

IDENTIFICAZIONE MEDIANTE *NEXT GENERATION SEQUENCING* DI SOTTOPOPOLAZIONI VIRALI IN UN VACCINO VIVO ATTENUATO PER METAPNEUMOVIRUS AVIARE SOTTOTIPO B E LORO IMPLICAZIONE NEL FENOMENO DI REVERSIONE A VIRULENZA

Franzo G.¹, Naylor C.J.², Drigo M.¹, Croville G.³, Ducatez M.F.³, Catelli E.⁴, Laconi A.⁴, Cecchinato M.¹

¹ *Dipartimento di Medicina Animale, Produzioni e Salute (MAPS), Università degli Studi di Padova, Agripolis - Viale dell'Università 16, 35020 Legnaro (PD);*

² *Department of Infection Biology, University of Liverpool, Leahurst Campus, Neston, Cheshire, CH64 7TE – United Kingdom;*

³ *INRA and Université de Toulouse, INP, ENVT, UMR 1225 IHAP, F-31076, Toulouse, France;*

⁴ *Dipartimento di Scienze Mediche Veterinarie, Università di Bologna, Via Tolara di Sopra, 50, 40064 – Ozzano dell'Emilia (BO)*

Summary

Avian metapneumovirus (aMPV) infects respiratory and reproductive tracts of domestic poultry, often involving secondary infections, and leads to serious economic losses in most parts of the world. While in general disease is effectively controlled by live vaccines, reversion to virulence of those vaccines has been demonstrated on several occasions. Consensus sequence mutations involved in the process have been identified in more than one instance. In one previous subtype A aMPV candidate vaccine study, small subpopulations were implicated. In the current study, the presence of subpopulations in a subtype B vaccine was investigated by deep sequencing. Of the 19 positions where vaccine and progenitor consensus sequences differed, subpopulations were found to have sequence matching progenitor sequence in 4 positions. However none of these mutations occurred in a virulent revertant of that vaccine, thereby demonstrating that the majority progenitor virus population had not survived the attenuation process, hence were not obviously involved in any return to virulence. However within the vaccine, a single nucleotide variation was found which agreed with consensus sequence of a derived virulent revertant virus, hence this and other undetected, potentially virulent subpopulations, cannot totally be discounted from being involved in reversion. Much deeper sequencing of progenitor, vaccine and revertant may clarify whether problematic virulent subpopulations are present and therefore whether these need to be routinely removed during aMPV vaccine preparation prior to registration and release.

INTRODUZIONE

Il metapneumovirus aviare (aMPV) è un patogeno di grande importanza per l'industria avicola, in quanto in grado di determinare sindromi respiratorie sia nel tacchino che nel pollo che se complicate da infezioni batteriche secondarie possono conseguire in rilevanti perdite economiche (7).

Alla fine degli anni '80 sono stati messi in commercio vaccini vivi attenuati che hanno permesso il controllo di tali infezioni. Tuttavia in campo si osservano ancora forme respiratorie riferibili ad aMPV (2, 5). Ciò è stato imputato a diverse cause tra cui la vaccinazione mal praticata, scarsa durata dell'immunità e incompleta protezione tra i sottotipi (11, 15).

Diversi studi hanno inoltre dimostrato la possibilità di fenomeni di reversione a virulenza di questi vaccini e riportato, sia per vaccini allestiti con aMPV sottotipo A e sottotipo B, i siti coinvolti in questo fenomeno (3, 4). Sebbene la presenza di sottopopolazioni vaccinali caratterizzate da una diversa virulenza sia stata riportata per il sottotipo A (12), questo non preclude il concomitante ruolo di una progressiva evoluzione con acquisizione di mutazioni dei ceppi vaccinali durante la catena di trasmissione *in vivo*. Nel presente studio, un vaccino basato su aMPV sottotipo B è stato sottoposto a *deep sequencing* tramite protocolli di *Next Generation Sequencing* (NGS) al fine di identificare e studiare la presenza di sottopopolazioni virali potenzialmente implicate nel fenomeno di reversione a virulenza.

MATERIALI E METODI

Vaccino aMPV sottotipo B

Un flacone di vaccino aMPV sottotipo B (ceppo VCO3/50) è stato risospeso in 2 mL di acqua per biologia molecolare. L'RNA è stato estratto da 200 μ L di sospensione virale utilizzando il kit High pure Viral RNA kit (Roche Diagnostics, Italia). Durante questa fase è stata anche prevista una fase di digestione del DNA.

