

ヒトゲノム研究の新しい地平

斎藤 成也*

国立遺伝学研究所集団遺伝研究部門

(平成21年5月7日受付, 平成21年5月11日受理)

要 約

20世紀末にはじまったゲノム研究は、ヒトゲノム解読をひとつの到達点とした。しかしそれは終わりではなく、始まりだった。ヒトゲノム配列をもとにして膨大なSNPやマイクロサテライト多型の研究が急激に進み、小数の古典多型マーカーを用いた従来の研究成果を追認しつつ、日本列島人の遺伝的多様性についても新しい光が当てられつつある。また個々人のゲノム配列を決定する研究も進展している。これらヒトゲノム研究の新しい地平を紹介した。

キーワード：個人ゲノム, DNA多型, SNP, 日本列島人, 集団ゲノム学

New Horizon of Human Genome Study

Naruya Saitou*

Division of Population Genetics, National Institute of Genetics

(Received 7 May 2009; accepted 11 May 2009)

Abstract

Genomic studies initiated at the end of the 20th Century reached one peak when the human genome was sequenced. It was, however, not the end, but the starting point. Studies based on enormous numbers of SNP and microsatellite DNA polymorphism exploded, and a new light also shines on genetic diversity of people of Japanese Archipelago, while confirming results based on small scale classic polymorphic markers. Sequencing individual human genomes is also going on. These new horizon of human genomics was introduced.

Key words: personal genome, DNA polymorphism, SNP, People on Japan Archipelago, population genomics

はじめに

われわれは、人間の顔を見ただけで、その人のおおよその出自集団を推定することができる。両親の出自集団

が、たとえば東ユーラシア人と西ユーラシア人（それぞれ、Saitou (1995) が提唱した East Eurasian と West Eurasian の日本語訳）というように大きく異なっていれば、混血であることをかなり高い精度で推定することができる。

*国立遺伝学研究所集団遺伝研究部門
〒411-8540 三島市谷田1111
E-mail: saitounr@lab.nig.ac.jp
©2009 The Anthropological Society of Nippon

これは人間が高いパターン認識能力を持っているからであるが、同時に顔の形態に関連している遺伝子数がかかなりあり、それらの多型が複雑に関与しているからだともいえるだろう。これらの、形態に関与する遺伝子がわかっていれば、遺伝子の情報から顔面形態を推定することが可能である。残念ながら、現在まだその段階にはいたっていない。動物の発生メカニズムがまだ完全には解明されていないからだ。

遺伝子のデータはどうだろうか。今から20年ほど前の1990年に行なわれた、先史モンゴロイドプロジェクト(赤澤威代表)に参加していた人類学者7名による座談会(石田ら, 1990)のなかで、著者は形態学的形質の高い個体認識力が「非常にうらやましい」と述べた。当時は、すでにDNA多型が少しずつ研究されてはいたが、まだ多くの場合、血液型や赤血球酵素、血清タンパク質といった、いわゆる古典的マーカー(classic marker)が数十座位調べられていたにすぎない。これらの情報によって、人類集団間の遺伝的近縁関係は推定できたものの(たとえばNei and Roychoudhury, 1993; Omoto and Saitou, 1997)、個人個人の出自集団の推定は困難だった。

21世紀に入ってヒトゲノムの塩基配列が大部分明らかにされてから、この状況は一変した。量が質に変換される典型である。ヒトだけでなく、DNA配列やアミノ酸配列を比較して生物の進化を研究する分子進化の研究は、現在ではゲノム進化学に発展しつつある(斎藤, 2007)。本稿では、このようなゲノム規模の膨大なDNAデータからどのようなことがわかりつつあるのかを紹介するとともに、近い将来研究がどのように展開してゆくのかを考えてみたい。

ヒトゲノムにおける遺伝的多型

生物が持つ遺伝的多様性は、種内・種間のどちらも存在するが、本稿では主としてヒトという種の中での種内変異を考える。ヒトゲノム内の種内変異は、突然変異の性質によって、大きく塩基置換と挿入欠失に分けられる。塩基置換は、4種類の塩基間の置き換えであり、最小単位は1塩基である。これがもっとも多く、1塩基多型(Single Nucleotide Polymorphism; SNP)の大部分は塩基置換タイプである。また、塩基が置き換わる突然変異率はきわめて低いので、たいていの塩基サイトでは、4種類の塩基のうち2塩基だけがヒト集団中に存在しており、2対立遺伝子多型(biallelic polymorphism)となっている。もっとも、地球上に生きている60億人以上の人間全員のゲノム配列を比較すれば、ほとんどのSNPは3種類以上

の塩基が存在しているだろう。

塩基置換と異なるもうひとつの突然変異のタイプが挿入欠失である。これはすくなくとも4種類に細分される。第一は微少な挿入欠失であり、1塩基からせいぜい10塩基程度である。塩基の長さが長くなるほど突然変異率が低いことが知られている(Saitou and Ueda, 1994)。第二は短い塩基の繰り返し(リピート)数の変化である。繰り返しの単位が数塩基である場合には、STR(Short Tandem Repeat)多型、あるいはマイクロサテライト多型を生じる。ヒトゲノム中に数万種類が知られている。リピートの長さが数十塩基の場合にはVNTR(Variable Number of Tandem Repeat)あるいはミニサテライト多型と呼ばれる。さらにもっと繰り返しの単位が長く、しかも繰り返し単位の間にも別の配列が入っている場合には、CNV(Copy Number Variation)と呼ぶ。

