

ごとう おさむ
 後藤 修
 o.gotoh@aist.go.jp
 生命情報科学研究センター

ホモロジーを用いた遺伝子構造予測

— その薬物代謝酵素遺伝子への応用 —

ヒトをはじめとする5種の真核生物全ゲノム配列の概要がすでに公開され、解読されるゲノムは今後ますます増えるものと予想される。生命の設計図であるゲノム配列を読み解き、有用な情報を得るための第一歩は、そこに含まれる遺伝子を同定することである。しかし、真核生物の遺伝子では、タンパク質をコードする領域(翻訳領域-エキソン)とそれ以外の領域(イントロンや5'、3'非翻訳領域)とが混在しており、計算機を用いてその正確な内部構造を推定することは容易でない。この問題は遺伝子発見問題とよばれ、この10年ほどの間に大きく進展した。通常、翻訳領域における3~6塩基出現頻度の偏りや、エキソン・イントロン境界近辺の配列の特徴を統計的に処理し、翻訳領域らしさを判断する手法が採られている。しかし、この方針による予測精度には限界がある。

予測精度を高め、産物であるタンパク質の立体構造や機能をより正確に推測するために、我々は既知のアミノ酸配列とのホモロジーを積極的に利用する方法を開発した¹⁾。この方法では、未知の遺伝子を含むゲノム配列を仮想的に翻訳して参照配列との対応関係を求め、対応関係が得られる領域を翻訳領域とみなす。このとき、配列の類似性ととも、上に述

べた統計情報も同時に考慮する(図1)。参照配列と50%以上アミノ酸が一致すれば、96%以上の精度で翻訳領域に含まれる塩基を予測できることが確かめられた。

ほとんどの動植物では、外来性の有毒な化学物質を代謝して除去する機能を、チトクロームP450を中心とする薬物代謝酵素が担う。これらの酵素の遺伝子多型が、薬物や発がん物質に対する感受性の個人差に深く関与している。我々の開発した方法を用いて、ゲノム配列が解読された生物種のすべての薬物代謝酵素遺伝子を見出す試みを始めている。ヒト(図2)、ショウジョウバエ、線虫²⁾では、それぞれ60~80の、シロイヌナズナでは250以上のP450遺伝子をすでに同定している。各酵素の基質、産物、遺伝子多型などの情報を含む総合的データベースの構築を現在進めている。

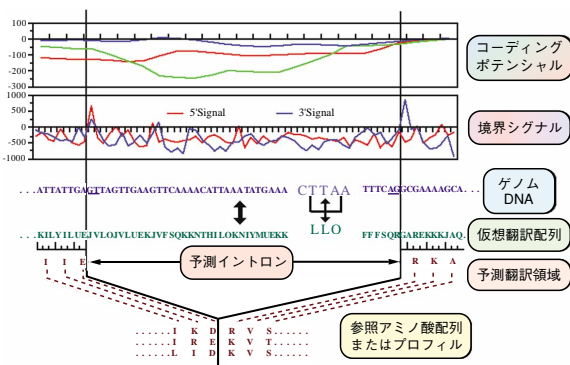


図1 遺伝子構造予測に用いる各種の情報

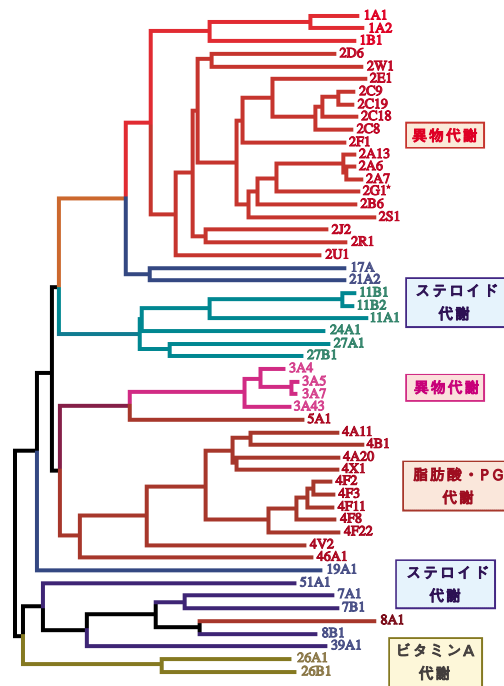


図2 ヒトチトクロームP450の系統樹

関連情報

- 1) O. Gotoh: *Bioinformatics* **16**, 3, 190-202 (2000).
- 2) O. Gotoh: *Mol. Biol. Evol.* **15**, 11, 1447-1459 (1998).