

# User Manual

## ExMEGA 2.0

### & Data Analysis

## <목 차>

1. ANCOM and Definitions of key terminologies -----3
2. ExMEGA Download and Setup -----4
3. Differential Abundance (DA) -----6
4. Diversity Analysis (ExMEGA GraphicPlus) -----10

# 1. Definitions of key terminologies

## ANCOM

**ANCOM**(analysis of composition of microbiomes)은 Metagenome 을 분석하는 검증된 통계적 방법입니다. ANCOM 을 사용하여 Relative abundance 로 Absolute abundance 를 추정할 수 있습니다. ANCOM 을 효과적으로 이용하기 위해서는 그룹 간에 차이가 있으며 서로 연관된 샘플들이어야 합니다.

## 기타 용어

ASV(Amplicon Sequence Variant)는 dada2 방법으로 생성되며, 기존 OTU(유사도를 기반으로 하는 서열 클러스터링 방법)보다 뛰어난 해상도를 보여줍니다.

Relative abundance 를 통한 분석만으론 Metagenome 생태계를 비교하기는 어렵습니다. 최근에는 다양한 통계방법(ANCOM 등)을 이용하여 Absolute abundance 를 추정하고 있습니다.

용어	정의
OTU	운영 분류 단위: 97% 유사성을 갖는 DNA 서열 그룹.
ASV	PCR 증폭 및 시퀀싱 중에 생성된 가짜 시퀀스를 제거한 후 고처리량 마커 유전자 분석에서 회수된 개별 DNA 시퀀스.
Absolute abundance	생태계의 단위 부피에서 관찰할 수 없는 실제 분류군의 풍부함.
Relative abundance	표본에 있는 모든 분류군의 합에 대한 특징표에서 관찰된 분류군의 비율. 0 과 1 사이

Reference <https://www.zymoresearch.com/blogs/blog/microbiome-informatics-otu-vs-asv>  
a Callahan, Benjamin J., Paul J. McMurdie, and Susan P. Holmes.

## 2. ExMEGA Download and Setup

(주)이바이오젠은 Metagenome data 를 엑셀 기반에서 쉽게 분석할 수 있도록 분석 결과 보고 시 ExMEGA (Excel based MetaGenome Analysis) tool 과 ExMEGA Graphic Plus 를 함께 제공한다.

ExMEGA 분석 툴은 (주)이바이오젠이 연구자들이 Metagenome 데이터를 보다 쉽게 다루고 원하는 데이터를 쉽게 얻을 수 있도록 사용자 편의를 최대한 반영한 분석 툴이고 엑셀 프로그램 안에서 다양한 분석을 직관적으로 수행할 수 있도록 개발되었다. ExMEGA 분석 툴은 사용자들의 요구사항을 지속적으로 반영하여 데이터 분석과 엑셀 사용에 익숙하지 못한 연구자들도 쉽게 사용이 가능하도록 계속 업데이트 될 예정이다.

이바이오젠에서 제공하는 Metagenome 를 열기 전에 **다운로드 폴더에서 함께 제공한**

**ExMEGA\_v(버전)\_Installer.zip** 파일의 압축을 풀고, **setup.exe** 를 실행하면 분석 툴이 설치된다(그림 1-1 A). 만약 설치가 되지 않을 경우, 압축을 푼 파일에 있는 ExMEGA 폴더를 컴퓨터의 로컬 C 드라이브 아래로 복사+붙여넣기 하면 ExMEGA Graphic Plus 프로그램이 설치 완료된다(그림 1-1 B). 설치가 완료되고 ExMEGA format 의 엑셀 데이터를 열면 자동으로 ExMEGA 분석 툴이 구동된다. 참고로 ExMEGA 설치 전에 실행 중인 엑셀 파일이 있으면 종료시킨 후 다시 실행해야 ExMEGA 를 사용할 수 있다.

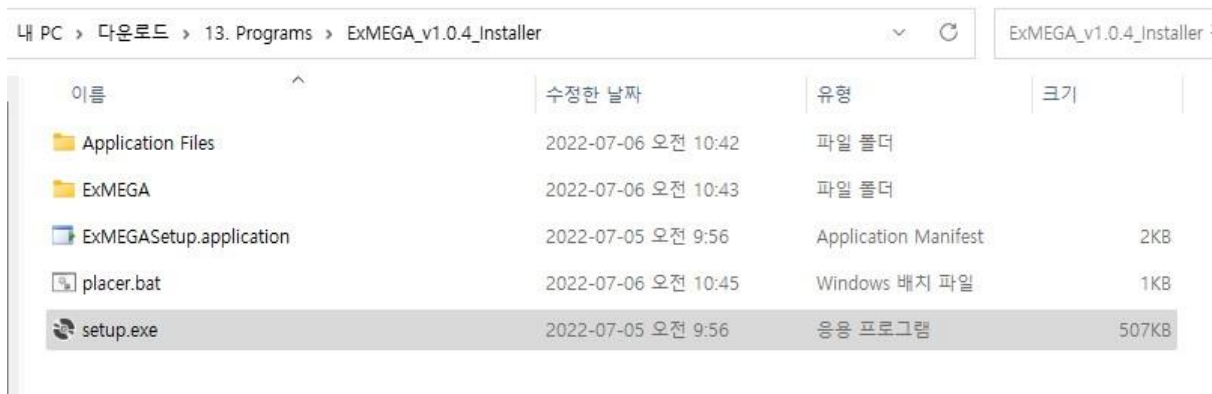


그림 1-1 A. ExMEGA set up

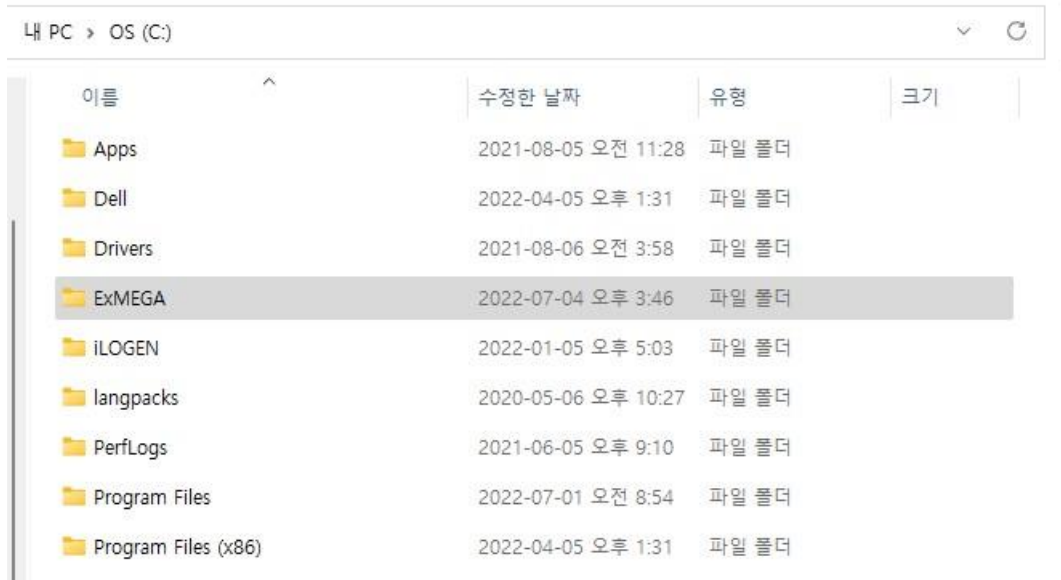


그림 1-1 B. ExMEGA Graphic Plus installation

ExMEGA format 의 엑셀 파일을 열면, 왼쪽에 Taxonomy 창과 가운데에 Metagenome data, 오른쪽에 DA(Differential Abundance) Analysis 창이 실행된다(그림 1-2). Taxonomy 창에서는 기본 설정된 Taxonomy 정보가 있고, 원하는 Taxon 들만 선택하여 데이터를 필터링할 수 있다. DA Analysis 창에서는 Relative Abundance 값을 선택하여 DA 선별을 쉽게 할 수 있고 DA 를 각 Taxon 별로 그래프를 작성할 수 있다. DA 분석 창에서 Pie Chart 뿐만 아니라 Venn Diagram 을 직접 그릴 수 있고 선별된 ASV(Amplicon Sequence Variant) 대상으로 Bar plot, Krona Chart, PCoA, Clustering heatmap, LDA plot 을 그릴 수 있다.

