

遺伝研大量研・DDBJセンター  
中村保一

[RSSを購読する](#)[DDBJ Twitter](#)

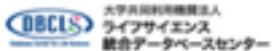
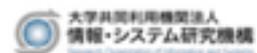
## DDBJ Service

[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](http://ddbj.nig.ac.jp)

## Hot Topics

[ニュースアーカイブ](#)[お知らせ](#) [データ公開](#) [スパコン](#) [メンテナンス](#) [運用情報](#) [一覧](#)

- 2014.04.21 [オオムギ \(\*Hordeum vulgare\* subsp. \*vulgare\*\) 由来 EST データの公開](#)
- 2014.04.18 [UniProt 2014\\_04 公開](#)
- 2014.04.18 [PDB 2014-04-12 公開](#)
- 2014.04.11 [Mass Submission System \(MSS\) 関連ツール更新のお知らせ](#)
- 2014.04.11 [getentry, ClustalW デザイン変更のお知らせ](#)
- 2014.04.10 [PDB 2014-04-05 公開](#)
- 2014.04.03 [PDB 2014-03-29 公開](#)
- 2014.04.02 [NRRP VM に関連した微生物ゲノム配列データの公開](#)



# 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



# 国際塩基配列データベースの一員

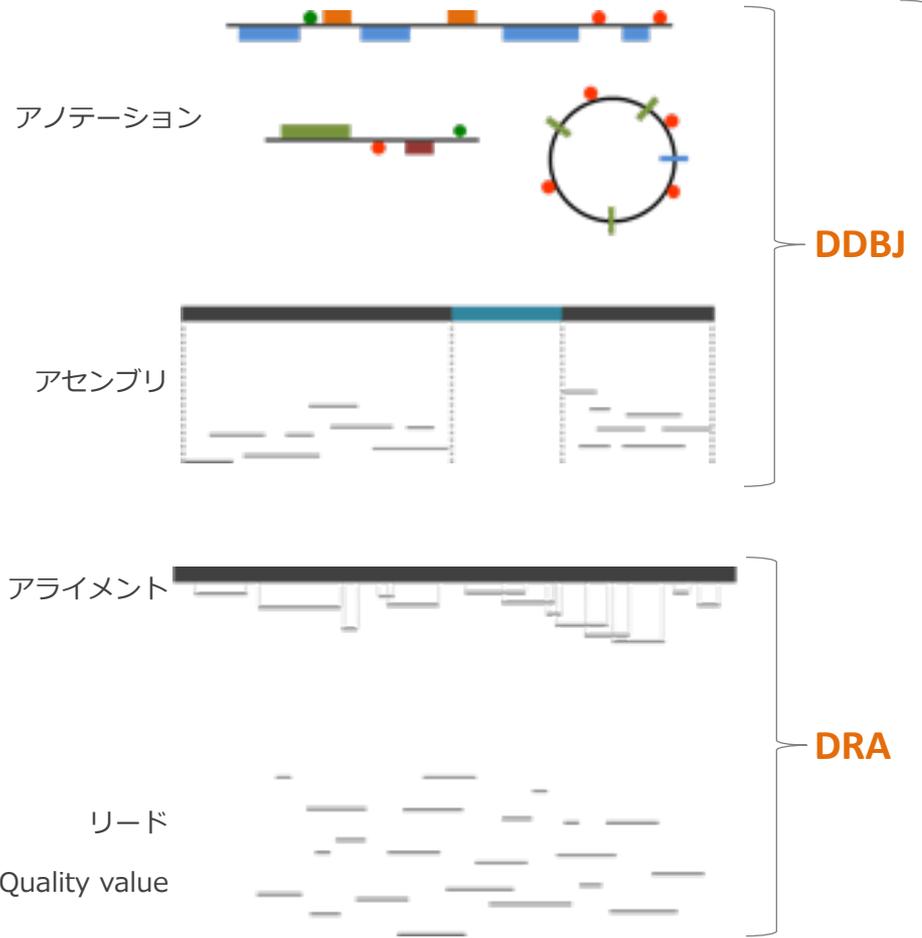
---

## International Nucleotide Sequence Databank Collaboration



# DDBJが運営しているデータベース

## DDBJ センター

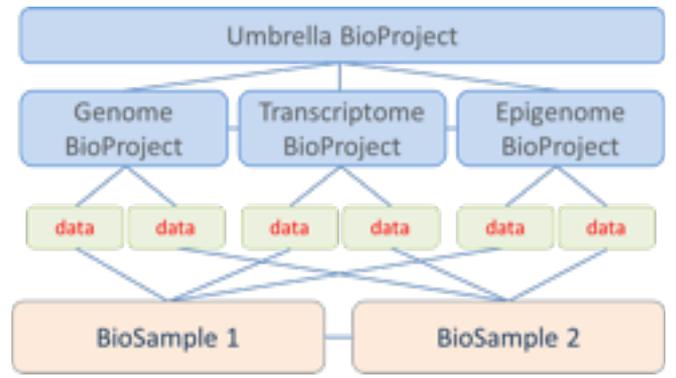


**アクセス制限**  
**JGA**  
個人の遺伝型と表現型

**NBDC**  
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

## 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

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



The screenshot shows the homepage of the NIG SuperComputer. The browser address bar displays [sc.ddbj.nig.ac.jp](http://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 on the right. The main content area features a "重要なお知らせ" (Important Notice) section with a list of dates and messages, and a "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. A large red arrow points to the "システム構成" (System Configuration) section in the left sidebar.

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 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タシステムサービスを提供しています。