Next Generation Sequencing

Al fine di limitare il possibile bias legato alla processione del campione, sono stati utilizzati 2 distinti protocolli di preparazione delle librerie utilizzate poi per il sequenziamento.

Protocollo 1. Il primo protocollo ha previsto l'amplificazione del campione di partenza tramite RT-PCR. L'intero genoma di aMPV è stato retro trascritto in un'unica reazione utilizzando il kit Maxima H Minus Reverse kit (Thermo Scientific, Italia). Il cDNA così ottenuto è stato poi amplificato utilizzando il kit Phusion Hot Start II High Fidelity Polymerase Kit (Thermo Scientific, Italia) con il fine di ottenere 2 amplificati parzialmente sovrapposti (approssimativamente 7000bp per ciascun frammento) che includessero l'intera sequenza codificante del genoma virale. I prodotti così ottenuti, previa purificazione, sono stati inviati a MACROGEN per il successivo sequenziamento. La libreria, preparata tramite il Nextera DNA XT kit, è stata sequenziata (100bp *paired-end*) utilizzando il sequenziatore Illumina HiSeq.

Protocollo 2. Il secondo protocollo ha previsto l'esecuzione di una two-step RT-PCR utilizzando *random hexamer primers* recanti una sequenza *tag* comune al fine di amplificare in modo casuale tutto l'RNA del campione, fornendo contestualmente gli "adaptor" necessari per la successiva fase di PCR. I prodotti di amplificazione sono stati sottoposti a elettroforesi su gel di agarosio e le bande di dimensione di approssi-

mativamente 500bp sono state escisse da gel e purificate. I prodotti così ottenuti sono stati sequenziati tramite lo strumento MiSeq (300bp *paired-end*).

Data analisi

I dati grezzi ottenuti con entrambi i protocolli sono stati elaborati al fine ottenere la sequenza dell'intero genoma virale e di investigare la presenza di siti caratterizzati da variabilità (*Single nucleotide variation* - SNV), indice della presenza di sottopopolazioni virali. Gli *adaptor* e le basi caratterizzate da un *quality score* inferiore a 30 sono stati rimossi utilizzando il programma Trimmomatic (<http://www.usadel-lab.org/cms/?page=trimmomatic>). Le *reads* sono state poi allineate ad un genoma di riferimento utilizzando il software Bowtie2 (8). Al fine di discriminare gli SNV genuini da quelli derivati dalla presenza di inevitabili errori generatisi durante la processazione del campione e il sequenziamento stesso, sono stati utilizzati due software open source, (17)ShoRAH {{833 Zagordi,O. 2011}} and QUASR (16), basati su differenti algoritmi. Il primo corregge gli errori di sequenziamento dividendo in cluster tutte le *reads* che si allineano ad una data porzione del genoma. La sequenza consenso di ogni cluster viene quindi assunta essere l'aplotipo originale da cui le restanti, ed erronee, *reads* sono derivate. Il numero di *reads* all'interno del cluster permette altresì di stimare la prevalenza di ciascun aplotipo all'interno della popolazione. Il secondo software sfrutta invece un approccio basato sulla qualità delle sequenze e sul *coverage* di ciascuna posizione per verificare l'accuratezza di ciascuna variante. Al fine di garantire la confidenza dei risultati ottenuti, è stato utilizzato un approccio molto conservativo nell'analisi dei risultati: sono stati accettati solamente le SNV rilevate da entrambi i software e analizzando i dati ottenuti da ambedue i protocolli di sequenziamento. QUASR è stato anche utilizzato per ottenere la sequenza consenso del genoma virale.

RISULTATI

Analisi dei dati

L'analisi delle *reads* ottenute tramite il protocollo 1 e 2 ha permesso di evidenziare un *coverage* medio di 200 e 272. In particolare, ciascuna base è stata sequenziata un numero di volte almeno superiore a 10 nella regione compresa fra il nucleotide 14 e 13427, con il protocollo 1, e da 6 a 13433, con il protocollo 2. L'analisi delle sequenze consenso ha permesso di evidenziare un pieno accordo fra i risultati ottenuti con i 2 metodi di preparazione del campione. Confrontando queste sequenze con quella vaccinale di riferimento (ottenuta con il metodo Sanger) è stata evidenziata una sola differenza in posizione 3566 (T↔C).

