

## 論文内容の要旨

論文題目 : Adaptive evolution revealed by the correlation between the two phases of molecular evolution in human: polymorphism and fixation

(進化の2つの段階、多型と固定の相関から明らかになったヒトの適応進化)

氏名 五條堀 淳

ヒトゲノム上のタンパク質をコードする遺伝子のうち、どれだけが適応進化をしたかという問題は分子人類学や人類集団遺伝学において長らく議論されてきた。適応進化を検証する方法として種内変異 (polymorphism) と種間変異 (divergence) を対比する方法があり、進化の段階を多型 (polymorphism) と固定 (fixation) に分けて考える事で、自然選択の検出力を高める事が期待される。ヒトの全ゲノム塩基配列に加えて、チンパンジーのゲノム概要塩基配列が決定され、さらにヒトのゲノム規模の一塩基多型 (Single Nucleotide Polymorphism, SNP) データベースも利用可能となった事で、塩基配列レベルやアミノ酸配列レベルでのヒトの適応進化の検証がゲノム規模で可能になった。本研究では2つの進化の段階、多型と固定を対比し、ヒトの適応進化をゲノム規模で検証した。

集団中に新たな変異が導入された時、その変異が有害であったならば負の自然選択によって集団より除去され、有害でなかった場合、その変異は消失または多型となり、遺伝的浮動によりその変異の頻度は時間とともに増減する。集団中に除去されずに残った変異体は、遺伝的浮動により偶然に集団中に固定することもあり、生存に有利な変異であった場合は、正の自然選択により集団中に固定する可能性や、その速度が大きくなる。このように、変異体が集団中に固定する過程を考え、その前半の段階を進化の多型の段階、後半を固定の段階とすると、多型の段階では負の自然選択が遺伝的浮動とともに重要な力であるのに対し、固定の段階では遺伝的浮動に加えて正の自然選択の働きも重要となる。このように多型と固定の段階において、主に働く自然選択の種類が違う事から、この2つの段階を対比する事で、変異に対して自然選択がどのように作用するかを検証できる。

この2つの進化の段階の相関を見るのに、これまでの単純な同義変異、非同義変異という分類では不十分であるが、非同義変異を1つの塩基の変異で起こりうる75種類のelementary amino acid change に分類す

る事でそれが可能になる。本研究では、75種類のelementary amino acid changeのそれぞれに対して、多型のなりやすさの指標であるPI (Polymorphism Index) と、固定のしやすさの指標であるFI (Fixation Index, ここではWright [1921] のFstとは違う指標である) を定義し、ヒトの多型と固定の相関を検証した。

多型データとしてヒトのSNPのデータベース、Perlegen Sciences と国際HapMap Project、Seattle SNPsとNIEHS SNPsを用いた。固定のデータはヒト-チンパンジー間での種間変異を使用した。またヒトのSNPについて、チンパンジーをアウトグループとして祖先型と派生型の対立遺伝子を決めた。解析には常染色体上にあるタンパク質をコードする領域のみを用いた。解析の対象となった遺伝子数はPerlegenで5008個、HapMapで5535個、SeattleとNIEHSで274個となった。

ヒトのSNPを派生型対立遺伝子の頻度別にみると、低頻度のSNPでは非同義変異/同義変異比 (A/S比) が高頻度のものと比べて高かった。これは有害突然変異がヒトの集団の中で低頻度に蓄積していることを示している。そこでヒトのSNPを、頻度20%を境界として、20%以下の頻度のSNPをレアなSNP、20%より高い頻度のSNPをコモンなSNPとした。コモンなSNPはレアのものに比べて有害突然変異が少ない事が期待され、それゆえに中立的な変異がより多い事が期待される。

表1はヒトのSNPデータベースの概要を示している。ここでMcDonald-Kreitmanテストを用いて多型と固定の段階のA/S比を比較し、ヒトゲノムにおける自然選択を検証した。ここでは多型にコモンなSNPだけ (より中立的な変異だけ) を使う事で検出力を高めた。ただし、Seattle+NIEHSのデータベースでは、高頻度のSNPのA/S比が大きく、hitchhikingの効果の可能性があるため、頻度20%-80%のSNPsをコモンとした。FI = (固定のA/S比)/(コモン多型のA/S比)と定義し、FIの観測値が中立の期待値を上回るかどうかで正の自然選択の有無を検証した。その結果、PerlegenとHapMapの両方のデータベースにおいては強い正の自然選択は検出されなかった。特にHapMapでは、Perlegenと比べても非同義変異が同義変異より多く、収集バイアス (ascertainment bias) の影響が強く見られた。収集バイアスが含まれないSeattle+NIEHSのデータベースでは正の自然選択を検出できたが、これはこのデータベースに多くの免疫関連の遺伝子が含まれているためと思われ、必ずしもヒトゲノム一般の傾向を反映していないものと考えられる。これ以後の解析はPerlegenとHapMapについてのみ行った。

