

書 評

Chinese surnames and the genetic differences
between North and South China

Du Ruofu, Yuan Yida, Hwang Juliana, Mountain Joanna, and
Cavalli-Sforza L. Luca

Journal of Chinese Linguistics, Monograph Series No. 5 (1992) 93pp.

齋藤 成也

国立遺伝学研究所 進化遺伝研究部門

本モノグラフ「中国の苗字と南北中国の遺伝的差異」は、中国科学院遺伝研究所人類群
体遺伝研究室（日本流に言えば人類集団遺伝学研究室）を主宰しているDu Ruofu（杜若
甫）

教授の研究グループと、米国スタンフォード大学医学部遺伝学科のCavalli-Sforza教授の研
究グループとの共同研究の成果である。共著者の一人である Yuan Yida氏は、中国の苗字
に関する研究論文をこれまでにいくつか発表しており、1992年にシンガポールで開かれた
第1回中国言語学会では、中国の苗字に関する研究を招待講演の形で発表している。ま
た、共著者の Joanna Mountainさんは、Cavalli-Sforza教授の大学院生である。

私はたまたま杜若甫教授およびCavalli-Sforza教授とも個人的に知っており、特に杜若甫
教授の研究グループとは、以前文部省海外学術調査で共同研究を行なったこともある。私
自身、日本の苗字の研究をしたことがあり（齋藤 1983）、このたび上記のモノグラフの
書評（というよりも、紹介）を行なうことになった。興味深い論文を紹介していただいた
平田昌司先生に感謝する。以下では、本論文の抄訳を中心にしながら、私のコメント（`{`
マーク内に示した）をはさんでゆくスタイルをとる。数式の記述や集団遺伝学的な議論の
部分に関しては、想像される読者の興味の範囲外だと思われたので、かなり割愛した。興
味のある方は原著をご覧いただきたい。また、略していない部分の訳も、原文に忠実に沿
っているわけではないことをお断わりしておく。

はじめに

苗字の分析は、遺伝学において、いとこ結婚の割合から近親交配率を推定する方法にか
わる簡便法として、また距離に応じた遺伝的隔離の程度を調べるための情報源として用い
られてきた。しかし、我々の興味はそれらとは少し異なる。集団内の苗字の分布をある遺
伝子座の対立遺伝子頻度の分布と考えることによって、遺伝的な集団構造を解きあかすこ
とができるのである（たとえば Piazza et al. 1987）。本研究は、この問題意識にたって
我々

{Cavalli-Sforza教授のグループを指す}が行なってきた研究の一環である。

多くの文化では、苗字 (surname) は男系で伝わる。つまり、父親の苗字を子どもが伝えるのである。そのため、苗字の伝達は、父親から息子に伝わるY染色体とふるまいが似ている。もっとも、苗字は父親から娘にも伝わる場所が異なるが、娘が結婚して夫の苗字を名乗ればそこで父の苗字がとぎれる。苗字伝達様式の例外として、アイスランド人 (かつては北欧人およびケルト人でも) は父親の名前を息子の苗字としたので、苗字は毎世代変化することになる。{日本の「養子」制度は、血 (遺伝子) の結びつきと苗字の伝達の関係が壊れる一要因であろう。}

苗字が使われるようになった時代には、世界の中で様々な差がある。欧州の大部分においては、現在と同じ苗字を用いるようになったのは中世後期であり、ルネッサンスの初期に苗字の使用が全集団へ広がった。日本では前世紀まで苗字は使われなかった。{これが明らかな間違いであることは、日本の知識人ならすぐに指摘できるだろう。明治新姓は確かに日本の苗字の歴史で重要な局面ではあったが、それ以前、遡れば1500年前ごろから日本でも苗字にあたるものが使われていた。} 中国では、苗字の歴史はおそらく4000年以上になるだろう。これは苗字の使用としては世界の中でもっとも古い。{本当にそうであろうか? 中華思想のにおいを感じる。}

本研究は、1982年に中華人民共和国政府が行なった国勢調査から1/2000を層別任意抽出した苗字データをもとにしたものである。標本の単位は、「行政村」あるいは市街地の一部にあたる「居民委員会」である。こうして得られた苗字分布は、単性生殖する半数体の単一遺伝子座における、淘汰上中立な対立遺伝子の分布と考えて分析が行なわれた。{苗字が厳密に父から息子に伝わると仮定すると、男性のみを考えればよい (単性生殖)。また、苗字は一人が1個しか持たないから、倍数体の場合のホモ接合・ヘテロ接合などは考えなくてもよい (半数体)。苗字という「遺伝形質」1個だけを考えるので「単一遺伝子座」の取り扱いとなる。様々な文化的理由 (結婚における差別など) から、個人の持つ苗字によってその人が残す子孫の数が異なるかもしれないが、その場合は淘汰があることになる。しかし、とりあえず苗字によって子孫数が異なることはないと考えている (淘汰上中立)。}

