

Big-data for microbiome researches: Data integration, analysis and applications

Kang Ning

Professor, PI of Microbial Bioinformatics Group
Director of the Department of Bioinformatics and Systems Biology
School of Life Science and Technology
Huazhong University of Science and Technology
Wuhan, China

Email: ningkang@hust.edu.cn



2018/08, Hayama, Japan

Microbiome Research

Key Challenge:

- Complexity of ecosystem
- Big-data



Microbiome Research

Key Challenge:

- Dynamics of ecosystem
- Big-data again!

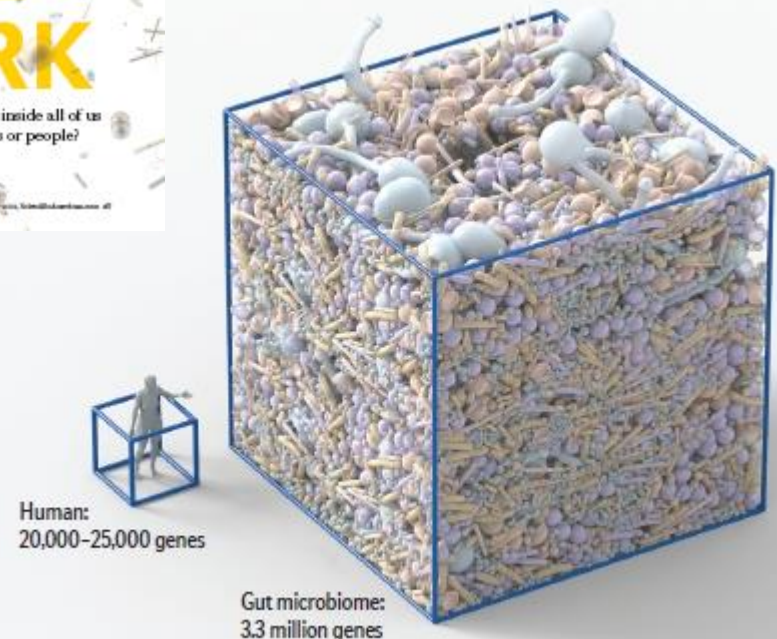
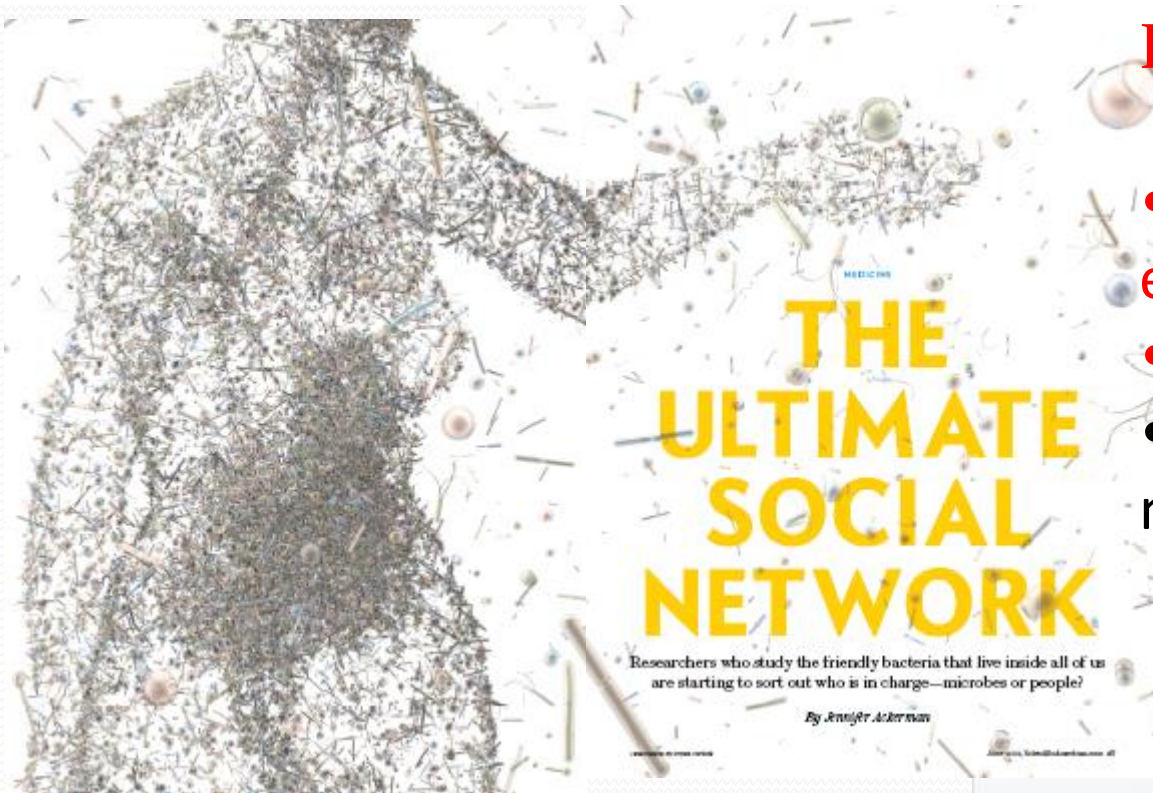


Microbiome Research

Key Challenge: “complexity and big-data”

Human microbiome:

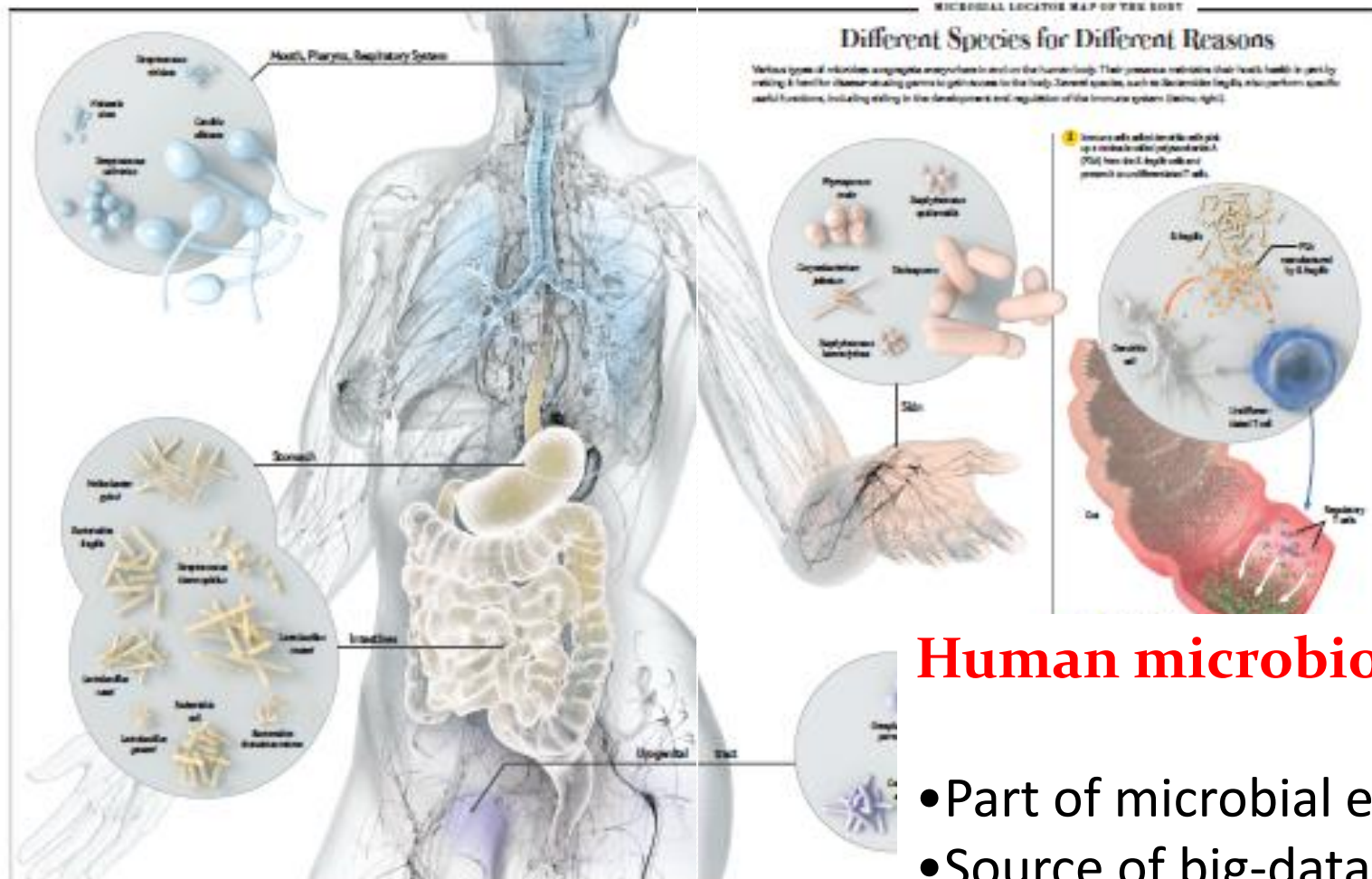
- Part of microbial ecosystem
- Source of big-data
- Important precision medicine applications



Ackerman, et al., *Scientific American*, 2012

Microbiome Research

Key Challenge: “un-culturable, complexity and big-data”



