ラン藻の全遺伝情報の解明

かずさDNA研究所 佐藤修正, 田畑哲之

DNA塩基配列決定技術や配列情報処理システムの改良によって、ゲノムサイズが比較的小さいバクテリアなどでは、全ゲノムの塩基配列を数ヵ月から2年程度の期間で決定することができるようになった。1995年7月に初めての生物のゲノムの全構造、インフルエンザ菌 *Haemophilus influenzae* Rd株のゲノムの全塩基配列が報告されて以来¹⁾、すでにマイコプラズマ *Mycoplasma genitalium*²⁾、酵母 *Saccharomyces cerevisiae*、メタン細菌 *Methanococcus jannaschii*³⁾の全ゲノムの構造が公開されており、現在も有用遺伝子の探索を主な目的として30種をこえるバクテリアについてゲノムの配列分析が進められている。ゲノムの全塩基配列を手にするによって、ある生物のもつ遺伝子や遺伝子の候補をすべてリストアップし、その中から目的の遺伝子やそれに関わる遺伝子群を探し出すという新たなアプローチをとることが可能になった。

ラン藻は植物型(酸素発生型)光合成能をもつバクテリアの総称である。植物細胞のプラスチド(葉緑体はその代表的なもの)は、ラン藻の祖先が進化の過程で真核細胞に共生することによって生じたと考えられており、事実ラン藻と植物の間には、光合成のメカニズムや調節機構において共通点や類似点が非常に多い。そのため、ラン藻は、光合成反応メカニズムや関連蛋白質の発現制御機構の解析など植物型光合成研究のモデル材料として広く利用されてきた⁴⁾。このような状況のもと、1996年3月に単細胞性のラン藻、*Synechocystis* sp. PCC6803(以下PCC6803)の全ゲノムの塩基配列が決定され、遺伝子領域の推定など基本的な分析が行われた後9月に全データが公開された⁵⁾。これは、光独立栄養生物の全ゲノム構造としては初めての報告であった。このデータをもとに、ラン藻のもつさまざまな生物学的特性を遺伝子のレベルから理解しようとする試

みが行われつつある。

1. ラン藻ゲノムの全遺伝情報

ゲノム全体をカバーする約140種類のクローンやlong PCR産物の塩基配列を連結することによって、PCC6803のゲノムが3,573,470bpの長さをもつ環状ゲノムであることが明らかになった。平均GC含量は47.72%であった。

遺伝子の探索は、主に相同性検索とコンピューター予測プログラムによって行われた。その結果、全ゲノム上に2コピーのrRNA遺伝子クラスター(16S rRNA-Ile-tRNA-23S rRNA-5S rRNA)と42個41種類のtRNA遺伝子が同定された。また、蛋白質遺伝子としては、3,166個の存在が明らかになった。ただし、50アミノ酸以下をコードする小さな遺伝子については見落としがある可能性が高く、実際の遺伝子数はこれを数十程度上回ると考えられる。推定遺伝子産物の平均長は326アミノ酸残基で、最大のものは4,199アミノ酸残基であった。すべてのRNA遺伝子と蛋白質遺伝子を合わせると、全ゲノムの87%以上に相当し、平均遺伝子密度は1遺伝子/1.1 kbでバクテリアの平均的な数値であった。

全蛋白質遺伝子の翻訳アミノ酸配列をデータベースで検索したところ、3,166個のうち、1,402個(44.2%)はこれまでに報告されている場合も含めてその機能が推定できるもの、340個(10.8%)は他の生物で報告されている機能未知の遺伝子(hypothetical gene)に類

表1 PCC6803, *H. influenzae*において同定された遺伝子の機能分類

	PCC6803	<i>H. influenzae</i> Rd
アミノ酸生成	84	68
補助因子、補欠分子団、担体生成	108	54
細胞表面	64	84
細胞の諸機能	68	53
中間代謝	31	30
脂肪酸、炭脂質、ステロール代謝	35	25
エネルギー代謝	86	84
光合成及び呼吸	138	21*
塩基、ヌクレオチド代謝	39	53
調節機能	135	64
DNA複製、修復、組換え	49	87
転写	24	27
翻訳	144	141
輸送	158	123
その他	248	93
機能未知遺伝子	340	347

