

16S rRNA Metagenome Sequencing 분석 결과 보고서

㈜세니젠 기술연구소

2023-01-02

Sample/Customer Information

Sample Information

- 샘플 종류 : Metagenome
- 샘플 구성 : 36 Samples
- 분석 종류 : 16S rRNA metagenomic sequencing을 이용한 샘플 내 균총 분석

Customer Information

- 소 속 : 국민바이오
- 성 함 : 정예진 선생님

Analysis Overview



Sequencing

- ✓ 16S rRNA V3-V4 region PCR amplification
- ✓ Library construction (Nextera XT library prep kit)
- ✓ Sequencing (Illumina MiSeq 300 bp paired end)

Quality Control

- ✓ Check sequencing quality (FastQC-v.0.11.9)
- ✓ Check overrepresented sequences

Denoising Process

- ✓ Trimming low quality reads (Trimmomatic v0.39)
- ✓ Error correction of noisy reads (DADA2-Qiime2.2021.11)
- ✓ Remove chimeric sequences (DADA2-Qiime2.2021.11)

Taxonomy Classification

- ✓ Taxonomic sequence classification
(reference database : SILVA 138)
- ✓ Visualization

Diversity Analysis (Qiime2 – 2021.11)

- ✓ Alpha diversity
- ✓ Beta diversity

Analysis Report

Quality Control

Raw Data Information

Sample ID	Yield (bp)	Read Number	GC (%)	Q20 (%)	Q30 (%)
5-ASA-1	83,313,135	277,384	54.63	80.92	68.8
5-ASA-3	57,670,940	192,012	54.55	80.49	68.21
5-ASA-5	81,890,622	272,616	54.32	80.91	68.62
5-ASA-6	56,776,518	189,022	55.01	79.94	67.59
5-ASA-7	56,363,216	187,632	55.34	80.48	68.4
5-ASA-8	37,429,378	124,594	54.81	78.31	65.24
Con-1	58,421,168	194,552	55.68	79.62	67.4
Con-2-Re	183,323,767	611,320	54.85	76.81	63.08
Con-3	59,563,777	198,346	54.99	80.98	68.64
Con-6	65,362,578	217,672	55.14	79.54	66.92
Con-7	60,963,870	202,988	55.21	79.78	67.39
Con-8	42,578,312	141,816	55.29	77.34	64.15

Sequencing 결과 yield는 충분하게 확보되었습니다.

- Yield (bp) : Sequencing된 total base pair 수입니다.
- Read Number : Sequencing을 통해 생산된 total read 수입니다.
- GC ratio : 전체 염기서열 중 GC contents의 비율입니다.
- Q20: 1/100 이하의 확률 (99% 정확도)로 error가 생길 것으로 추정되는 base pair 비율입니다.
- Q30: 1/1,000 이하의 확률 (99.9% 정확도)로 error가 생길 것으로 추정되는 base pair 비율입니다.

Quality Control

Raw Data Information

Sample ID	Yield (bp)	Read Number	GC (%)	Q20 (%)	Q30 (%)
DSS-2	71,556,843	238,280	54.79	80.48	68.25
DSS-3	76,384,436	254,308	54.54	80.57	68.23
DSS-5	67,342,106	224,146	55	81.09	68.89
DSS-6	64,760,121	215,590	55.07	80.07	67.68
DSS-7	73,797,329	245,696	54.42	81.35	69.28
DSS-8	78,360,574	260,868	54.57	80.96	68.68
Go-1	71,725,245	238,826	54.11	81.43	69.33
Go-2	76,382,847	254,302	54.73	80.96	68.84
Go-3	77,987,436	259,632	54.17	81.12	68.96
Go-4	53,554,312	178,320	54.57	80.02	67.54
Go-6	58,350,522	194,276	54.85	80.13	67.75
Go-9	44,684,193	148,760	54.6	79.35	66.78

Sequencing 결과 yield는 충분하게 확보되었습니다.

- Yield (bp) : Sequencing된 total base pair 수입니다.
- Read Number : Sequencing을 통해 생산된 total read 수입니다.
- GC ratio : 전체 염기서열 중 GC contents의 비율입니다.
- Q20: 1/100 이하의 확률 (99% 정확도)로 error가 생길 것으로 추정되는 base pair 비율입니다.
- Q30: 1/1,000 이하의 확률 (99.9% 정확도)로 error가 생길 것으로 추정되는 base pair 비율입니다.

Quality Control

Raw Data Information

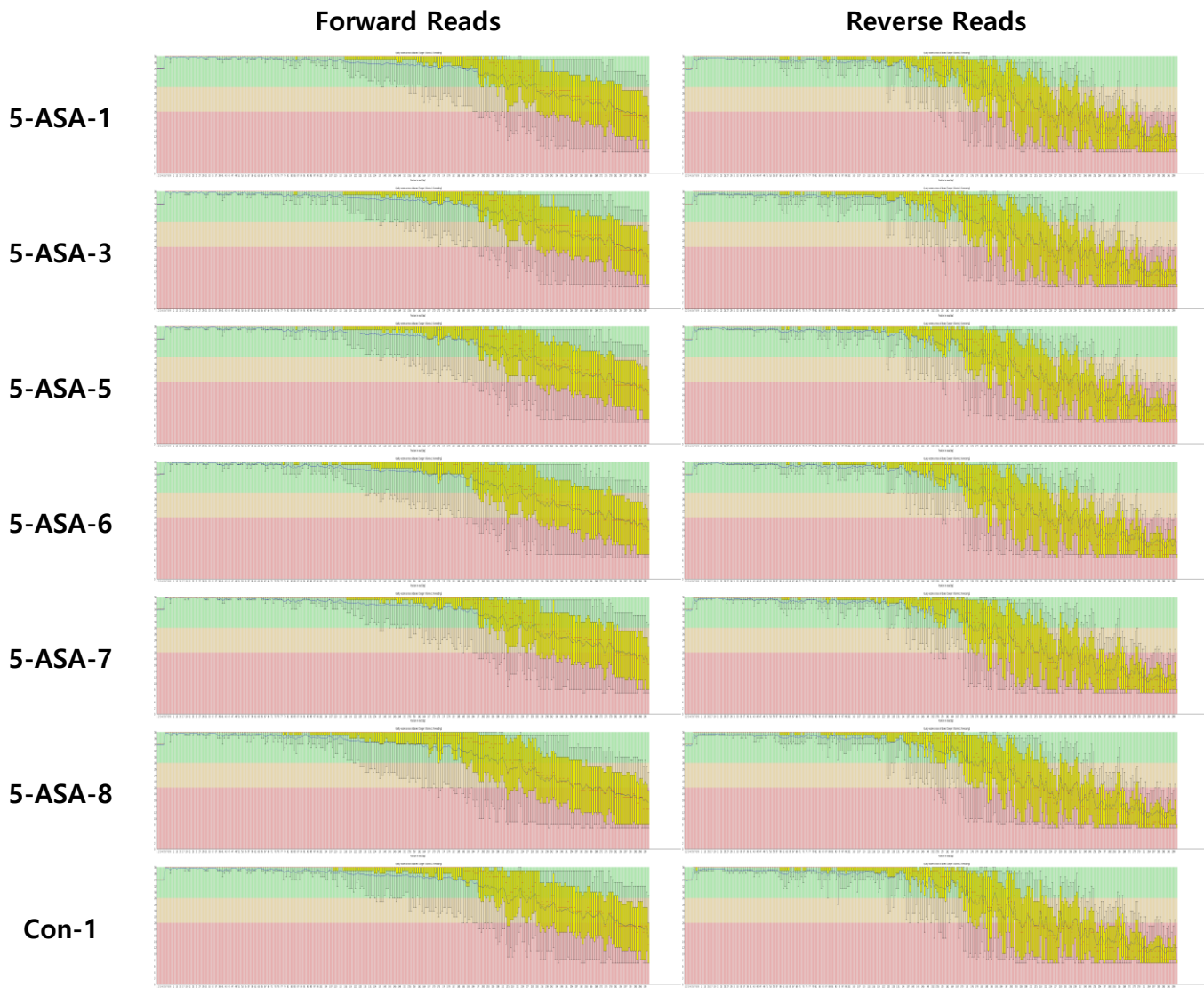
Sample ID	Yield (bp)	Read Number	GC (%)	Q20 (%)	Q30 (%)
Ju-3	63,843,824	212,556	54.39	80.57	68.18
Ju-4	65,326,352	217,454	54.7	80.4	68.08
Ju-5	62,004,115	206,420	54.42	81.01	68.81
Ju-6	61,330,773	204,156	54.67	79.96	67.45
Ju-8	90,767,990	302,200	55	80.76	68.76
Ju-9	91,479,840	304,556	54.08	80.94	68.86
Jung-1	63,902,229	212,708	54.58	80.39	67.88
Jung-2	52,663,840	175,320	54.45	80.25	67.76
Jung-6	54,015,144	179,880	54.11	81.03	68.74
Jung-7	57,367,976	190,978	55.08	78.15	65.22
Jung-8	56,460,213	187,994	54.36	81.21	69.15
Jung-9	46,496,796	154,808	55	79.69	67.44

Sequencing 결과 yield는 충분하게 확보되었습니다.

