資料　２

統合データベース支援：
バイオDBサーバー構築演習

森下　真一
中谷　洋一郎

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
- 次の1年で独創的サーバーを構築

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。

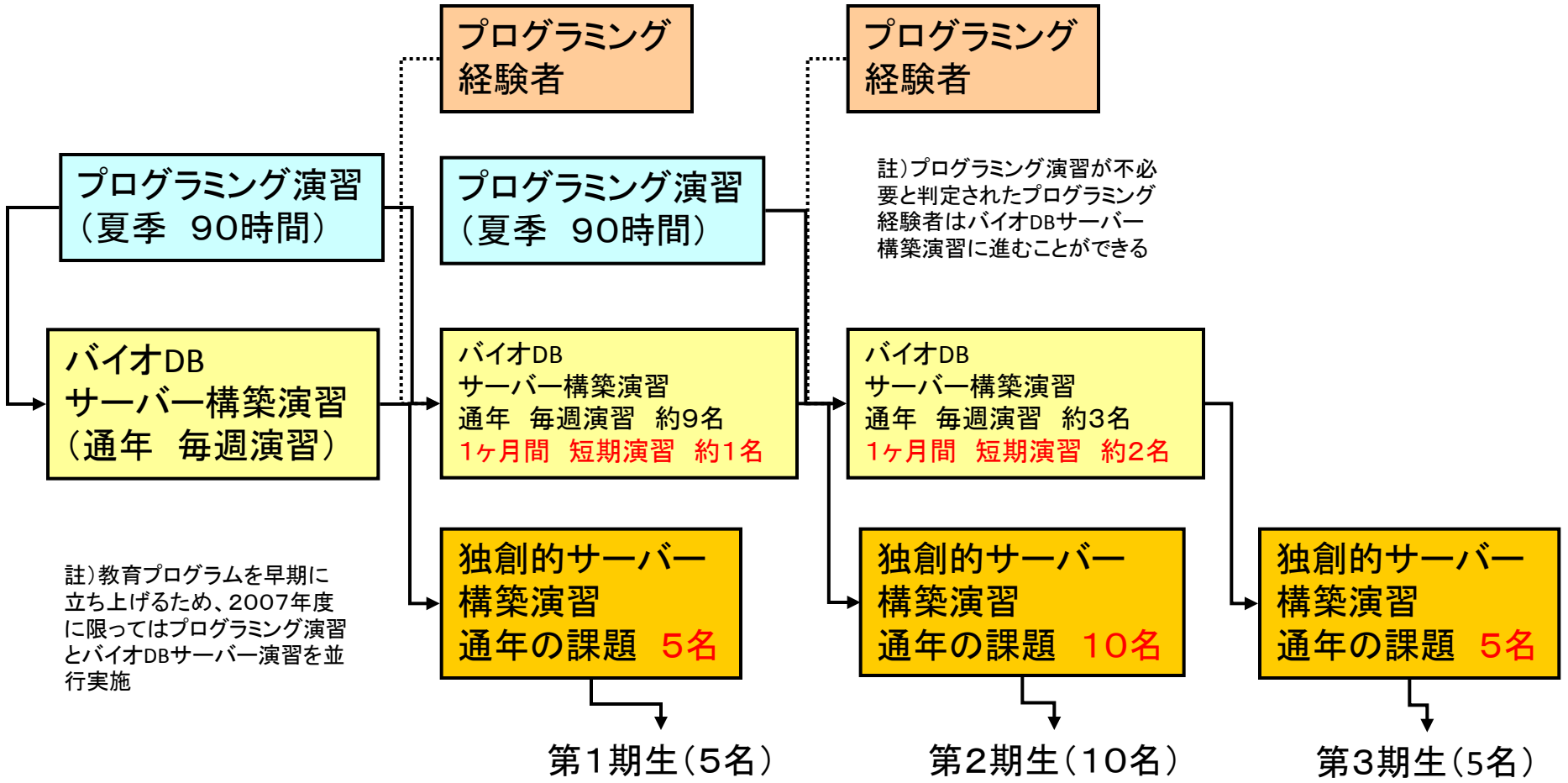
①バイオ DB サーバー構築演習
データベースサーバーのミラーサイトを構築する。OS, apache, MySQL 等の主要ソフトウエアのインストールおよびネットワークセキュリティに習熟することが目標である。参加者には各自にサーバー構築用ワークステーションを配布する。演習を完了するまでには、受講者の能力と受講可能時間に応じて最短で3ヶ月、最長で1年間の時間を予定している。

