



UNIVERSITA' DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biomediche Cliniche e Sperimentali

Dipartimento di Scienze biomediche per la salute

Corso di Dottorato di Ricerca in Sanità Pubblica
XXVI ciclo

***Sorveglianza delle infezioni respiratorie acute (ARI):
approcci innovativi per l'identificazione e la caratterizzazione
degli agenti virali coinvolti***

Tesi di:

Marianna MARTINELLI

Matricola: R09125

Tutor: Prof.ssa **Elisabetta TANZI**

Coordinatore del dottorato: Prof. **Alessandro ZANETTI**

Anno Accademico 2012/2013

Abstract	» 1
Introduzione	» 4
<u>1. Le infezioni respiratorie acute</u>	» 5
1.1 Epidemiologia	» 5
1.2 Manifestazioni cliniche	» 7
1.2.1 Le sindromi simil-influenzali	» 9
1.3 Agenti eziologici associati alle infezioni respiratorie	» 10
1.1.1 Funghi	» 10
1.1.2 Batteri	» 10
1.1.3 Virus	» 12
<u>2. I virus associati all'insorgenza di ARI</u>	» 13
2.1 Virus respiratori "classici"	» 14
2.1.1 I virus influenzali	» 14
2.1.2 I virus parainfluenzali	» 20
2.1.3 Virus respiratorio sinciziale	» 23
2.1.4 I picornavirus	» 27
2.1.5 I Coronavirus	» 32
2.2 Virus di recente identificazione	» 35
2.2.1 Human Metapneumovirus	» 36
2.2.2 Human Bocavirus	» 39
2.2.3 Human Polyomavirus Ki e Wu	» 44
2.2.4 Virus influenzale pandemico A(H1N1)pdm09	» 50
<u>3. Sorveglianza e diagnosi delle infezioni respiratorie virali</u>	» 53
3.1 Diagnosi di infezione virale	» 54
3.1.1 Campioni biologici di partenza	» 54
3.1.2 Isolamento virale	» 54
3.1.3 Saggi sierologici	» 55
3.1.4 Test rapidi	» 56
3.1.5 Saggi molecolari	» 56
3.2 Nuove tecniche di diagnosi virale	» 61
3.2.1 Next-generation sequencing	» 61
3.3 La caratterizzazione virale: i metodi filogenetici	» 66
3.3.1 Gli studi di evoluzione molecolare	» 67
3.3.2 Principali fattori in grado di promuovere l'evoluzione virale	» 68
3.3.3 La rappresentazione delle relazioni evolutive mediante alberi filogenetici	» 71
3.3.4 L'orologio molecolare	» 74

Disegno dello studio	» 76
<u>4</u> <u>Scopo</u>	» 77
<u>5</u> <u>Obiettivi specifici</u>	» 78
5.1 Identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo influenzale pandemico e post-pandemico (2009-2011)	» 79
5.2 Variabilità genetica ed evoluzione molecolare del virus respiratorio sinciziale (RSV)	» 79
Materiali e Metodi	» 80
<u>6</u> <u>Popolazione in studio</u>	» 81
<u>7</u> <u>Metodi</u>	» 84
7.1 Diagnosi molecolare	» 84
7.1.1 Tecniche di biologia molecolare per l'identificazione e la caratterizzazione molecolare dei virus influenzali	» 85
7.1.2 Tecniche di biologia molecolare per l'identificazione e la caratterizzazione molecolare di altri patogeni respiratori	» 91
7.2 Virus Discovery	» 92
7.2.1 Metodica VIDISCA-454	» 92
7.3 Sequenziamento genico del gene codificante per la proteina G del virus respiratorio sinciziale sottotipo A (RSV A) e B (RSV B)	» 100
7.3.1 Saggi biomolecolari per la determinazione della sequenza da analizzare	» 100
7.3.2 Rilevazione dei prodotti di amplificazione tramite elettroforesi su gel di agarosio	» 104
7.3.3 Purificazione e concentrazione dei prodotti di amplificazione	» 105
7.3.4 Sequenziamento genico del gene codificante per la proteina G	» 106
7.3.5 Analisi delle sequenze	» 107
7.3.6 Analisi filogenetica	» 108
7.3.7 Identificazione di sequenze ricombinanti	» 114
7.3.8 Studio evolutivo di RSV	» 115
7.3.9 Calcolo delle distanze sinonime e non sinonime	» 119
7.3.10 Analisi della pressione selettiva	» 120
7.3.11 Analisi dei siti di N- e O-glicosilazione	» 121
7.3.12 Analisi statistica	» 121

Risultati	» 122
<u>8 Identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo influenzale pandemico e post-pandemico (2009-2011)</u>	» 123
8.1 Indagine epidemiologico molecolare dell'infezione da virus influenzale A(H1N1)pdm09	» 123
8.2 Indagine epidemiologico molecolare delle infezioni sostenute da altri patogeni respiratori noti	» 124
8.3 Analisi molecolare mediante la tecnica di "virus-discovery" VIDISCA-454	» 126
<u>9 Variabilità genetica ed evoluzione molecolare del virus respiratorio sinciziale (RSV)</u>	» 128
9.1 Analisi delle sequenze del gene G di RSV A	» 128
9.2 Analisi delle sequenze del gene G di RSV B	» 131
9.3 Identificazione di sequenze ricombinanti	» 134
9.4 Stima dell'Evolutionary rate	» 134
9.5 Bayesian dated tree	» 135
9.6 Dinamiche di popolazione	» 139
9.7 Analisi sito specifica della pressione selettiva	» 141
9.8 Analisi dei siti di glicosilazione	» 143
Conclusioni	» 144
Bibliografia	» 153
Riassunto	» 162

Surveillance of acute respiratory infections (ARI): innovative approaches for the identification and characterization of viral agents

Introduction. Acute respiratory infections (ARI) are ubiquitous, air-borne transmitted and highly contagious infections, characterized by typical epidemic pattern. Though ARI causative agents may be bacterial or viral, viruses are by far the most common causes of ARI. In recent years, the global epidemiological scenario has been enlivened by the identification and emergence of many airborne pathogens (such as influenza viruses A/H5N1, A/H7N7, A/H7N9, coronaviruses SARS and MERS), which have been co-circulating along with the already known viral strains (i.e. seasonal influenza viruses, parainfluenza viruses, respiratory syncytial viruses, etc.), thus highlighting the public health concern about emerging infectious disease.

Objectives. The project aimed at applying new molecular biology techniques, "virus discovery" methodology and bioinformatics analyses to investigate the etiology of ARI.

The specific objectives were:

1) Identification of viral pathogens responsible of severe acute respiratory infections (SARI) and acute respiratory distress syndrome (ARDS) during the pandemic and post-pandemic (2009-2011). On purpose, the proportion of SARI/ARDS cases and deaths due to A(H1N1)pdm09 infection and the impact of other respiratory pathogens were evaluated during the pandemic and post-pandemic period in Lombardy. Additionally, unknown viruses were investigated in those cases for which diagnosis remained negative by using VIDISCA-454 methodology, a "virus discovery" technique. This analysis was performed at the Laboratory of experimental virology, Academic Medical Center (AMC), University of Amsterdam, where I completed an internship under the supervision of Prof. Lia van der Hoek.

2) Evaluating the genetic variability and molecular evolution of respiratory syncytial virus (RSV). In order to reconstruct the origin and phylodynamic history of RSV, the genetic diversity and evolutionary dynamics of RSV A and RSV B identified in respiratory specimens collected from children aged ≤ 3 years hospitalized in Lombardy for ARI over six epidemic seasons (2006 to 2012) were analyzed.

Materials and methods

1) From October 2009 to December 2011, 206 respiratory samples were collected from patients (61.2% males, median age: 44.3 years) hospitalized for SARI/ARDS. Nucleic acids were purified by NucliSENS® easyMAG® (bioMérieux, France), and analyzed by real-time RT-PCR assay to identify influenza virus. The clinical specimens that resulted negative to influenza virus detection were then screened by real-time RT-PCR/PCR for a panel of respiratory pathogens (Respiratory MWS r-gene™ Real-time PCR, bioMérieux, France) to detect: RSV A and B; human metapneumovirus (hMPV) A and B; human rhinovirus (hRV) and enterovirus (hEV); adenovirus (AdV); human bocavirus (hBoV) 1-4; human coronavirus (hCoV) 229E, NL63, OC43, HKU1; human parainfluenza virus (hPIV) 1-4; *Chlamydomonas pneumoniae*; *Mycoplasma pneumoniae*. Cases resulted negative to all diagnostic assays were further investigated by VIDISCA-454 (virus discovery cDNA-AFLP) technique. This is a virus discovery method based on recognition of restriction enzyme cleavage sites, ligation of adaptors and subsequent amplification by PCR combined with high-throughput sequencing 454 FLX/Titanium system of Roche.

2) RSV A (n=23) and RSV B (n=12) sequences obtained from oro-pharyngeal swabs of RSV-infected children aged ≤ 3 years hospitalized for ARI from 2006 to 2012 were considered for molecular characterization. Sequences were obtained by multiplex-PCR to amplify a fragment of RSV G gene and several bioinformatic programs were used for the phylogenetic and phylodynamic analysis. Phylogenetic trees of RSV A and RSV B sequences were constructed by *MEGA 5* program using the Neighbor-Joining method to identify RSV genotypes circulating and a bootstrap re-sampling analysis was performed to test tree robustness. To clarify RSV variability, amino acid mutations analysis was performed, and potential N-glycosylation and O-glycosylation sites were predicted by *NetNGlyc 1.0* and *NetOGlyc 3.1* programs, respectively. To evaluate site-specific selection pressure, different Maximum Likelihood approaches were applied (*SLAC*, *FEL/IFEL*, and *MEME*) by *DATAMONKEY*. To assess the evolutionary dynamics of RSV A and B, dated trees and evolutionary rates were estimated by *BEAST* with a Bayesian Markov Chain Monte Carlo (MCMC) approach. This analysis allowed identifying the time of the most recent common ancestor (tMRCA).

Results and discussion

1) During the pandemic and post-pandemic different pathogens have co-circulated and were associated with clinical cases in the study, but influenza virus A(H1N1)pdm09 showed the greatest impact. Influenza A(H1N1)pdm09 virus was detected in 58.3% (120/206) of SARI/ARDS cases (61.7% males; 13.3% aged ≤ 5 years, 67.5% aged 6-64 years). A(H1N1)pdm09 was identified in 77.8% of fatal ARDS cases. The impact of respiratory pathogens other than A(H1N1)pdm09 was 19.4% (40/206) (65% males; 30% aged ≤ 5 years, 47.5% aged 6-64 years). The influence of other respiratory viruses was significantly lower (19.4% vs. 58.3%, $p < 0.0000001$): hRV/hEV were the most commonly detected viruses, but also A(H3N2) influenza virus has played a significant role. Forty-six (46/206: 22.3%) SARI/ARDS cases (including two fatalities) resulted negative to all diagnostic assays and were further investigated by VIDISCA-454 that revealed no sequence reads that could belong to a novel virus or viral variant; however it enabled the identification of one case of undiagnosed measles. VIDISCA-454 methodology proved to be a sensitive and specific method, successfully applied to the monitoring of viral respiratory infections. Anyway, nearly 22% of SARI/ARDS cases did not obtain a definite diagnosis. In clinical practice, great efforts should be devoted to improve diagnosis of severe respiratory infections and to reduce such "diagnostic gap". The advantage from relying upon more accurate diagnosis could benefit the patient - in term of receiving the more appropriate antiviral drugs -, and could provide more detailed information on viruses circulating in the community, thus making public health authorities aware so as to adjust their policies accordingly.

2) From phylogenetic analysis resulted that 3 RSV A sequences clustered in genotype GA2 and all the other isolates clustered with NA1 genotype. Compared to the reference strain, 31 amino acid substitutions were identified. Phylogenetic analysis of RSV B sequences showed that study sequences were included in BA genotype tended to cluster within a clade including also reference sequence BA4 and 8 amino acid substitutions were identified among all of them. Similar mean evolutionary rates for RSV A and RSV B were estimated, 2.1×10^{-3} subs/site/year (95% highest density probability (HPD): $1.7-2.5 \times 10^{-3}$) and 3.03×10^{-3} subs/site/year (95%HPD: $2.1-3.9 \times 10^{-3}$), respectively. The tMRCA for the RSV A tree root was 71 years (95%HPD: 60-85), suggesting an origin of the currently circulating strains back to 1940s. The study strains within clade NA1 shared a single significant internal node with an estimated mean tMRCA of 7 years ago (2005). The three GA2 strains had a significant tMRCA

dating back to 16 years ago (95%HPD = 14-18). The dated tree obtained with RSV B strains showed a temporal-structure similar to RSV A. In particular, the tree root had an estimated mean tMRCA of 55 years ago (1957). All the studying strains clustered within a significant subclade, having a tMRCA estimate of 9 years ago (2003), including also a single BA4 strain. The RSV A Bayesian skyline plots (BSP) showed a first pick of the number of effective infections in the second half of 1980s, followed by a decrease of transmission events ending in about 2005, when a sharp growth restored the original viral population size. The RSV B BSP showed a similar trend, with a decrease in the effective number of infections occurring between the mid-1980s and 1990s, followed by a rapid growth in early 2000s. The RSV A site-specific selection analysis identified 10 codons under positive selection and a total of 39 sites were found to be under negative selection. The RSV B codon-specific selection analysis identified only one positively selected site and 27 negatively selected. Different patterns of O- and N-glycosylation sites were found by analyzing RSV A and RSV B studied sequences. Phylogenetic and phylodynamic analysis permitted to better understand the natural history of the virus, even if RSV remains difficult to control due to its antigenic diversity. G protein variability may play a significant role in RSV pathogenesis by allowing immune evasion. It is important monitoring changes in sequences coding this protein to permit the identification of future epidemic strains and to implement both vaccine and therapy strategies. This study thus contributed to have a better knowledge on the molecular epidemiology of RSV in Lombardy and provided data for a comparative analysis with other strains circulating in other regions of the world.

Conclusions. This project showed that respiratory virology needs a constant update of diagnostic procedures and surveillance scheme, essential support in the study and monitoring of viral infections of both classic and, above all, emerging or newly identified viruses. The application of cutting-edge technologies can be a successful tool for public health to face such emerging threats.

Introduzione

1. Le infezioni respiratorie acute

Le infezioni respiratorie acute (*Acute Respiratory Infection*, ARI), sono patologie causate da differenti agenti eziologici, che colonizzano diversi livelli delle vie aeree causando un quadro clinico con sintomatologia a danno dell'apparato respiratorio [1]. Sono malattie ubiquitarie, altamente contagiose e colpiscono individui di tutte le età, con manifestazioni cliniche più severe in bambini al di sotto dei 5 anni e adulti sopra i 55 anni di età. A causa inoltre dell'ampia varietà di microrganismi, dotati di un'elevata variabilità genetica e antigenica, che possono causare infezioni del tratto respiratorio, le ARI possono manifestarsi più volte e in diverse forme nell'arco della vita [1, 2]. Le ARI rappresentano un rilevante problema sanitario in quanto costituiscono una delle maggiori cause di morbosità e di mortalità a livello mondiale, rappresentando infatti la prima causa di ospedalizzazione dei bambini nei Paesi industrializzati ed una delle maggiori cause di mortalità nei Paesi in via di sviluppo. Oltre ad essere un importante problema sanitario, le infezioni respiratorie sono la principale causa di perdita economica associata ad invalidità temporanee, quali assenze dal lavoro e da scuola, rappresentando il 63% del danno finanziario totale. Costituiscono inoltre la causa del 28% di tutte le visite pediatriche e del 21% dei ricoveri ospedalieri [1, 3].

1.1 Epidemiologia

Le sorgenti di infezione sono rappresentate da individui infetti che eliminano il patogeno per via respiratoria. La diffusione degli agenti eziologici avviene per via aerea diretta, attraverso secrezioni respiratorie contaminate quali tosse e starnuti (*droplets*), o indiretta, mediante veicoli contaminati da secrezioni respiratorie infette (*fomites*) e tramite contatto diretto. Ciò facilita la circolazione dei patogeni soprattutto in popolazioni con alti livelli di densità e in spazi chiusi [3, 4]. Le infezioni respiratorie presentano un andamento tipicamente epidemico, per via della loro elevata contagiosità e diffusione, durante il periodo invernale, con diffusione geografica e stagionale per lo più nelle aree a clima temperato. La frequenza maggiore delle epidemie si trova nelle stagioni freddo-umide, da ottobre a marzo, quando maggiore è il tempo trascorso in ambienti chiusi senza alcun ricircolo dell'aria ed è più

elevato l'inquinamento atmosferico [2, 3]. Un episodio epidemico in una determinata area geografica segue generalmente uno schema standard: esordio brusco (solitamente l'episodio è anticipato da casi sporadici), raggiungimento di un picco massimo di incidenza della malattia in circa 2-3 settimane, declino rapido e, dopo circa 5-6 settimane, spegnimento quasi completo del fenomeno. Generalmente dapprima si ha un incremento del numero di bambini affetti da patologie respiratorie e/o febbrili che si evidenzia con l'assenteismo scolastico; segue l'estendersi di manifestazioni alla popolazione adulta con un conseguente assenteismo al lavoro ed un aumento di bisogno di assistenza medica comprovato da un aumento del numero di visite mediche e di ricoveri ospedalieri [2, 4].

L'elevatissima frequenza delle infezioni respiratorie è dovuta a molteplici fattori: l'enorme varietà di agenti eziologici coinvolti, la grande diffusibilità e la breve durata dell'immunità seguente le forme morbose più comuni. Alcuni degli agenti patogeni responsabili di queste infezioni vanno incontro in maniera ciclica a delle variazioni antigeniche che interessano un determinante maggiore. Queste variazioni possono dare origine a pandemie, particolari forme di epidemie che oltrepassano i confini di una nazione, espandendosi nei diversi continenti nel volgere di poco tempo, con conseguente aumento dei tassi di morbosità e mortalità.

L'immunità acquisita dopo un'infezione respiratoria è abitualmente di breve durata, in quanto i più comuni agenti eziologici presentano scarsa immunogenicità. Per questo motivo sono possibili reinfezioni, da parte dello stesso agente eziologico, nel corso della vita e, nel caso dei soggetti in età pediatrica, nell'arco di una medesima stagione [3, 4, 5]. Le infezioni più severe, che colpiscono le basse vie aeree, hanno massima frequenza nel primo anno di vita e diminuiscono progressivamente con l'età, fino ad aumentare nuovamente negli anziani. I bambini risultano essere più soggetti ad infezioni delle basse vie aeree. Con l'avanzare dell'età, lo sviluppo di difese immunitarie diversificate rendono gli adulti meno suscettibili e l'infezione rimane generalmente localizzata [6]. Le categorie a più alto rischio di infezione e di sviluppare complicanze sono i bambini, i soggetti in età avanzata e pazienti immunocompromessi o che soffrono di disturbi cronici [6]. Dati del WHO evidenziano come le ARI siano la causa di morte nel 13% dei casi nella popolazione al di sopra dei 55 anni e rappresentino il 20% di tutte le morti nei bambini al di sotto dei 5 anni [1].

1.2 Manifestazioni Cliniche

Le infezioni respiratorie causano una varietà di patologie, dovute alla colonizzazione dei patogeni a vari livelli dell'apparato respiratorio: lungo le vie aeree, nel parenchima polmonare o sulle sierose pleuriche, causando un quadro clinico con sintomatologia prevalentemente a danno dell'apparato respiratorio. La porta d'ingresso per la successiva localizzazione dei patogeni sono le cellule epiteliali delle alte e basse vie respiratorie. La variabilità di pH, umidità e temperatura, che caratterizza le diverse aree del sistema respiratorio, lo rendono habitat ideale per svariati microrganismi. La maggior parte delle infezioni respiratorie mostrano sintomatologia blanda e sono autolimitanti; tuttavia, soprattutto in individui a rischio, possono causare patologie più severe con gravi complicanze.

A seconda del distretto interessato dall'infezione, la ARI possono essere distinte in: URI (Upper Respiratory Infection - Infezioni delle alte vie respiratorie) e LRI (Lower Respiratory Infection - Infezioni delle basse vie respiratorie) (**Fig.1**). Spesso, la sede di impianto e moltiplicazione del patogeno coincide con le prime vie aeree, che rappresentano appunto la porta d'ingresso del patogeno nell'individuo. Esse infatti sono in continua comunicazione con l'esterno e, oltre a veicolare agenti patogeni, sono la sede di ingresso di sostanze irritanti (fumo di tabacco, aria fredda, polveri inquinanti) che facilitano l'instaurarsi delle infezioni. Tra le infezioni delle alte vie respiratorie si possono annoverare riniti, faringiti, tonsilliti, laringiti e otiti medie. Le infezioni del basso tratto invece, che comprendono bronchioliti, bronchiti e polmoniti, sono generalmente più gravi e spesso sono caratterizzate dalla presenza concomitante di diversi tipi di microrganismi. Non è rara la presenza di quadri clinici misti, come ad esempio broncopolmoniti (BCP) o tracheobronchioliti [3]. A causa della vasta gamma e l'aspecificità dei sintomi e la grande varietà di agenti causali risulta difficile orientare la diagnosi clinica verso uno specifico agente eziologico. E' inoltre noto che lo stesso agente può causare quadri clinici diversi, a seconda delle caratteristiche immunologiche e dello stato di benessere del soggetto colpito, e che la medesima sintomatologia può essere causata da numerosi agenti. Per tale motivo la classificazione delle ARI, effettuata secondo i criteri di diagnosi clinica, non è immediatamente verificabile dal punto di vista eziologico [9].

L'età e le condizioni generali di vita dei soggetti (ambiente, inquinamento, malnutrizione, stato immunitario compromesso, patologie croniche eventualmente presenti), possono favorire l'insorgere di complicanze a seguito di un'infezione. Nei soggetti con patologie croniche, nelle persone anziane, negli immunocompromessi e nei bambini al di sotto dei 5 anni possono insorgere gravi complicanze, che portano ad un aumento delle ospedalizzazioni e dei casi di morte [1, 3, 4].

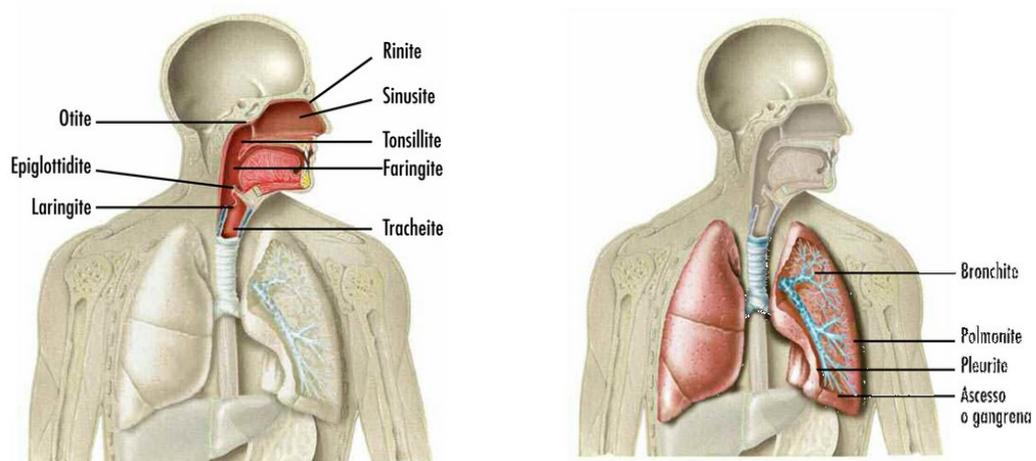


Fig. 1: *Suddivisione anatomica fra URI (Upper Respiratory Infection – Infezioni delle alte vie respiratorie) e LRI (Lower Respiratory Infection – Infezioni delle basse vie respiratorie).*

1.2.1 Le sindromi simil-influenzali

Una particolare menzione meritano le sindromi simil-influenzali (*Influenza-like illness, ILI*). La rete italiana di sorveglianza dell'influenza (*Italian Surveillance Influenza Network, InluNet*) definisce un caso di ILI come un brusco inizio di sintomi con febbre sopra i 38°C più almeno un sintomo sistemico (mal di testa, malessere, mialgia, sudorazione o brividi o fatica) e almeno un sintomo respiratorio (tosse e/o mal di gola) [7].

A differenza di molti agenti responsabili di ARI il virus influenzale conferisce al soggetto colpito un quadro clinico più specifico. Tuttavia, la diagnosi di influenza può essere offuscata dalla presenza concomitante di sintomi specifici (come ad esempio febbre, tosse, mal di gola, rinorrea) e non specifici (ad esempio disturbi addominali), che possono essere attribuibili ad altri agenti patogeni piuttosto comuni nella popolazione e spesso co-circolanti al virus influenzale. Inoltre, altri agenti eziologici possono essere responsabili dell'insorgenza di quadri clinici di ILI.

Per tali motivi è dato che nella pratica clinica la valutazione dei sintomi è la principale via utilizzata per la diagnosi, sindromi con quadri clinici sovrapponibili a quelli del virus influenzale vengono definite ILI, fino a che non viene effettuata una diagnosi confermata da dati di laboratorio. Pertanto risulta spesso necessario non solo confermare in laboratorio i casi di influenza, ma anche effettuare una sorveglianza di tutti gli altri agenti patogeni che possono essere causa di una sindrome simil-influenzale [8].

1.3 Agenti eziologici associati alle infezioni respiratorie

Gli agenti patogeni coinvolti nell'eziologia delle infezioni respiratorie acute possono essere di natura virale, batterica, protozoaria e micotica. Fra questi, i virus sono quelli dotati di una diffusibilità più elevata, essendo i patogeni più numerosi e dotati di una variabilità antigenica maggiore, mentre i batteri sono quelli caratterizzati da infezioni con maggiore mortalità, in quanto in grado di causare complicanze con maggior frequenza e associate ad un quadro clinico più grave.

1.3.1 Funghi

Le infezioni fungine, generalmente ad andamento subclinico, sono invece comuni in pazienti immunocompromessi, particolarmente riscontrate durante neutropenie severe, e in pazienti HIV-positivi. Le specie fungine più comunemente in causa sono *Candida* e *Aspergillus*. Le candidosi solitamente si limitano a infezioni del tratto aereo superiore ma possono in rari casi causare anche LRI, soprattutto in soggetti con AIDS conclamata. Fra le aspergillosi la specie più comunemente isolata è *Aspergillus fumigatus*. In pazienti affetti da asma o fibrosi cistica si può verificare anche una manifestazione polmonare, che colpisce anche in assenza di una grave compromissione immunitaria, nota come aspergillosi broncopolmonare allergica, manifestazione allergica nei confronti dei conidi inalati [9].

1.3.2 Batteri

Le sindromi più comunemente riscontrate associate ad infezione batterica dell'apparato respiratorio sono polmonite, faringite, bronchite e otite. Solitamente i batteri si impiantano su tratti della mucosa precedentemente infettati da virus, sebbene abbiano anche la capacità di infettare tratti dell'apparato respiratorio indenni. *Streptococcus pneumoniae* è l'agente batterico identificato con maggior frequenza (30-50% dei casi), seguito da *Haemophilus influenzae* tipo B (Hib) (10-30%), *Staphylococcus aureus* e *Klebsiella*

pneumoniae. Sono inoltre comunemente riscontrati come causa di ARI anche *H. influenzae* non tipizzabile, *Salmonella* spp. non tifoide, *Mycoplasma pneumoniae*, *Chlamydia* spp., *Pseudomonas* spp., *Legionella* spp., *Nocardia* spp. ed *Escherichia coli* [1]. Lo *Streptococcus pyogenes* risulta spesso essere un comune colonizzatore delle alte vie respiratorie e la sua presenza non implica necessariamente un ruolo patogenetico. Infatti è ben nota la presenza di portatori sani, la cui frequenza nella popolazione aumenta con il decrescere dell'età, fino a raggiungere il 20% nei bambini. Tuttavia, lo *Streptococcus pyogenes* è la principale causa di faringite (nei bambini con età compresa fra 5 e 10 anni il 25% delle faringiti è causata da tale agente microbico) e la sua ricerca colturale è di routine [10]. *S. pneumoniae*, Hib e streptococchi b-emolitici sono gli agenti che più frequentemente causano otiti medie, patologie che colpiscono con alta frequenza i bambini con età fino a 6 anni. E' stato stimato che *S. Pneumoniae* causi più di un terzo dei 2 milioni di morti a causa di ARI, soprattutto nei paesi in via di sviluppo, dove tale agente infettivo è uno dei più importanti patogeni di natura batterica dell'infanzia. Tale agente è prevalentemente causa di polmoniti acute sia in bambini che in adulti [1, 10]. La difterite è un'infezione acuta che coinvolge le alte vie respiratorie causata da ceppi tossigenici di *Corynebacterium diphtheriae*. Grazie alle campagne di immunizzazione di massa tale patologia è virtualmente assente nei paesi industrializzati [14]. Infine la tubercolosi o TBC (causata da *Mycobacterium Tuberculosis*) continua ad essere la principale causa di morte a livello globale; infatti si stima che circa un terzo dell'umanità sia stata infettata e che la TBC causi circa 1,7 milioni di morti ogni anno (circa 4650 morti ogni giorno) [1]. Grazie ad una serie di interventi preventivi di Sanità Pubblica tale patologia veniva considerata controllata e contenuta nei paesi industrializzati, nonostante continuasse ad essere endemica nei paesi in via di sviluppo. Tuttavia la creazione di nuovi flussimigratori, la creazione di nuove sacche di povertà (anche in paesi industrializzati) e l'aumento di patologie immunodepressive ne ha causato una nuova diffusione e oggi è considerata una patologia riemergente per quanto riguarda i paesi industrializzati. In alcuni casi possono essere causa di infezioni respiratorie acute anche agenti patogeni normalmente non coinvolti in queste patologie, quali *Enterobacteriaceae* (come ad esempio *E. coli* o *Proteus mirabilis*) e *Pseudomonas aeruginosa*. Questo ampliamento dello spettro eziologico è dovuto allo stato fisico dei soggetti: gli immunocompromessi, i bambini, gli anziani e i portatori di malattie croniche a carico

dell'apparato respiratorio risultano più suscettibili. Essi sono infatti considerati soggetti maggiormente a rischi di contrarre l'infezione e di sviluppare in forma grave la patologia [1].

1.3.3 Virus

I virus rappresentano gli agenti eziologici più diffusi nelle patologie del tratto respiratorio; attualmente, infatti, sono noti più di 200 sierotipi, ritenuti antigenicamente differenti, che possono essere responsabili di ARI. Le diverse famiglie principalmente in causa sono: *Orthomyxoviridae* (virus influenzali A e B), *Paramyxoviridae* (human parainfluenzavirus1-4, hPIV1-4; virus respiratorio sinciziale, RSV; human Metapneumovirus, hMPV), *Adenoviridae* (human Adenovirus, hAdV), *Picornaviridae* (human Rhinovirus o hRV, human Enterovirus o hEV, human Parechovirus o hPeV), *Coronaviridae* (human Coronavirus, hCoV-229E, hCoV-OC43, hCoV-NL63, hCoV-HKU1, SARS CoV e MERS-CoV) *Polyomaviridae* (KU-PyV e WI-PyV), *Parvoviridae* (human Bocavirus hBoV) [3, 5] (**Fig. 2**).

Nel paragrafo successivo verranno spiegate le caratteristiche di ciascuna famiglia in dettaglio.

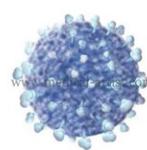
2. I virus associati all'insorgenza di ARI

Come accennato in precedenza i virus sono la causa principale nell'insorgenza di ARI. Sono patogeni di diversa natura in base alla quale riescono ad infettare diversi distretti dell'albero respiratorio provocando svariate sintomatologie. In base alla data di scoperta e alle conoscenze che si hanno a riguardo, si possono dividere in virus "classici" e di "recente identificazione".

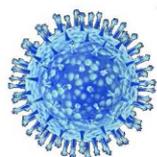
VIRUS A RNA



Influnza virus
 Orthomyxoviridae
 ssRNA-, 8 segmenti, 13-13,6Kb
 Envelope (HA, NA)
 3 tipi (A, B, C)



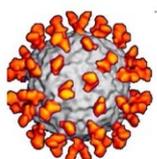
Human Parainfluenza virus (HPIV)
 Paramyxoviridae
 ssRNA-, 15Kb
 Envelope (HN,F)
 4 tipi (1-4)



Virus Respiratorio Sinciziale (RSV)
 Paramyxoviridae
 ssRNA-, 15Kb
 Envelope (F, G)
 2 tipi (A, B)



Human Metapneumovirus (HMPV)
 Paramyxoviridae
 ssRNA-, 13Kb
 Envelope (F, G)
 2 tipi (A, B)

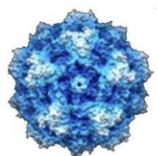


Picornavirus
 Picornaviridae
 HUman Rhinovirs (hRV), HUman Enterovirus (hEV), Human Parechovirus (hPeV)
 ssRNA+, 7,2Kb
 No envelope
 Circa 200 tipi

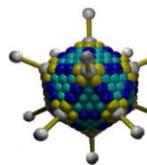


Human Coronavirus (HCoV)
 Coronaviridae
 ssRNA+, 30Kb
 Envelope (E1, E2, E3)
 5 tipi (229^e, OC43, SARS, NL63, HKU1)

VIRUS A DNA



Human Bocaviru (hBoV)
 Parvoviridae
 ssDNA-, 5Kb
 No envelope
 2 tipi (St-1, St-2)



Human Adenovirus (HAdV)
 Adenoviridae
 dsDNA, 30-35Kb
 No envelope
 Circa 50 tipi

Fig.2: Principali virus responsabili di ARI

2.1 Virus respiratori “classici”

Vengono definiti “classici”, i virus respiratori scoperti ormai da molto tempo, sui quali si sono fatti numerosi studi e si ha una conoscenza approfondita. I principali componenti di questo gruppo sono: i virus influenzali stagionali, il virus respiratorio sinciziale, i virus parainfluenzali e i coronavirus.

2.1.1 I virus influenzali

Struttura e organizzazione genomica

I virus dell'influenza appartengono alla famiglia delle *Orthomyxoviridae*, genere *Orthomyxovirus*; vengono classificati nei tre tipi: A, B e C, in base all'antigenicità delle proteine del *core*, ovvero gli antigeni maggiori NP (nucleoproteina) e M1 (proteina della matrice) [11]. Sono virus di forma sferica con un diametro di 80-120 nm, e hanno un *envelope* esterno costituito da un *bilayer* lipidico, dal quale sporgono due tipi di glicoproteine: l'emoagglutinina (HA, un trimero di 135 aa), e la neuroaminidasi (NA, un tetramero di 60 aa) (**Fig. 3**). L'**HA** è il principale determinante antigenico responsabile del riconoscimento e dell'adesione ai recettori specifici contenenti un residuo di acido sialico delle cellule target. L'HA media la fusione tra la membrana virale e quella cellulare, per permettere l'ingresso del *core* virale nella cellula ospite. L'**NA** premette l'ancoraggio del virus alla membrana ospite grazie alla sua attività neuroaminidasica. Inoltre ha un ruolo nella liberazione dei virioni maturi dalla cellula infetta e ne impedisce l'aggregazione al di fuori della cellula ospite. I virus in grado di infettare la specie umana sono caratterizzati da 3 tipi di HA (H1, H2, H3) e due tipi di NA (N1, N2). L'*envelope* circonda un *capside* costituito da proteine della matrice (**M1** e **M2**) che intervengono nell'assemblaggio del virus e nello stabilizzare la componente lipidica. La parte più interna è costituita dal *core nucleocapsidico* a simmetria elicoidale, in cui è contenuto il genoma virale: questo è costituito da 8 segmenti (7 nel sierotipo C) di RNA a singolo filamento a polarità negativa, associati alla nucleoproteina NP e a tre polimerasi (PB1, PB2 e PA) [11].

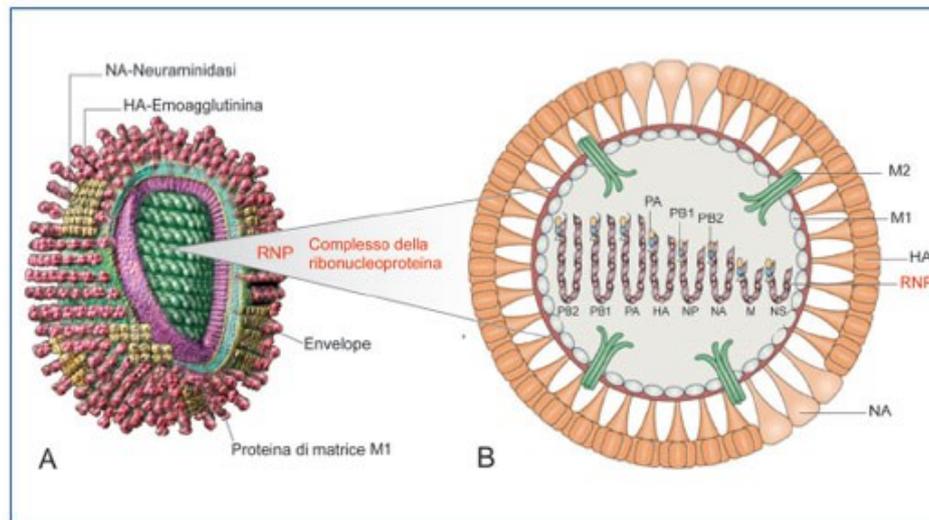


Fig.3: Struttura virus influenzale

Ciclo replicativo

Il ciclo replicativo ha inizio con il riconoscimento da parte del virus, tramite l'emoagglutinina, delle cellule target ovvero le cellule dell'epitelio respiratorio; in seguito avviene l'adesione alla cellula ospite attraverso le mucoproteine con residui di acido sialico e il successivo adsorbimento mediante endocitosi. Il pH acido delle vescicole provoca un cambiamento conformazionale delle HA tale da permettere la fusione dell'*envelope* con la membrana endosomica; in seguito il basso pH favorisce la dissociazione tra *nucleocapside* e proteina M1 all'interno dell'endosoma permettendo la liberazione del *nucleocapside* stesso nel citoplasma. Il genoma contenuto nel *core* raggiunge il nucleo della cellula ospite e qui va incontro a due processi: trascrizione e replicazione. Con la trascrizione viene prodotto mRNA virale. Questo ultimo fuoriesce nel citoplasma per essere poi tradotto dai ribosomi nelle nucleoproteine e negli antigeni HA ed NA (questi rimangono nel citoplasma o si associano alla membrana cellulare). Nel nucleo, durante la replicazione, vengono prodotte nuove copie del genoma virale che nel citoplasma si assoceranno alle nucleo-proteine neosintetizzate. Il nucleocapside, così assemblato, passa nel citoplasma verso la membrana, dove sono concentrate le proteine di matrice virale. Questi complessi vengono infine espulsi tramite esocitosi; la vescicola esocitotica con inglobati gli antigeni HA ed NA precedentemente associati alla membrana cellulare, costituirà il rivestimento del virione [12] (**Fig. 4**).

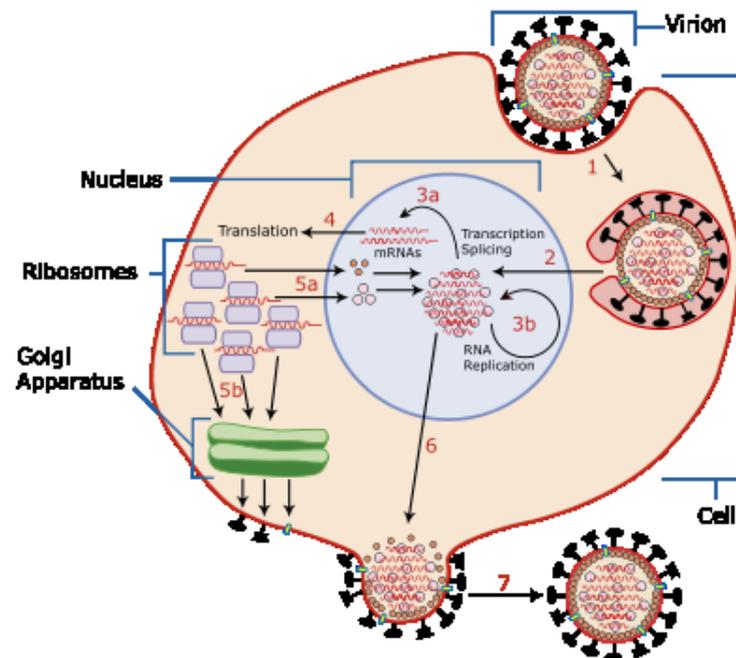


Fig. 4 : Ciclo replicativo del virus influenzale

Variazione antigenica dei virus influenzali

Il virus dell'influenza si distingue dalle altre infezioni respiratorie per due caratteristiche: la prima consiste nel passaggio dell'infezione direttamente dall'animale all'uomo; la seconda, e più importante, consiste nella variazione degli antigeni virali NA e, soprattutto, HA. Questa variazione avviene principalmente per il virus influenzale di tipo A ed è responsabile della formazione di nuovi ceppi virali verso i quali la popolazione non possiede ancora anticorpi. Per questo motivo l'influenza viene considerata la maggiore patologia a carattere epidemico. Due sono meccanismi in grado di generare variabilità antigenica: l'*antigenic shift* e l'*antigenic drift*.

L'*antigenic shift* è detto riassortimento antigenico maggiore. Esso si verifica solo nel virus di tipo A ed è responsabile degli episodi pandemici ogni 10-12 anni, verso i quali nessun soggetto della popolazione è immune. Alla base dell'*antigenic shift* vi è il riassortimento genico tra un ceppo animale e uno umano. Ciò è reso possibile dalla segmentazione del genoma virale, per la quale uno o più segmenti virali umani vengono sostituiti dal corrispondente segmento di origine animale. Il riassortimento può avvenire in un ospite intermedio, spesso rappresentato dal maiale. Se lo *shift* coinvolge simultaneamente sia gli

antigeni HA che quelli NA si può verificare una pandemia, cioè una particolare forma di epidemia che travalica i confini di una nazione, espandendosi nei diversi continenti nel volgere di poco tempo, con conseguente aumento dei tassi di mortalità. Se invece lo *shift* riguarda solo gli antigeni HA il nuovo virus causerà pandemia, ma sarà clinicamente meno grave.

L'*antigenic drift* invece coinvolge sia i virus di tipo A sia quelli di tipo B; esso si verifica nei periodi interpandemici, ogni 1-3 anni e genera nuovi ceppi nell'ambito dello stesso sottotipo. Le epidemie che si verificano quindi sono limitate nel tempo e meno gravi. L'*antigenic drift* è dovuto a ripetute mutazioni (anche puntiformi) che causano una o più sostituzioni aminoacidiche a livello degli antigeni HA o NA; si ha quindi una variazione nella struttura primaria dell'antigene, che può non venire più riconosciuto dagli anticorpi di una parte di popolazione, dando origine così ad epidemie locali (**Fig, 5**).

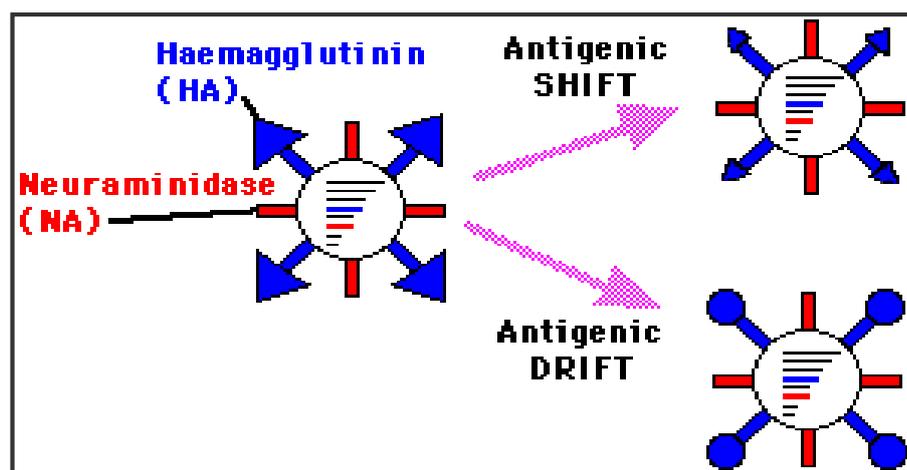


Fig. 5: Variazioni antigeniche dei virus influenzali.

Alla base dell'*antigenic drift* vi è la mancata attività di *proof reading* da parte dell'RNA polimerasi virale, ciò permette quindi l'insorgenza di ripetuti errori nella sequenza genica che non vengono pertanto corretti [12, 13].

Epidemiologia e manifestazioni cliniche

L'influenza costituisce un rilevante problema di sanità pubblica a causa della sua ubiquità e contagiosità, dovute alla sua variabilità antigenica (che rende parte della popolazione scoperta di fronte a nuovi ceppi o sottotipi), all'esistenza di serbatoi animali (che rende l'influenza una patologia non eradicabile) e alle possibili gravi complicanze sintomatologiche, che possono anche portare a morte. Il virus dell'influenza è infatti un'infezione virale altamente contagiosa, che si trasmette da individuo a individuo per via aerea, con andamento epidemico-pandemico, che può colpire individui di tutte le età più volte nel corso della vita. È il principale agente eziologico delle malattie respiratorie acute, costituisce una delle principali cause di morte per malattia infettiva nei paesi industrializzati, ed è causa di raffreddori comuni nel 10-15% dei casi [14]. Si stima che ogni anno l'epidemia influenzale sia responsabile di 3-5 milioni di casi di patologia severa, e di 250.000-500.000 casi di morte nel mondo. La morbosità per influenza è maggiormente presente nei bambini, negli adolescenti e nei giovani adulti, mentre gravi complicanze e mortalità sono maggiori negli anziani e in particolari gruppi a rischio.

È possibile distinguere due forme epidemiologiche di influenza. La prima è quella pandemica, conseguenza della comparsa di un nuovo sottotipo del virus A, contro la quale nessun soggetto della popolazione possiede ancora anticorpi specifici. Quando compare questa forma circa il 50% della popolazione mondiale si ammala di influenza nell'arco di 1-2 anni. Nel secolo scorso si è verificata tre volte: la pandemia Spagnola nel 1918 (sottotipo H1N1), l'Asiatica nel 1957 (sottotipo H2N2) e la Hong Kong nel 1968 (sottotipo H3N2) [15,16]. L'ultima pandemia, detta anche la pandemia del nuovo millennio, si è verificata nel 2009 con la comparsa del nuovo ceppo influenzale A/H1N1pdm09 [17].

La seconda forma epidemiologica è detta "interpandemica" e si verifica in corrispondenza di variazioni antigeniche minori del virus A e B. La malattia si può presentare come caso sporadico, focolaio localizzato o epidemia improvvisa; in questo caso si raggiunge il picco massimo di casi dopo circa un mese dai primi casi. Le varianti del sottotipo A danno epidemie ogni 2-4 anni, per via del tasso di mutazione più elevato rispetto al tipo B, che invece causa epidemie ogni 4-6 anni.

I focolai epidemici si verificano annualmente durante i mesi invernali, in particolare da novembre a marzo nell'emisfero boreale, da maggio a settembre in quello australe [11, 12].

L'influenza copre un ampio spettro di sintomi, si va dalle forme virali asintomatiche alle polmoniti, ma i più tipici sono la febbre e la tosse. I sintomi più diffusi sono comuni a molte altre infezioni respiratorie: malessere generale, cefalea, raffreddore, dolori muscolari ed articolari. Nei bambini si possono manifestare anche problemi a carico dell'apparato gastro-intestinale (nausea, vomito, diarrea), otite media nonché febbre più alta rispetto all'adulto. Nei neonati l'influenza compare solo come febbre molto alta e a causa della difficoltà di diagnosi può avere anche esito fatale. Nei periodi di *outbreaks* influenzale si registrano infatti maggiori percentuali di ospedalizzazione di bambini e infanti [11].

L'influenza si trasmette per via aerea, attraverso le goccioline di saliva (*droplets*) emesse con la tosse, lo starnuto o anche semplicemente parlando. Il virus manifesta elevato tropismo per le cellule ciliate dell'epitelio respiratorio, anche se causa maggior danno nelle basse vie aeree. L'influenza inoltre resiste molto bene nell'ambiente esterno in situazioni di bassa temperatura ed umidità e si diffonde facilmente negli ambienti affollati e chiusi; per questo motivo i picchi epidemici si verificano nei mesi invernali (quando la popolazione tende a trascorrere più tempo in luoghi chiusi) [11].

