

UST 2021-2학기 현장연구

Contents

1 강의목표	2
2 운영방식	2
3 일정	2
4 수업 (20210902 목)	3
4.1 주제선정	3
4.2 구현 방법	3
4.3 Rstudio 접속방법	3
5 수업 (20210906 월)	4
6 수업 (20210907 화)	4
7 수업 (20210913)	4
7.1 사전 공지	4
7.2 Rstudio 접속방법	4
8 수업 (20210920)	4
9 수업 (20210927)	4
10 수업 (20211004)	5
11 수업 (20211018)	5
12 수업 (20211025)	5
13 수업 (20211101)	5
13.1 Nanopore sequencing pipeline	5
13.2 DNA PCA 실험	5
DNA 준비 회석	5
PCA	5
PCR	6
Gel QC	6
14 수업 (20211108)	6
15 수업 (20211111)	6
15.1 Rstudio 접속	6
15.2 프로젝트	7
15.3 Rmarkdown 활용	7
15.4 Vector	7
15.5 Data frame	7
15.6 TODO	8

16 수업 (20211115)	8
17 수업 (20211118)	8
18 수업 (20211122)	9
19 수업 (20211125)	9
19.1 성원재	9
20 수업 (20211129)	9
21 수업 (20211202)	9
22 수업 (20211206)	9
23 수업 (20211213)	9
23.1 성원재	9
23.2 허성보	9
23.3 고민지	9
23.4 박성균	9
23.5 Aporva	9
23.6 박아영	10

<https://greendaygh.github.io/fieldresearch2021-2/>

1 강의목표

본 현장연구에서는 합성생물학 연구에 필요한 컴퓨터 관련 지식의 학습과 실습을 통해서 관련 데이터의 수집과 변환, 가시화 그리고 분석 능력을 습득하는 것에 목적이 있음. 특히 개인별 연구 주제에 따라서 필요한 컴퓨팅 기술들을 정리하고 각 기술들에 대해서 짧은 테크니컬 노트 수준의 보고서들을 작성하는데 목적이 있음.

2 운영방식

- 각 수강생들의 학위과정 동안의 주요 연구 주제와 연계한 수업임을 고려한 각 수강생별 연구주제 토의 및 선정
- 선정된 연구주제 중 컴퓨터 활용 관련한 기술을 정리하고 개인 수준별 타깃 기술 선정 및 목표를 수립함
- 선정된 타깃 기술에 대해서 수강생들의 실습 발표 위주의 진행하며 다른 수강생 및 지도교수의 코드 지도
- 필요한 경우 온라인 수업 툴을 활용함
- 연구 데이터 분석에 직접적으로 필요한 기술 위주로 진행하며 이를 위해 필요한 기반 기술 (리눅스 활용 등)은 필요한 경우 설명
- Docker 이미지 활용 (개인컴퓨터 활용 가능)
- 출석50%, 과제50%
- 각 수강생별로 본인의 연구와 관련된 간단한 도표 및 정리 문서 등이 수업의 목표에 따라서 과제로 나갈 수 있음
- 테크니컬 리포트는 길이에 따라서 1건 이상이 될 수도 있음
- 가능한 명확하고 짧은 길이의 리포트가 선호됨
- 과제는 Rmarkdown을 활용한 pdf 형식의 문서로 제출함을 원칙으로 함

3 일정

- 1주 - 강의 개요 및 연구 주제 토의
- 2주 - 각 수강생들의 연구 주제 발표 및 컴퓨팅 기술 도출

- 3주 - 연구 주제에 따른 기술 워크플로우 작성 및 토의
- 4주 - 각 기술별 필요한 세부 컴퓨팅 기술 도출
- 5주 - 수강생별 담당 컴퓨팅 기술 선정
- 6주 - 기술 학습/실습/운영 위한 기반 docker 설치 및 학습
- 7주 - 수강생별 기술 학습 및 실습 1~4
- 11주 - 테크니컬 노트 중간 정리
- 12~15주 - 수강생별 기술 학습 및 실습 5~8
- 16주 - 테크니컬 노트 최종 정리

4 수업 (20210902 목)