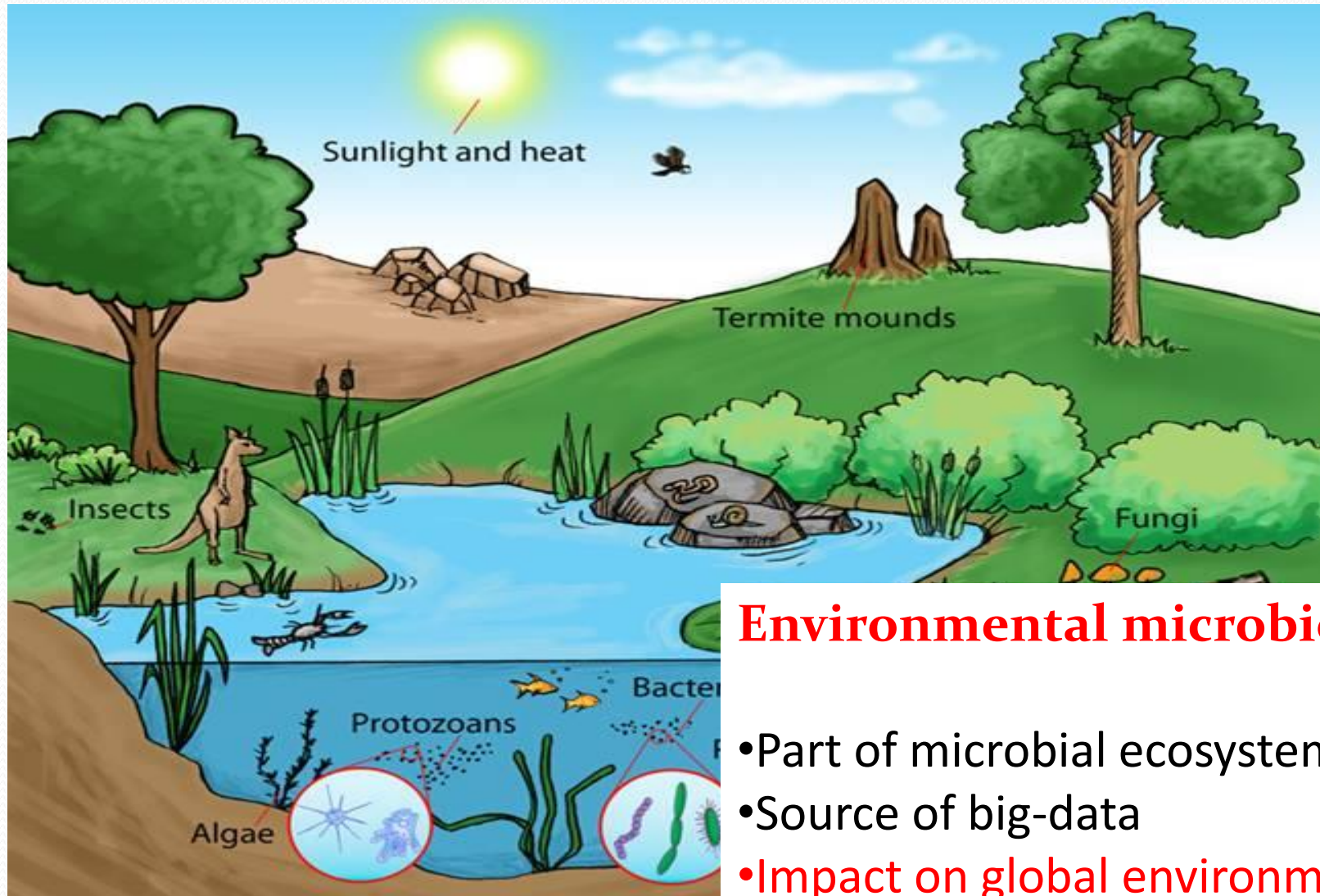
Human microbiome:

- Part of microbial ecosystem
- Source of big-data
- Important precision medicine applications

Ackerman, et al., *Scientific American*, 2012

Microbiome Research

Key Challenge: “un-culturable, complexity and big-data



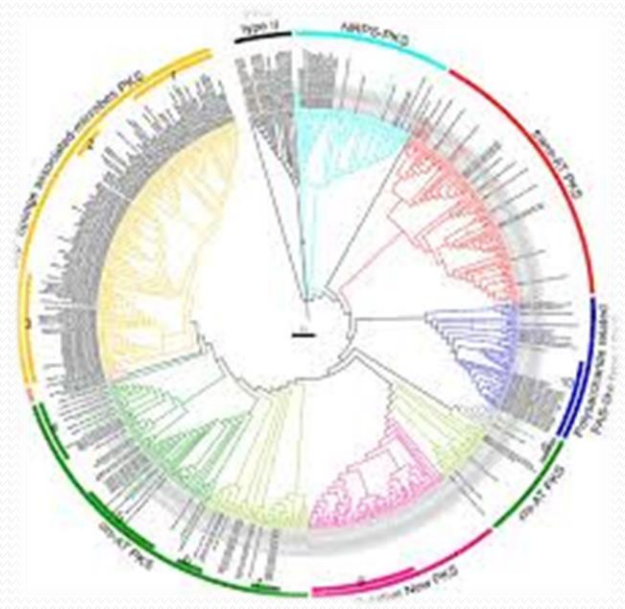
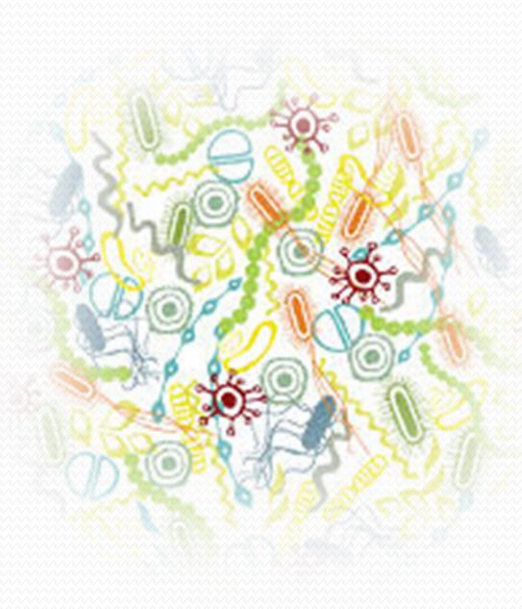
Environmental microbiome:

- Part of microbial ecosystem
- Source of big-data
- Impact on global environment

Microbiome Research

-- At present

From microbial communities to big-data...



EBI metagenomic portal (MGnify)

- <https://www.ebi.ac.uk/metagenomics>

The screenshot shows the EBI Metagenomics (MGnify) portal homepage. At the top, there is a search bar with a magnifying glass icon and the text "EBI Metagenomics". Below the search bar is a navigation menu with links: Home, Search, Sequence search, Submit data, Projects, Samples, Comparison tool, About, and Help. On the right side of the navigation menu, there are links for "Not logged in" and "Login".

The main banner area features the text "Submit, analyse, visualize and compare your data." and a prominent blue "SUBMIT DATA" button. Below the banner, there are several statistics and project categories:

- 101177 data sets** (represented by a red database icon)
- 79918 amplicons**, **2360 assemblies**, **1293 metabarcoding**, **16311 metagenomes**, and **1175 metatranscriptomes** (represented by a blue DNA double helix icon)
- 101177 runs** and **77796 samples** (Public), **1354 projects** (represented by a blue padlock icon)
- 6983 runs** and **6711 samples** (Private), **159 projects** (represented by a blue padlock icon)

The "Browse projects" section is divided into two columns:

- By selected biomes:** Five circular icons representing different biomes: Soil (422), Engineered (198), Host-associated human (169), Marine (145), and Human digestive system (130).
- Latest projects 1354:** Two project entries are shown:
 - Bos taurus breed:Angus Hereford Targeted Locus (Loci)**: Nasopharyngeal microbiota of feedlot cattle exposed to commingling and auction market stress. ... View more - 107 samples
 - Bos taurus Targeted Locus (Loci)**: Monitoring the nasopharyngeal microbiota in feedlot cattle from entry to exit ... View more - 1 sample

NODE SysEcoKit portal

[Home](#) [Project](#) [Sample](#)

[Login](#) | [Register](#)

NODE Microbiome



70

Total project number



45,328

Total sample number



11,393,850,212

Total sequence number



Soil



Human



Marine



Freshwater



Animal



Engineered

微生物组大数据的4色

(4 colors for big-data towards healthy microbiome)

人体健康大数据（红色）
Human microbiome

高性能的生物计算（蓝色）
High Performance
Computation

健康的生存环境（绿色）
Healthy environments



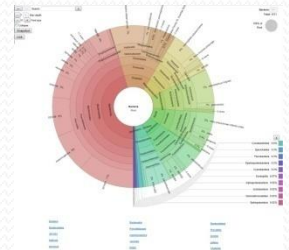
海量的健康大数据（褐色）
Big-data integration

From a sample to thousands of samples

I. Parallel-META

Su, et al, *BMC Sys. Bio.* 2012.
Su, et al, *PLoS ONE* 2014.
Zhou, et al, *Scientific Reports*, 2017

Parallel **META** Sample **analysis**

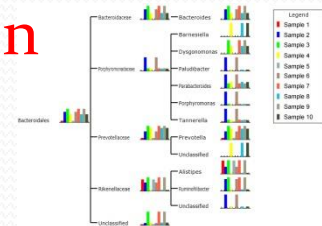


II. MetaSee

Song, Su, et al, *PLOS ONE* 2012

Meta-
See

Sample **visualization**

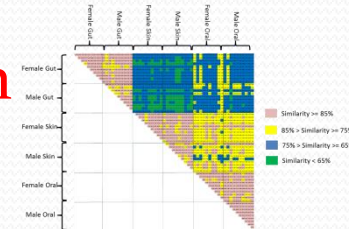


III. Meta-Storms

Su, et al, *Bioinformatics*, 2012;
Su, et al, *Bioinformatics*, 2014

Meta-
Storms

Sample **comparison**

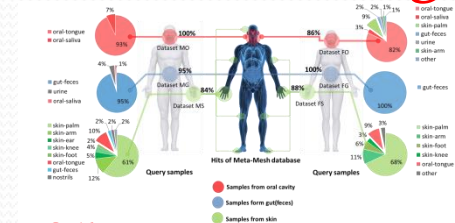


IV. Meta-Mesh

Su, et al, *Scientific Reports*, 2014c, d

Meta-
Mesh

Sample **search** and **data-mining**



V. MetaBoot

Wang, et al, *PeerJ*, 2014
Gabriel, et al, *Bioinformatics*, 2016

Biomarker discovery

From structure to function

高性能的生物计算（蓝色）
HPC

VI. Parallel-Prism

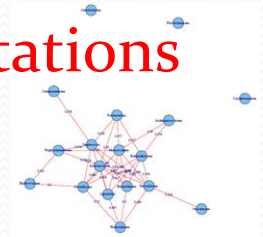
Su, et al, *Scientific Reports*, 2017
Yu, et al, *Bioinformatics*, in revision

Sample **indexing**

VII. Meta-Network

Yang, et al, *Bioinformatics*, in revision

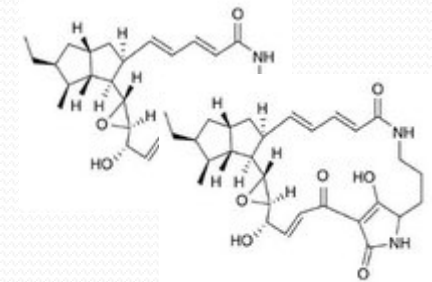
Species interaction **interpretations**



VIII. Meta-BGC

Zhang, et al, *NAR*, in revision
Chen, et al, in preparation

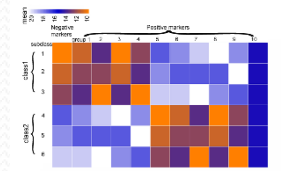
BGC **interpretations**



IX. Meta-Biomarker 2.0

Tan, et al, *Bioinformatics*, in revision

Biomarker for noisy data



X. Meta-Discover

GPU and cloud computational platform...