次にアミノ酸変異を1塩基変異のみでおこる75種類のelementary amino acid changeに分類し、それぞれのアミノ酸変異についてA/S比を多型と固定の段階で比較した。また、各アミノ酸変異について期待される変異の数 (new mutation の数) を解析に用いた遺伝子の塩基組成から計算した。ここで、PI = (レア多型のA/S比)/(new mutationのA/S比)とし、アミノ酸変異の多型のなりやすさの指標とした。FIはアミノ酸変

表1 SNPデータベースの概要 A = 非同義変異、S = 同義変異

Perlegen	A	S	A/S
New mutation	9742	3019	3.227
Polymorphism (rare, ≤20%)	2929	3019	0.970
Polymorphism (common, >20%)	2306	3082	0.748
Divergence (fixed)	26505	32582	0.813
FI (Fixation Index)	Observed = 1.087, Expected = 1.128		
HapMap	A	S	A/S
New mutation	5934	1849	3.209
Polymorphism (rare, ≤20%)	3772	1849	2.041
Polymorphism (common, >20%)	2937	2073	1.417
Divergence (fixed)	30158	35502	0.849
FI (Fixation Index)	Observed = 0.600, Expected = 1.023		
Seattle + NIEHS	A	S	A/S
New mutation	1589	690	2.303
Polymorphism (rare, ≤20%)	696	690	1.009
Polymorphism (common, 20% ~ 80%)	99	168	0.589
Polymorphism (high, >80%)	168	126	1.333
Divergence (fixed)	1277.5	1710.5	0.747
FI (Fixation Index)	Observed = 1.205, Expected = 1.079		

異ごとに再計算し、アミノ酸変異の固定のしやすさとした。表2は、PerlegenにおけるFIについて上位と下位5種のアミノ酸変異とPerlegenとHapMapの概要である。どちらのデータベースについても、75種類のアミノ酸変異のFIについて $\chi^2$ 検定を行うと有意差があることから、FIは75種類のアミノ酸変異について一様ではない事が分かった。

Perlegenについて、FIの上位34種のアミノ酸変異を除いて $\chi^2$ 検定を行うと有意差がない事から、下位41種のアミノ酸変異はFIについて一様であることがわかる。表2から上位34種のアミノ酸変異のFIは1.713であり、下位41種のアミノ酸変異のFIの0.948より大きいので、上位34種のアミノ酸変異は正の自然選択を受けている可能性が示された。同様にHapMapでは、FIについて上位30種のアミノ酸変異が正の自然選択を受けている可能性が示された。

表2からPerlegenについて、同義変異の固定/多型比を中立における期待値とすると、FIの上位34種のアミノ酸変異には  $(7589 - [419 \times 32582 / 3082]) = 3160$ 個だけ期待値以上に種間変異が蓄積していることになり、その割合は種間変異におけるアミノ酸置換の11.9%  $(= 3160 / [7589 + 18916])$ となる。この期待値以上のアミノ酸置換の蓄積は正の自然選択によるものと解釈できる。HapMapについては非同義変異にかたよった収集バイアスが強いので、Perlegenと同じ解析は不可能であるが、FIについて下位45種のアミノ酸変異の固定/多型比を中立における期待値とし、Perlegenと同様の解析をすると、3149個  $(7168 - [437 \times 22990 / 2500])$ のアミノ酸置換が期待値以上に蓄積しており、その全体のアミノ酸置換に対する割合は10.4%  $(= 3149 / [7168 + 2290])$ となる（同じ解析をPerlegenのデータを用いてすると結果は12.8%となる）。2種類のデータベースからの正の自然選択を受けているアミノ酸置換の割合の推定は、ほぼ同じ結果になり、ヒトゲノム全体で正の自然選択を受けているアミノ酸置換の割合は10.4% ~ 12.8%であると推定される。

どのような種類のアミノ酸変異が固定しやすいか、すなわち高いFIを持つのか？アミノ酸の変異や置換を議論するとき、しばしばアミノ酸の物理化学距離が使われる。物理化学距離が大きいアミノ酸の変異は、タンパ