4.1 주제선정

각자 한 학기동안 연구를 통해서 만들어볼 기술문서에 대한 주제 선정

- 허성보: Linux + docker (Seongbo)
 - Objective: To build Rstudio env
- 박아영: R language (A-young)
 - Objective: BT Codon design
- 성원재: Tensorflow, Keras (Wonjae) R based data analysis
 - Objective: To build tensorflow models (data)
- 아포바: Shiny, R package (Aporva)
 - Objective: To build R (shiny) interface for dsembler-script (Database)
- 박성군: PDB, NCBI DB API (SeongKun) Shiny
 - Objective: Download and preprocessing (FACS screening)
- 고민지: Rstudio (Minji)
 - Objective: Load data and plot graphs

4.2 구현 방법

위 목표 구현을 위한 필요 사항

- We'll use Cloud based Rstudio
- Once create an account, it'll ready to use Rstudio
- Rmarkdown for report (Rmarkdown study)
- Homework ? Create a Rmarkdown file after login Rstudio

4.3 Rstudio 접속방법

1. 터미널 접속을 통한 계정 생성 (강사나 Aporva 등 도움)

- KRIBB ID와 동일 ID 계정 생성 (카톡 발신 예정)
- putty 다운로드
- 설치/실행 후 접속 (Session -> Host name에 IP, Port에 8022 입력)
- 주어진 계정으로 터미널 로그인
- 로그인 후 패스워드 변경 (로그인 하면 다음 그림처럼 자동 변경 요청)
- 패스워드 변경 후 카톡이나 메일로 변경 알림

```
WARNING: Your password has expired.  
You must change your password now and login again!  
Changing password for myuser2.  
(current) UNIX password: █
```

2. 브라우저로 Rstudio 접속

- 브라우저 오픈
- 주소창에 공지된 IP 입력 (카톡 발신 예정)
- 설정된 계정으로 Rstudio 로그인

5 수업 (20210906 월)

6 수업 (20210907 화)

7 수업 (20210913)

7.1 사전 공지

월요일 10시 수업 관련해서 간단히 지금 하는 일들 ppt 2~3장으로 (5분 발표) 준비해 주시고 지난번 계획한 일들 진행을 어떻게 할지 다시 한 번 생각해 보세요

그리고 클라우드 계정은 보안 이슈로 단분간 사용하기 어려울 것 같습니다. 대신 우리 192.168.122.158 서버를 Linux/Rstudio 사용합니다. 연구원 내에서만 접속이 가능하며 월요일에 계정 등 사용법을 설명 드리겠습니다.

7.2 Rstudio 접속방법

1. 터미널 접속을 통한 계정 생성

- KRIBB ID와 동일 ID 계정 생성
- putty 다운로드
- 설치/실행 후 접속 (Session -> Host name에 192.168.122.158, Port에 8022 입력)
- 주어진 계정으로 터미널 로그인
- 로그인 후 패스워드 변경 (로그인 하면 다음 그림처럼 자동 변경 요청)
- 패스워드 변경 후 카톡이나 메일로 변경 알림

```
WARNING: Your password has expired.  
You must change your password now and login again!  
Changing password for myuser2.  
(current) UNIX password: █
```

2. 브라우저로 Rstudio 접속

- 브라우저 오픈
- http://192.168.122.158:8788 입력
- 앞서 설정된 계정으로 Rstudio 로그인

8 수업 (20210920)

연구 발표

9 수업 (20210927)

연구 발표

10 수업 (20211004)

연구 발표

11 수업 (20211018)

연구 발표

12 수업 (20211025)

연구 발표

13 수업 (20211101)

13.1 Nanopore sequencing pipeline

박성균 연구원 NGS pipeline 발표

13.2 DNA PCA 실험

2개 샘플 수행 기준 PCA 프로토콜

DNA 준비 회석

재료: 아이스, KOD-one, 샘플 (2개 기준), 1.5ml tube 7개, 3DW

- DNA polymerase KOD-one 꺼내 아이스에 넣어둠
- 50pmol powder 상태 DNA 샘플을 vortexing 후 spindown 2초
- 샘플에 3DW 10 μ l 를 넣고 vortexing 후 spindown 2초
- 1/10, 1/50, 1/100 농도 희석용 1.5ml tube 6개 준비 (2개 샘플)
- 원액 1 μ l를 각각 1/10, 1/50 에 넣고 vortexing, spindown
- 1/10 샘플 1 μ l를 1/100에 넣고 vortexing, spindown

