

## SỰ PHÂN BỐ CÁC ĐA HÌNH NUCLEOTIDE ĐƠN CỦA NHÓM ĐƠN BỘI C TRÊN NIỄM SẮC THỂ Y Ở NGƯỜI VIỆT NAM

Nguyễn Đăng Tôn, Nguyễn Thùy Dương, Nông Văn Hải

Phòng thí nghiệm trọng điểm Công nghệ gen, Viện Công nghệ sinh học, Viện Khoa học và Công nghệ Việt Nam

### TÓM TẮT

Các chi thị trên vùng không trao đổi chéo của nhiễm sắc thể Y ở người được ứng dụng trong rất nhiều lĩnh vực khác nhau, bao gồm: nghiên cứu về sinh học tiến hóa, khoa học hình sự, di truyền y học và tái lập cây phát sinh phả hệ ở người. Nhóm đơn bội C xuất hiện ngay sau khi loài người di cư ra ngoài châu Phi khoảng 50 nghìn năm trước. Nhóm đơn bội này có thể theo con đường từ Nam bán đảo Ả Rập qua Pakistan và Ấn Độ đến Sri Lanka, Đông Nam Á và Australia. Cận nhóm (paragroup) C\* được tìm thấy ở Ấn Độ, Sri Lanka, một số nơi ở Đông Nam Á. Nhánh C1 xuất hiện chủ yếu ở người Nhật Bản, trong khi đó, nhánh C2 có tần số phân bố cao ở New Guinea, Melanesia. Nhánh C3 có thể có nguồn gốc ở Đông Nam Á hoặc Trung Á, từ đó phân bố lên Bắc Á và di cư sang châu Mỹ. Để tìm hiểu về sự phân bố đa hình các nhóm đơn bội ở người Việt Nam, chúng tôi đã tiến hành phân tích sự đa hình các chi thị SNP C-M216, C-M217 và C-M38 thuộc các nhóm đơn bội C, C3 và C2 tương ứng ở người Việt Nam. Trong tổng số (n = 280) mẫu nghiên cứu thuộc dân tộc Kinh (Việt), sống tại miền Bắc Việt Nam, đã phát hiện thấy chi thị C-M38 (T→G) có 8 mẫu xuất hiện đa hình, với tần số phân bố của mỗi chi thị là 2,85%. Tương tự như vậy, chi thị C-M216 (C→T) cũng có 8 mẫu xuất hiện đa hình, với tần số phân bố là 2,85%. Trong khi đó, chi thị C-M217 (A→C) có 30 mẫu trong tổng số các mẫu nghiên cứu có đa hình, với tần số phân bố là 10,7%. Kết quả phân tích cho thấy, nhóm đơn bội C3 của nhiễm sắc thể Y có tần số phân bố tương đối cao, trong khi đó, nhóm đơn bội C và C2 có tần số phân bố trung bình thấp ở người Việt Nam.

**Từ khóa:** Đa hình nucleotide đơn (SNP), người Việt Nam, nhiễm sắc thể Y, nhóm đơn bội (haplogroup) C

### ĐẶT VẤN ĐỀ

Các nghiên cứu về di truyền quần thể và tiến hóa người đã tăng rất nhanh chóng trong khoảng 10 năm trở lại đây. Với những đặc điểm như: di truyền theo phụ hệ (dòng bố) và không xảy ra hiện tượng trao đổi chéo trên phần lớn chiều dài của nhiễm sắc thể (ngoại trừ 2 vùng trao đổi chéo với nhiễm sắc thể X ở phía 2 đầu của nhiễm sắc thể - PAR1 và PAR2), nhiễm sắc thể Y là công cụ hữu hiệu để nghiên cứu về di truyền quần thể theo phụ hệ, tương tự như mtDNA trong nghiên cứu di truyền quần thể theo mẫu hệ (dòng mẹ). Sự đa hình của nhiễm sắc thể Y không chỉ có ý nghĩa khoa học trong phân tích đa dạng di truyền quần thể, nghiên cứu nguồn gốc tiến hóa loài người, mà nó còn có ý nghĩa đối với việc điều tra tội phạm, giám định hình sự và pháp y. Nghiên cứu sự đa hình của nhiễm sắc thể Y ở người còn cho phép chúng ta tìm hiểu về các bệnh liên quan đến nhiễm sắc thể Y như chứng thiếu hoặc không có khả năng sinh tinh trùng, ung thư và một số bệnh thông thường khác...

