European Journal of Human Genetics (2001) 9, 765–772

www.nature.com/ejhg

ARTICLE

DNA sequence variability of IGHG3 alleles associated to the main G3m haplotypes in human populations

Patricia Dard[1], Marie-Paule Lefranc[2], Ludmilla Osipova[3] and Alicia Sanchez-Mazas*[,1]

[1]*Laboratory of Genetics and Biometry, Department of Anthropology and Ecology, University of Geneva, Switzerland, and URA 49 CNRS, Musée de l'Homme, Paris, France;* [2]*Laboratory of Molecular Immunogenetics, Institute of Human Genetics, UPR CNRS 1142, University of Montpellier II, France;* [3]*Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*

The present study investigates the molecular basis of the G3m polymorphism expressed by the heavy constant domains of human immunoglobulins gamma 3 chains. By using a new protocol allowing the specific cloning of IGHG3 genes, a total of 51 full-length IGHG3 genomic sequences (about 2 kb) isolated from African, Siberian, West Asian and European population samples were sequenced. IGHG3 sequences were assigned precise G3m haplotypes on the basis of specific associations between G3m allotypes and IGHG3 RFLPs. Specific DNA substitutions involved in the expression of G3m(5), G3m(6), G3m(15), G3m(16), G3m(21), G3m(24) and G3m(28) allotypes were then deduced, elucidating almost completely the determination of the G3m polymorphism at the DNA level. The molecular evolution of G3m haplotypes was investigated by a maximum likelihood phylogeny of IGHG3 sequences. Sequence clusters are shown to be G3m haplotype-specific, corroborating the Gm molecular model deduced from serology, and showing that populations differentiation is much more recent than G3m haplotypes differentiation. The widely distributed G3m^{5,10,11,13,14} haplotype is likely to be ancestral to the other G3m haplotypes presently found at high frequencies in different continental areas. *European Journal of Human Genetics* (2001) **9**, *765–772*.

Keywords: IgG immunoglobulins; Gm allotypes; IGHG3; DNA polymorphism; human genetic diversity; evolution; phylogeny

Introduction

The Gm polymorphism is associated to the constant domains of human immunoglobulins IgG1, IgG2 and IgG3 heavy chains.[1] It is defined serologically by 18 allotypes specific to γ1 (G1m 1, 2, 3, 17), γ2 (G2m 23) and γ3 (G3m 5, 6, 10, 11, 13, 14, 15, 16, 21, 24, 26, 27, 28). The IGHG1, IGHG2 and IGHG3 genes are physically linked on chromosome 14q32.3. Gm allotypes are thus usually inherited in fixed combinations called Gm haplotypes.

***Correspondence: Alicia Sanchez-Mazas, Laboratory of Genetics and Biometry, Department of Anthropology and Ecology, University of Geneva, 12 rue Gustave-Revilliod, 1227 Carouge, Geneva, Switzerland. Tel: +41 22 702 69 84; Fax: +41 22 300 03 51; E-mail: alicia.sanchez-mazas@anthro.unige.ch.**
Received 22 January 2001; revised 13 June 2001; accepted 26 June 2001

The Gm polymorphism has been extensively studied in human populations, showing that a limited number of haplotypes are observed worldwide, with high frequency variation among populations from different continents.[2–4] Gm diversity among populations appears to be highly correlated with geography at the world scale, and merely reveals significant correlation with linguistic relationships at continental levels.[3,5–7]

In spite of its interest for anthropology, the Gm system has seldom been investigated before at the molecular level in population studies. The reason is that the molecular basis of Gm allotypy is almost completely unknown due to the difficulty of isolating the highly homologous IGHG genes from each other. Punctual amino acid substitutions have been proposed for the determination of G1m and G2m allotypes,[8–15] but most amino acids responsible for the expression of the numerous G3m allotypes remain unknown.

The aim of this study is to determine the non-synonymous substitutions involved in G3m allotypy, and to investigate the molecular variation associated to G3m haplotypes. For this purpose, we design an original protocol that permits the specific cloning of IGHG3 genes. We apply it to the DNA sequencing of 51 full-length IGHG3 genes from different population samples for which G3m serological and IGHG3 RFLP typings were previously done. Specific associations between G3m haplotypes and IGHG3 RFLPs allow us to define the precise relationship between IGHG3 DNA sequences and G3m haplotypes deduced by serology. From these analyses, we determine the non-synonymous substitutions involved in the expression of most G3m allotypes, and we propose a tentative evolutionary model of G3m haplotypes.