PCA

재료: PCR tube 8개, 3DW

- 1.5ml tube에 KOD-one 100 μ l, 3DW 90 μ l 넣고 vortexing
- PCR tube 8개 준비
- 각 tube에 위 KOD 희석액 19 μ l 분주
- 각 tube에 원액, 1/10, 1/50, 1/100 샘플 1 μ l를 분주 (20 μ l 채움)
- vortexing 후 Thermal cycler 수행 (AY, PCA 프로토콜)
 - 94도 10초
 - Denaturation: 94도 10초
 - Annealing: 55도 20초 (lamprate -0.1/도)
 - Extension: 72도 20초 (kb당 5~10초, 1.5kb 기준)
 - Cycle: 20 cycle
 - Final extension: 72도 3분

PCR

재료: PCR용 primer 2set, PCR tube 8, 3DW

- Primer dry 샘플에 (20mer 기준) 300 μ l 넣고 vortexing spindown
- 1.5ml tube에 3DW 228 μ l + 두 개의 primer 각각 1 μ l를 넣음
- PCR tube 8개 준비
- 각 tube에 위 혼합액 23 μ l 분주
- 각 tube에 KOD 혼합액 25 분주
- PCA product 2 μ l 분주
 - 98도 10초
 - Denaturation: 98도 10초
 - Annealing: 55도 5초
 - Extension: 72도 20초 (kb당 5~10초, 1.5kb 기준)
 - Cycle: 30 cycle
 - Final extension: 68도 1분
 - 4도 무한
- 완료 시 PCR, PCA 각 8개 PCR tube

Gel QC

재료: 1.5ml tube, Bluejuice 10x, 3DW, PCR tube 8개, 아가로즈 젤

- 젤 준비 (17 칸)
- 1.5 ml tube에 Bluejuice 10x 10 μ l와 3DW 70 μ l를 섞음
- PCR tube 8개에 각각 8 μ l 분주
- PCR tube 8개에 앞서 PCA product 2 μ l 분주
- 위 PCR Product 8개 튜브에 Bluejuice 5 μ l 분주
- 가운데 lane Ladder 로딩
- Ladder 기준으로 왼쪽 한 샘플, 오른쪽 다른 샘플 각각 PCA, PCR 순서로 로딩
- 130v 20분 전기영동
- 사진 확인

14 수업 (20211108)

DNA PCA 실험 2nd

15 수업 (20211111)

온라인수업: <https://www.remotemeeting.com/reservation/share/2c908ad67cea7782017d0902b318133f>

15.1 Rstudio 접속

1. 터미널 접속을 통한 계정(비번) 생성
 - KRIBB ID와 동일 ID 계정 생성
 - putty 다운로드
 - 설치/실행 후 접속 (Session -> Host name에 192.168.122.158, Port에 8022 입력)
 - 주어진 계정으로 터미널 로그인
 - 비번은 모두 동일 (research2)
 - 로그인 후 패스워드 변경 (로그인 하면 다음 그림처럼 자동 변경 요청)

```
WARNING: Your password has expired.  
You must change your password now and login again!  
Changing password for myuser2.  
(current) UNIX password: █
```

- 로그아웃 (창닫기, Ctrl+D)

2. 브라우저로 Rstudio 접속

- 브라우저 오픈
- `http://192.168.122.158:8788` 입력
- 앞서 설정된 계정으로 Rstudio 로그인

3. 화면 구성

15.2 프로젝트

File > New project

15.3 Rmarkdown 활용

Rmarkdown은 Rstudio의 강의를 링크 참고하며 Markdown은 Pandoc의 문법을 링크 참고.

Rstudio를 사용한 Rmarkdown 만드는 법을 익힘. Rstudio는 R언어 외에도 다양한 언어를 이용한 프로그래밍을 지원하며 Rstudio의 철학 중 하나는 Rmarkdown, shiny 등을 활용한 사람들과의 소통임.

