

Nicotiana tabacum

Zea mays

Homo sapiens

Rattus norvegicus



1-1: OS, ハード構成

1-2: ネットワーク基礎

@東大農 Sep. 1, 2014

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

中村 保一

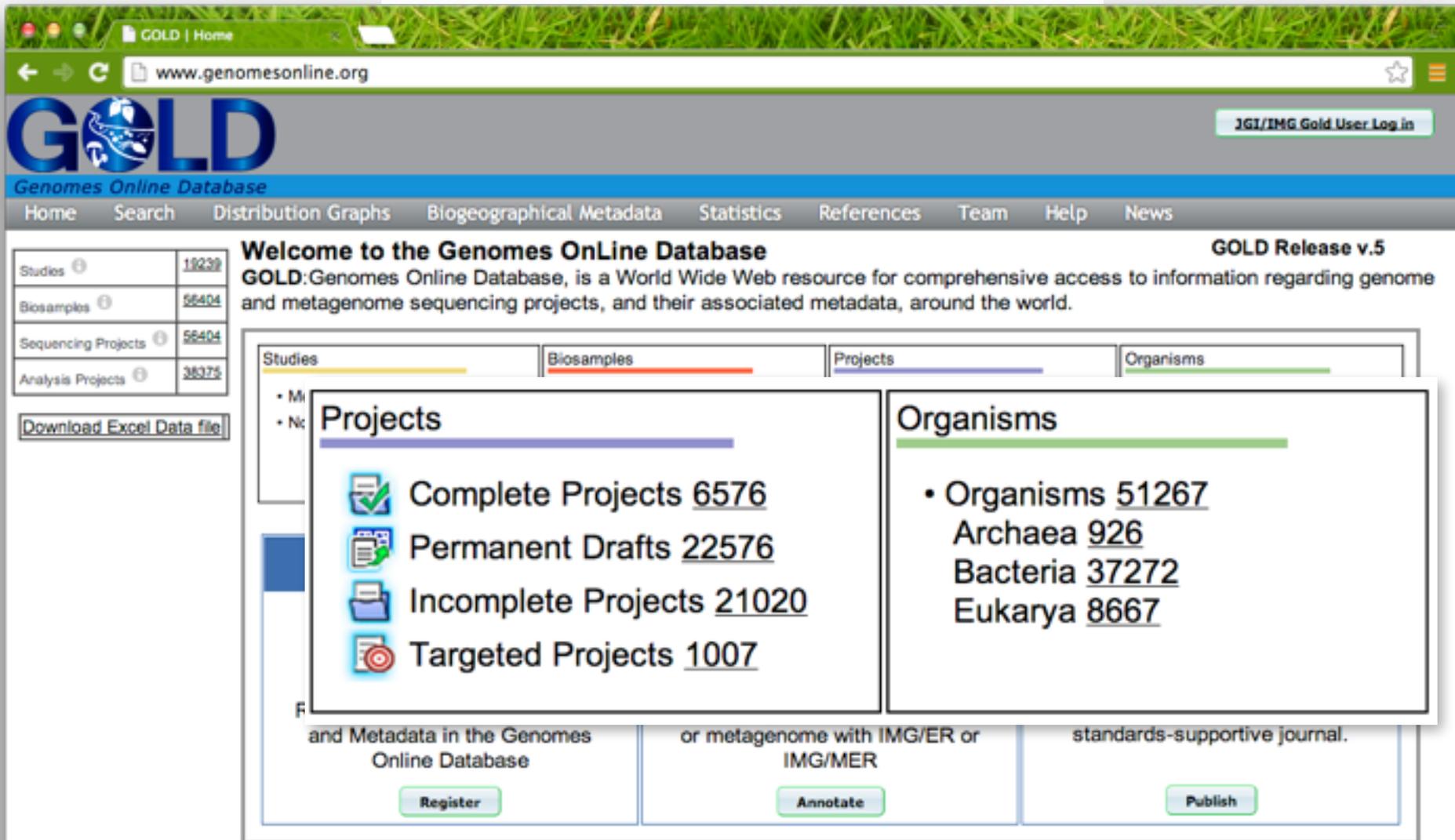
NAKAMURA Yasukazu, Professor

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

今や膨大なゲノム情報が入手可能な訳で

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



The screenshot shows the homepage of the Genomes Online Database (GOLD). The browser address bar displays www.genomesonline.org. The page features a navigation menu with links for Home, Search, Distribution Graphs, Biogeographical Metadata, Statistics, References, Team, Help, and News. A sidebar on the left provides statistics for Studies (19239), Biosamples (56404), Sequencing Projects (56404), and Analysis Projects (38375), along with a 'Download Excel Data file' button. The main content area is titled 'Welcome to the Genomes OnLine Database' and includes a 'GOLD Release v.5' label. It contains a summary of project types: Complete Projects (6576), Permanent Drafts (22576), Incomplete Projects (21020), and Targeted Projects (1007). A separate section lists the number of organisms: Archaea (926), Bacteria (37272), and Eukarya (8667). At the bottom, there are three buttons: 'Register', 'Annotate', and 'Publish'.

Studies	Biosamples	Projects	Organisms
19239	56404	6576 (Complete) 22576 (Permanent Drafts) 21020 (Incomplete) 1007 (Targeted)	51267 (Total) 926 (Archaea) 37272 (Bacteria) 8667 (Eukarya)

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



<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



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

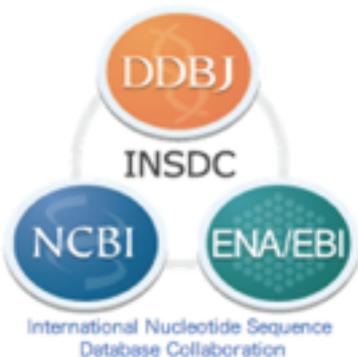
光合成細菌のゲノム解析+データベース

現在はDDBJのDB構築担当教員です

 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

▶ 一覧

- ▶ 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)

