

遺伝と生物情報 を考え直す

1. 遺伝現象の科学的理解はなぜ可能になったのか

近代科学の考え方の原理は、物質科学でも生物科学でも、基本的に同じ類型で理解できる。

量子化原理：遺伝現象を担う粒子状の単位（すなわち遺伝子）を想定し、その組み合わせで遺伝現象を理解しようとする。すなわち個々の遺伝子 G_i は特定の遺伝子型 a_{i1}, a_{i2}, \dots （対立形質 allele）をとり、それ以外の中間的な状態はとらない。ある生物個体（細胞）の遺伝子状態は、すべての遺伝子の状態の組み合わせ $\{G_i: i=1,2,\dots,n\}$ である。ただし n は遺伝子の総数。

可視的な現象を不可視的な状態から説明する手法：遺伝子 G_i について、不可視的な遺伝子型 a_{ij} をとる ($G_i = a_{ij}$) ものと想定し、その解釈 F_{G_i} によって i に関わる表現型 P_i が特定の表現型 p_{ik} をとることを説明しようとする。つまり

$$P_i = F_{G_i}(G_i)$$

である。この場合、解釈のルール F_{G_i} は遺伝子ごとに異なるとする。問題はこの解釈関数 F_{G_i} の中身とその理解の仕方である。また、表現型はもともと現象論的なものなので、それぞれの表現型に関わる遺伝子 G_i が予めわかっているわけではない。しかし、逆に、遺伝子 G_i に対応して表れうる表現型の種類 p_{ik} は限られている。

倍数体にあわせた修正：生物によっては、複数の遺伝子セットを持つ場合（2倍体など）があるので、遺伝子セットをゲノムと定義する。個々の遺伝子の状態は、各ゲノム N_m ($m=1,2, \dots, c$: c はゲノムの数つまり倍数性を表す) に含まれる同じ遺伝子の状態の組み合わせ $\sum_m G_{im}$ (和は m についてとる) として理解することにする。すなわち、

$$P_i = F'_{G_i}(\sum_m G_{im})$$

ここで、 F'_{G_i} は倍数体に対応した解釈関数である。

2. 遺伝子型と表現型が1対1対応する場合の理解

一個の遺伝子 G_i の状態がそのまま表現型 P_i に表れる場合は、1倍体細胞や1倍体生物で見られる。通常、遺伝子は特定のタンパク質（酵素）の情報を持っているので、もっとも単純な場合には、 P_i はある酵素 E_i の活性を持つか持たないかの2種類となる。つまり、

$$P_i = G_i$$

である。ただし、 G_i の値としては、機能あり(functional) a_{if} , 機能なし(non-functional) a_{in} の二つをとる。

稀には、奇妙なことが起きる。それは、機能がなくなった酵素が細胞に害を与える場合で、タンパク質が沈殿して固まるなどの場合が考えられる。この場合には、 G_i の値として、機能あり、機能なしで無害、機能なしで有害の3通りあることになる。

細胞がもつ代謝活性は、多数の酵素の連続した反応で構成されていることが多いので、表現型として表れるのは、ある代謝経路 MP_α が機能しているかどうかという場合が多い。また、タンパク質も複合体を形成して機能している場合が多く、その場合、どの成分がな

くなくても複合体 CX_β 全体の機能がなくなる。従って、こうした場合には、複数の遺伝子の機能不全が同一の表現型（代謝経路や複合酵素の機能不全）を示すことになる（表現型の冗長性）。