- Yield (bp) : Sequencing된 total base pair 수입니다.
- Read Number : Sequencing을 통해 생산된 total read 수입니다.
- GC ratio : 전체 염기서열 중 GC contents의 비율입니다.
- Q20: 1/100 이하의 확률 (99% 정확도)로 error가 생길 것으로 추정되는 base pair 비율입니다.
- Q30: 1/1,000 이하의 확률 (99.9% 정확도)로 error가 생길 것으로 추정되는 base pair 비율입니다.

Quality Control

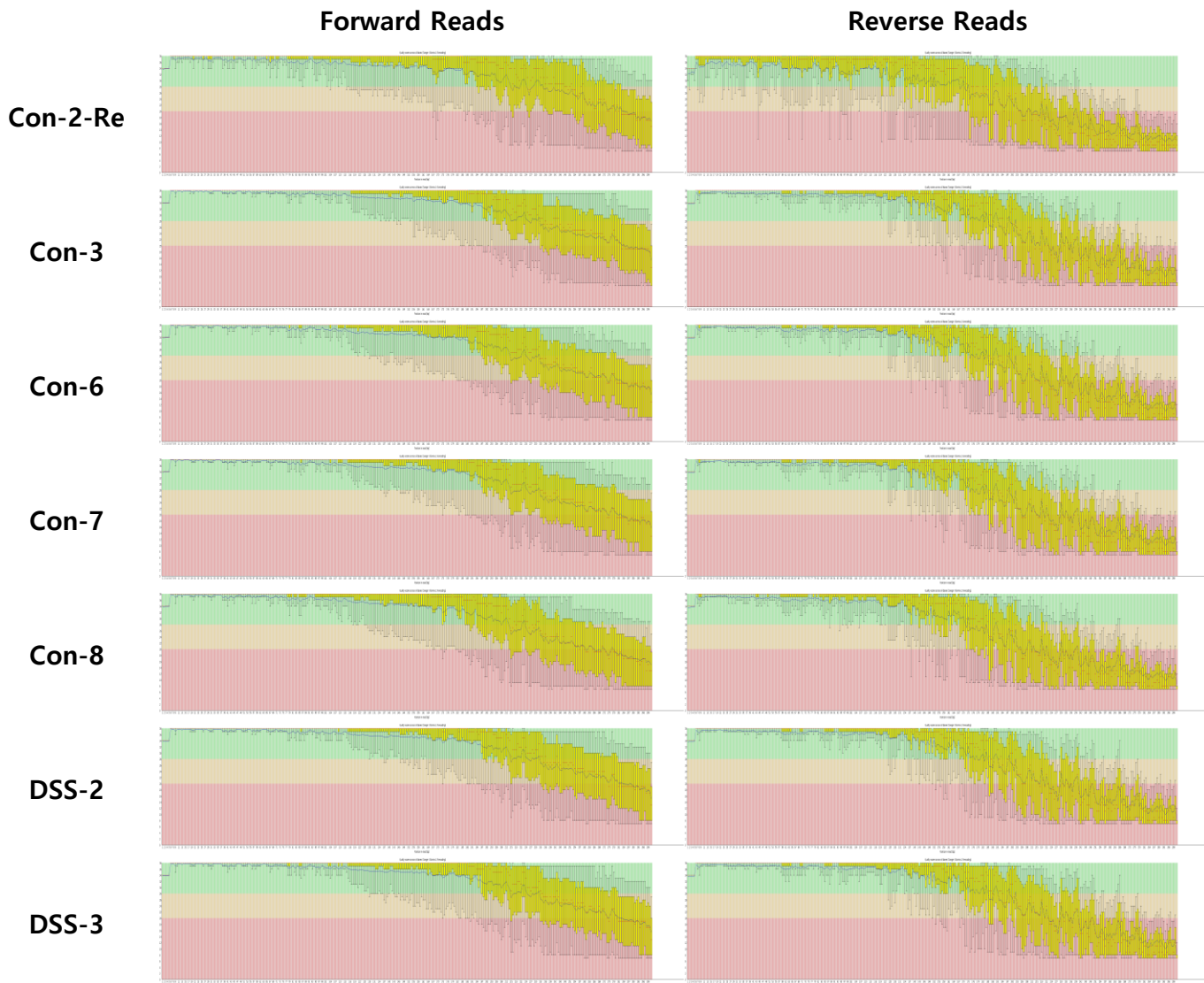
Per Base Sequence Quality



- Sequence 내 각 Position 에서의 평균 Data Quality를 나타냅니다
- Illumina Sequencing Platform의 특성상 Sequence 말단으로 갈수록 Sequencing Quality 가 낮아지는 경향이 있습니다

Quality Control

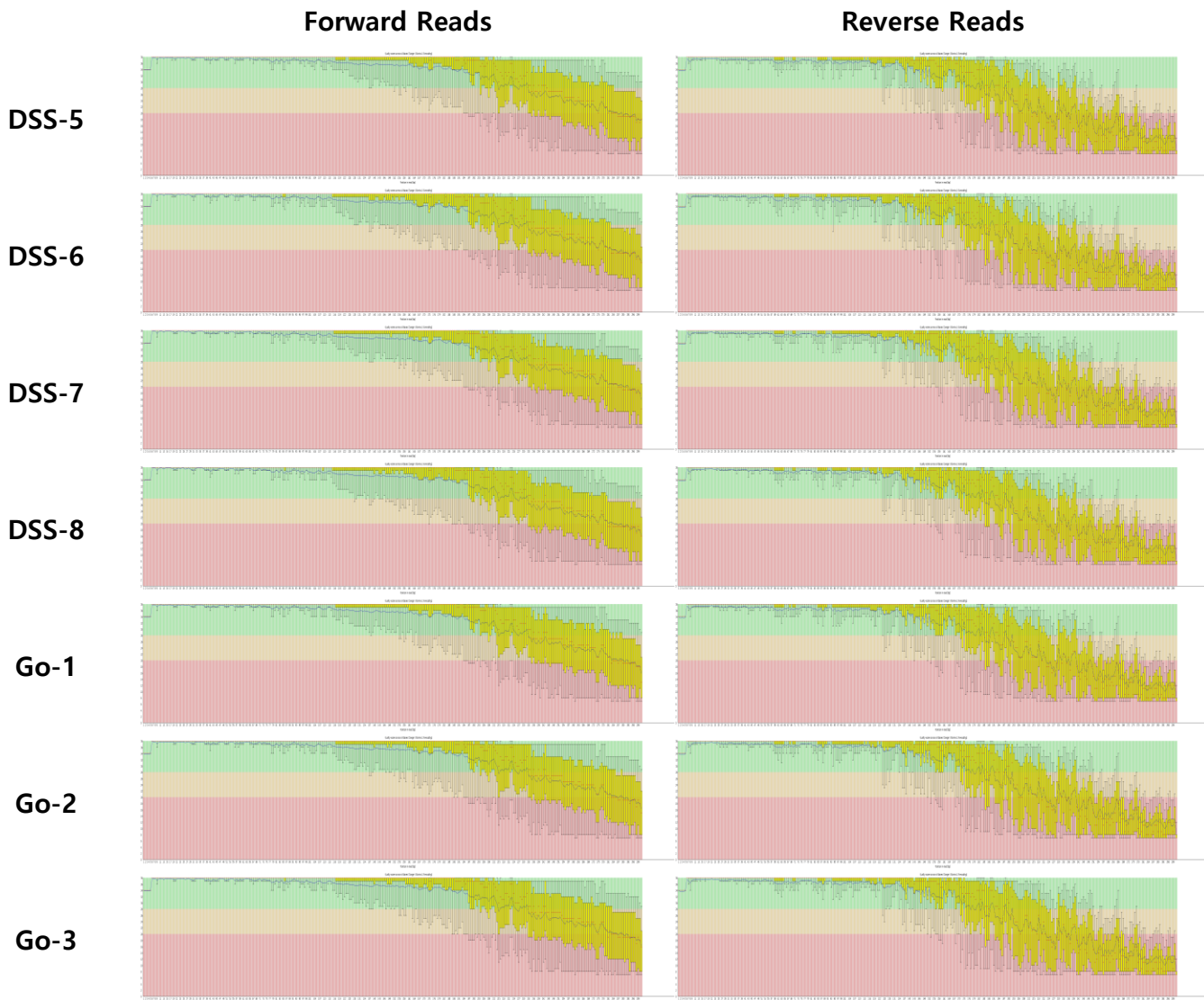
Per Base Sequence Quality



- Sequence 내 각 Position 에서의 평균 Data Quality를 나타냅니다
- Illumina Sequencing Platform의 특성상 Sequence 말단으로 갈수록 Sequencing Quality 가 낮아지는 경향이 있습니다