Subjects and methods

Studied populations

The studied individuals, all being volunteers, originate from two main populations, the Senegalese Mandenka and the Siberian Selkups. This choice was done because both samples were previously tested for Gm allotypes and IGHG RFLPs.[16–18] The Mandenka population is an endogamous community of about 3200 individuals living in the Niokholo region of eastern Senegal. Its language belongs to the Niger-Congo Mande family, which includes several millions people located between Nigeria and the Atlantic coast. Historical information indicates that the Mandenka came from the East and settled in West Africa during the 14th century. Many serological and DNA polymorphisms were already analysed in this population.[7,19–25] The northern Selkups are about 2000 people living in seven localities on a vast territory in the north of western Siberia. Their language belongs to the north-Samodian branch of the Uralic family. The Selkups represent a long-settled tribe in Siberia, but their more numerous ancestors lived further south before the Russian invasion in the 16th century. Under the pressure of nomads and Russian settlers, they migrated north towards the basins of the Turukhan and Taz rivers. The Selkups were previously tested for several serologic and molecular polymorphisms.[18,26–29]

Sampling strategy

As our aim was to define G3m allotypy at the DNA level, individuals were chosen from the two available Mandenka and Selkup samples according to already known Gm phenotypes. We planned to sequence the maximal number of different G3m haplotypes expressing different allotypes. For picking up the DNAs corresponding to precise G3m haplotypes, we used the results of RFLP typings. Indeed, we previously showed that very strong associations exist between IGHG3 hinge lengths, which are composed of 2, 3, or 4 exons, and G3m haplotypes in the Mandenka population.[17] We first checked the putative existence of these associations in 30 Selkups, 30 Lebanese and 13 French (results not shown). We then selected individuals whose IGHG3 genes were known to be associated to specific G3m haplotypes. Most individuals were chosen among heterozygotes for IGHG3 hinge lengths, to avoid any confusion between the sequenced gene and its associated G3m haplotype. We finally sequenced 24 Mandenka and 17 Selkup genes. To complete our study, a few additional samples were obtained from Lebanese (seven Christian Maronites sampled in Beyrouth), French (2), and Algerian Mzab (1) individuals who had previously been typed for both Gm allotypes and IGHG RFLPs. This gave us a final sample of 51 IGHG3 of known haplotype (Table 1).

PCR amplification, cloning and sequencing of IGHG3 genes

An original protocol was designed to specifically sequence IGHG3 genes (details available at http://anthro.unige.ch/~sanchez/ejhg01/). Briefly, IGHG genes were amplified after digestion with the Bsp*HI* enzyme, which cuts all IGHG in the CH2 exon (T$^{\downarrow}$CATGA), except IGHG3. This permitted to avoid almost completely the amplification of IGHG1, IGHG2, IGHG4 and IGHGP genes. After separation on gel, IGHG3 PCR products were ligated in BlueScript vectors and cloned in DH5 bacteria. Ten primers were used to sequence IGHG3 genes on both strands with an ABI 377 automatic sequencer. Each gene was sequenced from three independent clones in order to avoid artifacts that could be confused with polymorphism. Thus, a total of 30 sequences were necessary for each IGHG3 gene (10 primers for each clone).

Sequence alignment and phylogenetic analyses

The reconstruction of each IGHG3 gene from its 10 overlapping sequences and the alignment of the 51 reconstructed IGHG3 genes were done by using the Sequence Navigator™ software. Neighbor-Joining (NJ),[30] UPGMA,[31] and Maximum

Table 1 Ethnic and geographic origin of the individuals whose IGHG3 genes were sequenced, with their associated G3m haplotypes

G3m haplotype	*Ethnic origin*	*Geographic location*	*Number of sequenced IGHG3 genes*
G3m^{5,10,11,13,14}	Mandenka	Senegal (West Africa)	10
	Selkup	Siberia (East Asia)	4
	Lebanese	Lebanon (West Asia)	2
	French	France (Europe)	1
	Mzab	Algeria (North Africa)	1
G3m^{21,28}	Selkup	Siberia (East Asia)	7
	Lebanese	Lebanon (West Asia)	2
	French	France (Europe)	1
G3m^{5,6,10,11,14}	Mandenka	Senegal (West Africa)	2
G3m^{10,11,13,15,16}	Selkup	Siberia (East Asia)	6
	Lebanese	Lebanon (West Asia)	3
G3m^{10,11,13,15}	Mandenka	Senegal (West Africa)	4
G3m^{5,6,11,24}	Mandenka	Senegal (West Africa)	8

Likelihood (ML)[32] phylogenetic trees were constructed for these sequences by using the MEGA[33] and DNAML programs of PHYLIP.[34] Robustness of the NJ and UPGMA trees was assessed by a bootstrap procedure.[35,36]

Results

IGHG3 DNA sequence polymorphism

The alignment of the 51 IGHG3 genes sequenced in this study is shown in Table 2. The EZZ sequence (T1) of aTunisian (X03604[37]) is taken as a reference, and the LAT sequence of another Tunisian (T2) (X16110[38]) is also included. The G3m haplotype corresponding to each sequence, and the amino acids involved in the observed non-synonymous substitutions are indicated (right and bottom of Table 2, respectively). Nineteen different sequences involving 40 segregating sites are observed: EMBL (http://www.ebi.ac.uk/) and IMGT (http://imgt.cines.fr:8104/) accession numbers AJ390235 to AJ390284, X99549, X03604, X16110. IGHG3 alignments are available at http://imgt.cines.fr:8104/textes/IMGTrepertoire.html. Most substitutions are found in the CH3 domain. Seventeen substitutions are non-synonymous, defining a total of 15 different proteins. The deletions corresponding to the variable number of exons in the hinge region[17,39] are also shown.

