



# ゲノム情報解析基礎

## ～ バイオインフォマティクス基礎知識 (Win版) ～



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プログラム

門田幸二(かどた こうじ)

[kadota@iu.a.u-tokyo.ac.jp](mailto:kadota@iu.a.u-tokyo.ac.jp)

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

# 講義予定

- 4月11日月曜日(17:15-20:30)PC使用
  - 嶋田透:ゲノムからの遺伝子予測
  - 門田幸二:バイオインフォマティクス基礎知識、Rのイントロダクション
- 4月18日月曜日(17:15-20:30)PC使用
  - 門田幸二:Rで塩基配列解析1、multi-FASTAファイルの各種解析
- 4月25日月曜日(17:15-20:30)PC使用
  - 嶋田透:ゲノムアノテーション、遺伝子の機能推定、RNA-seqなどによる発現解析、比較ゲノム解析
  - 門田幸二:Rで塩基配列解析2、Rパッケージ、k-mer解析の基礎
- 5月02日月曜日(17:15-19:00頃)PC使用
  - 勝間進:非コードRNA、小分子RNA、エピジェネティクス
  - 講義後、小テスト

# 各講義科目へのアクセス

①教育プログラム、②各講義のページ、③「ゲノム情報解析基礎」の場合



# バイオインフォ関連情報

- ①「ゲノム情報解析基礎」のページ。
- ②前半はこのページを使います

## 2. ゲノム情報解析基礎

①

### 授業の目標・概要

次世代シーケンサーの普及により、ゲノム情報を基盤とした膨大な塩基配列情報を自在に解析するスキルが要求される時代になっています。フリーソフトウェアを用いて、配列決定後の基礎情報取得など各種配列解析の基本スキル向上を目指した実習を含む講義を行います。また、ウェブツールなどを用いて遺伝子領域の予測やアノテーションなどゲノム情報を比較または解析するための手法について解説します。

### 担当教員

嶋田 透 (東大・農・生産・環境生物学専攻 / 教授)  
勝間 進 (東大・農・生産・環境生物学専攻 / 准教授)  
門田幸二 (東大・農・アグリバイオ / 特任准教授)

### 参考図書

門田幸二 著 (金明哲 編)、「シリーズ Useful R ⑦ トランスクリプトーム解析」、共立出版、2014年。ISBN:978-4-320-12370-0

### お知らせ

講義では、Rの様々なパッケージを利用します。持ち込み用PC利用希望者は [インストール | について](#) を参考にしてR本体および必要なパッケージ群を必ずインストールしておいてください。

### 講義日程 (平成28年度)

- 平成28年04月11日 (PC使用)  
講師：嶋田 透  
講師：門田幸二  
[バイオインフォマティクス基礎知識](#)
- 平成28年04月18日 (PC使用)  
講師：門田幸二
- 平成28年04月25日 (PC使用)  
講師：嶋田 透  
講師：門田幸二
- 平成28年05月02日 (PC使用)  
講師：勝間 進

②

[ゲノム情報解析基礎](#) バイオインフォマティクス基礎知識 (2016年04月11日) 門田幸二

### バイオインフォマティクス系学会など

- JSBI(日本バイオインフォマティクス学会)
  - 「最新イベント情報」欄が有用。学会に入るとメーリングリストでも流れる。2016年09/29-10/01に第5回生命医薬情報学連合大会が開催される。
- 定量生物学の会
- 生命情報科学若手の会
- NGS現場の会
- オープンバイオ研究会
- CBI(情報計算化学生物学会)
- 情報処理学会 バイオ情報学研究会 (SIG BIO)
- 人工知能学会第二種研究会 分子生物情報研究会 (SIG-MBI)

### バイオインフォマティクス系?よろず相談所

- [ライフサイエンスQA](#)
  - 過疎ってる印象。。。
- [Bio Technical フォーラム](#)
  - 主に実験系だがときどきバイオインフォ系トピックも。。。活発な印象。

### 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)

- バイオサイエンスデータベースセンター(NBDC)運営委員会人材育成分科会で、2014年3月に策定されたNGS用カリキュラムが存在。このカリキュラムをベースにして、平成26年度よりNBDCと東大アグリバイオ主催でハンズオン(ノートPCを用いた実習型)講義を実施。おそらくアグリバイオ本体に次ぐ受講人数規模。
- NGS速習コース講習会(平成26年度)
  - NGS用カリキュラムに沿った内容を東大農で2014年9月1日~12日に実施
  - 座学(講義のみ)を含む。講師数10人。
  - 講習会映像は統合TVとYoutubeから公開
  - 報告書
- NGS/ハンズオン講習会(平成27年度)
  - 東大農で2015年07月22日~08月06日(A日程)および08月26-28日(B日程)に実施
  - 座学部分は全て自習にしてハンズオンに特化。講師数を減らして全体の連携を強化(10→4人)
  - 講習会映像は統合TVとYoutubeから公開

# 学会

① バイオインフォマティクスの名前を冠した学会があります。  
 ② 今年の年会は2016年9月末

ゲノム情報解析基礎 バイオインフォマティクス基礎知識 (2016年04月11日) 門田幸二

## バイオインフォマティクス系学会など

- ① JSBI(日本バイオインフォマティクス学会)
- 「最新イベント情報」欄が有用。学会に入るとメーリングリストでも流れる。
  - 2016年09/29-10/01に第5回生命医薬情報学連合大会が開催される。
- ②
- 定量生物学の会
  - 生命情報科学若手の会
  - NGS現場の会
  - オープンバイオ研究会
  - CBI(情報計算化学生物学会)
  - 情報処理学会 バイオ情報学研究会 (SIG BIO)
  - 人工知能学会第二種研究会 分子生物情報研究会 (SIG-MBI)

## バイオインフォマティクス系?よろず相談所

- ライフサイエンスQA
  - 過疎ってる印象。。
- Bio Technical フォーラム
  - 主に実験系だがときどきバイオインフォ系トピックも。。活発な印象。

## 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)

- バイオサイエンスデータベースセンター(NBDC)運営委員会人材育成分科会で、2014年3月に策定されたNGS用カリキュラムが存在。このカリキュラムをベースにして、平成26年度よりNBDCと東大アグリバイオ主催でハンズオン(ノートPCを用いた実習型)講義を実施。おそらくアグリバイオ本体に次ぐ受講人数規模。
- NGS速習コース講習会(平成26年度)
  - NGS用カリキュラムに沿った内容を東大農で2014年9月1日~12日に実施
  - 座学(講義のみ)を含む。講師数10人。
  - 講習会映像は統合TVとYoutubeから公開
  - 報告書
- NGS/ハンズオン講習会(平成27年度)
  - 東大農で2015年07月22日~08月06日(A日程)および08月26-28日(B日程)に実施
  - 座学部分は全て自習にしてハンズオンに特化。講師数を減らして全体の連携を強化(10 → 4人)
  - 講習会映像は統合TVとYoutubeから公開

JSBI 特定非営利活動法人 日本バイオインフォマティクス学会 Japanese Society for Bioinformatics

検索 English

ホーム お問い合わせ FAQ サイトマップ

バイオインフォマティクスに関する学問分野の発展    バイオインフォマティクスに関する技術および関連事業の振興    バイオインフォマティクス教育の基盤確立と充実に寄与

日本バイオインフォマティクス学会は、これらを目的として設立されました。詳しくは学会案内をご覧ください。

Japanese Society for Bioinformatics

お知らせ イベント案内 研究会・地域部会 バイオインフォマティクス技術者認定試験 学会案内 入会・継続案内

公募のお知らせ (人事関連)

- 山口大学大学院医学系研究科 バイオインフォマティクス・医学統計分野(仮称) 教授公募 (2016-02-23)
- 国立遺伝学研究所 助教募集 (2016-02-04)
- 九州大学マス・フォア・インダストリ研究所・カーボンニュートラル・エネルギー 国際研究所テニュアトラック制助教公募 (2016-01-12)
- 田辺三菱製薬株式会社 バイオインフォマティクス正社員の募集 (2015-11-30)
- 教員の公募(准教授) 東京工業大学 生命理工学院 生命理工学系 生命理工学コース (2015-11-04)
- 九州大学マス・フォア・インダストリ研究所准教授公募 (2015-07-06)

もっと記事を見る

公募のお知らせ (研究費、顕彰ほか)

- 京都大学化学研究所スーパーコンピュータシステム利用案内 (2016-03-08)
- 2016年度「統合データベース講習会:AJACS」受入れ機関募集 (2016-01-13)
- 平成28年度九州大学マス・フォア・インダストリ研究所 共同利用研究計画公募 (2015-12-03)
- X線自由電子レーザー施設SACLA 2016A期(平成28年3月~平成28年7月予定) 利用研究課題公募 (2015-10-06)
- [JST/JICA] 地球規模課題対応国際科学技術協力プログラム (SATREPS) 平成28年度 研究提案募集開始 (2015-09-14)

もっと記事を見る

最新イベント情報

- 2016年年会・第5回生命医薬情報学連合大会 (2016.9.29~10.1)
- 第371回CBI学会研究講演会「翻訳後修飾研究の「コンテニア」とそれを支える基盤技術-創薬R&Dへの期待-」 (2016.4.22)
- 第19回 創薬インフォマティクス研究会 (2016.4.15)
- 量子化学計算/分子動力学学連成計算ソフト「Platypus-QM/MM」講習会 (2016.3.16)
- 第20回バイオメディカル研究会「Precision Medicineへの展開」 (2016.3.25)
- 日本バイオインフォマティクス学会九州地域部会イベント (2016.3.21)
- 統合化推進プログラム(統合データ解析トライアル) 平成27年度研究成果報告会 (2016.3.10)
- 第3回創薬ワークショップ (2016.3.3, 4)
- 情報処理学会ITフォーラム:ビッグデータ活用実務フォーラム:機械学習とデータマイニングの最前線~500人大集会:そのツールと応用~ (2016.2.2)
- 統合データベース講習会:AJACS薩摩 (2015.1.26, 27)

もっと記事を見る

学会へのご意見・ご要望

ニュースレター >> バックナンバー

賛助会員一覧 >>

# 相談窓口

ゲノム情報解析基礎 バイオインフォマティクス基礎知識 (2016年)

バイオインフォマティクス系学会など

- ・ [JSB\(日本バイオインフォマティクス学会\)](#)
  - 「最新イベント情報」欄が有用。学会に入るとメール
  - 2016年09/29-10/01に [第5回生命医薬情報学連合](#)
- ・ [定量生物学の会](#)
- ・ [生命情報科学若手の会](#)
- ・ [NGS現場の会](#)
- ・ [オープンバイオ研究会](#)
- ・ [CBI\(情報計算化学生物学会\)](#)
- ・ [情報処理学会 バイオ情報学研究会 \(SIG BIO\)](#)
- ・ [人工知能学会第二種研究会 分子生物情報研究会 \(SIG\)](#)

バイオインフォマティクス系<sup>①</sup>よろず相談所

- ・ [ライフサイエンスQA](#)
  - 過疎ってる印象。。
- ・ [Bio Technical フォーラム](#)
  - 主に実験系だがときどきバイオインフォ系トピックも

バイオインフォマティクス人材育成カリキュラム(次世

- ・ [バイオサイエンスデータベースセンター\(NBDC\)運営委員](#)で、2014年3月に策定された [NGS用カリキュラム](#) が存在。ベースにして、平成26年度よりNBDCと東大アグリバイオ(ノートPCを用いた実習型) 講義を実施。おそらくアグリバイオ人数規模。
- ・ [NGS速習コース講習会\(平成26年度\)](#)
  - [NGS用カリキュラム](#) に沿った内容を東大農で2014年
  - 座学(講義のみ)を含む。講師数10人。
  - 講習会映像は [統合TV](#) と [Youtube](#) から公開
  - [報告書](#)
- ・ [NGS/ハンズオン講習会\(平成27年度\)](#)
  - 東大農で2015年07月22日~08月06日(A日程)お
  - 座学部分は全て自習にしてハンズオンに特化。講
  - 連携を強化(10 → 4人)
  - 講習会映像は [統合TV](#) と [Youtube](#) から公開

ここはじめてですか? FAQをチェックしましょう。 ×

[ログイン](#) [概要](#) [よくある質問](#)

QA

質問
タグ
ユーザー
バッジ
未回答
質問する

● 質問 ○ タグ ○ ユーザー

≡ **すべての質問** 有効 最新 ホット 注目の質問

1 投票	0 回答	437 閲覧	<b>1000人ゲノムデータのvcfファイルからのDAFの計算方法について</b> <span style="font-size: 0.8em;">1000genome population_genetics daf vcf af</span>	Jan 09 at 15:47 <b>suimye 296</b>
0 投票	3 回答	924 閲覧	<b>RNA-seqのAT bias</b> <span style="font-size: 0.8em;">stranded rna-seq miseq</span>	Jan 05 at 17:58 <b>suimye 296</b>
0 投票	1 回答	397 閲覧	<b>IGVの結果とFPKM値の違いについて</b> <span style="font-size: 0.8em;">rna-seq</span>	Jan 05 at 15:57 <b>suimye 296</b>
0 投票	1 回答	625 閲覧	<b>GEOからデータのダウンロード: MAX formatとは?</b> <span style="font-size: 0.8em;">geo</span>	Oct 22 '15 at 19:35 <b>kh 1</b>
0 投票	0 回答	429 閲覧	<b>癌のゲノム解析データベースに対する配列検索</b> <span style="font-size: 0.8em;">癌ゲノム</span>	Oct 20 '15 at 17:40 <b>Hiroya 1</b>
0 投票	0 回答	515 閲覧	<b>MacOS X 10.9.5に対するaugustus3.1のインストール時のエラーについて</b> <span style="font-size: 0.8em;">遺伝子予測 augustus</span>	Jul 19 '15 at 00:25 <b>fatman_2 1</b>
0 投票	0 回答	1.1k 閲覧	<b>RNA seqにおけるRを用いたマッピング時のエラーについて</b> <span style="font-size: 0.8em;">r rna-seq mapping</span>	Apr 06 '15 at 17:58 <b>brown 1</b>
0 投票	1 回答	2.2k 閲覧	<b>RNA-seq で得られるFPKMのグラフの作り方</b> <span style="font-size: 0.8em;">fpkm</span>	Mar 29 '15 at 18:27 <b>nob_fj ◆ 507</b>
0 投票	2 回答	1.5k 閲覧	<b>NCBIのMicrobial Nucleotide BLASTをLocal環境に構築したいのですが、</b> <span style="font-size: 0.8em;">blast</span>	Feb 24 '15 at 16:01 <b>TY 1</b>
0 投票	0 回答	1.0k 閲覧	<b>MISOアノテーションについて</b> <span style="font-size: 0.8em;">miso as splicing-junction</span>	Feb 17 '15 at 17:41 <b>Kent_allow 1</b>

welcome to #LSQA

ただいまベータテスト中です。そのため通知無く停止更新されることがあります。

質問するならライフサイエンスQA

質問する→回答がある→質問と回答が蓄積する→良い質問と回答が増える→よりよいライフサイエンス研究をする時間が増える

[概要](#)  
[よくある質問](#)

