

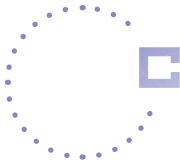
BIO-EXPRESS 2.0

USER MANUAL

바이오 익스프레스 2.0 사용자 매뉴얼

BIO-EXPRESS GBOX





BIO-EXPRESS 2.0 USER MANUAL

바이오 익스프레스 2.0 사용자 매뉴얼

Bio-Express GBox

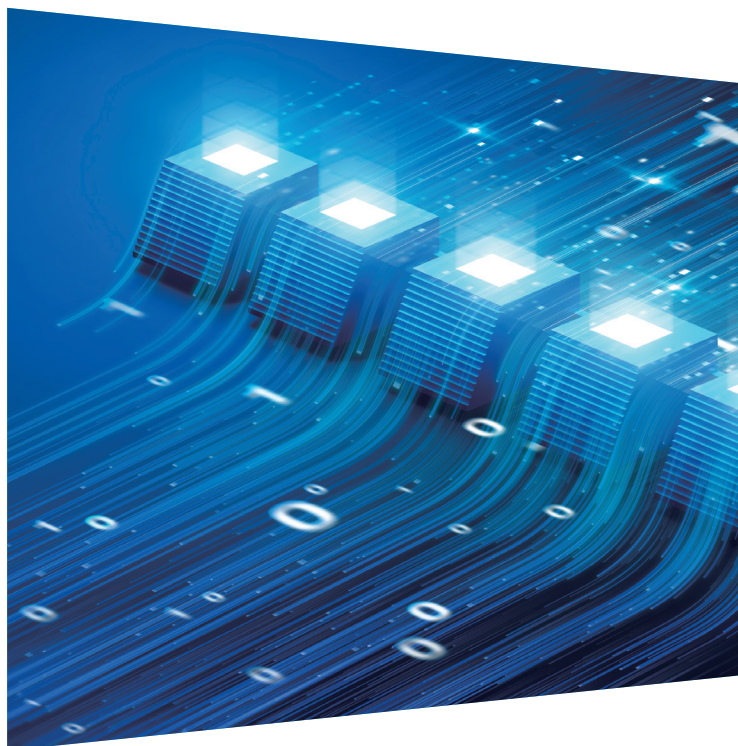
- 04 1.1 화면 및 기능 소개
- 05 1.2 빅데이터 고속 전송
- 07 1.3 데이터 파일 관리 기능
- 08 1.4 명령 줄 인터페이스 기반 GBox-CLI



BIO-EXPRESS 2.0

국가생명연구자원정보센터 (KOBIC)는

대규모 분석 인프라 또는
고도화된 분석 기술이 필요한
연구자를 위해 Bio-Express 유전체
빅데이터 분석 클라우드 서비스를
제공합니다.



EX BIO EXPRESS

Bio-Express 서비스는 동적 컨테이너 기반 자동화된 워크플로우 분석 플랫폼과 고속 데이터 전송 서비스를 통해 과학 분야의 빅데이터 분석을 가능하게 하는 클라우드 기반 통합 데이터 분석 서비스입니다.

Bio-Express는 코딩을 모르는 실무자부터 데이터 전문가에 이르기까지 간편한 데이터 분석이 가능한 다양한 클라우드 기반 데이터 분석 서비스와 고속 분석을 위한 빅데이터 플랫폼 기반 인프라 서비스를 제공합니다.

Cloud-Based Open Integrated Analytics Systems

클라우드 기반 개방형 통합 분석 시스템



CLOSHA

Bio-Express 워크벤치는 분석 알고리즘 구성 요소를 이용해 빅데이터 분석 워크플로우를 설계하고 분석을 고속으로 실행하는 플랫폼입니다. 사용자는 원하는 분석 파이프라인 또는 알고리즘의 세트 값을 정의하여 파이프라인을 재구성함으로써 간단하고 신속하게 분석할 수 있습니다. 동적 컨테이너 기반 사용자 코드 실행을 지원하여 자유도가 높은 서비스를 지원합니다.

