

## TÌM VỀ NGUỒN GỐC DÂN TỘC VIỆT NAM QUA DI TRUYỀN HỌC

## MỘT VÀI PHÁT HIỆN BAN ĐẦU VÀ ĐƯỜNG HƯỚNG NGHIÊN CỨU

NGUYỄN VĂN TUẤN(\*)



Người Việt Nam cuối thế kỷ XIX

Cách đây vài tháng, nhân điểm qua cuốn sách "Eden in the East" (tạm dịch: Địa đàng ở phương Đông) của Tác giả Stephen Oppenheimer, tôi có đặt lại vấn đề nguồn gốc dân tộc Việt Nam [1]. Bài điểm sách đã được nhiều bạn đọc quan tâm đến vấn đề ở trong và ngoài nước đón nhận một cách nồng nhiệt. Một số bạn đọc viết thư riêng đề nghị chúng tôi nên tiến hành một nghiên cứu sinh học để xác định rõ hơn mối quan hệ giữa dân tộc Việt và các dân tộc khác trong vùng Đông Nam Á châu. Một số bạn đọc khai triển thêm đề tài này và đề nghị nên cẩn thận trong khi diễn dịch các số liệu và dữ kiện khảo cổ học. Tuy nhiên bên cạnh đó, cũng có người cho

rằng không cần phải đặt lại vấn đề nguồn gốc dân tộc Việt, vì trong quá khứ đã có tác giả, chẳng hạn như Bình Nguyên Lộc, đã "chứng minh" nguồn gốc Mã Lai của dân tộc Việt [2].

Tuy nhiên, người viết bài này tin rằng một quan điểm mang tính khẳng định cao như thế rất nguy hiểm cho tiến bộ khoa học, vì nó hàm ý cho rằng câu chuyện đã chấm dứt, câu trả lời đã dứt khoát, không còn vấn đề gì để bàn thảo nữa, cũng như không có gì để phải nghiên cứu thêm. Thực ra, vấn đề [nguồn gốc dân tộc Việt] chưa được giải quyết thỏa đáng. Chưa ai chứng minh một cách khoa học rằng dân tộc Việt có nguồn gốc Mã Lai, và thậm chí câu hỏi vẫn chưa được đặt ra một cách có hệ thống. Tôi cho rằng vấn đề cần phải được đặt lại. Tại sao? Ngoài những dữ kiện trong sách "Eden in the East" của Oppenheimer mà tôi đã trình bày, tôi thấy còn có vài lý do quan trọng khác để chúng ta đặt lại vấn đề nguồn gốc dân tộc Việt.

- Thứ nhất, giả thuyết về nguồn gốc Mã Lai của dân tộc Việt chỉ là một giả thuyết trong nhiều giả thuyết khác nhau. Giả thuyết này, phần lớn, dựa vào các dữ kiện khảo cổ và ngôn ngữ. Những dữ kiện về đặc tính cơ thể và các chỉ số nhân trắc (như màu da, xương, sọ, khuôn mặt, v.v..) từng được dùng làm các đơn vị thông tin để nghiên cứu nguồn gốc dân tộc và sự tiến hóa của loài người. Nhưng các đặc tính này thay đổi theo thời gian, và chịu ảnh hưởng vào môi trường sinh

(\*)Tiến sĩ, Viện Nghiên cứu Y khoa Garvan Sydney, Australia

sống. Chẳng hạn như chiều cao của con người trong vòng 200 năm qua đã tăng một cách đáng kể do những cải thiện về dinh dưỡng và môi trường sinh sống. Ngay cả cấu trúc xương cũng thay đổi theo thời gian và môi trường. Do đó các đặc tính nêu trên không phải là những thông tin lý tưởng cho việc nghiên cứu lịch sử di truyền của con người.

Những dữ kiện về ngôn ngữ cũng có nhiều khiếm khuyết, vì mức độ tương đương về từ ngữ không thể nói lên một cách đầy đủ khuynh hướng di cư của các sắc dân. Ngay cả việc xác định mức độ tương đồng từ ngữ giữa các ngôn ngữ cũng là một vấn đề mang tính kỹ thuật mà các nhà nghiên cứu vẫn chưa đồng ý về phương pháp làm. Và lại, sự tương đồng giữa các ngôn ngữ có thể là một hằng số mang tính văn hóa, chứ không hẳn do các cơ chế sinh học và di truyền. Nói tóm lại, những bằng chứng khảo cổ và ngôn ngữ không phải là những loại thông số đáng tin cậy để xác định nguồn gốc dân tộc.