180 質問  
303 回答  
最近更新された質問

ただいまベータテスト運用中です。そのために通知無くコンテンツの変更やサービスの停止変更されることがあります。

質問するならライフサイエンスQA。このページのRSSフィード

@lsqa\_all · @lsqa\_ngs

世界に広がるQAサイト

バイオインフォマティクスの

# 相談窓口

気軽に質問していることがわかります

0	1	773	初学者的質問: paired end readのクオリティチェック	paired end quality check	Sep 18 '14 at 18:27 deer 1
0	2	824	ヒトの核酸が混入している場合のカバレッジは？	virus metagenome	Sep 14 '14 at 17:48 deer 1
2	2	1.4k	Togo picture gallery リクエスト	togopic dbcls 募集中	Sep 01 '14 at 12:00 suimye 286
1	3	1.7k	アライメントについて	ngs	Aug 11 '14 at 23:14 mn3 ◆◆ 5
1	1	1.5k	データベースにあるデータの再解析	データベース	Jul 14 '14 at 19:33 genetics1127
2	1	944	統合TVのキーワード検索について	キーワード	May 06 '14 at 19:23 かどた 1
0	1	2.2k	Rでfastaファイルを読み込む際におすすめのパッケージはありますか？	fasta r	May 04 '14 at 21:38 かどた 1
0	1	1.8k	RNA-seq データの定量と可視化の、ツールの組み合わせ	bowtie rpkm express cufflinks rna-star	Apr 16 '14 at 05:06 nob_fj ◆ 5
0	3	3.3k	大量の配列のアライメントについて	mapping alignment ngs	Apr 11 '14 at 18:27 izumi
0	3	918	Cufflinks (binary)が実行できない	cufflinks	Apr 08 '14 at 13:28 meguro
0	2	2.2k	SAMファイルのFASTAファイルへの変換について	fasta fastq sam ngs	Apr 08 '14 at 11:15 izumi



## 統合TVのキーワード検索について

旧バージョンの統合TVはNGSというカテゴリ中に20エントリー程度リストアップされていますが、統合TV curatedのほうでは、検索窓での"NGS"の結果は4エントリー、"次世代シーケンサ"でも4エントリー、"次世代シーケンサ"だと21エントリー、"次世代シーケンサー"だと4エントリー、と非常に不安定です。MeSH termっていうのかなんとかかわかりませんが、講義で使う側としてはユーザの意図を汲んでうまくやってくれとありがたいです。具体的に多くリストアップしてほしいです。まあ旧バージョンのほうを使えばいいのでしょうか。。

キーワード

質問日 May 04 '14 at 21:26

かどた  
186 ● 1 ● 2 ● 9

### One Answer:

回答順 最新 支持されている順

貴重なご指摘をいただき、どうもありがとうございます。表記ゆれ問題が起きており、ご迷惑おかけいたしました。講演タイトルで使用されているものを除き、用語を次世代シーケンサ(NGS)に統一しました。現在、旧統合TV Curated サイトのリニューアルを計画しており、エントリーのカテゴリ分け等を工夫して 検索性の向上に努めたいと考えています。今後ともどうぞよろしくお願いいたします。

回答日 May 06 '14 at 15:45

hono ◆◆  
196 ● 1 ● 4

GW中にもかかわらず早速の対応どうもありがとうございます。過去のタイトルとの兼ね合いもあって悩ましいというのは非常によくわかります。どこかのタイミングで過去との決別も必要かもしれませんが、それはそれでインターフェースが変わるとかえってわかりづらい、これまで見えていたものが見られなくなったなどのクレームがきたりとどっちを選択しても一定の割合でアンチが出ますからね。現状でもまだ多少エントリー数に差が出るようですが、追々修正していただければ幸いです。

かどた (May 06 '14 at 19:23)

# NGS用カリキュラム

①2014年3月に策定されたNGS用カリキュラム。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「速習以外」に分かれている

## 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)

- ・バイオサイエンスデータベースセンター(NBDC)運営委員会人材育成分科会で、2014年3月に策定されたNGS用カリキュラム<sup>①</sup>をベースにして、平成26年度よりNBDCと東大アグリバイオ(ノートPCを用いた実習型)講義を実施。おそらくアグリバイオ人数規模。

- ・NGS速習コース講習会(平成26年度)
  - ・NGS用カリキュラムに沿った内容を東大農で2014年
  - ・座学(講義のみ)を含む。講師数10人。
  - ・講習会映像は統合TVとYoutubeから公開
  - ・報告書
- ・NGSハンズオン講習会(平成27年度)
  - ・東大農で2015年07月22日～08月06日(A日程)および(B日程)に実施
  - ・座学部分は全て自習にしてハンズオンに特化。講師連携を強化(10 → 4人)
  - ・講習会映像は統合TVとYoutubeから公開
  - ・報告書
- ・NGSハンズオン講習会(平成28年度)
  - ・東大農で2016年07月19日～08月04日に実施予定
  - ・全てハンズオン。講師数は(ほぼ)限界レベル(4 → 2)
  - ・中～上級レベルの内容を新規提供

## 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】

大項目	日数	No.	項目	習得技術	レベル	形式
1. コンピュータリテラシーとサーバー設計	4日	1-1	OS、ハード構成	・コンピュータの基本的理解	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的理解	初級	講義
		1-3	UNIX I	・UNIXの基礎的理解 ・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
4. 次世代シーケンサ	2.5日	4-1	次世代シーケンサ基礎I	・原理の理解	初級	講義
		4-2	次世代シーケンサ基礎II	・応用分野とそれのための計測技術の理解(RNA-seq, ChIP-seq, がんゲノム, 個人ゲノム, 環境ゲノム, Hi-C)	初級	講義
	0.5日	4-3	次世代シーケンサ実習I	・ファイル形式、可視化、quality check、マッピング、アセンブル	初級	実習
	1.5日	4-4	次世代シーケンサ実習II	・代表的なパイプラインについての実習(多型解析(IGV)、RNA-seq、ChIP-seq、及び統合解析、定番のツールを利用)	初級	実習
5. ゲノム関連の倫理・法律	0.5日	5-1	ゲノム情報倫理概論	・ゲノム情報を扱う上で、プライバシー保護などの必要な倫理的問題、法的問題の国内外の状況を理解し、ゲノム情報を適切に利用できるようにする ・匿名化、暗号化、情報セキュリティ概要	初級	講義
6. 分子生命科学	0.5日	6-1	分子生命科学概論	・複製、転写、翻訳、代謝、シグナル伝達などの基礎知識	初級	講義
		6-2	オミクス概論	・ゲノム以外のオミクスデータの基礎知識	初級	講義
		6-3	遺伝/進化概論	・ゲノムデータを扱う上での遺伝学、進化学の基礎知識	初級	講義



①「速習」コースのほうを2014年9月に  
試行実施。平均約80名が10日間受講

# NGS速習コース講習会

## 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)

・バイオサイエンスデータベースセンター(NBDC)運営委員会人材育成分科会で、2014年3月に策定されたNGS用カリキュラムが存在。このカリキュラムをベースにして、平成26年度よりNBDCと東大アグリバイオ主催でハンズオン(ノートPCを用いた実習型)講義を実施。おそらくアグリバイオ本体に次ぐ受講人数規模。

- ・ **NGS速習コース講習会(平成26年度)**
  - ・ [NGS用カリキュラム](#)に沿った内容を東大農で2014年9月1日～12日に実施
  - ・ 座学(講義のみ)を含む。講師数10人。
  - ・ 講習会映像は[統合TV](#)と[Youtube](#)から公開
  - ・ [報告書](#)

- ・ **NGS/ハンズオン講習会(平成27年度)**
  - ・ 東大農で2015年07月22日～08月06日(A日程)および08月26-28日(B日程)に実施
  - ・ 座学部分は全て自習にしてハンズオンに特化。講師数を減らして全体の連携を強化(10 → 4人)
  - ・ 講習会映像は[統合TV](#)と[Youtube](#)から公開
  - ・ [報告書](#)

- ・ **NGS/ハンズオン講習会(平成28年度)**
  - ・ 東大農で2016年07月19日～08月04日に実施予定(申込み期限は5/31)
  - ・ 全てハンズオン。講師数は(まほ限界レベル(4 → 2人))
  - ・ 中～上級レベルの内容を新規提供

## 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)

・ シ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とを想定した「速習以外」に分かれています。

No.	項目	習得技術	レベル	形式
0-1	OS, ハード構成	・コンピュータの基本的理解	初級	講義
0-2	ネットワーク基礎	・インターネット、セキュリティの基本的理解	初級	講義
1	UNIX I	・UNIXの基礎的理解 ・Linux導入	中級	実習
1-4	スクリプト言語	・Perl ・シェルスクリプト	中級	実習
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
3-1	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
4-1	次世代シーケンサ基礎I	・原理の理解	初級	講義
4-2	次世代シーケンサ基礎II	・応用分野とそれのための計測技術の理解(RNA-seq, ChIP-seq, ゲノム、個人ゲノム、環境ゲノム, Hi-C)	初級	講義
0.25日				
0.5日				
1.5日				
4-3	次世代シーケンサ実習I	・ファイル形式、可視化、quality check、マッピング、アセンブル	初級	実習
4-4	次世代シーケンサ実習II	・代表的なパイプラインについての実習(多型解析(IGV)、RNA-seq、ChIP-seq、及び統合解析、定番のツールを利用)	初級	実習
5.	ゲノム関連の倫理・法律			
0.5日				
5-1	ゲノム情報倫理概論	・ゲノム情報を扱う上で、プライバシー保護などの必要な倫理的問題、法的問題の国内外の状況を理解し、ゲノム情報を適切に利用できるようにする ・匿名化、暗号化、情報セキュリティ概要	初級	講義
6.	分子生命科学			
0.5日				
6-1	分子生命科学概論	・複製、転写、翻訳、代謝、シグナル伝達などの基礎知識	初級	講義
6-2	オミクス概論	・ゲノム以外のオミクスデータの基礎知識	初級	講義
6-3	遺伝/進化概論	・ゲノムデータを扱う上での遺伝学、進化学の基礎知識	初級	講義

①ハンズオンに特化した講習会を2015年7-8月に実施。平均約60名が14日間受講

# NGSハンズオン講習会

## 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)

- ・バイオサイエンスデータベースセンター(NBDC)運営委員会人材育成分科会で、2014年3月に策定されたNGS用カリキュラムが存在。このカリキュラムをベースにして、平成26年度よりNBDCと東大アグリバイオ主催でハンズオン(ノートPCを用いた実習型)講義を実施。おそらくアグリバイオ本体に次ぐ受講人数規模。
- ・NGS速習コース講習会(平成26年度)
  - NGS用カリキュラムに沿った内容を東大農で2014年9月1日~12日に実施
  - 座学(講義のみ)を含む。講師数10人。
  - 講習会映像は統合TVとYoutubeから公開
  - 報告書
- ・NGSハンズオン講習会(平成27年度)
  - 東大農で2015年07月22日~08月06日(A日程)および08月26-28日(B日程)に実施
  - 座学部分は全て自習にしてハンズオンに特化。講師数を減らして全体の連携を強化(10→4人)
  - 講習会映像は統合TVとYoutubeから公開
  - 報告書
- ・NGSハンズオン講習会(平成28年度)
  - 東大農で2016年07月19日~08月04日に実施予定(申込み期限は5/31)
  - 全てハンズオン。講師数はほぼ限界レベル(4→2人)
  - 中~上級レベルの内容を新規提供



# NGSハンズオン講習会

①NGSハンズオン講習会の公式サイトはNBDCから提供。②ほぼ完全な講義資料や映像をフリーで視聴可能

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)

・バイオサイエンスデータベースセンターで、2014年3月に策定されたNGSデータベースにして、平成26年度よりNBDC(ノートPCを用いた実習型)講義を人数規模。

・NGS速習コース講習会(平成26年度)  
 ◦ NGS用カリキュラムに沿った座学(講義のみ)を含む。講習会映像は統合TVとYouTube

①

・NGSハンズオン講習会(平成27年度)  
 ◦ 東大農で2015年07月22日～23日(2日程)に実施  
 ◦ 座学部分は全て自習にして連携を強化(10 → 4人)  
 ◦ 講習会映像は統合TVとYouTube  
 ◦ 報告書

・NGSハンズオン講習会(平成28年度)  
 ◦ 東大農で2016年07月19日～20日に実施  
 ◦ 全てハンズオン。講師数は10名  
 ◦ 中～上級レベルの内容を新



～ 散在するデータベースを、まとめて、使い易く～

## バイオサイエンスデータベースセンター

National Bioscience Database Center



国立研究開発法人 科学技術振興機構

文字サイズ変更 大 中 小

English
サイトマップ
サイト内検索 
検索

ホーム
NBDCについて
研究開発
公券情報
採用情報
広報
人材支援
お問い合わせ
リンク

Home > 人材支援 > 支援 > 講習会 > 平成27年度NGSハンズオン講習会

### 平成27年度NGSハンズオン講習会

平成27年度は、平成26年度の実績を踏まえ、講義内容の改善等を行い、ハンズオンに特化した、より効果的なNGS講習会を開催しました。

- [H27年度概要](#)
- [H27年度講義日程・参考資料](#)
- [H26年度講習会の情報についてはこちらをご覧ください。](#)
- [H27年度実施報告書・講義資料・動画等](#)

- [講習会実施報告書 \(PDF: 2.17MB\)](#) および [受講者アンケート集計結果 \(データ集\) \(PDF: 662KB\)](#)
- [講義資料・動画](#) \*講義資料一覧のファイル名をクリックすると資料ファイル (PDF等) がダウンロードできます。

実施日	実施時間	大項目	項目	レベル	習得技術	担当講師(敬称略)	講義資料・動画(統合TV)
7月22日 (水)	10:30-12:00	PC環境の構築	Bio-Linux8とRのインストール状況確認		・Linux導入 ・R導入 ・NGS解析に必要な環境構築技術	門田 幸二 (東京大学)	<a href="#">事前予習資料一覧 (PDF:52KB)</a> <a href="#">講義資料一覧 (PDF:108KB)</a>
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
7月23日 (木)	10:30-12:00	UNIX/Linuxとスクリプト言語	Linux基礎	初級	UNIXの基礎の理解	門田 幸二 (東京大学)	<a href="#">講義資料一覧 (PDF:32KB)</a> <a href="#">統合TV</a>
	13:15-14:45			中級			
	15:00-16:30						

②

Apr 11 2016

11

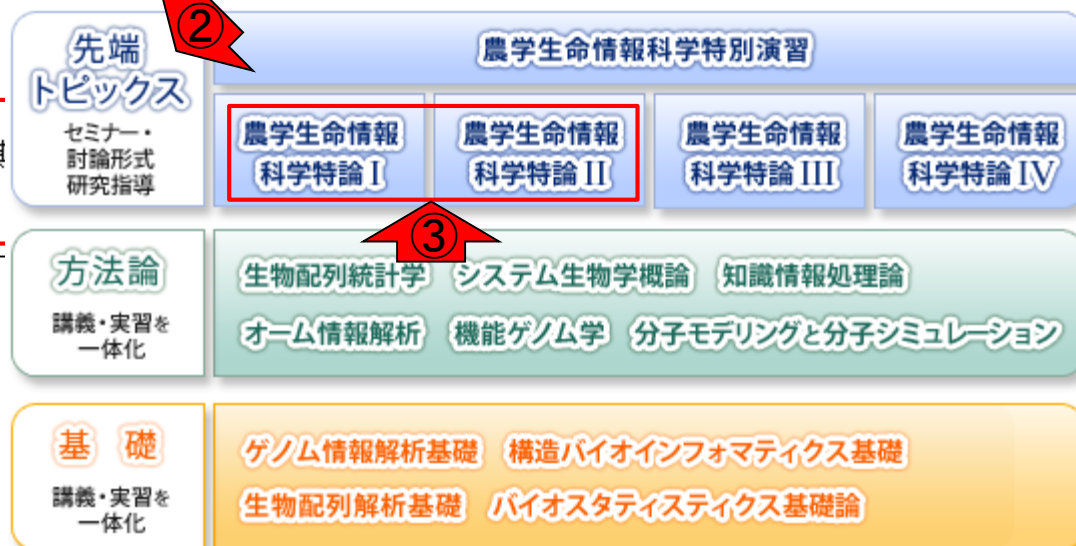
# NGSハンズオン講習会

①平成28年度も実施。門田担当分は②先端トピックスの③2科目(農学生命情報科学特論IとII)と兼ねている。①のリンク先が次のスライド

## 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)


- ・バイオサイエンスデータベースセンター(NBDC)運営委員会人材育成分科会で、2014年3月に策定されたNGS用カリキュラムが存在。このカリキュラムをベースにして、平成26年度よりNBDCと東大アグリバイオ主催でハンズオン(ノートPCを用いた実習型)講義を実施。おそらくアグリバイオ本体に次ぐ受講人数規模。
- ・NGS速習コース講習会(平成26年度)
  - NGS用カリキュラムに沿った内容を東大農で2014年9月1日~12日に実施
  - 座学(講義のみ)を含む。講師数10人。
  - 講習会映像は統合TVとYoutubeから公開
  - 報告書
- ・NGSハンズオン講習会(平成27年度)
  - 東大農で2015年07月22日~08月06日(A日程)および08月26-28日(B日程)に実施
  - 座学部分は全て自習にしてハンズオンに特化。講師数を減らして全体の連携を強化(10→4人)
  - 講習会映像は統合TVとYoutubeから公開
  - 報告書
- ① NGSハンズオン講習会(平成28年度)
  - 東大農で2016年07月19日~08月04日に実施予定(申込み期)
  - 全てハンズオン。講師数は(ほぼ)限界レベル(4→2人)
  - 中~上級レベルの内容を新規提供