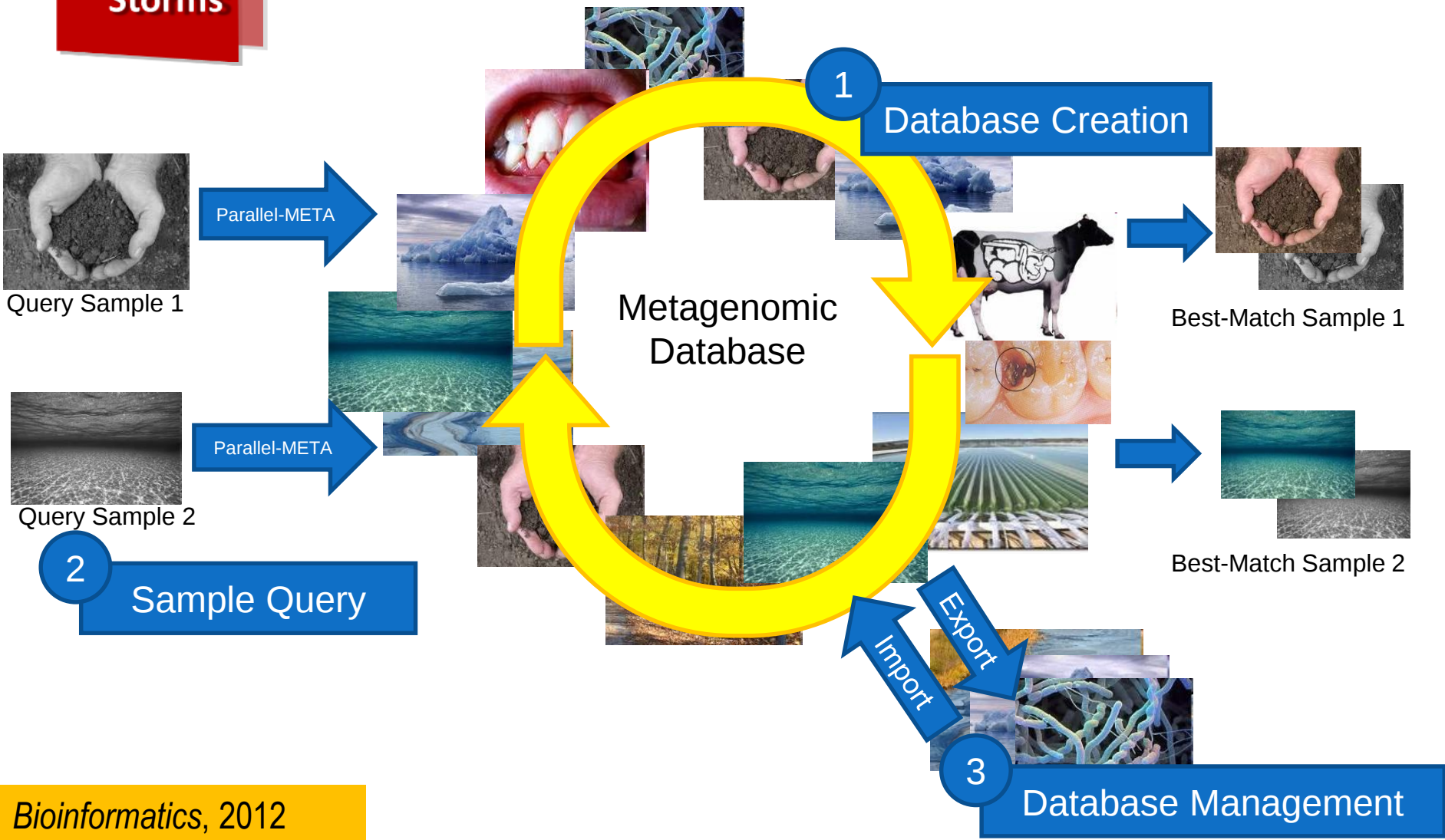
Integrated Metagenomic analysis

Meta-Storms

高性能的生物计算（蓝色）
HPC

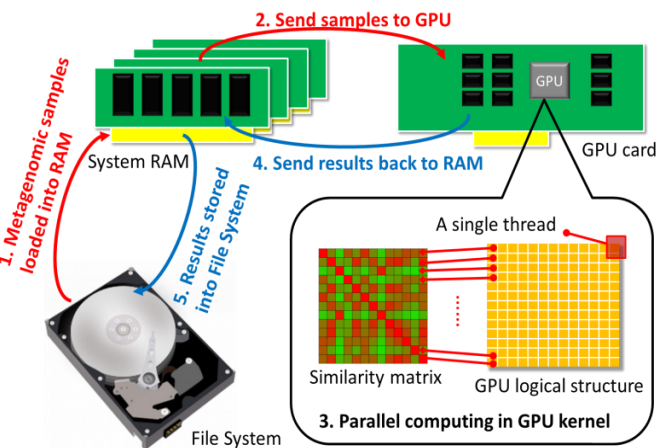


利用先进的数据库和索引技术处理群落比较和搜索

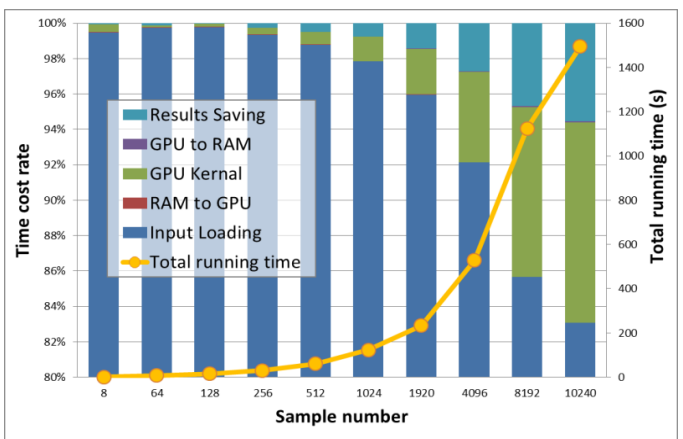


Meta-Storms

高性能的生物计算（蓝色）
HPC

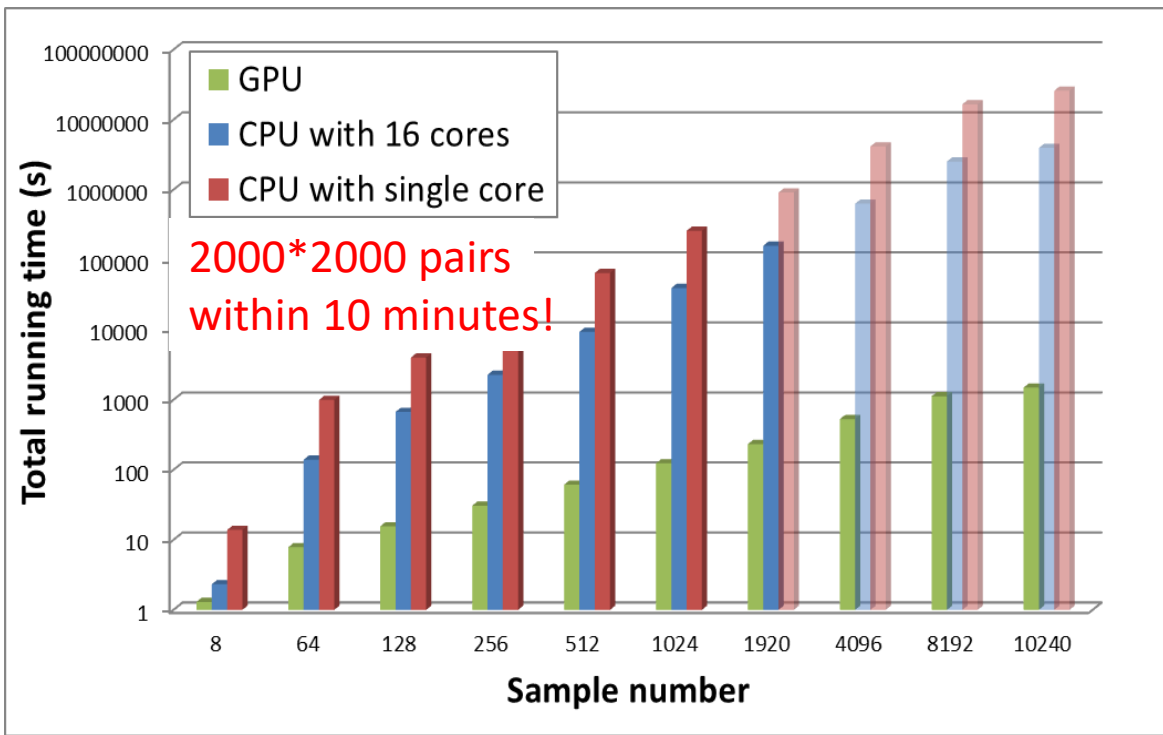


利用并行计算技术
加速大规模群落比较和搜索



Data source: MG-RAST database

Bioinformatics, 2014
Scientific Reports, 2017



1. Microbiome plasticity (菌群可塑性)

应用：
水土不服的机制

应用：
中国肠型的分类

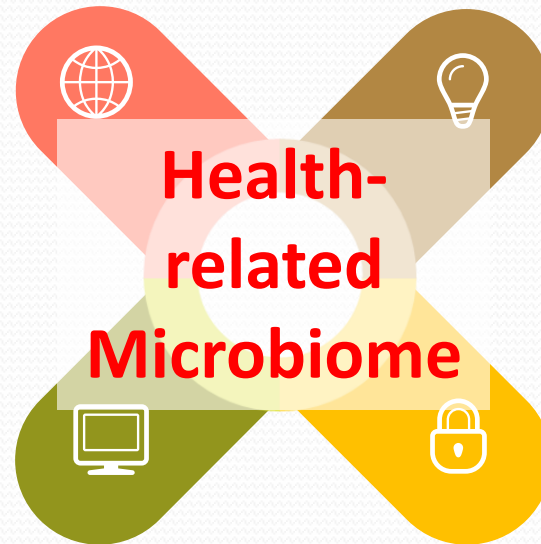
3. Microbiome dysbiosis (菌群失调)

