

미생물 유전체 분석 교육 수강자를 위한 준비 사항

KRIBB 감염병연구센터

정해영

hyjeong@kribb.re.kr

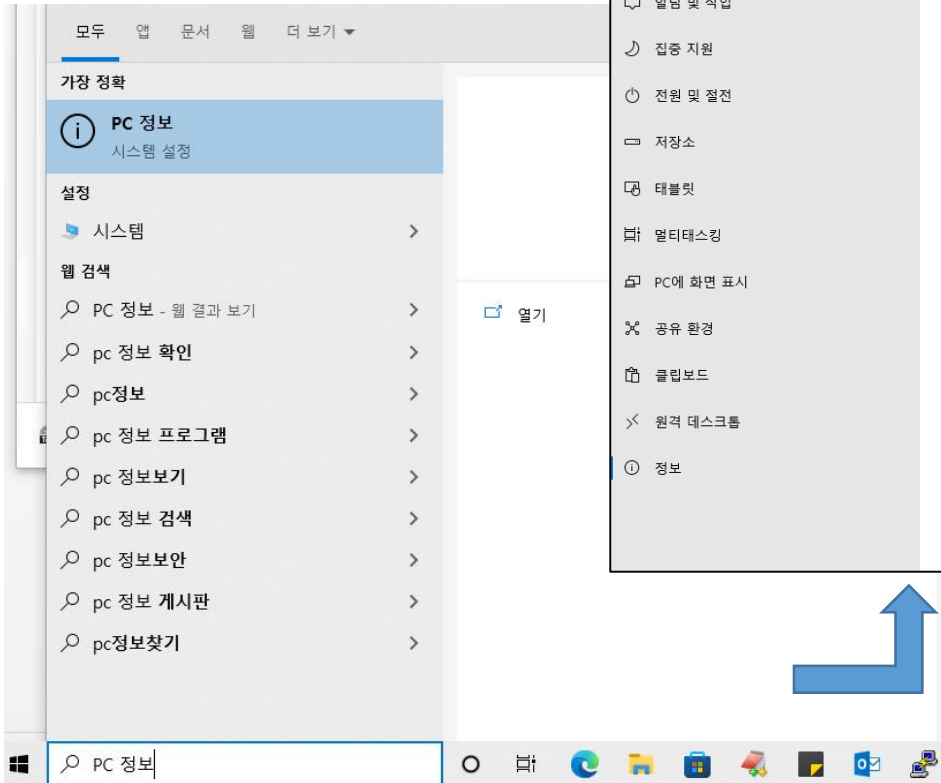
실습을 위한 필수 준비 사항

- Windows가 설치된 노트북 컴퓨터
 - Windows 10 버전 2004 이상(빌드 19041 이상) 또는 Windows 11
 - 64비트 시스템(x64)
- 미리 설치할 프로그램 및 파일(**필수**)
 - 1 {
 - **Linux용 Windows 하위 시스템**(Windows Subsystem for Linux, WSL)
 - **Windows Terminal**: 명령 프롬프트('cmd') 대신 추천함
 - 2 {
 - 별도 제공할 **실습용 Ubuntu distro(9.2GB)**를 WSL 환경에 설치
 - 미리 준비한 **data 파일(2.7GB)**을 Ubuntu distro 안에 복사
 - 3 {
 - Java 실행 환경 **AdoptOpenJDK 11(LTS)**
 - ✓ Oracle Java 대신 이것을 추천함
 - 실습에 사용할 Java 응용프로그램은 별도로 소개
 - ✓ 파일 용량이 크지 않아서 실습 당일에 설치해도 무관함
 - [Optional] 텍스트 편집기 Notepad++ v8.3.3 [다운로드 링크](#)

- 1: Windows 제어판 및 Microsoft Store 이용
- 2: 슬라이드 8~14번 참조
- 3: 인터넷에서 다운로드하여 설치(슬라이드 16번~ 참조)

PC 정보 확인

- Windows 10 버전 2004 이상(빌드 19041 이상) 또는 Windows 11
- 64비트 시스템(x64)



또는



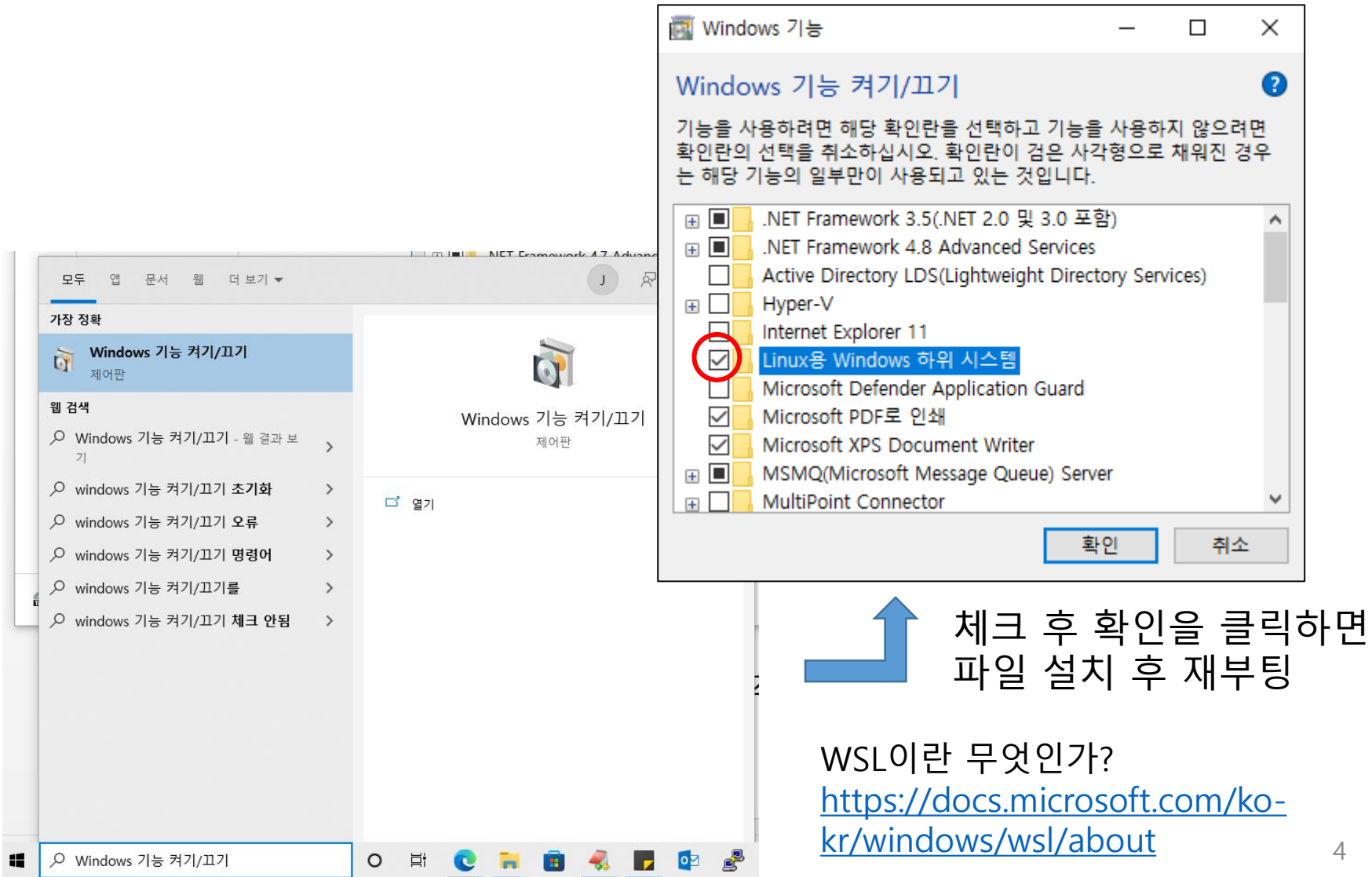
+ R →

열기:

winver

WSL 기능 켜기

Windows Subsystem for Linux



The image shows a Windows Settings window titled 'Windows 기능' (Windows Features). The 'Linux용 Windows 하위 시스템' (Windows Subsystem for Linux) feature is checked and highlighted with a red circle. Below the list are '확인' (OK) and '취소' (Cancel) buttons. An arrow points from the text below to the '확인' button.