## アグリバイオの教育プログラム



# NGSハンズオン講習会

①カリキュラムのページで、より詳細な予定講義内容を知ることができる

 **NBDC** National Bioscience Database Center

— 散在するデータベースを、まとめて、使い易く —

**バイオサイエンスデータベースセンター**

English [サイトマップ](#) [サイト内検索](#)

国立研究開発法人 **科学技術振興機構**  
JST LEAD SCIENCE AND KNOWLEDGE AGENCY

文字サイズ変更

[ホーム](#) [NBDCについて](#) [研究開発](#) [公衆情報](#) [採用情報](#) [広報](#) [人材支援](#) [お問い合わせ](#) [リンク](#)

[Home](#) ▶ [人材支援](#) ▶ [支援](#) ▶ [講習会](#) ▶ [平成28年度NGSハンズオン講習会](#)

## ■平成28年度NGSハンズオン講習会

### 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 (NGS) ハンズオン講習会 (2016年7月19日～8月4日) を開催します。

NGS (Next Generation Sequencer) は、基礎および応用の様々な分野で導入され、ライフサイエンス研究に大きな変貌をもたらしています。NGSから産生される塩基配列データには、多様な生物を特徴付ける重要な情報が膨大かつ複雑に存在しており、このような配列データを効率的に解析するバイオインフォマティクスの重要性が増しています。しかし現状では、量産されるデータ量に対し、多様な実験データの解析と解釈が出来るバイオインフォマティクス人材の数は圧倒的に少なく、基礎研究をはじめ、医療、創薬などの分野からもバイオインフォマティクス人材の育成が強く求められています。このような現状を鑑み、NBDCでは、より多くの生命科学系の研究者及び研究支援者がバイオインフォマティクススキルを身に付ける機会となるよう、引き続き、平成28年度もNGSハンズオン講習会を実施します。なお、平成27年度は、平成26年度の実績を踏まえ、ハンズオンに特化した[NGSハンズオン講習会](#)を実施しました。

#### ■平成28年度開催概要

統計解析や高度なNGS解析の内容を多く取り入れたNGSハンズオン講習会を2016年7月19日(火)～8月4日(木) (土日除く12日間)に開催致します。

平成28年度は、需要の多い「初～中級レベルのNGS解析(第2部:7/25-28)」を中核とした3部構成とします。第2部は、昨年度の講義内容を基本に「適度の増量」を予定しています。「統計解析(第1部:7/20-22)」の内容は、大幅な増加となる予定です。また、「中～上級レベルのNGS解析(第3部:8/1-4)」を新規に実施します。

これまでの本講習会の枠組みでは実施できなかった内容を多く含む予定です。第1部および3部は既に一定のスキルをもつ受講生を対象としておりますが、新規に受講される方でも、平成27年度の講習会資料を中心に指定された内容を実施すれば、問題なく理解していただけると思われま

詳しい内容や参考図書などの情報はリンクをご覧ください。[カリキュラム](#)、[参考図書](#)

\* 昨年度との違い、対応関係、想定受講者、および予習事項については [こちら \(PDF: 2.01MB\)](#) をご覧ください。

①カリキュラムのページ。②講義資料・動画は順次公開予定

# NGSハンズオン講習会

NBDC

- 散在するデータベースを、まとめて、使い易く -

バイオサイエンスデータベースセンター

国立研究開発法人 科学技術振興機構  
文字サイズ変更 大 中 小

National Bioscience Database Center

English サイトマップ サイト内検索 検索... 検索

ホーム NBDCについて 研究開発 公募情報 採用情報 広報 人材支援 お問い合わせ リンク

Home > 人材支援 > 支援 > 講習会 > 平成28年度NGSハンズオン講習会カリキュラム

H28年度 NGSハンズオン講習会カリキュラム

H28年度日程・講義資料・動画等

\*講義資料・動画については、順次公開いたします。

カリキュラム (1 KB)

実施日	実施時間	大項目	タイトル	内容 (予定)	担当講師 (敬称略)	講義資料・ 動画(統合TV)
7月19日 (火)	10:30- 18:15	PC環境の構築	Bio-Linux8とRのインストール状況確認	<ul style="list-style-type: none"> <li>・ Bio-Linux8 (第2部および3部で利用するovaファイル) の導入確認</li> <li>・ 共有フォルダ設定完了確認</li> <li>・ 基本的なLinuxコマンドの習得状況確認</li> <li>・ R本体およびパッケージのインストール確認</li> <li>・ 講師指定の事前予習内容の再確認</li> <li>・ 講習会期間中に貸与されるノートPCを用いた各種動作確認</li> </ul>	主催・共催機関	
7月20日 (水)	10:30- 18:15	第1部 統計解析 (農学生命情報科学特論I)	ゲノム解析、塩基配列解析	<ul style="list-style-type: none"> <li>・ NGS解析手段、ウェブツール(DDBJ Pipeline)との連携</li> <li>・ k-mer解析 (k個の連続塩基に基づく各種解析) の基礎と応用</li> <li>・ 塩基ごとの出現頻度解析(k=1)、2連続塩基の出現頻度解析(k=2)</li> <li>・ 塩基配列解析を行うための基本スキルの復習や作図</li> <li>・ de novoアセンブリ時のエラー補正やゲノムサイズ推定の基本的な考え方</li> </ul>	門田 幸二 (東京大学)	

# NGSハンズオン講習会

①昨年度との違い、想定受講者、予習事項などはこちら

NBDC National Bioscience Database Center － 散在するデータベースを、まとめて、使い易く－ **バイオサイエンスデータベースセンター** English サイトマップ サイト内検索   文字サイズ変更    国立研究開発法人 科学技術振興機構 LEIET SUKUNAWAKI AKAHIDE MACHIO

ホーム NBDCについて 研究開発 公券情報 採用情報 広報 **人材支援** お問い合わせ リンク

Home ▶ 人材支援 ▶ 支援 ▶ 講習会 ▶ 平成28年度NGSハンズオン講習会

## ■平成28年度NGSハンズオン講習会

### 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 (NGS) ハンズオン講習会 (2016年7月19日～8月4日) を開催します。

NGS (Next Generation Sequencer) は、基礎および応用の様々な分野で導入され、ライフサイエンス研究に大きな変貌をもたらしています。NGSから産生される塩基配列データには、多様な生物を特徴付ける重要な情報が膨大かつ複雑に存在しており、このような配列データを効率的に解析するバイオインフォマティクスの重要性が増しています。しかし現状では、量産されるデータ量に対し、多様な実験データの解析と解釈が出来るバイオインフォマティクス人材の数は圧倒的に少なく、基礎研究をはじめ、医療、創薬などの分野からもバイオインフォマティクス人材の育成が強く求められています。このような現状を鑑み、NBDCでは、より多くの生命科学系の研究者及び研究支援者がバイオインフォマティクススキルを身に付ける機会となるよう、引き続き、平成28年度もNGSハンズオン講習会を実施します。なお、平成27年度は、平成26年度の実績を踏まえ、ハンズオンに特化した[NGSハンズオン講習会](#)を実施しました。

#### ■平成28年度開催概要

統計解析や高度なNGS解析の内容を多く取り入れたNGSハンズオン講習会を2016年7月19日(火)～8月4日(木) (土日除く12日間)に開催致します。

平成28年度は、需要の多い「初～中級レベルのNGS解析(第2部:7/25-28)」を中核とした3部構成とします。第2部は、昨年度の講義内容を基本に「適度の増量」を予定しています。「統計解析(第1部:7/20-22)」の内容は、大幅な増加となる予定です。また、「中～上級レベルのNGS解析(第3部:8/1-4)」を新規に実施します。

これまでの本講習会の枠組みでは実施できなかった内容を多く含める予定ですので、第1部および3部は既に一定のスキルをもつ受講生を対象としておりますが、新規に受講される方でも、平成27年度の講習会資料を中心に指定された予習を実施すれば、問題なく理解していただければと思います。

詳しい内容や参考図書などの情報はリンクをご覧ください。[カリキュラム](#)、[参考図書](#)

\*昨年度との違い、対応関係、想定受講者、および予習事項については [こちら\(PDF: 2.01MB\)](#) **①** ご覧ください。



# 平成28年度概要

- ①統計解析を大幅に増加(0.5日 → 3日)
- ②よりハイレベルな内容を新規提供

## ～ 平成27年度 ～

7/22(水): PC環境の構築

7/23(木): Linux基礎

7/24(金): シェルスクリプト

7/27(月): Perl

7/28(火): Python

7/29(水): データ解析環境R

7/30(木): データ解析環境R

8/3(月): NGS解析(基礎)

8/4(火): NGS解析(ゲノムReseq、変異解析)

8/5(水): NGS解析(RNA-seq: 代表的なパイプライン)

8/5(水): NGS解析(RNA-seq: 統計解析)

8/6(木): NGS解析(ChIP-seq)

8/26(水): 予備日

8/27(木): 予備日

8/28(金): 予備日

## ～ 平成28年度 ～

7/19(火): PC環境の構築

7/20(水): 統計解析(塩基配列解析系)

7/21(木): 統計解析(発現解析系)

7/22(金): 統計解析(発現解析系)

7/25(月): NGS解析基礎

7/26(火): ゲノムReseq、変異解析

7/27(水): RNA-seq

7/28(木): ChIP-seq

8/1(月): Linux環境でのデータ解析1

8/2(火): Linux環境でのデータ解析2

8/3(水): ウェブツール、ロングリード

8/4(木): トランスクリプトーム解析系

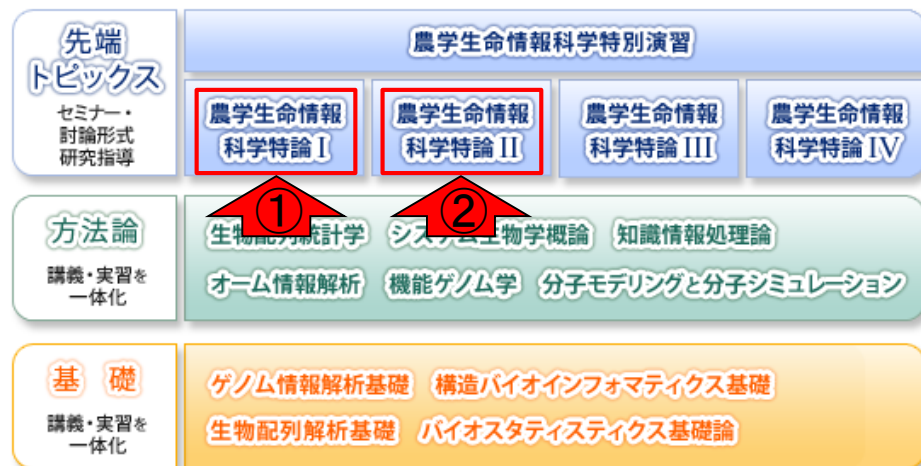




# 3部構成

①第1部は「農学生命情報科学特論I」、  
②第2部は「農学生命情報科学特論II」  
として開催

## アグリバイオの教育プログラム



～ 平成28年度 ～

7/19(火): PC環境の構築

第1部  
統計解析

7/20(水): 統計解析(塩基配列解析系)

7/21(木): 統計解析(発現解析系)

7/22(金): 統計解析(発現解析系)



第2部  
NGS解析  
(初～中級)

7/25(月): NGS解析基礎

7/26(火): ゲノムReseq、変異解析

7/27(水): RNA-seq

7/28(木): ChIP-seq

第3部  
NGS解析  
(中～上級)

8/1(月): Linux環境でのデータ解析1

8/2(火): Linux環境でのデータ解析2

8/3(水): ウェブツール、ロングリード

8/4(木): トランスクリプトーム解析系



# ゲノム情報解析=NGS解析

①「ゲノム情報解析基礎」で教えら  
れる内容は全体のごく一部。この科  
目では、フリーソフトRで塩基配列解  
析を行う基本スキルの伝授のみ

## アグリバイオの教育プログラム



～ 平成28年度 ～

7/19(火): PC環境の構築

### 第1部 統計解析

7/20(水): 統計解析(塩基配列解析系)

7/21(木): 統計解析(発現解析系)

7/22(金): 統計解析(発現解析系)

### 第2部 NGS解析 (初～中級)

7/25(月): NGS解析基礎

7/26(火): ゲノムReseq、変異解析

7/27(水): RNA-seq

7/28(木): ChIP-seq

### 第3部 NGS解析 (中～上級)

8/1(月): Linux環境でのデータ解析1

8/2(火): Linux環境でのデータ解析2

8/3(水): ウェブツール、ロングリード

8/4(木): トランスクリプトーム解析系

NGS解析(ゲノム情報解析)を自在に行う上で、Linuxを使いこなせるにこしたことはない。クラウド解析環境(ウェブツール)やRを組み合わせれば完璧

# 門田の主な活動

- 東大アグリバイオの大学院講義(バイオインフォ全般)
  - Rを中心としたハンズオン講義(平成16年度~)
    - 受講人数が多い(最大130名)ので、クラウド(ウェブツール)系実習は実質的に不可能
    - 講義補助員(TA)が数名のみなので、Linux系実習も困難
- NBDC/東大アグリバイオのNGSハンズオン講義(NGSに特化)
  - Linuxを中心としたハンズオン講義(平成26年度~)
    - 受講人数は多い(最大71名;おそらくアグリバイオ本体に次ぐ規模)が、**受講生の意識レベルが高く(きっちり予習をやるヒトが多数派)、環境構築済みノートPC数、TA数が充実しているため、本格的なLinux実習が成立しうる。**
- 日本乳酸菌学会誌のNGS連載
  - Linuxを中心とした自習用教材(平成26年度~)
    - バクテリア(乳酸菌)データを、主にBio-Linux上で解析するノウハウを提供。
    - 第6回(2016年3月予定)分以降は、DDBJ Pipeline(ウェブツール)の利用法も紹介
    - データ取得・インストール・実行に時間がかかるものも、自習できる。ハンズオン講義よりも心穏やか。



①受講者数と②スタッフ数の傾向が反比例。  
 アグリバイオ単体でクオリティの高い講義を  
 維持するのは困難な状況(個人の感想です)

# 自己紹介

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・CBRC
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月~ 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人材養成プログラム(科学技術振興調整費: 2004/10-2009/3)  
 アグリバイオインフォマティクス教育研究プログラム(特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	151
博士課程	3	7	11	13	6	8	12	21	16	19	24	25
社会人	5	3	8	4	1	0	11	19	32	26	55	34
合計	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>210</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20

1科目以上の合格者数



# NGSハンズオン講習会

NGSハンズオン講習会(主催:NBDCとアグリバイオ)の枠組みで実施することで、NBDCの協力を仰ぎながら、よりよいハンズオン講義を提供(するのが門田の戦略)

～平成28年度～

7/19(火):PC環境の構築

## 第1部 統計解析

7/20(水):統計解析(塩基配列解析系)

7/21(木):統計解析(発現解析系)

7/22(金):統計解析(発現解析系)