Maintenance

▶ 一覧

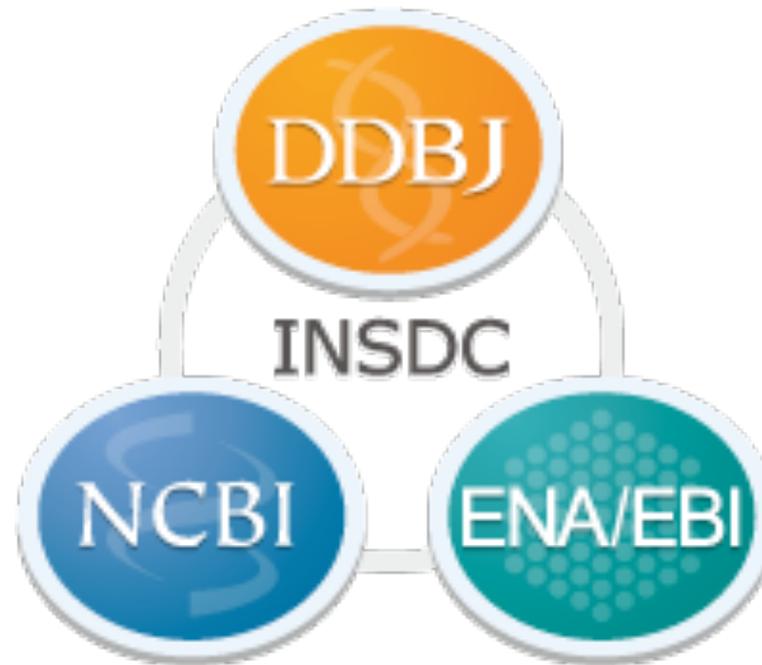
Information

塩基配列データベース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



コンテンツ

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>

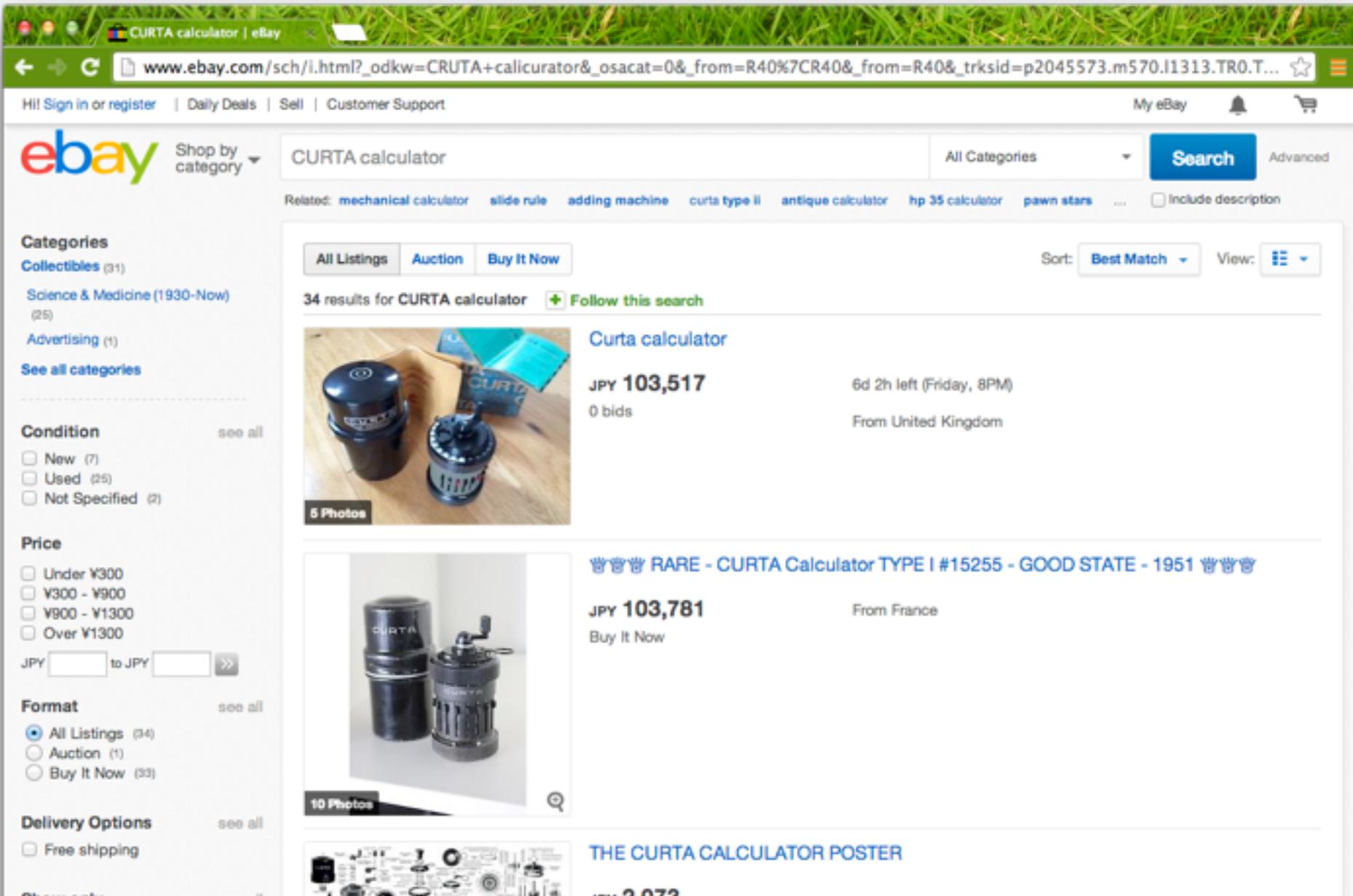
- (1) 数値の保持
- (2) 加算・減算の機能
- (3) 数値の表示

機械式計算機（詳細はリンク先をどうぞ）

- 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…



The screenshot shows an eBay search results page for 'CURTA calculator'. The browser address bar shows the URL: www.ebay.com/sch/l.html?_odkw=CRUTA+calicurator&_osacat=0&_from=R40%7CR40&_from=R40&_trksid=p2045573.m570.l1313.TR0.T... The page header includes navigation links like 'Sign in or register', 'Daily Deals', 'Sell', and 'Customer Support'. The search bar contains 'CURTA calculator' and a 'Search' button. Below the search bar, there are filters for 'Categories', 'Condition', 'Price', and 'Format'. The main content area displays search results for 'CURTA calculator', showing 34 results. The first result is a 'Curta calculator' for JPY 103,517, with 0 bids and 6 days 2 hours left. The second result is a 'RARE - CURTA Calculator TYPE I #15255 - GOOD STATE - 1951' for JPY 103,781, available for 'Buy It Now'. The third result is 'THE CURTA CALCULATOR POSTER' for JPY 2,072.

CURTA calculator | eBay

www.ebay.com/sch/l.html?_odkw=CRUTA+calicurator&_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

Shop by category

CURTA calculator

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-Present) (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

34 results for CURTA calculator [Follow this search](#)

All Listings Auction Buy It Now

Sort: Best Match View:

 5 Photos

Curta calculator

JPY 103,517

0 bids

6d 2h left (Friday, 8PM)

From United Kingdom

 10 Photos

👑👑👑 RARE - CURTA Calculator TYPE I #15255 - GOOD STATE - 1951 👑👑👑

JPY 103,781

Buy It Now

From France



THE CURTA CALCULATOR POSTER

JPY 2,072

- 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 ⇒ 歯車式計算機の電子版
 - プログラム内蔵式 ではない



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

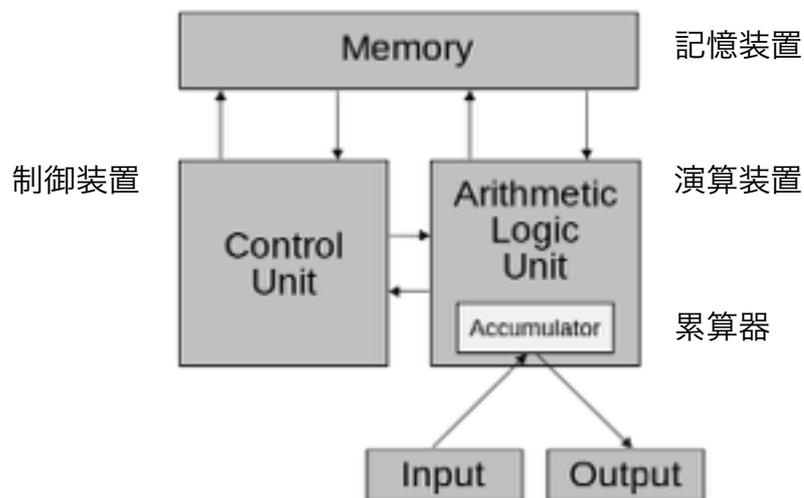
ノイマン型 計算機



[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
 - 結線では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）