表2 75種類のelementary amino acid changes (一部)

a) Perlegen			Rare		Common		
AA1	AA2	New mutation	Poly. $\leq 20\%$	PI	Poly. $> 20\%$	Divergence (Fixed)	FI
Leu	Trp	39	3	0.077	0	54.5	$> 5.0$
Lys	Ile	54	2	0.037	1	43	4.067
Lys	Met	80	8	0.100	3	85	2.680
Asn	His	95	10	0.105	7	196	2.649
Ser	Ile	104	12	0.116	4	108.5	2.566
...	...	...	...	...	...	...	...
Arg	His	58	125	2.141	90	754	0.792
Val	Ile	118	233	1.974	174	1445	0.786
Arg	Thr	69	14	0.204	15	108	0.681
Arg	Trp	69	49	0.710	41	282.5	0.652
Thr	Met	40	76	1.900	82	537	0.619
Top 34 classes		4457	622	0.140	419	7589	1.713
Bottom 41 classes		5286	2307	0.436	1887	18916	0.948
Synonymous		3019	3019		3082	32582	

b) HapMap		Rare		Common			
	New mutation	Poly. $\leq 20\%$	PI	Poly. $> 20\%$	Divergence (Fixed)	FI	
Top 30 classes		2260	734	0.325	437	7168	0.958
Bottom 45 classes		3673	3038	0.827	2500	22990	0.537
Synonymous		1849	1849		2073	35502	

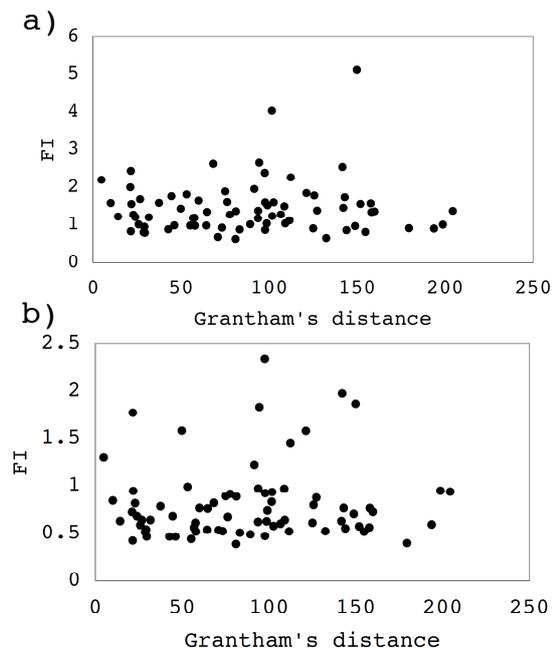


図1 アミノ酸の物理化学距離 (Granthamの距離) とFIの相関。  
a) Perlegen, b) HapMap.

ク質の構造や機能に与える影響が大きいと考えられ、このような種類のアミノ酸変異は自然選択により淘汰されるのか、または選択されるのかという問題は長く議論されてきた。図1はFIとアミノ酸の物理化学距離との相関図である。Perlegen とHapMap のどちらのデータベースにおいても両者に相関がないことから、アミノ酸の物理化学距離はFIを説明しない事が言え、アミノ酸の固定のしやすさはアミノ酸の物理化学距離によらないことが示された。

FIを説明するようなアミノ酸のパラメータは何か？図2はFIとPIの相関を示した図である。どちらのデータベースでもFIとPIはL字型の分布となり有意に負の相関を示した。特に、高いFIのアミノ酸変異のPIはすべて低かった（図2の点線枠内）。このことから、固定しやすいアミノ酸変異は多型になりにくいアミノ酸変異であること言え、アミノ酸変異の多型のなりやすさと固定のしやすさは、逆の傾向を示す事が分かった。また、多型になりにくいアミノ酸の変異は負の自然選択を受けやすい変異だと言え、したがって以上の事から、正の自然選択と負の自然選択は同じ種類のアミノ酸変異（PIが低い=多型になりにくい変異）に強く作用する事が示された。

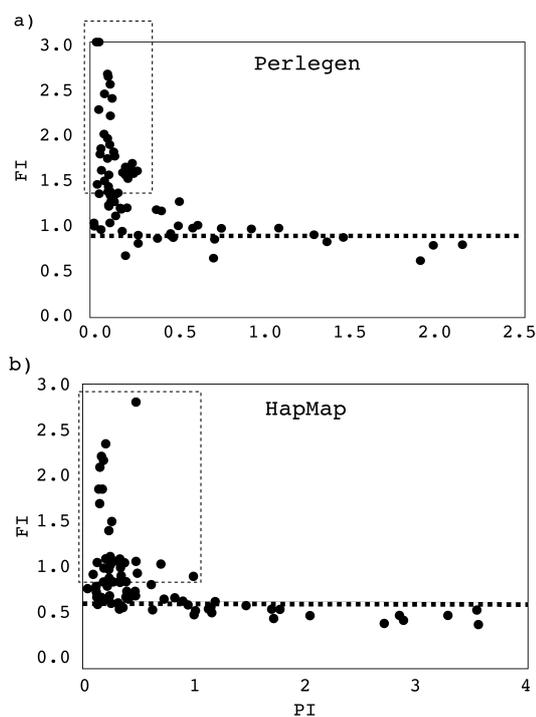


図2 FIとPIの相関図。a) Perlegen, b) HapMap  
点線はFIについて下位41(45)種のアミノ酸変異の平均FIを示す。