Genomic analysis of A(H1N1)pdm09 influenza virus in Finland

Triin Lakspere

Master’s thesis

Master’s Degree in Translational Medicine TRANSMED

Faculty of Medicine

University of Helsinki

2013
# Master's thesis abstract

**Faculty of Medicine** | **Master's Degree Programme in Translational Medicine**
---|---
**Subject:** Translational medicine | **Master's thesis**
---|---
**Author** | **Triin Lakspere**
---|---
**Title** | Genomic analysis of A(H1N1)pdm09 influenza virus in Finland
---|---
**Supervisor** | Denis Kainov, Laura Kakkola
---|---
**Month and year** | October 2013
---|---
**Number of pages** | 50
---|---
**Supervisor's affiliation** | Institute for Molecular Medicine Finland (FIMM), University of Helsinki

## Abstract

Influenza viruses are a group of pathogens in the family Orthomyxoviridae, which are classified into 6 genera (A,B,C, Thogotovirus, Isavirus and new unnamed genus). Type A influenza viruses are categorized based on their surface glycoproteins: hemagglutinin (HA) and neuraminidase (NA). So far 17 HA and 9 NA subtypes have been identified. Influenza genome comprises of eight single-stranded negative-sense RNA segments that encode ten to twelve proteins (HA, NA, NP M1/M2, NS1/NS2, PA, PB1 and PB2). Influenza replication cycle depends on the surface proteins binding to host cell receptors, pH mediated fusion and cell-mediated transcription and replication of the viral genome. Virus particles leave the host cell via budding.

Influenza viruses cause global epidemic infections each year, the peak is from December to March. These pathogens have also contributed to six global pandemics identified so far. The latest pandemic outbreak was announced by WHO in 2009 which caused over 5000 hospitalizations in Finland. Factors contributing to the severity of clinical outcome can be either genetic, environmental or caused by human host features.

This study aims to identify the susceptibility factors for severe influenza A infections and describe the phylodynamics of the latest pandemic A(H1N1)pdm09 from Finnish patient nasopharyngeal aspirates collected between 2009-2013.

One-step reverse-transcription PCR was used to amplify all of the 8 segments equally. Fast and precise next-generation sequencing with Illumina 2000 sequencer was used to generate the sequences. Results were bioinformatically analysed using Bayesian modelling with Markov Chain Monte Carlo algorithms of probability distributions. Models analysed showed highest mutation rate in hemagglutinin protein. Phylodynamic analysis revealed higher mutation rate of HA and NA compared to other proteins.

Subgroup specific polymorphisms (either in severe or mild cases) were not identified. In total 4657 amino acid substitutions were located in 135 pandemic A(H1N1)pdm09 patient isolates and 238 in 10 seasonal patient samples. Viral HA, NA and PB2 were more frequently mutated than other proteins. Interestingly this study identified double-resistant markers (E119K and S31N) to two antiviral drugs (amantadine and oseltamivir) in one patient isolate (A/Helsinki/598/2013). Previously reported D222 polymorphism (without the signalling peptide) causing more severe clinical outcome was not identified in any of the patient isolates in this study.

## Keywords

A(H1N1)pmd09, Finland, Bayesian, influenza

---

**Where deposited**

---

**Additional information**
Table of Contents

Abbreviations ........................................................................................................................................ 5

Review of the literature ........................................................................................................................ 6
Influenza virus ........................................................................................................................................... 6
Structure of influenza viruses .................................................................................................................. 6
Replication cycle ...................................................................................................................................... 8
  Attachment and nuclear transport ......................................................................................................... 8
  Transcription of viral RNA and translation of proteins ........................................................................ 9
  Replication of viral RNA genome ........................................................................................................ 10
  Export of viral genome from the nucleus ............................................................................................. 10
Molecular virus and host cell interactions ............................................................................................. 10
Assembly of virions .................................................................................................................................. 11
Function of neuraminidase ..................................................................................................................... 12
Human host factors in influenza replication .......................................................................................... 13
Immune response .................................................................................................................................... 13
Characteristics of influenza virus induced diseases .................................................................................. 14
  Genetic susceptibility and severity of influenza infection .................................................................... 15
  Treatment and prevention ..................................................................................................................... 17
Vaccine development .............................................................................................................................. 18
Evolution and epidemiology .................................................................................................................. 19
  Influenza virus pandemics and outcomes ......................................................................................... 19
  Influenza A(H1N1)pmd09 ..................................................................................................................... 20

Aim of the study ....................................................................................................................................... 23

Materials and methods ........................................................................................................................... 24
  Patient sample collection ....................................................................................................................... 24
  Propagation in MDCK cells ................................................................................................................... 24
  RNA extraction ....................................................................................................................................... 25
  RT-PCR ................................................................................................................................................ 26
  Library preparation for sequencing ......................................................................................................... 27
  Sequencing ........................................................................................................................................... 28
  Bioinformatic analysis ............................................................................................................................ 29
    Mutation analysis ............................................................................................................................... 29
    Phylodynamic analysis ....................................................................................................................... 30

Results ..................................................................................................................................................... 32
  Propagation of viruses ........................................................................................................................... 32
  One-step versus two-step RT-PCR methods .......................................................................................... 32
  Sequencing of the isolates ..................................................................................................................... 33
  Phylogenetic analysis ............................................................................................................................ 33
  Whole-genome mutational analysis ...................................................................................................... 35
    Structural analysis ............................................................................................................................. 36
    Antiviral resistance ............................................................................................................................. 37

Discussion ............................................................................................................................................... 39
Abbreviations

ACT     Auto-correction time
ARDS    Acute Respiratory Disease Syndrome
CaMK2B  Calcium/calmodulin kinase II beta
cRNA    Complementary RNA
COPI    Coat Protein Complex I
CPSF    Cleavage and polyadenylation specificity factor
ECDC    European Centre of Disease Prevention and Control
eIF4A,E,G Eukaryotic initiation factor 4A,E,G
ESS     Effective sample size
FDA     Food and Drug Administration
FGFR    Fibroblast Growth Factor Receptor
FIMM    Institute for Molecular Medicine Finland
GSK3    Glycogen synthase kinase 3
HA      Hemagglutinin
HPD     Highest posterior density
HUSLAB  Helsingin ja Uudenmaan sairaanhoitopiirin laboratoriokeskus
ICU     Intensive care unit
M1      Matrix protein 1
M2      Matrix protein 2
MCMC    Markov Chain Monte Carlo
NA      Neuraminidase
NAI     Neuraminidase inhibitor
NP      Nucleocapsid protein
NPA     Nasopharyngeal aspirate
NPC     Nuclear pore complex
NS1     Nonstructural protein 1
NS2     Nonstructural protein 2
PA      Polymerase activity protein
PABPII  Poly(A) binding protein II
PB1     Polymerase binding protein 1
PB2     Polymerase binding protein 2
PI3K    Phosphatidylinositol-3-kinase
PKR     Protein kinase R
RIG-I    Retinoic-acid inducible gene I
RT-PCR  Reverse transcription polymerase chain reaction
siRNA   Short interfering RNA
THL     Terveyden ja hyvinvoinnin laitos (National Institute for Health and Welfare)
TLR     Toll-like receptor
VLP     Virus-like particles
vATPase Vacuolar –type H⁺ - ATPase
vRNA    Viral RNA
vRNP    viral ribonucleoproteins
This Master’s thesis is based on the following publications:

1. Lakspere T., Tynell J. et al. Full-genome sequences of influenza A(H1N1)pdm09 viruses isolated from Finnish patients during 2009-2013, Genome Announcements, 2013 (submitted)

2. Lakspere T., Tynell J. et al. Genetic evolution of influenza A(H1N1)pdm09 virus (in preparation)

Review of the literature

Influenza virus

Influenza viruses are globally distributed pathogens in the family Orthomyxoviridae (name derives from their ability to bind sialic acid in mucoproteins). They are classified into six genera (influenza virus A, B and C and three other genera). Influenza type A and B cause annual epidemics while type C viruses are milder and not so prevalent. Three distinct genera are tickborne mammalian viruses (Thogotovirus), Atlantic salmon infecting genus (Isavirus) and recently discovered arboviral viruses (members include Quarantil, Johnston Atoll and Lake Chad viruses). Hosts of different genera include birds, various mammals and humans (Lambert, Fauci 2010). First isolates of influenza viruses in humans were collected in 1933 (Smith, Andrewes et al. 1933).

Type A influenza viruses are categorized based on sequences of two surface glycoproteins hemaglutinin (HA) and neuraminidase (NA). Up to this date 17 different HA (H1 to H17) and 9 NA subtypes (N1 to N9) have been identified (Lamb 2008, Briedis 2011). The World Health Organization (WHO) estimates the global disease burden to be up to 1 billion human infections, 3-5 million hospital cases and approximately 300 000-500 000 annual deaths (Girard, Cherian et al. 2005).

The outcome of influenza infection is a combination with viral, host and environmental factors resulting the infection to be either asymptomatic, mild, severe or lethal.

Structure of influenza viruses

Influenza virions are quasi-spherical or filamentous particles. The latter are responsible for cell-to-cell viral transmission while quasispherical virions may be part of human-to-human transmission of the virus. Diameter of the virus ranges between 80-120 nm. The viral particle consists of helical ribonucleoprotein complexes with viral membrane derived from host cell plasma membrane.
Influenza genome (total size 10-15 kbp varying in different genera) consists of eight distinct linear single-stranded negative sense RNA segments (from 890 to 2341 nucleotides long). Segments encode viral proteins: PB2, PB1-F1, PB1-F2, PA and PA-X (polymerase subunits), trimeric HA (hemagglutinin), NP (nucleoprotein), tetrameric NA (neuraminidase, also called sialidases), M1/M2 (matrix proteins) and dimeric NS1/ dimeric NS2 (non-structural proteins). Details of segments and encoded proteins are found in table 1. In addition, there has been studies on the identification of new influenza proteins like PA-N155, PA-182 which are N-terminally truncated forms of PA affecting the replication events in influenza infection (Muramoto, Noda et al. 2013).