- Rstudio를 활용한 파일 업로드 및 다운로드 방법
- Rmarkdown 작성을 위한 Visual mode 활용
- Visual mode에서 이미지, 코드블럭 넣기

```
a <- 2  
b <- 3  
c <- a*b  
c
```

15.4 Vector

vector는 R의 기본 데이터 구조. numeric vector, logical vector, character vector 등 저장되는 값의 타입에 따라 크게 세가지로 나눌 수 있으며 `class()` 함수를 이용해서 값의 타입을 알아낼 수 있음. Combine function인 `c()`를 활용하여 만들며 값을 순차적으로 붙일 수 있음.

```
v1 <- c(1, 2, 3, 4, 5)  
v2 <- c("a", "b", "c", "d", "e")  
class(v1)  
class(v2)
```

15.5 Data frame

데이터프레임은 매트릭스와 같은 형태로 컬럼 하나가 하나의 벡터 변수로서 각 변수들이 다른 모드의 값을 저장할 수 있음. \$ 기호를 이용하여 각 구성 변수를 참조할 수 있으며 컬럼 한 줄이 하나의 변수 이므로 새로운 변수도 컬럼 형태로 붙여 넣을 수 있음. 즉, 각 row는 샘플을 나타내고 각 column은 변수를 나타내며 각 변수들이 갖는 샘플의 개수 (row의 길이, vector의 길이)는 같아야 함. R 기반의 데이터 분석에서는 가장 선호되는 데이터 타입으로 최근 tidyverse 패키지의 tibble 타입은 dataframe의 업그레이드 버전으로 볼 수 있음.

```
v_df <- data.frame(v1, v2)  
v_df  
class(v_df)
```

```
v3 <- c(121, 22, 31, 98, 45)
v4 <- c("teamA", "teamB", "teamC", "teamD", "teamE")
v5 <- data.frame(v1, v2, v3, v4)
v5
```

15.6 TODO

다음 주제들에 대한 Rmarkdown 문서를 만드시오

Dr성원재

- 임의의 엑셀 파일 2개 이상 읽기
- tidyverse 패키지 이용 데이터들을 병합하기
- 해당 데이터 가시화용 그래프 그리기

박아영

- 엑셀 데이터 읽기
- plot 이용한 그래프 그리기

박성균

- Shiny 이용한 app 만들기 (NGS 파이프라인 관리?)

Aporva

- Prepare MySQL docker image
- Docker-compose for more than two docker images
- To access MySQL server in R

허성보

- 리눅스 디렉토리 구조 및 설명
- vi 에디터로 파일 생성
- Docker 이미지 만들기

고민지

- R 데이터 타입 정리
- 숫자형 벡터 만들기과 연산
- dataframe 만들기과 인덱싱

16 수업 (2021115)

- 각자 Topic 들에 대한 스터디 및 Rmarkdown 문서 작성 진행

17 수업 (20211118)

- 각자 Topic Rmarkdown 문서 업데이트 및 제출

18 수업 (20211122)

- Rmarkdown 문서 점검 및 코드 재작성
- Rstudio 프로젝트 활용
- Rstudio 목차 기능 활용 (document outline)
- DNA 서열 읽고 쓰기 Biostrings 패키지 활용
- pay211118 강의 Rmarkdown 코드

19 수업 (20211125)

19.1 성원재

20 수업 (20211129)

- 효소공학회 참가

21 수업 (20211202)

- 수업

22 수업 (20211206)

- 리포트 작성 실습

23 수업 (20211213)

23.1 성원재

- 보고서 구조 완성 및 pdf 형태로 rmarkdown 마무리 요청
- rmarkdown 사용법 등 배우는 것을 목적으로

23.2 허성보

- 보고서 구조 완성 리눅스에서 Rocker (docker) 실행하는 코드 실습

23.3 고민지

- Rstudio에서 엑셀 파일 읽고 그래프 그리기
- tidyverse 패키지 활용 실제 데이터 엑셀파일 막대그래프 그리기

23.4 박성균

- demultiplexing 코드 리뷰
- Bash script 형태로 진행, R CMD 활용

23.5 Aporva

- 단백질 엔지니어링 툴 개발 기본 방향 설명
- 보고서 Intro, Discussion 보강 요청

23.6 박아영