

Rapid and Highly Informative Diagnostic Assay for H5N1 Influenza Viruses

Nader Pourmand^{1*}, Lisa Diamond¹, Rebecca Garten², Julianna P. Erickson¹, Jochen Kumm¹, Ruben O. Donis², Ronald W. Davis¹

¹ Stanford Genome Technology Center, Stanford University, Palo Alto, California, United States of America, ² Molecular Virology and Vaccines Branch, Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

A highly discriminative and information-rich diagnostic assay for H5N1 avian influenza would meet immediate patient care needs and provide valuable information for public health interventions, e.g., tracking of new and more dangerous variants by geographic area as well as avian-to-human or human-to-human transmission. In the present study, we have designed a rapid assay based on multilocus nucleic acid sequencing that focuses on the biologically significant regions of the H5N1 hemagglutinin gene. This allows the prediction of viral strain, clade, receptor binding properties, low- or high-pathogenicity cleavage site and glycosylation status. H5 HA genes were selected from nine known high-pathogenicity avian influenza subtype H5N1 viruses, based on their diversity in biologically significant regions of hemagglutinin and/or their ability to cause infection in humans. We devised a consensus pre-programmed pyrosequencing strategy, which may be used as a faster, more accurate alternative to *de novo* sequencing. The available data suggest that the assay described here is a reliable, rapid, information-rich and cost-effective approach for definitive diagnosis of H5N1 avian influenza. Knowledge of the predicted functional sequences of the HA will enhance H5N1 avian influenza surveillance efforts.

Citation: Pourmand N, Diamond L, Garten R, Erickson JP, Kumm J, et al (2006) Rapid and Highly Informative Diagnostic Assay for H5N1 Influenza Viruses. *PLoS ONE* 1(1): e95. doi:10.1371/journal.pone.0000095

INTRODUCTION

The worldwide spread of high-pathogenicity H5N1 avian influenza A virus in poultry and wild birds has resulted in many human infections, with high fatality rates. Although sustained human-to-human transmission has not yet occurred, concern about a potential pandemic continues to mount. The current HA lineage of H5N1 avian influenza was first found among domestic poultry populations in 1996 in southern China [1]. A similar H5N1 influenza virus spread directly from poultry to humans in Hong Kong in 1997, causing death in 6 out of 18 persons diagnosed with infection with this virus [2]. While the massive culling of poultry in 1997 temporarily eradicated the virus in Hong Kong, the virus has continued to spread across Asia, causing human deaths in Thailand, Vietnam, Indonesia, China and elsewhere [2,3]. The rapid spread of H5N1 in birds from Asia into Europe and Africa in recent months has intensified efforts to control the virus and avert a pandemic. To address the recognized need for rapid, low-cost diagnosis, tracking critically important genetic changes in the virus among animal and human host populations, and identifying specific viral clades [4], we have developed high-throughput methods for monitoring viral mutations that may control virulence and transmissibility in humans [5]. Accurate and rapid detection and tracking of H5N1 will be critical to prevent or control a potential pandemic.

Diagnosis of influenza type A infections in clinical microbiology laboratories has traditionally been performed using cell culture and/or direct fluorescent antibody assays [5–7]. These methods are time-consuming and require biosafety level 3 enhanced biocontainment facilities and equipment to protect laboratory personnel from exposure to H5N1 cultured in the laboratory. Because these facilities not widely available, culture-based assays are increasingly being replaced in clinical settings by the various polymerase chain reaction (PCR) methods [8–11].

PCR is more sensitive than traditional tests and detection does not require viable virus or morphologically intact infected cells in the sample. A PCR-based molecular diagnostic test is currently the most widely used by public health laboratories to diagnose the presence of H5N1 in clinical specimens [12]. We hypothesized

that coupling a PCR assay to a rapid sequencing method would further increase the value of molecular techniques for virus identification and characterization, especially if implemented into automated robotic platforms in the near future. Nucleic acid sequencing is considered the most reliable and highest-resolution method for virus identification, but is typically too slow and costly to use as a primary assay. Samples can be prepared sequentially for PCR diagnosis of H5N1 influenza virus, and pyrosequencing, yielding results in approximately 90 minutes, with immediate availability of the viral sequence data. The speed, sensitivity, precision, low cost, and high throughput of this method give it substantial advantages in H5N1 influenza characterization. We have designed an assay that focuses on three biologically significant regions of the H5N1 hemagglutinin gene (Figure 1), including sites informative of viral ancestry.

Influenza type A viruses have an eight-segment negative-sense RNA genome complexed with nucleoprotein and polymerase surrounded by the matrix protein and a lipid envelope that contains two integral membrane glycoproteins, hemagglutinin (HA) and neuraminidase (NA), protruding from the virion surface. The HA binds to sialic acid-terminated glycan receptors on the host cell surface, triggering virion uptake by endocytosis. Human respiratory tract cells have predominantly alpha 2-6 linked sialic

.....
Academic Editor: Paul Digard, University of Cambridge, United Kingdom

Received October 18, 2006; **Accepted** November 20, 2006; **Published** December 20, 2006