このほかに、組換えや逆位、遺伝子変換も、広い意味での突然変異と考えてよいだろう。ヒトゲノムの遺伝的多様性は、これらさまざまな種類のDNA多型が複合されたものであり、それらの一部は表現型の個体差にも寄与している。たとえば、耳垢の乾型と湿型のちがいは、細胞膜貫通型タンパク質の一種であるABCC11の1アミノ酸の差によっており、非同義サイトにおけるSNPが原因である(Yoshiura et al., 2006)。

古典マーカーにもとづく日本列島人の位置

従来は、血液型や蛋白多型などの、いわゆる古典マーカーを調べて人類集団を比較することが行なわれていた。これらの遺伝的多型は、主として塩基置換に基づくものだが、欠失挿入などのタイプも含まれていた。例として、日本本土人、沖縄人、アイヌ人、韓国人4集団について、25個の多型座位の対立遺伝子頻度情報から集団間の遺伝距離を推定した距離行列(Omoto and Saitou, 1997)を系統ネットワークで表示した結果(Saitou, 2008)を図1に示す。横長の長方形は、左右の辺が{日本本土人+韓国人}グループと{沖縄人+アイヌ人}グループとの遺伝的違いを示す。この辺は同時に、沖縄人とアイヌ人がお互いに遺伝的に近いことを示している。すなわち、アイヌ・沖縄同系論およびその現代的名称である日本人の二重構造モデル(Hanihara, 1991)を支持する結果となっている。

短い前後の辺は{韓国人+アイヌ人}グループと{日本本土人+沖縄人}グループとの遺伝的違いを示し、辺の長さは遺伝距離に比例している。長方形の頂点から韓国人とアイヌ人に伸びる斜めの枝も、遺伝距離に比例し

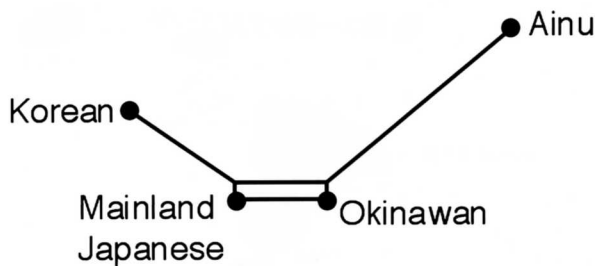


図1 日本本土人、沖縄人、アイヌ人、韓国人のあいだの系統ネットワーク (Saitou, 2008) より

た長さを示している。長方形の4頂点に日本本土人と沖縄人が位置しているのは、両者が韓国人で代表される大陸集団（渡来人の源郷）とアイヌ人で代表される縄文人集団との混血集団であることを示唆している。沖縄人は混血の程度が本土日本人ほど強くなかったため、アイヌ集団とグループを形成すると読み取ることができる。なお、図1を掲載した著者の論文 (Saitou, 2008) を、故塩原和郎氏に捧げた。

DNA 多型データに基づく日本列島周辺の人類集団間の遺伝的近縁関係

人類集団の遺伝的多型は、分子生物学技術の発展にとともに、1970年代にはDNAデータを少しずつ用いるようになっていった。当初は制限酵素断片長多型 (RFLP; Restriction Fragment Length Polymorphism) が用いられたが、やがて多数の集団で、進化速度が高いミトコンドリアDNAの部分配列、特に変異性の高いDループとかコントロール領域と呼ばれる非コード領域の塩基配列決定が進められるようになった。日本でも、故宝来聰のグループが活発に多数の人類集団について塩基配列決定を行なった。たとえば、台湾の原住民9集団の比較 (Tajima et al., 2003) や、日本人およびその周辺集団の比較 (Tajima et al., 2004) があげられる。アイヌ人のデータを周辺集団と比較した研究では、ニブヒヤコリヤークといった沿海州の原住民がアイヌ人とミトコンドリアDNAのタイプを共有している場合がかなりあることがわかった (Tajima et al., 2004)。

核DNAにおいては、遺伝的個体差が高く、法医学的研究も進められたマイクロサテライトDNA多型がよく用いられるようになった。我々の研究グループは、名古屋大学の法医学のグループとともに、東アジアを中心とする10人類集団の遺伝的近縁性を推定した (Li et al., 2006)。図2にその結果が示してある。沖縄と名古屋の日本列島2集団が、ブーツストラップ確率100%という緊

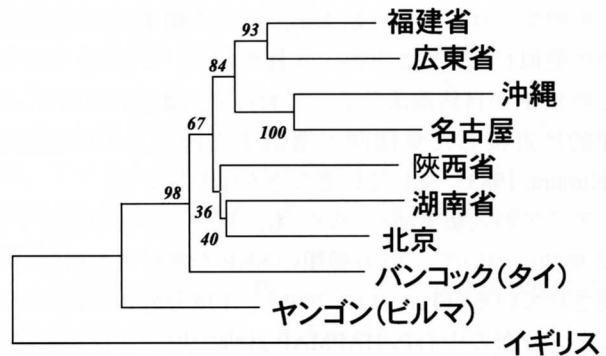


図2 東アジアと東南アジアを中心とする10人類集団の遺伝的近縁図 (Li et al., 2006 より)