Il periodo di contagiosità comincia 1-3 giorni prima che si manifestino i sintomi e si prolunga per 5-7 giorni; solitamente il periodo di contagiosità è più ampio nei bambini che negli adulti [8]. Il periodo di incubazione dell'influenza è molto breve, da 1 a 4 giorni (in media 2). La sintomatologia dura dai 3 ai 6 giorni circa, ma può permanere uno stato di astenia fino a 2 settimane. Tuttavia nei soggetti ad alto rischio come i bambini, anziani, soggetti immunocompromessi, individui con patologie croniche e donne in gravidanza possono insorgere complicanze in seguito all'infezione influenzale. Queste complicanze sono dovute a sovrainfezioni batteriche, favorite dallo stato di debilitazione del sistema immunitario e possono riguardare diversi distretti corporei:

- il tratto respiratorio (bronchiti, polmoniti, sinusiti, otiti soprattutto nei bambini)
- il sistema cardiovascolare (miocarditi, pericarditi)
- il sistema nervoso centrale (encefaliti, sindrome di Guillan-Barrè, sindrome di Reye)
- apparato muscolo-scheletrico (miositi, soprattutto nei bambini) [11,12].

Tutte queste complicanze fanno aumentare il numero di degenze ospedaliere, nonché la mortalità per influenza. Per questi motivi la prevenzione dell'influenza, basata per lo più sulla vaccinazione, deve essere un obiettivo primario delle Autorità di Sanità Pubblica a livello nazionale ed internazionale.

2.1.2 I virus parainfluenzali

I virus parainfluenzali appartengono alla famiglia *Paramyxoviridae*, sottofamiglia *Paramyxovirinae*, genere *Paramyxovirus*. Essi possono infettare animali terrestri, mammiferi acquatici, volatili e alcuni rettili; includono inoltre parecchi virus patogeni per il tratto respiratorio dell'uomo, simili per alcuni aspetti ai virus influenzali. Differiscono da essi per le caratteristiche degli antigeni di superficie e per le dimensioni.

I virus parainfluenzali umani (hPIV) si distinguono antigenicamente in 5 tipi: 1, 2, 3, 4A e 4B. Tutti quanti causano infezioni respiratorie nell'uomo, anche se con epidemiologia e quadro clinico differente [3]. L'hPIV 1, 2 e 3 sono più diffusi rispetto al tipo 4; in particolare hPIV 1 e 2 sono associati a "croup" (laringotracheobronchite) e malattie dell'alto e basso tratto respiratorio, mentre hPIV 3 causa maggiormente bronchioliti e polmoniti.

Struttura e organizzazione genomica

I virus parainfluenzali hanno forma generalmente sferica, con diametro variabile tra i 150 nm e i 250 nm. Possiedono un *envelope* lipidico, che deriva dalla membrana plasmatica della cellula ospite, contenente il nucleocapside a simmetria elicoidale; questo è costituito dalle nucleoproteine **NP**, presenti in maggior percentuale, e le proteine **P** ed **L**, ad attività RNA-polimerasica. Dall'*envelope* esterno sporgono 2 glicoproteine: **F** e **HN**. La prima media la fusione con la cellula ospite e possiede attività emolisinica, la seconda è invece responsabile dell'attacco ai recettori delle cellule ospiti e possiede attività emoagglutinante e neuroaminidasica. Al di sotto dell'*envelope* si trova la proteina di matrice **M** che interagisce con il nucleocapside interno e le 2 proteine di superficie [3,18,19]. Il genoma è costituito da un singolo filamento di RNA a polarità negativa di lunghezza variabile compresa tra i 15.000 e i 16.000 nucleotidi.

Ciclo replicativo

Il ciclo replicativo del virus ha inizio con l'adesione ai recettori delle cellule ospiti da parte della proteina di superficie HN, mentre la proteina F media la fusione della membrana plasmatica. La replicazione del genoma avviene totalmente all'interno del citoplasma,

dove viene rilasciato il nucleocapside; una parte degli mRNA prodotti viene tradotta nelle proteine del virione e una parte è trascritta in RNA(-) in grado di associarsi alle NP per formare i nuovi nucleocapsidi. Questi migrano successivamente in prossimità della membrana plasmatica, vengono inclusi per gemmazione assieme con le proteine di superficie F e HN, formando infine nuovi virus completi (**Fig. 6**).

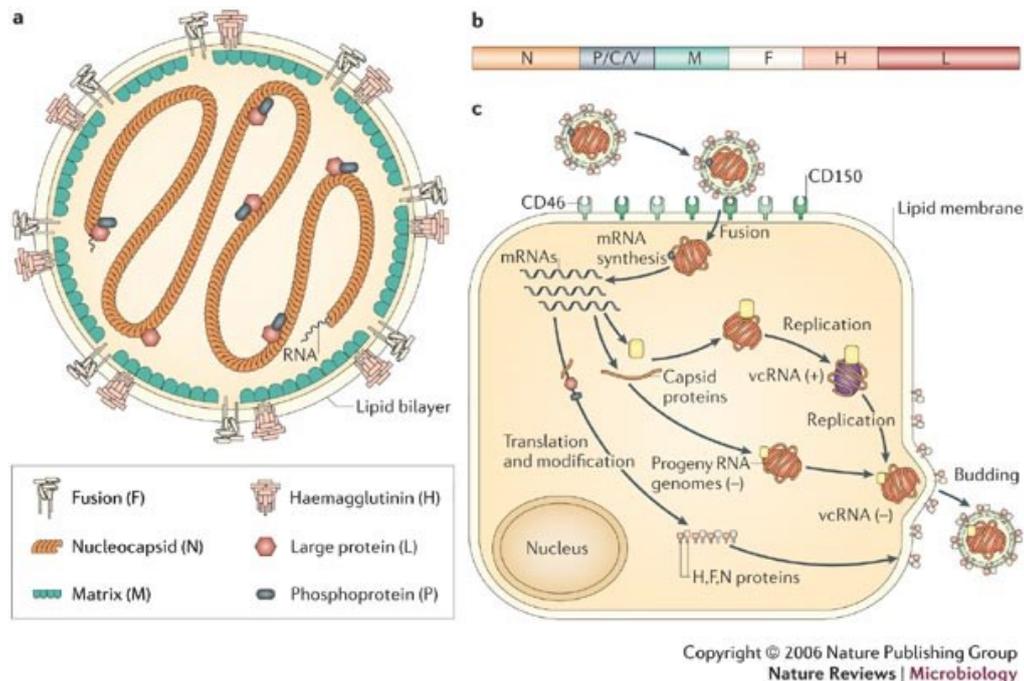


Fig. 6: Ciclo replicativo del parainfluenzavirus

Epidemiologia e manifestazioni cliniche

I virus parainfluenzali sono ubiquitari e assai comuni; infettano adulti ma soprattutto bambini, nei quali il virus si manifesta con sintomatologia più severa (bronchioliti, bronchiti, polmoniti). Gli hPIV, in particolare il sierotipo 3, sono, dopo il virus respiratorio sinciziale, la seconda causa di infezioni del basso tratto respiratorio in neonati ed infanti [18, 19]. L'infezione si presenta già nelle prime fasi della vita; dati sierologici hanno mostrato che il 50% dei bambini di un anno e il 90% a 5 anni presentano anticorpi protettivi per hPIV 3, il 75% per hPIV 1, il 60% per hPIV 2 e il 50% per hPIV 4 [18, 19]. La stagionalità delle infezioni varia a seconda del sierotipo: hPIV 1 e 2 provocano epidemie durante l'autunno e l'inverno, sono possibili comunque episodi di epidemie sovrapposte causate dai sottotipi hPIV 1 e hPIV 2; hPIV 3 causa invece infezioni sporadiche durante tutto l'anno con picco in primavera; hPIV 4 è presente durante tutto l'anno, ma è poco frequente e non causa sintomi rilevanti.

Dal punto di vista clinico i virus parainfluenzali coprono un ampio spettro di sintomi del basso e alto tratto respiratorio: croup, bronchite e bronchiolite sono i sintomi più comuni nei bambini, oltre a casi di corizza, otiti, polmoniti. Negli adulti non immunodepressi invece i virus parainfluenzali causano semplici raffreddori o rinolaringiti. Il 41% dei casi di croup è associato a questi virus [19]; in particolar modo hPIV 1 e hPIV 2 (in minor percentuale rispetto al tipo 1) sono i più coinvolti. In generale la bronchiolite è frequente nei primi 2 anni di vita e l'agente eziologico virale nel 10-15% dei casi è hPIV 1 e hPIV 3 [18]. Dopo il sesto anno di vita si nota un decremento delle infezioni da hPIV, che compaiono solo in maniera occasionale. Il virus colonizza le mucose nasali e della faringe, per poi replicarsi nell'epitelio ciliato respiratorio. I virus parainfluenzali si diffondono tramite le secrezioni respiratorie per diffusione di larghe particelle di aerosol, in cui rimane stabile per poco tempo, e tramite contatto diretto, modalità principale di trasmissione. hPIV 3 è altamente contagioso nelle comunità chiuse, come asili o nurseries, mentre i sierotipi 1 e 2 hanno infettività [19].

I sintomi compaiono dopo 1-7 giorni dal contagio, il periodo di contagiosità è compreso tra 3 e 10 giorni nella prima infezione, mentre nei casi di reinfezione l'intervallo di tempo è più breve [18, 19].

2.1.3 Virus respiratorio sinciziale

Il virus respiratorio sinciziale (RSV) appartiene alla famiglia delle *Paramyxoviridae*, genere *Pneumovirus* [18] e deve il suo nome al particolare tipo di crescita in coltura cellulare e alla localizzazione delle sue manifestazioni cliniche (**Fig. 7**). Vengono distinti in base alle caratteristiche degli antigeni di superficie, due gruppi di RSV: **RSV A e RSV B** [21] divisi a loro volta in vari sierotipi. L'uomo è conosciuto come l'unico serbatoio virale [20]. Il virus è stato isolato per la prima volta nel 1956 e riconosciuto successivamente come agente eziologico principale delle infezioni delle basse vie respiratorie nel neonato e nei bambini di età inferiore ai 5 anni ed è il maggior responsabile nell'insorgenza dei casi di bronchiolite e di polmonite [20].

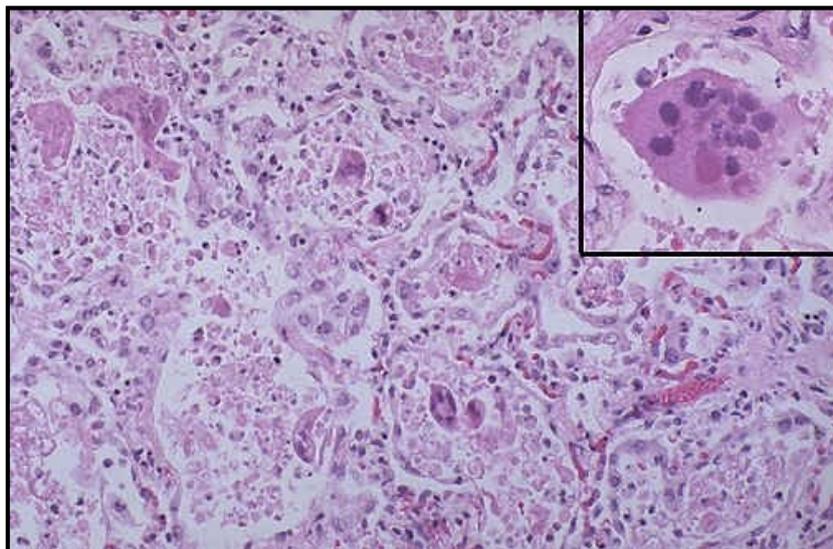


Fig. 7: Foto al microscopio di insieme di cellule giganti sinciziali, nel particolare in alto a destra è ben visibile la struttura di un sincizio, con tanti nuclei in un'unica cellula.

Struttura e organizzazione genomica

RSV, come i virus parainfluenzali, ha aspetto pleiomorfico, con diametro variabile compreso tra i 150 e i 300 nm. È composto da un *envelope* lipidico esterno a doppio strato, dal quale sporgono proiezioni proteiche, e racchiude il nucleocapside a simmetria elicoidale del diametro di 12-15 nm (minore rispetto agli altri *Paramyxovirus*), con genoma costituito da un singolo filamento di RNA a polarità negativa.

Sono presenti tre glicoproteine di superficie che sporgono per una lunghezza di 12 nm dall'*envelope*: **G** (media l'attacco alla cellula ospite), **F** (proteina di fusione responsabile anche della formazione dei sincizi costituita da due subunità F1 e F2), e **SH** (*Small Hydrophobic protein*, presente in varie forme a livello citoplasmatico). Al di sotto dell'*envelope* si trova il nucleocapside costituito dalla proteina di matrice **M**, dal piccolo dominio **M2** e che racchiude le proteine **N**, **P** (entrambe impegnate nella trascrizione), la proteina **L** (ad attività polimerasica) e **NS1** e **NS2** (proteine non strutturali) (**Fig. 8**). Nessuna proteina possiede proprietà emoagglutinante e neuroaminidasica.

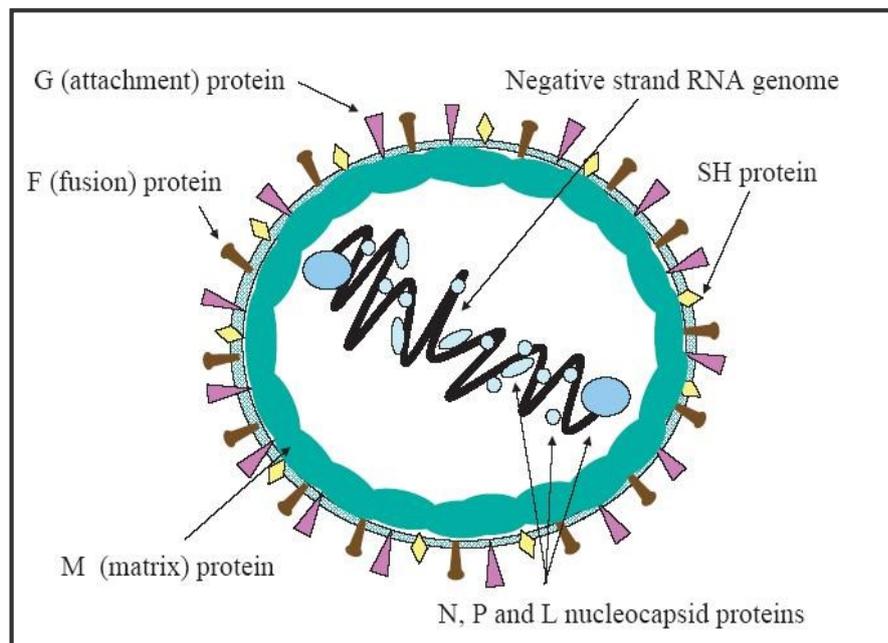


Fig. 8: Rappresentazione schematica di RSV.

Ciclo replicativo

RSV si attacca alla cellula ospite tramite la glicoproteina G, e la fusione tra la membrana cellulare e l'envelope avviene tramite l'intervento della proteina F, importante nella formazione del sincizio caratteristico di questo virus; l'*envelope* lipidico si fonde con la membrana della cellula ospite e il nucleocapside viene rilasciato nel citoplasma. A questo punto avviene la replicazione dell'RNA genomico, che verrà inserito nelle nuove particelle virali. Nello stesso tempo una parte dell'RNA è trascritto e tradotto nelle proteine del virione, che andranno ad accumularsi presso la membrana cellulare. I virioni neosintetizzati vengono poi inseriti dalla vescicola esocitotica (che costituirà il nuovo *envelope*) ed espulsi all'esterno (**Fig. 9**) [20, 22].

Epidemiologia e manifestazioni cliniche

RSV è un virus ubiquitario, la cui infezione mostra stagionalità nelle regioni temperate; nei centri urbani gli eventi epidemici avvengono nel periodo compreso tra tardo autunno e primavera, i picchi epidemici tra febbraio e marzo, con circolazione di entrambi i gruppi antigenici **A** e **B**. Nelle zone tropicali e subtropicali invece le epidemie coincidono con la stagione delle piogge, cioè nel periodo compreso tra giugno e dicembre.

Il virus si replica nella mucosa nasale e nel tratto respiratorio, si diffonde maggiormente tramite dispersione di *droplets*, per contaminazione di oggetti, ed in misura minore nelle particelle di aerosol [20].

RSV è l'agente eziologico più importante nelle infezioni del basso tratto respiratorio negli infanti e nei bambini di età inferiore ai 2 anni; si stima che metà dei bambini si infetti nel primo anno di vita, di cui il 25-40% sviluppi patologie a carico del basso tratto respiratorio; entro i primi 3 anni di vita tutti i bambini vengono in contatto col virus respiratorio sinciziale [20, 22]. La malattia inizia come infezione delle alte vie respiratorie (URI), caratterizzata da rinorrea e tosse. Entro 1-3 giorni, si verifica la comparsa di febbre, otite media e brividi, mentre dopo 7-12 giorni vi è la remissione della malattia. Nel caso di sintomi quali tosse persistente e comparsa di dispnea, l'infezione può raggiungere le vie inferiori e richiedere eventualmente il ricovero del soggetto (ciò avviene più frequentemente in bambini immunocompromessi o con patologie cardiache pre-esistenti).

Nei neonati fino ad un mese di vita l'infezione è rara (per via degli anticorpi materni ancora circolanti); a partire dal secondo mese fino al sesto, con picco massimo nei primi 3 mesi,

aumenta il rischio di sviluppare un'infezione severa con bronchiolite e polmonite. Al di sopra dei 18 mesi d'età i bambini presentano frequentemente bronchite e bronchiolite (questa ultima per lo più nei soggetti asmatici). Nell'adulto l'infezione si presenta in forma lieve soprattutto nei soggetti a stretto contatto con i bambini (genitori, personale infermieristico). Studi epidemiologici indicano una bassa mortalità nei paesi industrializzati pari all'1% dei casi; la percentuale però aumenta drasticamente fino al 37% nei bambini con malformazioni cardiache congenite. Le reinfezioni sono assai comuni nel corso di tutta la vita, sebbene nei bambini in età scolare presentino sintomatologia meno severa e negli adulti decorrano per lo più in maniera inapparente [3,5,18,21,22].

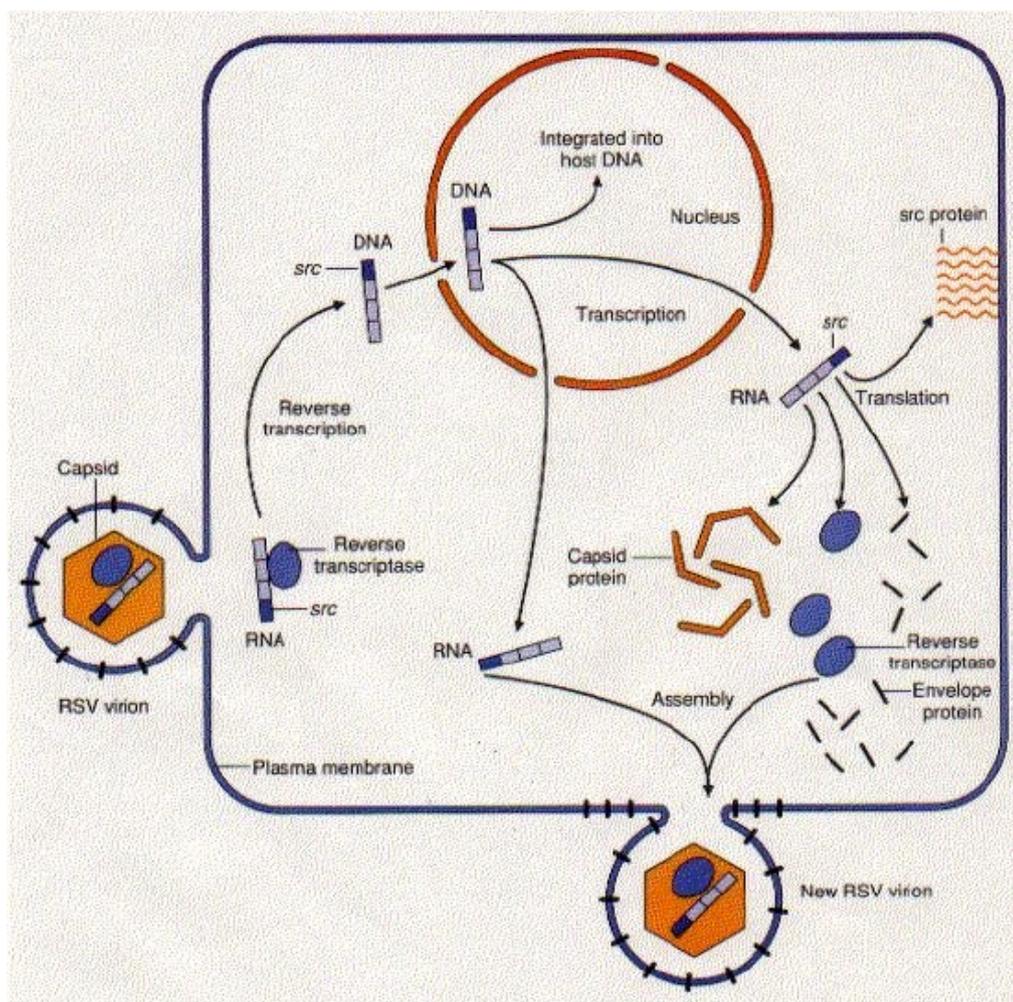


Fig. 9: Schema di replicazione del virus respiratorio sinciziale

2.1.4 Picornavirus

Il nome picornavirus deriva da “Pico” che significa piccolo e “RNA” con riferimento alla molecola, infatti a questa famiglia fanno parte piccoli virus ad RNA che hanno un capsidico icosaedrico di diametro di 28-30 nm e presentano un involucro pericapsidico. Il loro genoma è costituito da un unico filamento di RNA di polarità positiva. Il sito di replicazione e di assemblaggio del virione è il citoplasma della cellula infetta [23].

I Picornavirus contengono due principali generi che possono infettare le vie respiratorie: gli Enterovirus e i Parechovirus di più recente identificazione [24] (**Fig. 10**).

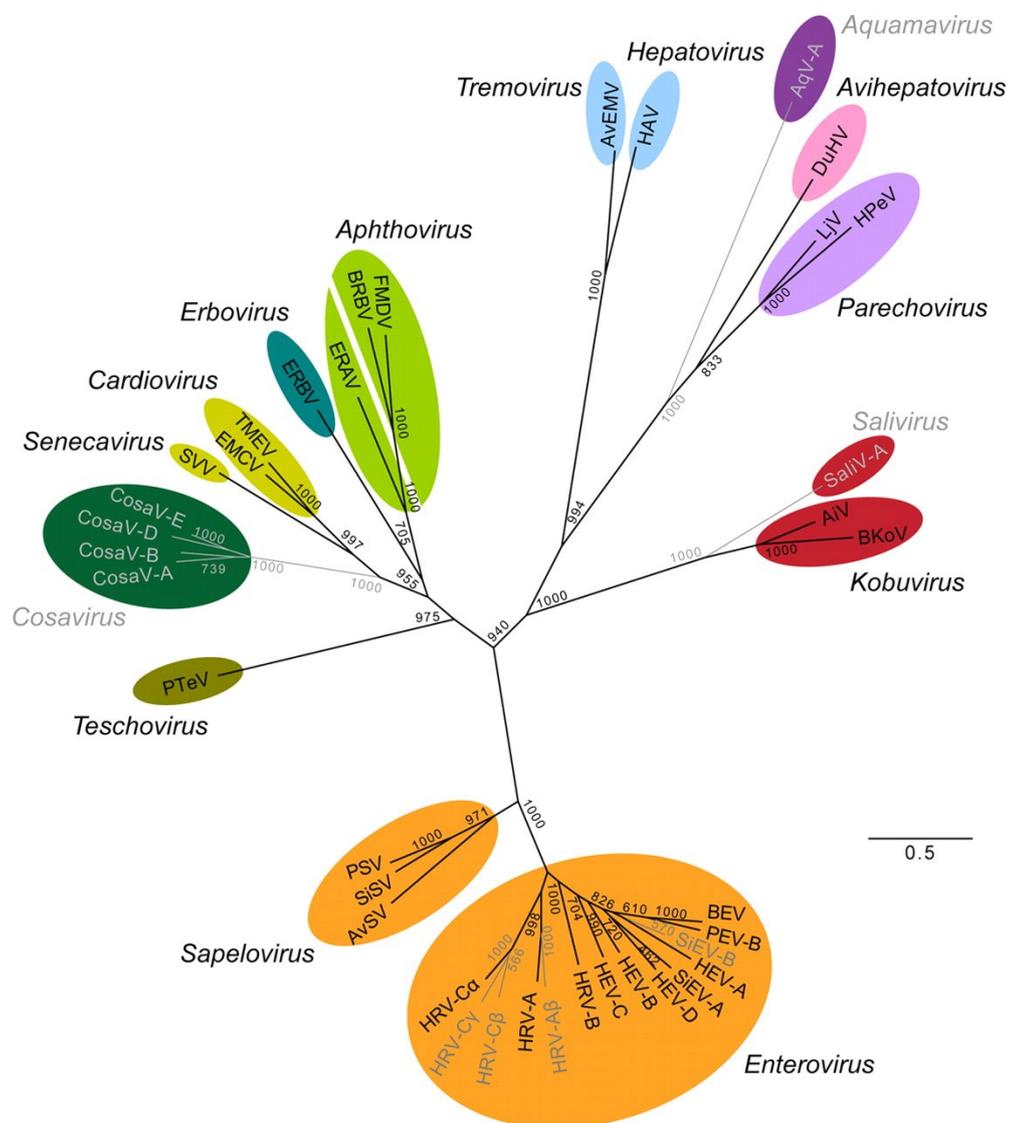


Fig. 10: Albero filogenetico dei virus appartenenti alla famiglia Picornaviridae

I Rhinovirus sono correlati tassonomicamente agli Enterovirus essendo una specie dello stesso genere *Enterovirus*. La tassonomia attuale dei nonpolio EVs si divide in quattro specie che infettano l'uomo (EVs umano da A a D), tra i quali si riconoscono svariati sierotipi. I diversi sierotipi hanno diversi schemi temporali di circolazione che può essere associata a diverse manifestazioni cliniche. Anche se la maggior parte delle infezioni umane da EV rimangono asintomatici, questi virus sono associati a diverse sindromi cliniche, che vanno da lievi a gravi malattie febbrili e patologie potenzialmente fatali, tra cui meningite asettica, encefalite, miopericarditis, paralisi flaccida acuta e grave malattia di sepsi neonatale-like). Inoltre può indurre malattie respiratorie nei bambini o adulti, tra cui infezioni delle alte vie respiratorie, ma anche infezioni del tratto respiratorio inferiore (LRTIs), con conseguente bronchite, bronchiolite e la polmonite. Diversi ceppi umani di enterovirus, compresi gli EV 68 [25] e 71, coxsackie A9, A21, B2 e B4 virus, e Echovirus 9, 11 e 22 sono state isolate da campioni di nasofaringeo, aspirati tracheali, lavaggi broncoalveolare, o tessuti polmonari da classici saggi di coltura cellulare e identificato come la causa di grave o fatale broncopolmonite virale.

I rhinovirus umani (hRV) comprendono molti sierotipi differenti divisi nei gruppi **A**, **B** e **C** (di recente identificazione) [26] in base ai recettori cellulari d'attacco. A differenza degli *Enterovirus*, i hRV sono molto labili a pH acidi ($3 < \text{pH} < 6$); in compenso sono termostabili a temperature tra i 24-37°C, e sono in grado di conservarsi a temperature molto basse. Sono state evidenziate cross-reazioni tra i vari sierotipi e anche infezioni di virus umani in animali e viceversa.

Struttura e organizzazione genomica

I *Picornavirus*, sono costituiti da piccole particelle con capsidi icosaedrico del diametro di 25-30 nm, non rivestite da *envelope*; il capside è costituito dall'impaccamento di 60 protomeri, ognuno dei quali è formato dai polipeptidi VP1-2-3-4 derivanti dal clivaggio del protomero d'origine VP0 (**Fig. 11**).

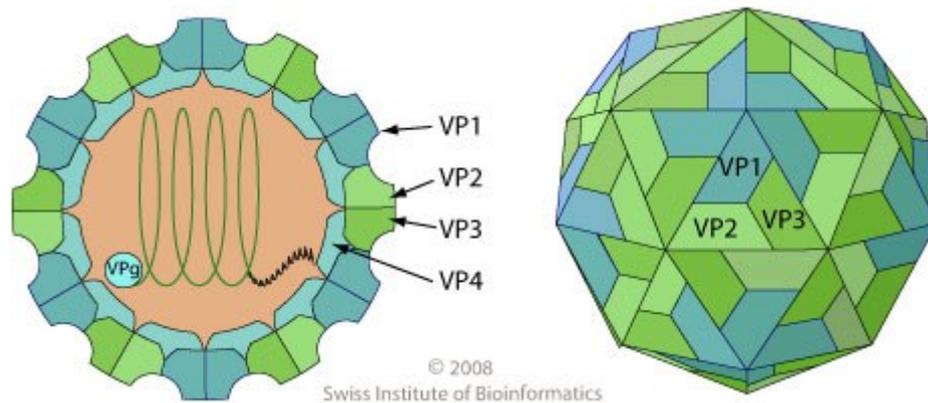


Fig. 11 : struttura *Picornavirus*

A causa dell'assenza di *envelope* e delle relative glicoproteine di superficie i rhinovirus non possiedono attività emoagglutinante. Il genoma è costituito da un singolo filamento di RNA a polarità positiva, di lunghezza pari a 2500 nm strettamente impaccato nel capside.

Ciclo replicativo

La replicazione avviene all'interno del citoplasma della cellula ospite (**Fig. 12**); una volta entrato il genoma viene prima tradotto e in seguito copiato in filamento a polarità negativa; questo servirà da stampo per sintetizzare un grande numero di copie genomiche: una parte sarà trascritta in mRNA e usata per sintetizzare le proteine virali, un'altra verrà invece retrotrascritta in genoma da inserire nei nuovi virioni [23].

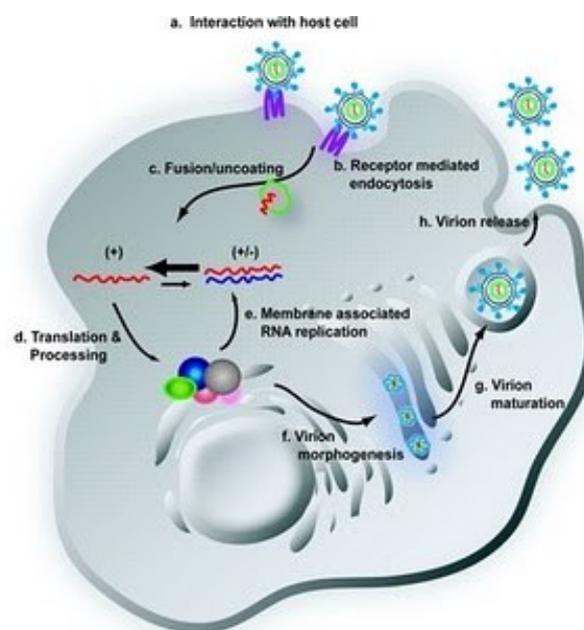


Fig. 12: Schema di replicazione dei Picornavirus.

Epidemiologia e manifestazioni cliniche

Gli *Enterovirus* causano infezioni ubiquitarie, rilevabili in ogni stagione nelle zone temperate del globo. Si presentano con maggior frequenza in autunno, a partire da settembre con picco ad ottobre, ed in primavera (con picco a marzo-aprile), quando cioè si rilevano i maggiori sbalzi di temperatura. Sono causa di infezioni respiratorie anche nei mesi invernali, sebbene in questo periodo i virus circolanti in percentuale maggiore siano i virus influenzale e RSV [27,28]. I rhinovirus rappresentano le infezioni respiratorie acute più comuni nei soggetti di

tutte le età. Si stima che più del 50% dei casi di raffreddore comune sia attribuibile ai rhinovirus; nei bambini queste infezioni sono ancora più diffuse e si presentano con una maggiore severità di sintomi. Studi recenti indicano come più del 90% dei bambini entro i 2 anni di età abbia già anticorpi circolanti contro i rhinovirus, e il 79% abbia contratto almeno un'infezione da hRV [28]. I virus si replicano e danneggiano per lo più l'epitelio ciliato del tratto nasale e, in misura minore, la mucosa tracheo-faringea, mentre le vie respiratorie inferiori sono molto meno suscettibili ai rhinovirus; per questa ragione il naso e il tratto respiratorio superiore rimangono il sito dove l'infezione si manifesta maggiormente. La sintomatologia è quella tipica del raffreddore comune: ostruzione nasale, corizza, mal di gola e tosse; possono tuttavia sopraggiungere leggera febbre, malessere generale e mal di testa. Studi indicano che l'otite media sia la complicanza più comune che può sopraggiungere nei bambini (fino al 40% di casi con infezione da hRV) [29]. In questi ultimi (ma anche nei soggetti con patologie croniche a carico dell'apparato respiratorio o negli anziani) possono insorgere complicanze riguardanti il tratto respiratorio inferiore, quali bronchite, bronchiolite e anche polmonite.

Studi epidemiologici indicano che più del 30% di bambini asmatici o con bronchite sviluppano LRI associate a hRV; inoltre, dai dati a disposizione sui virus respiratori più comuni in età pediatrica, si è notato che le LRI associate al virus respiratorio sinciziale sono prevalenti nei bambini di età < 12 mesi, mentre al di sopra di un anno prevale l'associazione con i rhinovirus [36-38].

I sintomi si manifestano con picco nei 3-4 giorni successivi all'infezione. La remissione avviene mediamente dopo 7 giorni; nel 25% dei casi la durata dei sintomi può estendersi fino alle 2 settimane e in rari casi fino ad un mese. Il virus si diffonde per contaminazione con secrezioni nasali o salivari, tramite contatto diretto con le mani (che rimangono il veicolo principale per gli hRV) o indiretto con fomite; in misura minore il contagio può avvenire anche per mezzo di particelle di aerosol [30].

2.1.5 Coronavirus

I Coronavirus appartengono alla famiglia delle *Coronaviridae* di cui fa parte solo il genere *Coronavirus*. Sono noti soltanto dal 1965, quando si riuscì a correlare un agente patogeno umano, causante raffreddore, con virus aviari causanti bronchiti o gastroenteriti. I Coronavirus infatti nell'uomo sono responsabili di infezione respiratorie lievi e aspecifiche, quali il raffreddore comune, e di infezioni dell'apparto gastroenterico. Negli animali (cani, gatti, topi, ratti, suini, bestiame, polli, tacchini) sono l'agente eziologico di gastroenteriti, infezioni dell'apparato respiratorio, del fegato e anche del sistema nervoso centrale. Non sono state evidenziate fino ad ora cross-infezioni animale-uomo.

Si distinguono 4 generi di CoV (**Fig.13**) e 4 sottotipi umani antigenicamente differenti, OC43 e 229E [3,30,31] e i più recenti CoV-NL63, CoV-HKU1.

Fanno parte dei *Coronavirus* anche le due più recenti specie:

- SARS-CoV, [32] l'agente eziologico identificato nel 2003, della Sindrome Respiratoria Acuta Severa, il cui *outbreak* è avvenuto nel 2002 in Cina e si è poi diffuso rapidamente in tutti i continenti causando epidemie, con gravi conseguenze socio-economiche, oltre che sanitarie.
- MERS-CoV, *Middle East Respiratory Syndrome*, segnalato per la prima volta il 24 settembre 2012 in Arabia Saudita è un coronavirus che probabilmente ha come *reservoir* i dromedari [33].

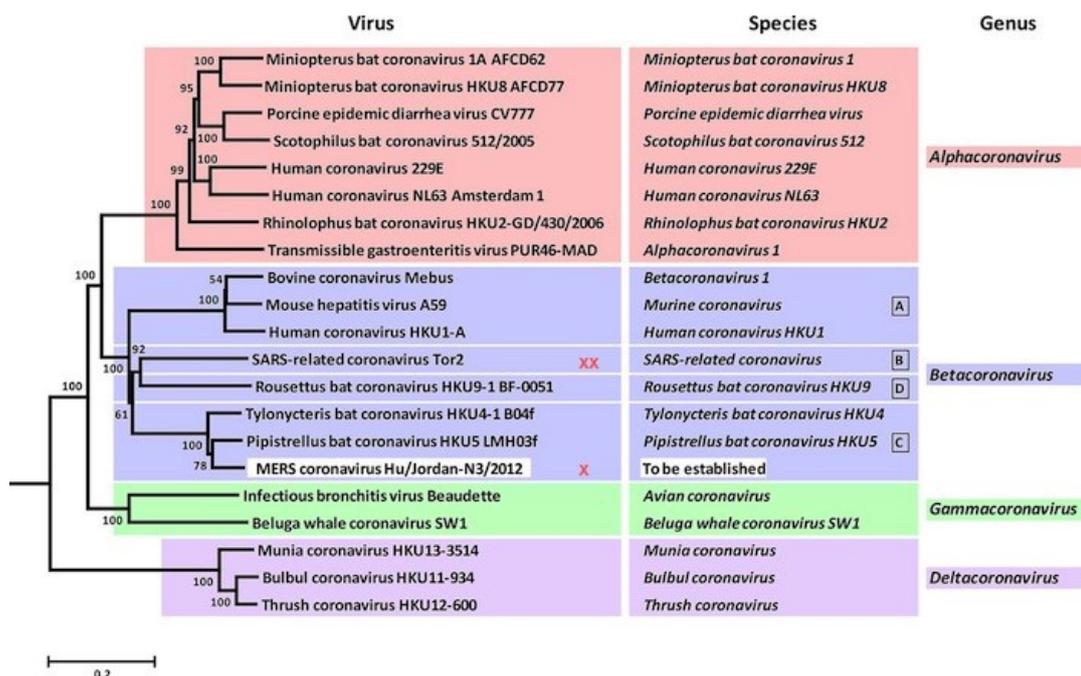


Fig. 13 : Classificazione Coronavirus

Struttura e organizzazione genomica

I coronavirus sono pleriomorfici, per lo più a struttura sferica, con diametro medio di 100-150 nm, e caratterizzati dal tipico aspetto a “corona”: dall'*envelope* lipidico sporgono numerosi peplomeri, ognuno dei quali di 20 nm di lunghezza (**Fig.14**). Dall'*envelope* lipidico sporgono 3 glicoproteine: la proteina **E1** detta di matrice (M), di piccole dimensioni (20-30 KD) e poco sporgente; la **E2**, di dimensioni maggiori (180-200 KD) detta di fusione (F) che conferisce al virus il caratteristico aspetto a corona, e infine la proteina **E3** con attività emoagglutinante (HE), di dimensioni medie (120-140 KD). Il nucleocapside interno possiede simmetria elicoidale e contiene un singolo filamento di RNA non segmentato a polarità positiva delle dimensioni di 27-31 Kb [2, 39].

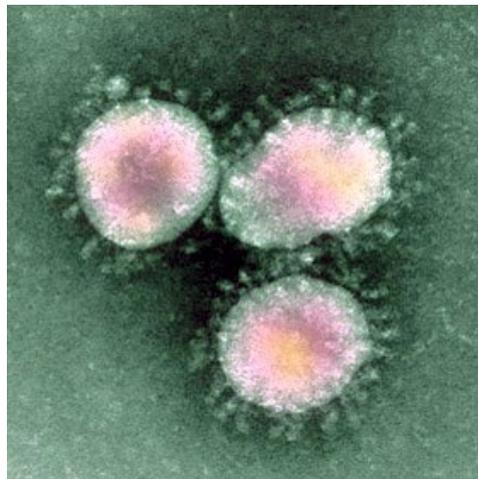


Fig. 14: Fotografia Coronavirus

Ciclo replicativo

Il ciclo replicativo è simile a quello dei Paramyxovirus, [3, 18] anch'esso avviene totalmente all'interno del citoplasma. L'adesione avviene tramite le spiccole e la proteina di fusione F. Successivamente il genoma rilasciato nel citoplasma va incontro a trascrizione: una parte degli mRNA a polarità negativa vengono ritrascritti in mRNA a polarità positiva, pronti per poi essere inclusi nel nuovo nucleocapside, un'altra parte è tradotta nelle proteine del virione; questo viene infine esocitato all'esterno.

Epidemiologia e manifestazioni cliniche

I coronavirus sono infezioni ubiquitarie e assai comuni, con andamento epidemico e picco nei mesi invernali e ad inizio primavera. Nell'uomo sono responsabili del 15-20% delle infezioni del tratto respiratorio superiore [1], ma possono interessare anche il tratto gastroenterico, soprattutto nei bambini d'età < 12 mesi. I coronavirus sono la seconda causa del raffreddore comune dopo i rhinovirus (5-17% dei casi). Numerosi studi hanno inoltre evidenziato la presenza di anticorpi a partire dal primo anno d'età fino al 50% dei bambini di 4 anni, e al 70% negli adulti [2]. La sintomatologia è quella tipica del raffreddore: tosse, rinorrea, mal di gola, lieve febbre, malessere generale. Negli infanti e nei bambini i coronavirus respiratori possono interessare anche le basse vie aeree, con sintomatologia più grave; ciò si è riscontrato soprattutto in bambini asmatici o con patologie a carico dell'apparato respiratorio già in corso. La replicazione avviene nell'epitelio delle vie respiratorie, del tratto gastroenterico, e nei macrofagi. La trasmissione si esplica tramite particelle di aerosol oppure per via feco-orale.

A causa dell'elevata variabilità genetica e della presenza di numerosi sierotipi sono possibili reinfezioni [2, 31].

2.2 Virus respiratori di recente identificazione

Fra il 2002 e il 2003, la comparsa e la rapida diffusione internazionale della SARS (*Severe Acute Respiratory Syndrome*) ha dimostrato quanto rapidamente possa diffondersi una malattia infettiva emergente a livello globale e diventare una vera e propria minaccia per la salute della popolazione mondiale. Il responsabile di questa malattia è il coronavirus CoV-SARS [32]. Nell'aprile 2009 una nuova variante del virus influenzale di tipo A/H1N1 (virus contenente frammenti genici di virus umani, aviari e suini e originatosi a seguito di riassortimento fra due virus, di cui uno già triplo riassortante), ha fatto il suo ingresso nella popolazione, affermandosi in pochi mesi come ceppo pandemico a tutti gli effetti (11 giugno 2009), dimostrando come l'emergenza di nuovi virus respiratori è un problema di grande attualità [34]. Più recentemente un altro coronavirus MERS-CoV ha fatto il suo ingresso nel panorama mondiale ed è approdato anche in Italia, causando 3 casi [33].

Accanto al fenomeno delle infezioni emergenti, è possibile ipotizzare che una percentuale di infezioni respiratorie acute ad eziologia ignota possa essere causata da un'infezione virale non emergente, ma tuttavia ancora sconosciuta. Un esempio di patogeno che è passato inosservato alle comuni indagini microbiologiche e di recente identificazione è lo Human metapneumovirus (hMPV) [35], isolato per la prima volta nel 2001, che oggi è considerato una delle cause principali di infezione respiratoria acuta ed è diffuso in tutto il mondo. Fra i virus di recente identificazione, definiti anche se impropriamente 'nuovi virus respiratori', possiamo annoverare alcuni sottotipi di Coronavirus umani (*Human Coronavirus HKU1* e *NL63*), caratterizzati da due diversi gruppi di ricerca nel 2004, due virus appartenenti alla famiglia delle *Polyomaviridae* (*PyV-KI* e *PyV-WU*) la cui caratterizzazione data 2007 e nuovi virus all'interno della famiglia dei *Picornaviridae* (*Rhinovirus* di tipo C, *Parechovirus* tipi 2-14). Di grande impatto inoltre è stata l'identificazione di Human Bocavirus nel 2005, primo virus della famiglia delle *Parvoviridae* ad essere associato a infezioni respiratorie.

2.2.1 Human Metapneumovirus

Human Metapneumovirus (hMPV) è stato isolato per la prima volta nel 2001 in Olanda, [35] in campioni prelevati da bambini (0-5 anni) che presentavano infezione respiratoria acuta. HMPV appartiene alla famiglia delle *Paramyxoviridae*, sottofamiglia *Pneumovirinae*, genere *Metapneumovirus* [36]. Esistono due genotipi principali, denominati A e B, ciascuno dei quali può essere ulteriormente suddiviso in due *lineage*, conosciuti come A1, A2, B1 e B2. Tuttavia, la classificazione di hMPV non è ancora definitiva.

Struttura e organizzazione genomica

Il Metapneumovirus ha la struttura tipica dei *Paramyxovirus*. Alle indagini in microscopia elettronica appaiono sia particelle sferiche, con un diametro medio di circa 200 nm, che particelle filamentose e pleiomorfiche. HMPV possiede un envelope lipidico, che deriva dalla membrana cellulare, in cui sono inserite tre glicoproteine di superficie: la glicoproteina di attacco (G), la proteina di fusione (F) e una piccola proteina idrofobica (SH). Al di sotto dell'envelope è presente un nucleocapside, a simmetria elicoidale di circa 17 nm di diametro e con lunghezza variabile compresa tra i 200 e i 1000 nm, composto dal genoma virale in associazione con alcune proteine: la proteina del nucleocapside (N) la fosfoproteina (P) e la polimerasi (L). Si pensa che anche la proteina M2-1, promotore trascrizionale, sia associata al nucleocapside. Questo è circondato dalla proteina di matrice (M), di natura idrofobica e non glicosilata, che forma un collegamento fra il nucleocapside e la membrana lipidica [45,47]. Il genoma a RNA a singolo filamento a polarità negativa (ssRNA-) è lungo circa 13kb e codifica per 9 proteine, i cui geni sono presenti sul genoma nel seguente ordine:

3' – N – P – M – F – M2 – SH – G – L – 5'

Il gene M2 codifica per due proteine distinte: M2-1 e M2-2, che presentano ORF (Open Reading Frames) sovrapposte. Inoltre, nel genoma sono presenti anche diverse sequenze non codificanti, rappresentate dalla 3'*leader*, contenente il promotore per la trascrizione virale, la 5'*trailer*, e altre regioni intrageniche che sono coinvolte nella regolazione della replicazione, della trascrizione e del processo di assemblamento del nucleocapside. [35,37], (Fig. 15).

