

Nicotiana tapacu

Zea

Salmo salar

Lotta giuridica

Autore: Giacomo Saccoccia

Rattus norvegicus @東大農 Sep. 1, 2

バイオインフォマティクス人材育成カリキュラム(次世代シークエンサ) | 速習コース

1-1: OS, ハード構成

1-2: ネットワーク基礎

中村 保一

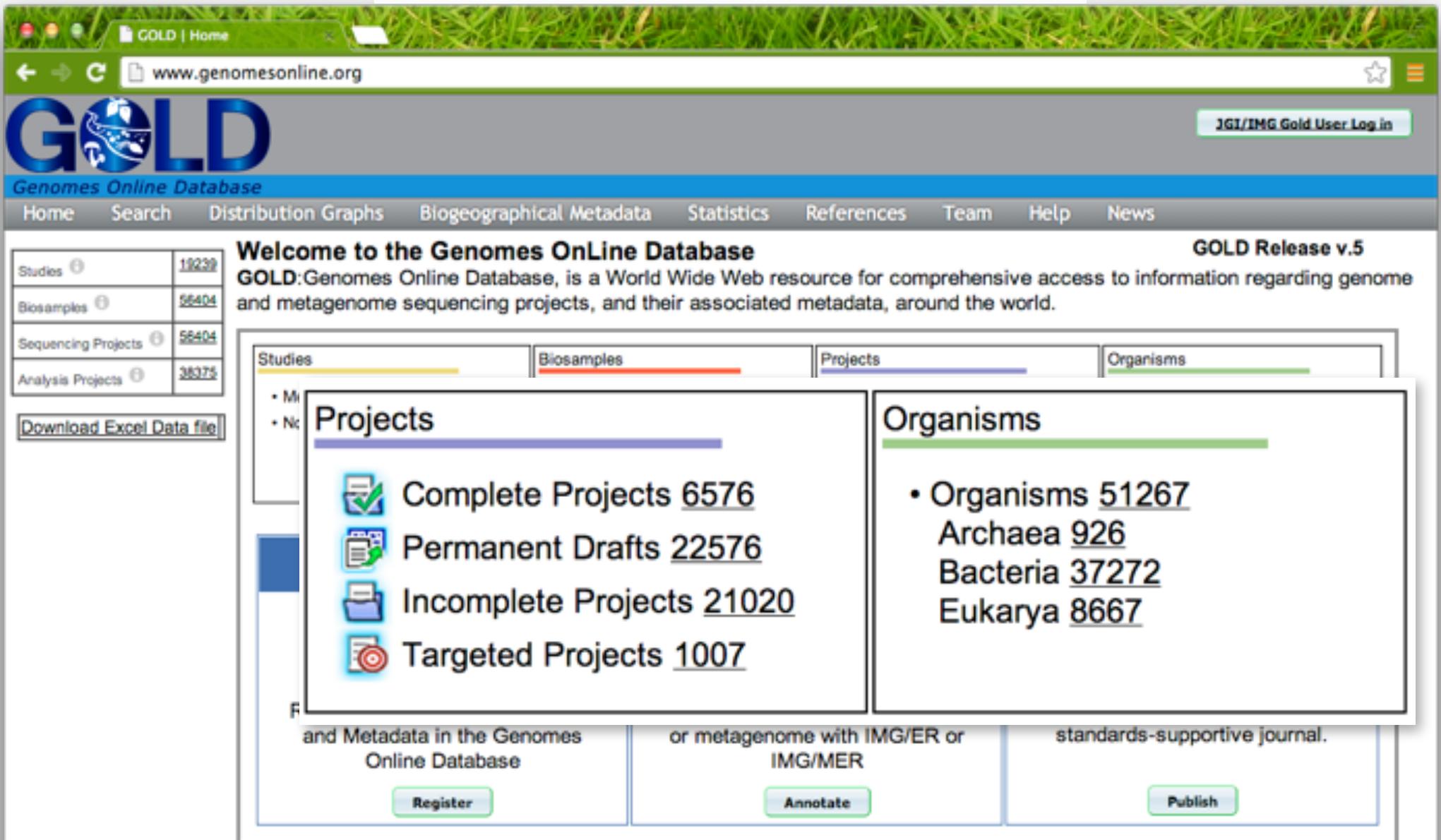
NAKAMURA Yasukazu, Professor

国立遺伝学研究所 大量遺伝情報研究室

<http://charles.genes.nig.ac.jp> yn@nig.ac.jp

今や膨大なゲノム情報が入手可能な訳で

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



GOLD | Home www.genomesonline.org JGI/IMG Gold User Log in

GOLD

Genomes Online Database

Home Search Distribution Graphs Biogeographical Metadata Statistics References Team Help News

Studies 19239
Biosamples 56404
Sequencing Projects 56404
Analysis Projects 38375

Welcome to the Genomes OnLine Database GOLD Release v.5

GOLD: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

Studies Biosamples Projects Organisms

• Mi
• Nc

Projects

- Complete Projects 6576
- Permanent Drafts 22576
- Incomplete Projects 21020
- Targeted Projects 1007

Organisms

- Organisms 51267
- Archaea 926
- Bacteria 37272
- Eukarya 8667

and Metadata in the Genomes Online Database or metagenome with IMG/ER or IMG/MER standards-supportive journal.

Register Annotate Publish

あなたに必要なスキルはなにか？

[要]

- ・ユーザとして UNIX が使いこなせる
- ・公開されているデータベースを使いこなせる
- ・公開されている解析ツールを使いこなせる
- ・簡単な言語でスクリプトが書ける

[不要]

- ・ハイレベルなプログラミング能力
→ゲノムアセンブラが作れなくても、大丈夫

元祖使い倒し系バイオインフォマティスト



<http://www.amazon.co.jp/dp/4758108110>

編集代表



<http://www.amazon.co.jp/dp/4897068746>

企画・編集



<http://www.amazon.co.jp/dp/4895924262>

翻訳分担（第9章・遺伝子予測と遺伝子調節）

植物関係のゲノム解析+DB屋でした



The Arabidopsis Genome Initiative (2000)
Analysis of the genome sequence of the flowering
plant *Arabidopsis thaliana*.
Nature, 408, 796-815.

シロイヌナズナゲノムプロジェクトで
全体の 1/4 (27 Mb, 6200 genes) の領域解析

A screenshot of the Cyanobase database interface. At the top, there is a navigation bar with links such as 'Home', 'About', 'Help', 'Contact', 'Logout', and 'User login'. Below the navigation bar, there is a search bar with placeholder text 'Search' and a dropdown menu for 'Search type'. The main content area features a table with several rows of data, each containing a blue circular icon, a cyanobacterium name like 'Synechocystis sp. PCC 6803', and other details. At the bottom of the page, there is a footer with links to 'About', 'Help', 'Contact', 'Logout', and 'User login'.

<http://genome.microbedb.jp/cyanobase/>

光合成細菌のゲノム解析+データベース

現在はDDBJのDB構築担当教員です


DNA Data Bank of Japan

English

DDBJ の紹介 利用の手引き レポート・統計 Q and A お問い合わせ

 Web Magazine
 RSSを購読する
 DDBJ Twitter

DDBJ Service

 登録 Data Submission

 検索・解析 Search / Analysis

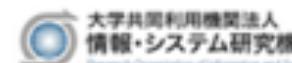
 スパコン Super Computer

 アーカイブ ftp://ddbj.nig.ac.jp

Hot Topics 

- 2013.06.26 WABI (Web API for Biology) の再開
- 2013.06.11 DDBJ リリース 93.0, DAD リリース 63.0 完成
- 2013.05.15 「第27回 DDBJing 講習会 in 三島(2013.7.4開催)」のご案内 (参加申込み受付中)

 大学共同利用機関法人 情報・システム研究機構
国立遺伝学研究所

 大学共同利用機関法人
情報・システム研究機構
Research Organization of Information and Systems

Maintenance 

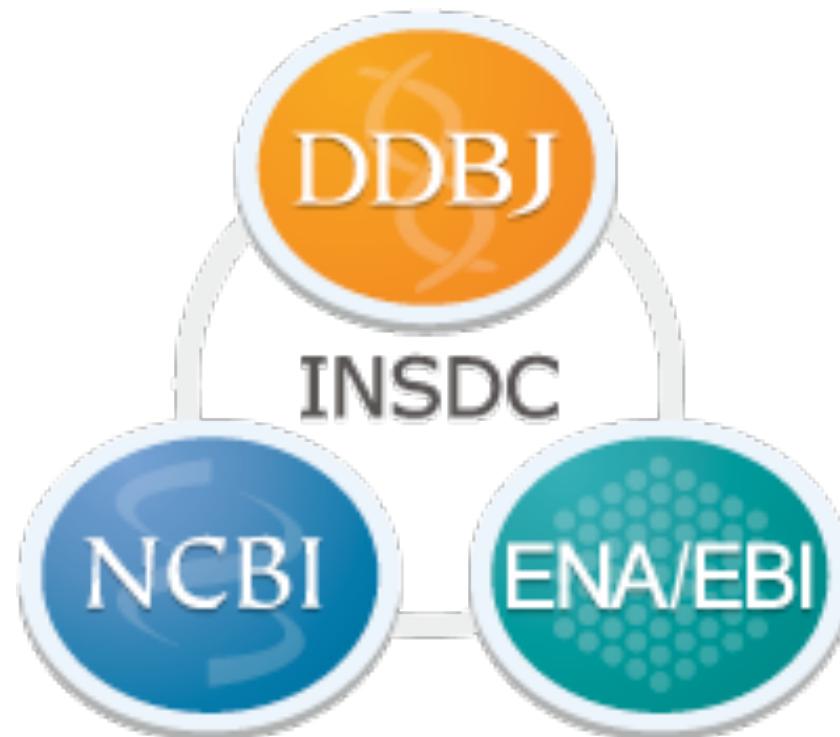
Information

塩基配列データバンクとはこのような事業

- 全世界で解読された塩基配列情報を
 - 査定して受入れ
 - データベースに蓄積し
 - 公開して共有する



- 米国: GenBank (NCBI)
- 欧州: ENA (EBI)
- 日本: DDBJ



ハードウェア

a) 計算機の歴史

b) 現代のハードウェア構成

ソフトウェア

a) オペレーティング・システム (OS)

b) 言語・プログラム・ライブラリ

ネットワーク

よく使うインターネットプロトコル



"Soroban" by Kowloonese - 投稿者自身による作品. Licensed under Creative Commons Attribution-Share Alike 3.0 via ウィキメディア・コモンズ - <http://commons.wikimedia.org/wiki/File:Soroban.jpg#mediaviewer/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Soroban.jpg>

- (1) 数値の保持
- (2) 加算・減算の機能
- (3) 数値の表示

- 1623年 シッカートの計算機（現存せず）
- 1642年 パスカルの計算機（歯車式／加算減算）
- 1672年 ライプニッツの四則計算機
- 1835年 バベッジの解析機関（実現せず）
 - 「ディファレンス・エンジン」
 - 穿孔カードでのプログラム、というアイデア

クルタ計算機 1948年 …ほしい…

CURTA calculator | eBay

www.ebay.com/sch/i.html?_odkw=CURTA+calculator&_osacat=0&_from=R40%7CR40&_from=R40&_trksid=p2045573.m570.l1313.TR0.T...