## 第2部 NGS解析 (初～中級)

7/25(月):NGS解析基礎

7/26(火):ゲノムReseq、変異解析

7/27(水):RNA-seq

7/28(木):ChIP-seq

## 第3部 NGS解析 (中～上級)

8/1(月):Linux環境でのデータ解析1

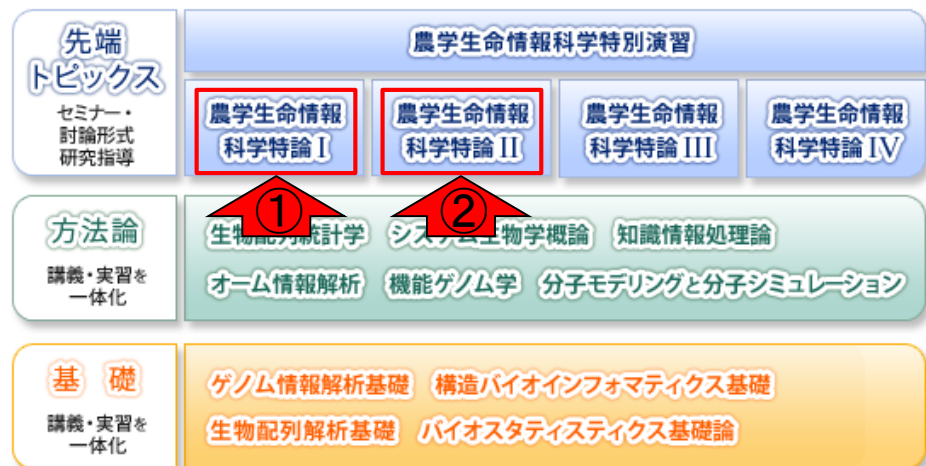
8/2(火):Linux環境でのデータ解析2

8/3(水):ウェブツール、ロングリード

8/4(木):トランスクリプトーム解析系



## アグリバイオの教育プログラム



# 講習会関連

①(おそらくこれ以外にも)各自の事情や感性に合った講習会があると思います。教え方はひとそれぞれなので色々出られてみてはいかがでしょうか

## 講習会関連

- ・ [イルミナ株式会社のiSchool](#)
  - NGS機器のシェアNo.1企業提供の講習会やウェビナー。クラウド解析環境[BaseSpace](#)の提供など意欲的。
- ・ [基礎生物学研究所のゲノムインフォマティクス・トレーニングコース](#)
  - おそらく受講生あたりのスタッフ数をもっとも充実しているコース。例年春と秋に開催しているので今年も開催されるであろう。
- ・ [NBDCの統合データベース講習会\(AJACS\)](#)
  - 基本はデータベース系の講習会だが、2016年2月に開催された[AJACAdvanced](#) (主催: [DBCLS](#))という枠組みではNGS解析中心。
- ・ [DDBJのDDBJing講習会](#)
  - DNA Data Bank of Japan (DDBJ)が提供する各種サービスの講習会。



## 自習用教材

- ・ [\(Rで\)塩基配列解析](#)
  - NGS系のそれっぽいキーワードで検索すると大抵上位に出現。RのインストールからLinux環境でのNGS解析まで幅広く解説。
- ・ [統合TV](#)
  - 有用なデータベースやウェブツールの活用法を動画で紹介してくれます。
- ・ [次世代シーケンスデータ解析情報共有フォーラム\(NGS Surfer's Wiki\)](#)
  - 実務に役立つ情報が豊富。
- ・ [Biopapyrus](#)
  - Linux、Perlなどの各種インストール系や用語説明などが豊富。

# 自習用教材

## 自習用教材



- ・ [\(Rで\)塩基配列解析](#)
  - NGS系のそれっぽいキーワードで検索すると大抵上位に出現。RのインストールからLinux環境でのNGS解析まで幅広く解説。
- ・ [統合TV](#)
  - 有用なデータベースやウェブツールの活用法を動画で紹介
- ・ [次世代シーケンスデータ解析情報共有フォーラム\(NGS Surfer's\)](#)
  - 実務に役立つ情報が豊富。
- ・ [Biopapyrus](#)
  - Linux, Perlなどの各種インストール系や用語説明などが豊

## (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 2016/04/08, since 2011)

### What's new?

- このウェブページは[インストール](#)についての推奨手順 ([Windows2015.04.04版](#)と[Macintosh2015.04.03版](#))に従ってフリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は[基本的な利用法](#)([Windows2015.04.03版](#)と [Macintosh2015.04.03版](#))で自習してください。本ウェブページを体系的にまとめた書籍もあります。(2015/04/03)
- [平成28年度ハンズオン講習会](#)が7/19-8/4の日程で開催されます。申込み期間は4/4-5/31です。追加申込みやキャンセル待ちの受付は絶対に行いませんので受講希望者はお気をつけください。門田担当部分は主にこれまでの受講生を対象としています(リクエストを多く反映させています)。「昨年度との違い、対応関係、想定受講者、および予習事項」については [こちらのPDF](#) (約2MB) をご覧になり、必要に応じてつまみ食いしてください。(2016/04/05) **NEW**
- [QuasR](#)でBowtieのマッピングを行う場合に、(内部的にはbowtie1が動いているため)リード長が1本でも1,024 bpを超えたものがあればコケマス(1024 bpはセーフで1025 bpはアウト)のでご注意ください(高橋 広夫 氏提供情報)。(2016/04/06) **NEW**
- [日本乳酸菌学会誌のNGS関連連載](#)の第6回ウェブ資料を更新しました。具体的には、W20-2のスライド256の解説内容を変更しました。(2016/03/29) **NEW**
- [RNA-Seq実験ハンドブック](#)(鈴木穰 編)が刊行されます。(2016/03/22) **NEW**
- 「書籍|トランスクリプトーム解析| [2.3.1 RNA-seqデータ\(FASTQファイル\)](#)」のgetFASTQinfoやgetFASTQfile関数実行部分でsra\_conという引数を追加しました。が、おそらくデータを取りに行っている[ENA](#)のURL ([http://www.ebi.ac.uk/ena/data/view/reports/sra/fastq\\_files/](http://www.ebi.ac.uk/ena/data/view/reports/sra/fastq_files/))がリンク切れになっているようであまり情報の取得ができなくなっているようですm(\_ \_)mもちろん私のせいではありませんw。ご指摘深謝m(\_ \_)m(2016/03/17) **NEW**
- [はじめに](#) (last modified 2015/03/31)
- [参考資料\(講義, 講習会, 本など\)](#) (last modified 2016/02/29)
- [過去のお知らせ](#) (last modified 2016/04/05) **NEW**
- [インストール](#)について (last modified 2015/11/12)
- [インストール](#) | R本体 | 最新版 | [Win用](#) (last modified 2015/03/22) 推奨
- [インストール](#) | R本体 | 最新版 | [Mac用](#) (last modified 2015/04/22) 推奨
- [インストール](#) | R本体 | 過去版 | [Win用](#) (last modified 2015/03/22)
- [インストール](#) | R本体 | 過去版 | [Mac用](#) (last modified 2015/03/22)

# (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 2016/04/08, since 2011)

### What's new?

- このウェブページは[インストール](#)についての推奨手順 ([Windows2015.04.04版](#)と[Macintosh2015.04.03版](#))に従ってフリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は[基本的な利用法](#)([Windows2015.04.03版](#)と[Macintosh2015.04.03版](#))で自習してください。本ウェブページを体系的にまとめた書籍もあります。(2015/04/03)
- 平成28年度ハンズオン②が7/19-8/4の日程②確定されます。申込み期間は4/4-5/31です。追加申込みやキャンセル待ちの受付は絶対に行いませんので受講希望者はお気をつけください。門田担当部分は主にこれまでの受講生を対象としています(リクエストを多く反映させています)。「昨年度との違い、対応関係、想定受講者、および予習事項」については[こちらのPDF](#) (約2MB)をご覧ください。必要に応じてつまみ食いしてください。(2016/04/05) **NEW**
- [QuasR](#)でBowtieのマッピングを行う場合に、(内部的にはbowtie1が動いているため)リード長が1本でも1,024 bpを超えたものがあればコケマス(1024 bpはセーフで1025 bpはアウト)のご注意ください(高橋 広夫 氏提供情報)。(2016/04/06) **NEW**
- [日本乳酸菌学会誌](#)の[NGS関連連載](#)の第6回ウェブ資料を更新しました。具体的には、W20-2のスライド256の解説内容を変更しました。(2016/03/29) **NEW**
- [RNA-Seq実験ハンドブック](#)(鈴木穰 編)が刊行されます。(2016/03/22) **NEW**
- 「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)」のgetFASTQinfoやgetFASTQfile関数実行部分でsra\_conという引数を追加しました。が、おそらくデータを取りに行っているENAのURL ([http://www.ebi.ac.uk/ena/data/view/reports/sra/fastq\\_files/](http://www.ebi.ac.uk/ena/data/view/reports/sra/fastq_files/))がリンク切れになっているようでうまく情報の取得ができなくなっているようですm(\_)\_mもちろん私のせいではありませんw。ご指摘深謝m(\_)\_m(2016/03/17) **NEW**

- [はじめに](#) (last modified 2015/03/31)
- [参考資料\(講義、講習会、本など\)](#) (last modified 2016/02/29)
- [過去のお知らせ](#) (last modified 2016/04/05) **NEW**
- [インストール](#)について (last modified 2015/11/12)
- インストール | R本体 | 最新版 | [Win用](#) (last modified 2015/03/22)推奨
- インストール | R本体 | 最新版 | [Mac用](#) (last modified 2015/04/22)推奨
- インストール | R本体 | 過去版 | [Win用](#) (last modified 2015/03/22)
- インストール | R本体 | 過去版 | [Mac用](#) (last modified 2015/03/22)

①貸与PCは、基本的にこのウェブページの推奨手順通りにR本体および必要なパッケージのインストールを行っています。この手順に沿ってインストールを行えば、来週以降は持込PCで講義を受けることができます。後半は、②「基本的な利用法」の一部を行います。



# Rの起動



起動直後は画面いっぱいになるので…。



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

```
R version 3.1.3 (2015-03-09) -- "Smooth Sidewalk"
Copyright (C) 2015 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

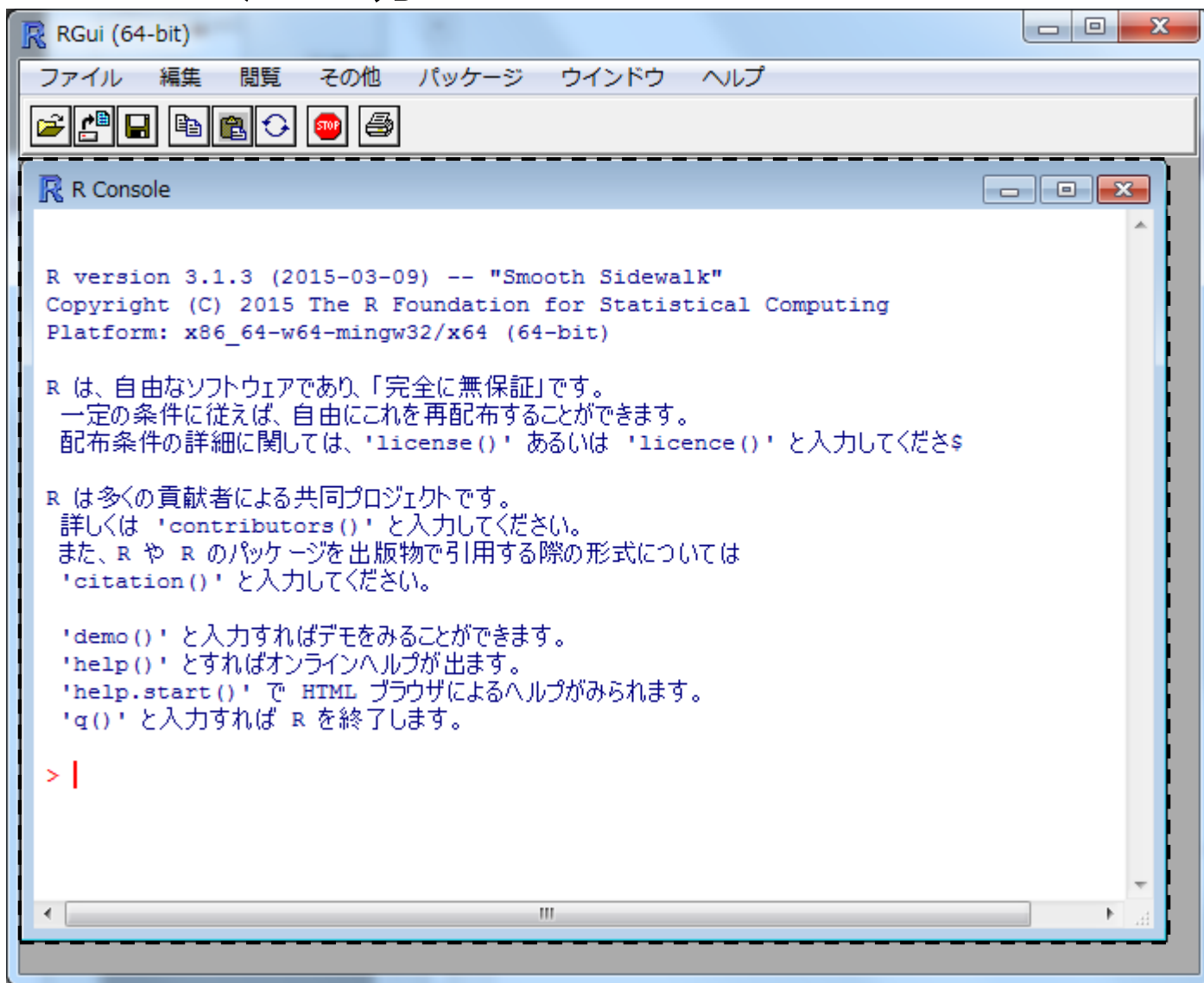
R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()' と入力してください。

'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。

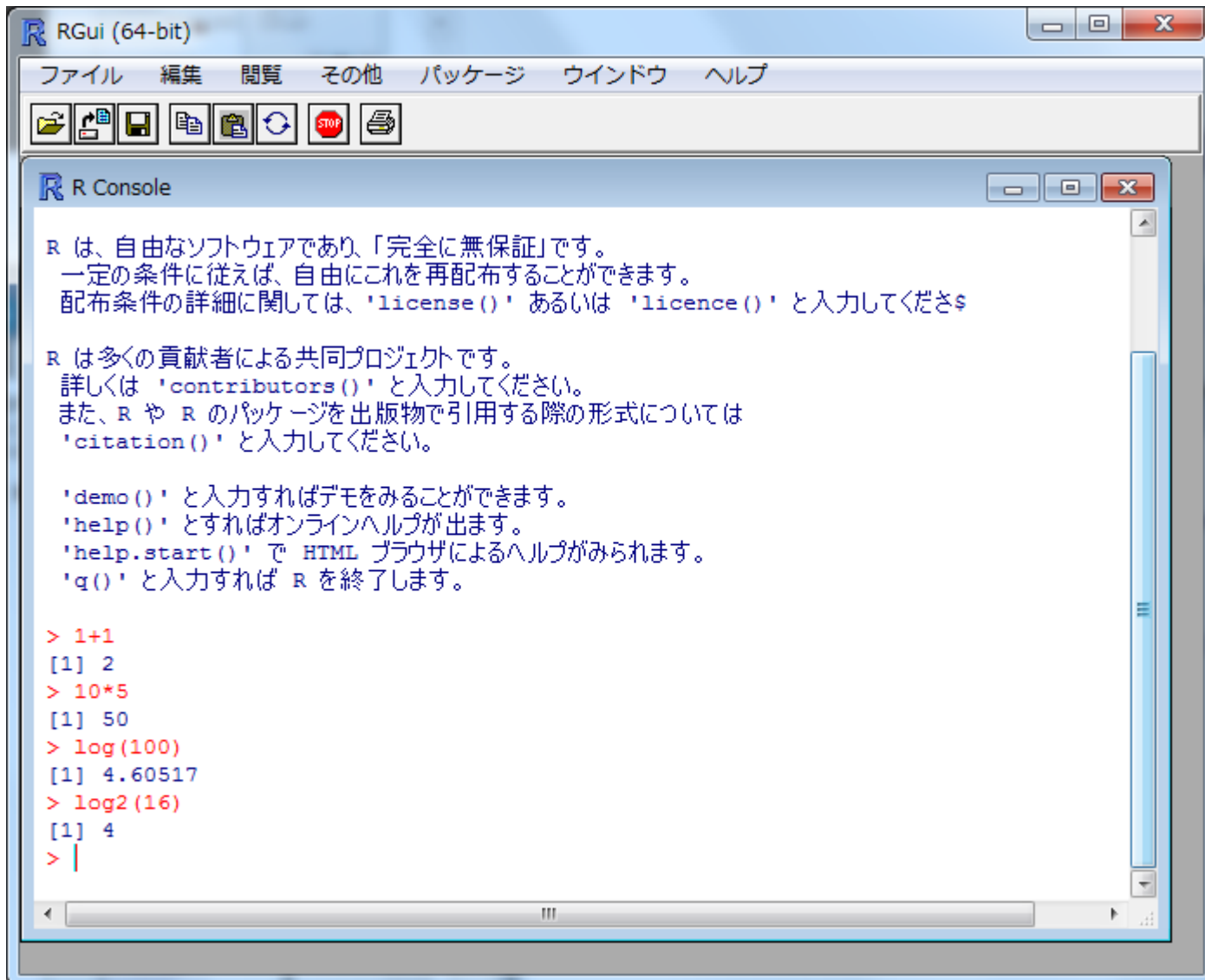
> |
```

