資料3－7

統合データベース支援:
DB 構築者の養成

森下　真一

東京大学

目的

- バイオDBを構築できる人材を育てる
  - 膨大なソフト外注費（150〜200万円/月）を回避
  - DBの保守・拡張が自前でできること
  - やむをえず外注する場合も、正確な仕様書を書ける力と、納入されたソフトの問題点を見抜く力を養う
- 必要スキルを1年間のカリキュラムで教え込むことができるか？
- 次の1年で独創的サーバーを構築できるか？
- 平成19年度：東大大学院生5名を対象

計画

DB 構築者を養成するために以下の3つの演習を実施する。

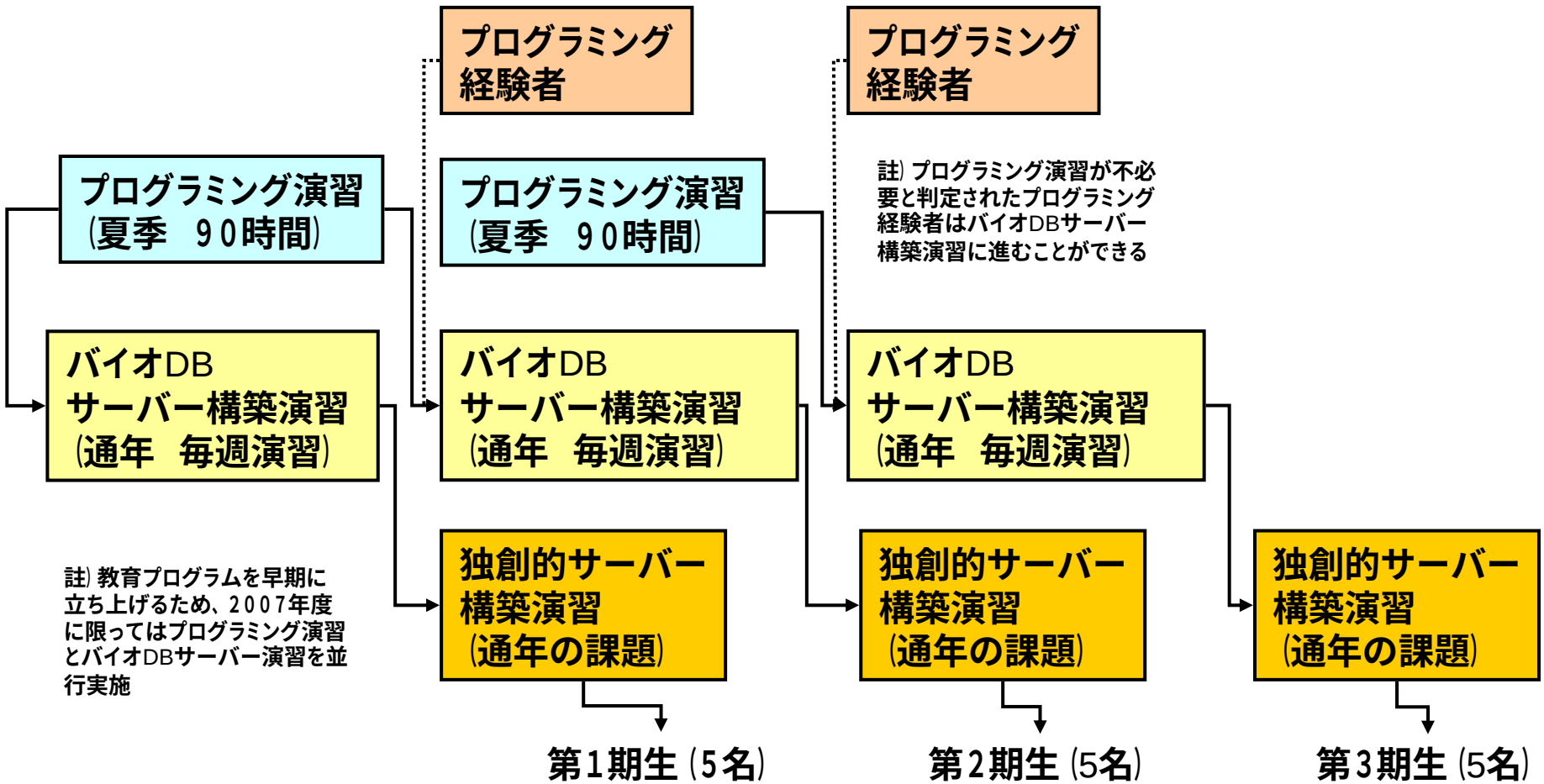
①バイオ DB サーバー構築演習
データベースサーバーのミラーサイトを構築する。OS, apache, MySQL 等の主要ソフトウエアのインストールおよびネットワークセキュリティに習熟することが目標である。参加者には各自にサーバー構築用ワークステーションを配布する。演習を完了するまでには、受講者の能力と受講可能時間に応じて最短で3ヶ月、最長で1年間の時間を予定している。