Analisi delle Single nucleotide variation (SNV)

Sono stati identificati nove siti nei quali, indipendentemente dal protocollo di sequenziamento utilizzato e dal software di analisi, è stata riscontrata la presenza di sottopopolazioni (Tabella 1). Fra questi, le SNV identificate in posizione 2082, 9571, 12290 e 12978 erano state precedentemente associate al processo di attenuazione vaccinale (4). Confrontando la sequenza del vaccino e del rispettivo progenitore è

stato possibile evidenziare come, invariabilmente, i nucleotidi presenti con la frequenza più alta corrispondano alla sequenza del ceppo vaccinale mentre quelli a frequenza minore coincidano con la sequenza del progenitore (Tabella 2). Similmente, la SNV riscontrata in posizione 6487 ripropone il polimorfismo precedentemente descritto fra la sequenza vaccinale e dei ceppi di origine vaccinale revertiti a virulenza e sequenziati in precedenti studi (4) (Tabella 3). In aggiunta, altri 4 e 2 siti, precedentemente riportati come implicati rispettivamente nei fenomeni di attenuazione e reversione a virulenza sono stati rilevati con almeno un metodo di analisi.

DISCUSSIONE E CONCLUSIONI

L'elevato tasso di mutazione dei virus a RNA (~1 mutazione/genoma/replicazione) caratterizza fortemente la biologia di questi patogeni (9, 14). Queste popolazioni virali sono sovente descritte come uno sciame di progenie mutanti che orbita attorno ad una sequenza consenso. Diversi autori hanno ipotizzato che questa struttura permetta la genesi di molteplici varianti in grado di esplorare rapidamente vaste regioni del *fitness landscape* (1, 6). Sebbene nella stragrande maggioranza dei casi le mutazioni risultino deleterie per il virus (13), in taluni casi queste possono risultare vantaggiose portando alla genesi di "individui" più adatti al presente ambiente o preadattati a nuovi ambienti (9). E' quindi possibile che, all'interno della popolazione di virus vaccinali, siano presenti delle varianti già parzialmente adatte alla replicazione *in vivo* e che queste possano quindi essere progressivamente selezionate in condizioni di campo. In particolare, l'assenza di un'adeguata copertura vaccinale negli allevamenti avicoli può favorire questo fenomeno, permettendo la circolazione del virus vaccinale all'interno dell'allevamento per prolungati periodi di tempo.

I risultati del presente studio hanno permesso di evidenziare l'effettiva presenza di sottopopolazioni virali nel vaccino aMPV sottotipo B studiato. Degno di nota è il riscontro di 4 siti caratterizzati da variabilità genetica precedentemente identificati come responsabili dell'attenuazione del virus vaccinale. Sarebbe stata quindi prevedibile una maggiore *fitness* di questi virus *in vivo* e una conseguente maggiore abilità dei ceppi latenti di queste mutazioni nel replicare e dare sintomatologia. Tuttavia, nessuno di questi siti è stato associato, nei precedenti studi, alla reversione a virulenza, ponendo in dubbio il loro contributo nel determinismo della virulenza.

Non è tuttavia possibile escludere che solo la simultanea presenza di queste mutazioni nel medesimo virus, evenienza attualmente non ricostruibile in modo accurato con gli strumenti di analisi disponibili, ne conferisca la patogenicità. Globalmente, i risultati di questo studio, paiono quindi confermare l'efficacia del processo di attenuazione.

Altre sottopopolazioni, inclusi virus recanti mutazioni in comune con ceppi revertanti, sono state identificate. È ipotizzabile che queste fossero già presenti nel ceppo progenitore e si siano mantenute nel vaccino o, addirittura, siano originate durante la produzione dello stesso. Due siti in particolare hanno presentato la medesima sequenza del ceppo revertante come sottopopolazione (posizione 6487), in un caso, e come popolazione unica (posizione 3566), nell'altro. Un'ipotesi suggestiva è rappresentata dall'idea che questi SNVs costituiscano, potenzialmente interagendo con siti

di attenuazione parziale, il substrato su cui la progressiva selezione ed evoluzione *in vivo* agiscono, sfociando nell'emergenza di ceppi revertanti. Difatti, il limitato numero di SNV identificato, sconfessa fortemente l'ipotesi di una popolazione virulenta già presente "tal quale" nel vaccino. Quanto finora riportato appare in accordo con i risultati di un precedente studio condotto su un vaccino AMPV sottotipo A nel quale è stata riportata l'importanza di sottopopolazioni presenti a bassa frequenza (1 in 10⁶) nel determinare la reversione a virulenza (12). Sebbene, sulla base di queste evidenze, il vaccino in questione sia stato sottoposto a purificazione mediante il metodo delle placche prima dell'immissione in commercio, episodi di reversione a virulenza, sebbene con minor frequenza rispetto alla formulazione non purificata (12), sono stati egualmente identificati (3, 10). Appare quindi impossibile negare il ruolo, quantomeno in termini di co-attore, della progressiva evoluzione virale, in condizioni di campo, nella riacquisizione del fenotipo virulento.