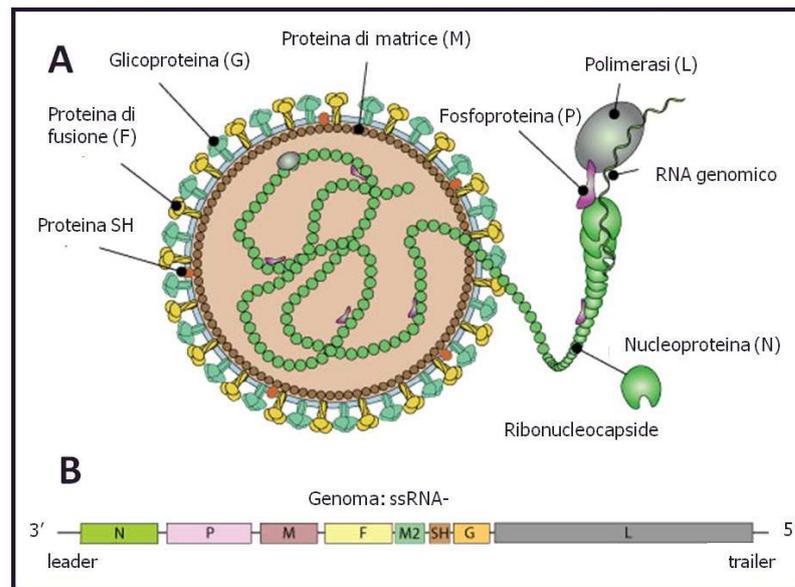


Fig.15: struttura e genoma di human Metapneumovirus

Ciclo Replicativo

Quando il virus raggiunge la cellula, le proteine G e F lavorano in sinergia per mediare l'attacco e la fusione dell'*envelope* virale con la membrana cellulare, consentendo in questo modo l'ingresso del virus in cellula. Inizialmente si forma il legame tra la glicoproteina G e recettori specifici sulla membrana (probabilmente residui di acido sialico); tale evento induce un cambiamento conformazionale che permette alla proteina F di interagire con la membrana cellulare e causare la fusione dell'*envelope* con la cellula ospite [38]. Durante questo processo le proteine M si dissociano dal nucleocapside, mediante un meccanismo ancora non noto, permettendone il rilascio di quest'ultimo nel citoplasma (processo definito svestimento o *uncoating*). In questo modo il genoma virale, in associazione stretta con le tre proteine del capsidico (N, L e P) a formare il complesso di replicazione, serve da template sia per la trascrizione dei geni virali, che per la replicazione del genoma. Le proteine L e P agiscono come RNA polimerasi e trascrivono mRNA subgenomici (poliadenilati a *capped*) richiesti per la sintesi delle proteine virali: le proteine vengono sintetizzate nello stesso ordine con cui sono disposti i loro geni sul genoma virale. I geni sono separati da sequenze non codificanti che portano i segnali di poliadenilazione e di inizio di trascrizione del gene successivo. Le due glicoproteine di superficie, che verranno inserite nell'*envelope*, vengono sintetizzate a livello del reticolo endoplasmatico ruvido (RER) e subiscono delle modificazioni post-traduzionali (glicosilazione) man mano che passano attraverso di esso e l'apparato di

Golgi, e vengono infine inserite all'interno della membrana cellulare. Tutte le altre proteine vengono sintetizzate libere all'interno del citoplasma e quelle che andranno a formare la matrice virale si localizzano a livello della membrana cellulare, formando uno strato continuo a stretto contatto con le glicoproteine G e F [38,39]. La replicazione del genoma avviene attraverso la sintesi di un 'anti-genoma' a polarità positiva, che funge da intermedio replicativo e quindi da stampo per la sintesi successiva di numerose copie di genoma virale a polarità negativa. A questo punto avviene la formazione del nucleocapside virale: le proteine neosintetizzate si associano a livello citoplasmatico con le nuove copie di genoma e le proteine M, a localizzazione sub-membranaria, veicolano il trasporto del nucleocapside verso la membrana cellulare. I nuovi virioni escono dalla cellula mediante gemmazione causando la lisi della cellula [38,39].

Epidemiologia e manifestazioni cliniche

La frequenza dell'infezione riportata nei diversi studi di letteratura è strettamente dipendente dalla casistica analizzata: hMPV è stato identificato con frequenze che variano dal 3 al 25% [40,41,42]. L'infezione da hMPV colpisce prevalentemente i bambini e, studi di sieroprevalenza hanno indicato come virtualmente tutti i bambini hanno già contratto l'infezione al compimento del quinto-decimo anno di età [41,42,43]. Infatti, già all'età di 5 anni più del 90% dei bambini mostra evidenza di avvenuto contatto con il virus, percentuale che tende ad aumentare fino a raggiungere valori prossimi al 100% negli adulti. I bambini con età inferiore ai 5 anni sono la classe di età più suscettibile all'infezione, mentre quelli con età inferiore ai due anni sono quelli a più alto rischio di sviluppare complicanze.

In generale, il virus causa una varietà di sintomi tipici dei *Paramyxovirus*, riguardanti sia le alte vie respiratorie (sindrome del raffreddore comune, rinofaringite, otite media), che le basse vie respiratorie, con sintomatologia variabile (bronchiolite, bronchite e polmonite nei casi più gravi).

Studi di sorveglianza virologica hanno dimostrato che il profilo di circolazione stagionale di hMPV coincide con quello dell'RSV. La maggior parte dei casi in aree a clima temperato, infatti, sono riportati nei mesi tardo-invernali e inizio-primaverili (dicembre-maggio), con un picco nei mesi di inizio primavera coincidente o leggermente successivo a quello dell'RSV, mentre nelle aree subtropicali il virus circola prevalentemente in primavera-estate. Tuttavia casi sporadici sono stati riportati tutto l'anno [43].

2.2.2 Human Bocavirus

Identificato solamente nel 2005, *human Bocavirus* (hBoV) è oggi considerato uno dei principali agenti virali associato alle infezioni del tratto respiratorio e gastroenterico, soprattutto nei bambini con età inferiore ai 3 anni [44,45].

Human Bocavirus appartiene alla famiglia delle *Parvoviridae*, alla sottofamiglia delle *Parvovirinae*, al genere *Bocavirus*. Ad oggi sono conosciute 4 specie di Bocavirus umani (hBoV1-4).

Struttura e organizzazione genomica

I Parvovirus sono fra i più piccoli virus a DNA degli animali; il virione, composto interamente da DNA e proteine, ha un diametro compreso fra i 18 e i 26 nm, e non sono dotati di *envelope* [46]. Il capside virale, composto esclusivamente da due proteine, VP1 e VP2, presenta simmetria icosaedrica ed è formato da 60 subunità proteiche riunite in capsomeri, la cui struttura è stata parzialmente risolta (**Fig. 16**). Nessun enzima, né di natura virale né di provenienza dall'ospite, è mai risultato associato al virione. L'interazione fra le proteine capsidiche e il sottostante DNA si pensa sia prevalentemente dovuto a legami idrogeno [47]. Probabilmente a causa della sua semplicità strutturale, e soprattutto dell'assenza dell'*envelope*, il virus risulta particolarmente resistente all'inattivazione.

Come tutti i *parvovirus*, il genoma di hBoV è formato da DNA a singolo filamento a polarità complementare all'mRNA, quindi negativo, della dimensione di circa 5kb. Il genoma di hBoV presenta la tipica struttura del genoma dei *parvovirus*: la parte in 5' del genoma contiene geni che codificano per proteine non strutturali, mentre quella in 3' comprende i geni delle proteine capsidiche. Possono quindi essere individuate due ORF principali: in 5' quella corrispondente al gene che codifica per la proteina non strutturale (NS) e in 3' quella invece corrispondente al gene che codifica per le due proteine capsidiche (VP1/VP2). La sequenza nucleotidica di quest'ultimo gene contiene l'informazione necessaria per la costruzione di entrambe le proteine capsidiche: la sequenza codificante la proteina VP1 comprende tutta la ORF, mentre quella codificante per la proteina VP2 rappresenta l'ultima parte (circa 2/3) della ORF; la prima parte della ORF codifica esclusivamente per VP1 e viene chiamata *VP1 unique region* [48]. E' inoltre presente nel genoma di hBoV una ORF centrale parzialmente

sovrapposta, tipica esclusivamente del genere Bocavirus, che codifica per la proteina non strutturale NP-1. I geni non strutturali (NS e NP-1) sono altamente conservati, mentre maggiori risultano i polimorfismi a livello delle proteine capsidiche, soprattutto nella porzione in 3', caratteristica che riflette il fatto che le proteine VP1 e VP2 sono quelle dotate di attività antigenica e quindi bersaglio della risposta anticorpale. Nonostante non siano ancora state individuate con certezza, è presumibile pensare che hBoV, come tutti i Parvovirus, possieda sequenze palindromiche ad entrambe le estremità (5' e 3') del genoma virale. Queste sequenze sono di fondamentale importanza in quanto giocano un ruolo primario nella replicazione del genoma virale: esse si ripiegano su loro stesse per formare delle strutture a forcina (stabilizzate da legami idrogeno tra le sequenze complementari) che serviranno da innesco per la DNA-polimerasi. Queste sequenze palindromiche sono intercalate da nucleotidi di collegamento che stabilizzano la struttura (**Fig. 15/B**) [47].

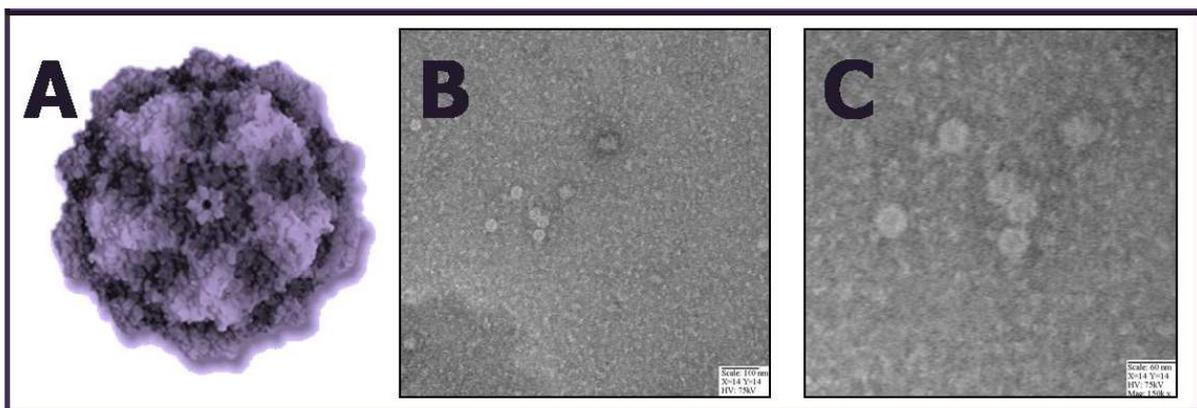


Fig. 16: Human Bocavirus: rappresentazione schematica (Jon Cornelius. Deutsche Krebsforschungszentrum)(A); immagini al microscopio elettronico [49] (B,C).

Ciclo Replicativo

I *Parvovirus* sono i virus a DNA più strettamente dipendenti dalle funzioni cellulari: hanno bisogno della DNA polimerasi e di altri fattori cellulari dell'ospite per poter replicare. I *Parvovirus* replicano efficientemente solamente in cellule in attiva replicazione (fase S), pur non essendo in grado di indurre la cellula ad entrare in fase replicativa (tuttavia non tutte le cellule che replicano attivamente sono sensibili all'infezione).

Tuttavia pare che questi virus siano in grado di infettare sia cellule permissive che non (probabilmente in quanto utilizzano recettori ubiquitari) ma un'infezione produttiva è possibile soltanto in cellule permissive.

Non è ancora ben chiaro quale sia il meccanismo di ingresso del virus in cellula, tuttavia si pensa che possa essere dovuta a endocitosi mediata da recettori. Nonostante il meccanismo con cui il genoma si sposta verso il nucleo non sia ancora noto, è possibile assumere che esso raggiunga il nucleo veicolato dal nucleocapside: il capsido parvovirale è sufficientemente piccolo per poter passare attraverso i pori nucleari e i segnali di ingresso, localizzati sulla superficie esterna del nucleocapside a livello della proteina NS, contengono motivi che riconoscono il macchinario di trasporto cellulare consentono al virus di poter essere trasportato all'interno del nucleo.

La replicazione del genoma virale avviene a livello nucleare. Tale caratteristica, combinata col fatto che il virus richiede una cellula in fase S per potersi replicare, indica che esiste una strettissima relazione fra replicazione virale e cellulare. Infatti, il virus utilizza la DNA-polimerasi e alcuni fattori cellulari per la replicazione del suo genoma.

La replicazione del DNA si svolge in due fasi. Durante la prima fase, la DNA-polimerasi comincia a duplicare il genoma virale (a livello della forcina localizzata in 3') e si ha la conversione del genoma virale ad una forma bicatenaria (dsDNA), denominata intermedio replicativo. La seconda fase è rappresentata dalla produzione della progenie a partire dall'intermedio replicativo; questo processo avviene attraverso la formazione di intermedi a doppio filamento attraverso un meccanismo di replicazione unidirezionale quasi circolare (le operazioni di duplicazione che avvengono durante la seconda fase, utilizzano come punto di partenza la forcina localizzata in posizione 5'). Questa fase prevede il clivaggio del DNA in diversi punti, causando la conseguente formazione di intermedi replicativi; ogni clivaggio specifico crea così un punto d'innescamento per la sintesi di un nuovo intermedio replicativo. Inoltre, perché tale processo si svolga correttamente, è necessaria la presenza dell'attività topoisomerasica dalla proteina NS, in grado di risolvere la struttura delle forcine [47]. Il DNA in fase replicativa presenta infatti una proteina legata covalentemente in posizione 5', rappresentata verosimilmente da NS1. La molecola bicatenaria viene così tagliata a livello della catena originale, in un punto opposto all'inizio della neosintesi, dando luogo a due catene polinucleotidiche lineari e complementari, di cui quella originale priva delle sequenze ripetute (trasferite alla molecola neoformata). La catena originale è infine completata in

direzione 3' della sequenza mancante. Le due catene complete vengono quindi separate tra loro ad opera di proteine virali e/o cellulari ad attività elicastica, portando alla formazione di due molecole in grado di iniziare un nuovo ciclo replicativo. La trascrizione del genoma può avvenire una volta che la molecola di DNA monocatenaria sia stata convertita nella forma bicatenaria ed avviene ad opera dell'apparato trascrizionale cellulare.

Il capsido vuoto si forma nel nucleo, riconosce le strutture a forcina che legano covalentemente la proteina NS e il genoma viene compattato all'interno del nucleocapside. I nuovi virioni contengono quindi il DNA genomico legato covalentemente alla proteina NS, che viene posizionata all'esterno della struttura del nucleocapside [47]. La liberazione della progenie virale avviene infine per lisi cellulare (**Fig. 17**).

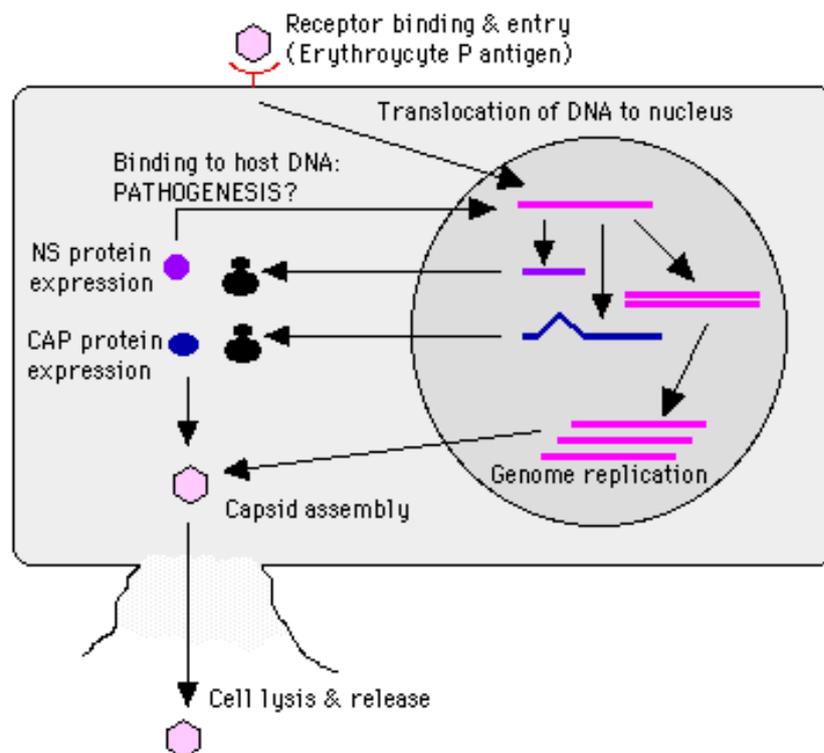


Fig. 17: Ciclo replicativo di un Parvovirus

Epidemiologia e sintomatologia

L'infezione colpisce prevalentemente i bambini. Le porzioni di campioni respiratori, raccolti da bambini sintomatici ospedalizzati, positivi alle indagini per hBoV riportate finora sono comprese fra l'1,5% e il 19% [50, 51], mentre percentuali più basse, comprese fra lo 0,9% e il 9,1%, sono riportate per quanto riguarda analisi di campioni fecali [52, 53]. Le discrepanze tra i dati di frequenza riportati tra i diversi studi possono probabilmente essere ricondotte al tipo di casistica analizzata (età e stato di ospedalizzazione dei pazienti) e dal metodo di rilevazione utilizzato per l'indagine (sensibilità e specificità).

La maggior parte dei bambini infettati ha un'età inferiore ai 24 mesi, ma possono essere colpiti anche bambini con età maggiore [54]. Sono stati riportati in letteratura anche studi che hanno considerato l'infezione negli adulti, tuttavia queste analisi riportano basse frequenze di infezione [45,55]. Nonostante non sia facile trarre conclusioni riguardo l'epidemiologia di un virus di così recente identificazione a partire da studi che analizzano la circolazione in aree geografiche limitate in singole stagioni, è tuttavia possibile avere un'indicazione di quale sia la circolazione di hBoV nella popolazione. Molti studi epidemiologici svolti in aree temperate mostrano come la maggior frequenza di infezione sia presente nei mesi invernali e primaverili, anche se il picco varia di anno in anno [56].

La diffusione di questo virus avviene per trasmissione aerea e per contatto diretto, come per gli altri virus respiratori. Inoltre, dato che è nota la presenza del virus nelle feci, va anche considerata un'eventuale via di trasmissione feco-orale [45].

I sintomi clinici più frequentemente riportati in associazione a infezioni coinvolgenti hBoV come singolo agente infettante includono tosse, rinorrea e febbre, che rappresentano i più comuni sintomi non specifici associate a infezioni respiratorie. Le più comuni diagnosi cliniche associate all'infezione, con o senza co-infezioni, risultano invece URI, bronchiti, bronchioliti, polmoniti ed esacerbazioni asmatiche acute [45].

2.2.3 Human Polyomavirus Ki e Wu

Le infezioni virali persistenti sono parte integrante della vita umana. La maggior parte delle persone è continuamente infetta con uno o più herpesvirus, papillomavirus, polyomavirus e rimanere sana. Tuttavia, molti di questi virus possono occasionalmente produrre una grave malattia. I Polyomaviruses sono piccoli virus a DNA in grado di infezione persistente e un potenziale oncogeno ("Polyoma" dal greco poli-molti e oma, cioè i tumori). Sono stati trovati in molti mammiferi e uccelli in tutto il mondo. Due polyomaviruses sono noti per infettare gli esseri umani: il virus JC (JCV) e virus BK (BKV), entrambe scoperti nel 1971 [57, 58, 59]. Sono geneticamente collegati gli uni agli altri, ed entrambi i virus mostrano il 70-80% sieroprevalenza negli adulti. I percorsi di acquisizione e di siti di infezione primaria sono in gran parte sconosciute, ma entrambi i virus è possibile stabilire una infezione latente nei reni e, in caso di JCV, anche nel sistema nervoso centrale. BKV è stato anche rilevato nelle feci di bambini. JCV e BKV sono altamente oncogeni negli animali da esperimento, ma un ruolo nello sviluppo dei tumori umani non è stato ancora stabilito. Una malattia causata da polyomaviruses umani è stata osservata in soggetti immunodepressi. JCV è l'agente eziologico della leucoencefalopatia multifocale progressiva, una malattia demielinizzante del cervello e una complicazione temuta di AIDS. BKV è stato associato alla nefropatia post-trapianto e la cistite emorragica nel trapianto di cellule staminali ematopoietiche (HSCT) destinatari.

Oltre a questi, nel 2007 sono stati identificati altri due polyomavirus umani che sono lontanamente imparentati con i primi e infettano l'apparato respiratorio. Questi sono denominati Polyomavirus KI e WU [57, 60, 61].

Struttura e organizzazione genomica

I Polyomaviruses (PyV) sono virus di forma icosaedrica senza envelope, di circa 45 nm di diametro (**Fig. 18**). Il capsidè di 72 capsomeri, 60 esagonali e 12 pentagonali ai vertici, è composto da 3 proteine VP1, VP2 e VP3. Oltre a queste il genoma codifica per altre 2 proteine, T e t, che sono implicate nel meccanismo della replicazione del DNA.

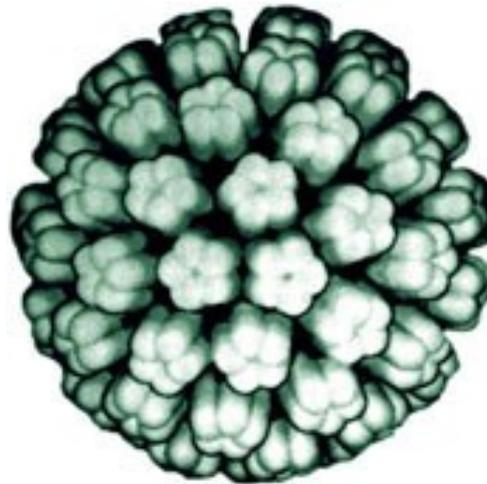


Fig. 18: *Struttura Polyomavirus*

All'interno delle particelle virali, il DNA assume una forma supercoiled, dove quattro istoni H2A cellulare, H2B, H3 e H4 sono associati con esso [61]. L'acido nucleico dei Poliomavirus è formato da una doppia elica (ds) circolare di DNA, di lunghezza di circa 5 kbp. Questa organizzazione genomica è stata progettata per dare il massimo pacchetto informativo (6 geni) in uno spazio minimo (5kbp). In particolare, il genoma di WUPyV ha una lunghezza di 5229 paia di basi suddiviso in: una regione "*early*" codificante per Small T Ag e per Large T Ag e una regione "*late*" che codifica per le proteine capsidiche VP1, VP2 e VP3. Queste due regioni sono separate da una regione regolatoria che contiene: una regione ricca di AT, tre ripetizioni del sito di binding GAGGC del pentanucleotide consenso LTag e una copia del sito di binding TAGGC di LTag non-consenso. A livello della regione *early* è presente un open reading frame di 194 aminoacidi che probabilmente codifica per STAg; invece nella regione *late* del genoma sono presenti delle ORFs per le proteine VP1-3 [61] (**Fig. 19**).

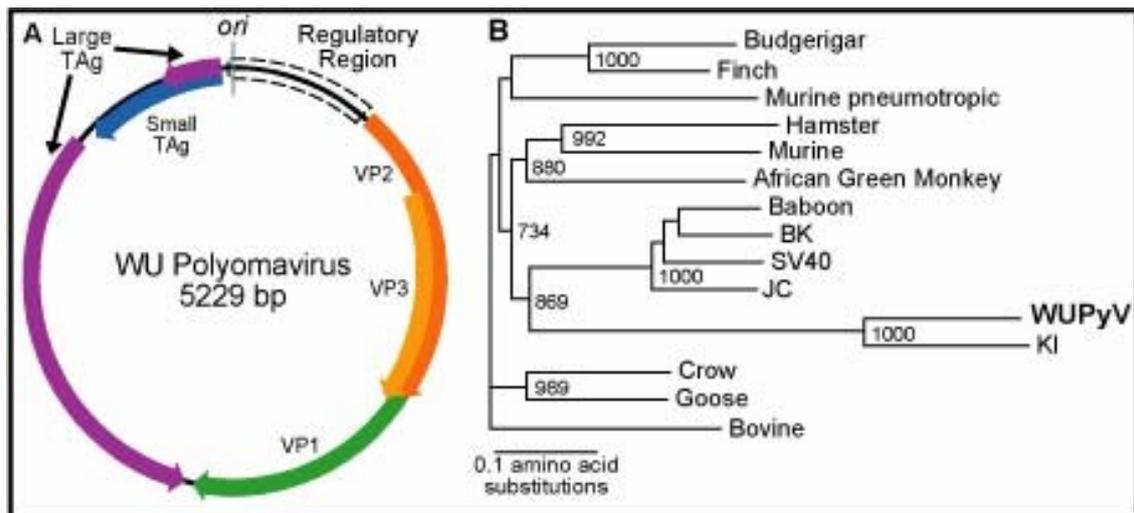


Fig. 19: A. Organizzazione genomica del Polyomavirus WU. B. Analisi filogenetica della proteina capsidica VP1

Il genoma di KIPyV è molto simile a quello di WUPyV, anch'esso suddiviso in una regione *early* codificante per Small T Ag e per Large T Ag, una regione *late* che codifica per le proteine capsidiche e una regione regolatoria che separa le prime due. A differenza del polyomavirus WU, il polyomavirus KI ha un genoma leggermente più piccolo di lunghezza di 5040 bp.

La regione regolatoria non codificante dei polyomavirus contiene i promotori per la trascrizione genica *early* e *late*, l'origine della replicazione e gli enhancers trascrizionali. Nella regione *early* si trovano delle presunte ORFs per le 2 proteine regolatorie (ST antigen e LT antigen) invece a livello della regione *late* sono presenti delle open reading frame per le proteine capsidiche [61] (**Fig.20**).

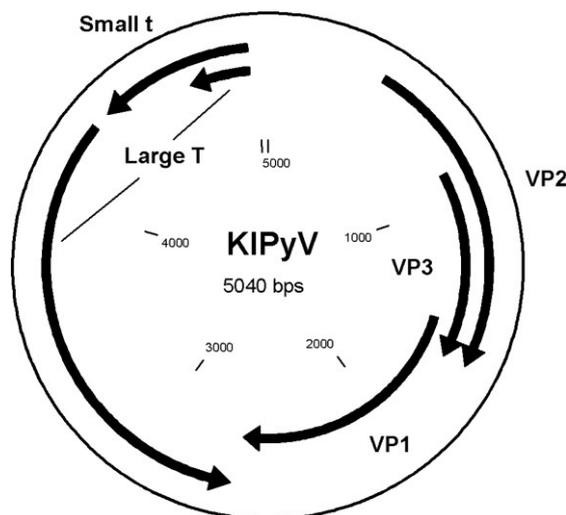


Fig. 21: Organizzazione genomica del Polyomavirus KI.

Ciclo Replicativo

L'attacco dei Polyomavirus alle cellule ospiti è mediato dalla proteina virale 1 (VP1) attraverso la regione di acido sialico. I virioni sono successivamente inglobati nella cellula per endocitosi e trasportati prima al reticolo endoplasmatico dove si verifica un cambiamento conformazionale della proteina VP2 e poi tramite un meccanismo poco conosciuto trasportati verso il nucleo dove avviene la replicazione. Essi utilizzano i meccanismi di replicazione della cellula ospite. La replicazione virale avviene in tre fasi distinte: espressione genica precoce (Early), replicazione del genoma e espressione genica tardiva (Late). La fase "Early" è responsabile della sintesi delle proteine non strutturali.

Durante la replicazione, il genoma virale è ripetutamente sintetizzato ed elaborato come se fosse il DNA cellulare, sfruttando macchinari dell'ospite. Il DNA virale sintetizzato viene associato a nucleosomi cellulari per formare strutture che sono spesso indicati come "minichromosomes"; in questo modo il DNA è confezionato in modo più efficiente. Nella fase "Late" si sintetizzano le proteine strutturali, responsabili della composizione delle particelle virali. Ciò si verifica durante e dopo la replicazione del genoma. Come nell'espressione genica precoce, nell'espressione genica late vengono generate una matrice di proteine a seguito di splicing alternativo. All'interno di ogni proteina virale ci sono dei "segnali di localizzazione nucleare" che portano le proteine virali ad accumularsi nel nucleo,

dove successivamente avviene l'assemblaggio. Il rilascio di particelle di nuova sintesi dalla cellula infetta può avvenire tramite due meccanismi:

- i virioni sono trasportati in vacuoli dal citoplasma alla membrana plasmatica e rilasciati all'esterno.
- Più frequentemente, vengono rilasciati quando la cellula si lisa a causa della citotossicità delle particelle virali presenti all'interno (**Fig.21**).

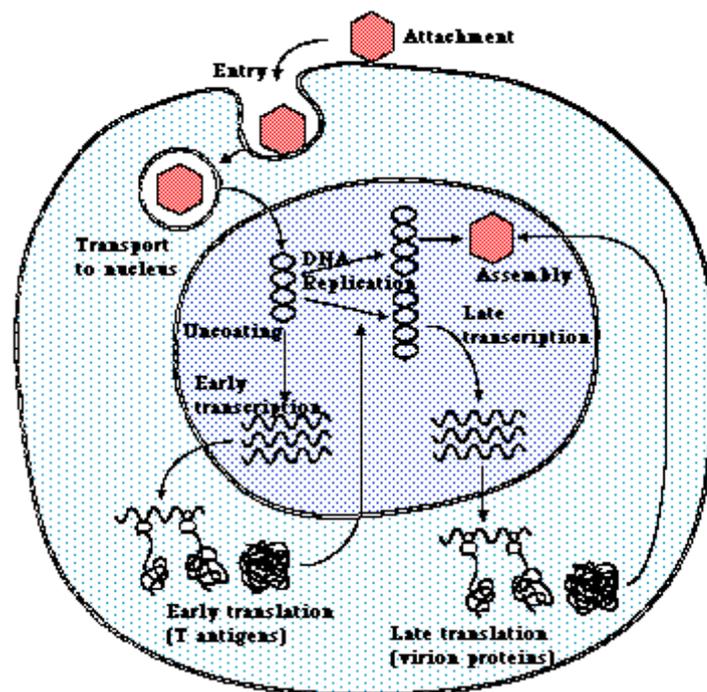


Fig.21: Ciclo Replicativo Polyomavirus

Epidemiologia e sintomatologia

Le percentuali di infezione da parte dei 2 polyomavirus è relativamente bassa [62]. Dalla letteratura è riportato che il virus WUPyV ha una percentuale di infezione compresa fra 1,4% e il 9% e che quella del virus KIPyV è più bassa ed è compresa fra lo 0,2% e il 3% [63]. Considerando però la recente scoperta dei virus e i pochi studi epidemiologici le percentuali sono relativamente diverse da studio a studio. La maggior parte dei bambini infettati ha un'età inferiore ai 18 mesi, ma possono essere colpiti anche bambini con età maggiore.

Un altro dato importante che è stato riportato in diversi articoli è che questi virus vengono ritrovati molto spesso in co-infezione con altri virus respiratori comuni come i virus influenzali, i Parainfluenzavirus, il Metapneumovirus e il Bocavirus.

Questo dato è molto importante da considerare per valutare la sintomatologia provocata dall'infezione di questi due virus. E' stato osservato inoltre che l'infezione è presente anche in molti soggetti asintomatici, di conseguenza è difficile capire effettivamente quale sia la causa scatenante la sintomatologia.

2.2.4 Virus influenzale pandemico A(H1N1)pdm09

I virus influenzali di tipo A, come visto anche in precedenza, sono approssimativi, incostanti e promiscui, e considerati dei veri e propri trasformisti [16]. L'elevato dinamismo genetico dei virus influenzali li rende particolarmente imprevedibili. Infatti, sebbene i piani pandemici mondiali fossero da anni focalizzati sul virus influenzale A/H5N1 quale probabile protagonista della prima pandemia del nuovo millennio [15], dall'aprile 2009 sono stati riportati casi di infezione nell'uomo da un nuovo virus influenzale di tipo A/H1N1 di origine suina [76]. La diffusione dell'infezione è diventata così capillare che, l'11 giugno 2009, l'OMS ha innalzato lo stato di allerta pandemica alla fase di massima allerta (fase 6) [17].

Variatione antigenica ed evoluzione del virus influenzale A(H1N1)pdm09

L'analisi genetica ha dimostrato che questo nuovo virus di tipo A/H1N1 presentava caratteristiche molto diverse rispetto a quelle dei virus responsabili delle comuni epidemie stagionali e definito, per la composizione del suo genoma, come virus quadruplo riassortante. Questo nuovo virus è, infatti, il prodotto di un riassortimento genico tra un virus triplo riassortante, circolante negli allevamenti suini del nord America già dalla seconda metà degli anni '90, ed un ceppo suino di origine euro-asiatica. In particolare, il suo assetto genomico è costituito da 5 segmenti derivati dal classico virus suino A/H1N1 nord-americano e da tre segmenti codificanti il complesso polimerasico sia di origine aviaria (PA e PB2) sia umana (PB1) [64] (**Fig. 22**). Secondo il sistema di sorveglianza dei CDC di Atlanta, già negli anni 2005-2008, undici persone tutte venute a stretto contatto con suini, erano state infettate con il virus triplo riassortante A/H1N1, indicando da un lato la capacità di questo virus di effettuare, seppur occasionalmente, il salto di specie suino-uomo e, dall'altro, la sua sostanziale inefficienza di trasmissione interumana [17, 64].

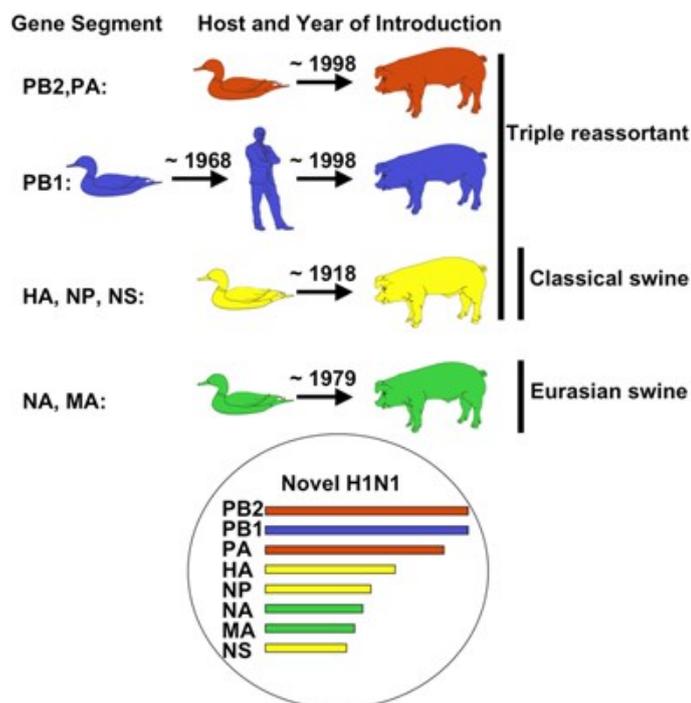


Fig. 22: Riassortimento genico del virus A/H1N1 v

Epidemiologia e sintomatologia

Secondo INFLUNET [65], il sistema di sorveglianza della sindrome influenzale coordinata dal Ministero della Salute che si avvale della collaborazione dell'Istituto Superiore di Sanità (ISS), del Centro Interuniversitario per la Ricerca sull'Influenza (CIRI), dei medici e pediatri di libera scelta, dei laboratori di riferimento per l'influenza e degli Assessorati regionali alla Sanità, il numero cumulativo dei casi di sindromi influenzali stimati nel nostro Paese dall'inizio della sorveglianza durante il periodo pandemico (dalla settimana 43/2009 fino alla settimana 18/2010) è stato di 5.582.000. La fascia d'età più colpita è stata quella pediatrica (0-14 anni), mentre quella meno colpita la fascia d'età ≥ 65 anni, probabilmente perché, come riportato dal CDC, una percentuale di essi sarebbe stata dotata di anticorpi cross-reattivi verso il nuovo virus A(H1N1)pdm09. La curva epidemica ha raggiunto il picco nella 46ma settimana con un'incidenza pari a 12,96 casi per mille assistiti per poi diminuire progressivamente nelle successive settimane (**Fig. 23**).

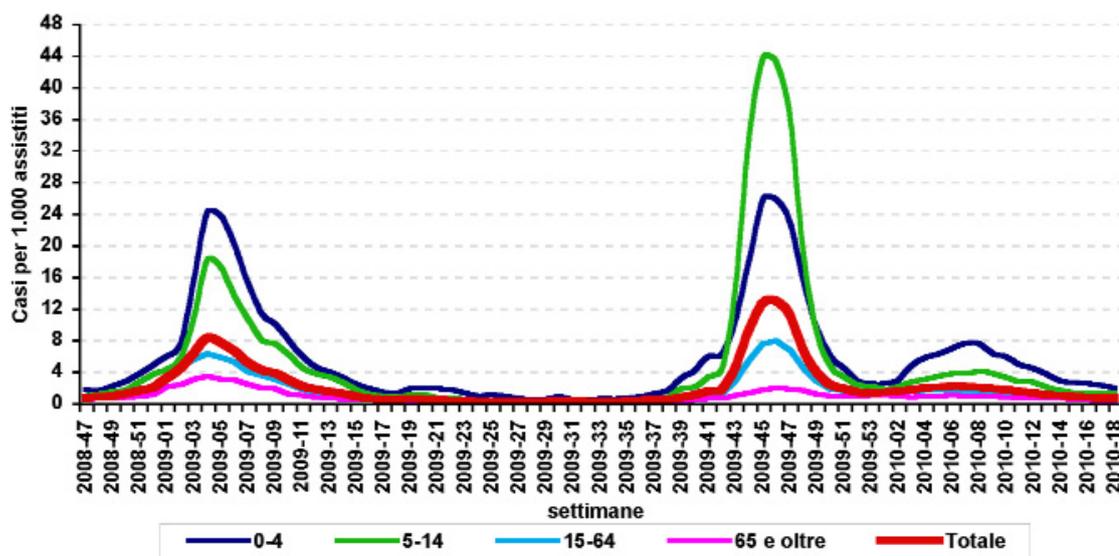


Fig. 23: Curva epidemica dei casi di influenza nelle stagioni 2008/2009 e 2009/2010

Per garantire la massima omogeneità di rilevazione, è stata fornita una definizione di caso clinico affetto da virus influenzale A(H1N1)pdm09 che includeva le manifestazioni acute con sintomi generali e respiratori [65].

La prima sintomatologia si presentava come infezione respiratoria acuta ad esordio brusco ed improvviso con febbre $\geq 38^{\circ}\text{C}$ accompagnata da almeno un sintomo tra i seguenti: cefalea, malessere generalizzato, sensazione di febbre (sudorazione, brividi), astenia e da almeno uno dei seguenti sintomi respiratori: tosse, faringodinia, congestione nasale.

Per la diagnosi clinica di influenza nel bambino è stato importante considerare quanto indicato per gli adulti tenendo conto che i bambini più piccoli non sono in grado di descrivere la sintomatologia sistemica che invece si può manifestare con irritabilità, pianto e inappetenza. Nel lattante l'influenza è stata spesso accompagnata da vomito e diarrea e solo eccezionalmente da febbre.

Nel bambino di 1-5 anni la sindrome influenzale invece si associava frequentemente a laringotracheite e bronchite e a febbre elevata.

3. Sorveglianza e Diagnosi di infezioni respiratorie virali

L'ampio spettro di sintomi comuni a molti agenti eziologici responsabili di ARI, rende difficoltosa una diagnosi eziologica precisa basata solamente sul quadro clinico del paziente. Gli esami biochimico-clinici di laboratorio riescono a fornire solo un'indicazione circa l'eziologia della malattia: un'elevata velocità di eritrosedimentazione associata a un alto valore di neutrofilia indica un'infezione batterica in corso, mentre uno scarso movimento degli indici infiammatori è tipico di un agente eziologico virale [66]. Quindi, considerando la grande complessità e la variabilità degli agenti responsabili di ARI, risulta di primaria importanza l'attuazione di metodologie diagnostiche sensibili e specifiche che siano in grado di identificare con alta efficienza tutti gli agenti patogeni coinvolti. Inoltre, nella pratica clinica, circa il 12-39% dei casi di ARI resta senza diagnosi nota, lasciando aperta l'ipotesi dell'esistenza di agenti eziologici ancora da determinare. Di grande attualità, infatti, sono le nuove identificazioni di virus coinvolti nella patologia respiratoria, nonché l'emergenza di nuovi patogeni del tratto respiratorio (grazie alla presenza di serbatoi animali) in grado di diffondersi rapidamente nella popolazione (ad esempio il virus influenzale A(H1N1)pdm09) o di causare patologia severa nell'uomo (ad esempio SARS-CoV/MERS-CoV). Appare quindi evidente come risultino fondamentali sia l'implementazione e il continuo aggiornamento delle procedure diagnostiche atte all'identificazione degli agenti respiratori già circolanti, con particolare riferimento a quelli di nuova identificazione, sia lo sviluppo di procedure efficaci per l'identificazione precoce e tempestiva dei ceppi emergenti e la caratterizzazione di agenti già circolanti nella popolazione ma non ancora identificati. È di grande importanza altresì, sia in ambito clinico, ma soprattutto in un contesto epidemiologico, valutare quale sia il singolo contributo dei numerosi agenti patogeni che possono essere causa di patologia respiratoria, monitoraggio che risulta possibile solo grazie all'attuazione di tecniche diagnostiche sensibili e specifiche.

3.1 Diagnosi di infezione virale

3.1.1 Campioni biologici di partenza

I campioni di materiale patologico da esaminare variano in relazione alla sintomatologia, alla localizzazione del processo infettivo ed ai particolari quesiti diagnostici.

Per le infezioni delle alte vie respiratorie il campione di partenza maggiormente utilizzato è rappresentato da essudato, muco o materiale necrotico prelevato mediante tampone dalla superficie della mucosa del distretto naso-faringeo (tampone naso-faringeo) o da quello faringeo-tonsillare (tampone oro-faringeo) [10].

Per la diagnosi delle infezioni delle basse vie respiratorie le tipologie di campioni principalmente utilizzate sono:

- espettorato opportunamente raccolto, facendo eseguire al paziente colpi di tosse profondi;
- broncoaspirato, ottenuto in corso di broncoscopia;
- lavaggio bronco alveolare (BAL), ottenuto mediante inserimento in un segmento polmonare di soluzione fisiologica sterile (tramite broncoscopio) e successiva ri-aspirazione della stessa.

Inoltre, per le diagnosi sierologiche dirette o indirette, possono essere utilizzati campioni di sangue [10].

Dato l'ampio spettro di sintomi comuni a molti agenti eziologici responsabili delle ARI, non è possibile effettuare una diagnosi precisa basandosi solamente sul quadro clinico del paziente.

3.1.2 Isolamento virale

Le prove di isolamento virale fino ad una decina di anni fa rappresentavano il "gold standard" nella rilevazione dei virus; attualmente vengono eseguiti solo in pochi centri specializzati, e non sono adatti ad una diagnosi rapida.

Generalmente, per i virus influenzali l'isolamento viene effettuato su cellule renali canine MDCK (*Madin-Darby Canine Kidney*), mentre per gli adenovirus e l'RSV viene utilizzata la linea Hep-2 (*Human larynx Epithelioma cell line*); un'ulteriore linea cellulare disponibile nell'isolamento dei virus respiratori è quella del rene di scimmia VERO (*Green Monkey Renal*

Fibroblast). Dopo l'inoculo le colture vengono controllate giornalmente per valutare la comparsa di un effetto citopatico (ECP).

Nonostante il metodo colturale sia ormai superato nella diagnosi virologica, rimane tuttavia utile nello sviluppo di vaccini, nella scoperta di nuovi virus, e come risorsa per l'analisi di cambiamenti genetici e antigenici nelle popolazioni virali [3].

3.1.3 Saggi sierologici

La diagnosi sierologica si basa sulla determinazione degli anticorpi IgM ed IgG diretti contro gli antigeni virali, presenti nel siero del paziente, ad indicare rispettivamente un'infezione presente o passata. Per questa metodologia è necessario avere il siero del paziente prelevato in fase acuta (esordio clinico) e dopo 20-30 giorni. La presenza di sierconversione o un incremento del titolo anticorpale di 4 volte tra la fase acuta e la fase convalescente indicano la presenza di un'infezione virale.

Le metodiche correntemente impiegate per la ricerca di anticorpi specifici sono l'inibizione dell'emoagglutinazione, la fissazione del complemento, l'immunofluorescenza diretta e i test immunoenzimatici EIA.

Il test di inibizione dell'emoagglutinazione (HI) si basa sulla capacità di legame tra le HA (emoagglutinine) di superficie di alcuni virus (quali quelli influenzali) e gli eritrociti di alcune specie animali (pollo, tacchino, montone), con la conseguente formazione di un reticolo di globuli rossi agglutinati sul fondo del pozzetto; gli anticorpi presenti nel siero del paziente, specifici per il virus, inibiscono tale reazione legandosi alle HA. Le emazie che non vengono agglutinate precipitano sul fondo della piastra sotto forma di un bottone rosso; il titolo inibente l'emoagglutinazione corrisponde al reciproco della diluizione massima dell'anticorpo in grado di inibire l'emogglutinazione.

Le indagini sierologiche possono essere utili nella sorveglianza epidemiologica e nel controllo dell'efficacia vaccinale. Il loro impiego è tuttavia limitato per la bassa sensibilità rispetto ad altri metodi (la presenza del 10-30% di falsi negativi) e perché sono indagini retrospettive. Inoltre, per quanto concerne l'RSV, alcune di queste tecniche, come la fissazione del complemento, non sono adatte alla ricerca degli anticorpi nei bambini piccoli, che rappresentano la popolazione più colpita da questo virus. La giusta collocazione della sierologia a scopo diagnostico potrebbe quindi essere quella di completamento ai test diretti, in grado di rilevare la presenza di antigeni virali.

La ricerca di antigeni virali avviene direttamente nelle secrezioni respiratorie e si basa su tecniche di immunofluorescenza (IFA, *Immunofluorescence assay*) diretta (per i virus influenzali, diretta ed indiretta per RSV) e su test immunoenzimatici.

Per le tecniche di immunofluorescenza, rilevante è la scelta degli anticorpi monoclonali, che devono essere specifici e sensibili, e nella presenza nel campione di un numero sufficiente di cellule esfoliative.

Per quanto riguarda i test immunoenzimatici, si tratta prevalentemente di test ELISA (*Enzyme-linked immunoabsorbent assay*).

3.1.4 Test Rapidi

Attualmente esistono in commercio test cosiddetti 'rapidi' perché consentono una rilevazione qualitativa in pochi minuti (15-30 minuti) degli antigeni virali. Attualmente sono in commercio test rapidi per la diagnosi di influenza di tipo A e B e dell'infezione da RSV. I metodi virologici rapidi potrebbero permettere una diagnosi specifica nelle fasi precoci di insorgenza della malattia, allo scopo di effettuare una terapia mirata dell'assistito, prevenire infezioni di tipo nosocomiale, e allertare, in tempo reale, i sistemi di sorveglianza sulla possibile insorgenza di epidemie. Tali sistemi però sono dotati di una scarsa sensibilità e quindi vengono solo impiegati come ausilio nella diagnosi rapida dell'infezione acuta.

3.1.5 Saggi molecolari

I saggi molecolari sono le nuove metodiche utilizzate nella rilevazione dei virus; la loro attuale diffusione è dovuta alla loro elevatissima sensibilità e specificità (vicine al 100%), e al fatto che riescono a discernere tra sottotipi virali non rilevabili con altri saggi. La tecnica dell'amplificazione genica è in grado di rilevare gli acidi nucleici presenti nel campione da analizzare, indipendentemente dallo stato dell'infezione (passata o in corso) e dalla quantità o tipologia del campione biologico, dal momento che tale metodo non necessita dell'integrità del virus.

Può complementare, se non in alcuni casi sostituire, l'approccio immunologico e colturale, grazie anche alla rapidità raggiunta attualmente dalle metodologie biomolecolari.

La PCR (*Polymerase Chain Reaction*) è la tecnica comunemente utilizzata per amplificare segmenti specifici dell'acido nucleico virale, sia DNA che RNA; il suo utilizzo presuppone la conoscenza della struttura e del genoma di ogni virus al fine di individuare con assoluta

specificità i singoli agenti eziologici. La specificità della reazione di PCR è garantita dall'uso di oligonucleotidi di sintesi (primers) complementari alle estremità 5' e 3' dei due filamenti stampo e specifici per la sequenza che si vuole amplificare.

I primers hanno una lunghezza di 20-30 nucleotidi e sono entrambi orientati con le estremità 3' verso l'interno della sequenza da amplificare in modo che la sintesi avvenga lungo il frammento di DNA fra loro compreso.

Nel disegno dei primer, è importante considerare i seguenti fattori:

- dovrebbero avere una lunghezza compresa fra 17 e 30 basi;
- il contenuto ottimale in GC è di circa 50%;
- sono da evitare primer che contengano più di 3-4 ripetizioni consecutive della stessa base (es.: AAAA);

La reazione di PCR utilizza la ripetizione di tre fasi che si svolgono in diverse condizioni di temperatura (**Fig. 24**):

- nella prima fase, detta di *denaturazione*, la sequenza di DNA viene denaturata al calore mediante incubazione ad alta temperatura (superiore ai 90°C);
- nella seconda fase, detta di *ibridazione*, che si svolge ad una temperatura calcolata sulla base della *T° di Melting*, i primer ibridano con i filamenti stampo ai quali sono complementari;
- nella terza fase, detta di *estensione*, che avviene alla temperatura ottimale per l'attività enzimatica (in genere 72°C), la DNA polimerasi sintetizza una catena complementare a quella originale, in direzione 5' → 3' a partire da ciascun primer ibridato alla sequenza stampo complementare, che funge da innesco per la sintesi.

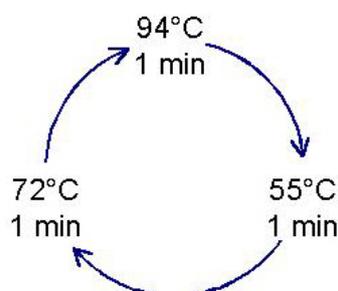


Fig. 24: Esempio di temperature nelle varie fasi di amplificazione.