- Thứ hai, có thể nói ngay rằng các giả thuyết về nguồn gốc dân tộc Việt Nam của một số tác giả người Việt trước đây thiếu tính khoa học hay "bán khoa học" [3]. Những dữ kiện dùng để đặt giả thuyết trước đây chưa bao giờ được xử lý bằng cách phương pháp khoa học, không đánh giá được mức độ biến thiên của dữ kiện, nghĩa là định lượng xem mức độ biến chuyển của những dữ kiện đó có tỷ lệ bao nhiêu là do các yếu tố ngẫu nhiên gây nên, và bao nhiêu là do các yếu tố không ngẫu nhiên tạo thành.

Người thông minh có thể có nhiều cách giải thích rất hay cho những quan điểm sai lầm. Ai cũng có thể phát biểu một giả thuyết, nhưng một giả thuyết khoa học khác với một giả thuyết phi khoa học ở chỗ một giả thuyết khoa học có thể thử nghiệm được. Điều quan trọng là một giả thuyết khoa học cần phải được thử nghiệm một cách khách quan, và cần phải được "phản nghiệm" (falsify) một cách độc lập.

Vì thế, cần phải nói rõ rằng việc phát triển một giả thuyết khoa học không chỉ đơn thuần gói gọn trong việc in vài ý kiến thành

sách và cho đó là "khoa học", là những giả thuyết mà ai cũng cần phải đọc, cần phải ghi nhận. Stephen Jay Gould (nhà tiến hóa học danh tiếng vừa mới qua đời), trước khi phát biểu một giả thuyết về sinh học tiến hóa đã từng làm nhiều nghiên cứu và từng công bố kết quả nghiên cứu trong hàng trăm bài báo khoa học trên các tập san chuyên môn. Stephen Oppenheimer cũng thế: trước khi cho in những giả thuyết của mình thành sách, ông ta đã phải bỏ ra hàng chục năm đi thu thập dữ kiện, thử nghiệm, phân tích, và công bố nhiều bài báo khoa học trên các tập san chuyên môn danh tiếng trên thế giới. Nhìn lại trường hợp Việt Nam, chúng ta có thể nói không một giả thuyết nào về nguồn gốc dân tộc Việt Nam được kiểm tra bởi các đồng nghiệp chuyên môn trên thế giới (peer review), hay được công bố trên các tập san hàn lâm trên thế giới, hay dựa vào những bằng chứng chặt chẽ về sinh học. Thành ra, chưa một giả thuyết nào của các tác giả Việt Nam được trình bày một cách có hệ thống và được đồng nghiệp trên thế giới công nhận.



Trống đồng Đông Sơn của Việt Nam

- Thứ ba, như đã nói trên nguồn gốc Mã Lai của dân tộc Việt Nam mà Bình Nguyên Lộc đề xướng là một giả thuyết. Không có ai có thể chứng minh một giả thuyết. Những gì mà người ta có thể "chứng minh" là các dữ kiện có nhất quán hay phù hợp với giả thuyết hay không. Khi có dữ kiện

mới giả thuyết sẽ được phát biểu lại. Do đó, cần phải nhấn mạnh rằng không có một giả thuyết nào được xem là bất biến. Ngay cả giả thuyết về nguồn gốc con người hiện đại cũng đã trải qua nhiều nghiên cứu, thảo luận hàng nửa thế kỷ nay, với hàng ngàn bài báo khoa học và hàng trăm cuốn sách đã viết về đề tài này. Điều này cũng không nên lấy làm ngạc nhiên, bởi vì quá trình phát triển khoa học nằm trong một chu trình đặt vấn đề hay giả thuyết, thu thập dữ kiện, thử nghiệm giả thuyết, và đặt lại vấn đề. Sự tích lũy dữ kiện theo thời gian sẽ soi sáng và góp phần kiểm tra lại giả thuyết ban đầu. Thành ra, cần phải đặt lại vấn đề, cần phải nêu những câu hỏi tốt để đi tìm những câu trả lời thích hợp và hữu dụng.

Trong quá khứ, giới khảo cổ học và nhân chủng học dựa vào các bằng chứng về khảo cổ, xương, hóa thạch, v.v... để phát triển lý thuyết, nhưng những đối tượng này hàm chứa nhiều hạn chế thông tin về tiến hóa, vì mối quan hệ phức tạp giữa môi trường và tiến hóa. Hậu quả của sự tập trung vào các đối tượng như thế trong một thời gian dài đã làm cho chúng ta xao lãng các dữ kiện cho chúng ta nhiều thông tin hơn: đó là gen. Không giống như xương sọ, những thay đổi trong gen thường xảy ra theo những qui luật mà chúng ta hiểu khá rõ, và vì thế gen và các đặc điểm của gen, như tần số gen, cấu trúc DNA, phân phối gen, v.v... cho chúng ta những thông tin cực kỳ quý giá về sự tiến hóa của con người. Ngày nay, những tiến bộ phi thường trong ngành di truyền học và sinh học phân tử (molecular biology) trong mấy năm gần đây đã cung cấp cho ngành nhân chủng học một phương tiện cực kỳ quan trọng trong việc xác định lịch sử tiến hóa của con người và mối liên hệ giữa các dân tộc. Nhưng trong khi khả năng và kỹ thuật càng ngày càng hiện đại thì những thay đổi trong xã hội trong các nước đang phát triển có nguy cơ làm nhòa đi tính "tinh khiết" của những người dân chính gốc ở Việt Nam. Vì thế, đặt vấn đề nguồn gốc dân tộc cũng có nghĩa là tìm một sự phối hợp có tổ chức để thu thập và duy trì những thông tin sinh học về quá khứ của chúng ta, những người Việt Nam.