関連論文として、「中国における遺伝子と言語の進化の合致」が中国語言語学報に発表されている (Mountain et al. 1992)。この論文は、中国語の研究に中国の苗字の分析が寄与できることについて論じている。言語との比較から得られた主要な結論は、言語距離と苗字距離の間に高い相関が見いだされたことである。これは、どちらの距離の場合にも人間の移住パターンが反映しているからであろう。遺伝子のデータは比較するにはまだまだ少ないが、遺伝子と苗字の伝達に平行関係があることを考えると、これもまた苗字のデータと高い相関のあることが期待される。この論文の第2の結論は、中国の苗字データと言語データが、地理的要因や歴史的要因よりもより高い相関が見いだされたことである。

1. 方法論：集団構造の研究への苗字の利用

a) 苗字数の分布

苗字は単一遺伝子座上の対立遺伝子として考えることができるので、対立遺伝子の分布を苗字分布に応用することができる。対立遺伝子の分布を研究するのは簡単ではない。それは、通常の方法では対立遺伝子数が5~6を越えることがないからであり、また、分析される個体数がふつうは少ないからである。しかし、苗字の場合にはこれらの制限はない。もっとも、苗字によって多元発生が起こっている場合には、これまで考慮されてこなかつ

た影響があるかもしれない。

単一遺伝子座上の対立遺伝子数の分布に関して、もっとも現実的なモデルは、Karlin と McGregor によって提唱されたものである。このモデルでは、固定された大きさ（染色体数あるいは半数体集団の大きさ） N の集団において、対立遺伝子数 r はランダムに分布している。各個体は一定の確率で死に、確率 $(1 - m)$ で同一の対立遺伝子を持つ個体が置き換わり、確率 m （個体あたり世代あたり）で突然変異により他の対立遺伝子のどれかを持つ個体が置き換わる。苗字の場合 r がきわめて大きい値となるので、対立遺伝子数が無限だと仮定することができる。

==中略==

Fisherが以前種数の分布を説明するのに提唱した対数分布が、Karlin-McGregor の分布のきわめてよい近似となっているので、実際の応用にはこのFisherの分布を用いた。 k 人が有する苗字の種類数 $n(k)$ の期待値は、以下の式で与えられる。

$$\text{式 [4]} \quad E[n(k)] = a (1 - v)^k / k$$

この式には a と v という二つの変数があるが、これらは標本数 N を介して以下の2式で示されるように関連している。

$$\text{式 [5]} \quad a = N v / (1 - v)$$

$$\text{式 [6]} \quad v = a / (N + a)$$

したがって、 a と v のどちらかひとつがわかれば、標本数 N を用いて他の変数を定めることができる。 a は「名字の多種類度」と呼ばれ、 v は突然変異率である。

b) a と v の推定

苗字の総種類数 S と標本数（個体数） N のあいだの関係は、対数分布においては以下のようなになる。

$$\text{式 [7]} \quad S / N = [v / (v - 1)] \log v$$

したがって、 S と N がわかれば、式 [7] を用いて v を定めることができる。ただし、式 [7] には解析解が存在しないので、数値解を求める必要がある。

==中略==

これまで、われわれは v を突然変異率と呼んできた。一般的には、突然変異を移住と置き換えたり、あるいは突然変異と移住の和と考えることができる。苗字に突然変異が起こることはまれにあるが、移住に比べるとずっと頻度が小さいので、無視することができる。

我々のこれまでの研究（例えば Piazza et al. 1987）によると、苗字の分布から推定された移住率は人口学的に推定された値と似ている。

vで推定される移住率は、苗字分布が平衡状態に達している場合には妥当性があるが、最近になって移住率に変化が見られた時には、この推定値は妥当ではない。移入が最近になって増えた場合には、まれな苗字の相対頻度が増加する。我々のサルジニア島の苗字分布の研究では、最近工業化が起こった地域でまれな苗字の数がやや増加していた。このような影響は、まれな苗字（1名から数名だけが持っている苗字）の部分を取り除いて分析すれば修正することができる。

