

SỰ DI CHUYỂN VÀ TÍNH BỀN VỮNG CỦA VIRUS CÚM A TẠI VIỆT NAM TỪ 2001 ĐẾN 2008

Các tác giả: Mai Quynh Le, Ha Minh Lam, Vuong Duc Cuong, Tommy Tsan-Yuk Lam, Rebecca A Halpin, David E Wentworth, Nguyen Tran Hien, Le Thi Thanh, Hoang Vu Mai Phuong, Peter Horby, Maciej F Boni

Các liên kết: Viện Vệ Sinh Dịch Tễ Trung Ương (M. Q. Le, V. D. Cuong, N. T. Hien, L. T. Thanh, H. V. M. Phuong); Đơn Vị Nghiên Cứu Lâm Sàng Đại Học Oxford thành phố Hồ Chí Minh (H. M. Lam, M. F. Boni); Trung Tâm Miễn Dịch, Truyền Nhiễm và Tiến Hóa, Đại học Edinburgh, Anh (H. M. Lam); Đại Học Oxford, Anh (T. TY. Lam, P. Horby, M. F. Boni); Viện The J. Craig Venter, Mỹ(R. A. Halpin, D. E. Wentworth); Đơn Vị Nghiên Cứu Lâm Sàng Đại Học Oxford Hà Nội (P. Horby);

Tóm lược

Sử dụng 240 trình tự bộ gen đầy đủ của cúm A được thu thập tại Việt Nam (từ 2001 đến 2008) chúng tôi muốn tìm hiểu quy luật tồn tại và mô hình di chuyển của virus cúm giữa Việt Nam và các nước khác. Chúng tôi đã thấy rằng các dòng virus ở Việt Nam chủ yếu du nhập đến/từ Trung Quốc, Hồng Kong, Đài Loan, Cam Pu Chia, Nhật Bản, Nam Hàn và Mỹ. Chúng tôi cố gắng giảm thiểu độ sai lệch do địa lý bằng cách tạo ra cây phân loài từ những mẫu con ở cấp độ năm và quốc gia. Mặc dù vậy, các sự kiện du nhập trong những cây này vẫn bị kiểm soát mạnh mẽ bởi sự hiện diện hoặc thiếu sót của dữ liệu các trình tự, xác nhận rằng cần phải thiết kế một nghiên cứu dịch tễ học kiểm soát tỉ lệ này để đẩy mạnh việc phân tích sự du nhập. Khi so sánh với kết quả phân tích của toàn bộ gen (8 phân đoạn), chúng tôi thấy rằng hầu hết các sự kiện du nhập của virus không được thể hiện đầy đủ trong cây phân loài của gen haemagglutinin (HA), mặc dù kết quả phân tích riêng lẻ từ phân đoạn này vẫn cho thấy mối liên kết tổng thể giữa các nước. Nghiên cứu của chúng tôi chỉ ra rằng nhiều khả năng các chủng virus của Việt Nam có thời gian tồn tại lâu hơn 1 năm.

Phần Giới Thiệu

Hiểu biết các động lực của cúm ở những khu vực Nhiệt đới là rất quan trọng để hiểu biết đại dịch cúm toàn cầu khi mà những động lực giữa các khu vực Ôn đới và Nhiệt đới là có liên kết chặt chẽ. Các nghiên cứu phát sinh loài đã chỉ ra rằng Đông Á, Đông Nam Á, và các khu vực Nhiệt Đới là những nguồn có tiềm năng sinh thái cho sự lưu thông cúm toàn cầu (1,2). Tuy nhiên, một số nghiên cứu khác lại cho rằng nhiều vùng địa lý đa dạng có thể hoạt động như các quần thể nguồn (3-5). Do đó, vai trò của các khu vực Nhiệt đới đối với đại dịch cúm toàn cầu vẫn chưa được khẳng định, và các dữ liệu chuỗi gen virus từ những nước Nhiệt đới đóng vai trò rất quan trọng cho việc hiểu biết các tuyến du nhập của bệnh cúm trong các khu vực Nhiệt đới cũng như giữa các nước Nhiệt đới và Ôn đới.

Việt Nam là ví dụ điển hình cho một quốc gia Nhiệt đới có thể đóng một vai trò quan trọng trong đại dịch cúm toàn cầu nhưng lại có ít thông tin về dịch cúm và cấu trúc quần thể virus được biết đến. Các kết quả khảo sát trọng điểm chỉ ra rằng bệnh cúm ở Việt Nam thường có 1 đến 2 đợt cao điểm mỗi năm, nhưng kể cả dữ liệu về các bệnh giống cúm (ILI) hay dữ liệu xác nhận virus cũng không chỉ ra được một mô hình theo mùa đơn giản; số trường hợp được xác nhận cúm có xu hướng biến động nhiều hơn số ca ILI (6,7). Các nghiên cứu huyết thanh học chỉ ra rằng tỉ lệ mắc cúm hằng năm ở Việt Nam vào khoảng 17% đến 26% (8). Do dân số Việt Nam có cấu trúc tương đối trẻ nên các mô hình tiên lượng có sử dụng tỉ lệ tiếp xúc (qua giao tiếp) đều cho thấy hầu hết các trường hợp nhiễm cúm nằm trong nhóm tuổi trẻ (9,10). Với mật độ dân số cao và các liên kết du lịch mạnh mẽ đến Đông Á, Đông Nam Á, Úc/New Zealand, Việt Nam (cũng giống như nhiều nước Đông/Đông Nam Á khác) có khả năng là một nguồn phát tán virus cúm toàn cầu vì đây là môi trường thuận lợi cho việc lưu thông liên tục và quanh năm của một dòng cúm duy nhất. Khả năng tồn tại và phát tán liên tục, sau đây, được chúng tôi định nghĩa như “tính bền vững” của chủng virus cúm.

Các nghiên cứu phát sinh loài trên cúm đã chứng minh sự pha trộn của nhiều dòng virus ở mức độ toàn cầu (3,4,11), virus ở vùng ôn đới thường thiếu “tính bền vững” (1,2,11,12), trong khi các chủng cúm ở vùng cận nhiệt đới có xu hướng ít bị thay thế hơn (5,13). Nghiên cứu ở các nước nhiệt đới (trừ nghiên cứu số [14]) đều cho thấy rằng cúm ở vùng nhiệt đới không có mùa rõ ràng như ở vùng ôn đới(15-19), cúm ở các nước nhiệt đới có thể lưu hành liên tục trong suốt cả năm

(20,21); tuy nhiên, trong 2 nghiên cứu gần [20] và [21] các phân tích phát sinh loài không được thực hiện. Ở đây, chúng tôi trình bày các phân tích phát sinh loài của 240 chuỗi trình tự cúm từ Việt Nam, mẫu được thu thập bởi Hệ Thống Giám Sát Trọng Điểm Quốc Gia từ năm 2001 đến 2008 (6). Chúng tôi xác định độ mạnh tương đối của các liên kết du nhập cúm giữa Việt Nam và phần còn lại của thế giới, và chúng tôi cố gắng giải thích những kết quả này trong bối cảnh của độ lệch thống kê gây do quá trình lấy mẫu gây ra (số lượng mẫu đại diện cho mỗi năm và mỗi vùng địa lý không đồng đều) là không thể tránh khỏi. Dựa trên việc thu mẫu thường xuyên trong năm 2007 và 2008, chúng tôi đánh giá liệu rằng chủng virus cúm có tính bền vững ở Việt Nam trong suốt giai đoạn này không. Mặc dù vậy, chúng tôi không thể kết luận chắc chắn liệu rằng Việt Nam nhận virus từ các nước khác hay là nguồn phát tán virus ra toàn cầu.

Hiểu biết các mô hình lan truyền và tính bền vững của cúm toàn cầu là rất quan trọng cho việc duy trì một tiến trình lựa chọn có tính phối hợp và hiệu quả đối với vac-xin cúm. Những lựa chọn cho các thành phần vac-xin trong tương lai sẽ phụ thuộc vào những mẫu được thu thập gần đây và những hiểu biết về sự đóng góp của từng vùng đối với sự lưu thông cúm toàn cầu sẽ giúp đưa ra các quyết định dựa trên những virus đến từ các khu vực có kết nối mạnh hoặc kết nối yếu.

