3P289

自動判定法による転写因子の網羅的同定と比較ゲノム解析

○峯崎 善章、西川 建（国立遺伝研）

これまでに我々は構造ゲノムプロジェクトのターゲット選択を目的として、自動的に転写因子を収集・分類する方法を開発してきた。この方法を用い公開されている全ゲノムの ORF から転写因子を網羅的に同定してデータベース (TTDB-TF) に収集・分類し、これを用いて比較ゲノム解析を行った。

本研究では直接 DNA に結合する転写因子 (activator, repressor など) に限定した。DNA 結合ドメインを持つことを必要条件とし、構造ドメイン (SCOP) と機能ドメイン (Pfam) を手がかりに、ドメインの組み合わせを経験ルールとして記述した。原核生物 (古細菌 18 種、真正細菌 136 種) にこの方法を適用し、約 20000 の転写因子を収集し、67 のファミリーに分類した。これまで転写因子の DB として真核生物では TRANSFAC があるが、原核生物の種横断的な DB はなく、個別のゲノムの文献や、実験データを集めた SwissProt などでは自動判定法の妥当性を確認した。

種ごとの転写因子の数をゲノムサイズ (全 ORF 数) に対してプロットすると、大まかに 3 つのグループに分かれた。第一は真正細菌の大部分を含むグループ (大腸菌や枯草菌を含む) で転写因子数は全 ORF 数にほぼ比例して増加していた。第 2 のグループは古細菌、シアノバクテリア、スピロヘータなどのグループで第一のグループと比較して明らかに転写因子は少ない。第 3 のグループは全 ORF 数は 1500 未満の寄生・共生細菌のグループで転写因子は極端に少なく高々数個から十数個であった。

詳細は当日報告する。

Y.Minezaki, K.Nishikawa : Automatic identification of transcription factors (TFs) and Comparative genomics of TFs

3P291

Power Law の進化モデル：トランスクリプトームと遺伝子ファミリー

○小笠原 理、久保田 功、大久保 公策（遺伝研・DDBJ・発現解析）

哺乳類から大腸菌にいたるまでの広範な生物種において、細胞内における遺伝子の発現量の分布が Power Law に従うことが、いくつかの研究で報告されている。この現象の原因についてこれまでにいくつかの数理モデルが提案されているが、それらは細胞内の遺伝子発現のダイナミクスで説明するモデルと、発現量の進化のパターンから説明するモデルの 2 種類に大別できる。本研究ではコンピュータシミュレーションを用いて、この現象が細胞時間ではなく進化時間を想定したモデルで説明すべきであることについて論ずる。このモデルは最近注目され始めている遺伝子発現の進化の中立説と密接なかわりを持つ。この説に従えば、構造の似た遺伝子の発現パターンは似通っていて然るべきとなる。一方、大野乾博士が提唱した遺伝子重複による進化の理論によれば、遺伝子重複の直後に、重複してできた 2 つの遺伝子のうちのいずれかが新しい発現パターンまたは機能を速やかに獲得しなければ、いずれは一方の遺伝子が偽遺伝子となり機能を失ってしまうであろうと考えられている。つまり構造の似た遺伝子の解剖学的発現パターンは異なっていなければならない。実際にはどうであるかについて実験データを用いて答えるためには正しく gene family が分類できていることが必要である。トランスクリプトームの Power Law の説明とほとんど同じ数理モデルにより、遺伝子ファミリーのメンバー数の分布が Power Law に従うことが説明できることが報告されている。正しく gene family が分類された結果はどのようなものか、について考えるために、このモデルを用いて遺伝子ファミリーの innovation rate や個々の遺伝子の進化速度などのパラメータが遺伝子ファミリーのメンバー数の分布にどのように影響するのかについて考察する。

O.Ogasawara, I.Kubota and K.Okubo : Evolutional model of power law behavior in transcriptome and gene family