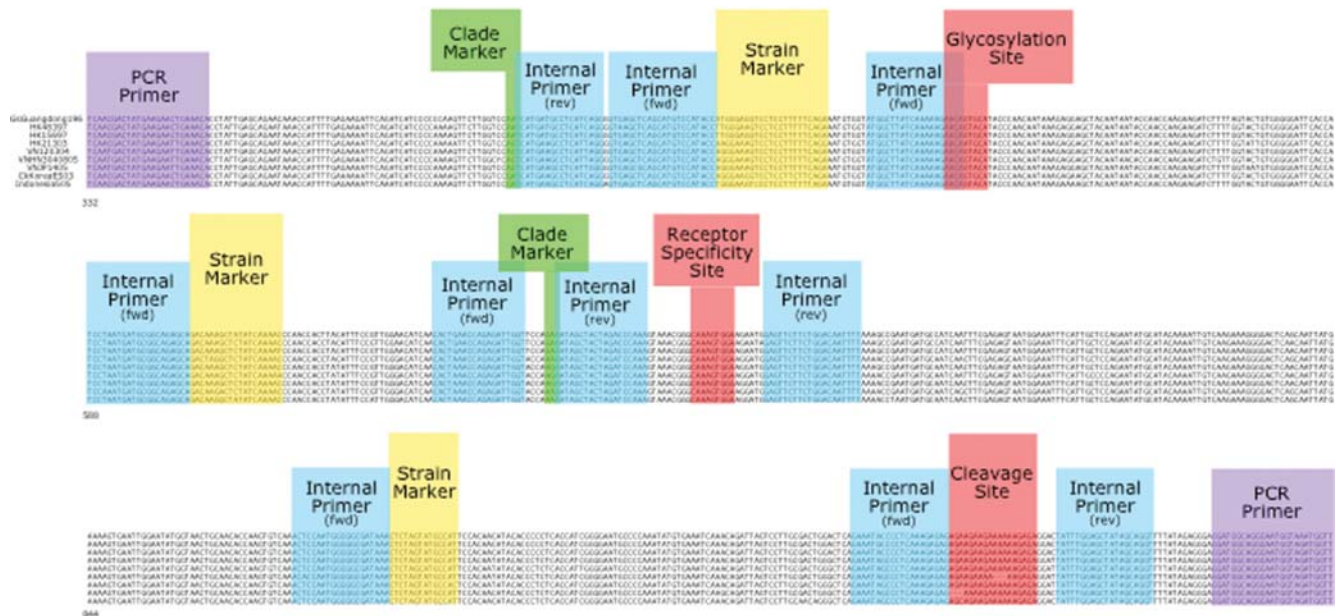
Copyright: © 2006 Pourmand et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: National Institutes of Health, 1R21 A1059499-01 and PO1-HG000205

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

*** To whom correspondence should be addressed.** E-mail: pourmand@stanford.edu

H5N1 Influenza A Hemagglutinin Gene Primers, Markers and Key Sites (Nine H5N1 strains, bases 332-1099)



A region of 768 bases of cDNA for the hemagglutinin gene of nine strains of H5N1 avian influenza was assayed to obtain sequence information for three key biologically significant sites for glycosylation, receptor specificity, and HA1/HA2 cleavage (sites indicated in red). Two clade markers (green) distinguish major and minor clades of H5N1. Three additional polymorphic sites (yellow) provide unambiguous strain identification. The PCR primers used to bracket this region are indicated in purple; internal primers are blue.

Figure 1. Primers, markers and key sites on H5N1 influenza A hemagglutinin gene. A region of 768 bases of cDNA of the hemagglutinin gene of nine strains of H5N1 was assayed to obtain sequence information for three key biologically significant sites (glycosylation, receptor specificity, and HA1/HA2 cleavage; sites indicated in red). Two markers (green) distinguish clades of H5N1. Three additional polymorphic sites (yellow) provide unambiguous strain identification. The PCR primers used to bracket this region are indicated in purple; internal primers are blue.
doi:10.1371/journal.pone.0000095.g001

acid receptors, whereas duck intestine has predominantly alpha 2-3 linked sialic acid receptors. Host-adapted human and avian influenza viruses selectively bind to homologous variant sialoside structures [13–15]. The HA receptor binding site is located at the convergence of one helix, two loops and three single residues near the top of the molecule; amino acid residues in the 184–186 (188–190 H3) helix as well as the 130–134 (134–138 H3) and 217–224 (221–228 H3) loops make up the rims, whereas residues 91 (98 H3), 149 (153 H3) and 179 (183 H3) form the floor of the cavity (H5 numbering; references: [16,17]). In the H3 subtype of HA, amino acids 222 and 224 (226 and 228 H3) are major determinants of human or avian host sialic acid specificity of the virus; alpha 2-6 linkages in human viruses and alpha 2-3 linkages in avian viruses [18]. Similar changes have been documented for the H1 HA that caused the 1918 so-called Spanish influenza pandemic [19,20]. Mutation at amino acid 223 from serine-to-asparagine (227 in H3) in a 2003 human H5N1 isolate (A/Hong Kong/213/03) maintained avian-like binding specificity of alpha 2-3 linked sialic acids but also had increased binding to alpha 2-6 sialic acid linkages [17].

The HA is also responsible for cell entry by mediating fusion of the endosomal and viral membranes. The HA requires proteolytic cleavage to become functional in membrane fusion. Cleavage, mediated by host proteases, results in two disulfide-linked subunits, HA1 and HA2. The HA1 region of the HA gene encompasses three sites of known biological significance:

- A glycosylation sequon at the top of the receptor binding site at amino acids 154–156 (of mature H5 HA1) (158–160 H3) has been linked to viral adaptation from waterfowl to domesticated poultry. The majority of H5N1 viruses isolated from humans since 2004 have this motif [4,21].
- The receptor binding site. Changes in the rims or floor of the receptor-binding site thought to favor binding of human forms of cell surface sialosides are considered critical for sustained transmission in the human population.
- Finally, the number of basic amino acids (lysine or arginine) in the cleavage site between HA1 and HA2 determines whether the virus is highly pathogenic for birds [22,23]. All Eurasian H5N1 viruses currently circulating are highly pathogenic and contain either five or six consecutive basic amino acids at this site [4].

Recent phylogenetic analysis of H5N1 evolution indicates that distinct major and minor clades have emerged among HA gene lineages; the two most important ones of these from a public health perspective are termed clades 1 and 2. These two clades are identifiable with two amino acids, 124 and 212 (129 and 216 in H3) [4]. These clade distinctions may be antigenically significant; therefore clade identification may aid selection of appropriate vaccines. Rapid H5N1 identification will assist in pinpointing the source of particularly virulent outbreaks, and in targeting limited supplies of vaccines and anti-virals to key regions.

Here, we report the results of a proof-of-principle experiment based on the application of pyrosequencing technology targeting the hemagglutinin gene of H5N1 influenza. The assay uses RT-PCR to amplify a known H5N1-specific region of 768 nucleotides. Subsequent pyrosequencing [24] of strains that are H5N1-positive with ten specific sequencing primers is used to determine HA clade and strain, receptor binding preference, low- or high-pathogenicity cleavage site, and glycosylation status. The assay presented here is considerably more informative than traditional techniques such as real-time PCR, as it not only identifies the H5N1 lineage but also predicts receptor binding properties that could herald the development of human-human transmissibility. This assay would be most advantageous as part of a screening method during high-volumes of H5N1 activity. Also, pyrosequencing has some advantages over Sanger sequencing-based methods, specifically the availability of sequence data directly following the sequencing primer, and the accessibility of the results in real time. Moreover, this assay is specific, rapid and cost-effective.