このようにして推定された移住率は、世代あたりの集団への移入率である。平衡状態で、かつ人口増加のない状態であれば、これは移出率に等しい。

c) 遺伝学で用いられるNmとa（苗字）との関係 ー略

d) 別々の時代に行なわれた国勢調査から得た苗字資料からの移住率推定 ー略

2. 中国の苗字の歴史と本研究の原資料

中国における苗字の歴史は4000年以上前の夏王朝（およそ紀元前21-16世紀）まで遡る。苗字は中国においてそれよりもずっと以前に出現していたが、より古い苗字の記録が残っていないのである。もともとは、苗字は氏族名であったが、氏族名はさらに氏族のトーマあるいは地名から由来している。この風習は、中国の少数民族によっては、1950年代まで続いていた。

苗字は、母系社会であった古代中国に疑いなくすでに存在した。中国の最古の苗字の漢字には、女扁をもつものがあるからである。この時代には、中国語では苗字は「姓」（XING）と呼ばれた。中国の古代の書にあるように、周王朝（紀元前1122年～255年）以前の時代には、「姓」は各人が由来してきた女性の氏族創始者を表わすものとして、きわめて重要であった。「姓」は結婚の規準として用いられた。同じ姓の男女は結婚できなかったのである。夏王朝の時代から、「氏」（SHI）という、氏族の支流のシンボルをその支流の男性の祖先で表したものが使われはじめた。戦国時代（紀元前475年～221年）には、「氏」と「姓」があまり差がないものとして用いられた。母系社会から父系社会へ転換した後、子供は父親のみからその「氏」または「姓」を受け継ぐようになった。最終的に、「氏」と「姓」はどちらも同様に重要なものとなり、どちらも家族名として父系で伝わることとなった。現代中国の苗字の大部分は「氏」に由来し、「姓」に由来するものはわずかである。

歴史的には、現代使われているよりもずっと多くの「氏」と「姓」が存在した。新しい「氏」と「姓」が誕生したが、それよりも古いものが絶えることのほうが多かった。奴隷所有者、高名な武将、貴族といった高い社会階級に属する人間の苗字はしばしば官僚などが名乗るようになり、他の苗字が消えて行くのに対してより増えていった。

文献から収集された中国の苗字の総種類数（漢族以外の少数民族のものを含める）は11000を越えるが、日本の苗字（7万以上 {実際には12万以上}）やヨーロッパの苗字 {3～4万} と比べるとずっと少ない。漢族が用いている苗字は3000～4000種類である。中国人がヨーロッパ人や日本人に比べるとずっと少ない数の苗字しかない、つまり同姓グループがはるかに大きい理由は、ヨーロッパや日本に比べて少なくとも3000年以上前に苗字が

生まれたからである。それ以来、人口が今よりずっと少なかったあいだに、多くの苗字が消滅してしまった。{苗字の種類の多少は苗字の突然変異率にも大きく依存する。文化によってこの苗字発生率が異なることは十分考えられるので、苗字の歴史だけにたよる上記の議論には疑問である。中国の苗字の古さを誇っているような感じを受ける。}

最初期の苗字は氏族のトーテムや地名に由来していたが、後には国の名前、出来事、官位、職業、死後の称号などの個人の特徴などから取られるようになった。同一の苗字であっても、多数の起源を持つものが知られている。一般に、同姓グループの大きさが大きいほど起源の種類も多い。

中国における苗字の突然変異(ある苗字が新しい苗字へ、あるいはすでに存在する別の苗字に変化する場合のどちらか)の原因には、以下のものがある: 誤記、複数文字の苗字を1文字苗字に単純化する、皇帝や孔子の個人名を忌避する、敵から逃れる願望、別の苗字を持つ家族が男子を養子とする、皇帝が高貴な苗字を与える。とはいえ、漢族において苗字の頻度を変化させる最大の要因は、中国の歴史の様々な時代にいくつかの少数民族が漢族の苗字と文化を受容したことである。漢族が高度に文明化されており、高い文化を持っていたために、少数民族は自分たちが漢族であると宣言していった。漢族の一員であることは、有利であり誇りであった。したがって、漢族が世界中で最大の民族にふくれ上がったのは、出生率が高かっただけではなく、周辺の多数の民族の部分あるいは全体を飲み込んでいったからである。満州族、壮族、土家族、苗族、リー族などの一部が漢族に合流したほか、もっと起源の古い民族には完全に吸収されていったものもあった。

