Microbial Diversity Using Gene-targeted Pyrosequencing

Design style of high-throughput data

Tae Kwon Lee

WCU Center for Green Metagenomics

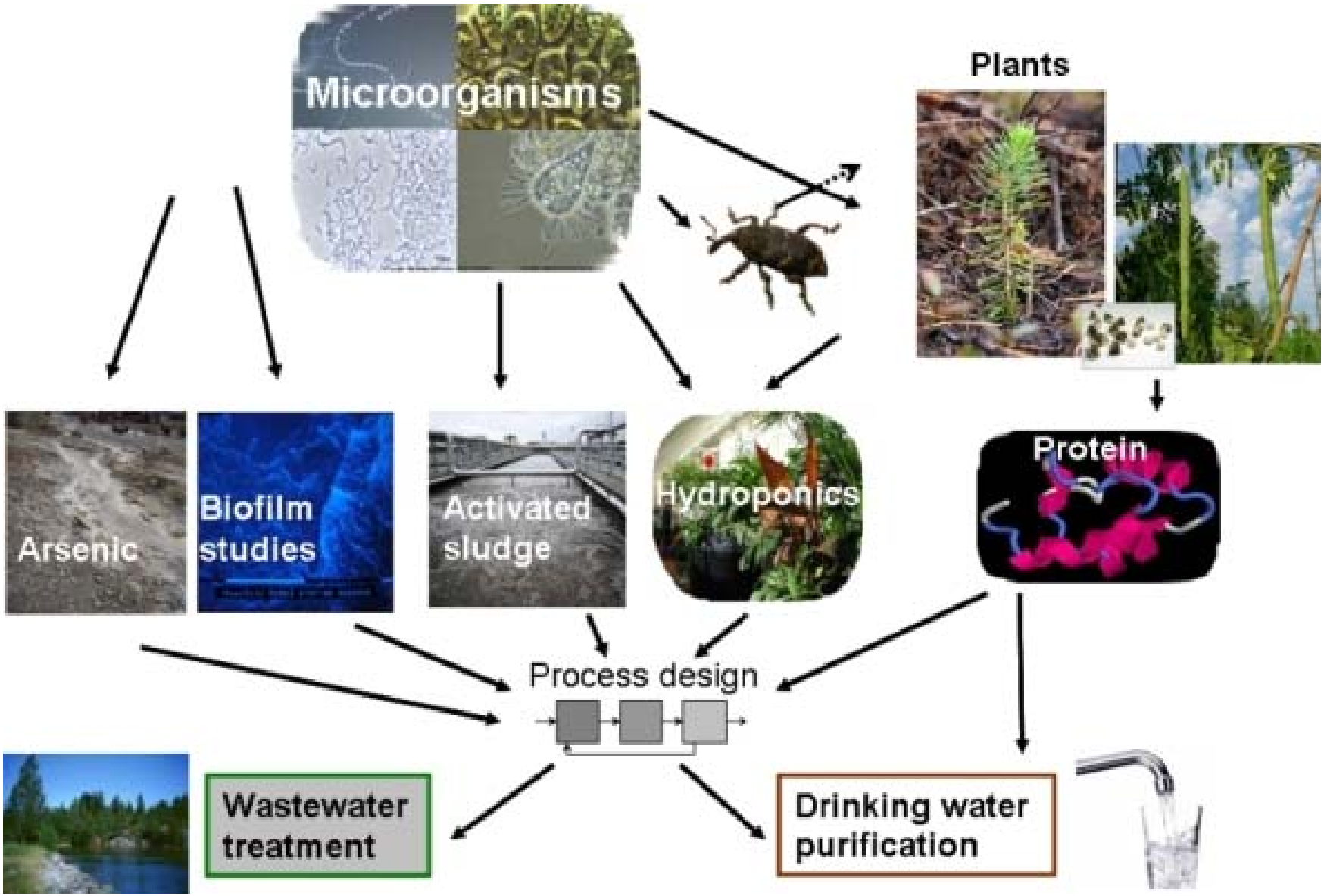
To Bore

No More

1. **Can we determine the true microbial diversity (phylogenetically and genetically)?**

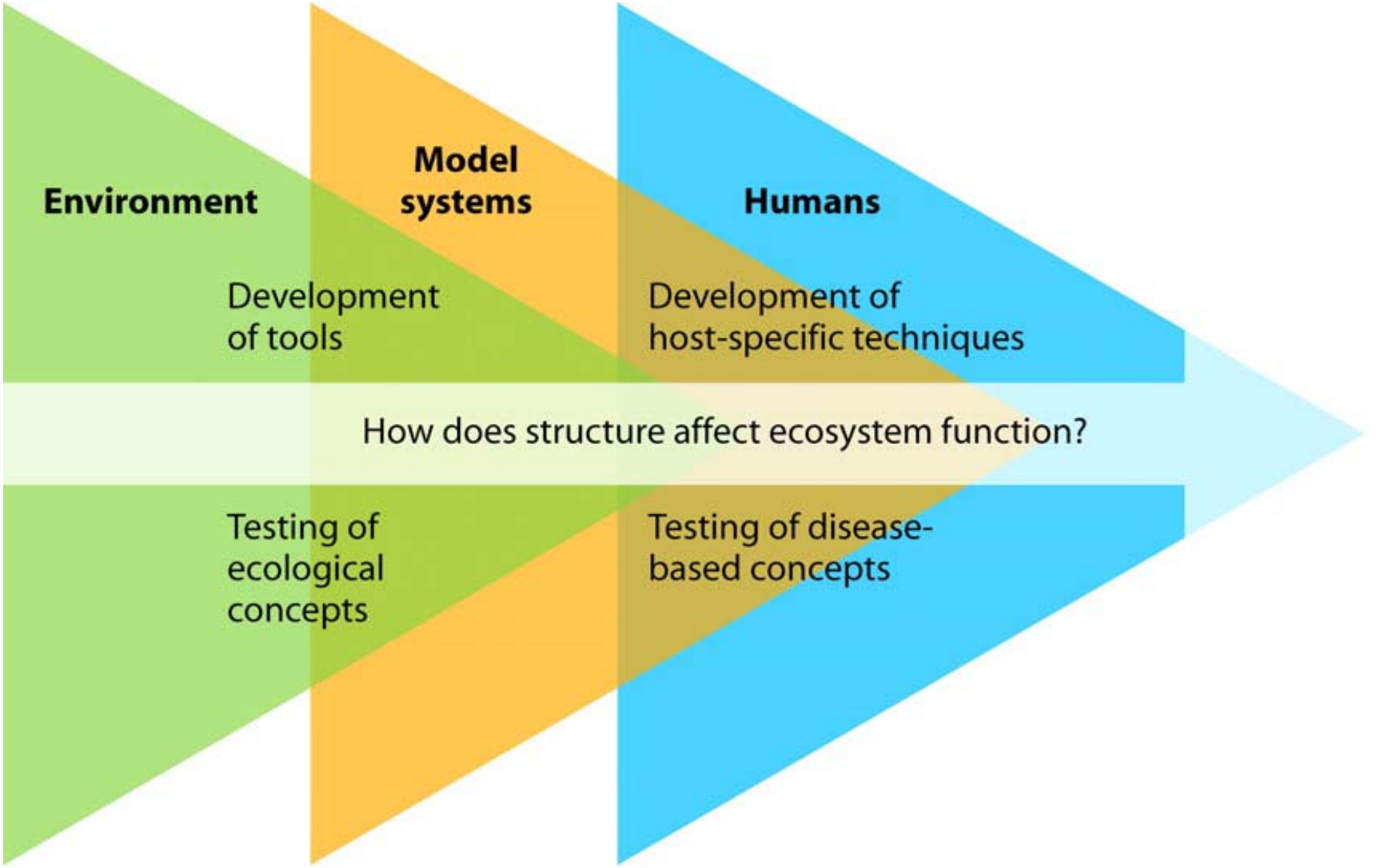
2. **Is there a core microbiome?**

3. **Does a structure impact ecological function? If so, how?**

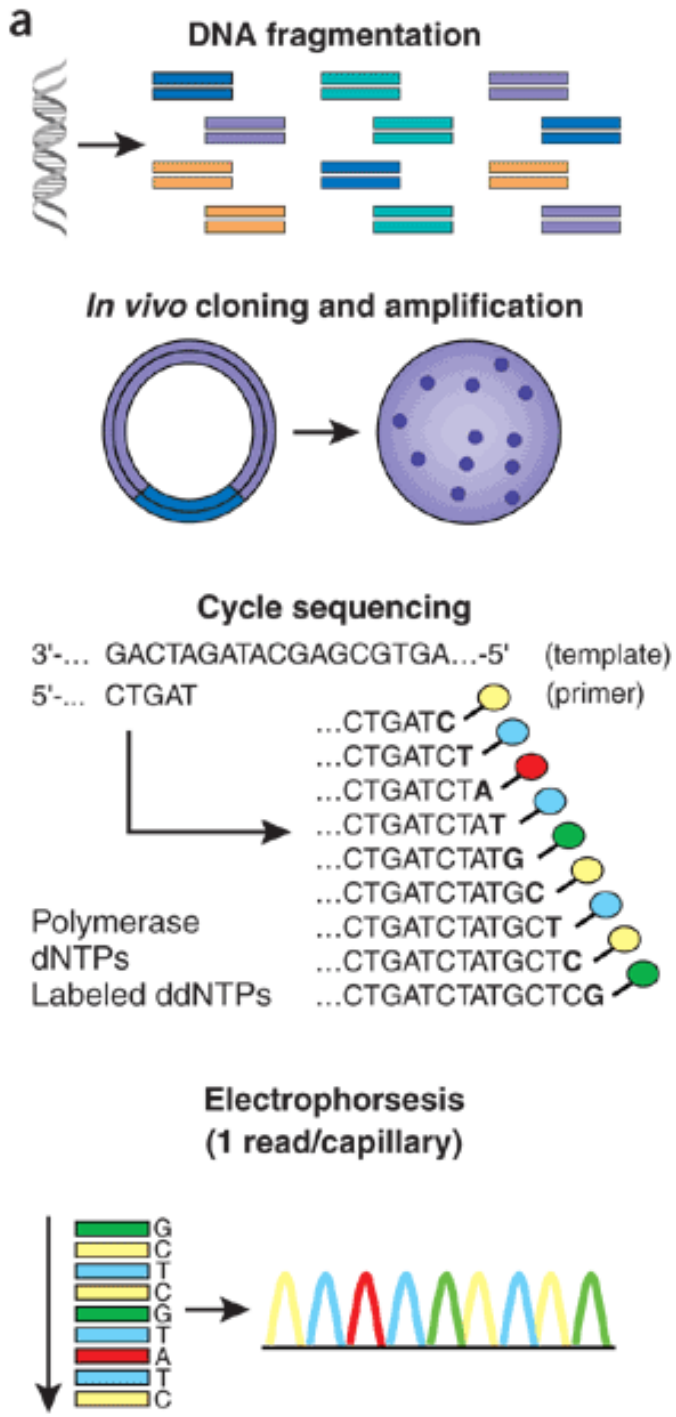


2011, Environmental microbiology, KTH biotechnology

Human Microbiome Project



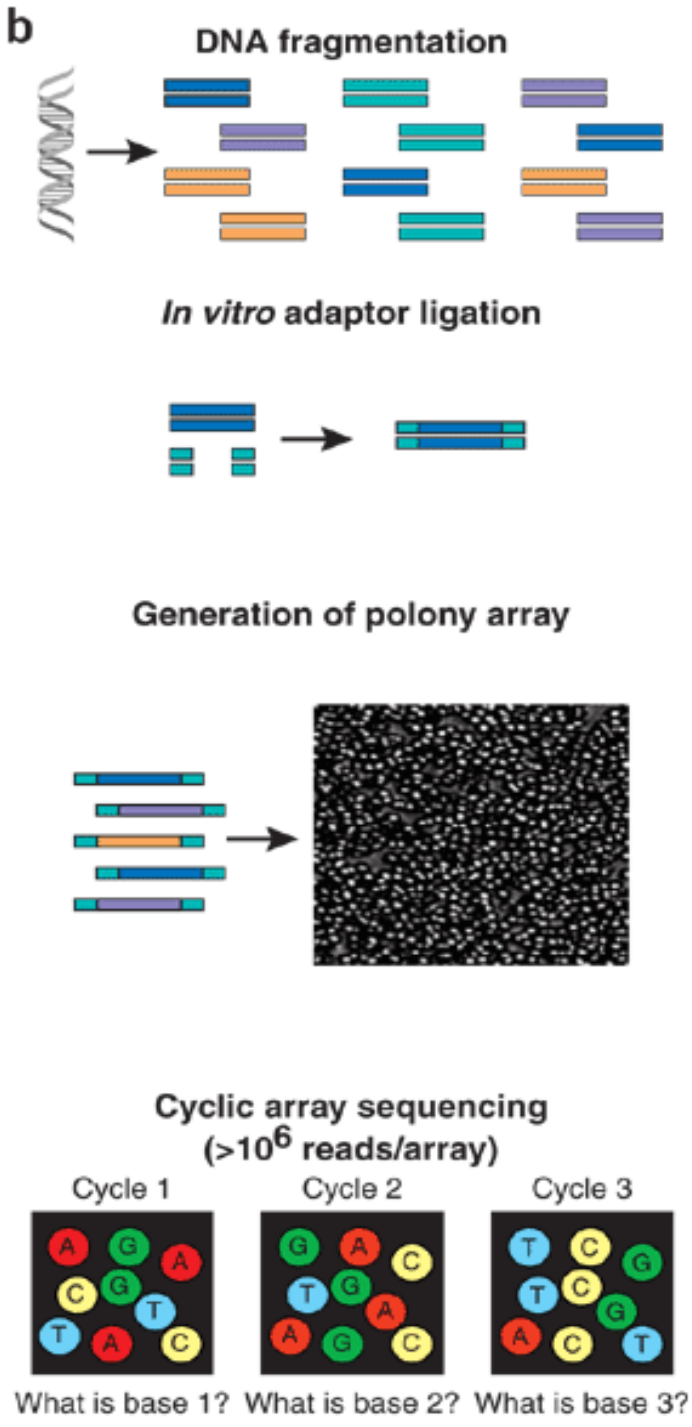
2010, Microbiol. Mol. Biol. Rev. Robinson et al.



