Evola：全転写産物解析に基づくヒト遺伝子の分子進化データベース

Evola: Evolutionary Annotation Database of the Whole Human Transcriptome

○坂手龍一 (1)、松矢明宏 (2, 3)、五條堀孝 (1, 4)、今西規 (1)
(1産総研・バイオメディシナル情報研究センター、2バイオ産業情報化コンソーシアム、3日立製作所、4遺伝研・生命情報・DDBJ研究センター)

Evola (http://hinv.jp/evola/)はヒトと脊椎動物（13種）とのオルソログ情報を解析し提供している。ヒト遺伝子の理解にはモデル動物との種間比較が不可欠であり、種間比較には種分化によって分かれた遺伝子であるオルソログ(ortholog)情報が必要である。Evolaではゲノムアラインメントの作成と全転写産物(transcript)のゲノム配列へのマッピングを独自に行い、種間でのゲノム保存領域と遺伝子との重なりを調べ、アミノ酸配列のアラインメントを行ってオルソログの同定を行っている。分子進化データベースとしてEvolaはヒトの22,496遺伝子のオルソログ情報を格納し（2009年3月更新のrelease 6）、アミノ酸配列のアラインメントや系統樹、スプライシングバリアントなどの種間比較情報を提供している。また、“Gene family/group browser”ではヒトとチンパンジー、マカクザル（アカゲザル）、マウス、ラットについて、遺伝子ファミリーレベルでの種間比較（オルソログとパラログのゲノム位置など）を可能としている。

Evolaとは

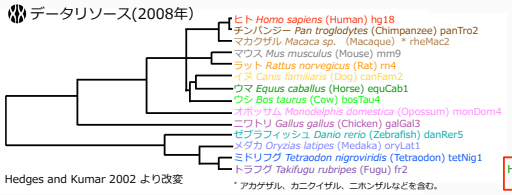
全転写産物解析で定義された、情報量に富むH-InvDBのヒト遺伝子セットについて、近縁な生物（13脊椎動物）とのオルソログを、ゲノムと転写産物の配列比較によって解析。最新かつ包括的なオルソログ情報を、学生から研究者まで幅広いユーザーに直感的にわかりやすく提供することを目的としている。

最新リリース：Evola 6.0（2009年３月更新）の主な特徴

- ➢H-InvDB 6.xに対応：ヒト遺伝子のオルソログが5.xより2,346増加
- ➢スプライシングバリアント(H-DBAS)レベルでの解析
- ➢偽遺伝子およびH-Inv予測遺伝子(eHIT, pHIT)についても解析
- ➢ユーザーインターフェースの改良

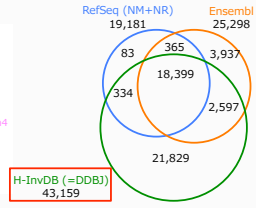
データ構築

データリソース(2008年)



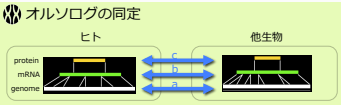
Hedges and Kumar 2002 より改変

解析対象生物種：ヒト＋13脊椎動物
（ゲノム：UCSC、転写産物：DDBJ, Ensembl, RefSeq）

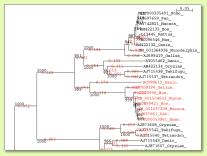


全転写産物解析に基づくヒト遺伝子数
（転写産物クラスター数）

オルソログの同定



1a. Genome alignment by BLASTZ
1b. mRNA – Genome alignment overlap: ≥240 nt
1c. Protein alignment: ≥1/2 of human (aa)



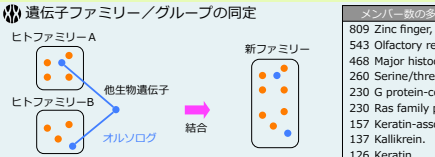
STEP 1. Computational analysis
ゲノムアラインメントと転写産物マッピングをもとに、ゲノム、転写産物、タンパク質の比較によるオルソログの同定