Windows taskbar: 13:58 2015/03/31

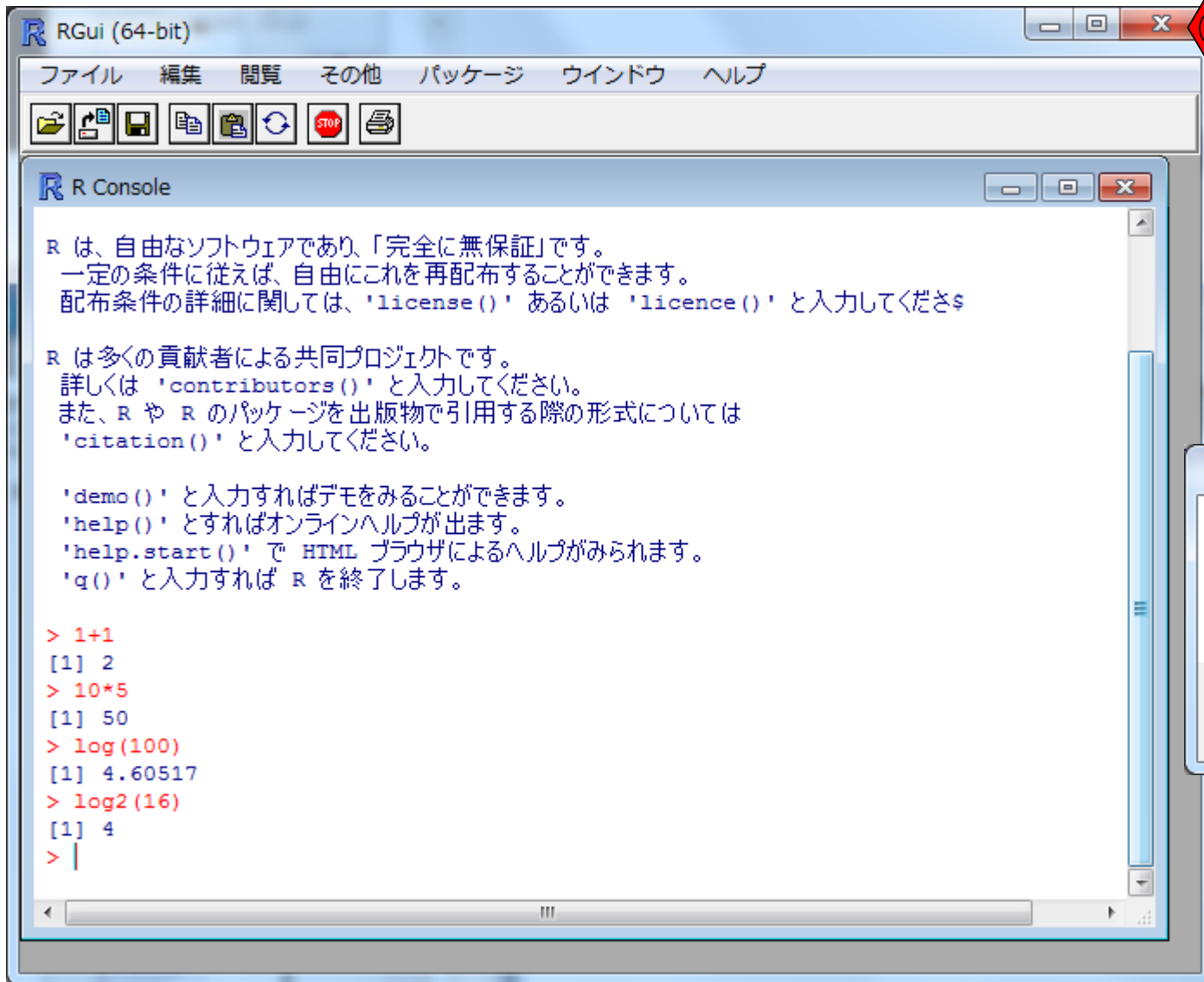
# Rの起動



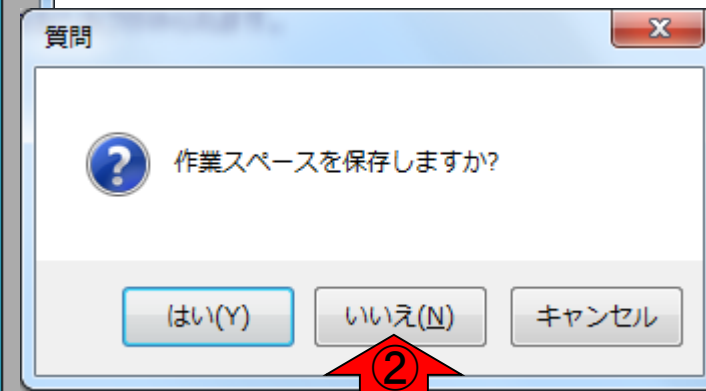
# 基本的な利用法



# Rの終了



①通常のソフトウェアと同様、右上の×ボタンを押せばよい。②「作業スペースを保存しますか?」というダイアログが出るが、最初のうちは「いいえ」でよい。(「はい」を押してしまっても.Rdataと.Rhistoryという2つのファイルが作業ディレクトリ上に作成されるだけなので特に問題はない。)





塩基配列を入力として、その翻訳されたアミノ酸配列を取得することができます

# 解析基礎1: 翻訳配列取得

- ・ [イントロ](#) | [一般](#) | [任意の位置の塩基を置換](#) (last modified 2013/09/12)
- ・ [イントロ](#) | [一般](#) | [指定した範囲の配列を取得](#) (last modified 2014/03/08)
- ・ [イントロ](#) | [一般](#) | [指定したID\(染色体やdescription\)の配列を取得](#) (last modified 2014/03/10)
- ・ [イントロ](#) | [一般](#) | [翻訳配列\(translate\)を取得\(基礎\)](#) | [Biostrings](#) (last modified 2015/03/09) **NEW**
- ・ [イントロ](#) | [一般](#) | [翻訳配列\(translate\)を取得\(応用\)](#) | [seqinr\(Chapman, 2005\)](#) (last modified 2015/03/09)
- ・ [イントロ](#) | [一般](#) | [相補鎖\(complement\)を取得](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [逆相補鎖\(reverse complement\)を取得](#)
- ・ [イントロ](#) | [一般](#) | [逆鎖\(reverse\)を取得](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [2連続塩基の出現頻度情報を取得](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [3連続塩基の出現頻度情報を取得](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [任意の長さの連続塩基の出現頻度情報を取得](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [Tips](#) | [任意の拡張子でファイルを保存](#)
- ・ [イントロ](#) | [一般](#) | [Tips](#) | [拡張子は同じで任意の文字を指定](#)
- ・ [イントロ](#) | [一般](#) | [配列取得](#) | [ゲノム配列](#) | [公共DBから](#)

## イントロ | 一般 | 翻訳配列(translate)を取得(基礎) | Biostrings **NEW**

[Biostrings](#)パッケージを用いて塩基配列を読み込んでアミノ酸配列に翻訳するやり方を示します。翻訳のための遺伝コード(genetic code)は、Standard Genetic Codeだそうです。もちろん生物種?!!によって多少違い(variants)があるようで、"Standard", "SGC0", "Vertebrate Mitochondrial", "SGC1"などいろいろ選べるようです。「ファイル」→「ディレクトリの変更」で解析したいファイルを置いてあるディレクトリに移動し以下をコピー。

### 1. FASTA形式ファイル(sample1.fasta)の場合:

multi-FASTAではないsingle-FASTA形式ファイルです。

```
in_f <- "sample1.fasta" #入力ファイル名を指定してin_fに格納
out_f <- "hoge1.fasta" #出力ファイル名を指定してout_fに格納

#必要なパッケージをロード
library(Biostrings) #パッケージの読み込み

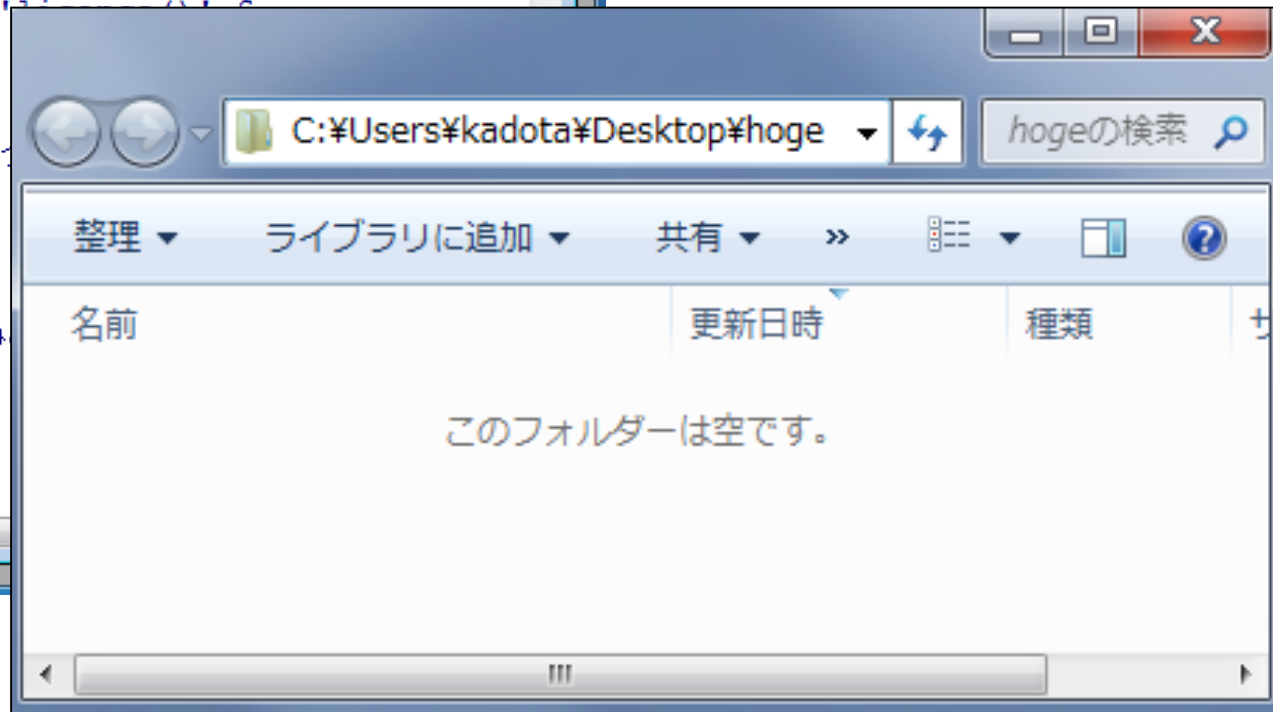
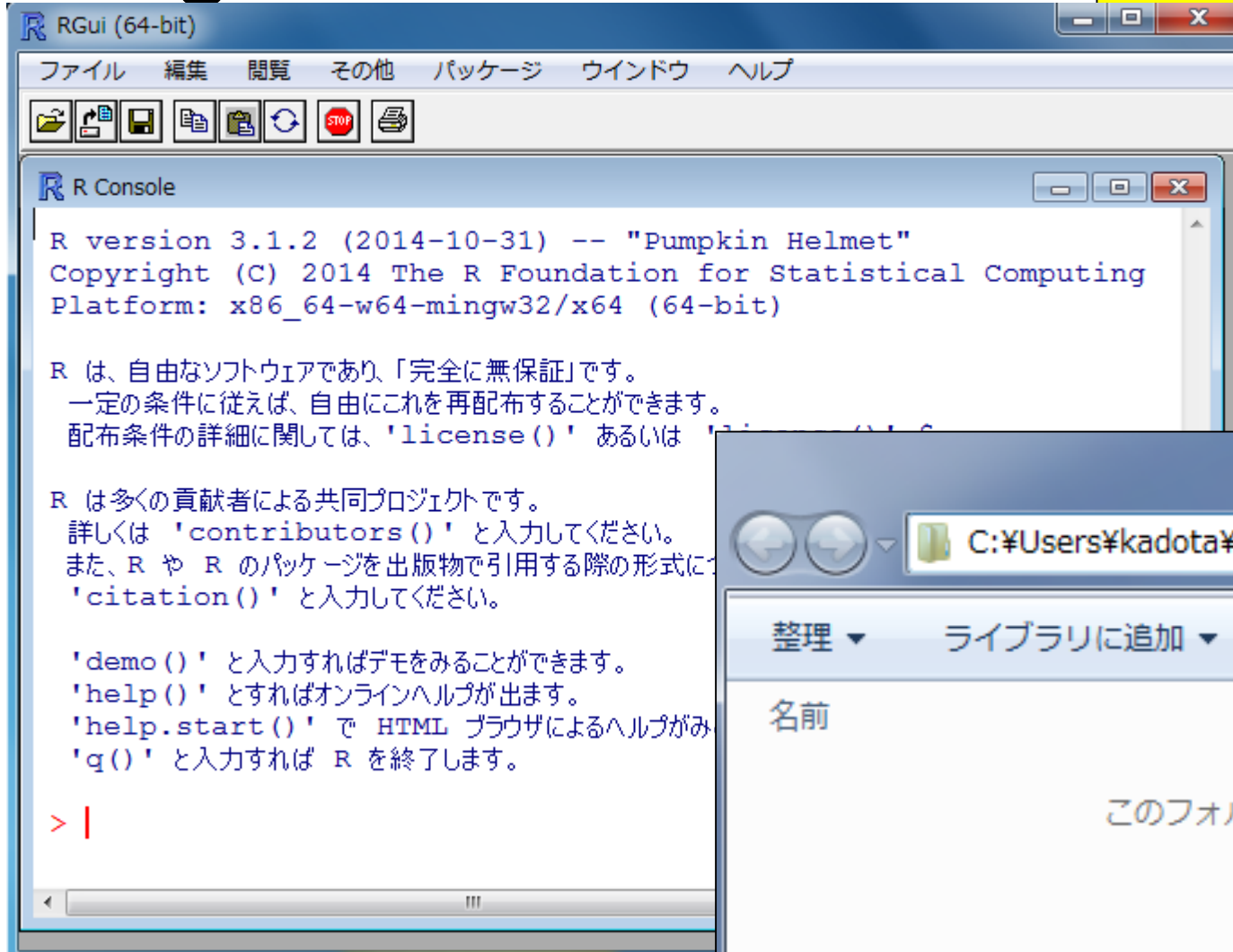
#入力ファイルの読み込み
fasta <- readDNAStringSet(in_f, format="fasta") #in_fで指定したファイルの読み込み
fasta #確認してるだけです

#本番
fasta <- translate(fasta) #アミノ酸配列に翻訳した結果をfastaに格納
fasta #確認してるだけです

#ファイルに保存
writeXStringSet(fasta, file=out_f, format="fasta", width=50) #fastaの中身を指定したフ
```