기능을 사용하려면 해당 확인란을 선택하고 기능을 사용하지 않으려면 확인란의 선택을 취소하십시오. 확인란이 검은 사각형으로 채워진 경우는 해당 기능의 일부만이 사용되고 있는 것입니다.

- .NET Framework 3.5(.NET 2.0 및 3.0 포함)
- .NET Framework 4.8 Advanced Services
- Active Directory LDS(Lightweight Directory Services)
- Hyper-V
- Internet Explorer 11
- Linux용 Windows 하위 시스템
- Microsoft Defender Application Guard
- Microsoft PDF로 인쇄
- Microsoft XPS Document Writer
- MSMQ(Microsoft Message Queue) Server
- MultiPoint Connector

확인 취소

체크 후 확인을 클릭하면
파일 설치 후 재부팅

WSL이란 무엇인가?
<https://docs.microsoft.com/ko-kr/windows/wsl/about>

Windows Terminal 설치

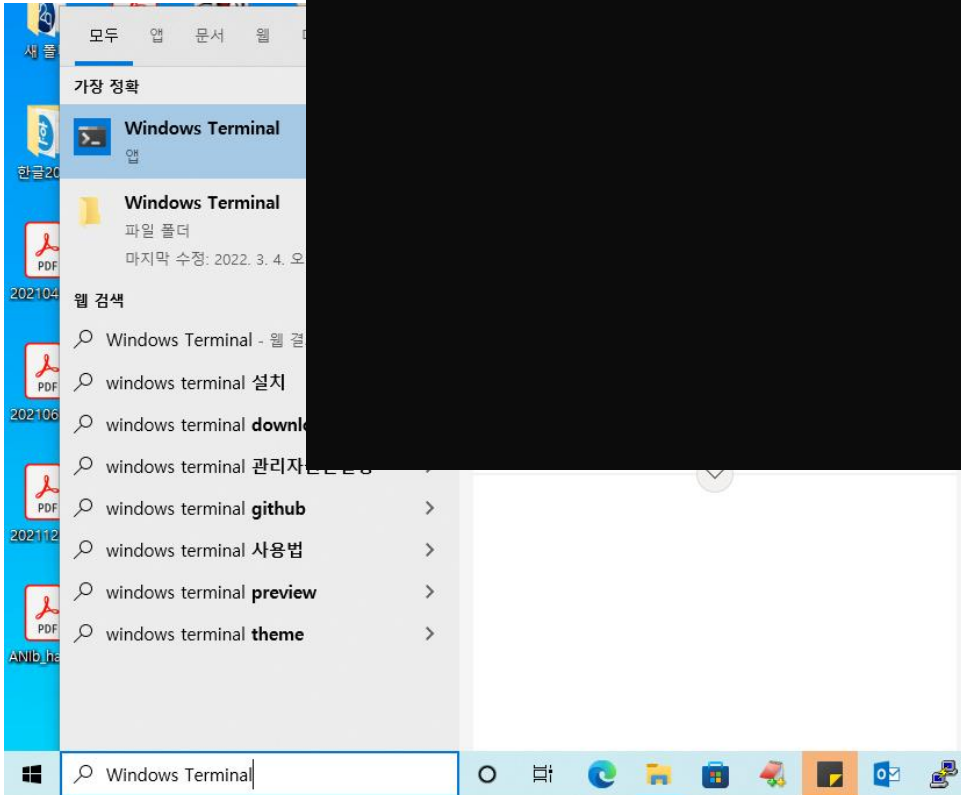
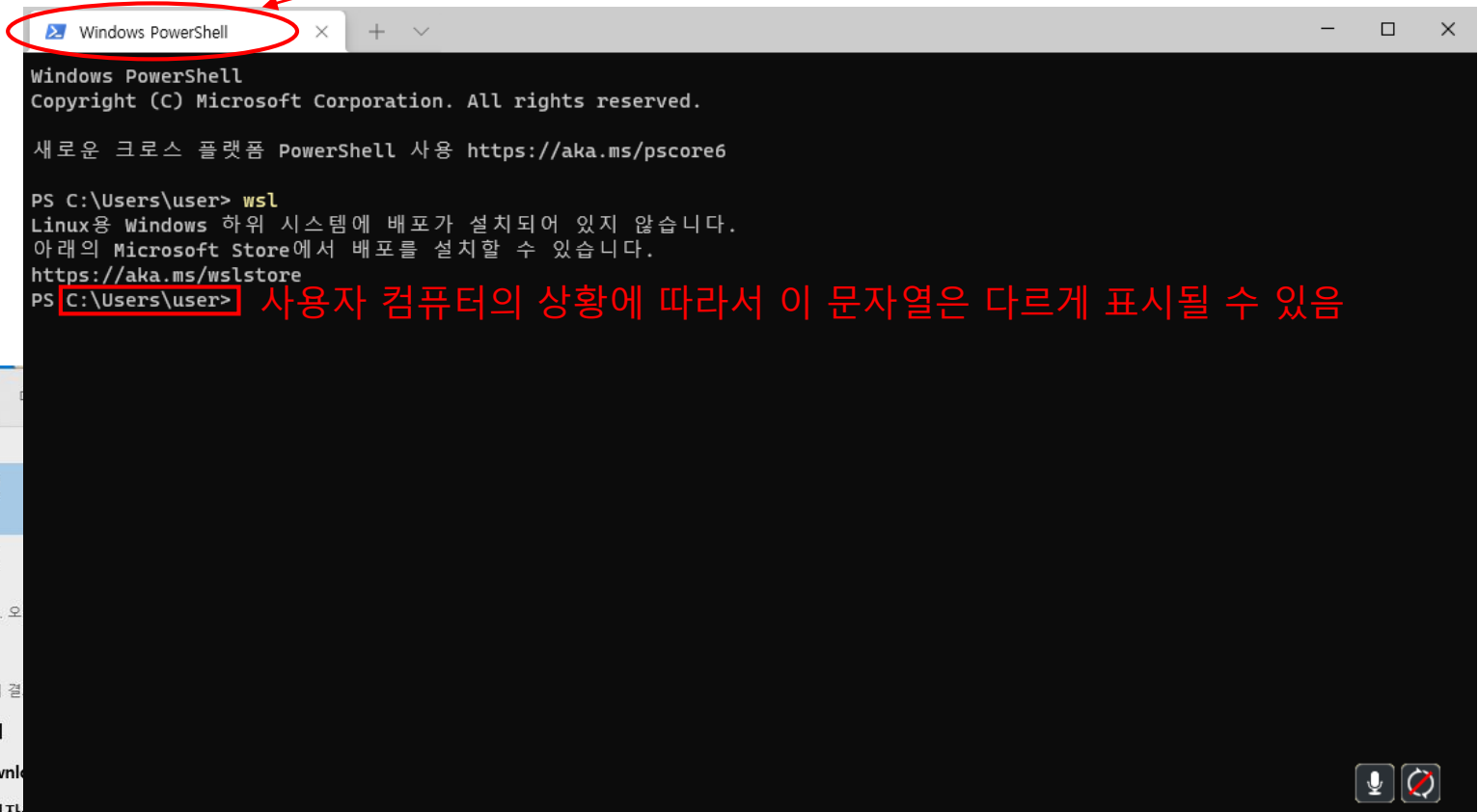
Windows 터미널 명령 프롬프트, PowerShell 및 bash(Linux용 Windows 하위 시스템(WSL)을 통해)와 같이 이미 좋아하는 명령줄 셸에 대한 최신 호스트 애플리케이션입니다([하이퍼링크](#)).