密なクラスターを形成し、この日本クラスターは、福建省と広東省の、いわば中国沿岸集団クラスターと近縁となっている。その外側には、北京、陝西省、湖南省の3集団から形成される、中国内陸部クラスターが位置する。すなわち、日本列島人の遺伝的多様性は、中国5集団の多様性の中に含まれてしまっていることになる。そのさらに外側に東南アジアの2集団（タイのバンコックとビルマのヤンゴン）が存在し、最後にアジアの集団ではないイギリスの集団が外群として位置づけられている。

膨大なSNPデータに基づく人類集団間の遺伝的近縁関係の推定

マイクロサテライトDNA多型研究の発展と平行して、DNAチップを用いた大規模SNP解析の研究も進展してきた。これはもちろんヒトゲノム塩基配列のほぼ全体が決定されたことによって可能となったのである (International Human Genome Sequencing Consortium, 2001, 2004)。2005年には日本人45名を含む4人類集団について、90万近いSNPが調べられた (International HapMap Consortium, 2005)。これらの成果を利用して、現在では膨大なSNPを調べて、それらと病気などの関連を調べる、ゲノム規模の連関研究 (GWAS; Genome Wide Association Study) が多数行なわれている。

人類集団間の遺伝的近縁性の研究も、大規模なSNPデータをもとに進められている。世界の51人類集団、合計938人のDNAを65万個のSNPについて調べて解析した研究 (Li et al., 2008) は、集団の系統関係についてはマイクロサテライトDNA多型の分析結果 (Rosenberg et al., 2002) とほぼ同じ結果を得ている。

欧州の多数の人類集団のSNPデータを生成し、解析した研究 (Novembre et al., 2008) では、SNPの対立遺伝子頻度の差異を主成分分析して、第1主成分と第2主成分

の座標をプロットした結果が、欧州人類集団の地理的分布に酷似していることが示されている。これは、ほとんどの SNP が自然淘汰を受けておらず、遺伝的浮動や、地理的に近接した集団間の遺伝子交流などの中立進化 (Kimura, 1983) をしていることを示している。

アジアの人類集団についても、東アジアを中心とする 22 集団について、20 万種類の SNP を解析した研究が発表されている (Tian et al., 2008)。主成分分析の結果を見ると、分布の中心に HAPMAP 計画で用いられた北京の中国人が位置し、日本人集団は、そこから左上に位置する。北京の中国人と日本人のほぼ中央に韓国人集団が位置している。一方、中国南部の少数民族集団 (ミャオ, シェ, トゥージェ) は、北京の中国人からは右側に分布しており、日本人からは遠くなっている。中国北部の少数民族集団 (ダウール, オロチョン, ホジェン) は逆に、北京の中国人からは左側に位置しており、左右を示す第 1 主成分からすると日本人や韓国人と同じ側に属するが、第 2 主成分は日本人や韓国人が上に位置するのに対して、これら北部の少数民族集団は下に位置している。この主成分分析の結果も、欧州集団と同じように、ほぼ集団の地理的分布に近いことが示されている。結局のところ、日本本土人にもっとも近いのは韓国人であり、ついで中国北部の漢族であるという、これまでの結果が追認されたことになる。

また Pan-Asian SNP Consortium がさらに多くのアジアの人類集団について 5 万種類の SNP データを生成し解析しているが、現時点 (2009 年 5 月) では論文発表がされていない。近い将来論文やデータが公開されれば、アジアにおける人類集団の遺伝的近縁性の研究が大きく進展することが期待される。

膨大な SNP データに基づく日本列島人集団間の 遺伝的近縁関係の推定

膨大な SNP データを用いて、日本列島人を詳細に解析した研究が、理化学研究所の山口由美、鎌谷直彦らによって、2008 年に発表された (Yamaguchi-Kabata et al., 2008)。この研究の示すことをここで考察してみたい。山口らは、自己申告で日本人と名乗った 7003 名の DNA について、140,387 箇所の SNP を比較した。まず、HAPMAP 計画で調べられた欧米人 (西ユーラシア人)、北京の中国人と東京の日本人 (東ユーラシア人)、ナイジェリア人 (アフリカ人) とこれら 7003 人の SNP を、主成分分析で解析した。その結果、図 3 で示すように、大部分は東ユーラシア人のクラスターに含まれたが、個体 1 と個体 2 はここ

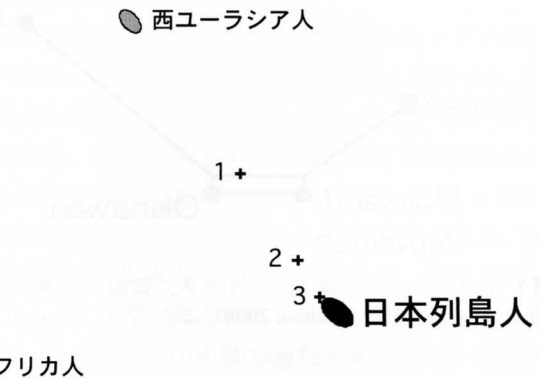


図 3 日本人と西ユーラシア人、アフリカ人の相対的位置関係 (Yamaguchi-Kabata et al., 2008 の図をもとにした)

から少し離れて、西ユーラシア人のクラスターに近づいていた。これらは個体 1 が両親のどちらかが西ユーラシア人であり、個体 2 は祖父母のうちの一が西ユーラシア人である可能性が高い。個体 3 も少しだけ西ユーラシア人の方向に位置しているが、東ユーラシア人クラスターに含まれているとも言える。そこで、山口らは個体 1 と個体 2 を以下の解析から除外した。