<table>
<thead>
<tr>
<th>Gene segment</th>
<th>Length (nucleotide)</th>
<th>Translated protein</th>
<th>Protein size</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2316</td>
<td>PB2</td>
<td>759</td>
<td>Regulation of transcription</td>
</tr>
<tr>
<td>2</td>
<td>2317</td>
<td>PB1</td>
<td>757</td>
<td>Polymerase activity, cleaving of pre-mRNAs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PB1-F2</td>
<td>87-90</td>
<td>Enhancing immune cell apoptosis, localized in mitochondria</td>
</tr>
<tr>
<td>3</td>
<td>2192</td>
<td>PA</td>
<td>716</td>
<td>Regulation of transcription and replication processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PA-X</td>
<td>252</td>
<td>Function in modulating host response to infection</td>
</tr>
<tr>
<td>4</td>
<td>1752</td>
<td>HA</td>
<td>56</td>
<td>Major surface glycoprotein; sialic acid-receptor binding; antigenic determinant</td>
</tr>
<tr>
<td>5</td>
<td>1541</td>
<td>NP</td>
<td>498</td>
<td>Forms viral capsid; regulator of RNA synthesis</td>
</tr>
<tr>
<td>6</td>
<td>1432</td>
<td>NA</td>
<td>454</td>
<td>Major surface glycoprotein; antigenic determinant; cleavage of sialic acid</td>
</tr>
<tr>
<td>7</td>
<td>1002(MP)</td>
<td>M1</td>
<td>252</td>
<td>Interactions with M2 and envelope proteins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>97</td>
<td>Ion channel activity</td>
</tr>
<tr>
<td>8</td>
<td>865(NS)</td>
<td>NS1</td>
<td>230</td>
<td>mRNA processing in host cell; suppression of immune response; reduction of apoptosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS2</td>
<td>121</td>
<td>Nuclear export of nucleocapsids by interacting with M1</td>
</tr>
</tbody>
</table>

Table 1: Negative sense RNA segments and the corresponding proteins of influenza A virus. Lengths of the segments taken from reference strain A/Helsinki/100/2013(H1N1) (Briedis 2011, Yewdell, Ince 2012).
Replication cycle

Attachment and nuclear transport

Unlike most RNA viruses, Orthomyxoviruses replicate in the cell nucleus. The replication cycle begins when type I transmembrane protein HA binds to the receptors on epithelial cells containing sialic acid. Hemagglutinin functions not only in cell attachment but also acts in viral entry into the cell via not fully understood fusion pathway. Proteases cleave HA into 2 subunits leading to their activation. HA1 contains sialic-acid binding domain and is presented on the surface of the virus particle while HA2 is attached to the viral envelope and contains hydrophobic fusion peptide (Briedis 2011).

After successful binding, virions in general can enter the cell by four mechanisms: via clathrin-coated pits, caveolae, macropinocytosis or by non-clathrin, non-caveolae pathway. Clathrin-mediated internalization is the primary pathway for influenza type A viruses. Once internalized, the particles travel via endosomal vesicles where the drop of pH induces conformational change in hemagglutinin (Figure 1). This states the beginning of the uncoating of viral particles. The conformational change of HA2 N-terminus moves the fusion peptide out of HA hydrophobic area to incorporate into the endosomal membrane. Resulting structural shift of many HA proteins leads to the formation of a pore and release of virus particle into the host cell cytosol (Briedis 2011, Palese, Shaw 2001).

The release of viral ribonucleotide complexes (consists of viral RNA, NP and three polymerase proteins) is facilitated by M2 protein, highly selective transmembrane ion channel, which allows H⁺ ions to penetrate the membrane. M2 channel is formed by four parallel α-helices lining in a tetrameric fashion that creates a small pore in the viral envelope (Figure 1). Its main functional element is a conserved (present also in influenza B and C viruses) His-XXX-Trp motif (X is any amino acid). Decline in endosomal pH induces proton entry via M2 ion channel and weakens interaction of M1 with vRNPs leading to the disruption of protein-protein interactions. The timing of the release of vRNPs depends on pH-mediated transition of the HA molecules involved (Briedis 2011, Chizhmakov, Geraghty et al. 1996).

vRNPs contain viral RNA segments that are wrapped in a helical conformation with NP proteins (one subunit binds ~20 nucleotides of viral RNA). RNA polymerase protein PB1 is bound to the 13 nucleotide long conserved regions at both ends of the viral RNA. NP and polymerase proteins interact with cellular importin-α followed by the binding to nuclear pore
complexes (NPC) facilitating the entry of influenza virus genomes into the nucleus (Briedis 2011).

The overall cascade of the replication cycle is the following: viral genomes are transported to nucleus for transcription, mRNAs are exported to the cytoplasm for translation, subsequent proteins are transported back to the nucleus where they direct viral genome replication and form vRNPs which are again transported back to the cytoplasm via NPC for virion assembly at the plasma membrane (Briedis 2011).

**Transcription of viral RNA and translation of proteins**

In the host cell nucleus, negative-strand vRNA is used for the synthesis of both capped, polyadenylated viral mRNAs and full-length positive-strand RNAs, also known as complementary RNA (cRNA) (Figure 1). Viral genomes are transcribed using fragments of cellular mRNAs as primers for mRNA synthesis initiation. Viral PB2 protein recognizes 5’ cap structure on eukaryotic mRNAs and binds to it. PB1 acts as a nuclease, cleaving ~10 nucleotides of the cellular mRNA starting from the 5’ cap structure. This process is also called “cap snatching”. The capped RNA segment (with PB2 still attached) acts as a primer for transcription initiation. PB1 acts as an RNA polymerase adding ribonucleotides to the template RNA. This process is regulated by a cleaved RNA segment bound to the 3’-terminal U residue on genomic RNA. Unlike other RNA viruses, Orthomyxoviruses can make capped mRNA without capping enzymes but are dependent on operating cellular RNA synthesis (Briedis 2011, Palese, Shaw 2001).

Termination of transcription is a result of the bound PB1 polymerase stuttering on the 5’ terminal of the template vRNA poly(A) stretch which blocks further RNA synthesis (Briedis 2011).

Six out of eight transcribed mRNAs are immediately exported to the cytoplasm and translated by cellular ribosomes. Two mRNAs are further processed: mRNAs encoding matrix and nonstructural proteins contain splicing sites that are recognized by nuclear splicing machinery. The ratio of spliced M2 and NS2 mRNAs ratio to their unspliced counterparts is approximately 9:1 resulting in higher amounts of M1 and NS1 proteins. Viral mRNAs are translated to proteins by cellular ribosomes. Host cell factors contributing to viral mRNA translation eIF4A, eIF4e, eIF4G are required for the binding of viral mRNA to ribosomal subunits (Briedis 2011).
Replication of viral RNA genome

During viral genome replication full-length vRNA (-) and cRNA (+) strands are synthesized. The switching from primed (mRNA transcription) to unprimed (RNA genome replication) RNA synthesis depends on the presence of free NP protein which interacts with viral RNA polymerase complex and 5’ and 3’ ends of the viral genome. The balance between them depends on the amount of NP present in the nucleus, the levels which are dropped during genome replication (Sidorenko, Reichl 2004).

Unlike in viral transcription where RNA polymerase subunits PB1 and PB2 mediate the synthesis of viral mRNAs, PB1 and PA are required for genome replication. The latter copies the viral genome directly from 3’ terminus of the genomic segment without an intervening primer. Synthesized plus-strand copy is complexed with NP protein to form antigenomic RNAs (positive sense). The synthesis of cRNAs is not terminated at the poly(U) stretch, generating full-length plus-strand copy of the segment. The cRNA lacks both the 5’ capped primer and 3’ poly(A) tail. Both the genomic minus-strand RNAs as well as antigenomic plus-strand RNAs can be used as templates for further vRNA synthesis as well as assembly of vRNA complexes. (Sidorenko, Reichl 2004, Briedis 2011).

Export of viral genome from the nucleus

Cellular nuclear export machinery aided by viral adaptor proteins is necessary to export vRNP complexes out of the nucleus. These complexes are bound to NP protein via sugar phosphate backbone and are too large for transport via passive diffusion. Nuclear export begins when imported M1 protein binds to newly formed nucleoparticles. NS2 with its nuclear export signal binds to the matrix/ribonuclear complex. It acts as a signal transducer to lead the protein export system to transport vRNPs to the cytoplasm via nuclear pores (Neumann, Hughes et al. 2000, Briedis 2011).

Molecular virus and host cell interactions

Viral NS1 plays a central role in counteracting host cell processes that try to interfere with viral replication. NS1 protein has several functional domains. The non-RNA binding domain inhibits cleavage and polyadenylation specificity (CPSF) and poly(A)-binding protein II (PABII) proteins responsible for the efficient 3’ end cleavage and polyadenylation of host
mRNAs in the nucleus. The synthesis of viral mRNA is not affected because their polyadenylation occurs by a separate mechanism. Most cellular pre-mRNAs are degraded due to the cap-binding and endonuclease activity of influenza PB1 and PB2 proteins (Nemeroff, Qian et al. 1995, Palese, Shaw 2001, Briedis 2011).

dsRNA-binding domain of NS1 interferes with host pathways in influenza infections. NS1 blocks innate cellular antiviral responses by inhibiting pattern recognition receptor pathways (RIG-1 interferon pathway) and suppresses double-stranded, RNA-dependent protein kinase (PKR) and 2',5'oligo(A)synthetase/Ribonuclease L). NS1 also activates the phosphatidylinositol-3-kinase (PI3K)/Akt pathway leading to suppression of apoptosis (Briedis 2011).