Hi! Sign in or register | Daily Deals | Sell | Customer Support My eBay My eBay Notification Cart

ebay Shop by category All Categories Search Advanced

Related: mechanical calculator slide rule adding machine curta type ii antique calculator hp 35 calculator pawn stars ... Include description

Categories
Collectibles (31)
Science & Medicine (1930-Now) (25)
Advertising (1)
[See all categories](#)

Condition see all
 New (7)
 Used (25)
 Not Specified (2)

Price
 Under ¥300
 ¥300 - ¥900
 ¥900 - ¥1300
 Over ¥1300
JPY to JPY >>

Format see all
 All Listings (34)
 Auction (1)
 Buy It Now (33)

Delivery Options see all
 Free shipping

CURTA calculator All Listings Auction Buy It Now Sort: Best Match View: grid

34 results for CURTA calculator + Follow this search


Curta calculator
JPY 103,517 6d 2h left (Friday, 8PM)
0 bids From United Kingdom


RARE - CURTA Calculator TYPE I #15255 - GOOD STATE - 1951
JPY 103,781 From France
Buy It Now


THE CURTA CALCULATOR POSTER

- 1946年 ENIAC
 - Electronic Numerical Integrator and Calculator
 - 「世界最初のコンピュータ」と目される
 - 約 18,000 本の真空管からなる
 - 10進法を採用 ⇒ 齒車式計算機の電子版
 - プログラム内蔵式 ではない



<http://commons.wikimedia.org/wiki/File:Eniac.jpg>

プログラム内蔵電子式計算機：ノイマン型

- 1949年 EDSAC (世界初の実用型)
 - Electronic Delay Storage Automatic Calculator
- 1951年 EDVAC (EDVAC 後継)
 - Electronic Discrete Variable Automatic Computer
- 二進数を使用
- メモリは 遅延記憶装置 を使用



EDSAC



EDVAC

ノイマン型 計算機

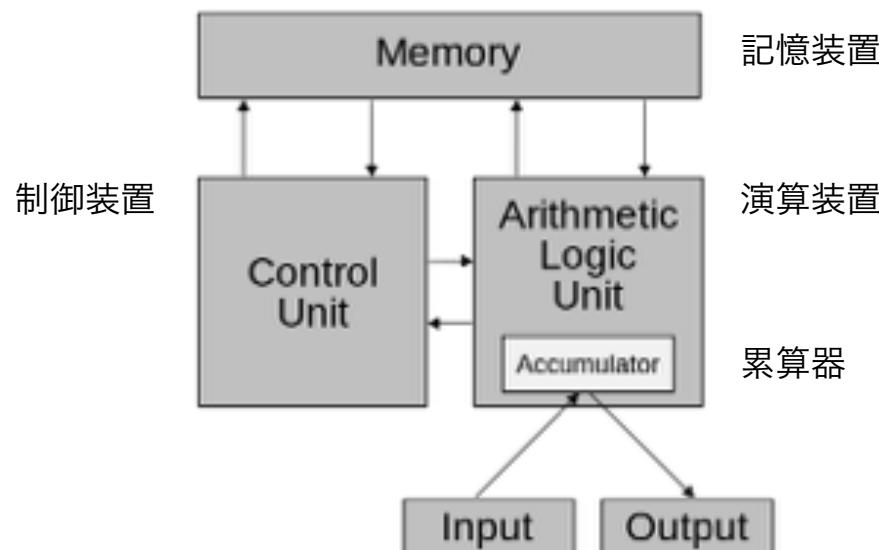


[John von Neumann \(1903 - 1957\)](#)

ハンガリー出身のアメリカ合衆国の数学者。

- プログラム内蔵型 電子計算機

- 結線でなく記憶装置に収めたプログラムを実行



ノイマン型 計算機

- プログラム内蔵型 電子計算機
 - 計算機のプログラムをコードで表し、数値データと同じ様に記憶装置に記憶させる
 - 一方、電気機械式計算機ではプログラムは記憶されず、一連の命令として紙テープに穿孔されていて、制御装置は順次テープからプログラムを読みだして実行する。
 - 命令が記憶装置に入っているので、演算を行って計算している途中で命令を変更することができる。

チューリングマシン



[Alan Mathison Turing \(1912 - 1954\)](#)

イギリスの数学者、論理学者、暗号解読者、計算科学者。



極めて単純な機械だが、これだけの機能があれば、必要最低限、人間が紙に書いて計算するのと同じことが、実現できる。

チューリングの仮想機械は、

1. 無限に長いテープ
2. その中に格納された情報を読み書きするヘッド
3. 機械の内部状態を記憶するメモリ

で構成され、内部状態とヘッドから読み出した情報の組み合わせに応じて、次の動作を実行する。

- ヘッド位置のテープに情報を書き込む
- 機械の内部状態を変える
- ヘッドを右か左に一つ移動する

上の動作を、機械は内部状態が停止状態になるまで反復して実行し続ける。

ハードウェア

- a) 計算機の歴史
- b) 現代のハードウェア構成

ソフトウェア

- a) オペレーティング・システム (OS)
- b) 言語・プログラム・ライブラリ

ネットワーク

よく使うインターネットプロトコル

スーパーコンピュータ

- プロセッサ、メモリ、ストレージ、ネットワーク等のハードウェアと、その上で動作する OS やアプリケーションなどのソフトウェアから構成される点では一般的なコンピュータと同じ
- ただし、本来画像処理のために開発された Graphics Processing Unit (GPU) を汎用的な計算に利用する General Purpose computing on GPU ([GPGPU](#)) などのコプロセッサによって浮動小数点演算性能を稼ぐ構成も多い
- 遺伝研スパコンは CPU 重視構成（でも GPU, Xeon Phi も一部に入ってますよ）