Các phương pháp

Mẫu. Mẫu phết mũi họng (NPS) hoặc mẫu phết họng (TS) được thu thập từ các bệnh nhân có biểu hiện giống cúm (ILI) tại các bệnh viện tham gia vào Hệ Thống Giám Soát Cúm Quốc Gia của Việt Nam, từ năm 2001 đến năm 2008. Mẫu bệnh phẩm được kiểm tra cúm A, B và các dòng cúm H1, H3, H5 bằng phương pháp RT-PCR sử dụng primers, probes và các hóa chất được khuyến cáo bởi các tổ chức CDC và WHO. Các mẫu dương tính sau đó được chọn lựa để phân lập virus và những virus được phân lập đạt đến nồng độ 1:8 trong thí nghiệm haemagglutination (8HAUs) sẽ được chọn để phân tích trình tự. Toàn bộ mẫu phân lập được phân dòng sử dụng thí nghiệm haemagglutination với các kháng nguyên và kháng thể từ bộ thử được khuyến dùng bởi WHO. Toàn bộ 242 mẫu được chuyển đến Dự án Trình Tự Gen Cúm tại Viện Sức Khỏe Quốc Gia Mỹ để giải trình tự bởi Viện J. Craig Venter (JCVI). Một mẫu không thể giải được trình tự và một mẫu bị nhiễm đã bị loại khỏi việc phân tích. Dữ liệu cuối cùng bao gồm 240 trình tự của chủng H3N2 (n=145) và H1N1 (n=95). Số ký hiệu nhập vào GenBank là CY103972-CY105893. Bảng 1 chỉ các số và vị trí của virus.

Bảng 1. Số lượng virus cúm được giải trình tự, 2001 - 2008								
Dòng virus, Khu vực địa lý	2001	2002	2003	Năm		2006	2007	2008
				2004	2005			
H3N2								
Miền Bắc	1	0	4	18	4	0	66	0
Miền Trung	0	0	9	2	19	1	8	13
Miền Nam	0	0	0	0	0	0	0	0
H1N1								
Miền Bắc	6	2	24	0	1	6	0	6
Miền Trung	0	0	1	0	2	4	0	36
Miền Nam	0	0	0	0	0	6	0	1

Thiết lập Dữ liệu. Để tiến hành các phân tích phát sinh loài theo vùng địa lý, các trình tự của virus cúm của cả 2 chủng H1N1 và H3N2 được đưa vào 2 nhóm dữ liệu: một “dữ liệu khu vực” gồm trình tự của toàn bộ 8 gen của virus từ Châu Á và Úc/New Zealand (ANZ), và một “dữ liệu toàn cầu” gồm 50 tập mẫu con của gen haemagglutinin (HA) (xem Các phương pháp bổ sung để biết thêm chi tiết). Các tập mẫu con được tạo ra bằng cách chọn ngẫu nhiên 12 trình tự trên một khu vực địa lý trên năm.

Phân tích phát sinh loài. Các dãy trình tự được giống cột bằng phần mềm MUSCLE v3.8 (23). Các cây phân loài được xây dựng bằng phương pháp tìm ước lượng hợp lý cực đại (Maximum Likelihood – ML) – thực hiện bằng phần mềm RAxML – với 2000 bootstrap (24,25). Đối với

các tập hợp dữ liệu của gen HA, thông tin ngày lấy mẫu được sử dụng để dựng cây phân loài (bằng phần mềm BEAST v1.6.2 (26)). Đồng hồ phân tử không nghiêm ngặt (phân phối lognormal không tương quan) được dùng để ước tính tốc độ đột biến của virus. Mô hình thay thế các nucleotide là SRD06+HKY85+ Γ ; các mô hình nhân khẩu học được sử dụng bao gồm: quy mô dân số không đổi và Bayesian skyride (xem phụ lục).

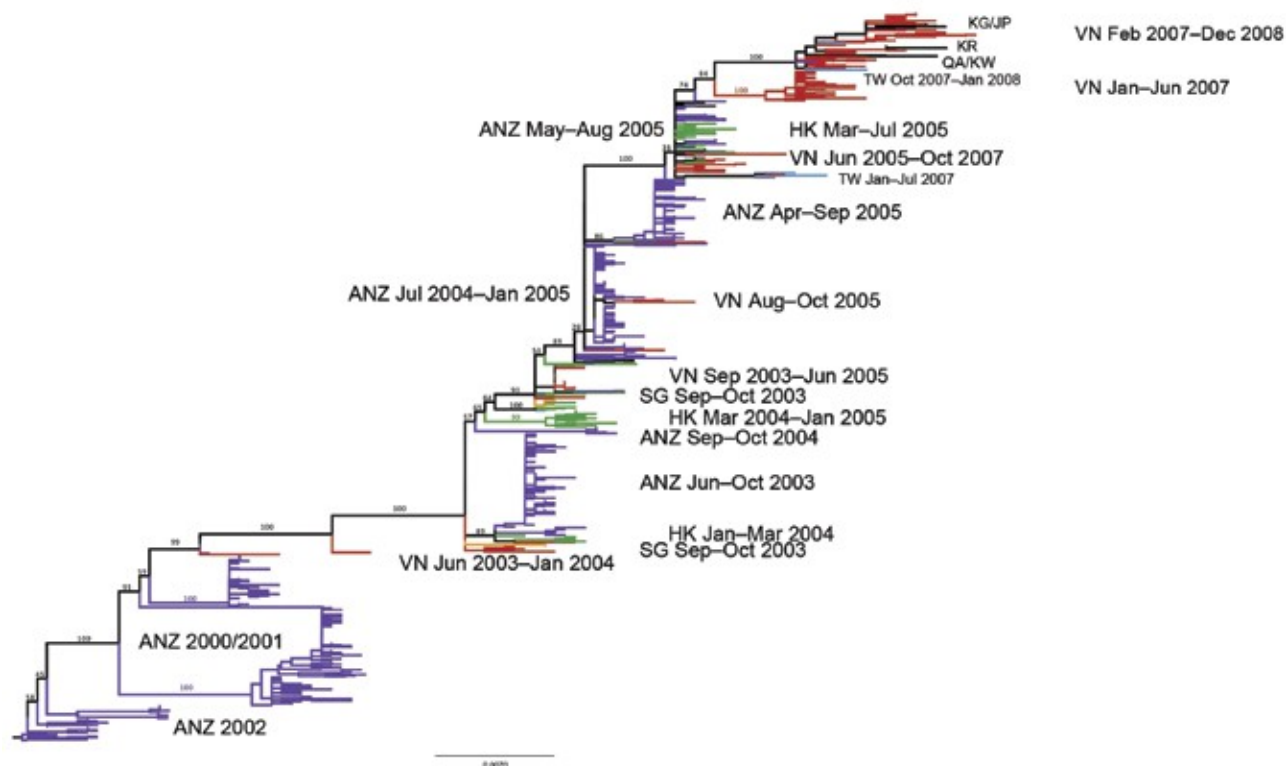
Phân tích di cư theo khu vực. Phân tích di cư được hoàn thành thông qua việc sử dụng một phương pháp hồi qui tuyến tính trong gói PAUP* (27,28). 2000 cây bootstrap và cây ML tốt nhất được suy ra bởi RAxML được đọc trong PAUP* với các trình tự nucleotide được thay thế bởi các mã vị trí kí tự đơn đã được định nghĩa (Bảng S2 và S3). Các thay đổi trong mã vị trí được ánh xạ lên các nhánh cây sử dụng tiêu chí hồi qui tuyến tính ACCTRAN (28). Do khả năng tái tổ hợp giữa các phân đoạn RNA của virus cúm, các phân tích trên được thực hiện với toàn bộ 8 phân đoạn gen để xác định xem lịch sử di cư có gì khác biệt khi quan sát những phân đoạn gen khác nhau. Khi sử dụng phương pháp xác định thể khảm lẫn phương pháp phân loài tiêu chuẩn, tín hiệu tái tổ hợp mạnh đều được xác định (29) (xem Tư liệu bổ sung và Hình S2).