PB1-F2 is a 90 amino acid long recently described protein that is translated from PB1 segment in an alternative reading frame. It resides in mitochondria and has been shown to overcome host immune response therefore enhancing pathogenicity. It is also believed that it increases apoptosis in host immune cells (Briedis 2011).

**Assembly of virions**

Viral HA, NA and M2 proteins are folded in host endoplasmic reticulum (ER) (HA and NA are also glycosylated) after which they are transported to the Golgi apparatus. Inside there, HA and M2 cysteine residues are palmitoylated, followed by the cleavage of hemagglutinin to HA₁ and HA₂ subunits by furin protease. Modified proteins are directed to the assembly site and assembled on the apical plasma membrane of polarized cells. They accumulate in the cholesterol rich membrane region named lipid rafts – believed to be the site of virion formation. Assembly and lipid raft association signals of HA and NA reside in their transmembrane domains. The mechanism behind the assembly of virions could be explained by the interactions of the cytoplasmic tails of HA, NA where M1 binds to serve as a docking site for cellular transport (Briedis 2011).

**Orthomyxovirus** virions contain one copy of each genome segment. Each copy is incorporated into a virus particle during viral budding. Viral proteins interact with specific vRNA sequences that are added to the formed viral particles one by one (Briedis 2011).
**Function of neuraminidase**

Neuraminidase is a type II transmembrane protein that acts as an antagonist to hemagglutinin by cleaving the terminal of N-acetyl neuraminic acid (sialic acid) from mucoproteins, cell-surface glycoproteins and glycolipids. NA reverses the binding of HA by cleaving the bound sialic acid residue. This relationship helps influenza viruses to bind and release abundant cell receptors in the respiratory tract to maximize infection efficacy (Briedis 2011).

**Figure 1: Influenza life cycle.** The virus binds to the sialic-acid containing receptor on the cell surface and is endocytosed. Low endosomal pH triggers a cascade that leads to viral membrane fusion and the release of the viral ribonucleoprotein complexes into the cytosol of the host cell after which vRNPs are transported into the nucleus via nuclear pore complex (NPC). Once in the nucleus, the negative-sense viral RNA (vRNA) is transcribed into messenger RNA (mRNA) by cellular own primer/oligonucleotide mechanism. Replication starts when a positive-sense copy of the vRNA (also complementary RNA or cRNA), is made and used as a template to produce more vRNAs. The viral protein complexes and ribonucleoproteins are assembled into viral particles. Abbreviations: **cRNP** – complementary ribonucleoproteins, **F2** – cleaved variant of PB1, also called PB1-F2, **HA** – hemagglutinin, **NA** – neuraminidase, **NS1** – nonstructural protein 1, **NP** – nucleocapsid protein, **NEP** – nuclear export protein, **ER** – endoplasmic reticulum, **mRNP** – messenger ribonucleoproteins, **vRNP** - viral ribonucleoproteins (Palese, Shaw 2001).
Human host factors in influenza replication

König et al. analysed cellular co-factors in early-stage influenza infection by whole-genome RNAi approach and identified 295 cellular proteins. These include proteins necessary for viral entry into the host cell: vacuolar ATPase (vATPase), four out of seven COPI complex subunits, FGFR and GSK3-beta. Other identified proteins include those responsible for virus replication inside the cell (proteases and CaM kinase II beta or CAMK2B) and the ones important in phosphatase activity, ubiquitination and kinase-regulated signaling pathways. Results showed that silencing of vATPase and CAMK2B had inhibitory effects on influenza virus replication. For successful virus replication, several host factors are needed (Konig, Stertz et al. 2010).

Immune response

Immunological innate cell mediated responses provide initial short-term protection against influenza infection. Main defense against influenza A virus infection is the production of chemotactic (RANTES, MIP-1-α, MCP-1, MCP-3 and IP-10), pyrogenic (fever inducing) pro-inflammatory (interleukins IL-1α/β, IL-6, IL-18 and TNF-α) and antiviral cytokines (interferons) in the site of infection by host leukocytes and epithelial cells. These factors are responsible for the onset of primary clinical symptoms. Important antiviral cytokines secreted are type I interferons (IFN-α/β) whose functions are to induce antiviral proteins such as protein kinase R (PKR) and to activate natural killer (NK) cells during the early stage of infection (Julkunen, Sareneva et al. 2001, Gazit, Gruda et al. 2006).

Immune response of respiratory epithelial cells is directly contributable to Toll-like receptor 3 (TLR3) expressed on pulmonary epithelial cells. These receptors sense double-stranded RNA which stimulates cytokine production. Primary innate immune response has also been shown to be dependent on TLR8 sensing single-stranded RNA (Guillot, Le Goffic et al. 2005).

In order to replicate vRNA in infected cells, influenza viruses counteract the innate immunity responses. One important protein for this is the viral NS1. For example downstream signalling pathways activating transcription factors IRF-3, NF-κB, and AP-1 leading to production of antiviral cytokines are prevented via by NS1. It binds to viral dsRNA therefore preventing recognition by innate immunity receptors (Briedis 2011, Donelan, Basler et al. 2003).
Cytokine secretion via innate immunity response is a prerequisite for the adaptive immune response. Lymphocytes are activated by co-stimulatory factors induced by the innate immune system (Kamps, Hoffmann et al. 2006, Pawelek, Huynh et al. 2012). Influenza infection eventually leads to both humoral (B-cell) and cellular immune (cytotoxic T-cell) responses. During humoral immunity B-cells produce antibodies (IgG and IgA) against viral surface glycoproteins HA and NA as well as against M and NP proteins. Cytotoxic CD8+ T-cells recognize and eliminate infected host cells (Kamps, Hoffmann et al. 2006).

Influenza viruses have also a strategy to counteract these adaptive immunity responses. Virus has important immune targets, five antigenic sites (antigenicity is defined as the feature of a substance or a gene to stimulate antibody production) determined by mouse monoclonal antibodies and defined as common Ca1, Ca2 and Cb and strain specific Sa and Sb. Influenza viruses mutate fast on these sites to avoid recognition by host antibodies, since 1918 these sites have shown most variable amino acid compositions due to antibody-mediated immune pressure. Many of these mutations have been associated with the enhanced viral replication (S183P, I191L) (numbering without signal peptide) and more virulent phenotype (D127E, S183P, D222G) (Strengell, Ikonen et al. 2011).

Characteristics of influenza virus induced diseases

Virus transmits in humans via the respiratory route by either coughing or sneezing. Infection can proceed asymptotically, be severe or even fatal. Deaths occur more frequently in elderly, infants and chronically ill patients and are increased by secondary infections. Symptoms of influenza infection are fever, headache, cough and sore throat, nasal congestion, sneezing and whole-body muscular pain. On rare occasions heart complications like myositis, myocarditis may arise, and also toxic shock syndrome and Reye’s syndrome may develop in patients diagnosed with influenza infection (Lamb 2008).

The virus replicates in the upper and lower respiratory tract mucosal cells causing the loss of ciliated epithelium and inducing systemic inflammation by the production of immune system signaling factors like interferons and cytokines. Strongest symptoms appear 2-3 days after infection and the virus is usually cleared from the system within a week. Primary infection in children lasts approximately 13 days (Lamb 2008, Briedis 2011).

Leading cause of severe morbidity and mortality in influenza cases is bacterial secondary infection (co-infections) resulting in enhanced pneumonic illness (Joseph, Togawa et al.
The most common causes for community-acquired pneumonia are *Streptococcus pneumoniae*, *Staphylococcus aureus* and *Haemophilus influenzae*. *S.pneumoniae* is mostly associated with influenza infections (Brundage 2006). In the year 2009 pandemic caused by A(H1N1)pdm09, 28% of the influenza cases in New York had bacterial co-infections, most common being *S.pneumoniae*, followed by *Staphylococcus pyogenes* (Lee, Wu et al. 2010).

**Genetic susceptibility and severity of influenza infection**

Influenza virus induced diseases can have different outcomes, however, the underlying factors determining the extent and severity are not fully understood. It has been suggested that viruses could be distinguished by “intrinsic virulence” (transmission efficacy or ability to cause symptomatic infection) (Dolin 2013).

There are many determinants that could affect the virulence, transmissibility and severity of influenza infection and contribute to more devastating outcome for the patient:

1. Highly virulent virus (e.g. H5N1, H7N9)
2. Fast replication of the virus
3. Excessive inflammatory response
4. Poor innate antiviral immunity
5. Bacterial secondary infection
6. Underlying chronic diseases

Determining the causative factors requires thorough understanding of the infection process and its outcomes (Julkunen, Ikonen et al. 2012).

One regulator of infection is host genetics. Identification of occurring polymorphism allows the assessment of populations with high-risk influenza infection. Knowledge on genetics can also provide targets for new therapeutics. There are some identified SNPs in host genomics that could increase susceptibility to influenza infection (table 1). Adverse outcomes can be associated with genetic polymorphisms that affects host immune responses (Keynan, Malik et al. 2013).
<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Polymorphism</th>
<th>Functional significance</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5 (chemokine-receptor 5)</td>
<td>Leukocyte chemotaxis in chemokine response</td>
<td>CCR5Δ32</td>
<td>Increased allele frequency in Canadian H1N1 Intensive care unit (ICU) patient cases</td>
<td>(Keynan, Juno et al. 2010)</td>
</tr>
<tr>
<td>KIR (killer-cell immunoglobulin like receptor)</td>
<td>Factor activating NK cells</td>
<td>2DL2/L3</td>
<td>Also shown in Canadian H1N1 ICU cases</td>
<td>(La, Czarnecki et al. 2011)</td>
</tr>
<tr>
<td>IFTIM3 (interferon-induced transmembrane protein 3)</td>
<td>Restricting influenza virus replication</td>
<td>rs12252 altered splice acceptor</td>
<td>Increased in hospitalized cases of A(H1N1)pdm09 in England and Scotland</td>
<td>(Boon, de Mutsert et al. 2002)</td>
</tr>
<tr>
<td>FcyRIIa, IGHG2 ( Immunoglobulin G Fc receptor subunit)</td>
<td>Complement activation</td>
<td>IGHG2 <em>n/</em>-n FcyRIIa-R131H</td>
<td>IgG2 subclass deficiency reported in association with severe H1N1 cases</td>
<td>(Zuniga, Buendia-Roldan et al. 2012)</td>
</tr>
</tbody>
</table>

**Table 1** Genetic polymorphisms associated with influenza type A H1N1 susceptibility and severity. *Note:* CCR5Δ32 was found in 55.6% of patients confirmed with severe H1N1 infection (Keynan, Malik et al. 2013).