应用：
ICU重症菌群失调

应用：
益生菌干预治疗

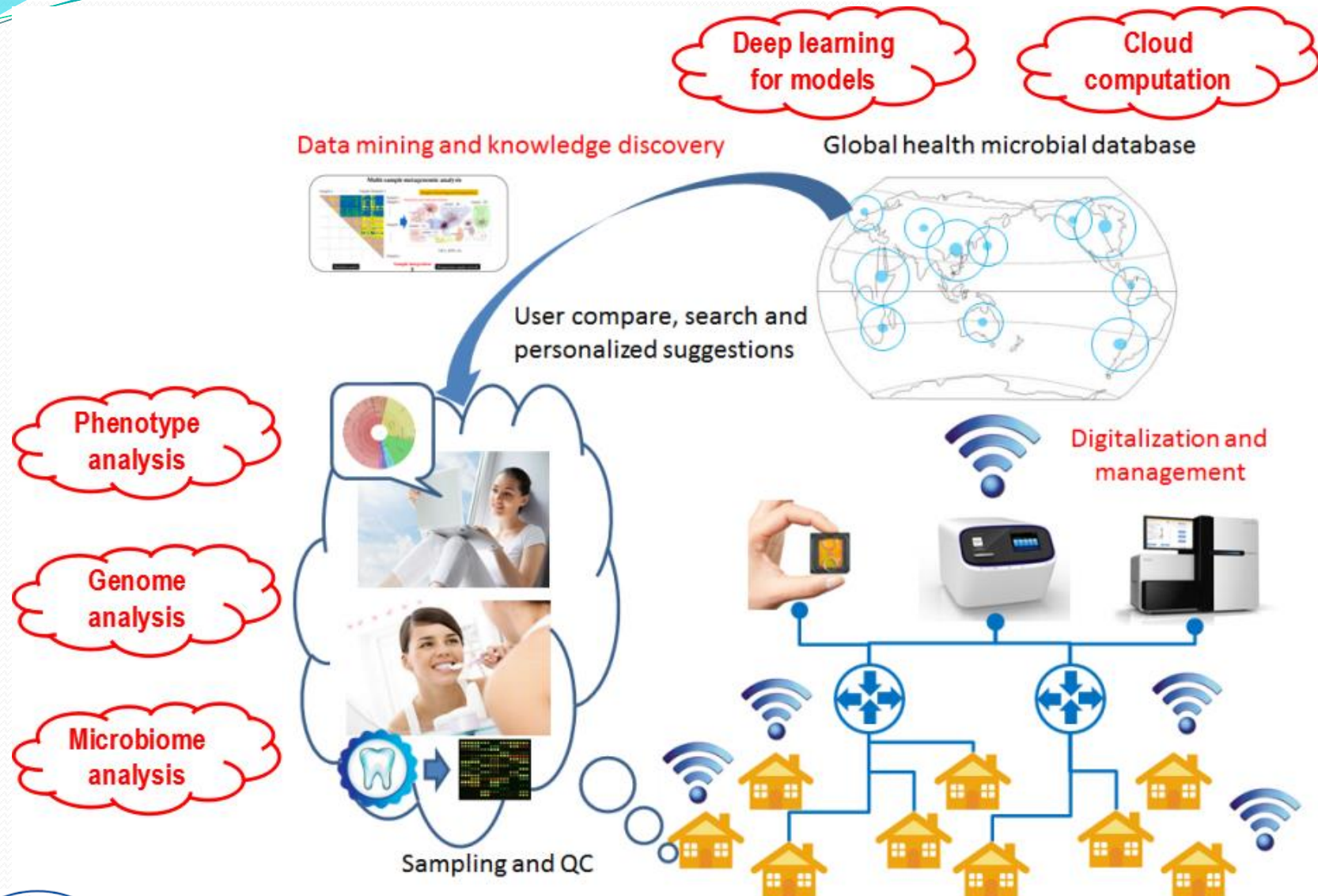
2. Microbiome enterotype (菌群分型)

4. Microbiome recovery (干预性菌群恢复)



Microbiome for Health - Applications

人体健康大数据 (红色)
Human microbiome



Plasticity of gut microbiome

Gut microbiome plasticity

Inadaptability :

desynchronosis (time)

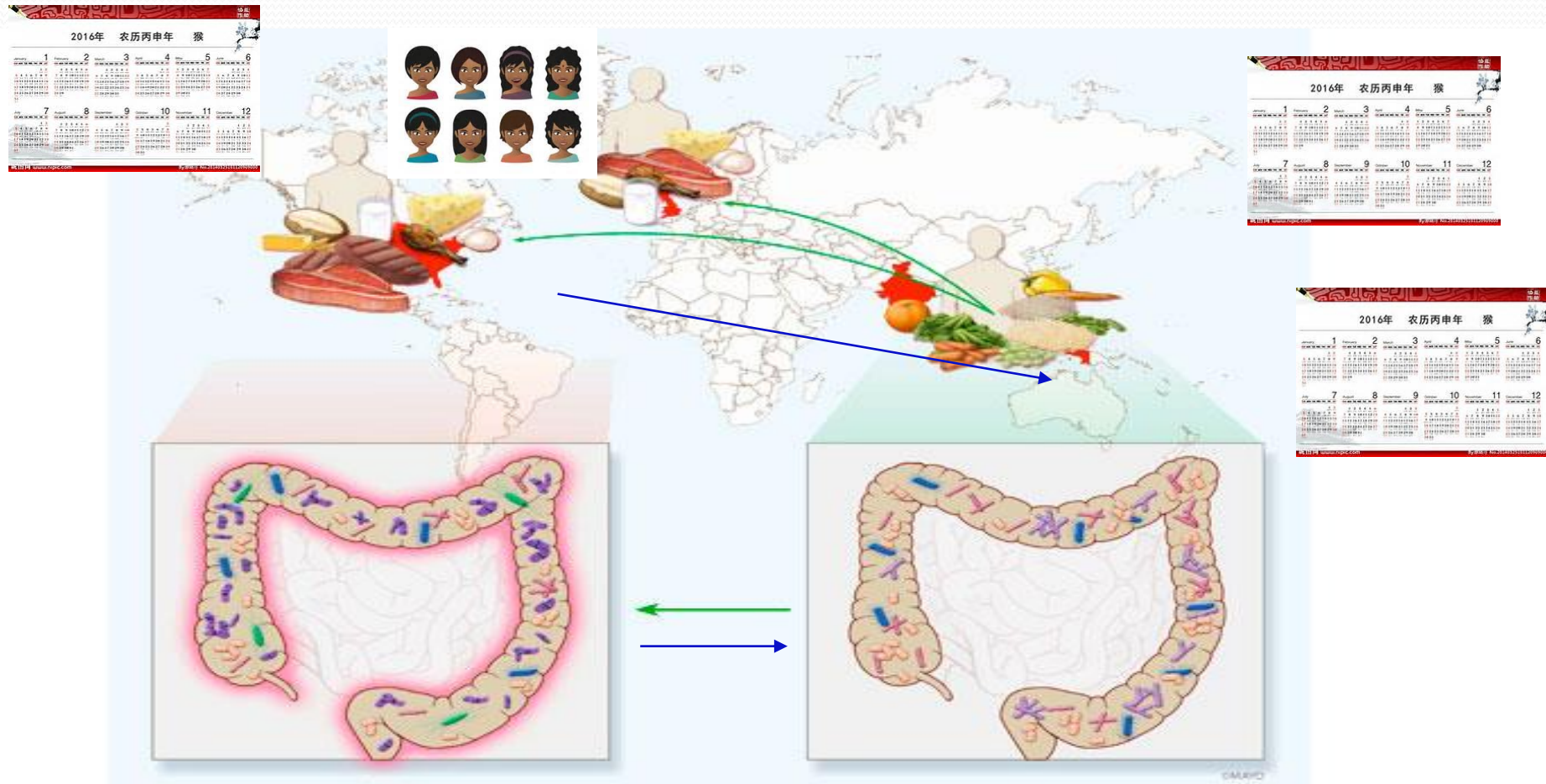
acclimatization (location)