なお、ここでいう「クラスター」とは、ある地域に居住する人間の大部分が主として分布している、主成分分析のプロットにおける領域を指している。地域名がついているが、その地域の人々がすべてそのクラスター内に位置していない場合もあり、また他の地域に居住している人間のなかの小数がそのクラスター内に位置する場合もある。

次に、「日本人」7,001 人と HAPMAP で解析された東ユーラシア人 90 人で共通する SNP データを主成分分析した結果、図 4 に示した結果が得られた。「日本人」の大部分は、本土クラスターと琉球クラスターの位置に分布したが、そのほかに中国クラスター (HAPMAP 計画で用いられた北京の中国人はみなここに位置する) に数名が、また中国クラスターと本土クラスターの間位置する部分に数十名が分布した。これらは本土日本人と中国人の混血である可能性もあるが、おそらく韓国系日本人ではないかと思われるので、図 4 では「韓国クラスター？」とした。もっとも多数の「日本人」が分布する本土クラスターのまわりには、3 種類の混血クラスター A, B, C が位置している。それぞれ、韓国系と思われる人々と本土人との混血、琉球系と本土人との混血、さらに、混血クラスター C は、以下に述べる理由から、本土人の系統に一部アイヌ人の系統が混血した個体である可能性がある。

この研究では、7000 人以上の「日本人」が以下の 7 地

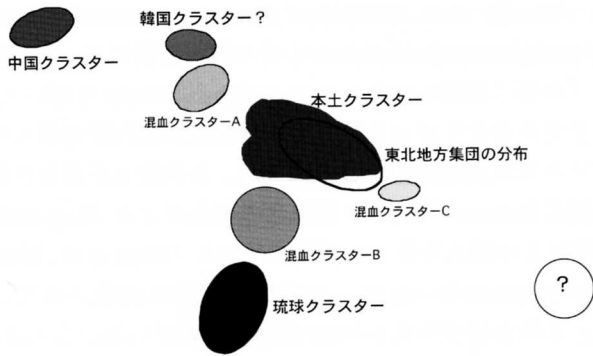


図4 日本列島人の SNP 解析の結果 (Yamaguchi-Kabata et al., 2008 の図をもとにした)

域に分割された：北海道，東北，関東甲信越，東海北陸，近畿，九州，沖縄。中国四国地域はサンプルが得られていない。沖縄以外の地域の「日本人」は大部分本土クラスターに分布したが，沖縄人だけは，琉球クラスターに分布した人が大部分であった。このために，このクラスターに「琉球」という名称を Yamaguchi-Kabata et al. (2008) は用いたのである。沖縄人の一部は，本土クラスターにも位置していた。これは，最近本土から沖縄に移住してきた人々の子孫と考えることができる。逆に，九州人（現在九州在住の人）の一部には，琉球クラスターに位置していた個体があった。これは沖縄出身の人が九州に移住した場合を示していると考えられる。同じ傾向が，近畿人と関東甲信越人でもみられた。それぞれ大都市を含んでいるので，沖縄からの移住者が多いことの反映だろう。

北海道，関東甲信越，東海北陸の各地域居住者については，本土クラスターの中央を中心に分布しているが，そのほかの地域は若干の偏りが見られている。近畿人は，本土クラスターの左側に偏り，東北人は逆に本土クラスターの右側に偏っている。本土クラスターは全体として三角形に近い形であり，右側がとんがっているが，この右端の部分に分布するのは，東北人のほかには，北海道人と関東甲信越人が小数存在するだけである。しかも，これら東日本あるいは北日本の集団に特異的な位置は，第1主成分でやや下になっており，琉球クラスターと同じ傾向である。

これらのことを考え合わせると，東北地方クラスターの下部分，さらには混血クラスターCは，図4の右下に疑問符で示したクラスターの影響を受けていると想定することができよう。このクラスターは，第1主成分の座標では琉球クラスターと似通っているが，第2主成分では中国クラスターと対極に位置している。第2主成分

では本土クラスターも琉球クラスターも，中国クラスターとこの謎のクラスターの間位置する。この謎のクラスターは，もちろん今回の7000名余の「日本人」サンプルには含まれていないので，仮想的なものではあるが，アイヌ人が同じ手法で調べられたとしたら，ここに位置するだろうと著者が考えて，想定位置を示したものである。

この理化学研究所の研究とは別に，東京大学医学部の徳永勝士の研究室は，日本人400人について，約91万個のSNPを決定した (Nishida et al., 2008)。今後，これら膨大なSNPデータが日本列島の周辺人類集団と比較できるようになれば，量が質に変換されて，従来のDNA多型の研究とははるかに解像度が増した日本列島人の遺伝的歴史が明らかになるだろう。

集団ゲノム学の誕生

SNPの研究とは視点が違うが，やはり大規模なDNA研究として，ゲノム全体を決定するという視野がある。52人のミトコンドリアDNAゲノム配列を完全に決定して比較した研究 (Ingman et al., 2000) のころから，集団ゲノム学 (population genomics) という呼び名が誕生したが，その嚆矢は，1個人のアフリカ人のミトコンドリアDNAゲノム配列を完全に決定して，当時知られていた唯一のヒトミトコンドリアDNA完全配列 (いわゆるケンブリッジ配列) と比較した，故宝来聰らの研究である (Horai et al., 1995)。その後，田中雅嗣ら (Tanaka et al., 2004) は841人の日本人についてヒトミトコンドリアDNA完全配列を決定し，詳細な解析を行なった。

