

隠れマルコフモデルによるホモロジー検索に基づくタンパク質機能推定法

Functional annotation of protein sequences by hidden Markov model

京都大学化学研究所 バイオインフォマティクスセンター 化学生命科学領域 荒巻 拓哉

研究成果概要

KEGG Orthology (KO) は遺伝子を機能と配列類似性を元に分類したもので、各 KO の ID (K 番号) と遺伝子を対応させることで機能アノテーションを行える。機能未知のタンパク質配列と KO を対応させるツールとしては、現在 BlastKOALA と GhostKOALA が主に使われており、一般に正確性では前者、計算時間では後者がより優れる。

本研究では既存のツールに代わる新たな手法として、HMMER を利用したタンパク質機能推定法の確立を目指している。各 KO のレファレンス配列からその KO 固有のプロファイル HMM を作成し、クエリのタンパク質の検索に用いた。検索結果のスコアが各 KO に決められた閾値を超えた場合にその配列は当該 KO に含まれると推定した。

本手法の評価のため、まず KEGG GENES の原核生物の遺伝子からいくつかの属を抜き取り、残りの配列からプロファイル HMM を作成した。そして抜き取った属のうち KO が手動でアノテーションされている生物種の遺伝子を選び、それをクエリとしてアノテーションを行った。また同じクエリを BlastKOALA、GhostKOALA を用いて同様にアノテーションした。

この結果、BlastKOALA、GhostKOALA はいずれも K 番号がついた遺伝子の 90% 以上を正しくアノテーションしたが、本手法では 87% だった。一方実行時間は本手法が BlastKOALA の約 60 倍、GhostKOALA の約 1.3 倍速かった。

以上より、HMMER を用いたタンパク質機能アノテーションは計算時間の点で従来手法よりも優れていることがわかった。

発表論文(謝辞あり)

発表論文(謝辞なし)

本年は共になし。