

Genetic Variations of the marine fishes and the freshwater fishes

メタデータ	言語: jpn 出版者: 公開日: 2016-08-23 キーワード (Ja): キーワード (En): 作成者: 八島, 亮子 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/257

海水魚類種と淡水魚類種における遺伝的多様性のパターンの比較

Genetic Variations of the marine fishes and the freshwater fishes

八島 亮子¹

Akiko Sato Yashima

1 要旨

遺伝子とは、親から子へと受け継がれる生命の設計情報である。遺伝情報は私たちの顔の特徴が人それぞれ異なるように、私たちが持つ遺伝子にも個々の違いや特徴があり、この遺伝子の同一生物種内における違いを「遺伝的多様性」と呼ぶ。古くはチャールズ・ダーウィンのべた通り、遺伝的多様性は生物の進化を考える上で非常に大事な概念でもある。ダーウィンは家畜を例に多様性を持つ遺伝子を持つ個体の中から、環境に最も適応的なもの、あるいは人為的に好まれて選択されたものが、子孫をより多く残し、種内全体にその遺伝子が広まることが進化の原動力であると予測した。このことを逆に捉えれば、遺伝的多様性が豊かな生物種は、環境の変化への適応能力に長けているということもできる。そのため、絶滅危惧種に指定されている生物種の遺伝的多様性の度合いは保全計画を立案する点からも重要であり、様々な生物種で遺伝的多様性の定量が行われてきた。遺伝的多様性、特に自然選択のかかかっていない中立的な遺伝領域の多様性は、主に集団サイズの大きさと突然変異率、集団の分断化などの要因によって変化することが理論的に予測されている。しかしながら、実際のデータを見たときにそれらの影響が実際にどの程度見られるのか、包括的な研究はあまり進んでいない現状がある。そこで本稿では、明らかに生息環境の異なる海水魚類種と淡水魚類種の遺伝的多様性のデータを比較することにより、生息域の違いに起因すると考えられる遺伝的多様性に対する影響が実際に観察されるのかを検討した。

2 導入

まず、簡単に遺伝子の基本的な概念について説明する。遺伝情報は RNA(ribonucleic acid: RNA) ウイルスなど一部例外を除いて、DNA(deoxyribonucleic acid: デオキシリボ核酸) がその物質的な基本となる。DNA は一見、はしごをねじったような構造をしており、はしごの支柱部分はリン酸と糖でできている。一方、はしごのステップにあたる部分は塩基と呼ばれる部分であり、アデニン (Adenine, 以下「A」と表記する), グアニン (Guanine, 以下「G」), シトシン (Cytosine, 以下「C」), チミン (Thymine, 以下「T」) の4つの種類で構成されて

¹ 武蔵野大学数理工学センター員 / 武蔵野大学工学部数理工学科助教

いる。遺伝情報として重要なのは、この4つの塩基の並び方である。4つの塩基のうち、AはTと、GはCと水素結合によってペア（塩基対）を成し、ひとつのステップにあたる部分を構成している。DNA配列はいったん mRNA (messenger ribonucleic acid: 伝令 RNA) に転写されたのち、ある一定の規則に従ってアミノ酸配列に翻訳される。アミノ酸配列は側鎖の化学的性質に従って折りたたまれ、複合体を形成し、タンパク質となる。タンパク質は生物の組織そのものを構成するだけでなく、酵素として細胞内の様々な化学反応を触媒することで生命活動そのものを司るなど、生物の体の構成と機能の両面から生命体を形成している。

DNAは複製され、細胞ひとつひとつに行き渡り、親から子へと伝わる。ヒトの遺伝情報の一セット（ゲノムと呼ばれる）のDNAの塩基配列は、約30億塩基対 (base pairs) であるが、父親からもらう一セットと、母親からもらう一セットで合計60億塩基対の遺伝情報をもつ。そして子どもに伝わる時に、2セットのゲノム情報はシャッフルされ（組換えとよばれる）、そのうち2セットのうち1セットが次世代に伝わる。30億塩基対あるゲノム中全ての領域が、タンパク質になる情報を含んでいるのではなく、翻訳されるのは一部の領域のみである。このような領域をコード領域 (coding region) とよばれるが、この領域はゲノム中のわずか5%以下にすぎない。あとのタンパク質に翻訳されない領域は、非コード領域とよばれる。非コード領域の一部はコード領域に書かれたタンパク質の発現量などを調節する役割を果たすが、大部分は意味をなさないと考えられている*1。

