出芽酵母のゲノム解析とその意義†

磯 野 克 己

(神戸大学理学部)

生物は、多数の遺伝子の機能を有機的に統合し、恒常性を保ちつつ増殖を続けるとともに、自己の生存する環境の変化を知覚し、それらの変化に対して、短期的あるいは長期的に、さまざまな方法で自らを適応させて生存している。生物を“ゲノムの塩基配列に情報を蓄えた情報系”として捉えてみると、このような適応過程は個々の遺伝子レベルでの変化にとどまらず、多くの場合、いわば“情報系の再組織化”を伴っているといえる。近年、遺伝子や遺伝子の産物である蛋白質の構造や機能を詳細に解析することが可能になり、膨大な種類の生物の膨大な数の遺伝子が解析・同定され、それらの構造や発現の制御機構が明らかになってきている。真核生物で現在最も解析の進んでいる出芽酵母ゲノムには、6000 あるいはそれ以上の遺伝子があるものと推定されている（後述）が、これらの多数の遺伝子は、出芽酵母のゲノム中に巧みに配置され、高度に連係して発現し、各種の代謝系や、シグナルの知覚や伝達などの情報ネットワークを構築し、全体として驚くべきほど高度に組織化された情報系を形成し、機能している。筆者は日頃、このゲノム DNA というハードウェアに、塩基配列の形で収録されている生物の情報系を明らかにすることにより、生物の“情報系の組織化の原理”の一端が解明され、生物学が新たな局面に展開していくのではないかと、そこに“ゲノム解析”の意義があるのではないかと考えている。

これまでに酵母は、主としてその扱いやすさと発酵工業的な有用性のゆえに、基礎・応用生物学で広く研究対象とされ、真核生物の代表的な材料として、生物学

の発展に大きく貢献してきた。出芽酵母のゲノムは、大きさが大腸菌のゲノムの3倍にあたる約 14 Mb (10^6 塩基対)と推定されており、250 kb から 2 Mb まで、大きさの異なる 16 本の染色体より構成されている。このうち、重複したものを含めて、6.67 Mb の塩基配列がこれまでに決定されている（表1）。さらに現在、ヨーロッパ、アメリカ、カナダおよび日本の研究者が共同して、出芽酵母のゲノム DNA の組織的な塩基配列決定を進めている（表2）。現在の進行状況からして、出芽酵母は、そのゲノムの全構造が最初に明らかにされる真核生物であることはまちがいない。この出芽酵母のゲノムの全構造の解明は、出芽酵母を一例とした“生物の情報系の組織化の素過程”の解析を可能にする。したがって、その成果は、たんに出芽酵母の個々の遺伝子や遺伝子群の構造や機能の研究にとどまらず、生物がゲノムに蓄えた情報を統合し組織化する原理がどのようなものであるかの解析を可能にする。さらに、このようなゲノムレベルでの解析が、一種類の生物だけでなく、多くの生物について進めば、それらの生物のゲノム間で共通する要素や異なっている要素を抽出することが可能になり、現在の生物のゲノムがどのように設計され、進化してきたのかを知るうえでの、重要な手掛かりが得られるものと期待される。

もちろん、特定の生物のゲノムの全塩基配列決定が完成したときに可能になるのは、上述したような解析ばかりではない。生物の多くの遺伝子は、多かれ少なかれ他の生物のどれかの遺伝子と構造的・機能的な類似性を示すので、ひとたび、ある生物（たとえば出芽酵母）の全塩基配列が決定されれば、自分の研究対象とする生物あるいは遺伝子が何らかの理由で取扱いの難しいものである場合には、直接その生物あるいは遺伝子を対象とする代わりに、たとえば出芽酵母の類似する（あるいは、相同な）遺伝子を仮の対象として解析し、得られた結果か

† Genome Analysis of the Budding Yeast and Its Perspective.