Phân tích di cư toàn cầu. Để làm phép so sánh, một phân tích di cư tương tự được thực hiện trên 50 tập mẫu con của dữ liệu toàn cầu của H3N2 và H1N1. Một ma trận di cư được xây dựng để diễn tả số lượng các kết nối giữa 27 (H3N2) hoặc 29 (H1N1) vùng địa lý được xác định trước. Phần mềm Gephi (30) được sử dụng để mô phỏng các kết nối trong ma trận. Đối với các vùng có đủ mẫu, khoảng cách tối thiểu đến các thân cây của sự phát sinh loài đầu tiên được tính toán cho toàn bộ 50 cây mẫu con (chỉ cho H3N2), để xác định nếu các virus từ các khu vực khác có thể được diễn tả như là tổ tiên (gần thân cây) hoặc là có nguồn gốc (xa khỏi thân cây).

Các kết quả

Sự di chuyển của H3N2 trong khu vực châu Á và châu Đại Dương (qua phân tích gen HA)

Mối quan hệ giữa trình tự H3N2 Việt Nam và các mẫu virus khác trong khu vực châu Á và châu Đại Dương được thể hiện trong Hình 1. Mẫu đại diện của VN có ở tất cả các năm từ năm 2003 đến 2008, ngoại trừ năm 2006 khi H1N1 chiếm ưu thế. Virus được phân lập của Việt Nam có mối quan hệ gần gũi với virus phân lập của Hồng Kông, Đài Loan, Singapore và Úc / New Zealand. Hai mươi sự kiện di cư “hồi quy tuyến tính” được suy ra từ cây này, chín cho thấy sự di cư VN-HK, bảy cho thấy sự di cư VN-ANZ, hai cho thấy sự di cư VN-TW, và có một sự kiện di cư duy nhất với Hàn Quốc và Qatar. Một cách rõ ràng, như HK (68 trình tự) và ANZ (> 500 trình tự) được thấy rõ trong dữ liệu khu vực, tỉ lệ lớn về sự di cư VN có liên quan đến hai khu vực này. Lúc đầu khi chúng tôi không chắc chắn mức độ chính xác mà cây phát sinh loài theo khu vực địa lý có thể phản ánh chân thực mô hình di chuyển của virus cúm tại Việt Nam, chúng tôi đã thực hiện một quy trình xác thực để kiểm nghiệm tỉ lệ di chuyển toàn cầu của cúm H3N2 Việt Nam được tìm ra trong cây phát sinh loài khu vực (xem Phụ lục). Có khoảng 70% sự kiện di chuyển của H3N2 trong phân tích toàn cầu cũng được tìm thấy trong cây khu vực.

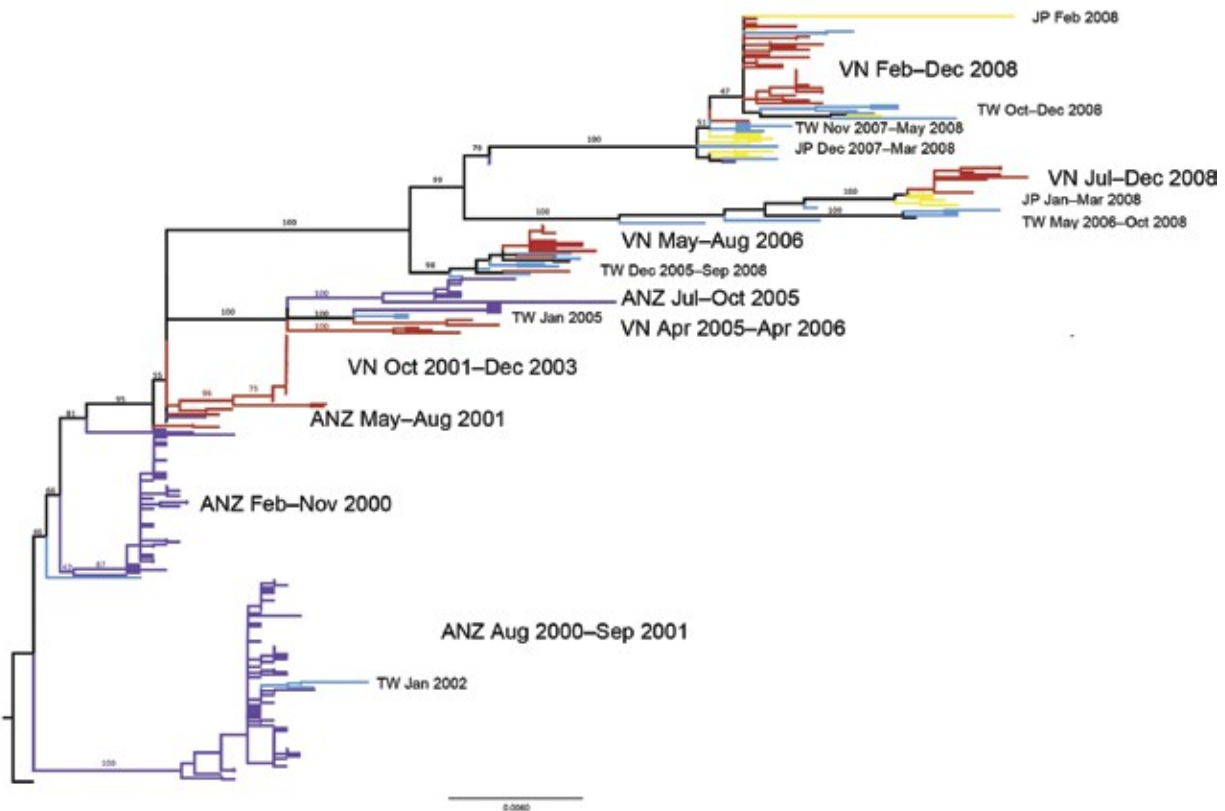


Hình 1. Cây phân loài (phân đoạn hemagglutinin – dựng bằng phương pháp tìm ước lượng hợp lý cực đại) của 787 trình tự H3N2 thu thập từ khu vực châu Á và châu Đại Dương. Dòng A/Canterbury/179/1999 được chọn làm gốc phát sinh loài. Giá trị bootstrap

được thể hiện trên các nốt trọng điểm. Các nhánh được tô màu theo vùng địa lý: đỏ – Việt Nam; tím – Úc hoặc New Zealand; xanh da trời – Hồng Kông; xanh lá – Đài Loan; cam – Singapo. Ngoài 2 nhãn “VN Jun 2003-Jan 2004” (đại diện cho các virus ở sát trên nó) và “ANZ 2000/2001” (đại diện cho 2 nhánh virus ở dưới và trên nó), các nhãn còn lại thể hiện thông tin cho dòng virus ở bên cạnh (trái hoặc phải) chúng. Tên các quốc gia được viết tắt: KG – Kyrgyzstan; VN – Việt Nam; JP – Nhật; KR – Hàn Quốc; QA – Qatar; KW – Kuwait; TW – Đài Loan; HK – Hồng Kông; ANZ – Úc và New Zealand. Thước tỉ lệ thể hiện tần suất biến đổi (trên 1 nucleotide) ứng với độ dài của mỗi nhánh trên cây.

Sự di chuyển của H1N1 trong khu vực châu Á và châu Đại Dương (quan phân tích gen HA)

Mối quan hệ giữa các trình tự H1N1 Việt Nam và các virus ở khu vực khác được thể hiện trong Hình 2. Mẫu đại diện cho VN có ở các năm từ năm 2001 đến năm 2008, ngoại trừ năm 2004 và 2007. Mười sự kiện di cư hồi qui tuyến tính được suy ra từ cây này, sáu cho thấy sự di cư giữa VN-TW, hai cho thấy VN-JP, và hai cho thấy sự di cư giữa VN-ANZ. Cũng như trong phân tích đối với H3N2, những liên kết di cư tương ứng với các virus đã được giải trình tự từ khu vực trong suốt giai đoạn 2001 - 2008.