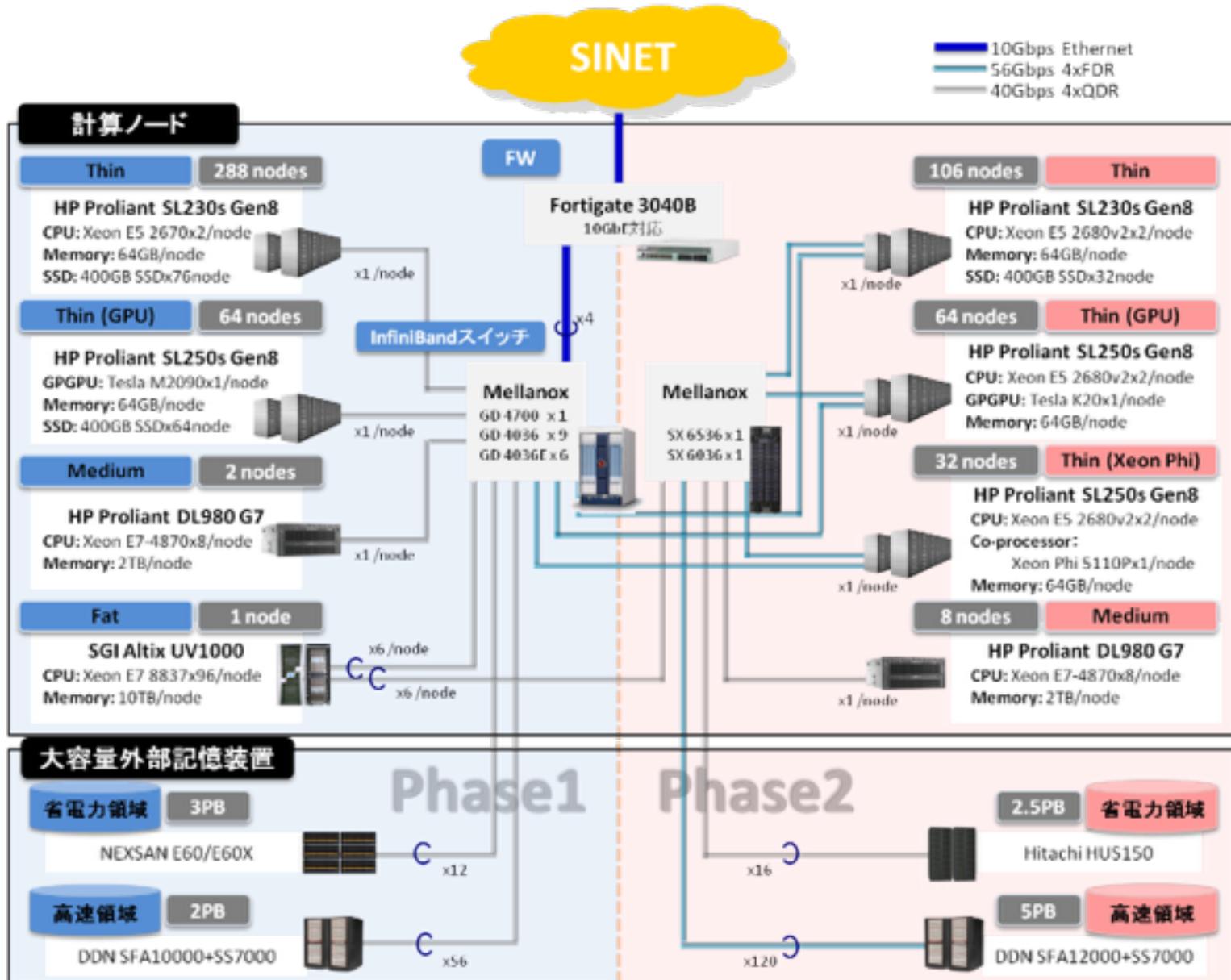
遺伝研スーパーコンピュータを例に

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

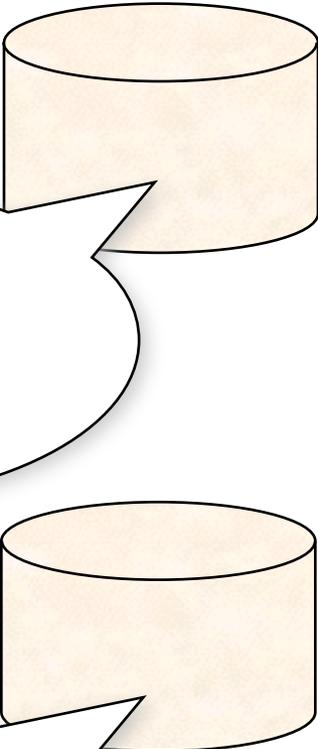


The screenshot shows the website for the National Institute of Genetics (NIG) SuperComputer. The browser address bar shows sc.ddbj.nig.ac.jp. The page header includes the NIG logo and the text "大学共同利用法人 情報・システム研究機構 国立遺伝学研究所 スーパーコンピュータシステム SuperComputer Facilities of National Institute of Genetics". There are navigation links for "サイトポリシー" and "サイトマップ", and a search box. The main content area is titled "重要なお知らせ" (Important News) and lists several announcements with dates and titles, such as "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". Below this 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" (About NIG SuperComputer System), which describes the system's capabilities for genome analysis. At the bottom, there are three images of server racks, with a large red arrow pointing to the middle image. The left sidebar contains navigation menus for "Language/言語", "このサイトへのログイン", "システム構成" (System Configuration), and "システム使用方法" (System Usage).

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)



7 PB
Lustre
高速HDD

5.5 PB
MAID
大容量省電力HDD



“thin”
64GB memory
x 554 台

“medium”
2TB memory
x 10 台

“fat”
10TB memory
(SGI UV) 1 台

コンテンツ

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った DEC-10 のかわりに部屋のスミに転がっていた借り物の PDP-7 で「えいやっ」と作ってみた「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





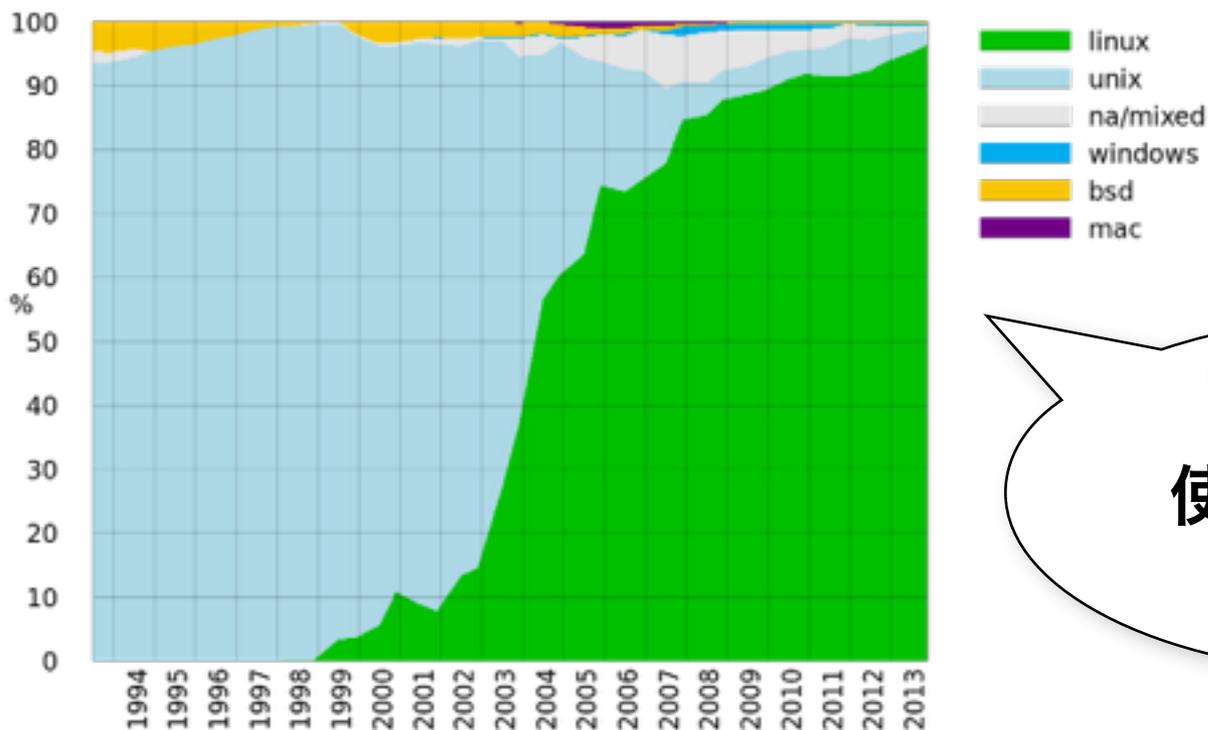
iOS の基礎部分は Darwin であり、NeXTSTEP を先祖に持つ由緒正しき Berkeley Software Distribution (BSD) UNIX の系譜に連なります。