Nhóm đơn bội C xuất hiện đầu tiên ở châu Á và

Tây Nam Thái Bình Dương, cách đây khoảng 50 nghìn năm. Nhóm đơn bội C được xác định bằng 5 chi thị SNP: RPS4Y711, M216, P184, P255 và P260. Nhóm đơn bội C được phân chia thành 19 nhánh với 29 chi thị SNP. Cho đến nay, nhóm đơn bội C vẫn không xác định được ở tiểu vùng Sahara của châu Phi, nên có giả thiết rằng, nhóm đơn bội này có nguồn gốc từ châu Á. Cận nhóm (paragroup) C\* và những nhánh nhỏ của chúng được tìm thấy phổ biến ở người châu Á, Australia và châu Đại Dương (Capelli *et al.*, 2001; Hammer *et al.*, 2001; Karafet *et al.*, 2001; Ke *et al.*, 2001; Kivisild *et al.*, 2003; Scheinfeldt *et al.*, 2006; Underhill *et al.*, 2001a). Hiện nay, một nhánh nhỏ của nhóm đơn bội C là C3b hoặc C-P39 được tìm thấy rất hiếm ở người Mỹ bản địa (Zegura *et al.*, 2004), trong khi đó nhánh C4 hay C-M347 chỉ quan sát thấy ở Australia (Hudjashov *et al.*, 2007).

Các nghiên cứu về đặc điểm phân tử của nhiễm sắc thể Y trên đối tượng người Việt Nam chủ yếu do các tác giả nước ngoài thực hiện. Từ năm 2001, trong một nghiên cứu về lịch sử của người Đông Á theo phụ hệ, người ta đã tiến hành phân tích tần số phân bố của các nhóm đơn bội trên nhiễm sắc thể Y

của 25 nhóm dân tộc phân bố ở Đông Á, trong đó có sử dụng mẫu là người Việt Nam sống ở khu vực Nam Trung Bộ với số lượng mẫu là 70 cá thể (Karafet *et al.*, 2001). Tiếp theo đó, cũng đã có một số tác giả khác, trong các nghiên cứu của mình có sử dụng mẫu nghiên cứu là người Việt Nam (Li *et al.*, 2008; Scheinfeldt *et al.*, 2006). Tuy nhiên, các công trình này đều được thực hiện ở nước ngoài, với số mẫu cá thể nhỏ.

Trong bài báo này, chúng tôi tiến hành xác định đa hình nucleotide đơn (SNP) của 3 chỉ thị là đại diện cho nhóm đơn bội C (M216) và hai nhánh của nhóm đơn bội này C2 (M38) và C3 (M217) trên các mẫu cá thể nam của người Việt Nam. Nhóm đơn bội này xuất hiện sớm nhất ở người châu Á.

## NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

### Nguyên liệu và hóa chất

Hai trăm tám mươi mẫu máu ngoại vi của các cá thể nam khỏe mạnh, có độ tuổi từ 18 đến 50, thuộc dân tộc Kinh (Việt) được thu thập ở khu vực miền Bắc Việt Nam và bảo quản trên giấy FTA ở nhiệt độ 4°C.

Bộ kit xác định trình tự nucleotide (BigDye 3.1 terminator) và kit thực hiện phản ứng kéo dài một nucleotide (Single Base Extension - SBE) (SNaPshot) đặt mua của hãng Applied Biosystems, Mỹ.

Các primer cho PCR (Bảng 1) và SBE primer (Bảng 3) được thiết kế dựa trên cơ sở công bố trước đây (Sanchez *et al.*, 2003), đặt tổng hợp tại hãng IDT (Canada). Các SBE primer được tinh sạch qua hệ thống sắc ký lỏng cao áp (HPLC).

Các hóa chất dùng cho PCR thông thường (dNTP, *Taq* DNA polymerase, MgCl<sub>2</sub>...) của hãng Fermentas.