Shotgun

Sanger Sequencing

2008, Nature Biotechnology, Shendure & Ji



Next Generation Sequencing

2008, Nature Biotechnology, Shendure & Ji

2nd NGS

The New
GS Junior

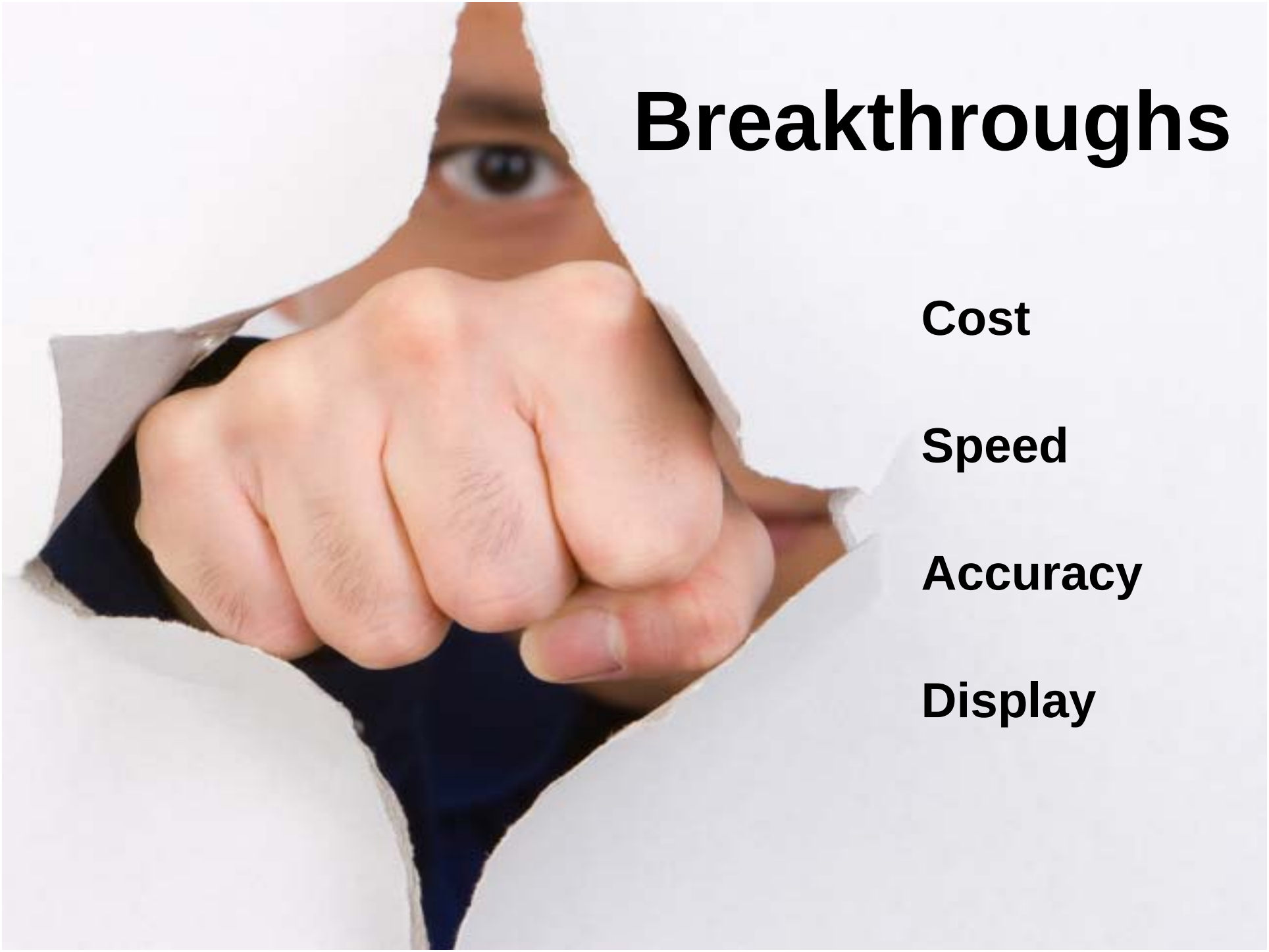
MiSeq

3rd NGS

ion torrent

PACBIO *RS*


Drowned in next generation sequencing data
HELP!
http://biocomicals.blogspot.com



Breakthroughs

Cost

Speed

Accuracy

Display

1. **Can we determine the true microbial diversity (phylogenetically and genetically)?**

2. **Is there a core microbiome?**

3. **Does structure impact ecological function? If so, how?**

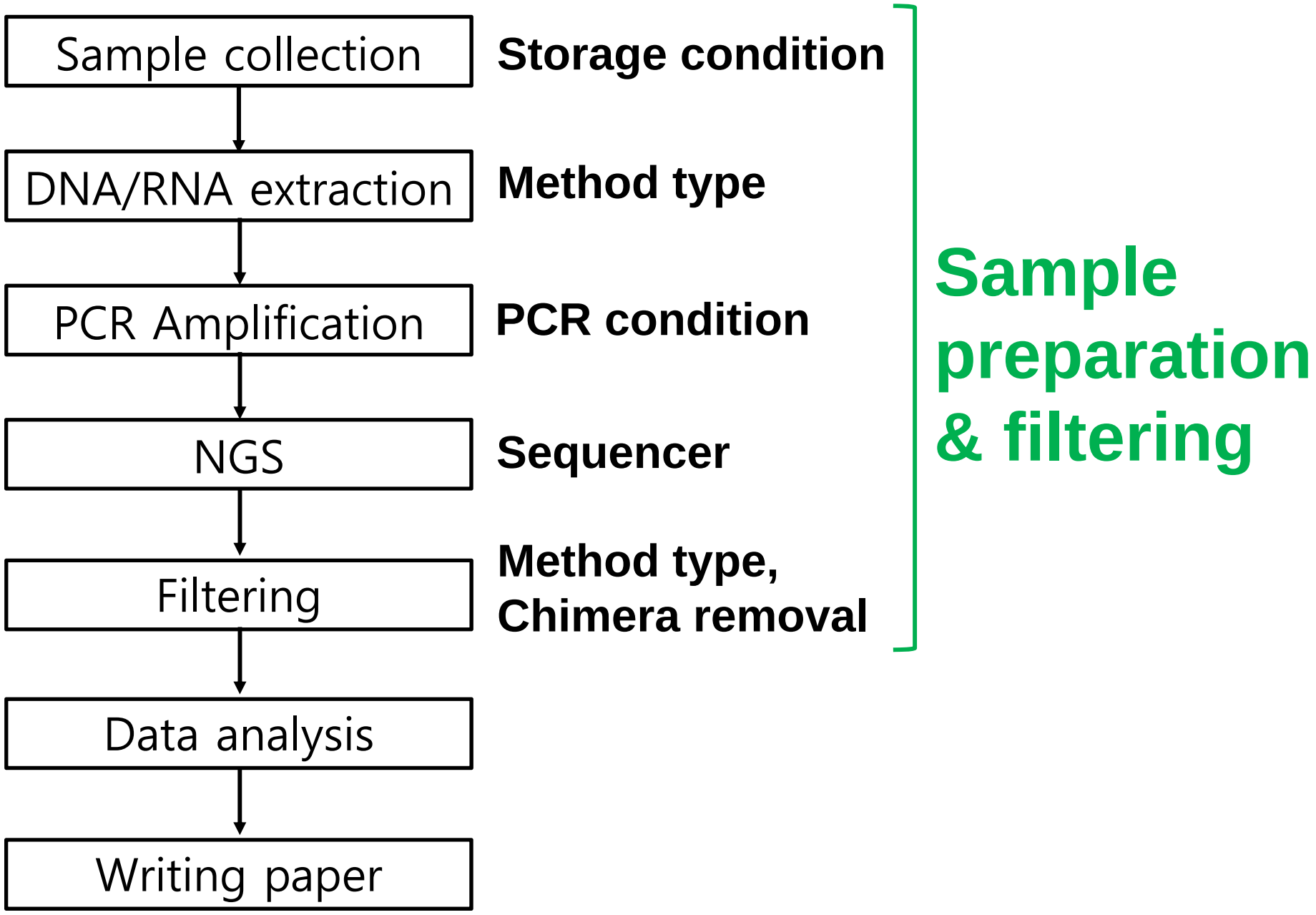
Steps of

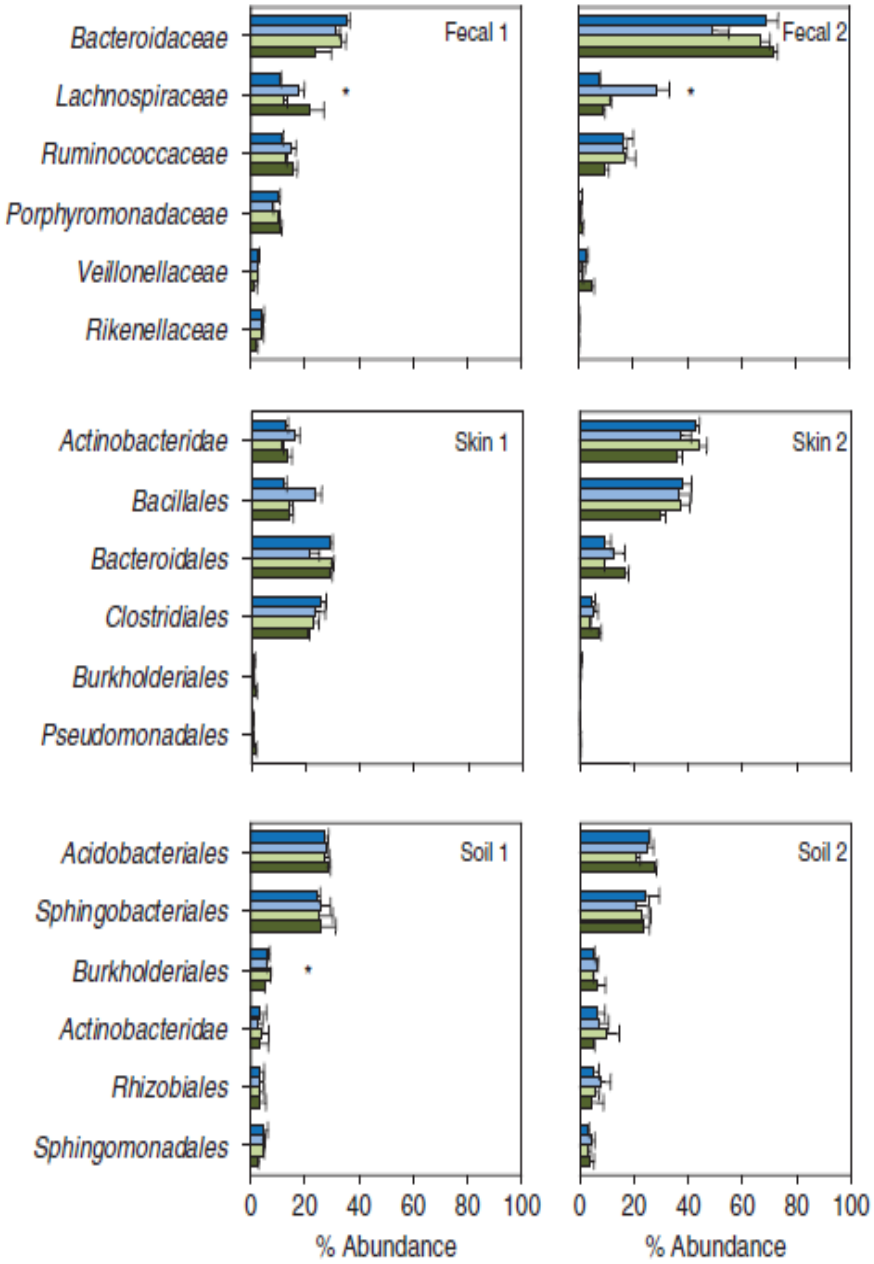
Microbial analysis

1. Sample preparation & filtering
2. Microbial structure & diversity
3. Interaction with function

Step 1

Sample preparation and filtering

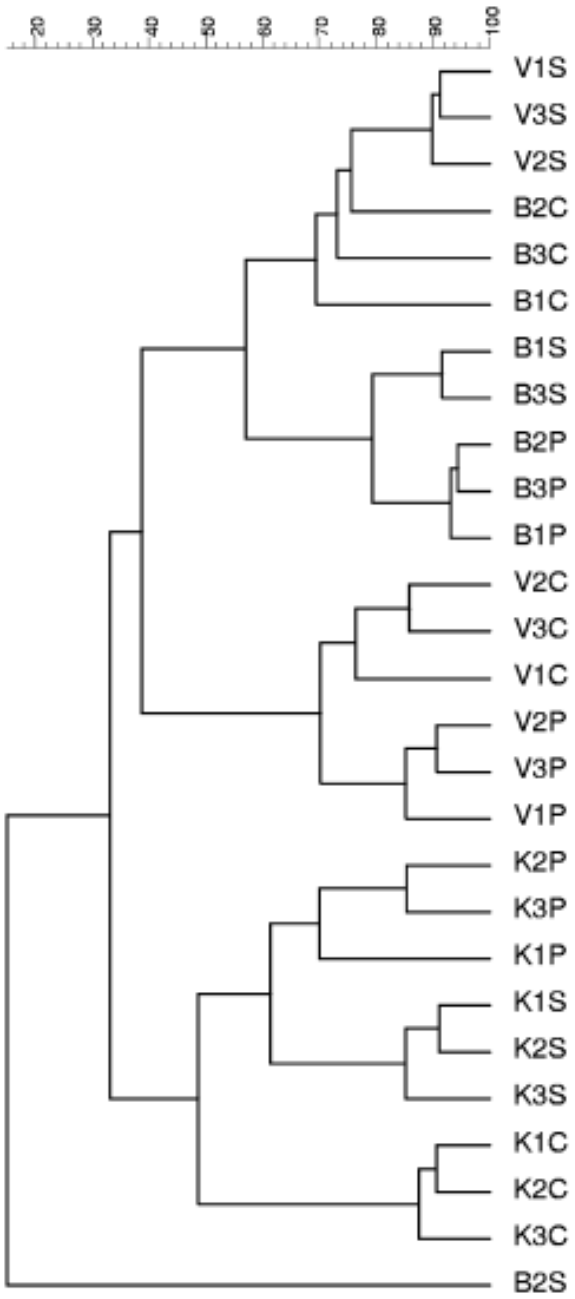

Sample collection
DNA/RNA extraction
PCR Amplification
NGS
Filtering
Data analysis
Writing paper
Storage condition
Method type
PCR condition
Sequencer
Method type,
Chimera removal
Sample
preparation
& filtering



Less effect of storage condition

For RNA, use RNALater

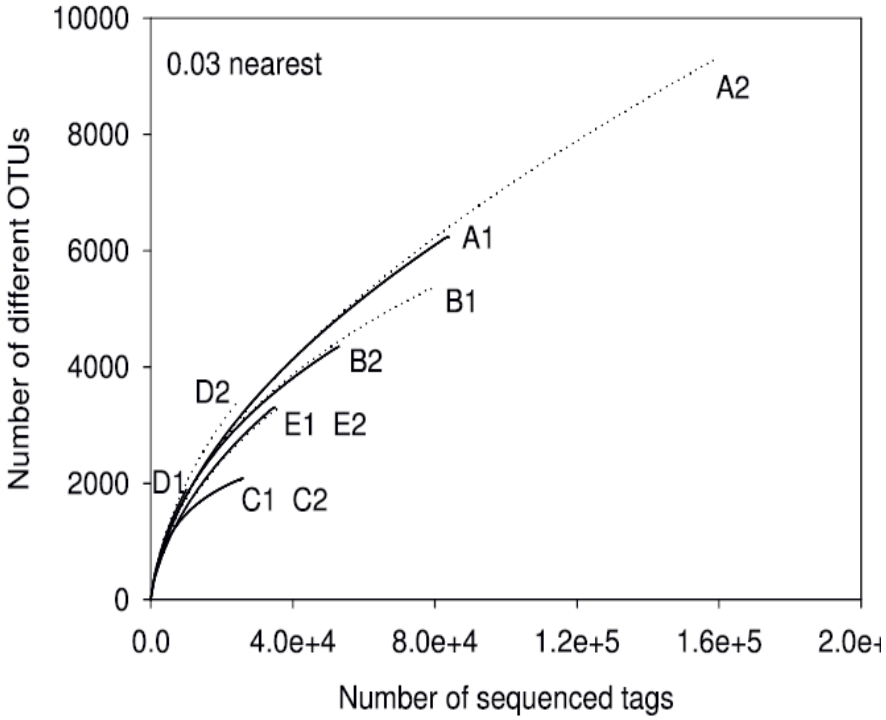
2010, FEMS Microbiol Lett. Lauber et al.