要するに組み込み系の UNIX (POSIX 準拠) で動作。

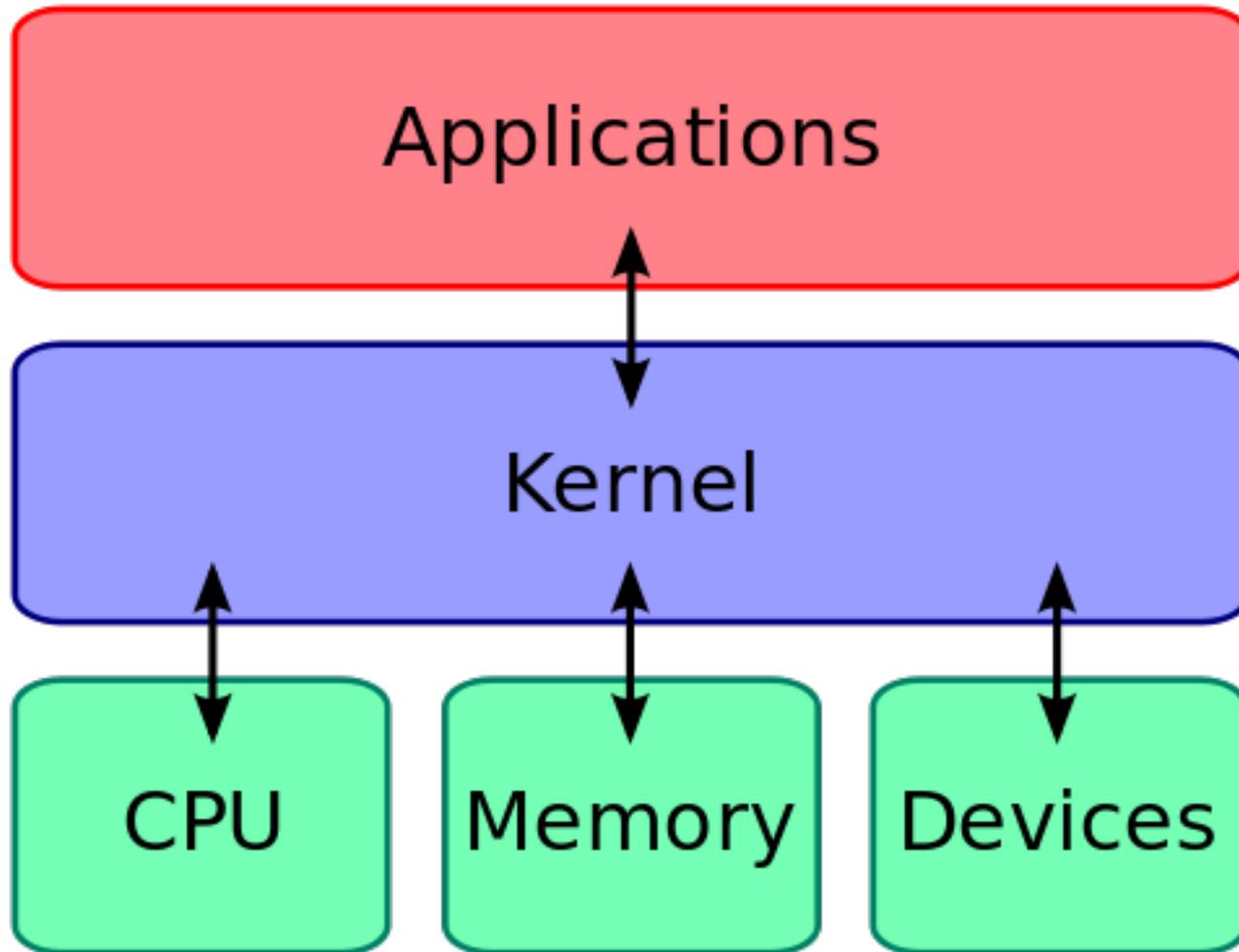
MacOS X も勿論 UNIX。

スパコンの OS だって

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

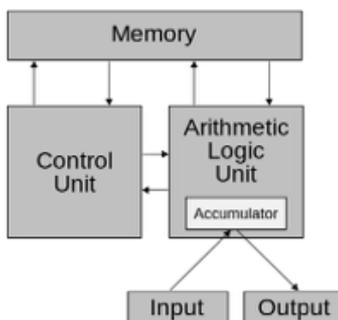


Top 500 で
使われている
OS の推移



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

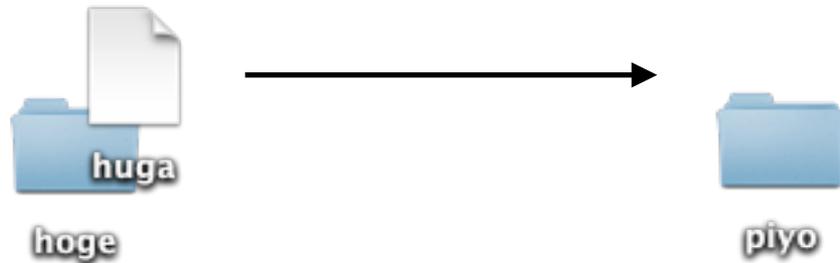


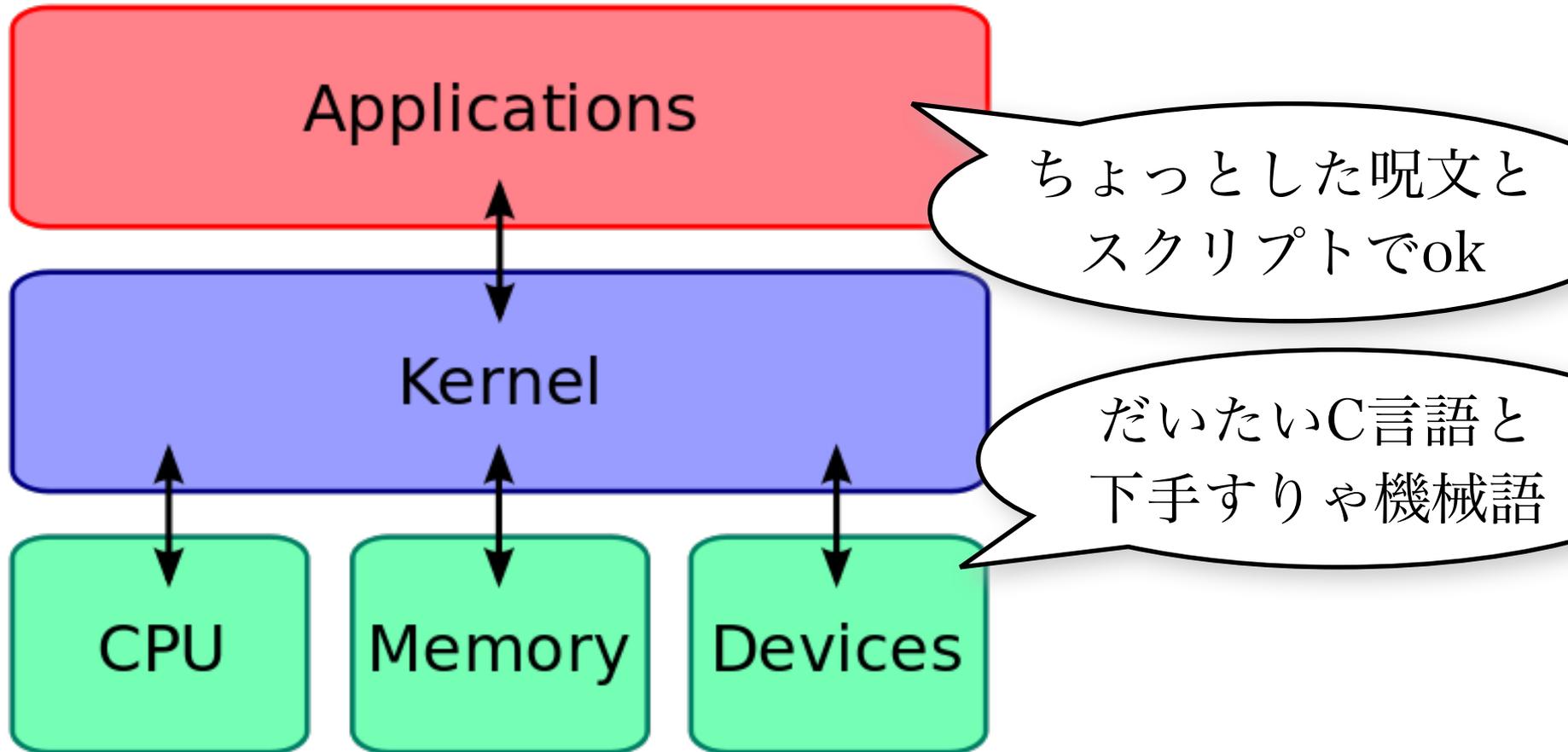
[Von Neumann architecture CC BY-SA 3.0](#)

- 「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、連続でとれなかっ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

僕ら対話するのは「アプリ（シェルを含む）」の層

コンテンツ

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 website for the National Institute of Genetics (NIG) Supercomputer. The browser address bar shows <https://sc.ddbj.nig.ac.jp/>. The page header includes the NIG SUPERCOMPUTER logo and 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". There are links for "サイトポリシー" and "サイトマップ", and a search box. The main content area is titled "重要なお知らせ" (Important News) and lists several announcements with dates, such as "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". Below this 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" (About NIG SUPERCOMPUTER), which describes the system's capabilities for genome analysis. A large red arrow points to a row of server racks in the background image. The left sidebar contains navigation menus for "Language/言語", "このサイトへのログイン", "システム構成" (System Configuration), and "システム使用方法" (System Usage).

- 「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](#), [tcsh](#), [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捗るようになりますよ。