Different levels of molecular diversity are observed for G3m haplotypes. Haplotype G3m^{5,10,11,13,14} is the most heterogeneous at the DNA level, as 11 different sequences are found among 19 G3m^{5,10,11,13,14} genes. A high level of diversity is found in the Senegalese Mandenka, where the 10 G3m^{5,10,11,13,14} genes analysed reveal seven different sequences (M1 to M7). By contrast, all Selkup, Lebanese and French G3m^{5,10,11,13,14} sequences (n=7) are identical to each other, except in the hinge region (S+L and L+F). The G3m^{21,28} haplotype is also highly variable at the DNA level, as four distinct sequences are found among 10 G3m^{21,28} genes, and the three G3m^{21,28} sequences observed in the Selkups are different from each other (S+F, S1, S2). The sequences associated to haplotype G3m^{10,11,13,15,16} vary among populations, as that found in Lebanese (L2, n=3) differ from that found in Selkups (S3, n=6) by one substitution. Finally, G3m^{5,6,10,11,14} (M8, n=2), G3m^{10,11,13,15} (M9, n=4), and G3m^{5,6,11,24} (M+T, n=9) sequences are monomorphic. These haplotypes were only sequenced in the Mandenka except one G3m^{5,6,11,24} in a Tunisian, which is identical to the Mandenka sequence (M+T).

Amino acids involved in G3m allotypy

The results presented in Table 2 indicate that γ3 chains coded by a given G3m haplotype sometimes differ from each other by non-synonymous substitutions. Indeed, substitutions found at positions 176 (CH1), 296, 309, 339 (CH2) and 392 (CH3) correspond to a polymorphism expressed within a given haplotype. Moreover, the mutations observed at positions 192 and 193 would not express γ3 allotypes as they are located in the CH1 domain.[40,41] This means that the 17 non-synonymous substitutions found among the 51 sequenced IGHG3 genes are not all involved in allotypy. We suggest that a total of 10 amino acid positions (291, 292, 379, 384, 397, 409, 419, 422, 435 and 436) are potentially involved in γ3 allotypy (Table 3):

- G3m(21) is likely to be determined by a Pro→Leu mutation at position 291 (CH2 domain). The same conclusion was drawn from the analysis of the OMM protein[42,43] and a previous sequencing of CH2-CH3 domains.[13] We exclude the involvement of amino acid Asn384, because it is coded by one G3m^{5,10,11,13,14} Mandenka sequence (M6), and of Arg435-Tyr436, because they are located on CH3: enzymatic digestions indicate that G3m(21) is detected on the Fc (CH2-CH3), but not on the Fc′ (CH3) fragments.[40] Arg435-Tyr436 would merely be involved in G3m(28) allotypy (see below).
- G3m(28) may be determined by Arg435-Tyr436 (CH3 domain), as previously postulated.[44] This is in keeping with studies of Van Loghem *et al*[41] which demonstrated that G3m(28) is present on Fc′ (CH3). We exclude the involvement of Asn384 because this amino acid is also present on the other γ sub-classes and is thus not specific for G3m allotypy. Asn384 would rather be responsible for the non-expression of allotypes G3m(10), G3m(13), and G3m(14) or G3m(11) in G3m^{21,28} haplotypes.
- G3m(16) is likely to be determined by an Arg→Trp substitution at position 292 (CH2 domain). This mutation discriminates G3m^{10,11,13,15,16} and G3m^{10,11,13,15} on CH2-CH3. We exclude Met379 postulated by Matsumoto *et al*[45] because this amino acid is shared by G3m^{10,11,13,15,16} and G3m^{10,11,13,15} haplotypes in our samples.
- G3m(15) is likely to be determined by a Val→Met substitution at position 379 (CH3 domain). We exclude His435, proposed by Matsumoto *et al*[45,46] because it is shared by γ1, γ2 and γ4 proteins. As G3m(15) is in fact not detected on Fc′ (CH3),[40] an allosteric effect with the CH2 domain is probably also necessary for its expression.
- G3m(6) is likely to be determined by a Gln→Glu substitution at position 419 (CH3 domain). As Glu419 is also present on γ4, an allosteric effect may again be involved. This time, we propose that this effect takes place on γ4 for the non-expression of G3m(6), because amino acids 355 and 445, which are γ4 sub-class specific, are in close contact with Glu419 on γ4 tri-dimensional structure. The expression of G3m(6) is also correlated to the non-expression of G3m(13) in G3m^{5,6,10,11,14}. These results agree with the location of G3m(6) and G3m(13) on CH3.[40]
- G3m(24) could be determined by Lys→Arg and Ile→Val substitutions at positions 409 and 422, respectively, which distinguish G3m^{5,6,11,24}. However, Arg409 and Val422 are both present on γ4 like Glu419 defining G3m(6). Amino acid 409 is buried inside the molecule, whereas amino acid 422 is on the surface and close to Glu419. The latter could