Microbiome for Health - Applications

人体健康大数据 (红色)
Human microbiome

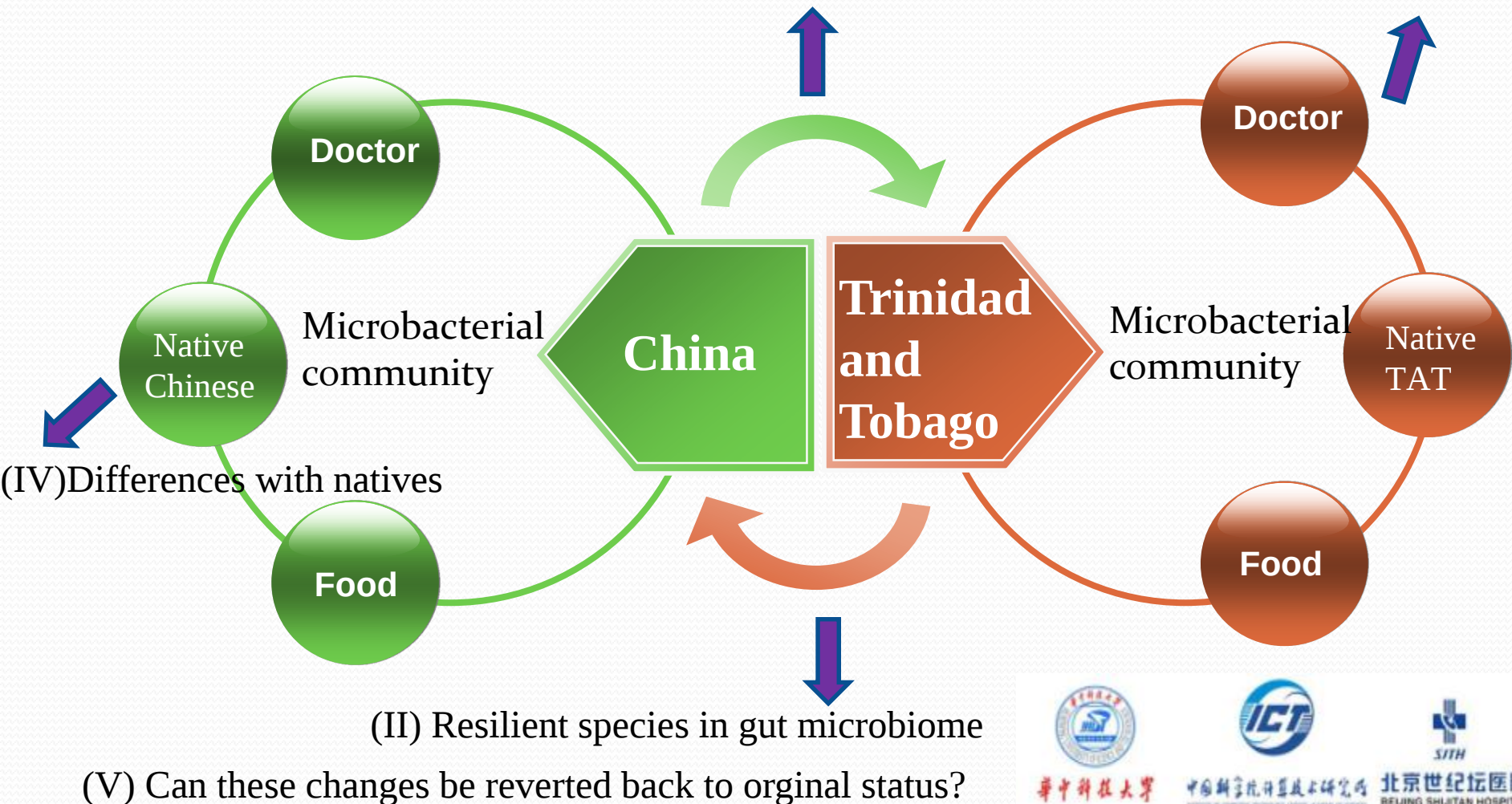
Plasticity of gut microbiome



Plasticity of gut microbiome

(I) Immediate microbiome change after dietar shift

(III) The relationship with environment



Trends of key genus



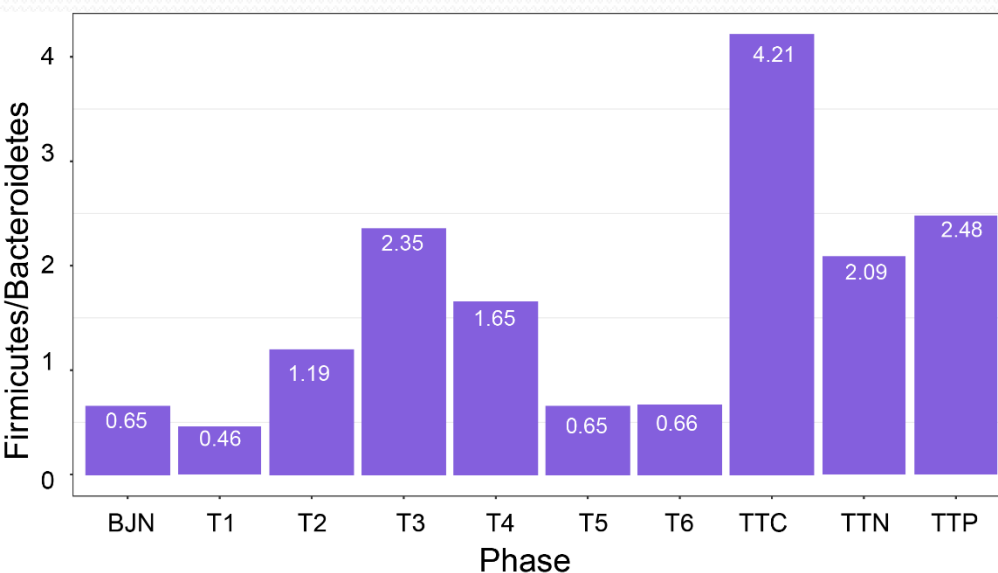
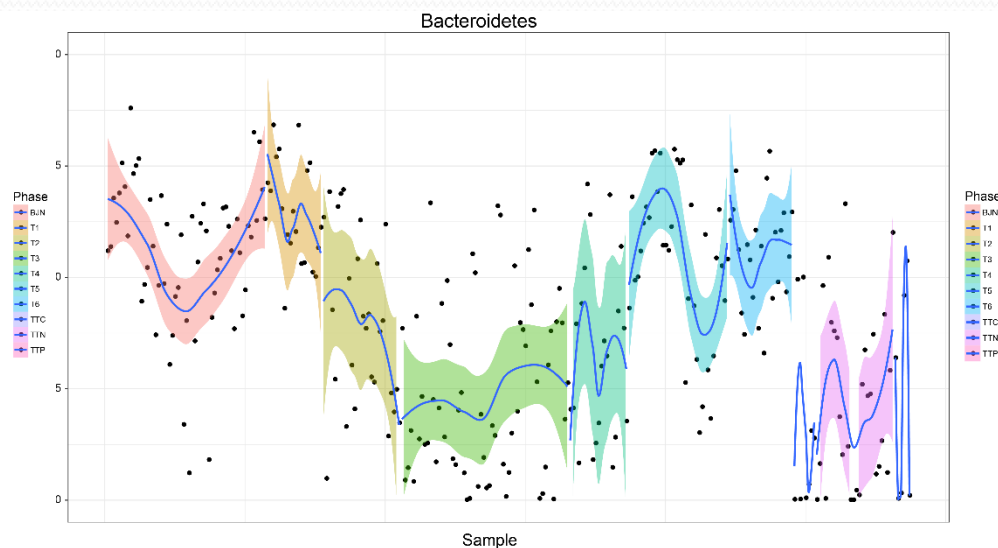
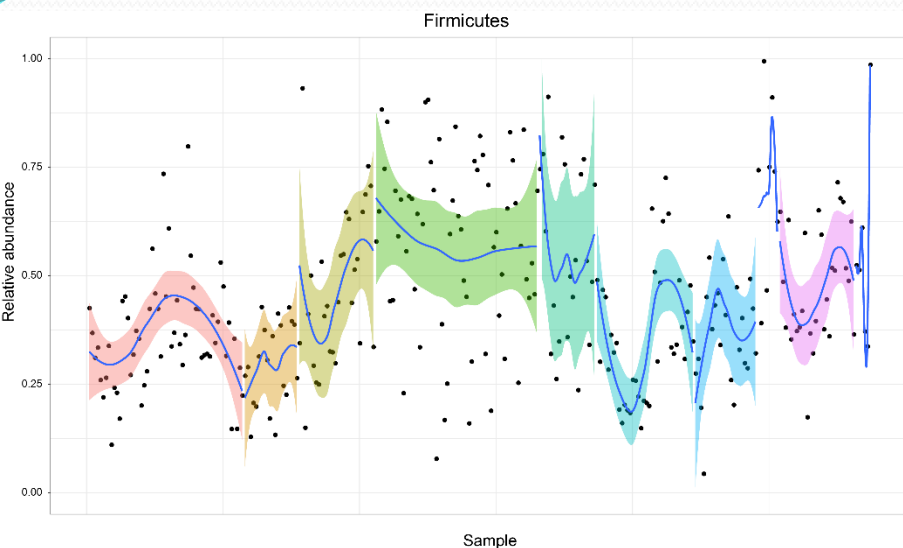
北京世纪坛医院
BEIJING SHIJITAN HOSPITAL