スクリプト言語

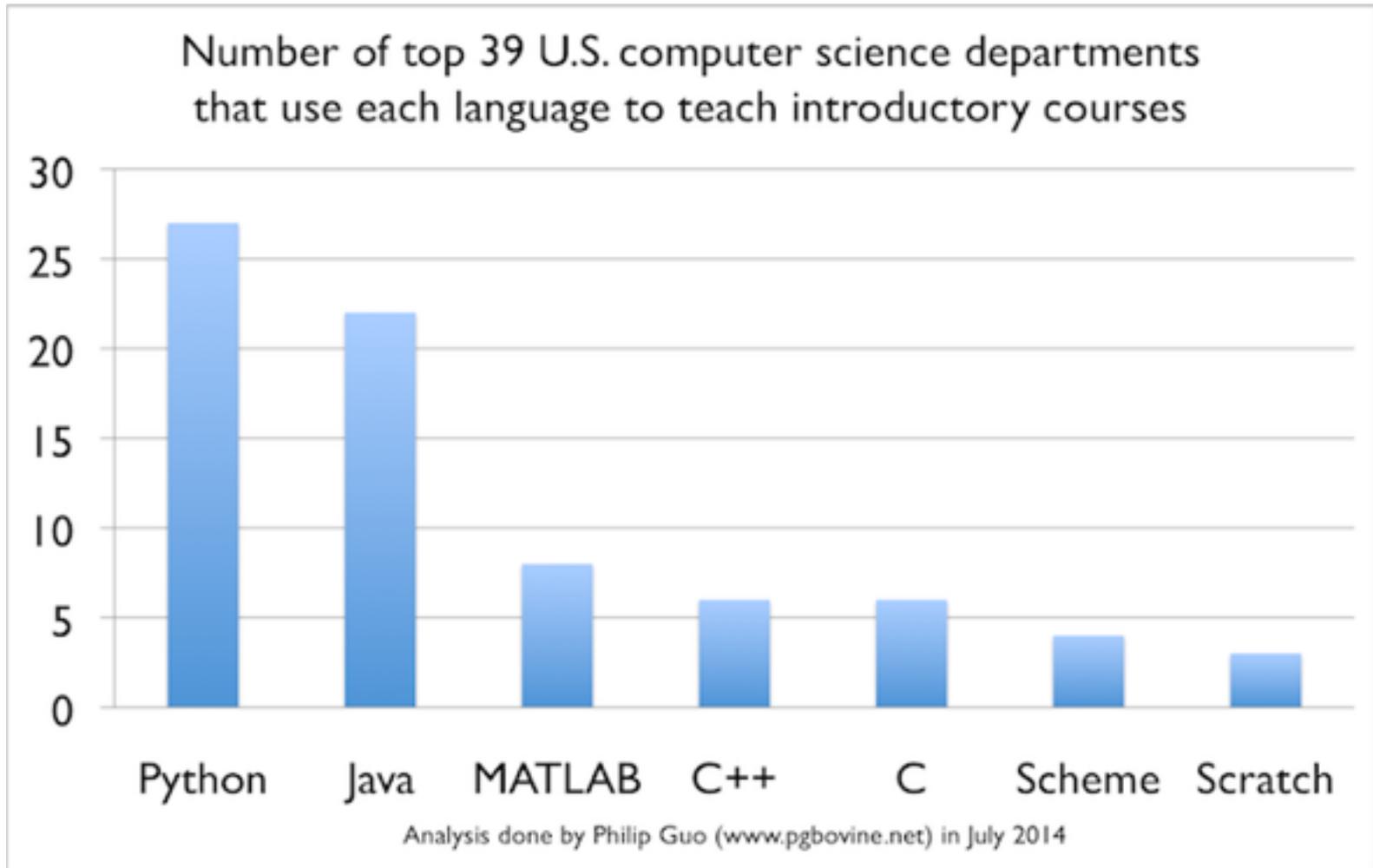
- 「一般ユーザが利用可能なスクリプト言語」
 - ログインノード、計算ノードで一般ユーザが利用可能なスクリプト言語環境は以下の通りです。

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>

コンテンツ

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

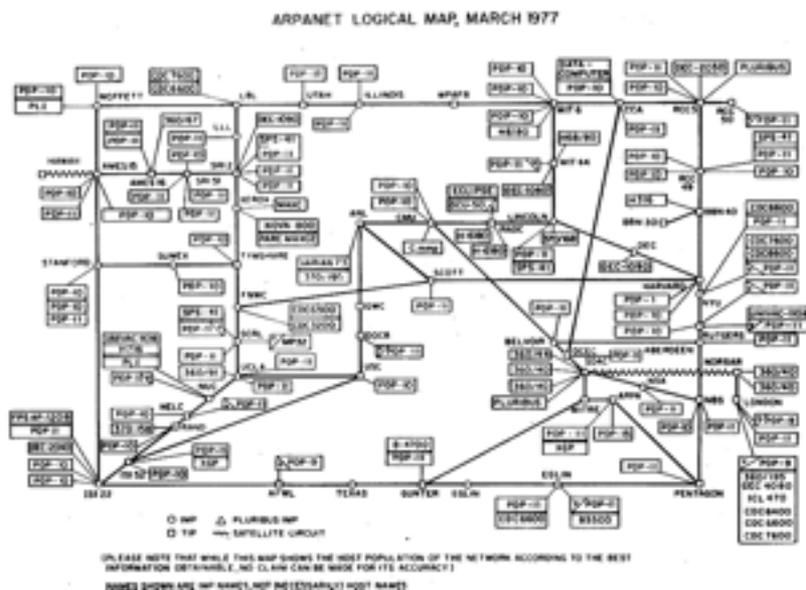
- a) オペレーティング・システム (OS)
- b) 言語・プログラム・ライブラリ

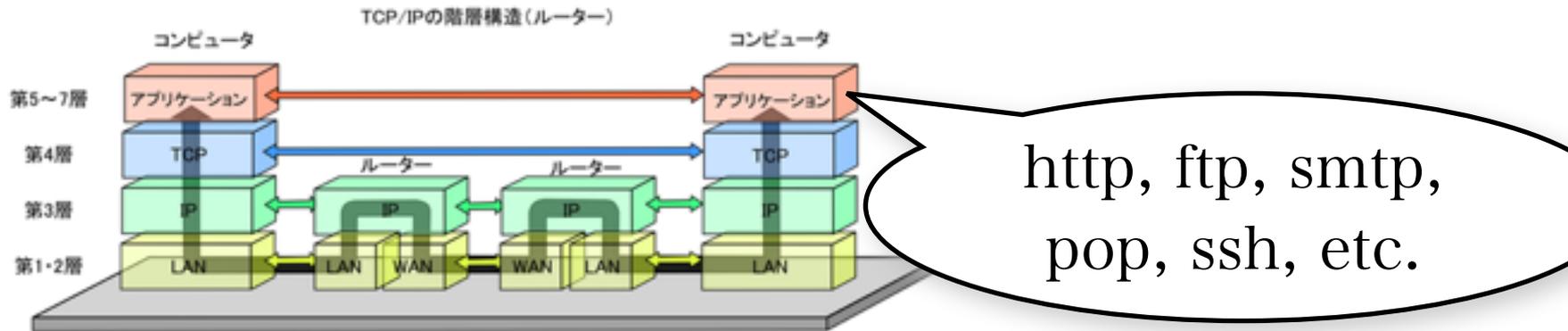
ネットワーク

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





- 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/>

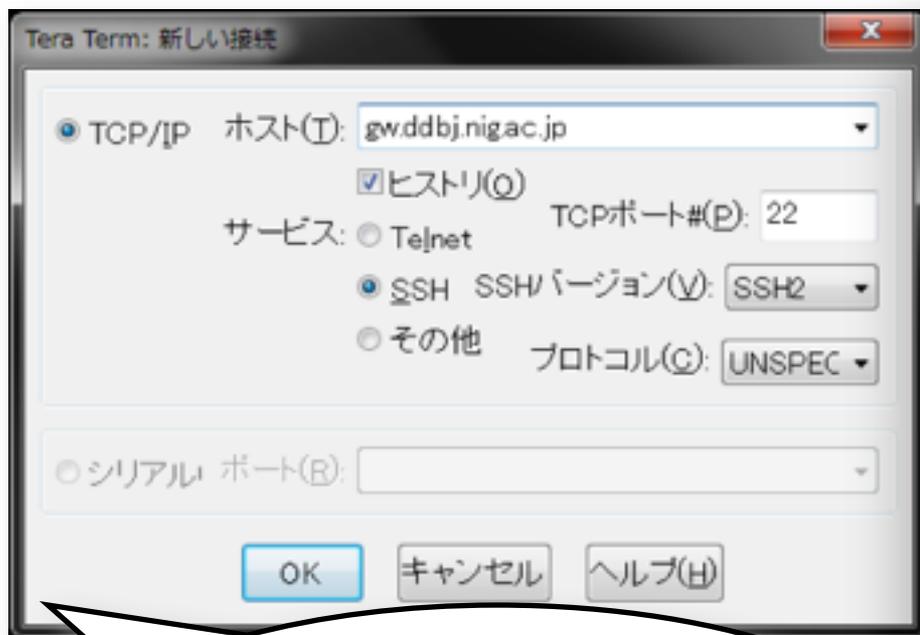


The screenshot shows the website for the National Institute of Genetics (NIG) SuperComputer. The browser address bar displays <https://sc.ddbj.nig.ac.jp/>. The page header includes the NIG logo and the text "大学共同利用法人 情報・システム研究機構 国立遺伝学研究所 スーパーコンピュータシステム SuperComputer Facilities of National Institute of Genetics". A search bar is located in the top right corner. The main content area features a "Language/言語" dropdown menu, a "ホーム" link, and a "このサイトへのログイン" section with a "Login" button. Below this is a "システム構成" (System Configuration) section with links for hardware, software, programming environment, and tools. A "システム使用方法" (System Usage) section is also present. A large red arrow points to the "システム構成" section. The "重要なお知らせ" (Important Notices) section lists several announcements from 2014, including system updates and service changes. The "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" (About NIG SuperComputer System) section provides a brief overview of the system's capabilities and update history. The bottom of the page features a grid of images showing server racks and a person working at a computer terminal.