Giá trị của di truyền học trong việc truy tìm nguồn gốc dân tộc đã được đánh giá cao về mức độ tin cậy. Di truyền học là một cửa sổ để chúng ta nhìn lại quá khứ của chúng ta. Tuy nhiên, những nghiên cứu về di truyền học trong người Việt còn cực kỳ khiêm tốn, nếu không muốn nói là không đáng kể. Trong thời gian khoảng 10 năm trở lại đây, đã có một số bằng chứng, tuy gián tiếp, nhưng cũng đủ để chúng ta có lý do để xem xét lại lịch sử tiến hóa của dân tộc Việt. Những bằng chứng này là:

(a) Năm 1998, Giáo sư Chu và đồng nghiệp (thuộc Trường Đại học Texas) phân tích 15 đến 30 mẫu "vi vệ tinh" DNA (microsatellites) để thử nghiệm sự khác biệt di truyền trong 24 nhóm dân từ nhiều tỉnh khác nhau ở Trung Quốc, 4 nhóm dân vùng Đông Nam Á (hai thuộc thổ dân Mỹ, một thuộc thổ dân Úc châu, và một thuộc Tân Guinea), 4 nhóm dân da trắng (Caucasian), và 3 nhóm dân Phi châu. Bằng một phương pháp phân tích thống kê có tên là "phân tích phát sinh chủng loại" (Phylogenetic analysis)", một số kết quả đáng ghi nhận như sau:

+ Hai nhóm dân có sự khác biệt rõ ràng nhất là Phi châu và các dân không thuộc Phi châu;

+ Tất cả các nhóm dân Đông Nam Á "tập hợp" thành một nhóm, và nhóm dân có đặc tính di truyền gần họ nhất là người thổ dân Mỹ châu, kế đến là thổ dân Úc châu, và Tân Guinea (Những kết quả này cũng phù hợp với thời gian định cư ở Úc châu (khoảng 60,000 đến 50,000 năm trước đây, và thời gian định cư ở Mỹ châu (từ 30,000 đến 15,000 năm trước đây);

+ Các nhóm dân miền nam Trung Quốc phân phối thành ba nhóm, gọi là S1, S2, và S3 (ngoại trừ nhóm S2 là người Hán từ tỉnh Henan, phần còn lại (S1 và S3) gồm các sắc dân trong vùng Yunnan); và

+ Các sắc dân miền bắc Trung Quốc phân phối thành hai nhóm, gọi là N1 và N2. Nhóm N1 gồm 6 sắc dân nói tiếng Altaic, một nhóm Hán tộc miền bắc từ tỉnh Yunnan.

Nhóm N2 gồm 4 sắc tộc thiểu số có lịch sử sinh sống lâu đời ở miền bắc, trong đó có một sắc tộc từ tỉnh Ninxia [4].

Từ những phát hiện trên, chúng ta có thể đặt ra một số mô hình để giải thích [5], nhưng mô hình thích hợp với dữ kiện của Giáo sư Chu và đồng nghiệp là các dân tộc miền Bắc Á được tiến hóa từ các dân tộc Đông Nam Á châu. Các dữ kiện liên quan đến răng, sọ [6,7] cũng nhất quán với mô hình này. Do đó, Giáo sư Chu và đồng nghiệp kết luận rằng: "Tổ tiên của các nhóm dân Đông Á ngày nay có nguồn gốc từ Đông Nam Á." Nói một cách khác, các dữ kiện di truyền học của Giáo sư Chu và đồng nghiệp cho thấy tổ tiên của những người nói tiếng Altaic từ Đông Á đã di cư vào Á châu từ ngã Đông Nam chứ không phải từ ngã Trung Á.