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年度に準備した。

年次計画

平成19年度	20年度	21年度	22年度



演習用WS15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15台であること、演習スタッフ1．5名による徒弟制度であるため、各年15名の受け入れが限度である

受講者

- 平成１９年度
  - 東大情報生命科学専攻から5名
- 平成２０年度
  - DBCLSから2名
  - 東大情報生命科学専攻から7名
  - 自治医大から1名（9月のみ受講）
  - 合計１０名

DBサーバー構築演習の目標設定

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を作成する
8: Ensembl core をインストールしミラーを作成する
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
10: バックアップを作成して即時復旧できる体制を作る

20年度バイオDBサーバー構築演習の概要

- OS (Linux) のインストール
- ネットワーク・ファイアーウォールの設定
- Web サーバーの設置・設定 (apache)
- RDBMSの設置・設定 (MySQL)
- Perl モジュールの設置・設定
- Ensembl の設置・設定
- Perl, Javaプログラミング
- CGIからのデータベース検索
- メンテナンス全般
  - 障害対応
  - ソフトウェアの Security fix やバージョンアップ等

演習日程	テーマ
4/03	イントロダクション、最初の準備
4/10	CentOSのインストールに向けて
4/17	Linux とネットワークの基礎
4/24	VMware Server 上で CentOS をインストール
5/15	CentOS 上でweb サーバーを設置する
5/22	web サーバーに動的なコンテンツを追加する
5/29	セキュリティと定期アップデート
6/05	Perl演習
6/12	CPANを使いこなす、BioPerlのインストール
6/19	ネットワークの設定、SSHによる外部からの接続
6/26	シェルスクリプト、Pukiwiki による情報共有
7/03	RDBMS を使ってみる
7/10	Perlからデータベースを扱う
7/17	CGIでデータベースを検索する
9/11	Ensemblデータ・モジュール等のインストール
9/18	Ensembleインストールの続き
9/25	Ensemblインストールの続き
10/09	UTGB toolkitの紹介
10/23	Javaからのデータベース検索、UTGB Shellの紹介
11/20	データベース検索、SQL
12/18	UTGBのインストール、DB設計
1/15	UTGB Shellによるウェブサイト作成

OSのインストール

- 講義日程：4/3, 10, 17, 24
- システム・ネットワーク・ウェブ・データベース等に関する基礎的な用語の解説。
- 各自のサーバーにLinuxをインストール。
  - CDイメージをダウンロードしCentOSをインストールする。
- VMwareのインストール。
  - VMwareをインストールすることで、1台の物理サーバーを複数の仮想マシンに分割する。
  - VMware上の仮想マシンにCentoOSをインストール。

ネットワーク接続の設定、Web サーバーの設置、CGIの作成

- 講義日程：5/15, 22, 6/19
- ネットワークの設定
- ウェブサーバーの設置。
  - Apacheのインストール、設定ファイルの編集。
- Firewallの設定。
- ウェブサイトに動的なコンテンツを作成。
  - プログラムによって動的に文書を生成し、ブラウザーに表示させる。
  - CGIを動作させるためのApacheの設定。
  - 現在時刻を表示するCGIプログラムを作成。

セキュリティと定期的な自動更新設定

- 講義日程：5/29
- サーバーをネットワークに接続する場合、セキュリティ上の問題が起きないよう対策をとることが重要。
- 脆弱性の例を説明。
  - Buffer overflow, SQL injection, Cross site scripting, Brute force attack, DNS spoofing, Man-in-the-middle attack, Social hacking, Memory snooping attack, RLO偽装, URL推測
- ソフトウェアの自動更新設定。
  - Yum-update, yum-cronによって定期的にセキュリティアップデートを行う。
- 公開鍵認証方式によるログイン。
  - パスワードを入力しない安全な方式で外部からssh接続を行う。