RESULTS

HA Amplification by PCR

The first step in the analysis of a clinical specimen or a viral isolate in our assay is the generation of a DNA copy of the viral RNA, which is accomplished by reverse transcription coupled to PCR (RT-PCR). This was performed using two different biotinylated combinations of PCR primers specific to the H5N1 regions of interest in order to achieve optimal sequencing flexibility of H5N1 isolates, as listed in Figure 2. As shown in Figure 3, either pair of primers provided reliable amplification of H5N1, and neither generated products when used in PCRs with negative controls (total genomic DNA from unrelated human cell lines) (confirmed by Sanger sequencing).

All of the H5N1 PCR products were sequenced at least twice both by Sanger dideoxy sequencing and pyrosequencing; PCR-positive amplicons generated correct sequence results independent of sequencing method or fragment size. Furthermore, no loops or primer-dimers were observed when primers were pyrosequenced in the absence of template. Representative and typical pyrograms of sample A/Vietnam/HN30408/2005, obtained by pyrosequencing, are shown in Figure 4; pyrograms for the remaining eight samples are available as supplementary information. The pyrosequencing run covering the required 14 bases took approximately 15 minutes.

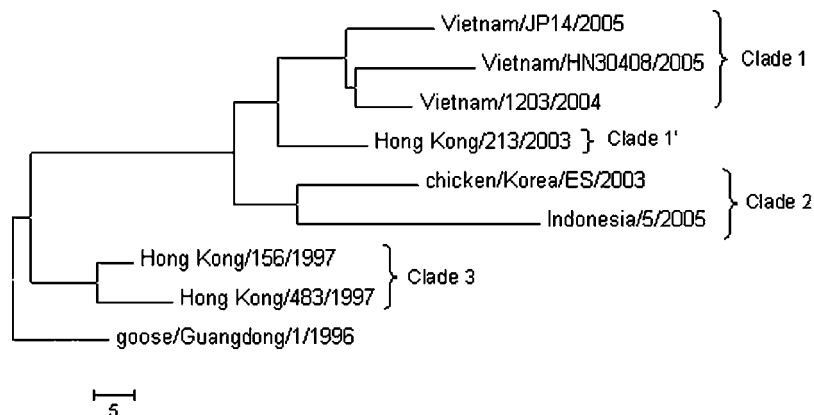


Figure 2. Phylogenetic relationships of H5N1 hemagglutinin (HA) genes from highly pathogenic H5N1 avian influenza viruses used in this study. Phylogenetic trees were inferred from nucleotide sequences by the neighbor joining method in the MEGA program. Horizontal distances are proportional to the number of nucleotide changes between the viruses. HA clade determinations are shown on the right. doi:10.1371/journal.pone.0000095.g002

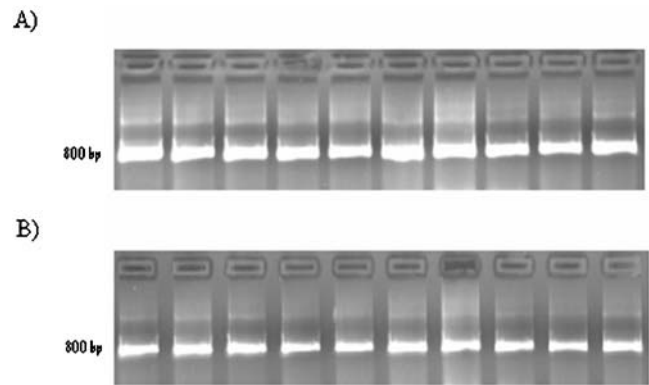
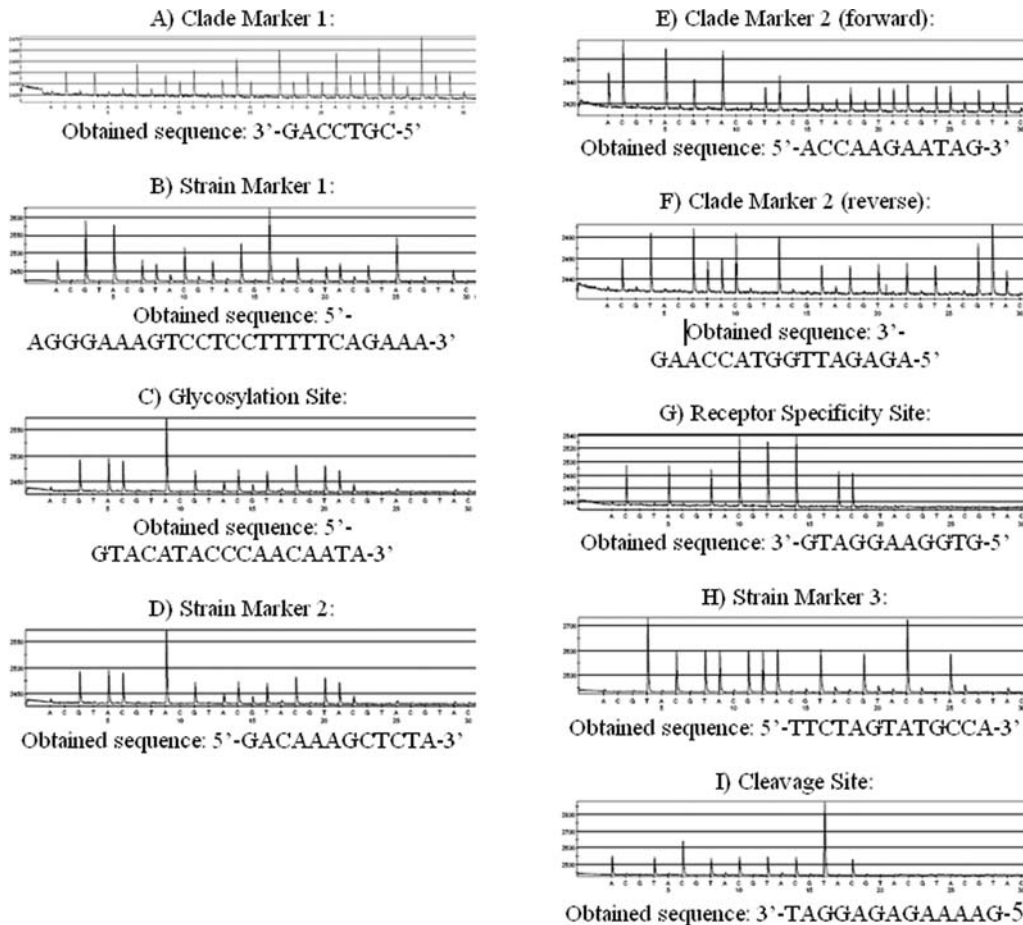