(b) Nhưng nghiên cứu của Giáo sư Chu và đồng nghiệp có một điểm yếu, đó là họ dựa vào vi vệ tinh DNA, một chất liệu di truyền rất "nhạy" (sensitive) và dễ bị đột biến (mutation [8]). Để khắc phục nhược điểm này, một nhóm nghiên cứu khác đã tiến hành một nghiên cứu độc lập và qui mô hơn để xác định nguồn gốc Đông Nam Á của dân tộc Trung Hoa. Nhóm nghiên cứu Mỹ - Trung Quốc phân tích DNA trong nhiễm sắc thể Y [9] trong các nhóm dân Hán (thuộc 22 tỉnh của Trung Quốc), 3 nhóm dân Đông Bắc Á (Buryat, Hàn Quốc, và Nhật Bản), 5 nhóm dân Đông Nam Á (Campuchia, Thái Lan, Malaysia, Batak, và Java), và 12 nhóm dân ngoài Á châu (3 nhóm từ Phi châu, 3 từ Mỹ châu, 2 từ Âu châu, và 4 từ châu Đại dương). Các nhà nghiên cứu ghi nhận rằng mức độ biến thiên đa hình thái (polymorphic variation) trong các nhóm dân Đông Nam Á cao hơn trong các nhóm dân thuộc vùng Bắc Á. Điều này có nghĩa là các sắc dân ở Đông Nam Á có một quá trình định cư lâu dài hơn là các nhóm dân Bắc Á. Dùng các phương pháp phân tích di truyền quần thể (population genetics), các nhà nghiên cứu kết luận rằng con người thời đó đã di cư từ Phi châu sang đến Đông Nam Á [10] vào khoảng 60 ngàn năm về trước, và sau đó đã di chuyển lên phía Bắc Á (kể cả Trung Quốc ngày nay) và Siberia [11]. Ngoài ra, còn có bằng chứng di

truyền cho thấy các nhóm dân Polynesians cũng có nguồn gốc từ Đông Nam Á [12].

(c) Trong một nghiên cứu trên 103 người ở Hà Nội [13], các nhà nghiên cứu Việt - Pháp phân tích DNA trong hai gien (HLA-DR và DQB1), và so sánh kết quả này với các sắc dân thuộc châu Đại Dương (Oceania) và Đông Á. Sau khi ước tính khoảng cách di truyền [14] giữa các sắc dân, các nhà nghiên cứu kết luận rằng cấu trúc di truyền của hai gien này trong người Việt gần với người Thái và người Hoa. Các nhà nghiên cứu này kết luận rằng dữ kiện của họ phù hợp với giả thuyết người Việt có nguồn gốc từ người Trung Hoa và Thái-Nam Dương.

Tuy nhiên, sau khi xem xét kỹ phần phương pháp nghiên cứu, tôi có lý do để cho rằng kết luận này rất có thể không đúng. Những lý do này là: Thứ nhất, nghiên cứu này chỉ dựa vào hai gien mà thôi (con người có khoảng 35 đến 39 ngàn gien), và với những yếu điểm của mtDNA, các ước đoán về khoảng cách di truyền không ổn định. Thứ hai, ngay cả trong bài báo các nhà nghiên cứu không xây dựng được một cây di truyền nào, và cũng chẳng phân tích phát sinh chủng loại thì không thể phán đoán về chiều hướng di cư hay nguồn gốc dân tộc được.

(d) Khoảng 2 năm sau, các nhà nghiên cứu này lại tiến hành một nghiên cứu khác trên 50 người cũng ở Hà Nội [15], và cũng qua dùng kỹ thuật PCR (polymerase chain reaction) họ phân tích 6 "restriction enzymes", và ghi nhận khoảng cách di truyền giữa người Việt và người Hoa rất thấp (chỉ 0.0022), nhưng giữa người Việt và Ấn Độ thì tương đối cao hơn (0.0468), có nghĩa là quan hệ giữa giống người Việt và Hoa gần nhau hơn so với quan hệ Việt và Ấn.

Nghiên cứu này cũng có những yếu điểm như nghiên cứu trình bày phần (c), tức là số lượng gien quá ít (trong trường hợp này chỉ có một gien), và tác giả cũng không tính toán mức độ biến thiên của chỉ số khoảng cách di truyền, nên không thể nào phát biểu khoảng cách giữa Việt - Hoa gần hơn khoảng cách giữa Việt - Ấn. Thực ra, sau khi tính toán lại, tôi thấy hai khoảng cách di truyền (Việt -

Hoa và Việt - Ấn) không có sự khác biệt đáng kể (non-significant)!



Người Hoa ở Việt Nam đầu thế kỷ XX

(e) Tuy nhiên, một nghiên cứu khác trên 5 gien trong nhiễm sắc thể Y trong hai nhóm dân: Bắc Á (Bắc Trung Quốc, Nhật, Hàn Quốc, và Mông Cổ), và Nam Á (Indonesia, Philippines, Thái Lan, và Việt Nam) cho thấy người Việt gần với các nhóm dân Bắc Á (nhất là Hàn Quốc) hơn là các nhóm dân Nam Á [16]! Chúng ta biết rằng, qua nghiên cứu của Giáo sư Chu, người Hoa phía Bắc Trung Quốc có cấu trúc di truyền khác với người Hoa phía Nam Trung Quốc. Do đó, phát hiện này quả rất khó giải thích. Càng khó giải thích hơn nữa khi phần lớn những người Việt trong nghiên cứu này là cư dân ở Hà Nội, tức gần miền Nam Trung Quốc.

