




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とRのイントロ～



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

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

講義予定

- 4月17日月曜日(17:15-20:30)
 - 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日火曜日(17:15-20:30)
 - 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24日月曜日(17:15-20:30)
 - 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01日月曜日(17:15-19:00頃)
 - 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R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 ⑨ 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。講義日程が当初の予定から変更になりました。 **NEW!! (3月30日更新)**

講義日程 (平成29年度)

1. 平成29年04月17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](#)
[講義資料PDF\(Win版 ; 完全版\)](#)
[講義資料PDF\(Mac版 ; Rの説明部分のみ\)](#)

2. 平成29年04月18日 (PC使用)

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 (2017年04月17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。学会に入っているとメーリングリストでも流れる。2017年09/27-09/29に第6回生命医薬情報学連合大会が北海道大学情報科学研究科棟で開催予定。
- ・ [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

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2017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 (2017年04月17日) [門田幸二](#)

① 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。学会に入っているとメールリストでも流れる。2017/09/29に[第6回生命医薬情報学連合大会](#)が北海道大学情報科学研究科棟で開催
- 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(次世代シークエンサ)

- [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9月3日 [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と共同主催でハンズオン(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この模。
- [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4人
 - 講習会映像は[統合TV](#)と[Youtube](#)から公開
 - [報告書](#)

ホーム	開催趣旨	プログラム
参加登録	発表登録	実行委員会
アクセス	リンク	問い合わせ
BoF参加登録		



開催概要

大会名：第6回生命医薬情報学連合大会(IIBMP 2017)
 大会長：遠藤俊徳(北海道大学大学院情報科学研究科)
 日程：2017年9月27日(水)～9月29日(金)
 会場：北海道大学情報科学研究科棟
 主催：
 日本バイオインフォマティクス学会(JSBI)
 日本オミックス医療学会
 情報計算化学生物学会(CBI学会)

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 (2017年04月17日) 門田幸二

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。学会に入っていると09/29に第6回生命医薬情報学連合大会が北海道大
- ・ 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-M)

バイオインフォマティクス系1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(次世代)

- ・ バイオサイエンスデータベースセンター(NBDC)運営委員会 [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主催でハンズオン(ノートPCを用いた実習型)講義を実施。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
 - 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
 - 座学部分は全て自習にしてハンズオンに特化。講師
 - 講習会映像は[統合TV](#)と[Youtube](#)から公開
 - [報告書](#)

ここははじめてですか? FAQをチェックしましょう。

ログイン 概要 よくある質問

質問 タグ ユーザー バッジ 未回答 質問する

検索する

●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

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

184 質問
308 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BioStar · 新型シーケンサーの

0 投票	6 回答	2.4k 閲覧	DRAのexperimentとRun	dra experiment run	二日前 moss 1
0 投票	2 回答	3.4k 閲覧	Cuffdiff を実行すると、Segmentation fault: 11	cuffdiff	Mar 28 at 10:56 moss 1
0 投票	0 回答	86 閲覧	cuffmergeの結果について	cufflinks	Mar 21 at 17:48 mlck 1
0 投票	1 回答	458 閲覧	BioMart V0.9を使い倒す のリンク不具合	biomart	Dec 23 '16 at 21:41 Hiromasa Ono 11
0 投票	1 回答	966 閲覧	RNA-seqデータを用いたエンリッチメント解析について	ma-seq エンリッチメント解析	Aug 15 '16 at 20:46 Ionicera 1
0 投票	1 回答	1.3k 閲覧	ホモログの検出について	ma-seq homolog	Jul 15 '16 at 14:03 Ionicera 1
1 投票	0 回答	1.4k 閲覧	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	1000genome population_genetics daf vcf af	Jan 09 '16 at 15:47 suimye 296
0 投票	3 回答	1.9k 閲覧	RNA-seqのAT bias	stranded ma-seq miseq	Jan 05 '16 at 17:58 suimye 296
0 投票	1 回答	1.4k 閲覧	IGVの結果とFPKM値の違いについて	ma-seq	Jan 05 '16 at 15:57 suimye 296
0 投票	1 回答	1.5k 閲覧	GEO からデータのダウンロード : MAX formatとは?	geo	Oct 22 '15 at 19:35 kh 1

相談窓口

気軽に質問していることがわかります

0	1	775	初学者的質問: paired end readのクオリティチェック	paired end quality check	Sep 16 '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	1,448	Togo picture gallery リクエスト	togopic abcis 編集済	Sep 01 '14 at 12:00 suimye 286
1	3	1,776	アライメントについて	ngs	Aug 11 '14 at 23:14 mn3 ♦♦ 5
1	1	1,544	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 genetica1127
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 186 ● 1 ● 2 ● 9
0	1	2,216	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 186 ● 1 ● 2 ● 9
0	1	1,884	RNA-seq データの定量と可視化の、ツールの組み合わせ	htseq rpkm express cufflinks ma-star	Apr 16 '14 at 05:06 nob_fj ♦ 5
0	3	2,316	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 tzumi
0	3	916	Cufflinks (binary)が実行できない	cufflinks	Apr 08 '14 at 13:28 meg
0	2	2,216	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 tzumi



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)連携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(①)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
- ・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
- ・中~上級レベルの内容を新規提供。講師数6人
- ・講習会映像は統合TVとYoutubeから公開
- ・報告書(もうすぐ公開2017.04.10現在)

・NGSハンズオン講習会(平成29年度)

- ・東大農で2017年08月下旬に実施予定



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

NGS速習コース講習会

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

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
 - ・ 中~上級レベルの内容を新規提供。講師数6人
 - ・ 講習会映像は統合TVとYoutubeから公開
 - ・ 報告書(もうすぐ公開2017.04.10現在)

- ・ **NGSハンズオン講習会(平成29年度)**
 - ・ 東大農で2017年08月下旬に実施予定

人材育成カリキュラム(次世代シーケンサ)

・ 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	OS、ハード構成	・コンピュータの基本的理解	初級	講義			
1	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, がんゲノム、個人ゲノム、環境ゲノム, Hi-C)	初級	講義			
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		
0.5日	5.	ゲノム関連の倫理・法律	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
0.5日	6.	分子生命科学	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



NGSハンズオン講習会

①ハンズオンに特化した講習会を実施。
2015年は約60名で計14日間、
2016年は約60-90名が12日間受講

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
 - 中～上級レベルの内容を新規提供。講師数6人
 - 講習会映像は[統合TV](#)と[Youtube](#)から公開
 - [報告書](#)(もうすぐ公開2017.04.10現在)
- ・ [NGSハンズオン講習会](#) (平成29年度)
 - 東大農で2017年08月下旬に実施予定



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(NBDC)で、2014年3月に策定されたNGS用カリキュラムに基づいて、平成26年度よりNBDCと東大ア(ノートPCを用いた実習型)講義を実施。おそ人数規模。

- 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(日程)に実施
 - 座学部分は全て自習にしてハンズオン(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日に中~上級レベルの内容を新規提供。講習会映像は統合TVとYoutubeから公開
 - 報告書(もうすぐ公開2017.04.10現在)

- NGSハンズオン講習会(平成29年度)
 - 東大農で2017年08月下旬に実施予定



— 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

科学技術振興機構
JST

文字サイズ変更 大 中 小

Search for... Search

ホーム
NBDCについて
研究開発
公募情報
採用情報
イベント
人材支援
アクセス
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

カリキュラム (PDF: 72KB)

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	はじめに(講習会参加者必読) PC環境の構築	Bio-Linux8とRのインストール状況確認	・ 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 ・ 共有フォルダ設定完了確認 ・ 基本的なLinuxコマンドの習得状況確認 ・ R本体およびパッケージのインストール確認 ・ 講師指定の事前予習内容の再確認 ・ 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	主催・共催機関	② 講義資料 (PDF: 4MB)
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1)	ゲノム解析、塩基配列解析	・ NGS解析手段、ウェブツール(DBJ Pipeline)との連携 ・ 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 ・ 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) ・ 塩基配列解析を行うための基本スキルの復習や作図 ・ 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	門田 幸二 (東京大学)	講義資料 (PDF: 7.3MB) 解析データ (ZIP: 2.2MB) 統合TV
7月21日(木)	10:30-18:15		トランスクリプトーム解析1	・ カウントデータ取得以降の統計解析(RNA-seq) ・ サンプル間クオスタリティ、結果の解釈		講義資料 (PDF: 6.6MB)

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

NBDC National Bioscience Database Center
 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
 Search for... Search

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

アグリバイオの教育プログラム



ハズオン講習会カリキュラム

資料・動画等
[カリキュラム \(PDF: 72KB\)](#)

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)
はじめに (講習会参加者必読)	Bio-Linux8とRのインストール状況確認	<ul style="list-style-type: none"> 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 共有フォルダ設定完了確認 基本的なLinuxコマンドの習得状況確認 R本体およびパッケージのインストール確認 講師指定の事前学習内容の再確認 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 	主催・共催機関	講義資料 (PDF:4MB)
PC環境の構築				
第1部 統計解析 (農学生命情報科学特論1)	ゲノム解析、塩基配列解析	<ul style="list-style-type: none"> NGS解析手段、ウェブツール(DDBJ Pipeline)との連携 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 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) 塩基配列解析を行うための基本スキルの復習や作図 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 	門田 幸二 (東京大学)	講義資料 (PDF:7.3MB) 解析データ (ZIP:2.2MB) 統合TV
	トランスクリプトーム解析1	<ul style="list-style-type: none"> カウントデータ取得以降の統計解析(RNA-seq) サンプル間クラスタリング、結果の解釈 		講義資料 (PDF:6.5MB)