土家族の人口が1964年の524,755人から1982年の2,836,814人に増加したのは、このような漢化の証拠である。土家族は湖南省の少数民族であるが、同一時期に中国全体では45%しか増加しなかったのに、441%の増加が見られたのである。この増加のほとんどは、漢族であるがもともと土家族であった人間が、実際に自分は土家族であると申し出たためである。中華人民共和国政府が少数民族に有利になる政策をとったために、中国解放{1949年}以降、他の少数民族でも同様の事態が生じている。

女真族(後に満州族と呼ばれた)は、漢族に吸収された民族のもうひとつの事例である。紀元後926年に遼王朝が女真族の建国した渤海国を滅ぼした。遼の皇帝は女真族の貴族数千氏族を今日の遼寧省南部に強制移住させた。これらの氏族はゆっくりとその地域の漢族と融合していった。紀元後1234年には、再び女真族の建国した金王朝をモンゴル族が滅ぼした。元王朝(モンゴル)の指導者は中国北部から現在の湖北省、湖南省、山西省、山東省に移動したすべての女真族を漢族として登録した。その後、女真族の人々は漢族に吸収されていった。全体として、漢族に流入した女真族は数十万以上にのぼるだろう。{中華民国以前の清朝がまさに女真族によって建国されたことに言及していないのは、漢族中心の中華思想に抵触するからであろうか?}

3. 苗字分布から得られた結果

今回我々が調査した国勢調査の標本537,429人の中には、1054個の苗字が見いだされた。図1{転載せず}にその頻度分布を示した。このデータは漢族だけのものであり、少数民族は入っていないので、特に中国西部の省では一部の人間だけが用いられたことになる。

全体で291の「県」(XIAN) - 米国の郡(county) にほぼ相当する -、28の「省」(SHENG) - 米国の州(state) にほぼ相当する - のデータが集められた。{「省」には、広西自治区や天津特別市のようなものも含まれている。}用いられた「県」と「省」の名前は、表1{転載せず}に示した。

a) 分布の適合

苗字分布をFisherの対数分布に適合させたところ、期待値と観察値のあいだにおおきなずれが見いだされた。その理由として3点が考えられる。ひとつは最近における移住の増加である。この場合、まれな苗字が増加するだけなので、それらのクラスを除いて再分析した。しかし、期待値と観察値のあいだのずれはあまり縮まらなかった。第2の理由としては、父子・兄弟といった血縁者で同一の苗字があらわれることである。このため、標本が任意抽出ではなくなってしまう。第3の要因は、最近になって2集団が混じりあったために生じた苗字の変異性である。たとえば過去1世紀のあいだに黒龍江省には移民が多数入植したが、この省は期待値と観察値のずれが最も大きいところのひとつである。黒龍江省の現人口の約1/3は過去1世紀に山東省や湖北省から来た移民の子孫であるが、この移住の波は2, 3百年前から起こっている。最近では北京や上海からの移民が来ている。また中国解放後、政府は多数の屯田兵を黒龍江省に入植させている。

期待値と観察値のずれがもっと大きい省には、シンジャン、湖南、湖北がある。1949年には、シンジャンの人口は約433万人だったが、このうち30万人が漢族であった。解放後、毎年10万人以上の入植者がシンジャンに入ってくるようになった。1986年には、シンジャンの人口は1384万人となった。この期間の平均の自然人口増加率は2.4%なので、1949年から1986年のあいだにおよそ500万人の移民がシンジャンに入ったことになる。移民のほとんどは漢族である。

湖南省の場合、4000年以上にわたってこのあたりはしばしば戦場となってきた。さらに、黄河の氾濫と飢饉もこの数千年間ひんぱんに起こった。このため、湖南省は移入・移出とも頻度が高い。特に、数百年前に山西省から大規模な移入があったので、湖南省のなかには祖先が山西省から来たという人が多い。湖北省では、現代の満州族の多くが元朝の時代に漢化され、モンゴル人の一部も漢族の苗字を採用した。

もっとも、変異性が苗字の分布に与える影響はそれほど大きくはないようである。ある「省」のすべての「県」のデータを合一すればデータ内の変異性が高まると予想されるが、実際には期待値からのずれがそれほど大きくはならないからである。

b) 苗字の多種類度 (a) の推定

苗字の多種類度 (a) には、調べられた「県」人口と都会化の程度が関係する。{人口が多ければ通常都会化していると呼ぶはずであるが?}

==中略==

aは、周辺の広い地域から常に移入があると、大きい値となる。移入なしで集団が人口増加を生じていれば、aはほぼ一定にとどまる。しかし、都市は一般に移入によって人口を増大させてきており、都市自身の内的自然増加率によって人口を維持できるようになったのは、最近のことである。{医療水準が低かった過去には、伝染病などにより常に都市人口が減少する可能性があったのだろう。} したがって、国勢調査で得られる都市人口は、過去にどれだけ移入があったかを示すものと考えることができる。商工業や文化の中心地は移入者を引きつけてきた。また、都市の開発が長くかかればかかるほど苗字の多様性が増大しただろう。この問題は将来研究する必要がある。