②プログラミング演習
Java および Perl プログラミングを演習した後に、アルゴリズムの知識を活かした配列処理やデータマイニングの実装を行う。上記①バイオ DB サーバー構築演習では実施がむずかしいプログラミング演習を行うことで、独自にソフトウエア構築ができる能力を身につけることをめざす。演習総時間は90時間で約2ヶ月間を予定している。

③ 独創的サーバー構築演習
大規模計算のためのクラスター利用技術を習得させ、他に類の無いバイオDBサーバーを設計、実装、公開することを目標とする。バイオDBサーバー構築演習およびプログラミング演習を修了した受講者に対して平成20年度より開講を予定しており、そのための計算機セットアップを平成19年度に準備する。

年次計画

平成19年度	20年度	21年度	22年度



演習用WS10台
（平成19年度予算申請）

註）　1期生と2期生が20年度には重なること（21年度は2，3期生）、WSが10台であること、演習スタッフ2名による徒弟制度であるため、各年5名の受け入れが限度である

教育内容と平成１９年度実績（5名 受講中）

- プログラミング演習
  - ✓ Java および Perl プログラミングを演習。
  - ✓ 独自のソフトウエア構築ができる基礎的能力を身につけることをめざす。
  - ✓ 演習総時間は９０時間で約2ヶ月間。
  - ✓ 本演習は東京大学理学部生物情報科学学部教育プログラムの生物情報演習。
  - ✓ 平成１９年度は３１名が受講（東大の学部生/大学院生）　１３名が単位
  - ✓ ＤＢ構築者養成では1名が単位取得（5名中2名は計算機科学科修了のため免除, 残り2名は一部の演習をこなすものの総合評定は不可）
- バイオ DB サーバー構築演習
  - ✓ 毎週2時間講義　宿題　次週にチェック
  - ✓ 演習を補足するノートを wiki として作成
  - ✓ 演習のステップ（１２月末の状況）
    (1) CentOS を自分のマシンにインストールする (2) ネットワークと接続する　(3) セキュリティアップデートを行う (4) web サーバーを立てる(ファイヤーウォールの設定を行う) (5) cgi を設置してみる (6) MySQL サーバーを立てる (7) 簡単なデータベース作成をする　（2名到達）　(8) Ensembl core をインストールしミラーを作成する　（2名到達）　(9) 複数種の実データをダウンロードして完全ミラーを作る　（1名到達）　(10) バックアップを作成して即時復旧できる体制を作る
- 独創的サーバー構築演習 (平成２０年度より開講)
  - ✓ 大規模計算のためのクラスター利用技術を習得させる。
  - ✓ 他に類の無いバイオDBサーバーを設計、実装、公開することが目標。
  - ✓ 演習に向けて１５台のＰＣ（DELL PowerEdge）を平成１９年度に購入しセットアップ
  - ✓ 5名のうち2名は独創的サーバーを構築へ

研究計画進行状況と変更

- 4月： バイオ DB サーバー構築演習開始
  受講者5名　2名の業務担当職員を募集
  その間は研究室内の人員によりカリキュラムを作成
- 7ー9月： プログラミング演習
- 8月： 業務担当職員1名採用　(東大　博士卒)
- 8月： 平成20年度の受講者数を増加する要請
  (15名程度)
- 10月： 平成20年度に必要な計算機資源の購入とセットアップ準備を申請
- 平成20年4月： 1名採用予定　(九大　博士卒)

演習ノート (はじめに)

統合データベース支援:DB構築者の養成におけるバイオDBサーバー構築演習

http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php?%C5%FD%B9%E7%A5%C7%A1%BC%A5%BF%A5%D9%A1%BC%A5%B9%BB%D9%B1%E7%A1%A7%A3%C4%A3%C2%B9%BD%C3%DB%BC%D4%A4%CE%CD%DC%C0%AE%A4%CB%A4%AA%A4%B1%A4%EB%A5%D0%A5%A4%A5%AA%A3%C4%A3%C2%A5%B5%A1%BC%A5%D0%A1%BC%B9%BD%C3%DB%B1%E9%BD%AC

[ホーム | 一覧 | 単語検索 | 最終更新 | ヘルプ] [新規 | 編集 | 添付] [no Trackback]

