

Ensembl

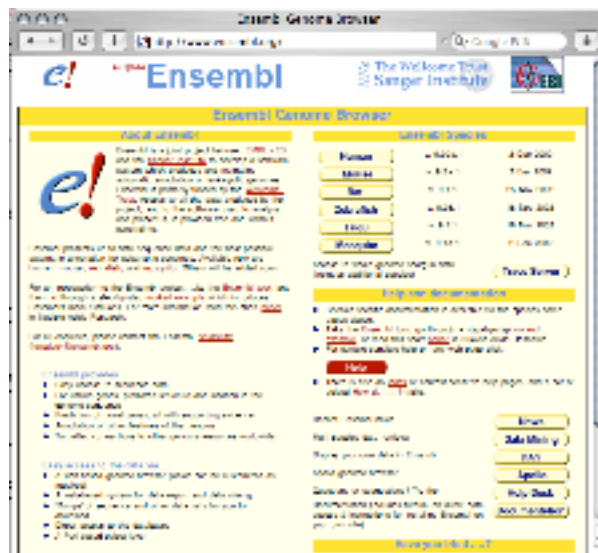
ディー・エヌ・エー マイクロアレイドットオルグ 坊農秀雅

真核生物のゲノムが簡単に検索、
閲覧できる

<http://www.ensembl.org/>

キーワード

ゲノム、ゲノムアノテーション、遺伝子
構造、データマイニング



1. 基本的な知識を学ぼう

1.1 Ensembl とは？

Ensemblのトップページ（上記 URL 参照）によると以下のような説明がある。

「Ensembl とは EBI(European Bioinformatics Institute) と Sanger Institute による共同プロジェクトで、真核生物のゲノムアノテーションを自動的に生成し、維持するソフトウェアシステムの開発を行っている。Ensemblは主に Wellcome Trust 財団により主に資金提供を受けている。Ensemblプロジェクトによって作成されたすべてのデータ、ならびにそれらのデータを解析し可視化するのに用いられるソフトウェアへのアクセスは無償で、なんの制約もなく提供されている。」

おそらく'ensemble'のもじりで命名されたと考えられる。リーダーズ英和辞典第2版によると、音楽用語で「アンサンブル《1) 2 部以上からなる合唱曲・合奏曲; その合唱[合奏]者たち 2) 演奏者・歌手・踊り手の一団; その演奏や演技の統一性》」という意味の他、数理用語で「(系の) 集合、集団」という意味があり、Ensemblが何たるかを暗示している。また、単語中に EBI で維持されている核酸配列データベース EMBL のスペルが含まれており、そういったウイットの利いたプロジェクト名となっている。

1.2 なぜ Ensembl を使うのか。その特徴とは？

究極の物理地図であるゲノム配列 (golden path と呼ばれる) を元にして、ゲノム中の任意の場所の特徴（例えばその生物種由来の EST の有無、類縁モデル生物種で同定されている遺伝子配列との配列相同性のある領域や GC 含量）を任意の解像度で見ることができる。それを助けているのが、Ensembl の大きな特徴となっている秀逸なウェブインターフェースである。

さらに martview の使用により簡単に手軽にデータマイニングできる。

また、ウェブインターフェースではアクセスすることが困難なぐらい多数の検索結果を求められる人向けに、UNIX のコマンドラインからもアクセスできる。Ensembl のデータベースは mysql と呼ばれるフリーの RDBMS(Relational Database Management System) で実装されているので、SQL での問い合わせが可能となっている。

最大の特徴は、フリーである、ということである。すなわち、使われているソフトウェアがすべてフリーウェアであるばかりでなく、フリーで使える上にその開発自体がオープン（本来の意味でのオープンソース）なので、自由に開発に参加できる。

1.3 どんな実験のどの段階で使うのか

いろいろな段階で使える。とくに、実験を組む際の候補遺伝子の *in silico* スクリーニングに威力を発揮する。また注目している遺伝子のゲノム上の位置関係がさまざまな解像度で閲覧可能なので、簡単に近傍のゲノム配列や遺伝子コード領域を検索できる。

2. 実際に使ってみよう

・ Ensembl を使うのに必要な準備とは

1. ウェブブラウザを介した利用

まずは何かを調べようという生物学的な興味である。あとはインターネットに接続された、コンピュータ

とそれにインストールされたウェブブラウザがあれば十分である。

2. UNIX のコマンドラインからの利用

mysql がインストールされた、インターネットに接続されたコンピュータ環境が最低限必要である。自前に Ensembl の mysql データベースを持つ場合は必ずしもインターネット接続が必要とはならないが、解析を進めるには他のサイトのウェブページを参照する際に必須であろう。さらに UNIX の基本的な知識に加え、現時点での Ensembl のスキーマを必要なレベルで把握しておく必要がある。

・ 具体的な手順

3. 知っているると便利な使い方

DAS サーバを自前で上げて、必要な情報を Ensembl 中に投影する使い方（二階堂氏の発表参照）。

参考文献

- 1) M. Clamp, D. Andrews, D. Barker, P. Bevan, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, T. Down, R. Durbin, E. Eyra, J. Gilbert, M. Hammond¹, T. Hubbard, A. Kasprzyk, D. Keefe, H. Lehvaslaiho, V. Iyer, C. Melsopp, E. Mongin, R. Pettett, S. Potter, A. Rust, E. Schmidt, S. Searle, G. Slater, J. Smith, W. Spooner, A. Stabenau, J. Stalker, E. Stupka, A. Ureta-Vidal, I. Vastrik and E. Birney: Ensembl 2002: accommodating comparative genomics, *Nucleic Acids Research*, **31**, 38-42 (2003)
- 2) 初心者でもわかるバイオインフォマティクス入門 坊農秀雅 羊土社 2002