### Phương pháp nghiên cứu

DNA tổng số được tách chiết theo phương pháp Chelex từ mẫu máu ngoại vi được lưu giữ trên giấy FTA (Walsh *et al.*, 1991).

PCR được tiến hành với tổng thể tích 25 µl với 10 ng DNA tổng số, 1X PCR buffer, 8 mM MgCl<sub>2</sub>, 700 mM mỗi dNTP, 5 mM primer và 1 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems), với chu trình nhiệt: biến tính 94°C trong 5 phút, 35 chu kỳ (95°C - 30 giây, 60°C - 30 giây và 65°C - 30 giây) và 7 phút ở 65°C. Phản ứng

được thực hiện trên máy GenAmp PCR system 9700. Sản phẩm PCR sau đó được tinh sạch bằng kit tinh sạch sản phẩm PCR MinElute (Qiagen).

Phản ứng SBE được tiến hành trong tổng thể tích là 8 µl, với 0,5 µl sản phẩm PCR đã tinh sạch, 1,5 µl hỗn hợp phản ứng SNaPshot, 0,5 µl hỗn hợp SBE primer (1 mM mỗi primer), với chu trình nhiệt: 30 chu kỳ (96°C - 10 giây, 50°C - 5 giây, 60°C - 30 giây). Sản phẩm SBE được tinh sạch bằng cách bổ sung 1 U enzyme Shrimp Alkaline phosphatase - (SAP) (Fermentas), ủ ở 37°C trong 45 phút, sau đó SAP bị bất hoạt bằng cách ủ ở nhiệt độ 75°C trong 15 phút.

Lấy 2 µl sản phẩm SBE đã được tinh sạch bằng SAP được trộn với 8 µl Hi-Di formamide, đưa lên phân tích bằng máy giải trình tự tự động ABI 3100 sử dụng capillary 36 cm và POP-4 polymer, với thời gian thu mẫu là 22 giây ở 2000 V. Kết quả được phân tích bằng phần mềm GeneMapper 4.0 (Applied Biosystems).

Các sản phẩm PCR tương ứng với các SNP được tinh sạch bằng cột MinElute PCR (Qiagen) theo quy trình của hãng sản xuất. Sản phẩm PCR sau khi tinh sạch được tiến hành xác định trình tự để kiểm tra tính chính xác của marker đã được nhân lên bằng bộ kit xác định trình tự BigDye 3.1 theo hướng dẫn của nhà sản xuất.

## KẾT QUẢ VÀ THẢO LUẬN

### Lựa chọn các marker SNP

Nhóm đơn bội C xuất hiện khoảng 50 nghìn năm trước, ngay sau khi loài người di cư ra ngoài châu Phi. Cận nhóm (paragroup) C\* được tìm thấy ở Ấn Độ, Sri Lanka, một số nơi ở Đông Nam Á. Nhánh C1 xuất hiện chủ yếu ở người Nhật Bản (Hammer *et al.*, 2006), trong khi đó nhánh C2 có tần số phân bố cao ở New Guinea, Melanesia, Polynesia (Bortolini *et al.*, 2003; Cinnioglu *et al.*, 2004; Kayser *et al.*, 2001). Nhánh C3 có thể có nguồn gốc ở Đông Nam Á hoặc Trung Á, từ đó phân bố lên Bắc Á và di cư sang châu Mỹ (Deng *et al.*, 2004; Karafet *et al.*, 2001; Underhill, Kivisild, 2007).

Nhóm đơn bội C được xác định bằng 5 chỉ thị (RPS4Y711, M216, P184, P255 và P260), nhánh C2 được xác định bằng chỉ thị M38 và C3 được xác định bằng các chỉ thị M217, P44 và PK2 (Karafet *et al.*, 2008). Để tìm hiểu về sự phân bố của nhóm đơn bội này ở người Việt Nam, chúng tôi tiến hành xác định

đa hình các marker SNP: M216, đại diện cho nhóm đơn bội C; M38, xác định nhánh C2 và M217, tham gia xác định nên nhánh C3. M38 có đa hình T→G ở vị trí 20201546, M216 có đa hình C→T ở vị trí 13946958 và M217 có đa hình A→C ở vị trí 13946727 trên trình tự của nhiễm sắc thể Y. Để xác định các điểm đa hình này, chúng tôi đã thiết kế được 3 cặp primer để nhân các đoạn có mang các SNP trên cơ sở trình tự nhiễm sắc thể Y đã được công bố trên ngân hàng Genbank, có kích thước M38 (177 bp), M216 (160 bp) và M217 (145 bp) (Bảng 1).

