

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

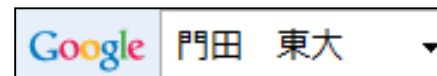
3. データ解析基礎 | 3-1. R 基礎1

東京大学・大学院農学生命科学研究科
アグリバイオインフォマティクス教育研究ユニット

門田幸二(かどた こうじ)

kadota@iu.a.u-tokyo.ac.jp

<http://www.iu.a.u-tokyo.ac.jp/~kadota/>



ファイルのダウンロード

← → http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html (Rで)塩基配列解析

(Rで)塩基配列解析

～NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル、バイオインフォマティクス～
(last modified 2014/08/22, since 2010)

What's new?

- このウェブページはフリーソフト R と必要なパッケージをインストール済みである前提で記述しています。初心者には、[1. Rのインストールと起動](#)および [2. 基本的な利用法](#)で自習してください。(2014/07/21)
- 2014年10月04日に [HPCIワークショップ「医療とビッグデータ解析」](#)(9:00-9:20)に引き続いて [中級者向けバイオインフォマティクス入門講習会@仙台国際センター](#)(10:50-12:20)で話します。興味ある方はどうぞ。(2014/07/23)
- 門田幸二 著 [シリーズ Useful R 第7巻トランスクリプトーム解析](#)刊行(共立出版)
- 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](#)で利用する計算機環境構築する一通りの手順を公開しました。(2014/08/11) **NEW**
- [日本乳酸菌学会誌](#)のNGS関連連載の [第1回分PDF](#)を公開しました。関連項目は [こちら](#)。(2014/08/03) **NEW**
- [参考資料\(講義、講習会、本など\)](#)の項目を更新しました。(2014/08/19) **NEW**

- [はじめに](#) (last modified 2014/01/30)
- [参考資料\(講義、講習会、本など\)](#) (last modified 2014/08/19) **NEW**
- [過去のお知らせ](#) (last modified 2014/08/03) **NEW**
- [Rのインストールと起動](#) (last modified 2014/07/31) **NEW**
- [基本的な利用法](#) (last modified 2014/07/20)
- [サンプルデータ](#) (last modified 2014/07/17)
- 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](#) (last modified 2014/08/22) **NEW**
- 書籍 [トランスクリプトームについて](#) (last modified 2014/05/12)
- 書籍 [トランスクリプトーム解析 | 2.3.1 RNA-seqデータ\(FASTQファイル\)](#) (last modified 2014/04/15) [トップページへ](#)
- 書籍 [トランスクリプトーム解析 | 2.3.2 リファレンス配列](#) (last modified 2014/04/16)

ファイルのダウンロード

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 NEW

2014年9月にJST-NBDCと東大農アグリバイオ主催で「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」が開催されます。主催機関のサイト上で情報提供したほうがいいだろうということで、受講者が各自でインストールするソフトウェアや、イメージファイルのダウンロードなど準備していただく計算機環境の情報などを示します。

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)関連:

- [NBDCの速習コース案内サイト](#) (速習コース主催機関)
- [HPCIの速習コース受講申込受付サイト](#) (速習コース共催機関)
- カリキュラムを策定した[NBDC](#)
- 「NBDCで実施した調査」
 - 「バイオインフォマティクス」
 - 「カリキュラムで習得するスキル」
 - 「カリキュラム フロア」

hoge.zipをデスクトップにダウンロードして解凍しておきましょう

計算機環境構築(Linux系):

8月初旬をめどにBio-Linux1にブ

- [VirtualBox](#)
 - [VirtualBoxをWind](#)
- [Bio-Linux: Field et al., Na](#)

- 2014年9月5日 15:00-18:15、「2-2. バイオ系データベース概論」、初級、実習
 - [小野浩雅 \(DBCLS\)](#)、統合TV、講義資料
 - 基本的な各種バイオ系データベースの理解、統合DBの利用法。
- 2014年9月8日 10:30-12:00、「3-1. R 基礎1」、初級、実習
 - [門田幸二 \(東京大学\)](#)、統合TV、[講義資料 \(20140822, 18:04版\)](#)
 - Rインストール自体は基本的に終了した状態を想定しているものの、最初に[library\(Biostrings\)](#)などいくつかの利用予定パッケージのロードを行い、パッケージのインストールがうまくいっているかどうかを確認(できていなかったヒトの同定および対処)。Rの一般的な利用法。log関数などの基本的かつ挙動を完全に把握できる関数を例として、関数内部のオプション変更や「?関数名」で利用法の幅を広げる基本テクを概観。exp, mean, median, sort, length関数。
 - 9/8-9の2日間で用いるファイル群:[hoge.zip](#) (20140822,17:24版)
 - Rコード:[rcode_20140908.txt](#)
- 2014年9月8日 13:15-14:45、「3-2. R 基礎2」、初級、実習
 - [門田幸二 \(東京大学\)](#)、統合TV、[講義資料 \(20140811, 0:11版\)](#)
 - 翻訳配列の取得を例に「(Rで)塩基配列解析」の基本的な利用法を紹介。塩基配列中にNを



- 2014年9月8日10:30-12:00、「3-1. R 基礎1」、初級、実習
- 門田幸二(東京大学)、統合TV、講義資料(20140822, 18:04版)
- Rインストール自体は基本的に終了した状態を想定しているものの、最初にlibrary(Biostrings)などいくつかの利用予定パッケージのロードを行い、パッケージのインストールがうまくいっているかどうかを確認(できていなかったヒトの同定および対処)。Rの一般的な利用法。log関数などの基本的かつ挙動を完全に把握できる関数を例として、関数内部のオプション変更や「?関数名」で利用法の幅を広げる基本テクを概観。exp, mean, median, sort, length関数。
 - 9/8-9の2日間で用いるファイル群: [hoge.zip](#) (20140822,17:24版)
 - Rコード: [rcode_20140908.txt](#)

- 2014年9月8日13:15-14:45、「3-2. R 基礎2」、初級、実習
- 門田幸二(東京大学)、統合TV、講義資料(20140822, 18:04版)
- 翻訳配列の取得を例に「(Rで)塩基配列解析」を含む場合のエラー例とその対処法。RGui画面タブ区切りテキストファイルからの情報抽出。Pとその対処法。

開く(O)
 新しいタブで開く(W)
 新しいウィンドウで開く(N)
 対象をファイルに保存(A)...
 対象を印刷(P)
 切り取り

名前を付けて保存

デスクトップ

整理 新しいフォルダー

お気に入り
 SkyDrive
 ダウンロード
 デスクトップ
 最近表示した場所

ライブラリ
 ドキュメント
 ピクチャ
 ビデオ
 ミュージック

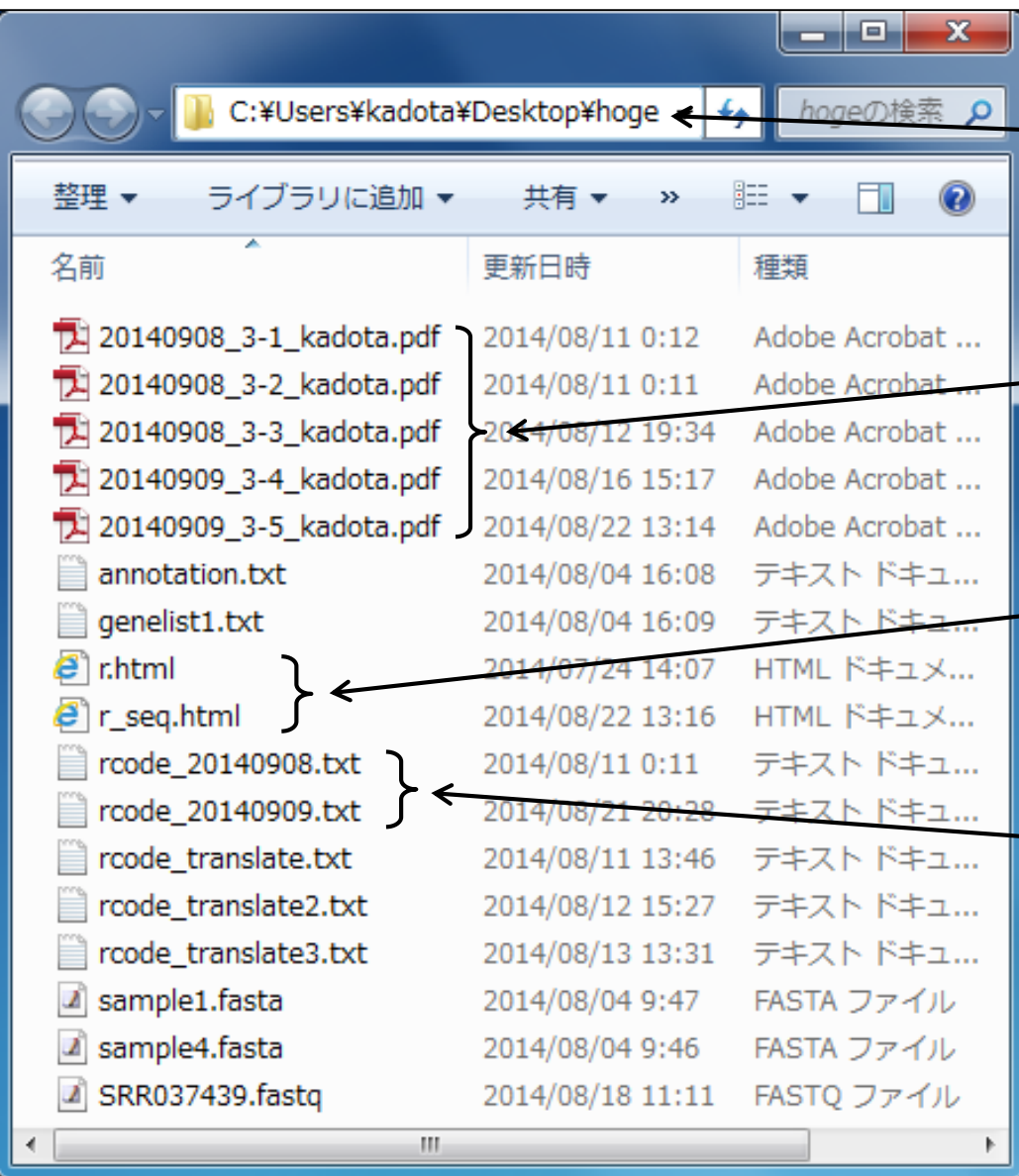
ライブラリ
 システム フォルダ
 ホームグループ
 システム フォルダ
 kadota
 システム フォルダ
 コンピューター
 システム フォルダ
 ネットワーク
 システム フォルダ

ファイル名(N): hoge.zip
 ファイルの種類(T): ZIP ファイル (*.zip)

フォルダーの非表示

保存(S) キャンセル

ファイルのダウンロード



デスクトップ上にhogeフォルダがあり、サンプルファイルが見られるという前提で行います。

9/8-9の講義資料PDF

ダブルクリックでローカルに参考ウェブページを開けます。ネットワーク不調時や負荷軽減用。

9/8-9の講義資料PDF中のRコード。コピー用。

NGS速習コース全貌のおさらい

(Rで)塩基配列解析
 ~NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計, モデル, バイオインフォマティクス~
 (last modified 2014/08/22, since 2010)

What's new?

- このウェブページはフリーソフト R と必要なパッケージをインストール済みである前提で記述しています。初心者には、[1. Rのインストールと起動](#)および [2. 基本的な利用法](#)で自習してください。(2014/07/21)
- 2014年10月04日に [HPCIワークショップ「医療とビッグデータ解析」](#)(9:00-9:20)に引き続いて [中級者向けバイオインフォマティクス入門講習会@仙台国際センター](#)(10:50-12:20)で話します。興味ある方はどうぞ。(2014/07/23)
- 門田幸二 著 [シリーズ Useful R 第7巻 トランスクリプトーム解析](#) 刊行(共立出版)
- 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](#) で利用する計算機環境構築する一通りの手順を公開しました。(2014/08/11) **NEW**
- [日本乳酸菌学会誌](#)のNGS関連連載の [第1回分PDF](#)を公開しました。関連項目は [こちら](#)。(2014/08/03) **NEW**
- [参考資料\(講義、講習会、本など\)](#)の項目を更新しました。(2014/08/19) **NEW**

- [はじめに](#) (last modified 2014/01/30)
- [参考資料\(講義、講習会、本など\)](#) (last modified 2014/08/19) **NEW**
- [過去のお知らせ](#) (last modified 2014/08/03) **NEW**
- [Rのインストールと起動](#) (last modified 2014/07/31) **NEW**
- [基本的な利用法](#) (last modified 2014/07/20)
- [サンプルデータ](#) (last modified 2014/07/17)
- 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](#) (last modified 2014/08/22) **NEW**
- 書籍 [トランスクリプトームについて](#) (last modified 2014/05/12)
- 書籍 [トランスクリプトーム解析 | 2.3.1 RNA-seqデータ\(FASTQファイル\)](#) (last modified 2014/04/15) [トップページへ](#)
- 書籍 [トランスクリプトーム解析 | 2.3.2 リファレンス配列](#) (last modified 2014/04/16)

NGS速習コース全貌のおさらい

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 **NEW**

2014年9月にJST-NBDCと東大農アグリバイオ主催で「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」が開催されます。主催機関のサイト上で情報提供したほうがいだろうということで、受講者が各自でインストールするソフトウェアや、イメージファイルのダウンロードなど準備していただく計算機環境の情報などを示します。

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)関連:

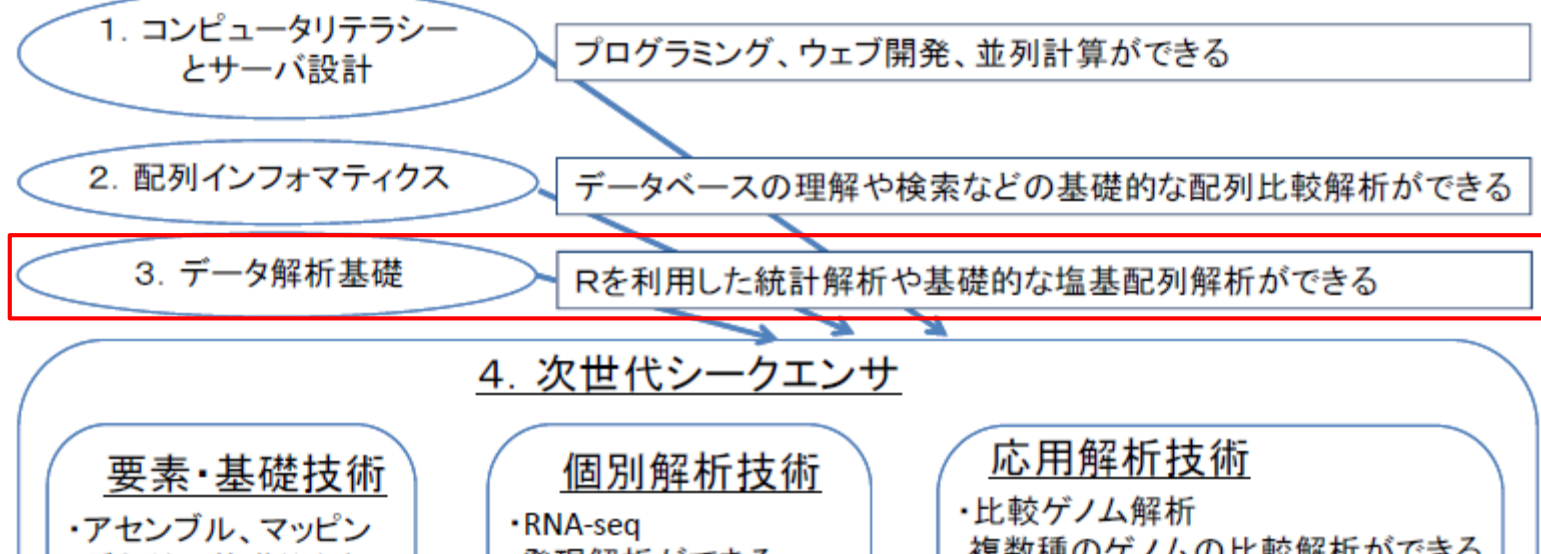
- ・ [NBDCの速習コース案内サイト](#) (速習コース主催機関)
- ・ [HPCIの速習コース受講申込受付サイト](#) (速習コース共催機関)
- ・ カリキュラムを策定した [NBDC運営委員会人材育成分科会](#)
- ・ 「NBDCで実施した調査」の [バイオインフォマティクス人材育成のためのカリキュラム](#)
 - ・ 「バイオインフォマティクス人材育成カリキュラム(次世代シーケンサ)」のPDF ([generation-sequencer.pdf](#))
 - ・ 「カリキュラムで習得できる技能」のPDF ([learning-skills.pdf](#))
 - ・ 「カリキュラム フロー図」のPDF ([flow-diagram.pdf](#))