Figure 3. Amplicon DNA yield from PCR with different combinations of biotinylated primers. All PCR products were Sanger-sequenced to confirm the identity of the sample. A) The nine H5N1 samples amplified with a biotinylated forward primer (B-F-H5N1-1) and nonbiotinylated reverse primer (R-H5N1-2). From left to right, the samples are A/goose/Guangdong/1/96 (AF144305), A/Hong Kong/156/97 (AF036356), A/Hong Kong/483/97 (AF046097), A/Hong Kong/213/2003 (AY575869), A/chicken/Korea/ES/03 (AY676035), A/Vietnam/1203/2004 vaccine candidate with deletion of polybasic cleavage site (AY651334), A/Vietnam/JP14/2005 (ISDN117778), A/Vietnam/HN30408/2005 (ISDN119678), A/Indonesia/5/05 (ISDN125873). B) The H5N1 PCR products amplified with a biotinylated reverse primer (B-R-H5N1-3) and a nonbiotinylated forward primer (F-H5N1-4). Samples are in the order stated for part (A). doi:10.1371/journal.pone.0000095.g003

HA Pyrosequencing

The results of our pyrosequencing assay clearly distinguished the nine different strains of H5N1 avian influenza, based on eight sites, as shown in Table 1. Furthermore, this approach provided accurate sequencing of regions of known biological significance (for supplementary data, see <http://bioel.stanford.edu/avianflu/index.html>). The results of this assay can be seen in Table 2, which shows the amino acid sequences characteristic of each strain tested.

In order to characterize the strains based on the HA sequence, pyrograms obtained via pyrosequencing were visually inspected to verify the sequences of functionally relevant sites and the presence of polymorphisms that could serve as lineage markers. These sequences were then compared to known H5N1 sequences for



J) Sequence Map

TCAATGACTATGAAGAATTGAAACACTTATTGAGCAGAATAAACCATTTTG
 AGAAAATTCAGATCATCCCCAAAAGTTCTTGGCYCAGTCATGAAGCCTCAT
 TAGGGGTGAGCTCAGCATGTCCATACCAGGGAAAGTCCTCCTTTTTCAGA
 AATGTGGTATGGCTTATCAAAAAGAACAGTACATACCCAACAATAAAGAGG
 AGCTACAATAACCAACCAAGAAGATCTGTTGGTACTGTGGGGGATTACCC
 ATCCTAATGATGCGGCAGAGCAGACAAAGCTCTATCAAAAACCAACCACC
 TATATTTCCGTTGGGACATCAACACTAAACCAGAGATTGGTACCAAGAATA
 GCTACTAGATCCAAAGTAAACGGGCAAAGTGAAGGATGGAGTTCTTCTG
 GACAATTTAAAACCGAATGATGCAATCAATTTGAGAGTAATGGAAATTC
 ATTGCCCCAGAATATGCATACAAAATTGTCAAGAAAGGGGACTCAACAATTA
 TGAAAAGTGAATTGGAATATGGTAACTGCAACACCAAGTGTCAAACACCAA
 TGGGGGCGATAAACTAGTATGCCATCCACAATATACACCCTCTACCA
 TCGGGGAATGCCCAAATATGTGAAATCAAACAGATTAGTCCTTGC GACTGG
 GCTCAGAAATAGCCCTCAAAGAGAGAGAAGAAA---
 AAGAGAGGATTATTTGGAGCTATAGCAGGTTTTA
 TAGAGGGAGGATGGCAGGGAATGGTAGATGGTT

Figure 4. Panels A–I: Pyrograms obtained by pyrosequencing of sample A/Vietnam/HN30408/2005 with the HS 96 system. Peaks above a given nucleotide sample indicate nucleotide incorporation (with height proportional to number of consecutive incorporation events). Initial pyrosequencing was performed with systematic nucleotide dispensation for *de novo* sequencing (pre-programmed dispensation runs can also be performed). Panel J: Map of the HA sequence of A/Vietnam/HN30408/2005 obtained by Sanger sequencing; pyrosequencing data is overlaid and color-coded to show correlation between the sequence data obtained by the two methods. Reverse-primed pyrosequencing results should be read as reverse complements. For clarity, primers and biologically significant sequence segments are colored to correspond to the sequence map presented in Figure 1.

doi:10.1371/journal.pone.0000095.g004

Table 1. Sequences of H5N1 clade markers and active sites, as determined by pyrosequencing.