*文献1)ではエネルギー代謝のカテゴリーに分類されている。

Whole genome analysis of a cyanobacterium *Synechocystis* sp. PCC6803

Shusei SATO, Satoshi TABATA

Kazusa DNA Research Institute

似のもので、残りの1,424個については機能に関する手がかりが得られなかった。類似性を示した1,742個を機能別カテゴリーに分類すると表1のようになる。

2. ラン藻ゲノムの特徴

42個のtRNA遺伝子のうち、重複しているのはrRNA遺伝子クラスターに含まれるIle-tRNAのみであり、他はすべて1コピーである。41種類のtRNAは必要なすべてのコドンに対応している。tRNA遺伝子数をPCC6803を含む5種類のバクテリアで比較すると(表2)、2種類のグラム陰性バクテリア*H. influenzae* Rdと大腸菌*Escherichia coli*で遺伝子数が特に増加していることがわかる。また、同様の傾向はrRNA遺伝子クラスターについても観察される。

H. influenzae Rdは全ゲノムの塩基配列が最初に決定されたバクテリアで、1,830,137bp (PCC6803の51%)のゲノム上に1,743個 (PCC6803の55%)の蛋白質遺伝子が同定されている¹⁾。これの機能別遺伝子構成をPCC6803のものと比較することにより(表1)、PCC6803ゲノムの特徴を知ることができる。転写、翻訳などのカテゴリーでは遺伝子数にそれほど大きな差がなく、これらの過程は類似のメカニズムで維持されている可能性が高い。一方、遺伝子数に大きな差があるカテゴリーでは、種特異的な遺伝子重複や多様化が起きていると考えられる。

PCC6803に最も特徴的な遺伝子群は、光合成と呼吸であろう。光合成反応は、光化学系I、II、電子伝達系、ATP合成酵素複合体や他の多くの蛋白質が複雑な制御のもとに協調して進められる。また、その一部は呼吸経路とオーバーラップしていることもわかっている。ゲノム配列の分析から、PCC6803では138個の光合成、呼吸関連遺伝子が同定された。今後、塩基配列データを利用した遺伝子の機能解析によって調節機構の解析が進むとともに、さらに多数の関連遺伝子が発見されていくことが予想される。

光合成関連遺伝子以外では、調節機能関連遺伝子群に属する遺伝子数が多い点が注目される。これは、

PCC6803がセンサー・レギュレーターからなる2成分環境応答系遺伝子を多数もっていることによる⁶⁾。2成分環境応答系はバクテリアで一般的な情報伝達系で、真核生物でもよく似た情報伝達系が存在することが知られている。この系は環境の変動にตอบสนองしてリン酸化を介した情報伝達を行い、遺伝子発現を制御する。この系を構成するセンサータンパク質およびレギュレータータンパク質には保存領域があり、それぞれリン酸の供与、受容に直接関係する。PCC6803では、79個の遺伝子がこれらの保存領域をもつ。ラン藻は、さまざまな環境下に存在することが知られており、多数の環境応答系遺伝子をもつことで環境の変化に対応しているのであろう。

最初にも述べたように、プラスチドとラン藻には系統進化的に関係が深い。これまでに全塩基配列が決定されたプラスチドゲノムのほとんどは120kbから190kb程度の環状分子であるが⁷⁾⁻¹⁶⁾、これは共生によって不要な遺伝子が脱落するとともに一部の遺伝子が核に移行したためであると考えられる。図1は、PCC6803ゲノムと6種類の陸上植物の葉緑体ゲノム上の遺伝子を比較したものである。葉緑体ゲノムには130個程度の遺伝子が存在するが、このうちの多数がPCC6803ゲノムで保存されていることがわかる。藻類のプラスチドゲノムとの類似性はさらに高く、約90%の遺伝子についてその相同遺伝子がPCC6803ゲノム上に存在する。

a: *Epifagus virginiana* プラスチド
b: クロミツクロプラスト
c: ゼニゴケクロプラスト
d: イネ及びトウモロコシクロプラスト
e: タバコクロプラスト
f: PCC6803