最新の20件

2007-12-25
- 演習ノート/20070913
- SeminarPowerpoints
- 演習ノート/20071011
- 統合データベース支援:DB構築者の養成におけるバイオDBサーバー構築演習

2007-12-13
- WebAppDevelopmentSchedule

2007-12-07
- 演習ノート/20070516その2

2007-11-08
- 演習ノート/20070705
- ソース課題1

2007-11-04
- 演習ノート/20070516その3
- Ensemblインストールのメモ

2007-11-02
- 演習ノート/20070418

2007-11-01
- GWTUserInterface

2007-10-30
- 演習ノート/20070802
- SeminarScribe
- 演習ノート/20070412

2007-10-25
- EnsemblMirror

2007-10-24
- 演習ノート/20070627

2007-10-23
- 演習ノート/20070425
- 演習ノート/20070530

2007-10-08
- 演習ノート/20071004

Top > 統合データベース支援:DB構築者の養成におけるバイオDBサーバー構築演習

Introduction

はじめに

生物学のような自然科学領域の学問における基本的な科学的態度とは、確かな観測とそこからの帰納的な推論であるといえる。特にDNA配列シークエンサーを中心とした近年のDNA観測技術の進歩は観測データの著しい増大を生み出し、生物の仕組みを理解するための帰納的推論の過程にはコンピューターを使った情報科学的・統計的なアプローチが欠かせなくなってきている。しかし、生物学的な観点からこのようなアプローチに向けてコンピューターを適切に扱える人材は不足しているのが現状であり、若手の育成が急務となっている。

また、近年はゲノム配列データだけではなく、cDNAを中心とした遺伝子やスプライシングのデータ、SNPや構造多型、比較ゲノムによる保存領域情報・シンテニー、SAGEやマイクロアレイによる発現情報やcis-element・発現局在情報、各種RNA、DNAのメチル化等のエピジェネティックな情報、クロマチン構造、発現や代謝のパスウェイ、オントロジー、たんぱく質相互作用・立体構造、遺伝学的マーカー、系統樹、疾患情報、家系情報・関連解析、解剖学的DB、フェノタイプ(画像・動画)、文献情報など数え切れない種類の生物学的・医学的データが産出され、それぞれデータベース化されている。しかし、これらの情報は適切に保守され、有機的に結びつけられるように相互互換性を保つ努力をし、優れた可視化を行わなければこれらのデータ上で分野横断的に生物学・医学的な知見を見出せるようにはならないと考えられる。

そこで、この演習では Ensembl データベースのミラーを作成し独自の拡張を加えることを通して、一般的なコンピュータースキルとデータベースのメンテナンス方法、プログラミングスキル、そして現在のバイオインフォマティクス業界で良く使われるデータの種類やフォーマット・生物学的意味を理解し将来に渡って役立つ基礎力を付けることを目指す。

基本的な演習の方針

ネットワークやLinux、プログラミングに関しては既に詳しく書かれた書籍やwebページがたくさんある。だから、毎週の講師が step by step で手法を解説したスライドを作るのでは無駄である。毎週の講師が頑張るのは「概要」を知ってもらう部分までで、それ以降の詳しい話は自力で学んでもらう。詳しいやり方を解説することより、詳しいやり方を調べるための方法を主に解説することを目指す。もちろん落とし穴などがある場合には講師がそれを解説する。基本的には、関連するman ページの見方や info の使い方、便利な web ページの紹介(@ITとか)、Google に入力する検索ワード、分かりやすい書籍、など自力でなんとかできる人を育てることを考えて演習を組み立てる。

自分で調べることを基本にはするが、「概要」を分かりやすく伝えることはサボらない。テクノロジーやソフトウェアは、「概要」が分かっていないと調べるためのキーワードすらわからない。調べるためのキーワードが閃く程度に分かってもらうのは毎週の講師の役目である。課題や内容については、最終目的に適した範囲で毎回や講師相当がご確認ください。

http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php

演習ノート (毎週の講義内容と宿題集)

SeminarScribe

http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php?SeminarScribe

[ホーム | 一覧 | 単語検索 | 最終更新 | ヘルプ] [新規 | 編集 | 添付] [no Trackback]