1 filter: 1539		Taxonomy						Average of Relative Abundance			Average of Feature Count			
ID	ASV	Kingdom	Phylum	Class	Order	Family	Genus	Species	A	B	C	A	B	C
1	4dda3a74f2b79a7d1a5be15fb8f9cd4	d	Bacteri	p	Bactero	c	Bacteroi	o	Bactero	f	Bacteroi	g	Bactero	.
2	889bfa9e582f8e8b58e7dfb4f7b66a7	d	Bacteri	p	Firmicu	c	Clostrid	o	Oscillo	f	Oscillo	.	.	.
3	0b2f9863db272dc65373dfc3d69097	d	Bacteri	p	Bactero	c	Bacteroi	o	Bactero	f	Muribac	g	Muribac	_uncultu
4	98cd478dce17c474061139e7505cfd	d	Bacteri	p	Firmicu	c	Clostrid	o	Clostrid	f	Clostrid	g	Clostrid	_uncultu
5	b4dca0048baf098b2063103c09819	d	Bacteri	p	Firmicu	c	Clostrid	o	Oscillo	f	Oscillo	g	Coldex	.
6	5e32e61daa94379223d221363003615	d	Bacteri	p	Bactero	c	Bacteroi	o	Bactero	f	Bacteroi	g	Bactero	.
7	cd6c8737913138e03a6536d8bba231e6	d	Bacteri	p	Firmicu	c	Bacilli	o	Achole	f	Achole	g	Anaero	_uncultu
8	cb0c5d4e25896d271f66dcb3f6d62	d	Bacteri	p	Firmicu	c	Bacilli	o	Lactob	f	Leucon	g	Leucon	.
9	029e0d19f9f94e0e97527cc0b9860a	d	Bacteri	p	Firmicu	c	Clostrid	o	Oscillo	f	Rumin	g	Rumino	_incertae
10	a322914b74187cc027862a9232206b2	d	Bacteri	p	Actinob	c	Corioba	o	Corioba	f	Figgerth	.	.	.
11	c6e4e4126581c5c459604d23c218ce	d	Bacteri	p	Desulfo	c	Desulfo	o	Desulfo	f	Desulfo	g	Desulfo	.
12	194f6f33dfe06056ec90050704860	d	Bacteri	p	Bactero	c	Bacteroi	o	Bactero	f	Muribac	g	Muribac	_uncultu
13	4r8000f3906a0a6014501411060a8a9	d	Bacteri	p	Desulfo	c	Desulfo	o	Desulfo	f	Desulfo	g	Desulfo	_uncultu

그림 1-2. Metagenome data in ExMEGA format

# 3. Differential Abundance (DA)

## Taxonomy 창 사용 방법

Metagenome 의 Clustering 최소 단위인 ASV(Amplicon Sequence Variant)는 정해진 분류군을 볼 수 있는 직접적인 방법이 아니다. 각 Taxonomy(Phylum, Class, Order, Family, Genus, Species)에서는 원하는 류군만 필터링하여 볼 수 있다. (그림 3-1)

그림 3-1. Taxonomy selection

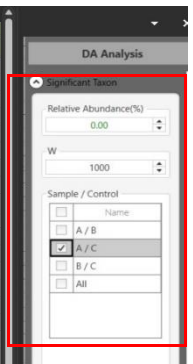
가장 왼쪽 상단에 'View All Data' 버튼을 누르면 필터를 모두 해제하여 다시 전체 결과를 볼 수 있다

Filter: 661		Taxonomy						
ID	ASV	Kingdc	Phylui	Class	Ordei	Famil	Genu	Speci
1	321592f084359113f23e3e38237c8093	d__Bacteri	p__Campilk	c__Campyl	o__Campyl	f__Helicob	g__Helicob	s__Helicob
2	8373a6f34bdfb5996c716e2c6e764ce2	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Rikenell	g__Alistipe	s__Alistipe
3	fa387857a369df7328f595391386555d	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Rikenell	g__Alistipe	s__Alistipe
4	8d0a3ca24c95ef79fb0fbbd91709f	d__Bacteri	p__Campilk	c__Campyl	o__Campyl	f__Helicob	g__Helicob	s__Helicob
5	e01914f46e26f1aaed24bd00c9c442b8	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__Muriba
6	c84137e57f8283dbc56df43da161dd1c	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Prevotel	g__Prevotel	s__Prevotel
8	82a938698291b68cefc9b9d46457229b	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Bacterg	g__Bacterg	s__Bacterg
10	d52e4e06b7fc88a48f5464555c0973f3	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Prevotel	g__Prevotel	s__Prevotel
11	e28749be518256d40762e358d5f48821	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Prevotel	g__Allopre	s__Allopre
12	edaebfecd88812ac54235907a571a16d	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__uncultu
14	3bfbf0a80abd58e7c6a16da6dec33154	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Rikenell	g__Alistipe	s__Alistipe
15	af5fe3a72c0bbba02162e60150c31887	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Bacterg	g__Bacterg	s__Bacterg
17	43ad07aa6643c9dd6d7fd103956c27d0	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__unident
18	ecb1f8a7d07e9f1e081f4f5faf2327de	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__uncultu
19	5eafb639b86bf607f38a817820e982a9	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__uncultu
20	4dda3a74f2b9a7d1a5be15fbf8fbc64	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Bacterg	g__Bacterg	s__Bacterg
21	d60740cdfec329165f0fbbc482494635	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Rikenell	g__Alistipe	s__Alistipe
22	84b1653758e1a862e7b599d66960273c	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__Muriba
23	34e8c93392eff3a8245f395589dbfcc2	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Rikenell	g__Alistipe	s__Alistipe
24	7a8c6354aaa1d4c6f0cb647e9fd91356	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__uncultu
25	3af92fc1a7e277a4a8322b86c7149c7e	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__uncultu
26	08f0f19726e994a36f5f5d0fb628cc3a	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muribac	s__Muribac
27	4240d1092b0c263df5f3c952a08fe75	d__Bacteri	p__Bacterc	c__Bacterc	o__Bacterc	f__Muribac	g__Muriba	s__Muriba

## Significant Taxon

오른편의 DA Analysis 부분에서 "Significant Taxon" 창은 그룹별로 비교한 결과에서 유의하게 발견 차이가 나는 분류군을 필터링 할 수 있도록 만들어 놓은 것이다. Relative Abundance(상대 풍부도) 와 W(통계 유의성)값을 이용해 필터링하여 각 샘플 및 그룹에 대한 조건이 적용된 결과를 확인할 수 있다. W 값을 설정하면 설정한 값 이상을 만족하는 Taxon 만 필터링 되어 통계적 유의성을 만족하는 분류군에 대한 분석이 가능하다. Relative Abundance 도 설정한 값 이상의 Taxon(ASV)만 필터링하여 낮은 분류군들은 제거하여 분석할 수 있다.