# hogeフォルダの作成

デスクトップにあるhogeフォルダ中のファイルを解析するやり方として説明します。  
デスクトップ上にhogeフォルダを作成



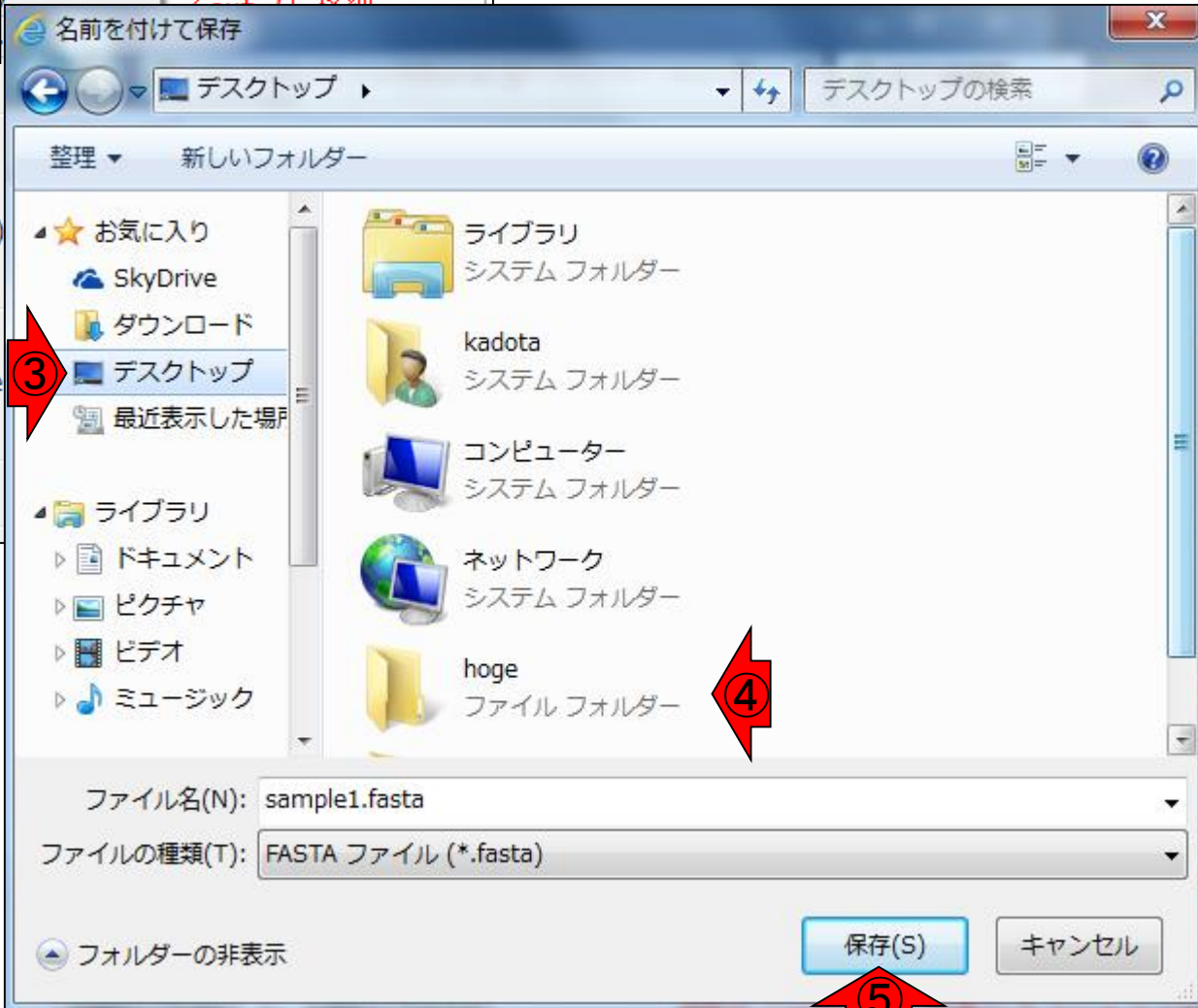
# ファイルの保存

①解析したいsample1.fastaのファイル名部分で  
右クリックして②対象をファイルに保存。③デスク  
トップ上に作成した④hogeフォルダに⑤保存

## I. FASTA形式ファイル(sample1.fasta)の場合:

```
multi-FASTAではないsingle FASTAの場合  
in_f <- "sample1.fasta"  
out_f <- "hoge1.fasta"  
  
#必要なパッケージをロード  
library(Biostrings)  
  
#入力ファイルの読み込み  
fasta <- readDNASTringSet(  
  fasta  
  
#本番  
fasta <- translate(fasta)  
fasta  
  
#ファイルに保存  
writeXStringSet(fasta, fil
```

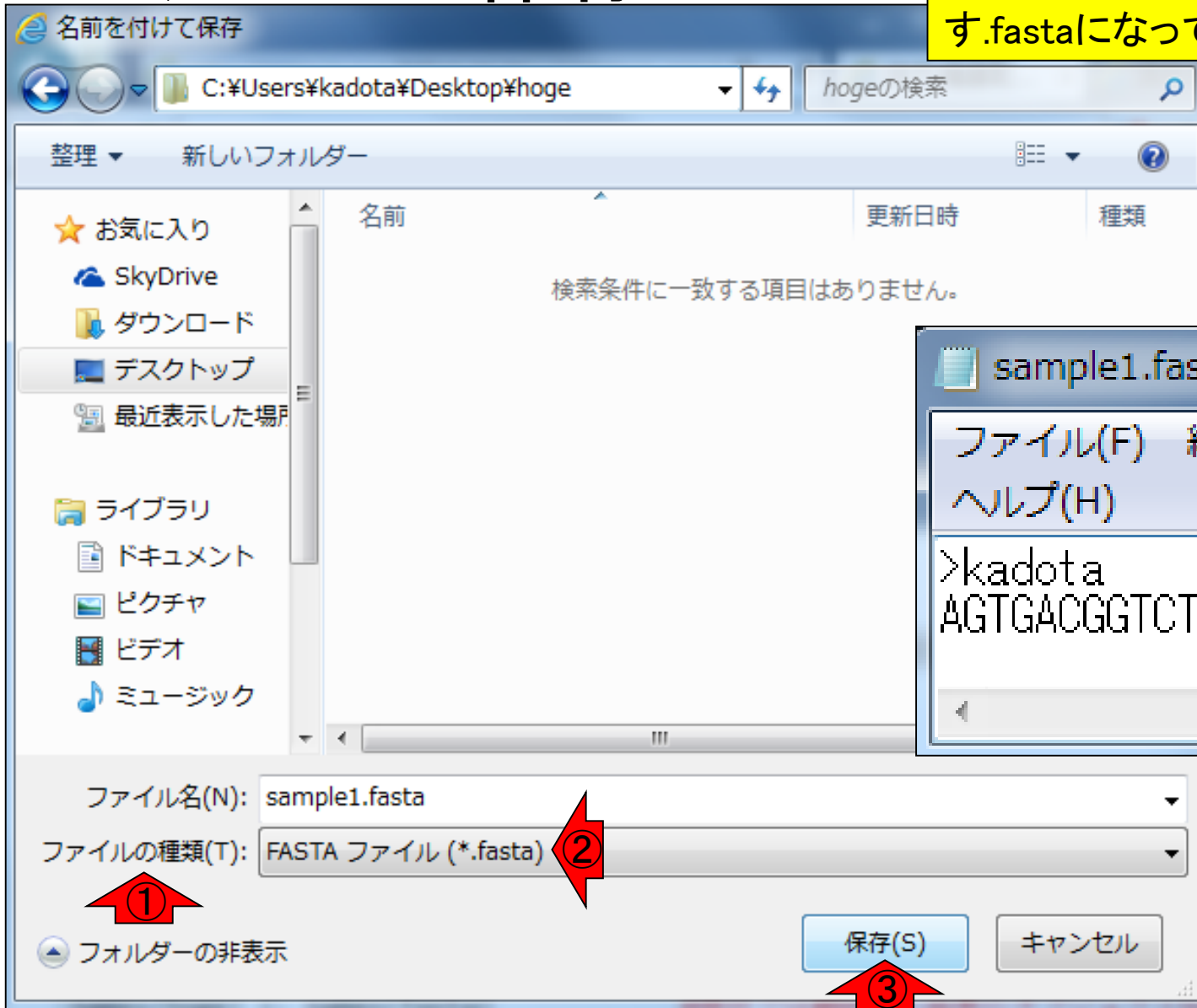
- 開く(O)
- 新しいタブで開く(W)
- 新しいウィンドウで開く(N)
- 対象をファイルに保存
- 対象を印刷(P)
- 切り取り
- コピー(C)
- ショートカットのコピー(T)
- 貼り付け(P)
- Bing で翻訳
- 電子メール (Windows Live)
- すべてのアクセラレータ
- 要素の検査(L)





# ファイルの保存

ときどき拡張子が\*.txtなどと勝手に変わっていることがあるので①ファイルの種類欄に注意。ここでは②FASTA形式ファイルであることを示す.fastaになっていることを確認して③保存



# 作業ディレクトリの変更

R起動直後のデフォルトの作業ディレクトリは、  
①ユーザ名 **kadota** のWindows環境では、「C:/Users/**kadota**/Documents」。その一方で、今解析したいディレクトリ(フォルダ)はデスクトップ上にある **hoge** なので、作業ディレクトリをそこに変更する必要があります。「`getwd()`」は、現在の作業ディレクトリを表示させるコマンド

R Console

```
R version 3.1.3 (2015-03-09) -- "Smooth Sidewalk"  
Copyright (C) 2015 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 'lic$
```

```
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$  
'citation()' と入力してください。
```

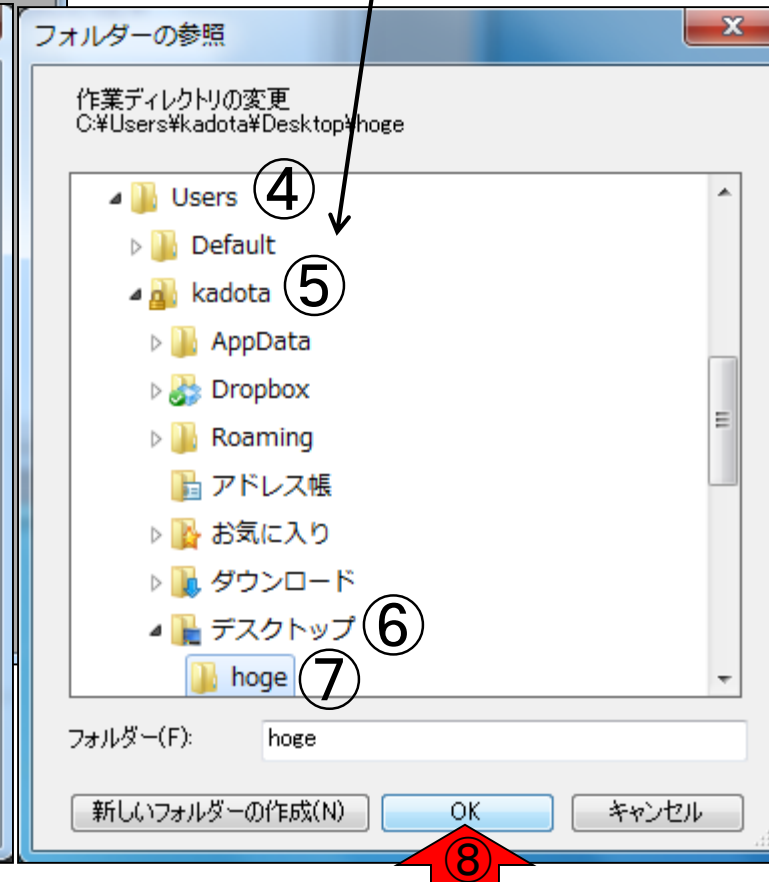
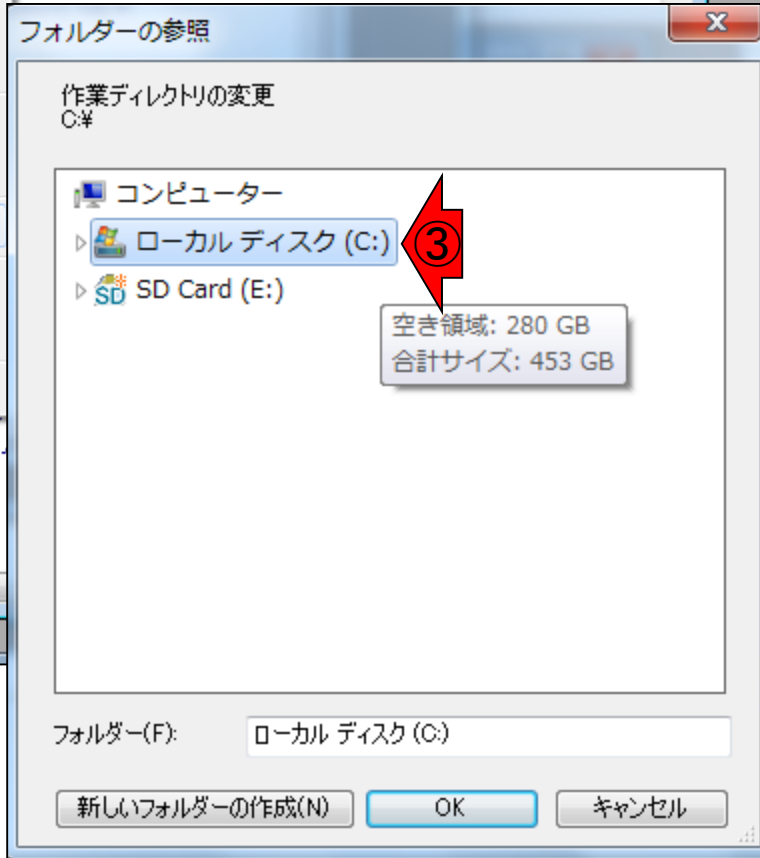
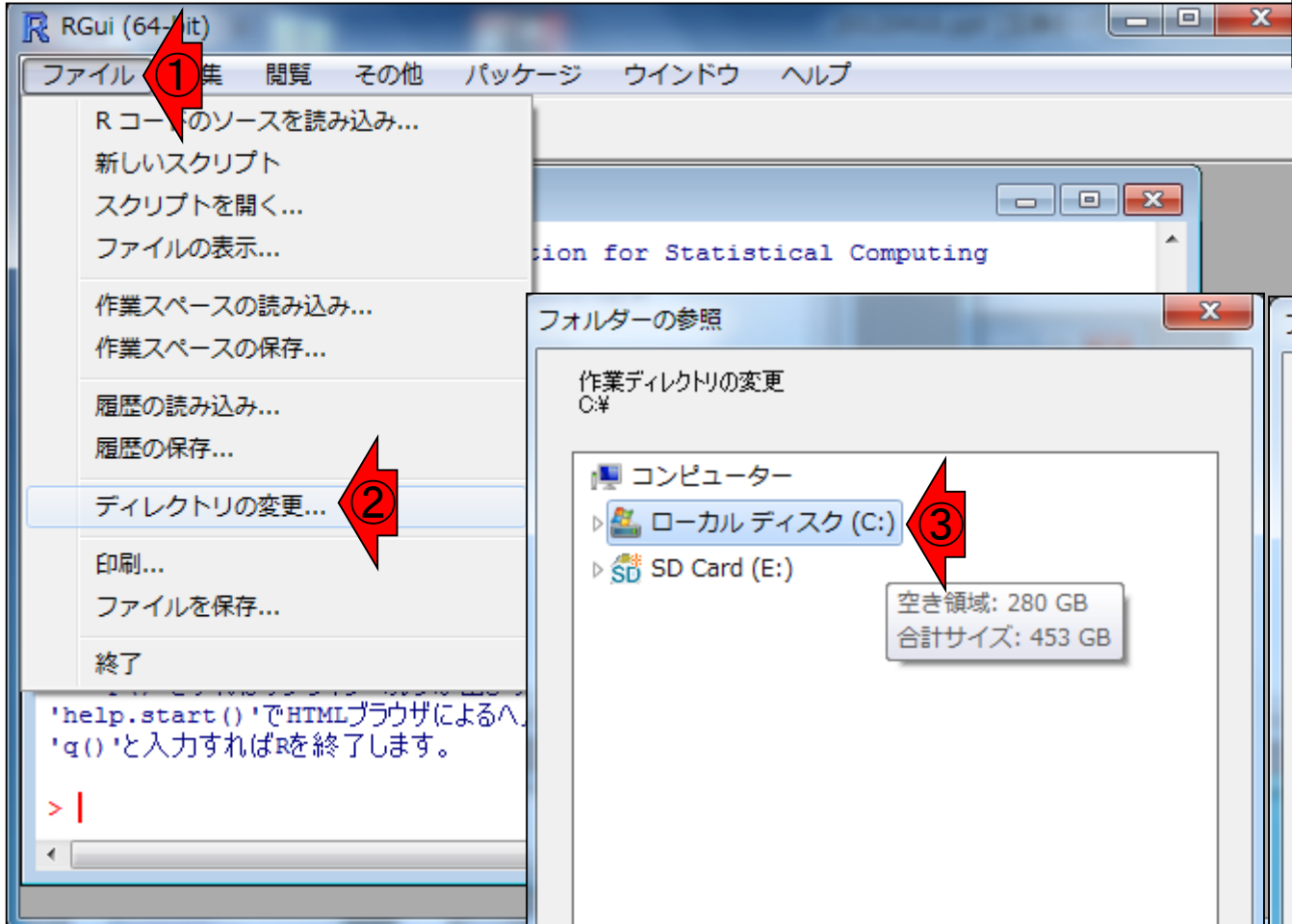
```
'demo()' と入力すればデモをみることができます。  
'help()' とすればオンラインヘルプが出ます。  
'help.start()' で HTML ブラウザによるヘルプがみられ$  
'q()' と入力すれば R を終了します。
```

```
> getwd()  
[1] "C:/Users/kadota/Documents"  
> |
```