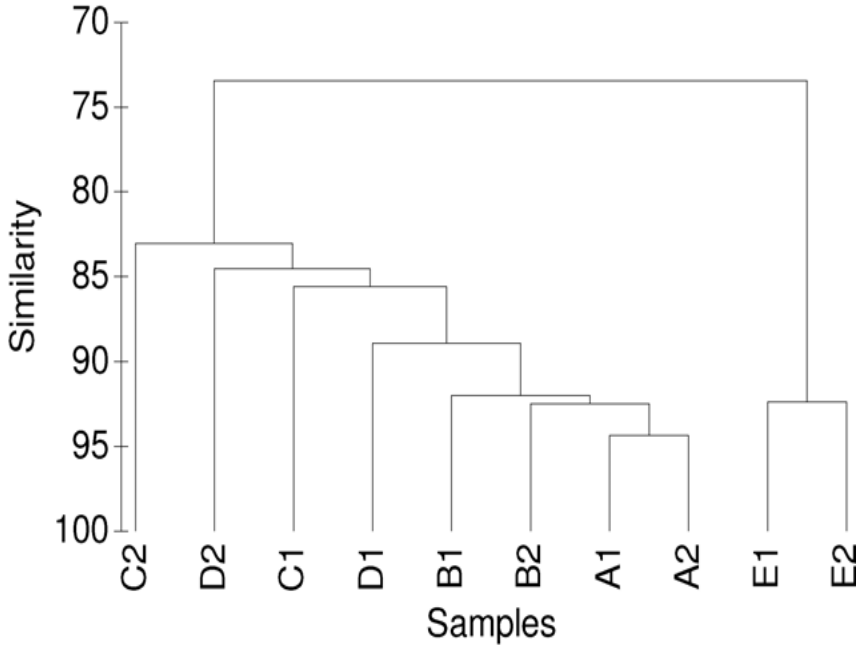


Use of same

extraction method

2010, Appl. Environ. Microbiol. Inceoglu et al.

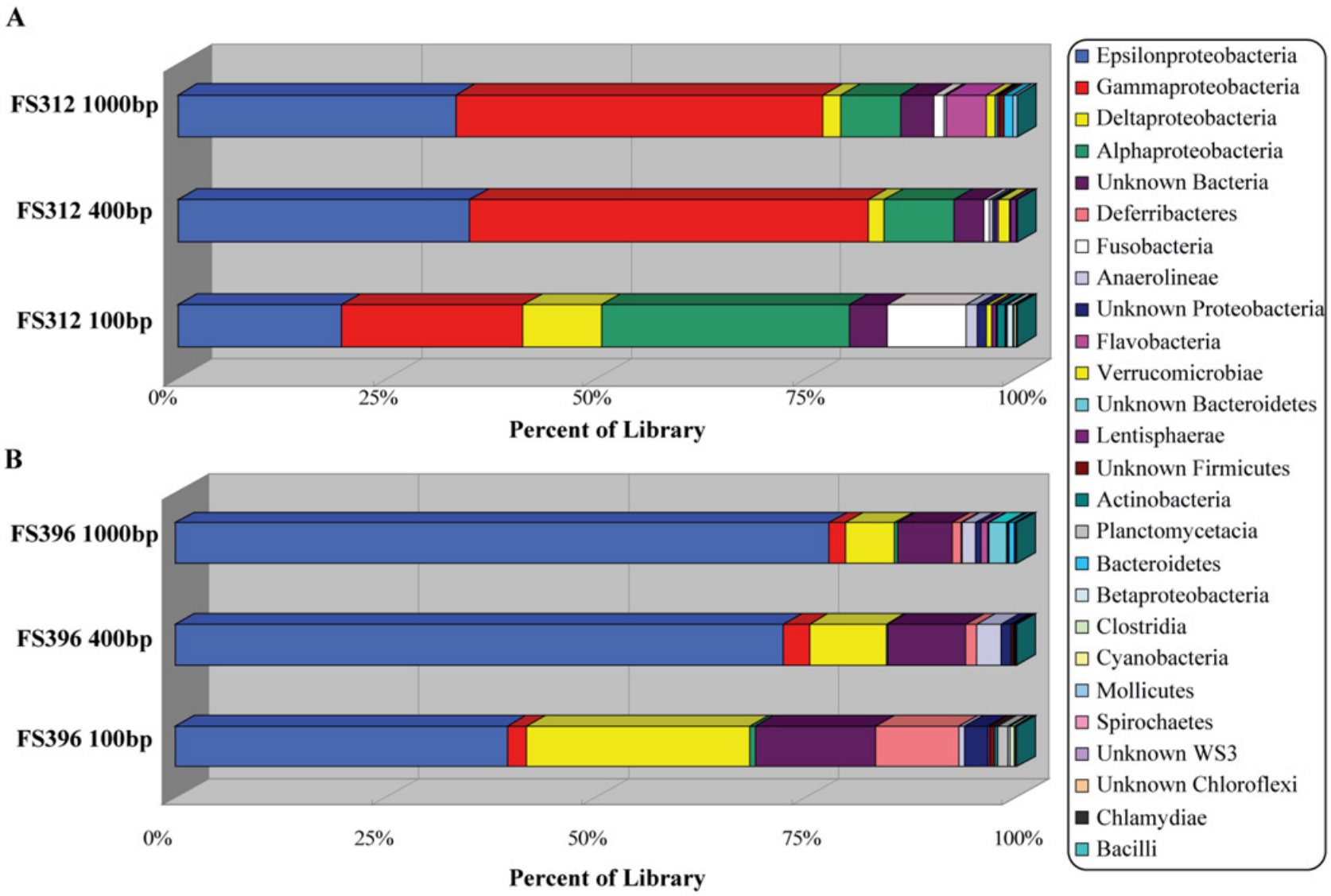




Less effect of Dilution/Cycle

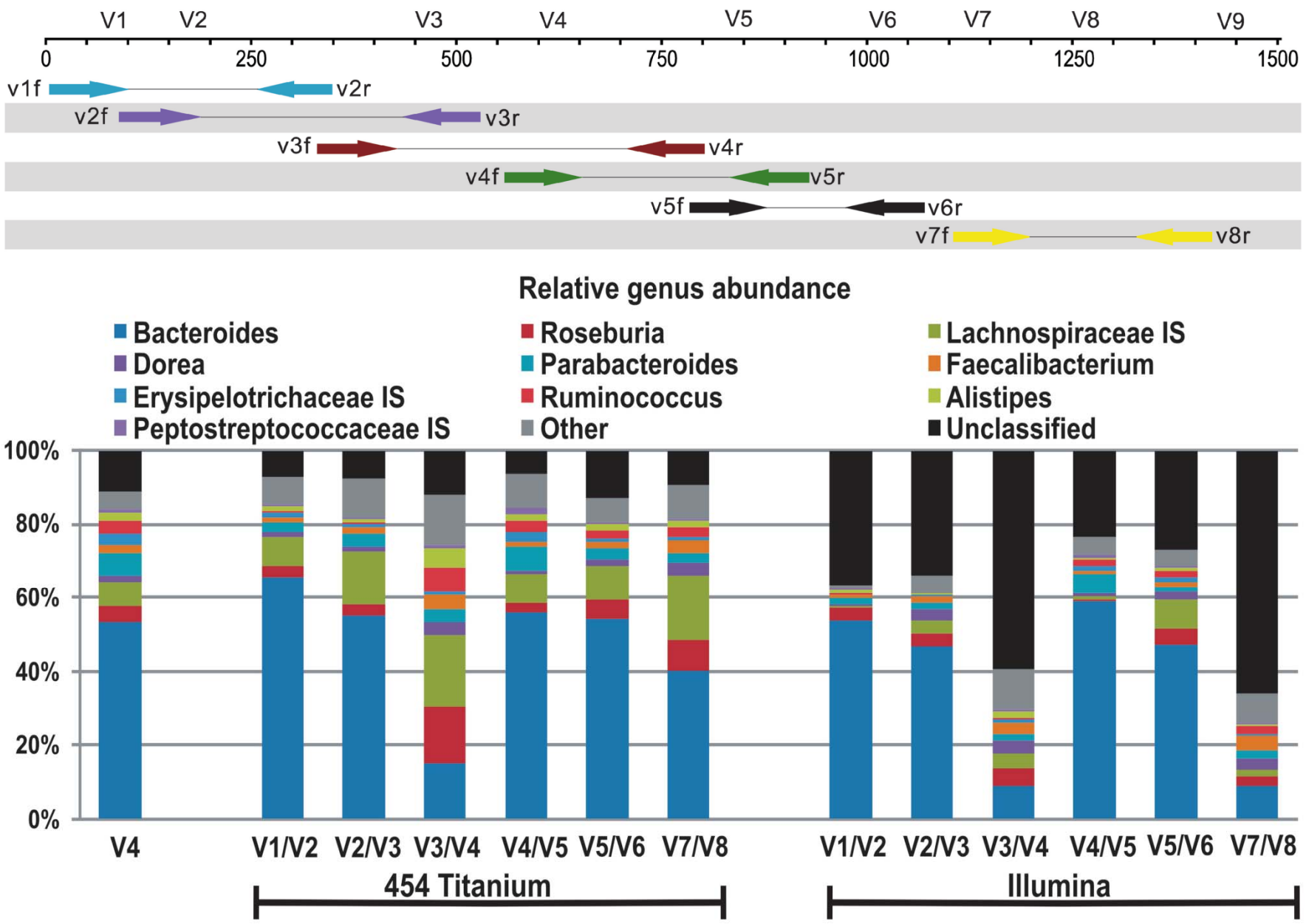
Much effect of Polymerase

2010, BMC Microbiol. Wu et al.



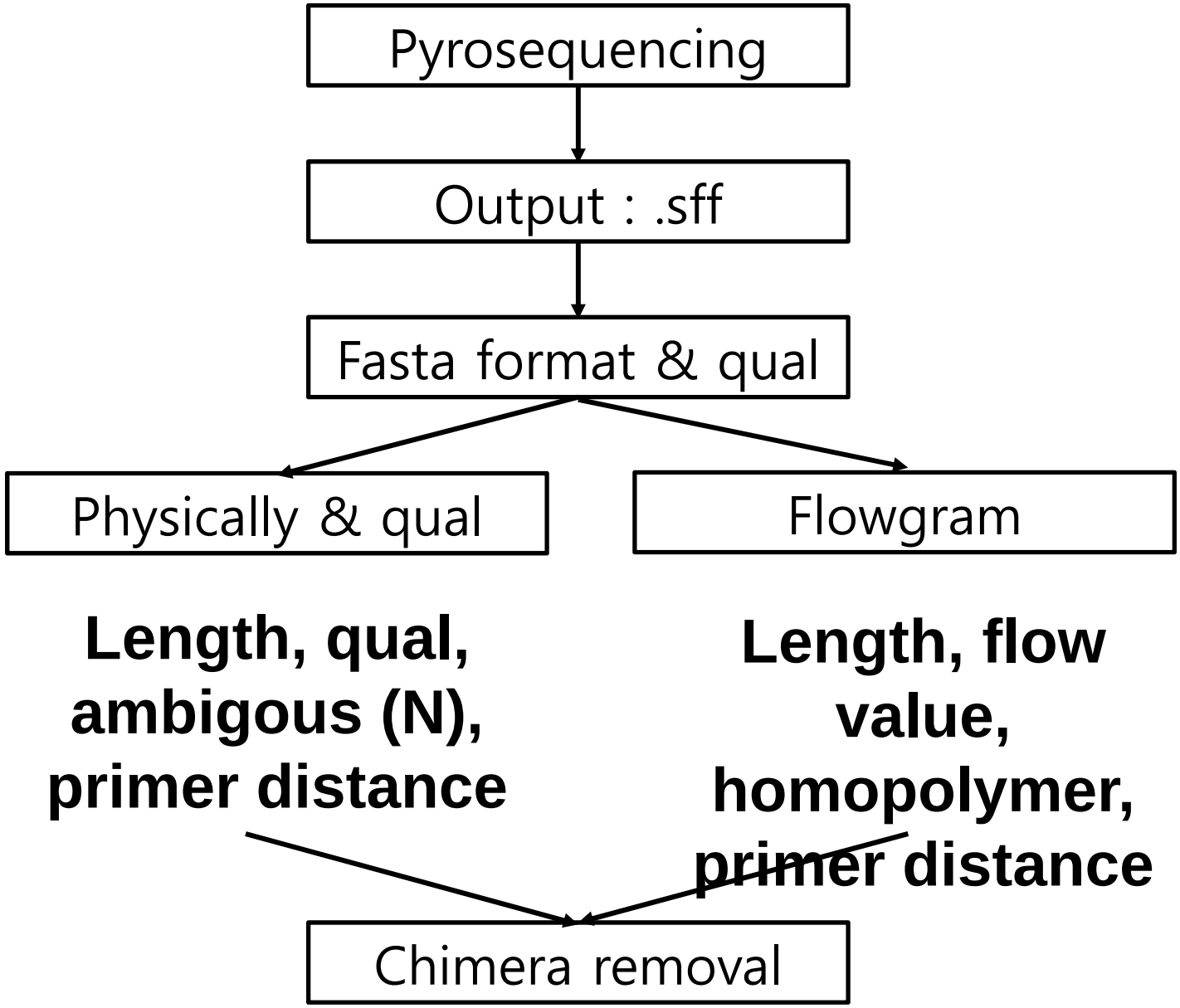
Same primer set and length

2009, Environ. Microbiol. Huber et al.

