4(29%)	8[HinfI]
Non structural protein 1 & 2 genes NS1, NS2	1;1 2;1 7;1	3(21%)	7[HinfI]

Table 5: List of 16 *Haemophilus* species genomes in the Database searched.

Completed *Haemophilus* ducreyi 35000HP proteins;
Unfinished *Haemophilus* influenzae 22.1-21 proteins;
Unfinished *Haemophilus* influenzae 22.4-21 proteins;
Unfinished *Haemophilus* influenzae 3655 proteins;
Completed *Haemophilus* influenzae 86-028NP proteins;
Unfinished *Haemophilus* influenzae PittAA proteins;
Completed *Haemophilus* influenzae PittEE proteins;
Completed *Haemophilus* influenzae PittGG proteins;
Unfinished *Haemophilus* influenzae PittHH proteins;
Unfinished *Haemophilus* influenzae PittII proteins;
Unfinished *Haemophilus* influenzae R2846 proteins;
Unfinished *Haemophilus* influenzae R2866 proteins;
Unfinished *Haemophilus* influenzae R3021 proteins;
Completed *Haemophilus* influenzae Rd KW20 proteins;
Completed *Haemophilus* somnus 129PT proteins;
Unfinished *Haemophilus* somnus 2336 proteins

Table 6: 16 Sequences producing significant alignments to the *H.somnus* RT Haso02000162 IDZP_00131874.1.

Sequences producing significant alignments:	Score Bits	EValue
ref|ZP_00131874.1| COG3344: Retron-type reverse transcriptase	1192	0.0
ref|ZP_01789906.1| magnesium/nickel/cobalt transporter CorA	40.4	0.003
ref|ZP_01788141.1| magnesium/nickel/cobalt transporter CorA	40.4	0.003
ref|YP_001291999.1| putative type I restriction-modification	33.5	0.31
ref|ZP_01787290.1| putative type I restriction-modification	33.5	0.31
ref|YP_247828.1| putative type I restriction-modification	33.5	0.31
ref|ZP_00157698.1| COG0286: Type I restriction-modification	33.5	0.31
ref|ZP_01789412.1| putative type I restriction-modification	32.3	0.69
ref|ZP_00154832.1| COG0286: Type I restriction-modification	32.3	0.69
ref|ZP_01797591.1| putative type I restriction-modification	32.0	0.90
ref|ZP_01788663.1| phosphorylcholine transferase	30.8	2.0
ref|NP_439689.1| lic-1 operon protein	30.8	2.0
ref|ZP_01785506.1| phosphorylcholine transferase ref|YP_249057.1|	29.6	4.5
phosphorylcholine transferase ref|YP_001291526.1| lic-1 operon protein	29.6	4.5
ref|ZP_01790600.1| lic-1 operon protein	29.3	5.8
	29.3	5.8

