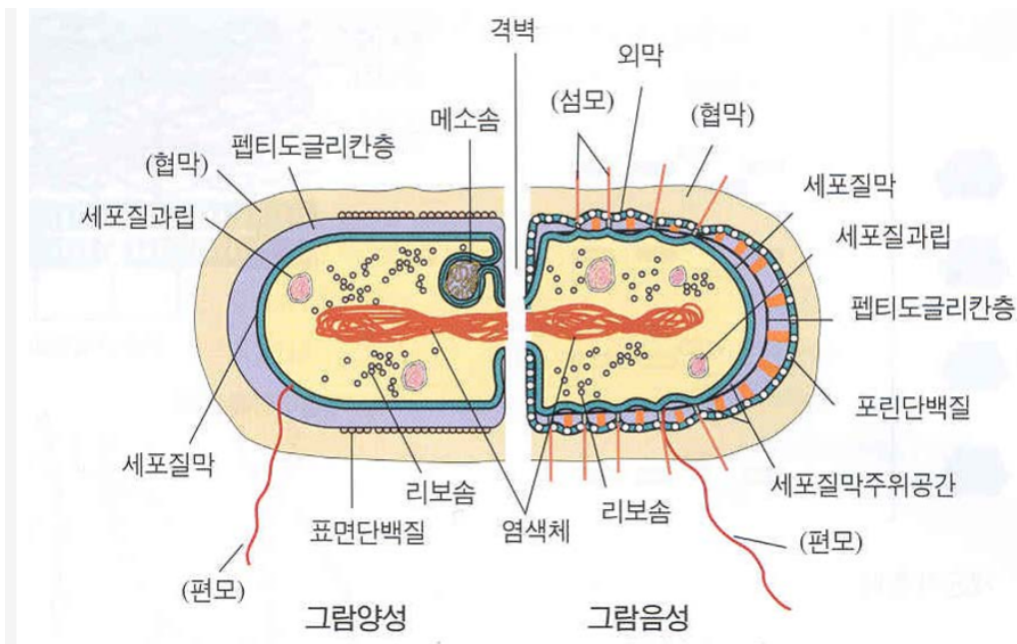


Bacteria mRNA-seq - NGS 기술을 통한 Bacterial Transcriptome

미생물(microorganism)은 흔히 육안으로 관찰할 수 없는 작은 생물, 즉 현미경으로 볼 수 있는 크기의 생물을 의미한다. 미생물은 Bacteria, Fungi, Virus 등을 포함하고 있는데, 이는 유전적 특성을 고려하지 않은 단순 크기의 개념으로만 정의한다. 미생물은 지구상에서 가장 오랫동안 진화해 온 생물로 생명 현상의 원리 및 생물의 다양성, 진화를 이해하는 기초가 되고 있으며, 계속해서 활발한 연구가 진행중이다. 그 중에서도 Bacteria 는 지구상에 2 번째로 번성한 생물군으로 병원성 세균 분야 이외에 산업 목적으로 많이 이용되고 있으며, 여기서는 Bacteria 유전적 특성에 따른 연구의 동향에 대해 소개하고자 한다.

Bacteria 는 단세포의 미생물로 대부분의 원핵생물을 뜻한다. 하지만 분류체계가 변화함에 따라 고세균 (Archaea) 과의 구분을 위해 Eubacteria, 진정세균으로 분류할 수 있으며, 여기서 후술할 Bacteria 는 모두 Eubacteria 를 의미한다. Bacteria 는 외막의 유무, 그리고 세포벽의 펩티도글리칸 구조 형태에 따라 그람 양성균 (Gram positive) 과 그람 음성균 (Gram negative)으로 구분 짓게 되는데, 이러한 외막의 특성 때문에 그람 음성균은 (Gram negative) 병원성을 특이적으로 갖는다. 외막의 일부 지질이 Lipopolysaccharide (LPS) 복합체를 포함하기 때문에, 내독소로 작용한 독성 반응을 일으킬 수 있으며, 면역 반응에 굉장히 중요하게 작용할 수 있다.



