

Les gènes des globines humaines : que nous apprend leur polymorphisme?

D. Charmot-Bensimon

Laboratoire de génétique et physiologie du développement, Centre universitaire de Marseille Luminy, Case 907, 13288 Marseille cedex 9, France. E-mail : bensimon@lgpd.univ-mrs.fr.

Summary: Human Globin Genes: What Can We Learn from Their Polymorphism?

Molecular analysis of human α and β globin genes reveals extensive polymorphism at these loci. Worldwide distribution of the sickle cell trait has been well known for some time. However, the molecular basis and distribution of thalassemia have been more recently studied. These are the commonest monogenic disorders. For most of them, α -thalassemia is due to single nucleotide substitutions, small deletions or insertions. They are very heterogeneous and widely dispersed in the Old World. β -thalassemia is mainly due to the deletion of one to four genes. On the whole, their distribution is quite similar to α -thalassemia. With some exceptions, both distributions coincide with present and past regions of malarious endemicity. On the other hand, when looking at individual mutations, no two regions are identical. The question of whether selection by malaria plays a role on observed allele frequencies is still a challenge. The only well clear instance is the S mutation, which causes sickle cell anaemia. The role of malaria is but one among other hypothesis for explaining thalassemia distribution and frequencies. A possible scenario could be the following: one (or a few) mutation happened in a population and spread because of its selective advantage, along with the founder effect and/or genetic drift. Migration, founder effect and genetic drift must be invoked to account for some observations. It is still difficult to say why a mutation is highly frequent in one population and not in another, even at equivalent malarial endemicity. On the other hand, many genes should contribute simultaneously, or in synergy, in the process of fighting against malaria. Fitness of each mutation could depend on its genetic background when the mutation arose. Selection must work on a set of genes. Populations who are living now, and genetically very different, could very well be the result of selection on many genes by many infectious agents.

Résumé :

L'analyse moléculaire des gènes α et β des globines humaines a souligné l'extrême polymorphisme de ces locus. Si la répartition géographique du trait drépanocytaire est connue depuis longtemps, la base moléculaire et la répartition des thalassémies sont des données plus récemment acquises. Les hémoglobinopathies constituent les maladies monogéniques les plus répandues. Les thalassémies sont essentiellement des mutations ponctuelles, très hétérogènes et largement répandues dans l'Ancien monde. Les thalassémies β sont essentiellement dues à une délétion d'un à quatre gènes alpha. Leur distribution recoupe celle des thalassémies α . Et l'ensemble coïncide avec la répartition actuelle ou passée du paludisme. Démontrer le rôle de la sélection par le paludisme sur la fréquence élevée atteinte par certaines mutations reste encore un challenge : cet effet n'est bien documenté que dans le cas de la drépanocytose. Ce rôle possible n'est ainsi qu'une hypothèse plausible pour expliquer la fréquence des thalassémies. Le scénario serait le suivant : une ou plusieurs mutations apparaissent dans une population, l'une d'elle se répand parce qu'elle présente un avantage sélectif, associée à un effet fondateur et/ou à la dérive génétique. D'autres observations ne peuvent être expliquées qu'en faisant intervenir migration, effet fondateur et dérive génétique. Il reste toutefois difficile d'expliquer pourquoi une mutation particulière atteint une fréquence appréciable dans une population et non pas dans une autre, à endémicité palustre comparable. Peut-être ne faut-il pas oublier que plusieurs gènes peuvent contribuer simultanément ou en synergie au processus de résistance au paludisme. La valeur adaptative de chaque mutation dépendrait alors du contexte génétique dans lequel elle survient. C'est une constellation de gènes qui serait soumise à la pression de la sélection. Les populations actuelles, dans leur diversité génétique, seraient le résultat de la sélection opérée par un ensemble d'agents infectieux sur un ensemble de locus.

Key-words: Human globin -
Genetics -
Polymorphism -
Natural selection -
Malaria -
Infectious disease

Mots-clés : Globine humaine -
Génétique -
Polymorphisme -
Sélection naturelle -
Paludisme -
Maladie infectieuse

Introduction

Depuis plusieurs décennies, l'hémoglobine, tétramère de chaînes protéiques 2×2 , est le paradigme de nombreux travaux en biologie. La drépanocytose (défaut qualitatif de la

chaîne β), les thalassémies α et β (défauts quantitatifs) comptent parmi les maladies monogéniques les plus répandues et, en cela, elles font exception à la règle de l'habituelle rareté de ces dernières. Depuis une cinquantaine d'années, l'idée d'un équilibre entre maladies infectieuses et maladies génétiques a

émergé. HALDANE (cité dans (3)), dès 1948, a émis l'hypothèse d'un polymorphisme équilibré pour interpréter la fréquence élevée des thalassémies dans le bassin méditerranéen. En effet, la coïncidence entre la répartition de la maladie thalassémique et la répartition du paludisme dans cette région était remarquable. En Afrique tropicale impaludée, la drépanocytose était soumise à la même analyse qui amenait à l'hypothèse suivante : l'avantage des hétérozygotes dans la résistance au paludisme serait le facteur maintenant le polymorphisme génétique. À la fin des années 60, cette idée était communément admise, mais difficile à transposer aux thalassémies. La mise au jour du remarquable polymorphisme des gènes de globine a relancé fructueusement la question du paludisme dans le maintien de la fréquence élevée de certains allèles thalassémiques. Seront donc examinées successivement : l'organisation des gènes de globine et les séries alléliques connues; puis la répartition et la fréquence de quelques variants informatifs seront présentées ainsi que l'interprétation que l'on est autorisé à fournir sur ces observations de terrain.