Quality Control

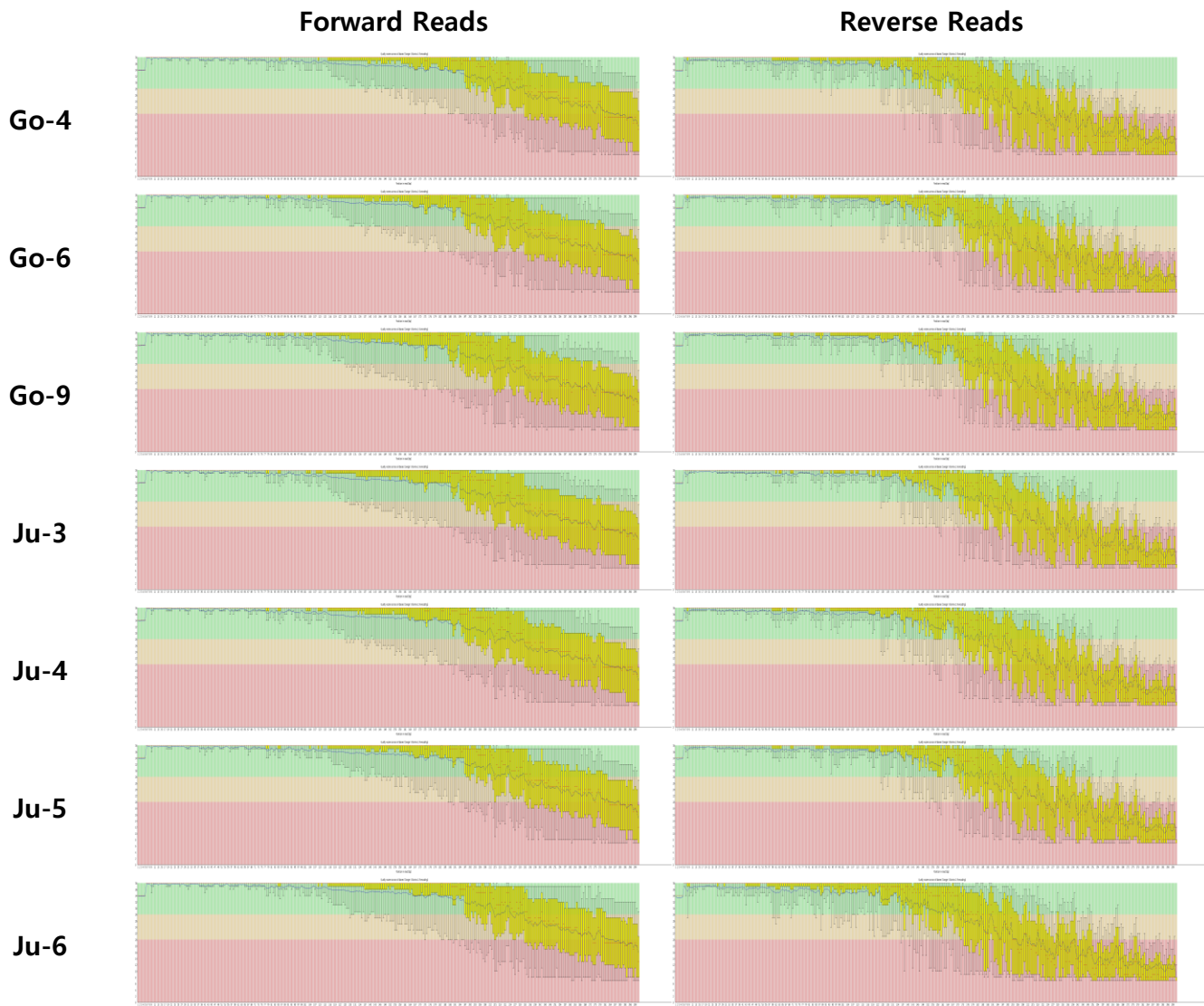
Per Base Sequence Quality



- Sequence 내 각 Position 에서의 평균 Data Quality를 나타냅니다
- Illumina Sequencing Platform의 특성상 Sequence 말단으로 갈수록 Sequencing Quality 가 낮아지는 경향이 있습니다

Quality Control

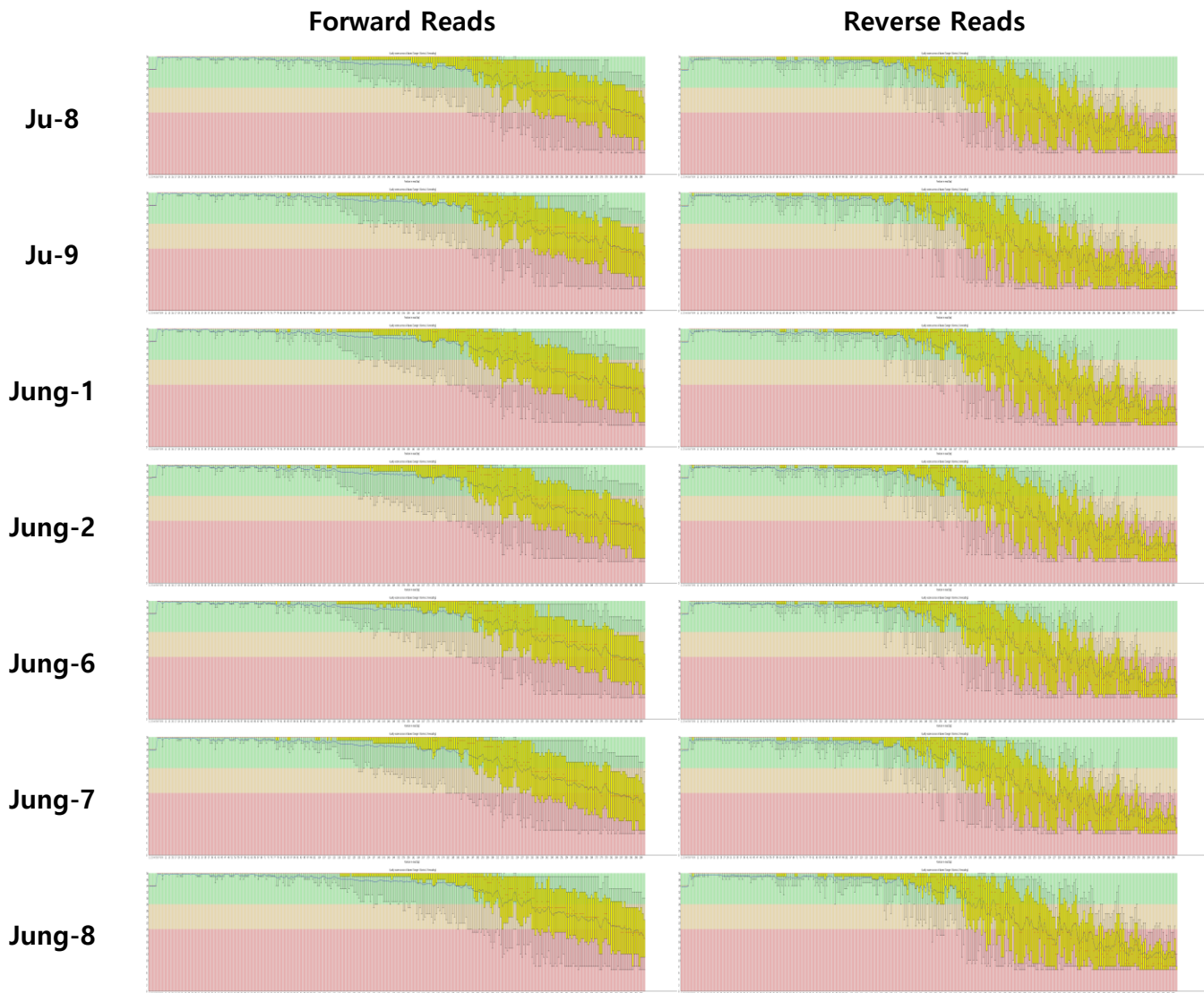
Per Base Sequence Quality



- Sequence 내 각 Position 에서의 평균 Data Quality를 나타냅니다
- Illumina Sequencing Platform의 특성상 Sequence 말단으로 갈수록 Sequencing Quality 가 낮아지는 경향이 있습니다