表1 主な生物種の塩基多様度の度合い (%)。

生物種名 (学名)	塩基多様度 (%)	参考文献
ヒト (<i>Homo sapiens</i>)	0.09–0.13	[1]
マウス (<i>Mus musculus</i>)	0.13–0.37	[2], [3]
イトヨ*2 (<i>Gasterosteus aculeatus</i>)	0.20–0.27	[4]
ショウジョウバエ (<i>Drosophila simulans</i>)	2.04–3.52	[5]
シロイヌナズナ (<i>Arabidopsis thaliana</i>)	0.55–0.65	[6]
マラリア原虫 (<i>Plasmodium falciparum</i>)	0.13–0.15	[7]

ヒトに最も近い生物種のひとつであるチンパンジーとの塩基の違いは平均して1.26%ほどである[8]。また、ヒトを種内でみたとき、ランダムに2セットのゲノムを比べた場合、平均して約0.1%程度の違いがあるといわれている[1]。この1000塩基に1つ現れる一塩基の違いを一塩基多型 (Single Nucleotide Polymorphism) と呼び、ランダムに2つのゲノムをサンプリングした時にどの程度の塩基が異なるかという指標は塩基多様度 (nucleotide diversity) と呼ばれる。塩基の違いは突然変異と呼ばれるDNAの複製ミスによって生じる。ある1塩基における突然変異率は、ヒトで1世代あたり 10^{-8} 程度である。代表的な生物種における塩基多様

*1 非コード領域の役割については、まだ分かっていないことが多い。

*2 北半球の亜寒帯に広く分布する魚の一種で、淡水にも海水にも生息する。生息環境によって形態が異なることが知られており、異なる環境への適応進化を調べるためによく用いられる生物種である。

度を (表 2) に示す.

表中の塩基多様度に幅があるのは, サンプルングした個体による誤差と, ゲノム中のどの部位を比較に用いたのかという違いのためである. 先に説明した具体的な機能のあるコーディング領域よりも, 非コーディング領域の方が多様性が高いことが知られている. コード領域に突然変異が起きた場合, タンパク質の機能に変化が現れ, 極端な場合は死に至る^{*3}. そのため, コード領域に突然変異が起きても, 集団中から排除され, 残らないことが多い. 一方, 非コード領域は突然変異が起きてもほぼ影響がないため, そのまま突然変異が蓄積し, 多様性が高くなる傾向がある. 非コード領域のような自然選択のかからない領域は主として集団サイズの大きさと突然変異率の大きさによって決定する. その理論的背景についての説明は付録資料にまとめた.

さて, 次に実際サンプルングしてきた遺伝子のタイプを観察し, その数を観測するものとする. 付属資料に説明する通り, ゲノム中の同一部位にある遺伝子の一方を「対立遺伝子」あるいは「アレル」と呼ぶが, 今, サンプルングされた集団から k 個のタイプのアレルが観測されたとする. サンプル個体数を n とすると, i 番目のアレル ($1 \leq i \leq k$) の頻度の観測値 p_i は, i 番目のアレルの数/ $2n$ である. 同じタイプのアレルを 2 つ持つ個体である確率は, おおよそ次のように予測される.

$$\sum_{i=1}^k p_i^2$$

これはサンプルングされた個体のアレルの頻度が, 母集団となるその集団全体のアレル頻度と一致すると仮定した場合, 片一方のアレルがアレル i である確率が p_i , かつもう片方のアレルが同じくアレル i を受け取る確率は p_i であるためである. この逆に, 異なるタイプのアレルを持つ個体である確率は,

$$H_e = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

である. これは集団からランダムにサンプルした 2 つの遺伝子が異なるタイプのアレルであった確率と同義であるため, (6) 式のヘテロ接合度の理論値と一致する. これは十分に大きな任意交配の集団において成り立つ予測である. この値はアレル頻度から予測される, 「ヘテロ接合度の期待値 (expected heterozygosity: H_e)」と呼ばれ, 遺伝的多様性の観測値として重要な指標のひとつである.