Filter: 17		W			SigW				Average of Relative Abundance			Average of Feature Count			AA	AB
ID	ASV	A/B	A/C	B/C	All	A/B	A/C	B/C	All	C	A	B	C	A	B	AI
454	4dd3a3a74f2b9a7d1a5be15fbf8bc64		1418	1534	1531	FALSE	TRUE	TRUE	TRUE	2.97%	0.00%	0.00%	750.875	0	0	0.00%
792	889bfa9e582f8e8b58e7dfb4f7b6ea7		1374	1439	39	FALSE	TRUE	TRUE	FALSE	0.00%	0.00%	0.00%	0.5	0	0	0.00%
69	0b2f98063db272dce6373defc3d00907		1360	1432	1466	FALSE	TRUE	TRUE	TRUE	0.00%	0.62%	0.36%	0	162.5	82.142857	0.67%
883	98cd478dcb17c474061139e7505cf0f		1302	21	24	FALSE	TRUE	FALSE	FALSE	0.01%	0.00%	0.00%	2.625	0	0	0.00%
1047	b4dca0048bafb098b206531033c09819		1269	1446	41	FALSE	TRUE	TRUE	FALSE	0.00%	0.00%	0.00%	0.25	0	0	0.00%
554	5e329e61daa94379223d221363003615		1200	1412	1448	FALSE	FALSE	TRUE	TRUE	1.01%	0.00%	0.00%	256	0	0	0.00%
1180	c6dc8737919138e03a6536d8bba231e6		1185	1408	39	FALSE	FALSE	TRUE	FALSE	0.00%	0.00%	0.00%	0.5	0	0	0.00%
1137	c0bc5dde25d896d271f66c9cbe3d6d62		1170	1396	37	FALSE	FALSE	TRUE	FALSE	0.00%	0.01%	0.00%	0	3.5	1	0.00%
1254	d29ed019f8f94bdeb975627ccb98d0a		1162	1406	45	FALSE	FALSE	TRUE	FALSE	0.00%	0.00%	0.00%	0.25	0	0	0.00%
942	a382f914b74187cc027862e9232206b2		1156	1394	40	FALSE	FALSE	TRUE	FALSE	0.00%	0.00%	0.00%	0	0	0.8571429	0.00%
1182	c6e4e41126581c5cf596b4d23c8218ce		1133	1319	22	FALSE	FALSE	FALSE	FALSE	1.87%	0.54%	0.51%	473.125	142.5	116	0.03%
141	194f6fa35dfc9b0650eca906507a486b		1130	1321	1338	FALSE	FALSE	FALSE	FALSE	0.30%	0.00%	0.00%	74.75	0	0	0.00%
450	4cb800503906a9e601d50141060e8a49		1122	1307	1340	FALSE	FALSE	FALSE	FALSE	0.32%	0.00%	0.00%	80.25	0	0	0.00%
1426	ef83a8dc040566537b1e812378fc048		1122	1343	41	FALSE	FALSE	FALSE	FALSE	0.00%	0.00%	0.00%	0.375	0	0	0.00%
1471	f6e2f231f56930f016a3edfd8706a60c		1082	1261	94	FALSE	FALSE	FALSE	FALSE	0.13%	0.19%	0.00%	32.25	49	0	0.12%
635	6bd6e4eb36e22f5346fd4b79d3b4636d		1020	1329	41	FALSE	FALSE	FALSE	FALSE	0.00%	0.00%	0.00%	0.25	0	0	0.00%
518	5ec9269c2280f20b2544ebf0d2bd1a	1042	1077	43	1192	TRUE	FALSE	FALSE	FALSE	0.00%	0.09%	0.00%	0	24.5	0	0.14%



예를 들어 A / C 를 기준으로 W 값을 1000 으로 설정하여 총 17 개의 ASV 가 필터링 된 것을 확인할 수 있다.

그림 3-2. Significant Taxon

## Pie Chart

Pie Chart 를 활용하여 각 단일 샘플 및 그룹에 대한 6 가지 Taxon 별 분포 비율을 알 수 있다. ASV 단위가 아닌 실제 Taxon 별로 합산한 상대적 비율을 Pie Chart 로 확인할 수 있다.

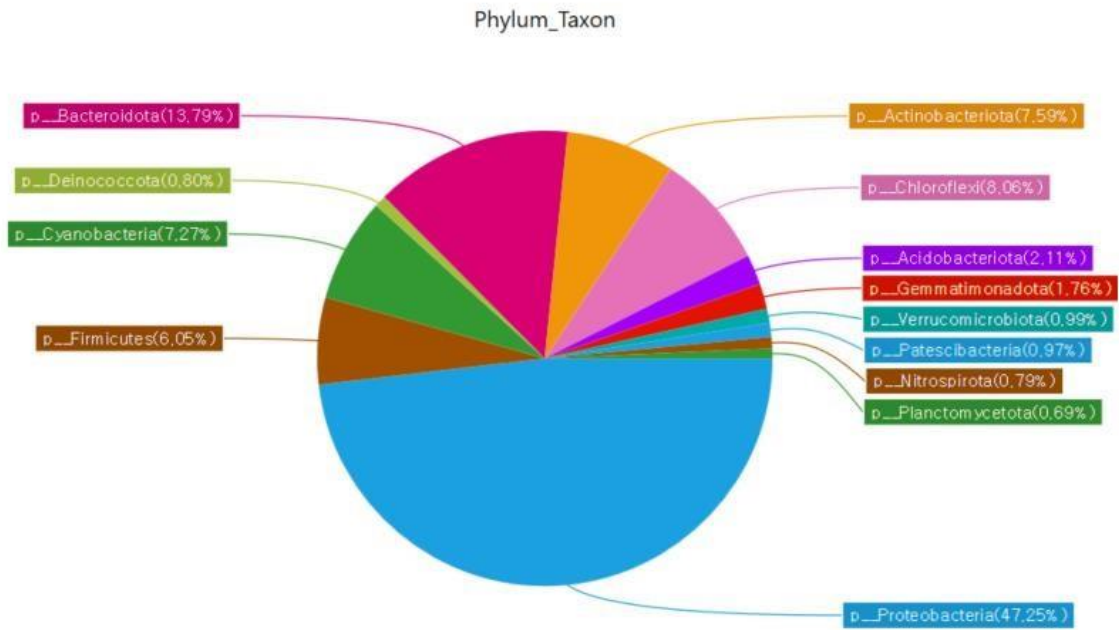


그림 3-3. Pie Chart

그림 3-3 은 단일 샘플에 대한 Phylum 에 대한 예시 파이 차트이다. 샘플 및 그룹에 대한 필터링이 적용된 풍부도를 알 수 있으며, Chart 내 taxa 를 클릭하면 필터링 된 결과만 엑셀창에서 확인할 수 있다. 단, 한 샘플만 선택해야 한다.



## Venn Diagram

원하는 샘플 및 그룹에서 ASV 에 대한 벤다이어그램을 제작할 수 있다. 최대 4 개까지 비교 가능하며, 선별 기준은 Relative Abundance 만으로 필터링할 수 있다.