Organisation des gènes de globines en famille

L'hémoglobine humaine adulte est un tétramère de deux chaînes α et deux chaînes β (hémoglobine $\alpha_2\beta_2$), mais aussi de façon minoritaire d'une autre combinaison polypeptidique : $\alpha_2\gamma_2$ (5). D'autres types d'hémoglobine sont synthétisés durant le développement embryonnaire (tétramères $\alpha_2\zeta_2$, $\alpha_2\epsilon_2$, $\alpha_2\delta_2$), puis foetal, (tétramère $\alpha_2\gamma_2$). Tous les gènes fonctionnels codant ces différentes chaînes protéiques sont regroupés en famille (cluster : "agrégat") : (α , β , γ , δ , ϵ) et (α , G, A, β , δ). Les gènes sont organisés sur le même modèle à 3 exons et dérivent par duplications successives d'un ancêtre commun existant il y a environ 450 millions d'années. L'ordre des gènes, de 5' en 3', au sein de chaque complexe, reflète l'ordre de leur expression séquentielle au cours de l'ontogénèse (fig. 1).

Figure 1.

Organisation des familles des gènes α et β des globines humaines (d'après KAZAZIAN et ANTONARAKIS, (5)).
Organization of α -like and β -like human globin genes (5).

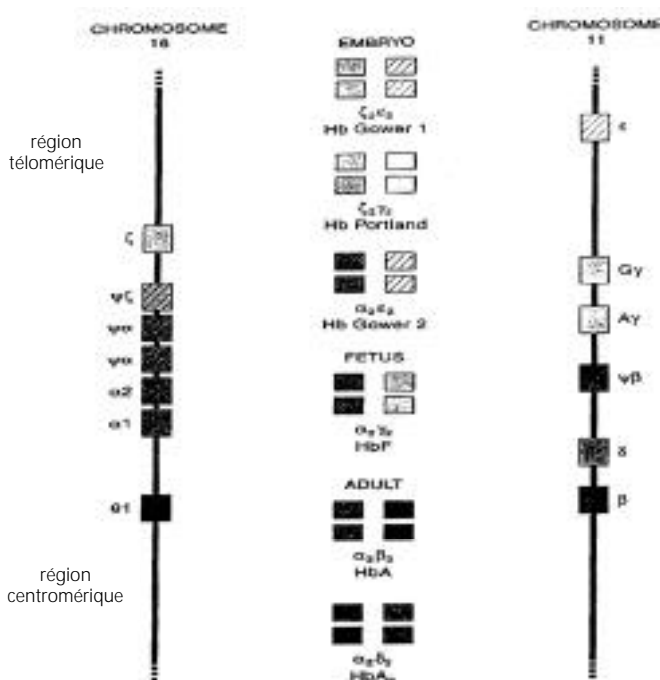
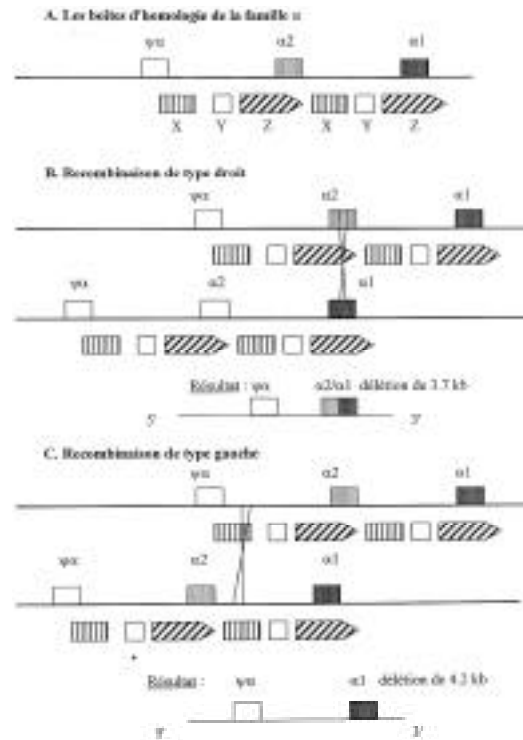


Figure 2.

Recombinaisons inégales entre les boîtes d'homologie Z et X au locus α .
A. Les boîtes d'homologie de la famille α ; B. Recombinaison de type droit; C. Recombinaison de type gauche.
Inequal crossing over within homology boxes in α genes cluster.
A. Homology boxes in α -like family; B. Right type of recombination; C. Left type of recombination.



La famille des gènes des α -globines

Cette famille est localisée sur la partie distale du bras court du chromosome 16 où elle occupe environ 30 kb. Le gène $\alpha 1$, le plus télomérique, est le premier exprimé durant l'embryogenèse. Les gènes $\alpha 2$ et $\alpha 1$ sont exprimés dès la vie foetale et continueront à fonctionner durant la vie adulte. Les séquences exoniques des gènes $\alpha 2$ et $\alpha 1$ sont identiques, ainsi que celles de leur 1er intron. Cette homologie de séquence serait le résultat de l'évolution concertée par conversion génique. Les gènes $\alpha 1$ et $\alpha 2$ sont eux-mêmes insérés dans deux régions de forte homologie, d'une taille de 4 kb, détaillées en trois boîtes X, Y et Z (fig 2A). Trois pseudogènes, $\alpha 3$, $\alpha 4$ et $\alpha 5$, s'intercalent entre $\alpha 2$ et $\alpha 1$. Une région cis-régulatrice a été identifiée à 40 kb en amont de $\alpha 1$. Nommée HS40, elle contrôle l'expression des gènes $\alpha 1$ et $\alpha 2$. Le phénomène de la commutation des gènes (le switch), c'est-à-dire le passage de l'expression du gène $\alpha 1$ à celle des gènes $\alpha 2$, au début de la vie foetale, n'est pas encore clairement décrypté.