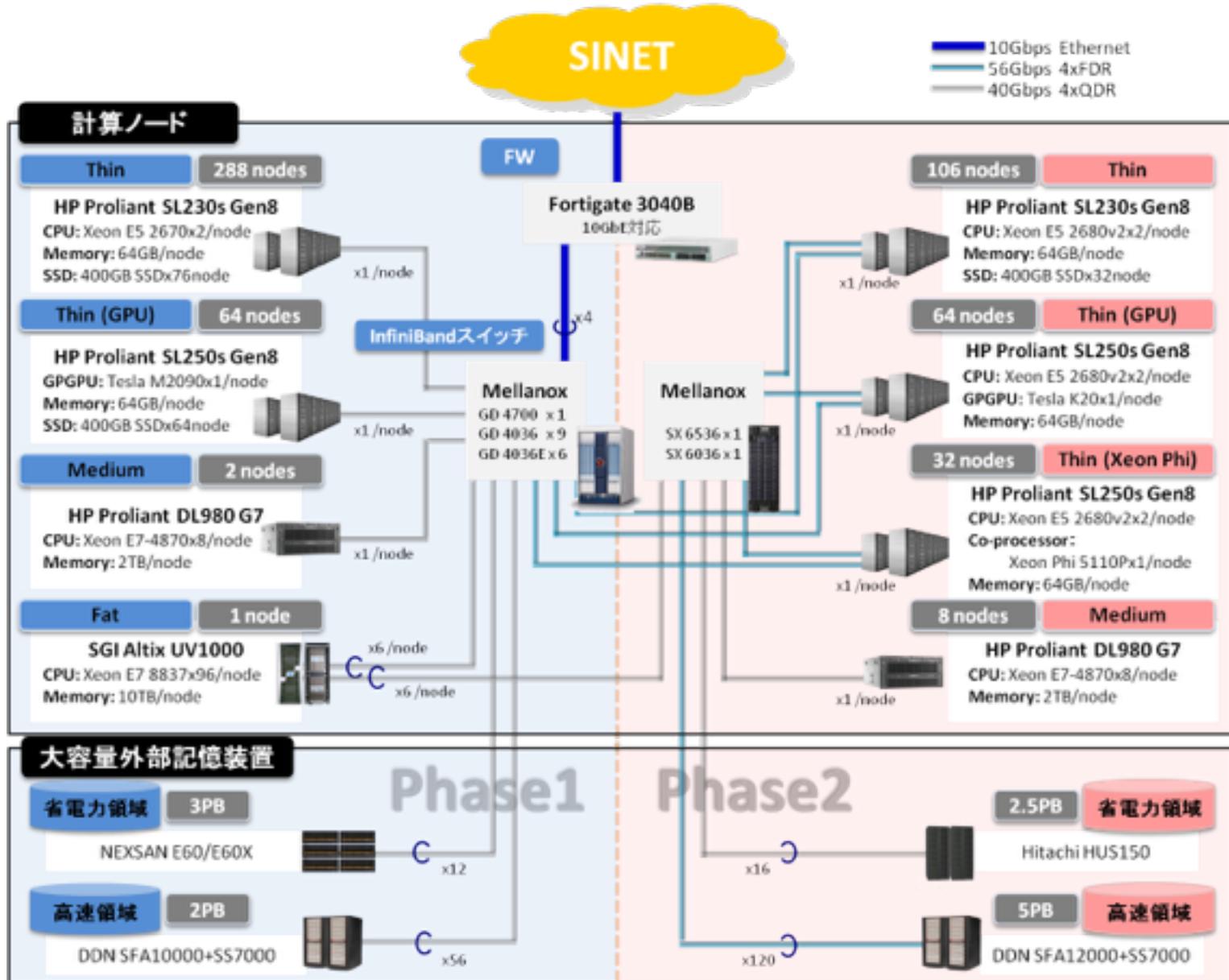


> システムハードウェア構成

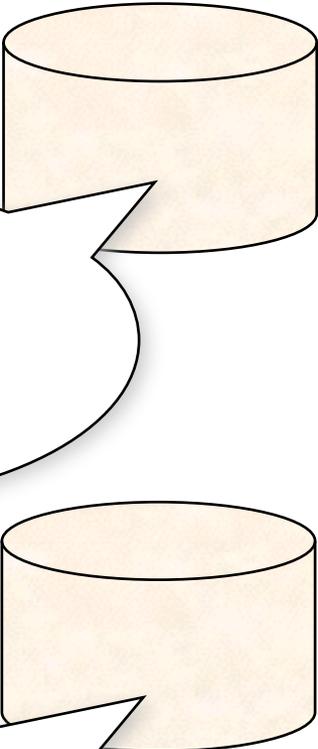
> システムソフトウェア構成

> システム稼働状況

# 遺伝研スーパーコンピュータ (全容)

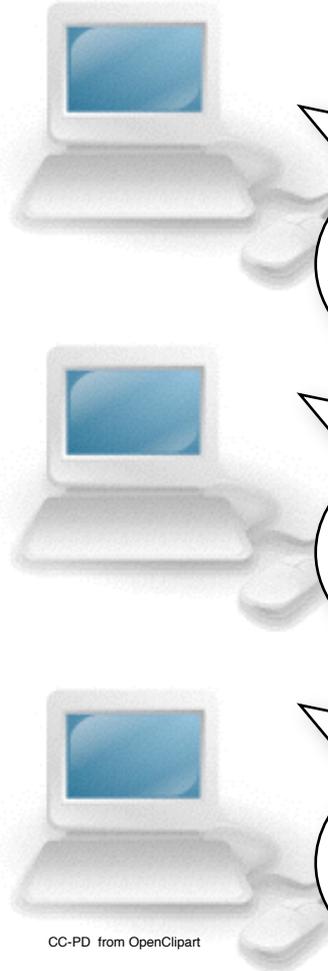


# 遺伝研スーパーコンピュータ (概要)



**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

# 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





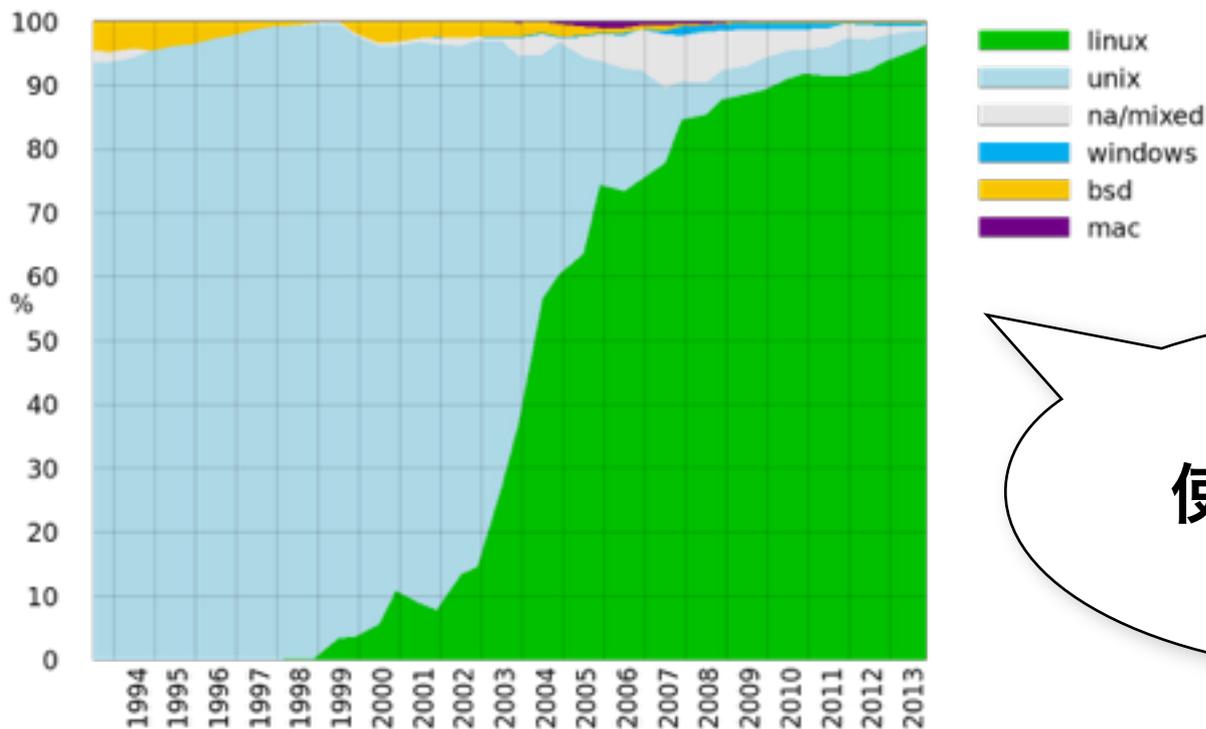
iOS の基礎部分は Darwin であり、NeXTSTEP を先祖に持つ由緒正しき Berkeley Software Distribution (BSD) UNIX の系譜に連なります。