最新の20件

2007-12-17
- 演習ノート/20071011
- 演習ノート/20070913

2007-12-14
- 統合データベース支援: DB構築者の養成におけるバイオDBサーバー構築演習

2007-12-13
- WebAppDevelopmentSchedule

2007-12-07
- 演習ノート/20070516その2

2007-11-08
- 演習ノート/20070705
- ソース課題1

2007-11-04
- 演習ノート/20070516その3
- Ensemblインストールのメモ

2007-11-02
- 演習ノート/20070418

2007-11-01
- GWTUserInterface

2007-10-30
- 演習ノート/20070802
- SeminarScribe
- 演習ノート/20070412

2007-10-25
- EnsemblMirror
- SeminarPowerpoints

2007-10-24
- 演習ノート/20070627

2007-10-23
- 演習ノート/20070425
- 演習ノート/20070530

2007-10-08
- 演習ノート/20071004

Top > SeminarScribe

- 演習ノート
  - ノート一覧

演習ノート

演習ノートはこのページからリンクしてください。下記一覧に自分の担当の演習ノートへのリンクが無い場合には、同じような書式で追加してください。

ノート一覧

- 4/6 イントロダクション
- 4/9 最初の準備
- 4/12 CentOSのインストールに向けて
- 4/18 Linux とネットワークの基礎
- 4/25 VMware Server 上で CentOS をインストールする
- 5/9 CentOS 上でweb サーバーを設置する
- 5/16 web サーバーに動的なコンテンツを追加する
- 5/16 その2 pukiwikiの設置
- 5/16 その3 シェルスクリプト
- 5/23 セキュリティと定期アップデート
- 5/25 Pukiwiki による情報共有
- 5/30 RDBMS を使ってみる
- 6/6-13 Perl 演習1-2
- 6/27 Perl 演習3
- 7/05 PerlでCGI演習
- 7/19 tarballからソフトのインストールをする
- 8/2 CPANを使いこなす
- 9/13 Ensemble core
- 10/4 ネットワークトラブルへの対処
- 10/11いろいろ

RSS

Last-modified: 2007-10-30 (火) 13:22:37 (55d)

Link: 演習ノート/20071011(8d) 演習ノート/20070913(8d) 演習ノート/20070516その2(17d) 演習ノート/20070705(46d) 演習ノート/20070516その3(50d) 演習ノート/20070418(52d) 演習ノート/20070802(55d) 演習ノート/20070412(55d) EnsemblMirror(60d) 演習ノート/20070627(62d) 演習ノート/20070425(62d) 演習ノート/20070530(63d) 演習ノート/20071004(78d) 演習ノート/20070525(80d) 演習ノート/20070516(81d) 演習ノート/20070719(90d) 演習ノート/20070523(103d) 演習ノート/20070509(104d) 演習ノート/20070606(104d) 演習ノート/20070409(157d) 演習ノート/20070406(167d)

http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php

演習ノート (毎週の宿題のヒント)

演習ノート/20070913

http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php?%B1%E9%BD%AC%A5%CE%A1%BC%A5%C8%2F20070913

[ホーム | 一覧 | 単語検索 | 最終更新 | ヘルプ] [新規 | 編集 | 添付] [no Trackback]

最新の20件

2007-12-17
- 演習ノート/20071011
- 演習ノート/20070913

2007-12-14
- 統合データベース支援: DB構築者の養成におけるバイオDBサーバー構築演習

2007-12-13
- WebAppDevelopmentSchedule

2007-12-07
- 演習ノート/20070516その2

2007-11-08
- 演習ノート/20070705
- ソース課題1

2007-11-04
- 演習ノート/20070516その3
- Ensemblインストールのメモ

2007-11-02
- 演習ノート/20070418

2007-11-01
- GWTUserInterface

2007-10-30
- 演習ノート/20070802
- SeminarScribe
- 演習ノート/20070412

2007-10-25
- EnsemblMirror
- SeminarPowerpoints

2007-10-24
- 演習ノート/20070627

2007-10-23
- 演習ノート/20070425
- 演習ノート/20070530

2007-10-08
- 演習ノート/20071004

Top > 演習ノート > 20070913

- Ensemblのインストール方法の概容を理解する
  - 文献
  - hxサーバーからEnsemlデータをダウンロードする
- MySQLにぶちこむ
  - 解凍
  - 壊れてたら
- サーバーが動かないとき
- PuTTYで日本語を表示する方法
- CVSについて
- DASについて
- Apache
  - Virtual Host設定方法
  - debugを表示させる

Ensemblのインストール方法の概容を理解する

文献

ensembleのサイトにインストールの方法が載っています
Instructions on how to install a local Ensembl website
またはトップ画面 から