Viral genetic modifications are another factor contributing to the severity of the disease. Many studies have reported genetic alterations in the virus genome that could cause more severe disease outcome. The mostly studied is the D222 polymorphism in HA protein (numbering without the signal peptide) being potential virulence marker associated with severe influenza infection worldwide - presumably due to the area’s binding affinity to α-2,3-linked sialic acid receptors present in human lung epithelial cells (Drews, Pabbaraju et al. 2011, Yasugi, Nakamura et al. 2012, Anton, Pozo et al. 2012). This substitution is in the antigenic site of hemagglutinin and other mutations have also been reported in the same area (Yasugi, Nakamura et al. 2012).

Mutations leading to antiviral resistance can also affect the outcome of the virus infection. H275Y and I223R mutations in neuraminidase protein have been shown to cause resistance to neuraminidase inhibitors (NAI). However, these are mostly found in immunocompromised patients (LeGoff, Rousset et al. 2012).
Treatment and prevention

Currently there are many antiviral drugs available that mostly target either NA or M2 proteins (Min, Subbarao 2010).

Oseltamivir (marketing name Tamiflu) is a selective neuraminidase inhibitor of the influenza viruses A and B. Upon exposure to the drug, the influenza particles aggregate on the surface of the host cell, thereby restricting the infection within the tissues and mucosal membranes (Figure 1) (McNicholl, McNicholl 2001). Oseltamivir is an ethyl-ester prodrug that results in an active form oseltamivir carboxylate [3R,4R,5S]-4-acetamido-5-amino-3-(1-ethylpropoxy)-1-cyclohexene-1-carboxylate phosphate (Lew, Chen et al. 2000).

Zanamivir (marketing name is Relenza) is a neuraminidase inhibitor consisting of 2-deoxy-2, 3-dihydro-N-acetylneuraminic acid. It is administered either intranasally or inhaled (Lamb 2008).

Peramivir is a FDA unapproved, neuraminidase inhibitor of influenza type A and B viruses. *In vitro* potency seems to be better than with zanamivir and oseltamivir carboxylate based on generally lower maximal effective concentration (EC(50)) (Sidwell, Smee 2002).

Amantadine (with the marketing name Symmetrel) is an inhibitor of M2 H⁺ proton-seletive ion channel. Rimantadine is a methyl derivative of amandatine. The use of ion channel blockers is not relevant because drug-resistant variants are globally abundant (Lamb 2008).

Alternative treatment options could arise from the knock-down studies identifying cellular proteins important in viral life cycle. Cellular targets for the treatment of influenza infections include viral particle binding, endosomal trafficking, viral replication inside the cell and vesicular transport (Figure 2) (Min, Subbarao 2010).
**Figure 2:** Viral and host cellular targets for anti-influenza drugs in the context of the replication cycle of influenza virus. Influenza A virus replication steps are shown in green. Cellular pathways essential for completion of the viral replication cycle are shown in red. Currently, the viral M2 ion channel protein and neuraminidase are the only two targets of influenza antiviral drugs (gray boxes) licensed by the US Food and Drug Administration. Adamantane drugs block M2 protein during uncoating of the virus. Zanamivir and oseltamivir target neuraminidase (Min, Subbarao 2010).

**Vaccine development**

Vaccination is the primary strategy for prevention and control of influenza virus. Even though both inactivated and attenuated vaccines are effective in the prophylaxis of influenza diseases, it is very important that the antigenic determinants match the circulating epidemic influenza viruses. Also important is the recipient’s age and health status. Inactivated vaccines can be divided based on their production method into egg-based, cell-based and with adjuvant. Their preventive efficacy is 60-80% in children and adults. Live attenuated vaccines are produced either using egg-based or cell-based method. Their advantage over inactivated vaccines is that they can be administered intranasally (Lambert, Fauci 2010).

Seasonal influenza vaccines are trivalent containing three viruses or their hemagglutinin proteins (influenza type A H3N2, influenza A H1N1 and influenza B strains). Hybrid and natural strains are combined in laboratory settings and adapted to grow in eggs (Matthews
Plasmid-based reverse-genetics is used to shorten the time frame of production (Hoffmann, Neumann et al. 2000). New technologies in vaccine development are necessary to increase the capacity in the event of a pandemic. Next generation vaccine production involves methods like the use of recombinant proteins (HA is cloned and purified), virus-like particles (vectors containing HA, NA and M1 genes are cloned and purified), viral vectors (use of “carrier” viruses that cannot cause disease), DNA-based vaccines (DNA plasmids containing viral segments) and universal vaccines (also known as “common-epitope” vaccine) (Lambert, Fauci 2010).

In Finland approximately 5.3 million vaccines were imported in the last pandemic years 2009-2010 (Lyytikainen, Kuusi et al. 2010).

Evolution and epidemiology

Influenza viruses cause annual epidemic (seasonal) outbreaks which are generally due to a subtype of a virus already circulating among people. Pandemics (outbreak of global proportions) however, are generally caused by novel subtypes and results in higher number of fatalities. One characteristic of a pandemic influenza subtype is that it needs to be easily transmittable among humans (Lyytikainen, Kuusi et al. 2010)

Genetic variability that results in a new virus strains emerges from antigenic changes. These evolutionary changes occur in two ways:
1) **Antigenic shift**- resassortment of segments due to mixture of different viral subtypes
2) **Antigenic drift**- accumulation of point mutations

Antigenic drifts result from adaption to neutralizing antibodies and are also called antibody escape mutants. These variants occur with the frequency of less than 1% per year. Antigenic shift however is more prevalent in the population (Briedis 2011, Lamb 2008).

Influenza virus pandemics and outcomes

First recorded global (and with the most fatalities) influenza viruses outbreak (also called the “Spanish Flu”) occurred in 1918-1919 causing death of at least 40 million people mostly due to secondary bacterial pneumonias (Figure 3) (Johnson, Mueller 2002). Detailed studies of the virus strain are limited due to the lack of sample material. However lung tissue sample obtained from 1918 suggest „unusual avian precursor“ for the virus. Eventhough the HA was
of avian origin, the 1918 influenza could still bind to human receptors (Reid, Taubenberger et al. 2004).

Asian influenza virus in 1957 (Figure 3) consisted of 3 segments from avian virus (H2, N2 and PB1) and other segments from previously circulating human virus. The global mortality was estimated to be 1-2 million deaths (Scholtissek, Rohde et al. 1978, Kamps, Hoffmann et al. 2006).

Hong Kong influenza virus (H3N2) in 1968-1969 had HA and PB1 segments of avian virus origin descending from H2N2 through antigenic shift (Figure 3). The acquisition of avian surface antigens allowed these viruses to escape from human immune response. The death rate was estimated to be the mildest with approximately 1 million deaths in the USA (Scholtissek, Rohde et al. 1978, Kamps, Hoffmann et al. 2006).

Figure 3: Recorded human influenza virus pandemics since 1885 (early sub-types inferred). Source: European Centre for Disease Prevention and Control (ECDC) 2009. Spanish influenza virus H1N1 pandemic in 1918 was caused by human- transmittable avian virus. Asian influenza subtype H2N2 in 1957 was a reassortant between H1N1 genes PB1, HA and NA. In 1968, another reassortant, H3 avian virus (segments PB1 and HA) and the remaining gene segments from an H2N2 human virus, emerged as a new pandemic strain (Hong Kong influenza; H3N2 subtype). Latest pandemic was caused by reassortant containing viral segments from human, duck and swine (subtype H1N1) (Nicoll 2010, Horimoto, Kawaoka 2005).

Influenza A(H1N1)pmd09

On June 2009, the World Health Organisation (WHO) indicated that a novel influenza virus had entered into pandemic stage (Briedis 2011). A(H1N1)pmd09 derives from the assortment of classical swine and avian influenza of North American lineage, human seasonal H3N2 lineage and Eurasian swine lineage therefore combining the genes of human, swine and avian origin (European Centre for Disease Prevention and Control 2009).
In Finland the first reported influenza A(H1N1)pmd09 infections in individuals returning from a trip from Mexico were diagnosed on May 10 2009. From May- July 2009 nearly 90% of infections, and in August 2009 ~60% of infections were diagnosed in individuals returning from abroad. By December 2009, the virus had spread all across the country (Ikonen, Haanpaa et al. 2010).

Basic epidemiological studies show that younger people (between ages 10-29) were more susceptible to the virus and nearly 80% of the cases were detected in patients under 30 years of age. This can be explained by the adaptive immune system in elderly acquired from previous influenza-like diseases (A(H1N1)v Ewgoi 2009, Dawood, Jain et al. 2009). There was equal distribution of influenza incidence in males and females (European Centre for Disease Prevention and Control 2009). The 2009 influenza pandemic had similarities with the previous pandemics. However, differences were found in the higher incidence of severe infections in children and pregnant women. Furthermore, residual immunity of elderly was shown to explain their lower mortality rates compared to other age groups.