**Nhân và xác định trình tự các đoạn mang các SNP**

Với mục đích xác định các đa hình nucleotide

đơn chúng tôi đã tiến hành nhân các đoạn chứa các SNP cần phân tích. Theo tính toán lý thuyết, đoạn chứa SNP M38 có kích thước là 177 bp, SNP M216 có kích thước là 160 bp SNP M217 có kích thước là 145 bp. Kết quả trên hình 1 cho thấy, đã nhân được các đoạn trình tự DNA mang các SNP với kích thước tương ứng với tính toán lý thuyết.

Để khẳng định các đoạn trình tự DNA nhân được có đúng là các đoạn mang đa hình SNP hay không, chúng tôi đã tiến hành xác định trình tự các sản phẩm PCR mang các SNP đã nhân được ở 3 mẫu nghiên cứu. Kết quả phân tích cho thấy, trình tự của các đoạn đã nhân được là đúng với các đoạn mang đa hình SNP (kết quả không đưa ra ở đây).

**Bảng 1.** Các chỉ thị SNP, trình tự primer và kính thước sản phẩm PCR cần nhân.

SNP	Đa hình	Vị trí trên NST Y	PCR primer (5' - 3')		Kích thước (bp)
			Forward	Reverse	
M38	T>G	20201546	AGATCTGTTGGCTACTGTTACCCCTA	GAGCTGGCACATCTGCATAATA	177
M216	C>T	13946958	CCAATGGAAATTTTATACCCACA	TGACTGCTAGTTATGTATACCTGTTG	160
M217	A>C	13946727	TCTGTTTCGAGATCATTCTAATTACTG	CTGCTGTGGCTTTCATCAAATA	145

**Bảng 2.** Phân tích đa hình các SNP của 280 mẫu nghiên cứu là người Việt Nam, có độ tuổi từ 18 - 50. (-): không xuất hiện đa hình ở mẫu nghiên cứu.

Mẫu	M38	M216	M217	Mẫu	M38	M216	M217	Mẫu	M38	M216	M217
626	-	T	G	667	-	-	-	707	-	-	-
627	-	-	-	668	-	-	-	708	-	-	-
628	-	-	-	669	-	-	-	709	-	-	-
629	-	-	-	670	-	-	-	710	-	-	-
630	-	-	-	671	-	-	-	711	-	-	-
631	-	-	-	672	-	-	-	712	-	-	-
632	-	-	-	673	-	-	-	713	-	-	-
633	-	-	-	674	-	-	-	714	-	-	-
634	-	-	-	675	-	-	-	715	-	-	-
635	-	-	-	676	-	-	-	716	-	-	-
636	-	-	-	677	-	-	-	717	-	-	-
637	-	-	-	678	-	-	-	718	-	T	G
638	-	-	-	679	-	-	-	719	-	-	-
639	-	-	-	680	-	-	-	720	-	-	-
641	-	-	-	681	-	-	-	721	-	-	-
642	-	-	-	682	-	-	-	722	-	-	-
643	-	-	-	683	-	-	-	723	-	-	-
644	-	-	-	684	-	-	-	724	-	-	-
645	-	-	-	685	-	-	-	725	-	-	-
646	-	-	-	686	-	-	-	726	-	-	-
647	-	-	-	687	-	-	-	727	-	-	-