7月20日 (水)	10:30-18:15
7月21日 (木)	10:30-18:15

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。
- ・ [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。遺伝研スパコンの使い方などについても丁寧な講習会資料あり。



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。
 - [日本乳酸菌学会誌](#)上で2014年から連載中のNGS記事も[こちら](#)で提供。
- ・ [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。
- ・ [Biopapyrus](#)
 - ここも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。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

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](#)
 - おそらく受講生あたりのスタッフ数かも
- ・ [DDBJのDDBJing講習会](#)
 - DNA Data Bank of Japan (DDBJ)が提
伝研スパコンの使い方などについても

自習用教材

- ・ [\(Rで\)塩基配列解析](#)
 - NGS系のそれっ(まいキーワードで検索
ストールからLinux環境でのNGS解析ま
 - [日本乳酸菌学会誌](#)上で2014年から連
- ・ [統合TV](#)
 - 有用なデータベースやウェブツールの
- ・ [Biopapyrus](#)
 - ここもNGS系のそれっ(まいキーワードで
Linux, Perlなどの各種インストール系

(Rで)塩基配列解析

(last modified 2017/04/10, since 2010)

このウェブページのR関連部分は、[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)

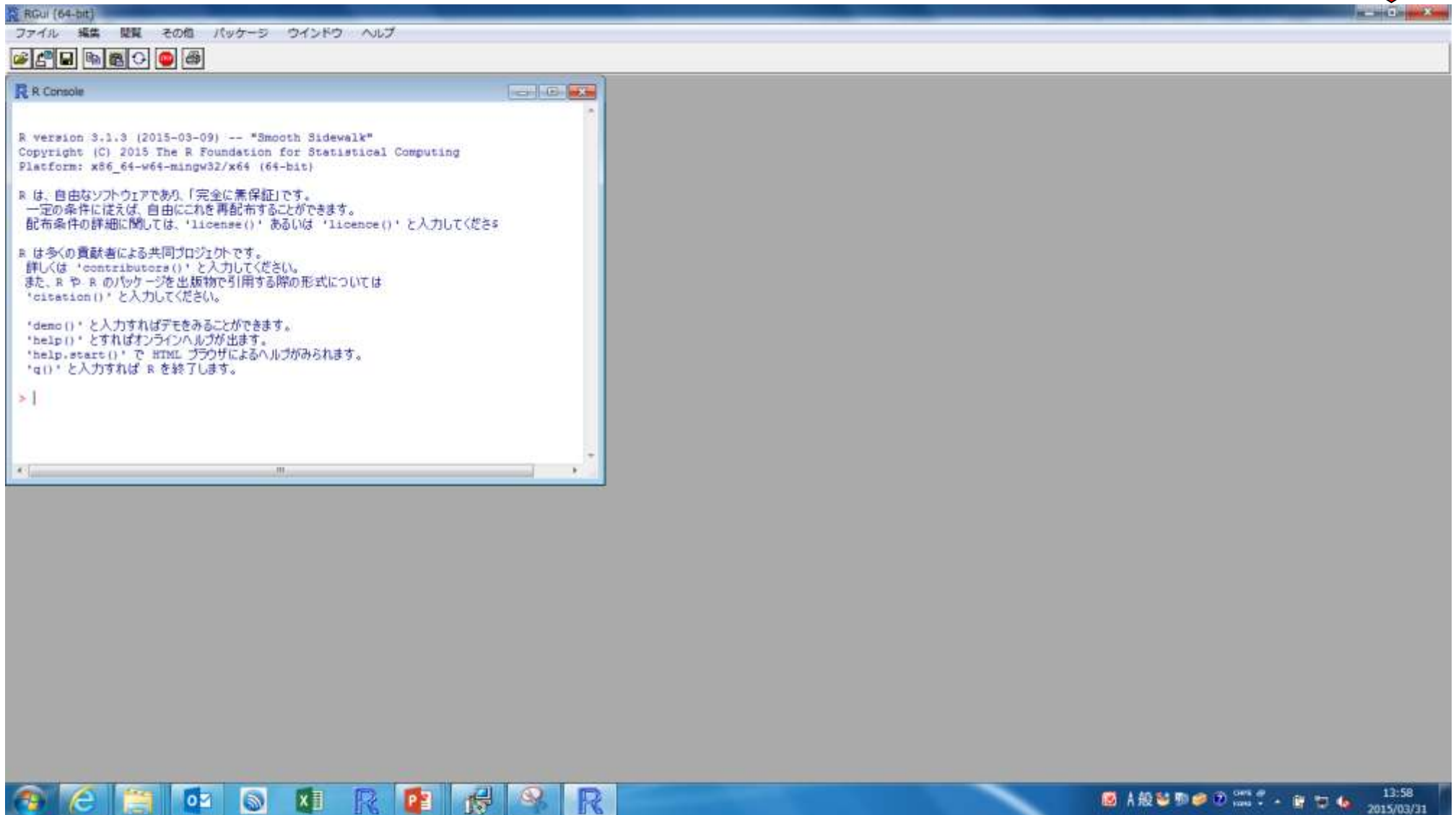
What's new?

- ・ [Galaxy](#)のウェブサイトのリンク先を<http://usegalaxy.org>から<https://galaxyproject.org>に変更しました。(2017/03/17) **NEW**
- ・ [日本乳酸菌学会誌](#)のNGS関連連載の[第9回分原稿PDF](#)を公開しました。(2017/03/13) **NEW**
- ・ 「[書籍](#) | [日本乳酸菌学会誌](#) | [第7回ロングリードアセンブリ](#)」でgrep -Aの説明が間違っていたので修正しました。「検索文字列の行を含む後ろx行を表示」みたいに書いていましたが、検索文字列の行は含まないようですm(_)m(2017/01/26)
- ・ [門田からメール返信をもらえない場合は](#) (last modified 2016/08/23)
- ・ [はじめに](#) (last modified 2015/03/31)
- ・ 参考資料 | [書籍、学会誌](#) (last modified 2017/03/13) **NEW**
- ・ 参考資料 | [講習会、講義、講演資料](#) (last modified 2016/12/07)
- ・ [過去のお知らせ](#) (last modified 2017/04/10) **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)
- ・ インストール | Rパッケージ | [ほぼ全て\(20GB以上?\)](#) (last modified 2015/05/25)
- ・ インストール | Rパッケージ | [必要最小限プラスアルファ\(数GB?\)](#) (last modified 2017/03/13)推奨 **NEW**
- ・ インストール | Rパッケージ | [必要最小限プラスアルファ\(アグリバイオ居室のみ\)](#) (last modified 2015/06/16)
- ・ インストール | Rパッケージ | [必要最小限\(数GB?\)](#) (last modified 2015/05/25)

Rの起動



起動直後は画面いっぱいになるので…



Rの起動



黒点線で囲まれた部分が「Rコンソール画面」

```
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)

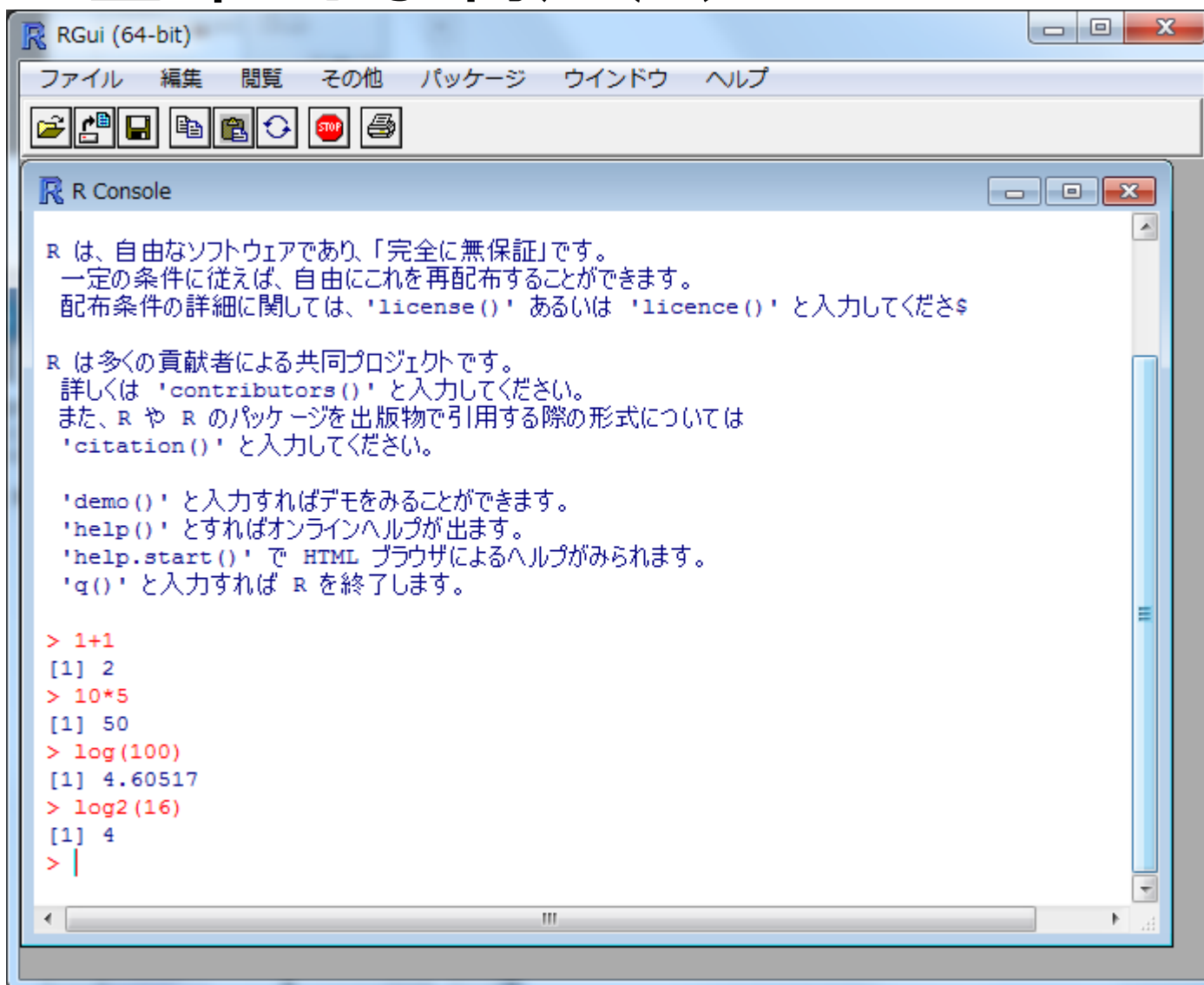
R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()' あるいは 'licence()' と入力してくださ$

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。

> |
```