Another uncommon feature of influenza infections compared to previous pandemics was the existence of sudden acute illnesses and deaths from acute respiratory disease syndrome (ARDS) (European Centre for Disease Prevention and Control 2009, Jain, Kamimoto et al. 2009). European Centre for Disease Prevention and Control (ECDC) has named risk groups for the severe infection related to A(H1N1)pdm09 pandemic in 2009. These include:

- People with chronic conditions such as cardiovascular, respiratory, metabolic (notably diabetes), renal and hepatic diseases, immunodeficiency or neuromuscular conditions
- Pregnant women
- Children under two years of age

The clinical features of mild and severe A(H1N1)pmd09 cases differ. Uncomplicated mild infections were often followed by gastro-enteric symptoms usually seen in seasonal influenza epidemics (Centre of Disease Control and Prevention 2009). In severe infections one notable feature was that the patients became very sick in a short time period. Co-infections (viral pneumonias and bacterial infections) were detected but less common in the pandemic cases than in seasonal ones. Usual cause of death in severe influenza infections was progressive organ failure (World Health Organisation 2009).

In Finland the peak of pandemic influenza A(H1N1)pdm09 cases occurred in 2009/2010 (Figure 4). The National Infectious Disease Register reported the highest incidence of A(H1N1)pdm09 in Northern-Finland where also the first cases of this strain were reported and lowest in the South-East region of the country, including Helsinki region. Before the
outbreak of A(H1N1)pdm09 in 2009, the most relevant influenza strain amongst the population was H3N2 (Lyytikainen, Kuusi et al. 2010).

A total of 7,669 laboratory-confirmed pandemic influenza A(H1N1)pdm2009 cases were identified in Finland from May 2009 until March 2010. The highest morbidity was amongst children, lowest in elderly (Lyytikainen, Kuusi et al. 2010). In the study by Ikonen et al. lower infection incidence in elderly was partly explained by the cross-reactive immunity when antibodies originating from earlier infections were discovered in their system (Ikonen, Haanpaa et al. 2010).

Figure 4: Yearly influenza epidemics in Finland. Strongest effect on the national Finnish health occurred in 2009 with the A(H1N1)pdm09 pandemic outbreak (graph from Ilkka Julkunen).

Genetic profiling has revealed mutations in the genome of A(H1N1)pmd09 that could have effect on the severity of the disease. One such is previously mentioned D222 found in hemagglutinin and statistically present more in severe cases than in mild ones (Ikonen, Haanpaa et al. 2010, Ruggiero, De Rosa et al. 2013).

To date over 40 isolates of the pandemic virus A(H1N1)pdm09 resistant to oseltamivir have been reported. All the isolates have been susceptible to zanamivir. Amantadine is not used in the treatment of influenza infections because all the emerging strains contain resistance marker (S31N) for amantadine (World Health Organisation 2009).
Aim of the study

The aims of this study:

• Find and apply fast and precise method for sequencing large amount of virus isolates from patient samples
• Identify susceptibility factors for severe influenza A infections
• Describe the phylodynamics of influenza A(H1N1)pdm09 virus in Finland.
Materials and methods

Whole-genome sequencing of virus samples is necessary to investigate the full genomes of influenza viruses for mutational and evolutionary studies. This process was done at the Institute for Molecular Medicine Finland (FIMM) Sequencing Unit at the Faculty of Medicine, University of Helsinki, using Illumina technology based next-generation sequencing. Samples were obtained from patients diagnosed with influenza A(H1N1)pdm09. Viruses were propagated, RNA extracted and amplified and two 96-sampled libraries were prepared for sequencing.

Patient sample collection

Ethical approval for handling patient samples was provided by different institutions:

✓ Handling of native and recombinant influenza virus strains in BSL2 laboratory at FIMM (21/M/09).

✓ Sample collections were approved by the ethical review committee of the University Helsinki Central Hospital, Finland.

✓ Anonymized blood and NPA samples of influenza A virus-positive individuals were collected and patients gave informed consent that their samples can be used in scientific research (DNROs 165/13/03/00/00/2011 and 244/13/03/01/2012).

Samples (total of 125 mild and 28 severe A(H1N1)pdm09 samples, out of which 6 severe cases were sequenced previously and 20 seasonal samples) were obtained from Haartman Institute Helsingin ja Uudenmaan sairaanhoitopiirin laboratoriokeskus (HUSLAB) Unit and from Finnish National Institute For Health and Welfare (THL). Nasopharyngeal aspirates (NPA) were collected and stored as aliquots at -80°C at FIMM. Annotations and patient information for each sample were provided.

Propagation in MDCK cells

The yield of viral RNA in NPA samples is too low for sequencing, thereby propagating viruses in MDCK (Madin-Darby canine kidney epithelial cells) is necessary to amplify the viral RNA (cells were provided by collaborative laboratory Professor Richard Elliott, University of St Andrews, United Kingdom). In order to propagate viruses, MDCK cells need
to be prepared to form a confluent monolayer on culture vessel. This is achieved by
passaging cells on 96-well plates and T75 (the floor area of the flask is 75cm²) flasks. For
this, MDCK cells were washed once with PBS followed by trypsin treatment at 37°C for few
minutes to detach cells from the bottom of the flask. Detached cells were suspended in
Dulbecco’s modified Eagle’s growth medium (DMEM Eagle; GIBCO, Carlsbad, California)
supplemented with 10% fetal bovine serum (FBS; GIBCO, Carlsbad, California), 4 mM L-
glutamine and 100 units/ml of penicillin/streptomycin antibiotic solution to support cell
growth ad prevent bacterial growth. For 96-well plates the volume was 100µl per well. With
cells diluted 1:10 (meaning 1 part of cell suspension and 9 parts of medium), it took 3-5 days
for them to reach confluency at +37°C humidified 5% CO₂/ 15% O₂ incubator. Confluency
was determined visually with light microscope (Leica DM Inverted Light).

For influenza virus propagation, 1:100 ratio of NPA samples in virus growth mediawas added
to the confluent MDCK cells. In order to support the growth of influenza viruses, N-p-Tosyl-
L-phenylalanine chloromethyl ketone treated trypsin (TPCK-trypsin) was added to make the
progeny virions infectious (cleavage of the HA so that it could bind to sialic acid present on
the surface of cells). Instead of FBS, bovine serum albumin (BSA) was added to viral growth
media (DMEM Eagle). To prevent the growth of bacterium, fungi or mycoplasma (general
contamination) 100 μg/ml normocin was added to the media, as well as antibiotics
streptomycin and penicillin.

Two days post infection, the efficacy of infection (replication efficacy indicated by infection-
induced cell death) was measured by quantifying the ATP (present in metabolically active
cells) with CellTitre Glo® (CTG) Luminescent Cell Viability Assay (Promega) using
manufacturer’s protocol. Smaller quantified result of ATP shows better viral infection on
MDCK cells. Survival of the cells was also visually monitored with light microscope (Leica
DM Inverted Light LED Fluorescence).

**RNA extraction**

High-throughput RNA extraction was done with RNeasy 96 kit (Qiagen) using
manufacturer’s protocol. This method enables fast and high-throughput extraction of viral
RNA from large number of samples using silica membranes and centrifugation system.
Before the lysis step of viral samples, 2 microlitres of proteinase K was added to each sample
to inhibit endogenous RNases. Proteinase K-treated samples were stored at -20°C to prevent
degradation of RNA. Viral genome extraction was proceeded and handled with care in a clean
environment to avoid any possible contaminations.

**RT-PCR**

Comparison of two-step reverse transcription polymerase chain reaction (RT-PCR) and one-step RT-PCR was conducted to test which method amplifies all of the 8 viral RNA (vRNA) segments equally (meaning with the same concentration of each segment). The difference of methods is in the number of stages: with one-step RT-PCR reverse transcription and PCR stage are done in the same tube. In two-step RT-PCR they are conducted separately. For one-step method, Invitrogen SuperScript™ One-Step RT-PCR system (Life Technologies) was used and for two-step method different solutions and enzymes were purchased from Life Technologies. T7 universal primers (5’- TAATACGACTCACTATAGGG-3’) were chosen for two-step PCR reaction as universal primers for full-length amplification of all segments, whereas primers from Watson *et al.* were used for one-step RT-PCR reaction (see below). For two-step RT-PCR 1000U Revert Aid reverse transcriptase was used with the strategy described by Baillie *et al.* (Baillie, Galiano et al. 2012). After primary tests, one-step RT-PCR was chosen for viral RNA reverse transcription and following amplification due to better amplification results. After optimization the final experiment was conducted with the following conditions:

<table>
<thead>
<tr>
<th></th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reverse transcription</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>94°C</td>
<td>2 min</td>
</tr>
<tr>
<td></td>
<td>42°C</td>
<td>15 min</td>
</tr>
<tr>
<td></td>
<td>55°C</td>
<td>15 min</td>
</tr>
<tr>
<td></td>
<td>60°C</td>
<td>5 min</td>
</tr>
<tr>
<td><strong>Denaturation</strong></td>
<td>94°C</td>
<td>2 min</td>
</tr>
<tr>
<td><strong>Primary amplification</strong></td>
<td>Denaturation</td>
<td>94°C</td>
</tr>
<tr>
<td></td>
<td>Annealing</td>
<td>45°C</td>
</tr>
<tr>
<td></td>
<td>45°C-68°C</td>
<td>46 sec; 0,5°C/sec, slow ramp</td>
</tr>
<tr>
<td></td>
<td>Extension</td>
<td>68°C</td>
</tr>
<tr>
<td><strong>Secondary amplification</strong></td>
<td>Denaturation</td>
<td>94°C</td>
</tr>
<tr>
<td></td>
<td>Annealing</td>
<td>57°C</td>
</tr>
<tr>
<td></td>
<td>Extension</td>
<td>68°C</td>
</tr>
</tbody>
</table>
Mixture:

- RNase-free water: 13.8 µl
- RNA: 10 µl
- Primers: 0.1 µl (each)
- Buffer: 25 µl
- Enzyme mix: 1 µl

The enzyme mix contains High-Fidelity Platinum® Taq DNA polymerase and the buffer doesn’t have magnesium salts in it. The mixture was prepared as a mastermix.