	Clade Marker 1		Strain Marker 1	Glycosylation Site	Strain Marker 2	Clade Marker 2 (F)	Clade Marker 2 (R)	Receptor Specificity Site	Strain Marker 3	Cleavage Site
goose/Guangdong/1/96	AACCTGGTCTTG AAACCCATCA	ATGGGAGGTCCTCC TTTTTTCAGAAA	GTGCATACCCA ACAATAAA	GACAAAGCTCTA	TACCAGAAATAG CTACTAGACCCA	AGACCTTGGTT	GTAAGAAGGTGAA ACGGGCAATG AAACCCAGA	CTCTAGTATGCCA	AGAAGAAGAA AAAAGAGAGAGGA	
Hong Kong/156/97	AACCTGGTCTT GAAACCCATCA	TGGGAGGTCCT CCTTTTTCAGA	GTACATACCCA CAATAAAGA	GACAAAGCTCTA	TCCAGAAATAG	AGACCTTGGTT AGAGA	GTAAGAAGGTGGA AACGGGCAATG	CTCTAGTATGCCA	AGAAGAAGAA AAAAGAGA	
Hong Kong/483/97	AACCTGGTCTT GAAACCCATCA	TGGGAAAGTCCT CCTTTTTCAGA	GTACATACCCA CAATAAAGA	GACAAAGCTCTA	TCCAGAAATAG	AGACCTTGGTT	ATAAGAAGGTGAA ACGGG	CTCTAGTATGCCA	AGAAGAAGAA AAAAGAGA	
Hong Kong/213/03	GACCTGGTCTT AAACCCATCA	AAGGAAAGTCCT CCTTTTTCAG	ATGCATACCCA CAATAA	GACTAGGCTCTA	ACCAAAA	AAACCATGGTTA GAGACCAAAATCACA	GTAGGAAGGT AAAA	CTCTAGTATGCC	AGAAGAAGAA AAGAGAGGA	
chicken/Korea/ES/03	AGCCTGGTCTT GAAACCCATCA	AGGGAAGGTCCT CCTTCTCAGAAA	GTGCACCTACCC AACAATAA	GACAAAGCTCTA	ACCAAAA	AAACCATGG TTAGAGACAA	GTAGGAAGGTG AACGGGCA	CTCTAGTA	AAAAGAAA AAGAGA	
Vietnam/1203/04	GACCTGGTCTT GAAACCCATCA	AGGGAAGGTCCT CCTTTTTCAGA	GTACATACCCA CAATAAAGA	GACAAAGCTCTA	ACCAAGAA	unreadable	GTAGGAAGGTG AACGGG	CTCTAGCATGCCA	unreadable	
Vietnam/JP14/05	GACCTCGGCTT	AGGGAAGGTCCT CCTTTTTCAGAAA	GTACATACCCA TAAGAAA	GACAAAGCTCTA	ACCAAGAA	GAAACCCATGG	GTAGGAGGGTG AACGGG	CTCTAGTATGCCAC	AGAAGAAG AAAAAAGAGA	
Vietnam/HN30408/05	CAGCCTGGTCTT GAAACCCATCA	AGGGAAGGTCCT CCTTTTTCAGAAA	GTACATACCCA CAATAAGA	GACAAAGCTCTA	ACCGAAGAA	GAAACCCATGGTT AGAGACCAAAAT	GTAGGAAGGTG	TTCTAGTATGCCA	AGAAGAAA AGAGAGG	
Indonesia/5/05	CAGCCTGGTCTT GAAAGGATCA	TGGGAAAGTCCT CCTTTTTAGA	GTACATACCCA CAATAAAGAA	GACAAAGGCTATA	ACCAAGAA	AAACC	GTAGGAAGGTG AACGGG	CTCTAGTATGCCA	AGCAGAAGAA AAAAGAGA	

Sample strains are listed in the leftmost column, site of sequence information in the top row. "F" or "R" indicates whether the primer is forward- or reverse-directed. The 5'-3' sequences presented here were obtained by reading pyrograms (as reverse complements when using a reverse-directed sequencing primer).

doi:10.1371/journal.pone.0000095.t001

Table 2. Characterization of the hemagglutinin from the H5N1 influenza viruses used in this study.

Virus Name	HA clade	Glycosylation Motif (NXT/S, X≠P) at aa 154	Receptor Binding Site (221-224)	Cleavage Motif	Host/Outcome
goose/Guangdong/1/96	ancestor	Present	GQSG	RRRKKR	Goose
Hong Kong/156/97	3	Absent	GQSG	RRRKKR	Human/Died
Hong Kong/483/97	3	Present	GQSG	RRRKKR	Human/Died
Hong Kong/213/03	1'	Present	GQNG	RRRKKR	Human/Died
Vietnam/1203/04	1	Present	GQSG	RRRKKR	Human/Died
Vietnam/JP14/05	1	Present	GQSG	RRRKKR	Human/Died
Vietnam/HN30408/05	1	Present	GQSG	RRKKR	Human/Survived
chicken/Korea/ES/03	2	Absent	GQSG	KRKKR	Chicken
Indonesia/5/05	2	Present	GQSG	SRRKKR	Human/Died

doi:10.1371/journal.pone.0000095.t002

clade and strain identification. Certain positions provided essential information for the identification of signatures important for public health, such as changes in the receptor binding site which could signal an increase in the ability of the virus to transmit from human- to-human. The prototype receptor binding site for avian isolates in amino acids 221-224 of the HA1 is GQSG. One human isolate, Hong Kong/213/03, has a mutation at amino acid 223, S->N, which switches the affinity of the hemagglutinin from alpha 2-3 linked sialic acid to the alpha 2-6 linked sialic acid preferred by human influenza viruses. This mutation was detected by a single nucleotide change, G->A, in the sequence of the receptor binding active site. Rapid characterization of the receptor binding active site and other active sites of the hemagglutinin is critical in identifying viruses with increased pandemic potential.

DISCUSSION

Our initial results indicate that targeted pyrosequencing approach can clearly distinguish among different strains of H5N1, and can accurately sequence regions of known biological significance such as receptor binding (supplementary data: <http://bioel.stanford.edu/avianflu/index.html>). The HA sequence results obtained by pyrosequencing were 100% identical to those obtained by the Sanger method (two replicates were performed to cross-verify results). Three pyrosequencing replicates, with nine samples at ten sites, were performed: two with a *de novo* dispensation and one with a pre-programmed dispensation order. Of the three pyrosequencing replicates performed, all successful runs verified one another. Because Sanger sequencing relies on molecular separation of the polymerase-mediated extension of the primer, the first ~20–50 bases 3' of the primer are not discernible. In contrast, pyrosequencing provides unambiguous sequence from the first nucleotide 3' to the primer, increasing the sequence information yield per assay. This feature of pyrosequencing was exploited by designing H5N1-specific primers adjacent to critical polymorphic sites such that the initial base-calls are usually sufficient to determine the lineage of a given HA gene.

A simpler detection assay giving a positive or negative identification of H5N1 will be helpful, but a more information-rich method is necessary to provide additional information to guide patient care or public health measures, especially as new strains emerge. Indeed, some laboratories are already using pyrosequencing to identify drug-resistant viruses [25]. As H5N1 influenza becomes more prevalent, it will be impractical to perform full sequencing of all isolates in time for analysis of rapidly

changing epidemiological trends. Our assay could be a valuable complement to full sequencing at public health laboratories. We have developed a rapid and inexpensive assay based on DNA sequencing for early detection of virus present in host cells. This assay permits rapid, simplified and accurate identification of avian H5N1 influenza A, and consists of simple procedures maintaining high sensitivity and specificity. As new sites of functional relevance are identified, new site-specific primers can be added to broaden the utility of the assay. Expanded knowledge of H5N1 sequence and evolution is likely to contribute to more effective diagnostic methods and treatments that require less investment of time and money.

An additional advantage of pyrosequencing over the Sanger method is its application in genotyping single and multiple HPV infections [26]. This technology is based on the use of multiple sequencing primers in the same pyrosequencing reaction. Pools of primers are chosen based on their position on the target, the sequence information they yield, and their cross-reactivity with other primers in the multiplex assay. Clinically relevant samples containing multiple infections can then be identified and characterized.