GBOX

다양한 과학 분야에서 빅데이터의 전송을 위해 KOBIC은 최대 전송 속도, 절감된 파일 전송 비용, 양방향 데이터 전송, 체계적 파일 관리, 신속한 동기화 및 빅 데이터 저장소 백업을 제공합니다.



화면 및 기능 소개

GBox 고속 전송 시스템은 가속기 등의 추가적인 하드웨어 구축이나 ActiveX와 같은 별도의 플러그인 설치 없이 소프트웨어 기술만으로 사용자의 대용량 데이터를 고속으로 전송한다. GBox를 통해 업로드 된 사용자 데이터는 CLOSHA에서 사용 가능하다.

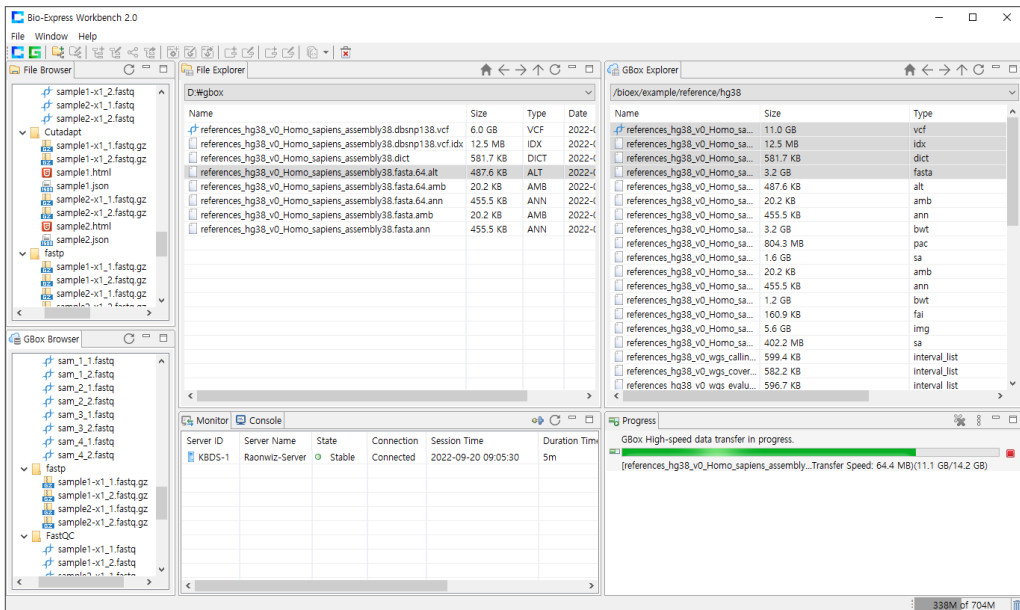


그림 1.1- 1바이오 익스프레스 GBox 화면

GBox는 쉽고 직관적인 사용자 인터페이스로 설계되어 별도의 교육 없이 사용이 가능하며, 파일 업로드 및 다운로드시 아래의 전송 조건을 만족해야 한다.

● 전송 조건

- 파일명을 포함한 전체 경로에 공백, 한글 또는 언더바(_), 하이픈(-), 마침표(.)를 제외한 특수 문자가 있을 경우 전송 장애 발생
- 외장하드나 USB 경로에서 전송시 속도 차이에 따른 전송 장애 발생
- 0B 빈파일의 경우 전송 장애 발생
- 이외에 손상된 파일 등 파일 자체에 문제가 있는 경우 전송 장애 발생