Mediante la ripetizione dei cicli di denaturazione, ibridizzazione dei primer ed estensione ad opera della DNA polimerasi, si ottiene un processo di amplificazione esponenziale (2^n) della sequenza di DNA bersaglio originariamente presente nel campione. La ripetizione di 20 cicli di PCR genera circa un milione di copie della sequenza bersaglio (**Fig. 25**).

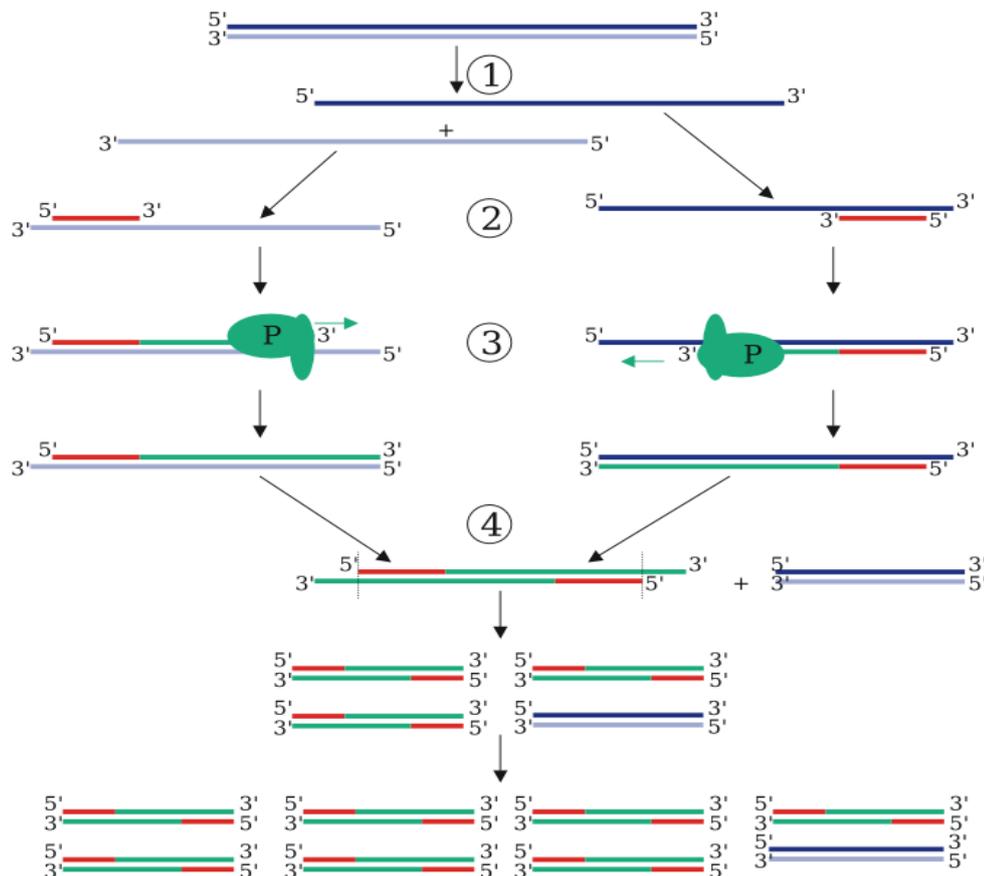


Fig. 25: Schema di amplificazione esponenziale (2^n) di una sequenza di DNA.

Esistono differenti tipologie di PCR. In ambito virologico, i saggi *nested-PCR*, *emi-nested PCR* e *multiplex-PCR* risultano essere tra i più utilizzati.

Il saggio *nested-PCR* consiste in due amplificazioni successive (*double step*) in cui vengono utilizzate due coppie diverse di primer: nel primo step vengono utilizzati due primer (detti *primer esterni*) che amplificano un frammento di DNA di una data grandezza. Nella seconda fase, partendo dall'amplificato già ottenuto, verrà utilizzata una diversa coppia di primer (detti *primer interni*) che si lega in posizione più interna rispetto alla prima; in questo modo si ottiene un frammento di dimensioni più ridotte (in genere tra le 100 bp e le 400 bp).

Il saggio *emi-nested* è una particolare nested PCR in cui, nel secondo step, viene utilizzato uno dei due primer utilizzati nella prima reazione di amplificazione.

Da ultimo, il saggio *multiplex-PCR* è una procedura utilizzata per la ricerca e la simultanea identificazione di diversi frammenti genici: vengono quindi utilizzate contemporaneamente coppie di primer diversi specifici per ognuno dei segmenti genici d'interesse.

I risultati ottenuti dall'amplificazione, sono poi rilevati mediante elettroforesi su gel d'agarosio. L'elettroforesi è un metodo di separazione basato sulla diversa velocità di migrazione di particelle elettricamente cariche attraverso una soluzione, sotto l'influenza di un campo elettrico. Il principio base è quello di un setaccio molecolare attraverso il quale le diverse molecole vengono fatte passare: la velocità di migrazione dipende dalla massa, dalla dimensione, dalla carica e dalla forma delle varie particelle, ossia dalla loro mobilità elettroforetica. Questa grandezza è il rapporto tra la velocità della particella (cm/s) e il campo elettrico utilizzato (Volt/cm). La mobilità elettroforetica, essendo una funzione del rapporto tra carica e raggio, è diversa da una particella ad un'altra: applicando un campo elettrico ad una miscela ionica le varie specie migreranno con velocità diversa a seconda delle rispettiva mobilità.

L'elettroforesi su gel separa i frammenti di DNA in base alla loro dimensione: frammenti di lunghezza minore sono meno ostacolati nella loro corsa attraverso le maglie del gel, quindi si troveranno nella parte bassa, mentre i frammenti più grossi si troveranno localizzati nella parte prossima ai pozzetti di caricamento.

Grazie alla presenza di gruppi fosfato (PO_4^{3-}), il DNA è carico negativamente e migrerà quindi verso il polo positivo (**Fig. 26**).

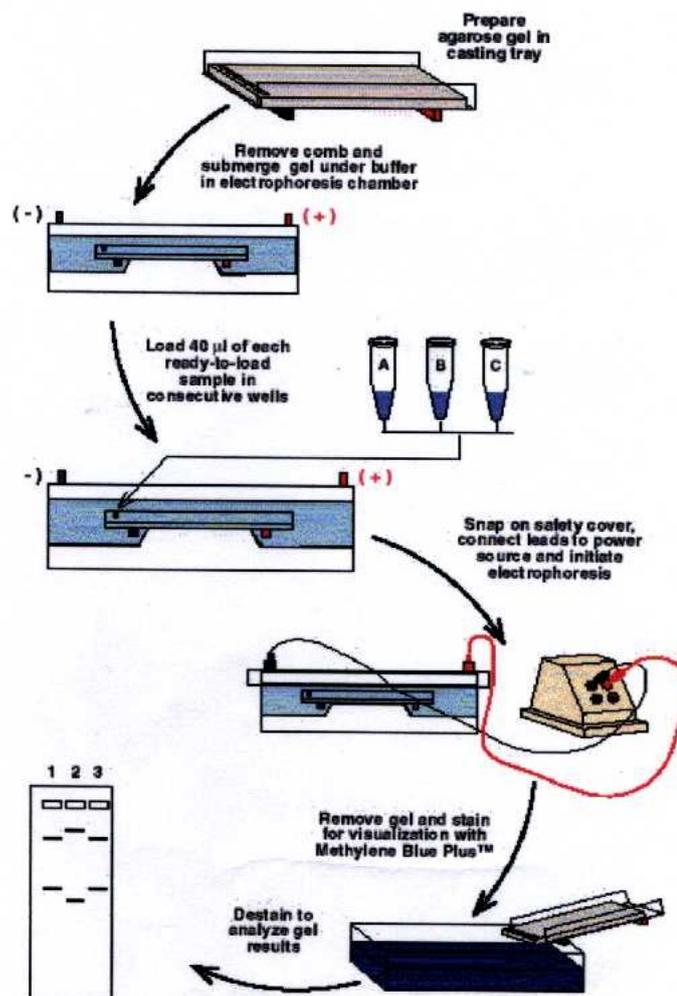


Fig. 26: Schema di preparazione di una corsa elettroforetica

Per concludere, le tecniche di amplificazione genica appaiono lo strumento ideale per attuare una sorveglianza epidemiologica in tempo reale e monitorare eventuali epidemie in corso [3].

3.2 Nuove metodiche di diagnosi di infezione virale

Il progresso della scienza è scandito dall'avvento delle tecnologie rivoluzionarie che forniscono tecniche innovative di investigazione per raggiungere sempre nuovi obiettivi. Sulla scia di microscopia elettronica, colture cellulari e PCR, il sequenziamento di nuova generazione (next-generation sequencing, NGS) è una delle metodologie che sta cambiando il modo in cui si studiano i virus, in particolare nei settori di analisi del genoma, evoluzione, ecologia, scoperta e trascrittomica virale. Ad oggi le possibilità di applicazione di queste metodologie sono limitate, in parte, dovuta al loro costo, che limita l'uso di queste ad un numero relativamente ridotto di campioni. Tuttavia, si stanno portando avanti diversi studi per perfezionare questo tipo di analisi per renderle disponibili ad un costo più accessibile e per sviluppare ulteriori applicazioni in virologia, compreso l'introduzione di queste nella diagnostica di routine.

3.2.1 Next-generation sequencing

Le nuove tecniche di sequenziamento del DNA, tra le quali la tecnica di "next-generation" sequencing, sono caratterizzate da elevata velocità e produttività in grado di produrre un grande volume di sequenze in breve tempo. Il primo strumento di sequenziamento ad alta velocità di NGS, è stata la piattaforma *454 FLX pyrosequencing* (<http://www.454.com>), sviluppata da 454 Life Sciences, poi comprato da Roche e resa disponibile nel 2005. All'inizio del 2007, Illumina (<http://www.illumina.com>) ha lanciato sul mercato la piattaforma *Genome Analyzer* sviluppato da Solexa GA, e più recentemente, SOLiD è stato rilasciato da Applied Biosystems (<http://www.appliedbiosystems.com>). Questo campo è in rapida espansione e nuove piattaforme sono in costante sviluppo e vengono continuamente lanciate sul mercato, come ad esempio la piattaforma *HeliScope*, sviluppata da Helicos (<http://www.helicosbio.com>), *Ion Torrent PGM* ideata dalla Life Technologies (<http://www.iontorrent.com/>) e la piattaforma di sequenziamento sviluppata da Pacific Biosciences (<http://www.pacificbiosciences.com>).

Questi metodi NGS hanno una diversa biochimica di base e si differenziano per il protocollo di sequenziamento e per la produzione di numero di sequenze e lunghezza. Il SOLiD può

essere più adatto per applicazioni che richiedono un alto rendimento di quantità di sequenze, ma non per analisi che richiedono il sequenziamento di sequenze con un elevato numero di basi nucleotidiche, come ad esempio i progetti di sequenziamento di un genoma completo; mentre sia il 454 che l'Illumina forniscono dati idonei per analisi di metagenomica virale e l'analisi di quasispecie virali.

Applicazioni nella diagnosi virologica

Le tecnologie di NGS sono attualmente utilizzate per:

- sequenziamento di genomi completi
- indagini di variabilità a livello dell'intero genoma
- meta genomica
- epigenetica
- gene-expression profiling by RNA sequencing

Le applicazioni tipiche di NGS in microbiologia e virologia oltre al sequenziamento del genoma intero, sono (**Fig.27**):

- identificazione di nuovi microrganismi utilizzando approcci di meta genomica;
- ricerca di comunità microbiche nell'ambiente e nelle nicchie del corpo umano in condizioni di salute e di malattia;
- analisi della variabilità del genoma virale all'interno dell'ospite (quasispecie);
- identificazione di mutazioni di farmaco-resistenza in pazienti affetti da virus dell'immunodeficienza umana (HIV) o epatite virale.

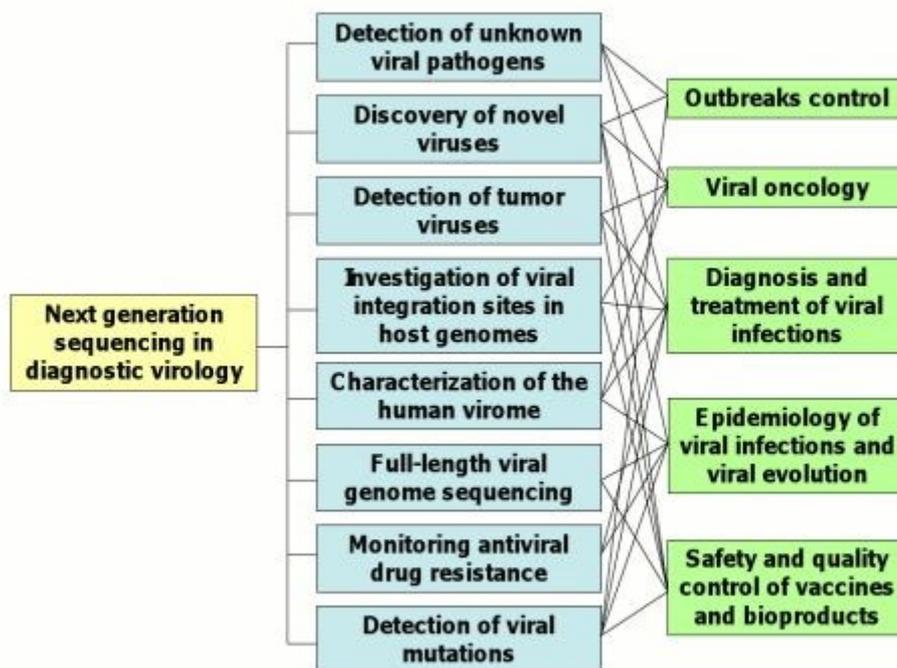


Fig. 27: Schema applicazioni della metodica di NGS virologia

- Identificazione di nuovi patogeni

La popolazione umana è esposta a un numero sempre più crescente di malattie infettive causate dalla comparsa di nuovi virus. I cambiamenti climatici, la globalizzazione, gli insediamenti che sorgono vicini ai diversi habitat animali e l'aumento del numero di persone immunocompromesse, probabilmente ha contribuito alla nascita e la diffusione di queste nuove infezioni [67]. Inoltre, sono diverse le sindromi cliniche che si sospetta siano di natura virale, tuttavia l'agente che le causa non riesce ad essere identificato e riconosciuto dai tradizionali metodi molecolari; quindi, è necessario migliorare continuamente le tecniche di identificazione di nuovi agenti patogeni virali.

A questo scopo, sono stati sviluppati metodi basati sulla NGS. Uno di questi metodi è definito VIDISCA-454, ossia **Virus Discovery cDNA AFLP** (Amplified Fragment Length Polymorphism) associato all'utilizzo della piattaforma **454** di NGS della Roche [68]. In questo metodo, i campioni vengono ultracentrifugati per ottenere una concentrazione maggiore delle particelle virali e trattati successivamente con DNasi e RNasi. Gli acidi nucleici virali vengono poi purificati, convertiti in doppio filamento DNA, digeriti con enzimi di restrizione e

legati ad adattatori oligonucleotidici, che vengono utilizzati come siti di legame di primer per l'amplificazione mediante PCR [69]. Questo metodo è stato descritto originariamente nel contesto la scoperta di un nuovo coronavirus CoV-NL63 [70]. Le tecniche di high throughput NGS rappresentano un potente strumento che può essere applicato anche a strategie di metagenomica per la rilevazione di virus associati a malattie sconosciute e per la scoperta di nuovi virus umani [71,72]. Rispetto ai saggi basati su microarray, metodi di NGS offrono il vantaggio di una maggiore sensibilità.

Una delle prime applicazioni di NGS per la scoperta degli agenti patogeni è stata effettuata su campioni di tre pazienti che erano morti a causa di una malattia febbrile poche settimane dopo il trapianto e per cui microbiologica convenzionale e test molecolari, così come l'analisi microarray per un'ampia gamma di agenti infettivi, non aveva portato a nessun risultato [21]. In questo studio, l'RNA è stato purificato da sangue, liquido cerebrospinale e campioni di tessuto da pazienti trapiantati e, dopo la digestione con DNasi per eliminare il DNA umano, l'RNA è stato retrotrascritto e amplificato con primer random. I prodotti di amplificazione sono stati raggruppati e sequenziati con l'utilizzo della piattaforma pyrosequencing 454. Dopo la sottrazione di sequenze di vertebrati e sequenze altamente ripetitive, sequenze contigue sono state assemblate e confrontate con sequenze riportate nei database, portando all'identificazione di sequenze che erano compatibili con un arenavirus. Ulteriori analisi di sequenza hanno dimostrato che si trattava di un nuovo arenavirus correlato a dei virus che causano coriomeningite linfocitaria. Indagini sierologiche e immunocistochemiche successive hanno documentato che il virus era trasmesso attraverso il trapianto di organi [73].

In studi recenti è stato riportato come l'associazione di tecniche di NGS con la metodica Rolling Circle Amplification (RCA), un processo di amplificazione unidirezionale di molecole circolari di DNA o RNA, ha aumentato notevolmente le sue prestazioni. Recentemente, questa tecnica ha portato all'identificazione del genoma di nuovi papillomavirus e poliomavirus umani [74], tra cui poliomavirus umano 6 e 7 (HPyV6 e HPyV7), rilevate in campioni cutanei di soggetti sani [75], e poliomavirus associati all'insorgenza di trichodysplasia spinulosa (TSPyV), rilevato in lesioni cutanee di pazienti immunocompromessi [76].

Oltre 454 pyrosequencing, anche la piattaforma Illumina GA è stata utilizzata per rilevare nuovi virus in campioni clinici. Illumina GA ha permesso di identificare agenti patogeni virali

in campioni di aspirato nasofaringeo da pazienti con infezioni acute delle basse vie respiratorie [33], come ad esempio un nuovo enterovirus, denominato enterovirus 109 (EV109) rilevato in una coorte di bambini nicaraguensi con malattie respiratorie virali [77].

Infezioni zoonotiche e infezioni trasmesse da vettori rappresentano un altro campo importante e stimolante per la scoperta virale.

I pipistrelli sono serbatoi di virus zoonotici emergenti che causano malattie negli esseri umani e animali, tra i quali: lyssavirus, filovirus e paramyxovirus. In uno studio di sorveglianza focalizzato alla scoperta di agenti patogeni pipistrello-trasmessa, è stato analizzato del tessuto gastrointestinale ottenuto da pipistrelli mediante PCR e successivamente sottoposto a high-throughput pyrosequencing portando alla rilevazione di sequenze di un nuovo coronavirus, legato filogeneticamente al CoV-SARS [78].

3.3 La caratterizzazione virale: i metodi filogenetici

I metodi filogenetici sono ampiamente utilizzati per analizzare e confrontare i dati di sequenza e sono particolarmente utili per l'analisi di geni dei virus a rapida evoluzione (quali i virus a RNA): tali metodi vengono estensivamente impiegati ad esempio per descrivere l'epidemiologia molecolare, le modalità di trasmissione e l'evoluzione di numerosi virus umani come HIV, SARS e Influenza.

I recenti progressi nei metodi filogenetici e nei software per l'inferenza della filogenesi permettono di ottenere informazioni sempre più approfondite sulle dinamiche dell'emergenza delle epidemie virali apportando un contributo fondamentale allo studio epidemiologico.

Esempi di informazioni che possono essere ottenute con questi metodi includono:

- la stima dell'effettivo momento di origine di un nuovo ceppo virale;
- la emergenza di un ceppo virale noto in una nuova specie (salto di specie);
- la presenza di eventi di ricombinazione e/o riassortimento tra genomi virali;
- il tasso di cambio della dimensione della popolazione nelle epidemie virali.

Altra importante applicazione dei metodi filogenetici riguarda lo studio delle modalità di diffusione ed evoluzione dei virus in una specifica popolazione (filodinamica) e/o in un'area geografica definita (filogeografia). Queste informazioni sequenza-derivate sono riassunte negli alberi filogenetici che rappresentano un valido strumento per la definizione ed il miglioramento degli interventi strategici per il controllo delle malattie infettive in territori definiti in ambito di sanità pubblica.

3.3.1 Gli studi di evoluzione molecolare

Gli studi di evoluzione molecolare, con il contributo di altre discipline biologiche più classiche, come la genetica, l'anatomia comparata e l'embriologia, sono uno strumento indispensabile per l'interpretazione dei processi alla base dell'evoluzione della materia vivente e forniscono uno strumento utile per lo studio degli organismi.

Questo tipo di analisi può fornire una serie di informazioni di grande aiuto alla comprensione del ruolo e della funzione di un gene o del suo prodotto. È proprio questo uno degli obiettivi più importanti degli studi di evoluzione molecolare.

Va sottolineato, tuttavia, che gli studi condotti finora dimostrano chiaramente che i meccanismi molecolari alla base dei processi evolutivi sono ancora in gran parte sconosciuti. L'interpretazione dei risultati richiede, quindi, molta cautela.

Le variazioni genetiche, presupposto fondamentale dell'evoluzione biologica, hanno origine spontaneamente in seguito a errori che hanno luogo nel processo di replicazione oppure a mutazioni accidentali, dovute anche a fattori ambientali, che in qualche modo alterano la sequenza nucleotidica. Perché tali variazioni possano risultare evolutivamente rilevanti e quindi significative devono, tuttavia, poter essere trasmesse alla progenie.

Il processo durante il quale la mutazione si propaga nella popolazione è definito fissazione. Da un punto di vista pratico una mutazione si considera fissata all'interno di una popolazione quando raggiunge una frequenza apprezzabile, cioè superiore all'1%.

Dato che una mutazione solitamente produce un cambiamento fenotipico, si preferisce utilizzare il termine di variante genetica quando non è noto se il cambiamento osservato nella sequenza è o meno neutrale.

3.3.2 Principali fattori in grado di promuovere l'evoluzione virale

I DNA virus e gli RNA virus differiscono tra loro sulla base della frequenza con cui vanno incontro a mutazioni spontanee. È necessario ricordare che le DNA polimerasi, sia cellulari che virali, sono dotate di un sistema per controllare il verificarsi di eventuali mutazioni, e per eliminarle, molto più efficacemente delle RNA-polimerasi RNA dipendenti e della trascrittasi inversa. Per questo motivo i retrovirus e gli altri virus a RNA sono geneticamente molto più variabili dei DNA virus.

L'elevata variabilità virale dipende, dunque, da fattori intrinseci al virus e si misura in termini di "*Mutation Rate*", cioè la velocità di comparsa di mutazioni nucleotidiche nell'acido nucleico. Quando una mutazione si afferma nella popolazione virale allora si parla di "*Evolutionary Rate*" o "*Fixation Rate*", che esprime la velocità di fissazione di un mutante, espressa come sostituzioni/sito/anno.

L' "*Evolutionary rate*", o probabilità di fissazione della mutazione, dipende da diversi fattori intrinseci ed estrinseci alla popolazione virale: il "*Mutation rate*", il tempo di generazione (*Generation time*), e le forze evolutive.

Nel 1967 Motoo Kimura enunciò la teoria neutrale dell'evoluzione molecolare [79]. L'elevato polimorfismo è dovuto ad un accumulo passivo di mutazioni adattativamente neutrale, che non aumentano né diminuiscono la probabilità di sopravvivenza della popolazione.

Gran parte delle mutazioni che si verificano non provocano variazioni evidenti (o funzionalmente importanti) per la proteina virale, e quindi sono silenti per la selezione naturale.

Queste mutazioni restano fissate nella popolazione per caso, non per azione della selezione (*genetic drift*); il ruolo della selezione è, dunque, minore: elimina le mutazioni deleterie e fissa quelle vantaggiose. Oggi gli studiosi indicano come veritiera la teoria della "*Nearly neutral evolution*" [80]. Secondo questa teoria esistono mutazioni neutrale e mutazioni solo debolmente vantaggiose o dannose. Solo sulle mutazioni non neutrale può agire parzialmente la selezione.

Le forze evolutive che condizionano la probabilità di fissazione di una mutazione sono: le pressioni selettive positive e negative (*fitness*) e la deriva genetica (*genetic drift*).

Alcuni fattori sono deterministici, ossia cambiamenti prevedibili sulla base della selezione, mentre altri stocastici, cioè fattori casuali che possono modificare la frequenza degli alleli. La dimensione della popolazione determina la relativa importanza di questi meccanismi.

La “*Fitness*” è l’adattamento replicativo di un organismo al suo ambiente, cioè la sua capacità di riprodursi. Il cambiamento radicale delle condizioni ambientali fa sì che un mutante non sia più selezionato e perda capacità riproduttiva/*fitness* (esempio: perdita di infettività nei ceppi attenuati; ceppi resistenti ai farmaci).

Una mutazione che si verifica in un gene può portare ad un cambiamento nel fenotipo, influenzando la *fitness*. Se la mutazione è deleteria sarà rimossa dalla popolazione (pressione selettiva negativa); se, invece, la mutazione è vantaggiosa sarà spinta ad affermarsi (pressione selettiva positiva). Se la mutazione è neutra non è soggetta a forze selettive.

Una mutazione nucleotidica che non provoca cambiamenti a livello aminoacidico è detta **mutazione sinonima** (o silente); viceversa, mutazioni che provocano un cambiamento nella sequenza aminoacidica vengono dette **non sinonime**.

La probabilità di mutazioni sinonime e non sinonime è diversa a seconda della posizione, relativamente al codone, in cui la mutazione si verifica. Una mutazione all’interno della 1° o 2° posizione di un codone ha una probabilità del 96-100% di essere non sinonima mentre una mutazione nella 3° posizione di un codone ha una probabilità del 30% di risultare non sinonima.

Le mutazioni sinonime sono spesso invisibili alla selezione naturale mentre le mutazioni non sinonime sono soggette a pressione selettiva (positiva o negativa).

Ogni mutazione che non causa cambiamenti significativi della *fitness* virale è soggetta a deriva genetica. La probabilità che una mutazione neutrale si affermi nella popolazione dipende dall’entità della stessa.

Il secondo fattore in grado di promuovere efficacemente l’evoluzione dei virus è la **ricombinazione virale**. Questo fattore si verifica quando due virus diversi infettano contemporaneamente la stessa cellula.

Al momento della sintesi degli acidi nucleici virali si possono verificare i seguenti fattori di ricombinazione:

- **ricombinazione intramolecolare**, consiste nello scambio di sequenze tra virus diversi, ma appartenenti alla stessa famiglia o genere. In questo modo, si possono originare nuovi virus. In condizioni sperimentali la ricombinazione può avvenire anche tra virus

appartenenti a famiglie diverse. Nei retrovirus la ricombinazione può avvenire tra le due catene di DNA retrotrascritto (DNA – e DNA +), come anche tra il DNA provirale e quello cellulare: questo ultimo evento può portare all'assunzione di un oncogene cellulare da parte di un virus.

- **riassortimento genetico**, si verifica principalmente nei virus dotati di genoma a RNA segmentato (come gli *Orthomyxovirus*). Quando due virus correlati si replicano nella medesima cellula può verificarsi uno scambio di segmenti con la produzione di svariati nuovi fenotipi. È un fenomeno molto importante che si verifica costantemente in natura.
- **riattivazione**, avviene quando due virus dello stesso ceppo, che hanno subito mutazioni letali, e pertanto non riescono a moltiplicarsi, vengono introdotti nella stessa cellula; se le mutazioni sono situate su geni diversi possono compensarsi consentendo la formazione di una nuova progenie infettiva.

La reale incidenza dei fenomeni di ricombinazione in natura non è, tuttavia, ancora ben definita anche se, visto l'importanza di questi fenomeni per l'evoluzione virale, è bene tenerne conto.

Infine, cambiamenti strutturali nei virus possono verificarsi, in infezioni miste, anche senza interscambio di materiale genetico, ma per scambio di proteine strutturali dell'*envelope* o del capsido. Si parla in questo caso di "**mixing**" fenotipico. Per esempio, i retrovirus *envelope* difettivi possono introdurre il loro capsido in un *envelope* preso "a prestito" da un virus *helper* non difettivo, acquisendo la possibilità di infettare altre cellule, tali virus sono detti **pseudotipi**.

Per quanto riguarda il capsido, il *mixing* fenotipico può essere parziale (scambio di alcuni capsomeri) o completo: in questo caso si parla di **transcapsidazione**. Tutti questi fenomeni possono comportare, come ovvia conseguenza, un cambiamento nel tropismo dei virus stessi [81-83].

3.3.3 La rappresentazione delle relazioni evolutive mediante alberi filogenetici

Le relazioni evolutive tra gli organismi, o più in generale tra geni omologhi, possono essere rappresentate attraverso **“alberi filogenetici”**. Un albero filogenetico è un grafico costituito da nodi e rami, in cui ogni ramo mette in relazione solo due nodi. I nodi rappresentano le unità tassonomiche mentre i rami definiscono le relazioni tra queste in termini di discendenza e ascendenza.

In un albero filogenetico si devono distinguere i nodi terminali da quelli interni. I nodi terminali rappresentano le unità tassonomiche attuali mentre i nodi interni rappresentano le unità tassonomiche ancestrali.

La struttura complessiva dell'albero si definisce **“topologia”**. Un albero in cui la lunghezza dei rami è proporzionale alla distanza evolutiva tra i nodi è definito **filogramma**. In questo tipo di albero, detto anche albero adattivo, la distanza evolutiva tra i nodi corrisponde alla loro divergenza genetica (espressa solitamente come numero di sostituzioni per sito e calcolata con uno dei metodi stocastici elencati precedentemente). Un albero che descrive semplicemente le relazioni filogenetiche tra i vari nodi, senza che la lunghezza dei rami sia proporzionale alla distanza evolutiva, viene detto, invece, **cladogramma**.

Gli alberi filogenetici possono essere rappresentati con (*rooted*) o senza (*unrooted*) radice. La radice di un albero è il nodo che rappresenta il comune progenitore di tutti i nodi rappresentati nell'albero. Un albero senza radice descrive esclusivamente le relazioni evolutive tra i nodi terminali senza fornire alcuna informazione circa il processo evolutivo in funzione del tempo. In altre parole non si può stabilire la collocazione temporale dei nodi interni, in termini di quali siano più antichi e quali più recenti in un albero senza radice.

Solitamente gli alberi vengono prodotti con una radice quando si assume l'ipotesi dell'orologio molecolare mentre se si suppone una diversa velocità di evoluzione lungo i diversi rami dell'albero possono essere determinati solo alberi senza radice. La radice può essere, tuttavia, collocata, anche nel caso di un albero senza radice, in modo indiretto, all'interno del ramo che congiunge i nodi terminali analizzati a un nodo terminale di riferimento che viene definito **“outgroup”**. L'*outgroup*, per un certo gruppo di nodi terminali, è rappresentato, dunque, da uno o più nodi terminali che, sulla base di dati provenienti da fonti non molecolari (paleontologia, morfologia...), ha iniziato il processo di divergenza dal suo progenitore prima che i nodi terminali in esame abbiano iniziato a divergere tra loro.

I metodi utilizzati comunemente per la costruzione degli alberi filogenetici possono essere classificati sulla base del tipo di metodologia utilizzata o della natura dei dati analizzati.

Si possono adottare due diverse metodologie:

- metodi che applicano algoritmi di *clustering* o raggruppamento dei nodi terminali analizzati, generalmente basati su misure di distanze genetiche;
- metodi che utilizzano algoritmi che massimizzano una funzione obbiettiva di qualità dell'albero (criterio di ottimalità).

Uno dei metodi maggiormente utilizzati in letteratura per lo studio dell'evoluzione genetica è il metodo ***Neighbor-Joining*** [84]. Questo metodo utilizza un algoritmo di *clustering* che parte da una topologia a stella (albero completamente non risolto) e, quindi, attraverso interazioni successive, determina i rami tra le coppie di nodi terminali più vicine (*Neighbors*) e i restanti nodi terminali dell'albero.

Se si considera un insieme di N nodi terminali si avrà $N(N-1)/2$ possibilità di scegliere la prima coppia di *Neighbors*. Tra tutte le possibili si sceglierà quella che minimizza la somma delle lunghezze di tutti i rami dell'albero. La coppia di nodi terminali prescelta costituirà un nodo terminale composito. La procedura viene ripetuta finché non siano determinati tutti i possibili $N-3$ rami interni dell'albero.

Il metodo di *Neighbor-Joining* è molto rapido e tiene conto delle diverse velocità di evoluzione lungo i diversi rami dell'albero ed è generalmente affidabile. L'applicazione di questo metodo consente di stimare la lunghezza di tutti i rami dell'albero filogenetico.

Tuttavia esistono altri metodi per la costruzione di alberi filogenetici.

Il metodo ***UPGMA*** [85] è basato su un algoritmo di *clustering* e presuppone la validità dell'orologio molecolare, ovvero che la velocità di evoluzione delle sequenze sia costante lungo tutti i rami dell'albero. Questo metodo utilizza un algoritmo di *clusterizzazione* iterativo che procede associando via via le sequenze o *cluster* di sequenze più simili tra loro. L'albero prodotto è detto "ultrametrico" in quanto impone che, a partire da un nodo interno, la lunghezza di tutti i rami discendenti sino ai nodi più esterni sia uguale. Questa costruzione formalizza l'assunzione dell'orologio molecolare dato che è ovvio che lo stesso tempo sia trascorso da un nodo progenitore fino a tutti i nodi esterni.

I metodi di ***Massima Parsimonia*** e di ***Massima Verosimiglianza*** sono basati sulle sequenze.

Il metodo della ***Massima Parsimonia*** [86] consiste nell'identificazione dell'albero filogenetico che richiede il minor numero possibile di sostituzioni che spieghino le differenze

osservate tra le sequenze in esame. Questo metodo è maggiormente indicato per l'analisi delle sequenze aminoacidiche tuttavia può essere utilizzato anche per l'analisi delle sequenze nucleotidiche. È un metodo essenzialmente qualitativo in quanto consente di determinare la topologia dell'albero che descrive le relazioni filogenetiche tra le sequenze omologhe in esame. La lunghezza di ciascun ramo dell'albero, pari al numero minimo di sostituzioni occorse tra i nodi che esso congiunge, non stima in modo accurato l'effettiva distanza genetica in quanto non tiene conto della possibilità di sostituzioni multiple o convergenti che in qualche modo sono, invece, considerate dai modelli stocastici. Per questa ragione il metodo della Massima Parsimonia è classificato come un metodo deterministico. Tuttavia questo metodo presenta diverse limitazioni: non considera alcun modello esplicito di evoluzione molecolare e non tiene conto di sostituzioni parallele, convergenti, o multiple che sono tanto più frequenti quanto maggiore è la distanza tra le sequenze; tutte le sostituzioni sono considerate equivalenti; non tiene conto della composizione nucleotidica delle sequenze in esame e non individua una soluzione univoca ma determina molti alberi, tutti ugualmente parsimoniosi.

Il metodo della **Massima Verosimiglianza** (*Maximum Likelihood*) [87] rappresenta probabilmente l'approccio che più di tutti è in grado di determinare l'albero più consistente con i dati molecolari forniti, cioè il multiallineamento delle sequenze nucleotidiche o aminoacidiche.

L'albero di Massima Verosimiglianza è, dunque, quello che meglio giustifica il set di dati in esame, ovvero il multiallineamento fornito. Praticamente, per ciascun sito del multiallineamento viene calcolata la probabilità che lo stato dei caratteri osservati sia prodotto da un particolare processo evolutivo, posto anche uno dei possibili alberi filogenetici. La probabilità complessiva dell'albero è data dal prodotto delle probabilità calcolate per tutti i siti del multiallineamento. Tutte le variabili vengono calibrate, nell'esecuzione del modello, in modo da massimizzare il valore di verosimiglianza. La principale limitazione del modello di Massima Verosimiglianza è di natura computazionale in quanto è richiesta una grande quantità di calcolo e questo rende di fatto impraticabile l'applicazione del metodo quando il multiallineamento comprende più di 20-30 sequenze [81, 82, 88-90].

3.3.4 L'orologio molecolare

Un'importante proprietà del processo d'evoluzione molecolare è la sua natura generalmente divergente. In altre parole il numero di sostituzioni nucleotidiche o aminoacidiche che si accumulano tra due geni o proteine omologhe è dipendente dal tempo intercorso dalla loro divergenza.

Questa proprietà è stata annunciata per la prima volta nel 1965 da *Pauling e Zuckerkandl* [91] i quali confrontando il numero di sostituzioni tra proteine omologhe con il corrispondente tempo di divergenza osservarono una relazione di proporzionalità diretta che denominarono “**orologio molecolare**”.

La possibilità dell'esistenza di un orologio molecolare implica conseguenze di grande rilevanza in ambito biologico. Infatti, se le proteine e i geni corrispondenti, evolvono a velocità costante, essi possono essere usati per datare la divergenza tra le specie e per ricostruire le relazioni filogenetiche tra gli organismi.

La disponibilità di un numero sempre maggiore di sequenze e anche di vari genomi completi ha permesso di verificare che non esiste un orologio molecolare universale. In realtà ogni proteina, e ogni gene, ha una propria dinamica evolutiva e spesso si osservano accelerazioni o decelerazioni nella dinamica evolutiva di uno stesso gene in alcune linee evolutive, soprattutto tra quelle più divergenti tra loro. Per questa ragione nell'analisi di un insieme di geni o proteine omologhi è più probabile riscontrare un orologio molecolare “locale”, nel senso che la sua validità è limitata a un gruppo tassonomico più o meno ristretto.

Se si considerano sequenze di acidi nucleici si deve, inoltre, considerare che l'osservazione di eterogeneità compositiva (non stazionarietà) è un indice della non applicabilità dell'orologio molecolare.

È dunque evidente che il concetto di orologio molecolare va utilizzato con molta cautela: le macromolecole biologiche non sono metronomi che possono misurare con precisione gli intervalli di tempo come il decadimento radioattivo. Inoltre l'evoluzione molecolare non è un processo solo divergente.

La validità dell'orologio molecolare può comunque essere sottoposta a verifica sperimentale attraverso il test della velocità relativa o altri approcci basati sul metodo della Massima Verosimiglianza [81, 82, 88, 89].

In figura 28 sono riassunte schematicamente le fasi principali di uno studio di evoluzione molecolare descritte precedentemente.

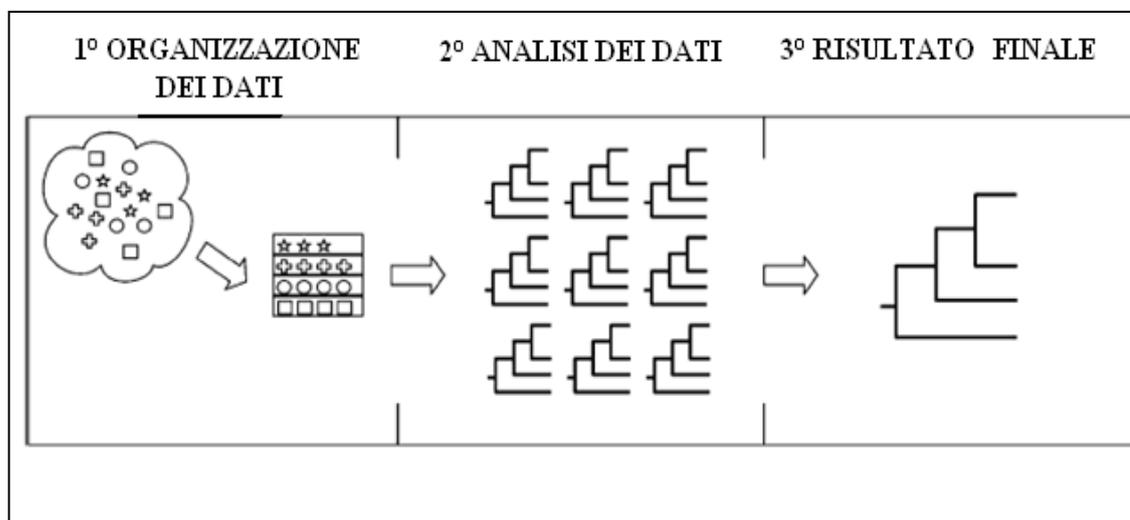


Fig. 28: Fasi principali di uno studio di evoluzione molecolare.

Disegno dello studio

Le infezioni respiratorie acute, di origine virale o batterica, rappresentano la principale causa infettiva di morte nel mondo.

Tra le infezioni virali, l'influenza è certamente la più importante in termini di morbilità, ma non si può sottovalutare il ruolo di altri agenti patogeni.

L'epidemia di SARS, il ripetersi di episodi sporadici di trasmissione di virus dell'influenza aviaria (H5N1, H7N7, H7N9 etc.) direttamente dai volatili all'uomo, il nuovo CoV-MERS evidenziano l'importanza del fenomeno costituito dalle infezioni emergenti.

Dal punto di vista epidemiologico, queste infezioni rappresentano una priorità di sanità pubblica, a causa dell'elevato coefficiente di riproduzione di base dell'infezione (R_0), che consiste nel numero di casi secondari generati da una persona infetta, dell'elevata frazione di popolazione esposta, nonché dall'estrema variabilità del virus.

In questo campo, le attività di ricerca e di sorveglianza epidemiologica non possono non concentrarsi:

- i) sull'identificazione dei virus respiratori circolanti in diversi periodi dell'anno;
- ii) sullo studio di filogenetica e filodinamica, ovvero lo studio dell'interazione fra aspetti evolutivisti e dinamiche epidemiche;
- iii) sull'identificazione di nuovi agenti o nuove varianti virali.

4. Scopo

Lo scopo del presente lavoro di tesi è stato quello di utilizzare nuove metodiche di biologia molecolare, “virus discovery” e bioinformatica al fine investigare e approfondire l’eziologia delle infezioni respiratorie acute in soggetti con sintomatologia respiratoria acuta grave.

Tale lavoro si inserisce all’interno di un progetto di sorveglianza delle infezioni respiratorie virali che coinvolge soggetti con patologie respiratorie, quali soggetti con ILI (*Influenza-like illness*) e pazienti pediatrici con età ≤ 3 anni ospedalizzati per ARI (*Acute Respiratory infections*).

Lo studio è stato condotto presso la Sezione di Virologia del Dipartimento di Scienze biomediche per la Salute in collaborazione con alcune aziende ospedaliere lombarde e con la rete di Medici Sentinella coinvolti nel Sistema di Sorveglianza epidemiologica e virologica dell’Influenza in Regione Lombardia.

La parte relativa all’impiego di una tecnica di “virus discovery” è stata effettuata presso l’Università di Amsterdam nel “Laboratory of Experimental Virology, Department of Medical Microbiology, Center for Infection and Immunity Amsterdam (CINIMA), Academic Medical Center (AMC)”, dove ho svolto un periodo di tirocinio formativo sotto la guida della Prof.ssa Lia van der Hoek.

5. Obiettivi Specifici

In particolare gli obiettivi specifici di questo lavoro sono stati:

5.1 Identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo influenzale pandemico e post-pandemico (2009-2011)

A tal fine sono stati analizzati campioni prelevati da soggetti con sindromi respiratorie acute gravi, andando a valutare:

- la quota di infezioni respiratorie gravi associate, *in primis*, all'infezione da virus influenzale pandemico A(H1N1)pdm09, agente patogeno a più elevata circolazione nel periodo in studio e quello maggiormente correlato all'insorgenza di sintomatologia grave del tratto respiratorio;
- il ruolo di altri virus respiratori conosciuti nella popolazione in studio;
- l'eventuale presenza di virus non ancora noti, nei campioni risultati negativi alle precedenti indagini utilizzando una tecnica di "virus discovery".

5.2 Variabilità genetica ed evoluzione molecolare del virus respiratorio sinciziale (RSV)

Al fine di caratterizzare i ceppi RSV circolanti in Lombardia nel periodo compreso tra il 2006 e il 2012 nella popolazione in studio è stato condotto uno studio filogenetico e filodinamico degli strain di RSV identificati.

Metodi filogenetici computazionali sono stati utilizzati per determinare alcune caratteristiche relative alla comparsa e alla diffusione di RSV nella popolazione umana. Attraverso lo studio dell'evoluzione molecolare del virus è stato possibile definire sia il momento in cui i diversi genotipi virali hanno cominciato a differenziarsi, sia valutare le caratteristiche dell'espansione spazio-temporale del virus. Inoltre, per valutare le forze di pressione selettiva che agiscono sull'evoluzione di RSV, è stata effettuata l'analisi specifica delle mutazioni sinonime e non-sinonime.

Materiali e metodi

6. Popolazione in studio

Obiettivo 1: Identificazione agenti virali

Per l'identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo influenzale pandemico e post-pandemico (Ottobre 2009 - Aprile 2011), sono stati raccolti 206 tamponi oro-faringei (TOF) prelevati da pazienti ospedalizzati per forme respiratorie gravi.

Le forme respiratorie gravi comprendono:

- **Infezione Respiratoria Acuta Severa** (Severe Acute Respiratory Syndrome, SARI): infezione polmonare che richiede il ricovero in terapia intensiva e che implica il trattamento del paziente con terapia ECMO (extra-corporeal membrane oxygenation)
- **Sindrome da Distress Respiratorio Acuto** (Acute Respiratory Distress Syndrome, ARDS): si tratta di una sindrome polmonare, associata a svariate eziologie, caratterizzata da esordio acuto, infiltrazioni polmonari bilaterali suggestivi di edema polmonare, infiammazione del parenchima polmonare con conseguente rilascio sistemico di mediatori infiammatori, rapporto $\text{PaO}_2/\text{FiO}_2 < 200$, ipossiemia e insufficienza polmonare. Questa condizione ha un tasso di mortalità del 90% in pazienti non trattati. Con il trattamento (in genere ventilazione meccanica in un reparto di terapia intensiva), il tasso di mortalità è di circa del 50% [92].

Ogni campione è giunto in laboratorio accompagnato da una scheda di registrazione dei dati demografici e clinici relativi al paziente: sesso, età, diagnosi al ricovero, eventuali patologie concomitanti, data di esordio dei sintomi, data del prelievo.

I 206 TOF analizzati nel presente studio provenivano da:

206 pazienti con età mediana di 44,3 anni (IQR: 49,7 anni; range d'età: 1 mese-89 anni).

In particolare, il 17,5% (36/206) era costituito da bambini con età ≤ 5 anni e il 23,3% (48/206) da adulti con età superiore ai 65 anni. I restanti 122 TOF sono stati raccolti da soggetti con età compresa tra 6 e 64 anni.

Il 61,2% dei soggetti era di sesso maschile e il 38,2% di sesso femminile.

In base alla diagnosi clinica effettuata al momento dell'ospedalizzazione, la popolazione era così distribuita:

- 91 (59,3% maschi; 18,7% età ≤5 anni, 58,2% età 6-64 anni) erano soggetti con SARI, per i quali si è reso necessario il ricovero in terapia intensiva e terapia ECMO.
- 115 (62,6% maschi; 16,5% d'età ≤5 anni, 60% età 6-64 anni) erano soggetti con ARDS.

Dei 115 pazienti con ARDS, 9 (età mediana: 35,6 anni, IQR: 21,4 anni) sono deceduti durante l'ospedalizzazione (case fatality rate CFR: 7,8%).

Nessun soggetto che presentava SARI è deceduto durante il ricovero.

I dati riguardanti la presenza di patologie concomitanti erano disponibili per circa il 70% dei soggetti in studio. Di questi il 67,3% ha mostrato la presenza di patologie concomitanti e, in particolare: il 25,6% erano soggetti immunocompromessi (a causa di cancro, HIV/AIDS, o trattamento prolungato con corticosteroidi), il 19,7% erano soggetti affetti da cardiopatie, il 11,6% presentava asma cronica o malattie polmonari croniche e il 10,4% presentava patologie neurologiche (**Fig. 29**).

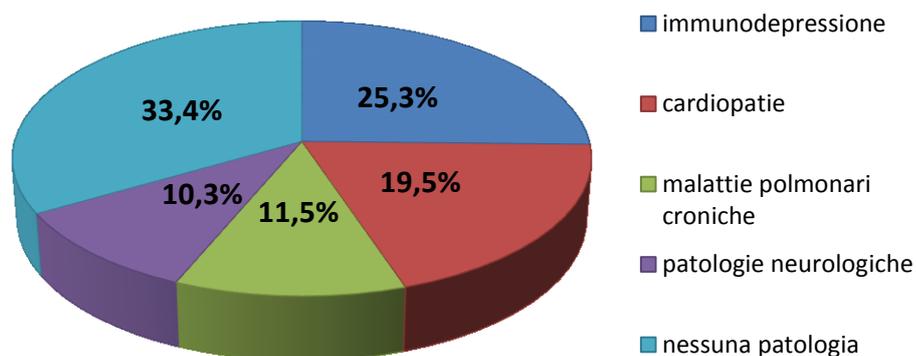


Fig. 29: Manifestazioni cliniche nei soggetti con patologie concomitanti alla sintomatologia respiratoria acuta severa.