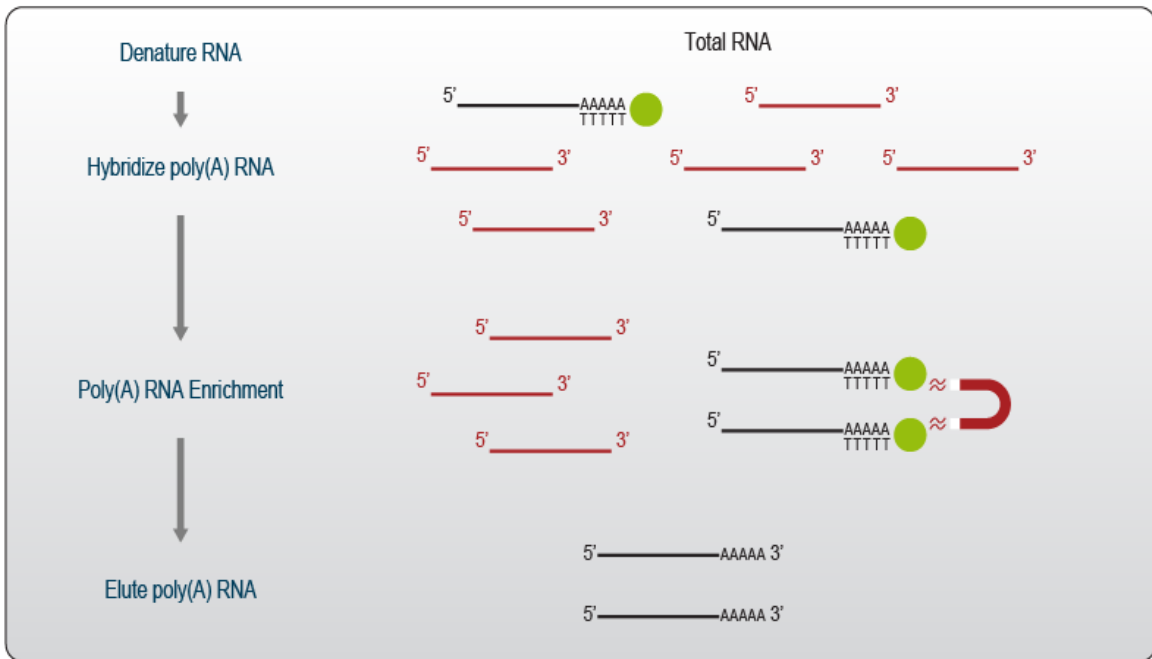
그래서 인간은 Bacteria 의 특성을 여러 방면에 활용하고 있다. 예컨대 유산균 (Lactic Acid Bacteria)을 이용하여 오래전부터 치즈나 요구르트 등 발효식품을 만들어왔고, 최근에는 첨단 생명공학적인 기술을 접목하고 있다. 대장균 (*Escherichia coli*) 은 유전자 복제에 많이 이용되며, 아그로박테리움 (*Agrobacterium tumefaciens*) 은 형질전환 식물을 만드는데 이용한다. 현재까지 약 100 종 가량의 Bacteria 가 형질 전환이 가능한 것으로 알려져 있다. 또한 directed mutagenesis, knockout, RNAi, CRISPR-Cas9 등을 이용한 유전자 조작을 통해서 병원성 유전자를 찾을 뿐 아니라 상업적으로 항생제와 같은 물질들의 대량생산, 혹은 기타 유용한 물질들을 얻기도 한다. 산업적으로는 유기물 분해를 통한 폐기물 처리 혹은 구리와 같은 금속을 채취하는 과정에서도 사용한다. 이러한 효율적인 활용방법은 유전체를 고속으로 분석하여 그 기능을 파악하게 되면서 가속화되었는데, 그것이 바로 NGS (Next Generation Sequencing) 이다.

인류 역사 최초로 헤모필루스 인플루엔자 (*Haemophilus influenzae*) 유전체 정보가 완전히 밝혀진 이래로 1 세대 시퀀싱을 거쳐 NGS 방식이 상용화되면서 많은 Bacteria 의 염기서열을 밝혀냈고, 유전체 정보를 구축하였다. 진핵생물과 같이 획득된 유전체 정보를 기반으로 각 유전자의 기능을 밝히거나, 유전자 발현 양상과 유전자에서 생산되는 단백질의 기능을 분석하여야 했지만, Bacteria 는 진핵생물과 조금 다른 특징을 갖기 때문에 쉽지 않았다. polyadenylation 은 진핵생물에서 mRNA 의 안정성에 기인하는 중요한 특징이지만, Bacteria 에서는 RNA 분해를 촉진한다. 더욱이 일반적으로 더 짧고 mRNA 분자 수가 적기 때문에, transcriptome 분석 시도 자체가 어려웠다. 따라서 효율적인 분석을 위해서는 ribosomal RNA (rRNA)를 제거하여 순도 높은 Bacteria mRNA 로 NGS 를 수행해야 한다.

rRNA Depletion - Bacteria mRNA-seq 의 핵심 Key point

rRNA 는 ribosome 을 구성하는 RNA 로, total RNA 내에서 약 80~98% 정도를 차지한다. 그러한 이유로 transcriptome 분석을 진행하고자 할 때, rRNA 는 시퀀싱 전체 데이터의 많은 공간을 차지하게 되고 결과적으로 분석하고자 하는 mRNA 분석의 효율성을 높이기 위해서는 rRNA 를 효과적으로 제거해야 한다. rRNA 제거는 아래와 같이 여러가지 방법을 사용할 수 있다.