The image shows a Windows search interface on the left and a Microsoft Store window on the right. The search results list 'Microsoft Store' as the top result. The Microsoft Store window displays the 'Windows Terminal' app page, including its icon, developer information (Microsoft Corporation), a 4.7 star rating, 82 reviews, and an IARC 3+ rating. A blue '설치' (Install) button is visible. Below the app page, there are search suggestions for 'microsoft store 설치', '재설치', '다운로드', 'download', 'file download', '오류', and '다운로드 오류'. A blue arrow points from the search results to the Microsoft Store window.

Microsoft Store에서 앱을 다운로드하려면 Microsoft 계정을 통해서 로그인해야 합니다([도움말](#)).

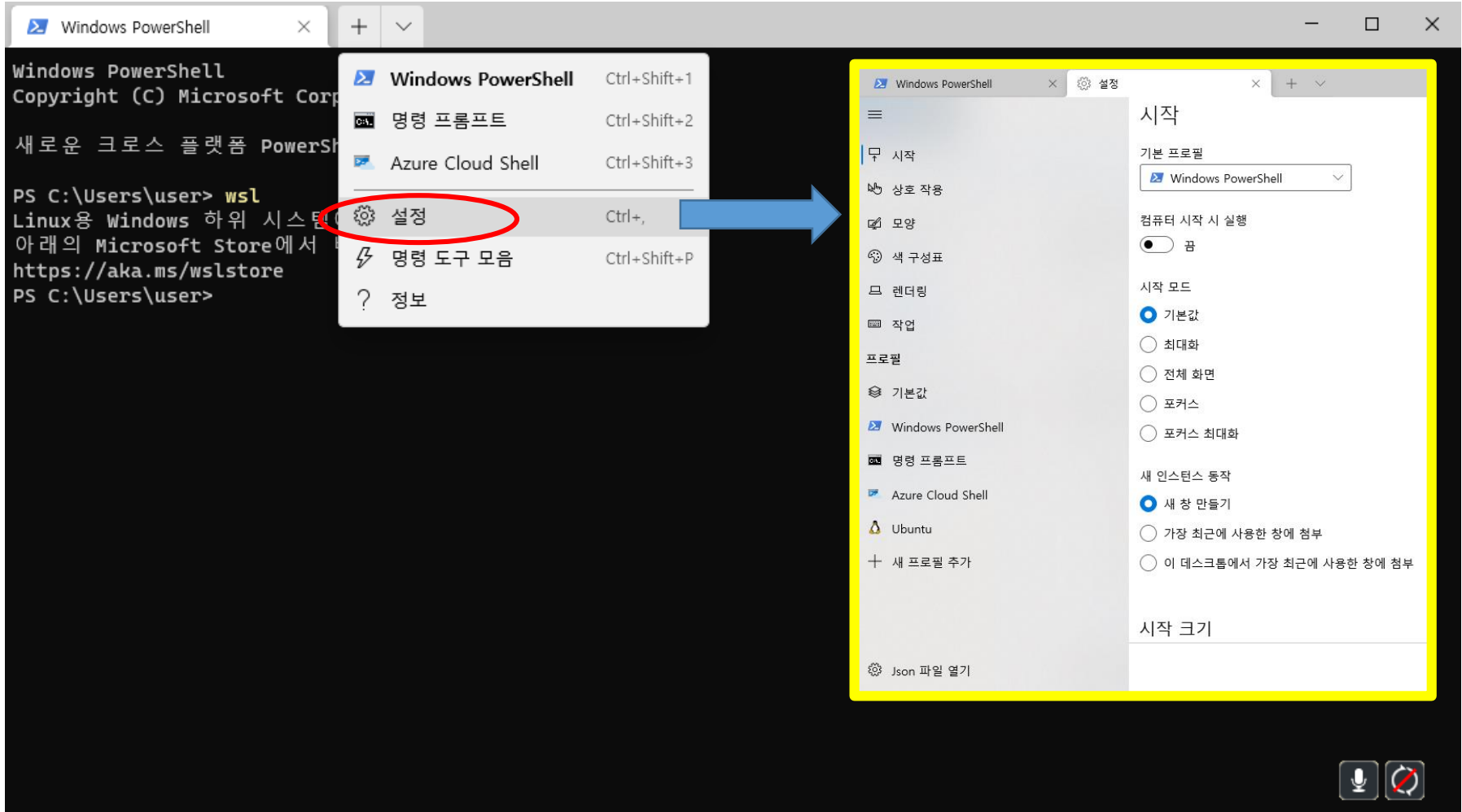
Windows Terminal 실행

Windows PowerShell(PS)이 Windows Terminal에 포함됨



<https://docs.microsoft.com/ko-kr/windows/terminal/>

Windows Terminal 설정 바꾸기



<https://cheatography.com/martinthogersen/cheat-sheets/windows-terminal/>

실습용 파일의 다운로드 위치

- 실습용 리눅스 배포용 파일: **KRIBBuntu-focal_2205a.tar** [9.24GB]
- 데이터 파일: **data_220615.tar** [2.65GB]

- KOBIC 3층 전산실습실이라면?
 - 실습용 PC의 바탕화면에 이미 복사되어 있습니다.
- KRIBB 내 유선 인터넷 환경이라면?
 - <http://192.168.130.248/share/kribbuntu/>에서 내려받으세요.
- 외부라면?
 - 상용 클라우드의 다음 링크를 이용하세요.
 - KRIBBuntu 배포 파일: <http://naver.me/59exlFvp> 용량을 줄이기 위하여 egg 파일[3.51GB]로 압축하였으므로, **반디집** 프로그램 무료버전으로 압축을 해제하여 tar 파일을 만드십시오
 - 데이터 파일: <http://naver.me/xmi3Ucyf>

실습용 배포(distro; tar 파일로 제공) 설치

1. 배포 파일 '**KRIBBuntu-focal_2205a.tar**'(9.24 GB)를 다운로드 하여 바탕화면에 복사합니다.
2. Windows Terminal을 열고 다음의 certutil 명령어를 입력하여 파일의 무결성을 점검합니다.

명령 대기 상태임을 나타내는 command prompt(자동으로 나타남)

```
> certutil -hashfile $env:USERPROFILE\Desktop\KRIBBuntu-focal_2205a.tar md5
```

```
PS C:\Users\user> certutil -hashfile $env:USERPROFILE\Desktop\KRIBBuntu-focal_2205a.tar md5
MD5의 C:\Users\user\Desktop\KRIBBuntu-focal_2205a.tar 해시:
016a0ef6429f49482344ebb75e870dc5
CertUtil: -hashfile 명령이 성공적으로 완료되었습니다.
PS C:\Users\user>
```

Hash 값: 016a0ef6429f49482344ebb75e870dc5

3. 배포를 설치할 폴더(설치 위치)를 만듭니다
4. 'wsl --import <배포> <설치 위치> <tar 파일 이름>' 명령을 이용하여 배포를 설치합니다.
 - <배포>와 <설치 위치>는 자유롭게 변경 가능합니다.