(f) Trong một nghiên cứu dùng mtDNA, Ballinger và đồng nghiệp [17] ghi nhận rằng chỉ số biến thiên (trong di truyền học gọi là F-value) trong người Việt cao nhất trong các sắc dân vùng Đông Nam Á. Tuy nhiên, các tác giả kết luận một cách mơ hồ rằng các dữ kiện này cho thấy "người Á châu có nguồn gốc từ nhóm dân Nam Mông" (nguyên văn: "The greatest mtDNA diversity and the highest frequency of mtDNA with HpaI/HincII morph 1 were observed in the Vietnamese suggesting a Southern Mongoloid origin of Asians").

Thực ra, các dữ kiện mà tác giả trình bày không cho phép họ kết luận như thế, bởi vì họ chỉ đưa ra 7 nhóm dân Á châu mà thôi, và cũng chỉ nghiên cứu trên vài mẫu gien rất nhỏ. Nhưng qua số liệu của các nhà nghiên cứu này, chúng ta có thể phát biểu rằng trong hai gien mà họ nghiên cứu, người Việt có lẽ là một sắc dân cổ nhất trong vùng Đông Nam Á.

Qua các nghiên cứu di truyền dựa vào nhiễm sắc thể Y và mtDNA mà tôi vừa tóm lược và nhận xét, chúng ta có thể rút ra một vài điểm chung như sau:

+ Thứ nhất, các nhóm dân Đông Á có cùng một nguồn gốc chung, đó là tất cả đều xuất phát từ Phi châu. Cũng dựa theo các dữ kiện mới này, các nhà nghiên cứu ước đoán rằng đợt người đầu tiên di cư đến Đông Nam Á xảy ra vào khoảng 18.000 đến 60.000 năm trước đây, và sau đó từ đây một đợt di cư về phía Bắc. Một đợt khác cũng từ Đông Nam Á di cư sang các quần đảo Thái Bình Dương qua ngã Mã Lai Á ngày nay. Tất nhiên, đây chỉ là "câu chuyện" mới được phát họa, nhiều chi tiết vẫn còn trong vòng nghiên cứu thêm.

+ Thứ hai, vấn đề nguồn gốc dân tộc Việt vẫn còn là một bí mật chưa được khai thác. Tất cả các nghiên cứu qui mô của họ đều không có dữ kiện gien của người Việt. Ngược lại những nghiên cứu có dữ kiện của người Việt thì lại chưa được tiến hành có hệ thống và qui mô, nếu không muốn nói là còn rất hạn chế. Vì thế, các nghiên cứu gần đây hoàn toàn không cho phép chúng ta phát biểu gì về nguồn gốc Mã Lai của dân tộc Việt như có người đặt giả thuyết.

+ Thứ ba, những kết quả của các nghiên cứu này đã cung cấp thông tin cho chúng ta để phát biểu một giả thuyết mới, rằng người Hoa ngày nay có thể xuất phát từ phía Nam, mà cụ thể hơn là Đông Nam Á; và có bằng chứng [tuy chưa đầy đủ] cho thấy người Việt có thể là một dân tộc cổ nhất trong vùng Đông Nam Á; cho nên có thể người Việt cũng có thể chính là tổ tiên của người Hoa ngày nay, nhất là các sắc dân ở miền Nam Trung Quốc. Cố nhiên, đó chỉ là một giả thuyết cần được thử nghiệm.

Do đó, vấn đề quan trọng trước mắt là cần xác định mối liên hệ di truyền (genetic relationships) giữa các nhóm dân trong vùng Đông Nam Á (và Việt Nam), các nhóm dân miền Nam và Bắc Trung Quốc, cũng như các nhóm dân thuộc Nam Đảo, Polynesians, v.v... Với những kết quả này cộng với các dữ kiện trình bày trong kho tàng khảo cổ và nhân chủng học cũng như trong cuốn "Eden in the East" chúng ta sẽ có một phối cảnh rõ ràng hơn về quá khứ. Hiện nay, chúng tôi đang xúc tiến một nghiên cứu di truyền học và cộng tác với các nhà nghiên cứu khác trên thế giới để nhằm tìm một câu trả lời cho vấn đề này. Chúng tôi dự tính sẽ thu thập và phân tích dữ kiện di truyền (chủ yếu qua phân tích DNA với khoảng 50 gen và các nhiễm sắc thể Y) trong khoảng 200 đến 300 người Việt, và sẽ dùng kết quả này để so sánh với các dân tộc khác trong vùng Đông Nam Á cũng như người Hoa để xác định xu hướng di cư của các nhóm dân này. Chúng tôi sẽ dùng những dữ kiện DNA để tiến hành một phân

tích phát sinh chủng loại nhằm thử giả thuyết giả thuyết "Bắc tiến" (người Hoa có nguồn gốc từ phương Nam) hay "Nam tiến" (người phương Nam xuất phát từ người Hoa). Hy vọng trong tương lai, chúng tôi sẽ có các dữ kiện di truyền học để đi đến một so sánh có ý nghĩa và có thể phát biểu một cách có cơ sở khoa học hơn.