3. 2倍体の場合の理解

これが通常メンデルの法則である。メンデルの法則は、現象を説明できる仮説の1つの集合体であって、それぞれの事項を分解して証明することは当時はできていなかった。

2倍体の場合、同じ遺伝子が2セットずつあるので、 G_{i1} と G_{i2} が異なる遺伝子型 a_{i1} , a_{i2} をとる場合にどのような表現型を生ずるのかについて、ルールを決める必要がある。上記の1倍体の場合の考察を当てはめると、遺伝子型 a_{i1} は正常な酵素、遺伝子型 a_{i2} は機能不全酵素を生み出すとする。正常な酵素の活性を $[Ei]_0$ とし、実測される酵素の活性を $[Ei]$ とすると、 G_{i1} と G_{i2} が異なる遺伝子型 a_{i1} , a_{i2} をとる場合は、

$$[Ei] = 1/2 [Ei]_0$$

となることが期待される。実際に、非メンデル遺伝として知られる花の色の濃さは、このように中間型となることがある。しかし、多くの細胞機能では、酵素は過剰量存在することが多く、その量が半分になっても大きな影響を受けないことが多い。そのため、この例のように一方の遺伝子が機能不全型酵素を作るとしても、表現型には表れない。これを**優性**と呼ぶ。優性は **dominance**（支配的な性質）ということで、優れた性質という意味ではない。この場合、遺伝子型 a_{i1} は優性、遺伝子型 a_{i2} は劣性(**recessive**)ということになる。なお、2で示した有害な遺伝子型は、通常優性を示す。こうした場合、**dominant negative** 変異と呼ぶ。

高校の生物の遺伝関係の問題は、ほとんどがこの倍数性と優性劣性の関係に関わるものであるが、これは遺伝現象の本質ではない。今回の教科書の第3章でも、これにまつわるヒトの遺伝学がたくさん説明されているが、もっとも本質的なのは、遺伝子と表現型の変換がルールに則って行われるということだけである。ルールそのものは、倍数性や問題の酵素の性質や細胞内での機能（代謝経路や複合体を形成しているか）によって異なってくる。

4. パターン化された遺伝法則

以上の前提を理解した上で、通常2倍体生物の遺伝学の大部分にあてはまる法則が、メンデルの3法則である。独立の法則は、上記の量子化原理の当然の帰結である。優性の法則は、上の説明の通り、特別な場合に成り立つ。分離の法則も量子化原理の帰結であるが、簡単に定式化しておく。

有性生殖する2倍体生物を考える。純系親 A が持つ遺伝子 G_i の遺伝子型は2個とも a_{iA} 、純系親 B が持つ遺伝子 G_i の遺伝子型は2個とも a_{iB} とする。それぞれの配偶子は1倍体であるが、それぞれの遺伝子型は a_{iA} と a_{iB} だけである。A, B の雑種1代(F1)では、 $\sum_m G_{im}$ の中身は、 $\{a_{iA}, a_{iB}\}$ となる。ここで、A が B に対して優性の場合、F1 の表現型は純系親 A の表現型と同一となる。優性でなければ、中間型となる。

ふつう F1 世代には、オスもメスも含まれるので、これらを交配することができる。または、植物ならば、自家受粉することもできる（もちろん、他家受粉でも話は変わらない）。F1 と F1 の子孫 F2 世代では、たくさんの子孫ができた場合、3通りの遺伝子型を持つものが表れる。つまり、 $\{a_{iA}, a_{iA}\}$, $\{a_{iA}, a_{iB}\}$, $\{a_{iB}, a_{iB}\}$ である。それぞれの出現頻度の期待値は、1:2:1 である。ただし、両配偶子の形成・生存・生殖能力等に差がなく、両配偶子がランダムに接合することを前提としている。優性がある場合には、表現型としては、A と B が 3:1 となる。また、優性がない場合、A, 中間型, B が 1:2:1 となる。ここで F1 では隠れていた B が F2 で再度表れる点が、量子化原理の証明となっている。

伴性遺伝は、2倍体を前提としたヒトの遺伝学の抜け穴のようなもので、X 染色体が男性では1倍体となっていることに起因している。普通、伴性遺伝は Y 染色体の遺伝子ではなく、X 染色体の遺伝子が問題となる。Y 染色体には男性決定遺伝子 *SRY* 以外、ほとんど遺伝子がないためである。