Mẫu	M38	M216	M217	Mẫu	M38	M216	M217	Mẫu	M38	M216	M217
648	-	-	-	688	-	-	-	728	-	T	G
649	-	-	-	689	-	-	-	729	-	-	-
650	-	-	-	690	-	T	G	730	-	-	-
651	-	-	-	691	-	-	-	731	-	-	-
652	-	-	-	692	-	-	-	732	-	-	-
653	-	-	-	693	-	-	-	733	-	-	-
654	-	-	-	694	-	-	-	734	-	T	G
655	-	-	-	695	-	-	-	735	-	-	-
656	-	-	-	696	-	-	-	736	-	-	-
657	-	-	-	697	-	-	-	737	-	-	-
658	-	-	-	698	-	-	-	739	-	-	-
659	-	-	-	699	-	-	-	740	-	T	-
660	-	-	-	700	-	-	-	741	-	-	-
661	-	-	-	701	-	-	-	742	-	-	-
662	-	-	-	702	-	-	-	743	-	-	-
663	-	-	-	703	-	-	-	744	-	-	-
664	-	-	-	704	-	-	-	745	-	-	-
665	-	-	-	705	-	-	-	746	-	-	-
666	-	-	-	706	-	-	-	747	-	T	-
748	-	-	-	788	-	-	-	830	-	-	G
749	-	-	-	789	-	T	-	831	-	-	G
750	-	-	-	790	-	-	-	832	G	-	G
751	-	-	-	791	-	-	-	833	G	-	G
752	-	-	-	792	-	-	-	834	-	-	-
753	-	-	-	793	-	-	-	836	-	-	-
754	-	-	-	794	-	-	-	837	-	-	-
755	-	-	-	795	-	-	-	838	-	-	-
756	-	-	-	796	-	-	G	839	-	-	-
757	-	-	-	797	-	-	-	842	-	-	-
758	-	-	-	798	-	-	-	843	-	-	-
759	-	-	-	799	-	-	-	844	G	-	-
760	-	-	-	800	-	-	-	846	G	-	G
761	-	-	-	801	-	-	G	848	-	-	-
762	-	-	-	802	-	-	-	849	-	-	-
763	-	-	-	803	-	-	-	850	-	-	-
764	-	-	-	804	-	-	-	851	-	-	G
765	-	-	-	805	-	-	-	852	-	-	-
766	-	-	-	806	-	-	-	853	-	-	G
767	-	-	-	807	-	-	-	854	G	-	G
768	-	-	-	808	-	-	-	856	-	-	-
769	-	-	-	809	-	-	-	857	-	-	-
770	-	-	-	810	-	-	-	858	-	-	-
771	-	-	-	811	-	-	G	859	-	-	-
772	-	-	-	812	-	-	-	860	-	-	-
773	-	-	-	813	-	-	-	861	-	-	-
774	-	-	-	814	-	-	-	865	-	-	-
775	-	-	-	815	-	-	-	867	-	-	G

Mẫu	M38	M216	M217	Mẫu	M38	M216	M217	Mẫu	M38	M216	M217
776	-	-	-	816	-	-	-	868	-	-	-
777	-	-	-	817	-	-	-	870	-	-	-
778	-	-	-	818	-	-	-	871	-	-	-
779	-	-	-	820	-	-	-	872	-	-	-
780	-	-	-	821	-	-	-	873	-	-	-
781	-	-	-	822	-	-	G	874	-	-	-
782	-	-	-	823	G	-	G	875	-	-	G
783	-	-	-	825	G	-	G	876	G	-	G
784	-	-	-	826	-	-	G	877	-	-	-
785	-	-	-	827	-	-	-	878	-	-	-
786	-	-	-	828	-	-	-	879	-	-	-
787	-	-	-	829	-	-	G	880	-	-	G
881	-	-	-	895	-	-	-	915	-	-	-
882	-	-	-	896	-	-	-	916	-	-	-
883	-	-	-	897	-	-	-	917	-	-	-
884	-	-	-	899	-	-	G	918	-	-	G
885	-	-	-	900	-	-	-	919	-	-	-
886	-	-	-	902	-	-	-	920	-	-	-
887	-	-	-	904	-	-	-	921	-	-	-
888	-	-	-	906	-	-	G	922	-	-	-
889	-	-	-	908	-	-	-	923	-	-	-
890	-	-	-	909	-	-	-	925	-	-	-
891	-	-	-	910	-	-	-	926	-	-	-
892	-	-	G	912	-	-	-	927	-	-	-
893	-	-	-	913	-	-	G				
894	-	-	-	914	-	-	-				



Hình 1. Sản phẩm PCR của các đoạn mang các SNP M38 (A), M216 (B) và M217 (C) ở một số mẫu nghiên cứu.