华中科技大学



中国科学院计算机技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

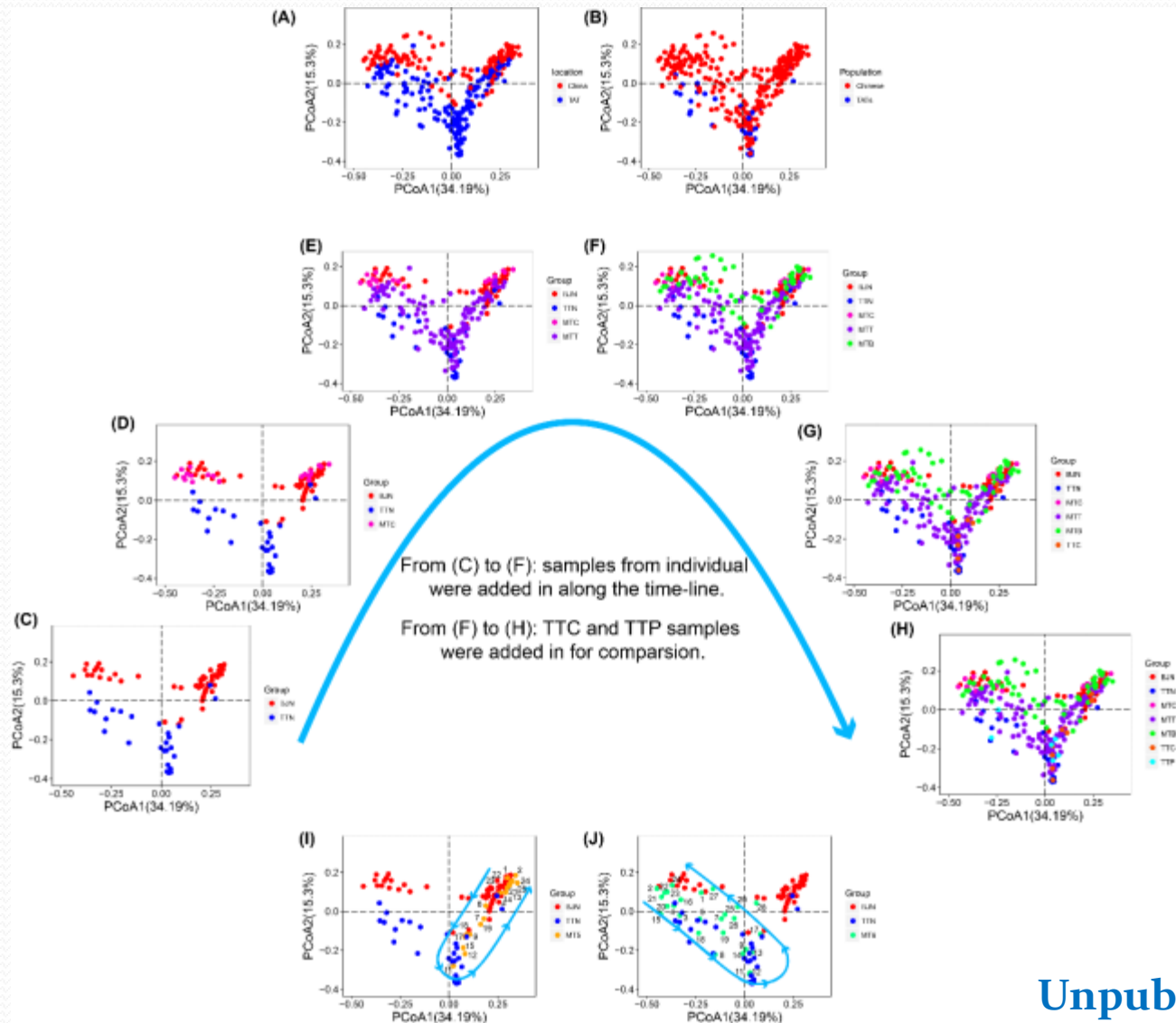


The ratio of *Firmicutes*:
Bacteroidetes, **considerable**
group-wise variations.

Unpublished data

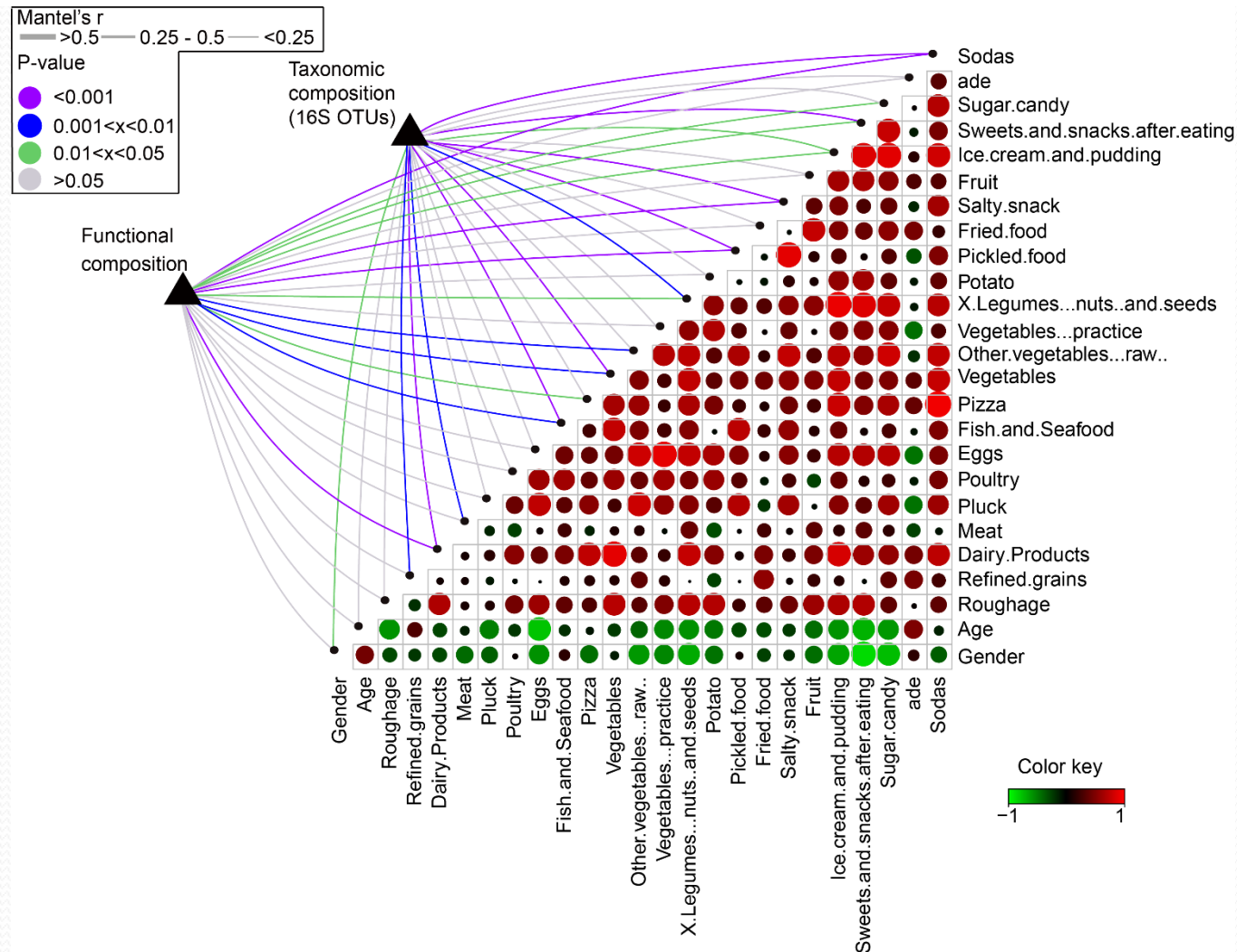
Sample similarity assessment

Clear separation of samples...



Unpublished data

Environmental drivers of community composition



Plasticity of gut microbiome

Gut microbiome plasticity

Summary:

(I) Obvious differences in gut microbiota during Chinese groups stayed at TAT.

- (i) The percentage of **Top 6 phyla**, shifted dramatically.
- (ii) The **ratio of Firmicutes: Bacteroidetes**, considerable individual variations.
- (iii) These **changes** might be associated with acclimatization.

(II) The stable states of the gut microbiota during stayed at TAT.

- (i) 70,544 OTUs, **18 OTUs**, all samples and 68 OTUs, more than 90% samples.
- (ii) Important: **correlations** between 18 core OTUs.

(III) The PCoA and CCA results also revealed the differences in gut microbiota.

- (i) **Taxonomical composition** of samples from the **same person with different time**, **immediate changes**.
- (ii) the samples from same person showed **obvious clustering**.



1. Microbiome heterogeneity (菌群异质性)

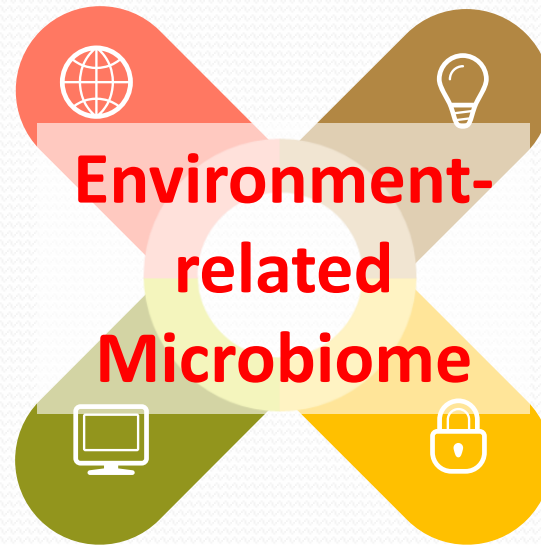
应用：
淡水水体生态

应用：
海洋抗性基因

3. Microbiome dynamics (菌群动态特征)

应用：
校园微生物季节性

应用：
城市环境微生物组



2. ARGs in microbiome (菌群中的抗性基因)

4. Microbiome for cities (菌群和健康中国)

Microbiome for environment - Applications

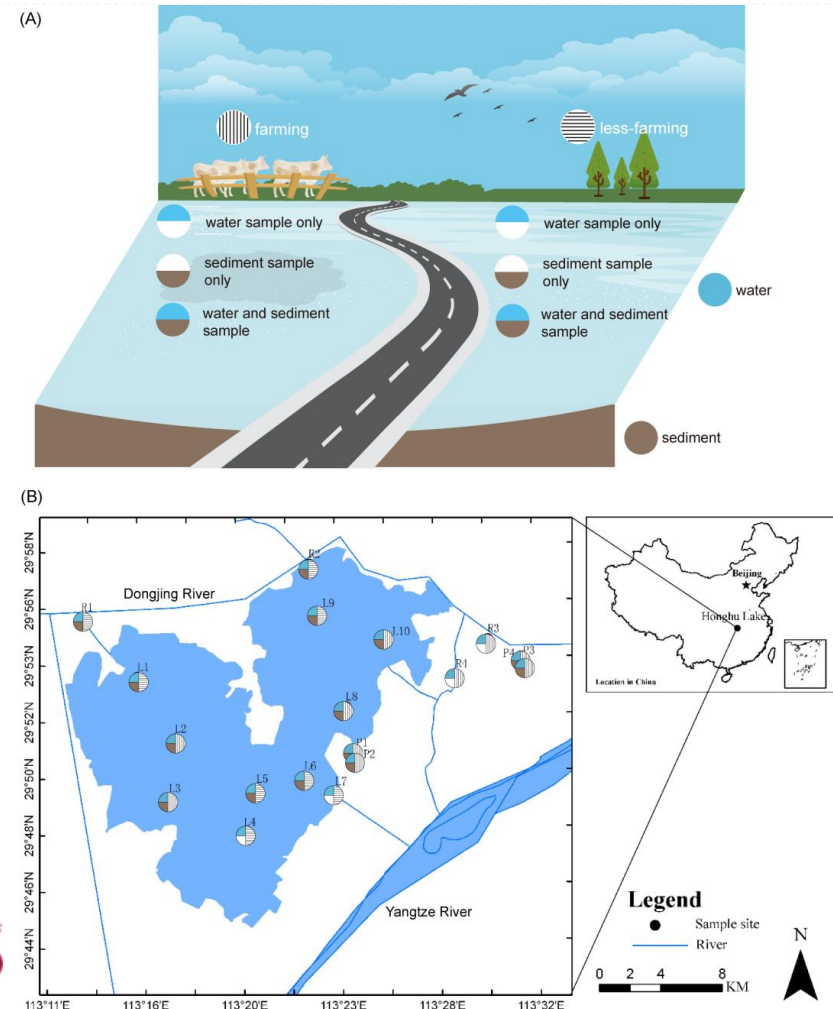
Freshwater lake data mining

健康的生存环境（绿色） Healthy environments

Focusing on spatial differences!