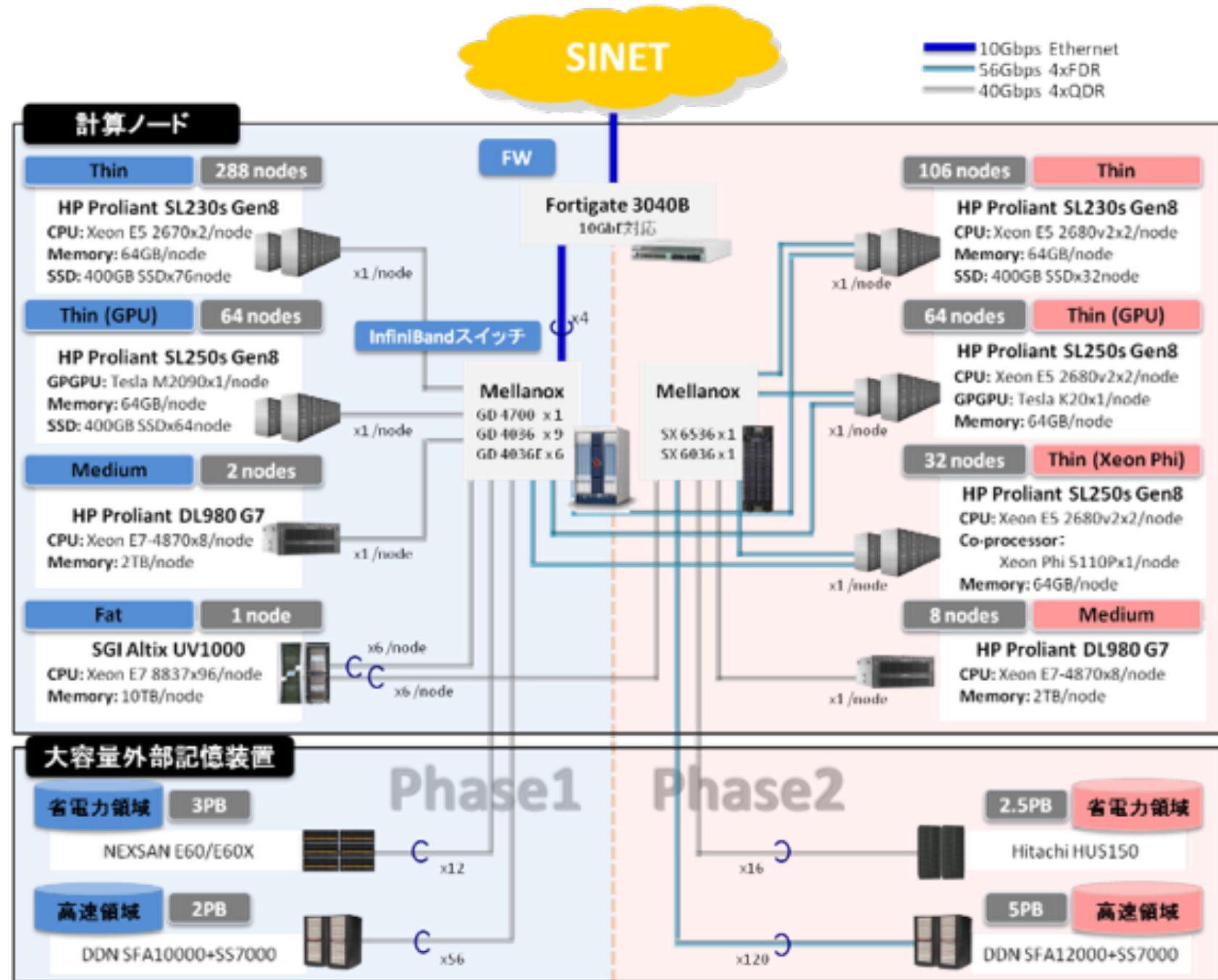
遺伝研スーパーコンピュータを例に

<http://sc.ddbj.nig.ac.jp/>

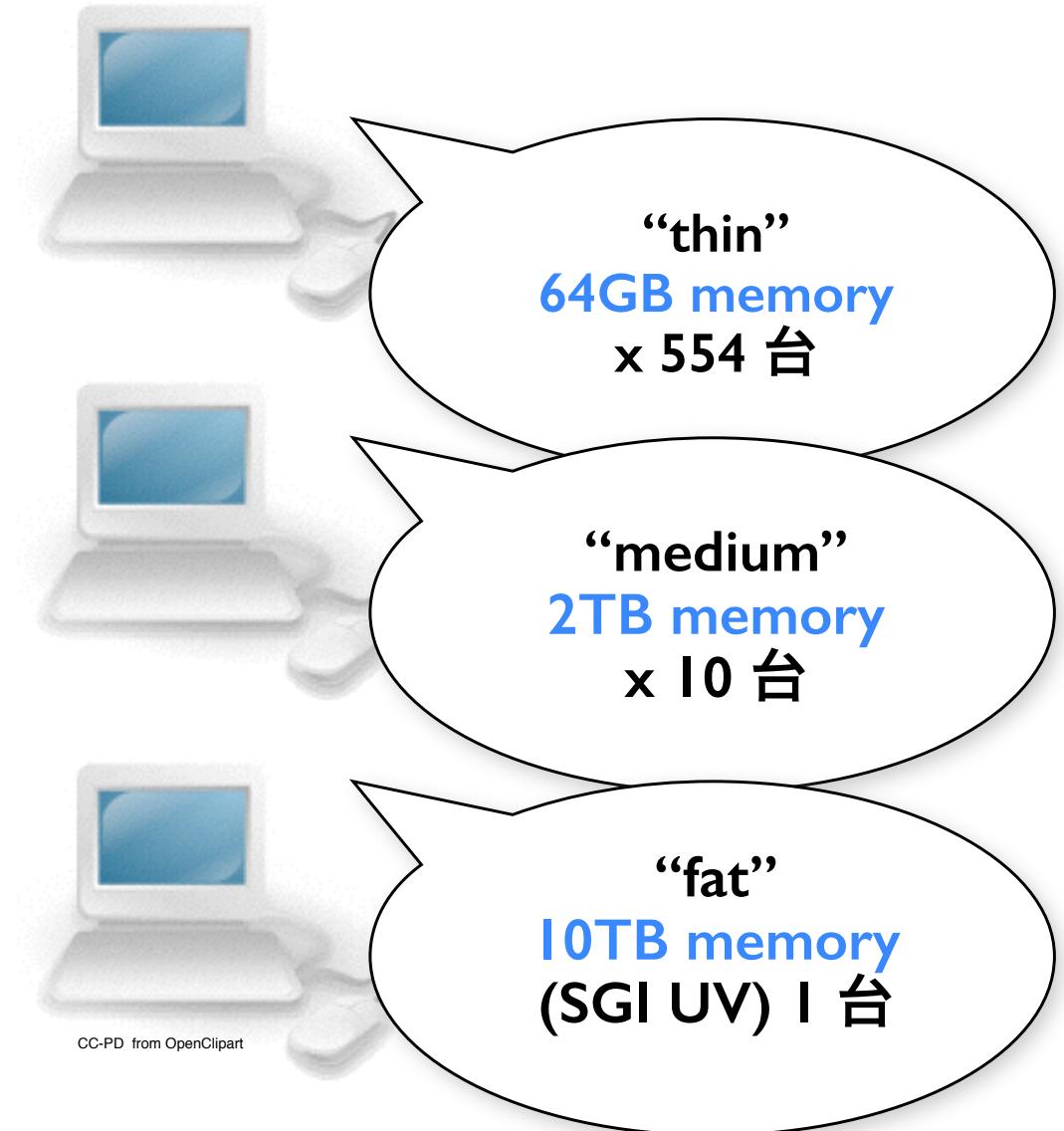
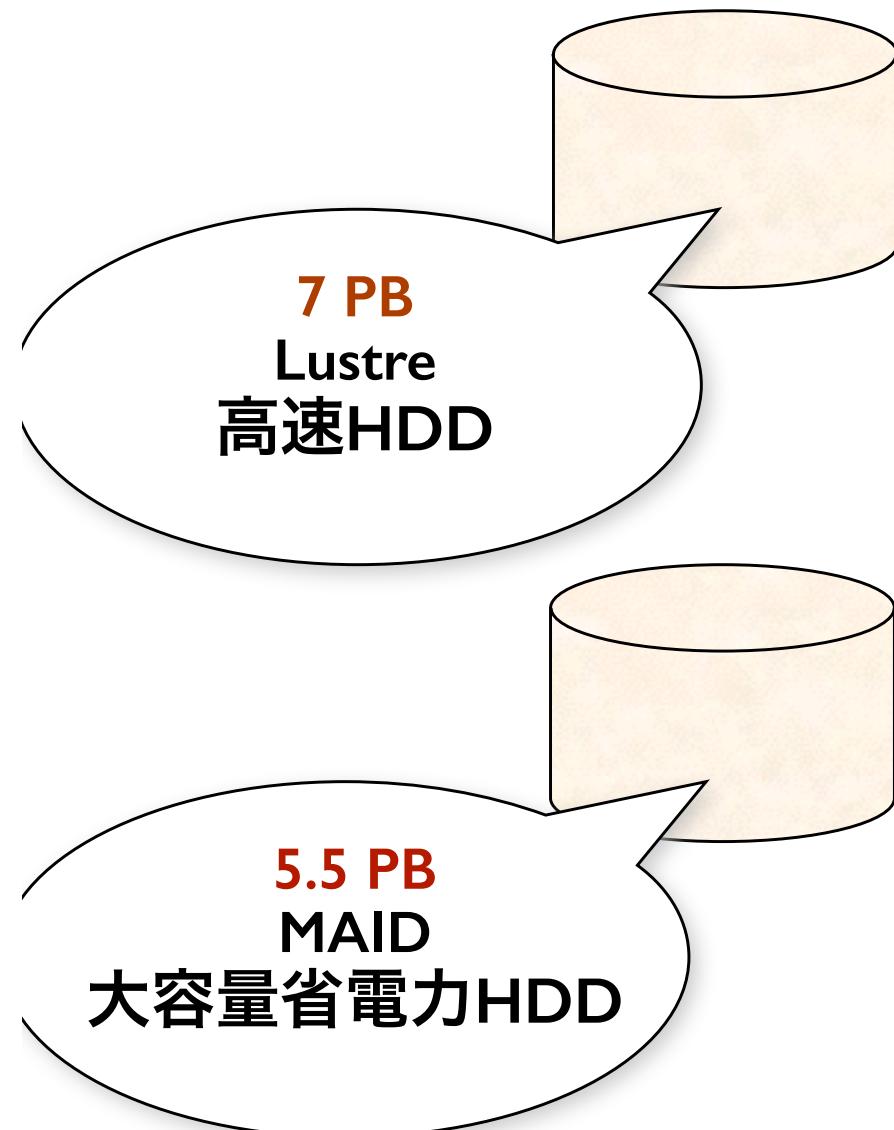


The screenshot shows the homepage of the NIG Supercomputer Facilities of National Institute of Genetics. The header includes the NIG logo, the text "大学共同利用法人 情報・システム研究機構 国立遺伝学研究所 スーパーコンピュータシステム SuperComputer Facilities of National Institute of Genetics", and links for "サイトポリシー" and "サイトマップ". The main content area has a sidebar with "Language/言語" (Japanese/English), "このサイトへのログイン" (Login), and "システム構成" (Hardware, Software, Programming Environment, Tools, OSS, DB). The "システム使用方法" (Usage Methods) sidebar lists "基本的利用方法", "その他UGE利用方法", "ファイル転送方法", and "システム利用TIPS". The main content area displays a news feed with entries from July 2014 to March 2014. Below the news is a section titled "国立遺伝学研究所 スーパーコンピュータシステム(NIG SUPERCOMPUTER)とは" which describes the system's architecture and capabilities. A large red arrow points to a photograph of a server room filled with server racks.

遺伝研スーパーコンピュータ (全容)



遺伝研スーパーコンピュータ (概要)



ハードウェア

- a) 計算機の歴史
- b) 現代のハードウェア構成

ソフトウェア

- a) オペレーティング・システム (OS)
- b) 言語・プログラム・ライブラリ

ネットワーク

よく使うインターネットプロトコル

UNIX Operating System のはじまり



Bell 研の Ken Thompson, Dennis Ritchieらが “Space Travel” で遊ぶために（の目的だけってわけでもないけど）高価で買ってもらえなかつた DEC-10 のかわりに部屋のスミに転がつていた借り物の PDP-7 で「えいやつ」と作つてみた「小さくて」「軽い」オペレーティングシステム (1968)

Life with UNIX (アスキー出版局; 1990) より引用／改変



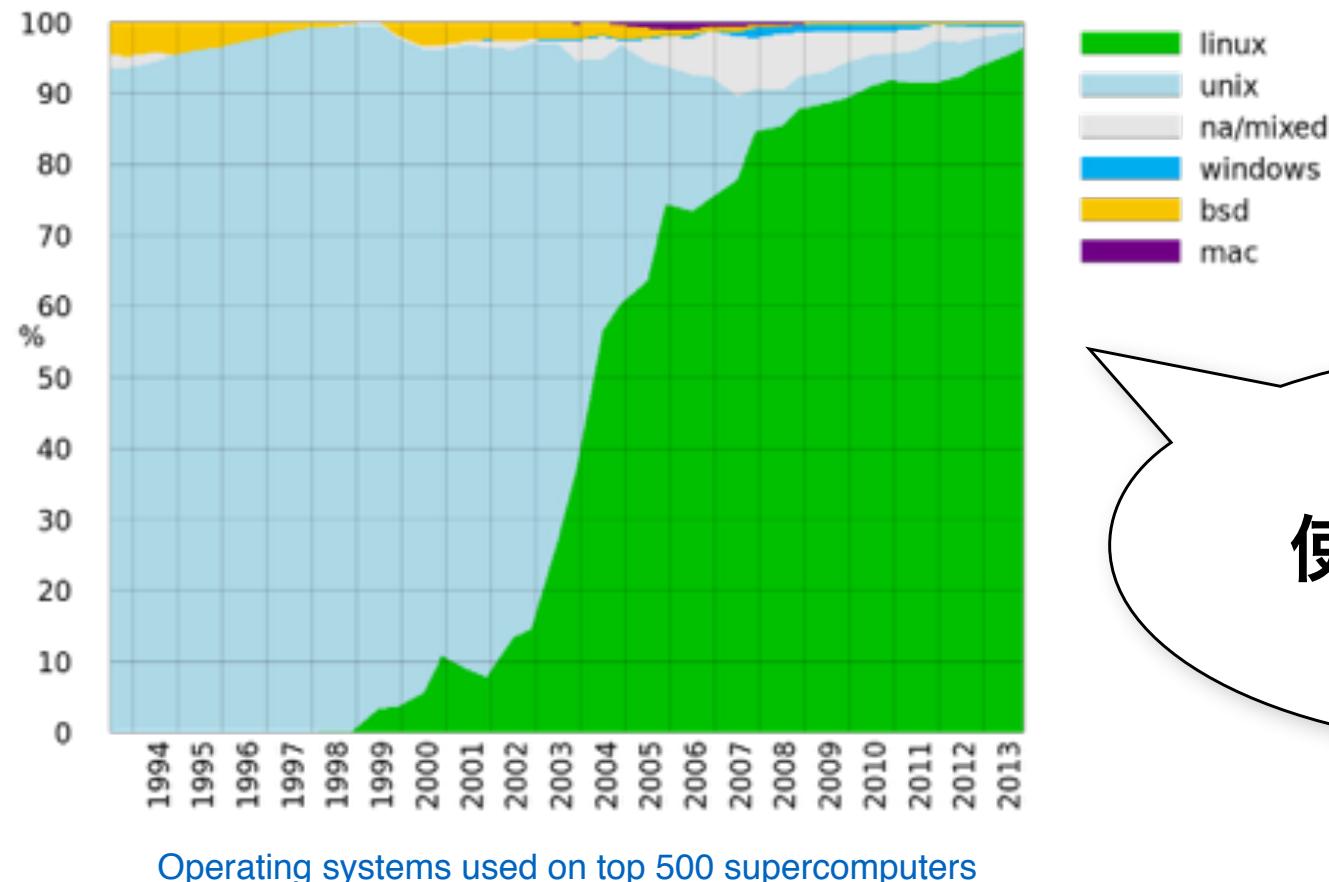
iOS の基礎部分は Darwin で
あり、NeXTSTEP を先祖に
持つ由緒正しき Berkeley
Software Distribution (BSD)
UNIX の系譜に連なります。