なお, ヘテロ接合度は配列全体が違うタイプか否かという確率であるため, どれだけの塩基数が違うかということも考慮した指標である塩基多様度 (表 2 に示す) よりも情報量が落ちる. 歴史的には, DNA の塩基配列の詳細な観察が実験的に困難だった時代に, タンパク質の

^{*3} 突然変異によっては全く影響のないものや, 影響が弱いものもある. しかし, おおむね新たに起こった突然変異を持つ個体の生存能力が低いことは, 様々な生物を用いた実験によって確かめられている. これらの実験結果については, Halligan と Keightley (2009) [9] の批評に詳しい. 逆に, 従来より生存に有利な機能をもつ突然変異が起きる場合もある.

電荷や性質の違い（例えば血液型）といった，DNA の塩基配列に違いがあった結果生じた，より下流の変化を元に遺伝的多様性を評価するためによく用いられていた指標である．今日でも後述するマイクロサテライトなどの遺伝マーカーによく用いられる．

さて，任意交配の集団について，遺伝的多様性の主要な指標であるヘテロ接合度の算出方法は (1) 式に述べた．しかしながら，任意交配が崩れるとき，例えば近親の個体や同じアレルを持つ個体と交配する傾向や，集団内に地理的な移動障壁が生じるなどした場合に，見かけ上のヘテロ接合度が低くなる現象が知られている．2倍体の生物において，異なるタイプのアレルを2つ持つ個体頻度の観察値は，(1) 式と一致することが期待されることはすでに述べた．しかし，近親交配など同じアレルを持つ個体と交配する傾向が強まると，次の「近交係数」とよばれる指標が高まる．

$$F_{IS} = 1 - \frac{H_o}{H_e} \quad (2)$$

ここで， H_o とは「ヘテロ接合度の観察値 (observed heterozygosity: H_o)」と呼ばれ，単純に $H_o = (\text{異なるタイプのアレルを2つ持つ個体数}) / (\text{サンプルされた個体数})$ で求められる．任意交配の集団では $H_o = H_e$ となるため， $F_{IS} = 0$ となる．

また，集団内に地理的な移動障壁が生じるなどして，ランダムな任意交配が阻害された場合に次の「集団分化係数」とよばれる指標が高まる．

$$F_{ST} = \frac{H_T - \overline{H_S}}{H_T} \quad (3)$$

H_T は「全集団 (total population) のヘテロ接合度」， $\overline{H_S}$ は「分集団 (subpopulation) のヘテロ接合度の平均値」である．ここでいう分集団とは任意に決めるもので，例えばある川の上流と下流で分集団を定義し，魚の F_{ST} を観察した場合，その2つの集団を混ぜてヘテロ接合度を求めたものが H_T ，上流，下流それぞれでヘテロ接合度を求めたものが， H_S である．もし，上流と下流で十分な行き来がある場合， $H_T = \overline{H_S}$ ， $F_{ST} = 0$ となる．上流と下流で個体の行き来が遮断されている場合，より小さな集団の中でだけ任意交配が進むため，分集団の見目の N が小さくなり， $\overline{H_S} < H_T$ ， $0 << F_{ST}$ となる．一般に， F_{ST} 値は集団間の個体の往来が多いほど低くなるため，2集団間の F_{ST} 値を比較した場合（「これを Pairwise F_{ST} と呼ぶ」），単純な直線距離が小さくなれば， $F_{ST} = 0$ に近づくと考えられる．実際に，人類集団の場合，日本人集団と中国人集団間では $F_{ST} = 0.007$ ，同じく日本人集団と中国人集団より地理的に離れたアフリカのヨルバ人集団間では $F_{ST} = 0.190$ であることが知られている [11]．

さて，以上に述べたような遺伝的多様性の指標であるヘテロ接合度 H_e や集団分化係数 F_{ST} は，1980年代より様々な生物で調べられてきた．付録に詳述するマイクロサテライトによる定量方法のみに限って見ると，その論文数は2010年まで約2万編ほどになる．しかしながら，これだけ多くの論文が出版されてきたにも関わらず，複数の研究結果をまとめて全体を俯瞰するようなメタ解析的な研究はほとんど例がない．例えば海水域に生息する魚に比べ，淡水域に生息する魚では明らかに集団間の個体の往来も生息域の大きさも限られている．こうした生活史の特徴は集団サイズの大きさに起因する遺伝的多様性の大きさや集団の分化度合いに違いが