ミトコンドリアDNAは16500塩基前後の短いものだが，細胞核内の染色体には，合計で30億余塩基 (狭い意味でのヒトゲノム) が2セット存在する。最初に決定されたヒトゲノムは，染色体のうちの片方だけをランダムに選んで決定された，1倍体 (haploid) で，しかも数人のDNAから由来するモザイクだった。したがって，次の段階は，二倍体であるヒト個人のゲノムを決定することになる。

ここで，ミトコンドリアDNAやY染色体という，人類の系統を研究するのにこれまでよく用いられてきたDNAと常染色体のDNAとの違いを簡単に説明しておきたい。ミトコンドリアDNAは母系遺伝であり，Y染色体は父系遺伝であるが，常染色体のDNAは双系遺伝である。このため，ある個人の常染色体2対は，それぞれが母親と父親から由来している。母親と父親も，同様に母方祖母と母方祖父から，父方祖母と父方祖父から常染色

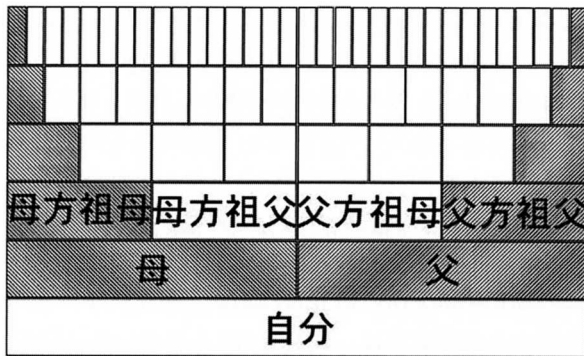


図5 ヒトが倍数体であるために生じる個人の系図と各祖先のゲノム貢献度

体を受け継いでいる。祖先を遡るとその繰り返しとなるので、 N 世代前には 2^N 人の祖先が存在する。単純な計算からすれば、それぞれの祖先が、現在の1個人に $1/2^N$ のゲノムDNAを伝えていることになる(図5)。実際には常染色体は22対であり、また組換えがランダムに生じるので、ひとりひとりの祖先が伝えるDNA量はばらつきが生じることになる。しかし、少なくとも数十人、あるいは数百人の祖先のDNAがモザイクになったものが、現在生きている人間ひとりひとりのゲノムDNAなのである。このような様相を考えると、人間社会でよく見られる世襲には、遺伝的な要素はほとんどないと言っていだろう。あくまでも世襲とは社会制度であり、生物学的な意義はないのである。ミトコンドリアDNAやY染色体については、遠い過去のある個人(それぞれ女性と男性)から受け継いだDNAが現在の個人にもそのままつなげられている可能性があるが(図5で影をつけた部分)、個人のゲノムの大部分は、多数の祖先のゲノムDNAがモザイクとなっているのである。

個人ゲノム配列の決定

最初に個人のゲノムDNA決定を行なったのは、民間企業としてヒトゲノム計画を推し進めたVenterのグループであり、しかも彼自身のゲノムを決定している(Levy et al., 2007; <http://huref.jcvi.org/>)。ここで用いられたのはこれまでどおり、サンガー法に基づくキャピラリーシーケンサーであるが、効率よく決定したので、予算は数億円だったとされている。次に、いわゆる第二世代シーケンサーとして最初に登場した454を用いたのが、米国政府を中心とする国際ヒトゲノム決定計画(最新版のヒトゲノム塩基配列データ取得は、たとえば以下のデータベースから:http://www.ensembl.org/Homo_sapiens/Info/Index)を当初率いていたWaston個人のゲノム決定であ

る(Wheeler et al., 2008; <http://jimwatsonsequence.cshl.edu/cgi-perl/gbrowse/jwsequence/>)。その数ヶ月後には、別のタイプの第二世代シーケンサーであるSolrexaを用いて、アフリカのナイジェリア人とだけ発表された匿名個人のゲノムである(Bentley et al., 2008)。この論文が発表された同じNatureの号には、癌患者の個人ゲノム(Ley et al., 2008)と中国人男性YHの個人ゲノム(Wang et al., 2008; <http://yh.genomics.org.cn/>)の論文も同時に掲載されている。これらはすべてSolrexaが用いられている。さらに、まだ現時点で論文は発表されていないが、オランダのある女性研究者および韓国人のある男性研究者の個人ゲノムがすでに決定されている。今後は、日本人も含めて、続々と個人ゲノムが決定されてゆくだろう。

これらの、個人個人のゲノム配列を決定する研究には、どのような意義があるのだろうか。まず、最終的なDNAにおける個人差を記述するためには全ゲノムを知る必要がある。現在ではゲノム配列を決定しなくても、100万種類以上のSNPをDNAチップなどの方法で簡易に調べることができるが、これらは個体差の一部にすぎない。SNPの基礎である塩基置換タイプの突然変異によって生じたヒトゲノムの個人差は、0.07%と推定されているので、血縁関係にない2個体間では、およそ420万塩基サイト(=0.07×0.01×3×10⁹×2)で塩基が異なっていることになる。比較する個体ペアによって異なるサイトが変動するので、100万ヵ所をいわば「つまみ食い」するDNAチップを用いた方法と異なり、すべての差を示すことができる。

また最初のほうで説明したように、DNAに生じる突然変異は塩基置換のほかには挿入欠失がある。これらの遺伝的個体差を知るには、個体それぞれのゲノム配列を決定して比較する必要がある。