5. 量子化された遺伝子から遺伝物質の構造まで

遺伝物質が DNA であることがわかり、その構造から遺伝子の本体が解明されたが、上記のような形で定式化された量子化の単位としての遺伝子は、実際には、構造を持つことになった。物理学で、原子がさらに構造を持ち、構成粒子がさらにクォークからできているようなものである。遺伝子は実際には、有限個の情報単位のならんだ文字列（配列）で表現されることがわかった。この情報単位が DNA を構成している4種類の核酸塩基である。従って、遺伝子型 a_{ij} は、原理的には非常に大きな自由度を持つことになる。ただし、実際に遺伝子として機能するための制約が大きいので、たいていは、機能する遺伝子として1つまたは複数の a_{iw} が定義され(wild type 野生型)、この遺伝子の塩基を置換して得られるほとんどの遺伝子は機能しない遺伝子としてすべて同等のもの a_{im} と見なされる(mutant 変異型)。つまり a_{iw} と a_{im} しか存在しないことが多い。

なお、酵素の構造を決めている遺伝子（構造遺伝子）と、その遺伝子が機能するかどうかを決めている別の遺伝子（調節遺伝子）があることが多い。調節遺伝子の機能はさまざまな中間的なものがあり得るので、調節遺伝子の変異は、中間的な表現型を示す場合がある。

血液型物質の場合は特別な例で、一つの遺伝子の支配でできる酵素の性質が少し変わることで、A 型物質を作ったり B 型物質を作ったり、または機能不全の結果として原料物質が O 型物質のままになったりする。つまり遺伝子型として3種類存在することになる。

6. 染色体の位置づけ

原理的な理解には、遺伝子とゲノムだけでよいが、現実の生物は、ゲノムを物理的にいくつかのブロックに分けた形で保持していることが多い。真核生物はすべて複数（3個以上）の染色体を持っている。染色体の1セットがゲノムに対応する。原核生物のゲノムは、ふつう、1個の DNA 分子となっていることが多く、これも染色体と呼ぶ。細菌の中には、複数の染色体 DNA を持つものや、環状ではなく線状の DNA を持つものもある。

染色体上の遺伝子の並び方を推定するために、古典遺伝学では、連鎖分析が行われた。これは、2個の遺伝子が物理的に近接して存在している場合には、かけあわせにより F2 世代で両者が共分離することを検定するものである。この理屈を理解するのははじめはとても難しいかもしれない。有性生殖では、配偶子形成の過程で起きる減数分裂で染色体の交叉・乗り換えによる組み換えが起きること、組み換えの頻度は遺伝子間の距離と関連している（単調増加）こと、などの前提をおけば、2個の遺伝子の組について組み換え率を求めることができる。組み換え率の加法性を前提として、3個の遺伝子の3通りの組み合わせについて求めた組み換え率を利用することで、3個の遺伝子の位置関係を定義することができる。これが遺伝子地図である。

現在では、塩基配列から求めた遺伝子地図（物理地図）と比べたときに、遺伝学的に作った遺伝子地図には伸び縮みがあることがわかっており、食い違いは、組み換えのホットスポット（組み換えが起りやすい染色体の領域）の存在により説明されている。また今でも、DNA のマーカーと表現型から取得した変異体の原因遺伝子の位置の関係を調べるために、連鎖分析が用いられている。この方法で原因遺伝子を特定することをポジショナルクローニングと呼ぶ。

7. 文字列としての生物情報

こうして遺伝子が4種類の情報単位の文字列として記述できることになった。このため、ある生物がもつ遺伝情報の総体であるゲノムも長い文字列として記述できる。この長い文字列情報のうちのどこに各遺伝子の情報が書かれているかという説明（アノテーション）をつけたものが、ゲノム情報として生物科学の研究に利用されている。ゲノム情報はコンピュータで扱うことができる。生物情報の取得、維持管理、利用処理に関わる情報科学をバイオインフォマティクスと呼ぶ。