- \$env:USERPROFILE은 Window PowerShell에서 쓰이는 환경 변수([참고](#))
- Md5 checksum 확인: md5sum(리눅스), certutil(Windows cmd/PS), get-filehash(Windows PS)

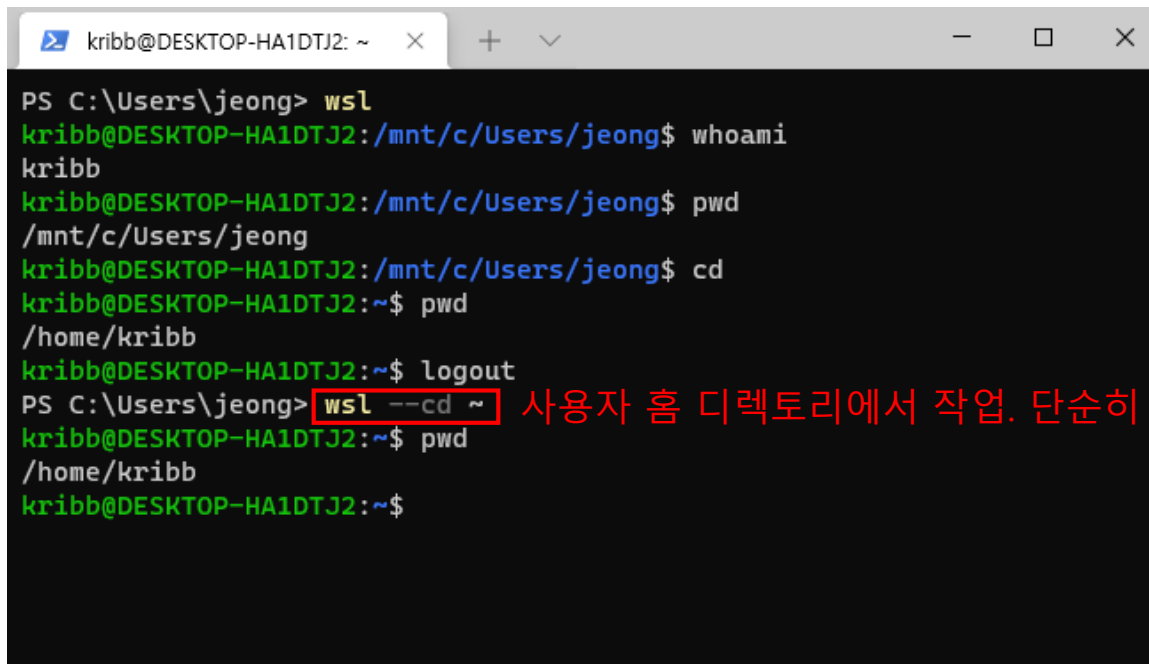
3.4번 과정의 실제

```
> mkdir C:\Distros\myUbuntu
> wsl --import myUbuntu C:\Distros\myUbuntu $env:USERPROFILE\Desktop\KRIBBuntu-focal_2205a.tar
> wsl --list
myUbuntu (기본값)
> wsl ← 아무 인수도 입력하지 않으면 기본값 배포를 실행
$
```

- 이미 설치한 다른 배포가 있는 경우 기본값이 아닌 것을 실행하려면 다음을 입력합니다.
 - > wsl -d <distro>
- '--cd <디렉토리>'로 특별히 지정하지 않았다면 \$env:USERPROFILE에서 시작합니다.
 - %USERPROFILE% 환경변수에 해당
- 사용자 홈 디렉토리에서 시작하려면 다음 중 하나를 입력합니다.
 - > wsl --cd ~ -d <distro>
 - > wsl ~ -d <distro>

처음 실행 화면

- 사용자 계정 **kribb**으로 로그인한 상태가 됩니다.
- 암호는 **Kribb#1234**입니다.
 - sudo 명령으로 관리자 권한을 행사할 때만 필요함
- 고급 설정 구성
 - <https://docs.microsoft.com/ko-kr/windows/wsl/wsl-config>



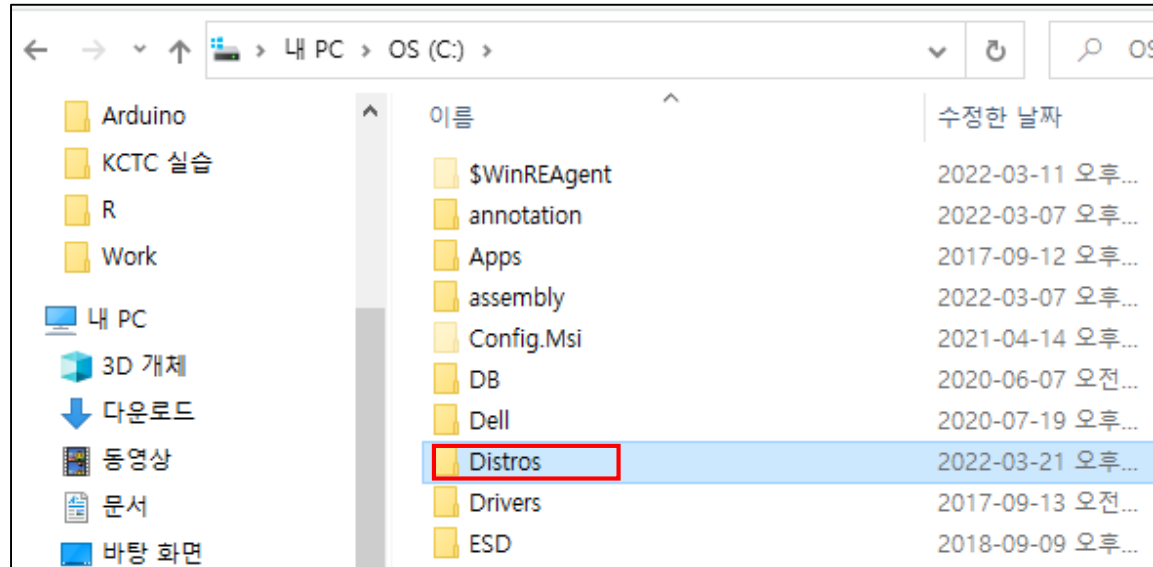
```
kribb@DESKTOP-HA1DTJ2: ~
PS C:\Users\jeong> wsl
kribb@DESKTOP-HA1DTJ2: /mnt/c/Users/jeong$ whoami
kribb
kribb@DESKTOP-HA1DTJ2: /mnt/c/Users/jeong$ pwd
/mnt/c/Users/jeong
kribb@DESKTOP-HA1DTJ2: /mnt/c/Users/jeong$ cd
kribb@DESKTOP-HA1DTJ2: ~$ pwd
/home/kribb
kribb@DESKTOP-HA1DTJ2: ~$ logout
PS C:\Users\jeong> wsl --cd ~
kribb@DESKTOP-HA1DTJ2: ~$ pwd
/home/kribb
kribb@DESKTOP-HA1DTJ2: ~$
```

사용자 홈 디렉토리에 작업. 단순히 'wsl ~'를 입력해도 됨

설치가 잘 안되었을 때에는..

