3PA125

「遺伝子型による種同定法」の確立——
ゲノム微生物生態学を目指して○西垣 功一、モハムド・ナイムディン、長谷川 孝、
齋藤 あゆむ、渡辺 雄大、マツシュビキニ (埼玉大・工)

ゲノム全塩基配列決定が進んでいるが、遺伝子型での種同定はまだ確立していない。今回、われわれはこれを実現する方法論を確立したので発表する。その方法は、基本的には、既に我々が発表してきた方法（ゲノムプロファイリング(GP)法 (1994)）に外ならないが、新たに、「種同定必要情報量」の考察と迅速定量的情報抽出法 (PaSS) の開発などにより、種同定一般に用いることが可能であることを示した。さらに、このことが今までにアプローチ不可能であった領域すなわち『ゲノム微生物生態学』の開拓も、夢ではなくなった。

現存する全生物のゲノムを一義的に指定するのには、ただか200ビットの情報があれば足りることと、GP法——ランダム PCR 産物を変性勾配ゲル電気泳動法により解析するもの——は、この程度の情報量ならば1回の試行により充足していることを示している。GP法の基礎技術のランダムPCRは、類縁種間あるいは個体間から、対応する「共通保存遺伝子領域断片 (CCGF)」を点突然変異寛容的に再現的に切り出してくるという優れた性質をもっている。このために、近縁種間については極めて精度の高い類縁性比較を可能としている。同時に、近縁性が低い生物間の比較においても同一の基準のもとで行うことができ、まさに広範な生物比較同定に適している。ゲノム微生物生態学は、「土壌細菌」、「腸内細菌」など種々な microflora の構成微生物の種類と数を数え上げることを可能とすることから始まる（微生物相互作用や環境影響はこれらのデータなしには求めるべくもない）。実にGP法はこれを可能とする。

K. Nishigaki, M. Naimuddin, T. Hasegawa, A. Saito, T. Watanabe, M. Biyani: Establishment of genotype-based identification of species toward *Genome Microbial Ecology*.

3PA127

オリゴスティックネスで見たゲノムの特徴

○斉藤 あゆむ、西垣 功一 (埼玉大・工)

近年、様々な生物の全塩基配列が決定され、ゲノム解析が進みつつある。我々は、ゲノム解析の手法として“*oligo-stickiness*”解析を提唱している。これは、特定のオリゴヌクレオチドがゲノムのどの位置にどれ位の安定性で結合するかを表わすもので、元々一般PCR産物解析プログラム(PCRAna-A1)として開発されたプログラムの応用でもある。

今回、大腸菌や枯草菌、およびその他の細菌類のゲノムと真核細胞生物の酵母などのゲノムを解析した結果、i)ある特定のオリゴヌクレオチドが特定の生物種の高い(または低い)“*oligo-stickiness*”を示すこと、ii)プロモータ領域配列と同じ配列のオリゴヌクレオチドなどが極端に低い“*oligo-stickiness*”を表わすこと、また iii)通常のゲノム解析(リピート配列、GC含量等)とは有意に異なる解析方法である。こと等が明らかになった。この他にも生物種固有の“*oligo-stickiness*”地図が得られている。ごく最近明らかになったゲノム塩基配列の“*oligo-stickiness*”も合わせて報告する。

A. Saito, K. Nishigaki :
Intrinsic genome structures observed by ‘*oligo-stickiness*’

3PA126

細胞間ゲノム‘準’距離測定のための
*Genome Profiling-PaSS*の導入

○渡辺 雄大, Mohammed Naimuddin, 西垣 功一 (埼玉大・工)

細胞を遺伝子型のみで同定する一般的方法論はまだ確立していない。今回、我々の研究室で開発したゲノム解析技術 *Genome Profiling*(GP)が、これを可能としている実証の下に、さらにそこでコンピュータを用いた迅速・高密度処理を行い、客観的、定量的情報を得ることを試みた。なお、GP はランダム PCR (任意のプライマーを用いてテンプレート中から複製 DNA を得る PCR 法)と TG/DG-GE (温度/変性剤濃度-勾配ゲル電気泳動法)からなる。

解析法として、まず GP の画像情報から DNA の熱変性 (又は変性剤変性) に由来する特徴点を抽出し、規格化したのちゲル電気泳動におけるその移動度と変性温度 (濃度) を得た。これらの点を種同定点群 (species identification dots); *spiddo* と名称した。近縁生物種の場合には、大部分これらは相互に対応する点として識別可能であり、この事実からこれらの点群を一種のパターンと見なし、その類似性を計量するパラメータ *PaSS* (Pattern Similarity Score) の導入が可能になった。

我々は、この *spiddo* 作成処理の客観化・自動高速化を検討した。現在、画像情報から「線分抽出」やそれらの線分の「関数表現」などに成功している。自動化のためにバンドパターン上の任意の点において、その近傍では色の濃淡情報が近いという「連形成による参照点シフト方式」という新しいアルゴリズムを導入し、自動抽出したバンドパターンの関数化を行った。また比較的解析の容易な GP データに関しては成功を収めている。これらのことについて報告する。

T. Watanabe, M. Naimuddin, K. Nishigaki:

Introduction of Genome Profiling-PaSS to measure the cell-to-cell genome sub-distance.

3PA128

アミノ酸レベルの特徴の利用による原核生物ゲノム塩基列の

蛋白質コード領域推定の高精度化

○広沢 誠 (かずさDNA研究所)

生物のゲノム構造データから機能解析を進めるための第一歩としては、蛋白質のコード領域推定の精度を上げることであり、そのためには特定の生物種において、計算機によるコード領域推定法をほぼ完全なものにする必要がある。この目的のため我々はラン藻をモデル生物と選択し、その方法を検討してきた。我々はこれまでに、GeneMark-RCと名付けた手法により98.4% (3122/3168)の正当率を達成している。この手法では、ゲノム配列からGC含量が異なる2種類の遺伝子群を自動的に分類し、それに基づいてコード領域を推量する。今回はこの方法を拡張し、核酸レベルの情報だけでなく、アミノ酸レベルの情報を同時に利用することにより、false-negativesの出現を低く抑えながら、さらに高い正当率を達成できた。

98.4%の正当率を達成してわかったことは、GeneMark-RCで採用している塩基のhexamer統計は、false-negativesの出現を低く抑えながら (122/3244)、非常に高い正当率を達成することができるが、これを用いて残りのコード領域も推定することは難しいということである。なぜならばそれらにおいては、連続する2アミノ酸の出現頻度が他のものとは異なるからである。したがって、これらを推定するためにはモノコドン統計のほうが有利である。しかし、モノコドン統計のみを用いて推計するとfalse-negativesの出現が高くなり、精度が落ちるという難点があった。そこで我々は、モノコドン統計でコード領域と検出したものをアミノ酸に翻訳した時に、特定の蛋白質としての特徴を持つ場合のみ、それをコード領域として加えることにした。今回は、膜貫通領域推定プログラムとして定評があるSosuiが膜蛋白質であると判定したもののみをコード領域として加えることにした。結果として、false-negativesの出現を137/3271に抑えながらも、98.9% (3134/3168) (注釈レベルでは99.1%)の正当率を達成することができた。この改良により拾いあげられた遺伝子の中には、ラン藻に必須である光合成関連の遺伝子が9個含まれているので、推定の質的な向上もあったと言える。今後は、膜蛋白質遺伝子以外も正しく抽出できるように拡張する予定である。

M. Hirose: Accuracy enhancement for the prediction of protein-coding regions by utilizing the amino acid-level analysis