Hình 2. Cây phân loài (phân đoạn hemagglutinin – dựng bằng phương pháp tìm ước lượng hợp lý cực đại) của 300 trình tự H1N1 thu thập từ khu vực châu Á và châu Đại Dương. Dòng A/ New Caledonia/20/1999 được chọn làm gốc phát sinh loài. Giá trị bootstrap được thể hiện trên các nốt trọng điểm. Các nhánh được tô màu theo vùng địa lý: đỏ – Việt Nam; tím – Úc hoặc New Zealand; vàng – Nhật; xanh lá – Đài Loan. Các nhãn thể hiện thông tin cho dòng virus ở bên cạnh (trái hoặc phải) chúng. Tên các quốc gia được viết tắt: JP – Nhật; VN – Việt Nam; TW – Đài Loan; ANZ – Úc và New Zealand. Thước tỉ lệ thể hiện tần suất biến đổi (trên 1 nucleotide) ứng với độ dài của mỗi nhánh trên cây.

Mô hình di chuyển của cúm qua phân tích toàn bộ gen. Khi cây khu vực chỉ bao gồm các trình tự mà toàn bộ bộ gen đã có sẵn, mô hình di cư được so sánh một cách hệ thống trên tất cả

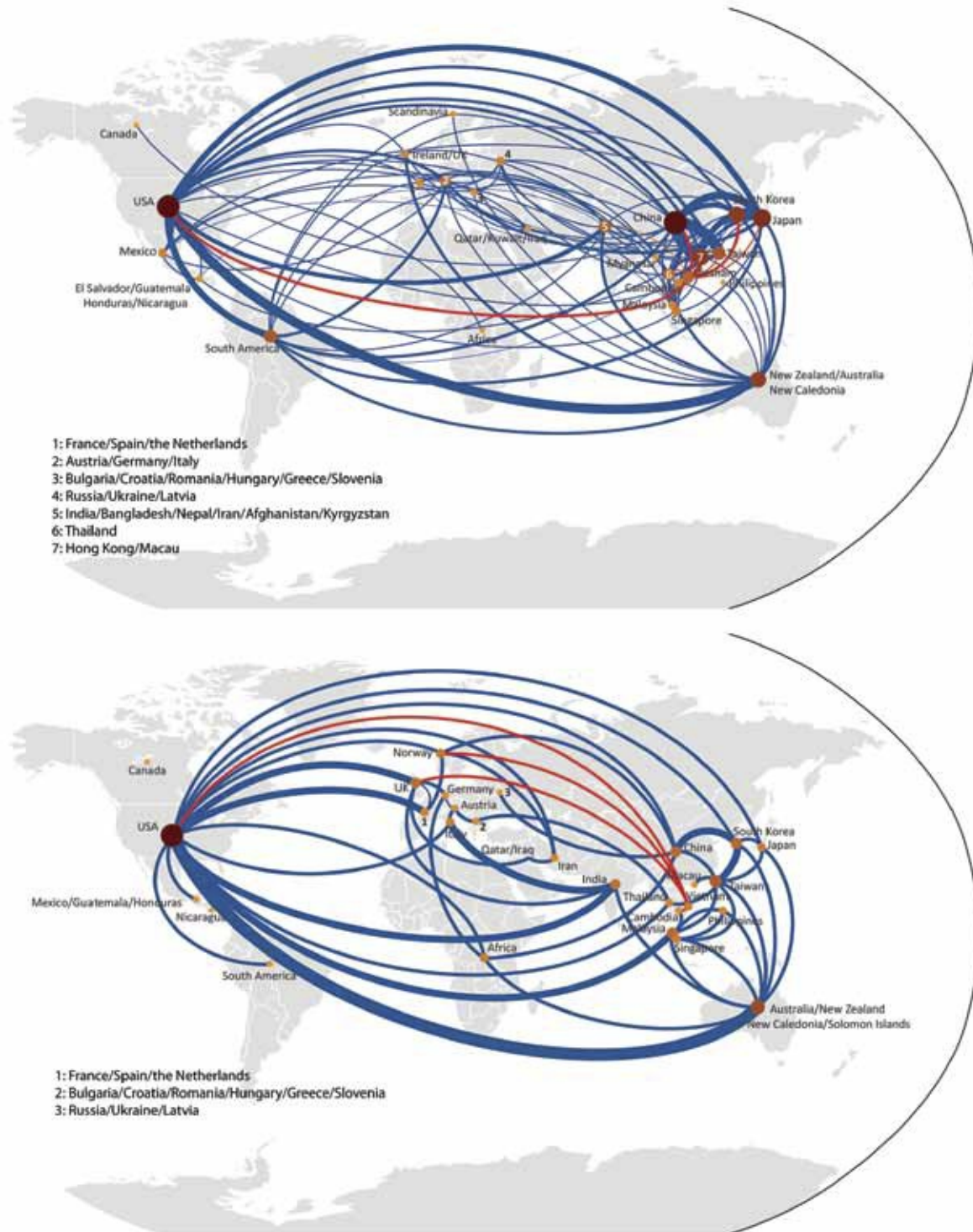
tám phân đoạn của virus cúm. Bởi vì virus cúm tái sắp xếp, các sự kiện lịch sử khác nhau sẽ được nhìn thấy trong các phát sinh loài suy ra riêng trong tám phân đoạn virus. Thật vậy, đối với dữ liệu H3N2, chúng tôi quan sát trung bình của 14 sự kiện di cư hồi qui tuyến tính đối với phân đoạn NA và trung bình của 41 đối với phân khúc MP, với các đoạn khác nằm đầu đó ở giữa (Bảng 2). Một lần nữa, phần lớn của sự di cư là với ANZ và Hồng Kông, điều này chỉ ra rằng mô hình di cư tương tự qua các phân khúc, thậm chí mặc dù có sự khác nhau về số lượng của sự di cư và sự khác nhau của các sự kiện di cư cá nhân nó vẫn có thể được nhìn thấy thông qua việc phân tích các phân khúc khác nhau. Đối với các phân đoạn MP và NS của virus H3N2, số lượng lớn các sự kiện di cư có thể là kết quả của số lượng lớn các nút hình học liên kết không chắc chắn và các nút polytomic trong những cây này phải được giải quyết một cách ngẫu nhiên để tính toán số lượng các sự kiện di cư – ví dụ các trình tự từ một quốc gia có thể nhầm lẫn trộn với các trình tự từ các nước khác, do đó tạo ra một số sự kiện di chuyển nhân tạo trong các phân tích hồi qui tuyến tính. Sự tự tin thấp trong các liên kết di cư VN-SG cho virus H3N2 có thể là do số lượng nhỏ các trình tự bộ gen có ở Singapore, tất cả đều được thu thập trong năm 2003. Đối với các vi rút H1N1, chúng tôi quan sát từ sáu đến mười sáu sự kiện di cư trên cây suy ra trong tám phân đoạn riêng lẻ (Bảng 2). Trong các dữ liệu bootstrapped, các kết nối di cư VN-TW và VN-JP xuất hiện xấp xỉ bằng nhau, mặc dù thực tế là các cây ML tốt nhất chỉ ra sáu kết nối VN-TW và hai kết nối VN-JP. Kết nối di cư giữa Việt Nam và ANZ có vẻ là hơi yếu. Điều này có thể là do việc thu mẫu ít hơn đáng kể đối của virus ANZ trong dữ liệu H1N1. Một số lượng tăng lên các sự kiện di cư đối với các phân đoạn MP và NS không được quan sát thấy trong các dữ liệu H1N1.