Amplification was done with two sets of primers: one with common influenza A primers named common_uni12 (5’-GCCGGAGCTCTGCAGATATCAGCRAAAGCAGG-3’) and common_uni13 (5’-CAGGAAACAGCTATGACAGTAGAAACAAGG-3’) that favorably amplify longer vRNA segments (PA, PA-X, PB1 and PB2) and one with common influenza A primers named common_uni12G (5’-GCCGGAGCTCTGCAGATATCAGCGAAAGCAGG- 3’) and common_uni 13 that amplify shorter vRNA segments (NA, NP, HA, M1/2, NS1/2). The primer sequences cover the conserved sequences on both ends of the viral segments. After amplification step the resulted two mixtures were combined (Watson, Welkers et al. 2013).

After RT-PCR all the samples were analyzed on 1% agarose gel on electrophoresis (120V) and visualized in a UV-chamber with ethidium bromide which incorporates into DNA. Proper precautions were used when handling ethidium bromide to minimize its harmful cancerogenous effects. Successfully amplified samples were purified from PCR mixtures, rechecked on 1% agarose gel on electrophoresis (120V) and submitted to FIMM Sequencing Unit.

**Library preparation for sequencing**

For sequencing all the samples were reverse transcribed into cDNA (complementary DNA) following dsDNA (double stranded DNA) synthesis using OneStep RT-PCR technology described previously. Libraries were prepared using Illumina Nextera® DNA Sample
Preparation Kit for generation of clustered paired-end libraries from genomic DNA. In total two libraries were generated in two 96-well plates. In general, the workflow was conducted using manufacturer’s protocol as following:

Samples were randomly fragmented and tagged with specially engineered transposomes. Tags are necessary to amplify the fragments in limited-cycled PCR reaction. After completion the fragments also contain primer sequences and indices. Ligated sequence markers on both ends of the fragments were added in order to identify and concatenate them later. The markers for each library were generated with the dual indexing strategy with different mixture of indexes N701-N712 and N501-N508 so that each sample in one library would have one index complex. Libraries were validated and concentration measured using PicoGreen dsDNA Picogreen assay (Invitrogen) which identifies the concentration with DNA-specific fluorescent dye method. Samples were placed on Illumina chip where the adaptors ligated on both ends of the fragments bind to the inside of the cell channels present on the chip. Upon completion the samples were demultiplexed (Illumina Proprietary, 2012).

**Sequencing**

Total of 140 samples were sequenced with Illumina HiSeq 2000 sequencer using Illumina deep sequencing technology at FIMM Sequencing Unit. Total runtime was 14 days, afterwards demultiplexing was performed by FIMM Bioinformatical Unit. Demultiplexing consisted of mapping the reads using GenBank reference strain (A/California/07/2009) and identifying variants using institute’s own developed Variant Calling Pipeline (VCP). This method identifies all single nucleotide polymorphisms compared to the reference strain with the statistical values showing the probabilities of different nucleotides in that specific position. For further analysis the most dominant variants were chosen.
**Bioinformatic analysis**

**Mutation analysis**

Sequences were rendered and concatenated to visualize whole-genomes using Ugene Unipro software. The order of the concatenated sequences was the following: hemagglutinin (HA), matrix protein (M2/M1), neuraminidase (NA), nucleocapsid protein (NP), nonstructural protein (NS2/NS1) and polymerase subunits (PA, PB1 and PB2). Sequence alignment was performed with commercially available ClustalW software: for A(H1N1)pdm09 isolates GenBank reference strain (A/California/07/2009) isolated from patient with severe disease, and for 10 seasonal samples GenBank reference strain (A/Sydney/5/1997) were used. Another alignment was performed with the translated sequences using Ugene UniPro software. Annotations and translated sequences were obtained from Influenza Virus Resource package annotation tool ([http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html](http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html)) provided by National Center for Biotechnology Information (NCBI). This tool is takes into consideration the correct reading frame for influenza genomes as well as the overlapping open reading frames in these viruses. Amino acid sequences of severe and mild cases were compared and cumulative results visualized as a histogram to determine possible correlative polymorphisms (appendix 1).

Interesting amino acid substitutions that could alter virulence were visualized using PyMol software downloadable online. The latest full structures of the proteins were downloaded from protein databank ([http://www.pdb.org/pdb/home/home.do](http://www.pdb.org/pdb/home/home.do)). The models were analyzed using Tracer program following Bayesian statistical parameters. The definitions for the statistics are (Cummings 2013):

- **Mean** – The mean value of the samples (excluded the burn-in).
- **Stdev** – The standard error of the mean.
- **Median** – The median value of the samples (excluding the burn-in).
- **95% HPD Lower** – The lower bound of the highest posterior density (HPD) interval.
- **95% HPD Upper** – The upper bound of the highest posterior density (HPD) interval.
- **Auto-Correlation Time (ACT)** – The average number of states in the MCMC chain that two samples have to be separated by for them to be uncorrelated (i.e. independent samples from the posterior).
- **Effective Sample Size (ESS)** – number of independent samples that the trace is equivalent to.
Phylodynamic analysis

To study the phylodynamics of H1N1pdm09 in Finland, phylogenetic analysis was done with BEAST program v.1.7.5 which uses the set of probability distribution algorithms called Markov Chain Monte Carlo (MCMC). The goal is to estimate posterior density under assumptions to the likelihood of the observed viral population and phylogenetic data (Ratmann, Donker et al. 2012).

In order for the program to generate the phylogenetic tree in relation with divergence times XML file was generated from the NEXUS alignments (exported from ClustalW). The generated file needs to contain the options of MCMC analysis and build proper evolutionary model. Generated input alignment was accomplished with BEAUti (Bayesian Evolutionary Analysis Utility) v1.7.5 program (Drummond, Rambaut 2007). Tip dates for the tree were set with the dates of the virus collection from patients. Sites were set with the FLU substitution model to generate protein phylogenetic tree and HKY substitution model to generate nucleotide alignment phylogenetic tree. FLU and HKY were chosen because only one organism (influenza virus) is present in this alignment. Four categories of gamma site heterogeneity model were set to allow rate variation between sites in the alignment (Goldman, Whelan 2000). A strict clock model (uncorrelated log-normal) was chosen because of influenza virus’es characteristically fast mutation rate. The initial value, shape for gamma prior clock rate was 0.01 with the scale of 1000 by taken into consideration the exponential growth in the model. Clades were determined by the following criteria (according to WHO):

- Sharing a common node
- Mean percentage distances between and within clades is 1.5%

Phylogenetic tree was annotated using the BEAST package program TreeAnnotator with burn-in of 100 due to high number of trees (1 000 000) generated meaning the first 1% of trees were eliminated. Phylogenetic tree was rendered and visualized using FigTree software. For a graphical output of the diagnostic information as well as summary of distributions of the tree Tracer program was used.

In order to unmask the evolutionary dynamics of influenza, Bayesian skyline plot using MCMC set of algorithms (to estimate a posterior distribution) with a strict clock model was used for its feature to fit a wide range of demographic scenarios. The difference between other methods and Bayesian skyline plot is that the latter includes credibility intervals for the
estimated population size at every time point. The goal is to determine tMCRAs (time of most common recent ancestry) or in other words the divergence dates of clades using FLU substitution model including gamma site heterogeneity model with four categories. Coalescent tree prior used constant Bayesian Skyline model in a piecewise manner (Drummond, Rambaut et al. 2005). XML file was generated using BEAUti v1.7.5 program and BEAST v1.7.5 for analysis for the output for Bayesian skyline plotting. The chain length for both tree generation and Bayesian Skyline was 10 000 000 to allow deep optimization and to reduce the likelihood of high error rate. Tracer program enabled visualizing the results by plotting them in a time wise manner.

During primary analysis of the phylogenetic trees and Bayesian plots, it became obvious that in addition to concatenated versions the clusters and mutation rates (interpreted as the substitution rate or the clock rate) of both hemagglutinin and neuraminidase were needed. To visually combine the Bayesian Skyline plots and phylogenetic trees, they were aligned to the same timescale and the start of the evolutionary change was highlighted in pink in every tree (appendix 2a-d).

All the programs (except Ugene Unipro) used in the phylodynamic analysis were obtained from the internet as freeware, no licences were required.
Results

Propagation of viruses

For influenza viruses, the infection cycle (new viruses produced in a cell) is approximately 12-18 hours (Sidorenko, Reichl 2004). Virus propagation was optimized in MDCK cells. The best virus yield was obtained when confluent MDCK cell were infected with 1:100 dilution of NPA sample in virus growth media and incubated for two days. Successful virus replication in cells was observed by visualizing cytopathic effect due to cell death by virus production. In addition, cell death was quantified with measurement of ATP content of infected cells. All NPA samples were grown on MDCK cells and supernatant from 4th passage was collected and used for amplification of viral RNAs.

One-step versus two-step RT-PCR methods

In order to identify the best method to amplify all of the eight influenza virus segments in equal amounts, two distinct RT-PCR methods were examined and optimized. One-Step RT-PCR proved to be more effective method than two-step RT-PCR. Using the primer pair common_uni12G and common_uni13 the long viral segments were equally amplified with the smaller ones and all 8 segments were visible on agarose gel when amplified with one-step RT-PCR (Figure 5). One-step RT-PCR method enabled the successful amplification of 84% of all the samples provided by HUSLAB and THL.

![One-step versus two-step RT-PCR methods](image)

Figure 5: Reverse-Transcription Polymerase Chain Reaction. One-step RT-PCR yielded better results than
two-step RT-PCR method. For one-step RT-PCR common_uni12, common_uni12G and common_uni13 primers were used. For two-step RT-PCR oligos for PCR step were T7 conserved primers. **Abbreviations**: PB1, PB2 and PA – polymerase subunits, HA – hemagglutinin, NP – nucleocapsid protein, NA – neuraminidase, M – matrix protein, NS – nonstructural protein, RT-PCR – reverse-transcription polymerase chain reaction, wt – wild type.