La famille des gènes des β -globines

La famille des β -globines s'étend, elle, sur environ 50 kb à l'extrémité distale du bras court du chromosome 11. Le gène β , le plus en 5' du complexe, est le premier à être exprimé, durant la vie embryonnaire. Les gènes G et A s'expriment durant la vie foetale (hémoglobine F : $\alpha_2\gamma_2$); l'adulte présentera normalement moins d'1 % d'hémoglobine F. Leurs séquences exoniques sont identiques à une position près : le codon 136 (glycine pour la chaîne G et alanine pour la chaîne A). À nouveau, l'évolution concertée est évoquée pour expliquer le maintien d'une si forte homologie. Entre les paires G / A et β / δ est localisé un pseudogène de type β ($\beta 1$). L'expression du gène β commence dès la vie foetale et atteindra son plateau d'expression quelques mois après la naissance. Le gène β , dont l'ex-

pression débute seulement après la naissance, est faiblement transcrit. Il n'intervient que pour 2 à 3 % des tétramères (hémoglobine A2 : 2 2). De 6 à 20 kb en amont du gène γ , quatre sites, HS1 à HS4 constituent la région *cis*-régulatrice distale, ou LCR (Locus control Region), du complexe γ . La commutation des gènes de la famille γ , sous le contrôle des éléments du LCR entre autres, se fait en deux étapes : vers G et A, au début de la vie foetale, puis δ et ϵ dans la période périnatale.

Quelques remarques sur l'origine de ces familles géniques

Comment ces gènes se sont-ils constitués en "agrégats"? Sur un même chromosome, la duplication d'une copie ancestrale du gène est le point de départ de l'émergence d'une famille. Il est vraisemblable que les duplications ont joué un rôle fondamental en tant que mécanisme évolutif. Les recombinaisons inégales consécutives à une erreur d'appariement conduisent à dupliquer une séquence sur un chromosome et à éliminer du chromosome homologue. Ces erreurs d'appariement sont favorisées par l'existence de courtes séquences répétées. Une fois la 1ère duplication établie, une erreur d'appariement sera encore plus probable. On verra l'importance de ces recombinaisons en examinant les bases moléculaires des thalassémies. La duplication peut disparaître ou se fixer dans la population soit par dérive génétique soit sous l'effet de la sélection naturelle. Les deux gènes récemment dupliqués peuvent garder la même séquence (c'est le cas des gènes des ARN ribosomiques, autre exemple de famille multigénique), rester quasi identiques (c'est le cas des gènes de globine β et δ , ou G et A), soit peuvent diverger (δ et ϵ , ou δ et γ).

Les bases moléculaires des principales mutations

La clinique, accompagnée depuis une vingtaine d'années des techniques toujours plus performantes de la biologie moléculaire, a permis la constitution d'une riche collection de mutations qualitatives ou quantitatives (les deux effets sont parfois associés) dans les gènes β et δ . La drépanocytose et les thalassémies β ou δ sont les illustrations les plus représentatives des deux catégories d'altérations géniques. Les chiffres cités dans le "syllabus" de T. HUISMAN (<http://globin/cse/psu/edu>) annoncent 1046 variants environ, répertoriés en 1997, dont 260 pour les deux gènes β , 531 pour le gène δ , 69 pour les gènes γ , 45 pour ϵ . Le nombre des remaniements chromosomiques, des variants des régions régulatrices et des différents syndromes de persistance héréditaire de l'hémoglobine foetale (HPFH) s'élève à 141. Les anomalies se répartissent en 693 variants de structure, 332 variants thalassémiques et 21 HPFH. Seules quelques-unes de ces mutations sont suffisamment fréquentes pour être trouvées à l'état homozygote. Il faut peut-être voir dans l'organisation particulière des gènes γ (gènes dupliqués au sein de bloc d'homologie) et δ (un seul gène) l'explication de la répartition des types d'altérations. Les anomalies les plus fréquentes au locus β sont de type ponctuel, réparties sur la région codante et les régions régulatrices; le signe distinctif du locus β est la délétion d'un gène, ou bien des deux, en *cis*. Autant l'homozygotie pour une mutation à un locus unique pourra être aisée à repérer phénotypiquement (par la maladie), autant la perte de deux gènes β sur un ensemble de quatre passera phénotypiquement quasi inaperçue et sera identifiée le plus souvent de façon fortuite.

Les mutations au locus des globines

Ce sont les délétions, nous venons de l'évoquer, qui sont le plus fréquemment rencontrées. La thalassémie β est tenue pour la maladie monogénique la plus fréquente de par le monde. Les délétions constituent 95 % des cas de thalassémie β . Plusieurs génotypes correspondant à des phénotypes de sévérité graduée sont rencontrés :

4 gènes β / δ situation normale
 3 gènes β - / δ thalassémie β + hétérozygote (appelée aussi thalassémie β de type 2, hétérozygote)
 2 gènes β — / δ thalassémie β ⁰ hétérozygote, (*cis*), (thalassémie β de type 1, mineure)
 - / - thalassémie β + homozygote, (*trans*), (thalassémie β de type 2, homozygote, mineure)
 1 gène β — / - hémoglobine H, excès de chaînes β ; thalassémie β +, intermédiaire
 0 gène β — / — *hydrops foetalis*, léthal, excès de chaînes β ; thalassémie β ⁰, majeure.

