

Barcoding & Demultiplexing

Seong-Kun Bak (sanekun@gmail.com)

2022 7 11

Introduce

For high-throughput colony genotyping, we conducted multiplexed sequencing with nanopore long-read sequencing. In this page, we subscribed 'how make DNA barcode sequence', 'Detailed method of multiplexed colony PCR', 'Barcode demultiplexing from sequencing data'.

Barcode primer generation

바코드 프라이머는 R bioconductor의 DNABarcodes 패키지를 사용하여 제작 되었다. 각 바코드는 서로 겹쳐 검출되지 않도록 Forward barcode 8개, Reverse barcode 12개로 총 20개의 barcode를 생성 하였으며 이는 Primer binding site (M13F, p15A_ori)와 연결되어 하나의 Primer로 제작 하였다.

사용전 우리는 colony PCR과 전기영동으로 해당 바코드가 정상적으로 PCR을 수행함을 확인 하였다.

[Primer Table]

Multiplexed colony PCR

1. 96 well plate에 row, column 별로 서로 다른 forward, reverse primer 와 PCR premix 분주.
 - 해당 Polymerase가 상대적으로 error rate가 높지만, 많은 read가 생성되어 part를 구분하는 정도는 문제가 없다.
2. Phenotyping으로 colony intensity가 계산 되고 Index가 부여된 colony를 pick 하여 넣기.
3. Bioneer PCR premix를 사용해서 PCR 진행 [Polymerase cycle table]
4. 최대 96개의 PCR product를 동일한 volume 별로 하나의 튜브로 모음.
5. ONT ligation sequencing kit(ONT#SQK-LSK109)와 barcoding kit(ONT#SQK-NBD104)를 사용하여 추가적인 ONT barcode 부착 및 sequencing library 제작.
6. RUN Sequencing.

Computational analysis (Barcode demultiplexing)

1. 모든 raw read는 guppy basecaller에 의해 ONT barcode가 분석됨.
2. 이후 Method I 의 pre-processing 과정에 의해 전처리됨.
3. 전처리 된 Read 들은 /script/tag_demultiplexing.py 를 사용하여 demultiplexing 되어 각각의 fastq file로 저장됨.
4. 각 fastq file은 minimap2를 사용해 조합 가능한 모든 library 서열(generated by /script/DB.py)에 mapping한 후 samtools와 awk을 사용해 reference 서열을 분석함.
 - 자세한 방법은 M2-single_molecule analysis 참고
5. 각 barcode(ONT, forward, reverse)별 score1, score2를 계산하여 서열을 확정 지음.
6. 해당 정보는 미리 index가 결정된 phenotyping 정보와 연결되어 함께 계산됨.