These proteins are found in viruses, archaebacteria, eubacteria, and eukaryotes. Interestingly, as with the LAGLI-DADG and the GIY-YIG motifs, the HNHc motif is often associated with endonuclease domains of self-propagating elements like inteins, Group I, and Group II introns(Gorbalenya, 1994; Dalgaard *et al.,* 1997). In at least two of the earlier studies on the HNHc group of proteins (with respect to Zn domains in group II introns), it has been suggested that the HNHc domain has a bacterial origin, being the site specificity determinant in REase activity (Zimmerly *et al*., 2001). Studies with both bacteria (*Lactococcus lactis*)(Zimmerly *et al*., 2001) and Eukaryotic (yeast mitochondrial) introns have found that introns with HNHc like motif have Restriction endonuclease activity in addition to their Maturase and Reverse transcriptase activity. Areas recognized by the HNHc motif such as the HHVR motif of bacteria REases and the CRS2 ORF recognized by plant derived mobile Group II Introns, have conserved regions of palindrome sequences that are recognizable by bacteria derived REases. For instance, the maize chloroplasts Group II intron CRS2 and CRS2 PCR primers CRS2D (-GCGGAATTCATGGAATACACGCCC-) and CRS2L (-GGA GGTCGACTTCAAACCCTG-) (Ostheimer *et al* 2003); and Yeast mitochondrial al1 and al2 EBS1-IBS1 Intron-exonic junctions sequence (-TTAATAATTTTCT-) (Morl & Schmelzer, 1990) can be noticed to contain similar palindrome sequences as those recognizable by the H. i*nfluenza* derived REases used here (Tables 1 – 4; supplementary files B and C1-3). For instance, the AATT and TTAA palindromes are targets for the H. *parainfluenza* derived DNases HpaI. Mutant retroposon RNP with partial or complete aberrations in the nt base content of this region have been found to have diminished or completely abolished reverse splicing abilities. A*t this point it may be hypothesized that Mobile Group II introns snRNA (orRNP complexes) may as well be an intra-genomic source of genetic diversity within RNA viruses (such as influenza, apart from or in addition to, the error prone Polymerase) and small DNA viruses* (Laura *et al*., 2005) *by virtue of their ability to splice and ligate both RNA and DNA.* While the lone RNA component has been observed in-vitro to have its splicing ability limited to only RNA in the appropriate alkali (Mg^{2+}) (Zimmerly *et al*., 1995[a]), the combination with a protein (RNP) seems to confer on the RNA the DNA catalytic potential (Zimmerly *et al*., 1995[a]; Matsuura *et al*., 1997). The first support for the hypothesis comes from our findings of a high prevalence of palindrome sequences recognizable to HNHc and HHVR REase motifs consistent with both REases and REases-like retroposons (Gorbalenya, 1994; Shub *et al*., 1994; Dalgaard *et al*., 1997; Zimmerly *et al*., 2001). Secondly, we also show that the innate error reading ability of the polymerase may in itself be primarily due to constant sequence diversity within its Polymerase Acidic protein A subunit (PA), which has the highest prevalence of palindrome sequences recognizable by sncRNAs of the Group II mobile introns (as demonstrated by the REase model). Thirdly, several epidemiological studies have revealed that Influenza pandemics in swine and other veterinary reservoirs are made worse by co-infections with *H. influenza* strains. Here, we show that the HapI REase recognition palindrome (-TTAA-) forms a major section of the HHVR motif consistent with Group II mobile introns, which leads us to question whether *H.inflenza* derived ncRNAs/Group II introns are not a major source of genetic diversity in Influenza virus genomes; in which case the HapI DNase may be considered the DNase component of a RNP complex derived from *H.parainfluenza*, and the notable severity of clinical disease due to the emergence of a new strain of virus to which the hosts have no pre-programmed immunity (naivety) rather than, or in addition to the currently held theory " that focuses on *H. influenza* inflammatory changes complicating Influenza viral disease (insert reference for this theory)." The later has been widely studied from the beginning of Influenza pandemics, with the isolation of *H.influenza* strains from nasal secretions of many of the patients during the great 1890 Influenza pandemic leading Pfeiffer (Andrewes *et al*., 1934; Burnet, 1935; Shope, 1931a; Shope, 1931b; Shope, 1937;) to wrongly believe that the isolated bacteria Haemophilus caused influenza. This was later to be disproved as subsequent studies failed to prove all three Koch's postulates of disease causation (Pittman, 1931; Smith *et al*., 1933; Andrewes *et al*., 1934; Burnet, 1935; Stuart- Harris, 1936; Shope, 1937; Chandler *et al*., 1939; Mote & Fothergill, 1940), and the current evidenced theory. We add our current hypothesis of infection with a new viral strain to which the host had no prior formed immunity, and that *H. influenza* associated inflammatory changes make the flue worse.