1.2

빅데이터 고속 전송

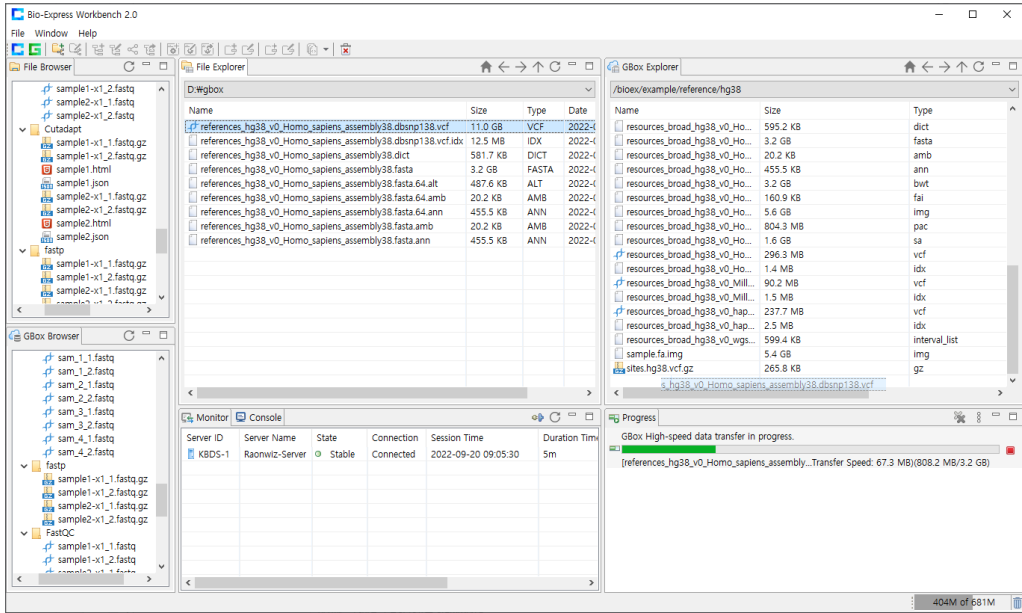


그림 1.2 - 1 바이오 익스프레스 GBox 데이터 다운로드 화면

● 업로드 방법 1

[File Explorer] view에서 전송하고자 하는 폴더 또는 파일 선택 하여 [GBox Explorer] view 로 Drag & Drop 한다.

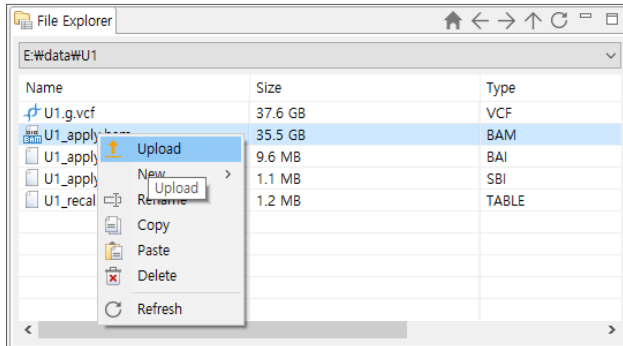


그림 1.2 - 2 컨텍스트 메뉴를 이용한 데이터 업로드

● 업로드 방법 2

[File Explorer] view에서 전송하고자 하는 폴더 또는 파일 선택 하여 컨텍스트 메뉴의 [Upload] 버튼으로 전송한다.

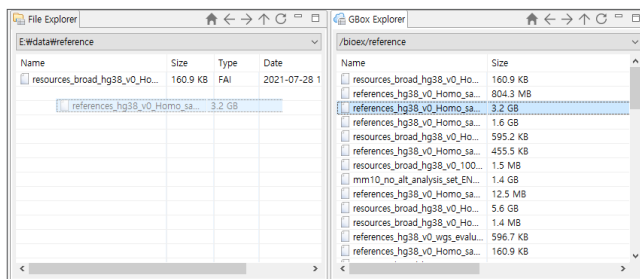


그림 1.2 - 3 Drag & Drop 을 이용한 데이터 업로드

● 다운로드 방법 1

[GBox Explorer] 에서 전송하고자 하는 폴더 또는 파일을 선택 하여 [File Explorer] view 로 Drag & Drop 한다.