La présence de zones de forte homologie englobant les gènes β et δ fournit une explication à l'origine des délétions d'un gène. En effet, des appariements irréguliers entre les boîtes X, Y et Z lors de la méiose, suivis de recombinaison, conduisent à la délétion d'un gène β sur un chromosome et à 3 copies β sur le chromosome homologue. L'appariement entre la boîte Z du gène β et celle du gène δ sur le chromosome homologue, dit du type droit (R : right; fig. 2B), la plus fréquente semble-t-il, entraîne une délétion de 3,7 kb et la formation d'un "gène de fusion", fonctionnel, $\beta\delta$ 1. L'appariement entre les boîtes X non homologues *stricto sensu*, dit de type gauche (L : left; fig. 2C), aboutit lui à une délétion de 4,2 kb, entraînant la totalité du gène β 2. Reste fonctionnel, sous le contrôle de son promoteur "faible", le gène δ 1. Huit exemples de délétions présentant des variantes des points précis de cassure sont dénombrés. La délétion la plus fréquente, - 3.7, est divisée elle-même en trois sous groupes I, II, et III en fonction de la localisation fine du point de cassure dans la boîte Z. À l'opposé, les délétions faisant perdre les deux gènes β et δ simultanément (en *cis*) sont de taille nettement plus hétérogènes et le mécanisme présidant à leur survenue est mal élucidé. Une vingtaine de ces délétions ont été caractérisées au niveau moléculaire. Les thalassémies non délétionnelles (une trentaine) sont rares. La mieux connue est due à la mutation Constant Spring (CS) : mutation dans le codon de terminaison (STOP -> glutamine) de la chaîne β 2. Toutes ces mutations sont récessives.

Les mutations au locus des globines

Elles présentent, nous l'avons évoqué, un spectre moléculaire très différent. Il s'agit essentiellement de mutations ponctuelles, récessives. La plupart de ces mutations, prises individuellement, sont rares. Seules trois atteignent des fréquences suffisantes pour être retrouvées à l'état homozygote : S₁, C₁ (hémoglobine C) et E₁ (hémoglobine E). Les patients présentant d'autres hémoglobinopathies seront des hétérozygotes composites, ou transhétérozygotes, pour deux mutations rares. Les thalassémies β (défaut quantitatif) sont donc majoritairement le résultat de mutations ponctuelles (179 en 1996). Les délétions sont minoritaires (n = 17, en 1996). Parmi elles, l'hémoglobine Lepore : à nouveau, du fait de l'homologie de séquences, un appariement erroné survient, au cours de la méiose, entre les gènes β et δ . Suivi de recombinaison, il produit un "gène de fusion" $\beta\delta$, fonctionnel et faiblement exprimé puisqu'il est sous le contrôle du promoteur β , dont on a vu qu'il était peu efficace. D'autres délétions ont apporté des infor-

mations essentielles à la compréhension de la régulation de l'expression et de la commutation géniques. Certaines d'entre elles éliminent simultanément les gènes α^1 et α^2 . Une région cis-régulatrice en 3' du gène α se trouve maintenant à proximité des gènes foetaux dont elle peut augmenter l'expression. Une thalassémie pourra alors être partiellement corrigée par un syndrome de persistance de l'hémoglobine foetale (HPFH). Les mutations ponctuelles qui entraînent un changement qualitatif ($n = 335$) ont été identifiées tout au long de la région codante. La plus fameuse est la substitution au codon 6, Glu \rightarrow Val, provoquant à l'état homozygote la maladie drépanocytaire. La mutation α^E , au niveau du codon 26 (1er exon; Glu \rightarrow Lys) active un site cryptique d'épissage. Lorsqu'il est utilisé, le messager auquel il donne naissance ne permet pas la traduction d'un polypeptide fonctionnel. Lorsque le site d'épissage habituel est utilisé, la protéine, qui comporte un acide aminé erroné, est cependant fonctionnelle et les homozygotes pour cette mutation sont asymptomatiques. Pour mémoire, parce que rares, des variants appelés "hyper instables" de la chaîne α existent et donnent des thalassémies à transmission dominante.

La répartition géographique des mutations

La possibilité d'effectuer l'identification des mutations par les techniques de la biologie moléculaire a donné l'impulsion à des recherches nouvelles et stimulantes dans le domaine médical bien sûr, mais aussi en génétique des populations dans une perspective évolutive (3, 4). À proximité des mutations, ont été répertoriés des polymorphismes des sites de coupures par les enzymes de restriction (RFLP) ou des polymorphismes de répétition de séquences (mini- et microsatellites...) qui constituent des marqueurs génétiques aisément manipulables. Cette nouvelle approche du génome a d'abord mis en avant la grande variabilité de l'espèce humaine. Elle a aussi permis d'étendre de façon considérable le nombre des marqueurs des familles α et β . Non seulement une mutation dans l'un des gènes pourra précisément être caractérisée, mais elle le sera dans un contexte génomique précis, par la détermination des marqueurs qui l'encadrent, sur quelques kilobases, et qui constituent un haplotype. Les populations humaines seront alors caractérisées par un ou des haplotypes, leur carte d'identité. On peut résumer de la façon suivante la répartition géographique des mutations les plus fréquentes :

La mutation α^S

Elle est fréquente surtout dans la bande équatoriale, zone où la pluie persiste la plus grande partie de l'année. Sa fréquence y dépasse 20 %, fréquence que l'on peut retrouver dans certaines oasis d'Arabie saoudite ou dans certaines parties de l'Inde. Dans les zones de savane africaine et le Sahel, la pré-

Figure 3.