Nguồn gốc con người hiện đại, mà đặc biệt là nguồn gốc dân tộc Việt, là một câu hỏi đơn giản, nhưng câu trả lời thì không đơn giản chút nào. Tổ tiên chúng ta xuất phát từ đâu, họ đến Việt Nam bằng cách nào, vẫn còn là những vấn đề khoa học "nóng", đòi hỏi nhiều nghiên cứu về di truyền học, khảo cổ học, và ngôn ngữ học. Tuy nhiên, với sự tiến bộ phi thường về khoa học di truyền và công nghệ sinh học trong khoảng mười năm qua, chúng tôi tin rằng việc nghiên cứu quá trình lịch sử di truyền của người Việt chúng ta sẽ đem lại nhiều kết quả thích thú và làm sáng tỏ hơn về nguồn gốc tổ tiên của chúng ta.

### **Chú thích và tài liệu tham khảo**

[1] Xem "Địa đàng ở phương Đông", Tập san Khoa học số 3(4)-2005, Đại học Mở Bán công TP.HCM.

[2] Trong Eden in the East, Stephen Oppenheimer không đưa ra một giả thuyết Nam tiến hay Bắc tiến nào cả, ông cũng chẳng đưa ra một giả thuyết nào về nguồn gốc dân tộc Việt Nam. Những luận án chính của Oppenheimer là: (a) Đông Nam Á là cái nôi của văn minh nhân loại ngày nay; (b) Những dân tộc thuộc quần đảo Polynesian không phải xuất phát từ Trung Quốc, mà có nguồn gốc từ Đông Nam Á; và (c) Người Đông Nam Á, chứ không phải Trung Hoa, là những người đã phát triển kỹ thuật trồng lúa đầu tiên trên thế giới, và là những nhà canh nông chuyên nghiệp đầu tiên của nhân loại. Dựa vào một số kết quả nghiên cứu gần đây và luận điểm của Stephen Oppenheimer, tôi muốn đặt lại vấn đề nguồn gốc Trung Hoa của dân tộc Việt Nam. Cần phải nói thêm rằng đặt lại vấn đề nguồn gốc dân tộc không phải vì những chuyện nhằm nhí như "lo xa", hay trẻ con như "bực bội", hay để, một cách cải lương, khôi phục "tự hào dân tộc", mà để đi tìm sự thật, và để đi tìm phương pháp cho một câu trả lời. Câu trả lời chỉ có ý nghĩa nếu câu hỏi được suy nghĩ cẩn thận.

[3] Thế nào là "khoa học"? Nói một cách ngắn gọn và theo quan điểm của người viết bài này, khoa học là tập hợp những phương pháp trí tuệ và phương pháp thực tiễn dùng để diễn tả và giải thích những hiện tượng quan sát được hay suy đoán được, trong quá khứ hay hiện tại, nhằm xây dựng một hệ thống tri thức mà chúng ta có thể bác bỏ hay xác nhận được. Những phương pháp trí tuệ bao gồm linh cảm, ước đoán, tư tưởng, giả thuyết, lý thuyết, và mô hình. Những phương pháp thực tiễn bao gồm bối cảnh của nghiên cứu, quá trình thu thập dữ kiện, hệ thống hóa dữ kiện, cộng tác cùng đồng nghiệp, thí nghiệm, tìm tòi và khám phá, phân tích thống kê, viết báo cáo khoa học, trình bày trong các hội nghị khoa học, và công bố kết quả nghiên cứu.

Khoa học dựa vào sự thật (facts). Những sự thật phải được quan sát hay thu thập bằng những phương pháp khách quan. Điều quan trọng và cần thiết trong khoa học là không chỉ sự thật, mà là sự thật có liên quan đến vấn đề đang được điều tra. Những sự thật này phải được thu thập (i) có tổ chức, trực tiếp, và khách quan; (ii) độc lập với lý thuyết; và (iii) một cách tin cậy để làm nền tảng cho suy luận. Một đặc tính quan trọng trong phát biểu trên về khoa học là những lý thuyết của khoa học có thể bác bỏ được. Cụm từ "có thể bác bỏ" ở đây cần phải hiểu một cách rộng rãi hơn, rằng tất cả các lý thuyết, nếu mang danh nghĩa khoa học, là những thuyết có thể thử nghiệm được, có thể tái xác nhận được bằng dữ kiện mới và độc lập. Và chính đặc tính này phân biệt rành mạch sự khác biệt giữa khoa học và ngụy khoa học (pseudoscience). Trong ngụy khoa học, các lý thuyết, chẳng hạn như thuyết tạo hóa (creationism) trong tôn giáo, đều không thể kiểm chứng hay thử nghiệm được. Nói một cách khác, một trong những tiêu chuẩn cho một lý thuyết khoa học là "khả năng phản nghiệm" (falsibility) của nó. (Có thể xem thêm vấn đề "phản nghiệm" này trong sách "The logic of scientific discovery," Karl Popper, London: Hutchinson, 1972).