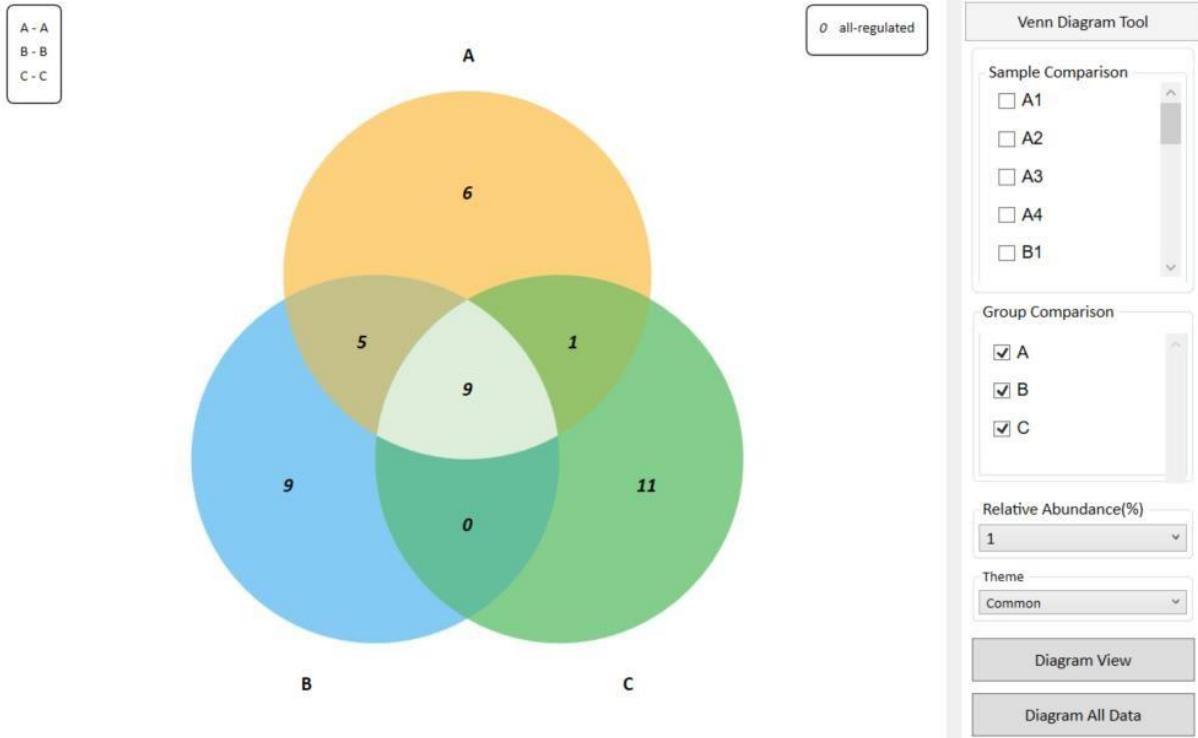


그림 3-5. Venn Diagram

그림 3-5 는 A, B, C 에 대해 Relative Abundance 1%로 선별하여 제작된 그림이다. 각 그룹별로 1% 이상인 ASV 행만 선별하여 공통된 region 과 그렇지 않은 region 에 대한 벤다이어그램이 제작된다. 특정 region 마우스 오른쪽 버튼으로 'Region All'을 클릭하면 해당 region 의 ASV 행만 선별하여 Excel 창에서 보여준다.

## 4. Graphic Plus

DA Analysis 부분에서 “GraphicPlus Start” 창을 펼치면 작동한다. Metagenome 분석 시각화에 대표적으로 사용되는 Bar Plot, Krona Chart, PCoA, Clustering heatmap, LDA effective size 을 제작할 수 있다.

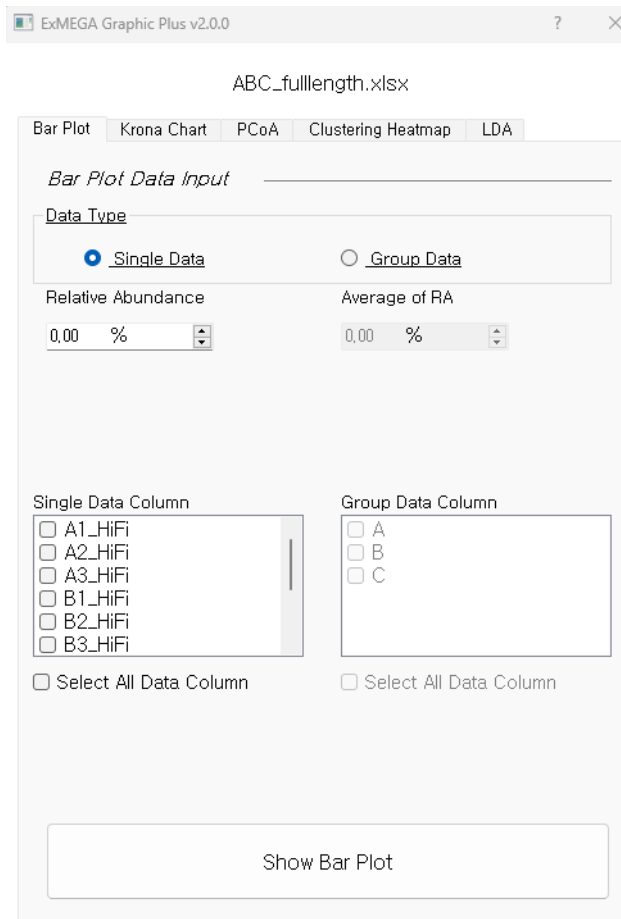


그림 4-0. Graphic Plus

모든 그래프는 단일 샘플 비교(Single Data)와, 그룹 비교(Group Data)가 모두 가능하도록 설계되어 있다. 단일 샘플비교는 Single Data 를 선택하고 Relative Abundance 를 threshold 로 데이터를 필터링을 진행하여 Bar plot 을 제작할 수 있다. 마찬가지로 그룹비교는 Group Data 를 선택하고 Average RA 로 원하시는 threshold 로 필터링을 진행하여 제작이 가능하다. Threshold 가 불필요하면 0.00% 로 설정하고 Bar plot 을 그리면 된다.

# Bar Plot

원하는 샘플 및 그룹에 대한 조건을 만족하는 Bar Plot 을 제작한다.

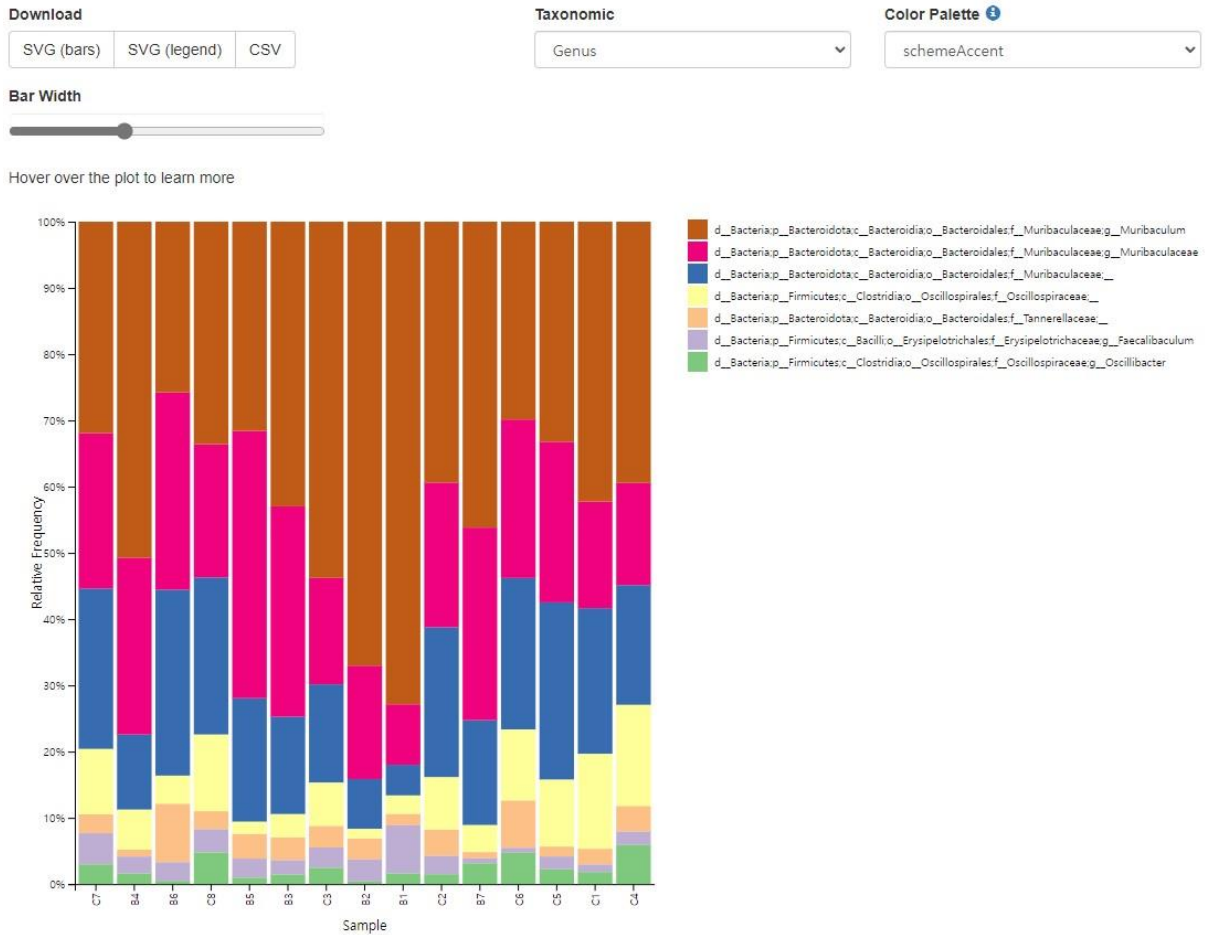


그림 4-1. Bar Plot

Bar Width 를 통해 Bar 크기를 조절할 수 있으며, Taxonomic 탭에서 모든 Taxonomy(총 7 가지 분류군)에 대한 그래프를 모두 확인할 수 있다. Bar Plot 을 필터링 없이 제작하면 낮은 Taxon 들로 시각화에 방해될 수 있다. 보통 Relative Abundance 값을 0.1~1%로 설정하여 낮은 분류군을 필터링하는 방법으로 많이 사용한다.