Quality Control

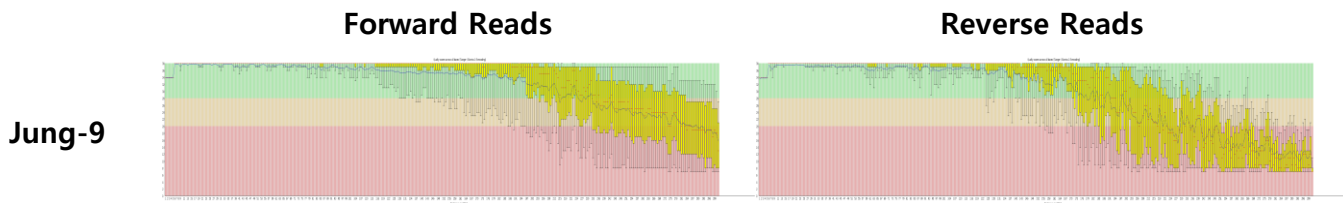
Per Base Sequence Quality



- Sequence 내 각 Position 에서의 평균 Data Quality를 나타냅니다
- Illumina Sequencing Platform의 특성상 Sequence 말단으로 갈수록 Sequencing Quality 가 낮아지는 경향이 있습니다

Quality Control

Per Base Sequence Quality



- Sequence 내 각 Position 에서의 평균 Data Quality를 나타냅니다
- Illumina Sequencing Platform의 특성상 Sequence 말단으로 갈수록 Sequencing Quality 가 낮아지는 경향이 있습니다

Denoising and Filtering Process

Denoising and Filtering results

Sample ID	Raw Reads	Denoised	Pair-Merged	Non-Chimeric	Chloroplast / Mitochondria Filtered
5-ASA-1	138,658	40,073	36,105	35,965	35,965
5-ASA-3	95,987	26,664	23,637	23,605	23,605
5-ASA-5	136,276	40,228	36,227	36,157	36,157
5-ASA-6	94,496	23,835	20,951	20,934	20,934
5-ASA-7	93,805	25,366	22,032	22,014	22,014
5-ASA-8	62,279	11,246	9,613	9,613	9,613
Con-1	97,260	22,946	19,807	19,741	19,741
Con-2-Re	305,607	70,424	63,407	62,999	62,999
Con-3	99,139	29,898	25,861	25,784	25,784
Con-6	108,815	26,684	23,360	23,309	23,309
Con-7	101,475	24,722	20,486	20,456	20,456

- **Denoising:** adaptor, index 등의 artificial sequence를 제거하고, 전술한 Read 말단의 Low Quality인 부분을 잘라낸 후, noise를 수정/제거합니다.
- **Pair-Merging:** 최종적으로 남은 신뢰할 수 있는 고품질의 forward read와 reverse read를 이어 붙여 하나의 긴 read로 만들어줍니다.
- **Remove Chimeric Sequences:** 복수의 16S rRNA gene으로부터 증폭되어 발생한 chimeric sequence를 제거합니다.
- **Chloroplast / Mitochondria Filtered:** Sequencing 과정에서 유입되는 Chloroplast / Mitochondria 유래 read를 제거합니다.

Denoising and Filtering Process

Denoising and Filtering results

Sample ID	Raw Reads	Denoised	Pair-Merged	Non-Chimeric	Chloroplast / Mitochondria Filtered
Con-8	70,889	11,284	9,210	9,210	9,210
DSS-2	119,120	32,392	29,296	29,248	29,248
DSS-3	127,130	36,333	32,921	32,856	32,856
DSS-5	112,058	33,784	29,735	29,579	29,579
DSS-6	107,776	27,783	23,785	23,763	23,763
DSS-7	122,838	38,069	34,486	34,416	34,416
DSS-8	130,408	39,406	35,301	35,241	35,241
Go-1	119,371	37,460	34,201	34,076	34,076
Go-2	127,112	37,654	33,058	33,036	33,036
Go-3	129,794	39,407	35,717	35,667	35,667
Go-4	89,122	22,155	19,026	19,018	19,018

- **Denoising:** adaptor, index 등의 artificial sequence를 제거하고, 전술한 Read 말단의 Low Quality인 부분을 잘라낸 후, noise를 수정/제거합니다.
- **Pair-Merging:** 최종적으로 남은 신뢰할 수 있는 고품질의 forward read와 reverse read를 이어 붙여 하나의 긴 read로 만들어줍니다.
- **Remove Chimeric Sequences:** 복수의 16S rRNA gene으로부터 증폭되어 발생한 chimeric sequence를 제거합니다.
- **Chloroplast / Mitochondria Filtered:** Sequencing 과정에서 유입되는 Chloroplast / Mitochondria 유래 read를 제거합니다.

Denoising and Filtering Process

Denoising and Filtering results

Sample ID	Raw Reads	Denoised	Pair-Merged	Non-Chimeric	Chloroplast / Mitochondria Filtered
Go-6	97,116	25,749	22,959	22,955	22,955
Go-9	74,352	17,039	15,274	15,272	15,272
Ju-3	106,249	30,273	27,508	27,483	27,483
Ju-4	108,710	30,326	25,909	25,870	25,870
Ju-5	103,175	31,034	27,718	27,700	27,700
Ju-6	102,055	26,244	22,797	22,784	22,784
Ju-8	151,069	43,212	39,437	39,316	39,316
Ju-9	152,245	45,454	42,140	41,731	41,731
Jung-1	106,331	28,825	25,395	25,394	25,394
Jung-2	87,639	24,121	21,172	21,169	21,169
Jung-6	89,911	26,819	23,883	23,873	23,873

- **Denoising:** adaptor, index 등의 artificial sequence를 제거하고, 전술한 Read 말단의 Low Quality인 부분을 잘라낸 후, noise를 수정/제거합니다.
- **Pair-Merging:** 최종적으로 남은 신뢰할 수 있는 고품질의 forward read와 reverse read를 이어 붙여 하나의 긴 read로 만들어줍니다.
- **Remove Chimeric Sequences:** 복수의 16S rRNA gene으로부터 증폭되어 발생한 chimeric sequence를 제거합니다.
- **Chloroplast / Mitochondria Filtered:** Sequencing 과정에서 유입되는 Chloroplast / Mitochondria 유래 read를 제거합니다.

Denoising and Filtering Process

Denoising and Filtering results

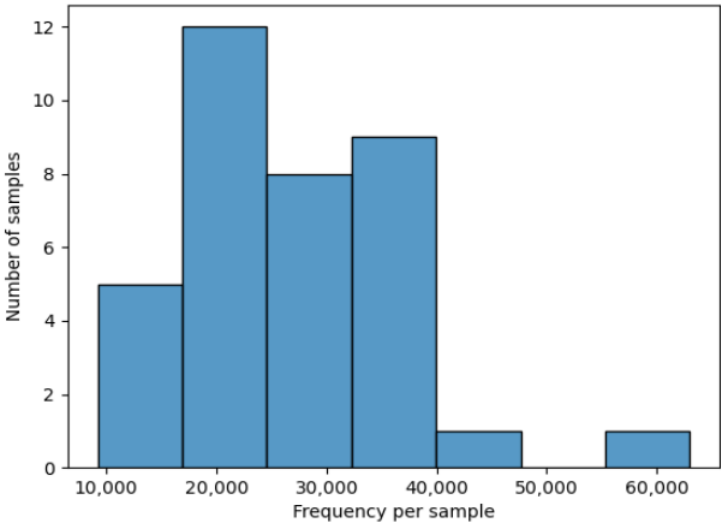
Sample ID	Raw Reads	Denoised	Pair-Merged	Non-Chimeric	Chloroplast / Mitochondria Filtered
Jung-7	95,470	16,579	13,093	13,090	13,090
Jung-8	93,974	28,064	24,947	24,893	24,893
Jung-9	77,393	18,120	15,490	15,490	15,490

- **Denoising:** adaptor, index 등의 artificial sequence를 제거하고, 전술한 Read 말단의 Low Quality인 부분을 잘라낸 후, noise를 수정/제거합니다.
- **Pair-Merging:** 최종적으로 남은 신뢰할 수 있는 고품질의 forward read와 reverse read를 이어 붙여 하나의 긴 read로 만들어줍니다.
- **Remove Chimeric Sequences:** 복수의 16S rRNA gene으로부터 증폭되어 발생한 chimeric sequence를 제거합니다.
- **Chloroplast / Mitochondria Filtered:** Sequencing 과정에서 유입되는 Chloroplast / Mitochondria 유래 read를 제거합니다.