要するに組み込み系の UNIX
(POSIX 準拠) で動作。

MacOS X も勿論 UNIX。

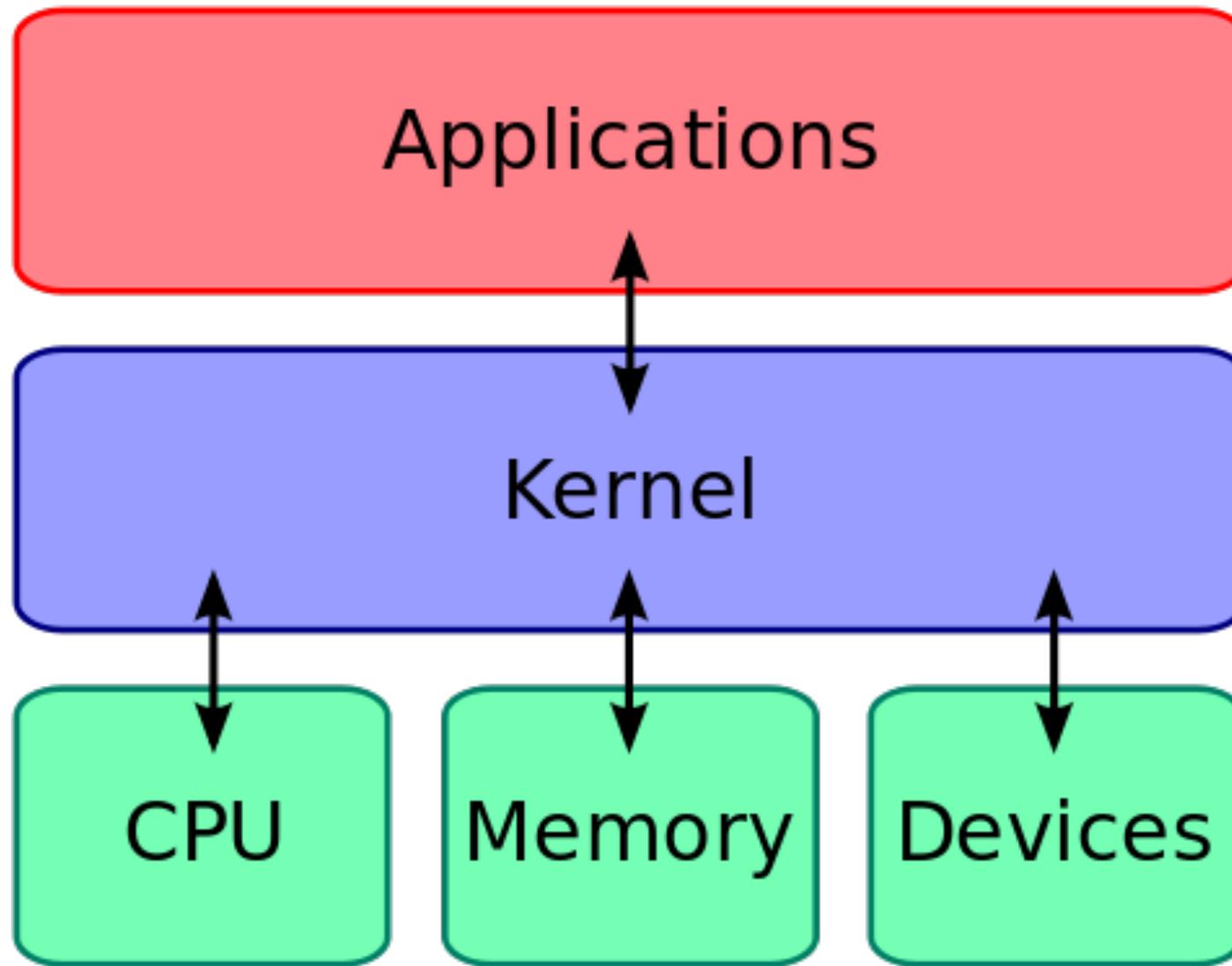
スパコンの OS って

- 遺伝研スパコン
 - Red Hat Enterprise Linux 6
 - C、C++、Objective-C、Fortran、Java、Ada



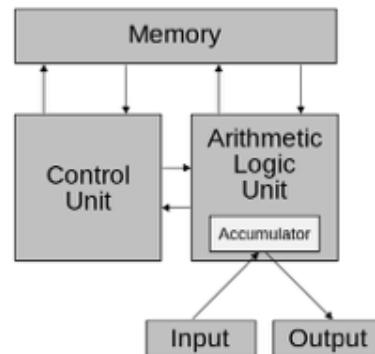
Top 500 で
使われている
OS の推移

UNIX の構成



Operating System のご利益

- 計算機とダイレクトにお話するのは大変に辛い

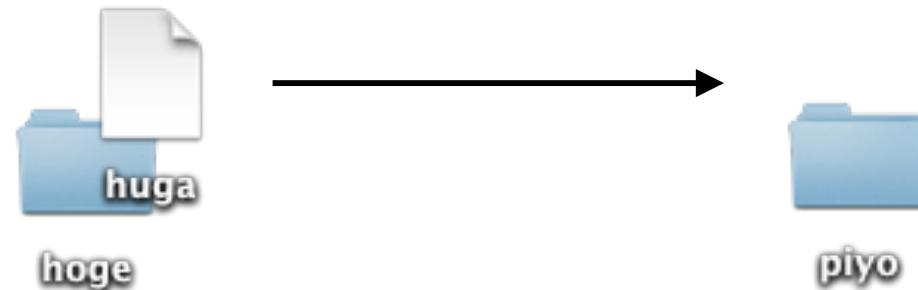


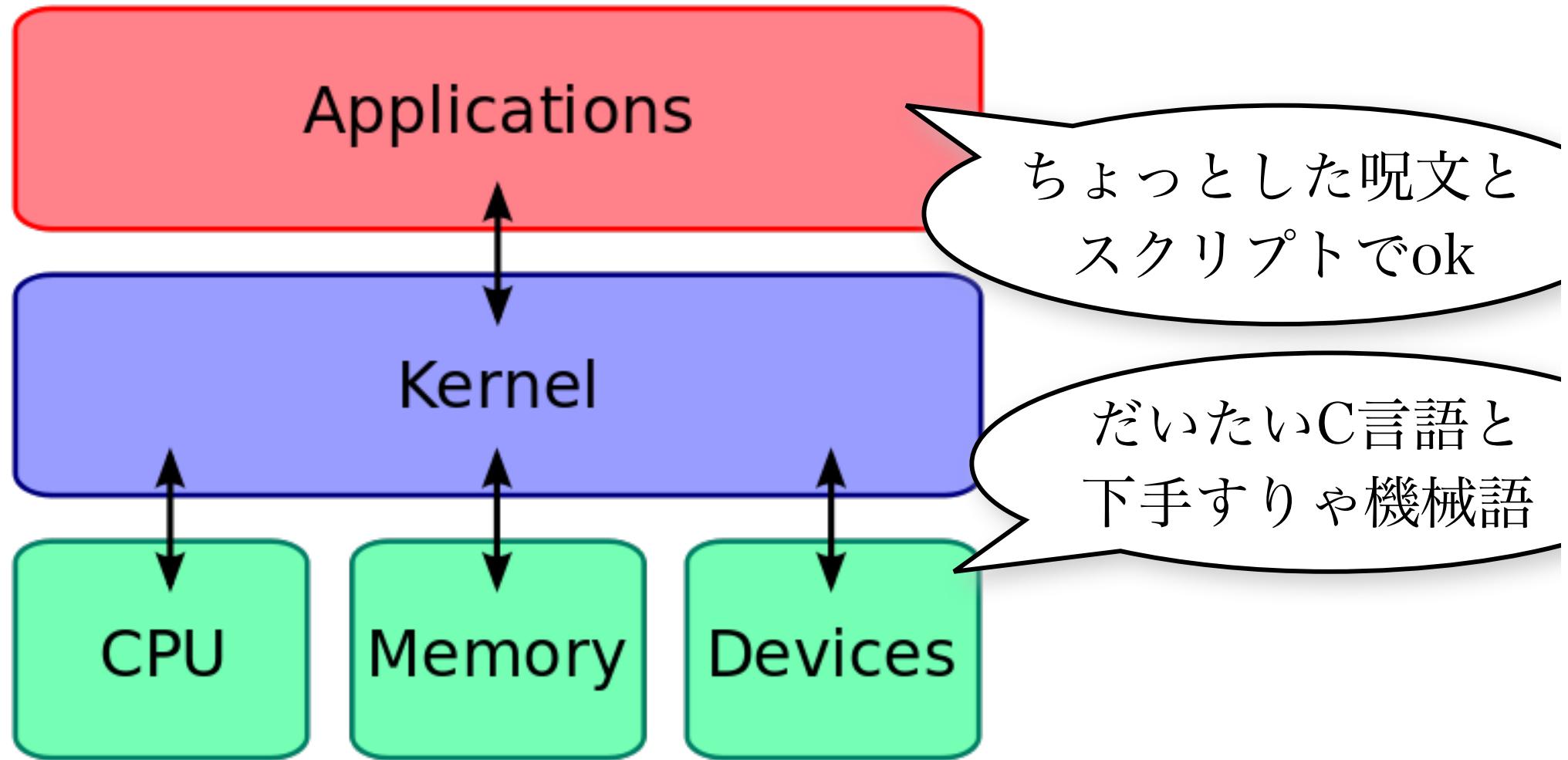
[Von Neumann architecture CC BY-SA 3.0](#)

- 「デバイスの何番地から何番地のデータを別のデバイスに書き移して跡地を空けよう。でも移す先に他のデータがあったらまずいので移すデータの空きがあるか調べて、連續でとれなかつたらデータを分割して書きこんで、分割してあることがわかるように印つけて…」
- みたいなことを逐一指示出すのはご勘弁願いたい

呪文みたいだけど、事細かに全部機械にお伝えするよりは、呪文のほうが楽な訳

```
$ mv /hoge/huga /piyo
```





http://commons.wikimedia.org/wiki/File:Kernel_Layout.svg

僕らが対話するのは「アプリ（シェルを含む）」の層

ハードウェア

- a) 計算機の歴史
- b) 現代のハードウェア構成

ソフトウェア

- a) オペレーティング・システム (OS)
- b) 言語・プログラム・ライブラリ

ネットワーク

よく使うインターネットプロトコル

遺伝研スーパーコンピュータの場合

<http://sc.ddbj.nig.ac.jp/>



The screenshot shows the homepage of the NIG Supercomputer Facilities of National Institute of Genetics. The header includes the NIG logo, the text '大学共同利用法人 情報・システム研究機構 国立遺伝学研究所 スーパーコンピュータシステム SuperComputer Facilities of National Institute of Genetics', and links for 'サイトポリシー' and 'サイトマップ'. The main content area features a sidebar with 'Language/言語' (Japanese/English), 'このサイトへのログイン' (Login), and sections for 'システム構成' (Hardware, Software, Programming Environment, Tools, OSS, DB) and 'システム使用方法' (Basic Usage, UGE, File Transfer, System Status). The central part displays a list of important notices (お知らせ) with dates and titles, such as '【スパコンユーザ会】三島開催 会場変更のお知らせ' (July 22, 2014). Below this is a section about the 'NIG SUPERCOMPUTER' system, mentioning its update in March 2012 and its role in genome analysis. At the bottom, there is a photograph of a server room.