3P290

原核ゲノム 44 種におけるタンデムリピートの解析とゲノム進化

水田 智史¹、アブリミティ エイメイティ¹、○木村 隆雄²、大沢 研二³、清水 俊夫¹ (¹ 弘前大・理工・電子情報システム工、² 青森中央学院大・経法・経法、³ 群大院・工・ナノ材料システム工)

タンデムリピートは、それ自体、遺伝子配置や遺伝子進化を議論する上で重要な役割を果たすものであるが、我々はそれをさらに押し進め、ゲノム上の異なる場所に位置する類似したタンデムリピートを解析することにより、全ゲノム重複等のゲノム進化に関する何らかの知見が得られるのではないかと考えた。

そこで、44 生物種のゲノムにおいて検出したタンデムリピートを解析した結果、2 回の全ゲノム重複がはっきりと見てとれるもの、1 回の全ゲノム重複がはっきりしているかまたは 2 回の全ゲノム重複の痕跡が曖昧なもの、ゲノム重複の痕跡がほとんど見られないもの、ゲノム重複の痕跡が全く見られないもの、の 4 種類に分類することができた。

なお、ゲノム進化を議論するためには、短く一致度の高いタンデムリピートだけでは不十分であると考え、本研究では、長くかつ一致度の低いタンデムリピートも検出可能なカラーコーディング法を用いて検出を行った。図は 2 回の全ゲノム重複がはっきり見て取れるものの一例で (*Streptococcus pneumoniae*)、円で表したゲノムにおいて相互に一定値以上の一致度をもつタンデムリピート同士を直線で結んだものである。

S.Mizuta, A.Abulimiti, T.Kimura, K.Oosawa and T.Shimizu : Analysis of tandem repeats found in 44 prokaryotic genomes and genome evolution

3P292

高度好塩性古細菌 Haloarcula 属のレチナールタンパク質ファミリーの進化

○井原 邦夫（名大・遺伝子）

これまで、高度好塩性古細菌において 4 種類のレチナールタンパク質が存在することが明らかとなっている。それらは、光プロトンポンプのバクテリオロドプシン (bR)、塩素イオンポンプのハロロドプシン (hR)、光センサーとして機能するセンサーロドプシン (sR)、フォボロドプシン (pR) である。昨年の本大会で、現在までに 6 属、24 種類の好塩菌において、少なくとも一つのレチナールタンパク質 (遺伝子) が存在する事を報告した。一方、Ha. marismortui のゲノム (URL: <http://zdna2.umbi.umd.edu/haloweb/hma.html>) に対して、bop をプローブに用いて BLAST 解析 (tblastn) を行くと、6 種類の有意な相同性を示す遺伝子が見いだされた。それらは、bop, hop, sop, pop の 4 種類と、bop に非常に近い遺伝子 (ここでは blp と呼ぶ) と、sop と pop の両方に似ている遺伝子 (ここでは slp と呼ぶ) であった。今回、8 種類の Ha 属の菌株から、遺伝子順序が保存されていると仮定して PCR 反応を行い、これらの 6 種類のレチナールタンパク質遺伝子の増幅を試みた所、6 種類の遺伝子すべてを検出できた株 (bop, hop, pop の 3 種類だけしか検出できなかった株 (sop が無い!) に分けられた。これらの遺伝子 DNA 配列を基にして分子系統樹解析を行った所、Ha 属の分子種の多様性は、(少なくともレチナールタンパク質の遺伝子においては) 外来遺伝子の取り込みやすさに因るものではなく、遺伝子重複 (欠失?) が生じやすい事に起因するものと推定した。また、遺伝子重複は、Ha 属が形成される時点ですでに生じていて、種分化していく過程において、いくつかの遺伝子 (sop, slp, blp) の欠失が起こった事が推定できた。本大会では、これらのレチナールタンパク質遺伝子の発現状況を含めて、機能・進化の関連を議論する予定である。

K.Ihara : Evolutionary analyses of archaeal type retinal proteins from a genus Haloarcula