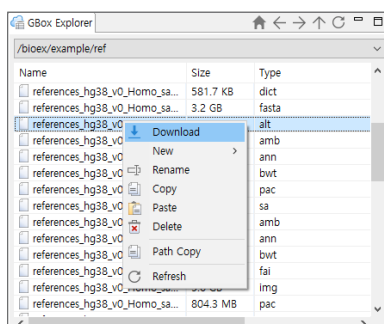


그림 1.2 - 4 컨텍스트 메뉴를 이용한 데이터 다운로드

● 다운로드 방법 2

[GBox Explorer] 에서 전송하고자 하는 폴더 또는 파일을 선택 하여 컨텍스트 메뉴 [Download] 버튼으로 전송한다.

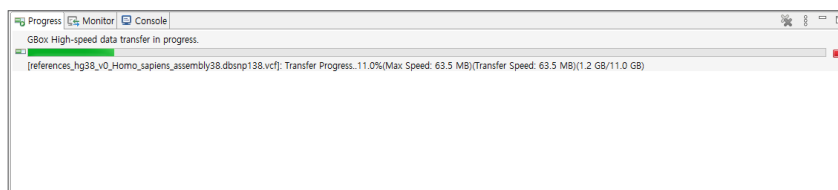


그림 1.2 - 5 데이터 전송 모니터링

파일 전송이 시작되면 어플리케이션 하단의 [Progress] view로 실시간 전송 상황 모니터링이 가능하다.

1.3

데이터 파일 관리 기능

- [File Explorer] view와 [GBox Explorer] view의 컨텍스트 메뉴로 데이터를 관리 할 수 있다.
([New], [Rename], [Copy], [Paste], [Delete], [Refresh])
- [File Explorer] view와 [GBox Explorer] view의 상단의 툴바 메뉴로 디렉토리 이동 및 바로가기 기능을 사용할 수 있다.