基本的な利用法



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 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

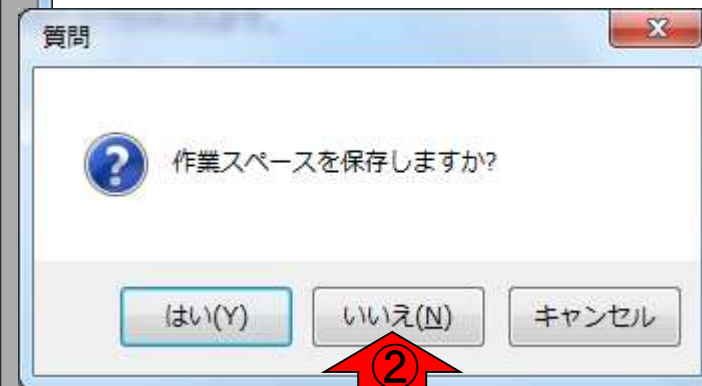
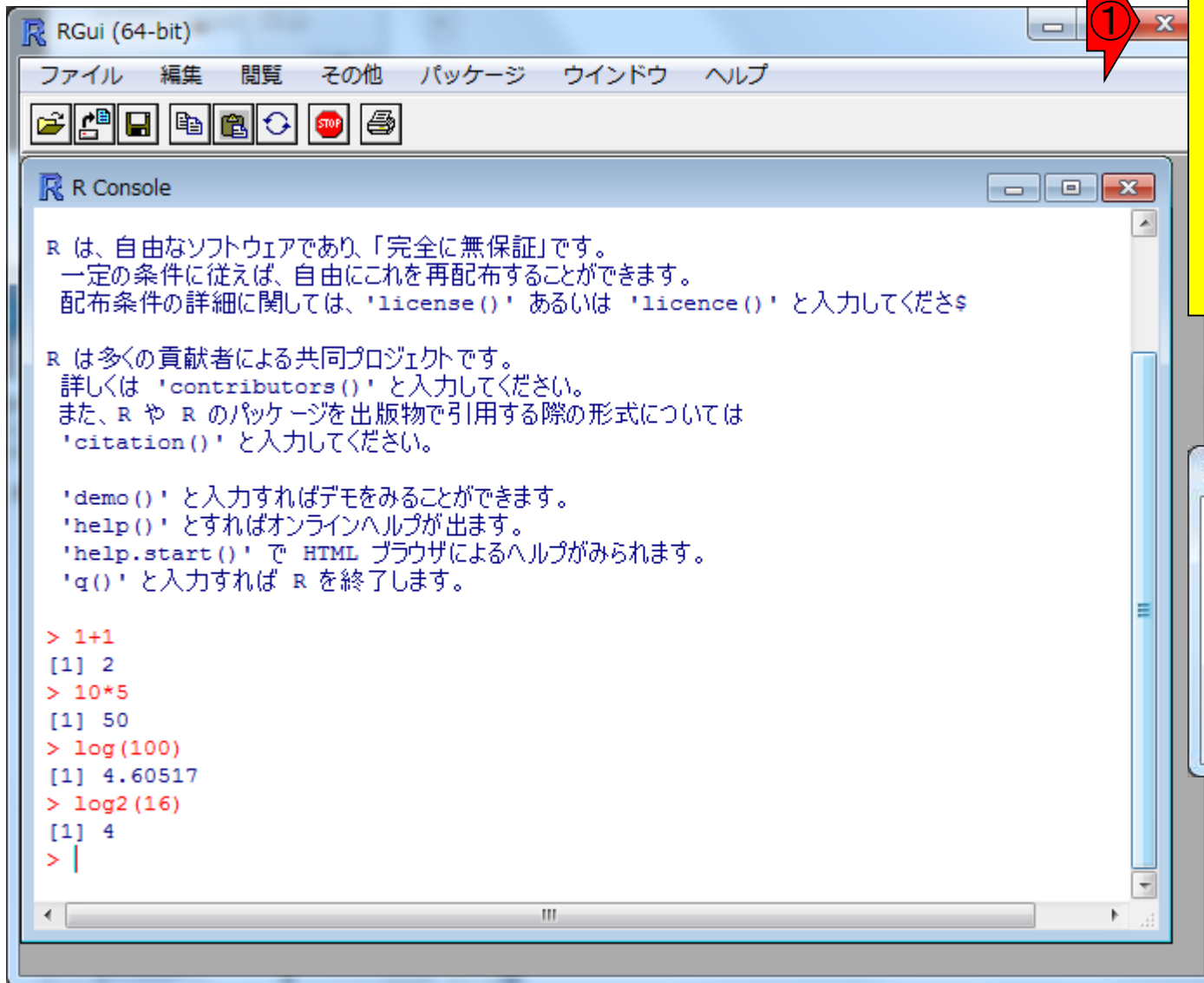
R は多くの貢献者による共同プロジェクトです。
詳しくは `'contributors()'` と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
`'citation()'` と入力してください。

`'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。

```
> 1+1
[1] 2
> 10*5
[1] 50
> log(100)
[1] 4.60517
> log2(16)
[1] 4
> |
```

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ない)



(Rで)塩基配列解析

(Rで)塩基配列解析

(last modified 2017/04/10, since 2010)

このウェブページのR関連部分は、[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)

What's new?

- Galaxyのウェブサイトのリンク先を<http://usegalaxy> **NEW**
- [日本乳酸菌学会誌](#)のNGS関連連載の[第9回分原](#)
- 「書籍 | [日本乳酸菌学会誌](#) | [第7回ロングリードア](#)文字列の行を含む後ろx行を表示」みたいに書いて (2017/01/26)
- [門田からメール返信をもらえない場合は](#) (last modified 2015/03/31)
- [はじめに](#) (last modified 2015/03/31)
- 参考資料 | [書籍](#)、[学会誌](#) (last modified 2017/03/11)
- 参考資料 | [講習会](#)、[講義](#)、[講演資料](#) (last modified 2017/03/11)
- [過去のお知らせ](#) (last modified 2017/04/10) **NEW**
- [インストールについて](#) (last modified 2015/11/12)

- 書籍 | [日本乳酸菌学会誌](#) | [第9回ゲノムアノテーションとその可視化](#)、[DDBJへの登録](#) (last modified 2017/03/13)
- 書籍 | [日本乳酸菌学会誌](#) | [第10回DDBJへの登録\(後編\)](#) (last modified 2017/04/07) **NEW**
- 書籍 | [日本乳酸菌学会誌](#) | [第11回統合データ解析環境Galaxy](#) (last modified 2017/03/24) **NEW**
- イントロ | 一般 | [ランダムに行を抽出](#) (last modified 2014/07/17)
- イントロ | 一般 | [任意の文字列を行の最初に挿入](#) (last modified 2014/07/17)
- イントロ | 一般 | [任意のキーワードを含む行を抽出\(基礎\)](#) (last modified 2016/04/20)
- イントロ | 一般 | [ランダムな塩基配列を生成](#) (last modified 2014/06/16)
- イントロ | 一般 | [任意の長さの可能な全ての塩基配列を作成](#) (last modified 2015/02/19)
- 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 2015/04/06)
- 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](#) **①** (last modified 2015/09/12)
- イントロ | 一般 | [翻訳配列\(translate\)を取得\(応用\)](#) | [seqinr\(Chang, 2005\)](#) (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](#) (last modified 2013/06/14)
- イントロ | 一般 | [逆鎖\(reverse\)を取得](#) (last modified 2013/06/14)
- イントロ | 一般 | [k-mer解析](#) | [k=1\(塩基ごとの出現頻度解析\)](#) | [Biostrings](#) (last modified 2016/04/27)
- イントロ | 一般 | [k-mer解析](#) | [k=2\(2連続塩基の出現頻度解析\)](#) | [Biostrings](#) (last modified 2016/01/28)
- イントロ | 一般 | [k-mer解析](#) | [k=3\(3連続塩基の出現頻度解析\)](#) | [Biostrings](#) (last modified 2016/01/28)
- イントロ | 一般 | [k-mer解析](#) | [k=n\(n連続塩基の出現頻度解析\)](#) | [Biostrings](#) (last modified 2016/05/01)
- (削除予定)イントロ | 一般 | [2連続塩基の出現頻度情報を取得](#) (last modified 2015/04/20)
- (削除予定)イントロ | 一般 | [3連続塩基の出現頻度情報を取得](#) (last modified 2015/02/19)
- (削除予定)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 2015/02/19)
- イントロ | 一般 | Tips | [任意の拡張子でファイルを保存](#) (last modified 2013/09/26)
- イントロ | 一般 | Tips | [拡張子は同じで任意の文字を追加して保存](#) (last modified 2013/09/26)
- イントロ | 一般 | [配列取得](#) | [ゲノム配列](#) | [公共DBから](#) (last modified 2014/05/28)