移出は、aの値を減少させるだろう。中国南東部の経済的に貧しかった海岸地域（福建・広東）は、台湾や他のアジア地域へ強い移民の波があったことがよく知られている。シン

ガポールとフィリピンの中国移民のほとんどは広東と福建から来ている。これらの移住はランダムには生じていない。海外に移住して住み良い場所をみつけた先駆者に続いて、その全家族だけでなくかなり遠縁の一族まで移住するのである（連鎖移住と呼ばれる）。広西自治区は、過去数百年にわたって大量の移入者がおらず、かわりに移出によって多数の人口を失った。また、文化的・経済的發展も周辺地域より遅れて生じた。

==中略==

4. 苗字データから推定された樹状図と主成分地図

a) 樹状図

28省間の苗字距離 (D_{name}) を以下の式で定義して用いた。

$$D_{name} = -\log [S_{piqi}] / \sqrt{[S_{pi2}S_{qi2}]}$$

ここで、 p_i と q_i は、それぞれ集団 p と q の i 番目の苗字の頻度であり、すべての苗字種類についての合計をとる。この苗字距離は、相対同姓率の自然対数をとってその符号を換えたものである。{上記の式は、遺伝子頻度から集団間の遺伝距離を推定する際に用いるもののひとつである、「根井の遺伝距離」とよく似ている。Saitou et al. (1994) を参照。} な

お、台湾省のデータは、他のデータと異なり、以前に台湾で行なわれた研究結果を用いた。

これら苗字距離から、平均リンケージ法 {距離の小さい集団から順にクラスターを作成してゆく方法} を用いて作成した樹状図が図7である。この樹状図によれば、最初の分岐はクラスターJとFのあいだで生じているが、これはほぼ中国の北と南に対応する。境界線は、黄河と揚子江のほぼ中間である。図8ではこの境界に省の境界を用いたが、これは伝統的な中国の南北境界にきわめてよく一致している。

サブクラスターも、地理的によくまとまっている。クラスターAとBは、それぞれ中国中央北部と西北部に対応する。やや周辺部に位置し、中央北部の南方にある山西は、クラスターA・B双方と近縁である。{クラスターA・クラスターB・山西はクラスターCを構成する。} クラスターDは中央北部クラスターの少し南方に位置するが、地理的にはやや散らばっている。シンジャンは、山西・甘肅・陝西と近い関係にある。このことは、これらの地域がシンジャンにおける漢族の形成に関与したという可能性を示唆するが、他の北部地域もシンジャンと近縁なのである。現在シンジャンに住む漢族の大部分は過去40年間の移住者である。

クラスターFは、クラスターCとDのほかにそれらの外側に天津・寧夏・青海を含む。この状況は、次のような相反する仮説のうちのどちらかで説明できるだろう。ひとつは、外側の地域が歴史の初期に他の地域から分岐し、独自の苗字分布を進化させてきたという仮説である。これは、人口密度が低ければ起こり得た。もうひとつの仮説は、クラスターCとDの地域と外側の3地域がかなり混血したというものである。

地域間の混血には、ふたつのモデルがある。ひとつは、ある地域がいくつかの特定の地域から移入者を引き寄せるタイプである。この場合、これら移入者を提供した地域と近縁性が生じるので、それらのクラスターの外側に位置することにある。このモデルは天津に

あてはまるだろう。天津は中国北部で最も重要な港であり、元朝（1279年～1368年）以来貿易が活発である。また、上海について第二の工業地区でもあり、人口では中国第3位である。天津は北京よりも移住者を引きつけてきたようである。北京が中国の首都となったのは比較的最近で、元朝のことである。その後、首都は北京と南京で交替した。北京は貿易や工業の中心地というよりは、文化的政治的中心地であった。このために、他の貿易・工業地帯よりも移住者にとって魅力が少なかったのだろう。長期間首都であり続けていればすべての地域から移住者が来ると期待されるのに、北京が遠方の地域と苗字分布の点であまり近縁でないのはこのせいかもしれない。