Table 2 Alignment of IGHG3 sequences associated to different G3m haplotypes

		CH1: 25–318					*IVS1: 319–715*							*Hinge: 716–1325*						*CH2: 1444–1773*					*IVS3: 1774–1870*		*CH3: 1871–2194*														*3′ region (stop codon: 2194)*			
#Nt[a]		*27*	*126*	*200*	*248*	*252*	*371*	*400*	*417*	*451*	*587*	*620*	*663*	*778*	*970*	*1122*	*1235*	*1310*	*1317*	*1625*	*1627*	*1640*	*1678*	*1768*	*1802*	*1847*	*1873*	*1985*	*2001–2*	*2026*	*2032*	*2039*	*2076*	*2080*	*2083*	*2105*	*2113*	*2114*	*2154*	*2157*	*2199*	*2201*	*2202*	*G3m haplotype*
#Id	*N*																																											
T1[b]	1	T	C	C	G	G	C	A	G	Δ	A	A	C	A	A	A	T	G	A	C	C	A	C	A	A	G	A	G	GC	C	G	A	A	C	C	C	C	A	G	T	G	C	A	5*
M1	1	-	-	-	-	-	-	-	C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5*
M2	1	C	-	-	-	-	-	-	C	G	G	-	-	C		Δ376	-	-	-	-	-	-	-	-	-	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
M3	1	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
Mza	1	-	A	-	-	-	-	-	C	G	G	-	T	C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
M4	1	-	A	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	-	-	-	-	-	-	-	G	-	--	G	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
M5	3	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	-	-	-	-	-	-	-	G	-	--	G	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
M6	1	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	AT	-	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
M7	2	-	-	-	-	-	-	G	C	G	G	-	-	C	G	-	-	-	-	-	-	-	G	-	-	-	-	-	--	-	-	-	-	-	T	-	-	-	-	-	-	-	G	5*
S+L	5	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	A	C	-	-	T	-	-	-	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
L+F	2	-	-	-	-	-	-	-	C	G	G	-	-	C	G	Δ188	-	A	C	-	-	T	-	-	-	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-	-	-	G	5*
M8	2	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	-	-	-	-	-	-	-	-	-	--	G	A	-	-	-	-	G	-	-	-	-	-	-	G	5,6,10,11,14
S+F	5	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	T	-	-	-	-	-	-	-	-	AT	-	-	-	-	-	-	-	-	-	-	A	C	T	G	21,28
S1	1	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	C	-	-	T	-	-	-	-	-	-	-	-	AT	-	-	-	-	-	-	-	-	-	-	A	C	T	G	21,28
S2	2	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	T	-	-	-	-	-	-	-	-	AT	G	-	-	-	-	-	-	-	-	-	A	C	T	G	21,28
L1	2	-	-	-	-	-	-	-	C	G	G	-	-	C	G	-	-	-	-	T	-	-	-	G	G	-	-	-	AT	-	-	-	-	-	-	-	-	-	-	A	C	T	G	21,28
M+T2	9	-	-	-	-	-	-	-	C	G	-	G	-	—	Δ188	-	-	-	-	-	-	-	-	-	-	-	-	-	--	-	-	G	G	A	-	G	T	G	-	-	-	-	G	5,6,11,24
M9	4	-	-	-	A	C	G	-	C	G	G	-	-	C	Δ188	G	-	-	-	-	-	-	-	-	-	-	-	A	--	G	-	G	-	-	-	-	-	-	A	A	-	-	G	10,11,13,15
L2	3	-	-	A	-	-	-	-	C	G	G	-	-	C	Δ188	G	-	-	-	-	T	-	-	-	-	A	-	A	--	G	-	G	-	-	-	-	-	-	A	A	-	-	G	10,11,13,15,16
S3	6	-	-	-	-	-	-	-	C	G	G	-	-	C	Δ188	G	-	-	-	-	T	-	-	-	-	A	-	A	--	G	-	G	-	-	-	-	-	-	A	A	-	-	G	10,11,13,15,16
EU index:				176	192	193														291	292	296	309	339				379	384	392		397	409			419		422	435	436				
Amino acid				S	S	L														P	R	Y	L	T				V	S	N		M	K			Q		I	R	F				
substitutions				Y	N	F														L	W	F	V	A				M	N	K		V	R			E		V	H	Y				

#Nt: nucleotide number; #Id: sequence identifier; N: number of sequences; T1: Tunisian EZZ; M1-M9: Mandenka; S1-S3: Selkup; L1-L2: Lebanese; S+L: Selkup + Lebanese; L+F: Lebanese + French; M+T2: Mandenka + Tunisian LAT; Mza: Mzab; Δ: 1 bp deletion; Δ188: 188 bp deletion; Δ376: 376 bp deletion; 5*: 5,10,11,13,14. [a]Nucleotide numbering starts with the first base located downstream the CH1 *Eco*RI primer, 24 bp upstream the first coding nucleotide; [b]Sequences T1 (EZZ), T2 (LAT), and M2 (Mand-541) were previously published ([37–39], respectively).

Table 3 Amino acids involved in the expression of G3m allotypes

Amino acid positions[a]*:*	*291*	*292*	*379*	*384*	*397*	*409*	*419*	*422*	*435*	*436*
γ1[b]	Pro	Arg	Val	Asn	Val	Lys	Gln	Val	His	Tyr
γ2	–	–	–	–	Met	–	–	–	–	–
γ4	–	–	–	–	–	Arg	Glu	–	–	–
G3m^{5,10,11,13,14}	–	–	–	Ser	Met	–	–	Ile	Arg	Phe
G3m^{5,6,10,11,14}	–	–	–	Ser	Met	–	Glu	Ile	Arg	Phe
G3m^{5,6,11,24}	–	–	–	Ser	–	Arg	Glu	–	Arg	Phe
G3m^{10,11,13,15}	–	–	Met	Ser	–	–	–	Ile	–	–
G3m^{10,11,13,15,16}	–	Trp	Met	Ser	–	–	–	Ile	–	–
G3m^{21,28}	Leu	–	–	–	Met	–	–	Ile	Arg	–
G3m allotypes:	21	16	15	–[c]	–		6[d]	24	28[e]	5