Denoising and Filtering Process

Feature Table Summary

Metric	Sample
Number of samples	36
Number of features	7,088
Total frequency	964,707



<Summary Table>

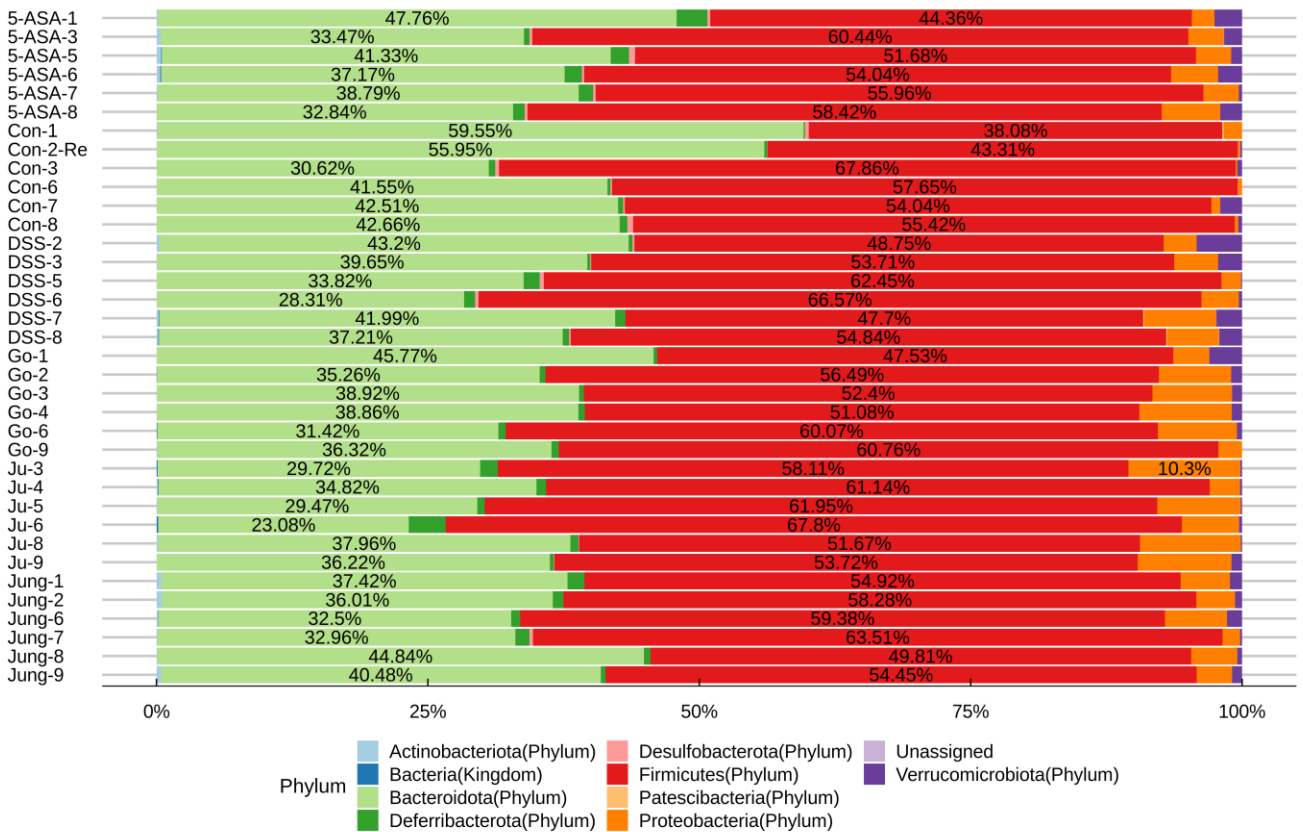
<Frequency per Feature>

- 최종적으로 선별 및 확보된 서열들은 Representative Sequence (Feature)로 정의하고, 확보된 전체 feature의 종류, 전체 빈도, feature 당 빈도를 확인했습니다.

Bacteria Taxonomy Classification

Charts for Classification visualization (Phylum)

Top 10 bar plot - Phylum



- 각 level에서 상위 10개 taxon의 taxonomic classification을 확인했습니다

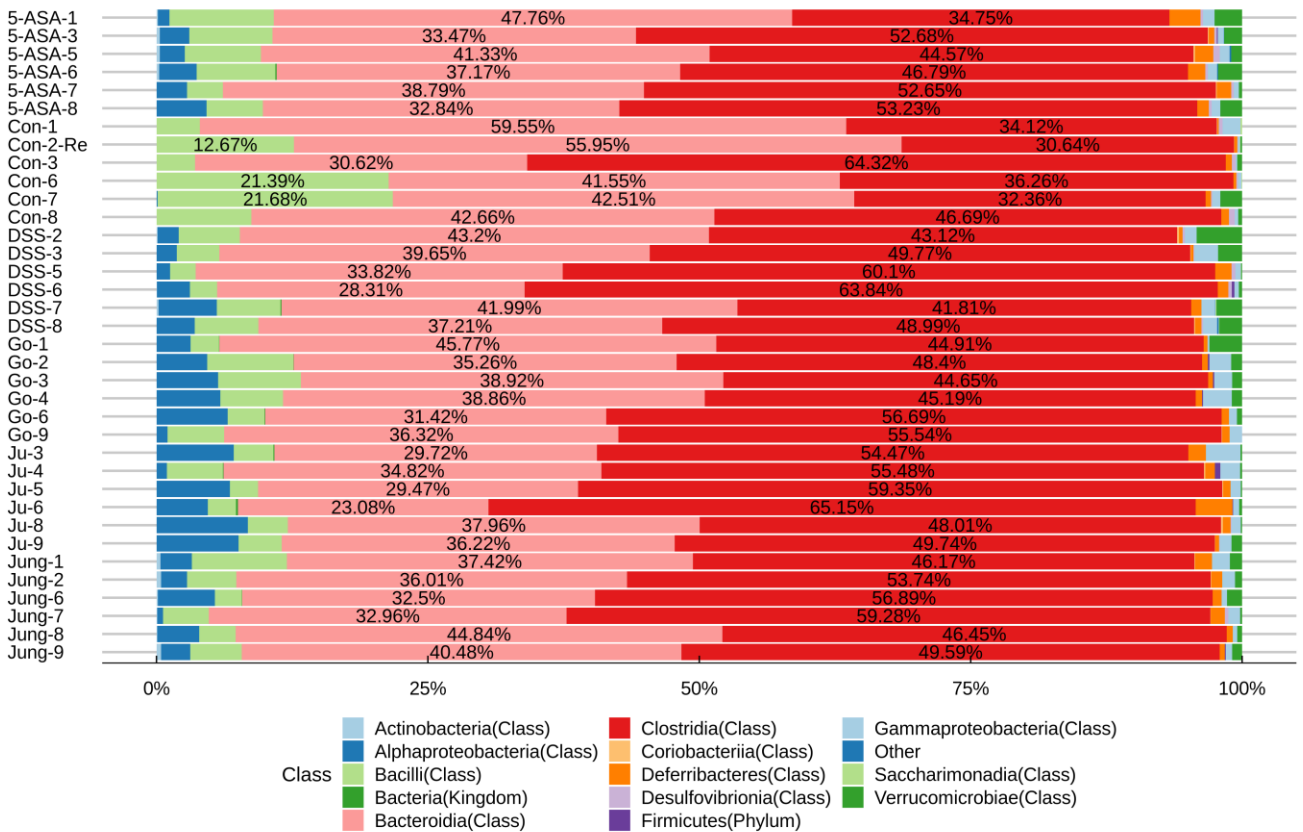
※ 상세 결과는 FTP 서버 상 **3. Taxonomy classification (Krona Plot)** 폴더 참고