- 「プログラミング環境について」

パッケージ名	バージョン	対応言語	インストール位置
Intel Composer XE Linux	14.0.0(XE2013)	C、C++、Fortran77, 90,IV,2003,2008	/opt/intel/composer_xe_2013_sp1.0.080/
PGI Accelerator CDK Cluster Development Kit	13.9	C、C++、C99、Fortran77 90,95,2003、HPF	/opt/pgi/linux86-64/current/
Gnu Compiler(OSS)	4.8.1	C、C++、Objective-C、 Fortran77, 90,2003,2008、Java、Ada	/usr/local/bin/

- 「科学技術計算ライブラリについて」

パッケージ名	バンドル製品	インストール位置
Intel Cluster Studio	Intel MKL、Intel IPP、Intel TBB、Intel MPI、Intel Trace Analyzer/Collector	/opt/intel
PGI Accelerator CDK Cluster Development Kit	PGDBG、PGPROF、ACML、ScalAPACK(v1.7)	/opt/pgi

<http://sc.ddbj.nig.ac.jp/index.php/system-software-config>

…まあ、普通に使うぶんにはとくに用はないか

シェル

- 「シェル環境について」
 - システムへのログインシェルは、bash, tcs, zsh が利用可能です。

<http://sc.ddbj.nig.ac.jp/index.php/system-software-config>

UNIX にコマンドを出すためのプログラムである「シェル」は複数のなかから選べます。お好きなものをどうぞ。

「シェルスクリプト」が書けるようになると、ぐつと大量の解析作業が捲るようになりますよ。

スクリプト言語

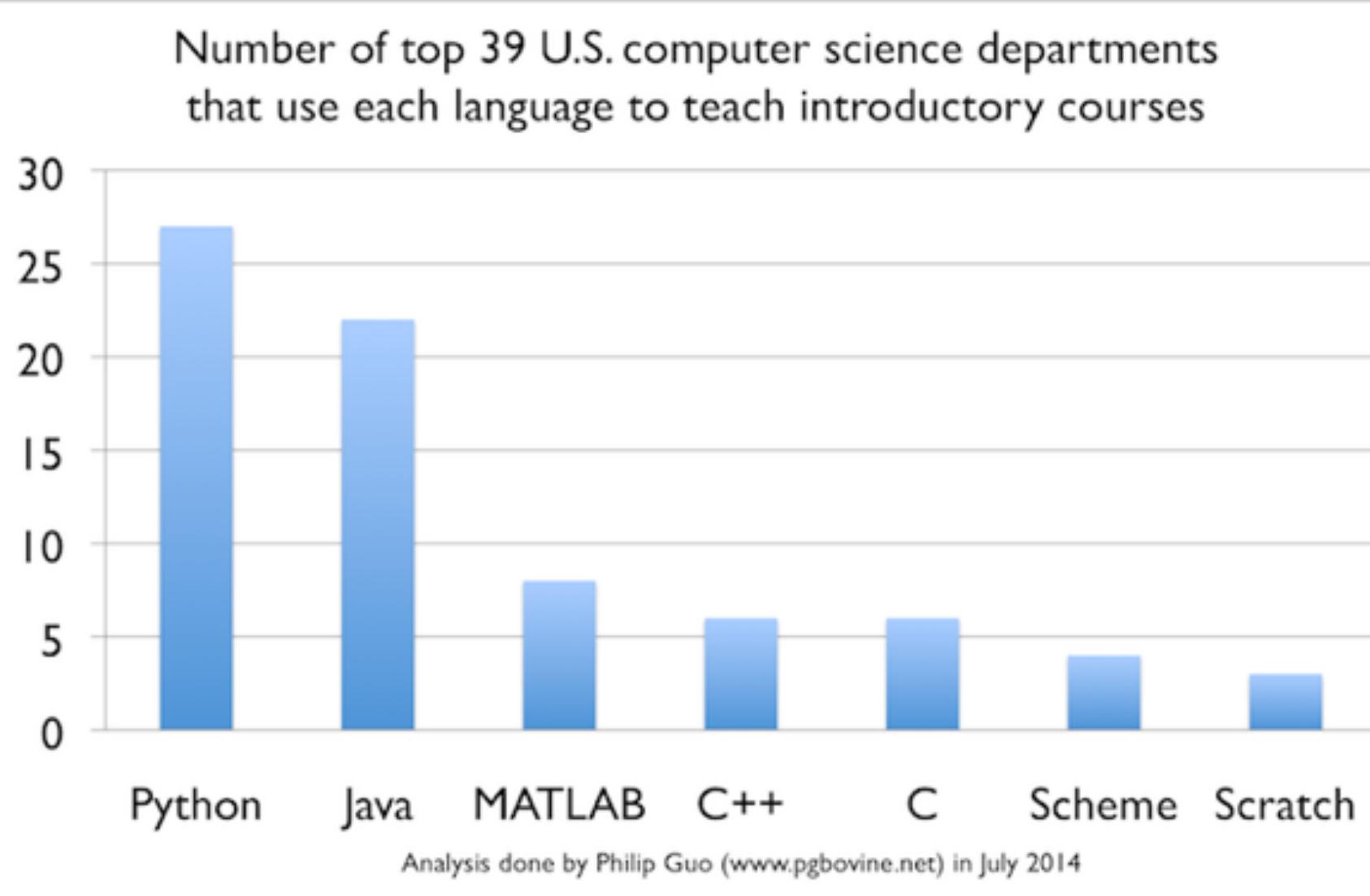
- 「一般ユーザが利用可能なスクリプト言語」
 - ログインノード、計算ノードで一般ユーザが利用可能なスクリプト言語環境は以下の通りです。

言語	バージョン	インストールパス
ruby	1.9.3p125	/usr/local/bin/ruby
python	2.7.2	/usr/local/bin/python
perl	5.14.2	/usr/local/pkg/perl

<http://sc.ddbj.nig.ac.jp/index.php/system-software-config>

さらに「スクリプト言語」が使えるようになると、一連の解析を効率的に繋いで高度化できますよ！

米国での導入は Python が人気らしいよ



<http://cacm.acm.org/blogs/blog-cacm/176450-python-is-now-the-most-popular-introductory-teaching-language-at-top-us-universities/fulltext>

Zea
Mays

Rattus norvegicus

バイオインフォマティクス人材育成カリキュラム(次世代シークエンサ) | 速習コース



中村 保一

NAKAMURA Yasukazu, Professor

国立遺伝学研究所 大量遺伝情報研究室

<http://charles.genes.nig.ac.jp> yn@nig.ac.jp

ハードウェア

- a) 計算機の歴史
- b) 現代のハードウェア構成

ソフトウェア

- a) オペレーティング・システム (OS)
- b) 言語・プログラム・ライブラリ

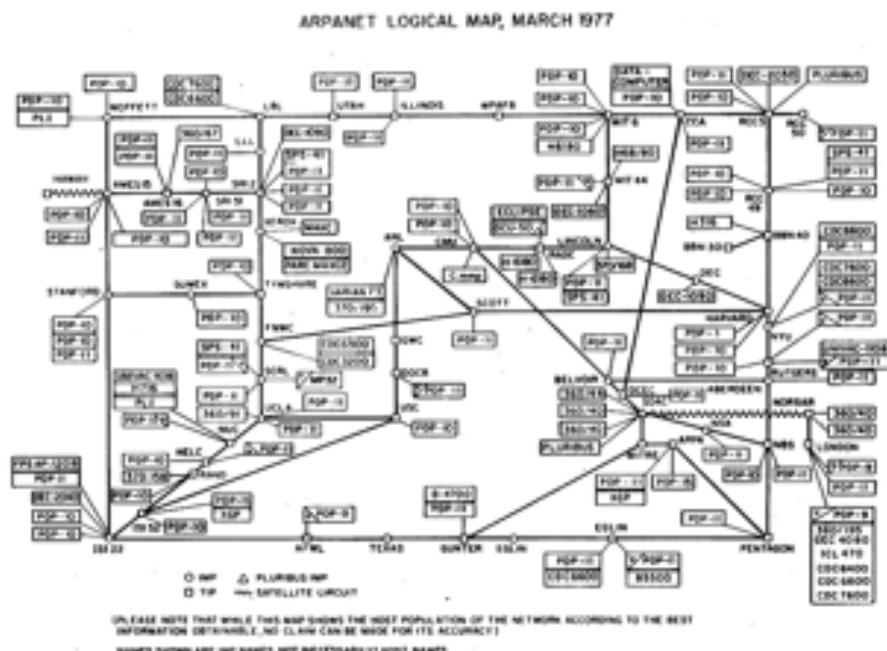
ネットワーク

よく使うインターネットプロトコル

インターネットのはじまり

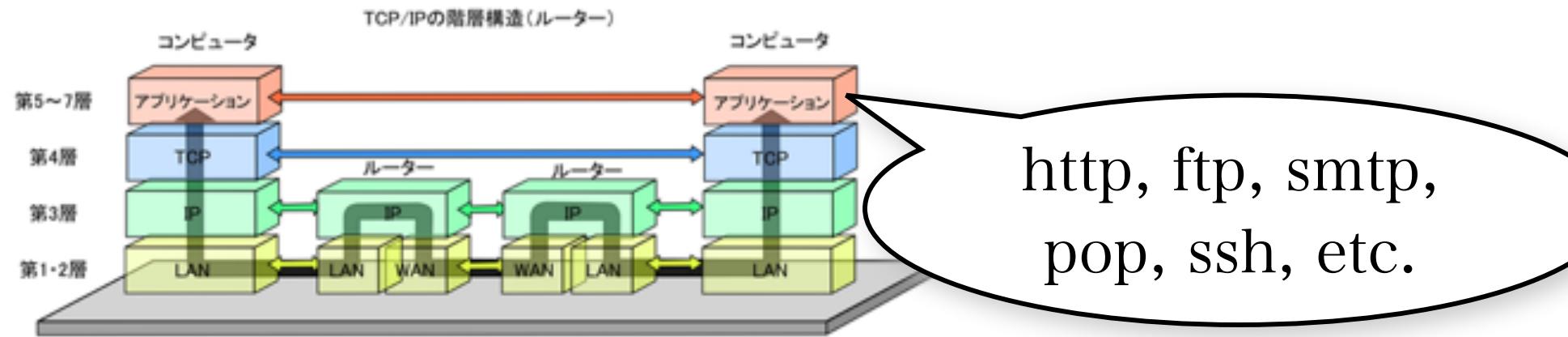


- Advanced Research Projects Agency Network
 - ARPANETは、世界で初めて運用されたパケット通信ネットワークであり、今日の世界的なインターネットの起源。アメリカ国防総省（Advanced Research Projects Agency、略称ARPA）が資金提供し、大学と研究機関を接続したプロジェクト



Arpanet logical map, march 1977

TCP/IP



- UNIX の構造と同じく、一般ユーザが利用する際には下部の階層を強く意識する必要はあまりなく、上位の階層である アプリケーション層の http, ftp などの通信プロトコルの指定 (稀に TCP ポート指定) と、IP アドレス の設定あたりがわかれれば良いと思う。

よく使う通信プロトコル

- HTTP: Hypertext Transfer Protocol
- FTP: File Transfer Protocol
 - ファイルの送受信に使う転送プロトコル
- SSH: Secure Shell
 - 認証を含むすべてのネットワーク上の通信を暗号化し安全に通信するプロトコル
- SFTP: SSH File Transfer Protocol
 - SSH を利用してセキュアにファイル転送
- SMTP/POP: 電子メール送受信に使うプロトコル

遺伝研スーパーコンピュータの場合

<http://sc.ddbj.nig.ac.jp/>



The screenshot shows the homepage of the NIG Supercomputer Facilities of National Institute of Genetics. The header includes the NIG logo, the text '大学共同利用法人 情報・システム研究機構 国立遺伝学研究所 スーパーコンピュータシステム SuperComputer Facilities of National Institute of Genetics', and links for 'サイトポリシー' and 'サイトマップ'. The main content area features a '重要なお知らせ' (Important Notices) section with a list of recent announcements. Below it is a section titled '国立遺伝学研究所 スーパーコンピュータシステム(NIG SUPERCOMPUTER)とは' (What is the NIG Supercomputer System?) which describes the system's purpose and components. At the bottom, there are five images showing server racks in a data center.