[4] Xem bài "Genetic relationship of populations in China," tác giả J. Y. Chu và đồng nghiệp, Tập san Proceedings of the National Academy of Science (USA) 1998; số 95, trang 11763-11768.

[5] Chẳng hạn như mô hình thứ nhất: các nhóm dân miền Đông Bắc di cư xuống miền Nam, và hòa hợp với các nhóm dân Nam Á (Australoids, tức là người thổ dân Úc và Tân Guinea), những người đã từng định cư tại vùng Đông Nam Á châu trước đây. Theo mô hình này, có một nhóm người "gây giống" hiện diện ở Đông Nam Á khoảng 30 ngàn đến 50 ngàn năm trước đây. Mô hình này cũng phù hợp với bản đồ phân phối các ngôn ngữ ở Trung Quốc. Từ bản đồ này, chúng ta có thể phỏng đoán rằng miền bắc Trung Quốc từng được các dân tộc nói tiếng Trung Hoa và các ngôn ngữ Trung Hoa - Tây Tạng (Sino-Tibetan languages) định cư, còn miền nam Trung Quốc thì được các dân tộc nói tiếng Nam Á (Austro-Asiatic languages), Austronesian, Miao-Yao, và Daic định cư. Sau đó (khoảng 3.000 năm trước đây, tức dưới triều đại Zhao) ngôn ngữ Trung Hoa - Tây Tạng dần dần thay thế các ngôn ngữ khác ở phía nam Trung Quốc, cũng giống như một quá trình truyền bá của các ngôn ngữ Ấn-Âu ở Âu châu.

Hay mô hình thứ hai: các nhóm dân miền bắc và nam Trung Quốc tiến hóa cùng một lượt kể từ thời đại Pleistocene (tức khoảng 10.000 năm trước đây). Mô hình này phù hợp với các dữ kiện khảo cổ học, và theo đó, có ít nhất là ba trung tâm nông nghiệp và các nền văn hóa Đồ đá mới cùng hình thành ở Trung Quốc. Đây là các văn hóa: (i) Yangshao (miền bắc, và miền trung sông Hoàng) khoảng 8.000 năm trước đây, nơi mà cây kê là một cây trồng phổ biến nhất, lợn, chó được thuần hóa; (ii) Văn hóa Ta-Pen-Keng, cũng hình thành vào khoảng 8.000 năm trước đây dọc theo bờ biển phía nam, kể cả Đài Loan, nơi mà sản phẩm chính là lúa gạo; (iii) Văn hóa Qing-Lien-Kang, hình thành vào khoảng 1.000 năm trước đây, trong vùng hạ lưu sông Dương Tử, nơi mà cây lúa được trồng và thú vật như chó, lợn được thuần hóa.

[6] Xem bài "Major features of Sundadonty and Sinodonty, including suggestions about East Asian microevolution, population history, and late Pleistocene relationships with Australian aboriginals", C. G. Turner, Tập san American Journal of Physical Anthropology, năm 1990; bộ 82, trang 295-317.

[7] Xem bài "Population prehistory of east Asia and the Pacific as viewed from cranio-facial morphology: the basic populations in east Asia, VII" T. Hanihara, American Journal of Physical Anthropology, năm 1990, năm 1993, bộ 91, trang 173-87.

[8] Đột biến (mutation) là một sự kiện sinh học xảy ra ở trong tế bào. Gien được cấu trúc bằng một chuỗi DNA gồm 4 mẫu tự A, G, C, T. Khi một chuỗi DNA bị thay đổi, tức đột biến (chẳng hạn như từ GCAATGGCCC thành GCAACGGCCC) thì các đặc tính sinh học liên quan đến gien, chẳng hạn như mật độ xương, có thể bị thay đổi.

[9] Con người được cấu tạo bằng nhiều tỷ tế bào. Tất cả các tế bào đều có cấu trúc giống nhau: trong mỗi tế bào đều có một cái nhân (nucleus) nằm chính giữa. Cái nhân này có chứa những chất liệu di truyền mà ta thường gọi là DNA (viết tắt từ chữ deoxyribonucleic acid). Mỗi nhân thường có hàng triệu DNA. DNA gồm có bốn mẫu tự (yếu tố hóa học): A (adeline), C (cytosine), G (guanine), và T (thymine). Một mảng DNA tạo thành một gien. Và nhiều gien tạo thành một bộ di truyền hay nhiễm sắc thể, còn gọi là chromosome. Con người có 23 nhiễm sắc thể.