Bảng 2. Sự di chuyển của virus cúm giữa Việt Nam và các nước khác*

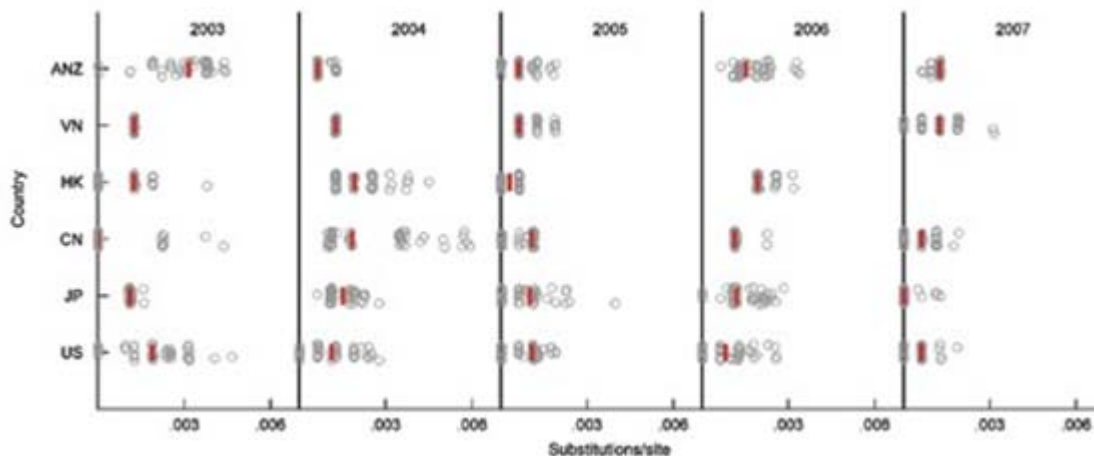
Dòng Virus, Quốc gia	Số lượng sự kiện di trú được tìm thấy: giá trị trung vị (khoảng dữ liệu 95%)							
	PB2	PB1	PA	HA	NP	NA	MP	NS
H3N2								
Úc/ New Zealand	9 (5-13)	9 (5-13)	7 (4-10)	7 (3-10)	11 (6-16)	5 (2-8)	26 (19-33)	27 (18-36)
Hong Kong	4 (1-7)	4 (1-8)	6 (2-10)	5 (2-8)	6 (2-10)	7 (3.5-12)	12 (6-18)	5 (1-10.5)
Đài Loan	2 (1-3)	3 (2-4)	3 (2-4)	3 (1-4)	2 (1-4)	1 (0-3)	3 (1-5)	2 (1-5)
Singapore	1 (0-3)	1 (0-3)	0 (0-2)	1 (0-3)	1 (0-6)	1 (0-3)	0 (0-2)	1 (0-3)
H1N1								
Đài Loan	3 (1-6)	3 (1-5)	4 (2-7)	5 (2-7)	5 (2-8)	4 (2-8)	5 (2-10)	5 (2-8)
Úc/ New Zealand	2 (0-5)	1 (0-4)	2 (0-5)	2 (0-5)	4 (1-8)	2 (0-4)	5 (2-9)	2 (0-6)
Nhật	2 (1-4)	2 (1-4)	2 (1-4)	3 (1-4)	4 (1-8)	1 (0-2)	6 (2-9)	4 (2-7)

*Dữ liệu từ 2.000 cây bootstrap cho toàn bộ 8 phân đoạn gen của các dòng cúm được thu thập từ châu Á và châu Đại Dương. Tên các phân đoạn gen được viết tắt: PB – polymerase basic protein; PA – polymerase acidic protein; HA – hemagglutinin; NP – nucleocapsid protein; MP – matrix protein; NS – nonstructural protein.

Sự di cư trong các cây mẫu con toàn cầu (qua phân tích gen HA). Do tương quan về số lượng mẫu giữa các vùng địa lý có ảnh hưởng lớn đến phân tích di chuyển, mô hình di chuyển được phân tích lại trên các dữ liệu mẫu con toàn cầu để giảm độ lệch địa lý từ việc có số lượng các mẫu cao hơn từ một số khu vực hơn những khu vực khác. Sử dụng cây HA toàn cầu cho cả trình tự của H3N2 và H1N1, chúng tôi xây dựng một ma trận di chuyển bao gồm tất cả các sự kiện di chuyển hồi qui tuyến tính giữa các khu vực được xác định trước (Bảng S2 và S3). Các mạng di chuyển được chỉ ra ở Hình 3, nó cho thấy Mỹ là một trung tâm lớn của sự di chuyển cúm, với khu vực Đông Á và ANZ cũng đóng một vai trò quan trọng. Dữ liệu H3N2 chỉ ra Việt Nam kết nối với hầu hết các nước khác trong khu vực Đông và Đông Nam Á, với Hoa Kỳ, và một cách yếu ớt với Nam Á. Dữ liệu H1N1 chỉ ra Việt Nam kết nối với Mỹ và châu Âu nhưng yếu với các nước châu Á khác. Đối với H3N2 và H1N1, tổng số các sự kiện di cư kết hợp với mỗi nút trong mạng là tương quan với số lượng mẫu cho nút đó (tất cả các giá trị $p < 10^{-5}$; các kiểm tra Kendall và Spearman). Lưu ý rằng số lượng mẫu không phải là giống hệt nhau cho mỗi nút bởi vì một số khu vực có ít hơn 12 trình tự mỗi năm và không cần phải tạo mẫu con trong những năm đó. Do đó, cả mẫu ít và mẫu nhiều đều có thể tạo ra mối tương quan này. Mặc dù nỗ lực của chúng tôi là làm giảm sai lệch địa lý trong các tập dữ liệu toàn cầu, suy luận trên các sự kiện di chuyển vẫn còn gắn liền với những vùng có mẫu; sai lệch này xuất hiện ảnh hưởng đến tất cả các nghiên cứu phát sinh loài địa lý. Đối với các trình tự H3N2, để xác định nếu có khu vực nào có những đặc điểm của một nguồn sinh thái, chúng tôi tính toán khoảng cách phát sinh loài của các trình tự từ sáu khu vực lấy mẫu nhiều (Trung Quốc, Hồng Kông, Nhật Bản, Việt Nam, ANZ, Mỹ) đến thân cây phát sinh loài ước lượng hợp lý cực đại toàn cầu (Hình 4). Trong năm 2003, ví dụ, qua toàn bộ 50 cây mẫu con, các trình tự được cô lập ở Trung Quốc dường như gần nhất với thân cây phát sinh loài điều đó cho thấy rằng những trình tự này là tổ tiên các trình tự virus khác được lấy mẫu trong năm 2003, điều này là phù hợp với sự thay thế toàn cầu chủng virus H3N2 bởi chủng A/Fujian/411/2002 đã xảy ra trong năm 2003. Nhìn chung, trong những năm 2003 đến 2007, không có một khu vực duy nhất nào có trình tự nhất quán với tổ tiên, xác nhận rằng không chỉ có một nguồn duy nhất virus cúm ở người trên toàn cầu. Mô hình di chuyển toàn cầu nhiều khả năng liên quan đến việc thay thế chuỗi trên toàn cầu một cách định kỳ ở các vùng khác nhau trong các năm khác nhau (3,4). Không có đủ mẫu từ tất cả các vùng/năm để thực hiện phân tích này trên các tập dữ liệu H1N1.



Hình 3. Bản đồ di chuyển của virus cúm được tái lập dựa trên cây phân loài của (A) H3N2, dựng từ 1,140 chuỗi trình tự, và (B) H1N1, dựng từ 554 chuỗi trình tự. Các chuỗi trình tự được chọn ngẫu nhiên từ bộ dữ liệu toàn cầu (xem phần Phương pháp nghiên cứu). Kích thước và màu sắc của các nốt trên bản đồ tương ứng với số lượng sự kiện di cư (giá trị trung vị từ 50 mẫu con) liên quan đến vị trí đó. Độ dày các đường nối ứng với mật độ di chuyển của virus - đường càng đậm thể hiện mật độ càng cao. Các đường di chuyển liên kết với Việt Nam được tô màu đỏ, các đường nối khác được thể hiện bằng màu xanh.