Obiettivo 2: Caratterizzazione RSV

Per valutare la variabilità genetica e l'evoluzione molecolare del virus respiratorio sinciziale (RSV) è stato analizzato un campione rappresentativo di sequenze ottenute da prelievi raccolti da soggetti pediatrici con età ≤ 3 anni, ospedalizzati tra il 2006 e il 2012 e risultati positivi alla diagnosi di infezione virale da RSV.

I campioni sono stati analizzati mediante l'utilizzo di una multiplex-PCR per l'identificazione di un frammento del gene F di RSV. Con l'utilizzo di questa metodica è stato anche possibile sottotipizzare i campioni, per RSV di tipo A o B.

Per le successive analisi bioinformatiche sono state utilizzate 23 sequenze ottenute da altrettanti campioni risultati RSV A positivi e 12 sequenze provenienti da campioni RSV-B positivi.

7. Metodi

7.1 Diagnosi molecolare

Il campione biologico, prelevato durante la fase sintomatica acuta, è stato raccolto utilizzando i tamponi Virocult (*MWE Medical Wire & Equipment*, England) [93]. Questo kit è costituito dal tampone per il prelievo a livello oro-faringeo e da una provetta contenente una spugnetta imbevuta di adeguato terreno di trasporto utile per preservare la vitalità del virus.

Il materiale clinico così raccolto è stato conservato a +4 °C fino alla consegna al laboratorio. Successivamente, sono stati aggiunti 2 ml di terreno di coltura (Minimum Essential Medium with Earle's salts, (MEM, *Sigma-Aldrich*, MO, USA) in grado di eluire il materiale presente nel tampone, dopo agitazione su vortex. Un'aliquota (200 µl) di ciascun campione così allestito è stato destinato alle analisi di biologia molecolare, mentre la parte restante è stata stoccata a -80 °C.

La diagnosi virologica è stata condotta mediante tecniche biomolecolari di amplificazione genica specifica.

7.1.1 Tecniche di biologia molecolare per l'identificazione e la caratterizzazione molecolare dei virus influenzali

- **Estrazione degli acidi nucleici**

L'utilizzo di un estrattore automatico (*NucliSENS® EasyMAG®*, Biomérieux, France) ha permesso l'estrazione contemporanea sia dell'RNA che del DNA contenuti nel campione. Questo metodo, basato sulla chimica di Boom, utilizza la silice magnetica per estrarre gli acidi nucleici dal campione biologico in esame. In ambiente con forti concentrazioni di sali, gli acidi nucleici legano le particelle di silice costituendo la fase solida di estrazione. Attraverso cicli di lavaggi i componenti diversi dagli acidi nucleici vengono eliminati e al termine dell'estrazione gli acidi nucleici vengono eluiti dalla fase solida.

Il protocollo ha previsto quattro diverse fasi (**Fig. 30**).

1. Lisi dell'involucro virale, con la conseguente liberazione degli acidi nucleici contenuti e inattivazione di Rnasi e Dnasi: 500 µl di campione sono stati aggiunti a 1,5 ml di NucliSens Lysis Buffer, una soluzione contenente prevalentemente tiocianato di guanidina (5 mol/l). Successivamente la miscela è stata incubata per 10 minuti a temperatura ambiente.
2. Legame alla silice: 100 µl di silice, costituita da biglie magnetiche, sono stati aggiunti a ciascun campione lisato.
3. Lavaggi: sono stati effettuati 4 lavaggi utilizzando tre diversi buffer a composizione non nota.
4. Eluizione: dopo 5 minuti di incubazione a 60°C in 100 µl di buffer di eluizione (a composizione non nota), la silice magnetica è stata eliminata e l'acido nucleico rimasto nel buffer di eluizione trasferito in provette da 1,7 ml.

L'estratto ottenuto è stato utilizzato direttamente per le analisi molecolari e successivamente conservato a -20°C.

Le fasi 1. e 4. Sono state condotte manualmente e in ambiente protetto (cappa a flusso laminare di classe II), mentre le restanti fasi sono state eseguite in modo automatizzato dallo strumento.

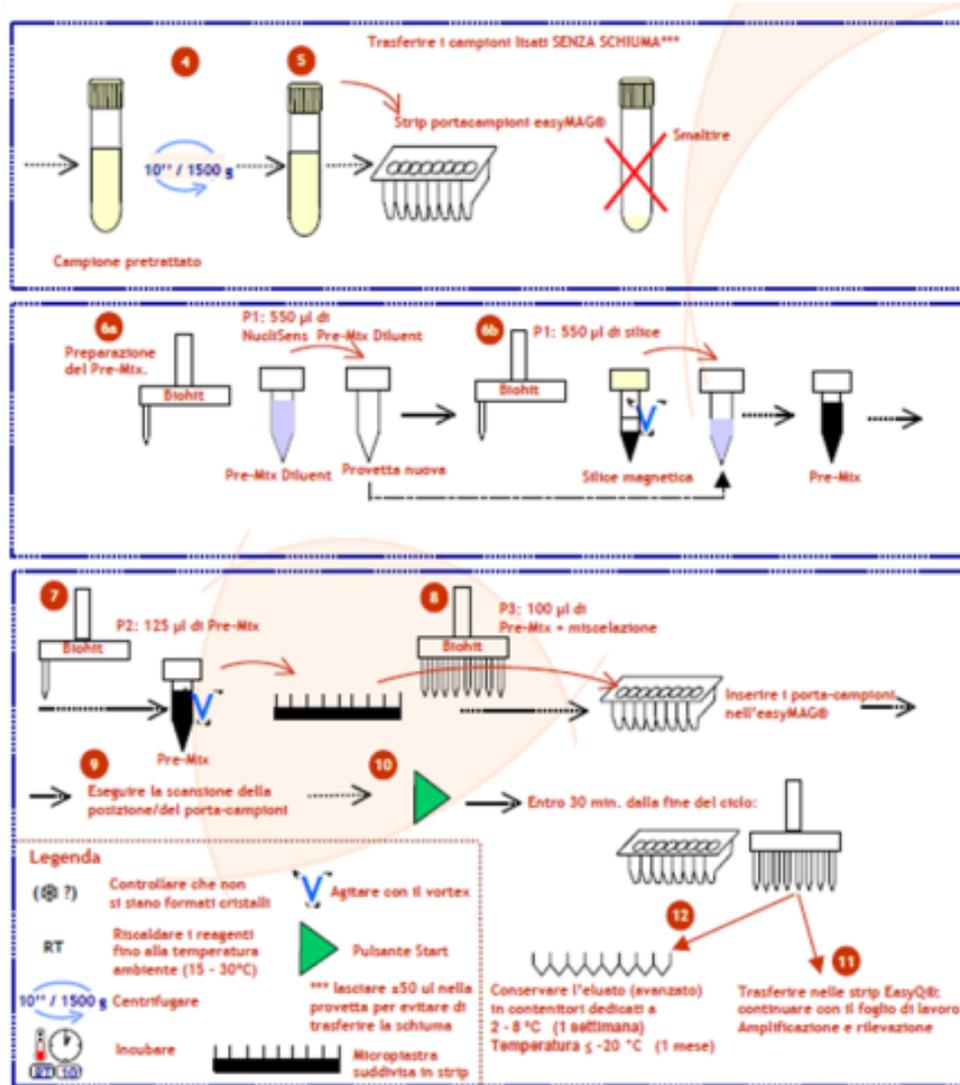


Fig. 30 Descrizione delle procedure del metodo d'estrazione NucliSENS® EasyMAG®, Biomérieux, France (biomerieux-usa.com).

- **Identificazione e tipizzazione dei virus influenzali mediante real-time RT-PCR**

Tutti i campioni sono stati sottoposti a amplificazione genica mediante real-time RT-PCR, con il duplice scopo di identificare e di tipizzare, cioè differenziare tra tipo A e tipo B, i virus influenzali eventualmente presenti. I campioni risultati positivi per influenza A sono stati ulteriormente sottotipizzati mediante real-time RT-PCR per distinguere il sottotipo A(H1N1)pdm09 dal sottotipo A(H3N2).

Protocollo

La tecnica di real-time RT-PCR è stata eseguita utilizzando *AgPath-ID one step RT-PCR kit* (Life Technologies, California, USA) che permette di effettuare in un unico saggio la retrotrascrizione dell'RNA in cDNA e l'amplificazione del cDNA, riducendo i tempi di analisi e le fasi di manipolazione del campione.

Il protocollo operativo prevede la preparazione di due mix di reazione (una per ciascun target, ovvero per virus influenzale di tipo A e B). Dopo aver determinato il corretto volume di ciascun reagente in base al numero dei campioni da saggiare in ogni seduta analitica, vengono allestite le due mix di reazione miscelando i seguenti reagenti: 12,5 µl di 2x RT-PCR buffer (una soluzione salina contenente tris base, dNTPs, glicerolo e un fluoroforo passivo ROX che permette la normalizzazione del segnale di fluorescenza), 0,5 µl di primer *forward* (40 µM), 0,5 µl di primer *reverse* (40 µM), 0,5 µl di sonda (10 µM), 1 µl di 25x RT-PCR mix di enzimi (contenente la trascrittasi inversa *Array Script* e la DNA polimerasi *AmpliTaq Gold*), e 5 µl di H₂O *nuclease-free*.

A seconda del tipo di virus influenzale da identificare si utilizzano set di primer e sonde specifici poiché è necessario amplificare target genici differenti. In particolare, per la ricerca di virus A si utilizzano i seguenti primer e sonda che identificano un frammento di 118 nt (nt 24-142) del gene M codificante per la proteina della matrice [94, 95]:

- FluA F: 5'-GACCRATCCTGCCACCTCTGAC-3'
- FluA R: 5'-AGGGCATTYTGACAAAKCGTCTA-3'
- FluA probe: 5'-FAM*-TGCAGTCCTCGCTCACTGGGCACG-BHQ1**-3'

*FAM: 6-carbossifluorescina, molecola *reporter*

**BHQ1: *Black Hole Quencher-1*, molecola *quencher*)

Per l'identificazione di virus B si utilizzano primer e sonda per l'amplificazione di un frammento di 102 nt (nt 13-115) del gene NP codificante per la nucleoproteina [94, 95]:

- FluB F: 5'-TCCTCAACTCACTCTTCGAGCG-3'
- FluB R: 5'-CGGTGCTCTTGACCAAATTGG-3'
- FluB probe: 5'-FAM-CCAATTCGAGCAGCTGAAACTGCGGTG-BHQ1-3'

Per il controllo della competenza del materiale estratto, ogni campione è stato anche sottoposto ad una reazione di amplificazione del gene endogeno Rnasi P (RNP). Il controllo RNP è un indice dell'efficacia del metodo di estrazione dell'RNA, introdotto per escludere che i campioni negativi alla real-time PCR possano in realtà essere dei falsi negativi a causa di una errata procedura di estrazione.

E' stata pertanto preparata la mix di reazione impiegando i medesimi reagenti sopra riportati, ma utilizzando primer e sonda specifici per l'amplificazione di un frammento di mRNA umano trascritto dal gene RNP [94, 95]:

- RNP F: 5'-AGATTTGGACCTGCGAGCG-3'
- RNP R: 5'-GAGCGGCTGTCTCCACAAGT-3'
- RNP probe: 5'- FAM-TTCTGACCTGAAGGCTCTGCGCG-BHQ1-3'

Nell'allestimento delle mix di reazione è stato necessario rispettare alcuni accorgimenti finalizzati alla corretta esecuzione del saggio. Tra questi è importante evitare l'illuminazione diretta della cappa di lavoro affinché la luce non vada ad inattivare la sonda fotosensibile, motivo per cui quest'ultima deve essere conservata al buio. Inoltre, il volume di mix di reazione deve essere sufficiente per introdurre almeno un controllo positivo di reazione (PTC, *positive template control*, rappresentato da estratti già riconosciuti positivi per il target ricercato) e un controllo negativo (NTC, *no template control*, costituito da acqua).

Dopo aver suddiviso ciascuna mix di reazione nei vari pozzetti (20 µl per pozzetto), si procede all'inoculo del campione. Cinque µl di ciascun estratto di RNA sono stati aggiunti all'interno dei pozzetti contenenti 20 µl delle mix specifiche, raggiungendo un volume finale pari a 25 µl. La piastra così allestita è stata caricata nel termociclatore (*ABI Prism 7300 real-time*, Life Technologies, California, USA) impiegando il seguente profilo di reazione:

	N° DI CICLI	TEMPERATURA	TEMPO
RETROTRASCRIZIONE	1	50 °C	30 minuti
INATTIVAZIONE DELL'RT E DELL'INIBITORE DELLA TAQ / INIZIALE DENATURAZIONE	1	95 °C	10 minuti
AMPLIFICAZIONE	45	95 °C	15 secondi
		55 °C	30 secondi

Durante la fase di amplificazione il segmento in via di polimerizzazione emette un segnale fluorescente che il software è in grado di rilevare e convertire in segnale digitale; pertanto l'andamento della reazione può essere controllato in tempo reale nel corso di tutta la fase di amplificazione. I risultati definitivi sono invece accessibili a fine reazione, dopo circa 2 ore e 12 minuti.

La **Fig.31** mostra un esempio della rappresentazione grafica emessa dallo strumento al termine della seduta analitica. Il grafico mette in relazione il numero dei cicli (sull'asse delle ascisse) e l'emissione di fluorescenza (sull'asse delle ordinate) per ciascun pozzetto di reazione. Sono considerati positivi quei campioni la cui curva di reazione determinata dal segnale di fluorescenza interseca la linea soglia (*threshold*) prima del quarantesimo ciclo di amplificazione. Il tracciato che identifica l'avvenuta amplificazione assume la tipica forma sigmoide il cui andamento è rappresentativo dell'emissione della fluorescenza, la quale dapprima aumenta in modo esponenziale, poi lineare fino al raggiungimento del *plateau*. Il tracciato interseca il *threshold* in corrispondenza del valore di Ct.

Diversamente, il tracciato emesso dallo strumento in riferimento ai pozzetti di reazione risultati negativi deriva dalla rilevazione del segnale di *background* aspecifico, e rimane al di sotto del *threshold*. Nella **Fig. 31** è stato inoltre riportato il tracciato che identifica il controllo interno RNP. Affinché la seduta analitica sia validata tutti i campioni clinici di origine umana devono presentare curve di RNP che intersecano il *threshold* prima del quarantesimo ciclo.

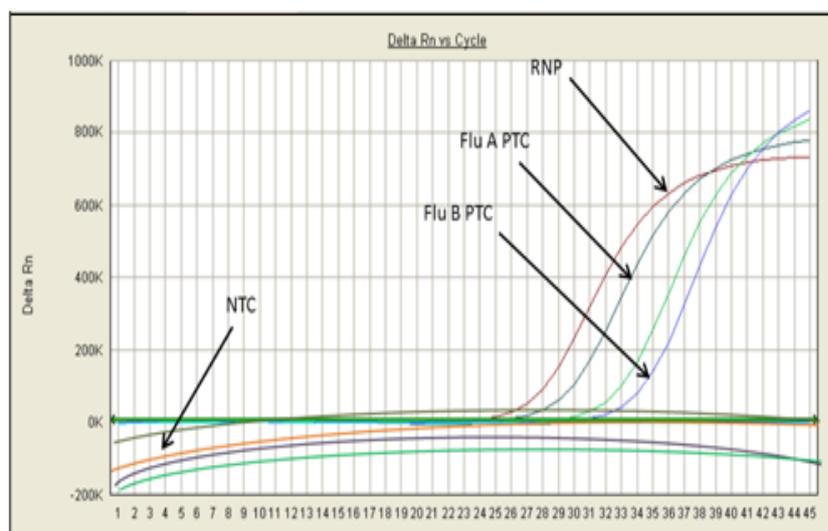


Fig. 31: Esempio di output emesso al termine della seduta analitica di real-time RT-PCR

- **Sottotipizzazione dei virus influenzali di tipo A mediante real-time RT-PCR**

Campioni che, in seguito a real-time RT-PCR, mostrano una positività per virus influenzale di tipo A sono stati sottotipizzati mediante un altro saggio di real-time RT-PCR che, grazie all'utilizzo di primer e sonde specifiche, permette di distinguere i sottotipi virali A(H1N1)pdm09 e A(H3N2). Ad eccezione dei primer e delle sonde specifiche, la metodica, i reagenti e il profilo di amplificazione sono identici a quelli impiegati nella real-time RT-PCR di tipizzazione sopra descritta. Poiché non si tratta di un'amplificazione di tipo multiplex, per ogni target indagato (H1 e H3) deve essere allestita la rispettiva miscela di reazione contenente i reagenti specifici.

In particolare per l'identificazione di virus A(H1N1)pdm09 sono stati utilizzati primer e sonde in grado di amplificare un frammento di 115 nt (nt 901-1016) nel gene HA [94, 95]:

FLUA/H1 F: 5'-GTGCTATAAACACCAGCCTYCCA-3'

- FLUA/H1 R: 5'-CGGGATATTCCTTAATCCTGTRGC-3'
- FLUA/H1 probe: 5'-FAM-CAGAAATACATCCRGTCACAATTGGARAA-BHQ1-3'

Per la ricerca di A(H3N2) sono stati utilizzati primer e sonde in grado di amplificare un frammento di 194 nt (nt 858-1052) nel gene HA [94, 95]:

- FLUA/H3 F: 5'-AAGCATTCCYAATGACAAACC-3'
- FLUA/H3/R: 5'-ATTGCRCRAATATGCCTCTAGT-3'
- FLUA/H3 probe: 5'-FAM-CAGGATCACATATGGGSCCTGTCCCAG-BHQ1-3'

7.1.2 Tecniche di biologia molecolare per l'identificazione e la caratterizzazione molecolare di altri patogeni respiratori

I campioni risultati negativi alle indagini specifiche per il virus influenzale A(H1N1)pdm09 sono stati sottoposti ad analisi per la determinazione dei seguenti agenti virali: Virus Respiratorio Sinciziale (RSV) A e B; human Metapneumovirus (hMPV) A e B; human Rhinovirus (hRV) e Enterovirus (hEV); Adenovirus (AdV); human Bocavirus (hBoV) 1-4; human Coronavirus (hCoV) 229E, NL63, OC43, HKU1; human parainfluenza virus (hPIV) 1-4; *Chlamydomphila pneumoniae*; *Mycoplasma pneumoniae*.

Tutti i campioni sono stati sottoposti a amplificazione genica mediante real-time RT-PCR, utilizzando il kit commerciale *Respiratory MWS r-gene™ Real-time PCR*, bioMérieux, France. Per la ricerca dei diversi agenti patogeni sono state utilizzate le specifiche mix di amplificazione:

- *RSV/hMPV r-gene™*
- *Rhino&EV/CC r-gene™*
- *HCoV/HPIV r-gene™*
- *AdVBoV r-gene™*
- *Chla/Myco pneumo r-gene™*

Le mix per la determinazione dei virus a RNA contenevano una trascrittasi inversa (SuperScript III®) per l'esecuzione del passaggio di retrotrascrizione precedente l'amplificazione.

Ogni kit conteneva un controllo negativo e un controllo positivo specifico per ciascuna tipologia di patogeno da identificare.

7.2 Virus Discovery

I campioni risultati negativi a tutte le indagini precedenti, e quindi privi di infezioni sostenute dai microrganismi ricercati, sono stati inviati al “Laboratory of Experimental Virology, Department of Medical Microbiology, Academic Medical Center (AMC)” dell’Università di Amsterdam, per ulteriori approfondimenti.

Presso questo laboratorio è stata impiegata, in particolare, la metodica VIDISCA-454, ossia **Virus Discovery cDNA AFLP** (Amplified Fragment Length Polymorphism) associato all’utilizzo della piattaforma **454** di NGS della Roche [68], allo scopo di individuare agenti virali non noti nei campioni in esame.

7.2.1 Metodica VIDISCA-454

In questo metodo, i campioni vengono ultracentrifugati per ottenere una concentrazione maggiore delle particelle virali e trattati successivamente con DNasi e RNasi. Gli acidi nucleici virali vengono poi purificati, convertiti in doppio filamento DNA, digeriti con enzimi di restrizione e legati ad adattatori oligonucleotidici (anchor), che vengono utilizzati come siti di legame di primer per l’amplificazione mediante PCR [69].

Un riassunto della metodica è riportata nella **Fig. 32**.

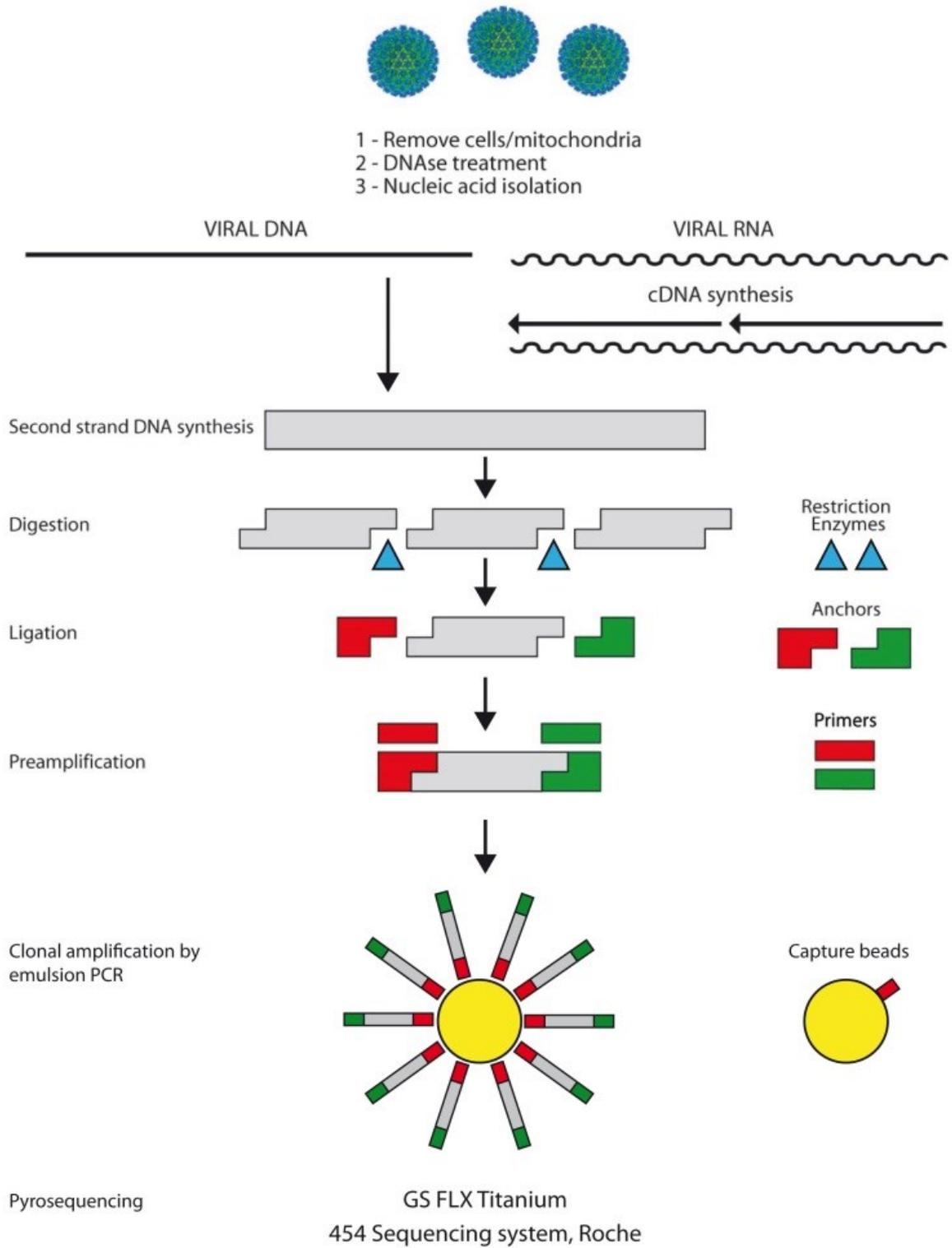


Fig. 32: Riassunto metodica VIDISCA-454

- **Pre-trattamento**

Prima di procedere all'estrazione degli acidi nucleici, 110 µl di ciascun campione sono stati centrifugati per 10 minuti a 10.000 g e 100 µl di surnatante sono stati quindi trasferiti in una nuova provetta facendo attenzione a non includere il pellet. I campioni sono stati poi trattati con enzima DNase, necessario per rimuovere eventuali residui di DNA che potrebbero interferire nelle successive analisi. A tal fine 10 µl Turbo Dnase (2U/µl; Ambion), 12 µl di Dnase buffer (Ambion) sono stati aggiunti al surnatante e incubati per 30 minuti a 37°C.

- **Estrazione acidi nucleici**

Tutti i campioni sono stati sottoposti ad estrazione degli acidi nucleici; anche in questo caso è stato utilizzato un metodo, basato sulla chimica di Boom, che impiega la silice. Come già descritto in precedenza, in ambiente con forti concentrazioni di sali, gli acidi nucleici legano le particelle di silice costituendo la fase solida di estrazione. Attraverso cicli di lavaggi i componenti diversi dagli acidi nucleici vengono eliminati e al termine dell'estrazione gli acidi nucleici vengono eluiti dalla fase solida. Anche questa procedura di estrazione ha previsto le medesime 4 fasi descritte nel paragrafo 7.1.1 ma, non essendo stata impiegata silice magnetica, i diversi lavaggi sono stati eseguiti attraverso centrifugazione e risospensione di ciascun campione. I campioni in questo caso sono stati eluiti in 50 µl di acqua sterile

- **Retrotrascrizione**

La retrotrascrizione è stata effettuata utilizzando l'enzima Superscript II (200 U, *Invitrogen, Life Technologies, California, USA*) in una mix contenente una ligasi di E.coli (5 U, *Invitrogen, Life Technologies, California, USA*). Sono state preparate due mix di reazione:

RT-mix I:

- 4,6 µl esamero 100 µM
- 3 µl 10X buffer (*Invitrogen*)
- 2,4 µl MgCl₂ (100mM)

RT-mix II:

- 2 µl 10X buffer, (*Invitrogen*)
- 1 µl SuperScript II (200U/µl *Invitrogen*)
- 0,8 µl dNTPs (25mM of each)
- 15,2 µl acqua (sterile)
- 1 µl DTT (0.1 M)

L'eluato ottenuto dopo estrazione è stato centrifugato per 30 secondi alla massima velocità per rimuovere qualsiasi residuo di silice. Successivamente, a 20 µl di questo sono stati aggiunti 10 µl di RT-mix I, incubando per 2 minuti a temperatura ambiente. Sono stati poi aggiunti 20 µl di RT-mix II, seguiti da incubazione per 90 minuti a 42 °C e da una seconda incubazione a 70 °C per 20 minuti.

- **Sintesi del Second strand DNA**

Dopo retrotrascrizione l'RNA in cDNA è stata effettuata la sintesi dello strand complementare mediante l'utilizzo di una polimerasi Klenow polymerase (3'-5' exo⁻) (5U/µl; *New England Biolabs, MA, USA*)

A tal fine a 50 µl di cDNA è stata aggiunta la seguente mix di reazione:

- 10 µl NEB2 10X buffer (*New England Biolabs, MA, USA*)
- 1 µl Klenow polymerase (3'- 5' exo⁻) (5U/µl; *New England Biolabs, MA, USA*)
- 1.5 µl RNaseH (5U/µl) (*New England Biolabs, MA, USA*)
- 1 µl dNTPs (25mM each)
- 86,5 µl acqua sterile

L'incubazione è avvenuta a 37 °C per 90 minuti (volume totale 150 µl)

- **Estrazione con fenolo/cloroformio**

Per eliminare ogni traccia di contaminanti, è stata effettuata l'estrazione con fenolo/cloroformio. 150 µl di mix di fenolo/cloroformio/alcol isoamilico (Ultra-pure PCI 25:24:1 v/v, *Invitrogen, Life Technologies, California, USA*) sono stati aggiunti al campione e sono stati vorticati vigorosamente. Successivamente i campioni sono stati centrifugati per 1 minuto alla massima velocità e 140 µl dello strato superficiale (fase acquosa) sono stati trasferiti in una nuova Eppendorf. Sono stati aggiunti 350 µl di etanolo al 100% e 14 µl di acetato di sodio (pH 5.2) e vorticati. Gli acidi nucleici purificati sono stati fatti precipitare a -20 °C overnight.

- **Digestione del cDNA con enzima di restrizione Mse I**

La fase successiva è stata quella di sottoporre il DNA double strand a digestione mediante l'utilizzo di dell'enzima di restrizione MseI, endonucleasi in grado di tagliare il DNA a livello

delle sequenze palindrome $5' \dots \overset{\nabla}{T}TAA \dots 3'$
 $3' \dots AAT \overset{\blacktriangle}{T} \dots 5'$, in modo tale da creare dei frammenti ai quali sarà possibile legare gli adattatori oligonucleotidici nella fase successiva.

A tal fine i campioni sono stati centrifugati per 25 minuti a 7500 g a +4 °C. E' stato scartato il surnatante, sono stati aggiunti 200 µl di etanolo al 70% e ed è stata effettuata un'ulteriore centrifugata per 25 minuti a 7500 g a +4 °C. E' stato poi rimosso l'etanolo e il pellet è stato fatto asciugare a temperatura ambiente per 15 minuti. Il pellet è stato poi dissolto in 30 µl di acqua sterile.

La mix di digestione era così composta:

- 4 µl 10 X NEB-2 buffer (*New England Biolabs, MA, USA*)
- 4 µl BSA diluito 1:10 (*New England Biolabs, MA, USA, supplied with MseI*)
- 1 µl acqua (sterile)
- 1 µl di enzima di restrizione Mse1 (10U/µl; *New England Biolabs, MA, USA*)

Infine 10µl di mix di digestione sono stati aggiunti ai campioni e incubati per 2 ore a 37 °C.

- **Legame degli anchors ai frammenti digeriti**

Il legame è stato effettuato utilizzando due tipi di anchors:

Anchor A: costruito sulla base di una sequenza A legata a un set di sequenze identificative specifiche di 10 nt, denominate MID (Multiplex Identifier) necessarie per il riconoscimento delle sequenze nella fase di analisi bioninformatica.

Anchor B: costruito sulla base di una sequenza B

Preparazione della mix di legame:

- 1 µl MIDX-A anchor (5 µM)
- 1 µl B anchor (5 µM)
- 2 µl 5x ligation buffer (Invitrogen)
- 1 µl T4 ligase (5U/µl; Invitrogen)
- 10 µl sterile water

15µl di mix di legame sono stati aggiunti a 40 µl di ciascun campione e incubati per 2 ore a temperatura ambiente.

- **PCR**

Una volta avvenuto il legame, è stata effettuata una reazione di PCR per amplificare i frammenti ottenuti utilizzando come primer (Tit-PCR-A e Tit-PCR-B) delle sequenze complementari agli anchor A e B.

Sono stati quindi trasferiti 10 μ l di campione ottenuto a 40 μ l di mix di PCR così composta:

- 25 μ l Green Buffer
- 1 μ l Tit-PCR-A (20 μ M)
- 1 μ l Tit-PCR-B (20 μ M)
- 13 μ l water

La reazione è stata effettuata utilizzando il seguente profilo termico:

5min 95 °C

1min 95 °C|

1min 55 °C| 40 cycles

2min 72 °C|

10min 72 °C

10min 4 °C

- **Purificazione con Agencourt Ampure XP**

La purificazione è stata effettuata utilizzando il protocollo del kit commerciale Agencourt Ampure XP (Beckman Coulter Company, USA) (**Fig. 33**)

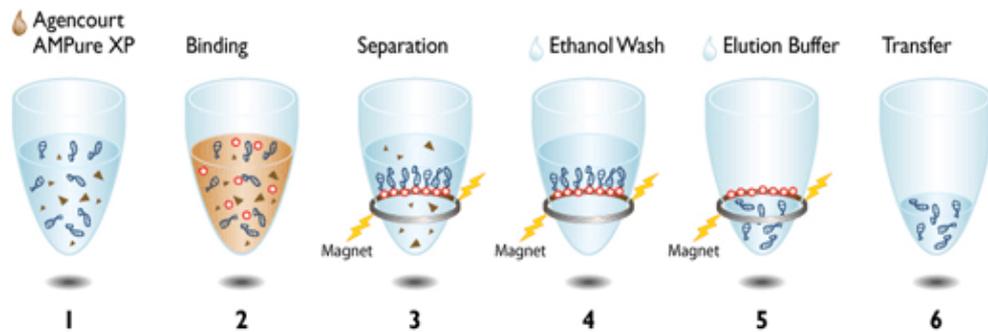


Fig. 33: Rappresentazione dei diversi step effettuati per la purificazione

Il protocollo di utilizzo di questo kit si articola su 6 fasi:

1. Aggiunta di 1,8 μ l di biglie AMPure XP per 1 μ l di prodotto di PCR
2. Legame del DNA alle biglie
3. Separazione delle biglie con legato il DNA dai contaminanti
4. Lavaggio con etanolo al 70% per rimuovere eventuali contaminanti residui
5. Eluizione
6. Trasferimento del DNA purificato in un'altra provetta

Per ottenere un risultato migliore il protocollo è stato ripetuto per due volte.

- **Quantificazione del DNA**

Il DNA è stato quantificato mediante l'utilizzo del kit Quant-iT™ dsDNA Assay utilizzando il fluorimetro Qubit (New Invitrogen, Life Technologies, California, USA).

- **Protocollo 454 sequencing**

Una volta quantificato il DNA ottenuto si è proceduto con il protocollo per il sequenziamento dei frammenti ottenuti utilizzando la piattaforma 454 GS FLX+ System (*Hoffmann-La Roche*, Basilea, Svizzera).

Per la preparazione dei campioni è stata effettuata una Emulsion PCR seguendo i protocolli: LIB-A SV emPCR kit, GS FLX Titanium PicoTiterPlate kit (70×75), GS FLX Titanium XLR 70 Sequencing kit (*Hoffmann-La Roche*, Svizzera).

I campioni sono stati sequenziati in 2 regioni Picotiterplate del sistema 454 Titanium system e processati secondo protocollo con 2 E6 di biglie come input. (**Fig. 34**)

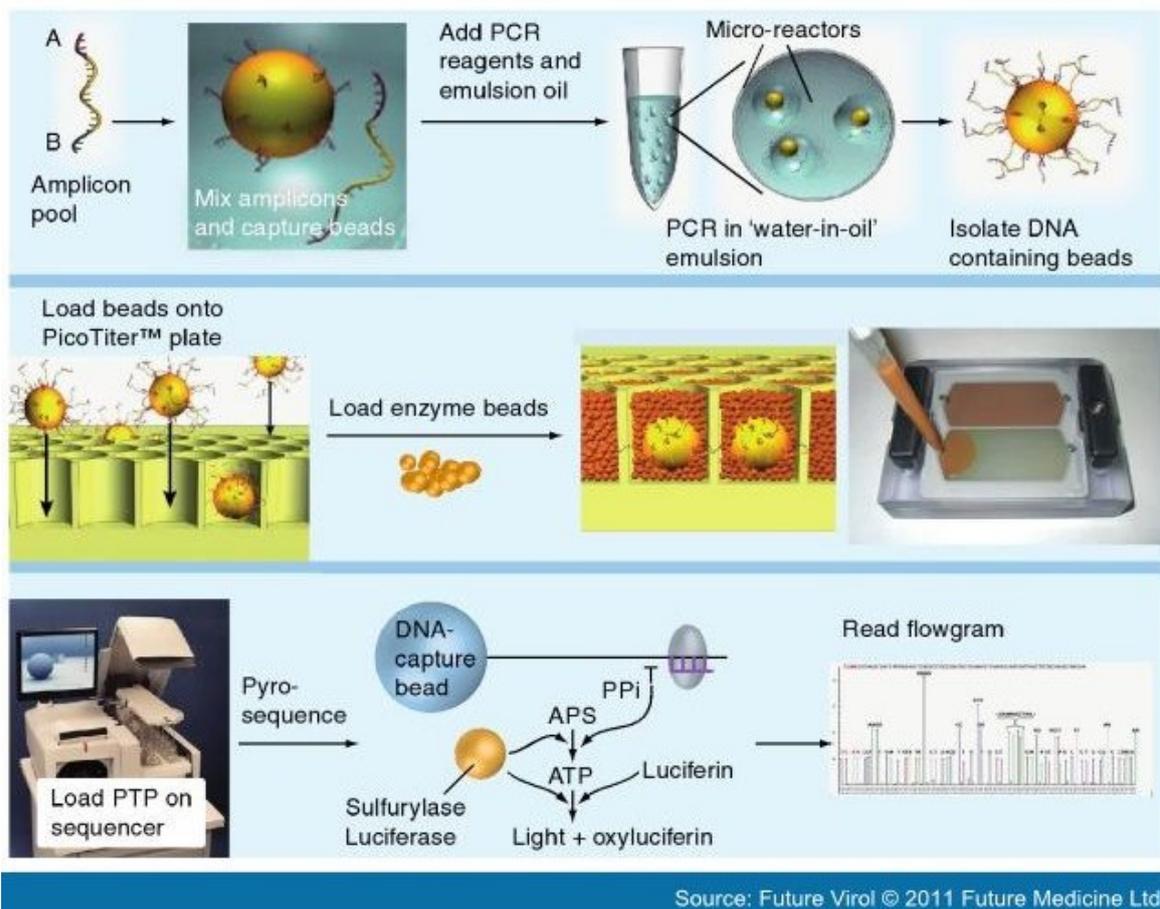


Fig. 34: Rappresentazione dei diversi step del protocollo di sequenziamento

- **Analisi delle sequenze**

Le sequenze ottenute sono state assemblate utilizzando il software CodonCode (www.codoncode.com) e la ricerca successiva delle sequenze virali è stata effettuata con Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

7.3 Sequenziamento genico del gene codificante per la proteina G del virus respiratorio sinciziale sottotipo A (RSV A) e B (RSV B)

7.3.1 Saggi biomelocolari per la determinazione della sequenza da analizzare

I campioni respiratori risultati positivi per la ricerca virus respiratorio sinciziale sottotipo A o sottotipo B sono stati amplificati per il sequenziamento genico del gene codificante per la proteina d'attacco G.

Le reazioni di amplificazione di frammenti del gene G di RSV A e RSV B dai campioni retrotrascritti, finalizzate al successivo sequenziamento, consistono in due *nested*-PCR. Le reazioni di amplificazione prevedono la preparazione di due distinte miscele di reazione di 2° *step* per ottenere due frammenti genici distinti la cui sovrapposizione permette di ottenere il frammento completo.

Nelle Fig. 35 e 36 sono rappresentate le strategie di amplificazione per il sequenziamento del gene G di RSV A e RSV B.

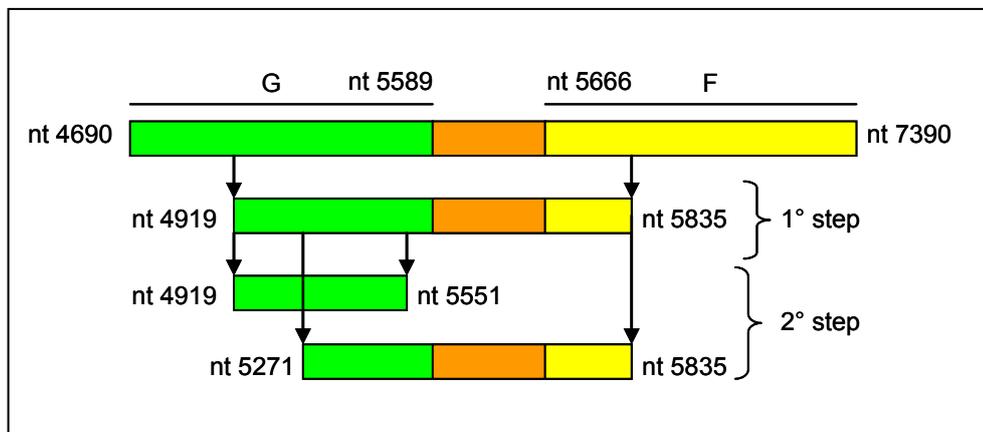


Fig. 35: Strategia di amplificazione finalizzata al sequenziamento del gene G di RSV A.

La loro parziale sovrapposizione consente di ottenere una sequenza della lunghezza complessiva di 915 nt (nt. 4919-5835)

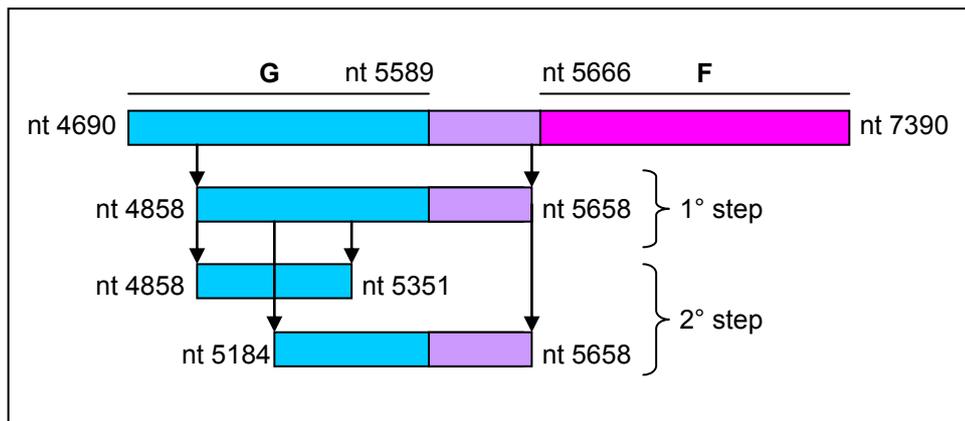


Fig.36 : Strategia di amplificazione finalizzata al sequenziamento del gene G di RSV B. La loro parziale sovrapposizione consente di ottenere una sequenza della lunghezza complessiva di 800 nt (nt. 4858-5658).

• **Saggio di nested-PCR per il sequenziamento del gene G di RSV A**

Nel 1° step di amplificazione sono stati utilizzati i seguenti *primers*, selezionati dalla letteratura scientifica [96]:

<u>Primer</u>	<u>Sequenza</u>	<u>Posizione nucleotidica</u>
GA1F	5'-GATGCAACAAGCCAGATCAAG-3'	4919-4939
GA1R	5'-GTTATCACACTGGTATACCAACC-3'	5813-5835

10µl di cDNA sono stati inoculati in 40µl di una miscela di reazione così composta: 10µl di *buffer* 5X (150mM di Tris Hcl, 500mM di KCl), 1µl di dNTP (10mM), 1µl di ciascun *primers* (GA1F e GA1R), 0,25µl di *Taq* DNA polimerasi (5U/µl) e 26,75µl di acqua.

Il 1° step è stato condotto in un termociclatore alle seguenti condizioni di amplificazione:

N° cicli	Denaturazione	Annealing	Estensione
1	95°C x 5'		
35	95°C x 1'	60°C x 1'	72°C x 1'
1			72°C x 7'

Nel 2° step di amplificazione sono stati utilizzati i seguenti *primers* [96]:

- *MIX 1*: frammento di 632 bp

<u>Primer</u>	<u>Sequenza</u>	<u>Posizione nucleotidica</u>
GA1F	5'-GATGCAACAAGCCAGATCAAG-3'	4919-4939
GA2R	5'-TTGTGATAGGTACTCGGATGT-3'	5531-5551

- *MIX 2*: frammento di 564 bp:

<u>Primer</u>	<u>Sequenza</u>	<u>Posizione nucleotidica</u>
GA2F	5'-GAAAGAAAACCACCACCAAGC-3'	5271-5291
GA1R	5'-GTTATCACACTGGTATACCAACC-3'	5813-5835

Per il 2° step sono state preparate due *mix* di reazione, una per ciascun frammento da amplificare. 5µl di amplificato di 1° step sono stati inoculati in ciascuna delle 2 *mix* di reazione così composte: 10µ di *buffer* 5X (150mM di Tris.Hcl, 500mM di KCl), 1µl di dNTP (10mM), 1µl di ciascun *primers* (*MIX 1*: GA1F e GA2R; *MIX 2*: GA2F e GA1R), 0,25µl di *Taq* DNA polimerasi (5U/µl) e 31,75µl di acqua.

Il 2° step, per entrambe le *mix* di reazione, è stato condotto in un termociclatore alle seguenti condizioni di amplificazione:

N° cicli	Denaturazione	Annealing	Estensione
1	95°C x 5'		
35	95°C x 1'	60°C x 1'	72°C x 1'
1			72°C x 7'

- **Saggio di nested-PCR per il sequenziamento del gene G di RSV B**

Nel 1° *step* di amplificazione sono stati utilizzati i seguenti *primers*, selezionati dalla letteratura scientifica [97]:

<u>Primer</u>	<u>Sequenza</u>	<u>Posizione nucleotidica</u>
GB1F	5'-GCAGCCATAATATTCATCATCTCT-3'	4858-4881
GB1R	5'-TGCCCCAGRTTTAATTCGTTC-3'	5637-5658

10µl di cDNA sono stati inoculati in 40µl di una miscela di reazione così composta: 10µl di *buffer* 5X (150mM di Tris.Hcl, 500mM di KCl), 1µl di dNTP (10mM), 1µl di ciascun *primers* (GB1F e GB1R), 0,25µl di *Taq* DNA polimerasi (5U/µl) e 26,75µl di acqua.

Il 1° *step* di amplificazione è stato condotto in un termociclatore alle seguenti condizioni di amplificazione:

N° cicli	Denaturazione	Annealing	Estensione
1	95°C x 5'		
35	95°C x 1'	60°C x 1'	72°C x 1'
1			72°C x 7'

Nel 2° *step* di amplificazione sono stati utilizzati i seguenti *primers* [97]:

- **MIX 1:** frammento di 493 bp:

<u>Primer</u>	<u>Sequenza</u>	<u>Posizione nucleotidica</u>
GB1F	5'-GCAGCCATAATATTCATCATCTCT-3'	4858-4881
GB2R	5'-TTGGCTRGTGTTTTGGGTCTCT-3'	5329-5351

- **MIX 2:** frammento di 474 bp:

<u>Primer</u>	<u>Sequenza</u>	<u>Posizione nucleotidica</u>
GB2F	5'-TGAAGTGTTCAAYTTYGTTCCCTGT-3'	5184-5208
GB1R	5'-TGCCCCAGRTTTAATTCGTTC-3'	5637-5658

Per il 2° *step* sono state preparate due *mix* di reazione, una per ciascun frammento da amplificare. 5µl di amplificato di 1° *step* sono stati inoculati in ciascuna delle 2 *mix* di reazione così composte: 10µl di *buffer* 5X (150mM di Tris.Hcl, 500mM di KCl), 1µl di dNTP (10mM), 1µl di ciascun *primers* (MIX 1: GB1F e GB2R; MIX 2: GB2F e GB1R), 0,25µl di **Taq** DNA polimerasi (5U/µl) e 31,75µl di acqua.

Il 2° *step*, per entrambe le *mix* di reazione, è stato condotto in un termociclatore alle seguenti condizioni di amplificazione:

N° cicli	Denaturazione	Annealing	Estensione
1	95°C x 5'		
35	95°C x 1'	65°C x 1'	72°C x 1'
1			72°C x 7'

7.3.2 Rilevazione dei prodotti di amplificazione tramite elettroforesi su gel di agarosio

La corsa elettroforetica su gel d'agarosio e il confronto con un *marker* di riferimento (100bp *DNA Ladder*, *BioLabs*) hanno permesso la rilevazione degli amplificati del gene G di RSV A e RSV B.

Il gel, contenete una matrice porosa, permette la migrazione in campo elettrico del DNA carico negativamente verso il polo positivo. La velocità di migrazione è correlata alle dimensioni dei frammenti di DNA: i frammenti più piccoli migrano a velocità maggiori.