Pukiwikiの設置、データベースの設置

- 講義日程：6/26, 7/3
- Pukiwikiの設置。
  - ウェブ上で情報の共有と整理を多人数で行える。
  - Pukiwikiをダウンロードし、サーバーにインストールする。
  - Pukiwikiの基本操作、文法の解説。
- データベースの設置。
  - MySQLのインストール。
  - MySQLの基本的なコマンドの解説。
  - MySQLを使用してデータベースの検索速度を比較。

Perl演習,PerlによるCGI作成

- 講義日程：6/5, 7/10, 17
- なぜPerlを学ぶのか？
  - “バイオインフォマティクスの分野で、最も広く使われているスクリプト言語。”
  - “ほとんどのプログラムはPerl で書かれているので、多くのバイオインフォマティクス研究者がPerlでプログラミングを学んでいる。”
- Perlのインストール。
- 基本的なPerl文法の解説。
- ゲノム配列データをダウンロードし、Perlを使用して簡単なデータ処理を行う。
- Perlスクリプトからデータベースにアクセスしデータ処理を行う。
- Perlを使用してCGIを作成し、データベースにアクセスするウェブサイトを構築する。

ソフトウェア・モジュールのインストール

- 講義日程: 6/12
- 他の研究者によって開発されたソフトウェア・ライブラリー・モジュールを使用することで、解析プログラム・解析パイプラインをすばやく簡単に作成することができる。
- ソースコードからのインストール。
  - configure, makeの解説。
  - Rubyをソースコードからインストールする。
- CPANからPerlモジュールをインストールする。
  - AppConfig, DBI, DBD::SQLite, File::HomeDir, YAML, Spreadsheet::ParseExcel, Spreadsheet::WriteExcel, Cwd, SVG, PostScript::Simple, File::HomeDir, HTML::Parser, XML::Parser, IO::Zlib, Term::ReadLine, Template, Digest::SHA::PurePerl, Bundle::BioPerl

UTGB Shellによる ゲノムブラウザー開発

- 講義日程: 10/9, 23, 11/20, 12/18, 1/15
- UTGB Shellを用いて新しいタイプのゲノムデータをトラックに表示する技術を習得することが目標。
- UTGBの紹介、UTGB Shellのインストール
- Java プログラミング、Javaでデータベース検索プログラムを作成
- SQL、データベース設計
- UTGB Shellを用いてウェブサイトにデータを表示する