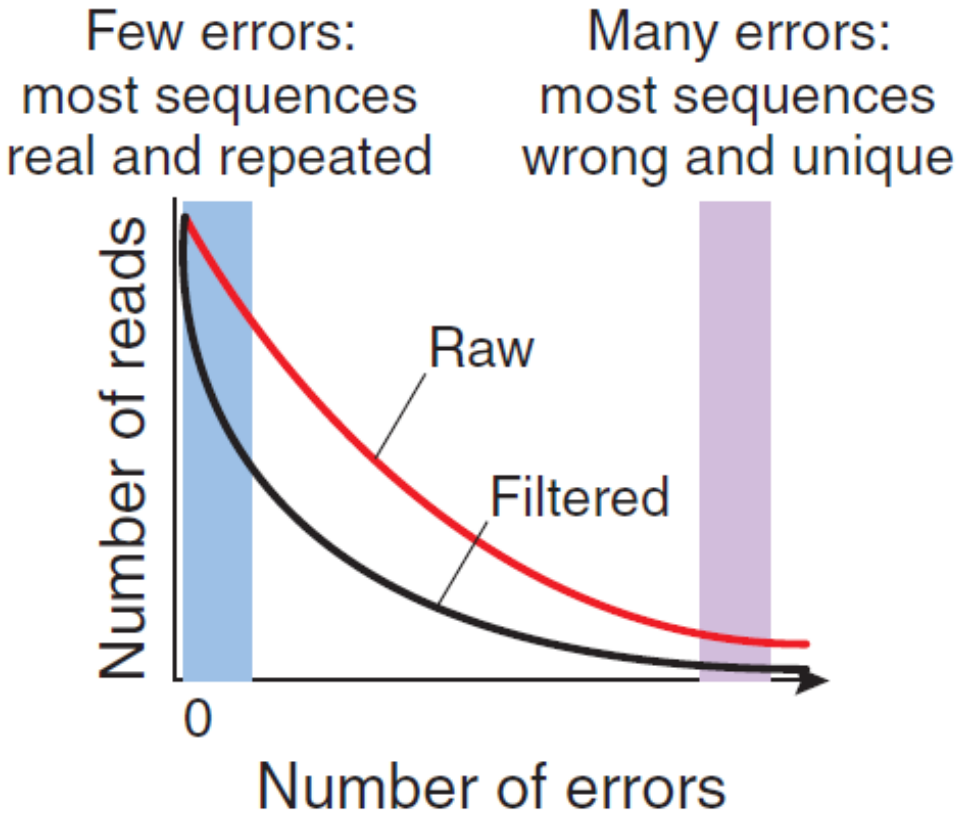


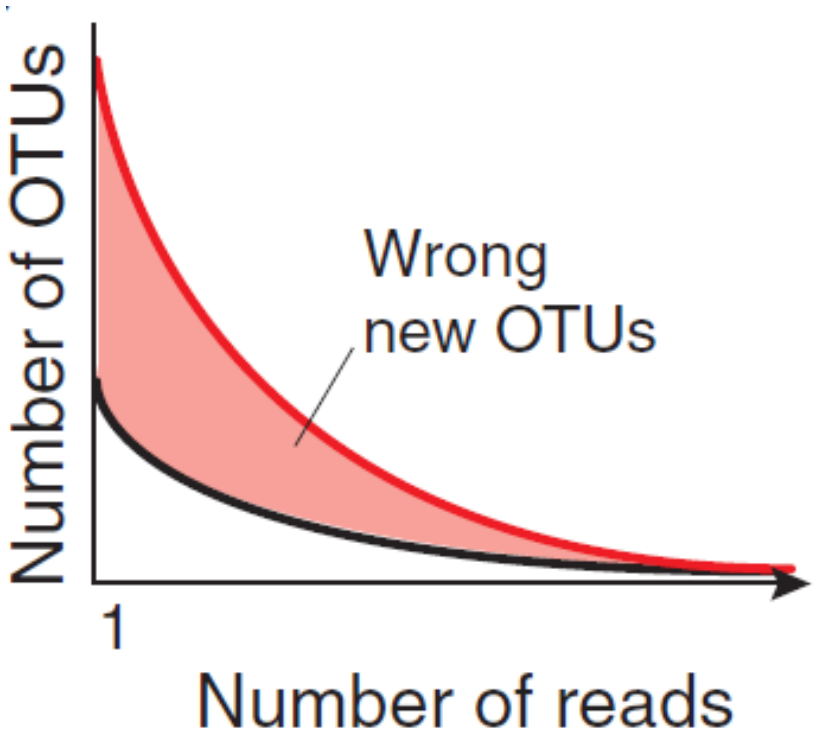
2010, Nucleic Acids Res. Claesson et al.

Filtering process



Overestimation of microbial diversity





2010, Nature Method, Reeder and Knight

Filtering effect on microbial diversity

Sample	# of reads	Subsample	Distances				
			0	**gap**	**0.03**	**0.05**	**0.1**
GM_A	10,729	1,000	**586**	**150**	196	123	51
GM_A_Flow	8,887	1,000	**436**		168	109	46
GM_D	4,650	1,000	**567**	**109**	227	161	85
GM_D_Flow	3,592	1,000	**458**		195	138	72
JM	9,929	1,000	**676**	**50**	260	195	135
JM_Flow	7,572	1,000	**626**		256	191	133
PM1	8,455	1,000	**635**	**44**	265	211	155
PM1_Flow	6,571	1,000	**591**		264	214	156
PM2	3,763	1,000	**731**	**1**	432	369	256
PM2_Flow	2,813	1,000	**732**		432	371	259
PM3	1,671	1,000	**850**	**1**	576	480	321
PM3_Flow	1,293	1,000	**849**		582	489	322

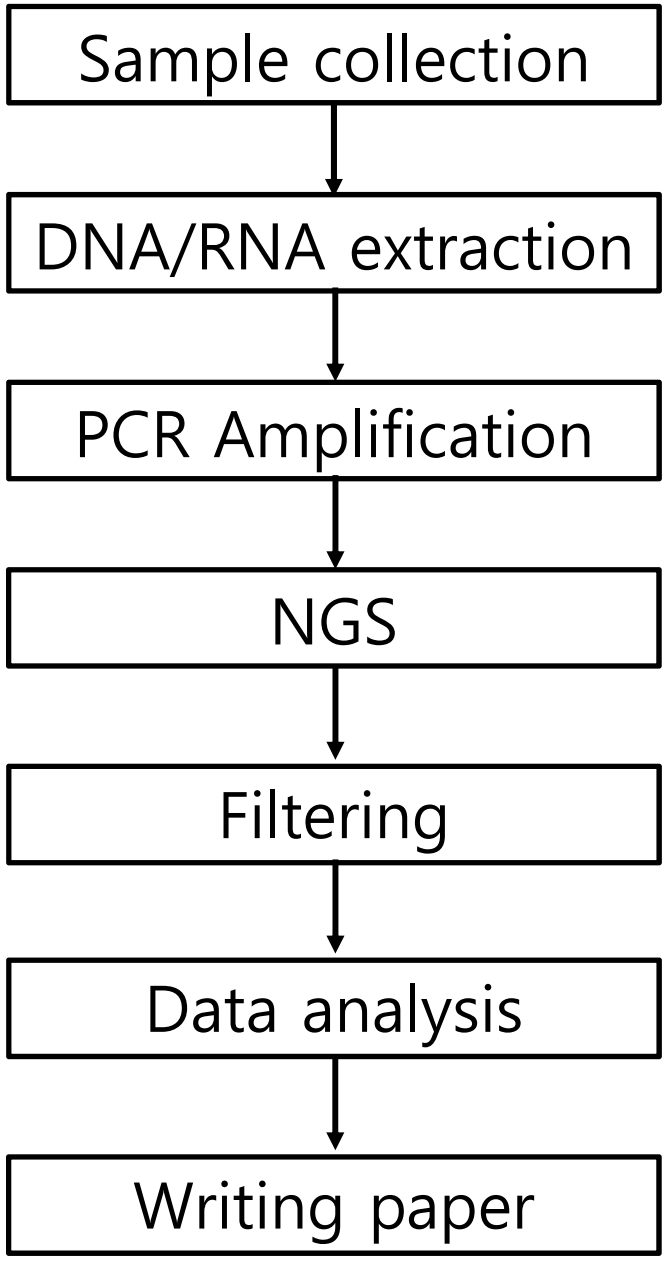
Subtle differences

Chimera slayer

2011, Genome Res., Haas et al.;2011, Bioinformatics., Edgar et al.

Step 2

Microbial structure & diversity



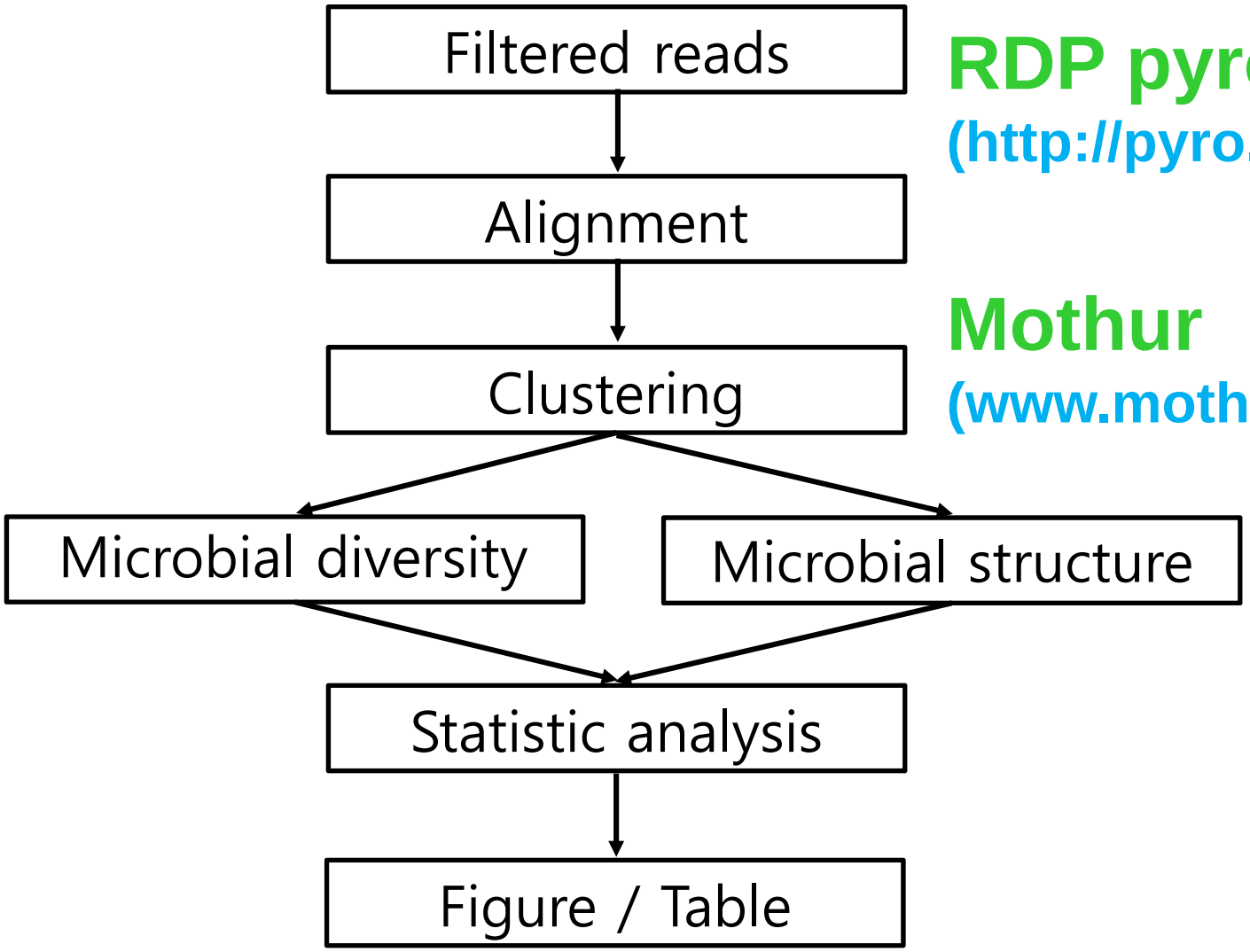
Microbial structure & Diversity

Pipeline

Statistics

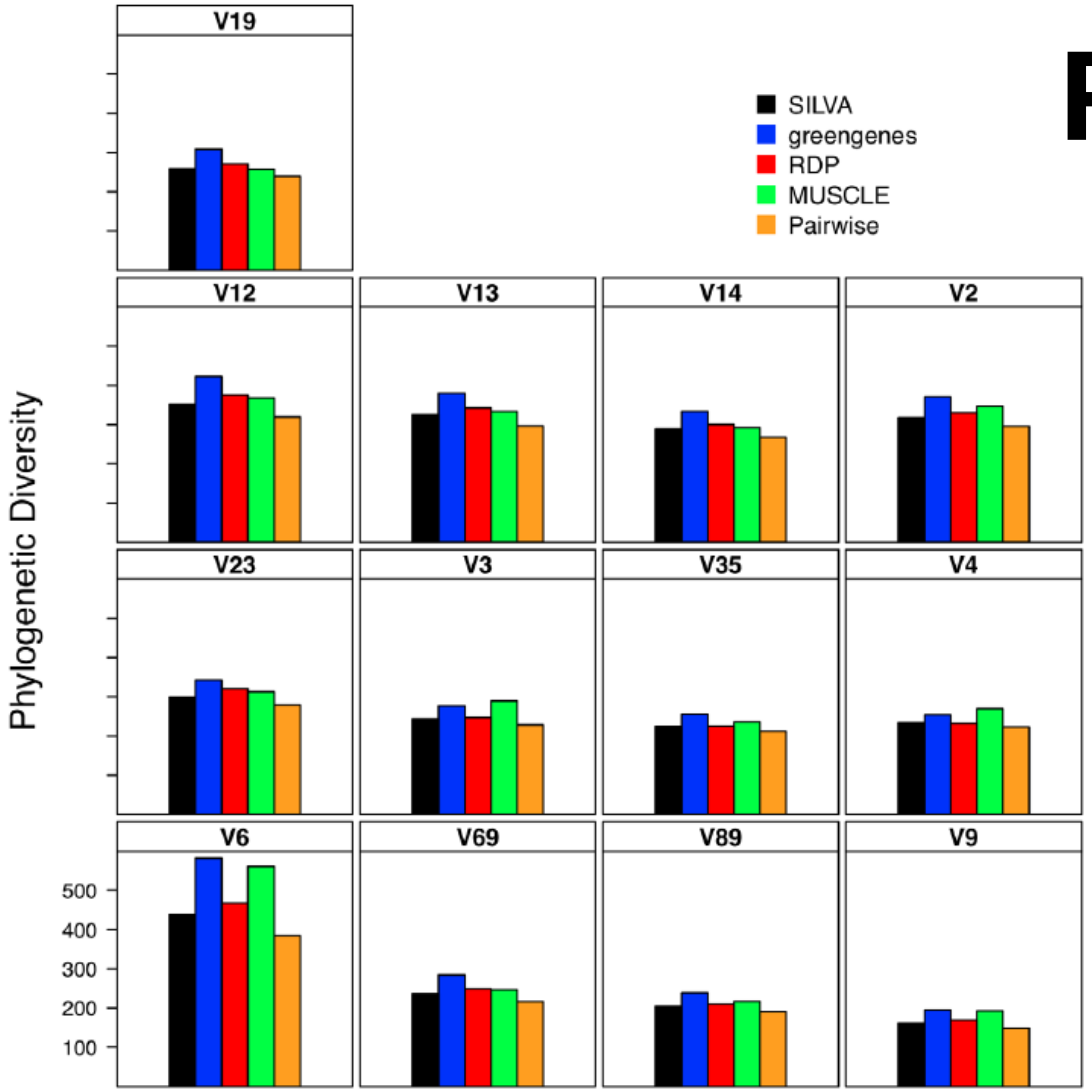
Display

Data analysis process



RDP pyro pipeline
(http://pyro.cme.msu.edu)

Mothur
(www.mothur.org)