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

# Linux とは

---

- Linuxは、狭義ではLinuxカーネルを意味し、広義ではそれをカーネルとして用いたオペレーティングシステム (OS) を意味する。Linuxカーネルを用いたオペレーティングシステムは、Unix に類似した振る舞いをするので、Unix系のオペレーティングシステムの一つに分類される。

<http://ja.wikipedia.org/wiki/Linux>

GNU

Applications

Linux  
Mach

Kernel

カーネルは  
人間と機械の媒

CPU

Memory

Devices



[リーナス・ベネディクト・トーバルズ \(Linus Benedict Torvalds; 1969-\)](#)

フィンランド、ヘルシンキ出身のプログラマー。Linuxカーネルを開発し、1991年に一般に公開した

- 1991年、ヘルシンキ大学の学生であったリーナス・トーバルズはオペレーティングシステムに好奇心を抱き Intel 80386 CPU を搭載した PC に UNIX 互換の独自のカーネルを開発した。最終的にこれが現在の Linux カーネルへと成長した。

<http://ja.wikipedia.org/wiki/Linux>

# 見えてるのは実は Linux でなくて GNU

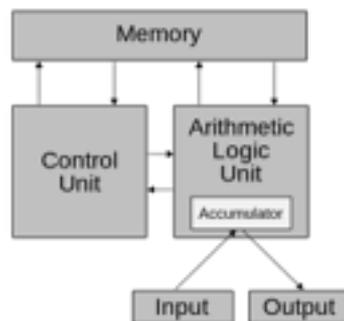
---

- GNU's Not Unix
- GNU は典型的には Linux と呼ばれるカーネルとともに使われます。この組み合わせが GNU/Linux オペレーティング・システムです。GNU/Linux は何百万人もの人に使われています。間違って“Linux”と呼ばれていますけれども。