生じると容易に予測できる。具体的には、付録資料に述べた通り集団サイズ N が大きく、分集団の存在しない生物種は遺伝的多様性が高いことが推測されるため、この特徴に該当する海水魚は淡水魚よりも H_e 値が高く、 F_{ST} 値が低いことが予測される。実際の先行研究としては、13種の淡水魚種と12種の海水魚種および7種の遡上性の魚種をマイクロサテライトで比較した研究では海水魚種の H_e が高い傾向が見られた [12]。また、タンパク質の電荷の違いを利用したアロザイム・マーカーを用いた方法ではあるが、49種の淡水魚種と57種の海水魚を比較した結果も同様に、海水魚種の H_e が高く、また F_{ST} 値が低い傾向が見られた [13]。しかしながら、前者の研究例はサンプル数は十分といえず、後者の研究例はタンパク質の性質の違いを定量しているため、コード領域にかかる自然選択の影響が考えられる。また、 F_{ST} 値の大きさは前述した通り、集団間の地理的な距離が影響することが知られているが、距離の影響を考慮に入れていないという問題がある。そこで本研究では、マイクロサテライトマーカーを用いた研究を比較し、距離の影響も考慮に入れてこの問題を考証した。

3 方法

1980年代から行われているマイクロサテライトによって遺伝的多様性を評価した論文を ISI Web of Science(<http://pcs.webofknowledge.com/>) を用いて検索した。検索キーワードは”microsatellite”, ”SSR”, ”STR” とし、検索は2011年の1月に行った。次に、膨大な論文数から魚類種の属名を含む論文をピックアップした。属名のリストは NCBI Taxonomy Database(<http://www.ncbi.nlm.nih.gov/taxonomy>) から取得した。論文はひとつずつ直接チェックし、各種ごとにもっともサンプル数と定量に用いたマーカー数の多い論文について、 H_e 値 F_{ST} 値および各サンプル地点の緯度経度を記録した。集団間の地理的な距離は緯度経度から Haversine Formula [15] によって求めた。次に、各魚類種の生息域の情報を、FishBase (<http://www.fishbase.org/>) より取得した。統計的な解析および図の描写は全て R 言語を用いた [14]。

4 結果と考察

各集団における H_e 値、 F_{ST} 値および直線距離を各種ごとに平均化し、各種の平均値について生息域ごとの項目に分類し、平均値と標準偏差を求めた (表 2)。なお、「淡水と海水」の項目はサケやウナギなどのように産卵のために海と河川を行き来する魚類種や、河口など汽水域に生息する魚類種を広く含む。

なお、断りをいれておくと、 H_e と F_{ST} 値は0から1までの値しか取らないため、母集団の等分散を仮定できない。よって、この表の値から値の大小に関する統計的な有意差を論じるのは適当ではないが、海水魚は淡水魚よりも H_e 値が高く、 F_{ST} 値が低いという予測と合致する傾向がうかがえる。集団間の距離については、淡水魚種より、海水魚種の方が大きい傾向があった。つまり、海水魚は集団間の距離が大きいにも関わらず、 F_{ST} 値が小さい傾向が見られ

表 2 魚類の生息域ごとの遺伝的多様性, 集団分化係数, 集団間の直線距離の要約統計量.

生息域	種数	ヘテロ接合度 (H_e)	集団分化係数 (F_{ST})	集団間の直線距離 (km)
淡水のみ	71	0.63 ± 0.18	0.12 ± 0.13	253.8 ± 367.2
海水のみ	67	0.77 ± 0.11	0.03 ± 0.06	1353.5 ± 1918.6
淡水と海水	34	0.67 ± 0.13	0.09 ± 0.11	801.27 ± 1333.4