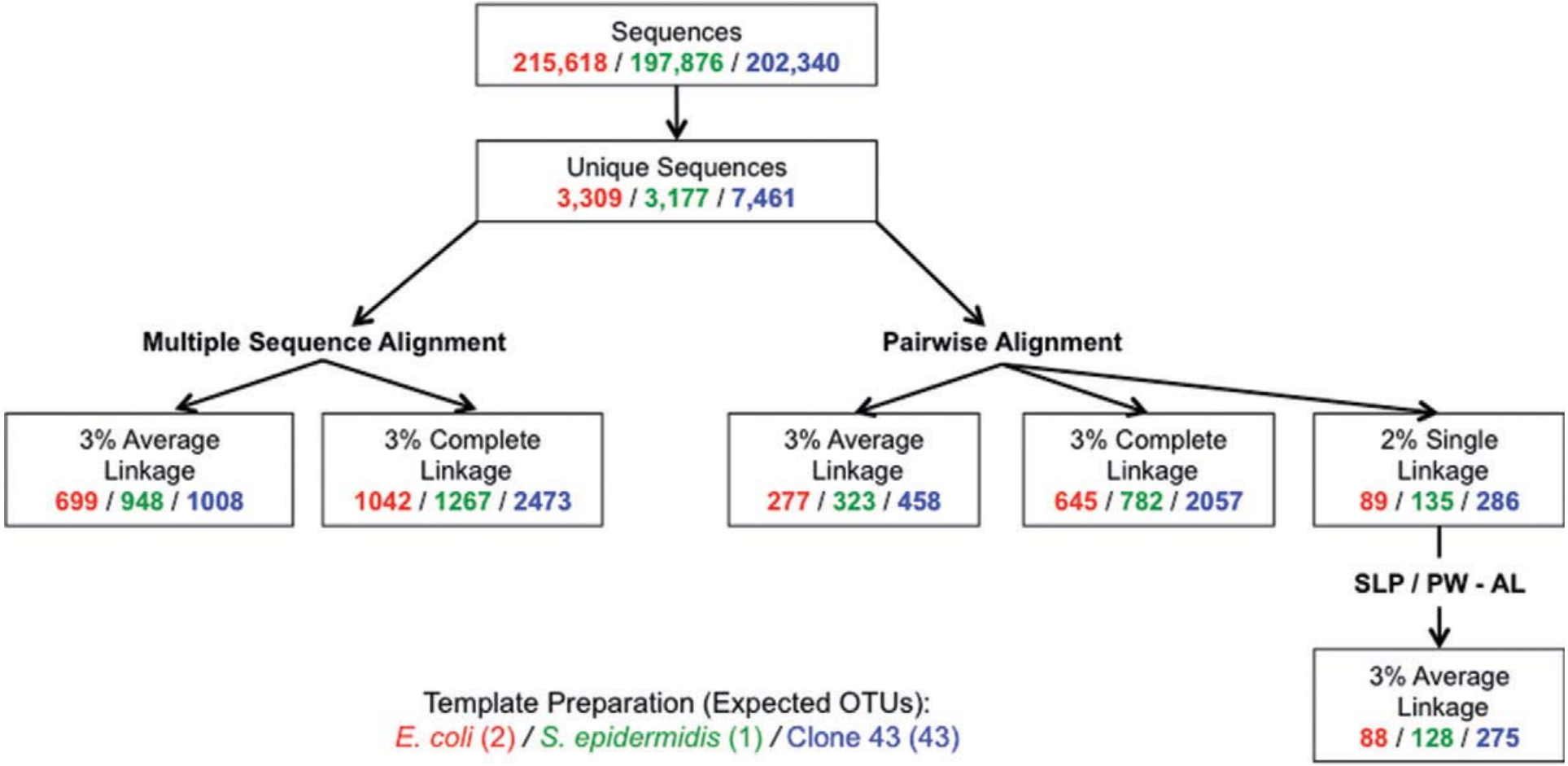
Recommand

SILVA

RDP

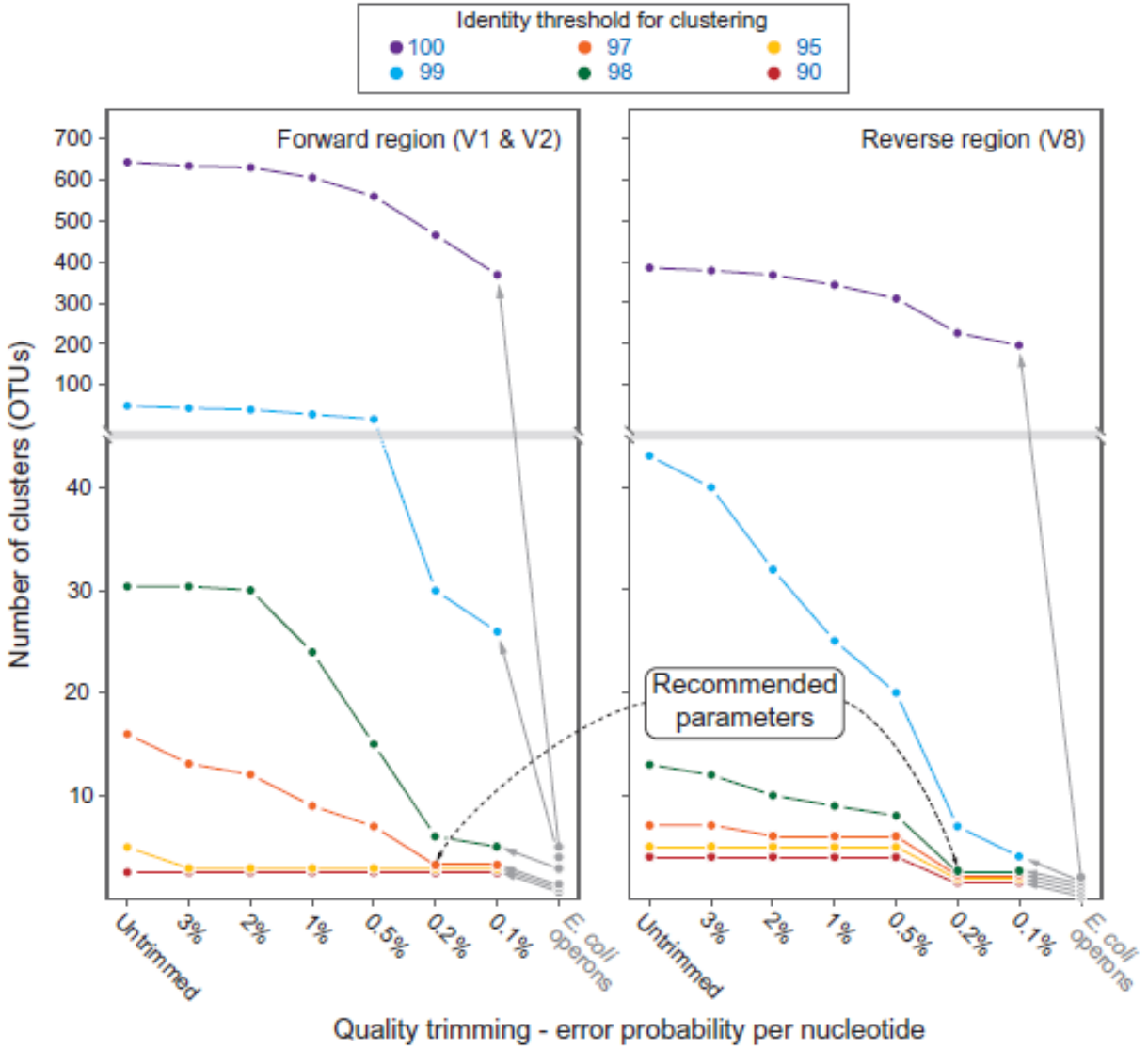
2010, PLOS Com. Bio., Schloss

Accuracy vs Time



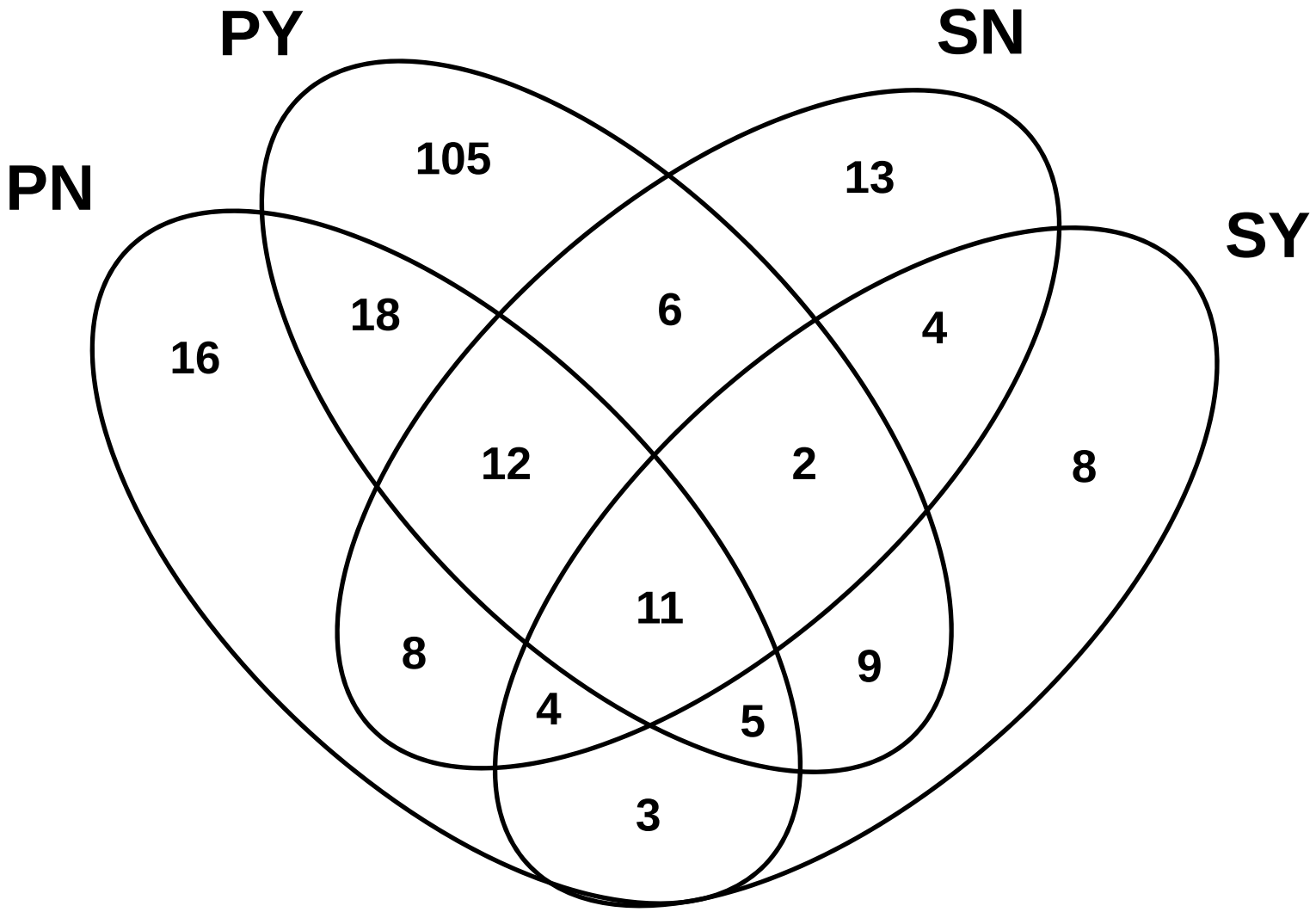
2010, Environ. Microbiol., Huse et al.

3% Complete linkage clustering



2010, Environ. Microbiol., Kunin et al.

Venn Diagram



2012, J. Endo. Res., Lim et al. (submitted)

Classification

Table 3 Bacterial identification of the commonly existing anode microbes detected in this study (the identified strain is the best match with each representative population from each OTU cluster [dissimilarity cut-off<5%])

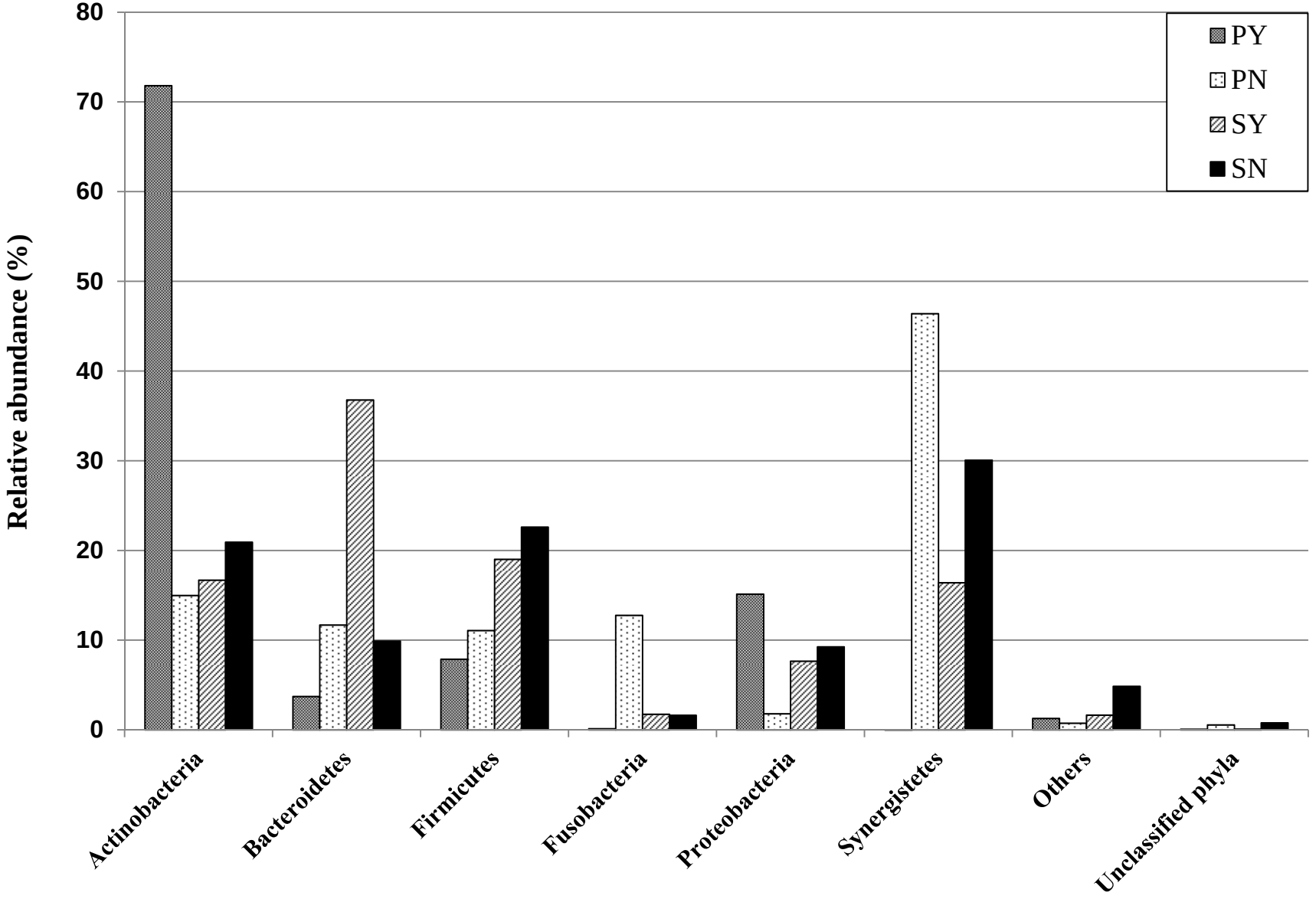
OUT ID	Accession number	RDP-II		Greengenes		NCBI BLAST	
CEAB3	GU594281	(D) *Geobacter psychrophilus* P39	81%	(D)*Geobacter psychrophilus* P35	95%	(D) *Geobacter psychrophilus* P35	95%
CEAB6	GU594282	(B) *Thauera* sp. 27	100%	(B) *Thauera* sp. 27	100%	(B) *Thauera* sp. 27	100%
CEAB8	GU594283	(B) *Variovorax limosa* EMB320	96%	(B) *Variovorax limosa* EMB320	96%	(B) *Polynucleobacter* sp. Fw70s-77	96%
CEAB9	GU594284	(G) *Thioflavicoccus mobilis* 83	92%	(G) *Thiococcus* sp. AT2204	95%	(G) *Thiococcus pfennigii* 4252	95%
CEAB11	GU594285	Unclassified bacteria	N.A.	(A) *Methylocystis* sp. KS7	81%	(A) *Oceanibaculum indicum* P24	81%
CEAB13	GU594286	(A) *Novosphingobium lentum* W-51	97%	(A) *Sphingomonas* sp. MT1	99%	(A) *Sphingomonas* sp. MT1	99%
CEAB14	GU594287	(A) *Afipia* sp. SP17	100%	(A) *Afipia* sp. SP17	100%	(A) *Afipia* sp. SP17	100%
CEAB15	GU594288	(F) *Fusibacter* sp. SA1	95%	(F) *Fusibacter* sp. SA1	95%	(F) *Fusibacter* sp. SA1	95%
CEAB17	GU594289	(G) *Pseudomonas mendocina PC12*	99%	(G) *Pseudomonas mendocina* PC12	99%	(G) *Pseudomonas mendocina* PC12	99%
CEAB18	GU594290	Unclassified bacteria	N.A.	(C) *Dehalococcoides* sp. BHI80-15	75%	(D) *Sorangium cellulosum* M5	85%
CEAB19	GU594291	(G) *Acinetobacter* sp. DZ0503SBS4	98%	(G) *Acinetobacter* sp. DZ0503SBS4	98%	(G) *Acinetobacter* sp. DZ0503SBS4	98%
CEAB21	GU594292	(F) *Clostridium litusburense*	93%	(F) *Clostridium* sp. C01-2409	97%	(F) *Clostridium litusburense*	96%

The letter in parenthesis indicates the identified phylum for each OUT group. The percent number indicates sequence similarity between the best-matched database sequence and the representative sequence of each OUT group

Type strain identification : Ez-Taxon

2010, Appl. Microbial. Biotech., Lee et al.