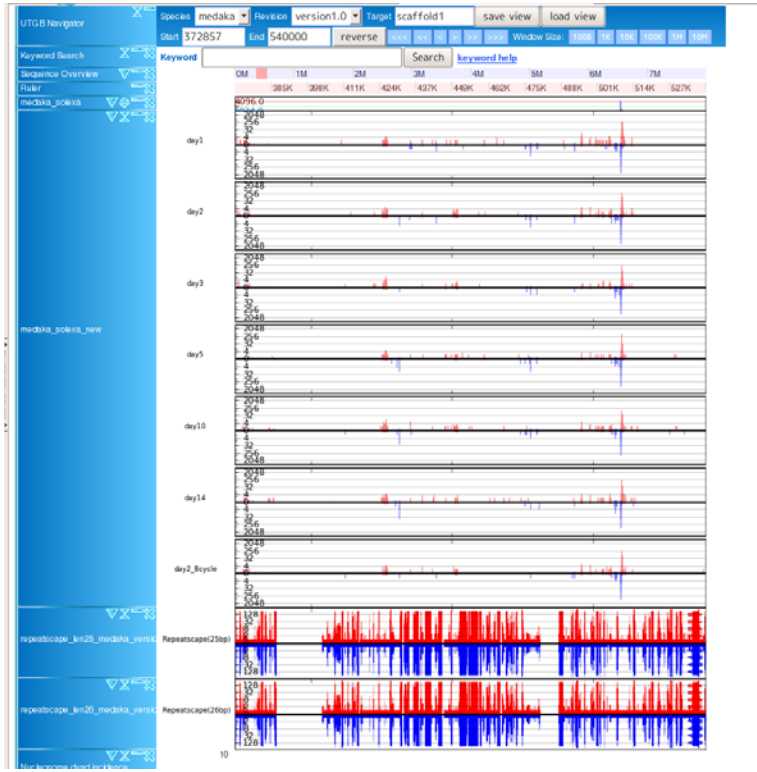
20年度受講者の進捗
（受講者数:4～7月8名、9月8名）

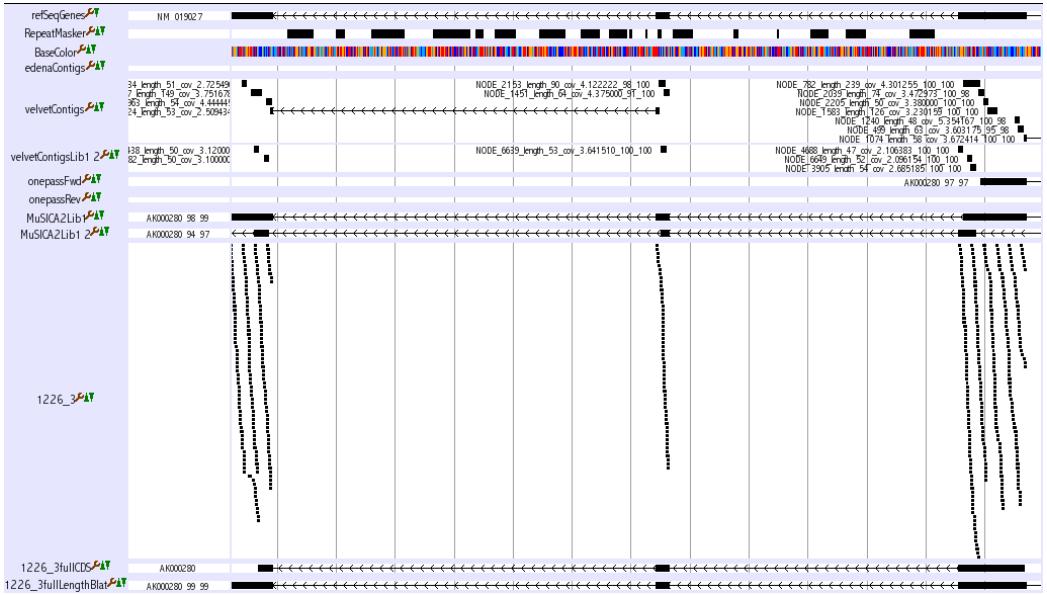
1. CentOS を自分のマシンにインストールする
2. ネットワークと接続する
3. セキュリティアップデートを行う
4. web サーバーを立てる(ファイヤーウォールの設定を行う)
5. CGIを設置してみる
6. MySQL サーバーを立てる
7. 簡単なデータベース作成をする
8. Ensembl core をインストールしミラーを作成する
9. 複数種の実データをダウンロードして完全ミラーを作る
10. バックアップを作成して即時復旧できる体制を作る

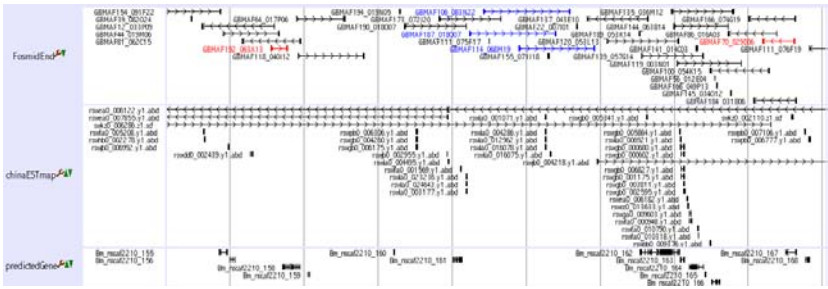
受講者中2名はEnsemblを用いたヒトゲノムデータの表示まで終了。残り6名は作業途中。
10月以降は独自データをUTGBを用いてブラウザー表示するため講習を行っている。

独創的サーバー構築演習

- 受講者が研究で使用する新規データをゲノムブラウザーに表示する。
  - 発現量データを表示するトラックの開発。
  - 配列特異性を視覚化するトラックの開発。
    - “RepeatScape”として公開。
  - Fosmid-end解析, 完全長cDNAアセンブリーの解析をブラウザーに表示。
- データ解析・論文作成に活用されている。