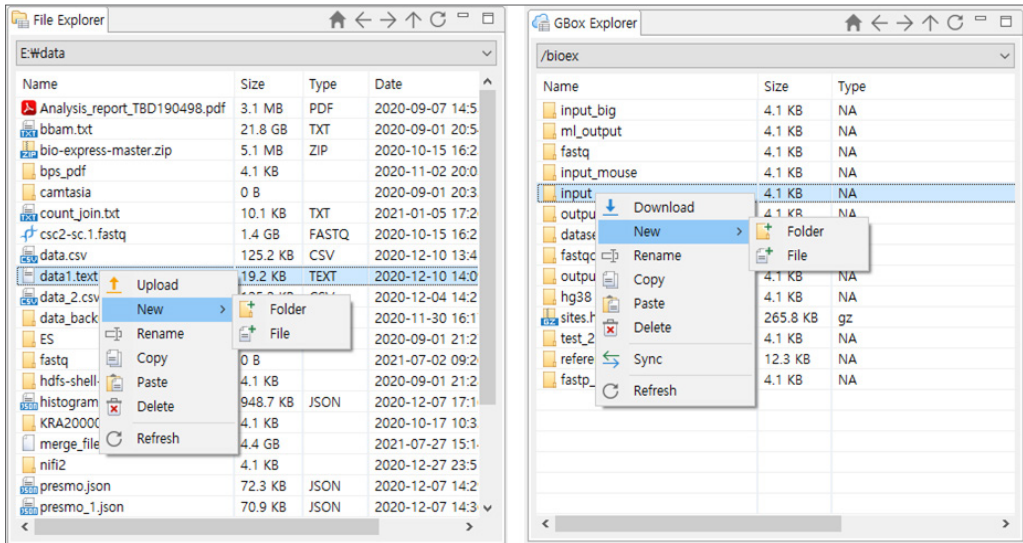


그림 1.3 - 1저장소 파일 관리 기능



명령 줄 인터페이스 기반 GBox-CLI

GBox 고속 전송 시스템은 GUI(Graphical User Interface) 뿐만 아니라 CLI(Command Line Interface) 기반의 GBox-CLI를 제공하여 다양한 클라우드 서버나 리눅스 환경에서도 다 이렉트로 고속 전송을 지원한다.

```

명령 프롬프트 - java -jar "gbox.cli-v.2.1.jar"
D:\g\gbox> java -jar "gbox.cli-v.2.1.jar"
BiO-Express GBox
Powered by KOBIC Cloud-Team v.1.0
11월 04, 2022 3:46:27 오후 org.jline.utils.Log logr
경고: The Parser of class org.springframework.shell.jline.ExtendedDefaultParser does not support the CompletingParsed
line interface. Completion with escaped or quoted words won't work correctly.
GBox> gbox -u USER_ID
Password:
2022-11-04 15:50:09 connect!! ==> USER_ID
GBox> cd pine_stem_20180605
/bioex/pine_stem_20180605
GBox> |
+-----+-----+-----+-----+
| Name | Size | Type | Date |
+-----+-----+-----+-----+
1. | pine_stem_20180605_N350_S1_L001_R1_001.fastq.gz | 24.4 GB | File | 2022-10-17 14:51:17 |
+-----+-----+-----+-----+
GBox> pwd
/bioex/pine_stem_20180605
GBox> receive -s /bioex/pine_stem_20180605/pine_stem_20180605_N350_S1_L001_R1_001.fastq.gz -t D:/gbox/
[##*****] 20.6%,transfer speed:[15MB] ,(4799MB/23309MB)
  
```

그림 1.4 - 1 GBox-CLI 구동 화면

1.4.1. GBox-CLI 로그인

- KOBIC 통합로그인(SSO)에 가입 되어 있는 아이디로 로그인 한다.

```

GBox> gbox -u USER_ID
Password:
2022-10-17 14:33:37 connect!! ==> 사용자ID
GBox>
  
```

그림 1.4.1 - 1 GBox-CLI 로그인

- 명령어
GBox] gbox -u <사용자 계정>

1.4.2. 명령어 확인

● GBox-CLI에서 제공하는 모든 명령어에 대한 내용을 확인한다.

- 명령어

GBox] help

GBox] help <명령어>

```
GBox] help
AVAILABLE COMMANDS

K Livepass CMD
ll: 현재 위치한 디렉토리 내용 상세 출력 명령어
cd: 디렉토리 위치 변경 명령어
disconnect: GBox 서버 접속 중지
receive: KOBIC 클라우드 저장소 데이터 다운로드 명령어
ls: 현재 위치한 디렉토리 내용 출력 명령어
health: 현재 접속 상태 표기
check: GBox 전송 테스트
exit: Exit the shell.
reconnect: GBox 서버 재접속
gbox: KOBIC 클라우드 저장소 접속 명령어
quit: Exit the shell.
state: 현재 접속 상태 표기
rm: 파일 혹은 디렉토리 삭제 명령어
pwd: 현재 디렉토리 절대 경로 출력 명령어
mkdir: 디렉토리 생성 명령어
send: KOBIC 클라우드 저장소 데이터 업로드 명령어
cd ..: 상위 디렉토리 이동 명령어
```

그림 1.4.2-1 전체 명령어 도움말 확인

```
GBox] help receive
NAME
    receive - KOBIC 클라우드 저장소 데이터 다운로드 명령어

SYNOPSIS
    receive [-s String] [-t String] -o String

OPTIONS
    -s String
    다운로드할 파일 및 폴더 경로
    [Mandatory]

    -t String
    다운로드 경로
    [Mandatory]

    -o String
    덮어쓰기
    [Optional, default = false]
```

그림 1.4.2-2 특정 명령어 도움말 확인

1.4.3. 데이터 업로드/다운로드

- 데이터를 고속 업로드/다운로드 한다.

