

2025 한국 식물분류학회 Technical Workshop Plant organelle genome assembly (두 번째)

순서

- 1. Code (4 pages)**
- 2. polap assembly 방법 (3 pages)**
- 3. Test (1 page)**

실습 방법:

<http://galaxy.sungshin.ac.kr:8787/>

```
source ~/miniconda3/bin/activate
```

```
bash polap/src/bolap setup conda
```

2. Install Bioconda packages: Install conda packages including polap and others in their own conda environments. We execute `bolap -y install polap` if we want the latest version of polap.

```
# option 1: to install all tools necessary
```

```
bash polap/src/bolap -y install all=0.4.3
```

```
bash polap/src/bolap setup polap
```

After the Bioconda package installation, log out and back into the terminal.

3. Benchamrk to reproduce the tables and figures in the manuscript: We create the report in PDF format.

```
cd
```

```
cp -pr /media/h1/share/5_H_hon 5_H_hon
```

```
cd 5_H_hon
```

```
mv o o2
```

```
git clone https://github.com/goshng/polap.git
```

```
cd polap
```

```
git lfs pull
```

Code

Polap (Plant Organelle Long-read Assembly Pipeline v0.5.3) is available at <http://github.com/goshng/polap>. A guide of the homology-based plastome assembly is provided for use on a Linux system with an Internet connection. It offers a step-by-step description of the procedures used for the results presented in the manuscript.

Requirements

- **Operating System:** Linux with Ubuntu 24.04 LTS (not compatible with macOS or Windows)
- **Dependencies:** Requires [Bash](#) (≥ 5.0) and [Miniconda](#)

Steps

To replicate the results presented in this manuscript on a Linux computer with `git` installed and an Internet connection, follow the steps below.

1. **Install Miniconda:** Open a new terminal in a Linux computer, such as one with Ubuntu. Make sure that the `Bash` version is at least 5.0.

```
bash --version
```

```
lsb_release -a
```

The expected screen output includes the following two lines.

```
GNU bash, version 5.2.21(1)-release (x86_64-pc-linux-gnu)
```

```
Description: Ubuntu 24.04 LTS
```

Use the following script to download and install **Miniconda**. It uses the [instructions](#) for the Miniconda installation available at <https://docs.anaconda.com/miniconda/#quick-command-line-install>. Use either `git` or `wget` (along with `unzip` and `mv`) to prepare a folder named `polap` at the current directory. Replace, if necessary, the version number 0.4.3.7.9 with something recommended at [Polap's github website](#).

```
# option 1: use git command
```

```
git clone https://github.com/goshng/polap.git
```

```
# Now, we have the polap folder.
```

```
bash polap/src/bolap.sh -y install conda
```

After installing, close and reopen the terminal application to log out and back into the terminal. Then, execute the followings to setup the conda channels for `polap`.

```
cd -

source ~/miniconda3/bin/activate

mv 11_all.fq 1.fastq

gunzip H-hon.R1.fq.gz

gunzip H-hon.R2.fq.gz

mv H-hon.R1.fq s_1.fastq

mv H-hon.R2.fq s_2.fastq

ls


bash polap/src/bolap config add Hemerocallis_hongdoensis 1 s

bash polap/src/bolap benchmark Hemerocallis_hongdoensis


conda activate polap

mv o o3

cp -pr Hemerocallis_hongdoensis/v5/0/polap-assemble-wga o

polap annotate view


cd

cd 6_D_pan

cd NEW

polap annotate view
```

polap assembly 방법

폴더에 l.fq, s1.fq, s2.fq 가 존재해야함 (l.fq: long read, s1.fq: short read foward, s2.fq: short read reverse)

\$ polap assemble1

o/0/30-contigger/graph_final.gfa 파일이 whole genome assembly 결과를 bandage로 볼 수 있는 파일

o/0/Contig-annotation-depth-table.txt 파일에 MT seed contig 결과가 존재

Contig	Length	Depth	Copy	MT	PT	Edge
edge_2782	207395	8	3	23	2	2782
edge_1716	176304	6	2	18	1	1716
edge_2781	87169	9	3	16	6	2781
edge_1717	111432	9	3	8	4	1717
edge_1617	91320	3	1	1	0	1617
edge_2289	61312	4	1	1	0	2289