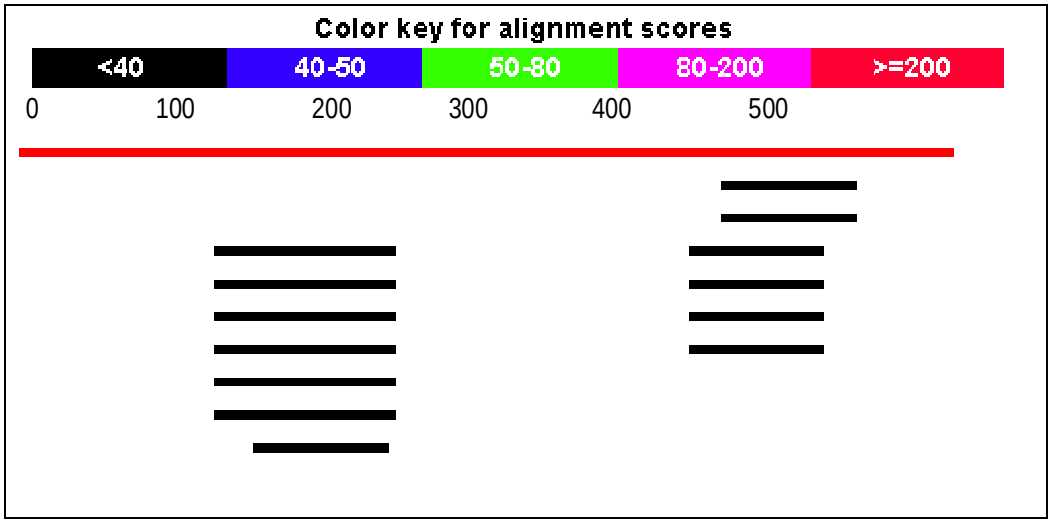
Lastly, despite the failure to demonstrate homologs of H.s.I1 in all other 15 Haemophilus spp genomes searched, the evolutionary relationship of Group II retroelement ORF to REases is demonstrated

by the low score hits with putative type 1 REase derived from *H. influenza*. Further still, previous studies that compared group II intron RNA structures with the predicted phylogenetic relationships of the ORFs encoded within them observed a primary pattern of coevolution, proposing the 'retroelement ancestor hypothesis' for the evolution of Group II introns (Toor *et al*., 2001). This hypothesis predicts that the ancestral group II intron for the data set was a bacterial group II intron RNA structure containing "nonstandard" or hybrid structural features and encoding a compact reverse transcriptase ORF (Zimmerly *et al*., 2001). The "standard"A and B structural forms of group II introns are predicted to have originated subsequently by coevolution with ORFs in the mitochondrial and chloroplast-like lineages+ ORF-less introns, which are mainly A and B forms, are predicted to be the result of ORF loss from mobile introns of the mitochondrial and chloroplast-like lineages(Toor *et al*., 2001). Perhaps the strongest evidence for coevolution came from the knowledge of the biochemical interactions between the intron RNA and RT protein. The L.l. RT (ltrA) binds very tightly to its intron (Kd 5 0+25 pM), with a primary binding site in intron domain IV and additional contacts with other intron domains (Wank *et al*., 1999). Moreover, both intron and RT subunits are required for each reaction of the RNP particle, including forward splicing, reverse splicing into DNA, DNA cleavage, and template specific reverse transcription [Lambowitz & Belfort, 1993; Zimmerly *et al*., 1995b; Matsuura *et al*., 1997; Zimmerly *et al*., 1999; Zimmerly *et al*., 2001). This high degree of biochemical cooperation between intron and RT would present a barrier to the reshuffling of introns and ORFs while retaining full splicing and mobility functions. The close cooperation contrasts with group I introns, for which the ORF's mobility activity (DNA nuclease) is biochemically independent of the intron's self-splicing activity(Zimmerly *et al*., 2001) and this functional distinction may provide the rationale for explaining why group II intron RNAs and ORFs predominantly coevolved whereas group I intron RNAs and ORFs did not.

This study has shown that several H.*influenza*, and other species (Herpesiphon *aphrophilus*) derived REase- site specific palindromes similar to those recognized by REase-like Group II mobile elements are present within Influenza genes. This finding has been used to argue the case that a number of related smaller Ribonucleproteins (REase-like retroposons and may be even RNA-DNases complexes from commensals such as Haemophilus spp) are likely exogenous biotic derivatives that influence the rate of antigenic drifts and shifts of the influenza virus, aside from the error prone-nature inherent within the large Influenza RNA polymerase. While the distribution of Palindromes supports the existence of HNHc/HHVR cleavage sites in Influenza RNA, evidence to support the likely role of Group II RTE snRNA in Influenza genomic drifts and shifts is limited by a lack of conclusive homologs of Reverse transcriptase (RT) ORFs in the 16 Influenza genomes studies.