[a]Amino acid positions correspond to the EU index; [b]The γ1 sequence is used as a reference; [c]Asn384 is likely to be involved in the non-expression of G3m(10),(13),(14) allotypes in G3m^{21,28} haplotypes, whereas non-expression of G3m(14) is probably associated to Val397 in G3m^{5,6,11,24}, G3m^{10,11,13,15} and G3m^{10,11,13,15,16} haplotypes; [d]Non-expression of G3m(13) correlates with the expression of G3m(6) in G3m^{5,6,10,11,14} and G3m^{5,6,11,24}; [e]Non-expression of G3m(11) is probably associated with the expression of G3m(28), and these two allotypes can therefore be considered as antithetical.

thus also be involved in the expression of G3m(24), whereas allosteric effects would prevent the expression of this allotype on γ4 proteins.

- G3m(5) may be determined by a Tyr→Phe mutation at position 436 (CH3 domain). However, the expression of this allotype probably also needs the interaction between the CH2 and CH3 domains, as it is detectable only on intact CH2-CH3 peptides.[40] This is possible through amino acid 436, which is in close contact with the CH2 domain.

For the remaining G3m allotypes, G3m(10), (11), (13), (14), the hypotheses are less straightforward. Amino acids Ser384, Ile422, Arg435 and Phe436 are specific for both G3m^{5,10,11,13,14} and G3m^{5,6,10,11,14} haplotypes, and are not present at the same positions on the other γ sub-classes. They may thus be essential for G3m(10), (11), (13), (14) located on CH3.[40] Tentative hypotheses concerning the non-expression of these allotypes are also given in Table 3 (footnotes).

A rare haplotype hidden in a common phenotype

One Mandenka individual, identified as Mand114, has been serologically typed as G3m[5,10,11,13,14]. A likely hypothesis is that this phenotype corresponds to a homozygous for G3m^{5,10,11,13,14}. However, an alternative genotype would be G3m^{5,10,11,13,14}/G3m^{5,11}. Indeed, Asn384, present in one of the two sequences (M6) of Mand114, is probably involved in the non-expression of G3m(10), G3m(13) and G3m(14) allotypes (Table 3, footnotes), and M6 would be related to G3m^{5,11}. In keeping with this hypothesis, the presence of G3m^{5,11} was previously postulated in the Mandenka from the observation of one homozygous G3m[5,11] individual.[7]

Molecular phylogeny of G3m haplotypes

A maximum likelihood phylogeny of G3m haplotypes is shown in Figure 1. All sequences associated to G3m^{5,10,11,13,14} (M1-M5, M7, MZA, S+L, L+F, T1) are closely related to each other and form a central node with short branches. The M6 (possibly G3m^{5,11}, see above) and M8 (G3m^{5,6,10,11,14}) sequences are also directly related to this node. The other sequences form three different clusters, each of them being characterised by a specific G3m haplotype and a long branch from the central node:

- the 10 G3m^{21,28} genes sequenced in the French, Lebanese and Selkups cluster together. Unfortunately, the DNA sequencing of the only G3m^{21,28} gene that we could isolate from a Mandenka did not succeed. Interestingly, the M6 sequence, postulated to be G3m^{5,11}, is found at an intermediate position between G3m^{5,10,11,13,14} and G3m^{21,28}.
- The 9 G3m^{5,6,11,24} genes sequenced in eight Mandenka and a Tunisian are monomorphic and form a long branch from G3m^{5,10,11,13,14} sequences.
- The nine G3m^{10,11,13,15,16} genes sequenced in the Lebanese and Selkups cluster together and are closely related to the four monomorphic G3m^{10,11,13,15} genes found in the Mandenka.

NJ and UPGMA trees constructed for the same data set confirm these results (data not shown). According to the former, G3m^{10,11,13,15} and G3m^{10,11,13,15,16} sequences cluster together with a very high bootstrap value (98%), as do G3m^{21,28} sequences (95%), the other bootstrap values being usually weak. The G3m^{5,6,10,11,14} sequence also segregates with G3m^{5,10,11,13,14} sequences in both trees.

Discussion

Thanks to the present work, the critical amino acids necessary for the expression of seven G3m allotypes are determined: G3m(5), G3m(6), G3m(15), G3m(16), G3m(21), G3m(24) and G3m(28) allotypes are shown to be defined by punctual mutations. However, the expression of four of them (5, 6, 15, 24) also probably depends on the conformation of γ3 or γ4 proteins, where other amino acids are involved in allosteric effects, as previously postulated.[47] On the other hand, G3m allotypy would not completely reflect the amino acid variability of γ3 molecules, as we have supposed that only 10 of the 17 positions found to be polymorphic at the protein level are involved in allotypy. The fact that we could not test some rarely determined G3m allotypes, like G3m(26) and G3m(27), due to the lack of antisera, should be taken into account to clarify this