These early robust results have validated our approach and selection of specific primers. We have moved toward a further refinement of the assay based on these results. After an initial *de novo* pyrosequencing run was performed on the H5N1 samples, sequence data for each strain were compiled for each nucleotide site. This information was integrated to produce consensus pre-programmed dispensation orders of nucleotides. This type of pyrosequencing would allow us to obtain sequence results of similarly high quality in 10 to 15 minutes (Figure 5). With this approach, all of the H5N1 subtypes included in this study could be positively identified. New variants would not be fully characterized, but would be heralded by truncated sequences.

In summary, we have developed a reliable, rapid, cost-effective, and information-rich diagnostic assay for H5N1 influenza. This sequence-based method could be extended to include amplicons from other genes of interest; for example, drug targets such as NA and the M2 ion channel [27,28], or the postulated virulence motifs on NS1 or PB2 [29,30]. Pyrosequencing is user-friendly, and permits significantly more efficient and rapid genotyping than traditional techniques. This method detects a wide range of influenza A H5N1 subtypes based on sequence information. The assay will allow further development of technology to directly detect H5N1 or specific strains of influenza A in clinical specimens without extensive sample preparation.

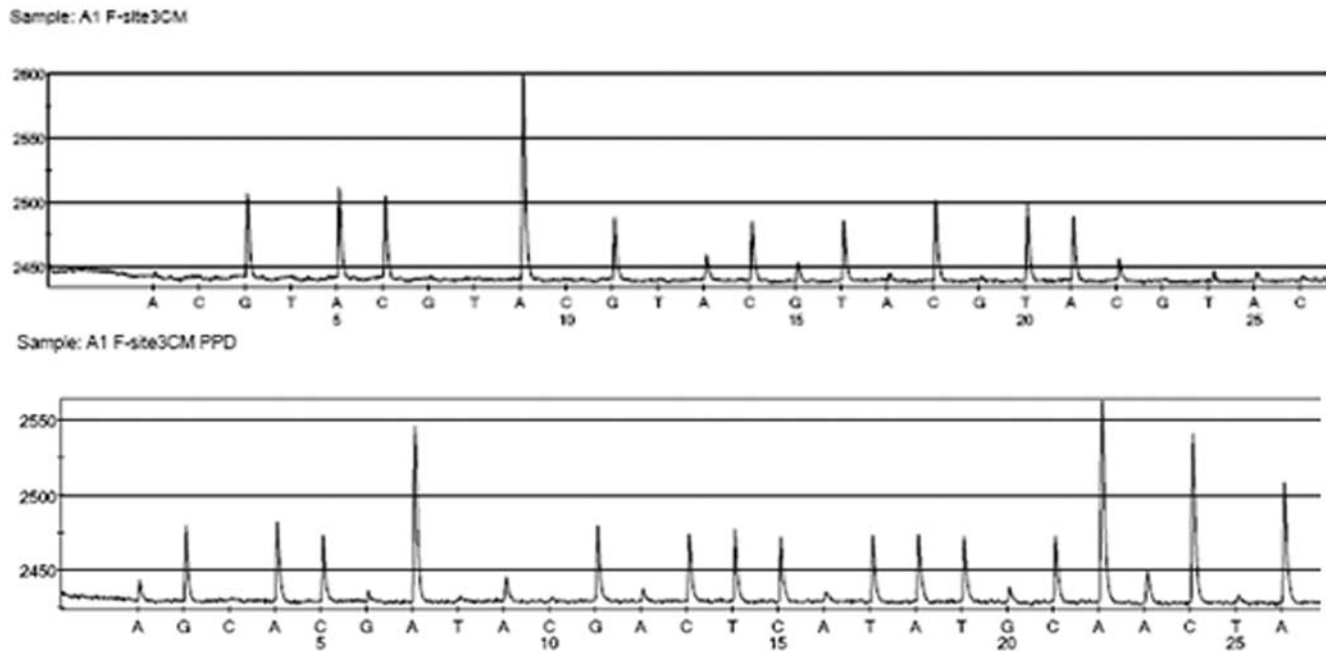


Figure 5. A representative pyrogram illustration. Comparison of pyrosequencing results obtained using the *de novo* (top) and consensus pre-programmed (bottom) sequencing methods for site 3 (a clade marker) in sample A1 (goose/Guangdong/1/96). Both methods give the sequence GACAAAGCTCTATCAAAC for our laboratory stock. In contrast the GenBank sequence (accession # AF144305) reads GACAAAGCTATATCAAAC. Database information was found to be based off wild-type sequence information, while the sample we analyzed was obtained via reverse genetics. A longer read length is obtained via the consensus pre-programmed dispensation method (22 versus 12 nucleotides). doi:10.1371/journal.pone.0000095.g005

MATERIALS AND METHODS

Avian Influenza A Subtypes and H5N1-Specific Primer Design

Sequences of 362 avian influenza A virus species were acquired from GenBank at the National Center for Biotechnology Information and Influenza Sequence Database at Los Alamos National Laboratory and aligned using Clustal X version 1.83 [31]. A specific sequencing primer was designed for H5N1, with no sequence similarity to the other influenza virus HA subtypes based on the database searches and alignments (Figure 1).

H5N1 Avian Influenza Virus Isolates

The viruses that formed the test set for this study were chosen on the basis of their diversity in biologically significant regions of hemagglutinin and/or their ability to cause infection in humans. Virus strain names are shown in Figure 2 and accession numbers are as follows: A/goose/Guangdong/1/96 (AF144305), A/Hong Kong/156/97 (AF036356), A/Hong Kong/483/97 (AF046097), A/Hong Kong/213/2003 (AY575869), A/chicken/Korea/ES/03 (AY676035), A/Vietnam/1203/2004 vaccine candidate with deletion of polybasic cleavage site (AY651334), A/Vietnam/JP14/2005 (ISDN117778), A/Vietnam/HN30408/2005 (ISDN119678), A/Indonesia/5/05 (ISDN125873).