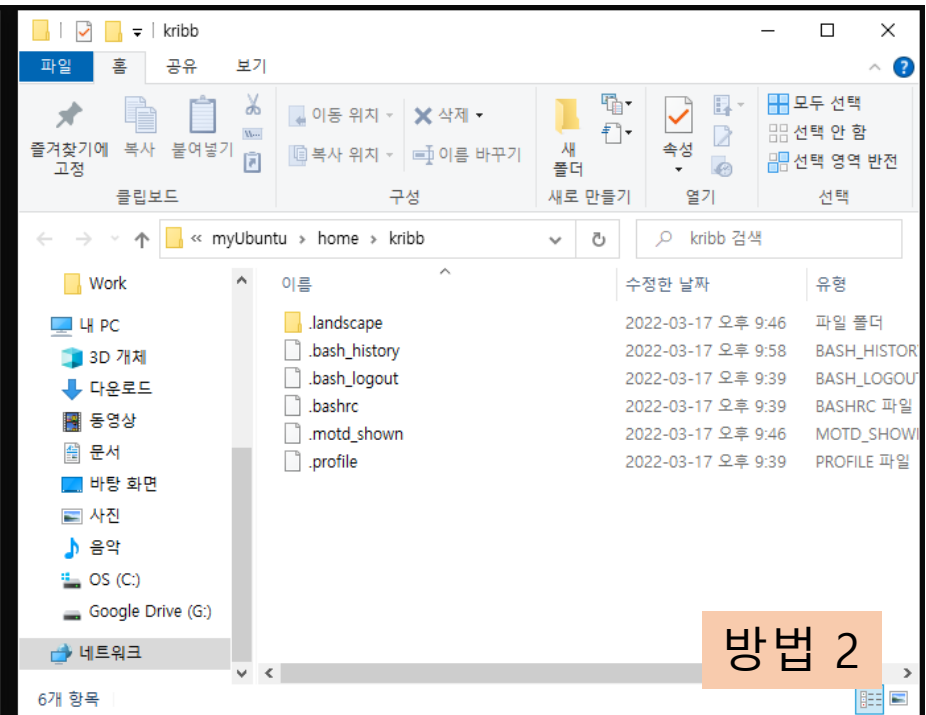
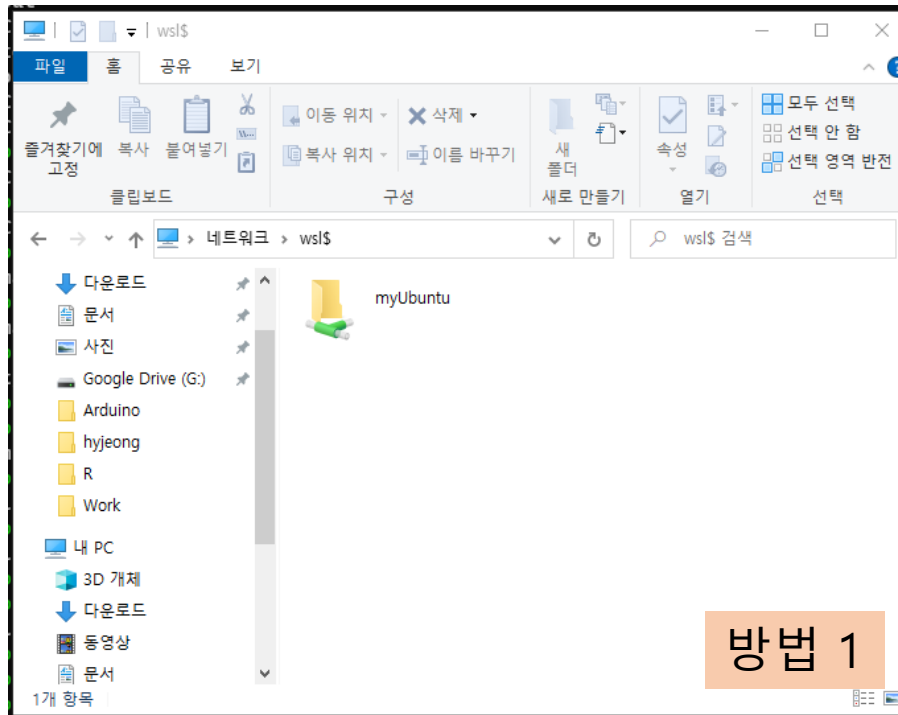
1. 'wsl --unregister <배포명>'으로 잘못 설치된 배포를 지웁니다.
 - 'wsl -l'은 현재 설치된 배포의 목록을 보여줍니다.
2. 탐색기를 열고 C:\WDistros 폴더를 지웁니다.
3. Windows Terminal에서 'mkdir C:\WDistros\myUbuntu' 명령부터 다시 실행합니다.

```
PS C:\Users\jeong> wsl -l
Linux용 Windows 하위 시스템 배포:
Ubuntu-20.04(기본값)
myUbuntu
PS C:\Users\jeong> wsl --unregister myUbuntu
등록 취소 중...
PS C:\Users\jeong> wsl -l
Linux용 Windows 하위 시스템 배포:
Ubuntu-20.04(기본값)
PS C:\Users\jeong> |
```



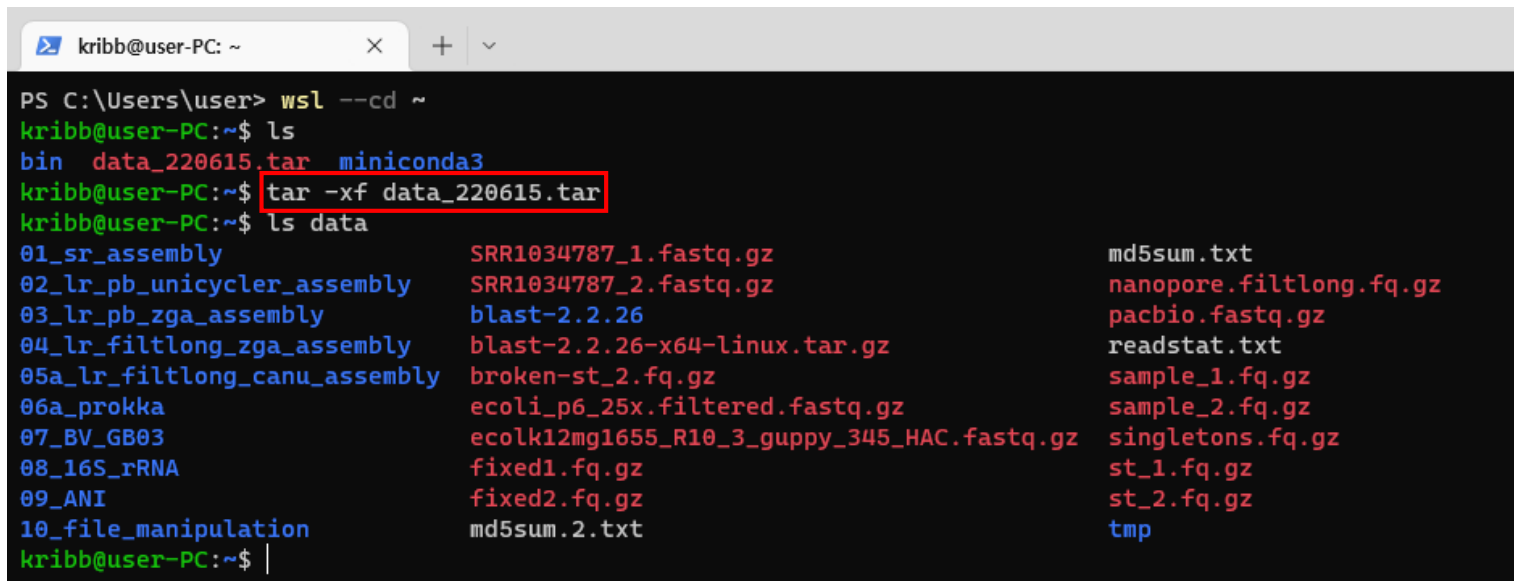
Ubuntu 내부 파일시스템에 접근하기

- Distro 실행 중에만 가능
- [방법 1] Windows 탐색기 주소창에서 '\\\\wsl\$' 입력
- [방법 2] Ubuntu 명령행에서 'explorer.exe .'을 입력
 - 반드시 명령어 다음에 점('.')을 입력해야 함