> Help & Documentation -> Software ->Ensembl Website->Installing Ensembl

にあります
Summary of InstructionsのBuild/Install まで読めば十分です

hxサーバーからEnsemlデータをダウンロードする

ensembl archiveのありかは

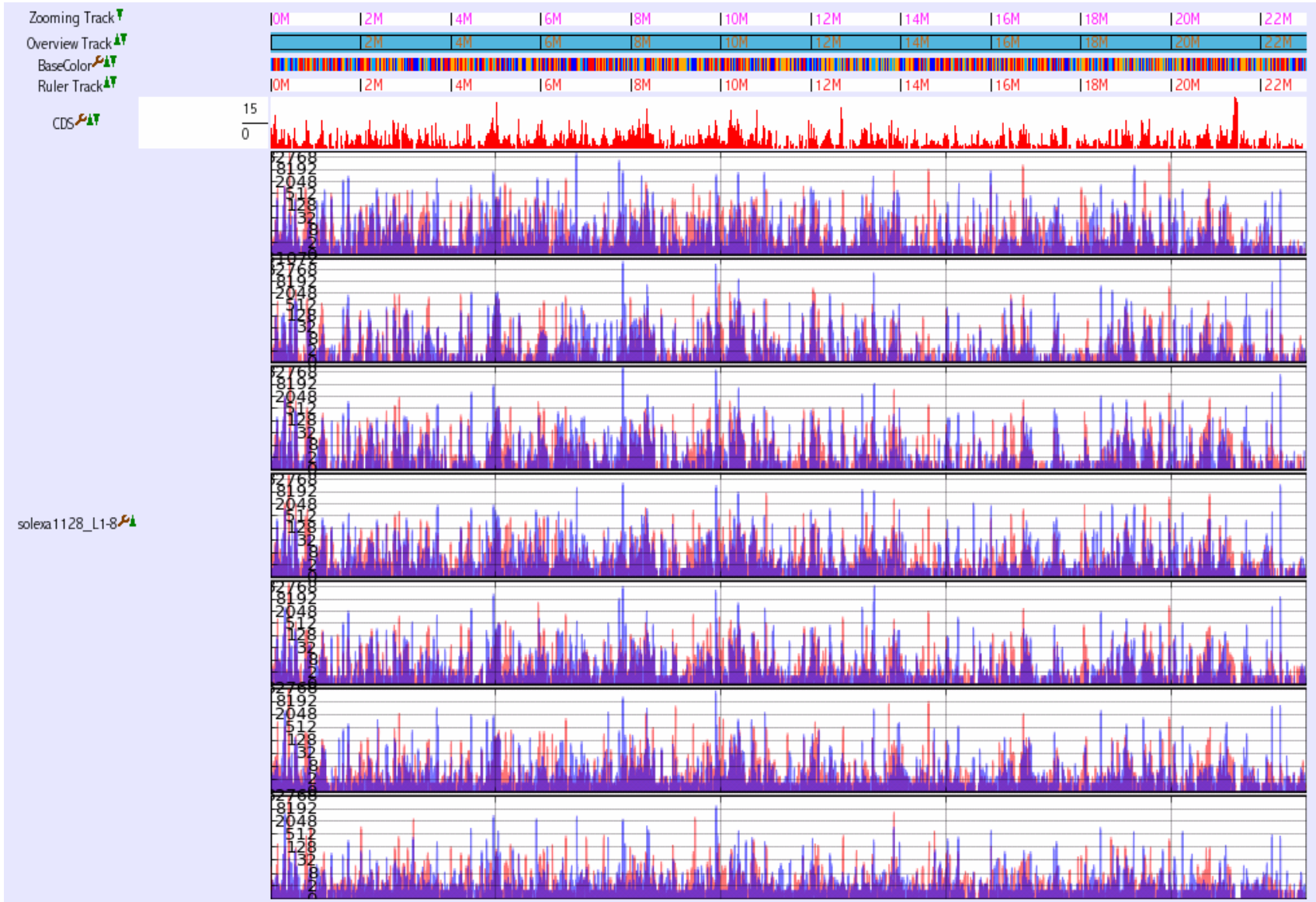
> mkasa@hx25:/home/mkasa/ensembl/archive

カレントディレクトリにダウンロード

http://mlab.cb.k.u-tokyo.ac.jp/~mkasa/ensemblmirror/index.php

独創的サーバー構築演習　例

- 進捗の早い2名が開始
  - ✓ 超高速シークエンサー Solexa の base call
  - ✓ 並列 BLAST / BLAT を使った短いタグ(25-36 nt)のアラインメント
  - ✓ 全長 cDNA 推定アルゴリズムの工夫
  - ✓ 5'end タグを使った遺伝子発現量解析
  - ✓ 解析パイプラインの研究開発
  - ✓ 表示ルールの研究開発
  - ✓ 従来の遺伝子情報に比べて数百倍のデータ量を適切な応答時間で処理する工夫



Solexaタグの表示例

従来の遺伝子情報に比べて数百倍のデータ量を適切な応答時間で処理する工夫