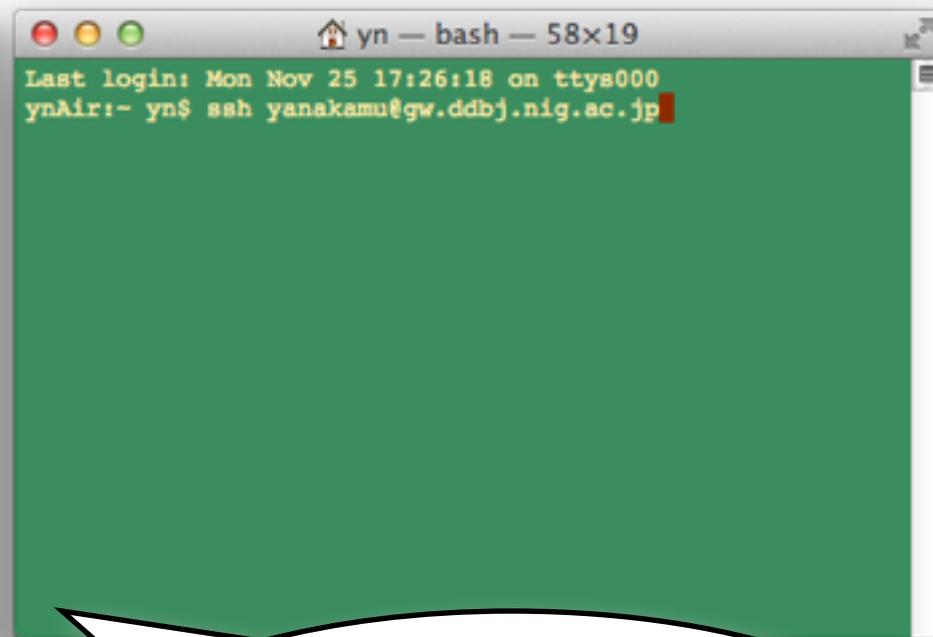
スパコンへのログイン (1)

- Windowsから
- Tera Term 起動

- Macから
- ターミナル 起動



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

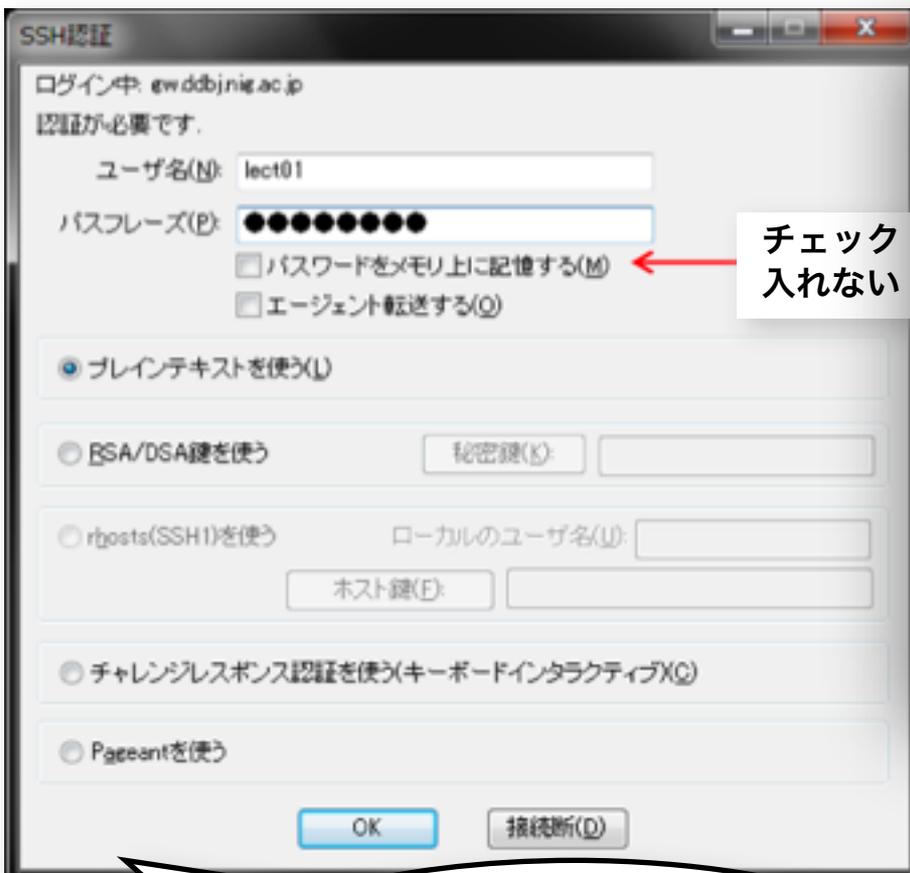


ssh **ユーザ名**@gw.ddbj.nig.ac.jp
 ⇒ Enter

スパコンへのログイン (2)

• Windowsから

• Macから



SSH認証

ログイン中: gwddbjinig.ac.jp

認証が必要です。

ユーザ名(U): lect01

パスワード(P): ●●●●●●●●

パスワードをメモリ上に記憶する(M) ← チェック
入れない

エージェント転送する(Q)

プレインテキストを使う(L)

RSA/DSA鍵を使う 秘密鍵(K):

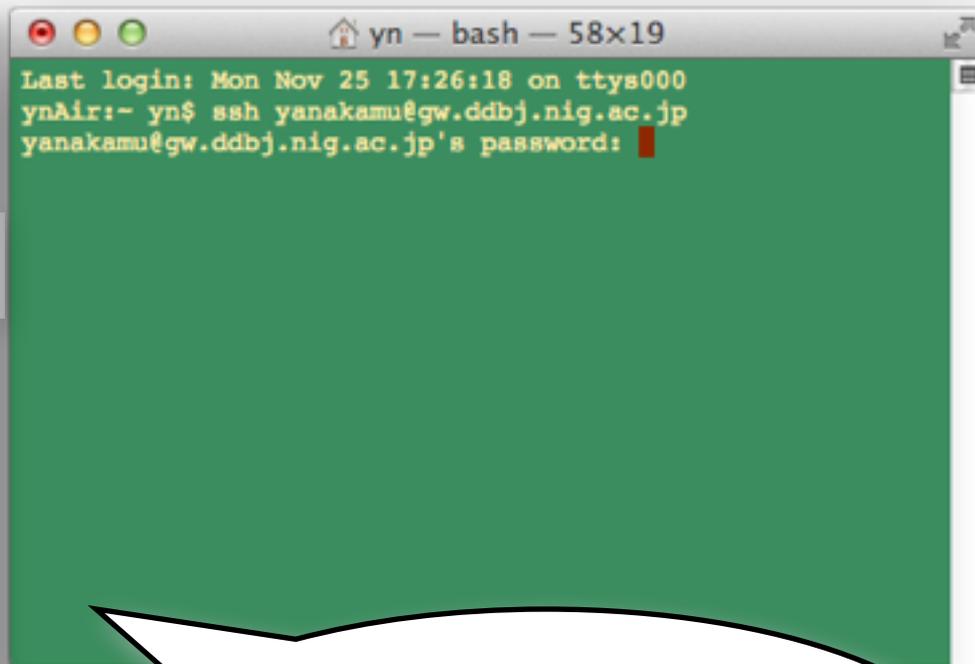
rhosts(SSH)を使う ローカルのユーザ名(U):

ホスト鍵(E):

チャレンジレスポンス認証を使う(キーボードインタラクティブ)(Q)

Pageantを使う

OK 接続断(D)



```
yn — bash — 58x19
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: 
```

パスワードを入力
⇒ Enter

ユーザ名とパスワードを入力
⇒ OK

スパコンへのログイン (3)

- Windowsから

- Macから

```

133.39.224.14.22 - New Termi...
コマンドプロンプト
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 ~]$
    
```

```

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 をする必要があります)

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