Il gel è stato preparato portando ad ebollizione una miscela contenente 100µl di soluzione tampone TBE (Tris HCl, EDTA-Na₂, 2H₂O, acido borico; pH 8,3) e 2g di agarosio. Sono stati successivamente aggiunti 2µl di etidio bromuro (1% in soluzione), agente intercalante gli acidi nucleici che, esposto ai raggi ultravioletti, emette fluorescenza, permettendo così la visualizzazione delle bande sul gel, corrispondenti al frammento amplificato.

2µl di blu di bromofenolo sono stati addizionati a 10µl di ciascun amplificato di 2° *step*, facilitandone la precipitazione nei pozzetti del gel. I campioni sono stati fatti correre in un campo elettrico con voltaggio di 120mV per circa 40 minuti.

7.3.3 Purificazione e concentrazione dei prodotti di amplificazione

I prodotti di amplificazione ottenuti, dopo corsa elettroforetica su gel di agarosio all'1%, sono stati purificati e concentrati mediante l'utilizzo di un *kit* commerciale *Wizard® SV Gel and PCR Clean-Up System (Promega, Italia)*.

Terminata l'elettroforesi, è stato exciso il gel in corrispondenza delle bande di DNA di interesse e posto in provette da 1,5mL, successivamente pesate su una bilancia di precisione. A questo punto sono stati addizionati 10µl di *Membrane Binding Solution* (contenente guanidina isotiocianato) ogni 10mg di gel exciso. Le provette sono state vorticate ed incubate a 50°C in un termoblocco fino a completa dissoluzione del gel. Il gel disciolto è stato poi trasferito in una colonnina (*SV Minicolumn*), assemblata con un tubo di raccolta, contenente una membrana di silice in grado di legare il DNA sulla sua superficie. Dopo un'incubazione a temperatura ambiente per 1 minuto, i tubi sono stati centrifugati a 16000xg per 1 minuto. Le colonne sono state sottoposte ad un primo lavaggio aggiungendo 700µL di *Membrane Wash Solution*, previa diluizione con etanolo 95%, e centrifugate per 1 minuto a 16000xg. Il secondo lavaggio è stato ripetuto con 500µL di *Membrane Wash Solution* e successiva centrifuga a 16000xg per 5 minuti. Una centrifuga a vuoto della durata di 1 minuto a 16000xg ha consentito di eliminare il *carry-over*.

Infine, nelle colonne, trasferite in provette da 1.5mL, sono stati addizionati 35µl di *Nuclease-Free Water*. Dopo un'incubazione a temperatura ambiente per 1 minuto sono state sottoposte ad'ultima centrifuga a 16000xg per 1 minuto. Il DNA è stato così eluito e conservato a 4°C fino al momento del sequenziamento.

5µL del prodotto di purificazione sono stati sottoposti a elettroforesi su gel al 2% d'agarosio per verificare l'efficienza della reazione.

7.3.4 Sequenziamento genico del gene codificante per la proteina G

Gli amplificati ottenuti sono stati sequenziati mediante l'utilizzo di un sequenziatore capillare *ABI PRISM® 3100 Genetic Analyzer* (*Applied Biosystems Inc, Life Technologies, California, USA*) con la collaborazione di un laboratorio esterno alla sezione di Virologia.

Il processo di sequenziamento ha inizio con una reazione di polimerizzazione.

A tale scopo è stato utilizzato un kit commerciale (*ABI PRISM™ Ready Reaction DyeDeoxy Terminator Cycle Sequencing*) contenente una soluzione composta da 4 dideossinucleotidi terminatori (ddNTPs) marcati con 4 fluorocromi diversi (dTTP 15.79 mM, Tris-HCl 48.42 mM pH 9.1, (NH₄)₂SO₄ 4.21 mM) e l'enzima termostabile AmpliTaq FS DNA-polymerase (0.42 U/ml).

I ddNTPs mancano del gruppo idrossile in posizione 3', coinvolto nella formazione del legame fosfodiesterico tra i nucleotidi che costituiscono le molecole di DNA. Pertanto, quando i ddNTPs marcati vengono incorporati al posto dei normali dideossinucleotidi la reazione di polimerizzazione si arresta.

Il processo di polimerizzazione è stato condotto mediante l'utilizzo di un enzima termostabile in grado di lavorare anche a temperature elevate, riducendo la formazione di strutture secondarie e la formazione di legami tra i primer e sequenze non specifiche.

L'AmpliTaQ DNA polimerasi FS, inoltre, facilita l'incorporazione dei ddNTPs nella catena del DNA in formazione. Essa, infatti, è una variante della consueta polimerasi *Thermos aquaticus*, contenente una mutazione puntiforme a livello del sito attivo, responsabile di una ridotta capacità di discriminazione nei confronti dei ddNTPs.

Una volta terminata la reazione di polimerizzazione, grazie ad un rilevatore laser che eccita le molecole fluorescenti, è possibile riconoscere i ddNTPs incorporati nella catena di DNA e risalire alla sequenza nucleotidica d'interesse.

7.3.5 Analisi delle sequenze

Le sequenze nucleotidiche del gene codificante la proteina G di RSV A e B sono state caratterizzate mediante analisi filogenetica.

In particolare sono state condotte due distinte analisi sulla base dei due sottotipi virali in studio. Per ciascuna analisi, le sequenze nucleotidiche in studio sono state confrontate tra loro con sequenze di riferimento del gene G presenti in banca dati *GenBank* [98].

- **Editing delle sequenze**

Per ogni singola sequenza è stata effettuata una verifica delle discordanze, mediante un confronto manuale dell'elettroferogramma e della sequenza interpretata automaticamente, e sono stati corretti manualmente i nucleotidi letti in modo non corretto dal sistema di lettura automatico. Successivamente, la sequenza dell'intero prodotto di amplificazione del frammento genico è stata creata unendo le due sequenze ottenute con le due sessioni di sequenziamento con i primer senso e antisense. Entrambi questi passaggi sono stati realizzati tramite il software bioinformatico BioEdit (versione 7.0.9.0) [99]. Ad ogni sequenza completa è stato quindi associato un codice identificativo alfanumerico: i primi due caratteri corrispondono alla provenienza geografica (MI: Milano), il terzo al sottotipo virale (RSV-A/RAV-B) mentre i successivi comprendono il numero identificativo del campione e l'anno di prelievo (ad esempio MI/A/511-11/11 rappresenta il campione 511-11 raccolto a Milano nel 2011 e RSV-A positivo). Tutte le sequenze sono state raggruppate in file di testo in formato FASTA. Questo particolare formato (.fas) comprende, per ogni singola sequenza, una riga di intestazione, che riporta il nome della sequenza preceduta dal simbolo >, e una serie di righe contenenti la sequenza nucleotidica vera e propria, identificata con le lettere delle 4 basi azotate.

- **Previsione delle sequenze aminoacidiche**

Le sequenze nucleotidiche sono state utilizzate per la previsione delle sequenze aminoacidiche corrispondenti mediante l'utilizzo del software bioinformatico BioEdit [99].

7.3.6 Analisi filogenetica

Tutte le sequenze nucleotidiche ottenute sono state quindi tipizzate mediante analisi filogenetica ed è stata così studiata la circolazione dei differenti ceppi virali nei diversi anni (2006-2012). Per lo studio filogenetico, ogni sequenza è stata confrontata con delle sequenze di riferimento mediante costruzione di un albero filogenetico.

- **Dataset di sequenze di riferimento utilizzate per lo studio filogenetico di RSV**

Per studiare la circolazione dei ceppi di RSV dal 2006 al 2012, le sequenze ottenute in questo studio sono state confrontate con altre di riferimento (riportati nelle **tabelle 1 e 2**) disponibili nella banca dati di GeneBank [98]. Sono stati quindi allestiti e utilizzati due dataset uno per ciascun sottotipo virale RSV-A e RSV-B.

Sequenza	Accession Number	Genotipo	Luogo di isolamento	Data di isolamento
ON67-12/10	JN257693	ON1	Ontario, Canada	2010
ON138-01/11	JN257694	ON1	Ontario, Canada	2011
A/WI/629-22/07	JF920049	NA1	Wisconsin, USA	2007
A/WI/629-DC9/08-09	JF920050	NA1	Wisconsin, USA	2008/2009
A/WI/629-21/07	JF920051	NA1	Wisconsin, USA	2007
A/WI/629-9-2/07	JF920048	NA1	Wisconsin, USA	2007
A/WI/629-2/07	JF920046	NA1	Wisconsin, USA	2008
A/WI/629-23/08	JF920047	NA1	Wisconsin, USA	2011
ON201-01/11	JN257702	NA1	Ontario, Canada	2010
ON69-03/10	JN257697	NA1	Ontario, Canada	2011
ON187-01/11	JN257701	NA1	Ontario, Canada	2010
ON45-10/10	JN257695	NA1	Ontario, Canada	2010
ON108-03/10	JN257699	NA1	Ontario, Canada	2010
ON66-12/10	JN257696	NA1	Ontario, Canada	2011
ON97-01/11	JN257703	NA1	Ontario, Canada	2011
ON160-01/11	JN257700	NA1	Ontario, Canada	2011
A/WI/629-Q0154/10	JF920052	NA1	Wisconsin, USA	2010
A/WI/629-Q0284/10	JF920053	NA1	Wisconsin, USA	2010

A/WI/629-Q0282/10	JF920054	NA1	Wisconsin, USA	2010
A/WI/629-Q0198/10	JF920055	NA1	Wisconsin, USA	2010
A/WI/629-4239/98	JF920057	GA2	Wisconsin, USA	1998
Se/12/97	AF193323	GA2	Seoul, South Korea	1997
Se/02/98	AF193325	GA2	Seoul, South Korea	1998
NY/CH57/94	AF065258	GA2	New York, USA	1994
NY/CH28/94	AF065256	GA2	New York, USA	1994
Mon/3/88	Z33425	GA2	Montevideo, Uruguay	1988
Mon/2/88	Z33424	GA2	Montevideo, Uruguay	1988
Mon/1/89	Z33422	GA2	Montevideo, Uruguay	1989
Se/12/94	AF193316	GA2	Seoul, South Korea	1994
Se/05/96	AF193321	GA2	Seoul, South Korea	1996
A/WI/629-3734/98	JF920059	GA5	Wisconsin, USA	1998
NY/CH17/93	AF065255	GA5	New York, USA	1993
Mad/8/92	Z33419	GA5	Madrid, Spain	1992
Mad/1/93	Z33414	GA5	Madrid, Spain	1993
Mad/6/93	Z33410	GA5	Madrid, Spain	1993
Mad/2/93	Z33493	GA5	Madrid, Spain	1993
Se/11/95	AF193318	GA5	Seoul, South Korea	1995
Se/03/91	AF193304	GA5	Seoul, South Korea	1991
Se/10/92	AF193310	GA5	Seoul, South Korea	1992
Mad/6/92	Z33418	GA5	Madrid, Spain	1992
ON89-01/11	JN257698	GA5	Ontario, Canada	2011
Mon/1/92	Z33423	GA5	Montevideo, Uruguay	1992
Bir/6190/89	X73352	GA5	Birmingham, UK	1989
Mon/1/90	Z33494	GA5	Montevideo, Uruguay	1990
Wv/12342/84	AF065409	GA5	West Virginia, USA	1984
Wv/5222/81	AF065406	GA5	West Virginia, USA	1981
NY/CH09/93	AF065254	GA4	New York, USA	1993
Se/05/91	AF193306	GA4	Seoul, South Korea	1991
Se/01/92	AF193308	GA4	Seoul, South Korea	1992
Uk/Rss-2/76	U39662	GA4	UK	1976
NY/CH34/94	AF065257	GA1	New York, USA	1994

Mon/9/92	Z33432	GA1	Montevideo, Uruguay	1992
Mon/5/90	Z33427	GA1	Montevideo, Uruguay	1990
Mon/5/91	Z33428	GA1	Montevideo, Uruguay	1991
Bir/1734/89	X73350	GA1	Birmingham, UK	1989
Wv/2780/79	AF065405	GA1	West Virginia, USA	1979
Bir/642/89	X73354	GA1	Birmingham, UK	1989
Wv/19983/87	AF065408	GA1	West Virginia, USA	1987
Mon/9/91	Z33431	GA1	Montevideo, Uruguay	1991
Wv/6973/82	AF065407	GA1	West Virginia, USA	1982
Aus/A2/61	M11486	Prototype	Australia	1961
Usa/Long/56	M17212	Prototype	Montevideo, Uruguay	1956
Mon/7/91	Z33429	Prototype	Montevideo, Uruguay	1991
Mad/1/89	Z33456	GA3	Madrid, Spain	1989
Se/01/95	AF193317	GA3	Seoul, South Korea	1995
Se/12/95	AF193319	GA3	Seoul, South Korea	1995
Mad/4/90	Z33416	GA3	Madrid, Spain	1990
Wv/23836/88	AF065410	GA7	West Virginia, USA	1988
Bir/6256/89	X73353	GA7	Birmingham, UK	1989
Mad/5/92	Z33417	GA7	Madrid, Spain	1992
Mad/3/92	Z33455	GA7	Madrid, Spain	1992
Mad/4/91	Z33420	GA7	Madrid, Spain	1991
Bir/6614/89	X73355	GA7	Birmingham, UK	1989
Se/09/92	AF193309	GA7	Seoul, South Korea	1992
Se/12/92	AF193312	GA7	Seoul, South Korea	1992
Se/08/96	AF193322	GA7	Seoul, South Korea	1996
Bir/5857/89	X73351	GA7	Birmingham, UK	1989
A/WI/629-4071/98	JF920065	GA7	Wisconsin, USA	1998

Tab. 1: sequenze di riferimento di RSV A

Sequence	Accession Number	Genotype	Place	Date
BA/164/02	DQ227367	BA1	Buenos Aires, Argentina	2002
BA/733/02	DQ227374	BA1	Buenos Aires, Argentina	2002
BA/619/02	DQ227372	BA1	Buenos Aires, Argentina	2002
BA/166/02	DQ227368	BA1	Buenos Aires, Argentina	2002
BA/167/02	DQ227369	BA1	Buenos Aires, Argentina	2002
BA/505/02	DQ227371	BA1	Buenos Aires, Argentina	2002
BA/1461/02	DQ227382	BA1	Buenos Aires, Argentina	2002
BA/1208/02	DQ227378	BA1	Buenos Aires, Argentina	2002
BA/1606/02	DQ227386	BA1	Buenos Aires, Argentina	2002
BA/770/02	DQ227373	BA1	Buenos Aires, Argentina	2002
BA/1214/02	DQ227379	BA1	Buenos Aires, Argentina	2002
BA/1271/02	DQ227380	BA1	Buenos Aires, Argentina	2002
BA/3833/99	AY333362	BA1	Buenos Aires, Argentina	1999
BA/3859/99	AY333363	BA1	Buenos Aires, Argentina	1999
BA/3931/99	DQ227365	BA1	Buenos Aires, Argentina	1999
BA/3997/99	DQ227366	BA1	Buenos Aires, Argentina	1999
BA/802/99	DQ227363	BA1	Buenos Aires, Argentina	1999
BA/1370/99	DQ227364	BA1	Buenos Aires, Argentina	1999
NY/CH10b/90-91	AF065250	GB1	New York, USA	1990
WV/B1/85	AF013254	GB1	West Virginia, USA	1985
WV/15291/85	M73542	GB1	West Virginia, USA	1985
WV/10010/83	M73541	GB1	West Virginia, USA	1983
WV/4843/80-81	M73540	GB1	West Virginia, USA	1960
SW/8/60	M73545	Prototype	Sweden	1960
CH/18537/62	M17213	Prototype	District of Columbia, USA	1962
NM/1355/89	M73543	GB4	New Mexico	1989
MON/15/90	AY333361	GB4	Montevideo, Uruguay	1990
NY/CH9b/92-93	AF065251	GB2	New York, USA	1992
NY/CH18b/92-93	AF065252	GB3	New York, USA	1992
NY/CH53b/92-93	AF065253	GB3	New York, USA	1993
Sap/4/00-01	AB117522	BA1	Sapporo, Japan	2000
BA/1161/02	DQ227377	BA2	Buenos Aires, Argentina	2002

BA/5140/03	DQ227393	BA2	Buenos Aires, Argentina	2003
BA/4826/03	DQ227389	BA2	Buenos Aires, Argentina	2003
BA/4862/03	DQ227391	BA2	Buenos Aires, Argentina	2003
BA/495/02	DQ227370	BA3	Buenos Aires, Argentina	2002
BA/1562/02	DQ227384	BA3	Buenos Aires, Argentina	2002
BA/1856/02	DQ227387	BA3	Buenos Aires, Argentina	2002
BA/1518/02	DQ227383	BA3	Buenos Aires, Argentina	2002
BA/998/02	DQ227375	BA3	Buenos Aires, Argentina	2002
BA/1004/02	DQ227376	BA3	Buenos Aires, Argentina	2002
BA/1441/02	DQ227381	BA3	Buenos Aires, Argentina	2002
BA/1565/02	DQ227385	BA3	Buenos Aires, Argentina	2002
BA/4915/03	DQ227390	BA3	Buenos Aires, Argentina	2003
BA/4974/03	DQ227392	BA3	Buenos Aires, Argentina	2003
BA/1889/02	DQ227388	BA3	Buenos Aires, Argentina	2003
BA/6564/03	DQ227394	BA3	Buenos Aires, Argentina	2003
BA/1607/04	DQ227397	BA3	Buenos Aires, Argentina	2004
BA/524/04	DQ227396	BA4	Buenos Aires, Argentina	2004
BA/100/04	DQ227395	BA4	Buenos Aires, Argentina	2004

Tab.2: Sequenze di riferimento di RSV B

- **Allineamento delle sequenze**

Per procedere con l'analisi filogenetica è stato necessario effettuare un allineamento delle sequenze (quelle ottenute in questo studio e quelle prelevate dalla banca dati online) utilizzando il software bioinformatico ClustalX (versione 2.0.11) [100]. Il software allinea tutte le sequenze a due a due, confrontando ogni possibile combinazione, e calcola la distanza genetica per ogni coppia. I risultati di questa analisi vengono quindi utilizzati per costruire un semplice albero iniziale di sequenze, che servirà da guida per l'allineamento multiplo. Sequenze simili tra di loro vengono così raggruppate insieme e viene generato un file di output, in formato .aln, che contiene l'allineamento finale: le sequenze sono posizionate una sopra l'altra, così da incolonnare i residui omologhi delle diverse sequenze.

- **Costruzione degli alberi filogenetici**

Per la costruzione di un albero filogenetico è necessario che tutte le sequenze nucleotidiche del dataset abbiano la stessa lunghezza e, per questo motivo, attraverso l'utilizzo del programma BioEdit [99], le sequenze del file di allineamento sono state tagliate alle estremità. Gli alberi filogenetici sono quindi stati costruiti utilizzando il programma bioinformatica MEGA [101]. Gli alberi filogenetici sono stati costruiti attraverso l'utilizzo del metodo Neighbor-Joining (NJ), che consente di generare alberi senza radice (definiti unrooted), basandosi sulle matrici di distanza (calcolate secondo il modello selezionato) con un'elevata efficienza e rapidità e in cui lunghezza dei rami (linea orizzontale che collega una sequenza al resto dell'albero) risulta proporzionale alla distanza genetica tra le unità tassonomiche (rappresentate dalle sequenze) [102]. Il software, inizialmente, costruisce un albero irrisolto a cespuglio, che non presenta nessun nodo interno (punto di divergenza fra due sequenze), che sarà il punto di partenza per la costruzione dell'albero definitivo. A questo vengono poi aggiunti tutti i possibili nodi interni, creando tutti i possibili raggruppamenti, e vengono fissati solo quei nodi che consentono di ottenere l'albero con il valore di somma di tutte le lunghezze dei bracci più basso possibile. Questo processo, ripetuto sequenzialmente, raggruppa due sequenze vicine (neighbor) introducendo il ramo interno più corto possibile fino a che non sono state raggruppate tutte le sequenze.

- **Valutazione dell'affidabilità degli alberi costruiti**

Al fine di stimare l'errore statistico del modello di sostituzione scelto e per valutare la significatività di specifici clades (gruppi di sequenze che fanno riferimento ad un unico braccio) è stata effettuata un'analisi di Bootstrap [103]. Questo metodo è in grado di calcolare una stima statistica dell'affidabilità dei vari raggruppamenti attraverso sotto campionamenti casuali dei dati. Il software crea N nuovi allineamenti simulati, a partire da quello originale, ricampionando in modo casuale le colonne (alcune di esse possono essere ripetute e alcune eliminate). Per ogni allineamento viene costruita una nuova matrice di distanza e quindi costruito un albero filogenetico. Infine viene calcolata l'affidabilità dei nodi, ovvero quante volte un certo raggruppamento è stato ottenuto sul totale degli alberi costruiti (valore di bootstrap). In questo studio sono state utilizzate campionature da 1000 set e sono stati considerati significativi valori di bootstrap superiori a 70.

7.3.7 Identificazione di sequenze ricombinanti

Dall'analisi degli alberi filogenetici ottenuti è possibile identificare delle presunte sequenze mosaico, originatesi cioè a seguito di eventi di ricombinazione fra ceppi differenti. Tali sequenze presentano una posizione topologicamente intermedia rispetto ai cluster principali e spesso assumono posizioni variabili, a seconda del numero e del tipo di sequenze di riferimento incluse nell'analisi.

Al fine di identificare o escludere la presenza di ceppi ricombinanti, sono stati utilizzati due metodi, il Single Break Point scan per la ricombinazione (SBP) e il Genetic Algorithm Recombination Detection (GARD), liberamente disponibile sul server Datamonkey (<http://www.datamonkey.org/>). Questi due metodi si basano sulla discordanza filogenetica e forniscono informazioni su eventi di ricombinazione e la posizione dei punti di rottura in un allineamento multiplo e non richiedono nessun riferimento.

7.3.8 Studio evolutivo di RSV

Metodi filogenetici computazionali possono risultare utili in un contesto di Sanità Pubblica per valutare caratteristiche virologiche importanti (come ad esempio le forze di pressione selettiva che agiscono su una popolazione virale o il suo andamento di crescita), per analizzare alcune caratteristiche dell'infezione (come il numero riproduttivo di base di un'epidemia) e di stimare con buona precisione il momento in cui un dato microrganismo ha iniziato a diffondersi nella popolazione. In questo contesto, grande rilevanza ha assunto la teoria dell'orologio molecolare. Tale teoria si basa sull'ipotesi che il tasso di evoluzione (evolutionary rate) di un dato gene o di una data proteina è costante nel tempo. Questo significa che esiste un orologio molecolare che scandisce il tempo evolutivo di ogni gene: la velocità evolutiva può essere diversa per geni differenti ma risulta costante per lo stesso gene in tutti i lineage. Mediante algoritmi raffinati applicati alla teoria dell'orologio molecolare e conoscendo l'evolutionary rate di un gene, è possibile datare il tempo di divergenza fra due lineage partendo da un allineamento di sequenze. In questo modo è possibile calcolare il tempo di coalescenza o tMRCA (Time of Most Recent Common Ancestor), che rappresenta il momento in cui un determinato lineage ha cominciato a divergere a partire dal più recente antecessore comune ai ceppi del lineage stesso. Inoltre, tale teoria implica che una deviazione di un gene da tale comportamento è indice di un'evoluzione di tipo adattativo (sotto l'influenza di una pressione selettiva positiva). La teoria dell'orologio molecolare è stata applicata per poter studiare l'evoluzione di RSV e per stimare alcune caratteristiche della dinamica evolutiva del virus e della infezione: tMRCA (stima del tempo di origine della divergenza dei due genotipi St-1 e St-2), dinamica di popolazione (R_0 o numero riproduttivo di base e modalità di crescita della popolazione virale) forze di pressione selettiva (positiva e negativa) che regolano l'evoluzione virale.

- **Dataset di sequenze utilizzate per lo studio evolutivo**

Per studiare la storia evolutiva di RSV, sono stati utilizzati gli stessi dataset allestiti per l'analisi filogenetica (**tabella 1-2**).

- **Scelta del modello evolutivo**

Per la ricostruzione filogenetica è sicuramente cruciale la scelta del modello evolutivo più adeguato. In generale, si potrebbe pensare che modelli più complessi si adattino meglio degli altri ai dati, tuttavia utilizzando modelli troppo complessi deve essere stimato un più grande numero di parametri e, di conseguenza, l'errore statistico tende ad aumentare. Quindi, risulta necessario scegliere il modello più adeguato a descrivere i dati che consenta di ottenere stime accurate (ottenibili altrimenti con dataset molto ampi che però richiedono tempi di analisi troppo lunghi). Il modello evolutivo più conforme ai dati è stato selezionato mediante l'ausilio di MODELTEST in combinazione con il software PAUP versione 4.0, secondo l'approccio del hLTR (Hierarchical Likelihood ratio test) [104, 105]. Lo hLRT utilizza come parametro per mettere a confronto più modelli il valore di likelihood, o di verosimiglianza, che risulta proporzionale alla probabilità di osservare i dati (sequenze) data una certa ipotesi (modello di sostituzione nucleotidica). Tale sistema permette di calcolare i valori di likelihood di un modello, ottenuto costruendo alberi filogenetici in NJ secondo quel modello, e confrontarlo con quelli ottenuti con altri modelli. Questi valori vengono confrontati a due a due e per ogni coppia viene calcolato il valore di LRT, secondo la formula:

$$LRT = 2(l_1 - l_0)$$

dove l_1 è il log del valore di likelihood ottenuto utilizzando il modello più complesso (ipotesi alternativa) e l_0 quello del modello più semplice (ipotesi nulla). Su tali valori viene anche calcolata una stima statistica (secondo il modello del chi quadro) e i relativi valori di p . Se l'LRT risulta significativamente elevato l'aggiunta di parametri nel modello alternativo aumenta significativamente la likelihood dei dati e quindi è favorito l'uso del modello più complesso. Al contrario, una piccola differenza dei valori di log likelihood indica che l'ipotesi alternativa non rappresenta i dati in modo significativamente migliore e il modello più semplice viene preferito. MODELTEST è in grado di paragonare i valori di likelihood di 24 differenti modelli evolutivi differenti definiti annidati, ovvero in cui il modello più semplice viene considerato come un caso particolare del modello più complesso, in maniera gerarchica e definisce il modello migliore da utilizzare per lo specifico dataset.

- **Studio evolutivo**

Gli alberi filogenetici, i parametri dei modelli, gli evolutionary rates e le modalità di crescita della popolazione sono stati co-stimati utilizzando il metodo Bayesian Markov Chain Monte Carlo (MCMC) implementato nel pacchetto del software BEAST [106, 107, 108]. L'inferenza Bayesiana della filogenesi è basata su una quantità chiamata probabilità posteriore di un albero. Il teorema di Bayes afferma:

$$P(\text{Albero/Dati}) = P(\text{albero}) * P(\text{Dati/Albero}) / P(\text{Dati})$$

Dove $P(\text{Albero/Dati})$ è la probabilità a posteriori di un albero filogenetico, $P(\text{Dati/Albero})$ è la verosimiglianza (likelihood), $P(\text{Albero})$ è la probabilità a priori (prior) di un albero (in assenza di dati specifici viene assegnata la stessa probabilità a probabilità non condizionata dei dati secondo un certo modello (costante normalizzante per assicurare che la somma delle probabilità posteriori sia pari ad uno) [109]. Le probabilità a posteriori vengono stimate successivamente all'acquisizione di conoscenze sui dati e la probabilità a posteriori di un albero è la probabilità che tale albero, secondo un certo modello con determinati parametri e dato un allineamento di sequenze, sia corretto. Il risultato delle analisi è una distribuzione delle probabilità a posteriori e l'albero con la più alta probabilità posteriore è la migliore stima della filogenesi; di solito, se i dati sono informativi, vengono generati solo pochi alberi individuati come i più probabili. Le probabilità a posteriori non possono essere calcolate analiticamente, ma possono essere approssimate mediante il metodo delle Catene di Markov Monte Carlo (Markov Chain Monte Carlo, MCMC). Il metodo Monte Carlo prevede l'utilizzo di simulazioni casuali, mentre le catene di Markov sono un modello di sequenze di eventi dove la probabilità di un evento dipende dall'evento precedente. L'analisi inizia a partire da un albero iniziale (i cui parametri sono definiti casualmente) con una data probabilità a posteriori; successivamente viene variato un parametro (la topologia oppure un parametro del modello), costruito un nuovo albero e ne viene calcolata una nuova probabilità a posteriori: se il nuovo albero generato ha una probabilità maggiore, il parametro variato viene accettato, in caso contrario viene accettato solamente entro un certo valore. Tutti questi passaggi costituiscono una così detta generazione. Il valore di likelihood dell'albero iniziale è molto basso ma, mano a mano che la catena si muove verso regioni con probabilità posteriori più elevate, i valori di likelihood incrementano

rapidamente (burn-in) fino a raggiungere un plateau (ad indicare che la catena sta convergendo sulla distribuzione target). I valori campionati durante il burn-in risultano eccessivamente influenzati dalla scelta dell'albero iniziale e quindi di solito vengono scartati [109]. In questo studio sono stati utilizzati come prior due modelli demografici di crescita di popolazione (dimensione costante e crescita esponenziale) e il Bayesian Skyline Plot (BSP, un modello che calcola le effettive dimensioni delle popolazioni nel tempo) e sono state analizzate entrambe le condizioni di orologio in senso stretto (strict molecular clock, che implica un tasso di evoluzione molecolare costante in tutto l'albero) e rilassato (relaxed molecular clock, che assume invece tassi di evoluzione molecolare differenti per i diversi lineage). Le catene MCMC sono state fatte funzionare per almeno 30 milioni di generazioni e campionate ogni 300 step. La convergenza è stata valutata sulla base dell'effettiva dimensione del campione (effective sampling size, ESS, ovvero il numero effettivo di campioni indipendenti a cui è equivalente la catena, bassi valori di ESS indicano un alto grado di correlazione tra i campioni che quindi non rappresentano in modo accurato la distribuzione) dopo un 10% burn-in mediante l'utilizzo del software Tracer versione 1.4 [110]. Solo le ESS ≥ 250 sono state accettate. L'incertezza nelle stime è stata indicata come highest posterior density (HPD) al 95%, intervallo della distribuzione che contiene il 95% dei valori campionati, e il modello best-fitting è stato selezionato mediante il Bayes Factor (BF) implementato in BEAST. Il BF rappresenta il rapporto tra i valori di likelihood di due modelli:

$$BF = P(X/M_0)/P(X/M_1)$$

Dove X rappresenta i dati e M_0 e M_1 i due modelli. Maggiore è tale valore più forte è l'evidenza contro M_1 [109]. In accordo con Kass e Raftery, la forza dell'evidenza contro la H_0 è stata definita come segue: $2\ln BF < 2$ nessuna, $2 < 2\ln BF < 6$ debole; $6 < 2\ln BF < 10$ forte; $2\ln BF > 10$ molto forte. Valori negativi di $2\ln BF$ indicano invece evidenze in favore della H_0 [111]. Solo valori > 6 sono stati considerati significativi. Il supporto statistico per i determinati clade è stato ottenuto calcolando la probabilità a posteriori di ogni clade monofiletico. Per valutare se un gruppo è monofiletico viene valutata la proporzione delle volte in cui gli alberi campionati contengono tale gruppo monofiletico. L'albero ottenuto è equivalente ad un albero di consenso di bootstrap, ma invece dei valori di bootstrap, ai nodi è possibile leggere le probabilità posteriori (posterior probability).

- **Costruzione dell'albero Bayesiano**

Gli alberi campionati sono stati riassunti in un unico albero target mediante il software TreeAnnotator incluso nel pacchetto di BEAST scegliendo l'albero con la somma massima di probabilità a posteriori (massima credibilità dei clade) dopo un burn-in del 10%.

7.3.9 Calcolo delle distanze sinonime e non sinonime

Uno dei metodi più utilizzati per valutare la presenza di pressione selettiva sul frammento genico di interesse è quello di calcolare la proporzione fra le mutazioni nucleotidiche che comportano un cambiamento dell'aminoacido corrispondente (sostituzioni non sinonime) e quelle che invece lasciano inalterata la proteina (sostituzioni sinonime). Le mutazioni sinonime sono quelle soggette a pressione selettiva perché potrebbero alterare la struttura proteica e quindi influire sulla fitness di un organismo. In condizioni di pressione selettiva negativa le sostituzioni non sinonime si accumulano più lentamente di quelle sinonime e il contrario accade in condizioni di pressione selettiva positiva.

Pertanto, un eccesso di sostituzioni non-sinonime ($dN > dS$) può essere interpretato come selezione positiva, mentre il prevalere di sostituzioni sinonime ($dN < dS$) starebbe ad indicare selezione negativa [112]. Nel presente studio le distanze relative alle mutazioni sinonime (dS) e non sinonime (dN) sono state stimate attraverso l'utilizzo del pacchetto di programmi bioinformatica MEGA (versione 5.05) utilizzando il modello Nei-Gojobori (1986) con correzione Jukes-Cantor (1969) [101, 113, 114].

Il metodo Nei-Gojobori considera, per ogni codone, quante delle 9 sostituzioni possibili sono potenzialmente sinonime e quante invece potenzialmente non sinonime. Per ogni posizione nell'allineamento, facendo una media di tali quantità per il codone corrispondente in ogni sequenza e sommando il valore ottenuto da tutti i codoni possibili si ottiene il valore S (numero di siti sinonimi) e N (numero di siti non sinonimi). Il numero effettivo di sostituzioni sinonime (Ds) e non sinonime (Dn) fra 2 sequenze viene invece stimato contando le differenze codone per codone (assumendo il numero minimo di mutazioni). A questo punto è possibile stimare i tassi di sostituzione sinonima ($dS = Ds/S$) e non sinonima ($dN = Dn/N$) e il rapporto corrispondente:

$$dN/dS = w$$

A questo punto viene applicata la correzione Jukes-Cantor che prende in considerazione la possibilità di sostituzioni multiple al medesimo sito. Il tasso di sostituzione sinonima corretto sarà dato dalla formula:

$$dS_c = -3/4 \log(1 - 4/3 dS)$$

Una correzione simile si applica al tasso di sostituzione non sinonima. Una stima dell'intervallo di confidenza può essere ottenuto mediante la tecnica del bootstrap [112].

7.3.10 Analisi della pressione selettiva

Il rapporto w è ormai diventato una misura standard per valutare la pressione selettiva. Per $w=1$ si ritiene pressione selettiva assente, per $w<1$ esiste pressione selettiva negativa e per $w>1$ esiste pressione selettiva positiva. In questo studio il tasso w è stato stimato utilizzando l'approccio ML (*Maximum Likelihood*) in un modello *global single-ratio* implementato nel programma HyPhy [115, 116]. Tale metodo consente di valutare valori di w sia globalmente, che per ogni singolo braccio dell'albero, contemplando in questo modo la possibilità che differenti *lineage* possano evolvere con valori differenti di w .

Inoltre, è stata valutata la pressione selettiva sito-specifica positiva e negativa utilizzando diversi approcci: single likelihood ancestor counting (SLAC), fixed-effects likelihood e internal fixed-effects likelihood (FEL/IFEL), e Mixed effects model evolution (MEME). Il metodo SLAC utilizza l'allineamento per costruire un Albero filogenetico Maximum Likelihood da cui è in grado di ricostruire (secondo i parametri stimati di tasso di sostituzione e lunghezza dei bracci) sito per sito la sequenza ancestrale, ottenuta valutando il numero minore possibile di mutazioni. Da questa sequenza vengono quindi calcolati i valori dN e dS e confrontati fra loro. Il test assume che in assenza di pressione selettiva una sostituzione sinonima random ha probabilità $p=S/S+N$ e calcola ad ogni sito quanto il valore osservato si discosta da quello atteso; se la frequenza di sostituzioni sinonime osservate ad un determinato sito si discosta significativamente (valore p di significatività) dal valore calcolato significa che in tale sito agisce pressione selettiva.

Il metodo FEL, invece, dopo aver costruito l'albero calcola, per ogni sito, due parametri: a (tasso di sostituzione sinonima) and b (tasso sostituzione non-sinonima). Ad ogni sito viene quindi effettuato un LRT per valutare se a è uguale a b e viene ricavato un valore di significatività p. Se la differenza risulta statisticamente significativa allora il sito è soggetto pressione selettiva [115]. I siti selezionati negativamente sono stati identificati comparando le stime di SLAC e FEL di dN e dS a livello di ciascun sito con un livello di significatività = 0,1 [117]. Il valore di IFEL va a valutare i valori descritti in precedenza ma a livello dei rami interni all'albero.

Il metodo MEME, è stato recentemente sviluppato per identificare la pressione selettiva positiva sia pervasiva che episodica. Questo metodo combina gli effetti fissi che avvengono a livello di un sito e gli effetti random che avvengono a livello dei rami. Questo modello è un'estensione di FEL, dove è permesso che i valori ω possono variare lungo i rami secondo una distribuzione 2-bin, vale a dire alcuni rami possono essere sotto pressione selettiva positiva, mentre altri sotto pressione selettiva negativa.

7.3.11 Analisi dei siti di N- e O-glicosilazione

Potenziati siti di N-glicosilazione (Asn-Xaa-Ser/Thr) e siti di O-glicosilazione sono stati calcolati utilizzando i programmi NetNGlyc 1.0 [118] e NetOGlyc 3.1 [119]. I risultati sono poi stati confrontati con le rispettive sequenze di riferimento.

7.3.12 Analisi statistica

Differenze nelle frequenze sono state verificate mediante il test del X^2 . Valori di $p \leq 0,05$ sono stati considerati indicazioni di significatività statistica (test a due code). Tutte le analisi statistiche sono state effettuate utilizzando il software OpenEpi, versione 3.01 [120].

Risultati

8.1 Identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo influenzale pandemico e post-pandemico (2009-2011)

8.1.1 Indagine epidemiologico molecolare dell'infezione da virus influenzale A(H1N1)pdm09

Nel corso del periodo pandemico e post pandemico (2009-2011) sono stati raccolti 206 tamponi orofaringei da pazienti ospedalizzati che presentavano sintomatologia respiratoria grave SARI/ARDS (*Severe Acute Respiratory Syndrome/Acute Respiratory Distress Syndrome*). Tutti i campioni sono stati analizzati per la determinazione dell'infezione causata dal virus pandemico A(H1N1)pdm09.

Dall'analisi mediante Real-time RT-PCR è emerso che il virus influenzale era presente nel 58,3% (120/206; IC95%: 51,4-64,9%) dei campioni raccolti dai casi SARI/ARDS in studio.

I soggetti con questo tipo di infezione erano per il 61,7% maschi con età mediana di 42,9 anni (IQR: 40,9 anni). Il 13,3% aveva un'età ≤ 5 anni e il 67,5% aveva un'età compresa fra i 6-64 anni.

Dal punto di vista clinico, si è rilevato che: il 51,7% (62/120; IC95%: 42,7-60,5%) dei soggetti A(H1N1)pdm09-positivi aveva una diagnosi di ARDS e il 48,3% di SARI.

La presenza di ulteriori condizioni patologiche che potevano aumentare il rischio di sviluppare complicanze influenza-correlate era nota per circa la metà dei soggetti risultati positivi al virus influenzale A(H1N1)pdm09. In particolare, il 25,4% era immunodepresso, il 15,2% aveva patologie cardiache, l'11,9% era obeso (body mass index BMI ≥ 40), il 10,2% aveva patologie respiratorie croniche o neurodegenerative e il 4,2% erano donne in gravidanza.

Degno di nota, il virus pandemico A(H1N1)pdm09 è stato identificato nel 77,8% (7/9) dei casi fatali di ARDS (42,9% maschi; età mediana: 30,4 anni, IQR: 15,4 anni). Quattro (4/7: 57,1%) di questi pazienti appartenevano a categorie a rischio: due erano malati di cancro, uno era patologicamente obeso e uno presentava condizioni patologiche a livello neurologico.

L'impatto del virus pandemico A(H1N1)pdm09 nei casi ARDS/SARI risultati fatali, calcolato come *Case Fatality Rate* (CFR), è risultato pari all'11,3% (7/62, IC95%: 5,1-21,1%).

8.2 Indagine epidemiologico molecolare delle infezioni sostenute da altri patogeni respiratori noti

I campioni risultati negativi all'indagine per infezione da virus influenzale pandemico A(H1N1)pdm09 (86/206) sono stati analizzati per la ricerca di altri patogeni in grado di sviluppare sintomatologia grave a livello del tratto respiratorio.

In particolare è stato effettuato un saggio di Real-time PCR per la ricerca dei virus: influenzali stagionali (A/H3N2, B); Respiratorio Sinciziale (RSV) A e B; human Metapneumovirus (hMPV) A e B; human Rhinovirus (hRV) e Enterovirus (hEV); Adenovirus (AdV); human Bocavirus (hBoV) 1-4; human Coronavirus (hCoV) 229E, NL63, OC43, HKU1; human parainfluenza virus (hPIV) 1-4. Sono stati inoltre ricercati due batteri associati ad infezioni respiratorie: *Chlamydomphila pneumoniae* e *Mycoplasma pneumoniae*.

Dalle analisi è emerso che il 46,5% (40/86; IC95%: 36,2-57,1%) dei campioni è risultato positivo ad almeno uno dei patogeni ricercati.

Dei soggetti positivi il 64,5% erano uomini (età mediana 44,2 anni; IQR: 57,1 anni); il 30% aveva un'età ≤ 5 anni, il 47,5% età compresa tra i 6 e i 64 anni e il 22,5% età ≥ 65 anni.

HRV/hEV sono stati i patogeni più frequentemente presenti, identificati nel 27,5% (11/40; IC95%: 15,4-42,8%) dei campioni analizzati, seguiti dal virus influenzale A(H3N2) identificato nel 20% (8/40; IC95%: 9,7-34,5%) dei casi.

Un riassunto dei risultati ottenuti è riportato in **Tabella 3**

Patogeni	N. di campioni positivi		
	SARI (N=91)	ARDS (N=115)	SARI+ARDS (N=206)
Influenza A(H1N1)pdm09 virus	58	62	120
hRV/hEV	7	4	11
Influenza A(H3N2) virus	7	1	8
RSV	3	1	4
hCoV	2	2	4
hPIV	2	2	4
AdV	2	0	2
Influenza B virus	1	0	1
hBoV	1	0	1
hMPV	0	0	0
<i>Chlamydophila pneumoniae</i>	0	0	0
<i>Mycoplasma pneumoniae</i>	0	0	0
Co-infections	2	3	5

Tab 3: *Impatto dei patogeni respiratori indagati nei pazienti con SARI/ARDS (N=206).*

Escludendo l'infezione da virus influenzale pandemico A(H1N1)pdm09, nessun altro agente patogeno è stato rilevato nei campioni provenienti da soggetti deceduti in corso di ospedalizzazione.

8.1.3 Analisi molecolare mediante la tecnica di “virus-discovery” VIDISCA-454

Quarantasei (46/206; 22,3%, IC95%: 17-28,4%) casi di SARI/ARDS (compresi due casi fatali) sono risultati negativi a tutti i saggi diagnostici descritti in precedenza. Di questi, il 58,2% erano maschi (età mediana: 49,6 anni, IQR: 61,3 anni); il 18,2% aveva un'età ≤ 5 anni, il 45,4% compresa fra i 6 e i 64 anni, il 36,4% età ≥ 65 anni.

I campioni prelevati da questi pazienti sono stati analizzati mediante l'utilizzo della tecnica di “virus discovery” VIDISCA-454 durante il periodo formativo svolto presso il Laboratorio di Virologia sperimentale dell'Università di Amsterdam.

I risultati dell'indagine hanno permesso di affermare che nessuna delle sequenze ottenute era attribuibile ad un nuovo virus o ad una nuova variante virale; tuttavia questa analisi ha permesso l'identificazione di un caso di infezione da virus del morbillo (Morbillivirus) non diagnosticato in precedenza, portando così la quota dei casi di SARI/ARDS con diagnosi virale nota al 78,2% (161/206, IC95%: 72,1-83,4%).

La quota dei casi rimasta a diagnosi ignota è risultata quindi del 21,8% (45/206, IC95%: 16,6-27,9%); per la maggioranza si trattava di casi di ARDS (34/45: 75,6%, 61,5-86,4%), 2 dei quali risultati fatali.

In **Figura 37** è riportato l'esito complessivo delle indagini svolte.

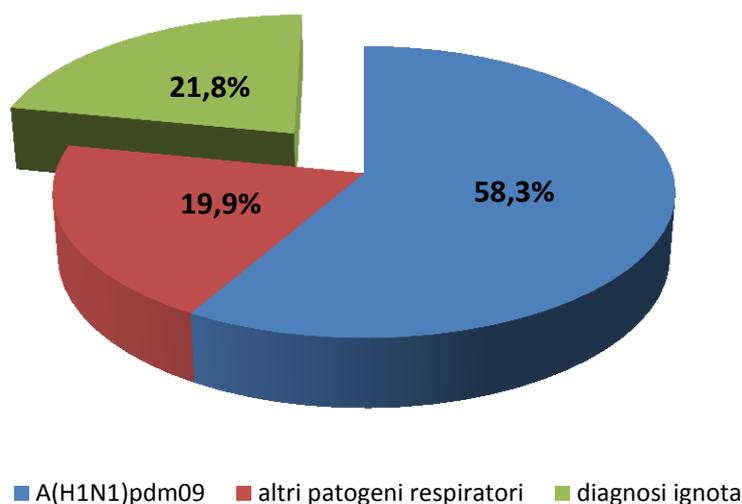


Fig. 37 . Diagnosi finale dei pazienti con SARI/ARDS in studio (N=206).

La **Figura 38** riassume i risultati dello studio condotto, riportando anche i CFR attribuibili a ciascuna tipologia di casi inclusi nelle indagini molecolari.

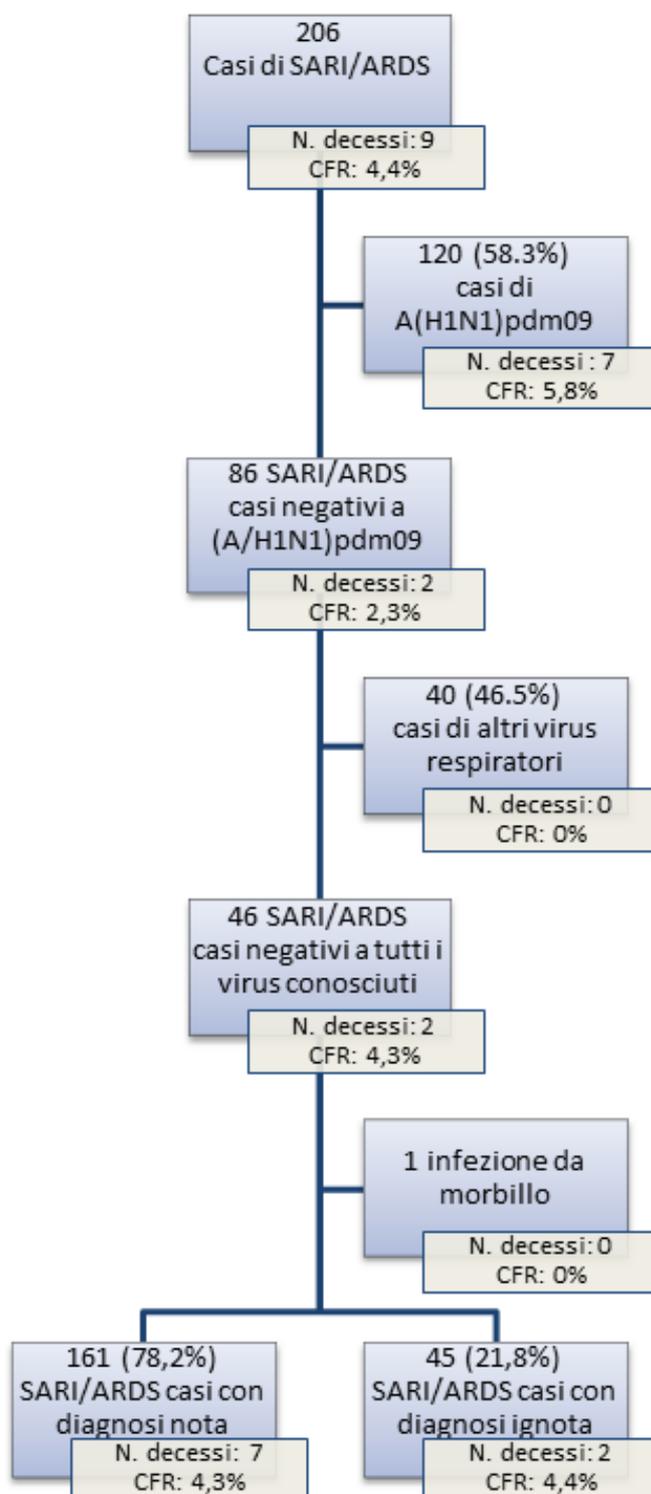


Fig. 38 Riassunto dei risultati ottenuti dello studio condotto. CFR = Case Fatality Rate

9. Variabilità genetica ed evoluzione molecolare del virus respiratorio sinciziale (RSV)

Per valutare la variabilità genetica e l'evoluzione molecolare del virus respiratorio sinciziale è stato analizzato un campione rappresentativo di sequenze ottenute da campioni di soggetti pediatrici ospedalizzati tra il 2006 e il 2012 con diagnosi positiva per infezione da RSV.