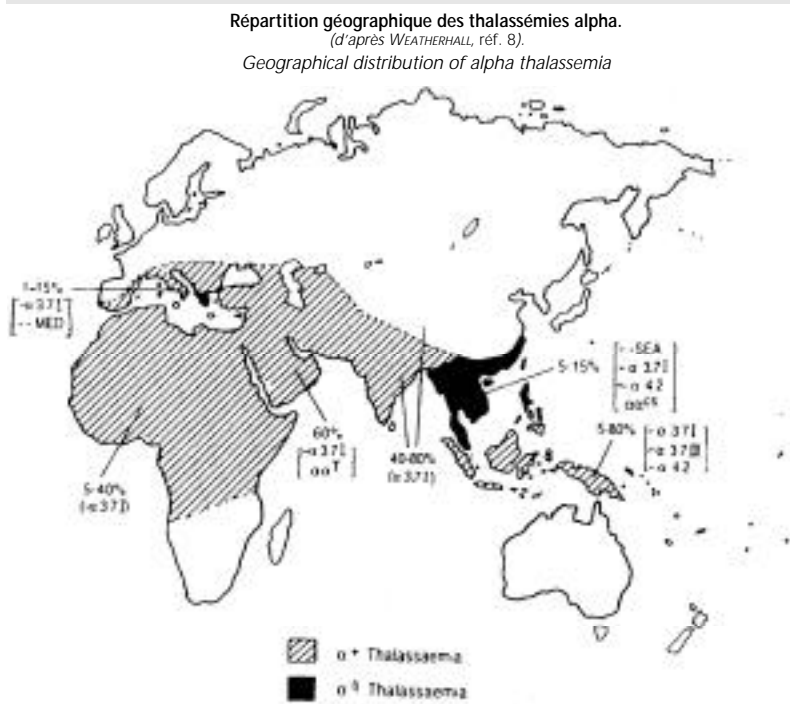
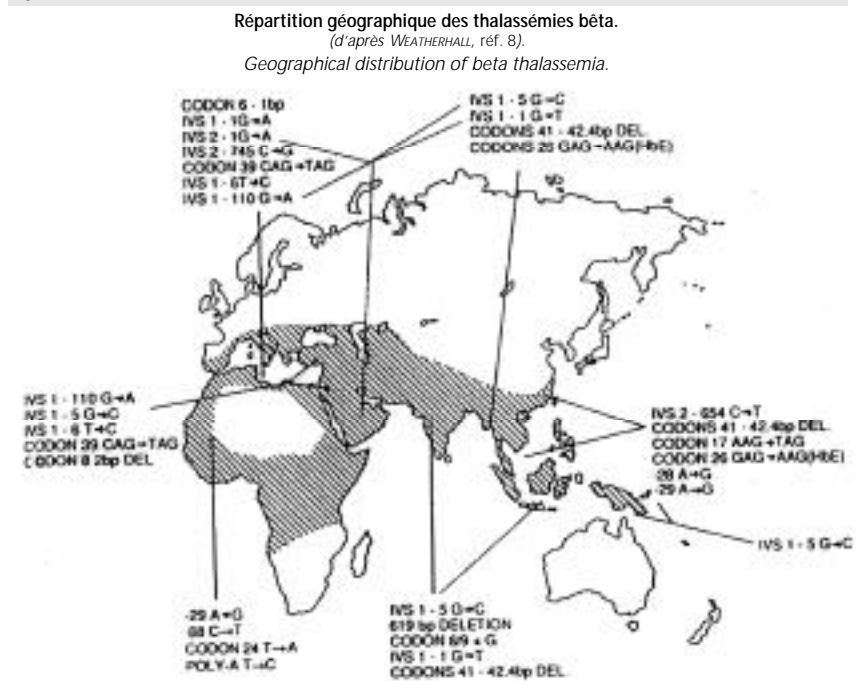


Figure 4.



valence de α^S est notablement plus faible. Sur le pourtour du bassin méditerranéen, au Moyen-Orient et en Iran, la fréquence ne dépasse pas 5 %. La mutation est aussi identifiée chez les Noirs américains, dans les Caraïbes et au Brésil. Elle est absente d'Asie.

La mutation α^C

On la retrouve quasi exclusivement en Afrique de l'Ouest, sur le plateau voltaïque, où sa distribution coïncide avec celle de la mutation α^S . Elle est également absente d'Asie.

Les allèles thalassémiques

Ils ont une distribution largement répandue dans l'Ancien monde (fig. 3). Chacune des mutations est prédominante dans une population donnée. Certaines sont communes à plusieurs

populations vivant dans des aires géographiquement et culturellement apparentées : la mutation IVS1 110 G- est caractéristique des pays du pourtour du bassin méditerranéen ; elle est inconnue en Asie du Sud-est. Les petites populations isolées ont un profil plus restreint de mutations, qui les différencie clairement par rapport aux autres populations.

La mutation E

Sa distribution concerne l'Asie, du nord de l'Inde à la Chine. Au nord de la Thaïlande, sa fréquence atteint presque 80 %, 50 % dans la population khmère.

Les allèles thalassémiques +

La délétion - est largement répartie elle aussi dans l'Ancien Monde (fig. 3). Sa fréquence atteint parfois 70 à 80 % (Népal, province d'Andhra Pradesh en Inde, côte nord de Papouasie-Nouvelle Guinée). Si l'on examine en détail la répartition des délétions -, l'apparente uniformité de sa distribution laisse place à une image en mosaïque: la délétion - 4.2 est caractéristique de l'Inde, de l'Asie du Sud-est et de la côte nord de Papouasie-Nouvelle Guinée; la délétion - 3.7III est trouvée exclusivement en Océanie (Micronésie, Polynésie); la délétion - 3.7I est particulièrement fréquente en Afrique, au Népal, en Inde et en Méditerranée.

Les allèles thalassémiques 0

La forme la moins rare de délétion affectant les deux gènes en cis, appelée —SEA, est caractéristique des populations thaï, vietnamienne, chinoise et des Philippines. La délétion dite —MED est la marque des populations italiennes, grecques et sardes.