習得できる技能は、「速習」と「速習以外」を含めたもの

計算機環境構築(Linux系):

8月初旬をめぐりにBio-Linuxに

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)で習得できる技能



NGS速習コース全貌のおさらい

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 **NEW**

2014年9月にJST-NBDCと東大農アグリバイオ主催で「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」が開催されます。主催機関のサイト上で情報提供したほうがいいたろうということで、受講者が各自でインストールするソフトウェアや、イメージファイルのダウンロードなど準備していただく計算機環境の情報などを示します。

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)関連:

- [NBDCの速習コース案内サイト](#) (速習コース主催機関)
- [HPCIの速習コース受講申込受付サイト](#) (速習コース共催機関)
- カリキュラムを策定した [NBDC運営委員会人材育成分科会](#)
- 「NBDCで実施した調査」の [バイオインフォマティクス人材育成のためのカリキュラム](#)
 - 「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)」のPDF ([generation-sequencer.pdf](#))
 - 「カリキュラムで習得できる技能」のPDF ([learning-skills.pdf](#))
 - 「カリキュラム フロー図」のPDF ([flow-diagram.pdf](#))

「速習」と「速習以外」の個別の項目での習得技術がおおまかに書かれている

計算機環境構築(Linux系):

8月初旬をめぐりにBio-Linuxにプリインストールされていないプログラムなども含めたイメージファイルの提供を行う予定です。

NGS速習コース全貌のおさらい

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)

本カリキュラムは、次世代シーケンサデータを扱うにあたり最低限必要とされる知識・技術を2週間程度で身につけることを想定した「速習」と、時間をかけて習得することを想定した「速習以外」に分かれています。

「速習」と「速習以外」の個別の項目での習得技術がおおまかに書かれている

【速習】

大項目	日数		No.	項目	習得技術		
1. コンピュータリテラシーとサーバー設計	4日	2日	1-1	OS, ハード構成	・コンピュータの基本の理解	初級	講義
			1-2	ネットワーク基礎	・インターネット、セキュリティの基本の理解	初級	講義
			1-3	UNIX I	・UNIXの基礎の理解 ・Linux導入	中級	実習
	2日	1-4	スクリプト言語	・Perl ・シェルスクリプト	中級	実習	
2. 配列インフォマティクス	1日		2-1	配列解析基礎	・配列、ゲノムデータ記述のフォーマット、アラインメント(DP)、データベース検索(BLAST, BLAT)等の基礎的な配列比較解析の原理と実習	初級	実習
			2-2	バイオ系データベース概論	・基本的な各種バイオ系データベースの理解、統合DBの利用法	初級	実習
3. データ解析基礎	2日		3-1	R 基礎1	・R言語の基礎(インストールから利用まで)	初級	実習
			3-2	R 基礎2	・ファイルの読み込み、行列演算の基本	初級	実習
			3-3	R 各種パッケージ	・Rの各種パッケージのインストール法と代表的なパッケージの利用法	中級	実習
			3-4	R bioconductor I	・bioconductorの利用法	中級	実習
			3-5	R bioconductor II	・FASTA and FASTQ形式ファイルの読み込み。ファイル形式の変換(FASTQ -> FASTA)、クオリティチェック、リード配列長分布、フィルタリングやトリミング、GC含量計算など	中級	実習

Contents (カリキュラム記載事項)

- 3-1. R 基礎1、2014/09/08 10:30-12:00、初級、実習
 - R言語の基礎(インストールから利用まで)
- 3-2. R 基礎2、2014/09/08 13:15-14:45、初級、実習
 - ファイルの読み込み、行列演算の基本
- 3-3. R 各種パッケージ、2014/09/08 15:00-18:15、中級、実習
 - Rの各種パッケージのインストール法と代表的なパッケージの利用法
- 3-4. R bioconductor I、2014/09/09 10:30-14:45、中級、実習
 - Bioconductorの利用法
- 3-5. R bioconductor II 、2014/09/09 15:00-18:15、中級、実習
 - FASTA and FASTQ形式ファイルの読み込み、ファイル形式の変換(FASTQ → FASTA)、クオリティチェック、リード配列長分布、フィルタリングやトリミング、GC含量計算など。

Contents (「3. データ解析基礎」全体)

- 3-1. R 基礎1、2014/09/08 10:30-12:00、初級、実習
 - Rおよびパッケージのインストール、インストール後の確認
 - 基本的な利用法。log, exp, mean, median, sort, length関数。?関数名など。
- 3-2. R 基礎2、2014/09/08 13:15-14:45、初級、実習
 - (Rで)塩基配列解析の基本的な利用法(翻訳配列の取得を例に)
 - 行列形式ファイルの解析基礎(アノテーションファイルを例に)
- 3-3. R 各種パッケージ、2014/09/08 15:00-18:15、中級、実習
 - Rの各種パッケージのインストール法と代表的なパッケージBiostringsの利用法
- 3-4. R Bioconductor I、2014/09/09 10:30-14:45、中級、実習
 - データの型、バージョンの違い、警告メッセージとその対処法、Bioconductorサイト概観。
 - setwd, translate, rm, ls, objects, sessionInfo, readDNAStringSet関数、プロモータ配列取得。
- 3-5. R Bioconductor II、2014/09/09 15:00-18:15、中級、実習
 - FASTA/FASTQ形式ファイルの操作。ファイル形式の変換、クオリティチェック、フィルタリング、クラスオブジェクト。alphabetFrequency, apply, rowSums, colSums, sample, set.seed関数。

「3. データ解析基礎」での目標

- Rの基本的な利用法を知る
- (Rで)塩基配列解析を使いこなす
 - できることの全体像を知る
 - 基本はコピーで実行(ファイル名や必要最小限のパラメータの変更のみ)
 - 原因既知状態でのエラーを沢山経験し、実データ解析時の複合的なエラーを着実に解決する基本的なノウハウを身につける
 - コードの中身のある程度知っておく
 - R本体やパッケージのバージョンの違い(関数名やデフォルトオプションの変更)
 - WindowsとMacintoshの違い
 - 得られた結果の合理的な解釈ができるようになる



Contents

- 3-1. R 基礎1、2014/09/08 10:30-12:00、初級、実習
 - Rおよびパッケージのインストール、インストール後の確認
 - 参考図書
 - パッケージが正しくインストールされているかどうかの確認
 - エラーメッセージとその対処法
 - 基本的な利用法
 - 四則演算、スクリプトファイルの作成とコピー、改行の有無に注意
 - コメント行、上下左右の矢印キーを有効利用、エラーメッセージ
 - 関数の利用マニュアル、ベクトル計算、オプションの変更
 - 条件判定、論理値ベクトル
 - ベクトル中の任意の要素を抽出(subsetting)
 - ベクトルの要素数、組合せ(ソート後に最初の3要素を抽出)の基本



ウェブ 画像 動画 ショッピング ニュース もっと見る ▼ 検索ツール

約 14,300,000 件 (0.22 秒)

(Rで)塩基配列解析

www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.htm

このページは、次世代シークエンシングの短い塩基配列(short read)を解析するためのRパッケージがあり、特にアグリバイオ

R+Bioconductor

togotv.dbcls.jp/20120925

2012/09/25 - 本日の統計学会: AJACSみちのく2か所による「R+Bioconductor

Learning R - The

cat.hackingisbelieving.com

著者: Itoshi NIKAIDO
NGS解析: 1限 Rの基礎
「RとBioconductorを使

RNA-seq analys

cat.hackingisbelieving.com

参考ウェブページ

http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.htm (Rで)塩基配列解析

(Rで)塩基配列解析

～NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル、バイオインフォマティクス～
(last modified 2014/07/31, since 2010)

What's new?

- このウェブページはフリーソフトRと必要なパッケージをインストール済みである前提で記述しています。初心者には、1. [Rのインストールと起動](#)および2. [基本的な利用法](#)で自習してください。(2014/07/21) **NEW**
- 2014年10月04日にHPCIワークショップ「医療とビッグデータ解析」(9:00-9:20)に引き続いて [中級者向けバイオインフォマティクス入門講習会@仙台国際センター](#)(10:50-12:20)で話します。興味ある方はどうぞ。(2014/07/23) **NEW**
- 門田幸二 著シリーズ [Useful R 第7巻トランスクリプトーム解析](#)刊行(共立出版)
- 2014年9月1日～12日に「[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](#)」を開催します。受講申込は6/24夕方に締め切りました。TA申込枠はあと数名です。(2014/07/21) **NEW**
- [参考資料\(講義、講習会、本など\)](#)の項目を追加しました。(2014/07/30) **NEW**

- [はじめに](#) (last modified 2014/01/30)
- [参考資料\(講義、講習会、本など\)](#) (last modified 2014/07/30) **NEW**
- [過去のお知らせ](#) (last modified 2014/07/31) **NEW**
- [Rのインストールと起動](#) (last modified 2014/07/07) **NEW**
- [基本的な利用法](#) (last modified 2014/07/20) **NEW**

[トップページへ](#)

Windows版を例としてR本体および各種パッケージのインストール手順やエラーへの対処法などを記載しています。動作確認はR ver. 3.1.0とBioconductor ver. 2.14で行っています。OSが違えど、基本的な手順は同じ。

(Rで)塩基配列解析

～NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル、バイオインフォマティクス
(last modified 2014/07/31, since 2014/07/31)

What's new?

- このウェブページはフリーソフトRとは、1. [Rのインストールと起動](#)および
- 2014年10月04日 [Rのインストールと起動](#)の更新
- 門田幸二 著 [シリアル Useful R 第1巻](#)
- 2014年9月1日～12日に「[バイオインフォマティクス](#)」を行います。受講申込は6/24夕方に締
- [参考資料\(講義、講習会、本など\)](#)
- [はじめに](#) (last modified 2014/01/30)
- [参考資料\(講義、講習会、本など\)](#)
- [過去のお知らせ](#) (last modified 2014/07/31)
- [Rのインストールと起動](#) (last modified 2014/07/31)
- [基本的な利用法](#) (last modified 2014/07/31)

http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.htm (Rで)塩基配列解析

Rのインストールと起動 NEW

基本的には[こちら](#)または[こちら](#)をご覧ください。
よく分からない人でWindowsユーザーの方は以下を参考にしてください。2014年7月31日にアップデートしたWindows用のインストール手順は[こちら](#)。2014年5月14日にアップデートしたMac版のインストール手順[こちら](#)(by [孫建強氏](#))もあります。注意点は、「Mac OS Xのバージョンに関わらず R-3.1.0-snowleopard.pkgをインストールしたほうがよい」です。

1. Windows release版のインストールの場合:

- ② [Rのインストーラ](#)を「実行」
- 聞かれるがままに「次へ」などを押してとにかくインストールを完了させる
- Windows Vistaの人**は(パッケージのインストール中に書き込み権限に関するエラーが出るのを避けるために)「コントロールパネル」-「ユーザーアカウント」-「ユーザーアカウント 制御の有効化または無効化」で、「ユーザーアカウント 制御(UAC)を使ってコンピュータの保護に役立たせる」のチェックをあらかじめ外しておくことを強くお勧めします。
- インストールが無事完了したら、デスクトップに出現する「R3.X.Y(32 bitの場合; XやY中の数値はバージョンによって異なります)」または「R x64 3.X.Y(64 bitの場合)」アイコンをダブルクリックして起動
- 以下を、「R コンソール画面」上でコピー＆ペーストする。10GB程度のディスク容量を要しますが一番お手軽です。(どこからダウンロードするか?と聞かれるので、その場合は自分のいる場所から近いサイトを指定)

```
install.packages(available.packages()[,1], dependencies=TRUE)#CRAN中にある全てのパッケージをインストール
source("http://www.bioconductor.org/biocLite.R")#おまじない
biocLite(all_group())#Bioconductor中にある全てのパッケージをインストール
biocLite("BSgenome.Athaliana.TAIR.TAIR9", suppressUpdates=TRUE)#Bioconductor中にあるBSgenome.Athaliana.TAIR.TAIR9をインストール
```

6. 「コントロールパネル」-「デスクトップのカスタマイズ」-「フォルダオプション」-「表示(タブ)」-「[トップページ](#)」ここで、「登録されている拡張子は表示しない」のチェックを外してください。

(Rで)塩基配列解析

～NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計, モデル, バイオインフォマティクス～

(last modified 2014/07/31, since 2014/07/31)

What's new?

- このウェブページはフリーソフトRとは、1. [Rのインストールと起動](#)および
- 2014年10月04日に[HPCワークショップ フォーマティクス入門講習会@仙台区](#)
- 門田幸二 著 [シリーズ Useful R 第1巻](#)
- 2014年9月1日～12日に「[バイオインフォマティクス入門](#)」を行います。 [受講申込](#)は6/24夕方に締め切ります。
- [参考資料\(講義, 講習会, 本など\)](#)
- [はじめに](#) (last modified 2014/01/30)
- [参考資料\(講義, 講習会, 本など\)](#)
- [過去のお知らせ](#) (last modified 2014/07/31)
- [Rのインストールと起動](#) (last modified 2014/07/31)
- [基本的な利用法](#) (last modified 2014/07/31)

Rのインストールと起動 NEW

基本的には[こちら](#)または[こちら](#)をご覧ください。

よく分からない人でWindowsユーザーの方は以下を参考にしてください。2014年7月31日にアップデートしたWindows用のインストール手順は[こちら](#)、2014年5月14日にアップデートしたMac版のインストール手順[こちら](#)(by 孫建強氏)もあります。注意点は、「Mac OS Xのバージョンに関わらず R-3.1.0-snowleopard.pkgをインストールしたほうがよい」です。

より詳細なインストール手順を示したPDFはこちら。NGS速習コース受講生はこちらを参考に正しくインストール済みという前提です。

1. Windows release版のインストールの場合:

- [Rのインストーラ](#)を「実行」
- 聞かれるがままに「次へ」などを押してとにかく進め
- Windows Vista**の人(パッケージのインストール時に「コントロールパネル」-「ユーザーアカウント制御(Windows Firewall) 設定」-「ユーザーアカウント制御(UAC)を使ってコントロールパネルの表示を制限する」をオフにしておくことを強くお勧めします。
- インストールが無事完了したら、デスクトップに出現する「R3.X.Y(32 bitの場合: XやY中の数値はバージョンによって異なります)」または「R x64 3.X.Y(64 bitの場合)」アイコンをダブルクリックして起動
- 以下を、「R コンソール画面」でコピー＆ペーストする。10GB程度のディスク容量を要しますが一番お手軽です。(どこからダウンロードするか?と聞かれるので、その場合は自分のいる場所から近いサイトを指定)

```
install.packages(available.packages()[,1], dependencies=TRUE)#CRAN中にある全てのパッケージをインストール
source("http://www.bioconductor.org/biocLite.R")#おまじない
biocLite(all_group())#Bioconductor中にある全てのパッケージをインストール
biocLite("BSgenome.Athaliana.TAIR.TAIR9", suppressUpdates=TRUE)#Bioconductor中にあるパッケージをインストール
```

- 「コントロールパネル」-「デスクトップのカスタマイズ」-「フォルダオプション」-「表示(タブ)」-「[詳細な表示オプション](#)」-「[詳細な表示オプション](#)」-「登録されている拡張子は表示しない」のチェックを外してください。

(Rで)塩基配列解析

～NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル、バイオインフォマティクス～

(last modified 2014/07/31, since 2014/07/31)

What's new?