(a) 海水棲魚類種の典型的なパターン (b) 淡水棲魚類種の典型的なパターン

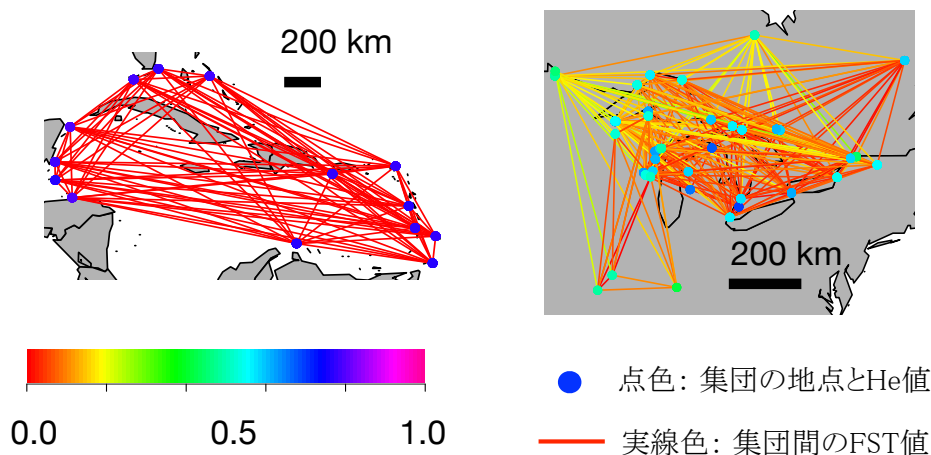


図 1 海水魚種と淡水魚種における典型的な遺伝的多様性の様相. 各サンプル地点を丸で示し, サンプル地点間を直線で示した. 丸の色はサンプル地点における H_e の高さを示し, 線の色はサンプル地点間の F_{ST} 値の高さを示す. (a) 海水魚の例として, *Haemulon flavolineatum* (イサキの一種) [16]. F_{ST} 値が低く, ほぼ 0 に近かった. (b) 淡水魚の例として, *Acipenser fulvescens* (チョウザメの一種) [17]. F_{ST} が高い傾向 (線が赤色でない) にあり, 同一水系かどうかで F_{ST} 値がばらつく傾向があることが見られた.

る. また, 図 1 に海水魚と淡水魚の典型的な遺伝的多様性のパターンを示した. ここでも同じく予測を指示する結果を示している. すなわち, 海水魚種は実線色が赤色で示されるとおり, ほぼ 0 に近い値をとることがわかる. H_e 値もおおよそ 0.8 前後と高い値をとっている. 一方, 淡水魚種では実線色が赤い部分もあるが, 黄色で示されている部分もあり集団の分化が海水魚種の例と比較して大きいことがわかる. H_e 値はおおよそ 0.4-0.8 程度でばらつき, 平均すると海水魚種より低い傾向が見られる. また, 淡水魚種では同一水系とみられる地域からサンプリングした集団同士では F_{ST} が低い傾向が見られた. 淡水魚種は海水魚種に比べて種内での H_e 値が大きくばらつく点も非常に興味深い. 現在, 著者は特に H_e 値の大きさについて, 統計的に比較できる方法を開発し (投稿中), 詳細な解析を行っている.

5 将来的な展望

冒頭に述べた通り、遺伝的多様性は進化の原材料である。進化は新たな突然変異が出現し、種内に固定することを繰り返して起きる。分集団化や小さな集団サイズは遺伝的浮動の効果が大きく、新たな突然変異が迅速に固定しやすい。このことは分集団化がおきやすい種は新たな生物種への分化をしやすい傾向があるといえるのかもしれない。今回は淡水魚でこのような傾向が見られたが、両生類など特殊な環境に依存する傾向の強い生物種でも同様な傾向が見られる可能性がある。こうした傾向のある生物種の傾向を見いだすことは生物進化を考える上で重要な知見になると筆者は考える。

参考文献

- [1] Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, 1000 Genomes Project, Sella G, Przeworski M. Classic selective sweeps were rare in recent human evolution. *Science* 331(2011), 920–924.
- [2] Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, Pritchard JK, Gilad Y. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research* 22(2012), 602–610.
- [3] Baines JF, Harr B.. Reduced X-linked diversity in derived populations of house mice. *Molecular Ecology* 175(2007), 1911–1921.
- [4] Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6(2010), e1000862.
- [5] Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, Pachter L, Myers E, Langley CH, Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology* 6(2007), e310.
- [6] Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J., The pattern of polymorphism in *Arabidopsis thaliana*., *Plos Biology* 3(2005), e196.
- [7] Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, Milner DA Jr, Daily JP, Sarr O, Ndiaye D, Ndir O, Mboup S, Duraisingh MT, Lukens A, Derr A, Stange-Thomann N, Waggoner S, Onofrio R, Ziaugra L, Mauceli E, Gnerre S, Jaffe