最初に個体のゲノムが決定されたVenterの二倍体ゲノム配列の場合、常染色体上に3,213,401個のヘテロ接合サイト(個体内のSNP)が存在し、それらの1/3以上にあたる1,288,319個はそれまでにSNPが知られていない場所にあった。挿入欠失については、1塩基から最長8万塩基になる欠失部位が85万ヵ所存在し、さらに逆位となっている部分が90ヵ所存在した(Levy et al., 2007)。

2番目に決定されたWatsonの個体ゲノム配列についても、SNPについてはほぼ同じ330万個発見され、そのうち1万余サイトはアミノ酸の変化を伴っていた。これらの違いは発現するタンパク質の個体差を生じ、ひいてはタンパク質の機能にも影響する可能性がある(Wheeler et al., 2008)。このあたりの解析は、Venterゲノムで詳細

に行われており、Watson とほぼ同じアミノ酸の変化を伴っていた1万余サイトの20%程度は人類集団中でまれなものだった (Ng et al., 2008)。頻度が低いということは、そのようなアミノ酸置換が生存に若干有害である可能性がある。実際に、1500 個の非同義置換はタンパク質の機能を変化させる可能性があり、これらはヘテロ接合になっていたり、これまでに見つかっていなかったサイトであったりした。

以上はかなり詳細に個人のゲノム配列を決定するという計画だが、HAPMAP 計画の拡大版として、1000 人ゲノム計画 (<http://www.1000genomes.org/>) も進められている。こちらは、一人一人については完全にゲノム配列を決定するのではなく、70～80%だけ決定するという方針である。

最後に、個人のゲノム規模における研究において生じ得る倫理的問題と個人情報の保護について論じる。遺伝性の病気の原因遺伝子を探る研究では、以前から患者個人の情報を保護する方策がなされてきた。日本では文部科学省、厚生労働省、経済産業省が合意した「ヒトゲノム・遺伝子解析研究に関する倫理指針」(http://www.mext.go.jp/b_menu/houdou/17/01/05012101/001.htm) を公開している。

しかし、理論的研究の進展により、これまで多数の人間のデータを統計的に示したにすぎないと考えられてきた SNP の対立遺伝子頻度データが、SNP のサイト数が膨大になることによって、その中にある特定個人の DNA が含まれているかどうかを調べることが可能になることが示された (Homer et al., 2008)。したがって、これまで公開されず厳密に管理されてきた個人個人の DNA データだけでなく、集団の対立遺伝子頻度情報も簡単には公開できなくなってきたのである。

この、ゲノム規模の連関研究における個人データ非公開方針は、ヒトゲノム計画の公開方針とはまったく反対である。ヒトゲノム計画が始まろうとしたとき、塩基配列を決定したら24時間以内に公的塩基配列データベースである DDBJ/EMBL/GenBank に登録するというルールが確立した。この、研究者や納税者に広く開かれた姿勢は、その後の個人ゲノム決定研究にも引き継がれており、しかも最初の2名 (Venter と Watson) を含む何名かの個人ゲノムデータは個人名が明らかにされている。個人情報の保護という観点からみると、これは大きな問題点なのである。DNA サンプルを提供した個人だけでなく、そのゲノム DNA の1/2は親や子、兄弟と共有しているからである。血縁の程度が離れるにしたがって、1/4, 1/8, 1/16

と低くなってゆくが、ある個人の血縁者のゲノム配列の一部も、その個人のゲノム配列を決定することで、わかってしまうのである。

このため、個人情報を守るためには、個人個人のゲノム配列データを公開することは、仮にその個人が匿名化されていても、するべきではない。なぜならば、将来別の個人のゲノムデータが決定されたとき、その人がたまたまゲノムデータが公開されているある匿名の個人の血縁者であれば、すくなくとも自分の血縁者だということと比較によって推定できてしまうからだ。ましてや、公開されているデータの個人名を明らかにしてしまっている Venter や Watson の場合は、きわめて問題であることは明白である。

一方で、貴重なデータを他の研究者と共有しないという姿勢も、将来の研究でのデータ利用を閉ざしてしまい、よくない。人間自身の問題になるので、倫理や個人情報の保護が大きな観点となるが、それをきちんと認識して適切に対応しつつ、かつ研究全体の利益を考えて行動するべきであろう。個人ゲノム研究の時代には、そのようなバランスのとれたデータの扱いが求められているのである。

謝 辞

本稿を執筆する機会を与えていただいた松浦秀治編集委員長に御礼申し上げます。いくつかの図の作画をしていただいた水口昌子さんに感謝します。

引用文献

- Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L., Bignell H.R., Boutell J.M., Bryant J., Carter R.J., Keira Cheetham R., Cox A.J., Ellis D.J., Flatbush M.R., Gormley N.A., Humphray S.J., Irving L.J., Karbelslavili M.S., Kirk S.M., Li H., Liu X., Maisinger K.S., Murray L.J., Obradovic B., Ost T., Parkinson M.L., Pratt M.R., Rasolonjatovo I.M., Reed M.T., Rigatti R., Rodighiero C., Ross M.T., Sabot A., Sankar S.V., Scally A., Schroth G.P., Smith M.E., Smith V.P., Spiridou A., Torrance P.E., Tzonev S.S., Vermaas E.H., Walter K., Wu X., Zhang L., Alam M.D., Anastasi C., Aniebo I.C., Bailey D.M., Bancarz I.R., Banerjee S., Barbour S.G., Baybayan P.A., Benoit V.A., Benson K.F., Bevis C., Black P.J., Boodhun A., Brennan J.S., Bridgham J.A., Brown R.C., Brown A.A., Buermann D.H., Bundu A.A., Burrows J.C., Carter N.P., Castillo N., Chiara E., Catenazzi M., Chang S., Neil Cooley R., Crake N.R., Dada O.O., Diakoumakos K.D., Dominguez-Fernandez B., Earnshaw D.J., Egbujor U.C., Elmore D.W., Etchin S.S., Ewan M.R., Fedurco M., Fraser L.J., Fuentes Fajardo K.V., Scott Furey W., George D., Gietzen K.J., Goddard C.P., Golda G.S., Granieri P.A., Green D.E., Gustafson D.L., Hansen N.F., Harnish K., Haudenschild C.D.,