- このウェブページはフリーソフト R とは、1. [Rのインストールと起動](#)および
- 2014年10月04日に[HPCワークショップ](#)
- [フォーマティクス入門講習会@仙台区](#)
- 門田幸二 著 [シリーズ Useful R 第1巻](#)

- [はじめに](#) (last modified 2014/01/30)
- [参考資料\(講義、講習会、本など\)](#)
- [過去のお知らせ](#) (last modified 2014/07/31)
- [Rのインストールと起動](#) (last modified 2014/07/31)
- [基本的な利用法](#) (last modified 2014/07/31)

Rのインストールと起動 NEW

基本的には[こちら](#)または[こちら](#)をご覧ください。

よく分からない人で Windows ユーザーの方は以下を参考にしてください。2014年7月31日にアップデートした Windows 用のインストール手順は[こちら](#)。2014年5月14日にアップデートした Mac 版のインストール手順[こちら](#)(by 孫建強氏)もあります。注意点は、「Mac OS X のバージョンに関わらず R-3.1.0-snowleopard.pkg をインストールしたほうがよい」です。

1. Windows release版のインストールの場合:

1. [Rのインストーラ](#)を「実行」
2. 聞かれるがままに「次へ」などを押して
3. **Windows Vista**の方は(パッケージのインストール時に「コントロールパネル」-「ユーザーアカウント制御(UAC)を使うかどうかを強くお勧めします」)
4. インストールが無事完了したら、デスクトップによって異なります)または「R x64 32bit」
5. 以下を、「R コンソール画面」上でコピーしてください。(どこからダウンロードするか?)

インストールができていなくても、実際にはできていなかったという事例が散見されます。パッケージ名のスペルミスもよく見受けられます。Macintosh のヒトもこちらを参考に、いくつかのパッケージについて適切にインストールされているか確認しておきましょう。

```
install.packages(available.packages()[,1], dependencies=TRUE)#CRAN中にある全てのパッケージをインストール
source("http://www.bioconductor.org/biocLite.R")#おまじない
biocLite(all_group())#Bioconductor中にある全てのパッケージをインストール
biocLite("BSgenome.Athaliana.TAIR.TAIR9", suppressUpdates=TRUE)#Bioconductor中にある
```

6. 「コントロールパネル」-「デスクトップのカスタマイズ」-「フォルダオプション」-「表示(タブ)」-「[詳細なページ](#)」-「登録されている拡張子は表示しない」のチェックを外してください。

(Rで)塩基配列解析

~NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計
(last modified 2014/07/31, since 2010)

What's new?

- このウェブページはフリーソフトRと必要なパッケージをインストール済みの環境は、1. [Rのインストールと起動](#)および2. [基本的な利用法](#)で自習してください。
- 2014年10月04日にHPCIワークショップ「医療とビッグデータ解析」(9:00-10:30)と「バイオインフォマティクス入門講習会@仙台国際センター」(10:50-12:20)で話します。
- 門田幸二 著 [シリーズ Useful R 第7巻 トランスクリプトーム解析](#) 刊行(共著)
- 2014年9月1日~12日に「[バイオインフォマティクス人材育成カリキュラム](#)」を実施します。受講申込は6/24夕方に締め切りました。TA申込枠はあと数名です。
- [参考資料\(講義, 講習会, 本など\)](#)の項目を追加しました。(2014/07/30)

- [はじめに](#) (last modified 2014/01/30)
- [参考資料\(講義, 講習会, 本など\)](#) (last modified 2014/07/30) **NEW**
- [過去のお知らせ](#) (last modified 2014/07/31) **NEW**
- [Rのインストールと起動](#) (last modified 2014/07/07) **NEW**
- [基本的な利用法](#) (last modified 2014/07/20) **NEW**

定期的なバージョンアップや予め殆ど全てのパッケージのインストールを推奨するポリシーも記載。ウェブページでは書きづらい統計的なものの考え方、数式が苦手な人向けに重みつき平均値を算出する手順を詳述し、用いるパラメータの常識・非常識などを丁寧に解説。



金明哲 編・門田幸二 著

シリーズ名 [シリーズ Useful R](#) 全10巻 【7】 巻
ISBN 978-4-320-12370-0
判型 B5
ページ数 240ページ
発売予定 2014年04月10日
本体価格 3,600円

新刊

今日では、インターネット上の検索エンジンでキーワード検索すれば、個別の情報は簡単に得られる。しかし、本書のメインターゲットである生命科学分野の実験系研究者やこれからバイオインフォマティクスを学ぼうとする大学院生にとっては、特に統計関連の記述は難解であろう。巷に溢れている統計関連書籍の記述内容もまた、意味不明だという声をよく聞く。

本書は、トランスクリプトーム解析を行うための一連のスク립ト集である著者の2つのウェブページ「(Rで)マイクロアレイデータ解析;
<http://www.iu.a.u-tokyo.ac.jp/~kadota/r.html>」および「(Rで)塩基配列解析;
http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html」を体系的にまとめた初の書籍である。

まず手元にある実際のデータやその解析結果を示し、解釈の仕方を述べてから一般論に導く記述形式を採用している。主な目的は、実データの解析結果を徹底的に眺めることで、統計的な感覚や数式感覚を身につけることである。一般に、書籍中に記載されているRパッケージや関数は、パッケージ自体がなくなってしまうたりオプションが変更されるなど比較的早期に陳腐化していく。本書は、もちろん執筆時点で最新の解析手順やRの関数を利用しているが、それらの賞味期限は短いことが予想される。そのため、最新の利用手順は2つのウェブページを参考にされたい。本書は、トランスクリプトーム解析のための二大技術であるマイクロアレイとRNA-seqをRで自在に解析するための基本的な考え方や注意点を体系的にまとめたものである。

その他参考書

羊土社HPより



書籍を購入

次世代シーケンス解析スタンダード

NGSのポテンシャルを活かしきるWET&DRY

二階堂 愛 / 編

定価 5,500円 + 税 2014年08月 発行 B5判 404ページ

ISBN 978-4-7581-0191-2

[SHARE] [ツイート](#) [シェア](#) [B!](#) [8+](#)

本書概要 | 目次詳細 | 立ち読み | 掲載広告・資料請求

エピゲノム研究はもとより、医療現場から非モデル生物、生物資源まで各分野の「NGSの現場」が詰まった1冊。コツや条件検討方法などWET実験のポイントが、データ解析の具体的なコマンド例が、わかる！

内容見本



PDFダウンロード

[目次](#) | [索引](#) | [目次詳細](#)



羊土社
YODOSHA

二階堂愛先生編集！現場目線で徹底的に丁寧な解説。



一連のインストール作業が終了すると、Biostrings, ShortRead, TCC, BSgenome.Athaliana.TAIR.TAIR9となどの各種パッケージが利用可能になっているはず。左記のように打ち込んでエラーが出ていなければオッケー。

```
RGui (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ
R Console
> library(Biostrings)
要求されたパッケージ BiocGenerics をロード中です
要求されたパッケージ parallel をロード中です

次のパッケージを付け加えます: 'BiocGenerics'

以下のオブジェクトはマスクされています (from 'package:parallel') :

clusterApply, clusterApplyLB, clusterExport, clusterMap, parLapplyLB, parRapply, parSapply

以下のオブジェクトはマスクされています (from 'package:BiocGenerics') :

xtabs

以下のオブジェクトはマスクされています (from 'package:BiocGenerics') :

anyDuplicated, append, as.data.frame, colnames, do.call, duplicated, intersect, is.unsorted, lapply, paste, pmax, pmax.int, pmin, Reduce, rep.int, rownames, save, union, unique, unlist

要求されたパッケージ IRanges をロード中です
要求されたパッケージ XVector をロード中です
> |
```

```
RGui (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ
R Console
要求されたパッケージ IRanges をロード中です
要求されたパッケージ XVector をロード中です
> library(ShortRead)
要求されたパッケージ BiocParallel をロード中です
要求されたパッケージ Rsamtools をロード中です
要求されたパッケージ GenomicRanges をロード中です
要求されたパッケージ GenomeInfoDb をロード中です
要求されたパッケージ GenomicAlignments をロード中です
要求されたパッケージ BSgenome をロード中です
> |
```

```
R Console
> library(TCC)
要求されたパッケージ DESeq をロード中です
要求されたパッケージ Biobase をロード中です
Welcome to Bioconductor

Vignettes contain introductory material; view
'browseVignettes()'. To cite Bioconductor, see
'citation("Biobase")', and for packages
'citation("pkgname")'.

要求されたパッケージ locfit をロード中です
locfit 1.5-9.1 2013-03-22

次のパッケージを付け加えます: 'locfit'

以下のオブジェクトはマスクされています (from 'package:GenomicRanges')
left, right

要求されたパッケージ lattice をロード中です
Welcome to 'DESeq'. For improved performance,
and functionality, please consider migrating
要求されたパッケージ DESeq2 をロード中です
要求されたパッケージ Rcpp をロード中です
要求されたパッケージ RcppArmadillo をロード中です

次のパッケージを付け加えます: 'DESeq2'

以下のオブジェクトはマスクされています (from 'package:DESeq2')
estimateSizeFactorsForMatrix, getVarianceStabilizationData,
plotPCA, varianceStabilizingTransformation
```

一連のインストール作業が終了すると、Biostrings, ShortRead, TCC, BSgenome.Athaliana.TAIR.TAIR9となどの各種パッケージが利用可能になっているはず。左記のように打ち込んでエラーが出ていなければオッケー。

```
R Console
要求されたパッケージ bayseq をロード中です
次のパッケージを付け加えます: 'bayseq'

以下のオブジェクトはマスクされています (from 'package:ShortRead')
$
rbind

以下のオブジェクトはマスクされています (from 'package:GenomicRanges')
$
rbind

以下のオブジェクトはマスクされています (from 'package:IRanges')
$
rbind

以下のオブジェクトはマスクされています (from 'package:BiocGenerics')
$
rbind

以下のオブジェクトはマスクされています (from 'package:base')
$
rbind

要求されたパッケージ ROC をロード中です
次のパッケージを付け加えます: 'TCC'

以下のオブジェクトはマスクされています (from 'package:edgeR')
$
calcNormFactors

> |
```

```
> library(BSgenome.Athaliana.TAIR.TAIR9)
```

```
要求されたパッケージ BSgenome をロード中です
要求されたパッケージ BiocGenerics をロード中です
要求されたパッケージ parallel をロード中です
```

```
次のパッケージを付け加えます: 'BiocGenerics'
```

```
以下のオブジェクトはマスクされています (from 'package:parallel') :
```

```
clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
clusterExport, clusterMap, parApply, parCapply, parLapply,
parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
以下のオブジェクトはマスクされています (from 'package:stats') :
```

```
xtabs
```

```
以下のオブジェクトはマスクされています (from 'package:base') :
```

```
anyDuplicated, append, as.data.frame, as.vector, cbind,
colnames, do.call, duplicated, eval, evalq, Filter, Find, get,
intersect, is.unsorted, lapply, Map, mapply, match, mget, order,
paste, pmax, pmax.int, pmin, pmin.int, Position, rank, rbind,
Reduce, rep.int, rownames, sapply, setdiff, sort, table, tapply,
union, unique, unlist
```

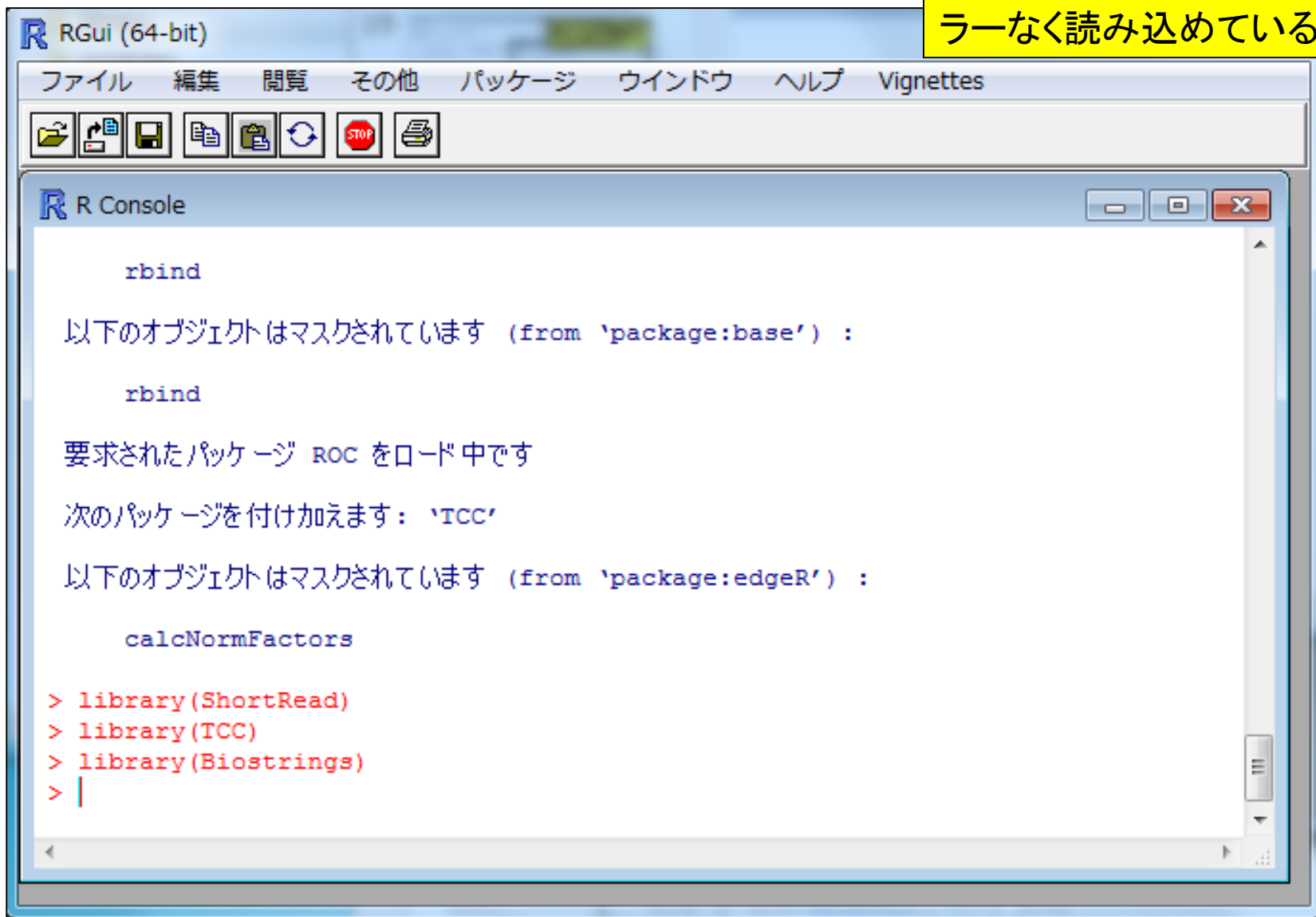
```
要求されたパッケージ IRanges をロード中です
要求されたパッケージ GenomicRanges をロード中です
要求されたパッケージ GenomeInfoDb をロード中です
要求されたパッケージ Biostrings をロード中です
要求されたパッケージ XVector をロード中です
```

```
> library(BSgenome.Athaliana.TAIR.TAIR9)
```

```
> |
```

一連のインストール作業が終了すると、Biostrings, ShortRead, TCC, BSgenome.Athaliana.TAIR.TAIR9となどの各種パッケージが利用可能になっているはず。左記のように打ち込んでエラーが出ていなければオッケー。

一度、library関数を用いて読み込んだパッケージをもう一度読み込むと、表示される文章がなくなります。しかしこれもエラーなく読み込めているので問題なしです



```
RGui (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ Vignettes

R Console
  rbind
 以下のオブジェクトはマスクされています (from 'package:base') :
  rbind
 要求されたパッケージ ROC をロード中です
 次のパッケージを付け加えます: 'TCC'
 以下のオブジェクトはマスクされています (from 'package:edgeR') :
  calcNormFactors
> library(ShortRead)
> library(TCC)
> library(Biostrings)
> |
```