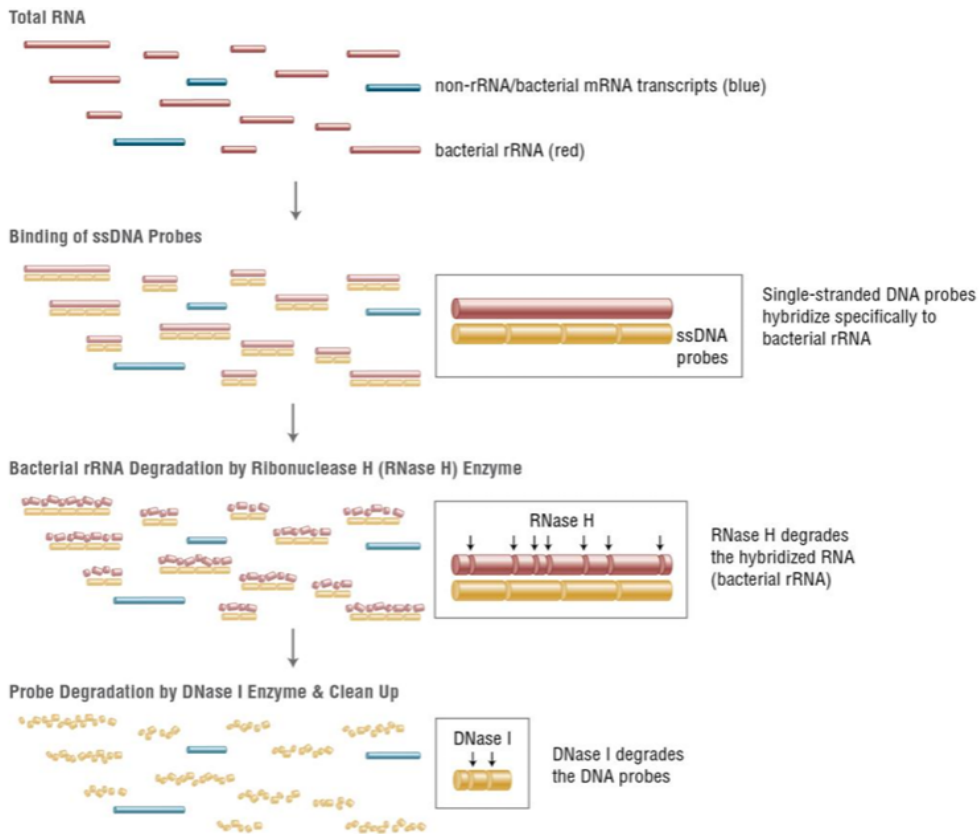
1. Poly(A) Enrichment / Capture



[1]

가장 강력한 rRNA 제거는 poly(A) enrichment를 통한 capture 방법이다. 정확하게는 poly(A)를 가진 mRNA를 선택적으로 분리하는 방법이지만, 하나의 전처리 과정으로써 간단하게 rRNA를 제거하는 것이라고 볼 수 있다. 이 방법은 Total RNA에서 polyadenylated 된 mRNA 만을 oligo(dT) magnetic bead에 혼성화시켜 끌어당긴다. 하지만 위에 언급한 것처럼, Bacterial mRNA는 진핵생물에 비해 poly(A) 길이가 짧고 RNA 분해 요소로 작용하며, 심지어는 polyadenylation 효소가 없어 poly(A) 구조를 가지지 않는 종들도 있다. 때문에 Bacteria transcriptome 분석에서는 poly(A) capture를 통한 rRNA 제거가 불가능하며, 결론적으로 NGS 효율성을 높이기 어렵다.

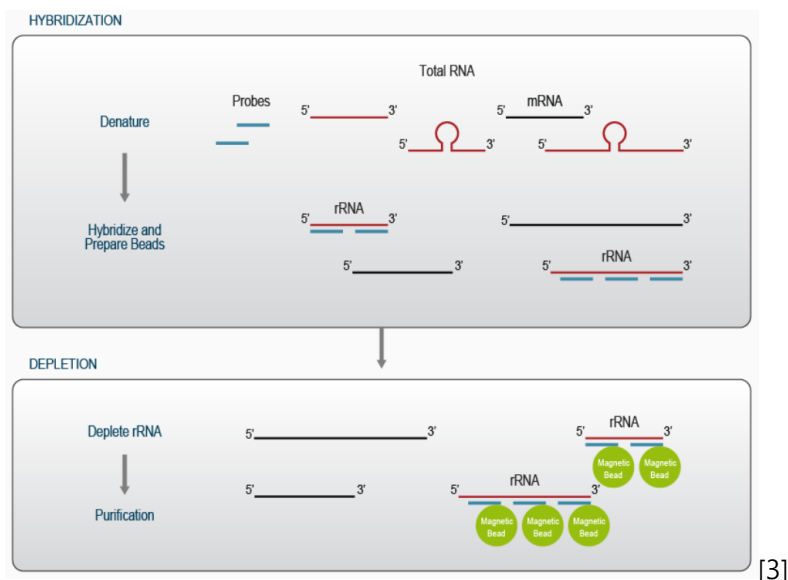
2. Enzymatic Removal of rRNA



[2]