Les allèles thalassémiques non délétionnels

Le plus répandu est l'allèle CS, connu en Chine du sud, Thaïlande, Cambodge, Vietnam et Laos.

Comment interpréter la variabilité des gènes de globines?

Globalement, la géographie des zones d'endémie actuelles ou passées du paludisme est superposable à la géographie des principales hémoglobinopathies de par le monde. La "malaria belt" des zones tropicales est bien connue. En Europe, *P. vivax* a sévi, avec une transmission saisonnière (été et début de l'automne) dans les zones marécageuses jusqu'au début, voire le milieu, du XXe siècle, entraînant une mortalité non négligeable. Dans le sud, il a coexisté avec *P. falciparum*. La répartition géographique même (la "malaria belt") suggérerait plutôt que seul ce dernier ait pu constituer une pression de sélection, réserve faite d'un rôle bénéfique précoce d'une infection par *P. vivax* sur le devenir d'une infection ultérieure par *P. falciparum* (vide infra).

Chaque région où sévit le paludisme a "son" anomalie moléculaire (ceci n'est pas valable pour le Nouveau Monde), mais deux régions d'endémicité équivalente n'ont pas les mêmes variants, ou la même association de variants. Par ailleurs, on relève des mutations (-, par exemple), à des fréquences notables, dans des régions où le paludisme n'a jamais été endémique. À l'opposé, en Amérique, où le paludisme est d'introduction récente (*P. vivax* à la fin du XVème siècle, amené par les Européens; *P. falciparum*, à la fin du XVIème siècle, apporté par le trafic d'esclaves), aucune hémoglobinopathie n'atteint de fréquence comparable à certaines de celles qui sont observées dans les régions impaludées. Aucune mutation thalassémique n'est observée chez les Indiens d'Amérique.

On dispose de nombreuses preuves de l'élimination rapide des mutations délétères. L'existence d'un polymorphisme dans

une population soulève alors la possibilité de son maintien par un mécanisme actif. La sélection naturelle est une hypothèse séduisante, mais les preuves en faveur d'une sélection positive restent rares (3, 4, 7, 8).

Les mutants de structure S, C, E

L'exemple princeps d'une mutation défavorable, létale avant l'âge de la reproduction, maintenue par un avantage sélectif de l'hétérozygote, est bien sûr celui de la mutation S. Sa prévalence forte en zones de pluviosité sur une longue période correspond à une modalité de transmission continue du paludisme. En revanche, cette prévalence est faible dans les régions aux saisons alternées ou au Sahel. L'avantage consisterait en une meilleure résistance aux accès sévères du paludisme à *Plasmodium falciparum*. En Afrique noire, la mortalité par *P. falciparum* affecte surtout les jeunes enfants, avant qu'ils n'aient acquis une immunité, au moins partielle, vis-à-vis des clones circulants dans la région considérée. C'est dans cette tranche d'âge que s'exercerait l'effet protecteur de l'allèle S. Les expériences *in vitro* fournissent une base physiologique vraisemblable aux mécanismes de résistance. Cette mutation est considérée comme l'exemple typique d'un polymorphisme équilibré, et c'est le cas le mieux documenté à ce jour. Elle serait survenue à 4 reprises sur le seul continent africain. Sa présence en Méditerranée, avec une fréquence modeste, est aussi explicable par les migrations de population (invasion, trafic d'esclaves) en provenance de l'Afrique. Le paludisme à *P. vivax* ou *P. falciparum* a existé dans plusieurs régions des Etats-Unis (essentiellement le sud) jusque vers 1943. Sa pression sélective était sans doute modeste car la fréquence des hétérozygotes (env. 7 %) est nettement inférieure à celle observée dans les régions d'où venaient les esclaves. En Arabie, dont le climat s'est progressivement désertifié durant les dix derniers mille ans, ce serait plutôt la persistance d'un niveau élevé d'hémoglobine F, et non plus l'endémicité paludéenne, qui maintiendrait la présence de l'allèle S dans la population. On ne peut fournir autant d'arguments pour expliquer la fréquence de deux autres mutations structurales du gène, à savoir C et E (défaut qualitatif et quantitatif). En effet, ni l'une ni l'autre ne sont létales à l'état homozygote, et les homozygotes E / E sont asymptomatiques. Dans le cas de l'hémoglobine C, on suppose que la fréquence de C est aussi due à l'avantage de l'hétérozygote, mais également en raison peut-être d'un avantage de l'hétérozygote composite C / S (cité dans 3). La fréquence élevée d'un second gène avantageux, dans une population où il existe déjà un gène (au moins) avantageux, trouverait ainsi une explication. Et, de fait, de nombreuses populations exposées au paludisme endémique présentent, associées, plusieurs anomalies génétiques de la structure et de la fonction des globules rouges.

Les thalassémies

Si le variant S est l'illustration classique d'un polymorphisme équilibré, dans quelle mesure ce modèle peut-il être étendu aux autres variants? La mutation E, la signature de l'Asie, est présente dans les régions impaludées, absente des régions limitrophes non impaludées, et elle est observée très fréquemment en association avec d'autres défauts génétiques du globule rouge (cité dans 3). Le rôle exercé par le paludisme sur la fréquence de la mutation E est donc singulièrement difficile à évaluer. La série des allèles thalassémiques est longue; les informations sur la répartition sont plus abondantes pour les populations européennes que pour le reste du monde. Des données précises sur la distribution et la prévalence du paludisme dans le passé font malheureusement défaut. Il est donc très difficile de relier la fréquence d'une mutation à la pré-