<http://www.gnu.org/home.ja.html>

# 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



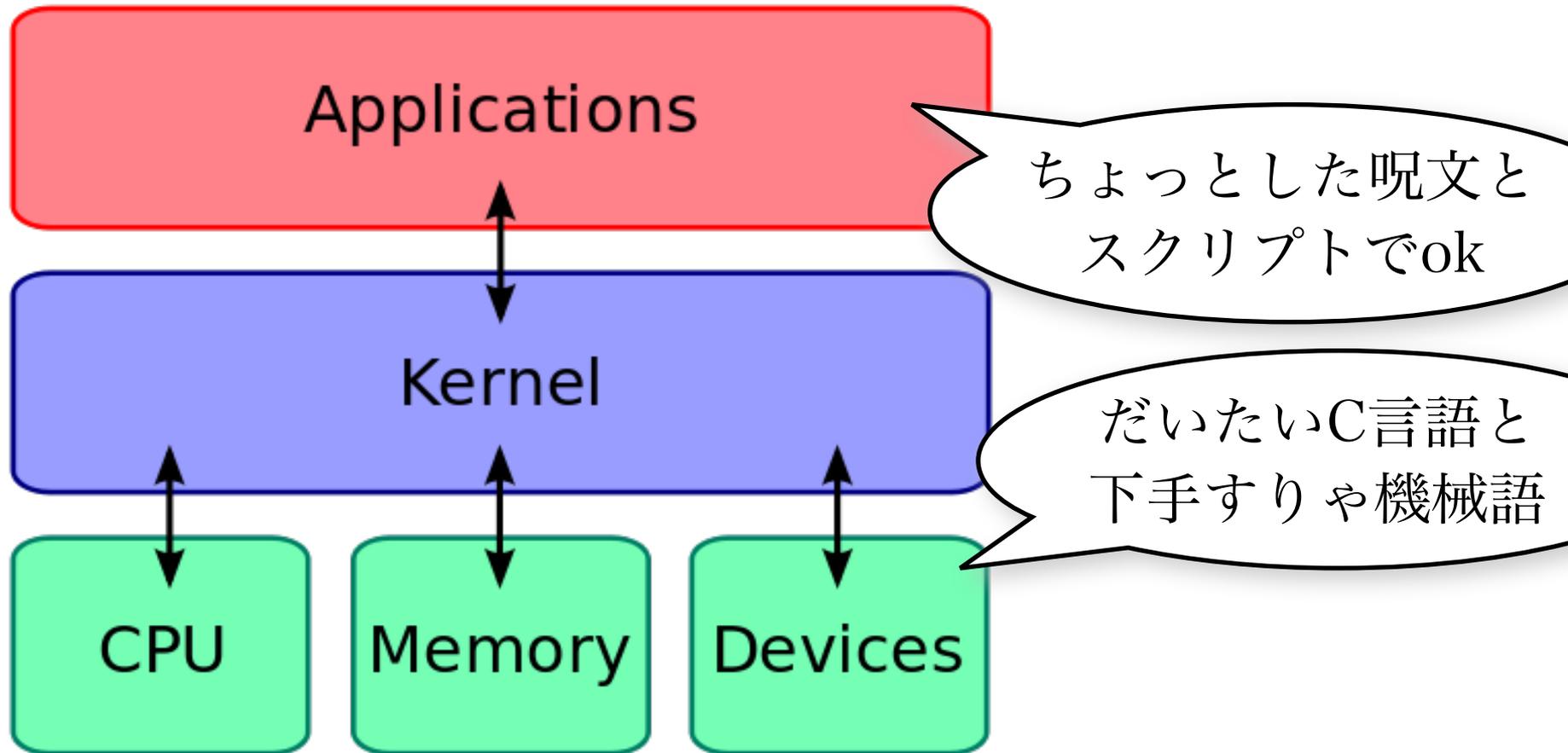
[Von Neumann architecture CC BY-SA 3.0](#)

- 「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](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

# 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 [sc.ddbj.nig.ac.jp](http://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 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 Announcements) and lists several notices 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 System). The page also features a sidebar with navigation links like "Language/言語", "このサイトへのログイン", and "システム構成". At the bottom, there are images of server racks and a large red arrow pointing to the left.

- 「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、ぐっ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！

# 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