Enzymatic Removal 방법은 말 그대로 Enzyme 을 이용하여 특정 RNA 에 결합하는 Enzyme 을 통해 rRNA 를 제거한다. RNase H, RNase III, RNase R 등을 주로 이용하며, 특히 RNA 와 DNA 복합체를 인식하여 특이적으로 분해할 수 있는 RNase H를 많이 활용하고 있다. Bacterial rRNA 에 특이적인 ssDNA probe 를 혼성화시키고 RNase H 를 처리하게 되면 혼성화된 RNA 는 분해되고 분석하고자 하는 mRNA transcript 는 남게 된다. 이후 남아있는 DNA probe 들에 대해 DNase I 를 처리하게 되면, Bacterial mRNA transcript 들만 enrichment 할 수 있게 된다. Enzyme 처리를 통해 비교적 간단하게 rRNA 를 특이적으로 제거할 수 있지만, 일부 Enzyme 이 특정 샘플이나 실험 조건에서 민감하게 반응할 수 있다.

3. Bead based Removal of rRNA



[3]

Bead based Removal 방법은 Enzyme Removal 방법과 비슷하다. 제거하고자 하는 rRNA 와 특이적인 probe 를 혼성화시키는 과정까지는 동일하지만, 이후에 Enzyme 을 사용하지 않고 probe 에 magnetic bead 가 결합할 수 있는 tag 를 포함하여 혼성화된 rRNA 를 capture 한다. 분해를 통한 rRNA 제거가 아닌, magnetic bead 와 결합한 rRNA capture 를 통해서 Bacterial mRNA transcript 들을 회수하여 enrichment 한다. Enzyme 반응에 의존하지 않는 방법으로 시퀀싱 등 여타 다른 실험에서 분석하고자 하는 Bacterial mRNA transcript 들의 분해를 최소화할 수 있다.

RiboCop rRNA Depletion Kit for Bacteria

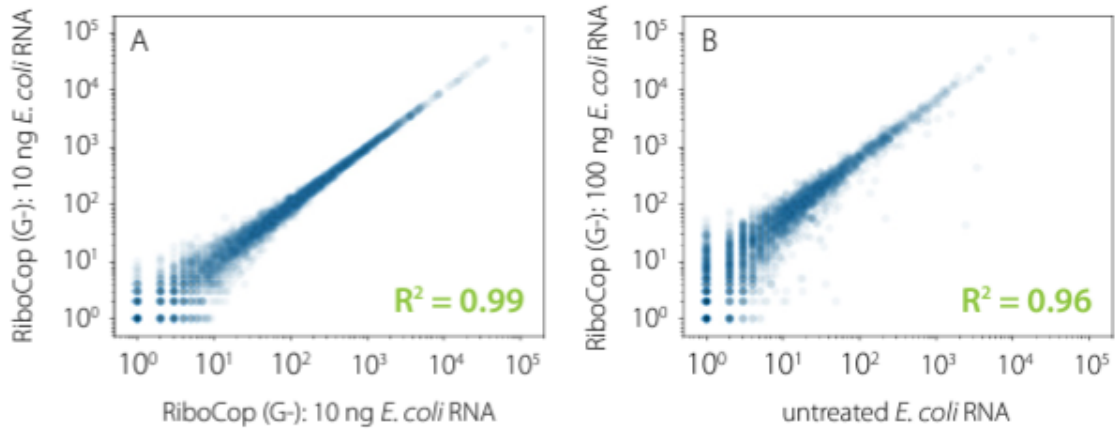
Lexogen 의 RiboCop rRNA Depletion Kit for Bacteria 는 bead based rRNA depletion 방법을 채택하여 사용하고 있다. RiboCop kit for Bacteria 는 23S, 16S, 5S rRNA 를 효과적으로 제거하기 위해 총 3 가지의 probe mix set 를 구성하였고, 아래와 같이 제공하고 있다.