①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 2015/04/06)
- ・ 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](#) **①** (last modified 2015/09/12)
- ・ イントロ | 一般 | [翻訳配列\(translate\)を取得\(応用\)](#) | [seqinr\(Charaffet 2005\)](#) (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](#) (last modified 2013/06/14)
- ・ イントロ | 一般 | [逆鎖\(reverse\)を取得](#) (last modified 2013/06/14)
- ・ イントロ | 一般 | [k-mer解析](#) | [k=1\(塩基ごとの出現頻度\)](#)
- ・ イントロ | 一般 | [k-mer解析](#) | [k=2\(2連続塩基の出現頻度\)](#)
- ・ イントロ | 一般 | [k-mer解析](#) | [k=3\(3連続塩基の出現頻度\)](#)
- ・ イントロ | 一般 | [k-mer解析](#) | [k=n\(n連続塩基の出現頻度\)](#)
- ・ (削除予定)イントロ | 一般 | [2連続塩基の出現頻度情報](#)
- ・ (削除予定)イントロ | 一般 | [3連続塩基の出現頻度情報](#)
- ・ (削除予定)イントロ | 一般 | [任意の長さの連続塩基の出現頻度情報](#)
- ・ 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追加](#)

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

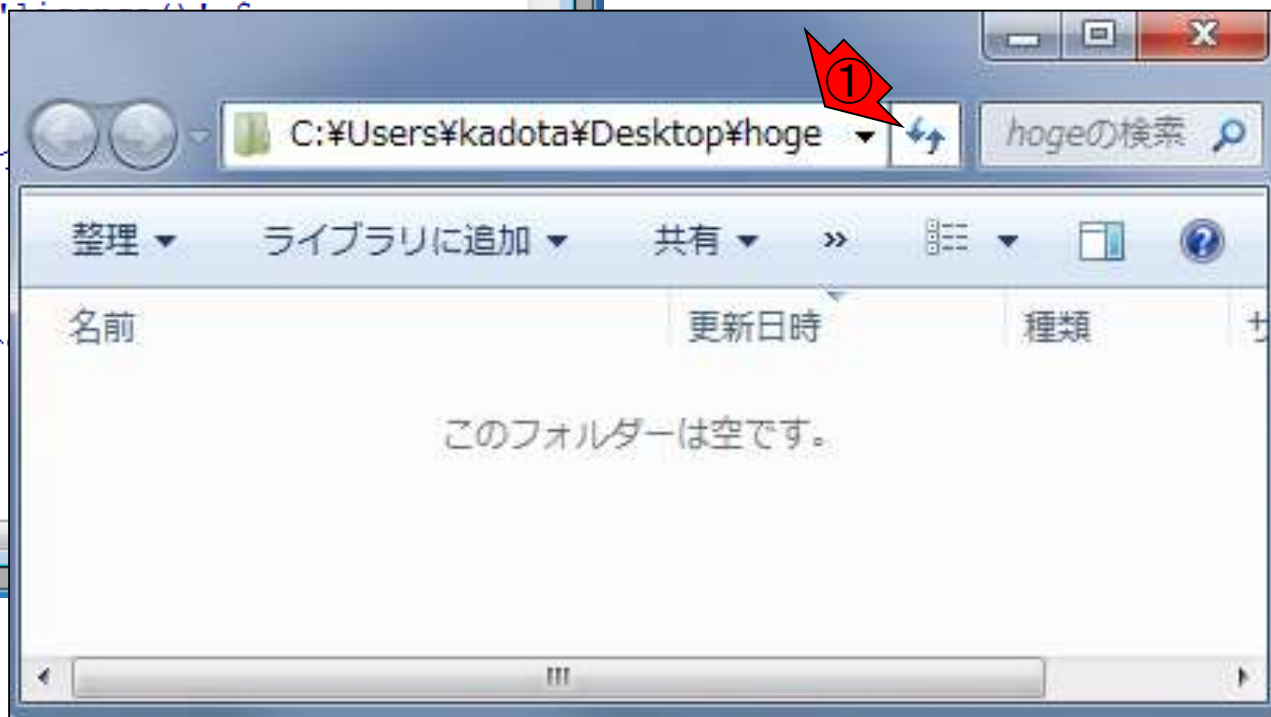
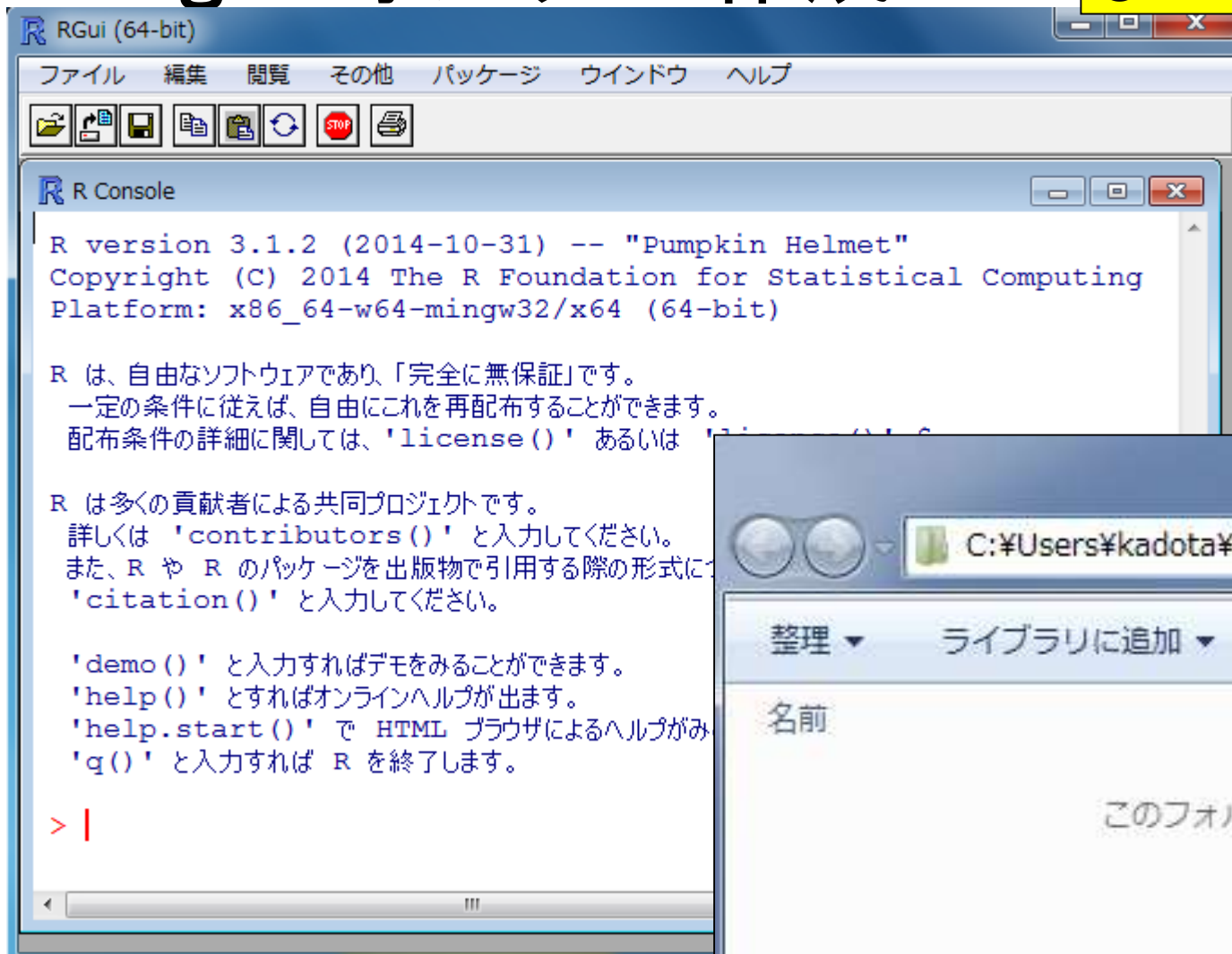
#本番
fasta <- translate(fasta)    #アミノ酸配列に翻訳した結果をfastaに格納
fasta                                #確認してるだけです

#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。
①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:

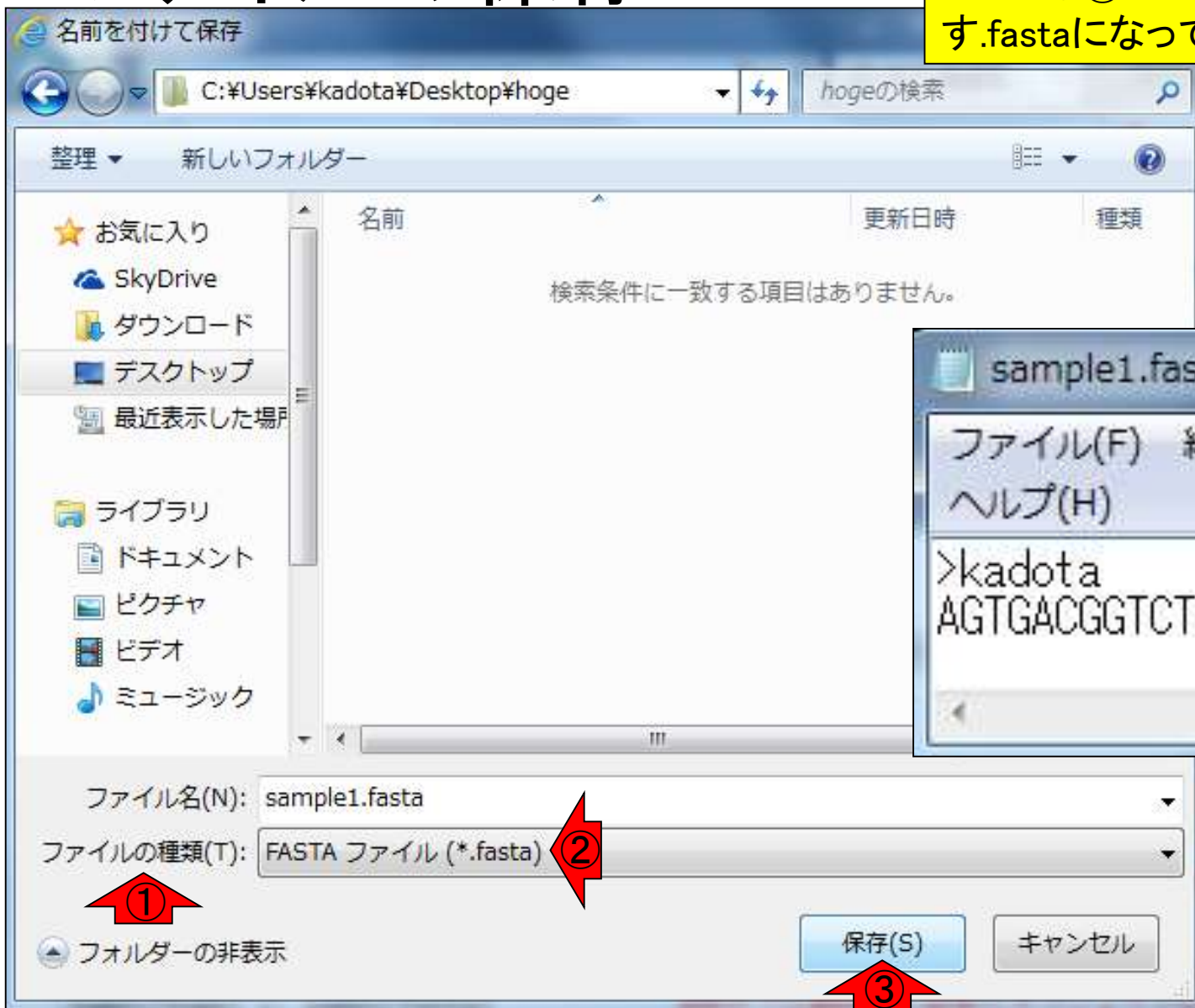
```
multi-FASTAではないsingle FASTAの場合  
in_f <- "sample1.fasta"  
out_f <- "hoge1.fasta"  
  
#必要なパッケージをロード  
library(Biostrings)  
  
#入力ファイルの読み込み  
fasta <- readDNASTringSet(  
  fasta  
  
#本番  
fasta <- translate(fasta)  
fasta  
  
#ファイルに保存  
writeXStringSet(fasta, file)
```

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(2)
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)

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(3)
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 フォルダー
kadota システム フォルダー
コンピュータ システム フォルダー
ネットワーク システム フォルダー
hoge ファイル フォルダー(4)
ファイル名(N): sample1.fasta
ファイルの種類(T): FASTA ファイル (*.fasta)
フォルダーの非表示
保存(S)(5) キャンセル

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。
```

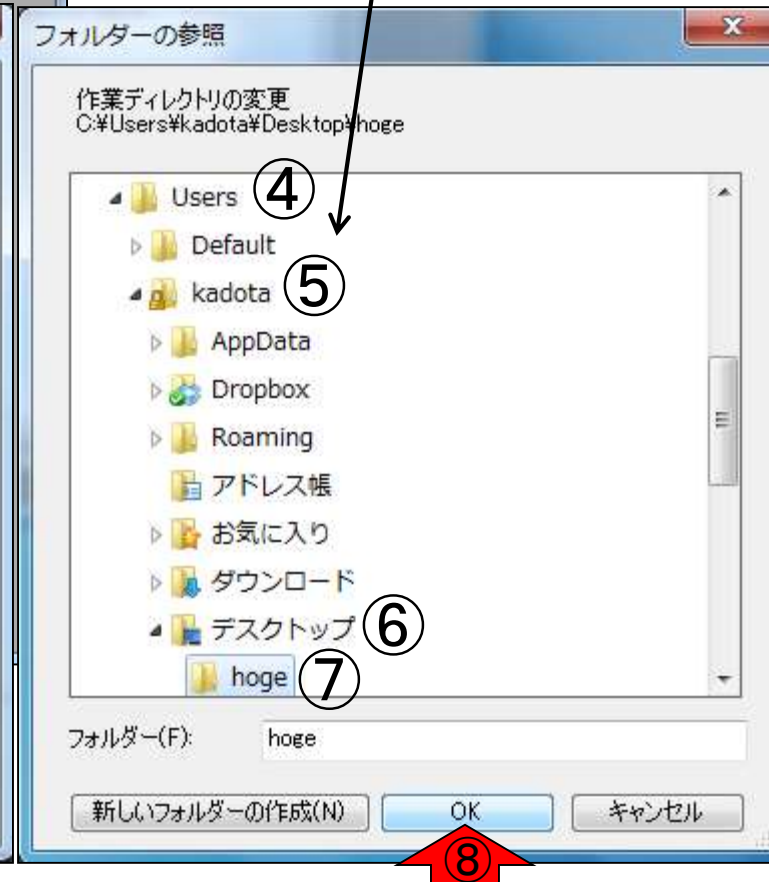
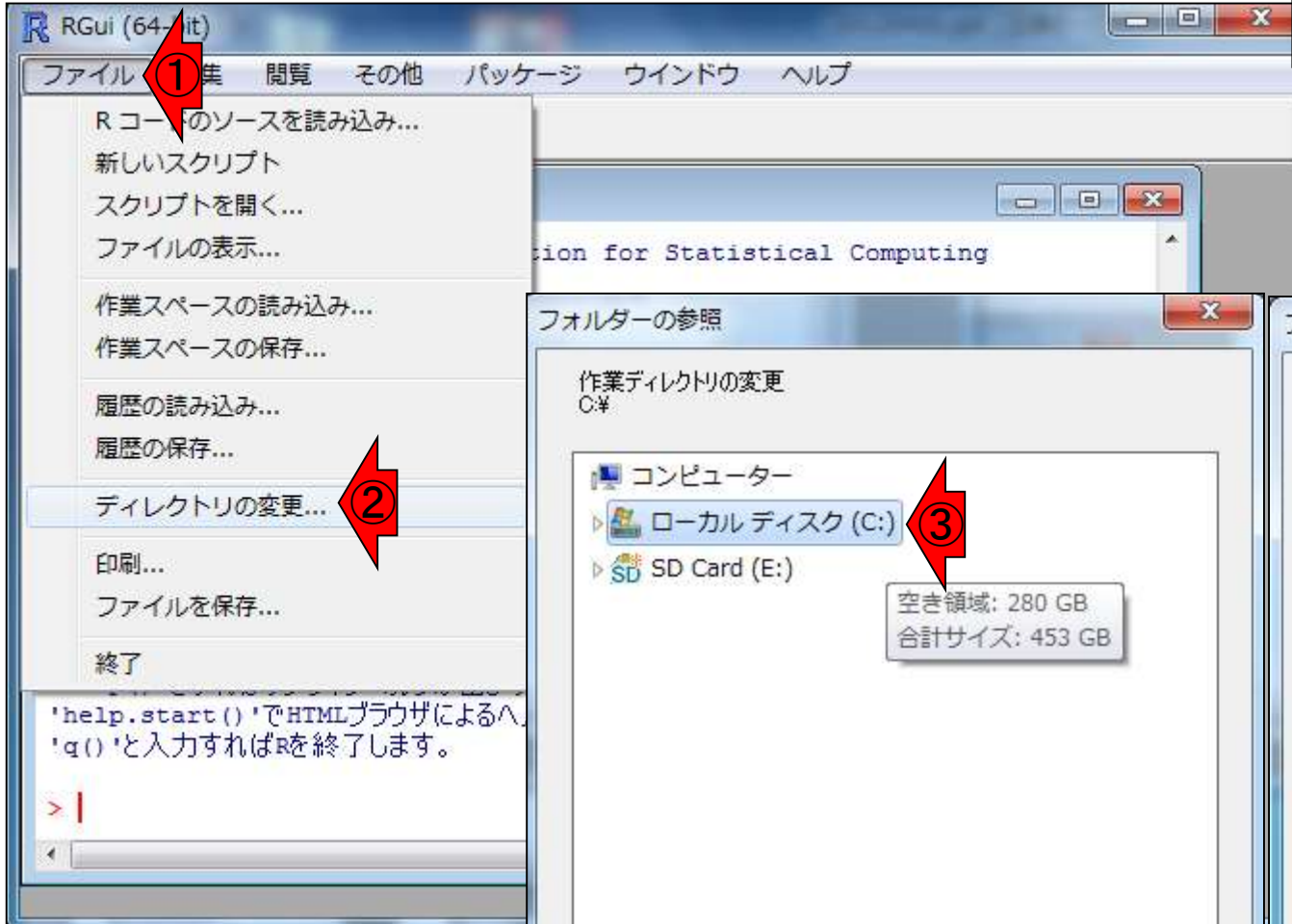
```
'demo()' と入力すればデモをみることができます。  
'help()' とすればオンラインヘルプが出ます。  
'help.start()' で HTML ブラウザによるヘルプがみられ$  
'q()' と入力すれば R を終了します。
```