# 作業ディレクトリの変更

①ファイル、②ディレクトリの変更。  
③「Windows(C:)」となっている場合もあるが、気にしない。⑤ヒトによって異なり、貸与PCの場合はiu



# getwd()と打ち込んで確認

当たり前ですが、解析したいディレクトリ(またはフォルダ)を正しく指定できていなければエラーに遭遇します。また、解析したいファイルが存在しない状態でもエラーが出ます。

```
R Console
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$
配布条件の詳細に関しては、'license()'あるいは 'lic$

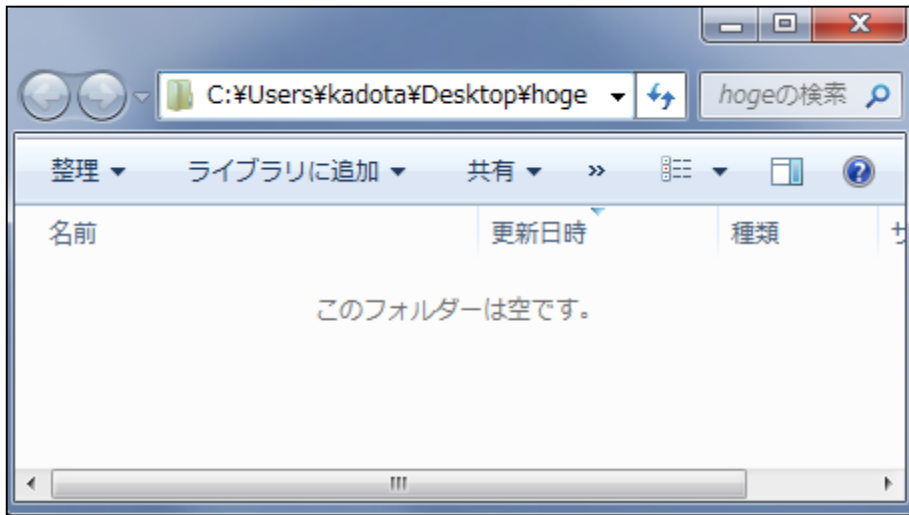
R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()'と入力してください。
また、R や R のパッケージを出版物で引用する際の形式$
'citation()'と入力してください。

'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$
'q()'と入力すれば R を終了します。

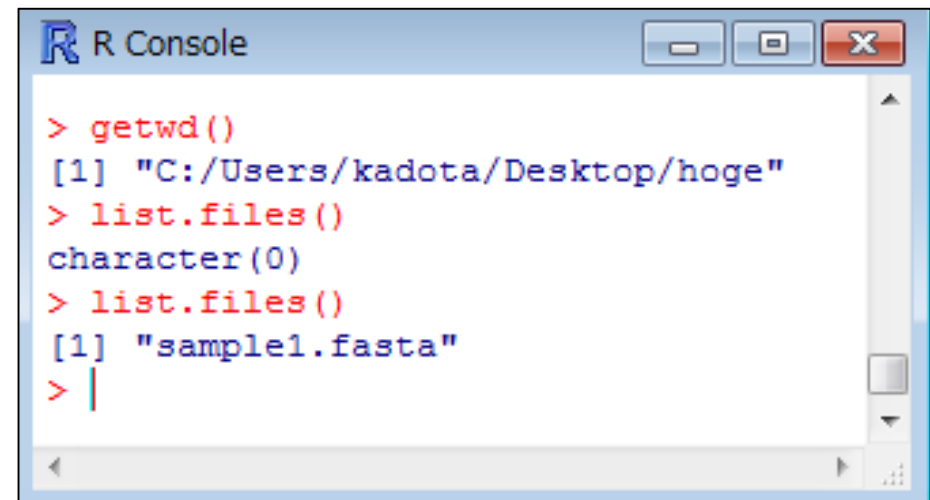
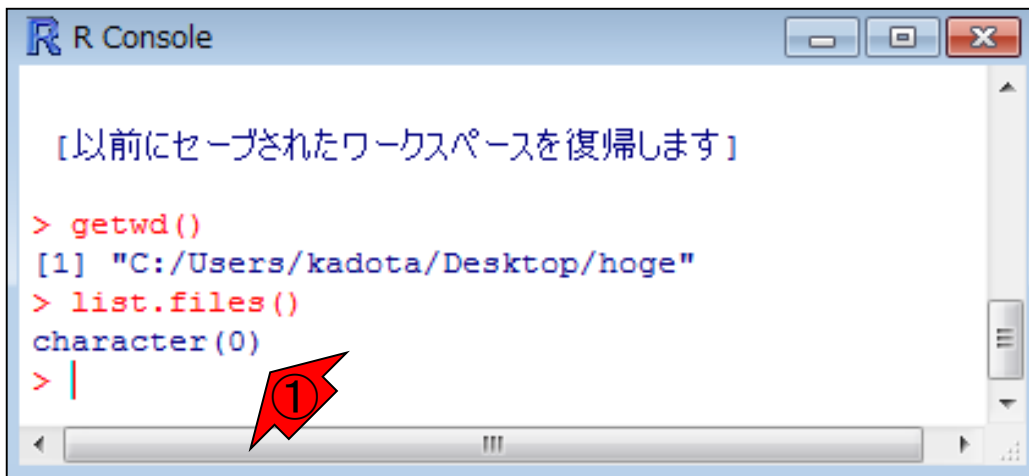
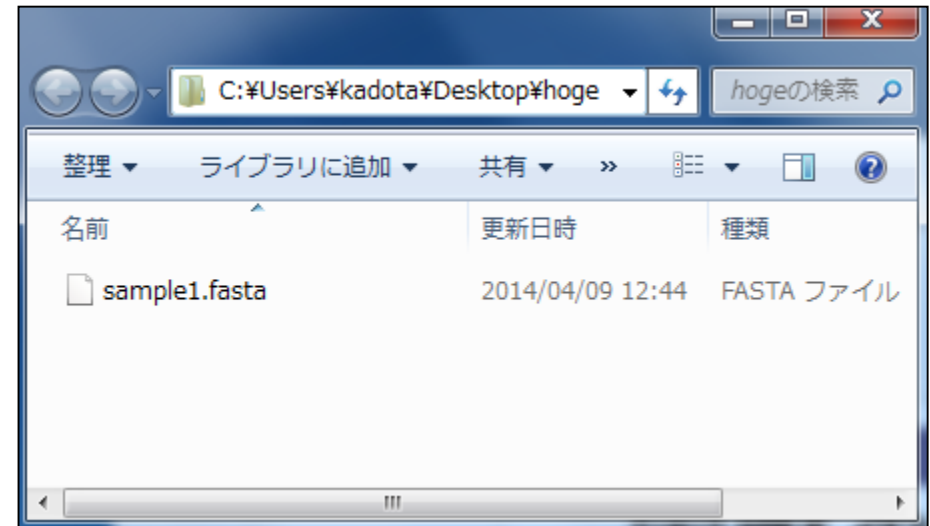
> getwd()
[1] "C:/Users/kadota/Documents"
> getwd()
[1] "C:/Users/kadota/Desktop/hoge"
> |
```

# 実際のhogeフォルダとR操作画面の関係

ファイル保存前



ファイル保存後



# 基本はコピー

## イントロ | 一般 | 翻訳配列(translate)を取得 **NEW**

塩基配列を読み込んでアミノ酸配列に翻訳するやり方を示します。

「ファイル」-「ディレクトリの変更」で解析したいファイルを置いてあるディレクトリに移動し以下をコピー。

### 1. FASTA形式ファイル(sample1.fasta)の場合:

multi-FASTAではない single-FASTA形式ファイルです。

```
in_f <- "sample1.fasta"
out_f <- "hoge1.fasta"
```

```
#必要なパッケージをロード
library(Biostrings)
```

```
#入力ファイルの読み込み
fasta <- readDNASTringSet(
  fasta
```

```
#本番
fasta <- translate(fasta)
fasta
```

```
#ファイルに保存
writeXStringSet(fasta, fil
```

切り取り(T)

コピー(C) **①**

貼り付け

すべて選択(A)

印刷(I)...

印刷プレビュー(N)...

Bing でマップ

Bing で翻訳

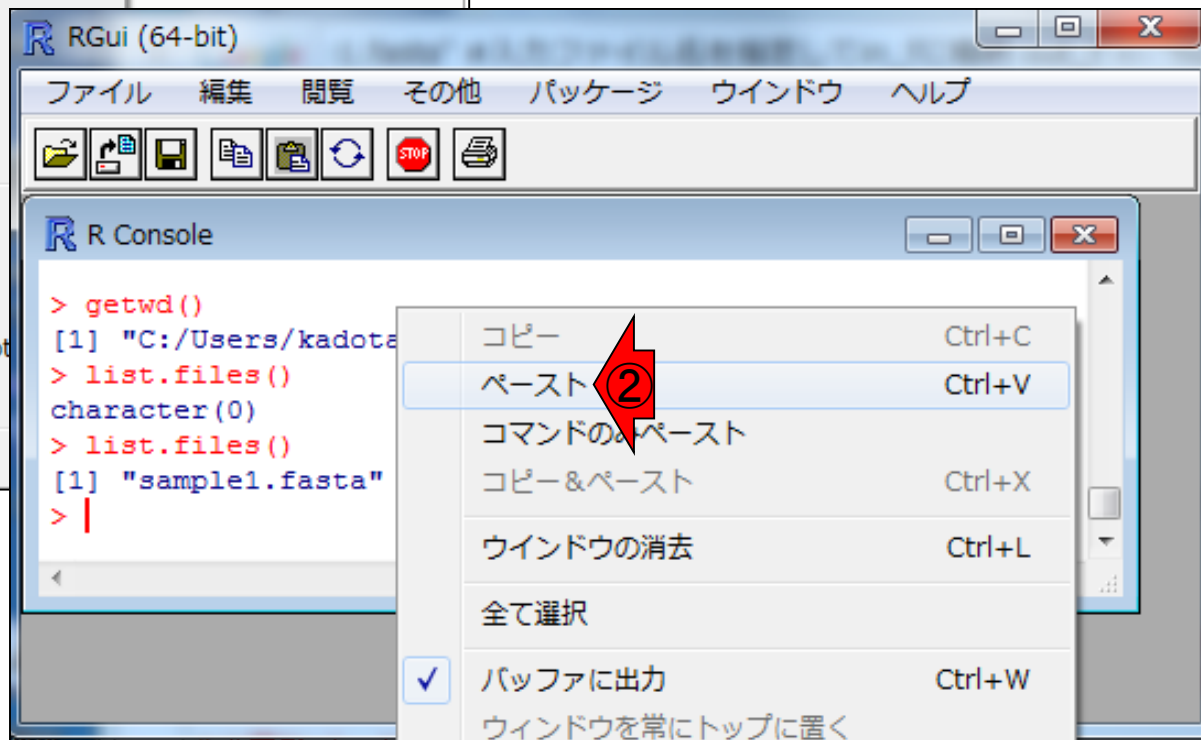
Google で検索

電子メール (Windows Live Hot

すべてのアクセラレータ

Send to OneNote

①一連のコマンド群をコピーして②R Console画面上でペースト。ブラウザがInternet Explorerの場合は、CTRLとALTキーを押しながらコードの枠内で左クリックすると、全選択できます。トリプルクリックでもよい。全選択の場合はできるかぎりこのやり方にしましょう



# 基本はコピペ

## イントロ | 一般 | 翻訳配列(translate)を取得 **NE**

塩基配列を読み込んでアミノ酸配列に翻訳するやり方を示します。  
「ファイル」-「ディレクトリの変更」で解析したいファイルを置いてあるディレクトリに移動してください。

### 1. FASTA形式ファイル(sample1.fasta)の場合:

multi-FASTAではないsingle-FASTA形式ファイルです。

```

in_f <- "sample1.fasta"      #入力ファイル名
out_f <- "hoge1.fasta"      #出力ファイル名

#必要なパッケージをロード
library(Biostrings)        #パッケージの読み込み

#入力ファイルの読み込み
fasta <- readDNASTringSet(in_f, format="fasta")#in_fで指定したファイルの読み込み
fasta                                     #確認してるだけ

#本番
fasta <- translate(fasta)    #アミノ酸配列に翻訳
fasta                         #確認してるだけ

#ファイルに保存
writeXStringSet(fasta, file=out_f, format="fasta", width=50)
    
```

エラーなく実行できた場合の全貌

```

R Console
> in_f <- "sample1.fasta"
> out_f <- "hoge1.fasta"      #出力ファイル名を指定してout_fに格納
>
> #必要なパッケージをロード
> library(Biostrings)        #パッケージの読み込み
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'

The following objects are masked from 'package:parallel':

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

The following object is masked from 'package:stats':

xtabs

The following objects are masked 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 XVector をロード中です

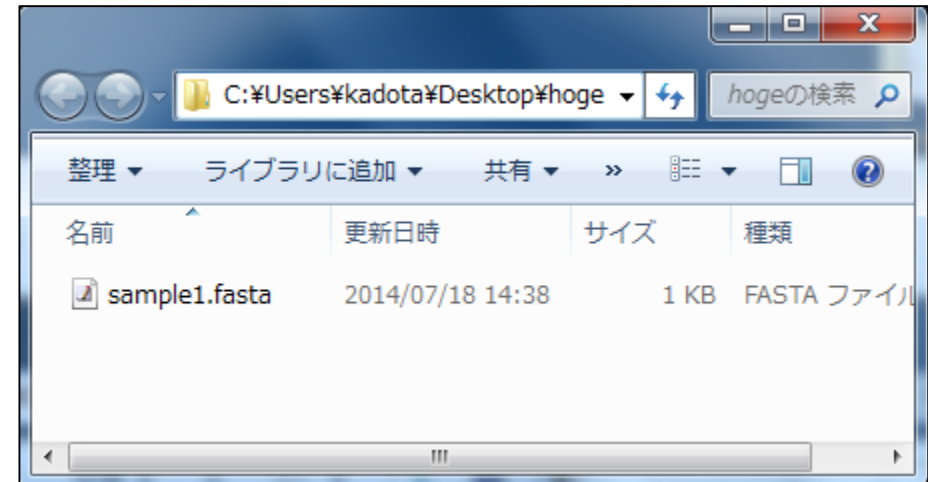
>
> #入力ファイルの読み込み
> fasta <- readDNASTringSet(in_f, format="fasta")#in_fで指定したファイルの読み込み
> fasta                                     #確認してるだけです
  A DNASTringSet instance of length 1
    width seq                      names
[1]    12 AGTGACGGTCTT             kadota
>
> #本番
> fasta <- translate(fasta)                #アミノ酸配列に翻訳した結果をfasta$
> fasta                                     #確認してるだけです
  A AAStringSet instance of length 1
    width seq                      names
[1]     4 SDGL                     kadota
>
> #ファイルに保存
> writeXStringSet(fasta, file=out_f, format="fasta", width=50)#fastaの中身を$
> |
    
```