- RiboCop rRNA Depletion Kit for Gram (-) Bacteria (Lexogen, 126)
- RiboCop rRNA Depletion Kit for Gram (+) Bacteria (Lexogen, 127)
- RiboCop rRNA Depletion Kit for META Bacteria (Lexogen, 125)

rRNA 를 제거하기 위한 probe 를 디자인하기 위해서는, 각 bacteria 특성에 맞는 특이적인 서열에 대한 정보를 토대로 해야 한다. RiboCop kit for Bacteria 는 이러한 효율을 극대화하기 위해 그람음성균과 그람양성균의 probe mix 를 구분하였고, environmental 혹은 microbiome 샘플과 같은 혼합 시료에 적합한 META probe mix 도 선택적으로 사용 가능하게 했다. 물론 META probe mix 를 단일배양 Bacteria 시료에도 사용할 수도 있다. 그람 염색법으로 명확하게 구분할 수 없는 Bacteria 샘플이나 특이성을 갖는 Bacteria 샘플에 적용하기에 탁월하다. 아래 그림에서 META probe mix 를 통한 rRNA 제거 이후에 *E. coli* (그람음성균), *B. subtilis* (그람양성균) 간에 높은 재현성을 보이는 것을 알 수 있다.

rRNA	<i>E. coli</i>			<i>B. subtilis</i>		
	Total	G-	META	Total	G+	META
23S	64.2 %	0.6 %	1.1 %	55.4 %	0.5 %	0.4 %
16S	32.2 %	0.3 %	0.4 %	40.9 %	0.4%	0.4 %
5S	0.5 %	0.02 %	0.06 %	0.01 %	<0.01 %	<0.01 %
Overall	96.9 %	0.9 %	1.6 %	96.3 %	0.9 %	0.8 %

또한 RiboCop kit for Bacteria 는 Off-target 을 효과적으로 제거한다. 아래 그림들은 각 샘플 간에 transcript mapping 을 통해 얻어진 유전자 수의 상관관계를 나타내고 있다. 그림 A 를 보면, Gram (-) kit 로 동일한 *E. coli* 에, 동량의 input RNA 을 사용하였을 때, 결정계수 $R^2=0.99$ 의 높은 상관관계를 가지는 것을 볼 수 있다. 이것을 토대로 같은 probe mix 를 사용하였을 때의 재현성이 높은 것을 확인할 수 있다. 옆의 그림 B 를 보면, Gram (-) kit 로 rRNA 를 제거한 *E. coli* 와 rRNA 를 제거하지 않은 *E. coli* 의 유전자 데이터를 비교하였을 때, $R^2=0.96$ 의 비교적 높은 상관관계를 확인할 수 있다. 이것으로 RiboCop kit for Bacteria 의 rRNA 제거 효율이 Off-target 에 영향을 주지 않도록 설계되었다는 것으로 판단할 수 있다.



Bacteria mRNA-seq Process

Sample Preparation

Bacteria Transcriptome 분석을 하기 위해서는 원하는 strain 의 Bacteria 를 배양한 후, total RNA 를 추출한다. 이바ิโอ젠에서는 total RNA preparation 을 진행할 때 Maxwell RSC 48 System (Promega, AS8500)을 사용하거나, TRIzol Reagent (Invitrogen, 15596018) method 를 이용하고 있다.

Library Preparation

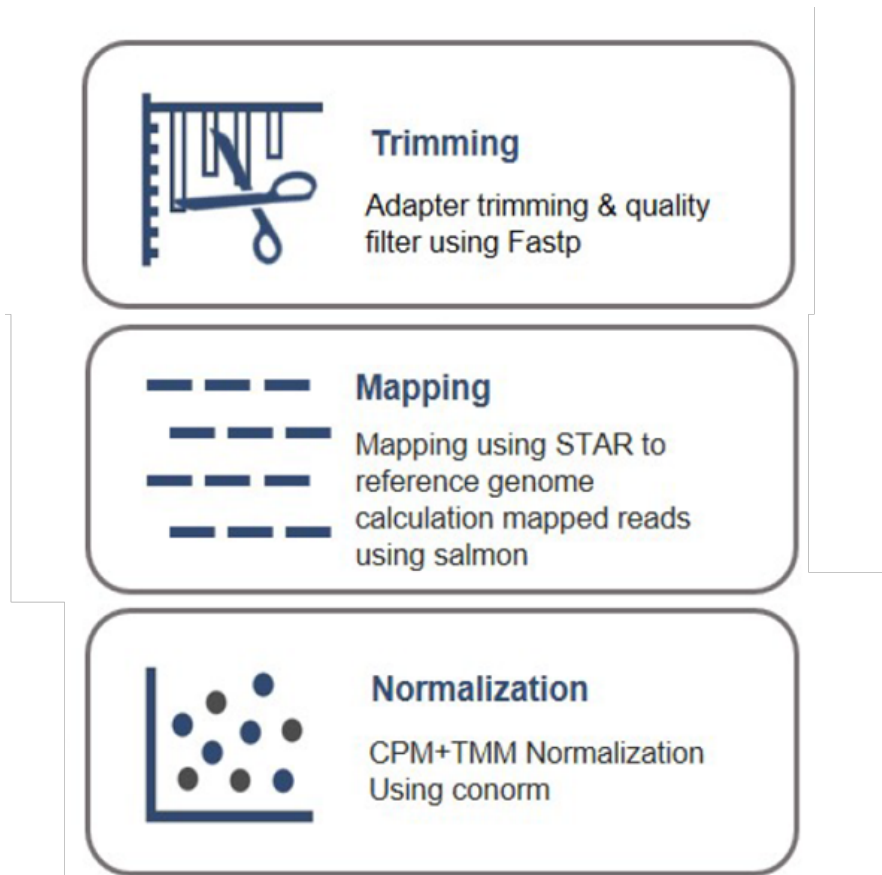
Bacterial total RNA 가 준비되었다면, 시퀀싱을 위한 library 를 구축한다. Lexogen 에서 제공하는 Selection Tool (<https://www.lexogen.com/support-tools/ribocop-for-bacteria-selection-tool/>)을 이용하면 Bacteria species 에 적용 가능한 RiboCop kit for Bacteria (META, G-, G+)를 확인할 수 있다. 이바ิโอ젠에서는 species 에 맞는 kit 를 이용하여 rRNA 제거를 한 Bacterial mRNA 를 CORALL RNA-seq Library Prep Kit (Lexogen, 176) 로 Illumina 시퀀싱에 적용 가능한 index 와 adapter 를 붙여 최종 library 를 준비한다.