Bar graph



2012, J. Endo. Res., Lim et al. (submitted)

Statistics analysis

Open
source
statistical
computing.

Pirate for
Statistics!

Contribute at cran.r-project.org

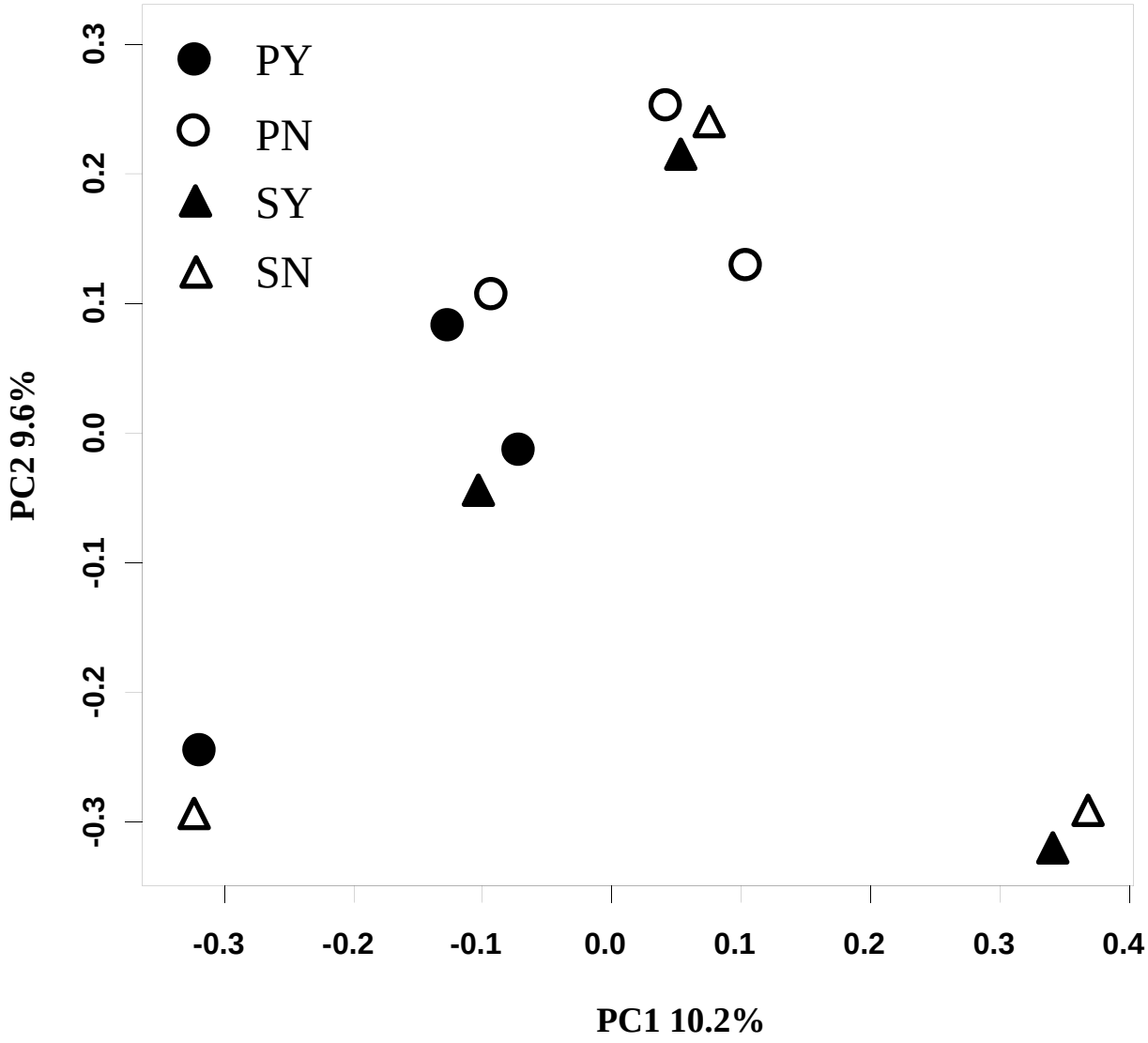
VEGAN

+

BiodersityR

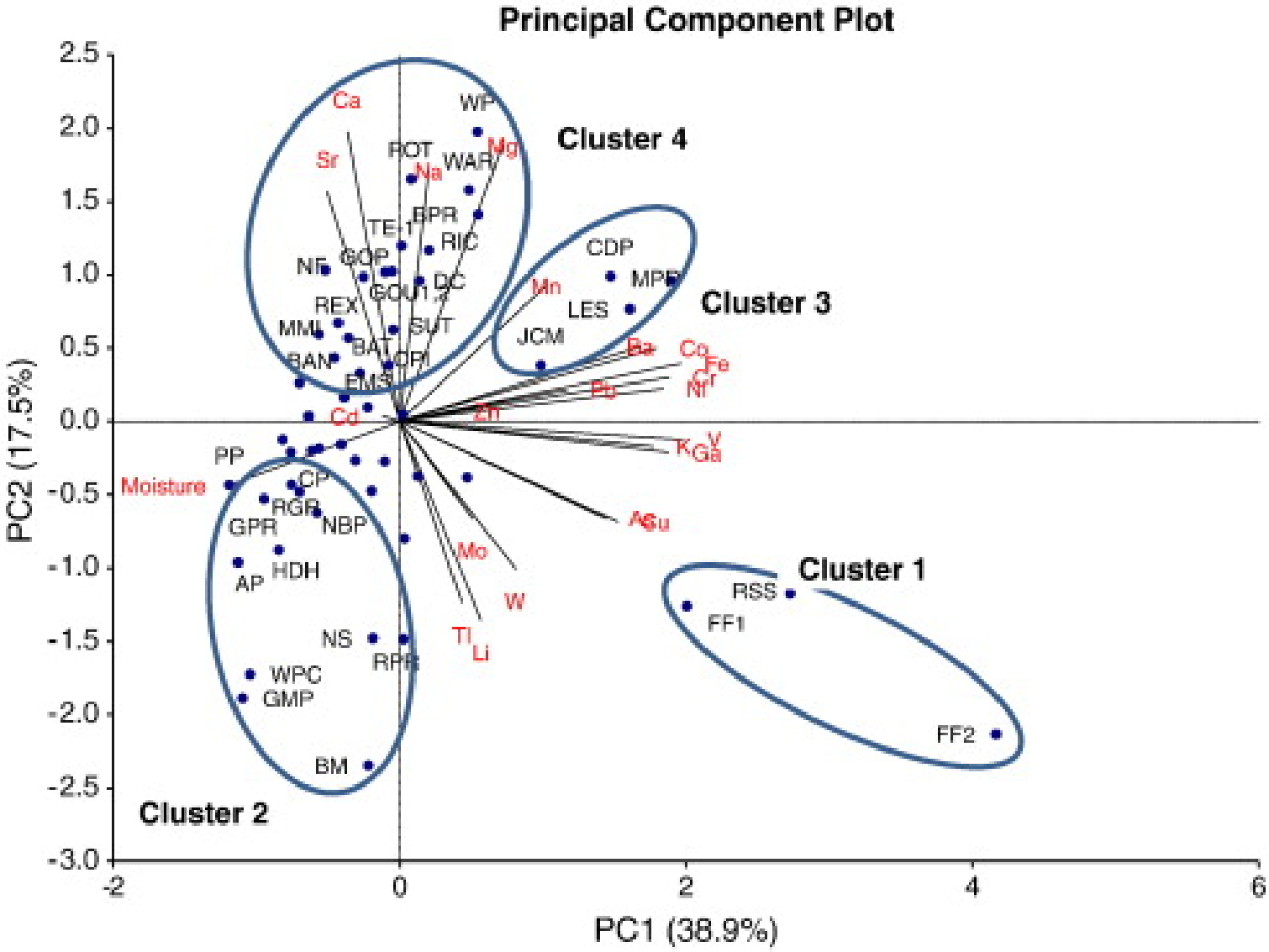
http://www.r-project.org

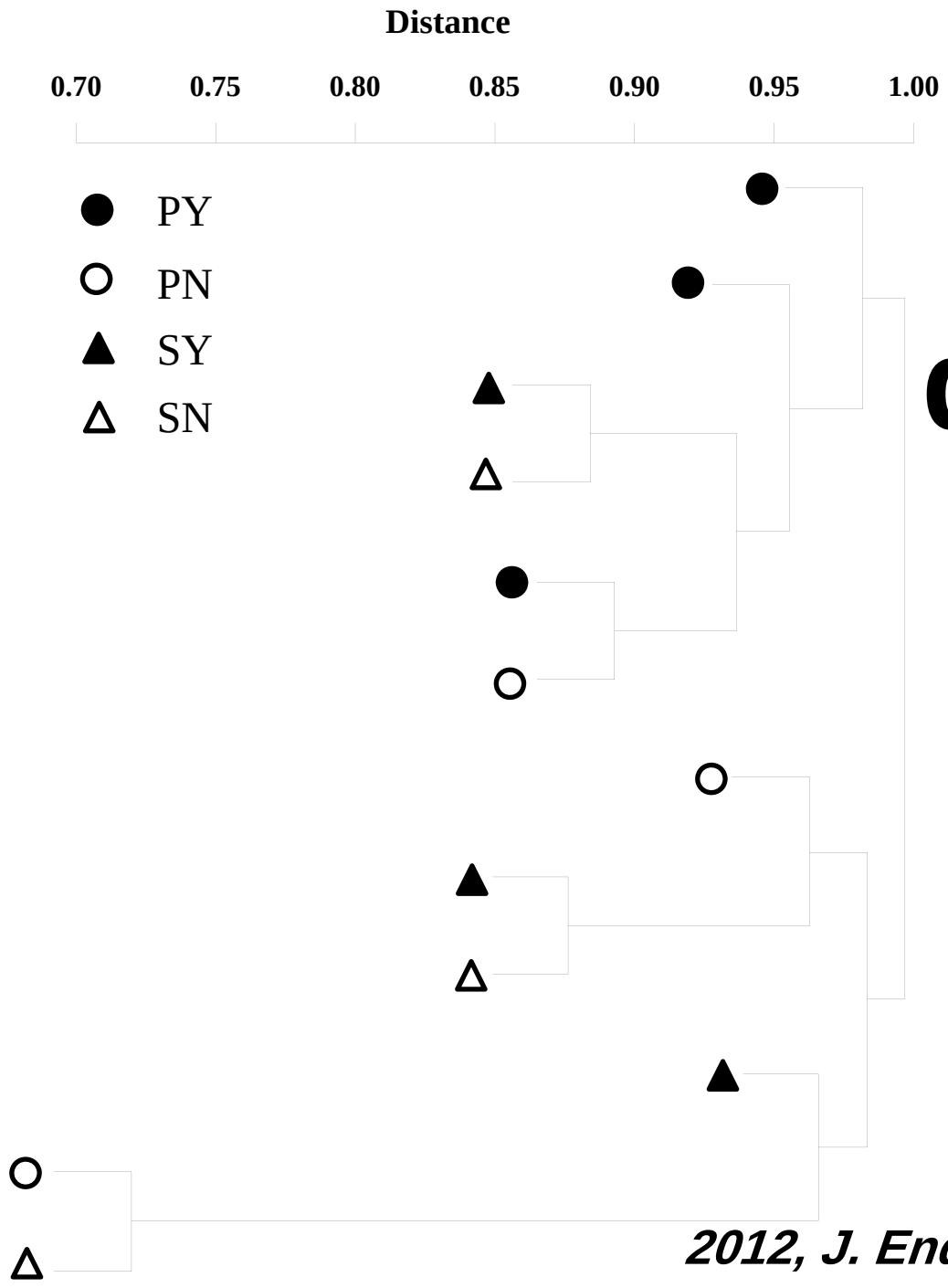
PCA analysis



2012, J. Endo. Res., Lim et al. (submitted)