Analysis Report

Bacteria Taxonomy Classification

Charts for Classification visualization (Class)

Top 10 bar plot - Class



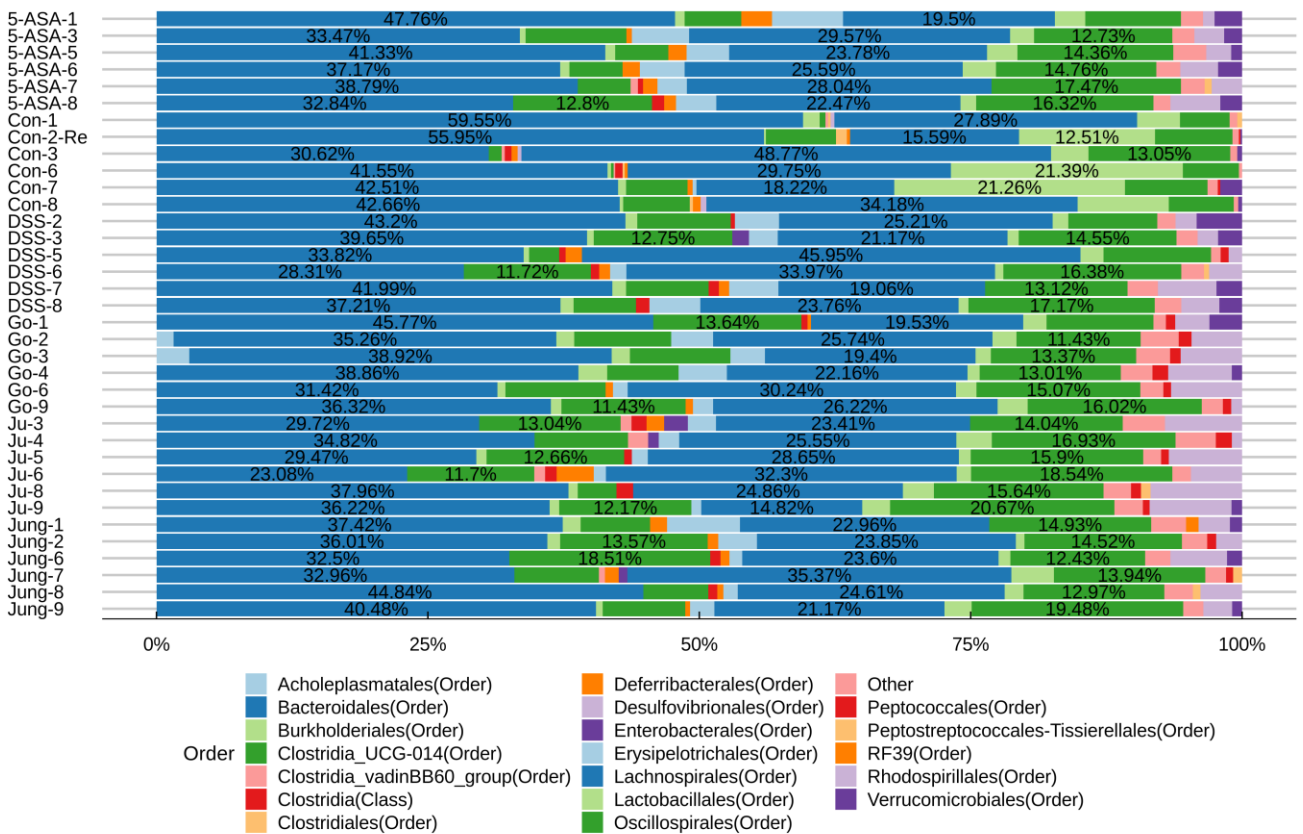
- 각 level에서 상위 10개 taxon의 taxonomic classification을 확인했습니다

※ 상세 결과는 FTP 서버 상 **3. Taxonomy classification (Krona Plot)** 폴더 참고

Bacteria Taxonomy Classification

Charts for Classification visualization (Order)

Top 10 bar plot - Order



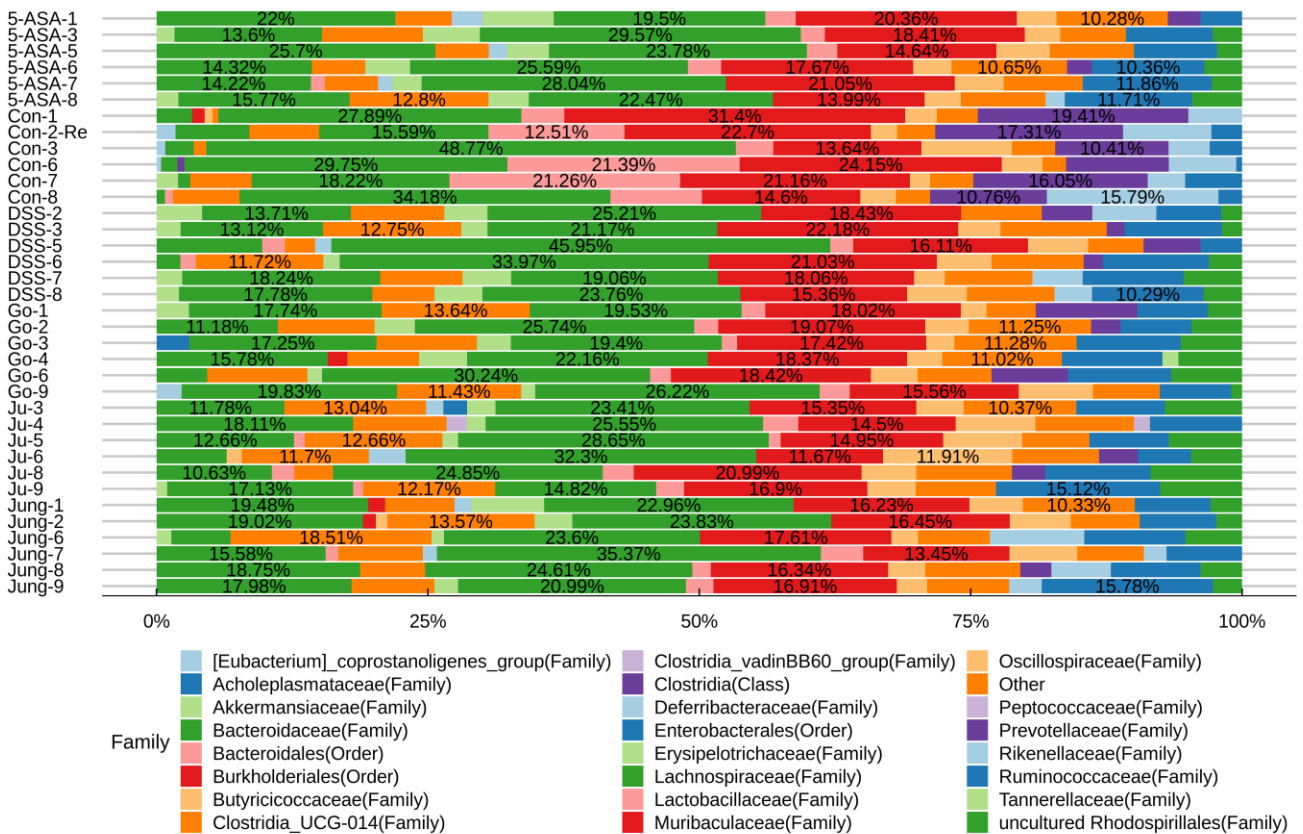
- 각 level에서 상위 10개 taxon의 taxonomic classification을 확인했습니다

※ 상세 결과는 FTP 서버 상 **3. Taxonomy classification (Krona Plot)** 폴더 참고

Bacteria Taxonomy Classification

Charts for Classification visualization (Family)