데이터 파일 설치하기

- 'data_220615.tar'(2.65GB) 파일은 실습 과정에서 다운로드한 대용량 파일 또는 분석을 통해 얻게 될 모든 결과를 미리 마련하여 묶은 것입니다.
- 탐색기를 이용하여 이 파일을 홈 디렉토리 (/home/kribb)로 복사한 뒤 우분투 내에서 다음과 같이 'tar -xf data_220615.tar' 명령으로 압축을 해제합니다.

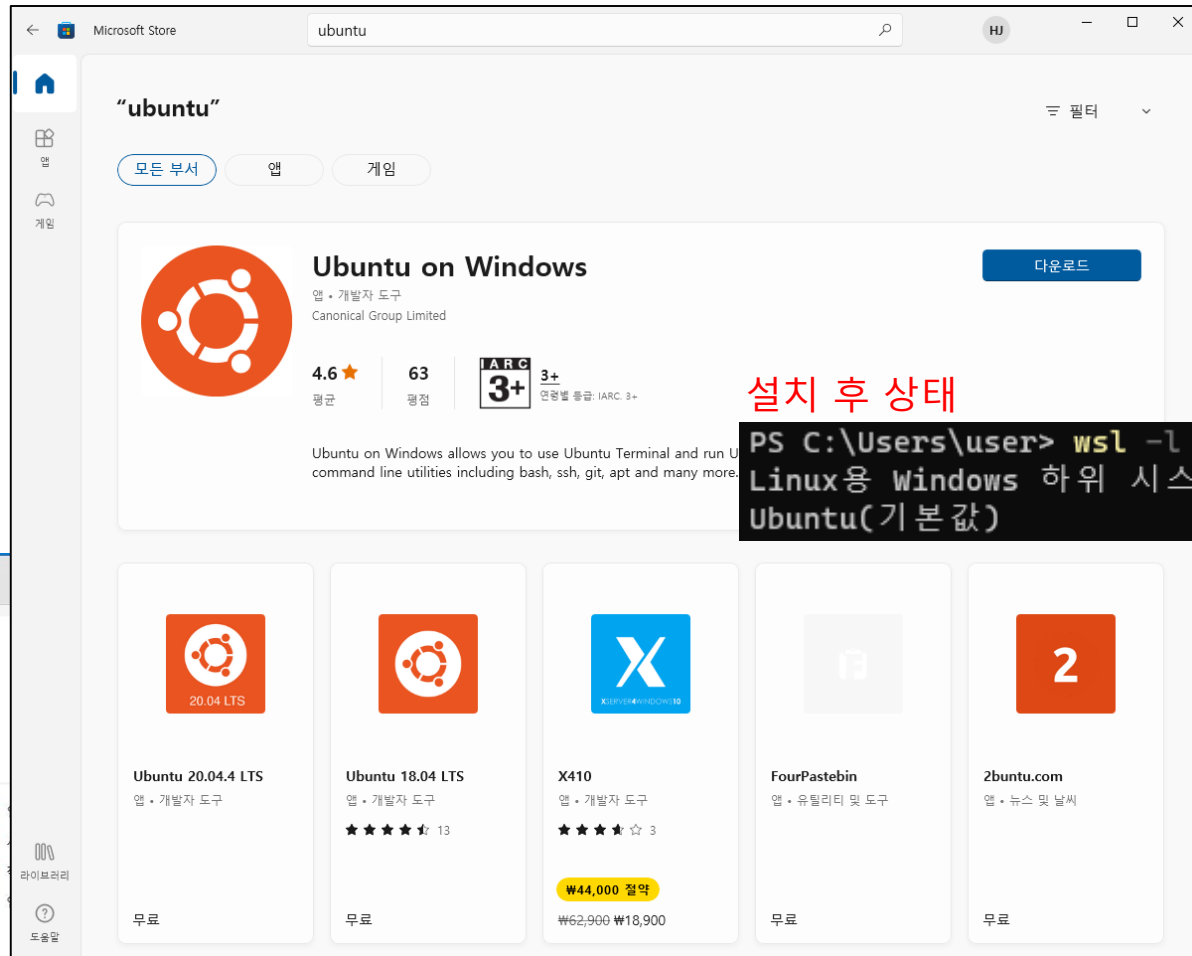


```
kribb@user-PC: ~  
PS C:\Users\user> wsl --cd ~  
kribb@user-PC:~$ ls  
bin data_220615.tar miniconda3  
kribb@user-PC:~$ tar -xf data_220615.tar  
kribb@user-PC:~$ ls data  
01_sr_assembly SRR1034787_1.fastq.gz md5sum.txt  
02_lr_pb_unicycler_assembly SRR1034787_2.fastq.gz nanopore.filtlong.fq.gz  
03_lr_pb_zga_assembly blast-2.2.26 pacbio.fastq.gz  
04_lr_filtlong_zga_assembly blast-2.2.26-x64-linux.tar.gz readstat.txt  
05a_lr_filtlong_canu_assembly broken-st_2.fq.gz sample_1.fq.gz  
06a_prokka ecoli_p6_25x.filtered.fastq.gz sample_2.fq.gz  
07_BV_GB03 ecolk12mg1655_R10_3_guppy_345_HAC.fastq.gz singletons.fq.gz  
08_16S_rRNA fixed1.fq.gz st_1.fq.gz  
09_ANI fixed2.fq.gz st_2.fq.gz  
10_file_manipulation md5sum.2.txt tmp  
kribb@user-PC:~$ |
```

용량을 줄이기 위해 FASTQ(.fastq, .fq) 파일은 gzip으로 압축하였습니다.

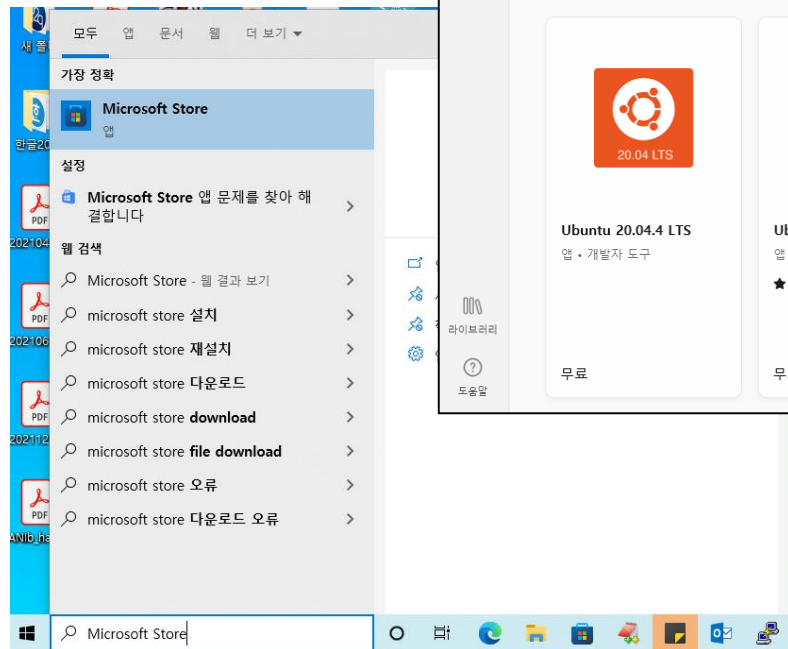
[Optional] 배포를 직접 설치해 보기

앞서 보인 작업은 tar file에 대한 import이고, 여기서 설명한 것은 install입니다.