PCA analysis with metadata

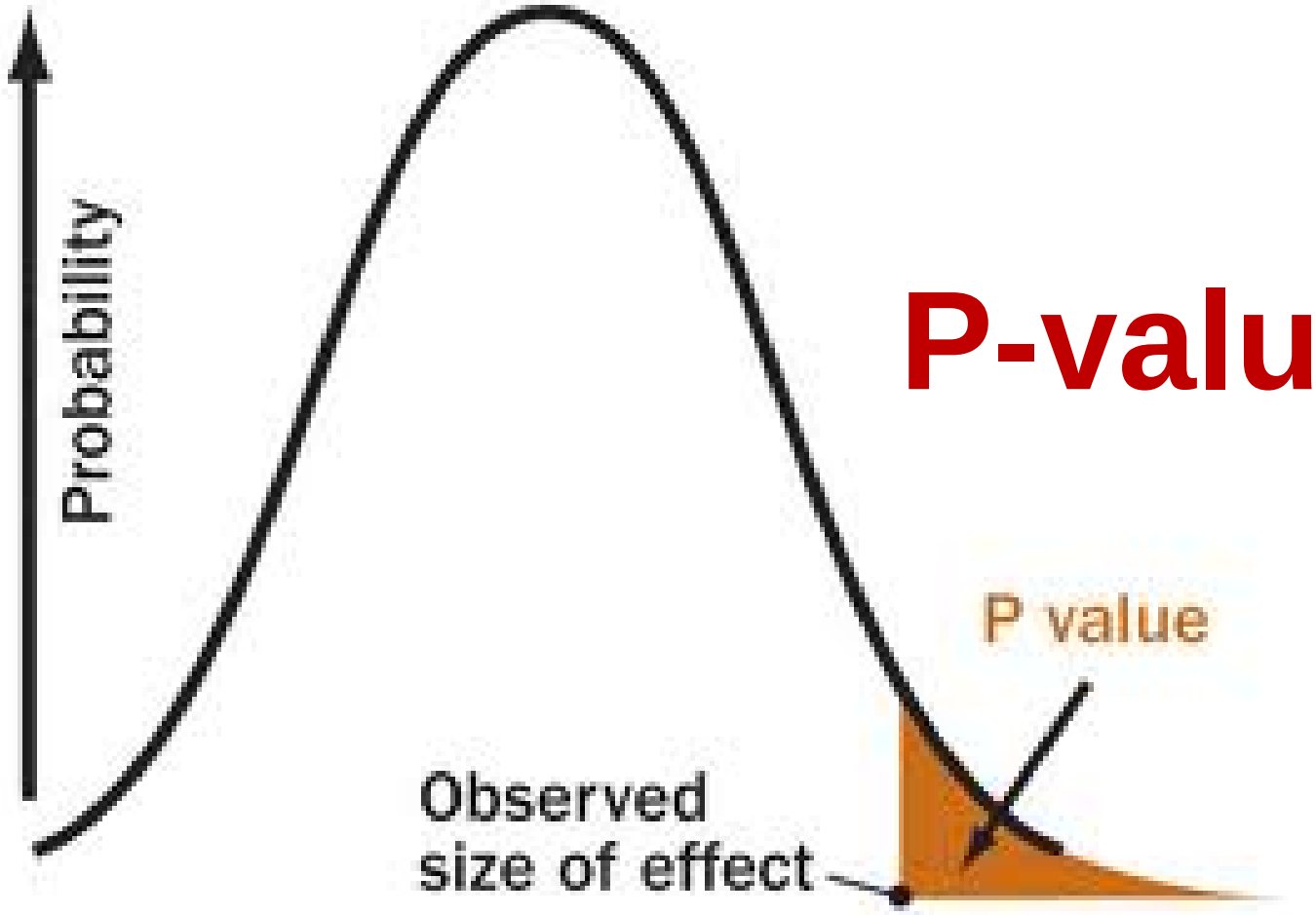




UMPGA dendrogram

2012, J. Endo. Res., Lim et al. (submitted)

Statistics analysis



P-value?

Basic significance

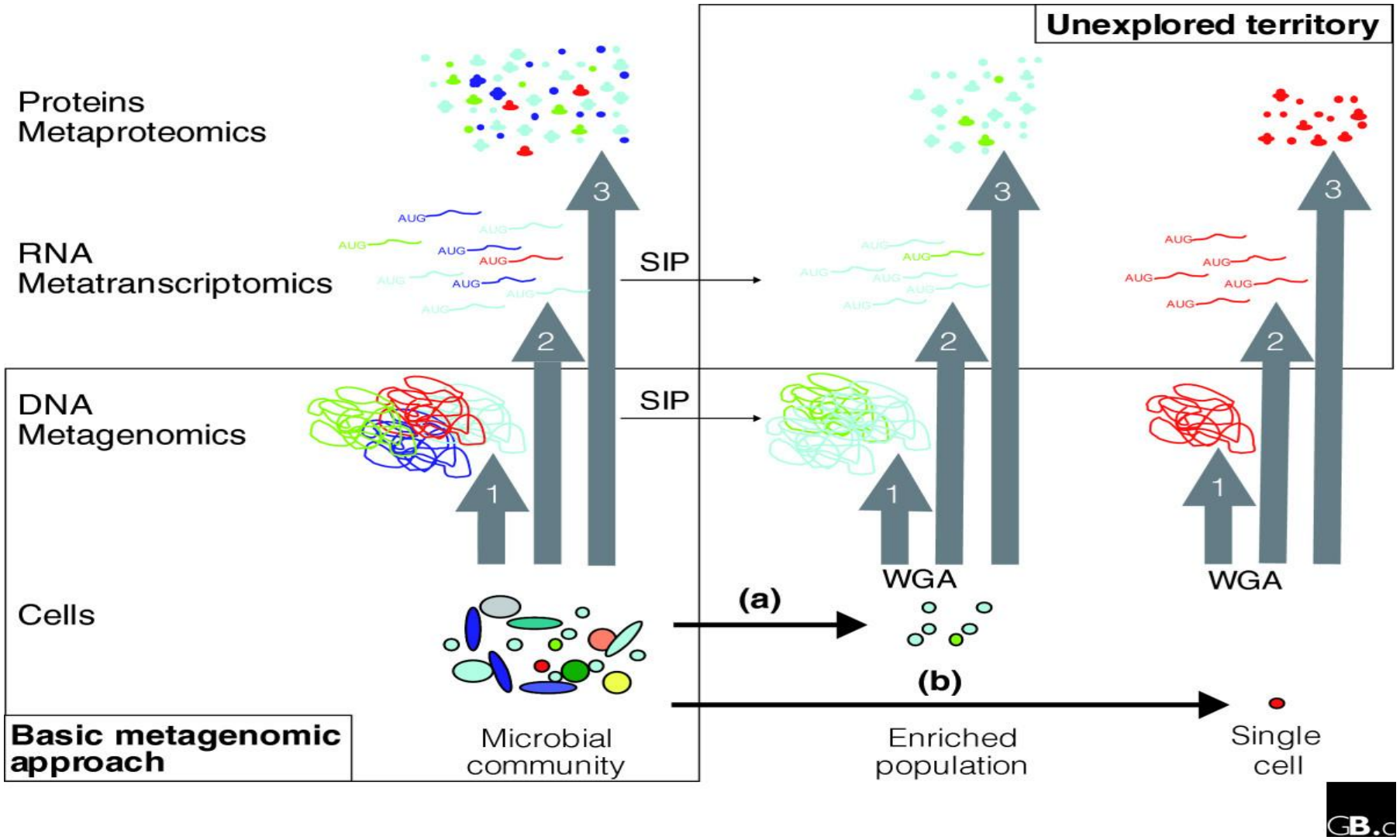
R & library compare (RDP)

Parsimony test (P-test)

Libshuff, ANOVA and HOMOVA

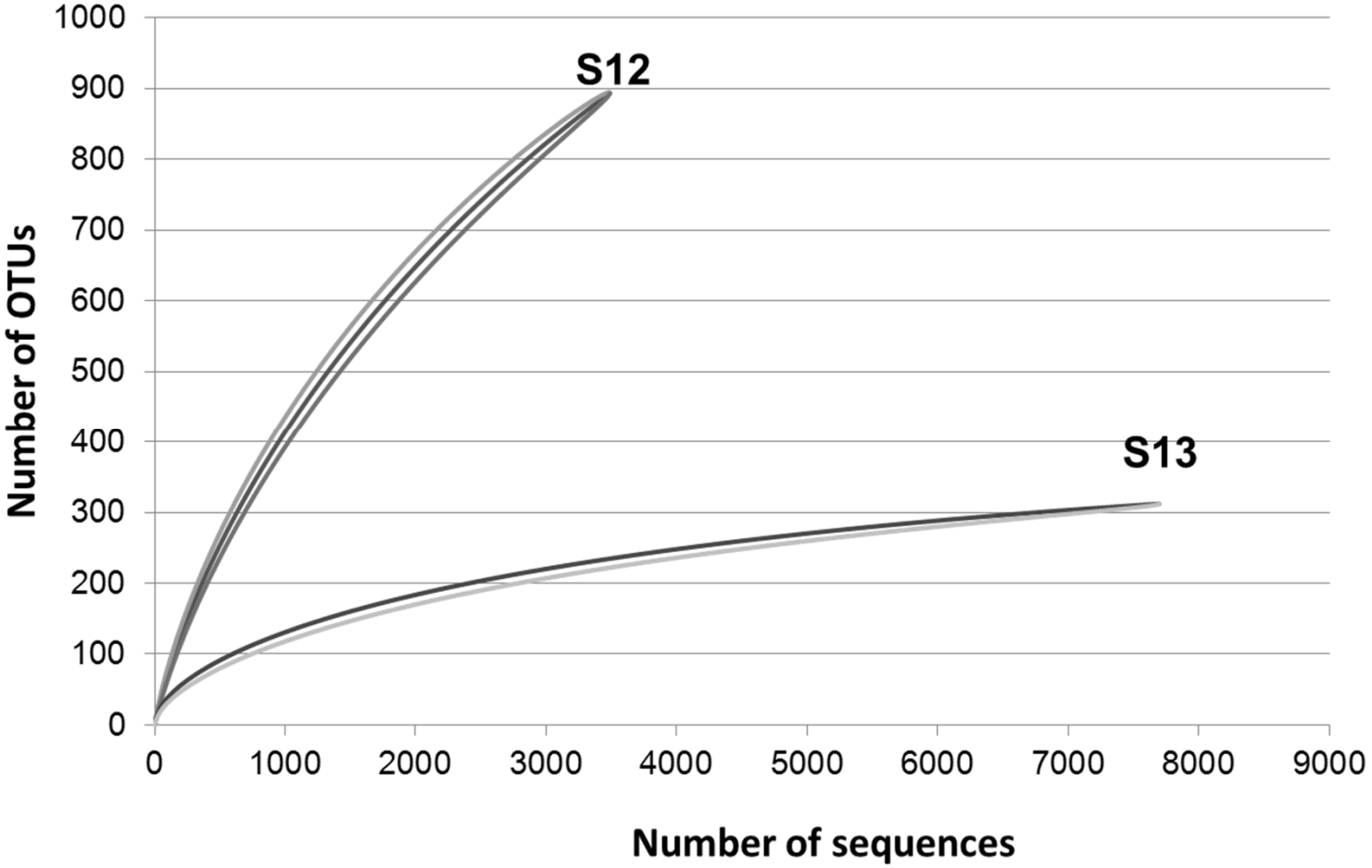

Step 3
Interaction with function

Function with microbial structure



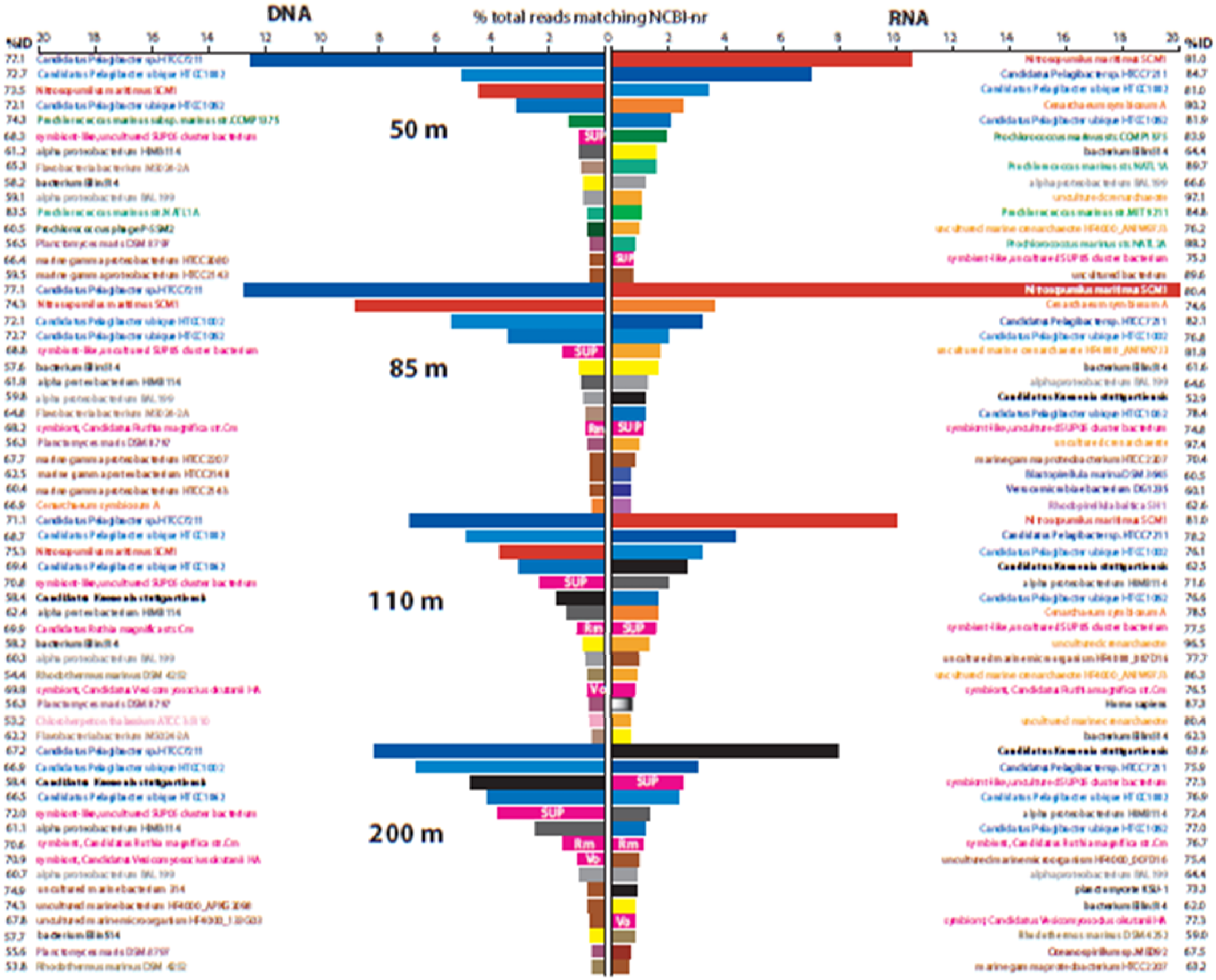
2007, Genome Biol., Warnecke and Hugenholtz

Stable isotope probing (SIP)