- DB, Zainoun J, Wiegand RC, Birren BW, Hartl DL, Galagan JE, Lander ES, Wirth DF. A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genetics* 39(2006), 113–119.
- [8] The Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome, *Nature* 437(2005), 69–87.
- [9] Halligan DL and Keightley PD, Spontaneous mutation accumulation studies in evolutionary genetics, *Annual Review of Ecology, Evolution, and Systematics* 40(2009), 151–172.
- [10] Spielman D, Brook BW, Frankham R, Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America* 101(2004)15261–15264.
- [11] Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, Toncheva D, ...& Jakkula E, Genetic Structure of Europeans: A View from the North?East. *PLoS ONE* 4(2009)e5472
- [12] DeWoody JA, Avise JC, Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of fish biology* 56(2000)461–473
- [13] Ward RD, Woodwark M, Skibinski DOF, A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of fish biology* 44(1994)213–232
- [14] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0
- [15] Sinnott, RW, Virtues of the Haversine. *Sky and telescope* 68(1984)158
- [16] Purcell JFH, Cowen RK, Hughes CR, Williams DA, Weak genetic structure indicates strong dispersal limits: a tale of two coral reef fish *PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES* 273 (2006)1483–1490
- [17] Welsh A, Hill T, Quinlan H, Robinson C, May B, Genetic assessment of lake sturgeon population structure in the Laurentian Great Lakes *NORTH AMERICAN JOURNAL OF FISHERIES MANAGEMENT* 28(2008)572–591
- [18] Fisher RA, The genetical theory of natural selection: a complete variorum edition, *Oxford University Press* (1930).
- [19] Wright S, Evolution in Mendelian populations, *Genetics* 16(1931) 97.
- [20] Kimura M, Evolutionary rate at the molecular level. *Nature* 217(1968)624–626.
- [21] Johnson WE, Onorato DP, Roelke ME, Land ED, Cunningham M, Belden RC, McBride R, Jansen D, Lotz M, Shindle D, Howard J, Wildt DE, Penfold LM, Hostetler JA, Oli MK, O'Brien SJ, Genetic restoration of the Florida panther. *Science* 329(2010)1641–1645
- [22] Waits, Lisette P and Paetkau, David, Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collec-

tion, *Journal of Wildlife Management* 69(2005), 1419–1433.

付録 1 : 非コード領域における遺伝的多様性の理論的背景について

ゲノム中の大半を占める非コード領域の多様性はどのように決まるのであろうか。突然変異が残ったままであれば、多様性は無限に大きくなりそうなものだが、実際にはある程度のところで止まってしまう。まず手始めに、ゲノム中のある一箇所に着目した時、ある 2 つのタイプの遺伝子があったとする。(例えば、片方を「ATAGC」、もう片方を「ATCGC」のように、3 番目の塩基が変わっていることに着目。このようにゲノム中の同一部位にある遺伝子の一方を「対立遺伝子」あるいは「アレル」と呼ぶ。) 突然変異が起きない場合に、集団において 2 つのタイプのアレルが集団中に占める割合がどの程度になっていくのかを考えたい。ここでの集団のモデルは、Wright-Fisher モデル [18][19] で考える。Wright-Fisher モデルは有限個体数の生物集団の数学的なモデルであり、その特性を端的に表すシンプルで優れたモデルとして、最もよく用いられている。まず、個体数 N の互いに任意交配を行う生物集団を考える。この集団には、微生物のような半数体では N 個、ヒトのように有性生殖を行う生物種では $2N$ 個のゲノムないし遺伝子が含まれる。以下説明では、後者の有性生殖を行う生物種のケースを考える。この集団は常に一定の子孫を残し、次世代の子供は $2N$ 個の配偶子のうち 1 個の配偶子をランダムに受け継ぐ。いずれの個体も配偶子を作る能力は等しく、どの親の配偶子を受け継ぐかは等確率であるとする。世代の重なりは考えない。以後、毎世代これを繰り返す。

集団からランダムにサンプリングした 2 つのアレルが、違うタイプのアレルである確率はヘテロ接合度と呼ばれ、 H で表される。さて、現在の 0 世代目におけるヘテロ接合度を H 、1 世代後のヘテロ接合度を H' で表すと、1 世代後のヘテロ接合度は次のように求められる。

$$\begin{aligned} H' &= \frac{1}{2N} \times 0 + \left(1 - \frac{1}{2N}\right)H \\ &= \left(1 - \frac{1}{2N}\right)H \end{aligned}$$