STEP 2. Manual curation*
系統樹での位置関係が生物種と遺伝子で一致するかを判定

*現在はツールにてほぼ自動化を実現

ヒトの22,496遺伝子（代表配列）のオルソログ（STEP1）
うち13,241遺伝子について１つ以上のManual curation判定を付加

遺伝子ファミリー／グループの同定



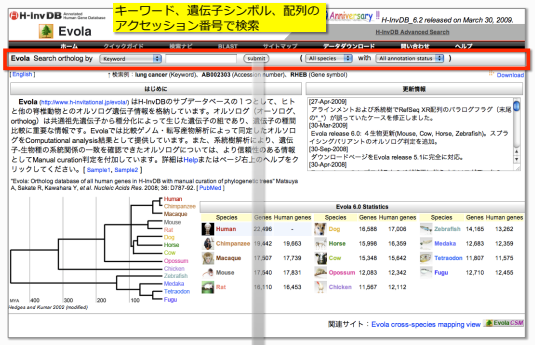
STEP 1. アミノ酸配列のシングルリンケージクラスタリング
（相同性条件はGu Z et al. *Mol. Biol. Evol.* 2002による）
STEP 2. 他生物１遺伝子がヒト複数遺伝子とオルソログの場合に結合

メンバー数の多い遺伝子ファミリー　トップ１０（名前は代表的な遺伝子）
809 Zinc finger, C2H2-type domain containing protein.
543 Olfactory receptor (OR).
468 Major histocompatibility complex.
260 Serine/threonine-protein kinase. Mitogen-activated protein kinase.
230 G protein-coupled receptor.
230 Ras family protein.
157 Keratin-associated protein.
137 Kallikrein.
126 Keratin.
105 Ring finger protein.

3,505遺伝子／グループ（ヒト２遺伝子以上を含む）を同定

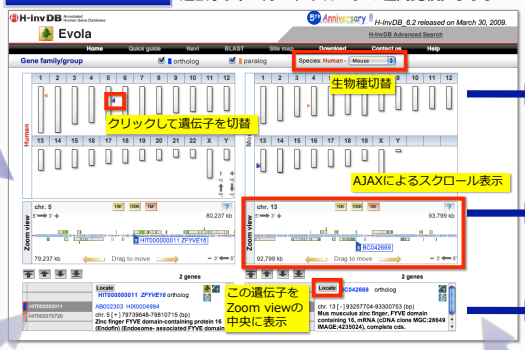
サイト構成

トップ／検索ページ



検索結果

Gene family/group　遺伝子ファミリー／グループの種間比較ブラウザ



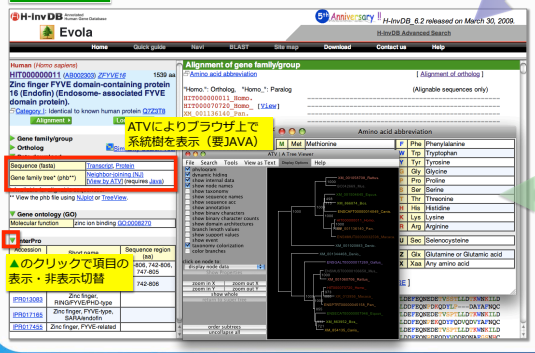
左にヒト、右に他生物（チンパンジー、マカクザル、マウス、ラット）を表示

①染色体一覧
▲：オルソログ
▲：パラログ

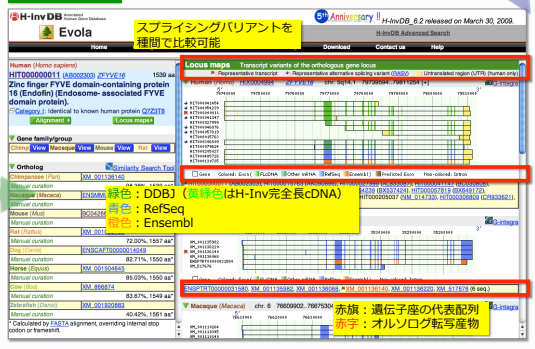
②遺伝子近辺の拡大図（“Zoom view”）

③遺伝子アノテーション情報

Alignment　オルソログ情報：アラインメント表示



Locus maps　オルソログ情報：遺伝子座表示



参考文献

オルソログ：
“Evola: Ortholog database of all human genes in H-InvDB with manual curation of phylogenetic trees.”
Matsuya A, Sakate R, Kawahara Y, et al.
Nucleic Acids Res. 2008 Jan; 36(Database issue):D787-92.

ゲノムアラインメント：
“A web tool for comparative genomics: G-compass.”
Fujii Y, Itoh T, Sakate R, et al.
Gene. 2005 Dec 30;364:45-52.

H-InvDB
Annotated Human Gene Database

National Institute of Advanced Industrial Science and Technology AIST
BIRC Biomedicinal Information Research Center
産業技術総合研究所　バイオメディシナル情報研究センター(BIRC)　分子システム情報統合チーム(IDB)
〒135-0064 東京都江東区青海2-42 産総研　臨海副都心センター別館　バイオ・IT融合研究棟7階　http://unit.aist.go.jp/birc2/idb/