# 実行結果

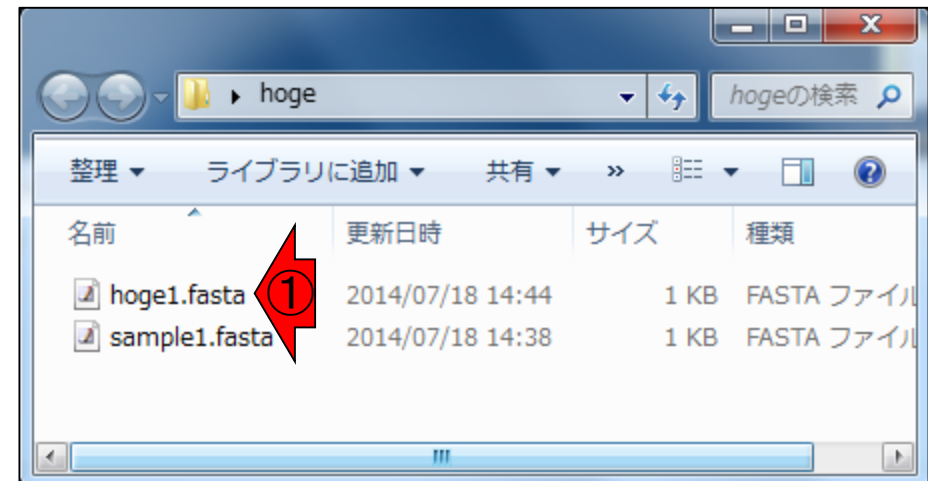
①出力ファイル名として指定したhoge1.fastaが生成されていることが分かります

```
R Console
> fasta <- readDNAStringSet(in_f, format="fasta")#$
> fasta #確認して$
A DNAStringSet instance of length 1
  width seq          names
[1]    12 AGTGACGGTCTT kadota
>
> #本番
> fasta <- translate(fasta) #アミノ酸$
> fasta #確認して$
A AAStringSet instance of length 1
  width seq          names
[1]     4 SDGL        kadota
>
> #ファイルに保存
> writeXStringSet(fasta, file=out_f, format="fasta$
> |
```

実行前のhogeフォルダ



実行後のhogeフォルダ





# 入出力の関係

## イントロ | 一般 | 翻訳配列(translate)を取得 **NEW**

塩基配列を読み込んでアミノ酸配列に翻訳するやり方を示します。

「ファイル」-「ディレクトリの変更」で解析したいファイルを置いてあるディレクトリに移動し以下をコピペ。

### 1. FASTA形式ファイル(sample1.fasta)の場合:

multi-FASTAではないsingle-FASTA形式ファイルです。

```
in_f <- "sample1.fasta" #入力ファイル名を指定してin_fに格納
out_f <- "hoge1.fasta" #出力ファイル名を指定してout_fに格納

#必要なパッケージをロード
library(Biostrings) #パッケージの読み込み

#入力ファイルの読み込み
fasta <- readDNAStringSet(in_f, format="fasta") #in_fで指定したファイルの読み込み
fasta #確認してるだけです

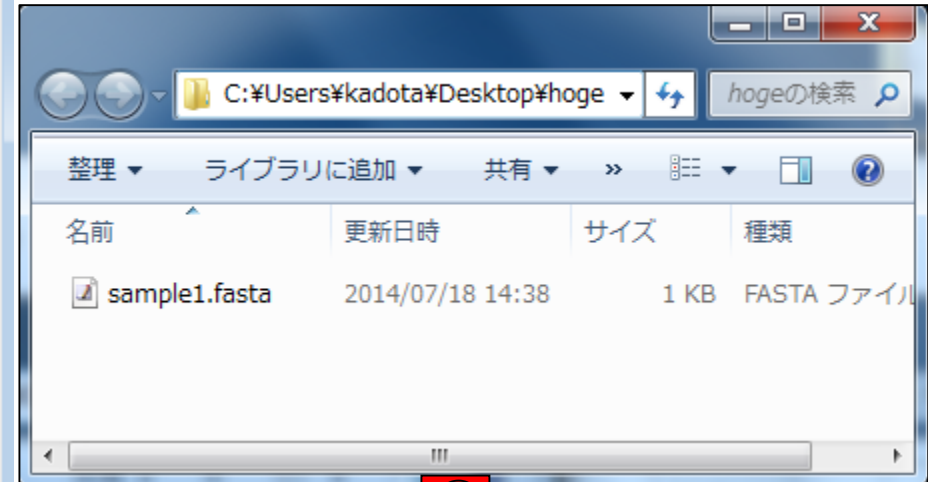
#本番
fasta <- translate(fasta) #アミノ酸配列に翻訳した結果をfastaに格納
fasta #確認してるだけです

#ファイルに保存
writeXStringSet(fasta, file=out_f, format="fasta", width=50) #fastaの中身を指定したフ
```

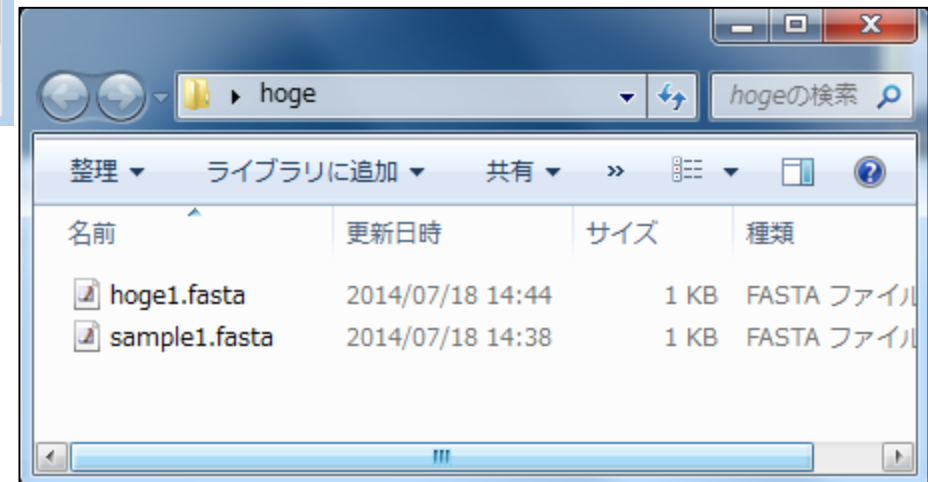
# 実行結果

①「list.files()で表示される結果」と②「実行後のhogeフォルダの中身」は当然同じ

実行前のhogeフォルダ



実行後のhogeフォルダ



```
R Console
> fasta <- readDNAStringSet(in_f, format="fasta")#$
> fasta                                     #確認して$
  A DNAStringSet instance of length 1
  width seq                                $
[1]    12 AGTGACGGTCTT                      $
>
> #本番
> fasta <- translate(fasta)                #アミノ酸$
> fasta                                     #確認して$
  A AAStringSet instance of length 1
  width seq                                $
[1]     4 SDGL                               $
>
> #ファイルに保存
> writeXString(fasta, file=out_f, format="fasta")
> list.files()
[1] "hoge1.fasta"  "sample1.fasta"
> |
```



入力ファイル中の塩基配列は、3の倍数の12塩基長、ACGTのみからなるので何のエラーもない

# 実行結果

入力: 塩基配列ファイル (sample1.fasta)

```
>kadota
AGTGACGGTCTT
```

出力: アミノ酸配列ファイル (hoge1.fasta)

```
>kadota
SDGL
```

# コドン表

<http://ja.wikipedia.org/wiki/%E3%82%B3%E3%83%89%E3%83%B3>

表1. 64コドンと各々に対応するアミノ酸を示したもの。mRNAの方向は5'から3'である。

		2nd base			
		U	C	A	G
1st base	U	UUU (Phe/F)フェニルアラニン	UCU (Ser/S)セリン	UAU (Tyr/Y)チロシン	UGU (Cys/C)システイン
		UUC (Phe/F)フェニルアラニン	UCC (Ser/S)セリン	UAC (Tyr/Y)チロシン	UGC (Cys/C)システイン
		UUA (Leu/L)ロイシン	UCA (Ser/S)セリン	UAA Ochre (終止)	UGA Opal (終止)
		UUG (Leu/L)ロイシン	UCG (Ser/S)セリン	UAG Amber (終止)	UGG (Trp/W)トリプトファン
	C	CUU (Leu/L)ロイシン	CCU (Pro/P)プロリン	CAU (His/H)ヒスチジン	CGU (Arg/R)アルギニン
		CUC (Leu/L)ロイシン	CCC (Pro/P)プロリン	CAC (His/H)ヒスチジン	CGC (Arg/R)アルギニン
		CUA (Leu/L)ロイシン	CCA (Pro/P)プロリン	CAA (Gln/Q)グルタミン	CGA (Arg/R)アルギニン
		CUG (Leu/L)ロイシン	CCG (Pro/P)プロリン	CAG (Gln/Q)グルタミン	CGG (Arg/R)アルギニン
	A	AUU (Ile/I)イソロイシン	ACU (Thr/T)スレオニン	AAU (Asn/N)アスパラギン	AGU (Ser/S)セリン
		AUC (Ile/I)イソロイシン	ACC (Thr/T)スレオニン	AAC (Asn/N)アスパラギン	AGC (Ser/S)セリン
		AUA (Ile/I)イソロイシン, (開始)	ACA (Thr/T)スレオニン	AAA (Lys/K)リジン	AGA (Arg/R)アルギニン
		AUG (Met/M)メチオニン, (開始) <sup>[3]</sup>	ACG (Thr/T)スレオニン	AAG (Lys/K)リジン	AGG (Arg/R)アルギニン
G	GUU (Val/V)バリン	GCU (Ala/A)アラニン	GAU (Asp/D)アスパラギン酸	GGU (Gly/G)グリシン	
	GUC (Val/V)バリン	GCC (Ala/A)アラニン	GAC (Asp/D)アスパラギン酸	GGC (Gly/G)グリシン	
	GUA (Val/V)バリン	GCA (Ala/A)アラニン	GAA (Glu/E)グルタミン酸	GGA (Gly/G)グリシン	
	GUG (Val/V)バリン, (開始)	GCG (Ala/A)アラニン	GAG (Glu/E)グルタミン酸	GGG (Gly/G)グリシン	

# (Rで)塩基配列解析

①の手順に沿ってインストールを行えば、来週以降は持込PCでも講義を受けることができます。貸与PC利用のヒトも一通り眺めておきましょう。

## (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 2016/04/08, since 2011)

### What's new?

- このウェブページはインストール | ① についての推奨手順 (Windows2015.04.04版とMacintosh2015.04.03版) | ① に従ってフリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は基本的な利用法(Windows2015.04.03版とMacintosh2015.04.03版)で自習してください。本ウェブページを体系的にまとめた書籍もあります。(2015/04/03)
- 平成28年度ハンズオン講習会が7/19-8/4の日程で開催されます。申込み期間は4/4-5/31です。追加申込みやキャンセル待ちの受付は絶対に行いませんので受講希望者はお気をつけください。門田担当部分は主にこれまでの受講生を対象としています(リクエストを多く反映させています)。「昨年度との違い、対応関係、想定受講者、および予習事項」については [こちらのPDF](#) (約2MB) をご覧になり、必要に応じてつまみ食いしてください。(2016/04/05) **NEW**
- QuasRでBowtieのマッピングを行う場合に、(内部的にはbowtie1が動いているため)リード長が1本でも1,024 bpを超えたものがあればコケマス(1024 bpはセーフで1025 bpはアウト)のご注意ください(高橋 広夫 氏提供情報)。(2016/04/06) **NEW**
- 日本乳酸菌学会誌のNGS関連連載の第6回ウェブ資料を更新しました。具体的には、W20-2のスライド256の解説内容を変更しました。(2016/03/29) **NEW**
- RNA-Seq実験ハンドブック(鈴木穰 編)が刊行されます。(2016/03/22) **NEW**
- 「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)」のgetFASTQinfoやgetFASTQfile関数実行部分でsra\_conという引数を追加しました。が、おそらくデータを取りに行っているENAのURL ([http://www.ebi.ac.uk/ena/data/view/reports/sra/fastq\\_files/](http://www.ebi.ac.uk/ena/data/view/reports/sra/fastq_files/))がリンク切れになっているようでうまく情報の取得ができなくなっているようですm(\_)\_mもちろん私のせいではありませんw。ご指摘深謝m(\_)\_m(2016/03/17) **NEW**

- [はじめに](#) (last modified 2015/03/31)
- [参考資料\(講義、講習会、本など\)](#) (last modified 2016/02/29)
- [過去のお知らせ](#) (last modified 2016/04/05) **NEW**
- [インストール | ① について](#) (last modified 2015/11/12)
- インストール | R本体 | 最新版 | [Win用](#) (last modified 2015/03/22) 推奨
- インストール | R本体 | 最新版 | [Mac用](#) (last modified 2015/04/22) 推奨
- インストール | R本体 | 過去版 | [Win用](#) (last modified 2015/03/22)
- インストール | R本体 | 過去版 | [Mac用](#) (last modified 2015/03/22)

# パッケージインストール確認

キーボードの上矢印キーを1回押すと直前に打ち込んだコマンドが表示される。もう一度リターンキーを押して実行すると、何のメッセージも表示されなくなる。これもエラーが出ていないのでOK。

```
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'

The following objects are masked from 'package:parallel':

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

The following object is masked from 'package:stats':

  xtabs

The following objects are masked 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, unsplit

要求されたパッケージ S4Vectors をロード中です
要求されたパッケージ stats4 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 XVector をロード中です
> library(Biostrings)
> |
```

# パッケージインストール確認

キーボードの上矢印キーなどを利用して、次にShortReadパッケージの確認を行う。エラーメッセージが出ていないことがわかる。

```
R Console

The following object is masked from 'package:stats':

  xtabs

The following objects are masked 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, unsplit

要求されたパッケージ S4Vectors をロード中です
要求されたパッケージ stats4 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 XVector をロード中です
> library(Biostrings)
> 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 GenomeInfoDb 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 GenomicAlignments をロード中です
> library(ShortRead)
> |
```

# パッケージインストール確認

```
R Console  
  
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, unsplit  
  
要求されたパッケージ S4Vectors をロード中です  
要求されたパッケージ stats4 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 XVector をロード中です  
> library(Biostrings)  
> 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 GenomeInfoDb 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 GenomicAlignments をロード中です  
> library(ShortRead)  
> library(Shortread)  
以下にエラー library(Shortread) :  
  'Shortread' という名前のパッケージはありません  
> library(shortread)  
以下にエラー library(shortread) :  
  'shortread' という名前のパッケージはありません  
> library(shortRead)  
以下にエラー library(shortRead) :  
  'shortRead' という名前のパッケージはありません  
> |
```