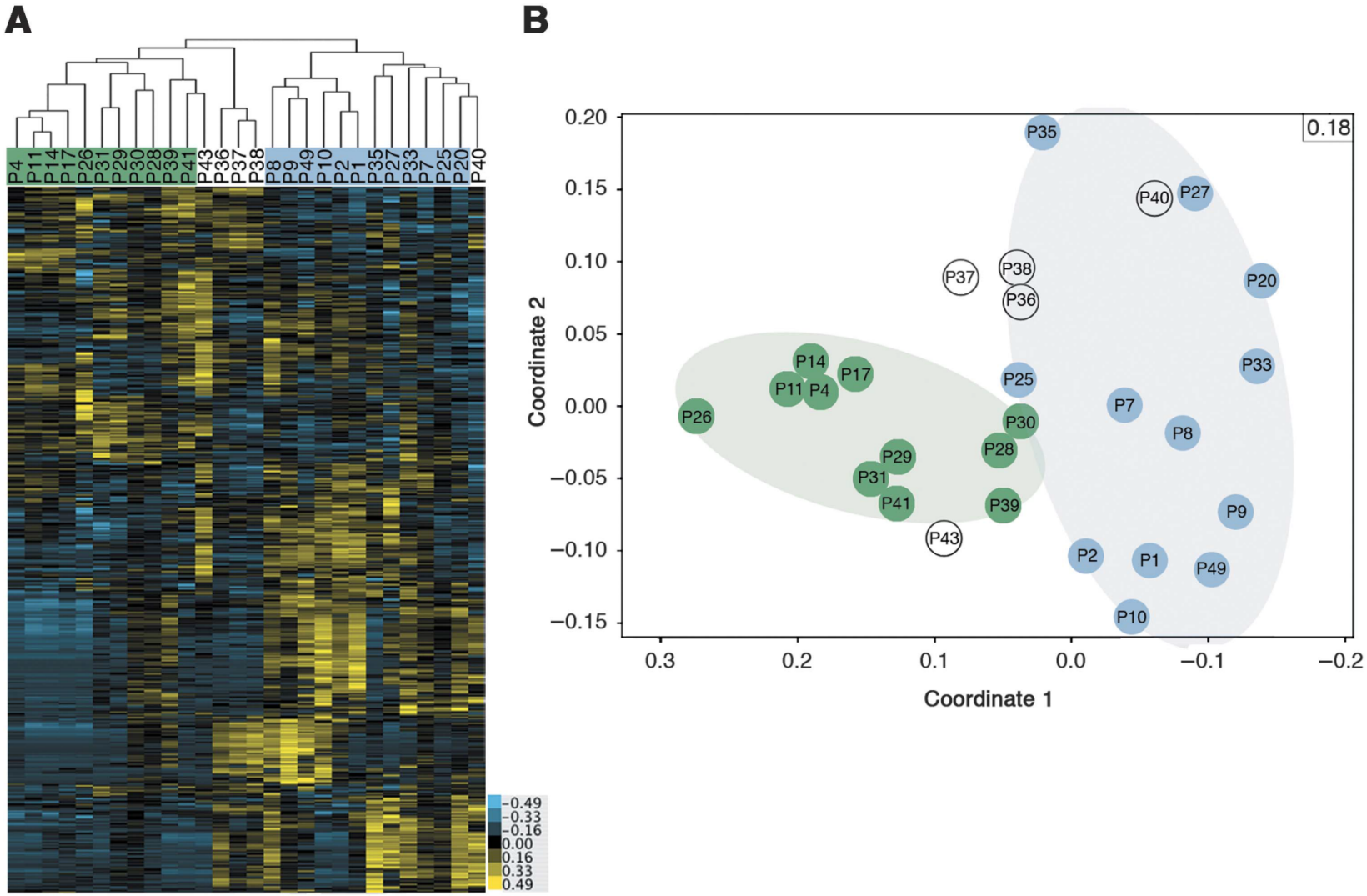
2011, Appl. Environ. Microbiol., Lee et al.

Metagenome & Metatranscriptome



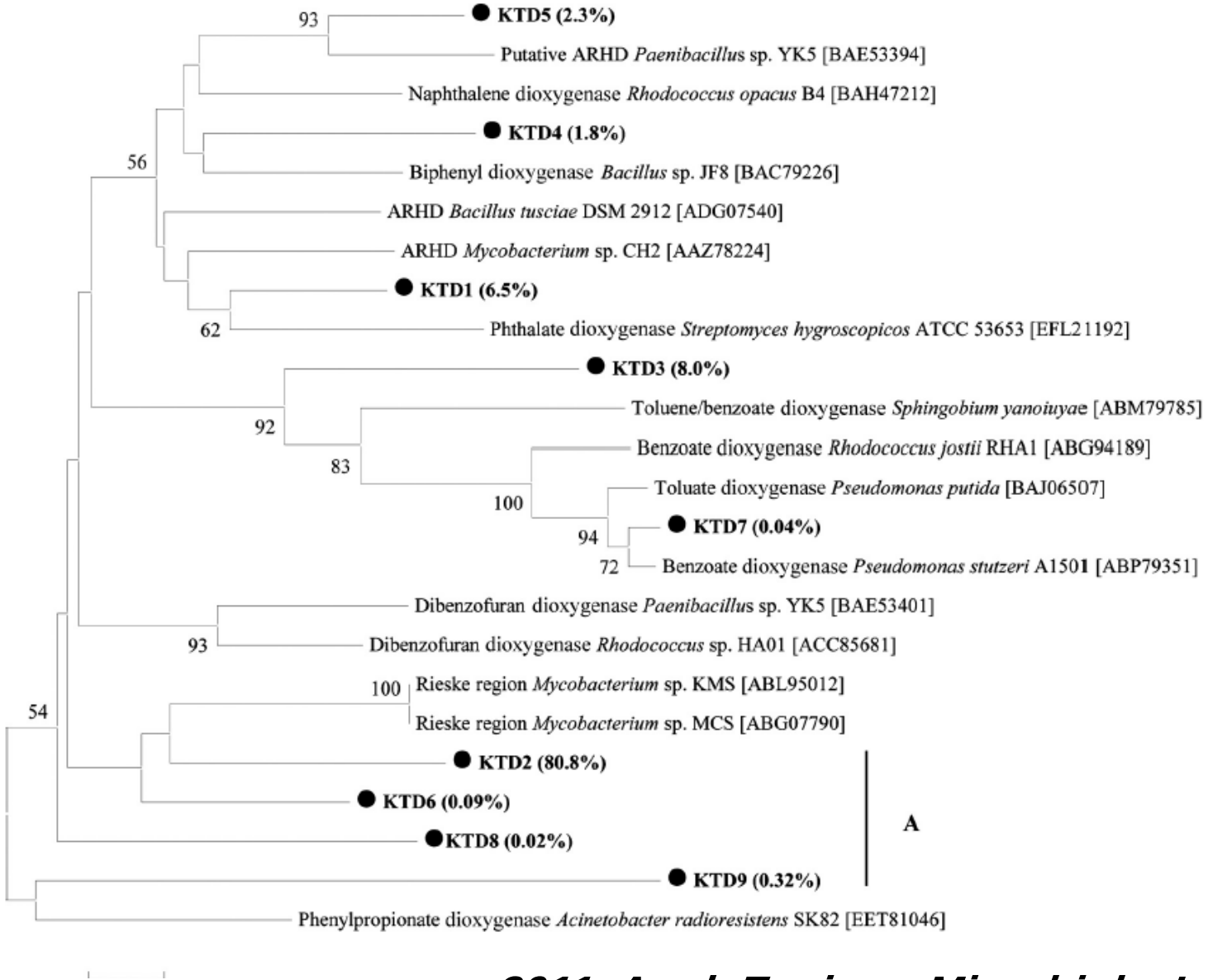
2011, Environ. Microbiol., Stewart et al.

Metaproteomics



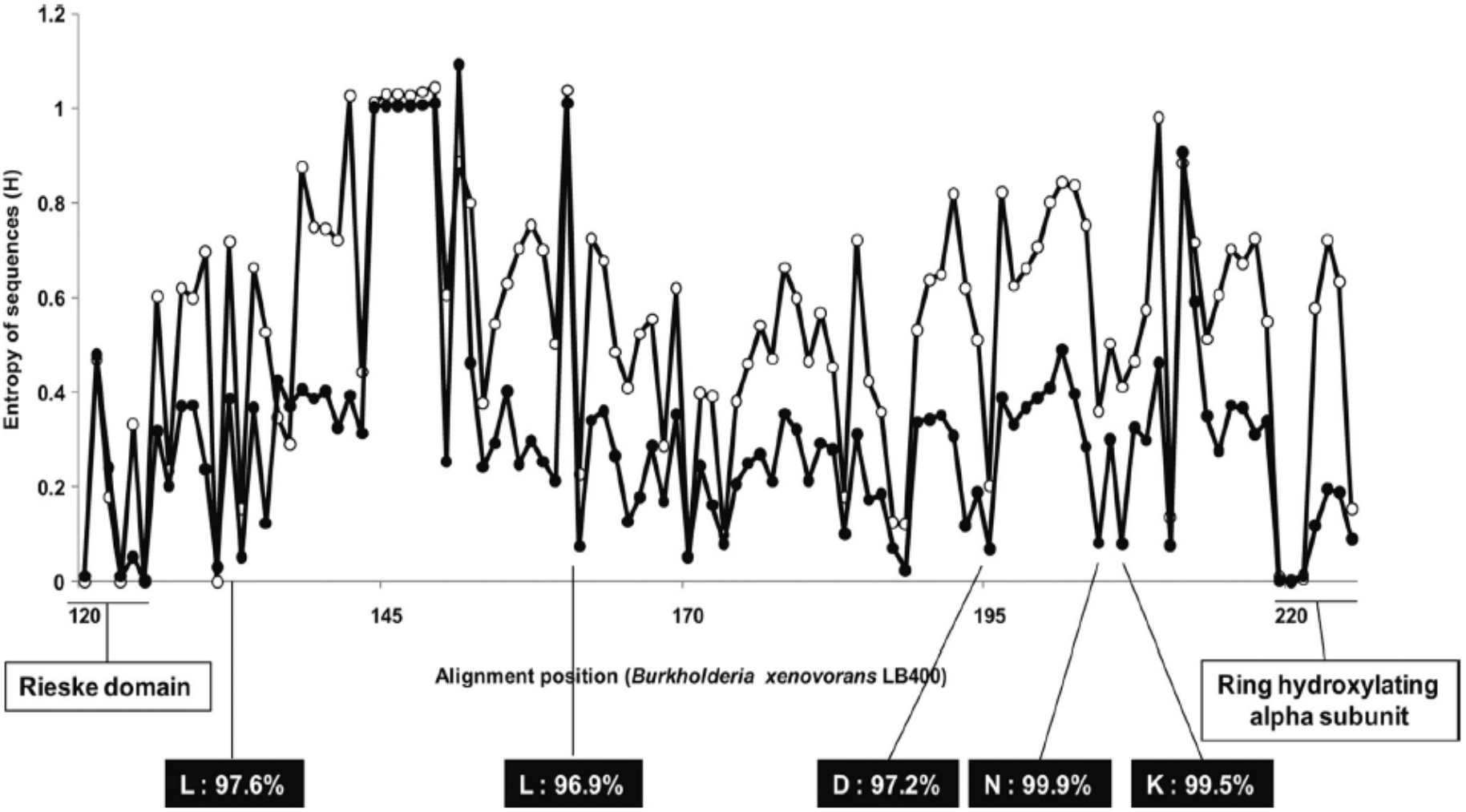
2010, Mol. Sys. Biol., Mueller et al.

Functional gene analysis 1



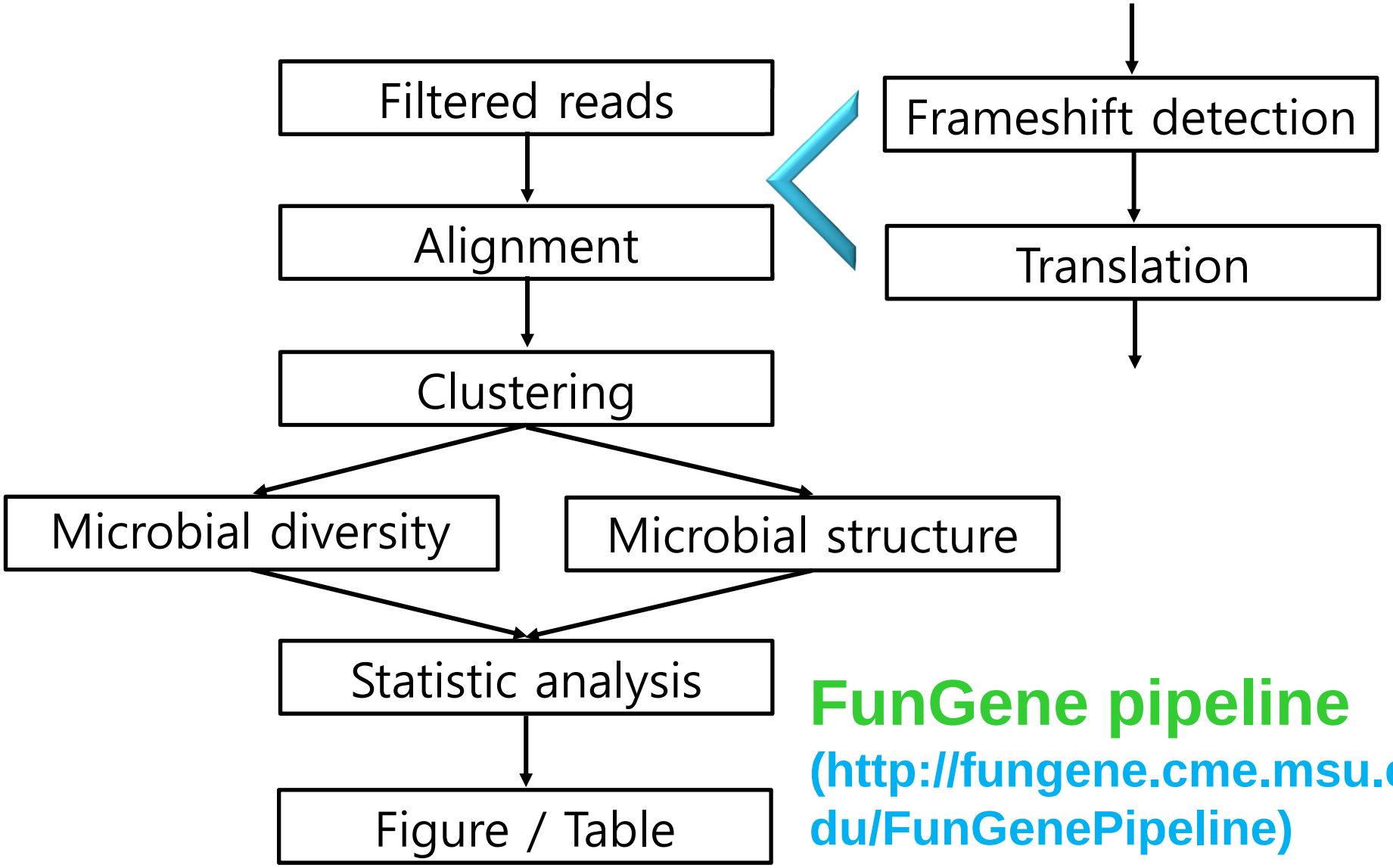
2011, Appl. Environ. Microbiol., Lee et al.

Functional gene analysis 1



2011, Appl. Environ. Microbiol., Lee et al.

Data analysis process



FunGene pipeline
(http://fungene.cme.msu.edu/FunGenePipeline)

Antibiotic resistances (6)

Biodegradation (18)

Biogeochemical cycles (26)

Phylogenetic markers (11)

Plant pathogencity (2)

11 genes

nifH, nirK …..

1. **Can we determine the true microbial diversity (phylogenetically and genetically)?**

2. **Is there a core microbiome?**

3. **Does structure impact ecological function? If so, how?**

Steps of

Microbial analysis

1. Sample preparation & filtering
2. Microbial structure & diversity
3. Interaction with function

Acknowledgement

WCU Center for Green Metagenomics @ Yonsei University
- Prof. Joonhong Park
- Jaejin Lee
- Jangho Lee

Center for Microbial Ecology @ Michigan State University
- Prof. James M. Tiedje

Ribosomal Database Project @ Michigan State University
- Prof. James Cole
- Qiong Wang

Marine Biological Laboratory @ Woods Hole
- Woo Jun Sul


Thank you.
Contact : blurple@live.co.kr