### Phân tích các đa hình SNP

Trong các nghiên cứu này, chúng tôi tiến hành

xác định các đa hình SNP theo nguyên lý kéo dài primer (primer extension) bằng bộ kit SNaPshot của hãng Applied Biosystems (ABI). Trên cơ sở, thiết kế một primer bắt cặp sát vị trí SNP theo chiều 5'-3' (Bảng 3), sau khi primer bắt cặp bổ sung với trình tự DNA đích, chúng sẽ được kéo dài thêm một nucleotide nhờ DNA polymerase. Vì trong phản ứng chỉ có ddNTP đã được đánh dấu huỳnh quang bằng các màu khác nhau, không có mặt dNTP, nên DNA polymerase chỉ kéo dài được một nucleotide tương ứng với vị trí SNP trên sợi khuôn. Dựa vào kích thước của primer và tín hiệu huỳnh quang khác nhau của các ddNTP, có thể xác định được đa hình của nucleotide đó.

Kết quả xác định các SNP được liệt kê ở bảng 2. Trong đó, chỉ thị M38 và M216 có 8 mẫu xuất hiện đa hình, M217 có 30 mẫu xuất hiện đa hình với tần số phân bố tương ứng là 2,85, 2,85 và 10,7% trong tổng số 280 mẫu nghiên cứu.

**Bảng 3.** Trình tự các SBE primer dùng để xác định các SNP.

Marker	Kích thước	SBE primer (5' - 3')	Chiều
M38	38	GTGAAAGTCTGACCAGTACTTATTATGGAAAACCAACT	F
M216	50	GGTGCCACGTCGTGAAAGTCTGACAAGCTAGTTATGTATACCTGTTGAAT	R
M217	30	ACAATTATGATTTTTCTCTCTGAAGAGTT	R

### Sự phân bố của nhóm đơn bội C-216, C-217 và C-38 ở người Việt Nam

#### Nhánh C-M38

M38 xác định nhánh C2, phân bố chủ yếu ở New Guinea, Melanesia, Polynesia (Jobling, Tyler-Smith, 2003; Karafet *et al.*, 2008; Kayser *et al.*, 2003 2006, Scheinfeldt, 2006; Underhill *et al.*, 2000; 2001b). Trong một nghiên cứu trước đây (Scheinfeldt *et al.*, 2006) về đa hình nhiễm sắc thể Y ở Melanesia, nhóm tác giả này đã không phát hiện được đa hình ở chỉ thị M38 trong 11 mẫu người Việt Nam. Cũng tương tự như vậy, trong một nghiên cứu về mối quan hệ giữa các nhóm người thuộc hai nhóm ngôn ngữ Austronesia và Dai, các tác giả khác cũng không phát hiện thấy đa hình ở chỉ thị C-M38 của 11 mẫu người dân tộc Chăm của Việt Nam (Li *et al.*, 2008). Trong nghiên cứu này, chúng tôi đã tìm thấy 8 cá thể xuất hiện đa hình của chỉ thị này với tần số phân bố là 2,85% ở người Việt Nam.

#### Nhóm đơn bội C-M216

Chỉ thị M216, cùng với 4 chỉ thị khác (RPS4Y711, P184, P255 và P260) xác định nhóm đơn bội C. Với tổng số cá thể là 280 được tiến hành phân tích, đã phát hiện được 8 cá thể có đa hình ở marker này, với tỷ lệ xuất hiện là 2,85%. Từ kết quả này cho thấy, C-M216 có tần số phân bố tương đối thấp trong cộng đồng người Việt Nam. Tần số phân bố của nhóm đơn bội C-M216 ở một số dân tộc sống tại các tỉnh phía Nam của Trung Quốc (Quảng Đông và Quảng Tây) từ 2,9% ở người Biao cho đến 10% ở người Man-Caolan. Trong khi đó, tần số phân bố của nhóm đơn bội này ở một số dân tộc sống ở khu vực Đông Nam Á và một số vùng của châu Đại Dương tương đối cao, từ 7,7% ở người Bangka cho đến 45,5% ở người Irian.