これは、ある親の遺伝子が次の世代に伝わったとして、他の子供の親も遺伝子も同じ親に由来する確率が $\frac{1}{2N}$ 、違う親から由来する確率は $1 - \frac{1}{2N}$ のためである。前者の場合、親から子に伝わる際に突然変異が起きない限り、両者が違うタイプのアレルを持つ確率は 0 である。後者の場合、両者が違うタイプのアレルを持つ確率は H である。この式から、最初の世代の H を H_0 とすると、 t 世代後の H_t は以下の式で表される。

$$H_t = H_0 \left(1 - \frac{1}{2N}\right)^t \quad (4)$$

この式からわかるのは、突然変異が起きない限り、ヘテロ接合度は毎世代 $1 - \frac{1}{2N}$ の割合で減衰するため、集団数 N が小さいと、この減衰率は大きいことになる。どれかひとつのタイプのアレルが集団中の全ての個体に行き渡り、他のタイプのアレルが消失してしまうことを「固

定」と呼ぶが、いったんあるアレルが固定しまうと、新たなアレルの出現がない限り、多様性は失われたままとなってしまう。このように、世代交代の際に偶然の効果によって集団中のアレル頻度が変化することを「遺伝的浮動」と呼ぶが、遺伝的浮動は長期的には多様性を失わせる効果を持つ。

しかし、実際の所、突然変異によって新しいタイプのアレルが供給されるため、ヘテロ接合度 H は、ある程度の大きさを保つ。無限アレルモデル^{*4}。という突然変異が毎回違ったアレルを生み出すようなモデル下における、遺伝的多様性に対する突然変異率と集団サイズの効果を考える。まず、1世代あたりの突然変異率を μ とおき、 $G = 1 - H$ (つまり、ランダムにサンプリングした2つのアレルが同じタイプである確率) とする。突然変異は G_t の1世代後の変化は以下のように表される。

$$G'_t = \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)G_t \right] (1 - \mu)^2 \quad (5)$$

$(1 - \mu)^2$ は、2つの兄弟に分けられたアレルの間で、突然変異が起きない確率である。 $(1 - \mu)^2 = 1 - 2\mu + \mu^2$ であるが、 μ は非常に小さな値であり、 N は大きな値であるため、 $\mu^2 \approx 0$ かつ、 $\frac{N}{\mu} \approx 0$ とする。この近似を使うと、1世代の G の変化量 ΔG は以下のように求められる。

$$\begin{aligned} \Delta G &= G_t - G'_t \\ &= G_t - \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)G_t \right] (1 - \mu)^2 \\ &\approx G_t - \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)G_t - 2\mu G_t \right] \\ &= 2\mu G_t - (1 - G_t) \frac{1}{2N} \end{aligned}$$

ここで、 $G = 1 - H$ より、

$$\Delta(1 - H) = 2\mu(1 - H_t) - \frac{H_t}{2N}$$

定常状態では、世代によるアレルの頻度の変化がないため、 $1 - \Delta H = 0$ を満たす。 $1 - \Delta H = 0$ を満たす H_t の期待値 \bar{H} は、

$$\bar{H} = \frac{4N\mu}{1 + 4N\mu} \quad (6)$$

である。

^{*4} これは、例えるならば、1000塩基からなるDNA配列を観察した場合、1回目の突然変異は100番目の塩基、2回目の突然変異は230番目の塩基に起きるといのように毎回違ったタイプの配列を生み出すような突然変異の状態を考える。また、複数回の突然変異の結果、元の遺伝子のタイプ（遺伝子のタイプのことをアレルと呼ぶ）と区別がつかない状態に戻ってしまうモデルも存在する。例としては、ある1塩基に着目した場合、そのアレルはA,G,C,Tの4つのいずれかであり、2回の突然変異が起きた場合、 $A \rightarrow T \rightarrow A$ のように、元のタイプに戻ってしまうことがある。しかし、同じ塩基に突然変異が複数回入ることは非常に稀なことであり、通常そのような現象が起きる時には、すでにその生物種は別種に分化するほど長い歴史を経ていると考えられるため、同種内での変異量について考える本モデルでは、通常そのような確率は考えない。