エラー遭遇例とその対処法1

ときどき必要なパッケージのインストールに失敗していて、任意のパッケージXXXの読み込みを行うlibrary(XXX)実行後にエラーが出てしまうことがあります。この例では、TCCパッケージが要求している「RcppArmadilloパッケージがないからダメ!」と文句を言われています。

```
R Console
> library(TCC)
要求されたパッケージ DESeq をロード中です
要求されたパッケージ BiocGenerics をロード中です
要求されたパッケージ parallel をロード中です

次のパッケージを付け加えます: 'BiocGenerics'
以下のオブジェクトはマスクされています (from 'package:parallel') :

  clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
  clusterExport, clusterMap, parApply, parCapply, parLapply,
  parLapplyLB, parRapply, parSapply, parSapplyLB

以下のオブジェクトはマスクされています (from 'package:stats') :

  xtabs

以下のオブジェクトはマスクされています (from 'package:base') :

  anyDuplicated, append, as.data.frame, as.vector, cbind,
  colnames, do.call, duplicated, eval, evalq, Filter, Find, get,
  intersect, is.unsorted, lapply, Map, mapply, match, mget, order,
  paste, pmax, pmax.int, pm
Reduce, rep.int, rownames
union, unique, unlist

要求されたパッケージ Biobase をロード中
Welcome to Bioconductor

Vignettes contain introductory
'browseVignettes()'. To c
'citation("Biobase")', an

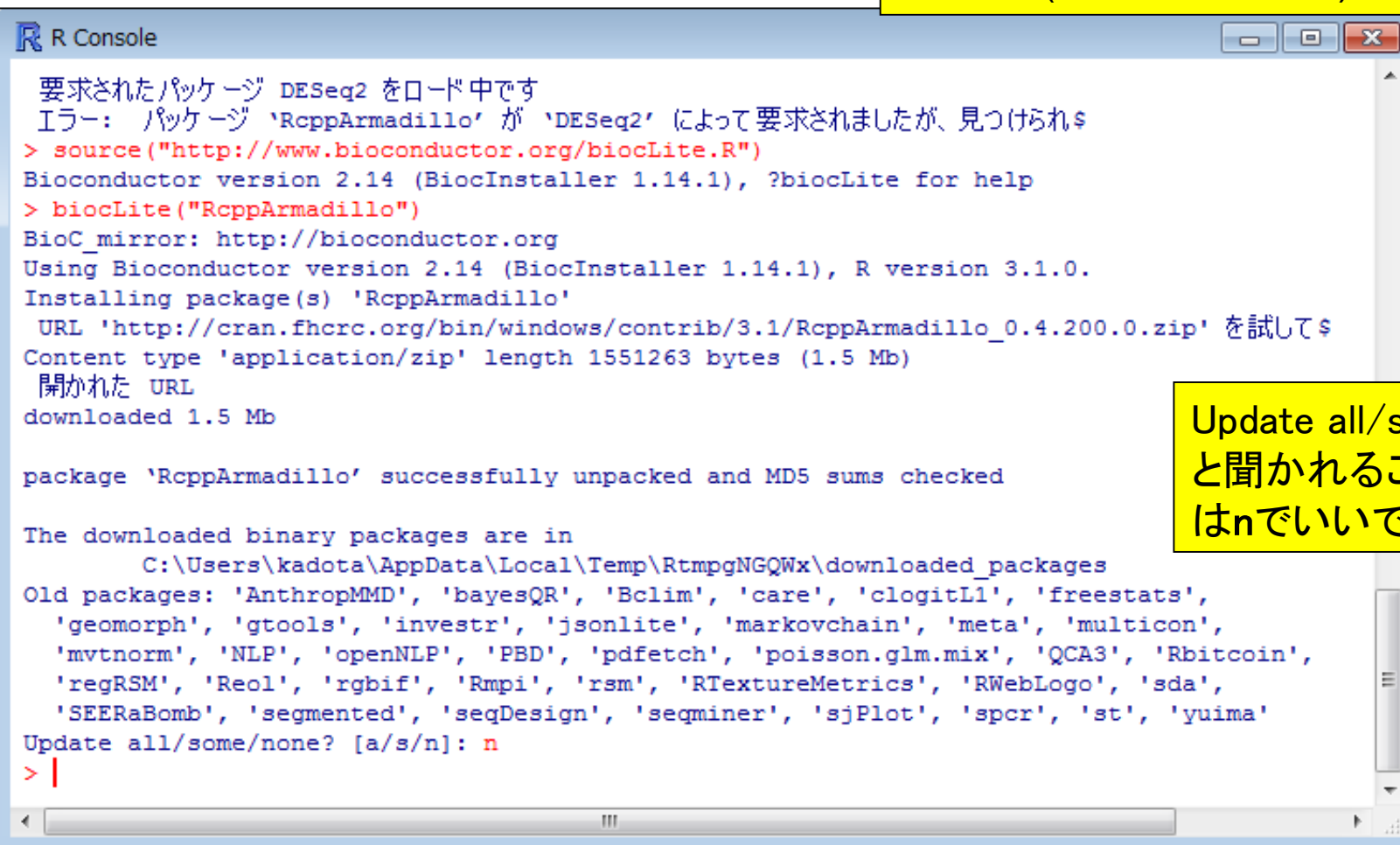
要求されたパッケージ locfit をロード中
locfit 1.5-9.1 2013-03-22
要求されたパッケージ lattice をロード中です
Welcome to 'DESeq'. For improved performance, usability and
functionality, please consider migrating to 'DESeq2'.
要求されたパッケージ DESeq2 をロード中です
要求されたパッケージ GenomicRanges をロード中です
要求されたパッケージ IRanges をロード中です
要求されたパッケージ GenomeInfoDb をロード中です
要求されたパッケージ Rcpp をロード中です
エラー: パッケージ 'RcppArmadillo' が 'DESeq2' によって要求されました
> |
```

```
R Console
> library(TCC)
要求されたパッケージ DESeq2 をロード中です
エラー: パッケージ 'RcppArmadillo' が 'DESeq2' によって要求されましたが、見つけれませんでした
> |
```


エラー遭遇例とその対処法1

基本的な対処法は、文句を言われたパッケージのみインストールすることです。RcppArmadilloパッケージを個別にインストールするためのコマンドの基本形は以下のとおりです：

```
source("http://www.bioconductor.org/biocLite.R")
biocLite("RcppArmadillo")
```



```
R Console
要求されたパッケージ DESeq2 をロード中です
エラー: パッケージ 'RcppArmadillo' が 'DESeq2' によって要求されましたが、見つけれ$
> source("http://www.bioconductor.org/biocLite.R")
Bioconductor version 2.14 (BiocInstaller 1.14.1), ?biocLite for help
> biocLite("RcppArmadillo")
BioC_mirror: http://bioconductor.org
Using Bioconductor version 2.14 (BiocInstaller 1.14.1), R version 3.1.0.
Installing package(s) 'RcppArmadillo'
URL 'http://cran.fhcrc.org/bin/windows/contrib/3.1/RcppArmadillo_0.4.200.0.zip' を試して$
Content type 'application/zip' length 1551263 bytes (1.5 Mb)
開かれた URL
downloaded 1.5 Mb

package 'RcppArmadillo' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\kadota\AppData\Local\Temp\RtmpgNGQWx\downloaded_packages
Old packages: 'AnthropMMD', 'bayesQR', 'Bclim', 'care', 'clogitL1', 'freestats',
'geomorph', 'gtools', 'investr', 'jsonlite', 'markovchain', 'meta', 'multicon',
'mvtnorm', 'NLP', 'openNLP', 'PBD', 'pdffetch', 'poisson.glm.mix', 'QCA3', 'Rbitcoin',
'regRSM', 'Reol', 'rgbif', 'Rmpi', 'rsm', 'RTextureMetrics', 'RWebLogo', 'sda',
'SEERaBomb', 'segmented', 'seqDesign', 'seqminer', 'sjPlot', 'spcr', 'st', 'yuima'
Update all/some/none? [a/s/n]: n
> |
```

Update all/some/none? [a/s/n]:
と聞かれることもありますが基本
はnでいいです。

```
R Console
Update all/some/none? [a/s/n]: n
> library(TCC)
要求されたパッケージ DESeq2 をロード中です
要求されたパッケージ RcppArmadillo をロード中です

次のパッケージを付け加えます: 'DESeq2'

以下のオブジェクトはマスクされています (from 'package:DESeq2') :

  estimateSizeFactorsForMatrix, getVarianceStabilizedData, plotPCA,
  varianceStabilizingTransformation

要求されたパッケージ edgeR をロード中です
要求されたパッケージ limma をロード中です

次のパッケージを付け加えます: 'limma'

以下のオブジェクトはマスクされています (from 'package:DESeq2') :

  plotMA

以下のオブジェクトはマスクされています (from 'package:DESeq2') :

  plotMA

以下のオブジェクトはマスクされています (from 'package:BiocGenerics') :

  plotMA

要求されたパッケージ baySeq をロード中です
次のパッケージを付け加えます: 'baySeq'

以下のオブジェクトはマスクされています (from 'package:GenomicRanges') :

  rbind

以下のオブジェクトはマスクされています (from 'package:IRanges') :

  rbind

以下のオブジェクトはマスクされています (from 'package:BiocGenerics') :

  rbind

以下のオブジェクトはマスクされています (from 'package:base') :

  rbind

要求されたパッケージ ROC をロード中です
次のパッケージを付け加えます: 'TCC'

以下のオブジェクトはマスクされています (from 'package:edgeR') :

  calcNormFactors

> |
```

RcppArmadilloパッケージのインストール後に、もう一度library(TCC)とやって、エラーが出なくなることを確認しています。

```
R Console

以下のオブジェクトはマスクされています (from 'package:$

  calcNormFactors

> library(TCC)
> |
```

エラー遭遇例とその対処法2

次の例では、TCCパッケージが要求している「XMLパッケージがないからダメ!」と文句を言われています。重要な点は、エラーメッセージ中に「パッケージ' DESeq' をロードできませんでした」と書いてありますが、原因はDESeqではなくXMLパッケージがないためであるということを読み解くことです。

```
R Console
> library(TCC)
要求されたパッケージ DESeq をロード中です
要求されたパッケージ BiocGenerics をロード中です
要求されたパッケージ parallel をロード中です

次のパッケージを付け加えます: 'BiocGenerics'

以下のオブジェクトはマスクされています (from 'package:parallel') :

  clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
  clusterExport, clusterMap, parApply, parCapply, parLapply,
  parLapplyLB, parRapply, parSapply, parSapplyLB

以下のオブジェクトはマスクされています (from 'package:stats') :

  xtabs

以下のオブジェクトはマスクされています (from 'pac

  anyDuplicated, append, as.data.frame
  colnames, do.call, duplicated, eval,
  intersect, is.unsorted, lapply, Map,
  paste, pmax, pmax.int, pmin, pmin.in
  Reduce, rep.int, rownames, sapply, s
  union, unique, unlist

要求されたパッケージ Biobase をロード中です
Welcome to Bioconductor

Vignettes contain introductory material; view with
'browseVignettes()'. To cite Bioconductor, see
'citation("Biobase")', and for packages 'citation("pkgname")'.

要求されたパッケージ locfit をロード中です
locfit 1.5-9.1 2013-03-22
要求されたパッケージ lattice をロード中です
Error in loadNamespace(j <- i[[1L]], c(lib.loc, .libPaths()), versionCheck = $
  'XML' という名前のパッケージはありません
エラー: パッケージ 'DESeq' をロードできませんでした
> |

Sep 8-9 2014 NGS速習コース
```

```
R Console
> library(TCC)
要求されたパッケージ DESeq をロード中です
Error in loadNamespace(j <- i[[1L]], c(lib.loc, .libPaths()), versionCheck = $
  'XML' という名前のパッケージはありません
エラー: パッケージ 'DESeq' をロードできませんでした
> |
```

エラー遭遇例とその対処法2

基本的な対処法は、文句を言われたパッケージのみインストールすることです。XMLパッケージを個別にインストールするためのコマンドの基本形は以下のとおりです：

```
source("http://www.bioconductor.org/biocLite.R")
biocLite("XML")
```

```
R Console
> library(TCC)
要求されたパッケージ DESeq をロード中です
Error in loadNamespace(j<-i[[1L]], c(lib.loc, .libPaths()), versionCheck = $
  'XML' という名前のパッケージはありません
エラー: パッケージ 'DESeq' をロードできませんでした
> source("http://www.bioconductor.org/biocLite.R")
Bioconductor version 2.14 (BiocInstaller 1.14.1), ?biocLite for help
> biocLite("XML")
BioC_mirror: http://bioconductor.org
Using Bioconductor version 2.14 (BiocInstaller 1.14.1), R version 3.1.0.
Installing package(s) 'XML'
URL 'http://cran.fhcrc.org/bin/windows/contrib/3.1/XML_3.98-1.1.zip' を試し$
Content type 'application/zip' length 4288694 bytes (4.1 Mb)
開かれた URL
downloaded 4.1 Mb

package 'XML' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\kadota\AppData\Local\Temp\RtmpY31Pk1\downloaded_packages
Old packages: 'AnthropMMD', 'bayesQR', 'Bclim', 'care', 'clogitL1',
'freestats', 'geomorph', 'gtools', 'investr', 'jsonlite', 'markovchain',
'meta', 'multicon', 'mvtnorm', 'NLP', 'openNLP', 'PBD', 'pdfetch',
'poisson.glm.mix', 'QCA3', 'Rbitcoin', 'regRSM', 'Reol', 'rgbif', 'Rmpi',
'rsm', 'RTextureMetrics', 'RWebLogo', 'sda', 'SEERaBomb', 'segmented',
'seqDesign', 'seqminer', 'sjPlot', 'sPCR', 'st', 'yuima'
Update all/some/none? [a/s/n]: n
> |
```

Update all/some/none? [a/s/n]:
と聞かれることもありますが基本
はnでいいです。

```
R Console
Update all/some/none? [a/s/n]: n
> library(TCC)
要求されたパッケージ DESeq をロード中です
Welcome to 'DESeq'. For improved performance, usability and
functionality, please consider migrating to 'DESeq2'.
要求されたパッケージ DESeq2 をロード中です
要求されたパッケージ GenomicRanges をロード中です
要求されたパッケージ IRanges をロード中です
要求されたパッケージ GenomeInfoDb をロード中です
要求されたパッケージ Rcpp をロード中です
要求されたパッケージ RcppArmadillo をロード中です

次のパッケージを付け加えます: 'DESeq2'
以下のオブジェクトはマスクされています (from 'package:DESeq2') :
  estimateSizeFactorsForMatrix, getVarianceStabilizedData,
  plotPCA, varianceStabilizingTransformation

要求されたパッケージ edgeR をロード中です
要求されたパッケージ limma をロード中です

次のパッケージを付け加えます: 'limma'
以下のオブジェクトはマスクされています (from 'package:DESeq2') :
  plotMA

以下のオブジェクトはマスクされています (from 'package:DESeq2') :
  plotMA

以下のオブジェクトはマスクされています (from 'package:BiocGenerics') :
  plotMA

要求されたパッケージ baySeq をロード中です
次のパッケージを付け加えます: 'baySeq'
以下のオブジェクトはマスクされています (from 'package:GenomicRanges') :
  rbind

以下のオブジェクトはマスクされています (from 'package:IRanges') :
  rbind

以下のオブジェクトはマスクされています (from 'package:BiocGenerics') :
  rbind

以下のオブジェクトはマスクされています (from 'package:base') :
  rbind

要求されたパッケージ ROC をロード中です
次のパッケージを付け加えます: 'TCC'
以下のオブジェクトはマスクされています (from 'package:edgeR') :
  calcNormFactors

> |
```

方法2

XMLパッケージのインストール後に、もう一度library(TCC)とやって、エラーが出なくなることを確認しています。

```
R Console
要求されたパッケージ ROC をロード中です
次のパッケージを付け加えます: 'TCC'
以下のオブジェクトはマスクされています (from 'package:edgeR') :
  calcNormFactors
> library(TCC)
> |
```

エラー遭遇例とその対処法3

シロイヌナズナ (*A. thaliana*) ゲノム配列情報を含む `BSgenome.Athaliana.TAIR.TAIR9` パッケージ読み込み時にエラーが出ている例です。対処法は以下の通りです。

```
source("http://www.bioconductor.org/biocLite.R")
biocLite("BSgenome.Athaliana.TAIR.TAIR9")
```

R Console

```
> library(BSgenome.Athaliana.TAIR.TAIR9)
以下にエラー library(BSgenome.Athaliana.TAIR.TAIR9) :
  'BSgenome.Athaliana.TAIR.TAIR9' という名前のパッケージはありません
> source("http://www.bioconductor.org/biocLite.R")
Bioconductor version 2.14 (BiocInstaller 1.14.1), ?biocLite for help
> biocLite("BSgenome.Athaliana.TAIR.TAIR9")
BioC_mirror: http://bioconductor.org
Using Bioconductor version 2.14 (BiocInstaller 1.14.1), R version 3.1.0.
Installing package(s) 'BSgenome.Athaliana.TAIR.TAIR9'
  URL 'http://bioconductor.org/packages/2.14/data/annotation/bin/windows/cont$
Content type 'application/zip' length 37765491 bytes (36.0 Mb)
  開かれた URL
downloaded 36.0 Mb

The downloaded binary packages are in
  C:\Users\kadota\AppData\Local\Temp\RtmpaoaPGj\downloaded_packages
Old packages: 'AnthropMMD', 'bayesQR', 'Bclim', 'care', 'clogitL1',
'freestats', 'geomorph', 'gtools', 'investr', 'jsonlite', 'markovch
'meta', 'multicon', 'mvtnorm', 'NLP', 'openNLP', 'PBD', 'pdffetch',
'poisson.glm.mix', 'QCA3', 'Rbitcoin', 'regRSM', 'Reol', 'rgbif',
'rsm', 'RTextureMetrics', 'RWebLogo', 'sda', 'SEERaBomb', 'segmente
'seqDesign', 'seqminer', 'sjPlot', 'spcr', 'st', 'yuima'
Update all/some/none? [a/s/n]: n
> |
```