Sequencing

구축된 library 는 NGS 방식으로 염기서열분석을 진행한다. 이바이오젠에서는 Illumina Novaseq 6000 플랫폼을 통해 paired end 100 base pair 를 읽어낸다. Bacteria 샘플 1 개 기준 총 3G base pair 정도의 데이터를 생산한다.

Data Analysis

시퀀싱이 완료된 raw data (FASTQ)는 적절한 trimming 을 거쳐 mapping, normalization 을 진행하게 된다. 이바이오젠에서는 아래와 같이 각각 Fastp, STAR, Salmon, Conorm 분석 방식을 이용하고 있다.



Normalization 이 완료된 발현분석 데이터는 이바이오젠 분석툴인 ExDEGA (Excel based Differentially Expressed Gene Analysis)를 통해 제공하며, DEG 분석 및 annotation 정보, alignment summary 등을 확인할 수 있다.

Our Performance of Technical Utilization

산업적 이용 사례 / Gram (-) RiboCop rRNA depletion

Sulfate-reducing bacteria(SRB)는 황산염 환원 세균으로, 미생물 식량사슬에서 굉장히 중요한 역할을 한다. 탄소, 질소, 황, 금속 등 자연 환경뿐만 아니라 토목 기반 시설과 같은 건축 환경의 부식에도 지대한 영향을 끼친다. 해당 논문에서는 실제로 미국이 이러한 문제를 해결하기 위해 약 40 억 달러 이상을 지출하고 있다고 소개했다. SRB 의 환경 스트레스에 대한 매커니즘을 분석하는 것이 산업적, 경제적 효과를 줄 수 있다는 것이다. 그림 1 은 이러한 매커니즘에 대한 4 종류의 SRB 분석을 유전적으로 접근하였다. DEG 분석을 통해 중요 전사 후보 유전자들을 예측하였고, Gene ontology 클러스터링을 통해 환경 적응 및 스트레스 반응에 대응하기 위한 대책연구를 마련하도록 제시했다.