#### Nhánh C-M217

C3 có thể có nguồn gốc ở Đông Nam Á hoặc Trung Á (Deng *et al.*, 2004; Karafet *et al.*, 2001; Underhill, Kivisild, 2007). Nhóm đơn bội C3 được xác định bằng các marker M217, P44 và PK2. Trong tổng số 280 mẫu nghiên cứu, đã phát hiện được 30

mẫu xuất hiện đa hình A→C, với tần số phân bố là 10,7%, cho thấy C-M217 phân bố khá rộng rãi ở Việt Nam. Kết quả này cũng phù hợp với những số liệu của các tác giả khác đã công bố về tần số phân bố của chúng ở khu vực Đông Á và Đông Nam Á. Trong một nghiên cứu về lịch sử khu vực Đông Á, trên cơ sở phân tích sự phân bố của nhiễm sắc thể Y, với 70 mẫu người Việt Nam sống ở khu vực Nam Trung Bộ, các tác giả đã tìm thấy có 3 cá thể thuộc nhánh C3-M217. Mặc dù, với số lượng mẫu nghiên cứu của từng quần thể tương đối nhỏ ( $n \leq 50$ ) nhưng các tác giả đã tìm thấy chỉ thị này có tần số phân bố không nhỏ ở Đông Nam Á, từ 2,3% ở người Yizu cho đến 18,3% (người Tujia). Ở khu vực Đông Bắc Á, nhóm đơn bội C3 cũng có tần số phân bố khá phổ biến, nhỏ nhất là 4,5% ở người Hán sống ở phía Bắc. Đặc biệt, trong tổng số 95 mẫu nghiên cứu là người Siberian Evenks, người ta đã tìm thấy 65 cá thể thuộc nhóm đơn bội C3, với tần số phân bố là 68,4% (Karafet *et al.*, 2001).

### KẾT LUẬN

Đã xây dựng được phương pháp xác định các đa hình nucleotide đơn (SNP) trên nhiễm sắc thể Y ở người. Bằng phương pháp này, đã phân tích được các SNP của các chỉ thị M38, M216 và M217, là các chỉ thị thuộc nhóm đơn bội C của nhiễm sắc thể Y ở người.

Tần số phân bố của các chỉ thị M38, M216 và M217 đại diện cho nhóm đơn bội C2 (M38), C (M216) và C3 (M217), ở người Việt Nam tương ứng là 2,85, 2,85 và 10,7%. Đây là những số liệu đầu tiên về đa hình các nucleotide đơn trên nhiễm sắc thể Y ở người Việt Nam được xác định với lượng mẫu đủ lớn ( $n = 280$ ).

**Lời cảm ơn:** Công trình này được hỗ trợ kinh phí từ đề tài "Ứng dụng công nghệ DNA để nghiên cứu một số đặc điểm cấu trúc nhiễm sắc thể Y của người Việt Nam" thuộc Chương trình "Nghiên cứu Cơ bản trong Khoa học Tự nhiên" giai đoạn 2006 - 2008 và nhiệm vụ nghiên cứu thường xuyên 2009 - 2010 "Nghiên cứu đặc điểm đa hình nucleotide đơn

(SNPs) của hệ gen đơn bội (DNA ty thể và nhiễm sắc thể Y) ở người Việt Nam”.

## TÀI LIỆU THAM KHẢO

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A (2003) Y-Chromosome Evidence for Differing Ancient Demographic Histories in the Americas. *Am J Hum Genet* 73(3): 524-539.

Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB (2001) A Predominantly Indigenous Paternal Heritage for the Austronesian-Speaking Peoples of Insular Southeast Asia and Oceania. *Am J Hum Genet* 68(2): 432-443.

Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114(2): 127-148.

Deng W, Shi B, He X, Zhang Z, Xu J, Li B, Yang J, Ling L, Dai C, Qiang B, Shen Y, Chen R (2004) Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. *J Hum Genet* 49(7): 339-348.

Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S (2006) Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* 51: 47-58.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18(7): 1189-1203.

Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA, Shen P, Oefner P, Renfrew C, Villems R, Forster P (2007) Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc Natl Acad Sci USA* 104(21): 8726-8730.

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4(8): 598-612.

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal Population History of East Asia: Sources, Patterns, and Microevolutionary Processes. *Am J Hum Genet* 69(3): 615-628.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA,

Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18(5): 830-838.