# Krona Chart

KronaTools 를 사용한 Krona chart 는 확대/축소가 가능한 multi-layered pie chart 이다. 이바이오젠에 선 필터링 조건을 적용한 Krona Chart 를 매크로 방법으로 제작한다. 여러 샘플을 선택할 경우 Chart 를 그리는데 시간이 오래 소요될 수 있다. 사용 방법과 자세한 설명은 해당 매뉴얼 페이지 에 자세히 설명되어 있다. <https://github.com/marbl/Krona/wiki/Browsing-Krona-charts>

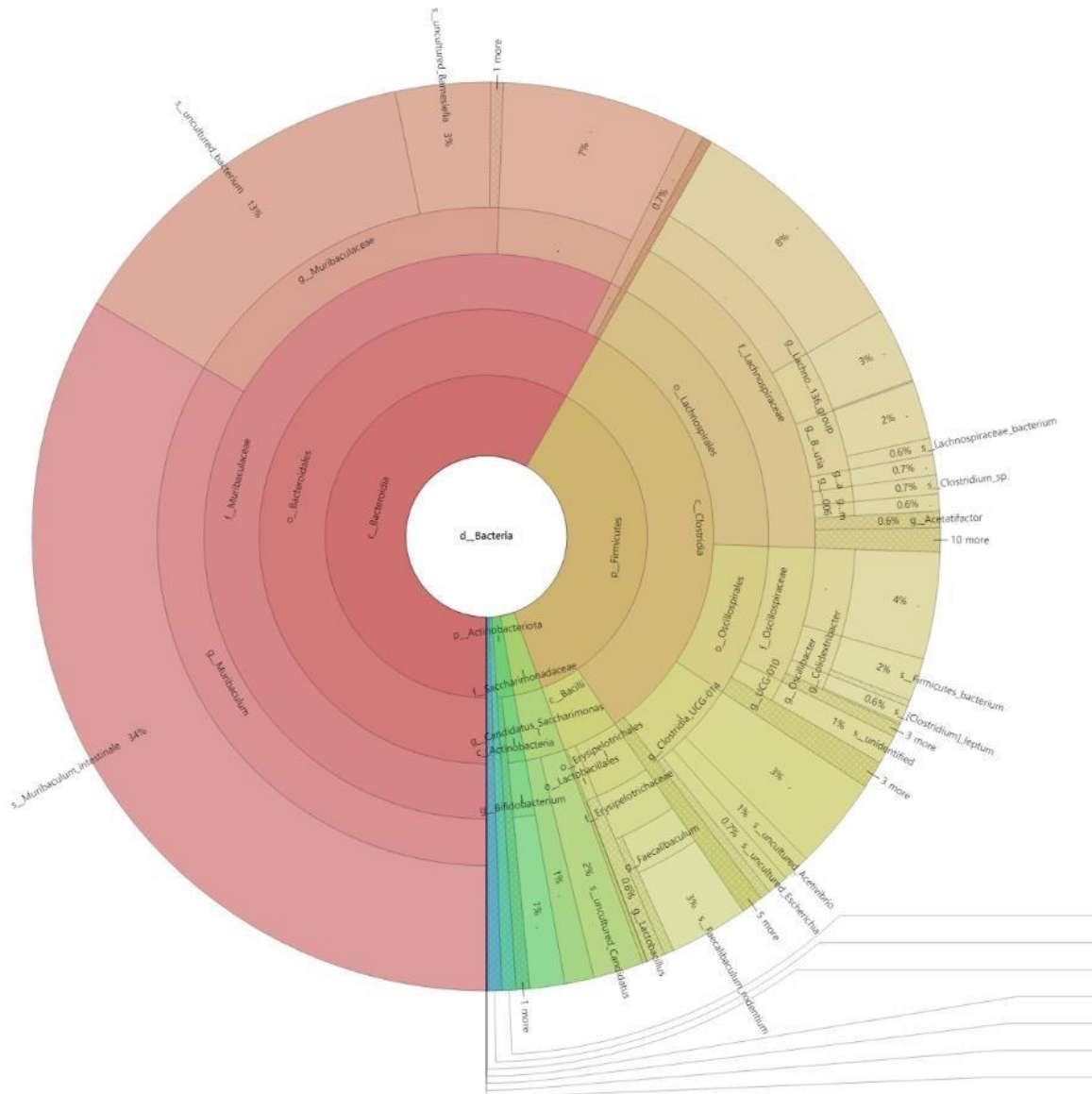


그림 4-2. Krona Chart

## PCoA

PCoA(Principal Coordinates Analysis)는 Metagenome 에 사용되는 대표적인 Beta diversity(샘플 간 분석) 분석으로 샘플간 유사성을 거리로 계산하는 차원 축소 개념이다. PCoA 차원은 2D/3D Type 중 선택하며, Metric type 은 distance matrix 알고리즘으로 대중적으로 사용하는 braycurits 와 jaccard 방법 중 선택하여 사용할 수 있다. Marker Type 을 Group 으로 설정하면 각 샘플에 대한 그룹 정보도 확인할 수 있다. "Save Value Options" 탭에서는 PCoA 에 활용한 각 수치 값들을 파일로 저장할 수 있다. PCoA 도 다른 그래프와 마찬가지로 필터링을 적용하여 유의미한 Taxon 분석할 수 있다.