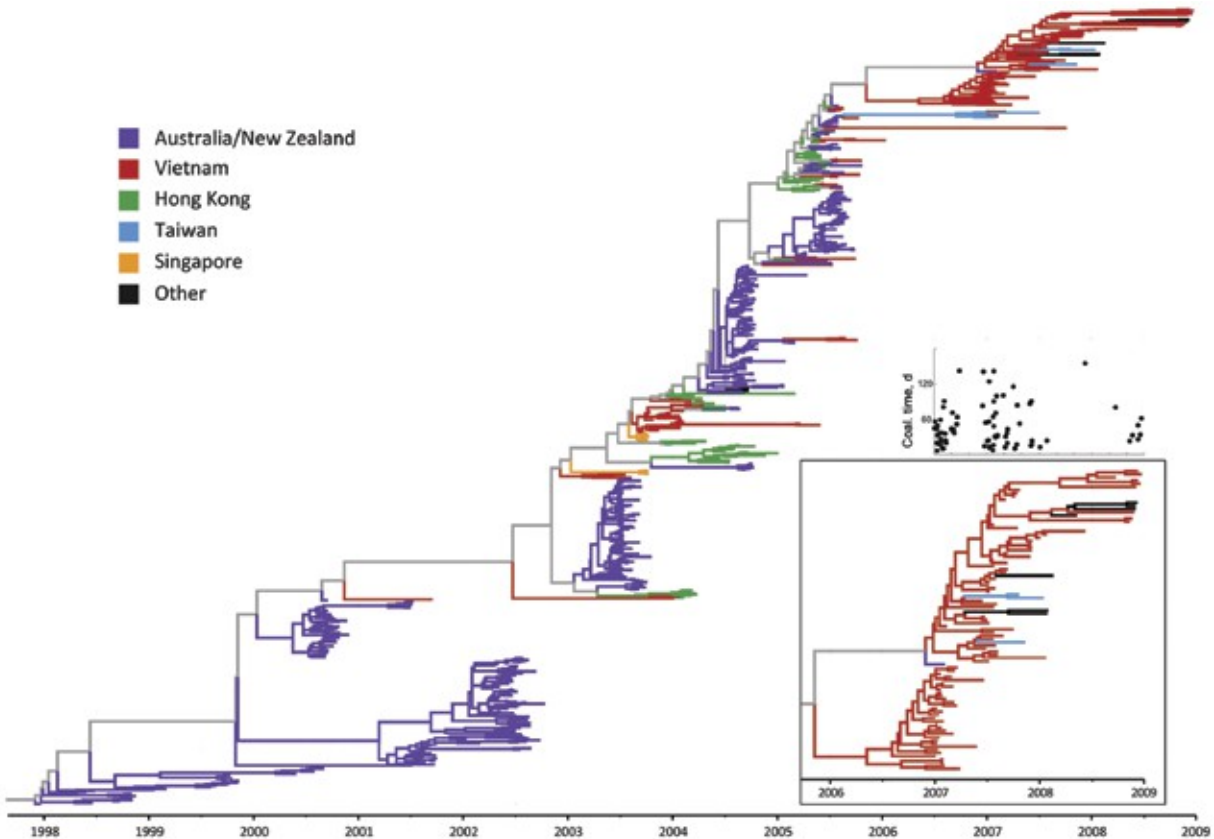


Hình 4. Khoảng cách tối thiểu của các nhóm virus (H3N2) đến thân cây phát sinh loài (tính từ 50 cây phân loài H3N2 toàn cầu). Virus được chia nhóm dựa vào năm (2003 – 2007) và vùng địa lý. Tên các nước được viết tắt: ANZ – Australia/New Zealand; VN – Việt Nam; HK – Hồng Kông; CN – Trung Quốc; JP – Nhật; US – Mỹ. Các đường màu đỏ thể hiện giá trị trung vị của khoảng cách đo được qua 50 cây. Phân tích này không được thực hiện với 2 nhóm virus Việt Nam – 2006 và Hồng Kông – 2007 do số lượng trình tự của các nhóm này quá ít.

Tính bền vững của các dòng cúm. Hình 5 cho thấy phát sinh loài H3N2 Bayes suy ra từ các dữ liệu trình tự khu vực theo thời gian. Những hình nhỏ trong Hình 5 chỉ ra chi tiết của phát sinh loài trong năm 2007-2008 (87 trình tự) cũng như những lần hợp lại của các đỉnh của những nhánh này. Rất khó để xác định đầy đủ tính bền vững của những chủng virus này vì lượng mẫu thu thập trong quý II năm 2007 và nửa đầu năm 2008 quá ít, mặc dù kết quả PCR thể hiện rõ bằng chứng về sự hoạt động cúm trong giai đoạn này (6). Thời gian phân nhánh trung bình của virus cúm ở Việt Nam trong giai đoạn này là 37 ngày (IQR: 21-72); các hình nhỏ trong Hình 5 cho thấy rằng có một dòng cúm tồn tại “bền vững” tại Việt Nam trong ít nhất 13 tháng (từ tháng 01/2007 đến tháng 01/2008). Do thiếu mẫu từ tháng 02/2008 đến tháng 05/2008 nên không thể kết luận liệu dòng này tồn tại ở Việt Nam trong thời gian bao lâu.

Để xác định liệu việc thiếu mẫu từ các nước khác có thể tạo ra một bức tranh sai lệch về tính bền vững của các chủng cúm ở Việt Nam trong năm 2007 đến năm 2008 hay không, chúng tôi tập hợp một bộ dữ liệu trình tự của toàn bộ các virus ở châu Á và ANZ từ năm 2006 đến 2008 ($n = 672$). Cây phân loài có độ tin cậy tối đa (maximum clade credibility tree - Hình S4) chỉ ra rằng các chủng của Việt Nam được phân chia thành hơn mười chủng khác nhau, khi được đặt cùng tất cả các chủng cúm của Châu Á / ANZ. Một trong những chủng này tồn tại ở Việt Nam trong 15 tháng (Hình S1 (a)), một chủng khác cũng tồn tại trong 10-12 tháng (Hình S1 (b)). Điều này cho thấy các chủng cúm có thể đã tồn tại bền vững (một năm hoặc hơn) ở Việt Nam trong khoảng

thời gian 2007 – 2008. Tuy nhiên, phương pháp phân tích này là rất nhạy cảm đối với các phát sinh loài không chắc chắn vì một số dòng virus thiếu số có quá ít mẫu nên một thay đổi nhỏ trên cấu trúc cây phân loài cũng có thể làm thay đổi vị trí của toàn bộ chủng trên cây phân loài.



Hình 5. Cây phân loài (phân đoạn hemagglutinin) của H3N2 thu được từ châu Á và châu Đại Dương (dữ liệu khu vực), được dựng bằng phần mềm BEAST phiên bản 1.6.2(26) dựa trên mô hình dân số không đổi. Các chuỗi trình tự đã dùng trong phân tích ở Hình 1 được dùng lại cho phân tích này, ngoại trừ 2 trình tự từ Việt Nam bị loại vì không xác định được ngày thu mẫu ($n = 785$). Cấu trúc phân loài của cúm ở Việt Nam trong năm 2007–2008 (được phóng đại ở góc dưới phải) cho thấy tính bền vững của virus trong suốt thời gian này. Đồ thị nhỏ (có trục hoành trùng với phần hình phóng đại) thể hiện thời gian phát sinh loài của các mẫu virus Việt Nam trong năm 2007–2008.