설치 후 상태

```
PS C:\Users\user> wsl -l
Linux용 Windows 하위 시스템 배포:
Ubuntu(기본값)
```

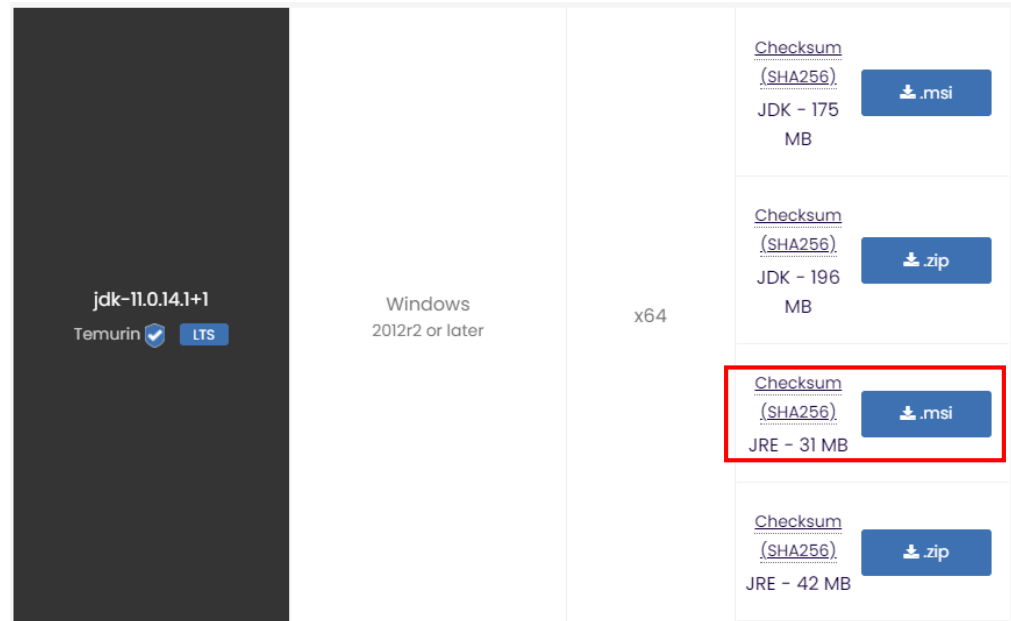
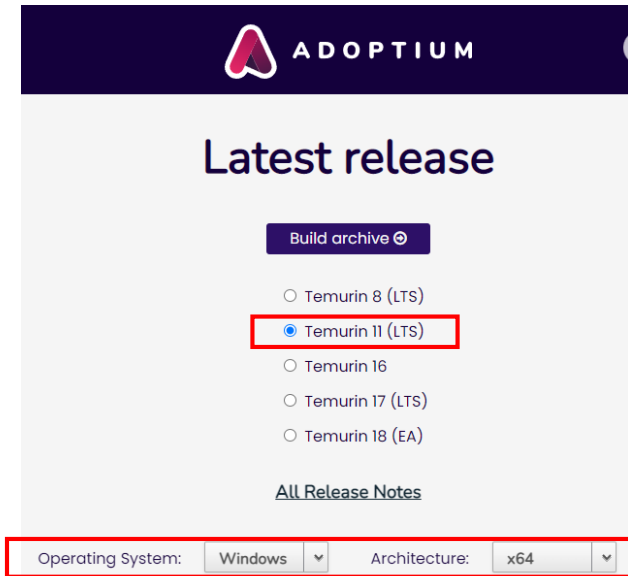


Windows Terminal에서 설치하는 것도 가능합니다.

```
> wsl --list --online
> wsl --install -d [distro]
```

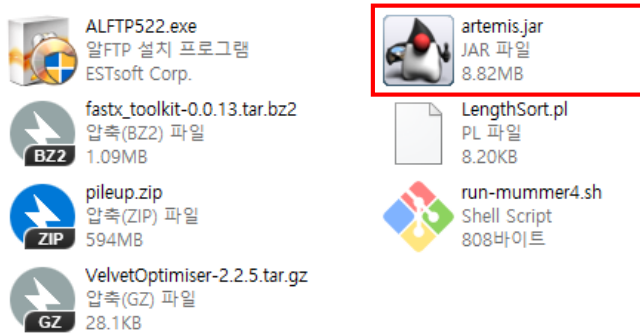
Java 실행 환경 설치

- 접속 URL: <https://adoptium.net/releases.html?variant=openjdk11>
- Version 11의 JRE Windows installer file을 다운로드하여 설치합니다.

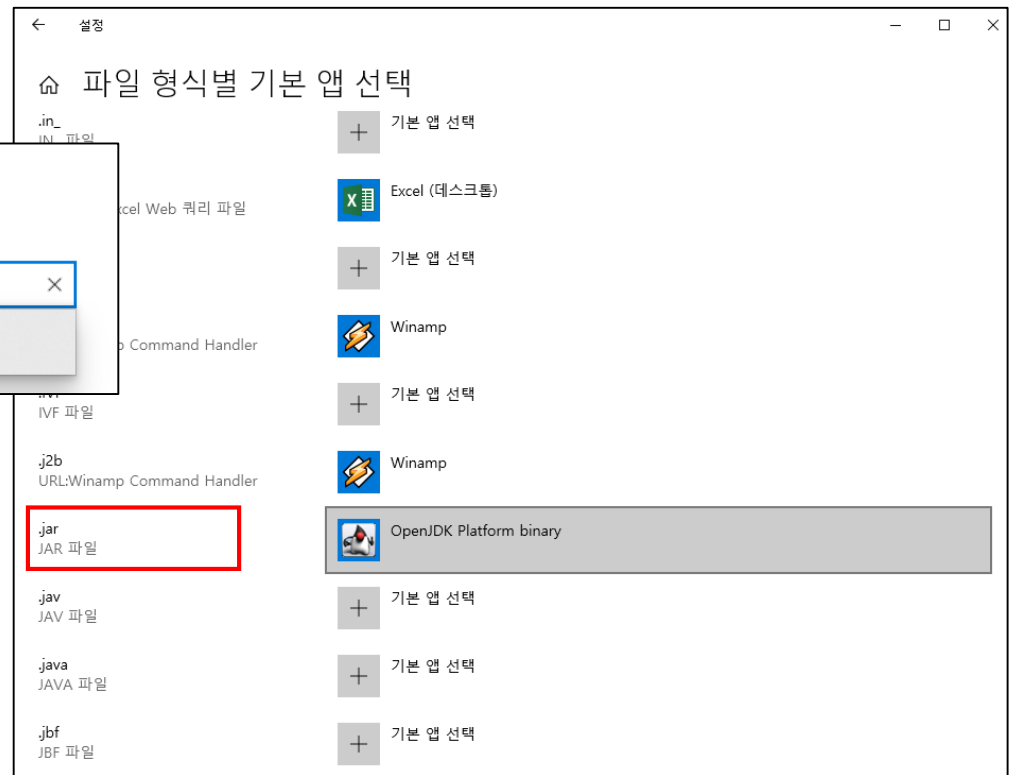
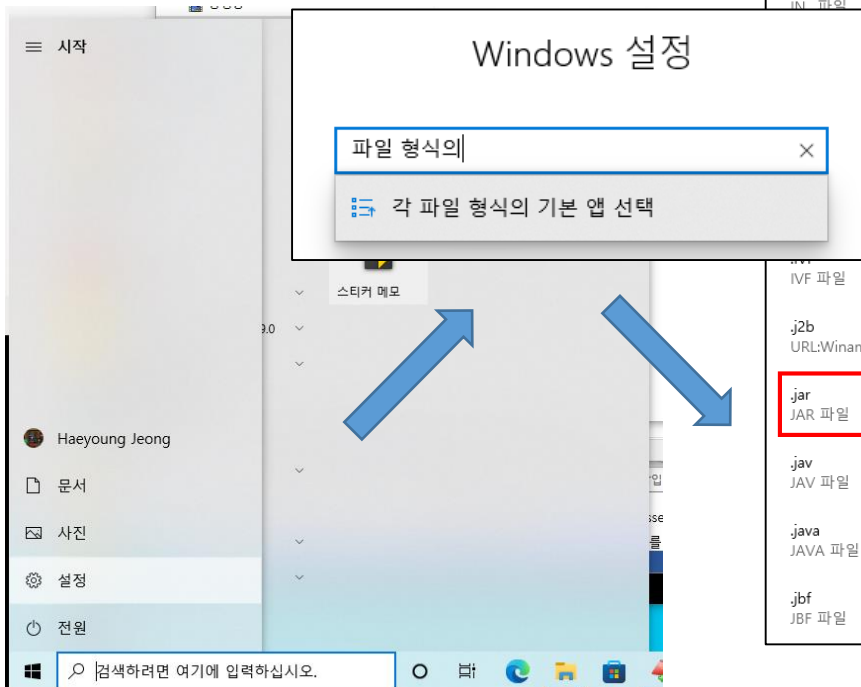