```
yn - bash - 117x12
iMac:~ yn$ sftp yanakamu@gw.ddbj.nig.ac.jp
yanakamu@gw.ddbj.nig.ac.jp's password:
```

または



[Cyberduck](#)

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

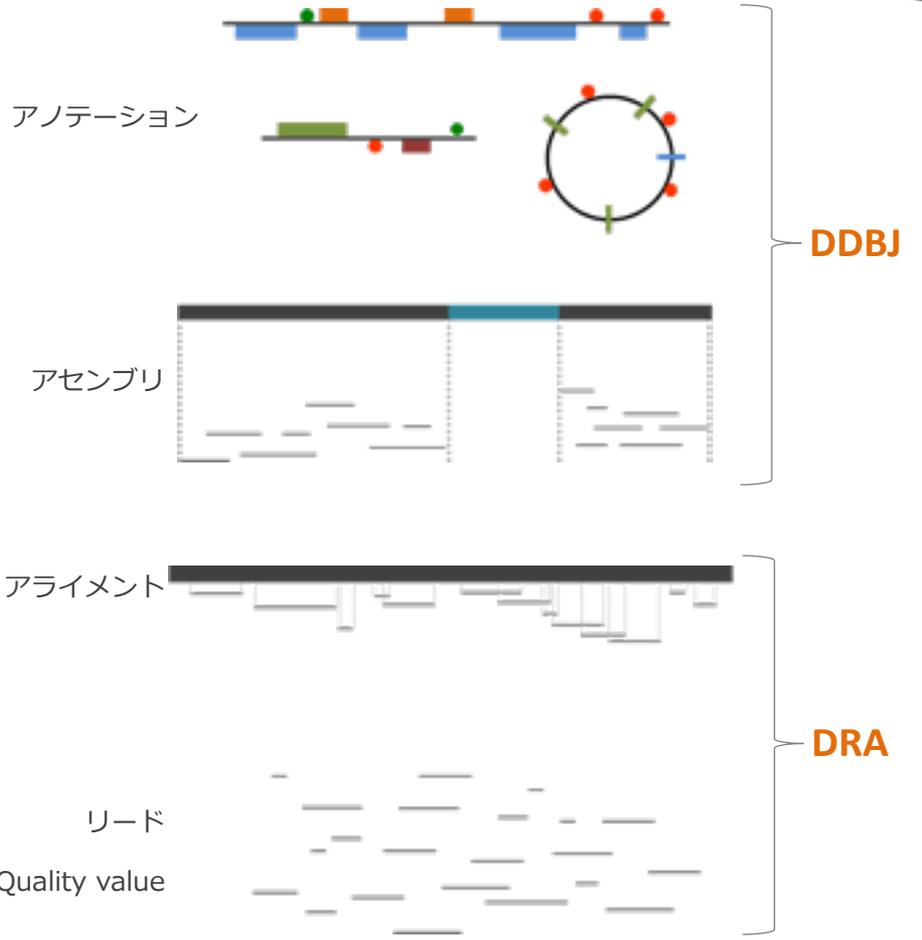
Cyberduck のような専用クライアントを使えばドラッグ&ドロップでファイル転送も。コマンドでの送受信もたいして難しくありませんが。

- 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

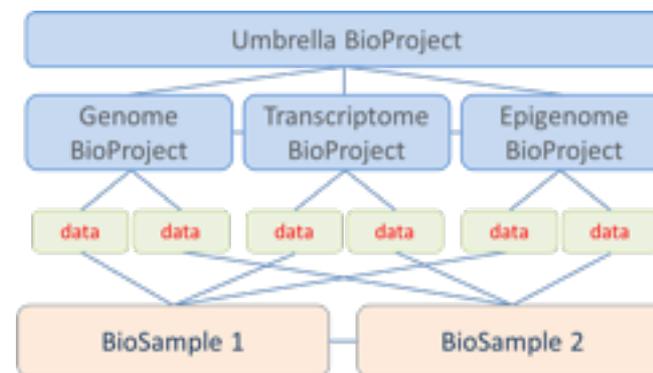
個人の遺伝型と表現型



ヒトデータ審査委員会



BioProject BioSample



DRA: DDBJ Sequence Read Archive

JGA: Japanese Genotype-phenotype Archive

INSDC: オープンアクセス

DDBJ Sequence Read Archive (DRA)

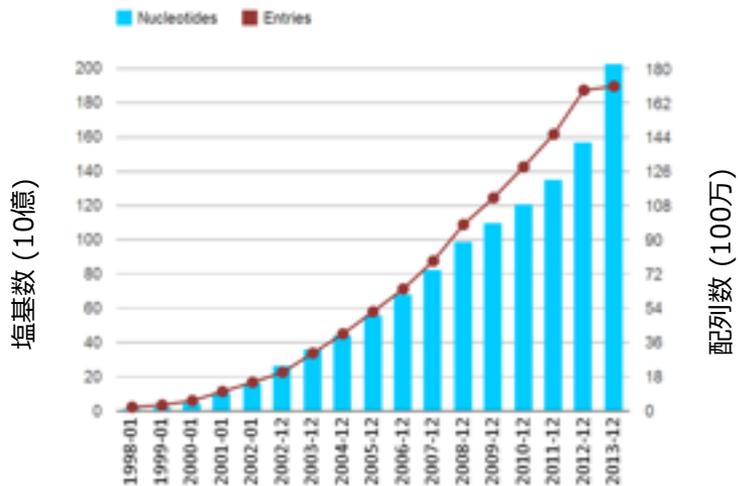
新世代シーケンサから出力される配列や
アライメントデータを登録・公開



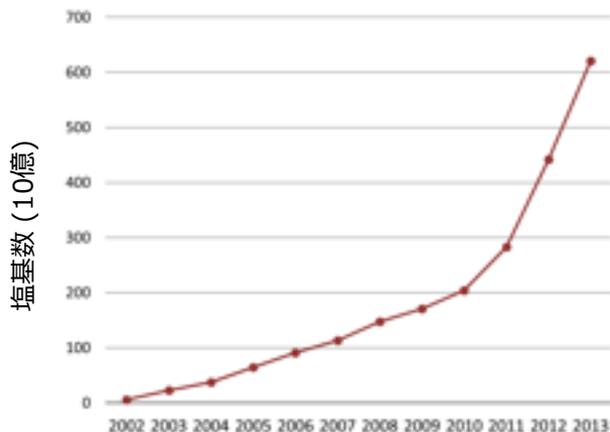
データ量の推移

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

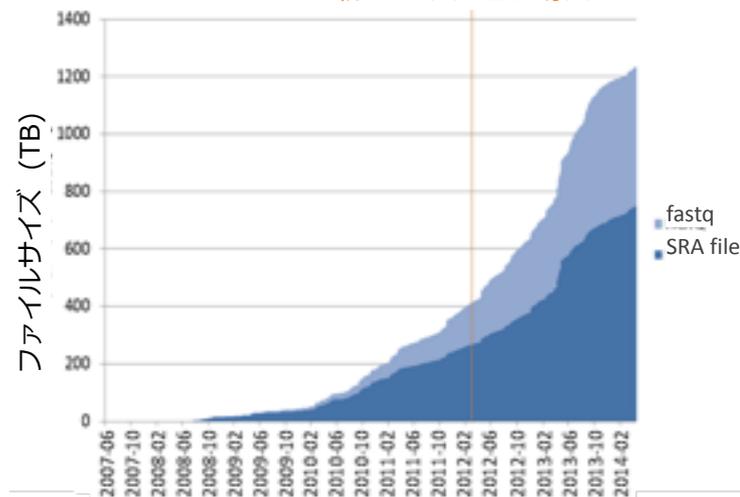


- 総計 8,200 億塩基 (30 万生物種)

SRA データベースの成長

次世代シーケンシングデータ、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



The screenshot shows the DDBJ Sequence Read Archive website. At the top left is the DDBJ logo. To the right are links for 'Login & Submit', 'Databases', 'English', and 'Contact'. Below these is a search bar with the text 'Google™ カスタム検索' and a magnifying glass icon. A navigation menu contains 'Home', 'Submission', 'Search', 'Download', 'Pipeline', and 'About'. The 'Pipeline' link is highlighted with a yellow callout box containing the text '解析パイプライン'. Below the navigation menu is a paragraph of text describing the archive, with two yellow callout boxes: 'データ検索' pointing to the word '検索' and 'データ取得' pointing to the word '取得'. At the bottom, there are three large buttons: '検索' (Search) with a magnifying glass icon, '登録' (Registration) with a database icon, and '動画マニュアル' (Video Manual) with a video camera icon. Each button has a short description below it.

Sequence Read Archive

Home Submission Search Download Pipeline About

解析パイプライン

登録関係情報

データ検索

データ取得

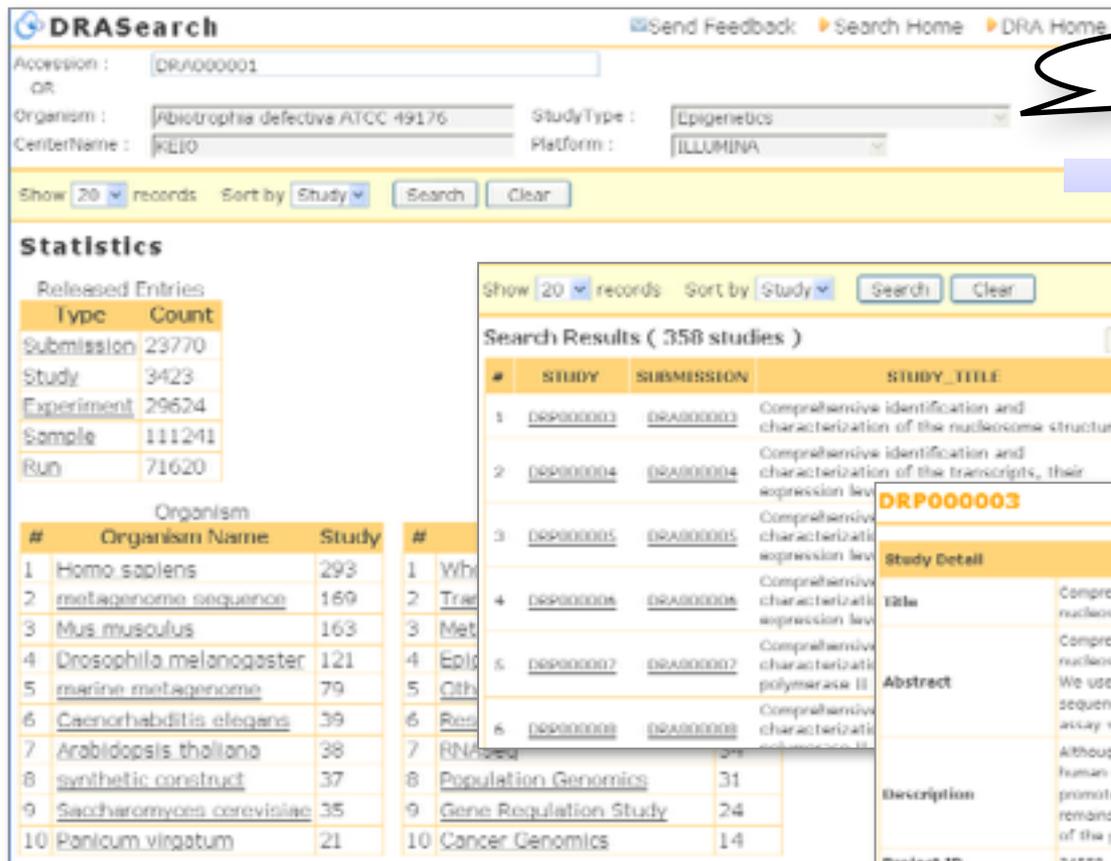
検索
データをキーワード、生物名、シーケンサなどで検索する

登録
新型シーケ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



DRASearch | Send Feedback | Search Home | DRA Home

Accession:
OR
Organism: StudyType:
CenterName: Platform:

Show records Sort by Search Clear

Statistics

Released Entries

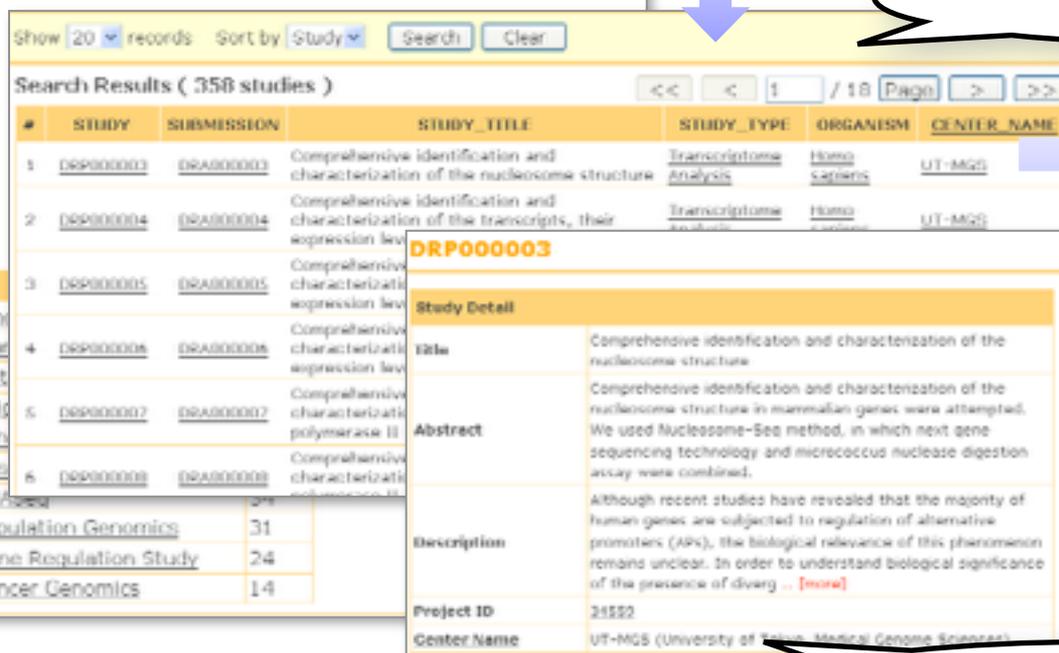
Type	Count
Submission	23770
Study	3423
Experiment	29624
Sample	111241
Run	71620

Organism

#	Organism Name	Study
1	Homo sapiens	293
2	metagenome sequenc	169
3	Mus musculus	163
4	Drosophila melanogaster	121
5	marine metagenome	79
6	Caenorhabditis elegans	39
7	Arabidopsis thaliana	38
8	synthetic construct	37
9	Saccharomyces cerevisiae	35
10	Panicum virgatum	21

生物名 etc での絞り込み

検索結果リスト



Show records Sort by Search Clear

Search Results (358 studies)

#	STUDY	SUBMISSION	STUDY_TITLE	STUDY_TYPE	ORGANISM	CENTER_NAME
1	DRP000003	DRA000003	Comprehensive identification and characterization of the nucleosome structure	Transcriptome Analysis	Homo sapiens	UT-MGS
2	DRP000004	DRA000004	Comprehensive identification and characterization of the transcripts, their expression lev	Transcriptome Analysis	Homo sapiens	UT-MGS
3	DRP000005	DRA000005	Comprehensive characterizat			
4	DRP000006	DRA000006	Comprehensive characterizat			
5	DRP000007	DRA000007	Comprehensive characterizat			
6	DRP000008	DRA000008	Comprehensive characterizat			

DRP000003

Study Detail

Title
Comprehensive identification and characterization of the nucleosome structure

Abstract
Comprehensive identification and characterization of the nucleosome structure in mammalian genes were attempted. We used Nucleosome-Seq method, in which next gene sequencing technology and microcococcus nuclease digestion assay were combined.