**Sequencing of the isolates**

Of the 140 samples submitted for sequencing 135 (including 6 previously sequenced samples) were used in bioinformatical analysis. The average reads per isolate was 3686912.27 with the read length of 100bp. All of the reads were trimmed and mapped with the reference strain (A/California/7/2009) using BWA software which reduced the number of reads by 46%. Variants detected with the institute’s own developed pipeline (VCP) had on average 180 variants per isolate. Sequences of A(H1N1)pdm09 were submitted to Genbank (accession numbers KF559358 - KF560309) and GISAID databases (GISAID Isolate ID-s: EPI_ISL_145286-145302). Previously submitted six A(H1N1)pdm09 (accession number JQ409139.1-JQ409246.1, JQ409131.1-JQ409238.1, JQ409123.1-JQ409230.1, JQ173145.1 - JQ173152.1, JQ173153.1 -JQ173160.1, JQ173161.1-JQ173168.1) were included into subsequent analyses. Seasonal isolates (10 samples in total) were submitted to GenBank.

**Phylogenetic analysis**

Tracer program was used to analyse both the phylogenetic trees and Bayesian Skyline Plots. Divergence was analysed on the basis of effective sampling size (ESS), the mean and credibility intervals with 10% burn-in. The statistical values of the substitution rates for the phylogenetic trees are shown in table 2. The difference between Bayesian probability interpreted by HPD (Table 2) and classical confidence intervals is that the latter enables the interval to cover the true parameter whereas Bayesian probability reflects a person’s subjective beliefs while modeling. Virus evolutionary dynamics can best be shown with Bayesian modeling (SAS Institute Inc 2009).

As shown in table 2, hemagglutinin has the highest substitution rate followed by neuraminidase. Wide credibility intervals (HPD) indicate that mutation rate could be even higher. Nucleotide sequences have changed less than amino acid sequences over time describing their smaller effect on the viral population size.
The phylogenetic analysis revealed that all of the phylogenetic trees separated into 5 clades (A-E). Clustering did not follow the date of isolation that representatives of different clades were co-circulating (appendix 1). Interestingly, viral sequences obtained from severe cases claded somewhat apart from the mild cases. The high standard error of the models analysed is due to low ESS. ESS itself is calculated by the chain length divided by the ACT. The latter is the estimation of samples in the trace. Low ESS can be increased when making the data set more dynamic by including more mild cases from different years. The burn-in has been excluded in the calculation of the statistical parameters. The mathematical interpretation of possible errors can expressed as alpha values:

<table>
<thead>
<tr>
<th></th>
<th>Nucleotide</th>
<th>Protein</th>
<th>Hemagglutinin</th>
<th>Neuraminidase</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>$1.30 \times 10^{-04}$</td>
<td>$3.16 \times 10^{-04}$</td>
<td>$1.66 \times 10^{-03}$</td>
<td>$1.59 \times 10^{-03}$</td>
</tr>
<tr>
<td>Standard deviation of mean</td>
<td>$3.29 \times 10^{-05}$</td>
<td>$3.39 \times 10^{-05}$</td>
<td>$1.32 \times 10^{-04}$</td>
<td>$7.16 \times 10^{-05}$</td>
</tr>
<tr>
<td>median</td>
<td>$1.14 \times 10^{-04}$</td>
<td>$3.19 \times 10^{-04}$</td>
<td>$1.64 \times 10^{-03}$</td>
<td>$1.51 \times 10^{-03}$</td>
</tr>
<tr>
<td>Lower 95% highest posterior density (HPD)</td>
<td>$1.14 \times 10^{-04}$</td>
<td>$3.07 \times 10^{-04}$</td>
<td>$1.60 \times 10^{-03}$</td>
<td>$1.52 \times 10^{-03}$</td>
</tr>
<tr>
<td>Upper 95% highest posterior density (HPD)</td>
<td>$4.70 \times 10^{-05}$</td>
<td>$1.77 \times 10^{-04}$</td>
<td>$9.11 \times 10^{-04}$</td>
<td>$8.23 \times 10^{-04}$</td>
</tr>
<tr>
<td>auto-correlation time (ACT)</td>
<td>$2.98 \times 10^{-04}$</td>
<td>$4.46 \times 10^{-04}$</td>
<td>$2.59 \times 10^{-03}$</td>
<td>$2.58 \times 10^{-03}$</td>
</tr>
<tr>
<td>effective sample size (ESS)</td>
<td>$2.04 \times 10^{05}$</td>
<td>$1.78 \times 10^{05}$</td>
<td>$88076.3$</td>
<td>$1.51 \times 10^{05}$</td>
</tr>
</tbody>
</table>

Table 2 Statistical parameters of substitution rate/clock rate of the four Bayesian models for the creation of phylogenetic trees.

Table 3 Alpha values of the Bayesian models for phylogenetic trees. The critical level for alpha value according to Fisher is 0.05 meaning that showing false-positive rate must not exceed 5% (Fisher 1960).

Alpha values for the Bayesian Skyline Plots were all below 5% value indicating their low false-positive rate.
Whole-genome mutational analysis

The total amount of mutations in relation with the length of the protein (Figure 6b) showed that HA, NA and PB2 have highest mutation rate over the period of 2009-2013. However when compared to the length of the protein, HA and NA have mutated more than PB2. The comparison graph shows also that PB1 segment is mutated the least, as assumed by the distance of the regression line. In total 18 590 nucleotide polymorphisms and 4657 amino acid substitutions were identified compared with the reference strain A/California/7/2009 in 135 patient isolates. Eight virus isolates originated from four patients (two samplings per patient), differing in amino acid sequence and thereby indicating fast mutation rate.

When comparing polymorphisms occurring in 2009-2013 (Figure 6c) it seems that HA has accumulated more amino acid substitutions than other proteins followed by M1 (involved in replication of the virus), NA (functioning at production of infectious virions) and PB2 (viral polymerase activity).

The analysis of the 10 seasonal samples (subtype H3N2) revealed in total 1024 nucleotide and 238 amino acid substitutions when compared to the reference strain (A/Sydney/5/1997). Majority of the polymorphisms was detected in protein encoding sequences and especially in the neuraminidase gene. Interestingly 50% of seasonal sequences had a mutation in PB2 protein (E627K) that has been related to high virulence of avian influenza viruses (Tian, Qi et al. 2012).
Figure 6: Mutation analysis for the cumulative values. Chart 6a shows the total number of amino acid substitutions in each protein sequence (n=135). Chart 6b reveals the correlation between the length of the protein and the number of mutation. Chart 6c shows the polymorphisms accumulated from 2009‐2013. Abbreviations: HA – hemagglutinin, M1 – matrix protein 1, NA – neuraminidase, NP – nucleocapsid protein, NS2 – non-structural protein 2, PA/PA‐X – polymerase activity proteins, PB1/PB2 – polymerase binding proteins.

A manuscript was submitted to a journal called “Genome Announcements” about the evolutionary phylogenetic and mutational analysis of pandemic A(H1N1)pdm09 Finnish patient isolates.

**Structural analysis**

Mutations that could lead to more severe disease outcome often change the structure and
function of the specific protein. Severe clinical outcome of influenza infection is connected with mutations in the different binding sites of the proteins to either cell receptors or antiviral drugs (Pan, Cheung et al. 2010).

Polymorphisms in hemagglutinin accumulate at the antigenic site of the protein affecting host immune response and resulting enhanced viral influenza replication. Several amino acid changes in HA were identified (Figure 7a,b) and localized on the protein structure to the sites potentially correlating to the severity of the disease (Pan, Cheung et al. 2010, Glinsky 2010). Also on neuraminidase several amino acid changes occurred, while interestingly on mutation identified in this study could cause oseltamivir resistance due to its close proximity to H275Y mutation (Figure 7c) (Pan, Cheung et al. 2010). The polymorphisms in polymerase subunit (PB2) located on structure on active sites (Figure 7d) and could affect the cap-binding activity during virus replication inside the host cell.

![Figure 7: Mutations mapped in protein structures obtained from Protein Databank using PyMol software. a) and b) show polymorphisms mapped in the structure of hemagglutinin (HA) protein. c) shows the structure of neuraminidase and d) has polymerase protein (PB2). Arrows indicate the location of the amino acid substitutions that could correlate with the severity of the disease in the 3D structure of the protein. Orange spheres show other identified amino acid polymorphisms.]

**Antiviral resistance**

Analysis of the amino acid alignment revealed that amantadine resistance S31N was found in all samples. Resistance markers for remantadine or zanamivir were not detected. However one isolate (A/Helsinki/598/2013) contained an amino acid substitution (E119K) that has been shown to cause oseltamivir resistance in N2 subtypes (Figure 8a). This is the first such amino acid...
change observed in N1 subtype and also first reported potentially double resistant strain (amantadine and oseltamivir) in Finland. THL verified the presence of this mutation with Sanger sequencing. They identified very small amount of the polymorphisms in the original NPA sample (close to background noise), however the mutation was present in the majority of sequences obtained from the propagated (4th passage) isolate which was also used in the library preparation for Illumina sequencing.

The change in the reading frame of the neuraminidase gene segment resulted from guanosine change to adenosine leading to the amino acid change from glycine (E) to lysine (K) (Figure 8a).

Plaque assay of sample 598 from 4th passage of in vitro propagated viruses revealed the mixture of different viruses which can be identified by the different size of spots (Figure 8b). Structural prediction with PyMol shows that polymorphism 119 E/K alters the properties of the binding pocket and could result in ineffective binding of oseltamivir (Figure 8c).

The European Center of Disease Prevention and Control (ECDC) has been informed of this finding.