Update all/some/none? [a/s/n]:
と聞かれることもありますが基本
はnでいいです。

エラー遭遇例とその対処法3

パッケージのインストール後に、もう一度 `library(BSgenome.Athaliana.TAIR.TAIR9)` とやって、エラーが出なくなることを確認しています。

R Console

```
> library(BSgenome.Athaliana.TAIR.TAIR9)
```

```
要求されたパッケージ BSgenome をロード中です  
要求されたパッケージ BiocGenerics をロード中です  
要求されたパッケージ parallel をロード中です
```

次のパッケージを付け加えます: 'BiocGenerics'

以下のオブジェクトはマスクされています (from 'package:parallel') :

```
clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
clusterExport, clusterMap, parApply, parCapply, parLapply,  
parLapplyLB, parRapply, parSapply, parSapplyLB
```

以下のオブジェクトはマスクされています (from 'package:stats') :

```
xtabs
```

以下のオブジェクトはマスクされています (from 'package:base') :

```
anyDuplicated, append, as.data.frame, as.vector, cbind,  
colnames, do.call, duplicated, eval, evalq, Filter, Find, get,  
intersect, is.unsorted, lapply, Map, mapply, match, mget, order,  
paste, pmax, pmax.int, pmin, pmin.int, Position, rank, rbind,  
Reduce, rep.int, rownames, sapply, setdiff, sort, table, tapply,  
union, unique, unlist
```

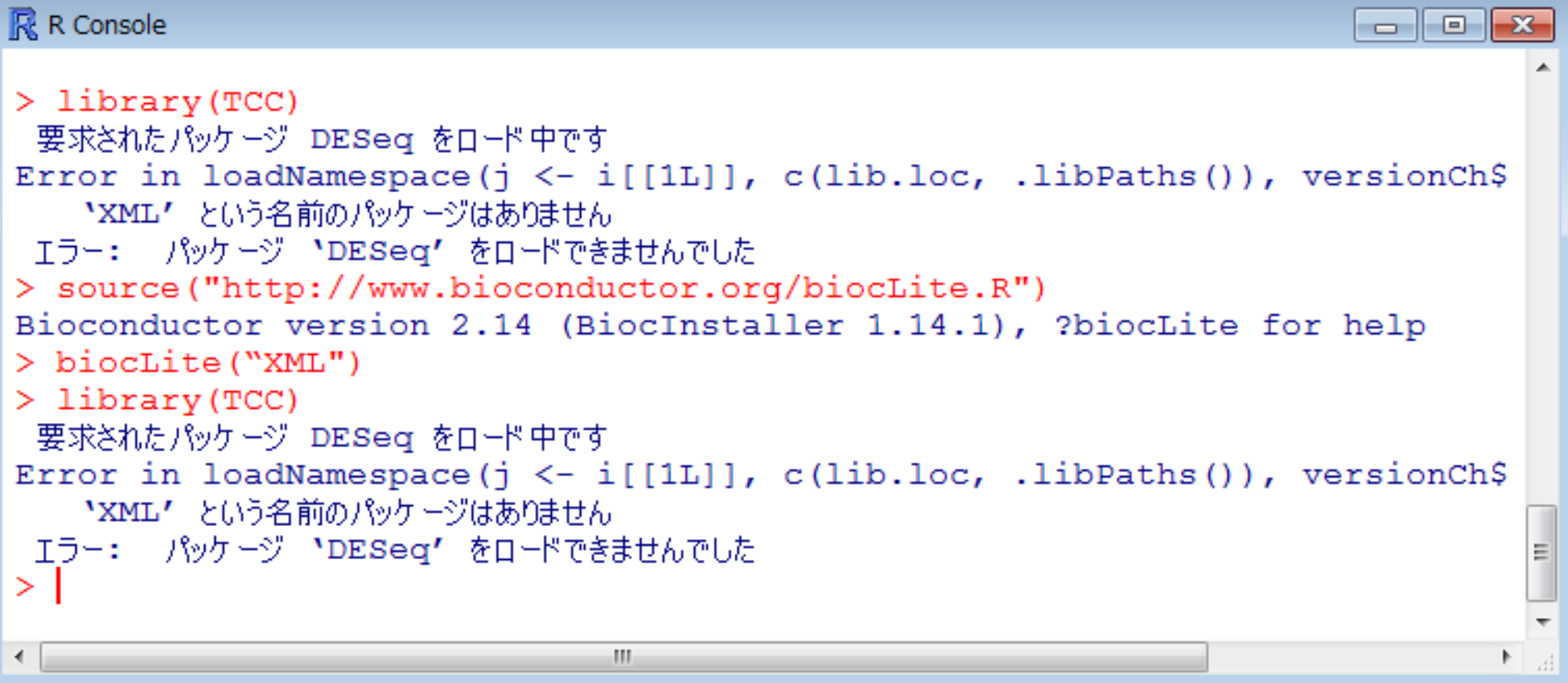
```
要求されたパッケージ IRanges をロード中です  
要求されたパッケージ GenomicRanges をロード中です  
要求されたパッケージ GenomeInfoDb をロード中です  
要求されたパッケージ Biostrings をロード中です  
要求されたパッケージ XVector をロード中です
```

```
> library(BSgenome.Athaliana.TAIR.TAIR9)
```

```
> |
```

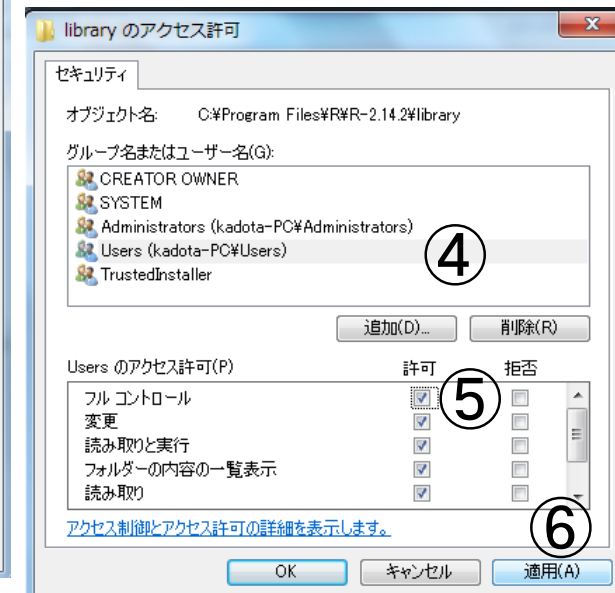
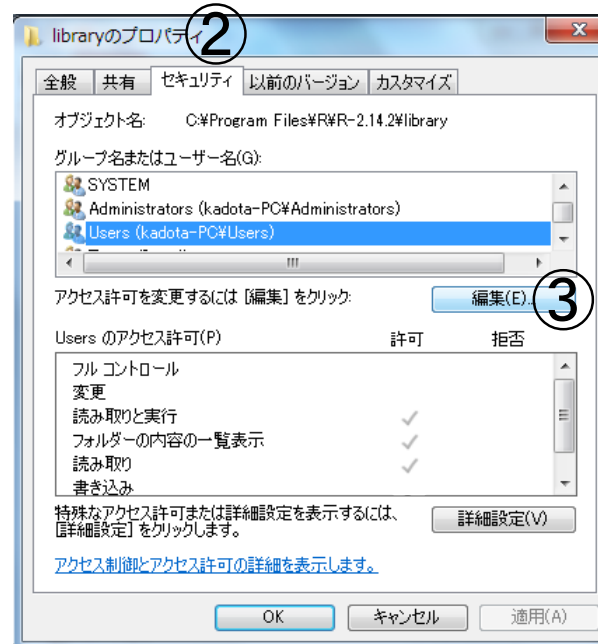
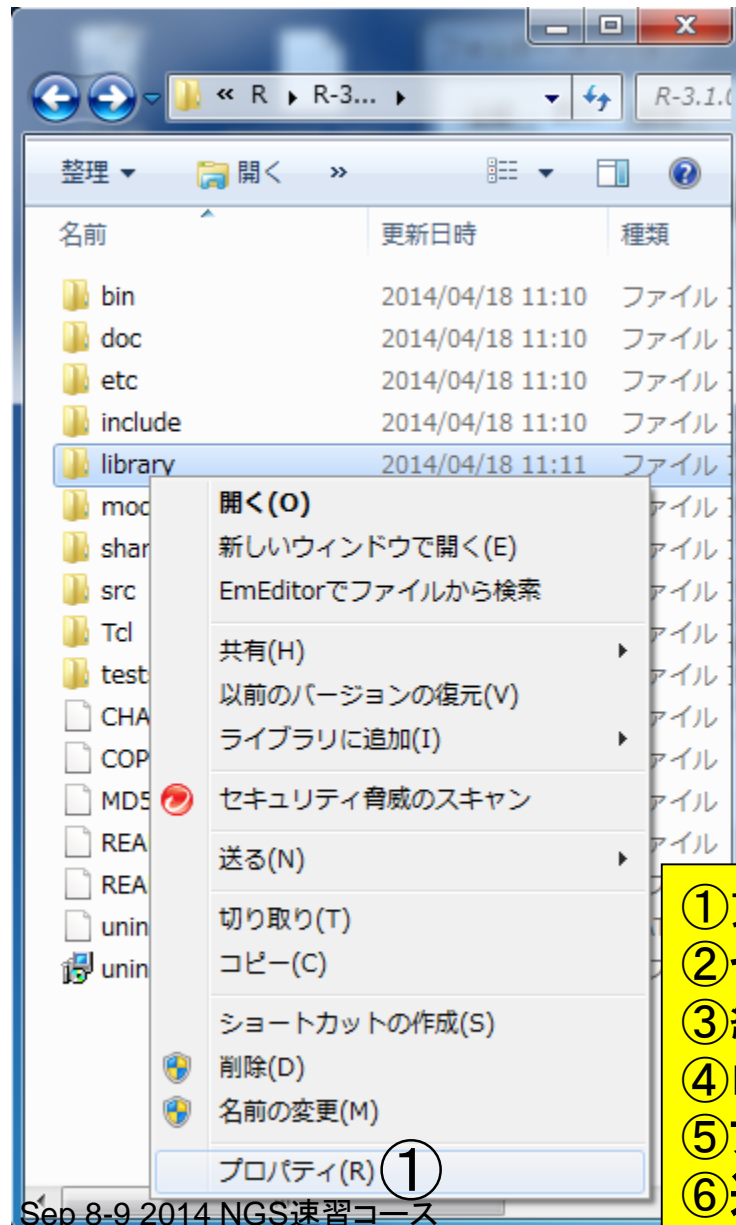
対処時の注意

パッケージを個別にインストールするためのコマンドの基本形は以下のとおりですが、**二重クォーテーションに注意!!** 以下はXMLの左側がダメな例です
source("http://www.bioconductor.org/biocLite.R")
biocLite("XML")



```
R Console
> library(TCC)
要求されたパッケージ DESeq をロード中です
Error in loadNamespace(j <- i[[1L]], c(lib.loc, .libPaths()), versionCh$
  'XML' という名前のパッケージはありません
エラー: パッケージ 'DESeq' をロードできませんでした
> source("http://www.bioconductor.org/biocLite.R")
Bioconductor version 2.14 (BiocInstaller 1.14.1), ?biocLite for help
> biocLite("XML")
> library(TCC)
要求されたパッケージ DESeq をロード中です
Error in loadNamespace(j <- i[[1L]], c(lib.loc, .libPaths()), versionCh$
  'XML' という名前のパッケージはありません
エラー: パッケージ 'DESeq' をロードできませんでした
> |
```


「“C:/Program Files/R/R-3.1.0/library”に書き込み権限がない」的なエラーが出てインストールできなかった人は、書き込み権限を取得してもう一度トライ



- ① 文句を言われたフォルダ上で右クリックでプロパティを選択
- ② セキュリティタブを選択
- ③ 編集をクリック
- ④ ログインしているユーザーを選択(フルコントロールにチェックなし)
- ⑤ フルコントロールにチェックを入れる
- ⑥ 適用をクリック

Contents

- 3-1. R 基礎1、2014/09/08 10:30-12:00、初級、実習
 - Rおよびパッケージのインストール、インストール後の確認
 - 参考図書
 - パッケージが正しくインストールされているかどうかの確認
 - エラーメッセージとその対処法
 - 基本的な利用法
 - 四則演算、スクリプトファイルの作成とコピー、改行の有無に注意
 - コメント行、上下左右の矢印キーを有効利用、エラーメッセージ
 - 関数の利用マニュアル、ベクトル計算、オプションの変更
 - 条件判定、論理値ベクトル
 - ベクトル中の任意の要素を抽出(subsetting)
 - ベクトルの要素数、組合せ(ソート後に最初の3要素を抽出)の基本



(Rで)塩基配列解析

～NGS, RNA-seq, ゲノム, トランスクリプトーム, 正規化, 発現変動, 統計, モデル, バイオインフォマティクス～
(last modified 2014/07/31, since 2010)

What's new?

- このウェブページはフリーソフトRと必要なパッケージをインストール済みである前提で記述しています。初心者
は、1. [Rのインストールと起動](#)および2. [基本的な利用法](#)で自習してください。(2014/07/21) **NEW**
- 2014年10月04日に[HPCIワークショップ「医療ゲノムデータ解析」](#)(9:00-9:20)に引き続いて [中級者向けバイオイン
フォマティクス入門講習会@仙台国際センター](#) (10:50-12:20)で話します。興味ある方はどうぞ。(2014/07/23) **NEW**
- 門田幸二 著 [シリーズ Useful R 第7巻トランスクリプトーム解析](#)刊行(共立出版)
- 2014年9月1日～12日に「[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](#)」を開催
します。 [受講申込](#)は6/24夕方に締め切りました。TA申込枠はあと数名です。(2014/07/21) **NEW**
- [参考資料\(講義, 講習会, 本など\)](#)の項目を追加しました。(2014/07/30) **NEW**

- [はじめに](#) (last
- [参考資料\(講](#)
- [過去のお知らせ](#)
- [Rのインスト](#)
- [基本的な利用](#)

基本的な利用法 **NEW**

[Rのインストールと起動](#)を参考にして必要なパッケージのインストールが完了済みのヒトを対象として、R Gui画面や
作業ディレクトリの変更、このウェブページの基本的な利用法を簡単に解説した[PDFファイル](#)を作成しました。2014年
10月4日の [中級者向けバイオインフォマティクス入門講習会@東北大学](#)(10:50-12:20)で [講習](#)希望者は、この[PDFファイ
ル](#)および [シリーズ Useful R 第7巻トランスクリプトーム解析](#)を参考にして、このウェブページの基本的な利用法をマ
スターしておいてください。