o/0/pt-contig-annotation-depth-table.txt 파일에 PT seed contig 결과가 존재

Contig	Length	Depth	Copy	MT	PT	Edge
edge_2786	74429	254	85	22	37	2786
edge_1761	12042	267	89	10	12	1761
edge_1762	20081	502	167	7	13	1672
edge_1763	18740	245	82	7	11	1763
edge_2661	32441	4	1	1	4	2661
edge_178	8612	4	1	0	1	178
edge_372	23722	3	1	0	2	372
edge_931	38973	3	1	0	1	931
edge_1124	30347	3	1	0	2	1124
edge_1273	6298	9	3	0	1	1273
edge_1297	40007	3	1	0	1	1297
edge_1405	41053	9	3	0	1	1405
edge_1594	28395	3	1	0	1	1594
edge_1754	10793	4	1	0	1	1754
edge_1760	10967	2	1	0	3	1760

edge_1832	44579	4	1	0	1	1832
edge_1923	48987	3	1	0	1	1923
edge_2055	12753	3	1	0	2	2055
edge_2070	30296	3	1	0	1	2070
edge_2787	5834	459	153	0	2	2787

\$ polap annotate view table (contig-annotation-depth-table.txt를 프롬프트 상에서 볼 수 있는 명령어)

bandage를 통해 위의 contig들이 어디에 존재하는지 확인

\$ polap seeds bandage

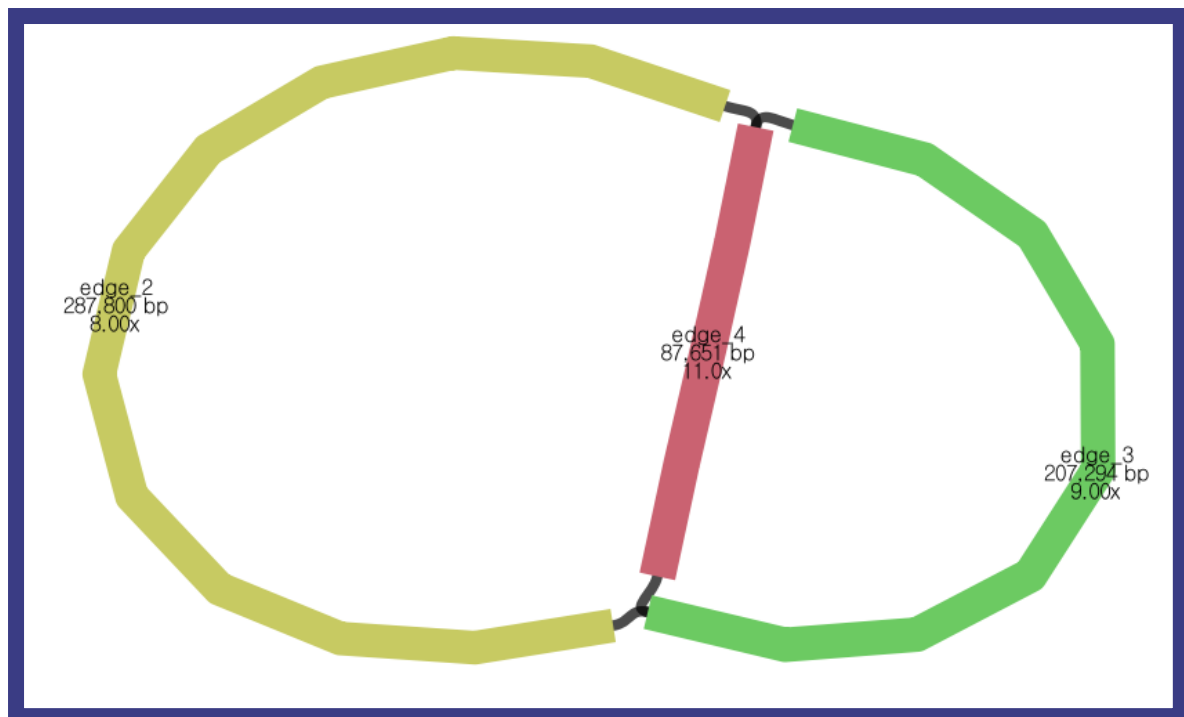
edge_1716, edge_1717, edge_2782, edge_2781

위의 edge를 선택

\$ polap assemble2 -m 6000

polap assemble2는 선택한 edge를 바탕으로 organelle assembly를 하는 명령어.

assembly 결과 gfa 파일을 bandage에 넣어서 확인 가능



Test

아래 내용을 순서대로 따라한다.

```
cd
```

```
mkdir test1
```

```
cd test1
```

```
cp /media/media/h1/share/anthoceros_angustus.tar.gz .
```

```
tar xzf anthoceros_angustus.tar.gz
```

```
git clone https://github.com/goshng/polap.git
```

```
cd polap
```

```
git lfs pull
```

```
cd ..
```

```
bash polap/src/bolap config add Anthoceros_angustus anthoceros_angustus_long  
anthoceros_angustus_short
```

```
bash polap/src/bolap benchmark Anthoceros_angustus
```