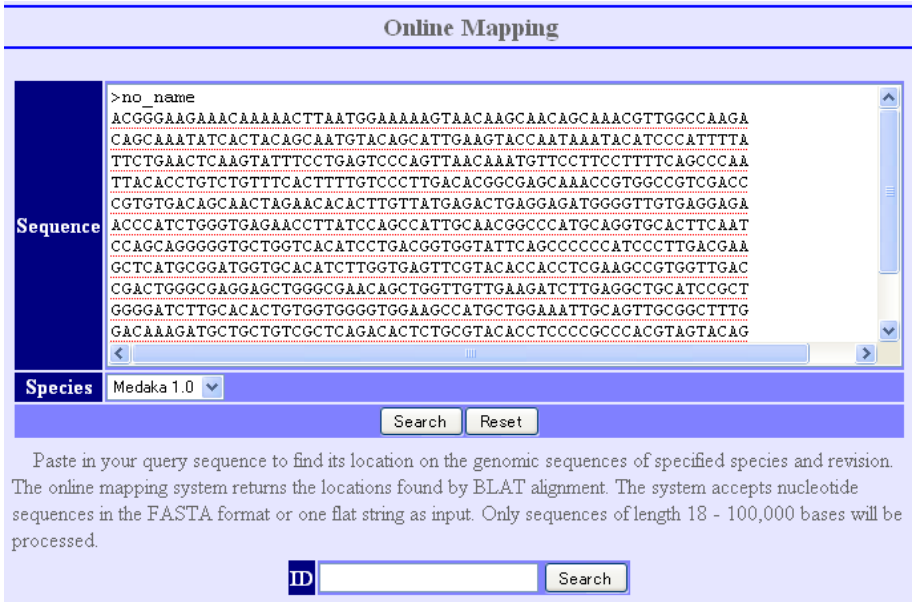


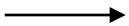




UTGB Medaka Online Mapping

- クラスターでアラインメントの計算
- ウェブブラウザーでマッピング結果を表示



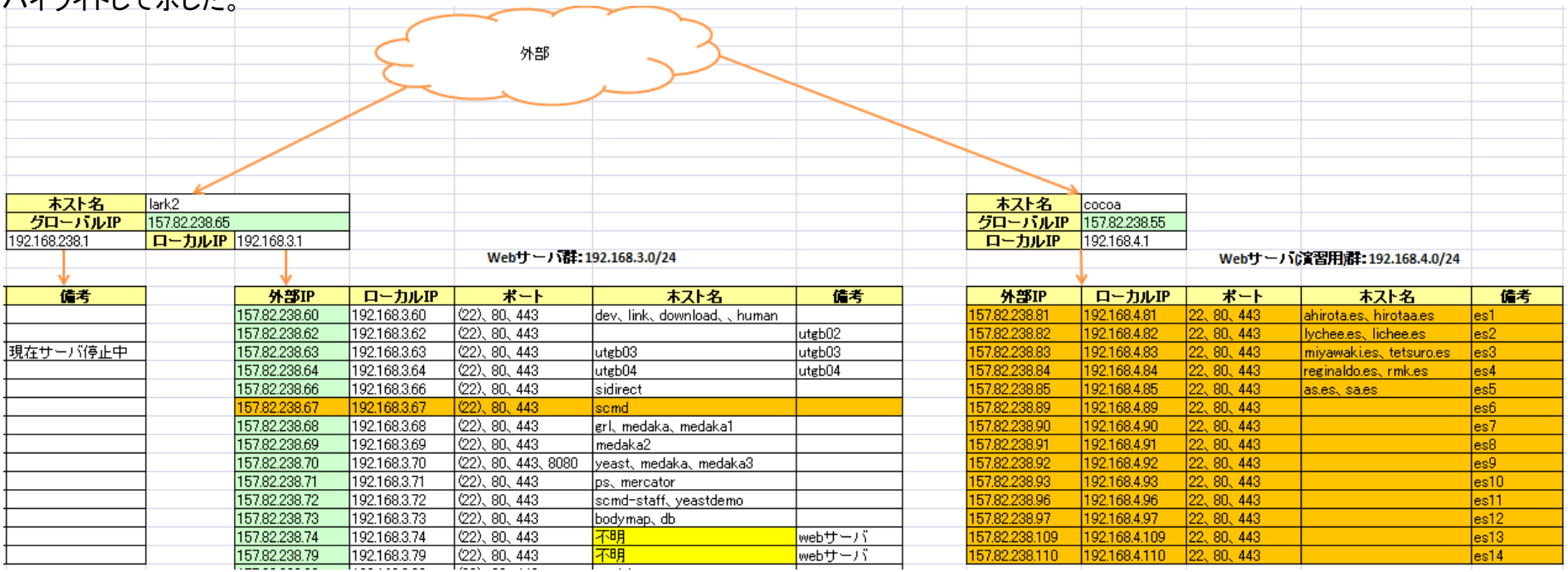


サーバー使用者氏名とネットワーク図

es1.gi.k.u-tokyo.ac.jp上野敏秀（20年度受講者）
es2.gi.k.u-tokyo.ac.jp 李鑫（19年度受講者）
es3.gi.k.u-tokyo.ac.jp 宮脇徹郎（19年度受講者）
es4.gi.k.u-tokyo.ac.jp レジナルド クロス（19年度受講者）
es5.gi.k.u-tokyo.ac.jp 佐々木惇（19年度受講者）

es6.gi.k.u-tokyo.ac.jp 仲里猛留（20年度受講者）
es7.gi.k.u-tokyo.ac.jp 小野浩雅（20年度受講者）
es8.gi.k.u-tokyo.ac.jp 劉晅晅（20年度受講者）
es9.gi.k.u-tokyo.ac.jp 宗永雅樹（20年度受講者）
es10.gi.k.u-tokyo.ac.jp 村中真人（20年度受講者）
es11.gi.k.u-tokyo.ac.jp 呉紅艶（20年度受講者）
es12.gi.k.u-tokyo.ac.jp 近藤修平（20年度受講者）
es13.gi.k.u-tokyo.ac.jp 白井和英（20年度受講者）
es14.gi.k.u-tokyo.ac.jp 中谷洋一郎（講師）
scmd.gi.k.u-tokyo.ac.jp 中谷洋一郎（講師）、斉藤太郎（講師補助）

下のネットワーク図では、15台のサーバーをオレンジでハイライトして示した。



備考	外部IP	ローカルIP	ポート	ホスト名	備考
	157.82.238.60	192.168.3.60	(22)、80、443	dev、link、download、、human	
	157.82.238.62	192.168.3.62	(22)、80、443		utgb02
現在サーバ停止中	157.82.238.63	192.168.3.63	(22)、80、443	utgb03	utgb03
	157.82.238.64	192.168.3.64	(22)、80、443	utgb04	utgb04
	157.82.238.66	192.168.3.66	(22)、80、443	sidirect	
	157.82.238.67	192.168.3.67	(22)、80、443	scmd	