```
> 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

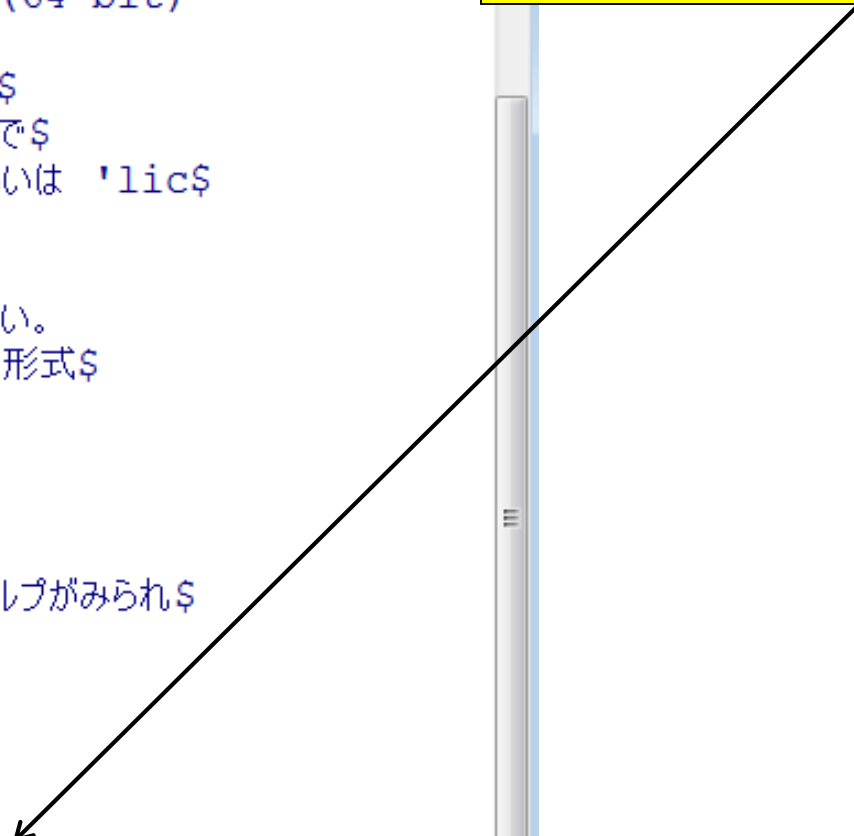
```
R Console
Platform: x86_64-w64-mingw32/x64 (64-bit)

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。

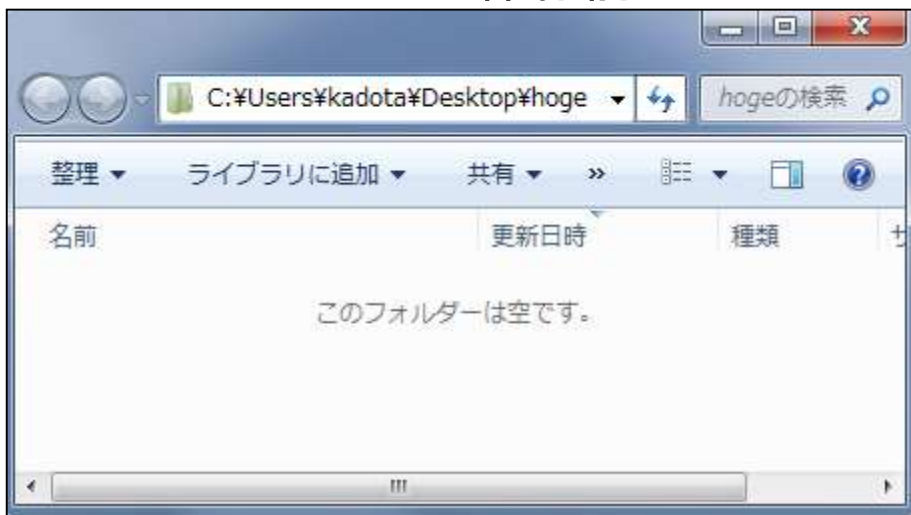
'demo()'と入力すればデモをみることができます。
'help()'とすればオンラインヘルプが出ます。
'help.start()'で HTML ブラウザによるヘルプがみられ$
'q()'と入力すれば R を終了します。

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

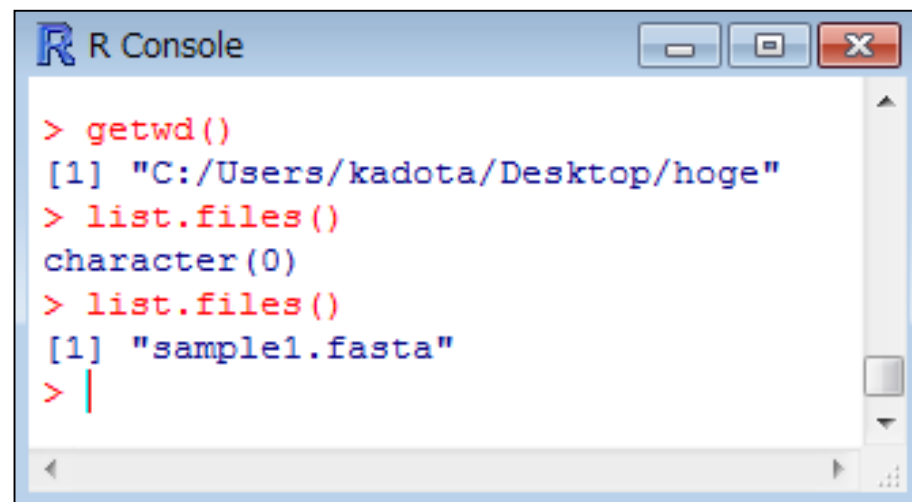
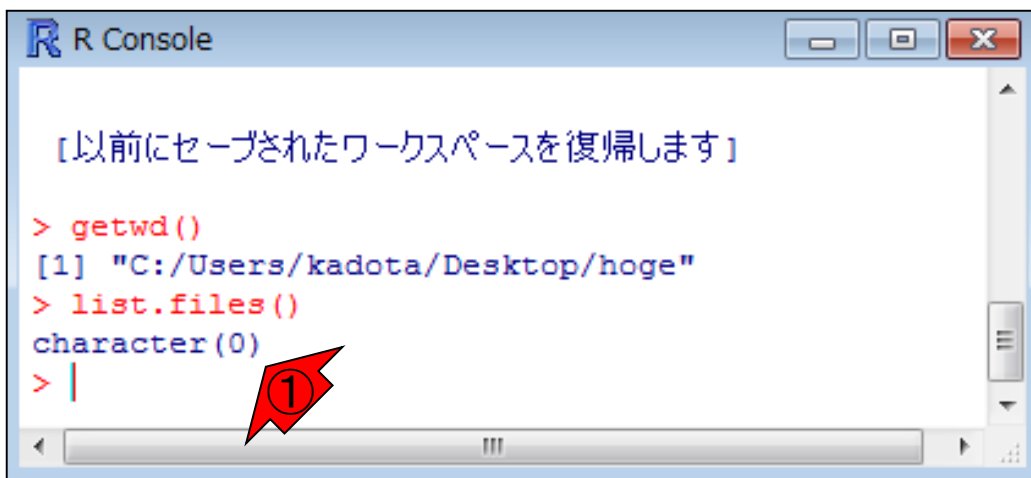
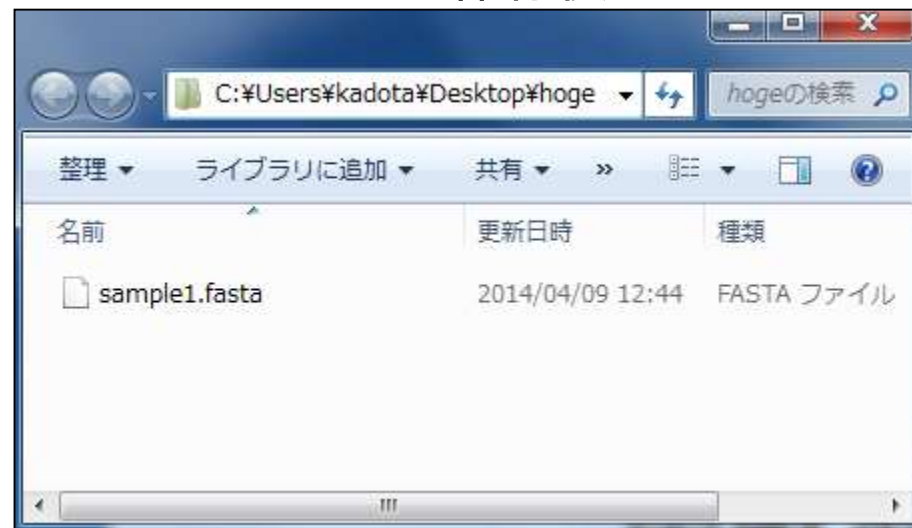


実際のhogeフォルダとR操作画面の関係

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

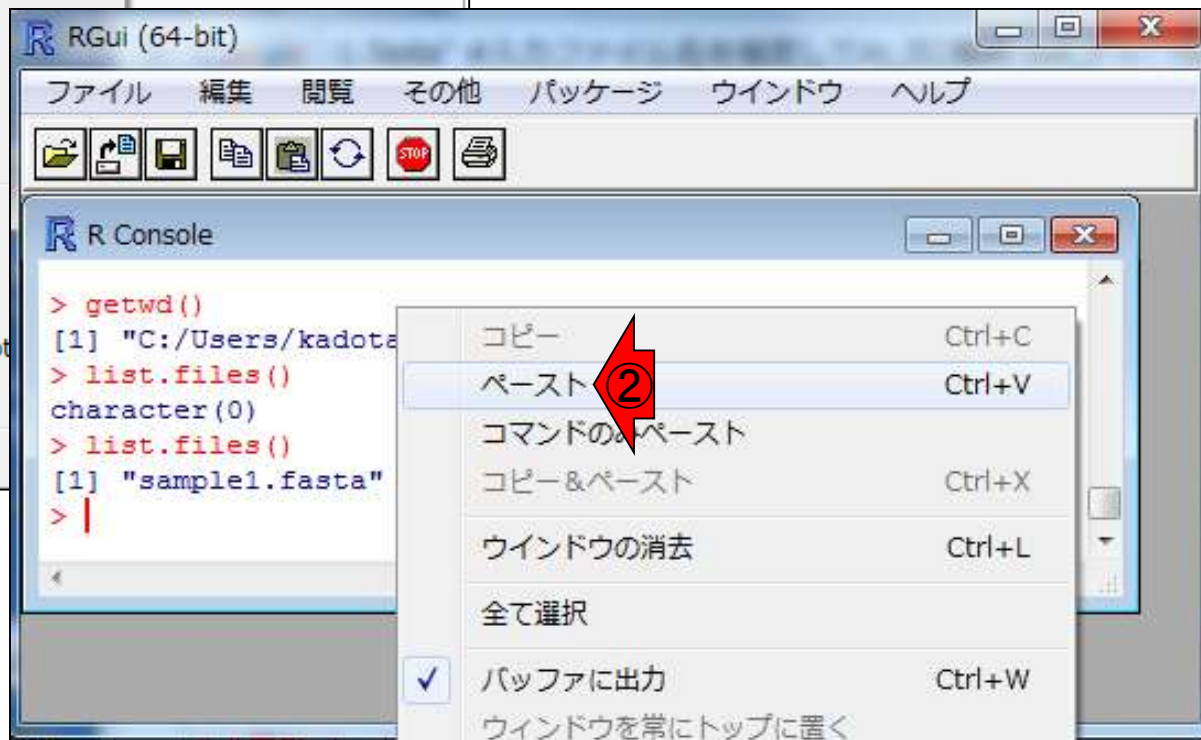
Google で検索

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

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
fasta                                #確認してるだけ

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

#ファイルに保存
writeXStringSet(fasta, file=out_f, format="fasta", w