9.1 Analisi delle sequenze del gene G di RSV A

Per l'analisi filogenetica del virus RSV A sono state utilizzate 23 sequenze ottenute da campioni RSV-A positivi e 78 sequenze ottenute dalla banca dati GenBank.

Dall'analisi è emerso che: 3 delle sequenze in studio si sono inserite nel cluster formato dalle sequenze appartenenti al genotipo GA2 con un valore bootstrap del 78%, mentre tutti e tutte gli altri isolati sono rientrati nel gruppo NA1 (**Fig.39**).

Le sequenze geniche del gene G appartenenti al genotipo NA1 hanno mostrato un elevato livello di somiglianza nucleotidica (mean p distance = $0,016 \pm 0,002$ sub/site). Le sequenze incluse nel genotipo GA2 hanno evidenziato una distanza genetica media significativamente superiore a quella di NA1 (mean p distance = $0,032 \pm 0,003$ sub/site) e simile alla distanza media tra i due genotipi (mean p distance = $0,044 \pm 0,006$ sub/site).

Comparando le sequenze in studio alla sequenza di riferimento RSV-A2 (Accession number: M11486), sono state identificate 31 sostituzioni aminoacidiche. Le mutazioni identificate sono state: P101F/L, P104L, E106G, I107T, I111S, G121S, V122A, K123E, L126P/S, T133I/V, T141I, P156Q, S157N, L208I, S222P, P226L, E233K, N237D/N, I244R, L258H, M262E, F265L, S269T, S280Y, P286L, P289S, S290P, P292S, P293S, P296T/S e R297K (**Fig.40**).

Tutte le sequenze analizzate hanno mostrato un codone di stop a livello dell'aminoacido 298 del gene G e solo in una sequenza in studio (MI/A/16-11/10) è stata identificata una sostituzione aminoacidica (N179K), a livello degli aminoacidi 159-186, corrispondente al dominio immunogenico della regione conservata del gene G.

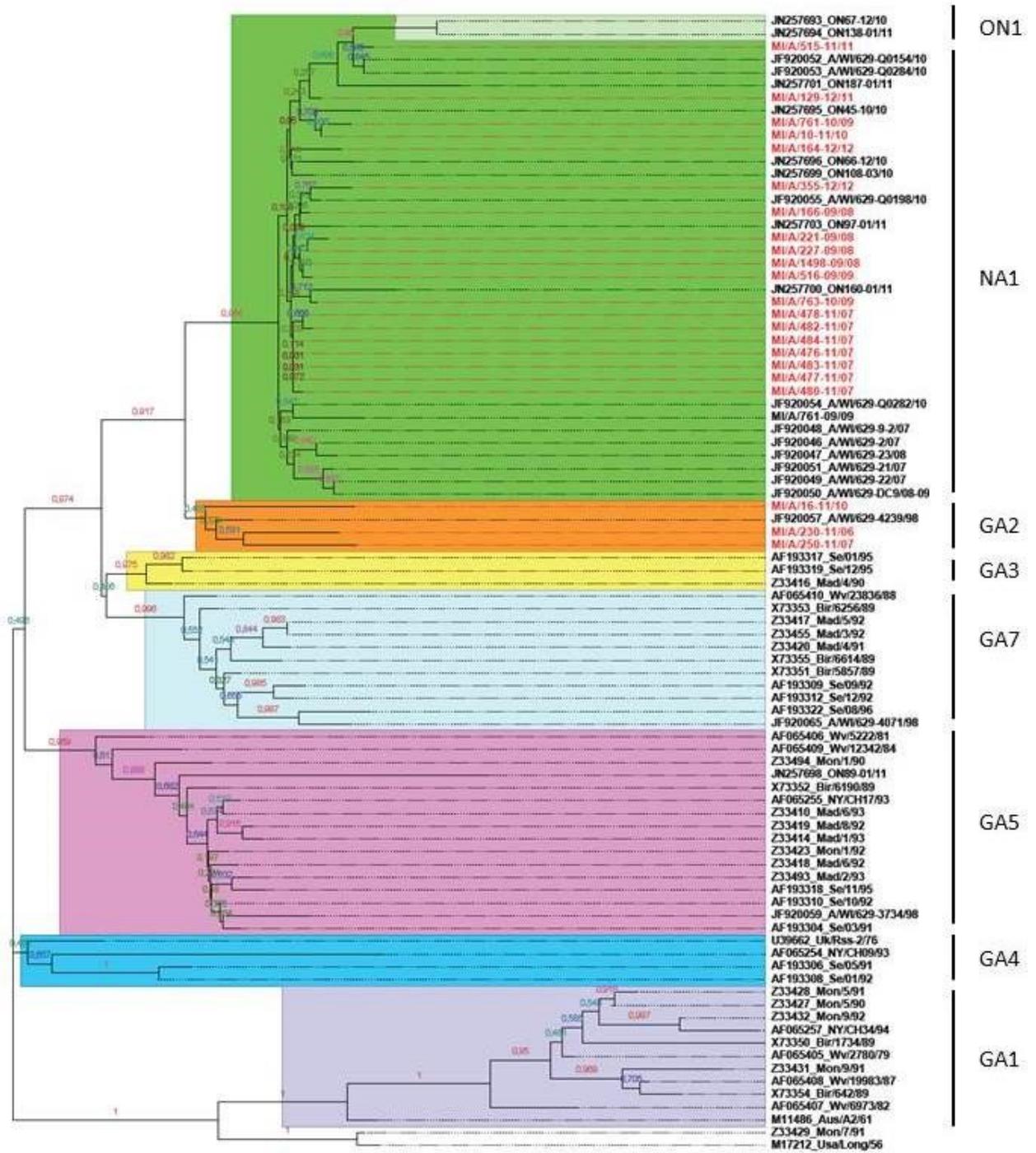


Fig. 39 Albero filogenetico costruito con il modello Neighbor-Joining delle sequenze di RSV-A.

9.2 Analisi delle sequenze del gene G di RSV B

L'analisi filogenetica del gene G di RSV B è stata effettuata utilizzando 12 sequenze isolate dai campioni in studio e 50 sequenze ottenute da GenBank.

I risultati hanno mostrato che tutte le sequenze in studio appartenevano al genotipo BA e, in particolare clusterizzavano all'interno di un sottogruppo non supportato da bootstrap significativo, insieme alla sequenza di riferimento del sottogruppo BA4 (**Fig. 41**).

Tutte le sequenze analizzate hanno mostrato una delezione di 6 nucleotidi dopo il residuo 489 del gene G, già descritto per la sequenza DQ227395 (Trento et al., 2006) e un codone di stop a livello dell'aminoacido 313.

Gli isolati appartenenti ai genotipi GB avevano una distanza genetica media (mean p distance = $0,043 \pm 0,005$ sub/site) superiore a quelli appartenenti al genotipo BA (mean p distance = $0,020 \pm 0,003$ sub/site); anche il sottogruppo BA4 (principalmente i ceppi italiani) ha presentato una distanza genetica media (mean p distance = $0,025 \pm 0,003$ sub/site) significativamente più alta rispetto agli altri ceppi BA (mean p distance = $0,011 \pm 0,002$ sub/site).

Rispetto alla sequenza di riferimento RSV-B1, 8 sostituzioni amminoacidiche sono state identificate fra tutte le sequenze in studio (**Fig.42**).

Alcune mutazioni amminoacidiche sono state identificate anche a livello della duplicazione di 60 nucleotidi situata tra gli aminoacidi 241 e 281 del gene G.

Anche in questo caso, è stata identificata una mutazione, H164Y, a livello del dominio immunogenico conservato della sequenza MI/B/220-09/09. Tutti questi dati sono indicati in Figura 2 .

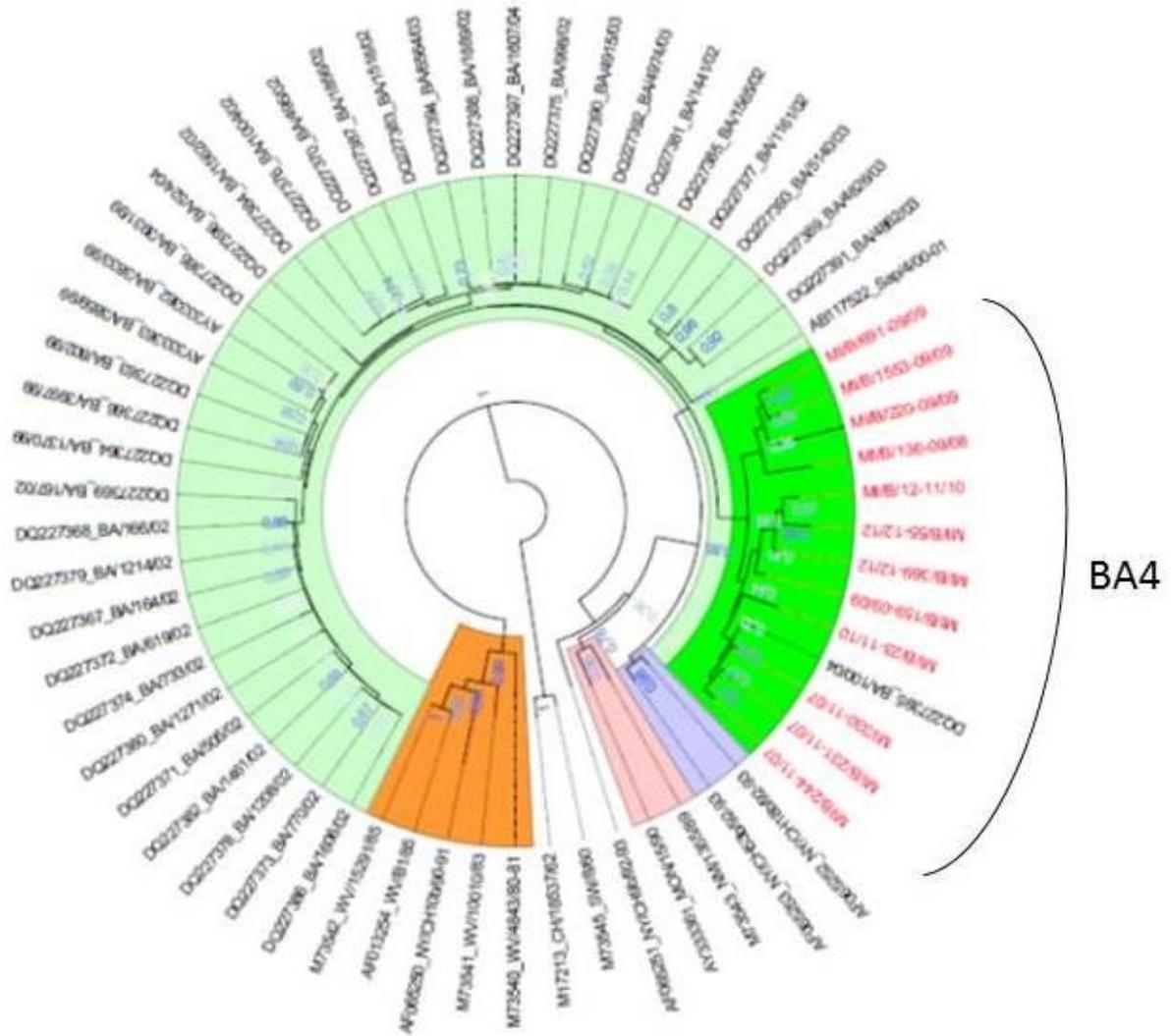


Fig. 41: Albero filogenetico costruito con il modello Neighbor-Joining delle sequenze di RSV-B

Fig. 42: Analisi aminoacidica delle sequenze di RSV-B.



9.3 Identificazione di sequenze ricombinanti

Dall'analisi di ricombinazione non è emersa la presenza di sequenze ricombinanti tra quelle in studio.

9.4 Stima dell'Evolutionary rate

Per l'analisi dell'evolutionary rate, entrambi i modelli di orologio molecolare (in senso stretto e rilassato) sono stati testati mediante l'impiego di due diversi modelli demografici (dimensione costante della popolazione e *non-parametric Bayesian skyline plot*).

Il confronto effettuato attraverso il Bayes factor test ha dimostrato che:

- (i) il modello di orologio molecolare che poteva descrivere significativamente meglio i dati era il modello di orologio rilassato
- (ii) il *non-parametric Bayesian skyline plot* è risultato il modello di popolazione più adatto per valutare gli allineamenti in studio

Questa scelta è risultata ottimale sia per l'analisi delle sequenze di RSV-A (2lnBF di orologio in senso stretto vs. orologio rilassato = 107,1; 2lnBF di BSP vs. dimensione costante = 31,9), che per quelle di RSV-B (2lnBF di orologio in senso stretto vs orologio rilassato = 30,8 e 2lnBF BSP vs dimensione costante = 27,4).

Utilizzando tali condizioni, sono stati calcolati i tassi medi evolutivi che sono stati stimati pari a:

- $2,1 \times 10^{-3}$ subs/site/year (95% HPD: $1,7-2,5 \times 10^{-3}$) per RSV-A
- $3,03 \times 10^{-3}$ subs/site/year (95% HPD: $2,1-3,9 \times 10^{-3}$) per RSV-B.

9.5 Bayesian dated tree

La **Figura 43** mostra l'albero filogenetico con radice ottenuto con metodo Bayesiano: la lunghezza dei rami è proporzionale al tempo. In corrispondenza dei nodi dei due alberi sono indicati i valori di posterior probability.

L'analisi dell'albero ottenuto mediante l'elaborazione delle sequenze di RSV-A ha mostrato alcuni valori altamente significativi, corrispondenti ai principali genotipi di RSV-A (GA1, GA5, GA7 e NA1).

Come osservato in precedenza nell'analisi filogenetica, la maggior parte dei ceppi italiani segregano all'interno del clade NA1, ma 3 isolati si inseriscono significativamente nel clade formato da alcune sequenze appartenenti al genotipo GA2.

In **Tabella 4** sono riportate le stime di tMRCA (time to Most Recent Common Ancestor) dei principali nodi interni.

In particolare, il valore di tMRCA per la radice albero di RSV-A è risultato risalire a 71 anni fa (95% HPD = 60-85), suggerendo un'origine dei ceppi attualmente in circolazione risalenti in media al 1940. Il genotipo GA1 è stato il primo genotipo a divergere tra 51 e 57 anni fa (1955-1961), tutti gli altri genotipi sembrano aver avuto origine tra 25 e 38 anni fa (1974-1987), mentre il più recente genotipo NA1 sembra aver iniziato a circolare nella popolazione tra i 7 e gli 11 anni fa (2001-2005). I ceppi isolati presso il nostro laboratorio appartenenti al clade NA1 condividevano un unico nodo significativo interno con un tMRCA medio stimato a 7 anni fa (2005; 95 % HPD tra il 2003 e il 2006). Invece i tre isolati del genotipo GA2 hanno mostrato un tMRCA significativo che risale a 16 anni fa (95 % HPD = 14-18), il che suggerisce un'origine risalente al 1996 (1994-1998).

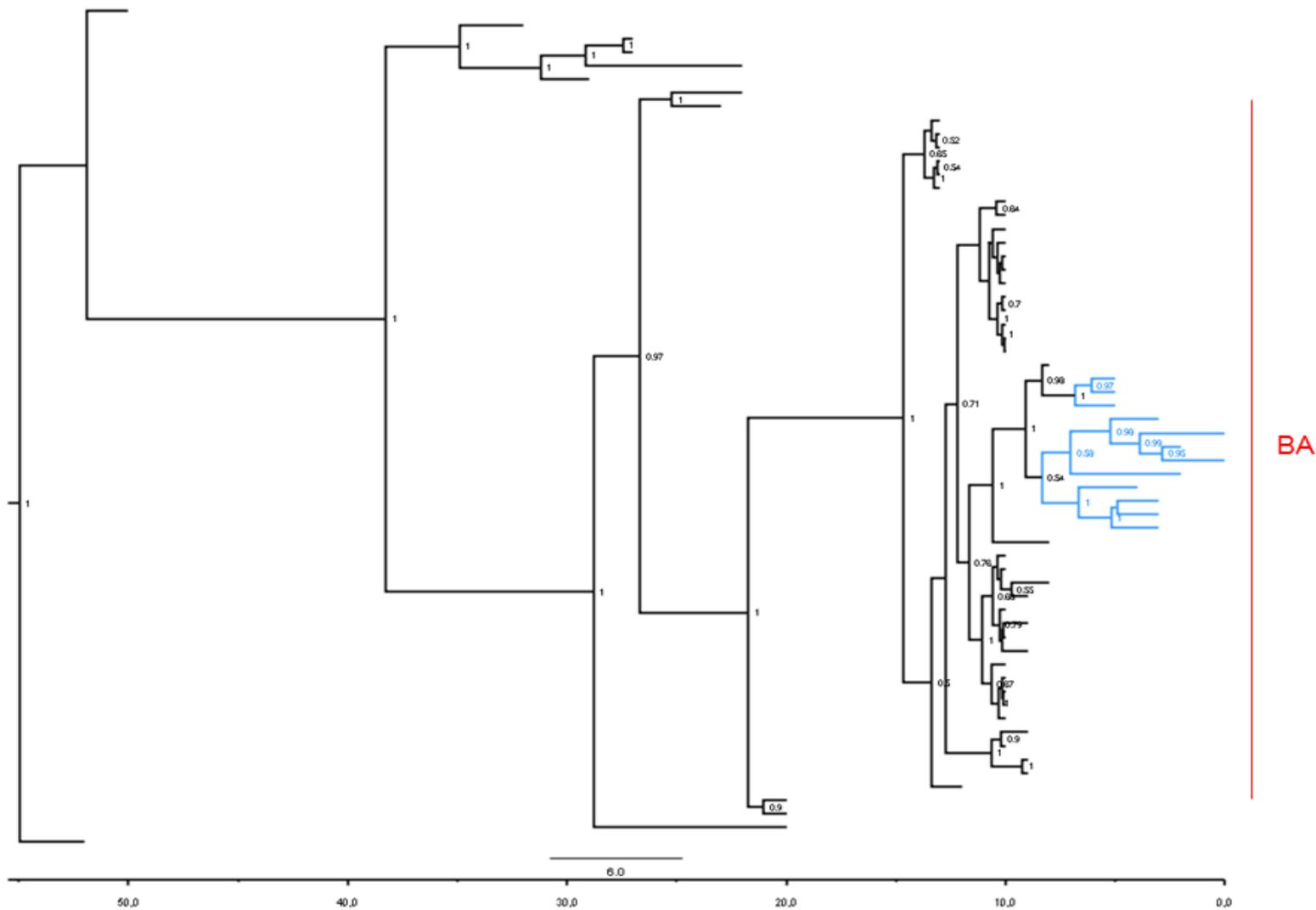
L'albero datato ottenuto dall'analisi delle sequenze di RSV B, ha mostrato una struttura temporale simile a quella dell'albero ottenuto con le sequenze di RSV A (**Fig.44**). In particolare, la radice dell'albero ha avuto un tMRCA medio stimato di 55 anni fa (1957). Tutti gli isolati in studio sono stati inclusi nel genotipo BA che è nato circa 15 anni fa (1997). In

particolare, essi clusterizzano all'interno di un subclade significativo, avente una stima tMRCA di 9 anni fa (2003), che comprende anche un unico sottogruppo il BA4.

Virus	Nodo	tMRCA	95%HPD L	95%HPD U	Anno	Intervalli
RSVA	Radice	71	60	85	1941	1927-1952
	GA1	53	51	57	1959	1955-1961
	GA5	35	32	38	1977	1974-1980
	GA4	32	26	39	1980	1973-1986
	GA7	28	26	31	1984	1981-1986
	GA3	30	27	34	1982	1978-1985
	GA2	27	25	29	1985	1983-1987
	NA1	9	7	11	2003	2001-2005
	Sequenze in studio NA1	7	6	9	2005	2003-2006
	Sequenze in studio GA2	16	14	18	1996	1994-1998
RSVB	Radice	55	52	61	1957	1951-1960
	BA	15	13	16	1997	1996-1999
	Sequenze in studio BA 4	9	8	10	2003	2002-2004

Tab. 4: Valori di tMRCA stimati per ciascun genotipo di RSV A e RSV B

Fig. 44 Albero filogenetico ottenuto dall'analisi delle sequenze di RSV-B con il modello Bayesiano.



9.6 Dinamiche di popolazione

Il Bayesian skyline plots (BSP) ottenuto dall'analisi delle sequenze in studio di RSV-A (**Fig. 45**) ha evidenziato che dal 1940 (anno della radice), una primo picco di infezioni è avvenuto nella seconda metà degli anni 1980, seguito da una diminuzione di eventi di trasmissione terminato nel 2005, quando una rapida crescita ha ripristinato la dimensione della popolazione virale originale.

Il BSP RSV-B ha mostrato un andamento simile, con una diminuzione del numero effettivo di infezioni che si verificano tra la metà degli anni 1980 e 1990, seguita da una rapida crescita nei primi anni 2000 (**Fig. 46**).

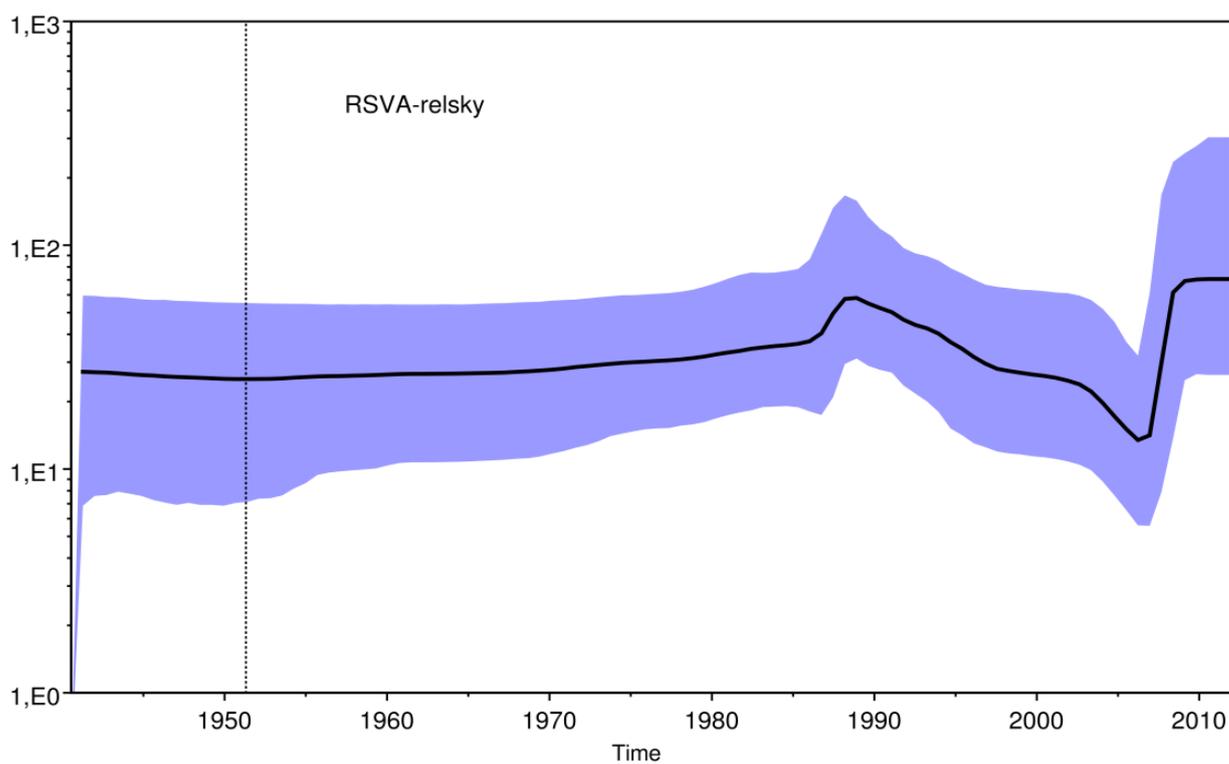


Fig. 45: Bayesian skyline plots (BSP) ottenuto dall'analisi delle sequenze in studio di RSV-A

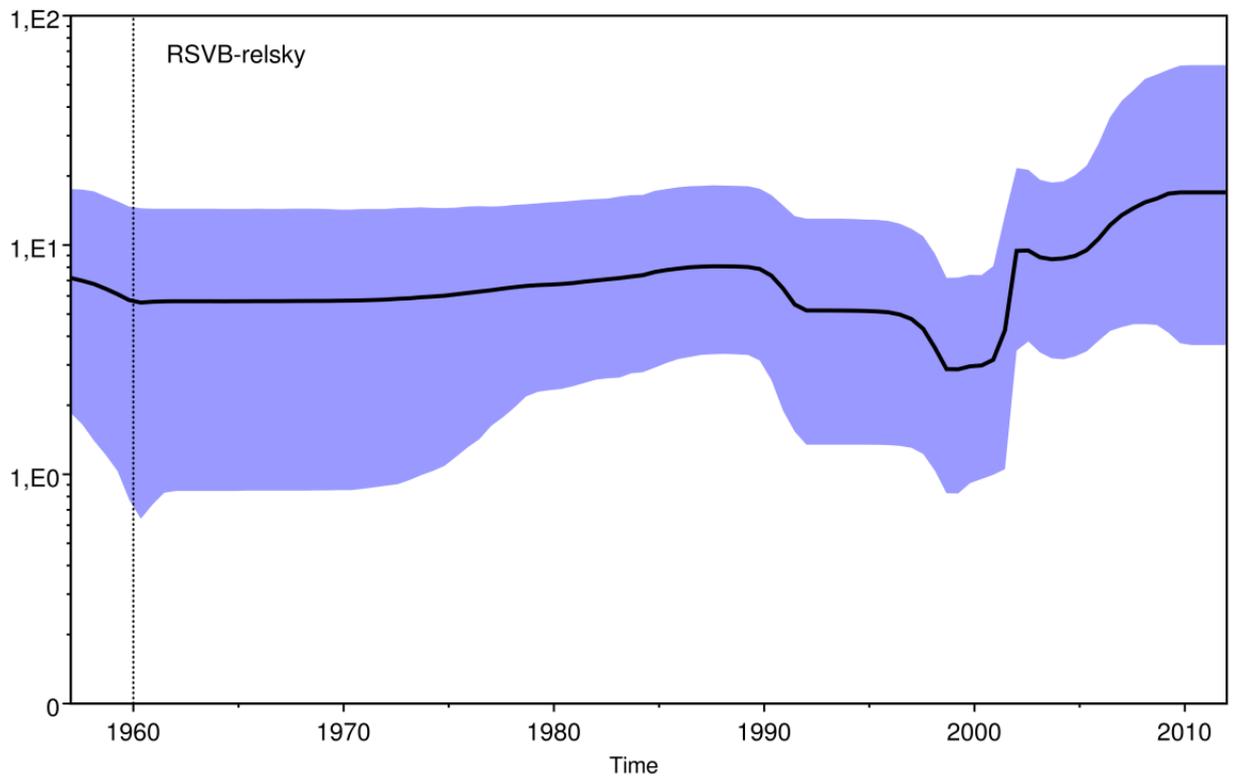


Fig. 46: Bayesian skyline plots (BSP) ottenuto dall'analisi delle sequenze in studio di RSV B.

9.7 Analisi sito specifica della pressione selettiva

Il valore $[dn-dS (\pm SE)]$ calcolato per le sequenze RSV A è risultato di $[0,020 (\pm 0,002)]$.

L'analisi di pressione selettiva sito specifica ha identificato, con almeno 3 metodi diversi, 10 codoni sottoposti a selezione positiva (115, 121, 132, 146, 215, 244, 250, 262, 286 e 297). Di questi, tutti tranne uno (146) hanno evidenziato positività anche a livello dei rami interni. Quattro siti selezionati positivamente si trovano nella prima regione ipervariabile (posizioni aminoacidiche: 115, 121, 132 e 146) e 6 nei siti selezionati positivamente si trovano nella seconda regione variabile del dominio C-terminale (posizioni aminoacidiche: 215, 244, 250, 262, 286 e 297) (**Tab. 5**).

Limitando l'analisi di pressione selettiva al clade NA1, è emerso che solo un codone (Q142P/L) è risultato sotto selezione positiva episodica (MEME). Indagando le dinamiche temporali di questo sito, inoltre, è stata osservata la graduale sostituzione del mutante maggiormente circolante nel 2007 (Q142), con uno diverso (L142), che è diventato prevalente nel 2010 e nel 2011 (**Fig. 47**).

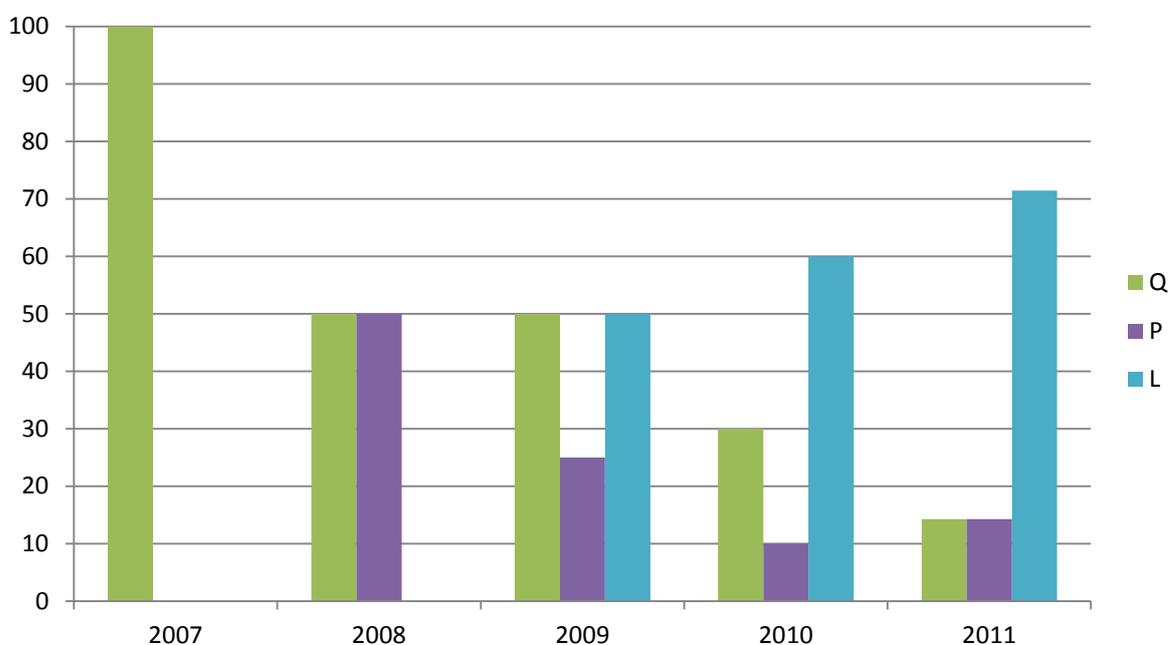


Fig. 47: Dinamiche temporali del sito 142 delle sequenze di RSV A

Un totale di 39 siti sono stati trovati essere sotto pressione selettiva negativa mediante analisi con FEL e 17 di loro sono risultati essere sotto selezione negativa anche mediante l'analisi con IFEL evidenziando questa pressione anche a livello dei rami interni.

GENOTIPO	MUTAZIONI CARATTERISTICHE
NA1	P (100% pos 111); T/I,V,S (90% pos 133); N/D (80% pos 237); P/S (100% pos292); N/K (100% pos297)
GA1	F/P (80% pos101); L/P (80% pos 104); T/ I (100% pos 107); I,T (100% pos 111); SA/GV (100% pos 121-122); I/A,T (100% pos 141); N/S (80% pos 157); I/L,P,F (100% pos 208); P/S (80% pos 222); K/E (100% pos 233); R/ I,T (100% pos 244); G/R (80% pos 254); H/L (100% pos 258); E/M (100% pos 262); L/F (100% pos 165); Y/S (100% pos 280); S/P (100% pos 293); T/S,P (90% pos 291) N/R (100% pos 292)
GA2	P/L (90% pos 290)
GA3	Q/P,L (90% pos 127); P/Q (90% pos 230)
GA4	I/T (90% pos 208); P/S (90% pos 234);
GA5	S/F (100% pos 102); T/A (90% pos 113); T/I (90% pos125); N/S (100% pos 191); T/I (90% pos 238); T/P (100% pos 241); N/S (100% pos 251); P/L (80% pos 256); P/T (80% pos 274); T/I (100% pos 295)
GA7	IL/TP (100% pos 114-115); H/Y (80% pos 258); S/P (90% pos 270)

Tab. 5: Mutazioni genotipo specifiche di RSV A

Il valore [dn-dS (\pm SE)] calcolato sull'allineamento ottenuto dalle sequenze di RSV-B era [0,034 (\pm 0,003)].

L'analisi di selezione codone specifica ha identificato un unico sito selezionato positivamente da tutti e tre i metodi utilizzati che si trova nella seconda regione ipervariabile (posizione aminoacidica 217). Considerando anche i risultati confermati da solo due dei metodi utilizzati, altri 4 siti sono stati individuati essere sotto pressione selettiva positiva significativa (138, 144, 227 e 265), di cui però solo il sito a livello della posizione 265 è risultato avere un valore significativo anche per la pressione selettiva a livello dei rami interni (**Tab. 6**).

Anche in questo caso, i siti sotto selezione negativa erano i prevalenti. Un totale di 27 siti sono stati rilevati da FEL, 17 dei quali sono stati anche confermati da IFEL.

GENOTIPO	MUTAZIONI CARATTERISTICHE
BA1	-
BA2	S/N (100% pos 188); I/K (100% pos 198)
BA3	T/I (100% pos 248); T/N (100% pos 290)
BA4	T/S (70% pos 138); L/P,T (100% pos 221); TV/IA (70% pos 249); H/Y (80% pos 265)
GB	P/S (70% pos 109); N/T (70% pos 144); P/T (70% pos 205); L/P (70% pos 217); I/T (100% pos 227); L/P (70% pos 235)

Tab. 6: Mutazioni genotipo specifiche di RSV B

9.8 Analisi dei siti di glicosilazione

Diversi siti di O- e N-glicosilazione sono stati identificati analizzando le sequenze di RSV-A in studio. Tutti gli isolati hanno mostrato un profilo simile con un elevato numero di potenziali siti di O-glicosilazione, tra 57 e 64 siti per sequenza (G score 0,5-0,8).

La stessa regione è stata esaminata per la ricerca di siti di N-glicosilazione (Asn-X-Ser/Thr) e sono stati individuati 6 possibili siti. In particolare, quasi tutte le sequenze appartenenti al genotipo NA1 hanno mostrato 4 siti putativi (AA 103, 135, 251, 294), al contrario, nelle sequenze appartenenti al genotipo GA2 (ad eccezione di MI/A/230-11/06) sono stati trovati cinque siti potenziali, gli stessi indicati sopra e uno a livello dell'aminoacido 237. La stessa caratteristica è stata trovata anche nella sequenza MI/A/515-11/11 appartenente al genotipo NA1. Solo una sequenza, MI/A/250-11/07, ha mostrato tutti e 6 i potenziali siti di N-glicosilazione (**Fig.40**).

Analizzando le sequenze di RSV-B con il server NetOGlyc 3.1, anche in questo caso sono stati trovati molti siti di O-glicosilazione potenziali, tra i 59 e i 69, a seconda della composizione aminoacidica della sequenza (G score: 0,5-0,7).

D'altra parte, analizzando le stesse sequenze con il server NetNGlyc 1.0, tutte le sequenze hanno evidenziato 4 putativi siti di N-glicosilazione (AA 86, 230, 296 e 310) e la sequenza MI/B/691-09/09 ha mostrato la presenza di 5 siti potenziali, i quattro descritti precedentemente e uno a livello dell'aminoacido 210 del gene G (**Fig.42**).

CONCLUSIONI

Le infezioni acute delle vie respiratorie (*Acute Respiratory-tract Infections, ARI*) rappresentano un problema sanitario di grande rilevanza in quanto costituiscono una delle maggiori cause di morbosità e di mortalità a livello mondiale [1-4]. Le ARI sono infezioni ubiquitarie, altamente contagiose che colpiscono soggetti di qualsiasi età, ripresentandosi più volte nell'arco della vita. Secondo una stima dell'Organizzazione Mondiale della Sanità (OMS), circa il 20% delle cause di morte dei bambini di età inferiore ai 5 anni è rappresentata dalle manifestazioni cliniche delle ARI [1].

Le infezioni respiratorie hanno un andamento tipicamente epidemico causato dalla loro elevata contagiosità e diffusione che avviene per via aerea diretta, indiretta e tramite contatto. La frequenza maggiore delle epidemie avviene nelle stagioni freddo-umide [2,3].

A seconda del distretto colpito, le ARI possono essere divise in URI (*Upper Respiratory Infection* – Infezioni delle alte vie respiratorie) e LRI (*Lower Respiratory Infection* – Infezioni delle basse vie respiratorie). Le URI comprendono infezioni quali raffreddore, faringite, laringite, tracheite mentre le LRI sono caratterizzate da patologie più gravi come bronchite, bronchiolite e polmonite. Le LRI colpiscono invece con meno frequenza, ma causano quadri clinici più severi con maggiori rischi di complicanze ed ospedalizzazioni.

I virus sono di gran lunga i microrganismi più comunemente in causa: attualmente sono noti più di 200 sierotipi, antigenicamente differenti, correlati ad infezioni respiratorie ed appartenenti alle grandi famiglie Adenoviridae, Coronaviridae, Orthomyxoviridae, Paramyxoviridae e Picornaviridae [5].

Tra gli agenti virali a trasmissione aerea di recente identificazione possiamo annoverare alcuni sottotipi di Coronavirus umani (SARS-CoV, hCoV-HKU1 e hCoV-NL63) identificati tra il 2002-2004 [31, 70], e nuovi virus della famiglia Picornaviridae (Rhinovirus di tipo C, Parechovirus tipi 2-14) scoperti nel 2007 [24, 26]. Di grande impatto, è stata l'identificazione di human Metapneumovirus (hMPV), isolato per la prima volta nel 2001 [35], di human Bocavirus (hBoV) nel 2005 [51], di due virus appartenenti alla famiglia Polyomaviridae (KIPyV e WUPyV) caratterizzati nel 2007 [57, 61], del virus influenzale A/H1N1v, identificato nel 2009 e responsabile della prima pandemia del nuovo millennio [64] e, più recentemente, il altro coronavirus MERS-CoV ha fatto il suo ingresso nel 2012 nel panorama mondiale.

Accanto ai virus, numerosi agenti batterici possono essere implicati in forme respiratorie anche gravi. Tra questi, in particolare, un ruolo rilevante assumono *Chlamydothila pneumoniae* e *Mycoplasma pneumoniae*.

Dato l'ampio spettro di sintomi comuni causati da molti agenti eziologici responsabili di ARI, non è possibile effettuare una diagnosi precisa basandosi solamente sulla sintomatologia. È noto, infatti, come lo stesso agente possa causare quadri clinici diversi e come il medesimo quadro possa essere attribuito a differenti agenti. Per tale motivo la precisa definizione del ruolo eziologico può avvenire solo tramite indagini specifiche di biologia molecolare.

Lo scopo del presente lavoro di tesi è stato quello di utilizzare nuove metodiche di biologia molecolare, di "virus discovery" e di bioinformatica al fine investigare e approfondire l'aspetto eziologico in soggetti con sintomatologia respiratoria acuta grave.

In particolare, uno degli obiettivi specifici è stato quello di identificare gli agenti patogeni responsabili di infezioni respiratorie gravi e complicate quali Infezione Respiratoria Acuta Severa (SARI) e Sindrome da Distress Respiratorio Acuto (ARDS) durante il periodo influenzale pandemico e post-pandemico (2009-2011).

A tal fine sono stati analizzati tamponi orofaringei prelevati da 206 soggetti ospedalizzati con diagnosi di sindrome respiratoria acuta grave, andando a valutare la quota di infezioni respiratorie gravi associate all'infezione da virus influenzale pandemico A(H1N1)pdm09 e da altri patogeni respiratori noti, così come la presenza, nei campioni negativi alle indagini precedenti, di virus non ancora conosciuti utilizzando una tecnica di "virus discovery".

Dalle analisi effettuate è emerso che durante il periodo pandemico e post-pandemico diversi agenti patogeni hanno co-circolato e sono risultati associati ai casi clinici in studio; tuttavia il virus influenzale A(H1N1)pdm09 ha mostrato il maggiore impatto risultando presente in oltre il 58% dei pazienti ed evidenziando un Case Fatality Rate dell'11,3% fra i casi di ARDS. I casi da A(H1N1)pdm09 risultati fatali si sono verificati in giovani adulti, per il 43% circa non appartenenti ad alcuna categoria a rischio.

La proporzione di casi di SARI/ARDS associati ad infezione sostenuta da altri virus respiratori è risultata significativamente più bassa (19,4% vs 58,3%, p-value < 0,0000001).

Gravi sintomi respiratori causati dall'infezione da A(H1N1)pdm09 sono stati più frequentemente individuati tra i pazienti di età compresa tra i 6 e i 64 anni (67,5 % vs 47,5%, p-value = 0,03), mentre quelli associati ad altri virus respiratori sono stati rilevati con maggior frequenza nei bambini d'età inferiore ai 5 anni (13,3% vs 30%, p-value = 0,02).

Rhinovirus e Enterovirus sono stati i virus maggiormente identificati, tuttavia anche il virus influenzale stagionale A(H3N2) ha svolto un ruolo significativo.

Più della metà (51,7%) delle infezioni da A(H1N1)pdm09 ha portato all'insorgenza di ARDS, mentre il virus influenzale A(H3N2) ha causato l'insorgenza di ARDS in un solo paziente, immunocompromesso in quanto infetto da HIV. Nel complesso, la percentuale di casi di SARI/ARDS-correlati ad un virus influenzale di tipo A è risultata del 62,1% (128/206), sottolineando così il ruolo centrale di questo virus nelle infezioni respiratorie gravi [121, 122].

La tecnica di "virus discovery" VIDISCA-454, applicata nel corso del periodo formativo svolto presso il laboratorio di Virologia sperimentale dell'Università di Amsterdam, si è dimostrata una metodologia sensibile e specifica, applicabile con successo alla sorveglianza delle infezioni respiratorie virali, un campo che necessita di attenzione perché in costante evoluzione, come ampiamente dimostrato dalla recente comparsa di virus emergenti, quali influenzavirus A(H5N1) e A(H7N9), MERS-Coronavirus, ecc.

L'esperienza condotta per questo lavoro di tesi ha permesso l'identificazione, mediante VIDISCA-454, di un caso di SARI associato ad infezione da virus del morbillo (Morbillivirus) sfuggito alla precedente diagnosi. Questo dato confermerebbe quanto già osservato da altri Autori [123], suggerendo l'inserimento della determinazione di tale infezione tra quelle incluse nel pannello diagnostico per le malattie respiratorie complicate.

Nonostante l'ampio spettro di agenti patogeni, virali e batterici, ricercati, quasi il 22% dei casi di SARI/ARDS non ha ottenuto una diagnosi definitiva.

Nella pratica clinica, particolare priorità dovrebbe essere dedicata al miglioramento dei protocolli di diagnosi delle infezioni respiratorie gravi, proprio per ridurre tale "gap diagnostico". Affidarsi ad una diagnosi più accurata potrebbe portare benefici sia al paziente, in termini di terapia mirata con l'utilizzo di farmaci antivirali appropriati, sia fornire

informazioni epidemiologiche più dettagliate riguardo gli agenti virali circolanti nella popolazione, allertando in modo tempestivo, in caso di necessità, le autorità sanitarie.

Un secondo obiettivo del presente lavoro è stato quello di indagare in modo più approfondito una delle infezioni respiratorie più rilevanti, in particolare nella popolazione pediatrica.

Allo scopo si è voluto valutare quali ceppi avessero circolato in Lombardia nel periodo compreso tra il 2006 e il 2012 in una popolazione di pazienti pediatrici ospedalizzati per ARI per poter delineare le dinamiche evolutive dell'infezione da RSV, sottotipi RSV A e RSV B.

E' stato quindi condotto uno studio filogenetico e filodinamico, analizzando sequenze del gene codificante per la proteina di superficie G, glicoproteina transmembrana di tipo II, coinvolta nella fase di legame tra RSV e recettori della cellula, nonché target principale della risposta anticorpale neutralizzante dell'ospite.

La variabilità antigenica di RSV, e in particolare della glicoproteina G, è la caratteristica chiave associata all'abilità del virus di re-infettare lo stesso ospite più volte nell'arco della vita e causare epidemie annuali [96, 97].

Dalle analisi condotte è stata osservata una circolazione prevalente di due genotipi (NA1 e GA2) del sottotipo RSV-A e di un unico genotipo (BA) per il sottotipo RSV-B. Questa tendenza è stata riportata anche in altri Paesi: dati pubblicati recentemente da studi condotti nelle Filippine [124], in Malaysia [125], e in Cina [126] hanno riportato la circolazione di tali genotipi nel continente asiatico.

Una predominanza dei genotipi NA1 e GA2 è stata rilevata anche in Europa: Adams et al. [127] hanno documentato come il genotipo GA2 fosse quello a maggior diffusione in Germania tra il 2003 e il 2008, mentre Forcic et al. [128] hanno riportato come il genotipo NA1 fosse quello identificato nell'82,9% dei ceppi RSV A isolati in Croazia tra il 2006 e il 2008 e come tutte le sequenze RSV-B fossero incluse nel clade BA.

Il genotipo ON1, di recente identificazione, è stato rilevato in molti Paesi, tuttavia nessuna delle sequenze da noi analizzate ha mostrato la duplicazione di 72 nucleotidi tipica di questo genotipo.

Ad oggi, in letteratura non ci sono dati per un confronto sulla circolazione in Italia di RSV nel periodo da noi considerato (2006-2012); solo uno studio, condotto in un arco di tempo

precedente (1997-2006), riporta come la maggior parte degli isolati italiani fosse riconducibile prevalentemente ai genotipi GA2 e GA5 [129].

Al fine di ricostruire la filodinamica dei genotipi di RSV A e B è stata effettuata un'analisi di due dataset costituiti da sequenze ottenute dai nostri isolati e da sequenze disponibili in banca dati (GenBank), per le quali era nota sia la data che la località di isolamento.

I tassi evolutivi del gene G stimati in questo studio, sono risultati essere vicini al limite superiore di quelli stimati in precedenza da altri Autori [130].

Sulla base di questi abbiamo stimato un tMRCA (time to Most Recent Common Ancestor) per la radice dell'albero che risale al 1940 (IC95% 1927-1951) per il sottotipo RSV A e al 1950 (IC95% 1951-1960) per RSV B, suggerendo così un'origine relativamente recente dei ceppi circolanti (il più vecchio isolato incluso nell'analisi è stata campionato nel 1956).

L'analisi delle dinamiche temporali dei genotipi appartenenti a RSV A hanno dimostrato quanto segue: a) il genotipo GA1 sembra essere il più antico, con un'origine risalente al 1950, seguito dal genotipo GA5, che ha avuto origine nel 1970; b) i genotipi GA2, GA3, GA4 e GA7 sono comparsi in un arco di tempo molto ristretto, tra il 1980 e il 1985; c) il genotipo di RSV A più recente, che è risultato anche il più diffuso tra i campioni in studio, è il genotipo NA1, con una stima del tMRCA risalente al 2003 (95%HPD: 2001-2005).

