

KEGG Orthology に関する研究  
Research about KEGG Orthology

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遺伝子機能データベースで、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。機能未知のアミノ酸配列に KO を割り当てるツールとしては、現在 KAAS、BlastKOALA、GhostKOALA の 3 つが存在する。

本研究では既存のツールに代わる新たな手法として、KO データベースの配列をプロファイル HMM 化した KOfam データベースを構築し、隠れマルコフモデル (HMM) による配列検索を用いた遺伝子機能予測法を開発した。またこの予測法に必要となる、ある配列がある KO に含まれるかの判定基準となる類似度スコアの閾値を KO ごとに設定する方法を開発した。具体的には、KO 配列の 3 分の 2 で作成したプロファイル HMM を用いて、残りの 3 分の 1 の配列および他 KO の配列を検索することで、同じ KO の配列のみをよく判別できる類似度スコアの閾値を探索するという方法である。

KOfam データベースによる遺伝子機能予測法の予測精度および計算時間を KAAS、BlastKOALA、GhostKOALA と比較したところ、計算時間の面で各ツールよりそれぞれ 1.9～3.4 倍、68～87 倍、1.05～1.7 倍の高速だった。また予測精度でも KAAS を上回り、BlastKOALA、GhostKOALA と同等の予測精度を示した。

開発した KOfam データベースおよびそれを用いた遺伝子機能予測ツールは、GenomeNet (<https://www.genome.jp/>) で公開する予定である。

発表論文(謝辞あり)

発表論文(謝辞なし)

本年は共になし。