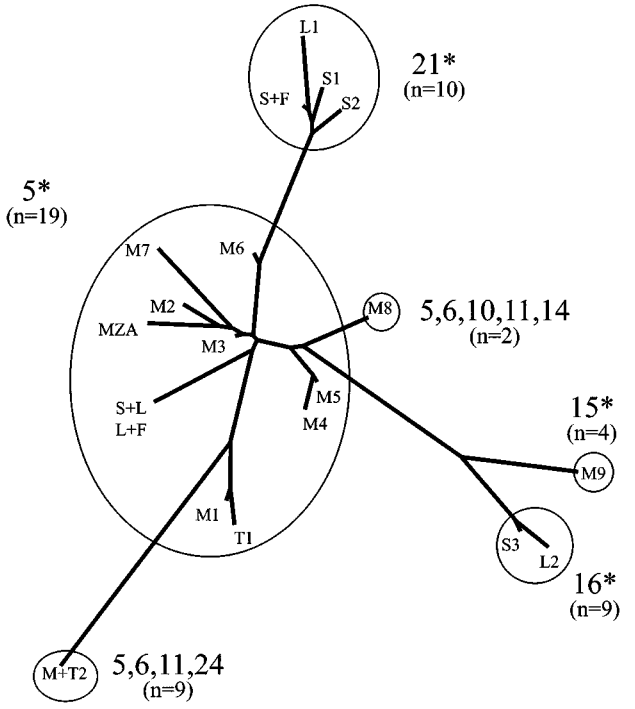

Figure 1 Maximum likelihood tree of IGHG3 DNA sequences. This phylogeny is based on all segregating sites reported in Table 2, including those located downstream the stop codon (3′ region). Deletions in the hinge region are not taken into account. Sequence identifiers are the same as in Table 2 (footnotes). Sequences associated to each G3m haplotypes are circled. 5*=G3m^{5,10,11,13,14}; 21*=G3m^{21,28}; 15*=G3m^{10,11,13,15}; 16*=G3m^{10,11,13,15,16}; *n*=number of IGHG3 sequences. The sequence M6 may correspond to haplotype G3m^{5,11}, instead of G3m^{5,10,11,13,14} (see text).

result. However, a likely explanation is that some mutations remain unrecognized by usual antisera if they are located out of reach of antibodies. Part of the γ3 amino acid variability would thus remain undetected by conventional serology.

Phylogenetic relatedness between DNA sequences associated to the different G3m haplotypes indicates that these haplotypes form distinct and relatively homogeneous groups. The serological definition of G3m haplotypes is thus corroborated by our molecular study. The most variable IGHG3 genes are those associated to G3m^{5,10,11,13,14}, and the other haplotypes may have derived independently from G3m^{5,10,11,13,14} sequences. Previous studies indicate that G3m^{5,10,11,13,14} is found on genes having either 2, 3 or 4 exons coding for the hinge, in contrast to other G3m haplotypes.[17,39] G3m^{5,10,11,13,14} is also present at very high frequency in most human population (except, for example, some East Asian, some Amerindian, and many Oceanian populations), although with different linkages to G1m (3 or 1,3 or 1,17) and G2m (23) alleles.[3,4] All these results suggest that G3m^{5,10,11,13,14} is an ancestral G3m haplotype. G3m^{21,28} also presents a high level of DNA diversity and is widely distributed, with very high frequencies in some human populations. This haplotype may thus have differentiated early after G3m^{5,10,11,13,14}, although its molecular differentiation among and within populations may be quite recent (short branches). On the other hand, the nine African G3m^{5,6,11,24} sequences (including one North African) are monomorphic. Although the long branch differentiating this haplotype suggests a long divergence time from G3m^{5,10,11,13,14}, the absence of variability among Mandenka and Tunisian suggests a recent diffusion of this haplotype within Africa. Huck *et al*[38] have interpreted the G3m^{5,6,11,24} sequence as a gene conversion between a G3m^{5,10,11,13,14} sequence and a IGHG4 gene, due to shared substitutions between G3m^{5,6,11,24} and IGHG4 genes. According to this hypothesis, we expect a high molecular divergence of this haplotype even in the case of a recent differentiation, which is compatible with our results. Interestingly, G3m^{5,6,11,24} and G3m^{5,6,10,11,14} (M8) are not monophyletic, in spite of the common presence of G3m(6). Previous studies indicate that their hinge structures are also different (3 and 4 exons, respectively[17]). G3m^{5,6,10,11,14} is probably a relatively recent differentiation of G3m^{5,10,11,13,14} (short branch) due to a single punctual mutation (Gln→Glu at position 419) that transformed G3m(13) allotype into G3m(6). Finally, G3m^{10,11,13,15} and G3m^{10,11,13,15,16} share a recent common ancestor. These haplotypes differ by a single allotype, G3m(16), and their associated IGHG3 genes possess a 3-exon hinge, so that serological and RFLP data also reflect their close DNA sequence relationship. G3m^{10,11,13,15} is essentially found in sub-Saharan Africans (with high frequencies in the Khoisan), whereas G3m^{10,11,13,15,16} is mostly frequent in North East Asians. This means that genetic distances among IGHG3 sequences do not correlate with geographic distances among sampled populations. Overall, the phylogeny presented in Figure 1 does not reflect population history. Most G3m haplotypes differentiate from each other by long branches, whereas short branches differentiate populations within each haplotype cluster, showing that the G3m polymorphism is much older than population differentiations.