もうひとつの移住モデルは、互いに近縁度の低い枝の複数の地域が新しい混血集団を形成するというものである。この新集団は、最大の割合の人間を提供した枝とクラスターを形成しするが、混血が生じたよりもずっと古い時代に対応する部分で枝分かれをするだろう。青海がこのモデルにあてはまるかもしれない。

中国南部のクラスターJは、サブクラスターG・H・Iからなる。Gは上海を中心とする揚子江下流地方にあたり、Hは揚子江流域のほぼ全域をカバーする。Iは南部海岸地域と主要島嶼を含む。IはGよりもHと近縁となっているが、IとHをつなぐ枝の長さが短いので、明瞭な差はない。他の方法によると、GとHがより近縁であるという結果が出ている。台湾への移住は、主に福建およびその周辺地域から起こった。また客家の移住も少数ながらあった。

{斎藤成也によるデータの再分析結果}

筆者（斎藤成也）は、原論文の図7（平均リンケージ法で作成した樹状図）が、近隣結合法を用いるとどうなるかに興味を持ち、原論文の表3のデータをもとにして、29地域間の苗字距離（ D_{name} ）を計算し、近隣結合法（斎藤 [1993]を参照）を用いて樹状図を作成してみた。それが図Aである。[枝の長さは苗字距離に比例している。破線は、作図上集団が混み合うために導入されたものであり、枝の長さはゼロである。]ところが、系統樹作成法が異なるといっても、クラスタリングがあまりにも原論文の図と違いすぎるので、原論文で用いられた平均リンケージ法を同一の苗字距離行列に適用して樹状図を作成した（図B参照）。図Bは、原論文の図7とはかなりの点で異なっている。これは、用いられたコンピュータプログラムの違いというよりも、用いられた苗字距離行列が異なるためのものである。筆者は、原論文の記述にしたがって、原論文の表3のデータ（相対同姓率を与えたとされている）から、自然対数による変換を行なって苗字距離行列を算出したのだが、そこでは黒龍江省と遼寧省の距離が最小となり、実際に図Bでもこれら2地域が最初にクラスターをなしている。ところが、原論文の図7では黒龍江省と山東省が最初にクラスターをなしている。

ところで、原論文の37頁には、台湾と他のいくつかの地域との「同姓率」が与えられているが、これらの数値は原論文の表3で「相対同姓率」とされている数値と同一であった。したがって、2種類の可能性が考えられる。（1）原論文の表3の数値は、「相対同姓率」ではなく、「同姓率」である。原論文の図7は「相対同姓率」から得られた正しい苗字距離に基づいている。（2）原論文37頁の数値は「同姓率」ではなく「相対同姓率」であり、したがって原論文の表3は正しい数値が与えられているが、図7を作成する際に、なんらかの理由で誤った距離が算出されてしまった。可能性（1）の場合には、筆者が作成した図A・図Bはあまり意味がない。なぜなら、これらは「同姓率」に基づいて作成されたものだからである。ところが、可能性（2）の場合には、むしろ図A・図Bのほうが正しいことになる。現時点ではどちらの可能性が正しいのか不明であるが、本論文の他の部分の数値を検討した結果、筆者はどうか後者の可能性のほうに分がりそうだと考えている。この場合、原論文の結論は大きな変更を余儀なくされる。つまり、南北の2大対

立の境界はもっと南にあり，ここでは「南」は，広東・広西・福建・台湾だけとなる（図B参照）。一方，近隣結合法による樹状図（図A参照）では，全体的には図Bの結果と似ているが，南北の2大対立というよりは，中国北部がよくまとまっているのに対して，南部の変異が大きいことが注目される。これは，少数民族の遺伝子データでも示唆された結果である（Saitou et al. 1994 を参照）。

いずれにせよ，近いうちに，筆頭著者である杜若甫教授および数値分析の中心であったと考えられるCavalli-Sforza教授に手紙を書いて，どちらの可能性が正しいのかを確かめる予定である。{齋藤成也によるデータの再分析結果：終わり}