Le nostre stime indicano come, periodicamente, nuovi genotipi di RSV appaiano, tendendo a predominare rispetto ai ceppi circolanti precedentemente, pur non notandosi una netta sostituzione. Pertanto, come già descritto da altri, si potrebbe ipotizzare una sorta di co-circolazione, a livello globale, di vecchi e nuovi genotipi per un tempo relativamente lungo [130]. Ciò è dimostrato dal fatto che, andando a valutare i genotipi più antichi di RSV A, si è potuto constatare che questi circolavano ancora all'interno della popolazione anche a distanza di 30 anni dalla loro origine.

Tale dinamica appare discostarsi da quella mostrata, ad esempio, dal virus influenzale, dove le varianti virali prevalenti causano l'estinzione dei ceppi circolanti durante le epidemie precedenti.

Più difficile è stata la ricostruzione delle dinamiche evolutive di RSV B, probabilmente a causa del numero inferiore di sequenze disponibili.

Per questa infezione, il genotipo maggiormente circolante negli ultimi decenni è risultato il BA (Buenos Aires) che, dalle analisi effettuate, ha mostrato una stima della datazione del tMRCA risalente al 1997 (95%HPD: 1996-1999).

Il genotipo BA, un ceppo virale caratterizzato da una duplicazione di 60 nucleotidi nel gene G, è stato isolato per la prima volta nel 1999 durante una attività di sorveglianza di RSV in Argentina [131] e, da allora, è diventato il genotipo prevalente in tutto il mondo [132].

I risultati delle nostre analisi hanno mostrato come gli isolati italiani formino un unico clade relativo al sottogruppo BA4 [133], con un tMRCA risalente al 2003, anno corrispondente, con ogni probabilità, al periodo di inizio circolazione di tale ceppo in Lombardia.

L'osservazione degli skyline-plot ottenuti dall'analisi dinamica ha rilevato che, dopo un lungo periodo di dimensione costante prolungatosi fino al 1990 (probabilmente imputabile all'esiguo numero di sequenze ottenute prima di questa data), sia RSV A che RSV B hanno mostrato una diminuzione del numero effettivo di infezioni: per RSV-A tra il 1990 e il 2005, per RSV-B dal 1990 al 1999, anni che corrispondono alla nascita degli attuali genotipi prevalenti.

Rispetto alla sequenza di riferimento Aus/A2/61 (AN), l'analisi aminoacidica delle sequenze di RSV-A ha mostrato diverse mutazioni, in particolare a livello delle due regioni ipervariabili, come atteso e già riportato in diversi studi. Botosso et al. [134] hanno riferito la presenza di 4 sostituzioni aminoacidiche selezionate positivamente (P226L, S269T, P289S e P293S) che definiscono il genotipo GA2, anche se nel nostro studio tali sostituzioni sono state contemporaneamente identificate solo nella sequenza MI/A/16-11/10. Le altre mutazioni aminoacidiche, tipiche del genotipo NA1, sono state identificate da Aaimir et al. [135] in Pakistan e da Eshaghi et al. [136] in Canada.

Quando il genotipo BA è stato identificato per la prima volta (1999), la duplicazione di 60 nucleotidi che lo caratterizzava era una replica esatta dei 60 nucleotidi precedenti [137]. Questa regione ha accumulato ulteriori mutazioni nel tempo ed è inoltre evidente come alcuni nucleotidi siano sottoposti a pressione selettiva positiva.

Rispetto alla sequenza di riferimento, alcune mutazioni aminoacidiche, quali T270I, L277S e E278K sono stati identificate anche nelle sequenze BA isolate dai nostri campioni. Tali

mutazioni potrebbero migliorare la fitness virale, il che spiegherebbe il motivo per cui, ad oggi, stiano diventando dominanti surclassando gli altri genotipi di RSV B [97, 134, 138].

La proteina G è fortemente glicosilata, sia da glicidi N-linked che da glicidi O-linked ed è stato dimostrato che ciò contribuisce all'evasione del virus dalla risposta immunitaria umorale [139]. L'analisi dei potenziali siti di N-glicosilazione (Asn-X-Ser/Thr) ha mostrato la presenza di 6 potenziali residui in sequenze appartenenti a RSV-A, già descritti anche da [130]. In particolare, la sostituzione aminoacidica S250N ha portato all'acquisizione di un potenziale sito di N-glicosilazione nella sequenza MI/A/250-11/07. Inoltre, confrontando le sequenze appartenenti al genotipo NA1 la sequenza di riferimento Aus/A2/61, è stato osservato che la mutazione N237D in queste sequenze provoca una perdita di uno dei potenziali siti di N-glicosilazione.

Tra le sequenze di RSV B in studio, solo 4 siti di N-glicosilazione sono invece stati rilevati, tutti già descritti in altri studi. Tuttavia, la mutazione aminoacidica K201N a livello della sequenza MI/B/691-09/09 ha portato all'acquisizione di un potenziale sito di N-glicosilazione.

Molti siti di O-glicosilazione sono stati identificati in corrispondenza del frammento di proteina G analizzato; in particolare tra le sequenze di RSV-A sono stati identificati tra i 57 e i 64 possibili siti di O-glicosilazione e tra le sequenze di RSV-B tra i 59 e i 69. Sia il numero che la posizione dei siti di glicosilazione è altamente variabile tra i diversi ceppi [136]; questa variabilità permette al virus di evadere la risposta immunitaria, dal momento che la proteina di superficie non viene più riconosciuta e neutralizzata da parte degli anticorpi preesistenti.

Per indagare le forze motrici che portano alla variabilità di RSV A che di RSV B, abbiamo valutato la pressione selettiva che agisce a livello della proteina virale G. È interessante notare che, nonostante il ruolo di proteina d'attacco alla cellula e, quindi, bersaglio della risposta anticorpale neutralizzante, solo pochi aminoacidi hanno dimostrato di essere sotto pressione selettiva significativa, suggerendo un'azione prevalente della selezione neutrale, sottolineato anche dal rapporto dn/ds , sempre <1 . Complessivamente, a livello dell'allineamento delle sequenze di RSV A sono stati trovati 10 siti sottoposti a pressione selettiva positiva, di cui 4 (244, 250, 286 e 146) già descritti in precedenza da altri autori

[130]. L'analisi della dinamica di sostituzione in uno di questi siti (146), ha mostrato quale è il meccanismo con cui la variante prevalente viene progressivamente sostituita da diversi mutanti che tendono a fissarsi nella nuova popolazione, senza tuttavia provocare l'estinzione del ceppo originale.

Tra le sequenze di RSV B analizzate, solo un codone (217) è stato trovato essere sotto pressione selettiva positiva, confermato da tutti e tre i metodi di indagine utilizzati. Per altri 4 siti si è invece riusciti ad ottenere la conferma di pressione selettiva positiva mediante due dei tre metodi di indagine. Il risultato finale ha dimostrato come l'effetto della pressione selettiva negativa sia stata prevalente fra le sequenze della proteina G di RSV B analizzate.

L'approfondimento delle caratteristiche geniche del virus in studio ha permesso di valutare la circolazione dei diversi genotipi e studiarne le caratteristiche molecolari. Metodi filogenetici computazionali sono stati utilizzati per determinare alcune caratteristiche relative alla comparsa di RSV e alla sua diffusione nella popolazione umana.

Attraverso lo studio dell'evoluzione molecolare del virus è stato possibile definire sia il momento in cui i diversi genotipi virali hanno cominciato a differenziarsi (mediante il calcolo del tasso evolutivo), sia di valutare le caratteristiche dell'espansione spazio-temporale del virus. Inoltre, per valutare le forze di pressione selettiva che agiscono sull'evoluzione di RSV, è stata effettuata l'analisi specifica delle mutazioni sinonime e non-sinonime.

L'analisi filogenetica e quella filodinamica hanno permesso di comprendere meglio la storia naturale del virus, anche se questo resta difficile da confermare, data la grande diversità antigenica presentata. La variabilità della proteina G sembra svolgere un ruolo significativo nella patogenesi di RSV consentendo l'evasione alla risposta immunitaria; è importante quindi monitorare i cambiamenti che avvengono a livello delle sequenze codificanti questa proteina, sia per consentire l'identificazione dei diversi ceppi epidemici, sia per sviluppare future strategie vaccinali e terapeutiche.

In conclusione, questo progetto di tesi ha messo in evidenza come la virologia respiratoria, in continua evoluzione, necessiti di un costante aggiornamento delle procedure diagnostiche e di sorveglianza, indispensabile supporto nello studio e nel monitoraggio sia delle infezioni virali classiche e, soprattutto, di quelle causate da agenti emergenti o di recente identificazione.

Bibliografia

-
- [1] <http://www.who.int/>
- [2] Schwarz T, Careddu P, Chiumello G, et al. **Manuale di pediatria**, Casa editrice Ambrosiana, 1993.
- [3] Moroni M, D'Arminio Monforte A, Milazzo L. **Infezioni dell'apparato respiratorio** in Moroni M, Esposito R, De Lalla F. **MALATTIE INFETTIVE**. Masson s.p.a.- Milano, 2003; 355-380.
- [4] Armstrong D, Cohen J. **Infectious Diseases**, MOSBY - Hartcourt Publishers Ltd, 1999.
- [5] Hayden F.G. **Respiratory viral treats**. *Curr Opin Infect Dis*, 2006; 19(2):169-178.
- [6] Ison M.G. Hayden F.G. **Viral infections in immunocompromised patients: what's new with respiratory viruses?** *Curr Opin Infect Dis*, 2002; 15(4):355-67.
- [7] **Sistema di sorveglianza dell'influenza INFLU-CIRI**. Disponibile al sito internet: <http://www.influciri.it>.
- [8] Jennings L.C, Skopnik H, Burckhardt I et al. **Effect of rapid influenza testing on the clinical management of paediatric influenza**. *Influenza Other Respi Viruses*, 2009; 3(3):91-98.
- [9] La Placa M. **Micobatteri** in La Placa M, **Principi di Microbiologia Medica**, decima edizione. *Società Editrice Esculapio, Bologna*, 2005; 257-268.
- [10] La Placa M. **Microbiologia Clinica** in La Placa M., **Principi di Microbiologia Medica**, decima edizione. *Società Editrice Esculapio, Bologna*, 2005;751-806.
- [11] B.R. Murphy, R.G. Webster, **Orthomyxoviruses** in: **VIROLOGY**, B. N. Fields - Raven Press – New York, 1990; 40: 1091-1137.
- [12] N.J. Cox, Y. Kawaoka. **Orthomyxoviruses: influenza** in: **Microbiology and microbial infections (Virology)**, L. Collier, A. Balows, M. Sussmann – Arnold Edt; 22: 386-418.
- [13] Steinhauer DA, Holland JJ. **Rapid evolution of RNA viruses**. *Annu. Rev. Microbiol.* 1987; 41: 409-433.
- [14] Dennis Wat, **The common cold: a review of the literature**. *European Journal of Int. Medicine* 2004; 15: 79-88.
- [15] Nicholson KG, Wood JM, Zambon M. **Influenza**. *Lancet* 2003; 362: 1733-1745.
- [16] Shortridge KF, Peiris JS, Guan Y. **The next influenza pandemic: lessons from Hong Kong**. *Journal of Applied Microbiology*. 2003;94 Suppl: 70S-79S.
- [17] Madhu Khanna, Neha Gupta et al. **Influenza A (H1N1) 2009: a pandemic alarm**. *J.Biosci*, 2009; 34:481-489.
- [18] Bellini WJ, Rota PA, Anderson LJ, **Paramyxoviruses** in: **Microbiology and microbial infections (Virology)**, L. Collier, A. Balows, M. Sussmann – Arnold Edt; 23: 435-452.
- [19] Chanock R.M, McIntosh K, **Parainfluenza Viruses** in: **VIROLOGY**, B. N. Fields - Raven Press – New York, 1990; 35: 963-984.
- [20] McIntosh K, Chanock R. M. **Respiratory Syncytial Virus** in: **VIROLOGY**, B. N. Fields - Raven Press – New York, 1990; 38: 1045-1066.
- [21] Mufson MA, Orwell C. Rafner B. **Two distinct subtypes oh human respiratory syncytial virus**. *J. Gen. Virol.* 1985; 66: 2111-2114.
- [22] Hacking D, Hull J. **Respiratory Syncytial Virus – Viral biology and the host response**. *Jour of infect.* 2002; 45: 18-24.
- [23] M. S Smyth, J. H. Martin. **Picornavirus uncoating**. *J Clin Pathol: Mol Pathol* 2002; 55: 214–219
- [24] Lauber C, Gorbalenya AE. **Toward genetics-based virus taxonomy: comparative analysis of a genetics-based classification and the taxonomy of picornaviruses**. *J Virol.* 2012;86(7):3905-3915.
-

- [25] Oberste M. S, Maher K. **Enterovirus 68 is associated with respiratory illness and shares biological features with both the enteroviruses and the rhinoviruses.** Journal of General Virology (2004), 85: 2577–2584.
- [26] Lau SK, Yip CC, Woo PC, Yuen KY. **Human rhinovirus C: a newly discovered human rhinovirus species.** Emerg Health Threats J. 2010; 3:e2.
- [27] Minor P, **Picornaviruses** in: Microbiology and microbial infections (Virology). L Collier, A Balows, M Sussmann – Arnold Edt – London 1998. Pag: 497-502.
- [28] Blomqvist S, Roivainen M, Puhakka T et al. **Virological and serological analysis of rhinovirus infections during the first two years of life in a cohort of children.** J. Med. Virol.2002; 66(2): 263-268.
- [29] Hendley JO. **Otitis media.** N. Eng J. Med. 2002; 347: 1169-1174
- [30] McIntosh K, **Coronaviruses** in: **VIROLOGY**, B. N. Fields - Raven Press – New York, 1990; 30: 857-863.
- [31] Baker S. C, **Coronaviruses: form common cold to Severe Acute Respiratory Syndrome.** Pediat. Infec. Dis. Journ. 2004; Vol.23 No.11: 1049-1050.
- [32] Denison M.R. **Severe Acute Respiratory Syndrome Coronavirus Pathogenesis, Disease and Vaccines An Update.** *Pediatr Infect Dis J*, 2004, 23(11): S207-14.
- [33] Haagmans BL, Al Dhahiry SH, Reusken CB, Raj VS, Galiano M, Myers R, Godeke GJ, Jonges M, Farag E, Diab A, Ghobashy H, Alhajri F, Al-Thani M, Al-Marri SA, Al Romaihi HE, Al Khal A, Bermingham A, Osterhaus AD, Alhajri MM, Koopmans MP. **Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation.** Lancet Infect Dis. 2013 Dec 16. pii: S1473-3099(13)70690-X.
- [34] Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team. **Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans.** *N Engl J Med*, 2009, 360(25): 2605-2615
- [35] Van den Hoogen B.G. de Jong J.C. Groen J. et al. **A newly discovered human pneumovirus isolated from young children with respiratory tract disease.** *Nat Med*, 2001, 7: 719-724.
- [36] **The universal virus Database - ICTVdb.** Disponibile al sito internet: www.ncbi.nlm.nih.gov/ictvdb.
- [37] Broor S. Bharaj P. Chahar H.S. **Human metapneumovirus: a new respiratory pathogen.** *J Biosci*, 2008, 33: 483-493.
- [38] Bossart K.N. Broder C.C. **Paramyxovirus Entry** in Pohlmann S. Simmons G. Viral Entry into Host Cells. Landes Bioscience, 2007: 1-20
- [39] La Placa M. **La moltiplicazione dei virus** in La Placa M., Principi di Microbiologia Medica, decima edizione. Società Editrice Esculapio, Bologna, 2005: 524-552.
- [40] Costa LF, Yokosawa J, Maltese OC et al, **Respiratory viruses in children younger than five years old with acute respiratory disease from 2001 to 2004 in Uberlandia, MG, Brazil.** Memorias do instituto Oswaldo Cruz 2006; 101(3): 301-306.
- [41] Madhi SA, Ludewick H, Abed Y et al. **Human Metapneumovirus-associated lower respiratory tract infections among hospitalized human immunodeficiency virus type 1 HIV-1-infected and HIV-1-uninfected African infants.** Clin. Infec. Dis. 2003; 37: 1705-1710.

- [42] Van Den Hoogen, Bernadette G, Osterhaus D, Fouchier RA. **Clinical impact and diagnosis of Human Metapneumovirus infection**. *Pediatr. Infect. Dis. Journ.* 2004; 23: S25-S32.
- [43] Williams John V, Harris Paul A, Tollefson Sharon J. et al. **Human Metapneumovirus and Lower Respiratory Tract disease in otherwise healthy infants and children**. *The New England Jour. Med.* 2004; 350: 443-450.
- [44] Schildgen O, Müller A, Allander T, et al. **Human Bocavirus: Passenge or Pathogen in Acute Respiratory Tract Infections?** *Clin Microbiol Rev*, 2008,21(2): 291-304.
- [45] Lindner J, Karalar L, Schimanski S, et al. **Clinical and epidemiological aspects of human bocavirus infection**. *J Cli Virol*, 2008, 43(4): 391-395.
- [46] Berns K.I. **Parvovirus replication**. *Microbiol Rev*, 1990, 54(3): 316-29.
- [47] Tijssen P. **Molecular and structural basis of the evolution of parvovirus tropism**. *Acta Vet Hung*, 1999, 47(3): 379-394.
- [48] Qu X.W, Liu W.P, Qi Z.Y, et al. **Phospholipase A2-like activity of human bocavirus VP1 unique region**. *Biochem Biophys Res Commun*, 2008, 365:158-163.
- [49] Brieu N, Gay B, Segondy M, et al. **Electron microscopy observation of human bocavirus (HBoV) in nasopharyngeal samples from HBoV-infected children**. *J Clin Microbiol*, 2007, 45(10):3419-3420.
- [50] Bastien N, Brandt K, Dust K, et al. **Human Bocavirus infection, Canada**. *Emerg Infect Dis*, 2006, 12(5):848-850.
- [51] Allander T, Jartti T, Gupta S, et al. **Human Bocavirus and acute wheezing in children**. *Clin Infect Dis*, 2007,44:904-910.
- [52] Lee J.I, Chung J.Y, Han T.H, et al. **Detection of Human Bocavirus in Children Hospitalized because of Acute Gastroenteritis**. *J Infect Dis*, 2007, 196: 994-997.
- [53] Vicente D, Cilla G, Montes M, et al. **Human Bocavirus, a Respiratory and Enteric Virus**. *Emerg Infect Dis*, 2007, 13(4): 636-637
- [54] Chung J.Y, Han T.H, Kim C.K, et al. **Bocavirus infection in hospitalized children, South Korea**. *Emerg Infect Dis*, 2006, 12:1254-1256.
- [55] Maggi F, Andreoli E, Pifferi M, et al. **Human bocavirus in Italian patients with respiratory diseases**. *J Clin Virol*, 2007, 38(4):321-325.
- [56] Weissbrich B, Neske F, Schubert J, et al. **Frequent detection of bocavirus DNA in German children with respiratory tract infections**. *BMC Infect Dis*, 2006, 6:109.
- [57] Allander T, Andreasson K et.al. **Identification of a Third Human Polyomavirus**. *Journal of Virology*, 2007, 4130-4136.
- [58] Gardner, S.D, A.M. Field, Et al. **New human papovirus (B.K) isolated from urine after renal transplantation**. *Lancet*, 1971, 1253-1257.
- [59] Padgett B.L, Walker D.L., el al. **Cultivation of papova-like virus from human brain with progressive multifocalleucoencephalopathy**. *Lancet*, 1971, 1257-1260
- [60] Wattier R.L, Vázquez M, et al. **Role of Human Polyomaviruses in Respiratory Tract Disease in Young**. *Emerging Infectious Diseases*, No. 11,2008, Vol.14: 1766-1768.
- [61] Gaynor A.M, Nissene M.D, et al. **Identification of a novel Polyomavirus from patients with Acute Respiratory Tract Infections**. *PloS*, 2007, Vol.3: 595-604
- [62] van de Pol A. C, F.W. Wolfs T, et al. **Human Bocavirus and KI/WU Polyomavirusse in Pediatric Intensive Patients**. *Emerging Infectious Disease*. Vol.15, No.3, 2009.
- [63] Marieke M. van der Zalm, W.A. Rossen J, e al. **Prevalence and Pathogenicity of WU and KI Polyomaviruses in Children, the Netherlands**. *Emerging Infectious Disease*. Vol.14, 2008, No.11: 1787-1789.

-
- [64] Carlos F. Arias, Marina Escalera-Zamudio et al. ***Molecular Anatomy of 2009 Influenza Virus A (H1N1)***. Elsevier, September 2009, 643-654
- [65] www.iss.it/iflu/
- [66] Henrickson K.J. ***Advances in the laboratory diagnosis of viral respiratory disease***. *Pediatr Infect Dis J*, 2007, 23(10): 6-10.
- [67] Morens D.M, Folkers G.K, Fauci A.S. ***The challenge of emerging and re-emerging infectious diseases***. *Nature* 2004, 430, 242-249.
- [68] de Vries M, Deijs M, Canuti M, van Schaik BDC, Faria NR, van de Gardeet MDB, et al. ***A sensitive assay for virus discovery in respiratory clinical samples***. *PLoS One* 2011; 6: e16118.
- [69] Pyrc K, Jebbink M.F, Berkhout B, van der Hoek L. ***Detection of new viruses by VIDISCA. Virus discovery based on cDNA-amplified fragment length polymorphism***. *Methods Mol. Biol.* 2008, 454, 73-89.
- [70] van der Hoek L, Pyrc K, Jebbink M.F, Vermeulen-Oost, W, Berkhout R.J, Wolthers K.C, Wertheim-van Dillen P.M, Kaandorp, J, Spaargaren, J, Berkhout B. ***Identification of a new human coronavirus***. *Nat. Med.* 2004, 10, 368-373.
- [71] MacConaill L, Meyerson M. ***Adding pathogens by genomic subtraction***. *Nat. Genet.* 2008, 40, 380-382.
- [72] Tang P, Chiu C. ***Metagenomics for the discovery of novel human viruses***. *Future Microbiol.* 2010, 5, 177-189.
- [73] Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan P.L, Hui J, Marshall J, et al. ***A new arenavirus in a cluster of fatal transplant-associated diseases***. *N. Engl. J. Med.* 2008, 358, 991-998.
- [74] Rector A, Tachezy R, Van Ranst M. ***A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification***. *J. Virol.* 2004, 78, 4993-4998.
- [75] Schowalter R.M, Pastrana D.V, Pumphrey K.A, Moyer A.L, Buck C.B. ***Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin***. *Cell Host Microbe* 2010, 7, 509-515.
- [76] van der Meijden E, Janssens R.W, Lauber C, Bouwes Bavinck J.N, Gorbalenya A.E, Feltkamp M.C. ***Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient***. *PLoS Pathog.* 2010, 6 (7): e1001024.
- [77] Yozwiak N.L, Skewes-Cox P, Gordon A, Saborio S, Kuan, G, Balmaseda A, Ganem D, Harris E, DeRisi J.L. ***Human enterovirus 109: A novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua***. *J. Virol.* 2010, 84, 9047-9058.
- [78] Quan, P.L, Firth C, Street C, Henriquez J.A, Petrosov A, Tashmukhamedova A, Hutchison S.K, Egholm M, Osinubi M.O, Niezgodna M, et al. ***Identification of a severe acute respiratory syndrome coronavirus-like virus in a leaf-nosed bat in Nigeria***. *mBio* 2010, 1, e00208-e00210.
- [79] Kimura M. ***The Neutral Theory of Molecular Evolution***. 1983, Cambridge University Press; Cambridge.
- [80] Ohta T. ***The Nearly Neutral Theory of Molecular Evolution***. *Ann. Rev. Ecol. Syst.* 1992, 23: 263-286.
- [81] Valle G., Helmer-Citteriche M., Attimonelli M, Pesole G. ***Introduzione alla bioinformatica***. 2003 Zanichelli Editore, Bologna.
-

- [82] Salemi M., Vandamme A. M. ***The Phylogenetic Handbook. A practical approach to DNA and protein phylogeny.*** 2003 Cambridge University Press, New York, USA.
- [83] Ponti W. ***Genetica ed evoluzione dei virus.*** In: Poli G., Cocilovo A., Dall'Ara P., Martino P. A., Ponti W. ***"Microbiologia e Immunologia Veterinaria"***. 2 ed. 2005, UTET, Torino, pag. 393-397.
- [84] Saitou N., Nei M. ***The Neighbor-Joining method: A new method for reconstructing phylogenetic tree.*** Mol. Biol. Evol. 1987; 4: 406-425.
- [85] Sneath P. H. A., Sokal R. R. ***Numerical Taxonomy.*** Freeman, 1973, San Francisco.
- [86] Eck R. V., Dayhoff M. O. ***Atlas of Protein Sequence and Structure.*** 1966 National Biomedical Research Foundation, Silver Springs, Maryland.
- [87] Felsenstein J. ***Evolutionary trees from DNA sequences: A maximum-likelihood approach.*** J. Mol. Evol. 1981; 17: 368-376.
- [88] Liò P., Goldman N. ***Models of Molecular Evolution and Phylogeny.*** Genome 1998; 8: 1233-1244.
- [89] DeSalle R. ***What's in a character?*** J. Biomed. Info. 2006; 39: 6-17.
- [90] Sarkar I. N. ***Phylogenetics in the modern era.*** J. Biomed. Info. 2006; 39: 3-5.
- [91] Zuckerkandl E., Pauling L. ***Molecules as documents of evolutionary history.*** J. Theor. Biol. 1965; 8: 357-366.
- [92] Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al. ***The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination.*** Am J Respir Crit Care Med 1994;149: 818-824.
- [93] Johnson F.B., Leavitt R.W., Richards D.F., ***Evaluation of Virocult transport tube for isolation of Herpes simplex virus from clinical specimens.*** J.Clin Microbiol 1984: 20: 120-122.
- [94] World Health Organization (WHO) Global Influenza Surveillance Network (GISN). Manual for the laboratory diagnosis and virological surveillance of influenza. 2011. Available from: http://whqlibdoc.who.int/publications/2011/9789241548090_eng.pdf
- [95] Center for Disease Control (CDC). 2009 protocol CDC protocol of real-time RT PCR for swine influenza A(H1N1), 30 April 2009. Available from: http://www.who.int/csr/resources/publications/swineflu/CDCrealtimerTPCRprotocol_20090428.pdf
- [96] Zlateva KT, Lemey P, Vandamme AM, Ranst MV. ***Molecular Evolution and Circulation Patterns of Human Respiratory Syncytial Virus Subgroup A: Positively Selected Sites in the Attachment G glycoprotein.*** Journal of Virology 2004; 78: 4675-4683.
- [97] Zlateva KT, Lemey P, Moës E, Vandamme AM, Ranst MR. ***Genetic Variability and Molecular Evolution of the Human Respiratory Syncytial Virus Subgroup B Attachment G Protein,*** Journal of Virology 2005; 79: 9157-9167
- [98] Benson D.A. Karsch-Mizrachi I. Lipman D.J. et al. ***GenBank.*** Nucleic Acids Res, 36:D25-D30, 2008.
Disponibile al sito internet: <http://www.ncbi.nlm.nih.gov/Genbank/>.
- [99] Hall T.A. ***Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows95/98/NT.*** Nucl Acids Symp, 1999, 41:95-98.
Disponibile al sito internet: <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>.

- [100] Thompson J.D. Gibson T.J. Plewniak F. et al. **The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res*, 1997, 25(24): 4876-4882. Disponibile al sito internet: <ftp://ftp.ebi.ac.uk/pub/software/clustalx>.
- [101] Kumar S. Tamura K. Jakobsen I.B. et al. **Molecular Evolutionary Genetics Analysis Software.** *Bioinformatics*, 2001, 17(12): 1244-1245. Disponibile al sito internet: <http://www.megasoftware.net/>.
- [102] Saitou N. and Nei M. **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol*, 1987, 4(4):406-425.
- [103] Felsenstein J. **Confidence limits on phylogenies: An approach using the bootstrap.** *Evolution*, 1985, 39:783-791.
- [104] Posada D. e Crandall K.A. **Modeltest: testing the model of DNA substitution.** *Bioinformatics*, 1998, 14(9):817-818.
Disponibile al sito internet: <http://darwin.uvigo.es/software/modeltest.html>.
- [105] Swofford D. L. **PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1** computer program distributed by the. *Illinois Natural History Survey*, Champaign, Illinois, 1989. Disponibile al sito internet: <http://paup.csit.fsu.edu/>.
- [106] Hastings W.K. **Monte Carlo sampling methods using Markov chains and their applications.** *Biometrika*, 1970, 57(1):97-109.
- [107] Drummond A.J. Rambaut A. **Bayesian evolutionary analysis by sampling trees.** *BMC Evolutionary Biology*, 2007, 7:214.
- [108] Metropolis N. Rosenbluth A.W. Rosenbluth M.N. et al. **Equation of State Calculations by Fast Computing Machines.** *J Chem Phys*, 1953, 21(6):1087-1092.
- [109] Ronquist F. van der Mark P. Huelsenbeck J.P. **Bayesian phylogenetic analysis using MrBayes. Theory** in Lemey P. Salemi M. Vandamme A.M., *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*, seconda edizione. *Cambridge University Press*, 2009, 210-236.
- [110] Rambaut A. Drummond A.J. **Tracer v1.4**, 2007. Disponibile al sito internet: <http://beast.bio.ed.ac.uk/Tracer>.
- [111] Kass R.E. e Raftery A.E. **Bayes Factors.** *J Am Stat Assoc*, 1995, 90(430):773-795.
- [112] Kosakovsky Pond S.L. Poon A.F.Y. e Frost S.D.W. **Estimating selection pressure on alignments of coding sequences** in Lemey P. Salemi M. Vandamme A.M., *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing, seconda edizione.* *Cambridge University Press*, 2009: 419-90.
- [113] Nei M. e Gojobori T. **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.** *Mol Biol Evol*, 1986, 3:418-26.
- [114] Jukes T.H. e Cantor C.R. **Evolution of protein molecules.** *Mammalian Protein Metabolism*, *Academic Press*, New York, 1969, 21-132.
- [115] Kosakovsky Pond S.L. e Frost S.D.W. **Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection.** *Mol Biol Evol*, 2005, 22(5):1208-1222.
- [116] Kosakovsky Pond S.L. Frost S.D.W. e Muse S.V. **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics*, 2005, 21(5):676-679.
- [117] Kosakovsky Pond S.L. Muse S.V. **Column sorting: rapid calculation of the phylogenetic likelihood function.** *Syst Biol*, 2004, 53(5):685-692.
- [118] Gupta R, Jung E and Brunak S. **Prediction of N-glycosylation sites in human proteins.**

- [119] Julenius K, Molgaard A, Gupta R, Brunak S. **Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites.** *Glycobiology*. 2005, 15(2):153-164.
- [120] Dean AG, Sullivan KM, Soe MM. **OpenEpi: Open Source Epidemiologic Statistics for Public Health**, Versione 3.01. www.Openepi.com, aggiornata 2013/04/06, accesso 2014/02/05.
- [121] Wiemken T, Peyrani P, Bryant K, Kelley RR, Summersgill J, Arnold F, et al. **Incidence of respiratory viruses in patients with community-acquired pneumonia admitted to the intensive care unit: results from the Severe Influenza Pneumonia Surveillance (SIPS) project.** *Eur J Clin Microbiol Infect Dis* 2013, 32:705-710.
- [122] Choi SH, Hong SB, Ko GB, Lee Y, Park HJ, Park SY, et al. **Viral infection in patients with severe pneumonia requiring intensive care unit admission.** *Am J Respir Crit Care Med* 2012, 186:325-332.
- [123] Do AH, van Doorn HR, Nghiem MN, Bryant JE, Hoang TH, Do QH, et al. **Viral etiologies of acute respiratory infections among hospitalized Vietnamese children in Ho Chi Minh City, 2004-2008.** *PLoS One* 2011, 6:e18176.
- [124] Ohno A, Suzuki A, Lupisan S, Galang H, Sombrero L, Aniceto R, Okamoto M, Saito M, Fuji N, Otomaru H, Roy C.N, Yamamoto D, Tamaki R, Olveda R, Oshitani H, **Genetic characterization of human respiratory syncytial virus detected in hospitalized children in the Philippines from 2008 to 2012.** *J Clin Virol*. 2013, 57(1):59-65.
- [125] Khor C.S, Sam I.C, Hooi P.S, Chan Y.F, **Displacement of predominant respiratory syncytial virus genotypes in Malaysia between 1989 and 2011.** *Infect Genet Evol.*, 2013, 14: 357-360
- [126] Cui G, Zhu R, Qian Y, Deng J, Zhao L, Sun Y, Wang F. **Genetic variation in attachment glycoprotein genes of human respiratory syncytial virus subgroups A and B in children in recent five consecutive years.** *PLoS One*. 2013, 17;8(9): e75020.
- [127] Adams O, Werzmirzowsky J, Hengel H. **Genetic analysis and antigenic characterization of human respiratory syncytial virus group A viruses isolated in Germany 1996-2008.** *Virus Genes*. 2013, 47(2): 210-218.
- [128] Forcic D, Ivancic-Jelecki J, Mlinaric-Galinovic G, Vojnovic G, Babic-Erceg A, Tabain I. **A study of the genetic variability of human respiratory syncytial virus in Croatia, 2006-2008.** *J Med Virol*. 2012, 84(12): 1985-92.
- [129] Montieri S, Puzelli S, Ciccozzi M, Calzoletti L, Di Martino A, Milia MG, Rossi A, Piro F, Rezza G, Donatelli I. **Amino acid changes in the attachment G glycoprotein of human respiratory syncytial viruses (subgroup A) isolated in Italy over several epidemics (1997-2006).** *J Med Virol*. 2007, 79(12): 1935-1942.
- [130] Tan L, Coenjaerts FE, Houspie L, Viveen MC, van Bleek GM, Wiertz EJ, Martin DP, Lemey P. **The comparative genomics of human respiratory syncytial virus subgroups A and B: genetic variability and molecular evolutionary dynamics.** *J Virol*. 2013, 87(14):8213-8226.
- [131] Trento A, Galiano M, Videla C, Carballal G, García-Barreno B, Melero J.A, Palomo C. **Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides.** *J Gen Virol*. 2003, 84: 3115-3120.

-
- [132] Van Niekerk S, Venter M. **Replacement of previously circulating respiratory syncytial virus subtype B strains with the BA genotype in South Africa.** *J Virol.* 2011, 85: 8789-8797.
- [133] Dapat I.C, Shobugawa Y, Sano Y, Saito R, Sasaki A, Suzuki Y, Kumaki A, Zaraket H, Dapat C, Oguma T, Yamaguchi M, Suzuki H. **New genotypes within respiratory syncytial virus group B genotype BA in Niigata, Japan.** *J Clin Microbiol.* 2010, 48: 3423-3427.
- [134] Botosso VF, Zanotto PM, Ueda M, Arruda E, Gilio AE, Vieira SE, Stewien KE, Peret TC, Jamal LF, Pardini MI, Pinho JR, Massad E, Sant'anna OA, Holmes EC, Durigon EL; VGDN Consortium. **Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus.** *PLoS Pathog.* 2009, 5(1):e1000254.
- [135] Aamir UB, Alam MM, Sadia H, Zaidi SS, Kazi BM. **Molecular Characterization of Circulating Respiratory Syncytial Virus (RSV) Genotypes in Gilgit Baltistan Province of Pakistan during 2011-2012 Winter Season.** *PLoS One.* 2013, 8(9):e74018.
- [136] Eshaghi A, Duvvuri V.R, Lai R, Nadarajah J.T, Li A, Patel S.N, Low D.E, Gubbay J.B, **Genetic variability of human respiratory syncytial virus A strains circulating in Ontario: a novel genotype with a 72 nucleotide G gene duplication.** *PLoS One.* 2012, 7:32807.
- [137] Trento A, Viegas M, Galiano M, Videla C, Carballal G, Mistchenko A.S, Melero J.A, **Natural history of human respiratory syncytial virus inferred from phylogenetic analysis of the attachment (G) glycoprotein with a 60-nucleotide duplication.** *J Virol.*, 2006, 80: 975-984.
- [138] Drummond A.J, Rambaut A, Shapiro B, Pybus O.G. **Bayesian coalescent inference of past population dynamics from molecular sequences.** *Mol Biol Evol.* 2005, 22(5):1185-1192.
- [139] Cane P.A, **Molecular epidemiology of respiratory syncytial virus.** *Rev Med Virol.*, 2001, 11: 103-116.

Riassunto

Le infezioni respiratorie acute (*Acute Respiratory Infections, ARI*) sono infezioni ubiquitarie, caratterizzate da un andamento tipicamente epidemico determinato dalla loro elevata contagiosità e diffusione che avviene per via aerea diretta, indiretta e tramite contatto. I virus sono tra i principali patogeni responsabili di ARI. Negli ultimi anni lo scenario epidemiologico delle infezioni respiratorie, oltre a essere stato movimentato dalla continua circolazione di ceppi virali già noti (*i.e.* virus influenzali stagionali, parainfluenzali, virus respiratorio sinciziale, *etc.*), è stato vivacizzato dall'identificazione, nonché dall'emergenza, di numerosi patogeni a trasmissione aerea. L'epidemia di SARS, il ripetersi di episodi sporadici di trasmissione di virus dell'influenza aviaria (H5N1, H7N7, H7N9 *etc.*) direttamente dai volatili all'uomo, il nuovo CoV-MERS evidenziano l'importanza del fenomeno costituito dalle infezioni emergenti.

Scopo

Lo scopo del presente lavoro di tesi è stato quello di utilizzare nuove metodiche di biologia molecolare, "virus discovery" e bioinformatica al fine investigare e approfondire l'eziologia delle infezioni respiratorie acute in soggetti con sintomatologia respiratoria acuta grave.

In particolare gli obiettivi specifici di questo lavoro sono stati:

1. Identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo influenzale pandemico e post-pandemico (2009-2011)

A tal fine sono stati analizzati campioni prelevati da soggetti con sindromi respiratorie acute gravi, andando a valutare la quota di infezioni respiratorie gravi associate, all'infezione da virus influenzale pandemico A(H1N1)pdm09, il ruolo di altri virus respiratori conosciuti e l'eventuale presenza di virus non ancora noti, nei campioni risultati negativi alle precedenti indagini utilizzando una tecnica di "virus discovery".

2. Valutazione della variabilità genetica ed evoluzione molecolare del virus respiratorio sinciziale (RSV)

Al fine di caratterizzare i ceppi RSV circolanti in Lombardia nel periodo compreso tra il 2006 e il 2012 è stato condotto uno studio filogenetico e filodinamico degli strain di RSV identificati utilizzando metodi computazionali per determinare alcune caratteristiche relative sia alle capacità mutageniche del virus, sia alla comparsa e alla diffusione di RSV nella popolazione umana.

Lo studio è stato condotto presso la Sezione di Virologia del Dipartimento di Scienze biomediche per la Salute e la parte relativa all'impiego di una tecnica di "virus discovery" è stata effettuata presso l'Università di Amsterdam nel "Laboratory of Experimental Virology, Academic Medical Center (AMC)", dove ho svolto un periodo di tirocinio formativo sotto la guida della Prof.ssa Lia van der Hoek.

Materiali e Metodi

Per l'identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo in studio, sono stati raccolti 206 tamponi oro-faringei (TOF) prelevati da pazienti ospedalizzati per forme respiratorie gravi. La diagnosi di infezione virale è stata condotta mediante indagini molecolari (real-time RT-PCR) allo scopo di identificare l'eventuale presenza del virus influenzale pandemico A(H1N1)pdm09 e la presenza dei seguenti agenti virali: Virus influenzale A/H3N2 e B, Virus Respiratorio Sinciziale (RSV) A e B; human Metapneumovirus (hMPV) A e B; human Rhinovirus (hRV) e Enterovirus (hEV); Adenovirus (AdV); human Bocavirus (hBoV) 1-4; human Coronavirus (hCoV) 229E, NL63, OC43, HKU1; human parainfluenza virus (hPIV) 1-4; *Chlamydomphila pneumoniae*; *Mycoplasma pneumoniae*. I campioni risultati negativi a tutte le indagini precedenti, sono

stati analizzati mediante la tecnica di “virus discovery” VIDISCA-454, allo scopo di individuare agenti virali non noti nei campioni in esame.

Per valutare la variabilità genetica e l'evoluzione molecolare di RSV è stato analizzato un campione rappresentativo di sequenze ottenute da prelievi raccolti da soggetti pediatrici con età ≤ 3 anni, ospedalizzati tra il 2006 e il 2012 e risultati positivi alla diagnosi di infezione virale da RSV. I campioni sono stati analizzati mediante l'utilizzo di una multiplex-PCR per l'identificazione di un frammento del gene F di RSV e diversi programmi bioinformatici sono stati utilizzati per l'analisi delle sequenze ottenute.

Risultati

1. Identificazione di agenti patogeni virali responsabili di infezioni respiratorie gravi e complicate durante il periodo influenzale pandemico e post-pandemico (2009-2011).

Dalle indagini molecolari è emerso che il virus influenzale A(H1N1)pdm09 era presente nel 58,3% (120/206; IC95%: 51,4-64,9%) dei campioni raccolti dai casi SARI/ARDS in studio. I soggetti con questo tipo di infezione erano per il 61,7% maschi con età mediana di 42,9 anni (IQR: 40,9 anni). Dal punto di vista clinico, si è rilevato che: il 51,7% (62/120; IC95%: 42,7-60,5%) dei soggetti A(H1N1)pdm09-positivi aveva una diagnosi di ARDS e il 48,3% di SARI. L'impatto del virus pandemico A(H1N1)pdm09 nei casi ARDS/SARI risultati fatali, calcolato come Case Fatality Rate (CFR), è risultato pari all'11,3% (7/62, IC95%: 5,1-21,1%). Dalle successive analisi sui campioni con esito negativo alla ricerca del virus A(H1N1)pdm09 è emersa positività ad almeno uno dei patogeni ricercati per il 46,5% (40/86; IC95%: 36,2-57,1%) dei campioni. HRV/hEV sono stati i patogeni più frequentemente presenti, identificati nel 27,5% (11/40; IC95%: 15,4-42,8%) dei campioni analizzati, seguiti dal virus influenzale A(H3N2) identificato nel 20% (8/40; IC95%: 9,7-34,5%) dei casi. Quarantasei (46/206; 22,3%, IC95%: 17-28,4%) casi di SARI/ARDS (compresi due casi fatali) risultati negativi a tutti i saggi diagnostici descritti in precedenza sono stati sottoposti ad indagine mediante VIDISCA-454. Nessuna delle sequenze ottenute tramite questa metodica è risultata attribuibile ad un nuovo virus o ad una nuova variante virale; tuttavia l'analisi ha permesso l'identificazione di un caso di morbillo non diagnosticato in precedenza, portando così la quota dei casi di SARI/ARDS con diagnosi virale nota al 78,2% (161/206, IC95%: 72,1-83,4%). La quota dei casi rimasta a diagnosi ignota è risultata quindi del 21,8% (45/206, IC95%: 16,6-27,9%); per la maggioranza si trattava di casi di ARDS (34/45: 75,6%, 61,5-86,4%), 2 dei quali risultati fatali.

2. Valutazione della variabilità genetica ed evoluzione molecolare del virus respiratorio sinciziale (RSV)

Dall'analisi delle sequenze di RSV A è emerso che: 3 delle sequenze in studio si sono inserite nel cluster GA2, mentre tutti gli altri isolati sono rientrati nel gruppo NA1. Le sequenze incluse nel genotipo GA2 hanno evidenziato una distanza genetica media significativamente superiore a quella di NA1 (mean p distance = $0,032 \pm 0,003$ sub/site) e simile alla distanza media tra i due genotipi (mean p distance = $0,044 \pm 0,006$ sub/site) e comparando le sequenze in studio alla sequenza di riferimento, sono state identificate 31 sostituzioni aminoacidiche. L'analisi filogenetica del gene G di RSV B ha mostrato che tutte le sequenze in studio appartenevano al genotipo BA e, in particolare clusterizzavano all'interno di un sottogruppo denominato BA4. Questo ha presentato una distanza genetica media (mean p distance = $0,025 \pm 0,003$ sub/site) significativamente più alta rispetto agli altri ceppi BA (mean p distance = $0,011 \pm 0,002$ sub/site). Rispetto alla sequenza di riferimento, 8 sostituzioni aminoacidiche sono state identificate fra tutte le sequenze in studio. Dall'analisi dell'evolutionary rate sono stati calcolati i tassi medi evolutivi che sono stati

stimati pari a: $2,1 \times 10^{-3}$ subs/site/year (95% HPD: $1,7-2,5 \times 10^{-3}$) per RSV A e $3,03 \times 10^{-3}$ subs/site/year (95% HPD: $2,1-3,9 \times 10^{-3}$) per RSV B. Il valore di tMRCA per la radice albero di RSV A è risultato risalire a 71 anni fa (95% HPD = 60-85) e i ceppi isolati presso il nostro laboratorio appartenenti al clade NA1 condividevano un unico nodo significativo interno con un tMRCA medio stimato a 7 anni fa (2005). Invece i tre isolati del genotipo GA2 hanno mostrato un tMRCA significativo che risale a 16 anni fa (1996). La radice dell'albero ha avuto un tMRCA medio stimato di 55 anni fa (1957). Tutti gli isolati in studio sono stati inclusi all'interno di un subclade significativo, il BA4, avente una stima tMRCA di 9 anni fa (2003). Il Bayesian skyline plots (BSP) ottenuto dall'analisi delle sequenze in studio di RSV A ha evidenziato che dal 1940, una primo picco di infezioni è avvenuto nella seconda metà degli anni 1980, seguito da una diminuzione di eventi di trasmissione terminato nel 2005, quando una rapida crescita ha ripristinato la dimensione della popolazione virale originale e il BSP RSV B ha mostrato un andamento simile. L'analisi di pressione selettiva delle sequenze di RSV A sito specifica ha identificato 10 codoni sottoposti a selezione positiva e un totale di 39 siti sono stati trovati essere sotto pressione selettiva negativa. Invece per RSV B ha identificato un unico sito selezionato positivamente e un totale di 27 siti sottoposti a pressione selettiva negativa. Diversi siti di O- e N-glicosilazione sono stati identificati analizzando le sequenze di RSV A e RSV B in studio.

Conclusioni

Durante il periodo pandemico e post-pandemico diversi agenti patogeni hanno co-circolato e sono risultati associati ai casi clinici in studio; tuttavia il virus influenzale A(H1N1)pdm09 ha mostrato il maggiore impatto risultando presente in oltre il 58% dei casi. La proporzione di casi di SARI/ARDS associati ad infezione sostenuta da altri virus respiratori è risultata significativamente più bassa (19,4% vs 58,3%, p-value < 0,0000001). Rhinovirus e Enterovirus sono stati i virus maggiormente identificati, tuttavia anche il virus influenzale stagionale A(H3N2) ha svolto un ruolo significativo. La tecnica di "virus discovery" VIDISCA-454, si è dimostrata una metodologia sensibile e specifica, applicabile con successo alla sorveglianza delle infezioni respiratorie virali. Questa ha permesso l'identificazione di un caso di SARI associato ad infezione da Paramyxovirus (virus del morbillo) sfuggito alla precedente diagnosi. Nonostante l'ampio spettro di agenti patogeni, virali e batterici, ricercati, quasi il 22% dei casi di SARI/ARDS non ha ottenuto una diagnosi definitiva. Nella pratica clinica, particolare priorità dovrebbe essere dedicata al miglioramento dei protocolli di diagnosi delle infezioni respiratorie gravi, proprio per ridurre tale "gap diagnostico".

L'approfondimento delle caratteristiche geniche di RSV ha permesso di valutare la circolazione dei diversi genotipi e studiarne le caratteristiche molecolari. L'analisi filogenetica e quella filodinamica hanno permesso di comprendere meglio la storia naturale del virus, anche se questo resta difficile da confermare, data la grande diversità antigenica presentata. La variabilità della proteina G sembra svolgere un ruolo significativo nella patogenesi di RSV consentendo l'evasione alla risposta immunitaria; è importante quindi monitorare i cambiamenti che avvengono a livello delle sequenze codificanti questa proteina, sia per consentire l'identificazione dei diversi ceppi epidemici, sia per sviluppare future strategie vaccinali e terapeutiche. In conclusione, questo progetto di tesi, ha messo in evidenza come la virologia respiratoria, in continua evoluzione, necessiti di un costante aggiornamento delle procedure diagnostiche e di sorveglianza, indispensabile supporto nello studio e nel monitoraggio sia delle infezioni virali classiche e, soprattutto, di quelle causate da agenti emergenti o di recente identificazione.