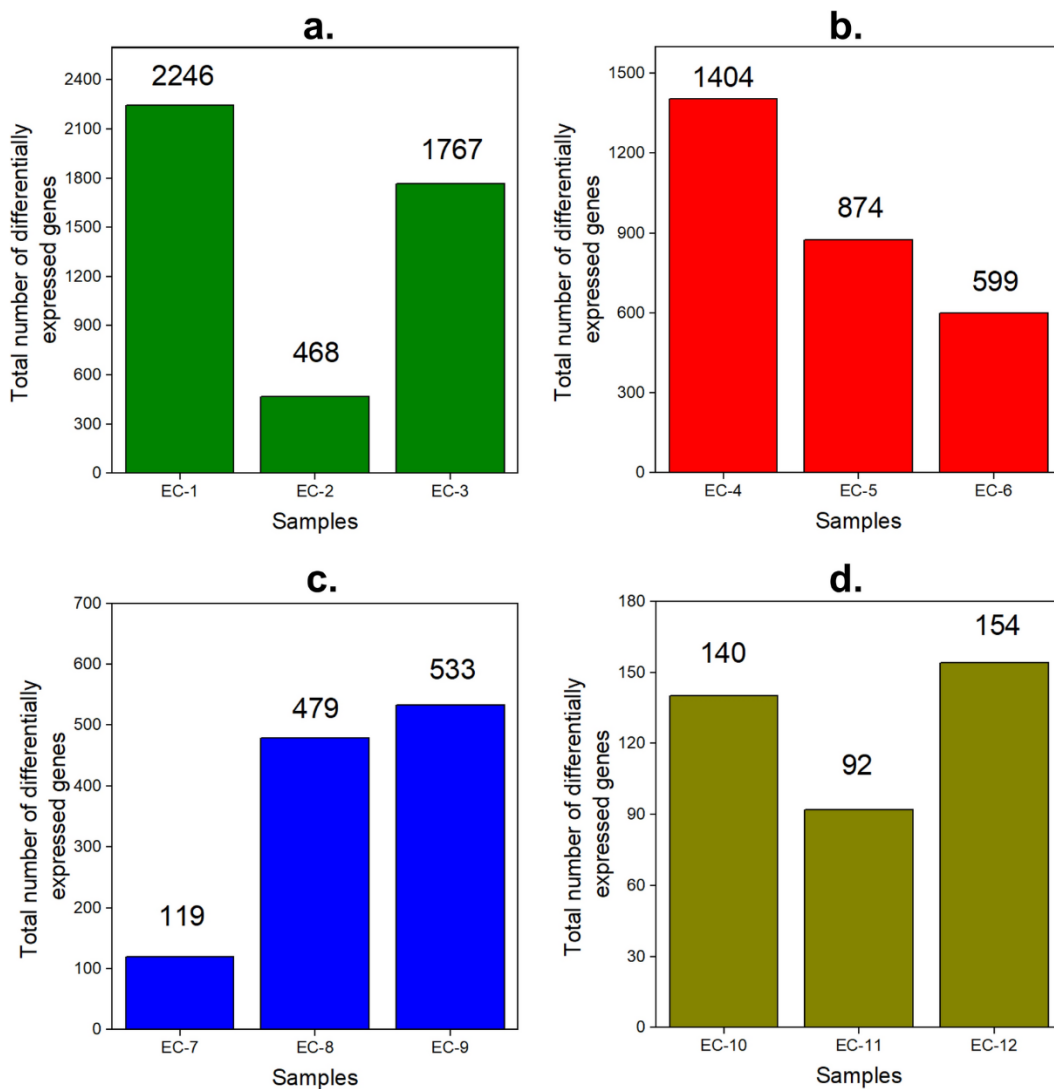


그림 1. Total number of DEGs of the RNA-seq datasets evaluated in this study. (a) OA-G20 exposed to copper stress (b) *M. hydrothermalis* exposed to hydrostatic stress (c) *P. piezophilus* exposed to hydrostatic stress and (d) *D. vulgaris* exposed to CuO stress. EC-1 to EC-12 corresponds to 12 different experimental stresses. [4]

산업적 이용 사례 / Gram (+) RiboCop rRNA depletion

엡실론폴리-L-라이신은 천연방부제로써 식품 및 화장품에 널리 사용되는 고가의 펩타이드 첨가제이다. 지금까지는 산업균주로 *Streptomyces albulus* M-Z18 를 이용해 주로 생산했는데, 지속적인 연구를 통해 고수율 산업균주 WG-608 을 개발, 이용하고 있다. 해당 논문에서는 M-Z18 과 WG-608 을 비교하여 엡실론폴리-L-라이신 생산의 매커니즘을 이해하기 위해 전사체 및 단백질체 분석을 진행하였다. DEG 분석을 통한 생합성 매커니즘을 규명함으로써 엡실론폴리-L-라이신 생합성을 촉진할 수 있는 ppk 유전자의 과발현과 polyP6 첨가를 통해 생산성을 증대하였다.