大学共同利用法人 情報・システム研究機構 国立遺伝学研究所
スーパーコンピュータシステム
SuperComputer Facilities of National Institute of Genetics

現在地: Home

2014年08月30日

Language/言語

ホーム

このサイトへのログイン

Login
(スパコンユーザでログイン可)

システム構成

ハードウェア構成

ソフトウェア構成

プログラミング環境

利用可能バイオツール

利用可能OSS

利用可能DB

システム使用方法

基本的利用方法

その他UGE利用方法

ファイル転送方法

システム利用TIPS

重要なお知らせ

公開日 表題

2014年7月22日 【スパコンユーザ会】三島開催 会場変更のお知らせ

2014年7月18日 ソフトウェアインストールのお知らせ

2014年7月15日 【fat障害:2】fat計算ノードサービス再開のお知らせ

2014年7月14日 【通信障害:2】Phase1システム通信不具合の復旧のお知らせ

2014年7月14日 【通信障害】Phase1システム通信不具合のお知らせ

2014年3月4日 2014年3月5日からのスパコンPhase2システムご利用方法について

一覧へ

国立遺伝学研究所 スーパーコンピュータシステム(NIG SUPERCOMPUTER)とは

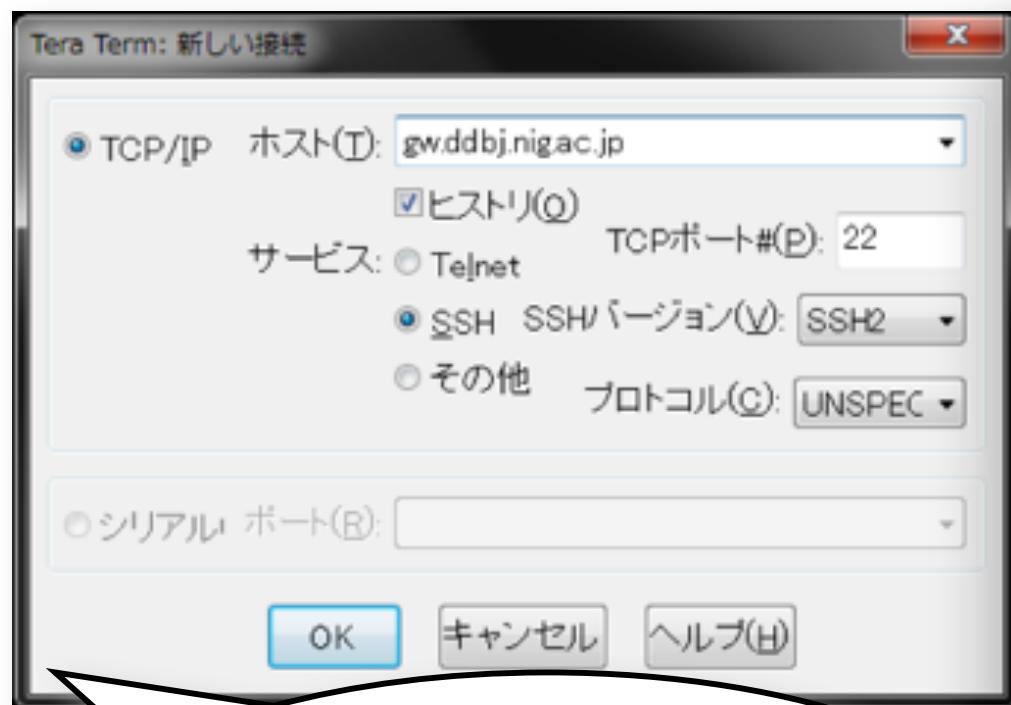
大学共同利用機関法人 情報システム研究機構 国立遺伝学研究所は、2012年3月にスーパーコンピュータシステムを更新しました。新しいスーパーコンピュータシステムはゲノム解析を主な目的とした大規模計算機利用拠点として 最新鋭の大規模クラスタ型計算機、大規模メモリ共有型計算機、および大容量高速ディスク装置で構成されたスーパーコンピューティングシステムサービスを提供しています。



スパコンへのログイン (1)

- Windowsから

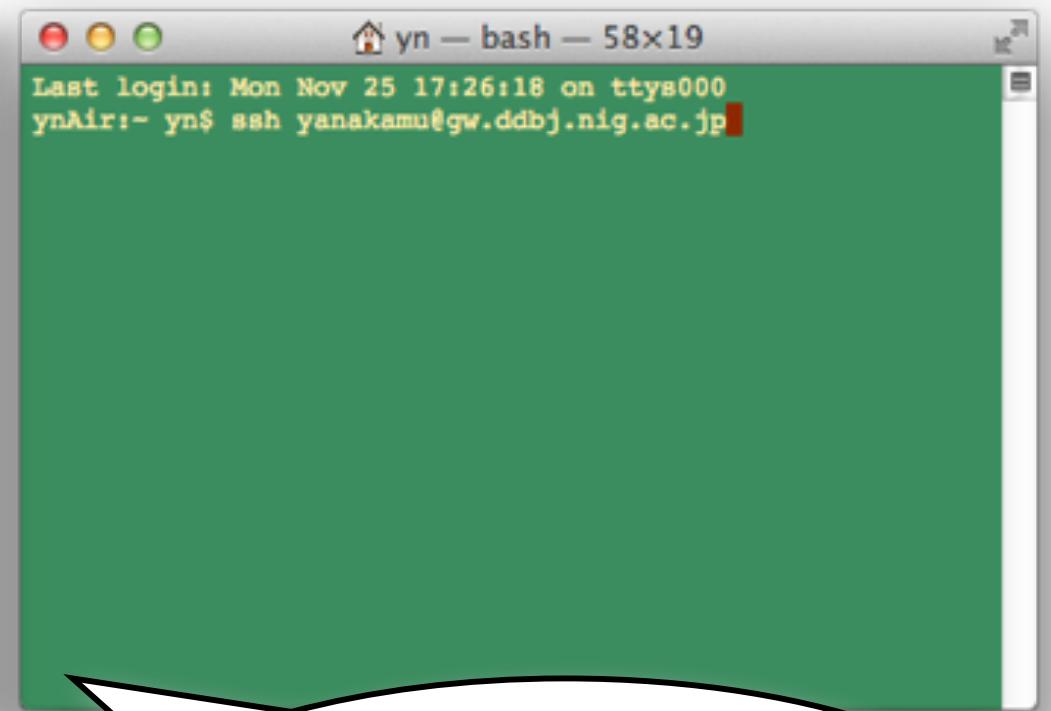
- Tera Term 起動



ホスト : gw.ddbj.nig.ac.jp
 サービス : SSH
 ⇒ OK

- Macから

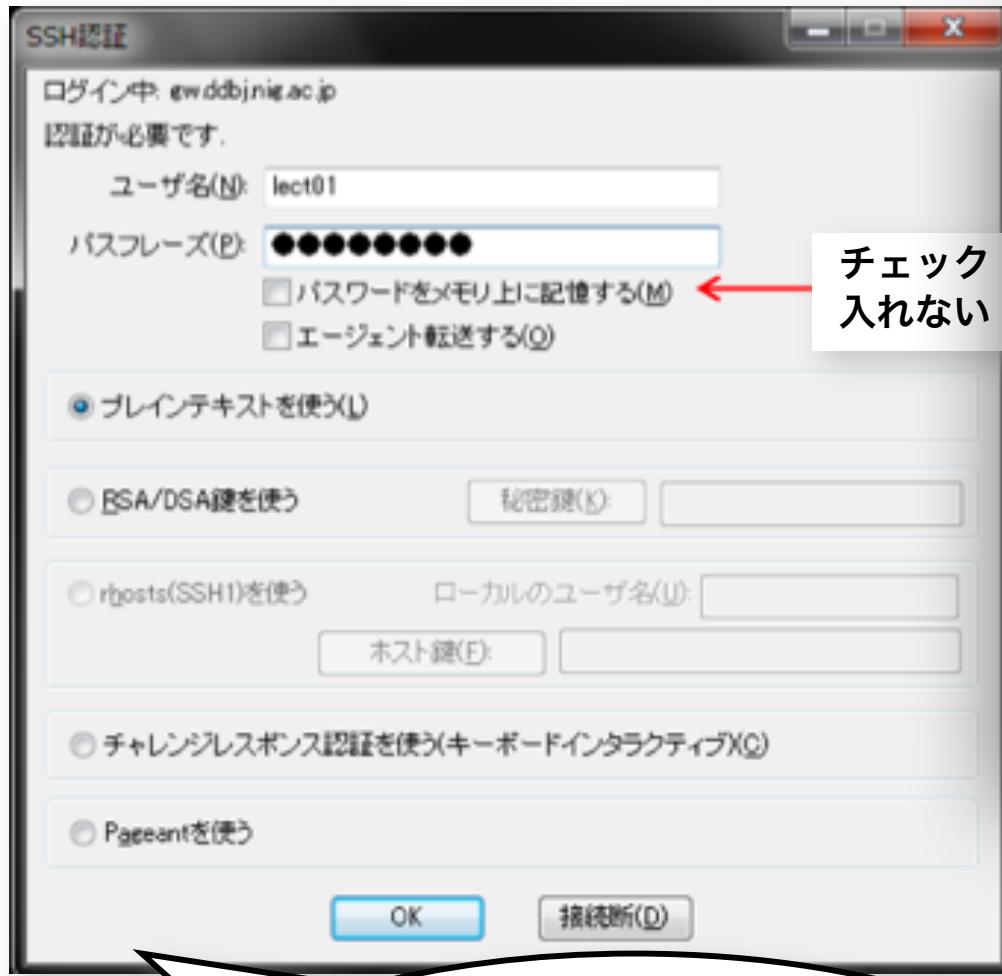
- ターミナル 起動



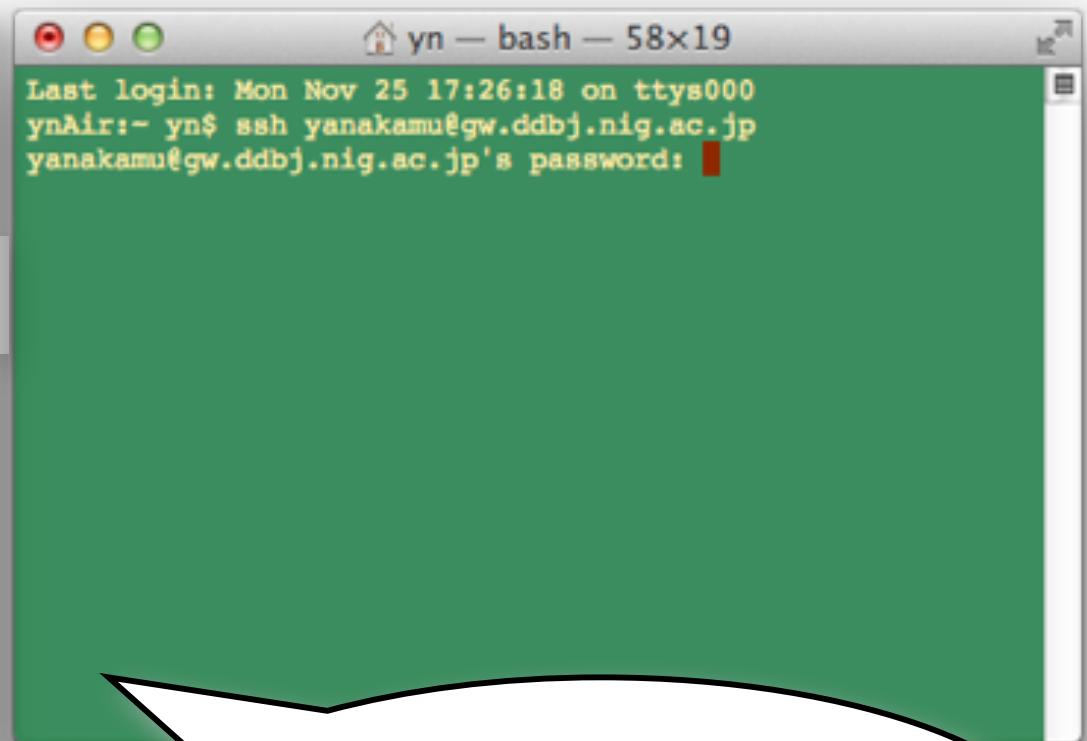
ssh ユーザ名@gw.ddbj.nig.ac.jp
 ⇒ Enter

スパコンへのログイン (2)

- Windowsから



- Macから



The screenshot shows a terminal window titled 'yn - bash - 58x19'. It displays the command `ssh yanakamu@gw.ddbj.nig.ac.jp` and a password prompt: 'yanakamu@gw.ddbj.nig.ac.jp's password:'. A large speech bubble on the right says 'パスワードを入力 ⇒ Enter' (Enter password ⇒ Enter).