- Heyer N.I., Hims M.M., Ho J.T., Horgan A.M., Hoschler K., Hurwitz S., Ivanov D.V., Johnson M.Q., James T., Huw Jones T.A., Kang G.D., Kerelska T.H., Kersey A.D., Khrebtukova I., Kindwall A.P., Kingsbury Z., Kokko-Gonzales P.I., Kumar A., Laurent M.A., Lawley C.T., Lee S.E., Lee X., Liao A.K., Loch J.A., Lok M., Luo S., Mammen R.M., Martin J.W., McCauley P.G., McNitt P., Mehta P., Moon K.W., Mullens J.W., Newington T., Ning Z., Ling Ng. B., Novo S.M., O'Neill M.J., Osborne M.A., Osnowski A., Ostadan O., Paraschos L.L., Pickering L., Pike A.C., Pike A.C., Chris Pinkard D., Pliskin D.P., Podhasky J., Quijano V.J., Raczky C., Rae V.H., Rawlings S.R., Chiva Rodriguez A., Roe P.M., Rogers J., Rogert Bacigalupo M.C., Romanov N., Romieu A., Roth R.K., Rourke N.J., Ruediger S.T., Rusman E., Sanches-Kuiper R.M., Schenker M.R., Seoane J.M., Shaw R.J., Shiver M.K., Short S.W., Sizto N.L., Sluis J.P., Smith M.A., Ernest Sohna Sohna J., Spence E.J., Stevens K., Sutton N., Szajkowski L., Tregidgo C.L., Turcatti G., Vandevondele S., Verhovskiy Y., Virk S.M., Wakelin S., Walcott G.C., Wang J., Worsley G.J., Yan J., Yau L., Zuerlein M., Rogers J., Mullikin J.C., Hurles M.E., McCooke N.J., West J.S., Oaks F.L., Lundberg P.L., Klenerman D., Durbin R., and Smith A.J. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456: 53–59.
- Hanihara K. (1991) Dual structure model for the population history of the Japanese. *Japan Review*, 2: 1–33.
- Homer N., Szlinger S., Redman M., Duggan D., Tembe W., Muehling J., Pearson J.V., Stephan D.A., Nelson S.F., and Craig D.W. (2008) Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*, 4: e1000167.
- Horai S., Hayasaka K., Kondo R., Tsugane K., and Takahata N. (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Sciences USA*, 92: 532–536.
- Ingman M., Kaessmann H., Pääbo S., and Gyllensten U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, 408: 708–713.
- International HapMap Consortium (2005) The haplotype map of the human genome. *Nature*, 437: 1299–1320.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, 409: 860–921.
- International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature*, 431: 931–945.
- 石田肇・尾本恵市・齋藤成也・徳永勝士・百々幸雄・山口敏・赤澤威 (1990) モンゴロイドの誕生と拡散—日本人とは (座談). *モンゴロイド*, 8号, pp. 2–9.
- Kimura M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F., Denisov G., Lin Y., MacDonald J.R., Pang A.W., Shago M., Stockwell T.B., Tsiamouri A., Bafna V., Bansal V., Kravitz S.A., Busam D.A., Beeson K.Y., McIntosh T.C., Remington K.A., Abril J.F., Gill J., Borman J., Rogers Y.H., Frazier M.E., Scherer S.W., Strausberg R.L., and Venter J.C. (2007) The diploid genome sequence of an individual human. *PLoS Biology*, 5: 2113–2144.
- Ley T.J., Mardis E.R., Ding L., Fulton B., McLellan M.D., Chen K., Dooling D., Dunford-Shore B.H., McGrath S., Hickenbotham M., Cook L., Abbott R., Larson D.E., Koboldt D.C., Pohl C., Smith S., Hawkins A., Abbott S., Locke D., Hillier L.W., Miner T., Fulton L., Magrini V., Wylie T., Glasscock J., Conyers J., Sander N., Shi X., Osborne J.R., Minx P., Gordon D., Chinwalla A., Zhao Y., Ries R.E., Payton J.E., Westervelt P., Tomasson M.H., Watson M., Baty J., Ivanovich J., Heath S., Shannon W.D., Nagarajan R., Walter M.J., Link D.C., Graubert T.A., DiPersio J.F., and Wilson R.K. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, 456: 66–72.
- Li J.Z., Absher D.M., Tang H., Southwick A.M., Casto A.M., Ramachandran S., Cann H.M., Barsh G.S., Feldman M., Cavalli-Sforza L.L., and Myers R.M. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319: 1100–1104.
- Li S.L., Yamamoto T., Yoshimoto T., Uchihi R., Mizutani M., Kurimoto Y., Tokunaga K., Jin F., Katsumata Y., and Saitou N. (2006) Phylogenetic relationship of the populations within and around Japan using 105 short tandem repeat polymorphic loci. *Human Genetics*, 118: 695–707.
- Nei M. and Roychoudhury A.K. (1993) Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution*, 10: 927–943.
- Ng P.C., Levy S., Huang J., Stockwell T.B., Walenz B.P., Li K., Axelrod N., Busam D.A., Strausberg R.L., and Venter J.C. (2008) Genetic variation in an individual human exome. *PLoS Genetics*, 4, e1000160.
- Nishida N., Koike A., Tajima A., Ogasawara Y., Ishibashi Y., Uehara Y., Inoue I., and Tokunaga K. (2008) Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. *BMC Genomics*, 9: 431–440.
- Novembre J., Johnson T., Bryc K., Kutalik Z., Boyko A.R., Auton A., Indap A., King K.S., Bergmann S., Nelson M.R., Stephens M., and Bustamante C.D. (2008) Genes mirror geography within Europe. *Nature*, 456: 98–101.
- Omoto K. and Saitou N. (1997) Genetic origins of the Japanese: A partial support for the “dual structure hypothesis”. *American Journal of Physical Anthropology*, 102: 437–446.
- Rosenberg N.A., Pritchard J.K., Weber J.L., Cann H.M., Kidd K.K., Zhivotovskiy L.A., and Feldman M.W. (2002) Genetic structure of human populations. *Science*, 298: 2381–2385.
- 齋藤成也 (2007) *ゲノム進化学入門*. 共立出版, 東京.
- Saitou N. (2008) Genetic Relationships of Human Populations in and around the Japanese Archipelago. In: Matsumura S., Forster P. and Renfrew C. (eds.), *Simulations, genetics and human prehistory*. McDonald Institute Monographs, Cambridge, pp. 89–92.
- Saitou N. and Ueda S. (1994) Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Molecular Biology and Evolution*, 11: 504–512.
- Tajima A., Cheih-Shan Sun C.-S., Pan I.-H., Ishida T., Saitou N., and Horai S. (2003) Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Human Genetics*, 113: 24–33.
- Tajima A., Hayami M., Tokunaga K., Juji T., Matsuo M., Marzuki S., Omoto K., and Horai S. (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal

- lineages. *Journal of Human Genetics*, 49: 187–193.
- Tanaka M., Cabrera V.M., González A.M., Larruga J.M., Takeyasu T., Fuku N., Guo L.J., Hirose R., Fujita Y., Kurata M., Shinoda K., Umetsu K., Yamada Y., Oshida Y., Sato Y., Hattori N., Mizuno Y., Arai Y., Hirose N., Ohta S., Ogawa O., Tanaka Y., Kawamori R., Shamoto-Nagai M., Maruyama W., Shimokata H., Suzuki R., and Shimodaira H. (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Research*, 14: 1832–1850.
- Tian C., Kosoy R., Lee A., Ransom M., Belmont J.W., Gregersen P.K., and Seldin M.F. (2008) Analysis of East Asia genetic substructure using genome-wide SNP arrays. *PLoS One*, 3: e3862.
- Wang J., Wang W., Li R., Li Y., Tian G., Goodman L., Fan W., Zhang J., Li J., Zhang J., Guo Y., Feng B., Li H., Lu Y., Fang X., Liang H., Du Z., Li D., Zhao Y., Hu Y., Yang Z., Zheng H., Hellmann I., Inouye M., Pool J., Yi X., Zhao J., Duan J., Zhou Y., Qin J., Ma L., Li G., Yang Z., Zhang G., Yang B., Yu C., Liang F., Li W., Li S., Li D., Ni P., Ruan J., Li Q., Zhu H., Liu D., Lu Z., Li N., Guo G., Zhang J., Ye J., Fang L., Hao Q., Chen Q., Liang Y., Su Y., San A., Ping C., Yang S., Chen F., Li L., Zhou K., Zheng H., Ren Y., Yang L., Gao Y., Yang G., Li Z., Feng X., Kristiansen K., Wong G.K., Nielsen R., Durbin R., Bolund L., Zhang X., Li S., Yang H., and Wang J. (2008) The diploid genome sequence of an Asian individual. *Nature*, 456: 60–66.
- Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X.Z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., and Rothberg J.M. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452: 872–877.
- Yamaguchi-Kabata Y., Nakazono K., Takahashi A., Saito S., Hosono N., Kubo M., Nakamura Y., and Kamatani N. (2008) Population structure of Japanese based on SNP genotypes from 7,001 individuals in comparison to other ethnic groups: Effects on population-based association studies. *American Journal of Human Genetics*, 83: 445–456.
- Yoshiura K., Kinoshita A., Ishida T., Ninokata A., Ishikawa T., Kaname T., Bannai M., Tokunaga K., Sonoda S., Komaki R., Ihara M., Saenko V.A., Alipov G.K., Sekine I., Komatsu K., Takahashi H., Nakashima M., Sosonkina N., Mapendano C.K., Ghadami M., Nomura M., Liang D.S., Miwa N., Kim D.K., Garidkhuu A., Natsume N., Ohta T., Tomita H., Kaneko A., Kikuchi M., Russomando G., Hirayama K., Ishibashi M., Takahashi A., Saitou N., Murray J.C., Saito S., Nakamura Y., and Niikawa N. (2006) A SNP in the ABCC11 gene is the determinant of human earwax type. *Nature Genetics*, 38: 324–330.