Không như DNA nằm trong nhân của tế bào, Mitochondrial DNA (mtDNA) nằm ngoài nhân của tế bào. MtDNA chỉ lưu truyền trong giống cái (ví dụ như từ mẹ sang con gái) do đó chúng rất có ích trong việc truy tìm những mối quan hệ di truyền giữa các nhóm dân. Nhưng mtDNA có hai nhược điểm: (i) mtDNA có nhiễm sắc rất nhỏ, chỉ khoảng 16.500 mẫu tự (tức chỉ khoảng 16 kilobases); và (ii) sự phân phối của các mẫu tự cũng không đồng đều.

Nhiễm sắc thể Y (hay "Y chromosome") là một đối tượng lý tưởng để truy tìm nguồn gốc dân tộc, bởi vì nó không có những nhược điểm như mtDNA. Ngược lại với mtDNA, nhiễm sắc thể Y chỉ lưu truyền trong giống đực. Nhiễm sắc thể Y chứa nhiều mẫu tự hơn mtDNA (khoảng 60 megebases, tức 4000 lần cao hơn mtDNA), và do đó cho chúng ta nhiều thông tin về di truyền hơn là mtDNA. Ngoài ra, khi nhiễm sắc thể Y được phân tích bằng tập hợp nhiều kiểu di truyền (tức "genotypes") thành "haplotype", các nhà nghiên cứu có thể biết được "dấu vết di cư" của một nhóm dân, và qua đó có thể ước đoán được tuổi của tổ tiên. Tuy nhiên, nhiễm sắc thể Y cũng có một vài nhược điểm, nhất là phân phối của haplotypes có thể bị ảnh hưởng bởi khả năng tái sản sinh của giống đực.

[10] Xem bài "African origin of modern humans in East Asia: a tale of 12000 Y chromosomes," tác giả Yuehai Ke và đồng nghiệp, [Tập san] Science, năm 2001, bộ 292, trang 1151-1153.

[11] Xem bài "Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last Ice Age," tác giả Bing Su và đồng nghiệp, [Tập san] American Journal of Human Genetics, năm 1999, bộ 65, trang 1718-1724.

[12] Xem bài báo khoa học "Polynesian origins: insights from the Y chromosome," tác giả Bing Su và đồng nghiệp, Tập san Proceedings of the National Academy of Science (USA), năm 2000, bộ 97, trang 8225-8228.

[13] Xem bài "HLA-DR and DQB1 DNA polymorphisms in a Vietnamese Kinh population in Ha Noi", tác giả A. Vu-Trieu và đồng nghiệp, Tập san European Journal of Immunogenetics, năm 1997, bộ 24, trang 345-356.

[14] Khoảng cách di truyền (hay còn gọi là genetic distance) là một thông số di truyền học đo lường mức độ khác nhau giữa các nhóm dân. Chỉ số này có giá trị tối thiểu là 0 (tức hai nhóm dân giống nhau về mặt di truyền), và giá trị tối đa là 1 (tức hai nhóm dân hoàn toàn khác nhau).

[15] Xem bài "Mitochondrial DNA polymorphism in the Vietnamese population", tác giả

R. Ivanova và đồng nghiệp, Tập san European Journal of Immunogenetics, năm 1999, bộ 26, trang 417-422.

[16] Xem bài "Y chromosomal DNA variation in East Asian populations and its potential for inferring the peopling of Korea", tác giả W. Kim và đồng nghiệp, Tập san Journal of Human Genetics, năm 2000; bộ 45, trang 76-83.

[17] Xem bài "Southeast Asian mitochondrialDNA analysis reveals genetic continuity of ancient Mongoloid migration," tác giả S. W. Ballinger và đồng nghiệp, Tập san Genetics, năm 1992, bộ 130, trang 139-45.

### **TÓM TẮT**

Người thông minh có thể có nhiều cách giải thích rất hay cho những quan điểm sai lầm. Ai cũng có thể phát biểu một giả thuyết, nhưng một giả thuyết khoa học khác với một giả thuyết phi khoa học ở chỗ một giả thuyết khoa học có thể thử nghiệm được. Điều quan trọng là một giả thuyết khoa học cần phải được thử nghiệm một cách khách quan, và cần phải được "phản nghiệm" (falsify) một cách độc lập.

### **SUMMARY**

An intelligent person can have many ingenious ways to defend for erroneous viewpoints. Anyone can talk about an assumption, but a scientific assumption is different from a non-scientific one in that the former can be tested. The important thing is that a scientific assumption needs to be objectively tested and "falsified" independently.