Thảo luận

Theo phân tích của chúng tôi, các tuyến di chuyển chính của cúm đi qua Hoa Kỳ, Đông Á, và Úc / New Zealand. Châu Âu - mặc dù mật độ dân số của nó và các trận dịch cúm luôn diễn ra vào mùa Đông – lại có ít kết nối đến những nơi khác trên thế giới hơn so với Hoa Kỳ. Kết quả này phù hợp với các nghiên cứu trước đây cho thấy khu vực Đông Á (2) và khu vực Châu Á nhiệt đới (1) là các nguồn cúm trọng điểm, và Hoa Kỳ như là một khu vực đóng góp chính (3). Các dữ kiện về trình tự mới trong phân tích này cho thấy các kết nối di cư mạnh mẽ giữa Việt Nam và các nước láng giềng, Mỹ, và châu Âu. Phân tích phát sinh loài trong khu vực của chúng tôi cho thấy một kết nối mạnh mẽ giữa Việt Nam và ANZ, nhưng các phân tích toàn cầu cho thấy trình tự ANZ lại liên quan chặt chẽ hơn với các trình tự từ các nước châu Á khác. Thêm vào đó, phân tích phát sinh loài cũng cung cấp bằng chứng về khả năng tồn tại ổn định hơn 1 năm (mà không bị thay thế) của virus cúm ở Việt Nam. Đây là một phát hiện quan trọng bởi các liên kết di chuyển và khả năng tồn tại bền vững của cúm là hai nhân tố chính để xác định nguồn phát tán virus. Việc tồn tại lâu dài của cúm ở khu vực nhiệt đới có thể có mối tương quan với sự tăng cường khả năng lẩn tránh miễn dịch của virus. Điều này có thể được chứng minh nếu ta có thể chỉ ra rằng việc tồn tại lâu dài hơn cũng đồng thời cho quần thể virus thêm thời gian để tích lũy đột biến và thay đổi kháng nguyên (2,31,32).

Nhìn chung, phân tích tính bền vững thật gặp khó khăn ngay cả khi việc lấy mẫu trình tự diễn ra thường xuyên và việc xác nhận virus được thực hiện hàng tuần. Để đánh giá tính bền vững của cúm trong một khu vực trọng điểm (như Việt Nam), chúng tôi phải lấy mẫu bên ngoài khu vực này để xác định xem virus địa phương thực sự bắt nguồn từ đâu. Tuy nhiên, khi càng nhiều mẫu được lấy từ khu vực bên ngoài, (i) khả năng tìm thấy virus (ở khu vực bên ngoài) đồng chủng với những dòng virus trong khu vực nghiên cứu càng cao, (ii) tính đa dạng sinh học của các khu vực bên ngoài cũng tăng lên. Cả hai điều này đều khiến cho liên kết di cư của virus giữa khu vực nghiên cứu với các khu vực bên ngoài trông có vẻ mạnh hơn. Không có tiêu chí rõ ràng để xác định lượng mẫu thích hợp cho khu vực nghiên cứu lẫn các khu vực ngoài nghiên cứu, do vậy, rất khó để khẳng định chắc chắn rằng một chủng cúm chỉ tồn tại “bền vững” ở một khu vực duy nhất. Đối với các trình tự cúm Việt Nam 2007 và 2008 (có thời gian thu mẫu trải khá đều và gần như phủ khắp 2007-2008), thời gian hợp nhất của những virus này trên cây phân loài tương đối ngắn, chỉ ra rằng hầu hết những loại virus này đều xuất hiện ở Việt Nam chưa lâu. Những dữ liệu này phù hợp và cũng là bằng chứng cho tính bền vững của các chủng virus ở Việt Nam trong

suốt thời gian này. Tuy nhiên, chúng tôi biết rằng không có kiểm nghiệm thống kê nào có thể bác bỏ hoàn toàn giả thiết xuất hiện chủng virus mới trong khoảng thời gian 2007-2008. Tập dữ liệu hoàn hảo để chứng minh tính bền vững của dòng virus có lẽ phải là 52 mẫu virus được thu trong 52 tuần liên tiếp, với các mẫu virus liên kế khác nhau ở 0 hoặc 1 vị trí nucleotide.

Phương pháp lấy mẫu. Một trong những hạn chế lớn của tất cả các phân tích di cư thực hiện với dữ liệu trình tự là độ lệch gây ra bởi lượng mẫu thu thập ở mỗi khu vực địa lý rất khác nhau. Càng nhiều trình tự để chỉ ra vị trí, thì càng nhiều khả năng là một trong những trình tự này nhập cư gần đây, và được nhận dạng bởi sự hiện diện của các trình tự tương tự từ các địa điểm khác. Để khắc phục sai lệch này, việc lấy mẫu con thường được sử dụng (3,5), để đảm bảo rằng các số lượng mẫu được sử dụng cho mỗi khu vực tương đối đồng nhất. Trong tình huống khi có quá ít trình tự từ một địa điểm cụ thể, thì sẽ chỉ có một số lượng nhỏ hơn các liên kết di cư có thể được suy ra từ vị trí đó. Sự sai lệch này không thể được sửa chữa bằng phương pháp thu mẫu con.

Phân tích của chúng tôi ở cấp độ toàn cầu trên các tập dữ liệu con cho thấy rằng số lượng mẫu và trọng số của các kết nối di cư có tính tương quan cao. Tuy nhiên, chúng tôi không thể xác định quan hệ nhân quả trong mối tương quan này. Một tín hiệu di chuyển có thể yếu vì thiếu mẫu. Ngược lại, số lượng nhỏ các mẫu có thể là kết quả của hoạt động cúm thấp và kết nối di cư yếu với các vùng khác. Để xác định nguyên nhân, hệ quả trong tương quan nói trên, một chiến lược lấy mẫu trình tự phải được đưa ra trong bối cảnh của một hệ thống giám sát cúm, sau đó các dữ liệu dịch tễ và dữ liệu trình tự phải được phân tích cùng nhau. Tỷ lệ nhiễm bệnh và dữ liệu trình tự nên được sử dụng đồng thời để xác định số lượng mẫu phù hợp cho từng vùng địa lý.

Mặc dù khá dễ để nhận ra những dấu hiệu của việc thu mẫu quá nhiều điều này dường như là một điểm rõ ràng về việc thu mẫu quá nhiều, đối âm là việc quá nhiều mẫu trong dữ liệu trình tự cúm xảy ra với một mức độ cao của sự sao chép giả. Phương pháp thu mẫu trình tự cúm hiện được dùng trong hầu hết các nghiên cứu khoa học và trong bối cảnh y tế cộng đồng thường dẫn đến việc mỗi mẫu được thu thêm, thay vì phải là một quan sát độc lập, lại có tính hệ số tương quan cao với những mẫu khác được thu cùng thời gian / cùng khu vực địa lý (33). Những mẫu “sao chép giả” này, về nguyên tắc, sẽ không tạo ra các sự kiện di chuyển nhân tạo bổ sung vào kết quả phân tích bởi cấu trúc của cây phát sinh loài không chịu nhiều ảnh hưởng từ mức độ độc lập lẫn nhau của mẫu trình tự.

Tuy nhiên, mối tương quan giữa số lượng mẫu và sức mạnh di chuyển vẫn tồn tại trong dữ liệu một phần, bởi vì khi số lượng mẫu tăng lên, xác suất một chủng virus mới được phát hiện cũng tăng theo.

Một cách tiếp cận mới cần được phát triển để có thể xem xét mọi biến số như không gian, sự tiến hóa, và dịch tễ học trong phân tích phát sinh loài. Các cách tiếp cận Bayesian đã được lựa chọn làm phương pháp tiêu chuẩn cho các nghiên cứu phát sinh loài theo vùng địa lý gần đây (1,3,4,34-37), chủ yếu là do khả năng của chúng để bù đắp cho sự không chắc chắn của các biến số tiến hóa, dân số và di cư, đặc biệt là bởi vì khả năng của chúng để kết hợp các nút hình học tô-pô không chắc chắn vào các phân tích phát sinh loài. Phương pháp Bayesian có thể phục vụ như một công cụ phụ trợ mạnh mẽ trong phân tích phát sinh loài theo vùng địa lý, cho phép xem xét độ lệch thống kê do quá trình lấy mẫu gây ảnh hưởng như thế nào đến từng vùng địa lý. Một yếu tố quan trọng nữa của phương pháp Bayesian trong nghiên cứu bệnh cúm là khả năng áp dụng của nó lên toàn bộ bộ gen trong các phân tích phát sinh loài theo vùng địa lý (1). Đối với các bộ dữ liệu có xác suất tái tổ hợp cao, sự hiện diện của tín hiệu di chuyển độc lập trong tám cây phát sinh loài khác nhau (cho tám đoạn RNA của virus) có thể được sử dụng để giảm sự không chắc chắn cho các tham số di cư được tìm ra.