Rの起動



起動直後は画面いっぱいになるので…



```
R version 3.1.0 (2014-04-10) -- "Spring Dance"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()' あるいは 'licence()' と入力してください。

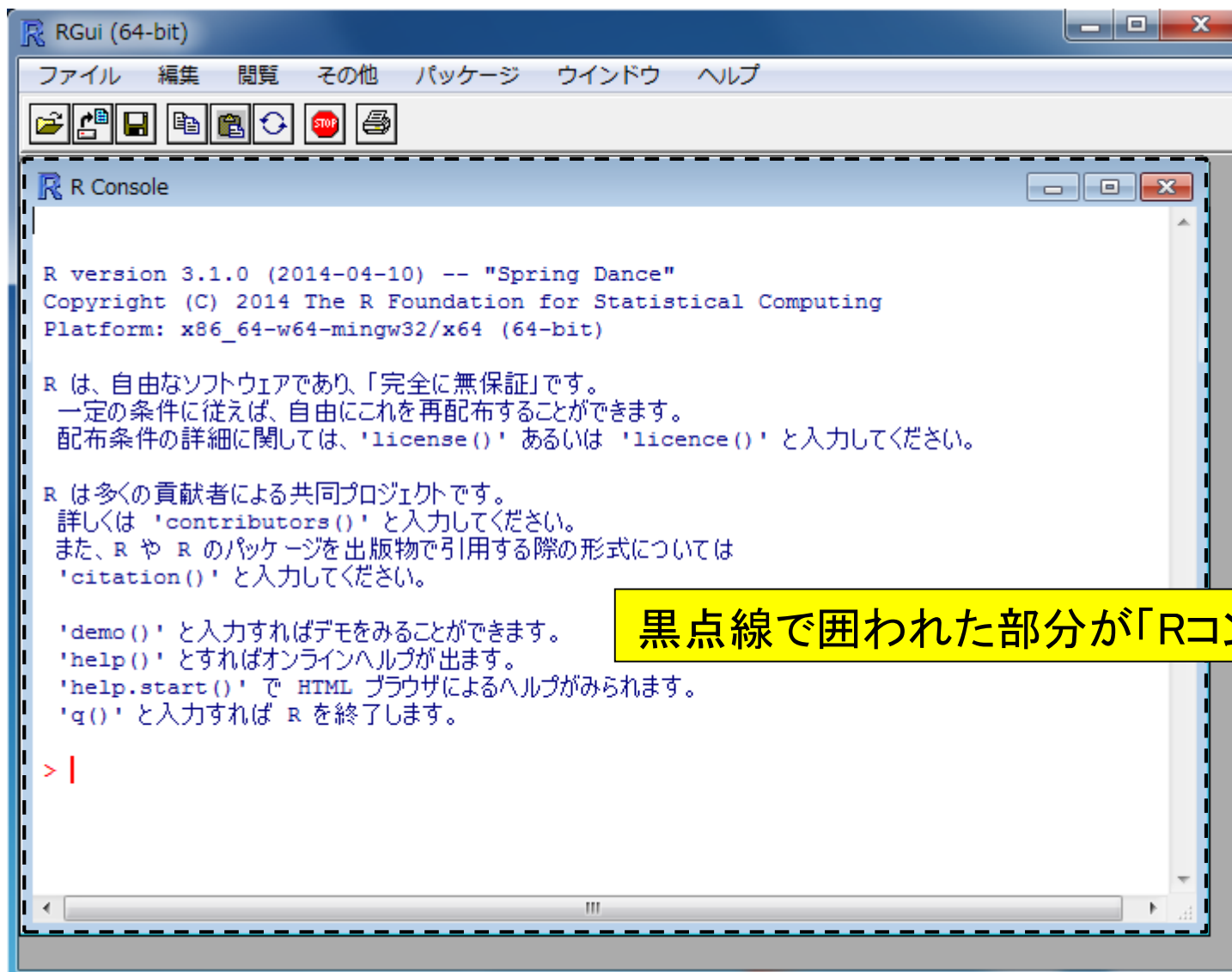
R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()' と入力してください。

'demo()' と入力すればデモをみることができます。
'help()' とすればオンラインヘルプが出ます。
'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

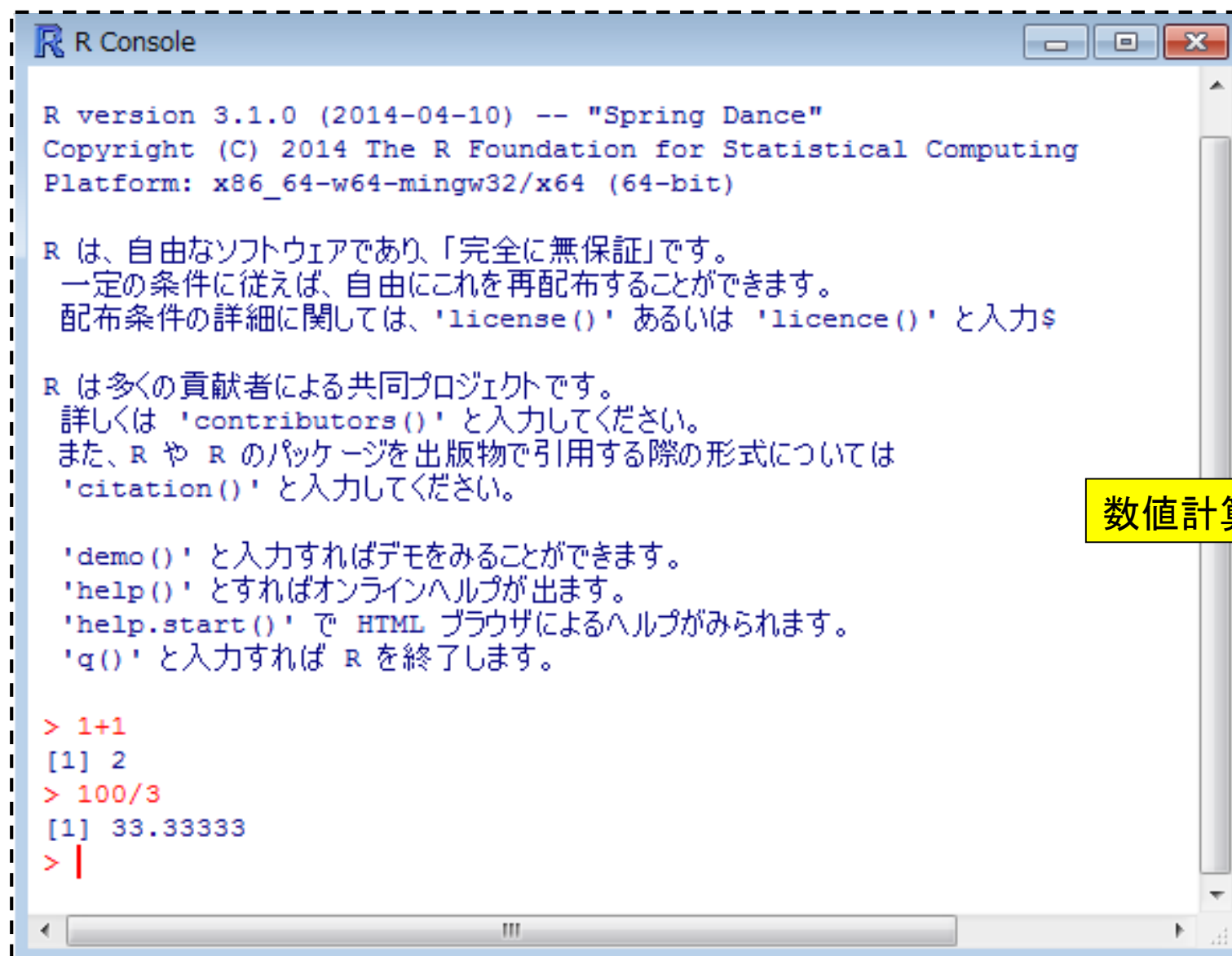
> |
```

Rのバージョン番号が多少異なるヒトもいるかもしれないが、とりあえずは気にしないでよい。「R i386 3.1.0」みたいなアイコンもできると思うが、基本的には「R x64 3.1.0」のような64 bit版のほうを利用すべし。

Rの画面説明



基本的な利用法



```
R Console

R version 3.1.0 (2014-04-10) -- "Spring Dance"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()' あるいは 'licence()' と入力$

R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()' と入力してください。

'demo()' と入力すればデモをみることができます。
'help()' とすればオンラインヘルプが出ます。
'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

> 1+1
[1] 2
> 100/3
[1] 33.33333
> |
```

数値計算ができます

基本的な利用法

```
R Console

'citation()' と入力してください。

'demo()' と入力すればデモをみることができます。
'help()' とすればオンラインヘルプが出ます。
'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

> 1+1
[1] 2
> 100/3
[1] 33.33333
>
> uge <- 100
> uge/3
[1] 33.33333
> |
```

ugeというもの(オブジェクト)に100という数値を代入しておいて、それを利用することもできます。

1	→
2	←
3	矢印
4	↓
5	⇒
6	↑
7	⇔
8	やじるし
9	野次るし

基本的な利用法

```
R Console

'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

> 1+1
[1] 2
> 100/3
[1] 33.33333
>
> uge <- 100
> uge/3
[1] 33.33333
>
> age = 100
> age/3
[1] 33.33333
> |
```

← の代わりに = を利用するのもよいです。が、門田は”←”派です。

基本的な利用法

```
R Console
>
> uge <- 100
> uge/3
[1] 33.33333
>
> age = 100
> age/3
[1] 33.33333
>
> uge<-100
> age=100
> uge
[1] 100
> age
[1] 100
> |
```

スペースを開けなくてもよいが、スペースを空けるのが一般的です。

基本的な利用法

```
R Console
[1] 33.33333
>
> uge<-100
> age=100
> uge
[1] 100
> age
[1] 100
>
> age <- 36
> uge - age
[1] 64
>
> hoge - age
エラー: オブジェクト 'hoge' がありません
> |
```

100が代入されていたageオブジェクトに対して、別の数値を新たに代入すると上書きされます。100-36 = 64となります。また、何も代入されていないhogeというオブジェクトを利用しようとしてもエラーが出るだけです。

バイオインフォ系研究者の実験ノート

- ・ [過去のお知らせ](#) (last modified 2014/08/03) **NEW**
- ・ [Rのインストールと起動](#) (last modified 2014/07/31) **NEW**
- ・ [基本的な利用法](#) (last modified 2014/07/20) **NEW**
- ・ [サンプルデータ](#) (last modified 2014/07/17) **NEW**
- ・ [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](#) (last modified 2014/07/30) **NEW**
- ・ [書籍 | トランスクリプトームについて](#) (last modified 2014/05/12)



- ・ 書籍 | トランスクリプトーム解析 | 2.2.1 RNA-seq データ (FASTQファイル) (last modified 2014/04/15)

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 **NEW**

2014年9月にJST-NBDCと東大農アグリバイオ主催で「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」が開催されます。主催機関のサイト上で情報提供したほうがいいだろうということで、受講者が各自でインストールするソフトウェアや、イメージファイルのダウンロードなど準備していただく計算機環境の情報などを示します。

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)関連:

rcode_20140908.txt(はhogeフォルダ中にあります。

- ・ [NBDCの速習コース](#)
- ・ [HPCIの速習コース](#)
- ・ カリキュラムを策定
- ・ 「NBDCで実施した」
 - 「バイオイン
 - 「カリキュラム
 - 「カリキュラム

- ・ 2014年9月5日15:00-18:15、「2-2. バイオ系データベース」
- ・ [小野浩雅 \(DBCLS\)](#)、統合TV、講義資料
- ・ 基本的な各種バイオ系データベースの理解、統合DBの利用法。
- ・ 2014年9月8日10:30-12:00、「3-1. R 基礎1」、初級、実習
- ・ [門田幸二 \(東京大学\)](#)、統合TV、[講義資料](#) (20140822, 18:04版)
- ・ Rインストール自体は基本的に終了した状態を想定しているものの、最初にlibrary(Biostrings)などいくつかの利用予定パッケージのロードを行い、パッケージのインストールがうまくいっているかどうかを確認(できていなかったヒトの同定および対処)。Rの一般的な利用法。log関数などの基本的かつ挙動を完全に把握できる関数を例として、関数内部のオプション変更や「?関数名」で利用法の幅を広げる基本テクを概観。exp, mean, median, sort, length関数。
 - 9/8-9の2日間で用いるファイル詳:[hoge.zip](#) (20140822,17:24版)
 - Rコード:[rcode_20140908.txt](#)



計算機環境構築(Linux系)

8月初旬をめどにBio-Linux

- ・ [VirtualBox](#)
 - [VirtualBoxを](#)
- ・ [Bio-Linux: Field et](#)

コピペを推奨(爆)

ファイル名: rcode_20140908.txt

```
#####  
### 基本的な利用法  
#####  
1+1↓  
100/3↓  
↓  
uge <- 100↓  
uge/3↓  
↓  
age = 100↓  
age/3↓  
↓  
uge<-100↓  
age=100↓  
uge↓  
age↓  
↓  
age <- 36↓  
uge - age↓  
↓  
hoge - age↓  
↓  
#####  
### 基本的な利用法↓  
#####  
1+1↓  
100/3↓  
↓  
uge <- 100↓  
uge/3↓  
↓  
age = 100↓  
age/3↓  
↓  
uge<-100↓  
age=100↓  
uge↓  
age↓  
↓  
age <- 36↓  
uge - age↓  
↓  
hoge - age↓  
↓
```

① コピペ

```
R Console  
[1] 33.33333  
>  
> uge<-100  
> age=100  
> uge  
[1] 100  
> age  
[1] 100  
>  
> age <- 36  
> uge - age  
[1] 64  
>  
> hoge - age  
エラー: オブジェクト 'hoge' がありません  
> |
```

- コピー Ctrl+C
- ② ペースト Ctrl+V
- コマンドのみペースト
- コピー&ペースト Ctrl+X
- ウインドウの消去 Ctrl+L
- 全て選択
- ハッファに出力 Ctrl+W
- ウインドウを常にトップに置く

スペルミスを避けるため、私は「メモ帳」などのテキストエディタに、必要なコマンド群(スクリプト、コード、Rコードなどという表現が一般的)をあらかじめ作成しておき、コピペで実行しています。このファイルがバイオインフォ系研究者の再現可能な実験ノートのようなものです。

コピペを推奨

ファイル名: rcode_20140908.txt

```
#####  
### 基本的な利用法  
#####  
1+1↓  
100/3↓  
↓  
uge <- 100↓  
uge/3↓  
↓  
age = 100↓  
age/3↓  
↓  
uge<-100↓  
age=100↓  
uge↓  
age↓  
↓  
age <- 36↓  
uge - age↓  
↓  
hoge - age↓  
↓  
#####  
### 基本的な利用法↓  
#####  
1+1↓  
100/3↓  
↓  
uge <- 100↓  
uge/3↓  
↓  
age = 100↓  
age/3↓  
↓  
uge<-100↓  
age=100↓  
uge↓  
age↓  
↓  
age <- 36↓  
uge - age↓  
↓  
hoge - age↓  
↓  
#####
```

①コピー

コピペで実行したところ。コピーする際に最後のコマンドの改行まできちんと含めるところがポイントです。

```
R Console  
[1] 100  
>  
> age <- 36  
> uge - age  
[1] 64  
>  
> hoge - age  
エラー: オブジェクト 'hoge' がありません  
> age <- 36  
> uge - age  
[1] 64  
>  
> hoge - age  
エラー: オブジェクト 'hoge' がありません  
>  
> |
```

コピペを推奨

ファイル名: rcode_20140908.txt

```
#####  
### 基本的な利用法 #####  
#####  
1+1↓  
100/3↓  
↓  
uge <- 100↓  
uge/3↓  
↓  
age = 100↓  
age/3↓  
↓  
uge<-100↓  
age=100↓  
uge↓  
age↓  
↓  
age <- 36↓  
uge - age↓  
↓  
hoge - age↓  
↓  
#####  
### 基本的な利用法↓  
#####  
1+1↓  
100/3↓  
↓  
uge <- 100↓  
uge/3↓  
↓  
age = 100↓  
age/3↓  
↓  
uge<-100↓  
age=100↓  
uge↓  
age↓  
↓  
age <- 36↓  
uge - age↓  
↓  
hoge - age↓  
↓
```

①コピー

```
R Console  
[1] 64  
>  
> hoge - age  
エラー: オブジェクト 'hoge' がない  
> age <- 36  
> uge - age  
[1] 64  
>  
> hoge - age  
エラー: オブジェクト 'hoge' がありません  
>  
> age <- 36  
> uge - age  
[1] 64  
>  
> hoge - age|
```

コピペで実行したところ。コピーする際に最後のコマンドの改行までしっかり含めるところがポイントです。この場合、改行が含まれていないので、最後のコマンドが実行されていないことがわかります。リターンキーを押して、コマンドを実行させましょう。

基本的な利用法2

ファイル名: rcode_20140908.txt

```
#####  
### 基本的な利用法2  
#####  
uge <- 100  
log(uge)  
log(uge) # デフォルトの底は自然対数  
↓  
log(uge, base=10) # 底を10にしたい場合↓  
log(uge, base = 2)#底を2にしたい場合↓  
↓  
log10(uge) # log10という関数も存在する  
log2(uge) # log2関数もある↓  
loge(uge) # 底がeの関数は存在しない。  
log(uge) # log関数のデフォルトがeだ  
↓  
log10(uge, base=2)# baseというオプションはな  
↓
```

①コピー

#から始まる行をR Console画面にコピペしてもかまいません。また、コマンドの右側に#から始まるコメントを書いてもかまいません。#のあとに関数などを書いても無視されます

```
R Console  
> #####  
> ### 基本的な利用法2  
> #####  
> uge <- 100  
> log(uge)  
[1] 4.60517  
> log(uge) # デフォルトの底は自然対数  
[1] 4.60517  
> |
```