JDK: Java Developmental Kit
JRE: Java Runtime Environment

알집 쓰시는 분들은 주의하십시오



- .jar 파일을 압축 파일로 인식해서는 안됩니다.

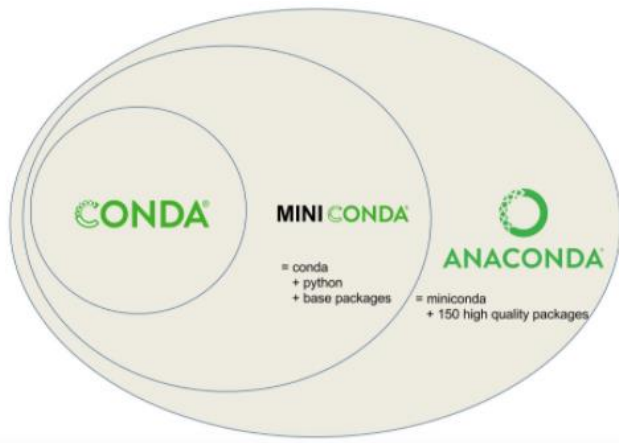


사전 설치할 Java 응용프로그램

- **Artemis** <http://sanger-pathogens.github.io/Artemis/Artemis/>
 - <https://github.com/sanger-pathogens/Artemis/releases/download/v18.2.0/artemis-windows-release-18.2.0.zip>
- **Gepard** <https://cube.univie.ac.at/gepard>
 - <https://github.com/univieCUBE/gepard/blob/master/dist/Gepard-2.1.jar?raw=true>
- **FigTree** <http://tree.bio.ed.ac.uk/software/figtree/>
 - <https://github.com/rambaut/figtree/releases/download/v1.4.4/FigTree.v1.4.4.zip>
- 압축을 풀고 .jar 파일을 더블 클릭하면 실행

중요 개념

- **Anaconda:** 패키지 관리와 디플로이를 단순케 할 목적으로 과학 계산을 위한 파이썬과 R 프로그래밍 언어의 자유-오픈소스 배포판[위키백과 설명]; 유료 및 무료 버전이 있음
- **Conda:** Windows, macOS 및 Linux에서 작동하는 오픈소스 패키지 및 관리 시스템(단일 명령어)
- **Miniconda:** Conda를 위한 최소한의 설치파일(minimal installer)
- **Bioconda:** 생명정보학 소프트웨어에 특화된 conda 패키지 관리자 위한 채널



```
PATH=/home/kribb/bin:$PATH

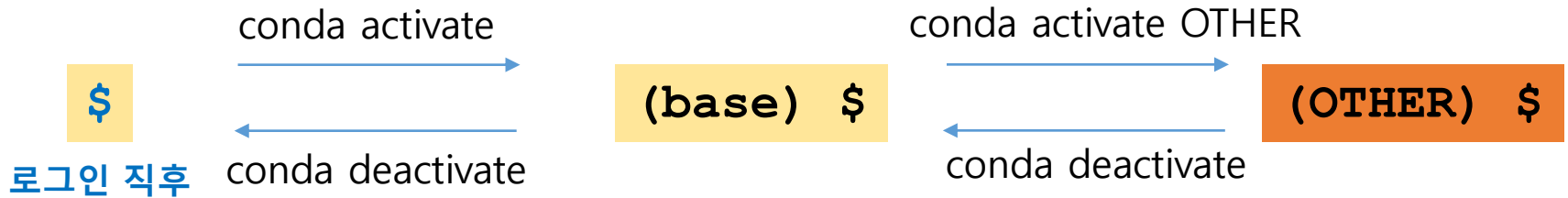
# >>> conda initialize >>>
# !! Contents within this block are managed by 'conda init' !!
__conda_setup="$('/home/kribb/miniconda3/bin/conda' 'shell.bash' 'hook' 2> /dev/null)"
if [ $? -eq 0 ]; then
    eval "$__conda_setup"
else
    if [ -f "/home/kribb/miniconda3/etc/profile.d/conda.sh" ]; then
        . "/home/kribb/miniconda3/etc/profile.d/conda.sh"
    else
        export PATH="/home/kribb/miniconda3/bin:$PATH"
    fi
fi
unset __conda_setup
# <<< conda initialize <<<

export REQUESTS_CA_BUNDLE=/etc/ssl/certs/ca-certificates.crt
export CURL_CA_BUNDLE=/etc/ssl/certs/ca-certificates.crt
```

<https://bioconda.github.io/tutorials/gcb2020.html>

Conda 실행을 위해 .bashrc 스크립트에 추가한 부분

Conda 환경 간의 순차적 전환



~/.condarc 설정 파일의 'auto_activate_base: false' 라인을 지우면 로그인 후 base 환경으로 곧장 들어감

Environment	설치된 주요 conda package ('conda list'로 확인)
Base (python 3.8.12)	biopython (1.77) ncbi-genome-download fastani sra-tools filtlong fasttree trimal mafft mummer pyani('average_nucleotide_identity.py') merge-gbk-records zga의 dependency: [fastp spades unicycler checkm-genome dfast bbmap blast nxtrim mash flye minimap2 racon samtools emboss]
System	blast-2.2.26(legacy) 수동 설치 필요

- 빨간색으로 표시한 것은 특별히 중요한 범용 프로그램
- 일부 파이썬 프로그램은 base environment에서 pip로 설치