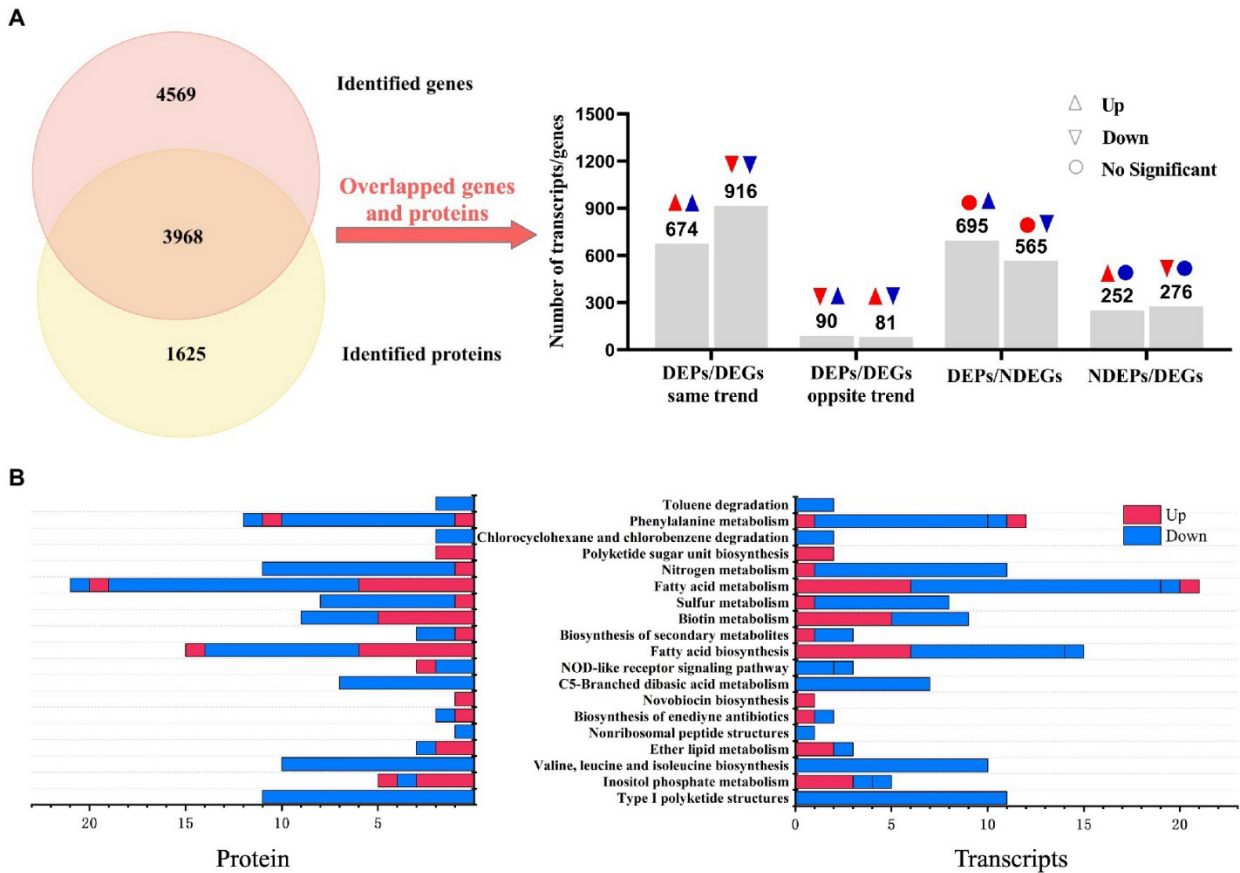


그림 2. Association analysis of genes and proteins between the high ϵ -PL-producing mutant *S. albulus* WG-608 and its progenitor strain *S. albulus* M-Z18. (A) Numbers of overlapping gene/protein pairs that show the same or opposite trends in expression, or that show differential expression only at the protein or mRNA level between WG-608 and M-Z18. The red indicates transcript, blue indicates protein. (B) KEGG enrichment analysis of overlapping gene/protein pairs between *S. albulus* M-Z18 and *S. albulus* WG-608. Red represents pathways enriched with upregulated genes/proteins; blue represents pathways enriched with downregulated genes/proteins.

[5]

E-biogen's Bacteria RNA-Seq Service

Sample Requirement	> 2ug total RNA , Bacteria Cell Pellet
Library Method	RiboCop rRNA Depletion kit (META, Gram(-), Gram(+)) -> CORALL RNA-seq Library Prep Kit
NGS Run Format	NovaSeq 6000, PE100bp
Data Yield	>~3Gb
Turnaround Time	4 weeks after RNA QC
Available species	NCBI DB에 등록되어 genome sequencing이 완료된 Bacteria

< 참고 문헌 >

1. N.A.S.Taufieq, et al. Isolation and Identification of *Desulfovibrio* sp. Bacteria from Acid Sulfate Soil, *Environmental Science, Biology*, 2015
2. Lexogen, Poly(a) Selection Kit Workflow
3. New England Biolabs, NEBNext rRNA Depletion Kit (Bacteria) Workflow
4. Lexogen, RiboCop rRNA Depletion Kit for Bacteria Technical Note
5. Kalimuthu Jawaharraj, et al. Transcriptome-wide marker gene expression analysis of stress-responsive sulfate-reducing bacteria, *Scientific Reports*, 2023
6. Wang L, et al. Integrative transcriptome and proteome revealed high-yielding mechanisms of epsilon-poly-L-lysine by *Streptomyces albulus*, *Frontiers in Microbiology*, 2023