b) 主成分座標 一略

5. 頻度による苗字の分類 一略

6. 距離による隔離 一略

7. 遺伝子データとの比較

苗字と遺伝子では伝達様式が異なるし，苗字は最近生じたものではあるけれども，我々は，苗字頻度が遺伝子頻度に似かよった情報を与えると考えている。この仮説は，現在用いることのできる遺伝子頻度データを用いて，ある程度まで検討することができる。28地域についてわずか3遺伝子座（A B O，M N，R h）のデータしか存在しないが，これらから遺伝距離を計算し，主成分座標分析を行なった。その結果，第1・第2座標軸で見ると，南東沿岸部（広東・広西・福建・台湾）が他の地域と大きく離れている。{これは，苗字頻度分布を再分析した齋藤の結果と同一である。} また，苗字・遺伝子ともに，北方集団のクラスターはそれほどまとまってはいない。地域によっては，雲南や青海のように苗字・遺伝子ともに他地域と離れているものもある。苗字と遺伝子のクラスターは全体的に似ているが，遺伝子では苗字ほど南北の違いが明確ではない。{齋藤の再分析によれば，苗字でも南北の違いは明確ではない。このため，むしろ苗字と遺伝子のデータの類似性は高まる。}

==中略==

G m遺伝子座の遺伝子頻度が中国のなかで，特に南北で大きい変異のあることは，以前から知られていた。たとえば，対立遺伝子fabは，北で20%，南で70%である。この差は，各集団がこうむった感染症・寄生虫症に違いがあったからかもしれない。しかしこの南北差は他の遺伝子でも見いだされたので，生態遺伝学的な要因だけでなく，もともと南北の集団は遺伝的に異なっていたようである。われわれが行なった世界全体の集団の比較によると，東アジア地域は南北2グループに別れるが，南部中国人は東南アジアクラスターと，北部中国人は東北アジアクラスターと結ばれる。

1986年に行なわれたアジアオセアニアH L Aシンポジウムの発表には，中国におけるH L A遺伝子のデータが2つ含まれている。どちらの報告でも，中国の南北差が明瞭に示されている。{Saitou et al. (1992) は，このうちのひとつのデータを中心に再分析を行ない，中国の南北のちがいが漢族でも少数民族と同じように生じていること，また日本人と朝鮮人は中国北方クラスターに属することを示した。}

{Saitou et al. (1994) は、海南島の少数民族を中心とした遺伝子頻度データをアジアの他
の他
集団と比較した。その結果、図A・Bと同様に、東アジア北部の集団間の変異が小さいの
に対して、南部集団および東南アジアの集団の変異が大きいことが示された。また、東ア
ジアの南北の集団の違い（集団の分岐年代）は過去1万年に遡ると推定された。}

{遺伝子データの解説については、例えば斎藤（1994）を参照されたい。}

8. 先史時代と歴史時代

中国における南北差は、古人骨のデータでも示されているし、後期旧石器時代にはすで
に南北の差が生じていたことが、考古学資料から示されている。北部では細石器中心であ
るのに対して、南部では石斧などで特徴づけられる東南アジアのホアビン文化複合の影響
を受けていた。中国における新石器時代の発達は複雑であり、最近の研究速度も著しいが、
以下の3点に要約される。（1）北部では、紀元前6千年紀にヤンシャオ文化が始まった。
主要農作物は雑穀であり、いくつかの家畜が用いられた。絹もすでに生産されていた。

（2）

南部では、ユエ文化（台湾を含む）がやはり紀元前6千年紀に始まった。主要農作物は米
であり、この文化の影響は雲南にもおよんだ。（3）揚子江下流域では、チンリンカン文
化がやや遅く始まった。米は灌漑された水田で生産され、犬・豚・水牛が家畜化された。
これら3地域の分化は、遺伝子および苗字の地理的構造とよく似ている。

中国の南北差は、東南アジアに起源を持ち中国南部でさきに始まったと考えられる青銅
器時代にも引き継がれた。北部では、紀元前約2700年にロンシャン（龍山）文化が始ま
ったが地域的な変異がきわめて大きかった。夏文明は湖南地域で紀元前約2100年に始まった、
それらの変種のひとつであるが、500年後には黄河下流域全体に広がった。

中国における現在の集団構造において重要な特徴のひとつに、少数民族の存在がある。
特に南部に多く、雲南省だけで24族いる。彼らは通常族内婚を行なうが、漢族とのあいだ
に遺伝子交流があったことは明らかである。

新石器時代に生じた現象は、遺伝学的には重要である。なぜなら、農耕が導入されたた
めに、人口が急増したからである。このような人口増によって、遺伝子頻度の機会的変動
は凍結され、過去に生じた遺伝的な差をを保った。これに対して、ヨーロッパの新石器時
代は、中近東から栽培植物と家畜を持ち込んだ農民が広がっていった。この移民はゆっく
りと起こり、ヨーロッパの最端までゆくのに約4000年経過した。農民がこうして移動す
るあいだに、彼らよりも人口密度のずっと低い先住の採集狩猟民と部分的に混血した。これ
によって、今日なお観察できる遺伝的勾配が生じたのである。中国では、北部・南部・東
部でほぼ同時に新石器時代の文化中心が発達したので、それ以前の遺伝的差異をある程度
維持することができた。