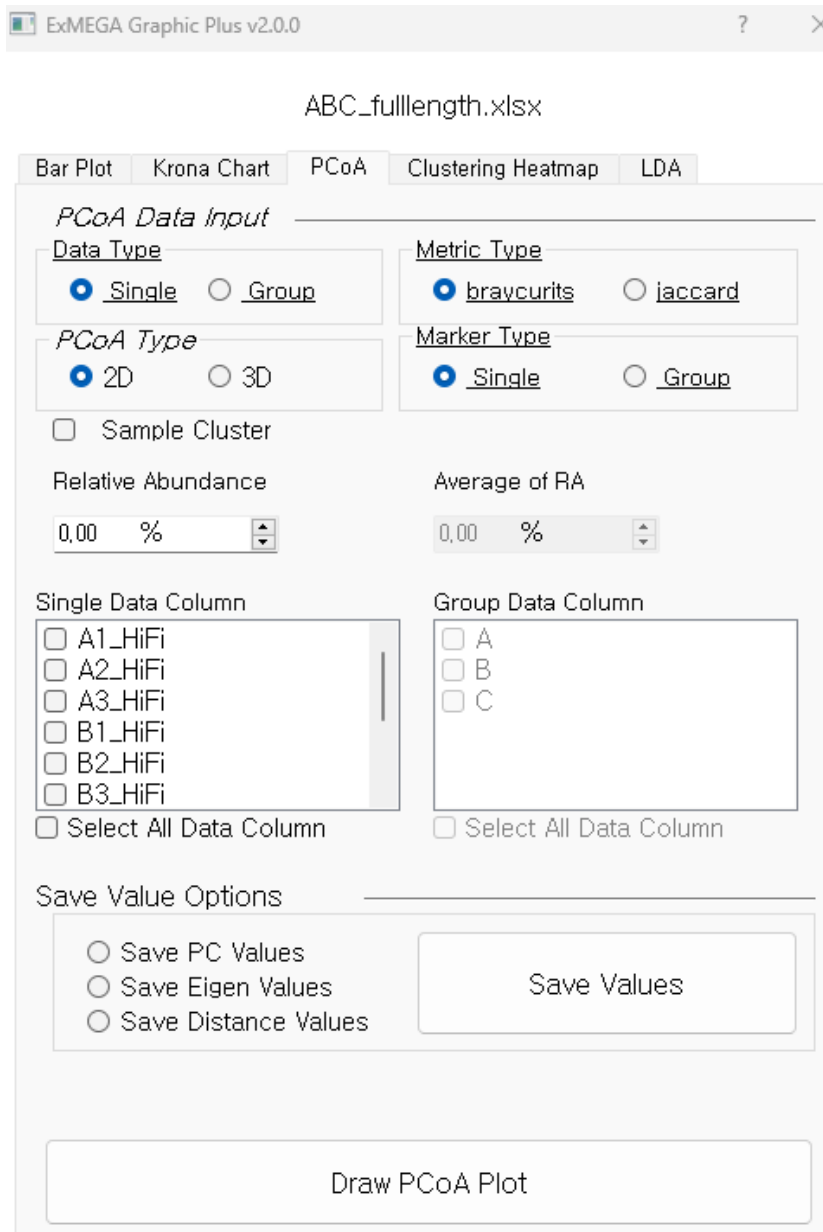


그림 4-3-A. PCoA

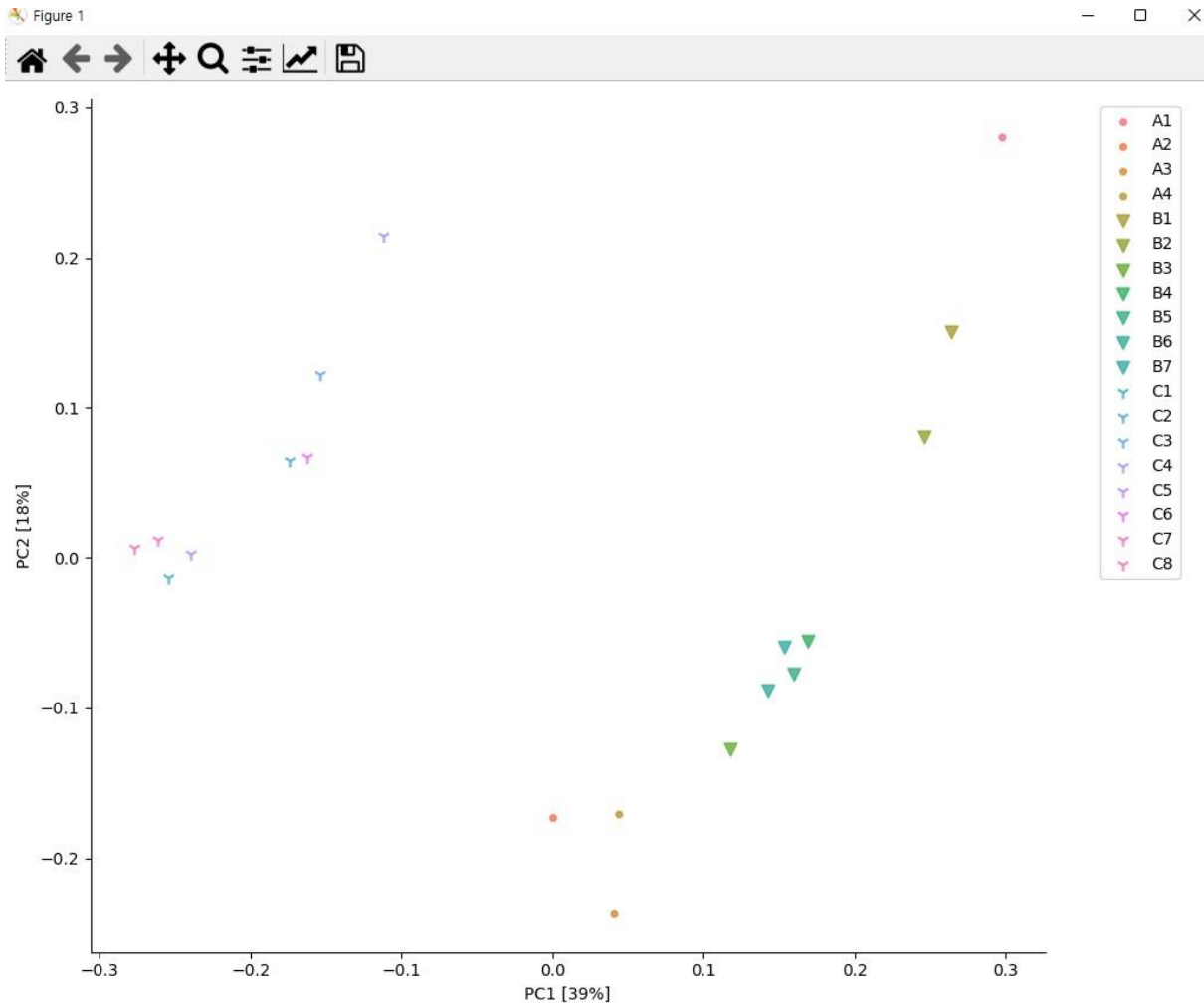
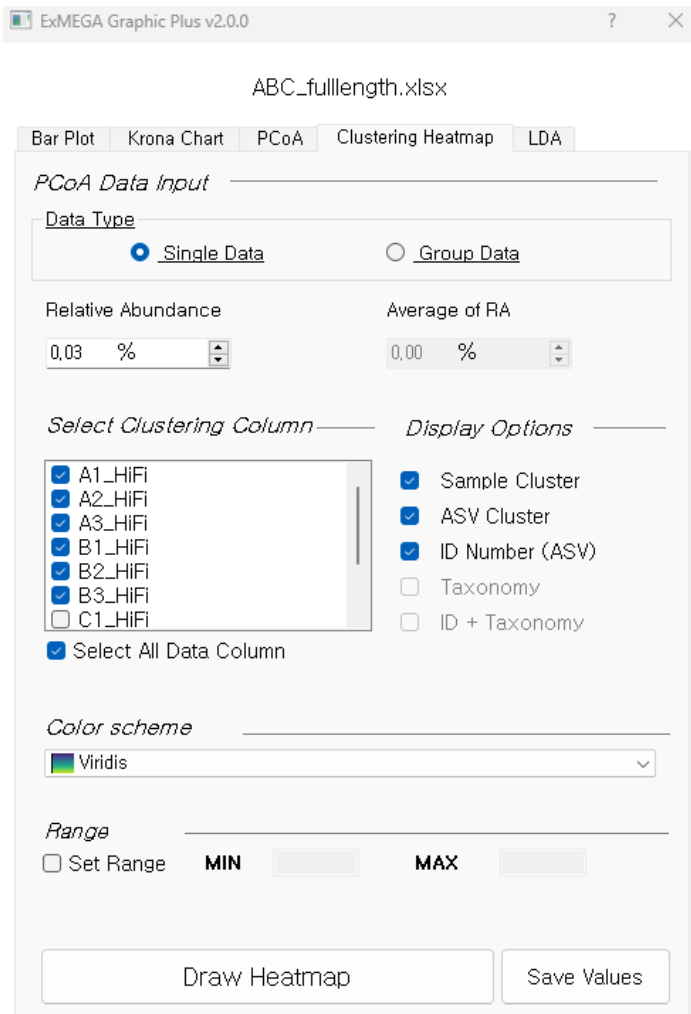


그림 4-3-B. PCoA2

그림 4-3-B 는 2D 로 분석한 샘플들 간의 PCoA 결과이다. 각 샘플들은 A / B / C 샘플들끼리 유사함을 알 수 있다.