基本的な利用法2

ファイル名:rcode_20140908.txt

```
R Console
> #####
> ### 基本的な利用法2
> #####
> uge <- 100
> log(uge)
[1] 4.60517
> log(uge)
[1] 4.60517
> uge <- 36
```

デフォルトの底は自然対数

100が代入されていたugeに別の数値をいれて挙動確認したい場合は、最初から全部打ち込み直すのではなく、上矢印キーで以前に打ち込んだコマンドを有効利用すべし！

上下左右の矢印キーを有効に利用し最小限の労力で打つべし！

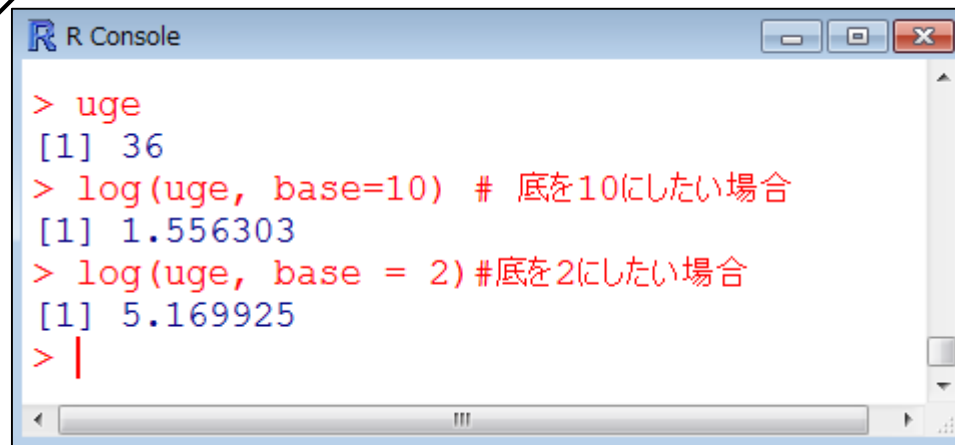


基本的な利用法2

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法2↓  
#####↓  
uge <- 100↓  
log(uge)↓  
log(uge)           # デフォルトの底は自然対数↓  
↓  
uge  
log(uge, base=10) # 底を10にしたい場合↓  
log(uge, base = 2)#底を2にしたい場合↓  
↓  
log10(uge)         # log10という関数も存在する↓  
log2(uge)          # log2関数もある↓  
loge(uge)          # 底がeの関数は存在しない...↓  
log(uge)           # log関数のデフォルトがeだから  
↓  
log10(uge, base=2)# baseというオプションはない↓  
↓
```

log関数のデフォルトの底はe (=2.718282)。しかし、baseというオプションを利用することで、任意の数値を底とすることができます。



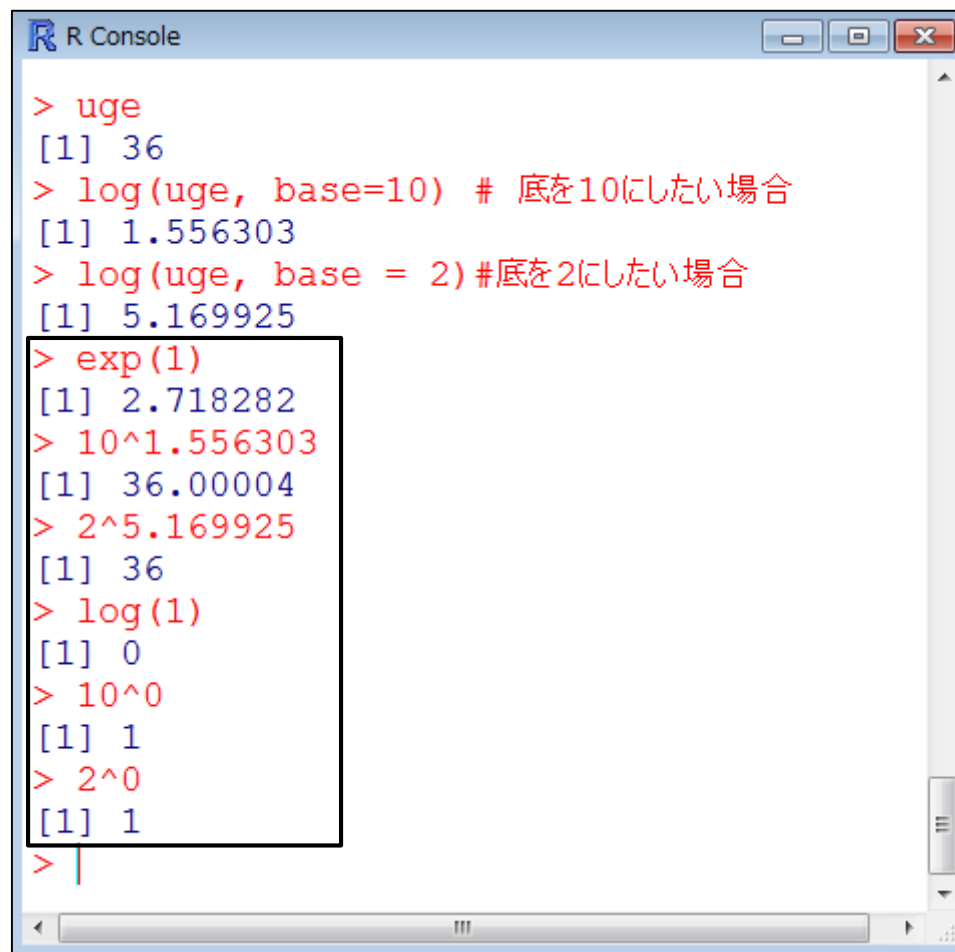
```
R Console  
> uge  
[1] 36  
> log(uge, base=10) # 底を10にしたい場合  
[1] 1.556303  
> log(uge, base = 2) #底を2にしたい場合  
[1] 5.169925  
> |
```

基本的な利用法2

おさらいです

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法2↓  
#####↓  
uge <- 100↓  
log(uge)↓  
log(uge)      # デフォルトの底は自然対数↓  
↓  
uge  
log(uge, base=10) # 底を10にしたい場合↓  
log(uge, base = 2)#底を2にしたい場合↓  
↓  
log10(uge)     # log10という関数も存在する↓  
log2(uge)     # log2関数もある↓  
loge(uge)     # 底がeの関数は存在しない...↓  
log(uge)     # log関数のデフォルトがeだから  
↓  
log10(uge, base=2)# baseというオプションはない↓  
↓
```



```
R Console  
> uge  
[1] 36  
> log(uge, base=10) # 底を10にしたい場合  
[1] 1.556303  
> log(uge, base = 2) #底を2にしたい場合  
[1] 5.169925  
> exp(1)  
[1] 2.718282  
> 10^1.556303  
[1] 36.00004  
> 2^5.169925  
[1] 36  
> log(1)  
[1] 0  
> 10^0  
[1] 1  
> 2^0  
[1] 1  
> |
```

基本的な利用法2

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法2↓  
#####↓  
uge <- 100↓  
log(uge)↓  
log(uge)           # デフォルトの底は自然対数↓  
↓  
uge↓  
log(uge, base=10) # 底を10にしたい場合↓  
log(uge, base = 2)#底を2にしたい場合↓  
↓  
log10(uge)         # log10という関数も存在する  
log2(uge)          # log2関数もある  
loge(uge)          # 底がeの関数は存在しない...  
log(uge)           # log関数のデフォルトがeだから  
↓  
log10(uge, base=2)# baseというオプションはない↓  
↓
```

一つの目的を達成するうえでも多くのやり方が存在します。例えば、log関数利用時にbase=10を利用するのとlog10関数を利用するのは同じです。また存在しない関数を実行すればエラーが出ます。重要なのはエラーにうろたえるのではなく、出たメッセージとの関連を適切に把握し、対処すること。

```
R Console  
> uge  
[1] 36  
> log(uge, base=10) # 底を10にしたい場合  
[1] 1.556303  
> log(uge, base = 2) # 底を2にしたい場合  
[1] 5.169925  
> log10(uge)         # log10という関数も存在する  
[1] 1.556303  
> log2(uge)         # log2関数もある  
[1] 5.169925  
> loge(uge)         # 底がeの関数は存在しない...  
エラー: 関数 "loge" を見つけることができませんでした  
> log(uge)         # log関数のデフォルトがeだから  
[1] 3.583519  
> |
```

基本的な利用法2

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法2↓  
#####↓  
uge <- 100↓  
log(uge)↓  
log(uge)          # デフォルトの底は自然対数↓  
↓  
uge↓  
log(uge, base=10) # 底を10にしたい場合↓  
log(uge, base = 2)#底を2にしたい場合↓  
↓  
log10(uge)        # log10という関数も存在する↓  
log2(uge)         # log2関数もある↓  
loge(uge)         # 底がeの関数は存在しない...↓  
log(uge)          # log関数のデフォルトがeだから  
↓  
log10(uge, base=2)# baseというオプションはない↓  
↓
```

log10という関数は底が10の対数をとるものなので、baseオプションで底を指定するという行為自体が非論理的。数値のみを入力とするのが妥当。これは1個の引数(argument)のみをlog10関数に与えるべきということに相当する。

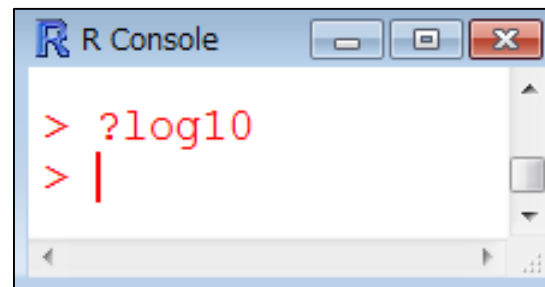
```
R Console  
> log10(uge, base=2) # baseというオプションはない  
以下にエラー log10(uge, base = 2) :  
  2 個の引数が 'log10' に渡されましたが、1 が必要とされています  
> |
```

基本的な利用法3

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法3↓  
#####↓  
?log10          # log10関数の使用法を調べる↓  
↓  
help(log10)     # log10関数の使用法を調べる↓  
↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
x               # xの中身を表示↓  
mean(x)        # 平均値↓  
(3+2+9+2)/4    # 平均値↓  
average(x)     # エクセルと違ってaverageはない↓  
median(x)      # 中央値↓  
log2(x)        # ベクトルの要素ごとに計算↓  
↓  
sort(x)↓  
sort(x, decreasing = FALSE)↓  
sort(x, decreasing=FALSE)↓  
sort(x, decreasing=F)↓  
sort(x, decreasing=TRUE)↓  
sort(x, decreasing=T)↓  
↓
```

関数ごとに利用可能なオプションや、どのようなデータを入力として受け付けるのかなどを記したhtmlマニュアルが存在します。それを見る手段は「?関数名」または「help(関数名)」です。



```
R Console  
> ?log10  
> |
```

Logarithms and Exponentials

Description

log computes logarithms, by default natural logarithms, log10 computes common (i.e., base 10) logarithms, and log2 computes binary (i.e., base 2) logarithms. The general form log(x, base) computes logarithms with base base.

log1p(x) computes $\log(1+x)$ accurately also for $|x| \ll 1$.

exp computes the exponential function.

expm1(x) computes $\exp(x) - 1$ accurately also for $|x| \ll 1$.

Usage

```
log(x, base = exp(1))
logb(x, base = exp(1))
log10(x)
log2(x)
```

```
log1p(x)
```

```
exp(x)
expm1(x)
```

Arguments

x a numeric or complex vector.

base a positive or complex number: the base with respect to which logarithms are computed. Defaults to $e = \exp(1)$.

Descriptionには関数の概要が書かれている。

Usageには基本的な利用法が書かれている。log関数には引数が2つ、log10関数には引数が1つしかないことはここでもわかる。

Argumentsには引数に関する説明が書かれている。baseオプションのデフォルトがeであることはここでもわかる。

全く知らない関数のマニュアルを見てもさっぱりわからない(個人の感想です)ため、この例のようにある程度挙動既知の関数を眺め、記述形式に慣れておくとよい。

Details

続き...

All except `logb` are generic functions: methods can be defined for them individually or via the [Math](#) group generic.

`log10` and `log2` are only convenience wrappers, but logs to bases 10 and 2 (whether computed *via* `log` or the wrappers) will be computed more efficiently and accurately where supported by the OS. Methods can be set for them individually (and otherwise methods for `log` will be used).

`logb` is a wrapper for `log` for compatibility with S. If (S3 or S4) methods are set for `log` they will be dispatched. Do not set S4 methods on `logb` itself.

All except `log` are [primitive](#) functions.

Value

A vector of the same length as `x` containing the transformed values. `log(0)` gives `-Inf`, and `log(x)` for negative values of `x` is `NaN`. `exp(-Inf)` is 0.

For complex inputs to the log functions, the value is a complex number with imaginary part in the range $[-\pi, \pi]$: which end of the range is used might be platform-specific.

S4 methods

`exp`, `expm1`, `log`, `log10`, `log2` and `log1p` are S4 generic and are members of the [Math](#) group generic.

Note that this means that the S4 generic for `log` has a signature with only one argument, `x`, but that `base` can be passed to methods (but will not be used for method selection). On the other hand, if you only set a method for the `Math` group generic then `base` argument of `log` will be ignored for your class.

詳細な説明。関数によってあったりなかったりします。

関数の返り値に関する説明。この関数を実行するとどんな結果が返されるかという説明書き。

...

`log1p` and `expm1` may be taken from the operating system, but if not available there then they are based on the Fortran subroutine `dlnrel` by W. Fullerton of Los Alamos Scientific Laboratory (see <http://www.netlib.org/slatec/fnlib/dlnrel.f> and (for small `x`) a single Newton step for the solution of $\log_{1p}(y) = x$ respectively.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole. (for `log`, `log10` and `exp`.)

Chambers, J. M. (1998) *Programming with Data. A Guide to the S Language*. Springer. (for `logb`.)

See Also

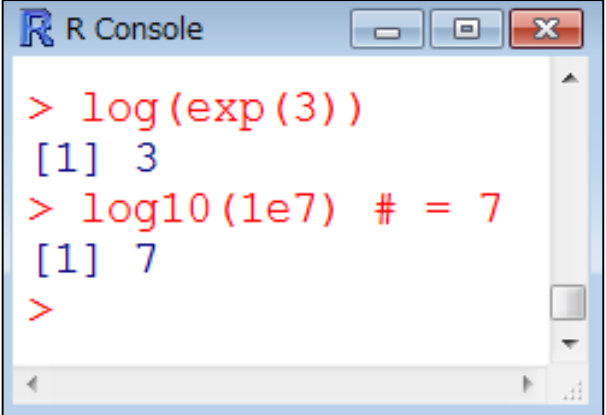
[Trig](#), [sqrt](#), [Arithmetic](#).

Examples

```
log(exp(3))
log10(1e7) # = 7

x <- 10^-(1+2*1:9)
cbind(x, log(1+x), log1p(x), exp(x)-1, expm1(x))
```

[Package *base* version 3.1.0 [Index](#)]



```
R Console
> log(exp(3))
[1] 3
> log10(1e7) # = 7
[1] 7
>
```

利用例。この記述をコピーして挙動確認したりします。

基本的な利用法3

ベクトル演算ができます

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法3↓  
#####↓  
?log10          # log10関数の使用法を調べる↓  
↓  
help(log10)     # log10関数の使用法を調べる↓  
↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
x               # xの中身を表示↓  
mean(x)         # 平均値↓  
(3+2+9+2)/4    # 平均値↓  
average(x)      # エクセルと違ってaverageはない↓  
median(x)       # 中央値↓  
log2(x)         # ベクトルの要素ごとに計算↓  
↓  
sort(x)↓  
sort(x, decreasing = FALSE)↓  
sort(x, decreasing=FALSE)↓  
sort(x, decreasing=F)↓  
sort(x, decreasing=TRUE)↓  
sort(x, decreasing=T)↓  
↓
```