valence de la maladie, à l'exception de l'étude menée par SINISCALCO en Sardaigne, dans les années 1960 (cité dans 3), montrant la variation simultanée des deux paramètres. On ne peut pas procéder autrement qu'en comparant les fréquences actuelles des mutations aux prévalences hypothétiques du paludisme dans le passé. Chaque variant est spécifiquement associé à un haplotype de restriction, ou un très petit nombre, 2 à 3, dans une région donnée. Il paraît vraisemblable que la plupart des mutations ne sont survenues qu'une seule fois et récemment. La redistribution de chacune sur quelques haplotypes est le résultat de la recombinaison, réciproque ou non réciproque, durant un laps de temps assez bref. Mais la comparaison des haplotypes de restriction portant l'une ou l'autre des deux mutations les plus fréquentes en Méditerranée est d'interprétation malaisée. Deux cent soixante et onze chromosomes sur 311, de 12 pays européens, portent la mutation IVS1 : 110 sur le même haplotype de restriction défini par 7 sites. La mutation au codon 39, elle, est distribuée sur 14 haplotypes de restriction (4). Faut-il voir dans l'existence actuelle de ces haplotypes la marque de l'intervention d'autres facteurs, génétiques ou non? Des différences phénotypiques subtiles expliqueraient-elles que des haplotypes soient distribués plus rapidement que d'autres? Peut-on imaginer que la composition haplotypique d'une population influence la fréquence de recombinaison ou de conversion génique? Autant de questions non résolues à ce jour illustrant les difficultés d'interprétation des données lorsque l'on sort du "cas" de la drépanocytose. On ne peut ni exclure ni démontrer le rôle de la sélection dans le maintien sur le pourtour de la Méditerranée des allèles thalassémiques. D'autres paramètres interviennent très certainement : taille de la population, dérive génétique, migrations...

Les allèles thalassémiques

L'étude de la distribution des thalassémies est liée aux progrès des techniques de biologie moléculaire, puisque les tests hématologiques standards ne permettent pas une estimation précise du génotype. Les populations d'Asie du Sud-est et du Pacifique sont la principale source d'informations sur les délétions des gènes. On dispose pour le Pacifique de données sur la prévalence du paludisme avant la mise en place des programmes d'éradication. FLINT, en 1986 (cité dans 3), a montré que la prévalence du paludisme et des délétions variaient dans le même sens, rappelant les observations faites en Sardaigne 20 ans plus tôt, tandis que d'autres polymorphismes moléculaires variaient aléatoirement. Ces délétions ont des répartitions géographiques précises. Elles sont portées par des haplotypes de restriction très différents, et il semble bien que ces mutations soient apparues à plusieurs reprises dans des fonds génétiques différents; elles peuvent atteindre des fréquences remarquables dans certains cas. La fréquence de délétion - / atteint 70 % dans le nord de la Papouasie-Nouvelle Guinée (PNG), 39 % dans l'île de Espiritu Santo de l'archipel des Vanuatu, tandis qu'elle est égale ou voisine de zéro dans certaines îles de Polynésie ou Micronésie. Trois sous-types de délétion ont été trouvés, chacun prédominant dans une région. La délétion - 4.2 prédomine sur la côte nord de la Papouasie-Nouvelle Guinée, tandis que - 3.7I est au sud de l'île et que - 3.7III occupe l'archipel de Vanuatu. Ces délétions - existent dans certaines régions du Pacifique, où il n'y a jamais eu de paludisme; elles peuvent y atteindre une fréquence de 12 %. En Polynésie, c'est la délétion - 3.7III qui est présente sur 95% des chromosomes mutants. La préhistoire et l'histoire du peuplement des îles du Pacifique peuvent apporter quelques clés. Les peuplements de la Polynésie et de la Micronésie sont récents : 3 à 4 000 ans sans doute, tandis qu'en Mélanésie, l'occupation humaine remonterait peut-être à 40 000

ans. L'hypothèse est donc la suivante : des vagues de migration depuis l'Asie du Sud-est ont dispersé les populations soit vers Madagascar (où l'on retrouve la mutation ^E) soit vers la Papouasie-Nouvelle Guinée, la Mélanésie et la Polynésie. Le nord de l'archipel de Vanuatu est considéré comme le lieu de naissance de la délétion - 3.7III (3, 4) où elle aurait été sélectionnée pour l'avantage qu'elle confère à ses porteurs. Ce seraient ensuite les migrants qui l'auraient disséminée sur leur chemin vers la Polynésie jusque dans sa partie la plus orientale. Au niveau de chaque île, où l'effectif colonisateur a dû être réduit, la fréquence de la mutation - 3.7III, en l'absence de pression de sélection dans ces îles non impaludées, peut présenter des fluctuations aléatoires tout à fait importantes, allant jusqu'à l'élimination de l'allèle. Dans ces îles, ce serait donc l'effet fondateur qui expliquerait les fréquences observées sur le terrain. Il faut être prudent quant à la possibilité de généraliser ce modèle qui repose sur l'analyse de populations dans un contexte historique et géographique précis. On ne dispose pour les thalassémies — (délétion en cis) et les - thalassémies non délétionnelles, d'aucune donnée comparable. Autant l'hypothèse d'un polymorphisme équilibré rend compte de la mutation ^S, autant la situation des délétions en Asie semble correspondre à un état transitoire, allant vers la fixation de la délétion dans les populations concernées (cité dans 3, 4). Trancher entre un avantage sélectif et un effet fondateur reste hardi. Des enquêtes récentes sur le terrain ont apporté un nouvel éclairage à la question. De façon apparemment paradoxale, la délétion homozygote favorise l'infestation précoce des enfants par *P. falciparum* et *P. vivax*. Mais c'est ce qui permettrait le développement ultérieur d'une meilleure immunité non spécifique vis-à-vis de *P. falciparum* et aussi, en "bénéfices secondaires", envers d'autres agents infectieux, et ce, chez les homozygotes uniquement. C'est en ce sens que concluent de façon originale, WILLIAMS en 1996 (9) et ALLEN en 1997 (1). Mais les bases biologiques de la résistance au paludisme restent étayées moins solidement, comparativement à celles de la mutation ^S. Le locus ^S, exprimé dans les cellules érythrocytaires, hôtes de parasites, semble particulièrement plastique. Son organisation propre (duplication des gènes ^S, présence de séquences répétées, existence de boîtes d'homologie) permet, sans dommage majeur pour la physiologie du globule rouge, la perte de deux gènes sur quatre; événement dont les conséquences s'avèrent à terme bénéfiques pour la survie. La protection vis-à-vis du paludisme est dans ce contexte un "produit dérivé", inattendu, de la structure du matériel génétique à ce locus. La répartition géographique des mutations n'est pas complètement claire. Comment expliquer l'absence de ^S en Asie? Celle de ^E en Afrique? Faut-il invoquer un assortiment d'allèles particulier à la population où survient la nouvelle mutation, participant de la valeur adaptative de ^S et ^E? C'est en effet, très probablement, sur l'ensemble des allèles présents que la pression de sélection s'exerce.