ACKNOWLEDGEMENTS: Author is grateful to Prof. Wilson B., Head of the Division of Human Genetic & Genomics, Makerere University, and Mr. Henry K., of the Department of Microbiology for their rigorous reviews and revisions of this manuscript.

Figure 1 (below): Graphic distribution of 15 blast hits on the query sequence. (Legend: The Query H.s.I1 RT ORF was used to search for ORF within 15 Influenza Genomewide protein databases(inclusive of that of H.somunus 2336 from whivh the H.s.I.1 RT is obtained). This figure shows the distribution of the hits to DB. Note the only homolog (>80% identity in red). The lower hits are shown by the blue and black bars. The figure was generated by the NCBI Genomic BLAST software http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi)



REFERENCES

Andrewes CH, Liadlaw PP, Smith W, 1934. The susceptibility of mice to viruses of human and swine influenza. Lancet 2: 859-862.

Barcus VW. and Murray N, 1995. Barriers to recombination. In, 'Population genetics of bacteria' (eds. Baumberg S, Young J, Wellington E. and Saunders J.) pp 31-38. Cambridge University Press, Cambridge.

Belshe RB, 2005. The origins of pandemic influenza—Lessons from the 1918 virus. N. Engl J Med. 353: 2209–2211.

Burnet FM, 1935 Influenza virus isolation from an Australian epidemic. Med J Australia 2: 651-653.

CDC, 2007a. Key facts about avian Influenza and Human influenza virus A. CDC Facts sheets http://www.cdc.gov/flu/avian/gen-info/facts.htm Accessed 21st AUG 2007.

CDC, NIAIDS & NCBI. 2007b. Influenza virus ressource. http: //www.ncbi.nlm.nih.gov/genomes/FLU/Database/request.cgi. Accessed 21st AUG 2007.

Chandler CA, FothurGill L D, Dingle GE, 1939. The pattern of dissociation in Haemophilus *nfluenza* J Bact. 37: 415-427.

Dalgaard JZ, Moser MJ, Klar AJ, Holley WR, Chatterjee A, Mian IS, 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res. 25: 4626–4638.

Gorbalenya AE, 1994. Self-splicing group I and Group II introns encode homologous putative DNA endonucleases of a new family. Protein Sci. 3: 1117–1120.

Gould AR, 2004. Virus evolution: disease emergence and spread *Australian Journal of Experimental Agriculture.* 44(11):1085–1094.

Horimoto T. and Kawaoka Y, 2005. Influenza: Lessons from past pandemics, warnings from current incidents. Nat Rev Microbiol. 3: 591–600.

Janulaitis PM, Maneliene Z, Klimasauskas S, Butkus V, 1992. Purification and properties of Eco 57I restriction endonuclease and methylase, a prototype of a new class (type IV). Nucleic Acids Res. 20: 6043-6049.

Kaye D. and Pringle CR, 2005. Avian influenza viruses and their implication for human health. Clin Infect Dis. 40: 108–112.

Kessler C. and Manta Y, 1990. Specificity of restriction endonucleases and DNA modification methyltransferases, a review gene. Gene. 92:1-248.

Lambowitz AM. and Belfort M, 1993. Introns as mobile genetic elements. Annu Rev Biochem 62: 587–622.

Laura AS, Colin RP, Uwe T, Edward CH, 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus P. Proc. Natl. Acad. Sci. USA . 102(2): 379-384

Matsuura M, 1997. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. Genes Dev. 11(21):2910-24.

Michel F. and Ferat JL, 1995. Structure and activities of group II introns. Annu. Rev. Biochem. 64: 435-461.

Mohr G, Perlman PS, Lambowitz AM, 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. Nucleic Acids Res. 21: 4991-4997.

Monto AS, 2005.The threat of an avian influenza pandemic. N Engl J Med. 352: 323–325.

Morl M. and Schmelzer C, 1990. Integration of group II intron bl1 into a foreign RNA by reversal of the self-splicing reaction in vitro. Cell 60: 629-636.

Mote JR. and Fothergill LD, 1940. The Effect of Human Strains of *Hemophilus influenzae* on Influenza Virus Infections of Swine. J Bacteriol. 40 (4): 505–516.