Chúng tôi dự định làm sáng tỏ con đường di chuyển của cúm vào và ra khỏi Việt Nam, cũng như khả năng tồn tại bền vững của virus ở Việt Nam. Đối với cả hai mục tiêu, chúng tôi đề nghị rằng các nghiên cứu trong tương lai liên kết phân tích phát sinh loài với dữ liệu dịch tễ nhằm sửa chữa những thiên lệch thống kê đã nêu trên và cung cấp bằng chứng dịch tễ bổ sung quan trọng cho việc di chuyển và tính bền vững. Nếu mô hình nguồn – bể chứa thực sự là sự đơn giản hóa của mô hình lưu thông cúm toàn cầu (3-5), các liên kết chặt chẽ giữa Việt Nam với các nước khác ở châu Á, châu Âu và Hoa Kỳ cho thấy đất nước này có thể đóng cả hai vai trò trong những thời điểm khác nhau.

Acknowledgements

We would like to thank WHO Vietnam, the US CDC, Prof H. Suzuki and Prof R. Saito (Niigata University, Japan) for helping us establish influenza sentinel surveillance in Vietnam. Thanks to E. C. Holmes for critical review of this manuscript.

HML, PH, MFB are supported by the Wellcome Trust (089276/B/09/7, 098511/Z/12/Z, WT/093724). TTL is supported by a Royal Society Newton International Fellowship (UK). Computing and hardware costs were funded by the Li Ka Shing - University of Oxford Global Health Programme (LG05). This project has been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN272200900007C.

Author Bio: Dr Mai Quynh Le, MD PhD, is the head of the Virology Department in Vietnam's National Institute for Hygiene and Epidemiology. She obtained her MD from Vietnam's Military Academy of Medicine in 1989, and she has been leading research on virology, public health, and human and avian influenza viruses in Vietnam for more than 20 years.

References

1. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. *Nature* 2008; 453:615–620.
2. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, et al. The global circulation of seasonal influenza A (H3N2) viruses. *Science* 2008;320:340-346.
3. Bedford T, Cobey S, Beerli P, Pascual M. Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2). *PLoS Pathog* 2010; 6:e1000918.
4. Bahl J, Nelson MI, Chan KH, Chen R, Vijaykrishna D, Halpin RA, et al. Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proc Natl Acad Sci USA* 2011;108:19359-19364.
5. Cheng X, Tan Y, He M, Lam TT-Y, Lu X, Viboud C, et al. Epidemiological dynamics and phylogeography of influenza virus in southern china. *J Inf Dis* 2013; 207:106–14.
6. Nguyen HT, Dharan NJ, Le MTQ, Nguyen NB, Nguyen CT, Hoang DV, et al. National Influenza Surveillance in Vietnam 2006-2007. *Vaccine* 2009;28:398–402.
7. Oxford University Clinical Research Unit (2013) Influenza-like illness (ILI) in Ho Chi Minh City, www.ili.vn, accessed March 4, 2013.

8. Horby P, Mai LQ, Fox A, Thai PQ, Yen NTT, Thanh LT, et al. The Epidemiology of Interpandemic and Pandemic Influenza in Vietnam, 2007-2010: The Ha Nam Household Cohort Study I. *Am J Epidemiol* 2012;175:1062–1074.
9. Boni MF, Manh BH, Thai PQ, Farrar J, Hien TT, Hien NT, et al. Modelling the progression of pandemic influenza A(H1N1) and the opportunities for reassortment with other influenza viruses. *BMC Med* 2009;7:43.
10. Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, Thoang DD, et al. Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases. *PLoS One* 2011;6:e16965.
11. Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC. Phylogenetic Analysis Reveals the Global Migration of Seasonal Influenza A Viruses. *PLoS Pathog* 2007;3:e131.
12. Nelson MI, Simonsen L, Viboud C, Miller MA, Taylor J, St. George K, et al. Stochastic processes are key determinants of short-term evolution in influenza a virus. *PLoS Pathog* 2006;2:e125.
13. Tang JW, Ngai KKL, Lam WY, Chan PKS. Seasonality of influenza A (H3N2) virus: a Hong Kong Perspective (1997–2006). *PLoS One* 2008;3:e2768.
14. Doraisingham S, Goh KT, Ling AE, Yu M. Influenza surveillance in Singapore: 1972-86. *Bull World Health Org* 1988;66:57–63.
15. Simmerman JM, Uyeki TM. The burden of influenza in East and South-East Asia: a review of the English language literature. *Influenza and Other Respiratory Viruses* 2008;2:81–92.
16. Wong CM, Yang L, Chan KP, Leung GM, Chan KH, Guan Y, et al. Influenza-associated hospitalization in a subtropical city. *PLoS Med* 2006;3:e121.
17. Yap FHY, Ho P-L, Lam K-F, Chan PKS, Cheng Y-H, Peiris JSM. Excess hospital admissions for pneumonia, chronic obstructive pulmonary disease, and heart failure during influenza seasons in Hong Kong. *J Med Virol* 2004;73:617–23.
18. Shek LPC, Lee BW. Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric Resp Rev* 2003;4:105–111.
19. Finkelstein BS, Viboud C, Koelle K, Ferrari MJ, Bharti N, Grenfell BT. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS One* 2007;2:e1296.
20. Viboud C, Alonso WJ, Simonsen L. Influenza in tropical regions. *PLoS Med* 2006;3:e89.
21. Alonso WJ, Viboud C, Simonsen L, Hirano EW, Daufenbach LZ, Miller MA. Seasonality of

- influenza in Brazil: a traveling wave from the Amazon to the subtropics. *Am J Epidemiol* 2007;165:1434–42.
22. Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, et al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 2005;437:1162–6.
 23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004;32:1792–1797.
 24. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 2008;57:758–71.
 25. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;22:2688–2690.
 26. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7: 214.
 27. Swofford DL, Maddison WP. Reconstructing ancestral character states under Wagner parsimony. *Math Biosci* 1987;87:199–229.
 28. Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). 2003.
 29. Boni MF, Jong MD de, Doorn HR van, Holmes EC. Guidelines for identifying homologous recombination events in influenza A virus. *PLoS One* 2010;5:e10434.
 30. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media* 2009.
 31. Boni MF, Gog JR, Andreasen V, Feldman MW. Epidemic dynamics and antigenic evolution in a single season of influenza A. *Proc R Soc B* 2006;273:1307–1316.
 32. Adams B, McHardy AC. The impact of seasonal and year-round transmission regimes on the evolution of influenza A virus. *Proc R Soc B* 2011;278:2249–2256.
 33. Harvey PH, Pagel MD. The comparative method in evolutionary biology. Oxford, UK: Oxford University Press, 1991.
 34. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci USA* 2012;109:15066-15071.
 35. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009;5:e1000520.

36. Rabaa MA, Hang VTT, Wills B, Farrar J, Simmons CP, Holmes EC. Phylogeography of recently emerged DENV-2 in southern Viet Nam. *PLoS Negl Trop Dis* 2010;4:e766.
37. Raghwani J, Rambaut A, Holmes EC, Hang VT, Hien TT, Farrar J, et al. Endemic dengue associated with the co-circulation of multiple viral lineages and localized densitydependent transmission. *PLoS Pathog* 2011;7:e1002064.

Address for correspondence: Maciej F Boni, Oxford University Clinical Research Unit, Hospital for Tropical Diseases, 764 Vo Van Kiet Street, Q5, Ho Chi Minh City, Vietnam; email: mboni@oucru.org, tel: +84 8 3923 7954, fax: +1 509 562 2296;