以上、本節をまとめると、集団数 N の逆数によって規定される遺伝的浮動の効果と、突然変異率 μ によって決まる新たな遺伝子の出現の、二つの効果によって遺伝的多様性は左右されることがわかる。このように、自然選択によらず確率的な変動によって、特別な機能を持たない遺伝子が集団中に固定されることがしばしば起こりえることを、木村資生は一貫した理論として初めて定式化した^{*5}[20]。

(6) 式に示した N を観測することは現実的には困難だが、ヘテロ接合度 H と突然変異率 μ は実験によって、ある程度求めることができる^{*6}。そのため、何サンプルかの個体の DNA を調査することで、残りのおおよその個体数^{*7}を把握し、絶滅危惧種の保護に役立てようという取り組みもなされている。(4) 式にしめした通り、集団数 N が小さいと、偶然の遺伝子変動の効果がより大きくなる。このことは生存に不利なアレルが集団中に広がりやすくなることを示している。また、片親から生存に不利な遺伝子の正常な機能を損ねたアレルを受け継いでも、もう片親から受け継いだアレルが正常な働きを持つ場合は、その機能により救済される場合が多い。しかし、多様性を損なうと、同じアレルを引き継ぐ確率が高くなり、生存に不利な遺伝子の影響が出てきやすくなる（この現象を近交弱勢と呼ぶ）。これらのことから、絶滅の危機に瀕している生物の遺伝的多様性を調査することは保全の上で重要な基礎調査となる。実際に絶滅の危機に瀕している生物は、その近縁種に比べて遺伝的多様性が低いことが報告されている [10] また、次節で述べるように、地域集団間の分集団化や近親交配が進んでいる集団かをある程度推測することができる。開発による分集団化と近親交配により遺伝的多様性を消失した集団の遺伝子を調べ、計画的に交配させることにより、近交弱勢を起こして遺伝病を頻発していた絶滅危惧種の生存率を上昇させた試みも存在している [21]。

付録 2：マイクロサテライトを用いた野生生物調査について

ヒトやマウスなどのよく研究された生物は、DNA 配列がゲノム上のどの位置にあり、コード領域の場合はその機能が詳細に調べられている。そのため、多様性を調べるにも DNA チップなどのツールが開発されている。しかし、多くの生物ではこうした情報やツールが未整備であり、DNA 配列が未知な状態でも容易にゲノム全体にまたがる多様性解析を可能にするようなテクニックが用いられる。マイクロサテライトの調査はそのようなテクニックのひとつである。マイクロサテライト (microsatellite) とはゲノム中に複数存在する、ATATATAT... のような単純な繰り返し配列領域のことである。一世代にある一つの塩基に突然変異が起きる確率

^{*5} この考えは”最も環境に適応した者だけが生き残る”というダーウィンの進化説に凝り固まっていた当時の進化学者を驚かせた。

^{*6} H は、体毛などの組織片から DNA やタンパク質を抽出し、PCR 反応やシーケンシング反応という化学反応や電気泳動によって観察し、遺伝子のタイプを定量する。 μ は親子間の遺伝子のタイプの違いを観察することで求められる。突然変異はごく稀な事象であるため、 H に比べて μ はより多くの配列を観察しなければならず、実験はやや困難である。そのため、 μ については近縁種で求められた値を代入するケースが多い。

^{*7} ここでは説明を省いたが、個体数 N は現在の集団数をそのまま反映しない。有効集団サイズとよばれる、集団数が過去から現在にわたって一定と仮定した場合の理想集団サイズである。各世代における実際の集団数の調和平均で求められる。

は 10^{-8} ほどであるが、マイクロサテライトのような繰り返し配列はすべりモデルと呼ばれる特殊な変異のメカニズムにより、 10^{-4} ほどの突然変異率があると言われている。この高い変異率により、マイクロサテライトの違いは個体間で違いが生じやすいため、個体識別や親子関係の解析、分集団間の違いなどを調べやすい^{*8}。また、DNA シークエンサー (配列を一文字ずつ読むことができる) を使用しなくても検出することができるため、安価であるという利点があり、野生生物の遺伝的多様性調査によく用いられる [22]。

(原稿提出: 2016 年 1 月 9 日; 修正稿提出: 2016 年 1 月 27 日)

^{*8} 余談ではあるが、マイクロサテライトは法医学や人類遺伝の分野では STR(Short Tandem Repeat) とも呼ばれ、15 個ほどの遺伝子座が犯罪捜査に用いられている。