- 명령어

GBox] send -s <사용자 로컬 저장소 경로> -t <GBox 저장소 경로>

GBox] receive -s <GBox 저장소 경로> -t <사용자 로컬 저장소 경로>

```

gBox> cd pine_stem_20180605
/bioex/pine_stem_20180605
gBox> ll

```

	Name	Size	Type	Date
1.	pine_stem_20180605_N350_S1_L001_R1_001.fastq.gz	24.4 GB	File	2022-10-17 14:51:17

```

gBox> pwd
/bioex/pine_stem_20180605
gBox> receive -s /bioex/pine_stem_20180605/pine_stem_20180605_N350_S1_L001_R1_001.fastq.gz -t D:/gbox
[#####] 20.6%,transfer speed:[15MB],(4799MB/23309MB)

```

그림 1.4.3 - 1 데이터 다운로드

```

gBox> pwd
/bioex/upload
gBox> ll

```

	Name	Size	Type	Date
1.	500GB	536.9 GB	File	2022-10-18 23:10:42
2.	gbox_cli-v.1.6.jar	24.3 MB	File	2022-10-18 10:01:16

```

gBox> send -s D:/gbox/pine_stem_20180605_N350_S1_L001_R1_001.fastq.gz -t /bioex/upload
[#####] 100.0%,transfer speed:[67MB],(23309MB/23309MB)
gBox> ll

```

	Name	Size	Type	Date
1.	500GB	536.9 GB	File	2022-10-18 23:10:42
2.	gbox_cli-v.1.6.jar	24.3 MB	File	2022-10-18 10:01:16
3.	pine_stem_20180605_N350_S1_L001_R1_001.fastq.gz	24.4 GB	File	2022-11-04 16:09:17

```

gBox>

```

그림 1.4.3 - 2 데이터 업로드

1.4.4. md5sum 데이터 전송 무결성

- 클라우드 저장소에 업로드한 데이터의 무결성을 확인한다.

- 명령어

GBox] checksum <GBox 저장소에 위치한 파일 경로>

```

$ pwd
/bioex/example/program/cellranger_10k/10k_PBMC_3p_nextgem_Chromium_Controller_fastqs
$ ll

```

Name	Size	Type	Date
1. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_I1_001.fastq.gz	752.4 MB	File	2021-07-17 05:50:51
2. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_I2_001.fastq.gz	754.6 MB	File	2021-07-17 05:50:53
3. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_R1_001.fastq.gz	2.1 GB	File	2021-07-17 05:50:58
4. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_R2_001.fastq.gz	4.8 GB	File	2021-07-17 05:51:10
5. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L002_I1_001.fastq.gz	756.1 MB	File	2021-07-17 05:51:11
6. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L002_I2_001.fastq.gz	762.8 MB	File	2021-07-17 05:51:13
7. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L002_R1_001.fastq.gz	2.1 GB	File	2021-07-17 05:51:18
8. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L002_R2_001.fastq.gz	4.8 GB	File	2021-07-17 05:51:31
9. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L003_I1_001.fastq.gz	767.1 MB	File	2021-07-17 05:51:33
10. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L003_I2_001.fastq.gz	772.8 MB	File	2021-07-17 05:51:35
11. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L003_R1_001.fastq.gz	2.1 GB	File	2021-07-17 05:51:39
12. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L003_R2_001.fastq.gz	4.9 GB	File	2021-07-17 05:51:51
13. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L004_I1_001.fastq.gz	767.0 MB	File	2021-07-17 05:51:53
14. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L004_I2_001.fastq.gz	772.8 MB	File	2021-07-17 05:51:55
15. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L004_R1_001.fastq.gz	2.1 GB	File	2021-07-17 05:52:00
16. 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L004_R2_001.fastq.gz	4.9 GB	File	2021-07-17 05:52:13

```

$ checksum 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_I1_001.fastq.gz
출대 경로: /bioex/example/program/cellranger_10k/10k_PBMC_3p_nextgem_Chromium_Controller_fastqs/10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_I1_001.fastq.gz MD5 Hash Value: [55629f56b7e6272c419fe4ad2a5d6b56]
$ checksum 10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_R2_001.fastq.gz
출대 경로: /bioex/example/program/cellranger_10k/10k_PBMC_3p_nextgem_Chromium_Controller_fastqs/10k_PBMC_3p_nextgem_Chromium_Controller_S2_L001_R2_001.fastq.gz MD5 Hash Value: [a1323c2629a7112a31b49a0ab3d72ee9]

```

그림 1.4.4 - 1 데이터 전송 무결성 확인

1.4.5. 디렉토리 및 파일 조회

- 현재 위치한 클라우드 저장소 경로의 내용을 출력한다.