Katsumi ISONO (Department of Biology, Faculty of Science, Kobe University, Rokkodai, Nada-ku, Kobe 657)

表 1 各種の生物のゲノム塩基配列決定の現状*

塩基配列の長さ	大腸菌	出芽酵母	ショウジョウバエ	線虫	シロイヌナズナ	マウス	ヒト
1～ 500	859	454	728	4685	2525	6758	32,627
501～ 1000	415	306	369	180	233	1530	3867
1001～ 2000	758	686	492	100	220	1730	3493
2001～ 4000	514	937	484	59	142	1055	2366
4001～ 6000	150	196	183	32	31	259	533
6001～ 8000	50	56	57	9	8	77	181
8001～10,000	17	22	27	6	7	341	90
10,000<	28	33	31	71	1	41	152
最大の長さ	176,196	315,338	80,423	54,962	10,014	94,647	180,388
最小の長さ	5	15	30	67	74	12	1
塩基配列の数	2791	2690	2371	5142	3167	11,491	43,309
総 計	5,570,115	6,666,606	4,607,305	4,369,651	2,014,945	11,099,881	30,970,041
平均の長さ	1996	2478	1943	850	636	943	715

* GenBank Release 80.0 (1993 年 12 月) に収録されているデータから抽出したもの。

ら本来対象とした遺伝子の機能を推測する、というような実験が可能になる。

このような理由から、ゲノム解析が生物学の新しい分野として定着しつつあり、現在、出芽酵母も含めて、大腸菌からヒトに至るまでのさまざまな生物を対象として、ゲノム DNA の組織的な塩基配列決定が行われている。出芽酵母のゲノムの塩基配列決定はかなり進んでおり、一昨年発表された第 3 染色体の全塩基配列決定⁽¹⁾をはじめ、多くの解析結果が発表されている。ところで、出芽酵母の最新の遺伝子地図⁽²⁾には 1046 個の遺伝子が登録され、それぞれの染色体上に位置づけられている。このほか、すでにクローン化はされているものの、染色体上の位置が不明のものも多数知られている。ちなみに、すでに塩基配列が決定され、データベースに収録されている遺伝子（機能不明な ORF を含む）は 2882 ある⁽³⁾。それでは一体、出芽酵母のゲノムにはどのくらいの数の遺伝子が存在するのだろうか？ 第 3 および第 6 染色体の整列クローンを用いて行った、組織的なノーザンハイブリダイゼーションの結果（すなわち、蛋白質をコードし、発現している遺伝子の解析）^(4,5)、および rDNA を除いたゲノムの大きさが約 12.5 Mb であることから推測すると、出芽酵母のゲノムには 5000 以上もの蛋白質をコードする遺伝子が存在すると推測される。したがってこれまでに知られている遺伝子の数は、この予想される数よりかなり少ないことになる。そこで新しい遺伝子探索法が問題になる。

その 1 つの方法としては、第 3 染色体および第 6 染色体の例^(4,5)のように、整列クローンを用いて組織的なノ

ーザンハイブリダイゼーションを行うことであり、他の方法としては、第 3 染色体の例⁽²⁾のように、まず染色体ごとに塩基配列を決定して、得られたデータから類推することであろう。前者の方法では、用いる mRNA をそれぞれ特定の条件下に置いた細胞から調製することにより、細胞周期、熱ストレス、培地の栄養素の変動、あるいは増殖の時期等により転写量が変動する遺伝子の探索が可能である。たとえば Price ら⁽⁶⁾は Olson らの作製した整列クローンバンクを利用して、細胞周期特異的に発現する遺伝子の網羅的探索を進め、300 のクローンから得られたデータを解析し、20 の細胞周期特異的遺伝子を同定している。筆者らも、筆者らが作製した整列クローンを用いて、熱ストレス、培地の糖源の変動、あるいは増殖の時期等により転写量が変動する遺伝子を数個同定し、解析している⁽⁷⁾。このように整列クローンは、特定の機能を持った遺伝子を網羅的に探索するのにきわめて有効であり、それによって機能的に関連する遺伝子や、相補的な遺伝子の同定や解析が可能である。

一方、すでに述べたように、出芽酵母の全ゲノムの塩基配列決定がヨーロッパの研究者を中心に、国際共同研究として進められている（表 2）。これまでに第 3 染色体の全塩基配列が発表されており⁽¹⁾、間もなく第 2、第 6 および第 11 染色体の塩基配列データも発表されることになっているという⁽⁸⁾。このうち、第 3 染色体には、塩基配列データから類推すると、長さが 100 アミノ酸残基以上の蛋白質をコードできる 182 個の ORF が存在し、遺伝子破壊実験により、少なくともそれらのうちの 3 遺伝子は生存に必須であるという。また、すでに遺伝