RNA Extraction, RT-PCR and PCR Amplification

Viruses were isolated and propagated by inoculation into the allantoic sac of 10-day old chicken embryos, as described previously [32]. Viral infectivity was determined by endpoint dilution, injection into chicken embryos and hemagglutination assay. All work involving infectious H5N1 influenza was performed in government-approved biosafety level 3-enhanced

containment facilities as required by the U.S. Department of Agriculture and the Select Agent Program (see guidance at www.cdc.gov/flu/h2n2bsl3.htm). Viral RNA was extracted using the QIAmp vRNA Kit (Qiagen, Valencia CA). Extractions were performed according to manufacturer's instructions. QIAGEN OneStep RT-PCR kit (Qiagen, Valencia CA) was used to perform RT-PCR from 30 ng of RNA (approximate concentration of 10 ng/ μ l by spectrophotometer) in a 50 μ l reaction volume. The RT-PCR amplification primers, which are biotinylated F-H5N1-1/3 (5'-TCAAYGACTATGAAGAAYTGAAACA-3'), and R-H5N1-2/4 (5'-AACCATCTACCATTCCCTGCCATCC-3) were synthesized by IDT (Coralville, IA, USA). RT-PCR was performed with a DNA Engine (PTC-200) Peltier Thermal Cycler (BIO-RAD, Hercules, Calif.) as follows: 50°C for 10 minutes, 95°C for 15 minutes, 35 cycles of 95°C, 55°C, and 72°C for 1 minute each, and finally a 10 minute final extension at 72°C. To confirm proper amplification, PCR products were electrophoresed in a 1% agarose gel and visualized by ethidium bromide staining under UV illumination.

Sanger dideoxy DNA Sequencing of PCR Products

The amplified DNA from all isolates was cycle sequenced in both directions using the BigDye Terminators Reaction Kit v. 3.1 (Applied Biosystems, Foster City, CA) on an ABI automated DNA sequencer (3730 XL DNA Analyzer).

Hemagglutinin Sequencing Primers

An entropy-based analysis of genetic variation among H5N1 strains (Figure 1) was used to design PCR primers [33]. The approach focused on amplicons comprising regions for receptor binding sites, cleavage site and glycosylation sites, along with markers to identify clade and individual strain (Figure 1). Highly

conserved sites were selected as internal sequencing primers to sequence each region of interest. The selected region is bracketed by sequences unique to and highly conserved in the H5N1 subtype. Based on sequence alignments of H5N1 and sequencing results (data not shown) from all the isolates, a set of sequencing primers were designed spanning a region of the HA informative with regards to clade, strain, receptor binding motif, cleavability and glycosylation (Figure 1) and which specifically hybridize to H5N1. Sequences for relevant influenza viruses with known properties and virulence status were compiled for each sample from pyrograms. These sequences were then compared to known H5N1 sequences by alignment and visual inspection.

Pyrosequencing

Biotinylated PCR product (5 ng/μl) from RT step from viral RNA (Figure 1) was immobilized onto 2.5 μl streptavidin-coated High Performance Sepharose beads (Amersham Biosciences, Piscataway, NJ) by incubation at room temperature for at least 10 minutes with agitation at 1400 rpm. Single-stranded DNA was obtained by washing the immobilized PCR product with 70% EtOH, denatured with 0.2 M NaOH, and washed with TA-Buffer (0.1 M Tris-Acetate, pH 7.6) using a Vacuum Prep Tool and Vacuum Prep Worktable (Biotage, Uppsala, Sweden). The beads were then suspended in 12 μl annealing buffer (10 mM Tris-acetate pH 7.75, 5 mM Mg-acetate) containing 0.3 pmol sequencing primer. Single-stranded DNA was hybridized to the sequencing primer by incubation at 90°C for 2 minutes, at 60°C for 5 minutes and at room temperature for 5 minutes.

Primed single-stranded PCR products were sequenced using PSQ™ HS96A System (Biotage). Sequencing was performed in a total volume of 12 μl using the PSQ 96 Gold kit (Biotage). Consensus pre-programmed dispensation orders were determined by integrating sequence information for each strain at a given sequencing site [34]. See figure 5 for a comparison of *de novo* and pre-programmed sequencing strategies. As shown, approximately 10 more nucleotides of sequence information are obtained with the pre-programmed method, because nucleotide incorporations are anticipated. This reduces the amount of time required as well as the accumulation of sequencing byproducts. Negative-control

nucleotide dispensations were also included in the program to check for insertions and to measure background signals. The identity and number of nucleotide extension events were determined by automated measurement of the amount of light generated after incorporation of each dNTP.

Raw data were interpreted using software developed specifically for this purpose, "Classifier." Classification of samples by strain is straightforward using the short sequence segments obtained. We use a Support Vector Machine [35] approach implemented in the statistical programming language R [36] to classify a given sample with statistical accuracy. This provides an automated sample identification tool, designed for eventual use with large numbers of source sequences [37,38].

Supplementary information

See <http://bioel.stanford.edu/avianflu/index.html> for detailed pyrosequencing results for all nine strains.

ACKNOWLEDGMENTS

We thank Alexander Klimov, Amanda Balish, Rick Bright, Catherine Smith, Michael Shaw, Yumi Matsuoka, and Nancy Cox for their invaluable help. Viruses and sequences were kindly provided through the WHO Global Influenza Program, and its affiliates, including Nyoman Kandun, Endang Sedyaningsih, Patrick Blair, Huluan Chen, Wilina Lim, Peter Cheng and Le Thi Quynh Mai.

Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, is provided for information only and does not constitute or imply its endorsement, recommendation, or favoring by the Centers for Disease Control and Prevention and the U.S. Department of Health and Human Services.

The views and opinions expressed herein are those of the authors and do not necessarily reflect the views and opinions of the Centers for Disease Control and Prevention and the U.S. Department of Health and Human Services.

Author Contributions

Conceived and designed the experiments: RD NP. Performed the experiments: JE. Analyzed the data: NP LD JE. Contributed reagents/materials/analysis tools: JK LD RG RD. Wrote the paper: NP LD RG JE RD.