Kayser M, Brauer S, Weiss G, Schiefenovel W, Underhill P, Shen P, Oefner P, Tommaso-Ponzetta M, Stoneking M (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72(2): 281-302.

Kayser M, Brauer S, Weiss G, Schiefenovel W, Underhill PA, Stoneking M (2001) Independent histories of human Y chromosomes from Melanesia and Australia. *Am J Hum Genet* 68(1): 173-190.

Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P, Xiao C, Shriver M, Lell J, Wallace D, Wells RS, Seielstad M, Oefner P, Zhu D, Jin J, Huang W, Chakraborty R, Chen Z, Jin L (2001) African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* 292(5519): 1151-1153.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 72(2): 313-332.

Li H, Wen B, Chen SJ, Su B, Pramoongjago P, Liu Y, Pan S, Qin Z, Liu W, Cheng X, Yang N, Li X, Tran D, Lu D, Hsu MT, Deka R, Marzuki S, Tan CC, Jin L (2008) Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol* 8(146): doi:10.1186/1471-2148-1188-1146.

Sanchez JJ, Børsting C, Hallenberg C, Buchard A, Hernandez A, Morling N (2003) Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Sci Int* 137(1): 74-84.

Scheinfeldt L, Friedlaender F, Friedlaender J, Latham K, Koki G, Karafet T, Hammer M, Lorenz J (2006) Unexpected NRY chromosome variation in Northern Island Melanesia. *Mol Biol Evol* 23(8): 1628-1641.

Underhill P, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41: 539-564.

Underhill P, Passarino G, Lin A, Shen P, Lahr M (2001a) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65: 43.

Underhill PA, Passarino G, Lin AA, Marzuki S, Oefner PJ, Cavalli-Sforza LL, Chambers GK (2001b) Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mut* 17(4): 271-280.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26(3): 358-361.

Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a

medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10: 506-513.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 21(1): 164-175.

## DISTRIBUTION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) OF Y CHROMOSOME HAPLOGROUP C IN VIETNAMESE

Nguyen Dang Ton, Nguyen Thuy Duong, Nong Van Hai\*

Key Laboratory of Gene Technology (KLG T), Institute of Biotechnology (IBT), Vietnam Academy of Science and Technology (VAST)

### SUMMARY

Markers on the non-recombining portion of the human Y chromosome continue to have applications in many fields including evolutionary biology, forensics, medical genetics, and genealogical reconstruction. Y-chromosome haplogroup C appears to have arisen shortly after modern humans left Africa and is estimated to be approximately 50,000 years old. This haplogroup can be traced across the southern Arabian Peninsula through Pakistan and India into Sri Lanka and Australia, and Southeast Asia. Paragroup (C\*) is found on the Indian subcontinent, Sri Lanka, and in parts of South East Asia. The rare C1 lineage appears to be restricted to Japan. C2 is found predominantly in New Guinea, Melanesia, and Polynesia. The successful C3 lineage is believed to have originated in Southeast or Central Asia, spreading from there into Northern Asia and the America. To investigate a distribution of Y chromosome variation in Vietnamese, three SNP markers C-M38, C-M216, and C-M217, belonging to haplogroups C2, C, and C3, respectively, were typed in 280 male samples from Kinh (Viet) ethnic group living in the North Vietnam. Of samples genotyped, marker C-M38, with T→G transversion at polymorphism site, 8 samples presented as G with distribution frequency of 2.85%. Eight samples presented as T at the polymorphism site by C-M216 marker, with C→T transition with distribution frequency of 2.85%. Thirty samples presented as C at polymorphism site by C-M217 marker (with A→C transversion), with 10.7% frequency. The distribution frequency of the haplogroup C3 of Y chromosome in Vietnamese was relatively high, while the C and C2 were quite low in the Kinh (Viet) ethnic group living in the North Vietnam.

**Keywords:** Haplogroup C, single nucleotide polymorphism (SNP), Vietnamese, Y chromosome

---

\* Author for correspondence: Tel: 84-4-37562934; Fax: 84-4-38363144; E-mail: [vhnong@ibt.ac.vn](mailto:vhnong@ibt.ac.vn)