== 中略 ==

9. 議論と結論

苗字は中国の南北で大きく異なるが、この対照は遺伝子の場合と一致しており、南北の
境界はどちらのデータでも、揚子江と黄河のあいだに来る。{これは誤りではないだろう

か。最初の2分岐は、広東・広西・福建・台湾の4省と他の地域であろう。著者らは、「南船北馬」に代表される伝統的な南北の文化境界に一致させたいからこう言っているのではないだろうか?}

苗字のような、文化的に伝達され、人類史のなかで遅く登場した「遺伝子」が、どうしてそれよりはるかに古い起源を持つ遺伝子と似た振る舞いをするのだろうか。ひとつには、中国の苗字が他のどの文化よりも古い起源を持っていることがあげられよう。歴史的な証拠から、少なくとも3000年前、おそらく4000年前まで起源をたどれる苗字が多い。その当時、中国は技術的には発展した青銅器時代だったが、人口の大部分は農村にいた。貴族だけが都市に住み、社会構造は氏族社会のものと大差なかった。新石器時代あるいはそれ以前に起源を持つ氏族名が、現存する多数の苗字の起源となったかもしれない。苗字が古いことを示唆するもうひとつの論拠は、苗字によっては母系の痕跡を残しているものがあることである。伝統的な中国の家族制度が父系構造を獲得したのは、ずっと後になってからである。

==中略==

頻度の高い苗字は、中国の歴史の特定時期に頻度を増大させた可能性が高い。中国で最も一般的な苗字である「李」は全漢族中7.9%の頻度を占めるが、これは唐王朝(紀元後618年-907年)の皇帝の苗字である。「王」も一般的な苗字のひとつであるが(頻度7.4%)、これは王様を意味する。「趙」の頻度は2.3%であるが、これは宋王朝(紀元後960年-1279年)の皇帝の苗字である。明王朝(紀元後1368年-1644年)の皇帝の苗字である「朱」の頻度は1.3%である。このように皇帝の苗字の頻度が高いことは、苗字頻度から移住を推定する正当性を減じる。しかし、同一の移住パターンが継続すれば、頻度の高い苗字を用いて推定しても、それほど偏りを生じることはないだろう。

==中略==

まとめると、中国の苗字は集団の遺伝的構造をより深く知るための貴重な情報を提供する。男性の移住のみを測定することからくる限界や、支配者の苗字と関連する苗字の頻度変化からくる限界を詳細に予測することはまだできないが、これらの限界はそれほど深刻ではないようである。将来研究が進めばこれらの点について明らかにされるだろう。また、豊かでしかも長期にわたって保存されてきた中国の古文書を用いることにより、将来過去の人口調査からも貴重な結果がもたらされるだろう。

引用文献

Mountain J.L., Wang W. S.-Y., Du R., Yuan Y., Cavalli-Sforza L.L. (1992) Congruence of genetic and linguistic evolution in China. *Journal of Chinese Linguistics*, Vol. 20, pp. 315-330.

斎藤 成也 (1983) 苗字資料による国内の移住パターン推定の試み. *人類学雑誌*, 91巻3号, 309-322頁.

Saitou N., Tokunaga K., and Omoto K. (1992) Genetic affinities of human populations. In Roberts D.L., Fujiki N., and Torizuka K. (eds.), *Society for the Study of Human Biology Symposium Series 33: Isolation and Migration.*, Cambridge University Press, Cambridge, pp. 118-129.

齋藤 成也 (1993) 近隣結合法. 五條堀孝他共編, 新生化学実験講座第16巻『分子進化実験法』, 400-410頁. 東京化学同人, 東京.

齋藤 成也 (1994) モンゴロイド諸集団の遺伝的近縁関係. 赤澤威編, 『先史モンゴロイドを探る』 (学術新書), 11-25頁. 日本学術振興会, 東京.

Saitou N., Omoto K., Du C., and Du R. (1994) Population genetic study in Hainan Island, China. II. Genetic affinity analyses. *Anthropological Science*, Vol. 102, No. 2, pp. 129-147.

Piazza A., Rendine S., Zei G., Moroni A., and Cavalli-Sforza L.L. (1987) Migration rates of human populations from surname distributions. *Nature*, Vol. 329, pp. 714-716.