REFERENCES

- Xu X, Subbarao, Cox NJ, Guo Y (1999) Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology* 261: 15–19.
- Subbarao K, Klimov A, Katz J, Regnery H, Lim W, et al. (1998) Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* 279: 393–396.
- Smith GJ, Fan XH, Wang J, Li KS, Qin K, et al. (2006) Emergence and predominance of an H5N1 influenza variant in China. *Proc Natl Acad Sci U S A* 103: 16936–16941.
- WHO (2005) World Health Organization Global Influenza Program Surveillance Network, "Evolution of H5N1 avian influenza viruses in Asia," *Emerg Infect Dis* 11, 2005. (<http://www.cdc.gov/ncidod/cid/voll1no10/pdfs/05-0644.pdf>).
- Shinya K, Hatta M, Yamada S, Takada A, Watanabe S, et al. (2005) Characterization of a human H5N1 influenza A virus isolated in 2003. *J Virol* 79: 9926–9932.
- Effler PV, Ieong MC, Tom T, Nakata M (2002) Enhancing public health surveillance for influenza virus by incorporating newly available rapid diagnostic tests. *Emerg Infect Dis* 8: 23–28.
- Schmidt NJ, Emmons RW, American Public Health Association. Committee on Laboratory Standards and Practices (1989) Diagnostic procedures for viral, rickettsial, and chlamydial infections. Washington, DC: American Public Health Association. xv, 1225 p.p.
- Templeton KE, Scheltinga SA, Beersma MF, Kroes AC, Claas EC (2004) Rapid and sensitive method using multiplex real-time PCR for diagnosis of infections by influenza A and influenza B viruses, respiratory syncytial virus, and parainfluenza viruses 1, 2, 3, and 4. *J Clin Microbiol* 42: 1564–1569.
- Fan J, Henrickson KJ, Savatski LL (1998) Rapid simultaneous diagnosis of infections with respiratory syncytial viruses A and B, influenza viruses A and B, and human parainfluenza virus types 1, 2, and 3 by multiplex quantitative reverse transcription-polymerase chain reaction-enzyme hybridization assay (Hexaplex). *Clin Infect Dis* 26: 1397–1402.
- Whiley DM, Sloots TP (2005) A 5'-nuclease real-time reverse transcriptase-polymerase chain reaction assay for the detection of a broad range of influenza A subtypes, including H5N1. *Diagn Microbiol Infect Dis*.
- Habib-Bein NF, Beckwith WH 3rd, Mayo D, Landry ML (2003) Comparison of SmartCycler real-time reverse transcription-PCR assay in a public health laboratory with direct immunofluorescence and cell culture assays in a medical center for detection of influenza A virus. *J Clin Microbiol* 41: 3597–3601.
- (CDC) CfDCaP (2006) New laboratory assay for diagnostic testing of avian influenza A/H5 (Asian Lineage). *MMWR Morb Mortal Wkly Rep* 10: 127.
- Paulson JC (1985) *The Receptors*. Conn M, ed. Orlando, FL: Academic. 131–219 p.
- Connor RJ, Kawaoka Y, Webster RG, Paulson JC (1994) Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* 205: 17–23.
- Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, et al. (2000) Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J Virol* 74: 8502–8512.
- Ha Y, Stevens DJ, Skehel JJ, Wiley DC (2002) H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. *Embo J* 21: 865–875.
- Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, et al. (2006) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* 312: 404–410.

18. Vines A, Wells K, Matrosovich M, Castrucci MR, Ito T, et al. (1998) The role of influenza A virus hemagglutinin residues 226 and 228 in receptor specificity and host range restriction. *J Virol* 72: 7626–7631.
19. Stevens J, Blixt O, Glaser L, Taubenberger JK, Palese P, et al. (2006) Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J Mol Biol* 355: 1143–1155.
20. Reid AH, Fanning TG, Hultin JV, Taubenberger JK (1999) Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. *Proc Natl Acad Sci U S A* 96: 1651–1656.
21. Banks J, Plowright L (2003) Additional glycosylation at the receptor binding site of the hemagglutinin (HA) for H5 and H7 viruses may be an adaptation to poultry hosts, but does it influence pathogenicity? *Avian Dis* 47: 942–950.
22. 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.
23. Ohuchi M, Orlich M, Ohuchi R, Simpson BE, Garten W, et al. (1989) Mutations at the cleavage site of the hemagglutinin after the pathogenicity of influenza virus A/chick/Penn/83 (H5N2). *Virology* 168: 274–280.
24. Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. *Science* 281: 363, 365.
25. Bright RA, Medina MJ, Xu X, Perez-Oroz G, Wallis TR, et al. (2005) Incidence of adamantane resistance among influenza A (H3N2) viruses isolated worldwide from 1994 to 2005: a cause for concern. *Lancet* 366: 1175–1181.
26. Gharizadeh B, Zheng B, Akhras M, Ghaderi M, Jejelowo O, et al. (2006) Sentinel-base DNA genotyping using multiple sequencing primers for high-risk human papillomaviruses. *Mol Cell Probes* 20: 230–238.
27. Scheffner M (1998) Ubiquitin, E6-AP, and their role in p53 inactivation. *Pharmacol Ther* 78: 129–139.
28. Li KS, Guan Y, Wang J, Smith GJD, Xu KM, et al. (2004) Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* 430: 209–213.
29. Hatta M, Asano Y, Masunaga K, Ito T, Okazaki K, et al. (2000) Mapping of functional domains on the influenza A virus RNA polymerase PB2 molecule using monoclonal antibodies. *Arch Virol* 145: 1947–1961.
30. Obenauer JC, Denson J, Mehta PK, Su XP, Mukatira S, et al. (2006) Large-scale sequence analysis of avian influenza isolates. *Science* 311: 1576–1580.
31. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.
32. Kendal AP, Pereira MS, Skehel JJ (1982) Hemagglutination inhibition. p. B17–35, *In* PMKendal AP, Skehel JJ, eds. *Concepts and procedures for laboratory-based influenza surveillance*. Centers for Disease Control and Prevention and Pan-American Health Organization, Atlanta.
33. Cover TM, Thomas JA (1991) *Elements of information theory*. New York: Wiley. xxii, 542 p.p.
34. Gharizadeh B, Akhras M, Unemo M, Wretling B, Nyren P, et al. (2005) Detection of gyrA mutations associated with ciprofloxacin resistance in *Neisseria gonorrhoeae* by rapid and reliable pre-programmed short DNA sequencing. *Int J Antimicrob Agents* 26: 486–490.
35. Cristianini N, Shawe-Taylor J (2000) *An introduction to support vector machines and other kernel-based learning methods*. Cambridge, U.K.; New York: Cambridge University Press. xiii, 189 p.
36. R Statistical Package <http://www.r-project.org/>.
37. Duda RO, Hart PE, Stork DG (2001) *Pattern classification*. New York; Chichester [England]: Wiley. xx, 654 p.p.
38. Meyer D (2006) *Support Vector Machines*. Technische Universität Wien, Austria. pp. Interface to libsvm in package e1071.