- 명령어

GBox] ll

GBox] ls

```

$ pwd
/bioex/example/ref
$ ll

```

Name	Size	Type	Date
1. BasesCounts.interval_list	582.2 kB	File	2022-05-10 09:05:27
2. hg38_annotar	0 B	Dir	2022-05-10 09:03:04
3. Homo_sapiens.GRCh38.dna.primary_assembly.fa	3.2 GB	File	2022-05-10 09:03:11
4. Homo_sapiens.GRCh38.dna.primary_assembly.fa.fai	6.4 kB	File	2022-05-10 09:00:59
5. Homo_sapiens.GRCh38.dna.primary_assembly.fa.index	49.2 kB	File	2022-05-10 09:01:55
6. Homo_sapiens.GRCh38.dna.toplevel.fa	64.2 GB	File	2022-05-10 09:05:21
7. Homo_sapiens.GRCh38.dna.toplevel.fa.fai	27.9 kB	File	2022-05-10 09:05:29
8. Homo_sapiens.GRCh38.dna.toplevel.fa.gz	1.1 GB	File	2022-05-10 09:01:04
9. Homo_sapiens.GRCh38.dna.toplevel.fa.index	98.3 kB	File	2022-05-10 09:01:31
10. references_hg38_v0_Homo_sapiens_assembly38.dbsnp138.vcf	11.0 GB	File	2022-05-10 09:02:37
11. references_hg38_v0_Homo_sapiens_assembly38.dbsnp138.vcf.idx	12.5 MB	File	2022-05-10 09:01:02
12. references_hg38_v0_Homo_sapiens_assembly38.dict	581.7 kB	File	2022-05-10 09:01:13
13. references_hg38_v0_Homo_sapiens_assembly38.fasta	2.2 GB	File	2022-05-10 09:05:27

그림 1.4.5 - 1 디렉토리 내용 출력

1.4.6. 디렉토리 생성/삭제

- 현재 위치한 클라우드 저장소 경로에 폴더를 생성하고 삭제한다.

- 명령어
 - GBox] mkdir <생성할 폴더명>
 - GBox] rm <삭제할 폴더명 또는 파일명>

```
GBox] pwd
/bioex/upload
GBox] ll

```

	Name	Size	Type	Date
1.	500GB	536.9 GB	File	2022-10-18 23:10:42
2.	gbox_cli-v.1.6.jar	24.3 MB	File	2022-10-18 10:01:16
3.	pine_stem_20180605_N850_S1_L001_R1_001.fastq.gz	24.4 GB	File	2022-11-04 16:09:17

```
GBox] mkdir folder1
폴더 생성을 완료 하였습니다.
GBox] ll

```

	Name	Size	Type	Date
1.	500GB	536.9 GB	File	2022-10-18 23:10:42
2.	folder1	0 B	Dir	2022-11-04 16:26:42
3.	gbox_cli-v.1.6.jar	24.3 MB	File	2022-10-18 10:01:16
4.	pine_stem_20180605_N850_S1_L001_R1_001.fastq.gz	24.4 GB	File	2022-11-04 16:09:17

```
GBox]
```

그림 1.4.6 - 1 디렉토리 생성 및 삭제

1.4.7. 경로 조회/이동

- 현재 위치한 클라우드 저장소의 경로를 조회하거나 이동한다.

- 명령어
 - GBox] pwd
 - GBox] cd <이동할 하위 경로>
 - GBox] cd..

```
GBox] pwd
/bioex/upload
GBox] cd folder1
/bioex/upload/folder1
GBox]
GBox] cd ..
/bioex/upload
GBox] pwd
/bioex/upload
GBox]
```

경로 조회

하위 경로 이동

상위 경로 이동

그림 1.4.7 - 1 경로 조회 및 이동



BIO-EXPRESS 2.0 USER MANUAL

바이오 익스프레스 2.0 사용자 매뉴얼