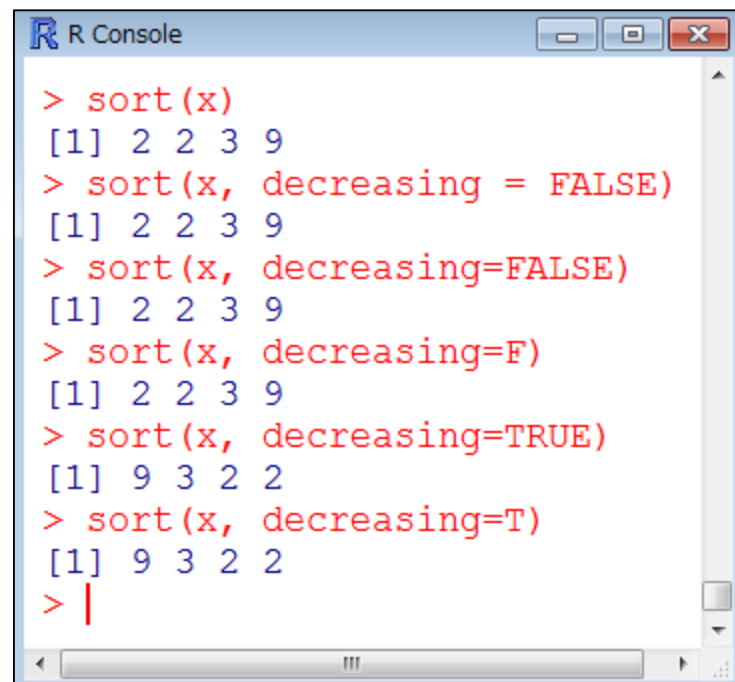
```
R Console  
  
> x <- c(3, 2, 9, 2) # 数値ベクトルxを作成  
> x                 # xの中身を表示  
[1] 3 2 9 2  
> mean(x)           # 平均値  
[1] 4  
> (3+2+9+2)/4      # 平均値  
[1] 4  
> average(x)       # エクセルと違ってaverageはない  
エラー: 関数 "average" を見つけることができませんで$  
> median(x)        # 中央値  
[1] 2.5  
> log2(x)          # ベクトルの要素ごとに計算  
[1] 1.584963 1.000000 3.169925 1.000000  
> |
```

基本的な利用法3

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法3↓  
#####↓  
?log10          # log10関数の使用法を調べる↓  
↓  
help(log10)     # log10関数の使用法を調べる↓  
↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
x               # xの中身を表示↓  
mean(x)        # 平均値↓  
(3+2+9+2)/4    # 平均値↓  
average(x)     # エクセルと違ってaverageはない↓  
median(x)      # 中央値↓  
log2(x)        # ベクトルの要素ごとに計算↓  
↓  
sort(x)↓  
sort(x, decreasing = FALSE)↓  
sort(x, decreasing=FALSE)↓  
sort(x, decreasing=F)↓  
sort(x, decreasing=TRUE)↓  
sort(x, decreasing=T)↓  
↓
```

最初の2行分で、sort関数のデフォルトはdecreasing = FALSEであることが分かります。decreasingオプションは、「ソートの際降順にするか否か」なので、FALSEを指定することは「降順にしない」つまり「昇順でソート」を意味します。



```
R Console  
> sort(x)  
[1] 2 2 3 9  
> sort(x, decreasing = FALSE)  
[1] 2 2 3 9  
> sort(x, decreasing=FALSE)  
[1] 2 2 3 9  
> sort(x, decreasing=F)  
[1] 2 2 3 9  
> sort(x, decreasing=TRUE)  
[1] 9 3 2 2  
> sort(x, decreasing=T)  
[1] 9 3 2 2  
> |
```

基本的な利用法4

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法4↓  
#####↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
x > 2.5           # 2.5より大きいかな否か↓  
x < 2.5           # 2.5より小さいかな否か↓  
↓  
k <- c(-3, 2, 9, 2) # 数値ベクトルxを作成↓  
k > -2.5          # -2.5より大きいかな否か↓  
k < -2.5          # -2.5より小さいかな否か↓  
↓  
obj <- k < -2.5    # -2.5より小さいかな否かという論理値↓  
obj              # objの中身を表示↓  
obj <- (k < -2.5) # ()で囲って見やすくしてもよい↓  
obj              # objの中身を表示↓  
↓
```

指定した条件を満たす要素の位置をTRUEまたはFALSEからなるベクトル(論理値ベクトル)で表すことができます。「~より大きい」の”>”や「~より小さい」の”<”以外にも、「~以上」の”>=”とか「~以下」の”<=”なども指定可能です。

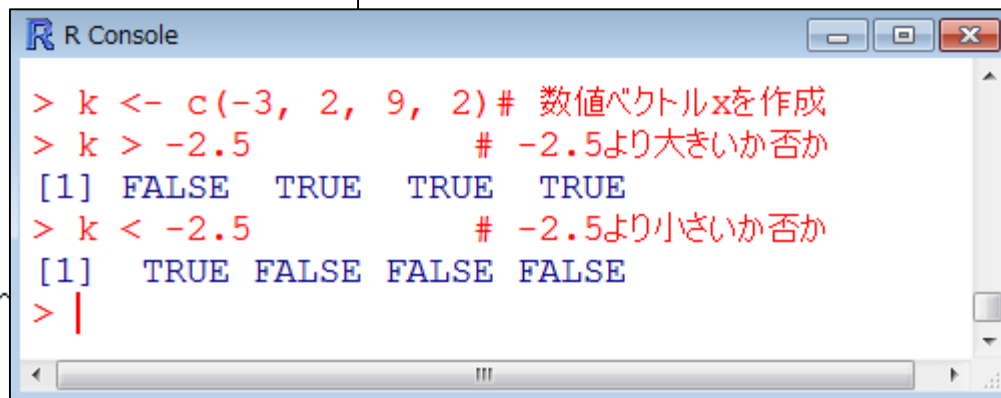
```
R Console  
> x <- c(3, 2, 9, 2) # 数値ベクトルxを作成  
> x > 2.5           # 2.5より大きいかな否か  
[1] TRUE FALSE TRUE FALSE  
> x < 2.5           # 2.5より小さいかな否か  
[1] FALSE TRUE FALSE TRUE  
> |
```

基本的な利用法4

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法4↓  
#####↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
x > 2.5           # 2.5より大きいかな否か↓  
x < 2.5           # 2.5より小さいかな否か↓  
↓  
k <- c(-3, 2, 9, 2) # 数値ベクトルxを作成↓  
k > -2.5          # -2.5より大きいかな否か↓  
k < -2.5          # -2.5より小さいかな否か↓  
↓  
obj <- k < -2.5    # -2.5より小さいかな否かという論理値↓  
obj              # objの中身を表示↓  
obj <- (k < -2.5) # ()で囲って見やすくしてもよい↓  
obj              # objの中身を表示↓  
↓
```

もちろんマイナスの要素を含むベクトルkでも問題なく条件を満たすかどうかを判定可能です。



```
R Console  
> k <- c(-3, 2, 9, 2) # 数値ベクトルxを作成  
> k > -2.5           # -2.5より大きいかな否か  
[1] FALSE TRUE TRUE TRUE  
> k < -2.5           # -2.5より小さいかな否か  
[1] TRUE FALSE FALSE FALSE  
> |
```

基本的な利用法4

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法4↓  
#####↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
x > 2.5           # 2.5より大きいかな否か↓  
x < 2.5           # 2.5より小さいかな否か↓  
↓  
k <- c(-3, 2, 9, 2) # 数値ベクトルxを作成↓  
k > -2.5          # -2.5より大きいかな否か↓  
k < -2.5        # -2.5より小さいかな否か↓  
↓  
obj <- k < -2.5  # -2.5より小さいかな否かという論理値ベクトルをobjに格納  
obj              # objの中身を表示↓  
obj <- (k < -2.5) # ()で囲って見やすくしてもよい↓  
obj             # objの中身を表示↓  
↓
```

赤色の下線部分が論理値ベクトルの部分。これをobjに格納することもできる。ただし、条件判定を意味する”<“か代入を意味する”<-“がわかりづらいため、見やすくすることを目的として括弧をつけたりすることもあります。

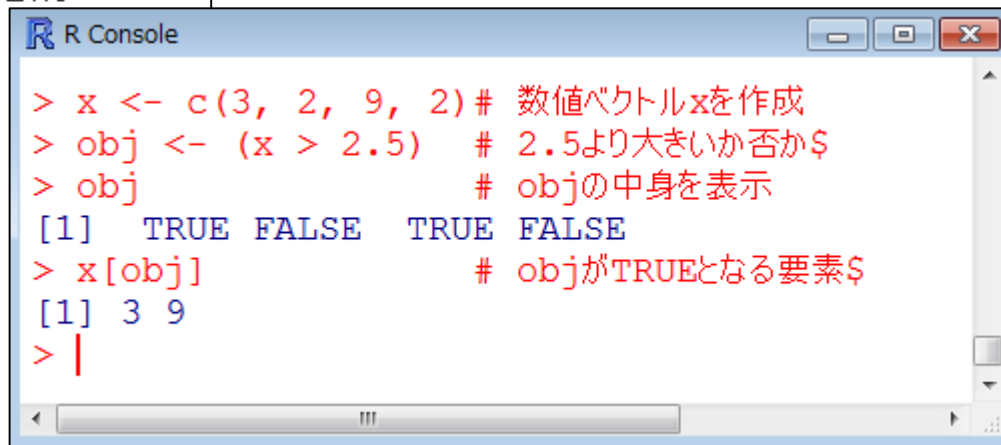
```
R Console  
> obj <- k < -2.5 # -2.5より小さいかな否か$  
> obj            # objの中身を表示  
[1] TRUE FALSE FALSE FALSE  
> obj <- (k < -2.5) # ()で囲って見やすく$  
> obj            # objの中身を表示  
[1] TRUE FALSE FALSE FALSE  
> |
```

基本的な利用法5

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法5↓  
#####↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
obj <- (x > 2.5) # 2.5より大きいかな否かの結果をobjに格納↓  
obj # objの中身を表示↓  
x[obj] # objがTRUEとなる要素のみを表示↓  
↓  
posi <- c(1,3) # 数値ベクトルposiを作成↓  
posi # posiの中身を表示↓  
x[posi] # posiで指定した要素番号を表示↓  
↓  
posi <- 2:4 # 数値ベクトルposiを作成↓  
posi # posiの中身を表示↓  
x[posi] # posiで指定した要素番号を表示↓  
↓
```

論理値ベクトルobjの実際の利用例です。「4つの要素からなるベクトル」を「4つの遺伝子からなる、あるサンプルの発現ベクトル」と考えて、発現レベルが一定以上を満たすもののみを抽出するようなイメージです。



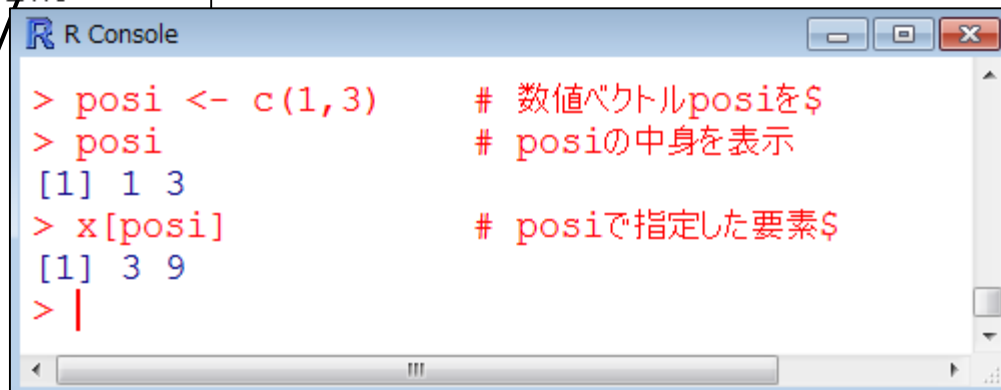
```
R Console  
> x <- c(3, 2, 9, 2) # 数値ベクトルxを作成  
> obj <- (x > 2.5) # 2.5より大きいかな否か$  
> obj # objの中身を表示  
[1] TRUE FALSE TRUE FALSE  
> x[obj] # objがTRUEとなる要素$  
[1] 3 9  
> |
```

基本的な利用法5

posiで指定した特定の位置の要素のみ抽出する基本形です。

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法5↓  
#####↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
obj <- (x > 2.5) # 2.5より大きいかなの結果をobjに格納↓  
obj # objの中身を表示↓  
x[obj] # objがTRUEとなる要素のみを表示↓  
↓  
posi <- c(1,3) # 数値ベクトルposiを作成↓  
posi # posiの中身を表示↓  
x[posi] # posiで指定した要素番号を表示↓  
↓  
posi <- 2:4 # 数値ベクトルposiを作成↓  
posi # posiの中身を表示↓  
x[posi] # posiで指定した要素番号を表示↓  
↓
```



```
R Console  
> posi <- c(1,3) # 数値ベクトルposiを$  
> posi # posiの中身を表示  
[1] 1 3  
> x[posi] # posiで指定した要素$  
[1] 3 9  
> |
```

基本的な利用法5

「2:4」は「c(2,3,4)」と同じ意味です。「1:1000」のような数値範囲が大きくなる場合に便利な表記法です。

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法5↓  
#####↓  
x <- c(3, 2, 9, 2) # 数値ベクトルxを作成↓  
obj <- (x > 2.5) # 2.5より大きいかなの結果をobjに格納↓  
obj # objの中身を表示↓  
x[obj] # objがTRUEとなる要素のみを表示↓  
↓  
posi <- c(1,3) # 数値ベクトルposiを作成↓  
posi # posiの中身を表示↓  
x[posi] # posiで指定した要素番号を表示↓  
↓  
posi <- 2:4 # 数値ベクトルposiを作成↓  
posi # posiの中身を表示↓  
x[posi] # posiで指定した要素番号を表示↓  
↓
```

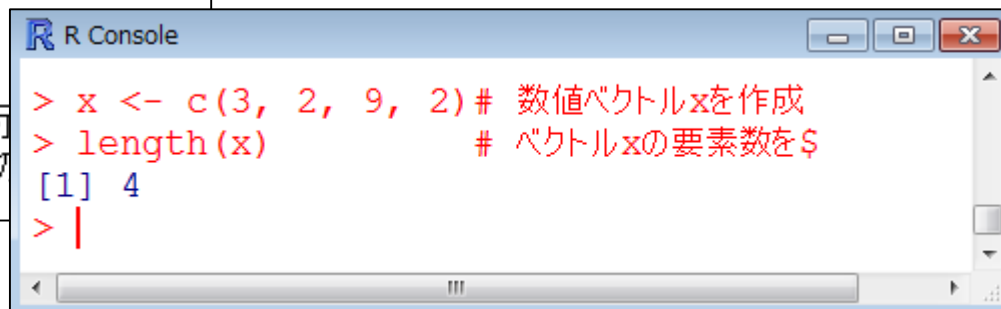
```
R Console  
> posi <- 2:4 # 数値ベクトルposiを$  
> posi # posiの中身を表示  
[1] 2 3 4  
> x[posi] # posiで指定した要素$  
[1] 2 9 2  
> |
```


基本的な利用法6

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法6↓  
#####↓  
x <- c(3, 2, 9, 2)# 数値ベクトルxを作成↓  
length(x)          # ベクトルxの要素数を表示↓  
↓  
posi <- 1:3        # 数値ベクトルposiを作成↓  
sort(x)[posi]      # xを昇順にソートし、最初の3つを表示  
sort(x, decreasing=T)[posi]# xを降順にソートし、最初の  
↓
```

length関数はベクトルの要素数を返します。ベクトルの要素数分だけ何かをしたいときにlength関数を利用することがあります。



```
R Console  
> x <- c(3, 2, 9, 2)# 数値ベクトルxを作成  
> length(x)          # ベクトルxの要素数を$  
[1] 4  
> |
```

基本的な利用法6

ファイル名: rcode_20140908.txt

```
#####↓  
### 基本的な利用法6↓  
#####↓  
x <- c(3, 2, 9, 2)# 数値ベクトルxを作成↓  
length(x)        # ベクトルxの要素数を表示↓  
↓  
posi <- 1:3      # 数値ベクトルposiを作成↓  
sort(x)[posi]    # xを昇順にソートし、最初の3つを表示↓  
sort(x, decreasing=T)[posi]# xを降順にソートし、最初の3つを表示
```

sort関数で昇順にソートし、最初の3つの要素を表示する組合せ(発展形)です。発現レベルの低い順や高い順に並べ、上位3つを表示させることと同義。

```
R Console  
> posi <- 1:3      # 数値ベクトルposiを$  
> sort(x)[posi]    # xを昇順にソートし、$  
[1] 2 2 3  
> sort(x, decreasing=T)[posi]# xを降順に$  
[1] 9 3 2  
> |
```