Following several decades of Gm typings in human populations, the present work represents a necessary step towards a better understanding of this complex polymorphism. Our aim was to define the molecular basis of the G3m polymorphism, and this has been almost totally achieved. The molecular determination of G3m allotypes proposed here may encourage people to design a protocol for the molecular typing of Gm specificities. Such an issue would ensure that useful Gm analyses of new population samples are done even when Gm antisera are no longer available.

Acknowledgements

This work was supported by FNRS grant 32-39720.93 to A Sanchez-Mazas, Switzerland. The expeditions in Siberia were financially supported by the rasnoselkup's Administration. We also thank Pr A Langaney for stimulating the present work, Dr L Excoffier for involvement in the study of Mandenka, Pr G Lefranc for revising the manuscript, Dr V Wiebe for long assistance in the study of Selkup populations, and anonymous reviewers for their helpful comments. Once again, we acknowledge the local populations (Mandenka and Selkup) and the other typed individuals for their consented collaboration.

References

1 Lefranc M, Lefranc G: Molecular genetics of immunoglobulin allotype expression; in Shakib F (ed): *The human IgG subclasses, molecular analysis of structure, function and regulation*. Pergamon Press, 1990, 43–78.
2 Cavalli-Sforza L, Menozzi P, Piazza A: The history and geography of human genes. Princetown, New Jersey, Princetown University Press, 1994.
3 Sanchez-Mazas A: Polymorphisme des systèmes immunologiques Rhesus, Gm et HLA et histoire du peuplement humain. PhD Thesis. University of Geneva: Geneva, 1990.
4 Steinberg A, Cook C: The Distribution of Human Immunoglobulin Allotypes. Oxford, New-York, Oxford University Press, 1981.
5 Excoffier L, Pellegrini B, Sanchez-Mazas A *et al*: Genetics and history of sub-Saharan Africa. *Y Phys Anthrop* 1987; **30**: 151–194.
6 Excoffier L, Harding R, Sokal R *et al*: Spatial differentiation of RH and GM haplotype frequencies in Sub-Saharan Africa and its relation to linguistic affinities. *Hum Biol* 1991; **63**: 273–307.
7 Blanc M, Sanchez-Mazas A, Hubert Van Blyenburgh N *et al*: Interethnic Genetic differentiation: Gm polymorphism in Eastern Senegal. *Am J Hum Genet* 1990; **46**: 383–392.
8 Cook C, Steinberg A: An amino acid substitution in the gamma 1 chain of human immunoglobulin G associated with the G1m(2) allotype. *Mol Immunol* 1979; **16**: 555–558.
9 Thorpe N, Deutsch H: Studies on papain produced subunits of human gamma-G-globulins. I. Physicochemical and immunochemical properties of gamma-G-globulin Fc-fragments. *Immunochemistry* 1966; **3**: 317–327.
10 Press E, Hogg N: The amino acid sequences of the Fd fragments of two human gamma-1 heavy chains. *Biochem J* 1970; **117**: 641–660.
11 Ito S, Suzuki K, Miyazaki T *et al*: An amino acid substitution determining G1m(x) allotypic marker. *Nippon Hoigaku Zasshi* 1989; **43**: 155–160.
12 Edelman G, Cunningham B, Gall W *et al*: The covalent structure of an entire gammaG immunoglobulin molecule. *Proc Natl Acad Sci USA* 1969; **63**: 78–85.
13 Balbin M, Grubb A, De LG *et al*: DNA sequences specific for Caucasian G3m(b) and (g) allotypes: allotyping at the genomic level. *Immunogenet* 1994; **39**: 187–193.
14 Brusco A, De Lange G, Boccazzi C *et al*: Molecular characterization of Gm(n+) and G2m(n-) allotypes. *Immunogenet* 1995; **42**: 414–417.
15 Brusco A, Saviozzi S, Cinque F *et al*: Molecular characterization of immunoglobulin G4 gene isoallotypes. *Eur J Immunogenet* 1998; **25**: 349–355.
16 Ghanem N, Lefranc M-P, Lefranc G: Definition of the RFLP alleles in the human immunoglobulin IGHG gene locus. *Eur J Immunol* 1988; **18**: 1059–1065.
17 Dard P, Sanchez-Mazas A, Dugoujon J *et al*: DNA analysis of the immunoglobulin IGHG loci in a Mandenka population from eastern Senegal: correlation with Gm haplotypes and hypotheses for the evolution of the Ig CH region. *Hum Genet* 1996; **98**: 36–47.
18 Osipova L, Posukh O, Wiebe V *et al*: BamHI-SacI RFLP and Gm analysis of the immunoglobulin IGHG genes in the Northern Selkups (west Siberia): new haplotypes with deletion, duplication and triplication. *Hum Genet* 1999; **105**: 530–541.
19 Dard P, Schreiber Y, Excoffier L *et al*: Polymorphisme des loci de classe I HLA-A, -B, et -C dans la population Mandenka du Sénégal oriental. *CR Acad Sci III* 1991; **314**: 573–578.
20 Dard P, Sanchez-Mazas A, Tiercy J-M *et al*: Joint report: HLA-A, -B, and -DR differentiations among North and West African populations; in Tsuji K, Aizawa M, Sasazuki T (eds): *Proceeding of the Eleventh International Histocompability Workshop and Conference, Yokohama (Japan), 6-13 november 1991*. Oxford, Oxford University Press, 1992, **vol 1**, pp 632–636.
21 Graven L, Passarino G, Semino O *et al*: Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* 1995; **12**: 334–345.
22 Tiercy J, Sanchez-Mazas A, Excoffier L *et al*: HLA-DR polymorphism in a Senegalese Mandenka population: DNA oligotyping and population genetics of DRB1 specificities. *Am J Hum Genet* 1992; **51**: 592–608.
23 Martinson J, Excoffier L, Swinburn C *et al*: High diversity of alpha-globin haplotypes in a Senegalese population, including many previously unreported variants. *Am J Hum Genet* 1995; **57**: 1186–1198.
24 Poloni E, Excoffier L, Mountain J *et al*: Nuclear DNA polymorphism in a Mandenka population from Senegal: comparison with eight other human populations. *Am Hum Genet* 1995; **59**: 43–61.
25 Poloni E, Semino O, Passarino G *et al*: Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet* 1997; **61**: 1015–1035.
26 Sukernik R, Osipova L, Schanfield M: Distribution of GM allotypes and deleted IGHG1 haplotypes in the Sel'kups in comparison to the Forest Nentsi of northwestern Siberia. *Exp Clin Immunogenet* 1992; **9**: 15–23.
27 Fefelova V, Osipova L, Yagotin E *et al*: HLA antigens in Selkups. *Sib J Ecol* 1994; **2**: 131–138.
28 Karafet T, Zegura S, Vuturo-Brady J *et al*: Y chromosome markers and Trans-Bering Strait dispersals. *Am J Phys Anthropol* 1997; **102**: 301–314.
29 Kazakovtseva M, Voevoda M, Osipova L: Polymorphism of mitochondrial DNA in Northern Selkups [article in Russian]. *Genetika* 1998; **34**: 416–422.
30 Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**: 406–425.
31 Sneath H, Sokal R: Numerical Taxonomy. San Francisco, WH Freeman, 1973.
32 Felsenstein J: Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 1981; **17**: 368–376.
33 Kumar S, Tamura K, Jakobsen I *et al*: MEGA (Molecular Evolutionary Genetics Analysis), ver 1.0. The Pennsylvania State University, 1993.
34 Felsenstein J: PHYLIP (Phylogeny Inference Package), ver 3.572. Distributed by the author, Department of Genetics, University of Washington, Seattle, 1995.
35 Felsenstein J: Confidence limits on phylogenies: an approach using the bootstrap. *Evol* 1985; **39**: 783–791.