```

エラーなく実行できた場合の全貌

```

> 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$
> 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$
> |

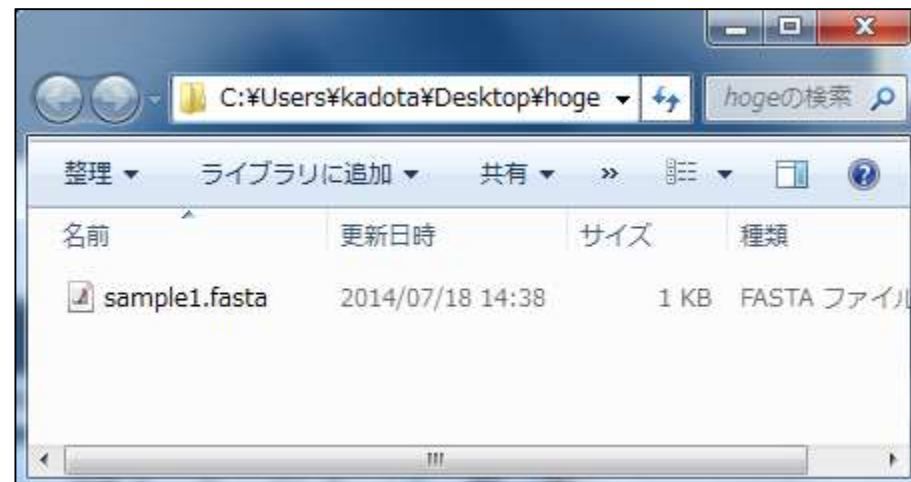
```

実行結果

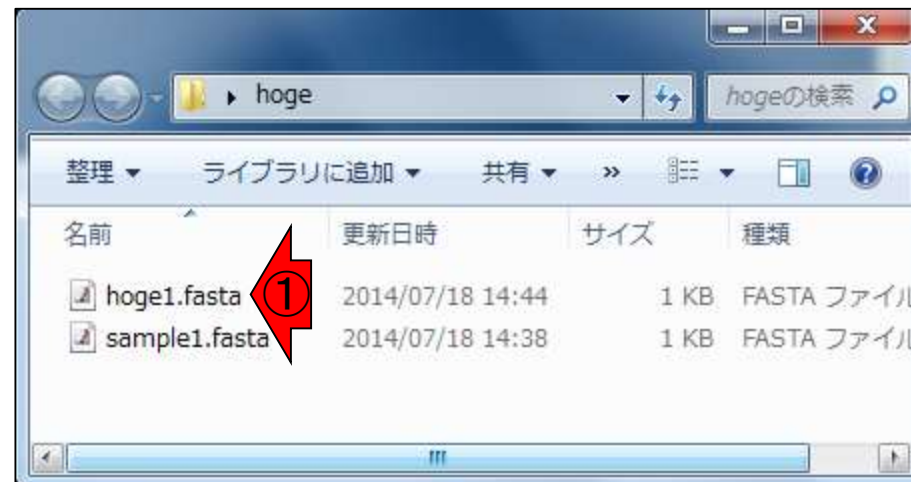
①出力ファイル名として指定したhoge1.fastaが生成されていることが分かります

```
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。

```

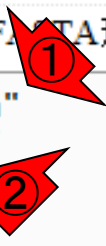
in_f <- "sample1.fasta"      #入力ファイル名を指定してin_fに格納
out_f <- "hoge1.fasta"      #出力ファイル名を指定してout_fに格納

#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

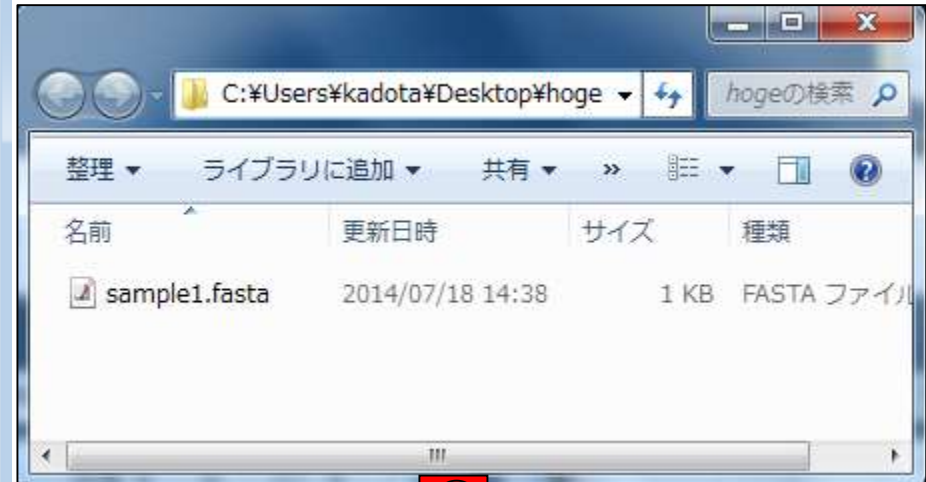
#ファイルに保存
writeXStringSet(fasta, file=out_f, format="fasta", width=50)#fastaの中身を指定したフ
    