Murray N, 2000. Type 1 Restriction systems. Sophisticated Molecular Machines (a legacy of Bertani and Weigle). Microbial Mol.Biol. Rev. 64: 412-434.

Nelson M. and McClelland M, XXX. Site specific Methylation effect on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res. 19:2045-2071.

Nelson M, Yuan R, Heywood J, 1972. Bacterial Restriction Modification systems. Annual Review of Biochem. 41: 447.

Nicholson KG, 1998. Human influenza. In: *Textbook of Influenza*(eds. Nicholson KG, Webster RG, Hay AJ) pp219–264 (Blackwell Science Oxford, England)

No author Listed. 2007. Webcutter Version two. Accessed 22 May 2007. http://rna.lundberg.gu.se/cgi-bin/cutter2/cutter Accessed 22 May 2007

Ostheimer GJ, Williams-Carrier R, Belcher S, Osborne E, Gierke J, Barkan A, 2003. Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. EMBO J. 22(15):3919-29.

Pittman M, 1931. Variation and type specifity in the bacteria species Haemophilus Influenza J Exptl Med. 53:471-492.

Radasci NW. and Bickle T, 1991. DNA Restriction and modification. In Escherchia coli, and salmonella. Cellular and molecular biology. (ed Neidhort) pp773-781. American Society for Microbiology, Washington DC, 1996.

Roberts RJ. and Macelis D, 1991. Restriction enzymes and their isoschizomes. Nuclei Acids Res. 19: 2077-2109.

Russell CJ. and Webster RG, 2005. The genesis of a pandemic influenza virus. Cell 123: 368–371.

Shope RE, 1931a. Swine Influenza.I Experimental transmission and pathology. J Exptl Med. 54:349-352.

Shope RE, 1931b. Swine Influenza. III Filtration experiments and etiology. J Exptl Med. 54: 373-385.

Shope RE, 1937. Immunologic relationship between swine and human influenza viruses in swine J Exptl Med. 66:151-168

Shope RE. and Francis T. Jr., 1936. Susceptibility of swine to the virus of human influenza. J Exptl Med. 64: 791-801.

Shub DA, Goodrich-Blair H, Eddy SR, 1994. Amino acid sequence motif of group I intron encoded endonucleases is conserved in open reading frames of group II introns. Trends Biochem. Sci. 19: 402–404.

Smith W, Andrewes CH, Liadlaw PP, 1933. A virus isolated from influenza patients. Lancet. 2: 66-68.

Stuart-Harris CH, 1936. The transmission of influenza virus to hedge hog. Brit J Exptl Path. 17:324-328

Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG, 2005. Characterization of the

1918 influenza virus polymerase genes. Nature 437: 889–893.

Toor N, Hausner G, Zimmerly S, 2001. Co-evolution of Grroup II intron RNAs with their intron encoded Reverse transcriptases RNA 7:1142–1152.

Treanor JJ, 2004. Influenza virus. In, 'Principles and Practice of Infectious Diseases' (eds. Mandell GL, Bennett JE, Dolin R.) pp 2060–2085. Churchill Livingstone Inc., New York.

Tumpey TM, 2005. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science 310: 77–80.

Vilchez RA, Fung J, Kusne S, 2002. The pathogenesis and management of influenza virus infection in organ transplant recipients. Transpl Infect Dis. 4: 177–182.

Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM, 1999. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. Mol Cell 4: 239–250.

WHO, 2007. Influenza fact sheets. http://www.who.int/mediacentre/factsheets/fs211/en/index.html Accessed 21st AUG 2007

Zimmerly S, Hausner G, Wu XC, 2001. Phylogenetic relationship among group II intron ORFs. Nucleic Acids Res. 29: 1238–1250.

Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM, 1995a. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell. 83(4): 529-38.

Zimmerly S, Guo H, Perlman PS, Lambowitz AM, 1995b. Group II intron mobility occurs by target DNA-primed reverse transcription. Cell 82: 545–554.

Zimmerly S, Moran JV, Perlman PS, Lambowitz AM, 1999. Group II intron reverse transcriptase in yeast mitochondria: Stabilization and regulation of reverse transcriptase activity by the intron RNA. J Mol Biol 289: 473–490.