ユーザ名とパスワードを入力
⇒ OK

パスワードを入力
⇒ Enter

スパコンへのログイン (3)

• Windowsから



```
133.39.224.14-22 - Tera Term VT
[ファイル] [編集] [選択] [コントロール] [ウィンドウ] [ヘルプ]
Last login: Fri Mar  9 16:26:46 2012 from hitach27.genes.nig.ac.jp
Thank you for using supercomputer system.
This node is in use for login service only. Please use 'qlogin'.
[lect01@gw ~]$
```

• Macから



```
yn — yanakamu@t351:~ — bash — 72x30
>Last login: Mon Nov 25 17:26:18 on ttys000
ynAir:- yn$ ssh yanakamu@gw.ddbj.nig.ac.jp
yanakamu@gw.ddbj.nig.ac.jp's password:
Last login: Mon Nov 25 18:17:43 2013 from 133.39.20.15
Thank you for using supercomputer system.
This node is in use for login service only. Please use 'qlogin'.
[yanakamu@gw ~]$
```

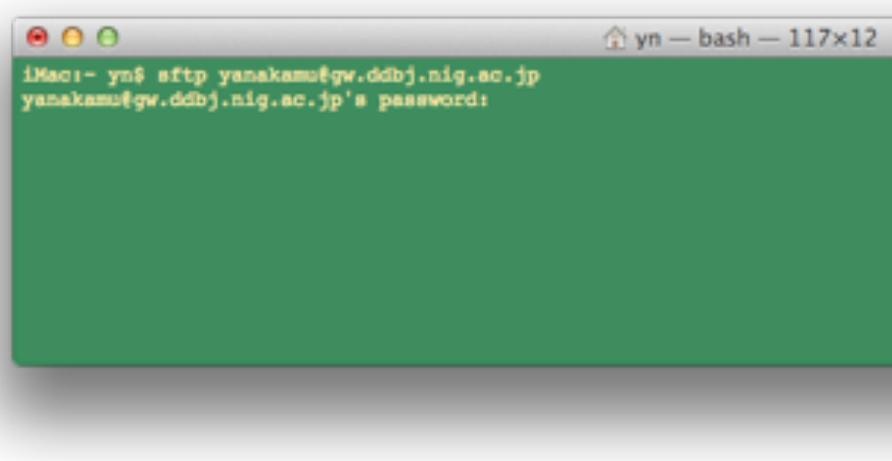
gwにログインできました。
(まだ玄関先なのでもう一手間
qlogin をする必要があり)

パスワードを4回間違えると
アカウントロックされます

ロックされた場合には
sc-info@nig.ac.jp まで

外部サーバとのファイル送受信 (1)

- sftpを利用したファイル転送方法
 - 遺伝研スパコンでは、ゲートウェイサーバ (gw.ddbj.nig.ac.jp) に対して sftp で接続可



```
iMac:~ yn$ sftp yanakama@gw.ddbj.nig.ac.jp
yanakama@gw.ddbj.nig.ac.jp's password:
```

または



<http://sc.ddbj.nig.ac.jp/index.php/ja-file-transfer>

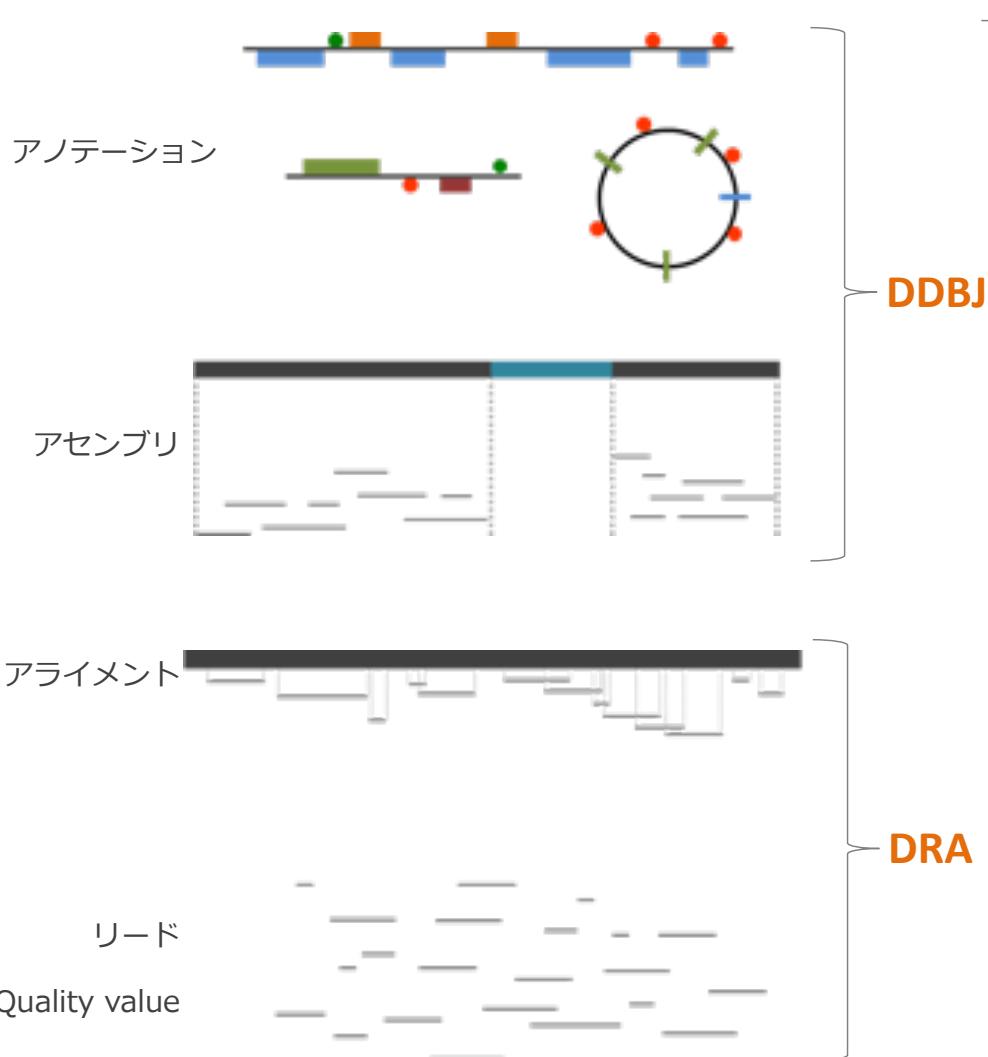
Cyberduck のような専用クライアントを使えば
ドラッグ&ドロップでファイル転送も。コマンド
での送受信もたいして難しくありませんが。

- Asperaを利用したファイル転送方法
 - 遺伝研スパコンではインターネット帯域を使い倒して大容量ファイルを高速に転送する商用ソフト Aspera が使用可能です。
 - Asperaを用いたファイル転送には、ユーザの端末に クライアント Aspera Connect を導入します。対応 OSは、Windows、Mac、Linux。
 - データ暗号化にも対応。

<http://sc.ddbj.nig.ac.jp/index.php/ja-file-transfer>

DDBJが運営しているデータベース

DDBJ センター



アクセス制限

JGA

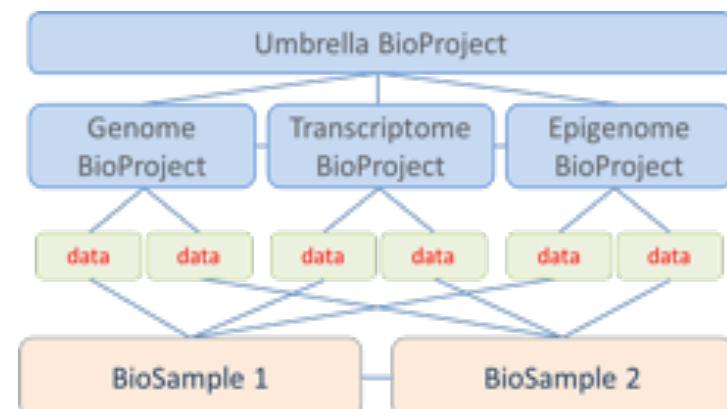
個人の遺伝型と表現型



ヒトデータ審査委員会



BioProject
BioSample



INSDC: オープンアクセス

DRA: DDBJ Sequence Read Archive

JGA: Japanese Genotype-phenotype Archive

DDBJ Sequence Read Archive (DRA)



新世代シークエンサから出力される配列や
アライメントデータを登録・公開

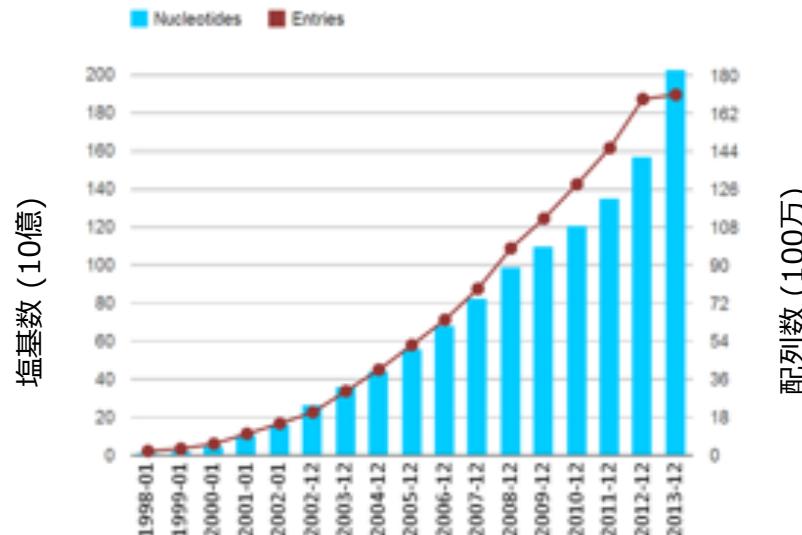


International Nucleotide Sequence Databank Collaboration

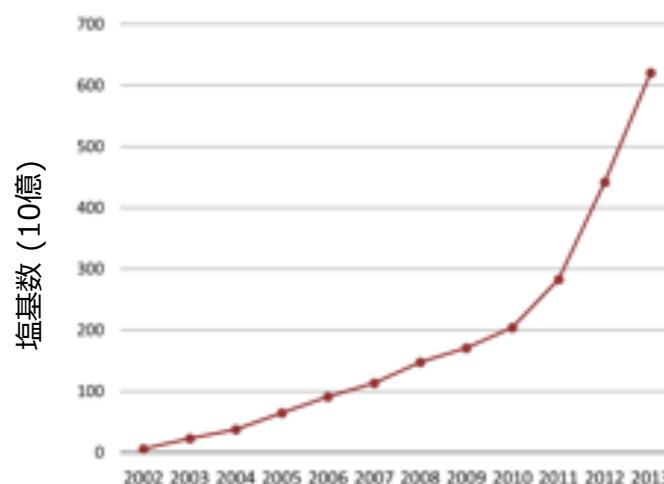
データ量の推移

DDBJ/ENA/GenBank データベースの成長

アノテーションされた配列 (WGS 以外)