```



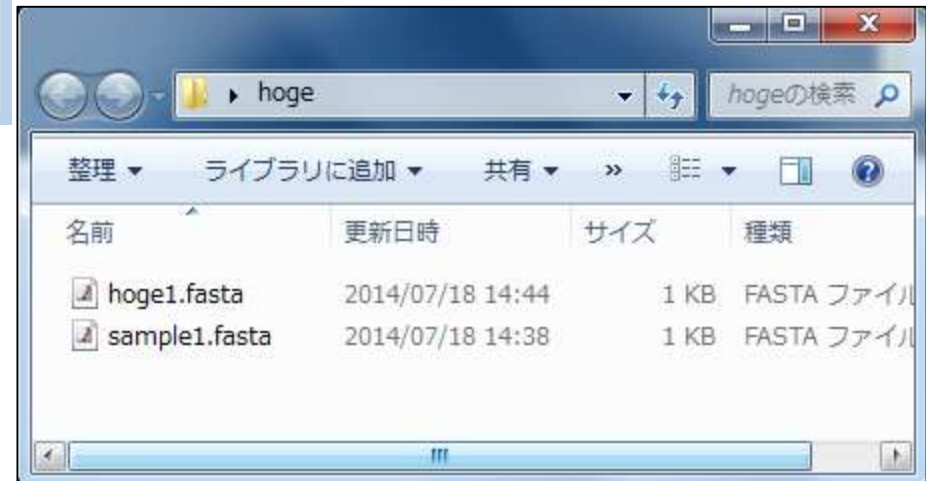
実行結果

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) ^[3]	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の手順に従ってインストールを行えば、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

(last modified 2017/04/10, since 2010)



このウェブページのR関連部分は、[インストール](#) | [R](#) | [R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)

What's new?

- Galaxyのウェブサイトのリンク先を<http://usegalaxy.org>から<https://galaxyproject.org>に変更しました。(2017/03/17) **NEW**
- [日本乳酸菌学会誌](#)のNGS関連連載の[第9回分原稿PDF](#)を公開しました。(2017/03/13) **NEW**
- 「書籍 | [日本乳酸菌学会誌](#) | [第7回ロングリードアセンブリ](#)」でgrep -Aの説明が間違っていたので修正しました。「検索文字列の行を含む後ろx行を表示」みたいに書いていましたが、検索文字列の行は含まないようですm(_ _)m (2017/01/26)

- [門田からメール返信をもらえない場合は](#) (last modified 2016/08/23)
- [はじめに](#) (last modified 2015/03/31)
- 参考資料 | [書籍、学会誌](#) (last modified 2017/03/13) **NEW**
- 参考資料 | [講習会、講義、講演資料](#) (last modified 2016/12/07)
- [過去のお知らせ](#) (last modified 2017/04/10) **NEW**
- [インストール](#) | [R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)
- インストール | Rパッケージ | [ほぼ全て\(20GB以上?\)](#) (last modified 2015/05/25)
- インストール | Rパッケージ | [必要最小限プラスアルファ\(数GB?\)](#) (last modified 2017/03/13) 推奨 **NEW**
- インストール | Rパッケージ | [必要最小限プラスアルファ\(アグリバイオ居室のみ\)](#) (last modified 2015/06/16)

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

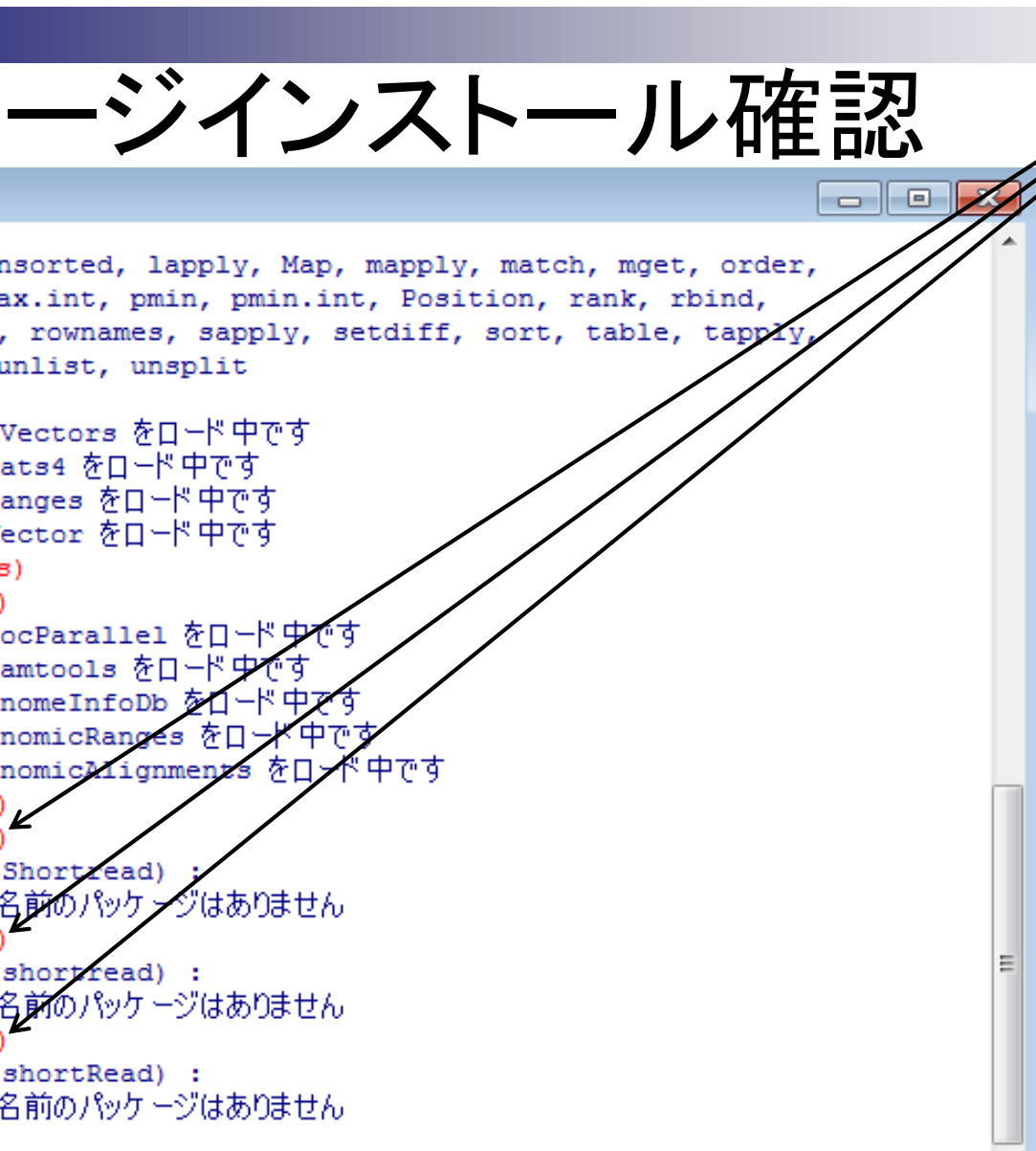
The following objects are masked from 'package:base':

  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  
> |
```



①Biostringsというパッケージは、②この部分でロードして使っていますが...同じ結果が得られる別のやり方も存在します

おまけ1: Biostrings

- ・ [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 2015/04/06)
- ・ [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](#) ① (last modified 2015/09/12)
- ・ [イントロ](#) | [一般](#) | [翻訳配列\(translate\)を取得\(応用\)](#) | [seqinr\(Charaff 2005\)](#) (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 | 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
#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
#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

```

- ・ [イントロ](#) | [一般](#) | [逆相補鎖\(reverse complement\)](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [逆鎖\(reverse\)を取得](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [k-mer解析 | k=1\(塩基ごとの出現\)](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [k-mer解析 | k=2\(2連続塩基の出現\)](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [k-mer解析 | k=3\(3連続塩基の出現\)](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [k-mer解析 | k=n\(n連続塩基の出現\)](#) (last modified 2013/06/14)
- ・ (削除予定) [イントロ](#) | [一般](#) | [2連続塩基の出現](#) (last modified 2013/06/14)
- ・ (削除予定) [イントロ](#) | [一般](#) | [3連続塩基の出現](#) (last modified 2013/06/14)
- ・ (削除予定) [イントロ](#) | [一般](#) | [任意の長さの連続塩基の出現](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [Tips](#) | [任意の拡張子でファイル名を指定](#) (last modified 2013/06/14)
- ・ [イントロ](#) | [一般](#) | [Tips](#) | [拡張子は同じで任意の位置](#) (last modified 2013/06/14)

これです。①seqinrというパッケージを、
②この部分でロードして使っています

おまけ1: seqinr

- ・ 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 2015/04/06)
- ・ 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 2015/09/12)
- ・ イントロ | 一般 | 翻訳配列(translate)を取得(応用) | [seqinr\(Charif 2005\)](#) ① (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 | 翻訳配列(translate)を取得(応用) | [seqinr\(Charif_2005\)](#)

[seqinr](#)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。本気で翻訳配列を取得する場合にはこちらの利用をお勧めします。翻訳できないコドンにはアミノ酸X(不明なアミノ酸)に変換してくれたり、`translate`関数のオプションとして`ambiguous=T`とすると、翻訳できるものは可能な限り翻訳してくれます(高橋 広夫 氏提供情報)。`lapply`関数を用いるやり方(高橋 広夫 氏提供情報)と`sapply`関数を用いるやり方(甲斐 政親 氏提供情報)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"      #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(seqinr)            #パッケージの読み込み

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

#本番
hoge <- lapply(hoge, function(x){ #アミノ酸配列に翻訳
  translate(x, ambiguous=T)      #アミノ酸配列に翻訳
})                                #アミノ酸配列に翻訳
hoge                                #確認してるだけです

#ファイルに保存
write.fasta(hoge, names=names(hoge), file.out=out_f, nbchar=50) #hogeの中身を指定したフ
    
```

おまけ2

①平成27年度NGSハンズオン講習会では、②Rのパッケージについてや、バージョンの違いに起因する問題など、より詳細な情報が7/29および7/30分の講義資料として提供されています

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

・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。このカリキュラムに基づいて、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～
 - 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

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～08月06日(A日程)および08月26～09月01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
 - 中～上級レベルの内容を新規提供。講師数6人
 - 講習会映像は統合TVとYoutubeから公開
 - 報告書(もうすぐ公開2017.04.10現在)

- ・ **NGSハンズオン講習会**(平成29年度)
 - 東大農で2017年08月下旬に実施予定

②

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(東工大) 寺田 透(東工大)	事前学習資料一巻(PDF:52KB) 講義資料一巻(PDF:108KB)
	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の基礎的理解	門田 幸二(東工大)	講義資料一巻(PDF:32KB) 統合TV
	13:15-14:45			中級			
	15:00-16:30						
	16:45-18:15						
7月24日(金)	10:30-12:00		スクリプト言語	中級	シェルスクリプト	服部 恵美(アメリエフ)	講義資料(PDF:1.8MB) 統合TV
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
7月27日(月)	10:30-12:00			中級	Perl		講義資料(PDF:1.9MB) 統合TV
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
7月28日(火)	10:30-12:00			中級	Python		講義資料一巻(PDF:52KB) 統合TV
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
7月29日(水)	10:30-12:00	データ解析環境R	R基礎 1	初級	R言語の基礎(インストールから利用まで)	門田 幸二(東工大)	講義資料一巻(PDF:37KB) 統合TV
	13:15-14:45		R基礎 2	初級	・ファイルの読み込み ・行列演算の基本		
	15:00-16:30		R各種パッケージ	中級	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		
	16:45-18:15						
7月30日(木)	10:30-12:00		Bioconductorの利用法 1	中級	データの型やバージョンの違い		講義資料一巻(PDF:53KB) 統合TV
	13:15-14:45						
	15:00-16:30		Bioconductorの利用法 2	中級	FASTA/FASTQファイルの各種解析		
	16:45-18:15						