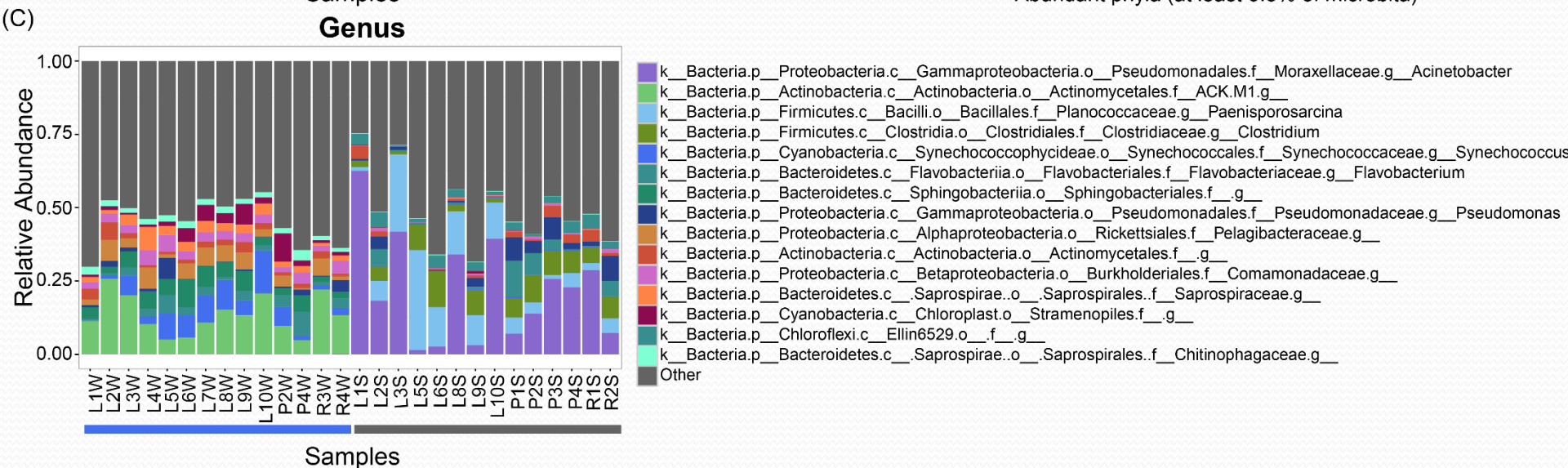
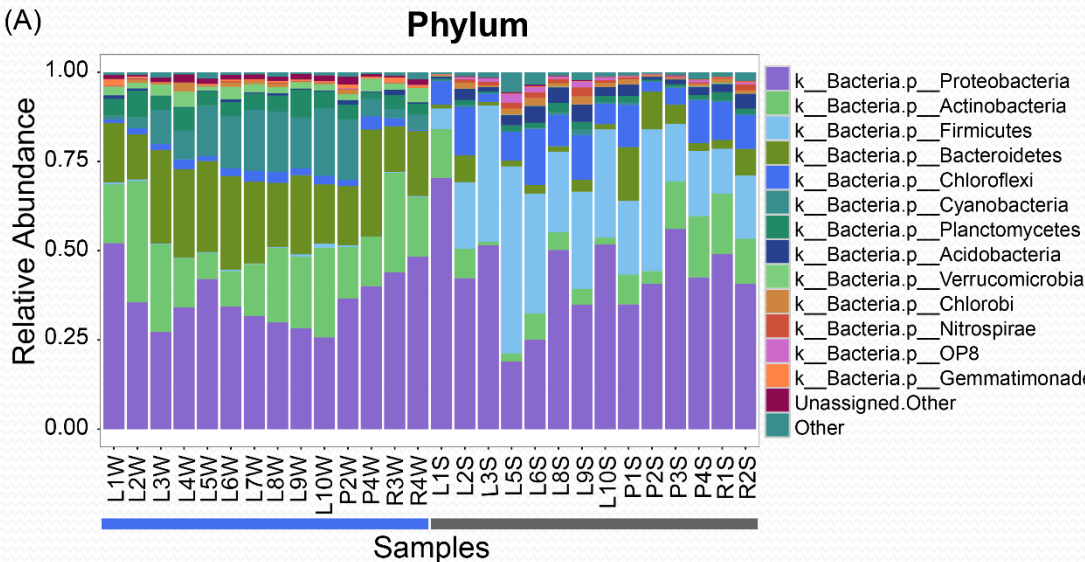
We have more questions:

- (i) The relationship of lake microbiome with eutrophication
- (ii) The relationship of lake microbiome with **antibiotics**
- (iii) The relationship of lake microbiome with **herbicides**
- (iv) The dynamics of lake microbial communities **in time-series**
- (v) ...



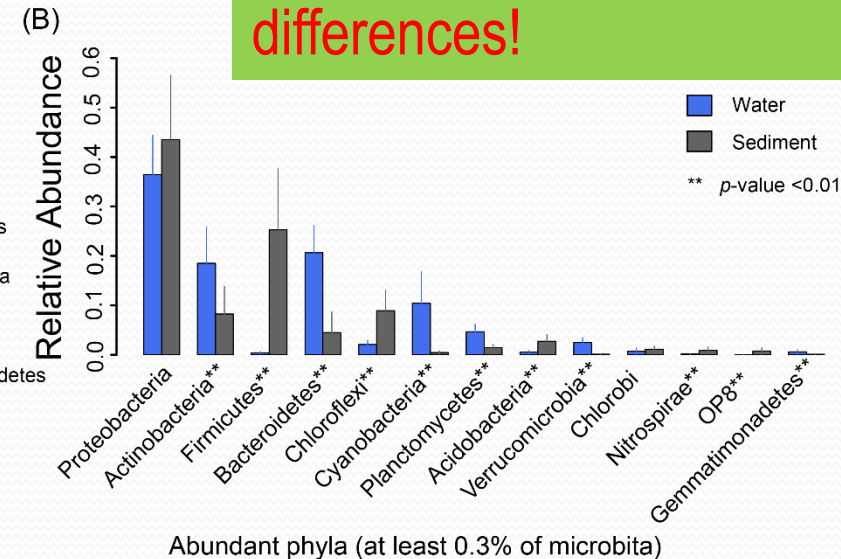
Microbiome for environment - Applications

Freshwater lake data mining



健康的生存环境（绿色） Healthy environments

Focusing on spatial differences!

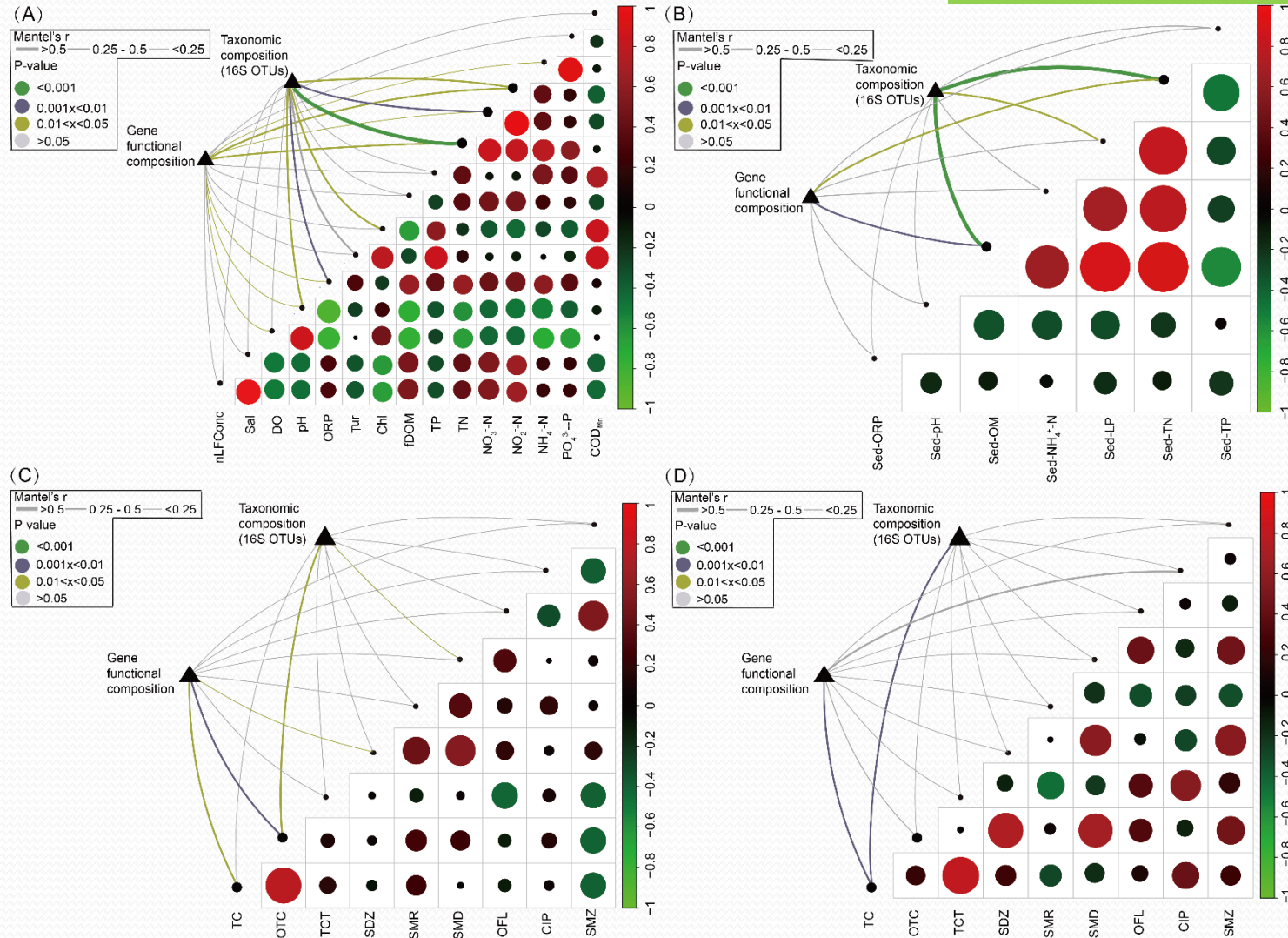


Microbiome for environment - Applications

健康的生存环境（绿色）
Healthy environments

Freshwater lake data mining

Focusing on spatial differences!