Description
Although recent studies have revealed that the majority of human genes are subjected to regulation of alternative promoters (APs), the biological relevance of this phenomenon remains unclear. In order to understand biological significance of the presence of diverg .. [\[more\]](#)

Project ID
24552

Center Name
UT-MGS (University of Tokyo, Medical Genome Research)

Navigation

- [SubmissionDRA000003](#) [FASTQ](#)
- [ExperimentDRP000003](#) [FASTQ](#) [SRALink](#)
- [Sample DRP000003](#)

ダウンロード

詳細 (メタデータ記述)

解析パイプラインも提供しています

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



Sequence Read Archive

Home

Submission ▾

Search

Download ▾

Pipeline

About

解析パイプライン

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シーケンサからの出力データのためのデータベースです。DRA は International Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、NCBI Sequence Read Archive (SRA) と EBI Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャピラリー式シーケンサからの出力データは DDBJ Trace Archive にご登録ください。



検索

データをキーワード、生物名、シーケンサなどで検索する



登録

新型シーケ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

DRA pipeline: ソフトウェア

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

DDBJ
DNA Data Bank of Japan

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.



Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK NEXT

Reference Genome Mapping

	Tool	Help	Version	Input data			Evaluation			Analysis		Output format			Comment
				Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM	
<input type="checkbox"/>	BLAT	Help	34	✓						✓					Single-end analysis only
<input type="checkbox"/>	Maq	Help	0.7.1	✓		✓				✓	✓	✓	✓	✓	
<input type="checkbox"/>	bwa	Help	0.5.9	✓		✓				✓				✓	
<input type="checkbox"/>	SOAP	Help	2.21	✓		✓				✓	✓			✓	
<input type="checkbox"/>	Bowtie	Help	0.12.7	✓	✓	✓				✓	✓			✓	
<input type="checkbox"/>	TopHat	Help	1.0.11	✓		✓				✓				✓	
<input type="checkbox"/>	Bowtie2	Help	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	Help	1.05			✓		
<input type="checkbox"/>	ABySS	Help	1.3.2			✓		Maximum K-mer value is 64.
<input type="checkbox"/>	Velvet	Help	1.2.03			✓	✓	We severe recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.

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 displays the DDBJ DRA pipeline interface. The main content area is titled "Specifying Database of Ref" and contains a "Major genome sets" section. This section has two dropdown menus: "Organisms" and "Genome sets". The "Organisms" dropdown is currently open, showing a list of species including Arabidopsis thaliana, 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. A red arrow points to this dropdown. The "Genome sets" dropdown is also open, showing a list of genome builds for the selected organism, including TAIR8, TAIR9, and TAIR10. The interface also includes a "User original sets" section and a "Download or upload reference" section. The left sidebar contains navigation links for ACCOUNT, ANALYSIS, and JOB STATUS. The top navigation bar shows the steps: Select Query Files, Select Tools, and Set QuerySet.

スパコン利用申請はこちら

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... 検索

現在地: Home 2014年07月24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UGL利用方法
ファイル転送方法
システム利用TIPS
稼働スケジュール
各種申請
新規ユーザ登録申請
大規模利用申請
MIGAP利用申請

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

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ー型計算機、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。

あわせてご参照ください

- 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 の生い立ち・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/)
コンピュータ黎明期からのハッカーの歴史・生態