Top 10 bar plot - Family



- 각 level에서 상위 10개 taxon의 taxonomic classification을 확인했습니다

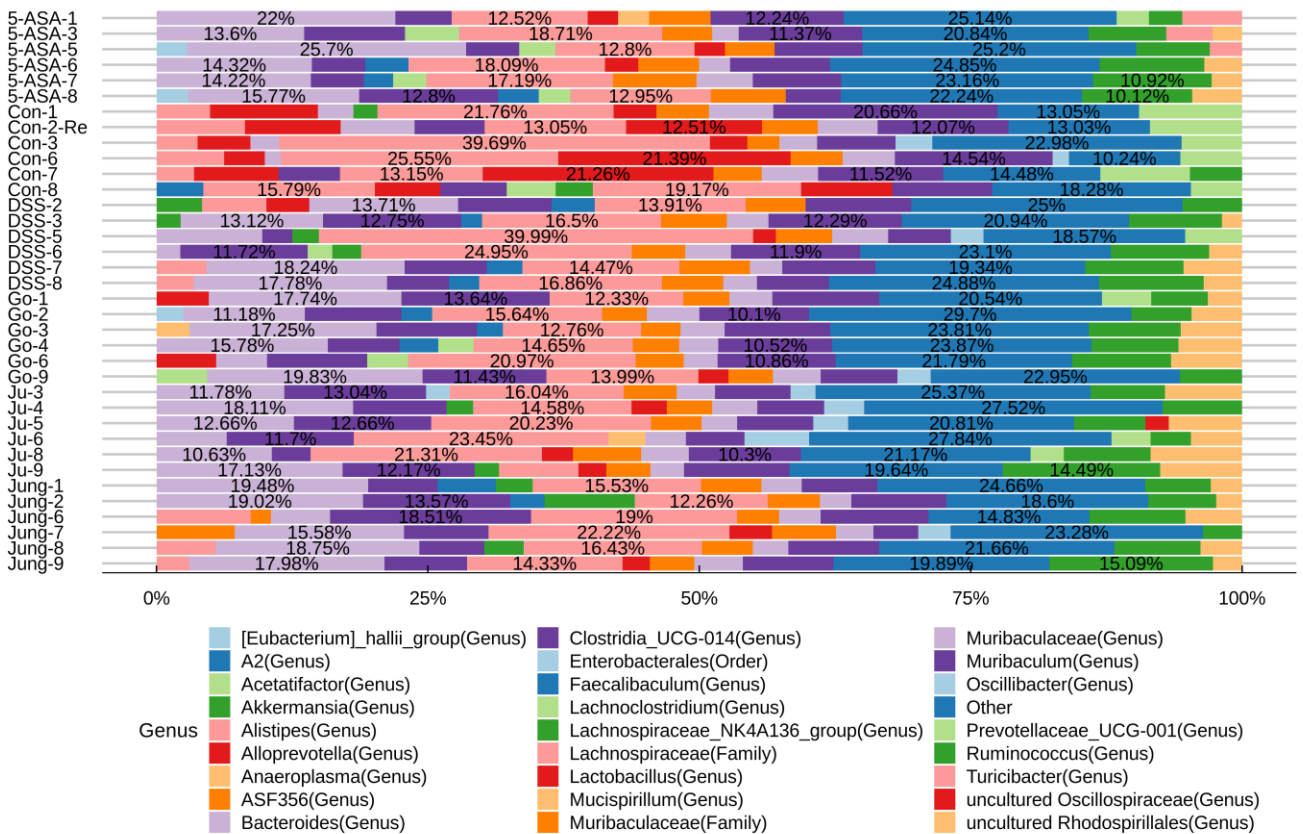
※ 상세 결과는 FTP 서버 상 **3. Taxonomy classification (Krona Plot)** 폴더 참고

Analysis Report

Bacteria Taxonomy Classification

Charts for Classification visualization (Genus)