# Clustering Heatmap

Clustering Heatmap 은 ASV 마다 Relative Abundance(%) 비교를 시각화 할 수 있다. Relative Abundance 값은 0.05~0.5%까지 추천하며, 너무 높은 값을 필터링하면 많은 ASV 가 제거되어 그래프가 제작되지 않는다.



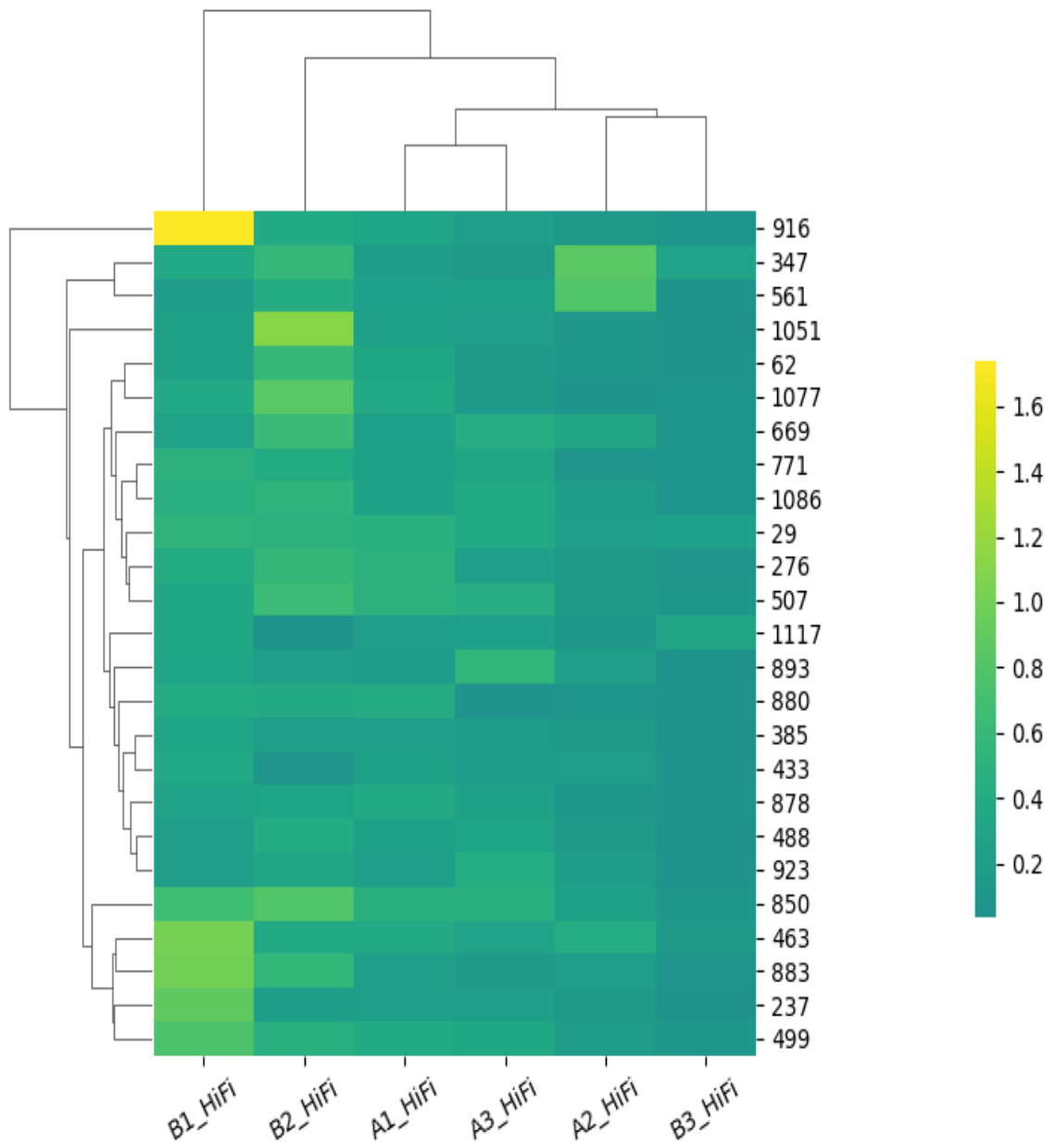
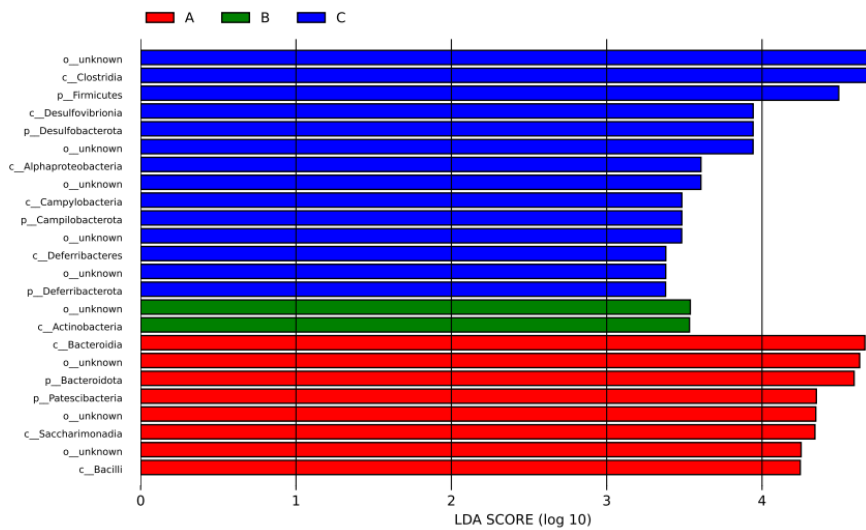
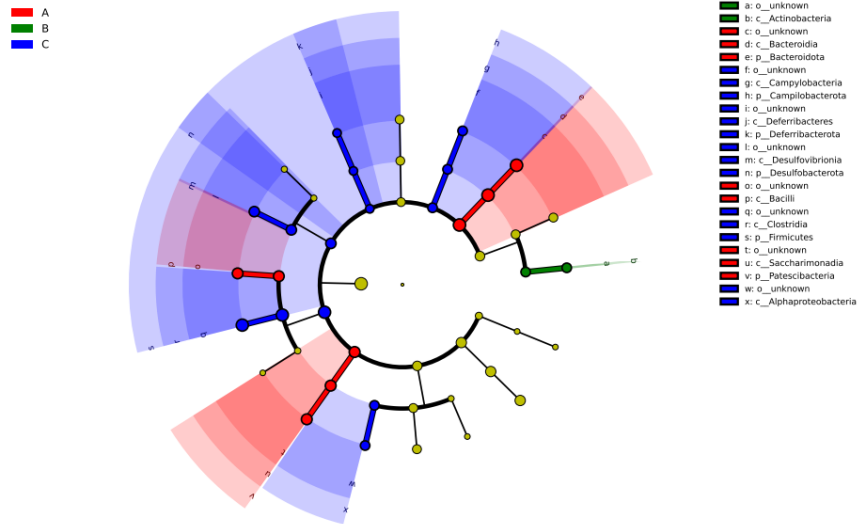