In conclusione, i risultati preliminari riportati nel presente studio, associati a quanto già descritto in letteratura, lasciano intravedere uno scenario caratterizzato dall'azione della pressione selettiva imposta dall'ospite su una popolazione di virus vaccinali di per se non omogenea. Sottopopolazioni parzialmente adattate alla replicazione *in vivo* potrebbero essere preferenzialmente selezionate e costituire la base per la successiva evoluzione nell'ospite e la riacquisizione del fenotipo virulento. Ulteriori studi saranno necessari per valutare la relazione fra la variabilità della popolazione presente nel ceppo progenitore e nel vaccino nonché per analizzare, mediante infezioni sperimentali, l'evoluzione del ceppo vaccinale *in vivo*.

BIBLIOGRAFIA

1. Belshaw, R., A. Gardner, A. Rambaut, and O. G. Pybus. (2008). Pacing a small cage: mutation and RNA viruses. 23:188-193.
2. Catelli, E., M. Cecchinato, M. Delogu, P. de Matteo, G. Ortali, C. Franciosi, M. A. de Marco, and C. J. Naylor. (2004). Avian Pneumovirus infection in turkey and broiler farms in Italy: A virological, molecular and serological field survey | Infezione da pneumovirus aviare nel tacchino da carne e nel broiler in italia: Indagini di campo mediante tecniche virologiche, sierologiche e molecolari. 3:287.
3. Catelli, E., M. Cecchinato, C. E. Savage, R. C. Jones, and C. J. Naylor. (2006). Demonstration of loss of attenuation and extended field persistence of a live avian metapneumovirus vaccine. Vaccine 24:6476.
4. Cecchinato, M., E. Catelli, C. Lupini, E. Ricchizzi, S. Prosperi, and C. J. Naylor. (2014). Reversion to virulence of a subtype B avian metapneumovirus vaccine: Is it time for regulators to require availability of vaccine progenitors? Vaccine 32:4660-4664.
5. Cecchinato, M., C. Lupini, E. Ricchizzi, M. Falchieri, A. Meini, R. C. Jones, and E. Catelli. (2012). Italian field survey reveals a high diffusion of avian metapneumovirus

- subtype B in layers and weaknesses in the vaccination strategy applied. *Avian Dis.* 56:720-724.
6. Elena, S. F., R. Miralles, J. M. Cuevas, P. E. Turner, and A. Moya. (2000). The two faces of mutation: extinction and adaptation in RNA viruses. *IUBMB Life* 49:5-9.
7. Gough, R. E., and R. C. Jones. (2008). Avian metapneumovirus. In: *Diseases of poultry*, 12th ed. Y. M. Saif, A. M. Fadly, J. R. Glisson, L. R. McDougald, N.L. Nolan, and D. E. Swayne. Wiley-Blackwell, Ames, IA.
8. Langmead, B., and S. L. Salzberg. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
9. Lauring, A. S., J. Frydman, and R. Andino. (2013). The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microbiol.* 11:327-336.
10. Lupini, C., M. Cecchinato, E. Ricchizzi, C. J. Naylor, and E. Catelli. (2011). A turkey rhinotracheitis outbreak caused by the environmental spread of a vaccine-derived avian metapneumovirus. *Avian Pathol.* 40:525-530.
11. Naylor, C., K. Shaw, P. Britton, and D. Cavanagh. (1997). Appearance of type B avian pneumovirus in Great Britain. *Avian Pathol.* 26:327-338.
12. Naylor, C. J., and R. C. Jones. (1994). Demonstration of a virulent subpopulation in a prototype live attenuated turkey rhinotracheitis vaccine. *Vaccine* 12:1225-1230.
13. Sanjuan, R., A. Moya, and S. F. Elena. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 101:8396-8401.
14. Sanjuan, R., M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw. (2010). Viral mutation rates. *J. Virol.* 84:9733-9748.
15. Van de Zande, S., H. Nauwynck, C. Naylor, and M. Pensaert. (2000). Duration of cross-protection between subtypes A and B avian pneumovirus in turkeys. *Vet. Rec.* 147:132-134.
16. Watson, S. J., M. R. Welkers, D. P. Depledge, E. Coulter, J. M. Breuer, M. D. de Jong, and P. Kellam. (2013). Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368:20120205.
17. Zagordi, O., A. Bhattacharya, N. Eriksson, and N. Beerenwinkel. (2011). ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. *BMC Bioinformatics* 12:119-2105-12-119.