Cela implique la prise en compte d'un autre paramètre : l'impact des interactions géniques sur le phénotype et sur la fréquence de tel allèle ou (2, 6). À titre d'illustration, on se rappellera que, dans certaines régions d'Asie du Sud-est, seulement 15 % de la population est dépourvue d'une anomalie génétique du globule rouge. On a évoqué plus haut la présence simultanée en Afrique de l'Ouest des variants ^S et ^C. Il semble aussi que la présence d'une thalassémie favorise le maintien dans une population de l'allèle ^S, de même que le syndrome HPFH (par mutation dans l'un des promoteurs), améliore le devenir des enfants drépanocytaires. La co-transmission des thalassémies et est également observée (cité dans 2). L'impact d'autres gènes est soupçonné et devrait être évalué :

le déficit en G6PD, le polymorphisme des groupes sanguins, le déficit en bande 3, entraînant l'ovalocytose héréditaire des Mélanésiens... Des gènes du système immunitaire, du métabolisme du fer, des cations (7, 8) doivent aussi être examinés. La variabilité génétique de l'agent de sélection putatif, le *Plasmodium*, s'ajoute à la variation probable de son pouvoir infectieux dans l'espace et dans le temps. L'analyse de ces interactions hôte - parasite s'avère particulièrement complexe.

L'étude des globines montre la diversité génétique remarquable de l'espèce humaine sur 80 kb de son génome (3 x 10⁹ paires de base pour le génome haploïde). L'image actuelle de la distribution des allèles à ces locus est-elle le résultat de leur rôle joué dans la défense vis-à-vis du paludisme dans les quelques derniers 5000 ans? Il n'est pas invraisemblable de l'imaginer. Est-ce valable pour certains? Pour tous? La preuve reste encore à apporter car, exception faite du cas de *S*, les arguments restent basés plus sur l'observation de données en termes de population que sur des données expérimentales. De plus, cette défense contre *P. falciparum* et (?) *P. vivax* s'est sans doute exercée et, à nouveau, il n'est pas interdit de l'imaginer, par le maintien de polymorphismes à d'autres locus. Le "modèle des globines" est-il généralisable? Dans ce cas, d'autres pathogènes auront entraîné eux aussi le maintien d'une diversité génétique, à d'autres locus. On pense bien sûr à l'exemple de la mucoviscidose, de la maladie de TAY SACHS... Il est plausible que les agents pathogènes aient contribué, en introduisant une pression de sélection, à la variabilité de notre fond génétique : cette exploration démarre.

Références bibliographiques

1. ALLEN SJ, O'DONNELL A, ALEXANDER NDE, ALPERS MP, PETO TEA *et al.* - α -thalassemia protects children against disease caused by other infections as well as malaria. *Proc Natl Acad Sci USA*, 1997, **94**, 14736-14741.
2. CAO A, GALANELLO C & ROSATELLI MC - Genotype-phenotype correlations in α -thalassemia. *Blood Review*, 1994, **8**, 1-12.
3. FLINT J, HARDING RM, BOYCE AJ & CLEGG JB - The population genetics of the haemoglobinopathies. In: HIGGS D & WEATHERHALL DJ (Eds) - *The Haemoglobinopathies*. Bailliere's Clinical Haematology, Bailliere Tindall, London, 1993, **6**, pp 215-262.
4. FLINT J, HARDING RM, CLEGG JB & BOYCE AJ - Why are some genetic diseases common? Distinguishing selection from other processes by molecular analysis of globin gene variants. *Hum Genet*, 1993, **91**, 91-117.
5. KAZAZIAN HH & ANTONARAKIS S - Molecular genetics of the hemoglobin genes. In: SINGER M & BERG P (Eds) - *Exploring genetic mechanisms*. University Science Book, Sausalito, California, 1997, pp. 301-336.
6. LABIE D & ELION J - Modulation polygénique des maladies monogéniques : l'exemple de la drépanocytose. *Méd/Sc*, 1996, **12**, 341-349.
7. NAGEL RL & ROTH EF Jr - Malaria and red cell genetic defects. *Blood*, 1998, **74**, 1213-1221.
8. WEATHERHALL DJ - Host genetics and infectious disease. *Parasitology*, 1996, **112**, S23-S29.
9. WILLIAMS TN, MAITLAND K, BENNETT S, GANCZAKOWSKI M, PETO TEA *et al.* - High incidence of malaria in α -thalassemia children. *Nature*, 1996, **383**, 522-525. Résistance de *M. tuberculosis* aux antibacillaires dans la Province Ouest du Cameroun