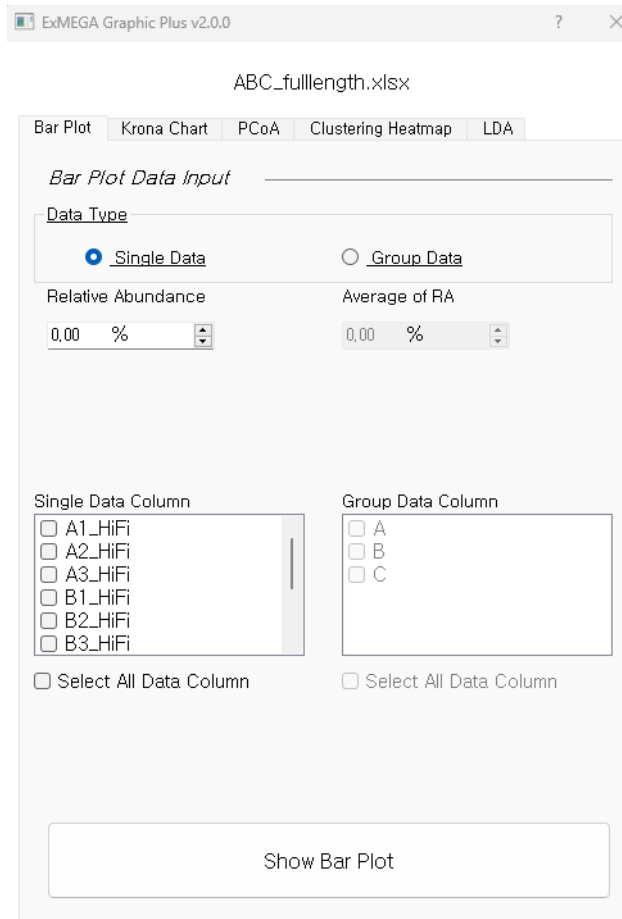
그림 4-4-A. Clustering Heatmap



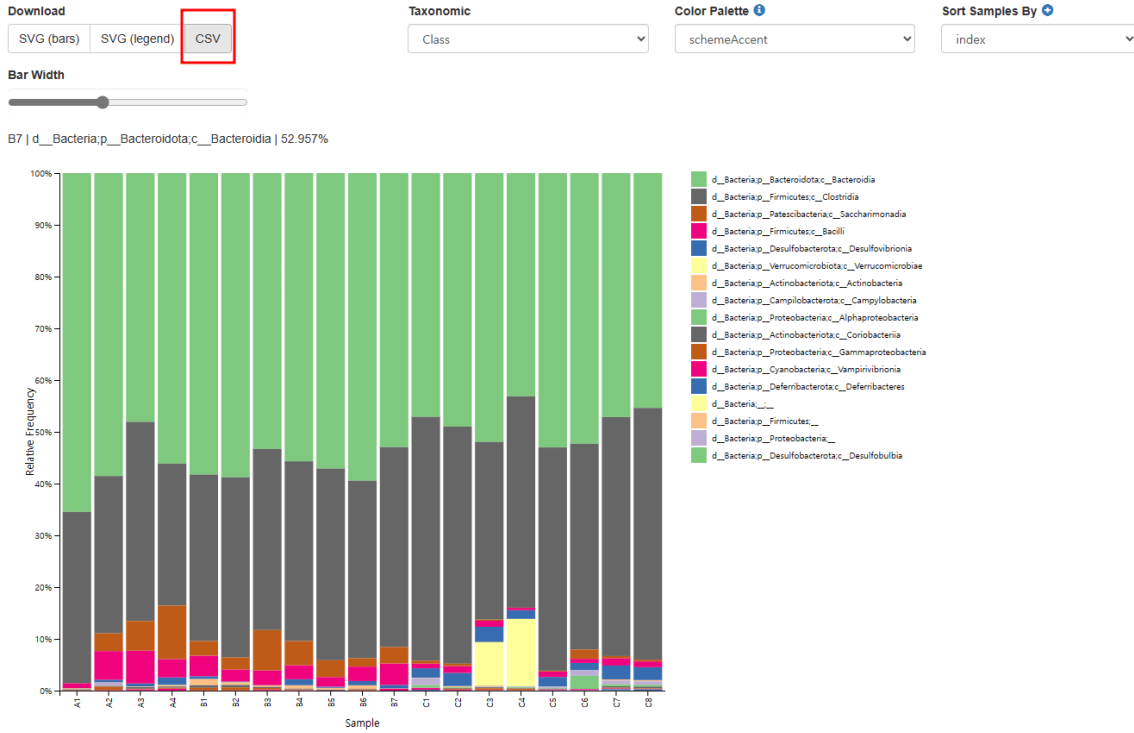
# LEfSe (Linear discriminant analysis Effect Size)

LDA 는 차원축소 기법 중 하나이고, LEfSe 는 LDA 기법을 이용하여 그룹을 비교할 때, 각 그룹에서의 특징적인 미생물들 즉, 클래스 간의 차이를 설명할 가능성이 가장 높은 특징을 가지고 있는 미생물들이 어떤 미생물인지를 찾아내는 분석 방법이다.

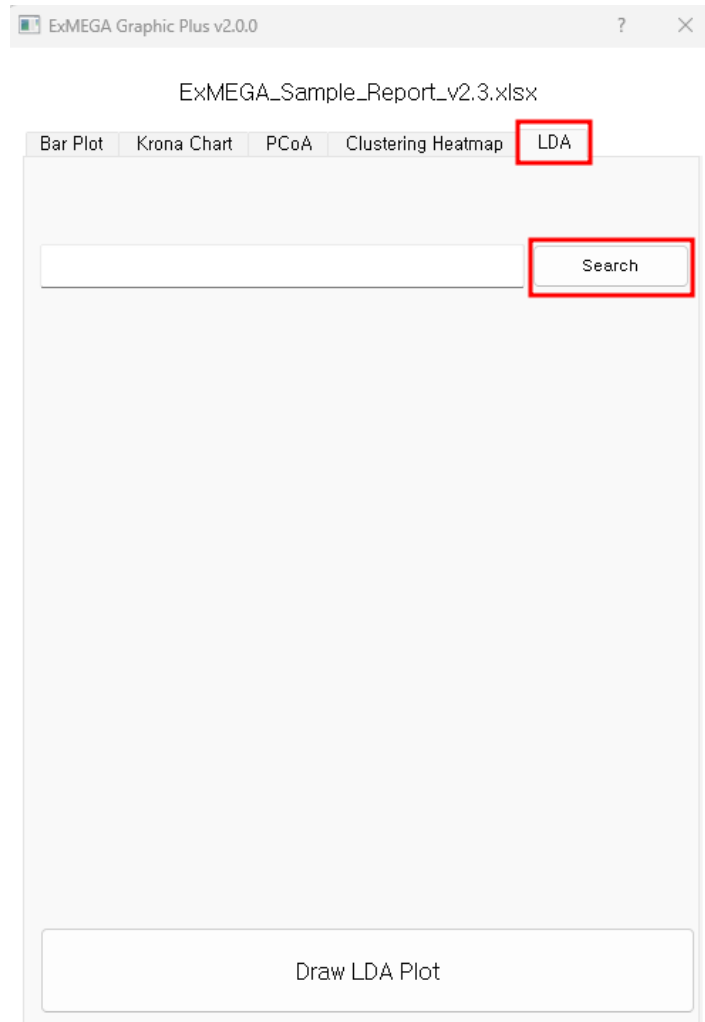





첫 버튼인 Bar Plot 을 클릭하고 원하는 설정을 선택하고 Bar plot 을 제작한다. 그 다음 CSV 버튼을 클릭하여 relative abundance matrix 를 다운로드 한다.



그리고 LDA 에서 search 로 다운로드한 level-x.csv 파일을 불러온후 Draw LDA Plot 을 클릭하면 분석결과를 선택할 곳을 지정할 수 있고 선택이 완료되면 다시 한번 Draw LDA Plot 을 클릭하면 분석결과를 지정된 폴더에서 확인할 수 있다.



분석결과는 LDA\_level-x\_ExMEGA 레포트 파일명으로 생성된 폴더에서 확인이 가능하며 LDA effective size 분석결과는 cladogram.pdf, LDA score 는 the\_list\_of\_biomarkers.pdf 에서 확인할 수 있고, biomarkers\_raw\_images 에서는 biomarker 별로 각 그룹에서의 분포를 확인이 가능하다.

 LDA\_level-3\_ExMEGA\_Sample\_Report\_v2