表 2 国際協力による出芽酵母ゲノム DNA の塩基配列決定プロジェクト^a

染色体番号	大きさ	担当国 ^b	統括者	開始年
I	200 kb	カナダ	H. Bussey	91 年 1 月
II	830	ヨーロッパ	H. Feldmann	91 年 1 月
III	310	ヨーロッパ	S. Oliver	(完了)
IV (右)	600	アメリカ	R. Davis	94 年 1 月
IV (中)	600	イギリス	B. Barell	94 年 1 月
IV (左)	600	ヨーロッパ	C. Jacq	94 年 1 月
V	610	アメリカ	R. Davis	92 年 1 月
VI	280	日本	Y. Murakami	90 年 1 月
VII	1200	ヨーロッパ	F. Foury	93 年 1 月
VIII	550	アメリカ	M. Johnston	93 年 1 月
IX	450	イギリス	B. Barell	91 年 1 月
X	760	ヨーロッパ	F. Galibert	93 年 1 月
XI	690	ヨーロッパ	B. Dujon	91 年 1 月
XII (右)	750	アメリカ	M. Johnston	94 年 1 月
XII (左)	450	ヨーロッパ	J. Koheisel	95 年 1 月
XIII	920	イギリス	B. Barell	93 年 10 月
XIV	800	ヨーロッパ	P. Philippsen	93 年 1 月
XV	1200	ヨーロッパ	B. Dujon	93 年 1 月
XVI (右)	500	イギリス	B. Barell	94 年 1 月
XVI (左)	500	カナダ	H. Bussey/J. Friesen	94 年 1 月

^a A. Goffeau からの私信による。^b ヨーロッパとは E C を中心とする諸国であり、イギリスとは連合王国をさす。

学的方法で知られていた 4 個の tRNA 遺伝子に加え、新たに 6 個の tRNA 遺伝子も見つかっている。今後の問題としては、このような塩基配列データから、どこにどのような機能を持った遺伝子が存在しているか、ということをおいかにして効率よく推測するかということであり、そのためのアルゴリズムを構築していくことであろう。これはいわば遺伝子発現に際して、RNA ポリメラーゼがどのようにして遺伝子の転写開始点を見出しているのか、という問題にもつながっていく。

DNA の塩基配列から遺伝子を探索することを、筆者らは現在、GENMARK というプログラム⁽⁹⁾を用いて行っており、いろいろと興味ある結果を得ている。このプログラムは、遺伝子（翻訳領域）内の“連続する塩基の出現頻度”には、遺伝子の塩基配列の翻訳が開始コドンから 3 塩基ずつ区切って行われ、かつ遺伝暗号には「縮退」(degeneracy)があり、それぞれのコドンの“使用頻度”が異なっていること、等が原因となって、統計的に大きな偏りがあることに着目し、一群の遺伝子の翻訳領域内の連続する塩基の出現頻度、および翻訳されない領域内の同様な連続する塩基の出現頻度、を基準データとして、未知の塩基配列データの、どこにどの程度の確率で遺伝子が存在するかという問題を、マルコフ過程

を用いて統計的に計算するというものである。筆者らは、このプログラムによる推測に加えて、fasta などのプログラムを用いたいわゆるホモロジー解析を行うことにより、かなり高い確率で遺伝子を見出している。

以上概括したように、ゲノム解析はまだ始まったばかりであるが、それは単に学問として面白いというだけでなく、農学や医学などの応用的な分野も含めて、今後の生物学を大きく変えていくポテンシャルを持っていることは明らかである。本稿が農芸化学の分野の読者に、ゲノム解析への関心を少しでも呼び起こすきっかけになれば幸いである。なお、出芽酵母のゲノム解析の詳細については、別の総説⁽¹⁰⁾を参照されたい。

- (1) S. G. Oliver, K. Isono, A. Yoshikawa, *et al.*: *Nature*, **357**, 38—46 (1992).
- (2) R. K. Mortimer, C. R. Contopoulou, and J. S. King: *Yeast*, **8**, 817—902 (1992).
- (3) GenBank (Genetic Sequence Data Bank) Release 80.0 (1993 年 12 月 15 日).
- (4) A. Yoshikawa and K. Isono: *Yeast*, **6**, 383—401 (1990).
- (5) A. Yoshikawa and K. Isono: *Nucleic Acids Res.*, **19**, 1189—1195 (1991).
- (6) C. Price, K. Nasmyth, and T. Schuster: *J. Mol. Biol.*, **218**, 543—556 (1991).

- (7) Y. Masuda-Kasai and K. Isono : (投稿準備中).
 - (8) B. Dujon : (Institut Pasteur, Paris, France) (私信).
 - (9) M. Borodovsky and J. McIninch : *Comput. Chem.*, **17**, 123—133 (1993).
 - (10) 田中誠司, 磯野克己 : 蛋白質 核酸 酵素, **39**, 718—728 (1994).
-