Figure 8: First potentially oseltamivir resistant A(H1N1)pdm09 strain identified in Finland. 8a) alignment of nucleotide sequences, change in the nucleotide sequence resulting in the amino acid shift E→K marked in a red box. 8b) shows the plaque assay with different titres of the virus. 8c) amino acid change affecting oseltamivir (structure in green) binding properties visualized with PyMol software.
Discussion

Sequencing methods in virological research

Molecular virology techniques for genomic analysis and the creation of recombinant viruses are important in understanding of influenza viruses. In order to amplify the viral genes or segments for further studies, viral RNA is converted to double-stranded DNA from negative-sense RNA using reverse transcription (RT) and PCR. The resulting amplicons can be probed, sequenced, or cloned into a variety of vectors for further analysis and for creating recombinant influenza A viruses by plasmid-based reverse genetics (Zhou, Wentworth 2012).

Up to this date many varieties of RT-PCR have been developed taking into consideration further testing of the amplicons. However, the major obstacle with this method has been that equal amplification of all the 8 viral segments is not often achieved. Equal amplification is extremely important for proper whole-genome sequencing. Of the available methods, multi-segment RT-PCR is useful because it doesn’t require in vitro propagation of viruses to increase the titer of the virus and is mainly used on diagnostic microarray studies (Zhou, Donnelly et al. 2009). When comparing two-step and one-step RT-PCR methods, one-step RT-PCR was found to be undoubtedly faster, cheaper and more efficient than two-step RT-PCR method. However, in order to achieve strong amplification of all eight segments, viruses from NPAs were first propagated in vitro in MDCK cells.

Since the development of sequencing methods that enable fast and large data sets to be processed, viral populations can also be analysed in unprecedented resolution. Massively parallel next generation sequencing provides such output. Illumina platform is leader in the sequencing market providing the cheapest, most time-efficient and easy to use service. The average run time is 11.5 days with the highest read length compared to other sequencing-platforms, 600GB. The negative side of this method is that it produces some sequence-related errors which increases the risk of missing important polymorphisms (Gullapalli, Desai et al. 2012, Watson, Welkers et al. 2013).

In this thesis work 160 patient nasopharyngeal isolates were grown and amplified with optimized one-step RT-PCR method for full-genome sequencing. In total 135 sequences (including the 6 previous samples) were used in bioinformatical analysis. This study validated an efficient one-step RT-PCR method in the pipeline of sequencing of large data (total runtime was 14 days).
Evolutionary changes of influenza virus in Finland

First whole-genome sequences of influenza A(H1N1)pdm09 were submitted to GenBank in 2012 by Medical systems virology research group lead by Dr. Denis Kainov at FIMM, University of Helsinki, Finland. Analysis of 6 sequenced genomes originating from severe cases revealed C-terminal truncation of NS1 protein in one of the sequences (A/Helsinki/P14/2009). Other significant polymorphisms were not identified. The purpose of the work in this thesis was to analyse full-length sequences from substantial amount of influenza A(H1N1)pdm09 isolates collected since the pandemic year 2009 and perform extensive evolutionary and mutational analyses on the data set obtained.

Bayesian modeling in virological research

Bayesian interference analysis with interpreting mechanisms using MCMC simulation algorithms is widely used method in viral phylodynamic analysis. Advantages of this method include:

- Combining past information to form a prior distribution for future analysis
- Small and large sample cohorts is processed similarly
- It follows the likelihood principle
- It enables the identification of statistical parameters such as distribution in 95% credible intervals

Disadvantages include:

- It requires skills to mathematically formulate the correct prior
- Posterior distributions are heavily influenced by the priors
- High computational cost

Markov Chain Monte Carlo provides the algorithms for sampling the posterior distributions and visualizing the results (SAS Institute Inc 2009).

Bayesian modeling on influenza A virus isolates obtained from Finnish patients

Phylogenetic analysis and Bayesian Skyline plots of 135 virus isolates showed that amino acid changes in viral proteins have stronger effect on the mutation rate of the virus than nucleotide polymorphisms in the corresponding genes. Regardless of high standard deviation and low
ESS (that is in correlation with standard deviation), the obvious trend is that the biggest change in the virus occurred sometime in 2010/2011. When viral proteins were analysed separately, neuraminidase and hemagglutinin had been mutating more than others. This could affect the virulence of H1N1 due to polymorphisms on antigenic sites of these proteins and their role in evading immune responses during the infection. Interestingly, NA had started evolving sooner than HA. This tendency could have been affected by the vaccinations which are targeting neuraminidase rather than hemagglutinin protein. As described by Strengell et al., two vaccinations were offered to Finnish healthcare professionals against influenza A(H1N1)pdm09 virus: monovalent AS03-adjuvant pandemic influenza vaccine (marketing name Pandemrix™) in October 2009 and non-adjuvant trivalent seasonal influenza vaccine (marketing name Fluarix™) in 2010. Both of the vaccines contained influenza subtype from severe case, A/California/7/2009 (Strengell, Ikonen et al. 2012). However when influenza virus A(H1N1)pdm09 mutation rate changes were compared to the national vaccination timings in Finland, it became apparent that there is no significant connection between increase in virus mutation rate and vaccinations.

Even though so far the studies on accumulating amino acid changes of A(H1N1)pdm09 strain have not caused antigenic distinction from the type A influenza strain used in the vaccine strain (A/California/7/2009) produced in 2009-2012 the differences in evolving influenza isolates compared to the vaccine isolate could be significant enough to change the antigenicity of the circulating influenza isolates (Anton, Pozo et al. 2012). As Guarnaccia et al. experimentally demonstrated, the fast antigenic drift occurred in virus to escape from a low level vaccine response thus in response to increasing infection-induced immunity and vaccinations antigenic drift is more likely to occur, pushing circulating strain to change even more (Guarnaccia, Carolan et al. 2013).

The results presented in this Master’s thesis suggests that the high substitution rate in protein sequences of circulating A(H1N1)pdm09 strain indicates the need for a new influenza A strain to be included in the upcoming vaccine.

**Viral polymorphisms affecting clinical outcome**

Many studies have identified possible mutations that could correlate with the severity of the disease. The most studied is the D222 polymorphism (numbering without the signal peptide) where substitutions from aspartic acid to glycine or to asparagine have been suggested to
lead to severe pathological outcome (Drews, Pabbaraju et al. 2011). In this thesis study no D222 substitutions were identified but polymorphisms in the close proximity sites were found (220S/T, 224S/N, 226K/E).

Polymorphisms found to be present in 2013 but not in 2009 seem to not have significant effect on the virulence of A(H1N1)pdm09. However, abundant amount of distinct mutations were discovered which could relate to the severity of disease. Polymorphisms in the polymerase subunits have shown to increase the pandemic potential of an influenza virus. Notably large amount of mutations (in total 1671 on 135 patient isolates) in the polymerase subunits identified in this thesis study indicates A(H1N1)pdm09 virus potentiality to escape from host barriers (Manz, Schwemmle et al. 2013). In addition, first reported double-resistant (oseltamivir and amantadine) strain is an alarming indication of changes occurring in the circulating A(H1N1)pdm09 that could have severe implication on general health of the Finnish population during the forthcoming epidemics.

Analysing only the virus genetics does not give complete overview on the effects of viral mutations on the health of individuals. Host genomics and underlying conditions contribute to the outcome of the disease and should be taken into consideration. In order to draw comprehensive conclusions, further studies on host immune responses to different virus isolates, in different genetic background, should be performed and combined with the medical data.

The role of E119 polymorphisms in NA in decreasing the susceptibility to antiviral drugs is showed by the study of Pan et al (Pan, Li et al. 2013).

**Conclusions and future prospects**

One-Step RT-PCR was shown to be the most effective way to equally amplify all 8 influenza H1N1 vRNA segments. The most rapid method for sequencing and analyzing large data set is next-generation sequencing. Influenza A(H1N1)pdm09 strain has evolved since 2009, with the most dramatic change occurring 2010/2011, possibly explaining re-emergence of this virus to the population on influenza season 2012/2013. Extensive polymorphisms were detected in virus isolates, especially in HA and NA proteins. The biological importance of amino acid substitutions needs to be assessed with functional studies.

To improve the statistical parameters (decrease in standard deviation and increase posterior densities) in evolutionary modeling of influenza viruses more dynamic data set is required. Obtaining more severe (hospitalized patients in the ICU unit) and mild cases (home treated)
from the years 2009, 2010, 2011 and 2012 would improve statistical significance and reduce the level of type I error (alpha).

In order to fully understand why some viruses affect one individual more than the other, host genomics and immunology need to be studied in parallel with virus genomics.
Acknowledgements

This Master’s thesis was done at the Institute for Molecular Medicine Finland (FIMM) at the Faculty of Medicine, University of Helsinki. I would like to thank Dr. Denis Kainov for offering me the opportunity to work on this influenza project. Also I’m very grateful to Dr. Laura Kakkola for amazing supervisory skills and the months full of patience and good words. Also I appreciate all the help and knowledge about conducting virological research both of my supervisors have given me.

My special thanks goes to Miko Valori for helping me with the bioinformatical analysis. I owe you a few litres of coca-cola.

I would also like to thank all my classmates and close friends for the moral support and writing this current thesis study.
References


57. RUGGIERO, T., DE ROSA, F., CERUTTI, F., PAGANI, N., ALICE, T., STELLA, M.L., MILIA, M.G., CALCAGNO, A., BURDINO, E., GREGORI, G., URBINO, R., DI PERRI, G., RANIERI, M.V. and GHISSETTI, V., 2013. A(H1N1)pdm09 hemagglutinin D222G and D222N variants are frequently harbored by patients requiring extracorporeal membrane oxygenation and advanced respiratory assistance for severe A(H1N1)pdm09 infection. Influenza and other respiratory viruses, .


Supplementary data

All the mutations located in mild and severe cases in reference to GenBank strain (A/California/7/2009) in appendix 1 and Bayesian Skyline Plots and phylogenetic trees of concatenated nucleotide and protein sequences as well as hemagglutinin and neuraminidase protein sequences in appendix 2a-2d.