36 Efron B: The Jacknife, the Bootstrap, and other Resampling Plans. Regional Conference Series in Applied Mathematics. Philadelphia, 1982.
37 Huck S, Fort P, Crawford D *et al*: Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: comparison with the other human Cg genes. *Nucl Ac Res* 1986; **14**: 1779 – 1789.
38 Huck S, Lefranc G, Lefranc M-P: A human immunoglobulin IGHG3 allele (Gmb0,b1,c3,c5,u) with an IGHG4 converted region and three hinge exons. *Immunogenet* 1989; **30**: 250 – 257.
39 Dard P, Huck S, Frippiat J *et al*: The IGHG3 gene shows a structural polymorphism characterized by different hinge lengths: sequence of a new 2-exon hinge gene. *Hum Genet* 1997; **99**: 138 – 141.
40 Natvig J, Turner M: Localisation of Gm markers to different molecular regions of the Fc fragment. *Clin Exp Immunol* 1971; **8**: 685 – 700.
41 Van Loghem E, Blanc M, De Lange G: Human IgG3 allotypes, with special reference to a new allotype related to G3m(g) (G3m21). *J Immunogenet* 1977; **4**: 371 – 383.
42 Michaelsen T, Frangione B, Franklin E: The amino acid sequence of a human immunoglobulin G3m(g) pFc' fragment. *J Immunol* 1977; **119**: 558 – 563.
43 Alexander A, Steinmetz M, Barritault D *et al*: Gamma Heavy chain disease in man: cDNA sequence supports partial gene deletion model. *Proc Natl Acad Sci USA* 1982; **79**: 3260 – 3264.
44 Ito S, Suzuki K, Miyazaki T *et al*: Amino acid substitutions determining G3m(g5). *Exp Clin Immunogenet* 1990; **7**: 91 – 100.
45 Matsumoto H, Ito S, Miyazaki T *et al*: Structural studies of another human IgG3 myeloma protein (Kam) carrying the allotypic markers Gm(s,t) and its alteration induced by chemical modification; in Brinkmann BaH, K (ed): *Advances in Forensic Haemogenetics*. Berlin Heidelberg, Springer-Verlag, 1986, vol 1, pp 153 – 156.
46 Matsumoto H, Ito S, Miyazaki T *et al*: Structural studies of a human gamma 3 myeloma protein (Jir) bearing the allotypic marker Gm(st). *J Immunol* 1983; **131**: 1865 – 1870.
47 Williams RJ, Malone C, Solomon A: Conformational dependency of human IgG heavy chain-associated Gm allotypes. *Mol Immunol* 1993; **30**: 341 – 351.