図1 PCC6803とプラスチドゲノムを含む遺伝子数の比較

PCC6803と6種類の陸上植物のプラスチドゲノムに共通に存在している蛋白質遺伝子の関係を示した。数字は各集団に含まれる遺伝子の個数を示している。

表2 5種類のバクテリアゲノムに存在するrRNAおよびtRNA遺伝子数の比較

		rRNA遺伝子 クラスター	tRNA遺伝子
PCC6803	(藍色細菌)	2	42 (41種類)
<i>H. influenzae</i> Rd	(グラム陰性菌)	6	56 (55種類)
<i>M. genitalium</i>	(グラム陽性菌)	1	33 (33種類)
<i>M. jannaschii</i>	(古細菌)	2	37 (35種類)
<i>E. coli</i>	(グラム陰性菌)	7	79 (46種類)

3. おわりに

ゲノムプロジェクトから得られる大量の遺伝情報は、いうまでもなく生物学に大きなインパクトを与えている。ゲノム上の遺伝情報のどの部分に着目するかは研究者の興味にかかっており、PCC6803の場合にも、本稿で紹介した以外にさまざまな視点から数多くの興味深い知見が得られている。しかしながら、塩基配列データからどのような情報をどのようにして抽出するか、さらに個々の遺伝子の機能をいかにして明らかにしていくかが今後の重要な問題として残されている。PCC6803では、全遺伝子の約55%、1,700個あまりについては機能に関する手掛りが得られていない。したがってこれをいかにして解明していくかが今後の大きな課題である。他の生物の場合も含めて、一次構造のみがわかっている数千の遺伝子について機能の解析を進めようとするには、方法論を含めた発想の転換が必要であろう。一方、遺伝子の機能解析のための技術的ブレイクスルーがない現状では、コンピュータープログラムによって遺伝子領域や遺伝子産物の機能を推定することが最も有効な手段のひとつとなっている。そして、今後この分野の重要性はさらに高まると予想されている。塩基配列データを実験材料としてピペットマンのかわりにキーボードを道具として生物学を行う研究者が求められている。

なお、PCC6803の塩基配列データや種々の解析結果は、インターネット上のデータベース CyanoBase [<http://www.kazusa.or.jp/cyanobase/>] で公開されている。

文 献

- 1) Fle. schmann, R. D., Adams, M. D., White, O. *et al.* (1995) *Science* **269**, 496-512.
- 2) Fraser, C. M., Gocayne, J. D., White, O. *et al.* (1995) *Science* **270**, 397-403.
- 3) Bult, C. J., White, O., Olsen, G. J. *et al.* (1996) *Science* **273**, 1058-1073.
- 4) Briant, D. A. (ed.) (1994) *The Molecular Biology of Cyanobacteria*, Kluwer Academic Publishers.
- 5) Kaneko, T., Sato, S., Kotani, H. *et al.* (1996) *DNA Res.* **3**, 109-136.
- 6) Mizuno, T., Kaneko, T. and Tabata, S. (1996) *DNA Res.* **3**, 407-414.
- 7) Shinozaki, K., Ohme, M., Tanaka, M. *et al.* (1986) *EMBO J.* **5**, 2043-2049.
- 8) Ohyama, K., Fukuzawa, H., Kohchi, T. *et al.* (1986) *Nature* **322**, 572-574.
- 9) Hiratsuka, J., Shimada, H., Whittier, R. *et al.* (1989) *Mol. Gen. Genet.* **217**, 185-194.
- 10) Wolfe, K. H., Morden, C. W. and Palmer, J. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10648-10652.
- 11) Hallick, R. B., Hong, L., Drager, R. G. *et al.* (1993) *Nucl. Acids Res.* **21**, 3537-3544.
- 12) Wakasugi, T., Tsudzuki, J., Ito, S. *et al.* (1994) *Proc. Natl. Acad. Sci. USA* **91**, 9794-9798.
- 13) Maier, R. M., Neckermann, K., Igloi, G. I. *et al.* (1995) *J. Mol. Biol.* **251**, 614-628.
- 14) Stirewalt, V. L., Michalowski, C. B., Löffelhardt, W. *et al.* (1995) *Plant Mol. Biol. Reptr.* **13**, 327-332.
- 15) Reith M. and Munholland, J. (1995) *Plant Mol. Biol. Reptr.* **13**, 333-335.
- 16) Kowallik, K. V., Stoebe, B., Schaffran, I. *et al.* (1995) *Plant Mol. Biol. Reptr.* **13**, 336-342.