	157.82.238.68	192.168.3.68	(22)、80、443	grl、medaka、medaka1	
	157.82.238.69	192.168.3.69	(22)、80、443	medaka2	
	157.82.238.70	192.168.3.70	(22)、80、443、8080	yeast、medaka、medaka3	
	157.82.238.71	192.168.3.71	(22)、80、443	ps、mercator	
	157.82.238.72	192.168.3.72	(22)、80、443	scmd-staff、yeastdemo	
	157.82.238.73	192.168.3.73	(22)、80、443	bodymap、db	
	157.82.238.74	192.168.3.74	(22)、80、443	不明	webサーバ
	157.82.238.79	192.168.3.79	(22)、80、443	不明	webサーバ

外部IP	ローカルIP	ポート	ホスト名	備考
157.82.238.81	192.168.4.81	22、80、443	ahirota.es、hirotaa.es	es1
157.82.238.82	192.168.4.82	22、80、443	lychee.es、lichee.es	es2
157.82.238.83	192.168.4.83	22、80、443	miyawaki.es、tetsuro.es	es3
157.82.238.84	192.168.4.84	22、80、443	reginaldo.es、rmk.es	es4
157.82.238.85	192.168.4.85	22、80、443	as.es、sa.es	es5
157.82.238.89	192.168.4.89	22、80、443		es6
157.82.238.90	192.168.4.90	22、80、443		es7
157.82.238.91	192.168.4.91	22、80、443		es8
157.82.238.92	192.168.4.92	22、80、443		es9
157.82.238.93	192.168.4.93	22、80、443		es10
157.82.238.96	192.168.4.96	22、80、443		es11
157.82.238.97	192.168.4.97	22、80、443		es12
157.82.238.109	192.168.4.109	22、80、443		es13
157.82.238.110	192.168.4.110	22、80、443		es14