Whole Genome Shotgun (WGS) 配列

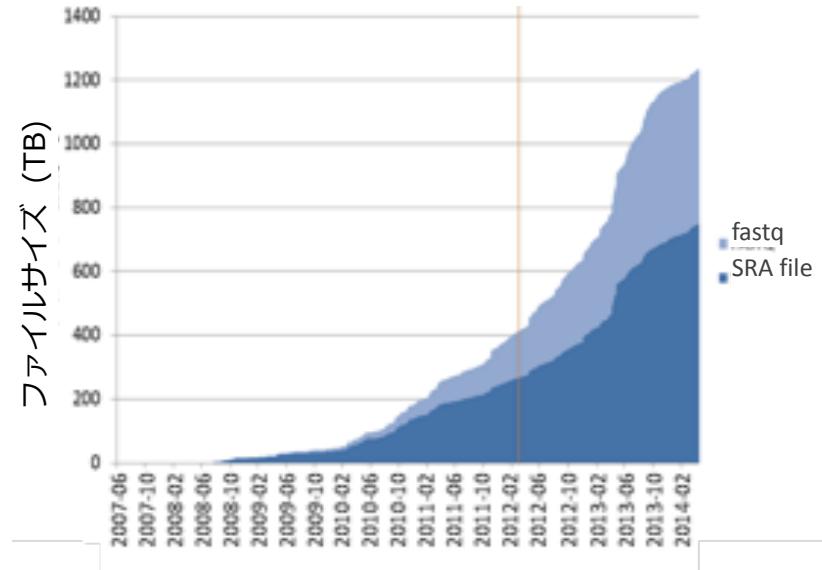


- 総計 8,200 億塩基 (30 万生物種)

SRA データベースの成長

次世代シークエンシングデータ、アライメント

新 NIG スパコン導入



- 1,000 兆塩基
- 1.4 PB
- 2017年3月の次期リプレースまでに
割り当てられている 4.5 PB が一杯になる計算

DRAウェブサイト ⇒ [DRA] で検索



<http://trace.ddbj.nig.ac.jp/dra/>

登録関係情報

解析パイプライン

データ検索

データ取得

検索

登録

動画マニュアル

DDBJ Sequence Read Archive (DRA) は、Illumina 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シークエンサからの出力データを登録するためのデータベースです。また、International Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、DRA は、International Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャビラリーによる測定データも、DRA にご登録ください。

検索
データをキーワード、生物名、シークエンサなどで検索する

登録
新型シークエンサからの生データやアライメントデータを登録する

動画マニュアル
DRA の利用方法や登録方法を解説している動画を見る

公開データの DRA Search での検索

公開データは EBI SRA / NCBI SRA と共有されています

生物名 etc での絞り込み

検索結果リスト

ダウンロード

詳細（メタデータ記述）

The screenshot shows the DRA Search interface. At the top left, there's a search form with fields for Accession (DRA000001), Organism (Pseudotrophe defector ATCC 49176), StudyType (Epigenetics), Platform (ILLUMINA), and a Statistics section. Below the search form is a 'Search Results' table with 358 studies. One study is highlighted: DRP000003. The study details page for DRP000003 includes sections for Study Detail (Title: Comprehensive identification and characterization of the nucleosome structure, Abstract: A detailed description of the study using the Nucleosome-Seq method), Description (A note about alternative promoters), Project ID (21552), and Center Name (UT-MGS). On the right, a 'Navigation' sidebar lists Submission (DRA000003), Experiment (DRP000003), and Sample (DRS000003). A large blue arrow points from the search form down to the search results table.

解析パイプラインも提供しています

<http://trace.ddbj.nig.ac.jp/dra/>

Screenshot of the DDBJ Sequence Read Archive (DRA) website.

The page title is "Sequence Read Archive". The navigation bar includes links for Home, Submission, Search, Download, Pipeline (which is highlighted with a yellow box), and About. A search bar is located above the main content area, featuring a "Google" logo and a search icon.

The main text describes DRA as a database for next-generation sequencing data from Roche 454 GS System, Illumina Genome Analyzer, Applied Biosystems SOLiD System, and other sequencing platforms. It is operated by INSDC in collaboration with NCBI SRA and EBI ERA.

Below the text are three call-to-action boxes:

- 検索**: Search for data using keywords, organism names, or sequencing instrument names.
- 登録**: Register new sequencing data or alignment files from new sequencing instruments.
- 動画マニュアル**: Watch videos explaining how to use DRA and how to register data.

よく用いられる
解析用ソフトウェアを
用意。クリックだけで
実行可能

DRA pipeline: ソフトウェア

Select Query Files → Select Tools → Set QuerySet → Set Genome

Running Status

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK **NEXT**

Reference Genome Mapping

	Tool	Help	Version	Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	Comment
<input type="checkbox"/>	BLAT		34	✓					✓						Single-end analysis only
<input type="checkbox"/>	Mag		0.7.1	✓		✓			✓	✓	✓	✓	✓	✓	
<input type="checkbox"/>	bwa		0.5.9	✓		✓			✓					✓	
<input type="checkbox"/>	SOAP		2.21	✓		✓			✓	✓	✓			✓	
<input type="checkbox"/>	Bowtie		0.12.7	✓	✓	✓			✓	✓				✓	
<input type="checkbox"/>	TopHat		1.0.11	✓		✓			✓					✓	
<input type="checkbox"/>	Bowtie2		2.0.0	✓	✓	✓			✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.

de novo Assembly
Total limit = 22 Gbp

Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment
<input type="checkbox"/>	SOAPdenovo		1.05		✓		
<input type="checkbox"/>	ABYSS		1.3.2		✓		Maximum K-mer value is 64.
<input type="checkbox"/>	Velvet		1.2.03		✓	✓	We severely recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.
<input type="checkbox"/>	TIGAL		(2013-02-14)				

ACCOUNT
login ID [yaskaz]
 Logout
 Change password

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start
step-1
Preprocessing
Mapping /
de novo Assembly
step-2
Workflow
Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.
Preprocessing
step1.
Mapping
step1.
de novo Assembly
step2-All status

HELP
[HELP](#)
TUTORIAL
 Contact Us.
DDBJ Read Annotation
Pipeline.
Development Team.

DRA pipeline: 比較対象

イネ、マウスなど
解析比較対象となる
配列を多数用意

The screenshot shows the DRA pipeline interface with the following components:

- Left Sidebar (ACCOUNT):** Includes login ID [yaskaz], Logout, Change password.
- Left Sidebar (ANALYSIS):** Data setup, DRA Start (highlighted), FTP upload.
- Main Panel:** A list of organisms with a blue checkmark next to **Arabidopsis thaliana**. Other listed organisms include Oryza sativa japonica, Oryza sativa indica, Zea mays B73, Sorghum bicolor, Homo sapiens, Mus musculus, Pan troglodytes, Caenorhabditis elegans, Xenopus (Silurana) tropicalis, Oryzias latipes, Solanum lycopersicum Heintz 1706, and Saccharomyces cerevisiae. Below this is a section for ChIP-seq (frag counts) and ChIP-seq.
- Central Panel (Specifying Database of Ref):**
 - Major genome sets:** Shows Organisms (Arabidopsis thaliana) and Genome sets (TAIR8, TAIR9, TAIR10). A red arrow points from the Organisms dropdown to the expanded list of genome sets for Oryza sativa japonica.
 - User original sets:** Options for all.fa, chr01.fa, chr02.fa, chr03.fa, chr04.fa, chr05.fa, chrC.fa, and chrM.fa.
 - Download or upload reference:** Options for all.fa, chr1.fa, chr10.fa, chr11.fa, chr12.fa, chr13.fa, chr14.fa, chr15.fa, chr16.fa, chr17.fa, and chr17_cnv5_haplotype.
- Right Panel (Major genome sets):**
 - Oryza sativa japonica:** Genome sets available include IRGSP Releases Build 4.0 (selected), Build 5.0, Build 5.0 masked by RepeatMasker with tigr version5.0, tigr version6.0, tigr version6.1, tigr mitochondrial, and tigr chloroplast.
 - Homo sapiens:** Genome sets available include Homo sapiens Feb. 2009 (hg19) (selected), Mar.2006 (hg18), May.2004 (hg17), NCBI build 36.1_CRA, NCBI build 36.1_Celera, NCBI build 36.1_ref, NCBI build 36.2_CRA, NCBI build 36.2_Celera, NCBI build 36.2_ref, NCBI build 36.3_CRA, NCBI build 36.3_Celera, NCBI build 36.3_ref, NCBI build 36.3_HuRef, NCBI build 37.1_CRA, NCBI build 37.1_Celera, NCBI build 37.1_GRCh, and NCBI build 37.1_HuRef.

スパコン利用申請はこちちら

無料です！

- [DDBJ スーパーコンピュータ] で検索



大学共同利用法人 情報・システム研究機構 国立遺伝学研究所
スーパーコンピュータシステム
SuperComputer Facilities of National Institute of Genetics

サイトポリシー サイトマップ

検索...

2014年07月24日

現在地: Home

Language/言語

» ● »

このサイトへのログイン

Login
(スパコンユーザでログイン可)

システム構成

ハードウェア構成

ソフトウェア構成

プログラミング環境

利用可能なオフィツル

利用可能なOSS

利用可能なDB

システム使用方法

基本的な利用方法

その他UGE利用方法

ファイル転送方法

システム利用TIPS

移動スケジュール

各種申請

各種申請について

新規ユーザー登録申請

入室権利申請

MIGAP実用申請

重要なお知らせ

一覧へ

公開日	表題
2014年7月22日	【スパコンユーザ会】三島開催 会場変更のお知らせ
2014年7月18日	ソフトウェインストールのお知らせ
2014年7月15日	【fat障害: 2】fat計算ノードサービス再開のお知らせ
2014年7月14日	【通信障害: 2】Phase1システム通信不具合の復旧のお知らせ
2014年7月14日	【通信障害】Phase1システム通信不具合のお知らせ
2014年3月4日	2014年3月5日からのスパコンPhase2システムご利用方法について

国立遺伝学研究所 スーパーコンピュータシステム(NIG SUPERCOMPUTER)とは

大学共同利用機関法人 情報システム研究機構 国立遺伝学研究所は、2012年3月にスーパーコンピュータシステムを更新しました。新しいスーパーコンピュータシステムはゲノム解析を主な目的とした大規模計算機利用拠点として 最新鋭の大規模クラスタ型計算機、大規模メモリ共有型型計算機、および大容量高速ディスク装置で構成されたスーパーコンピューティングシステムサービスを提供しています。



- » システムハードウェア構成
- » システムソフトウェア構成
- » システム稼働状況

本サイトは国立遺伝学研究所スーパーコンピュータシステムが提供する計算機リソース、各種アプリケーション、それらの利用方法についての各種情報を提供します。DDBJセンターとして提供する各種サービスについてはDDBJセンターのホームページからご参照ください。

ディスク利用状況

lustreのsize,file項目は、現在のディスク使用量（全ユーザ合計）／ディスク容量で表現しています。
quotaのsize項目は、申請ディスク使用量（全ユーザ合計）／ディスク容量で表現しています。
sizeの単位は、Tbyteです。

- 遺伝研スパコンの過去の講習会資料（利用詳細）
 - http://sc.ddbj.nig.ac.jp/images/stories/meetingdoc/20120510/ja/ja_Basic_usage-1.pdf
- Unix / Linux の生き立ち・思想に触れる書籍
 - UNIXという考え方—その設計思想と哲学（www.amazon.co.jp/dp/4274064069/）
 - それがぼくには楽しかったから（www.amazon.co.jp/dp/4796880011）リーナス・トーバルス（Linux 作者）の自著
 - ハッカーズ（[www.amazon.co.jp/dp/487593100X/](http://www.amazon.co.jp/dp/487593100X)）コンピュータ黎明期からのハッカーの歴史・生態