QUASR							ShoRAH					
Posizione	Protocollo1			Protocollo2			Protocollo1			Protocollo2		
	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage
2082	T	C(12%)	198	T	C(14%)	469	T	C(15%)	100	T	C(16%)	281
4921	A	G(4%)	197	A	G(4%)	246	A	G(7%)	101	A	G(4%)	190
4934	C	T(17%)	197	C	T(4%)	247	C	T(25%)	101	C	T(5%)	186
6346	T	A(10%)	196	T	A(14%)	227	T	A (24%)	101	T	A(16%)	122
6487	T	G(22%)	197	T	G(31%)	427	T	G (38%)	101	T	G(24%)	238
9571	A	T(9%)	194	A	T(11%)	124	A	T(22%)	101	A	T(23%)	71
9610	T	C(7%)	194	T	C(5%)	162	T	C(9%)	101	T	C(14%)	77
12290	A	T(14%)	193	T	A(39%)	166	T	A(68%)	99	T	A(39%)	111
12978	A	G(6%)	197	A	G(38%)	124	A	G(30%)	101	A	G(40%)	73

Tabella 1. Siti per i quali tutte le combinazioni di protocolli di sequenziamento e programmi di analisi hanno evidenziato la presenza di SNV. Le posizioni nucleotidiche sono numerate sulla base del genoma di referenza (vaccino VCO3/50) [4]. Per ciascuna di esse sono riportati il *coverage*, il nucleotide di referenza e la corrispondente variante, la cui frequenza è indicata fra parentesi.

QUASR							ShoRAH					
Posizione	Protocollo1			Protocollo2			Protocollo1			Protocollo2		
	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage
1220												
2082	T	C(12%)	198	T	C(14%)	469	T	C(15%)	100	T	C(16%)	281
2422												
2554	T	C (3%)	192									
2654												
3210				T	G (3%)	281						
3391												
5434	G	A (11%)	194				G	A (9%)	101			
5777	T	A (4%)	199				T	A (4%)	101			
6104												
6712												
9571	A	T(9%)	194	A	T(11%)	124	A	T(22%)	101	A	T(23%)	71
9817												
12290	A	T(14%)	193	T	A(39%)	166	T	A(68%)	99	T	A(39%)	111
12978	A	G(6%)	197	A	G(38%)	124	A	G(30%)	101	A	G(40%)	73
13463												
13464												

Tabella 2. Siti implicati nel processo di attenuazione [4] identificati come SNV da almeno un metodo di analisi. Le posizioni nucleotidiche sono numerate sulla base del genoma di referenza (vaccino VCO3/50) [4]. Per ciascuna di esse sono riportati il *coverage*, il nucleotide di referenza e la corrispondente variante, la cui frequenza è indicata fra parentesi.

QUASR							ShoRAH					
Posizione	Protocollo1			Protocollo2			Protocollo1			Protocollo2		
	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage	Genoma di Referenza	Variante	Coverage
1461	A	C (1%) T(4%)	190	A	G (1%)	280						
3508	-	-	-									
3566	C	T(100%)	201	C	T(100%)	398	C	T(100%)	101	C	T(100%)	190
4348				A	G (5%)	342						
5614												
6487	T	G(22%)	197	T	G(31%)	427	T	G (38%)	101	T	G(24%)	238
12277												
12847												

Tabella 3. Siti implicati nel processo di reversione a virulenza [4] identificati come SNV da almeno un metodo di analisi Le posizioni nucleotidiche sono numerate sulla base del genoma di referenza (vaccino VCO3/50) [4]. Per ciascuna di esse sono riportati il *coverage*, il nucleotide di referenza e la corrispondente variante, la cui frequenza è indicata fra parentesi.