Top 10 bar plot - Genus

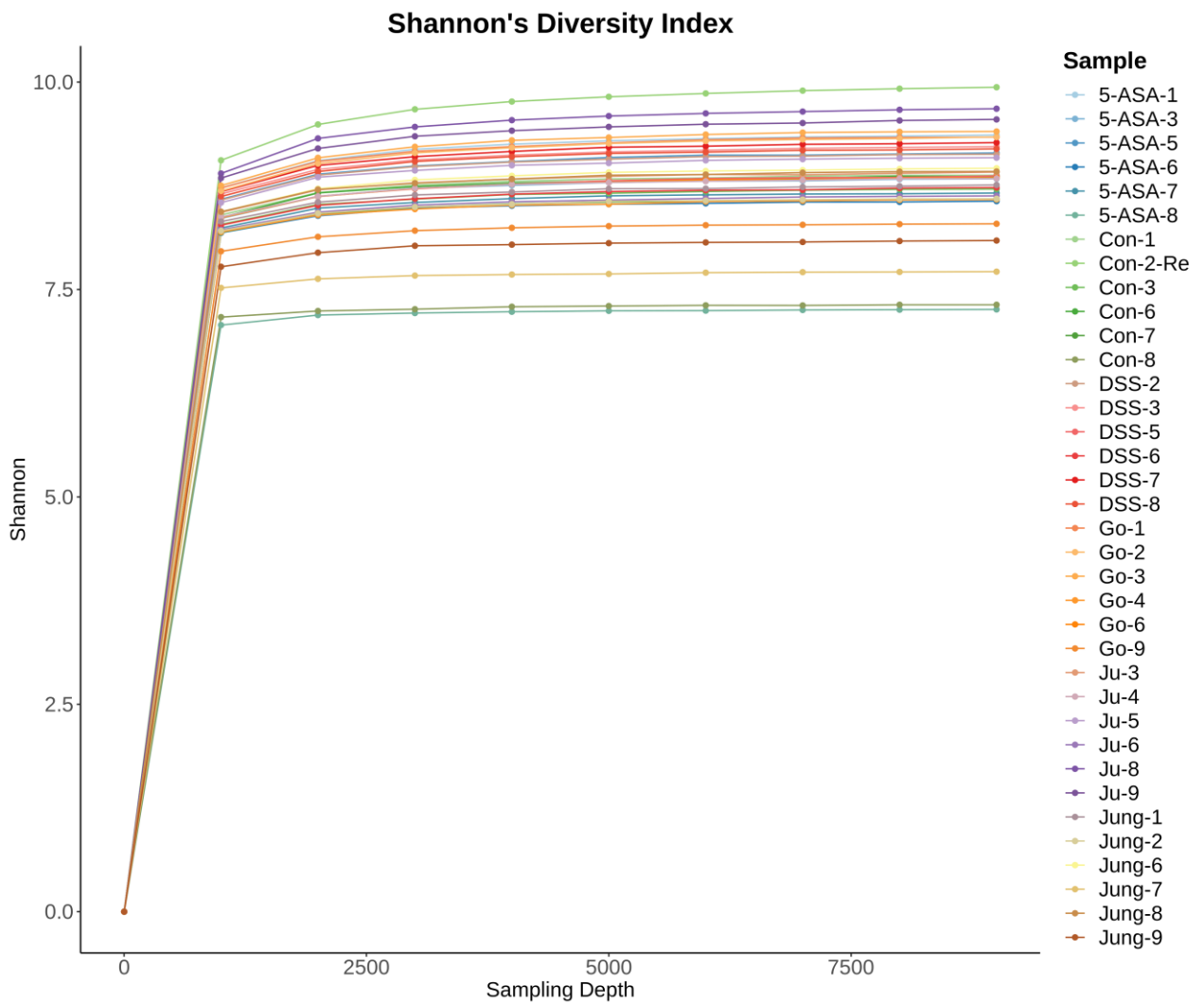


- 각 level에서 상위 10개 taxon의 taxonomic classification을 확인했습니다

※ 상세 결과는 FTP 서버 상 **3. Taxonomy classification (Krona Plot)** 폴더 참고

Diversity Analysis

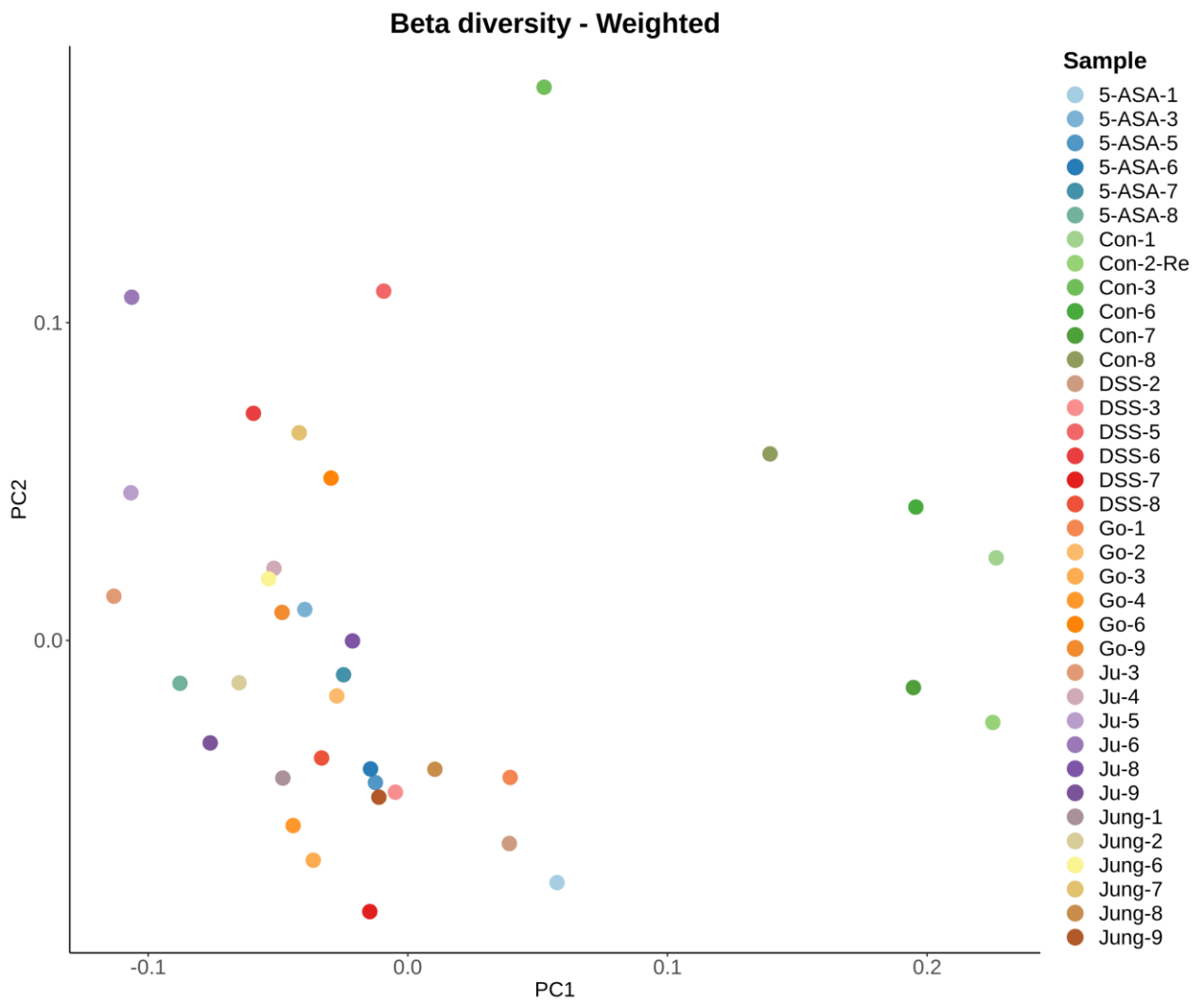
Alpha diversity - Rarefaction



- Shannon's Diversity Index (α -diversity) : Sample 내 미생물 다양성을 나타내는 지표입니다.
- Sampling Depth가 올라감에 따라 일정수준에서 지표가 Saturation이 되는 현상을 확인할 수 있습니다. Saturation이 되는 구간은 각 Sampling을 수행한 표본집단이 모집단을 충분히 반영하는 것을 의미합니다.
- 군집간 분석(Beta diversity)의 경우, 다양성이 급격하게 변하는 구간에서 수행할 경우 Sampling Bias로 인해 실제 군집분포와 상이할 가능성이 있습니다. 따라서 Sample 내 다양성 지표가 완만한 구간에서 Sampling 수행하여 군집간 분석을 진행합니다.

Diversity Analysis

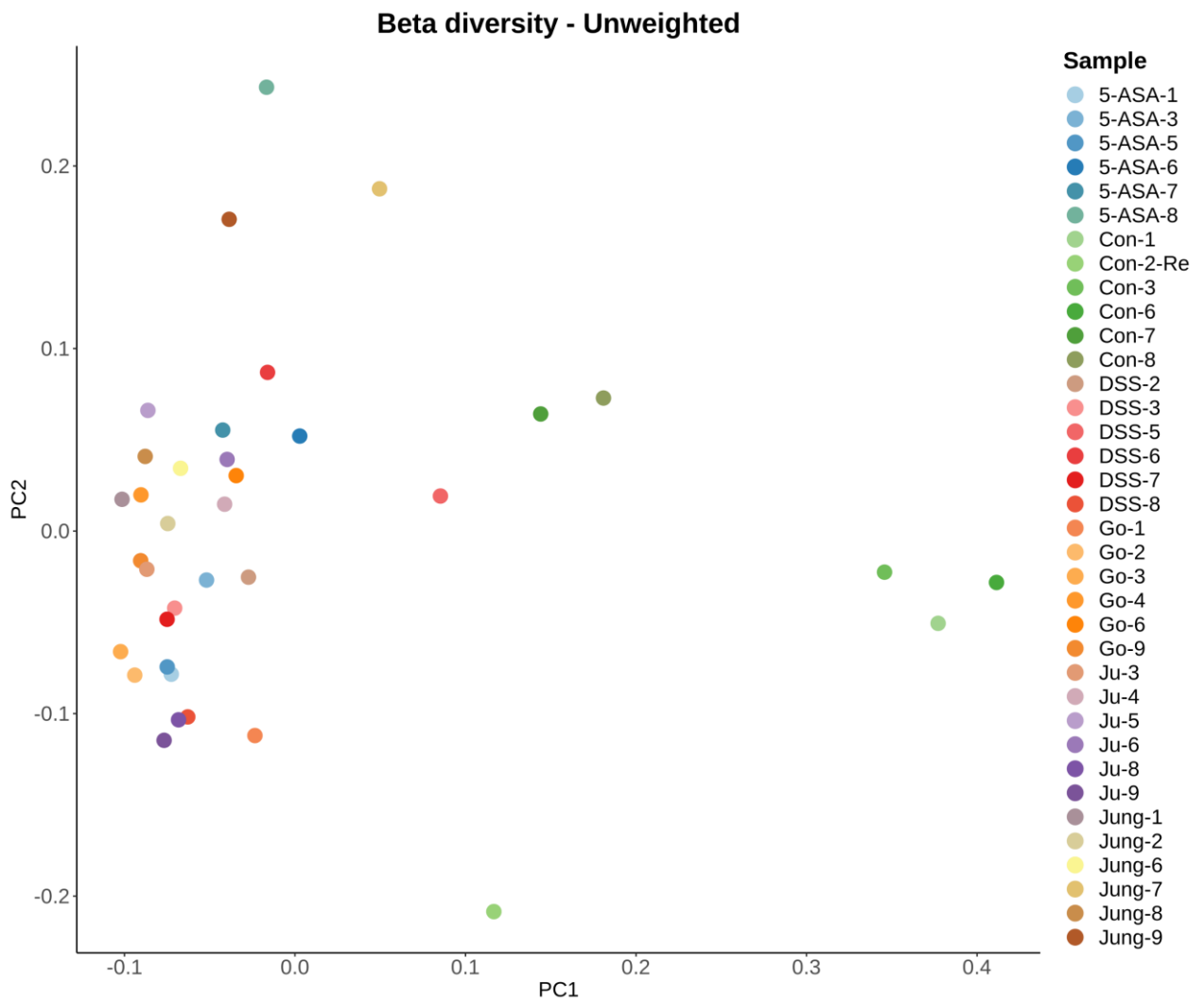
Beta diversity (Weighted Unifrac)



- 미생물 군총의 양(weighted)을 이용하여 Sample 별 거리 계산 후 좌표공간에 나타냈습니다.

Diversity Analysis

Beta diversity (Unweighted Unifrac)



- 미생물 군총의 유무(unweighted)를 이용하여 Sample 별 거리 계산 후 좌표공간에 나타냈습니다.

Comment / Contact

Comment

- Raw data는 안내 드린 FTP 주소에서 다운로드 받으시길 바라며, 다운로드 가능 기간은 90일 입니다.
- 각 Sample의 classification 정보는 FTP 서버 "3. Taxonomy Classification (Krona Plot)" 폴더 내 파일들을 참조하시기 바랍니다.

FTP

- FTP : <http://gofile.me/5YZtx/VEQ2V3tEk>