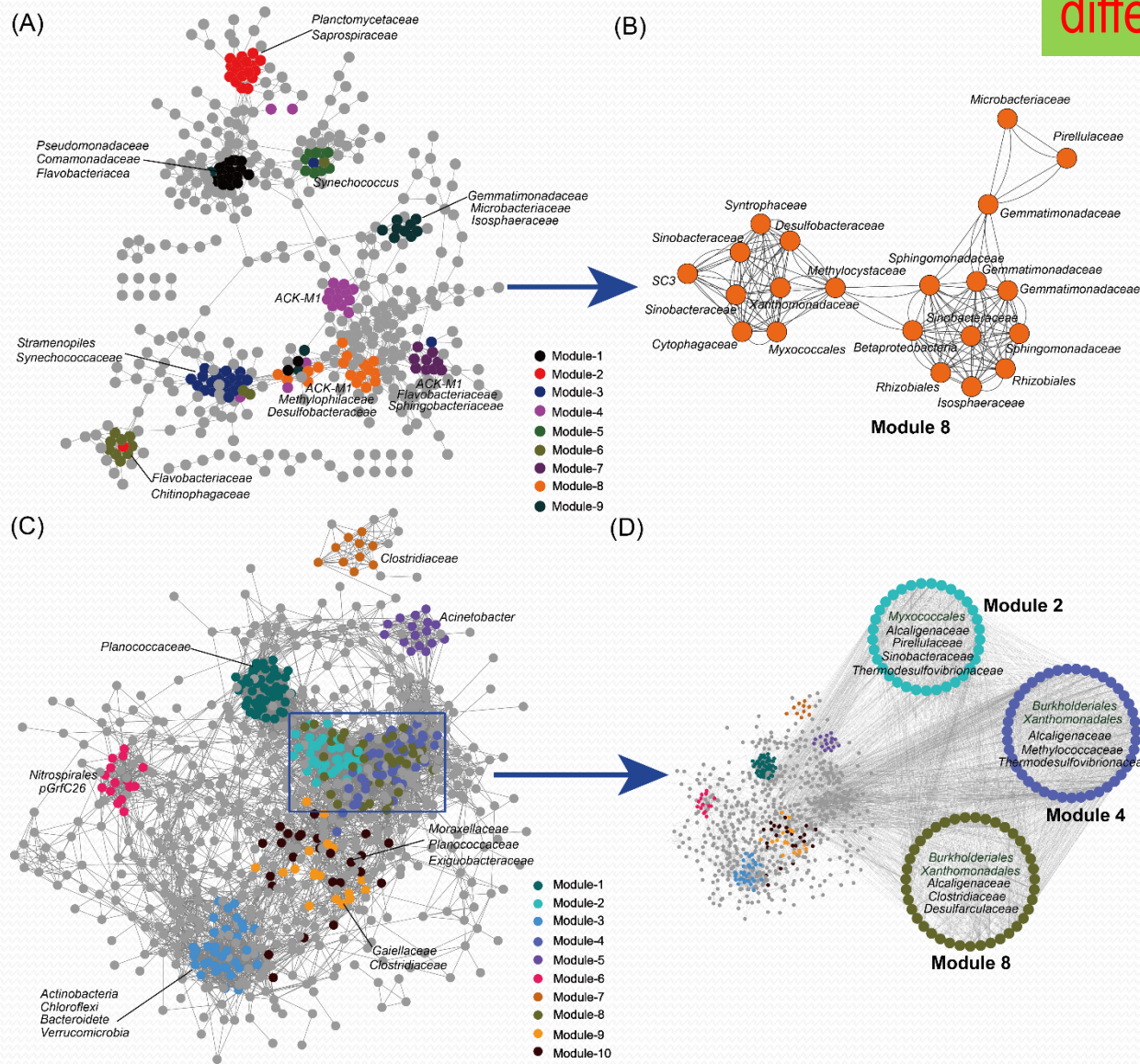


Microbiome for environment - Applications

健康的生存环境（绿色）
Healthy environments

Freshwater lake data mining

Focusing on spatial differences!



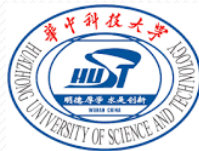
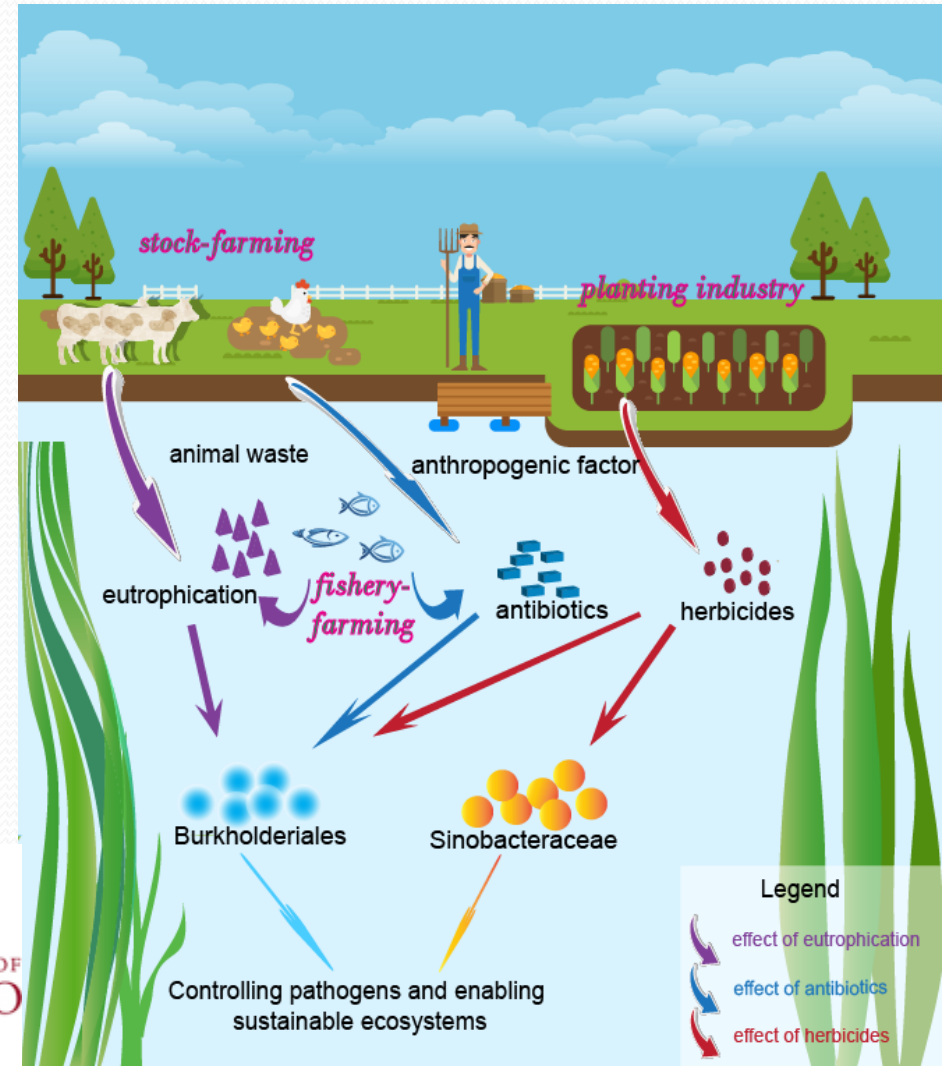
Microbiome for environment - Applications

Freshwater lake data mining

- (i) The dynamics of microbial communities in Honghu: water-body and sediments
There are significant differences among them.
- (i) The species-species co-occurrence ecological patterns
There exist core- and pan-microbiota in freshwater lake.
- (ii) The relationship of microbial communities with environmental factors
The relationships with eutrophication, herbicide, antibiotics.

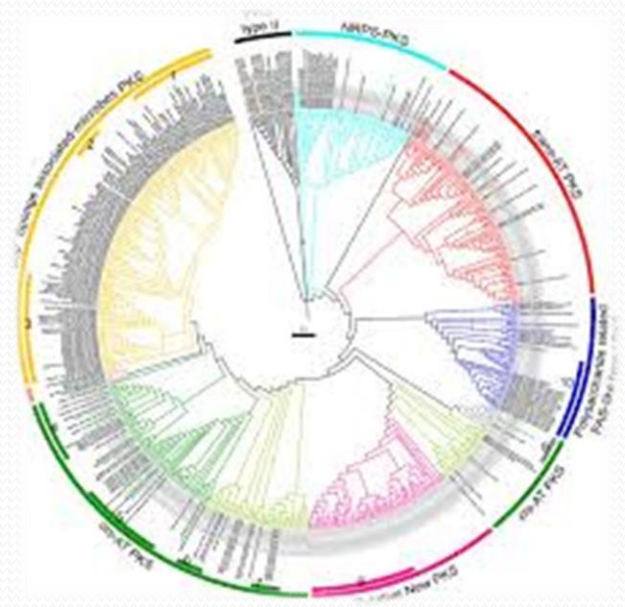
健康的生存环境（绿色） Healthy environments

Focusing on spatial differences!



How to push them forward?

From microbial communities to big-data...



From data to knowledge and treatment
...
in which collaborations would be essential!

Modified based on:
Lauren Davis with sources via Shutterstock

Collaboration network



Acknowledgements

<http://www.microbioinformatics.org/>



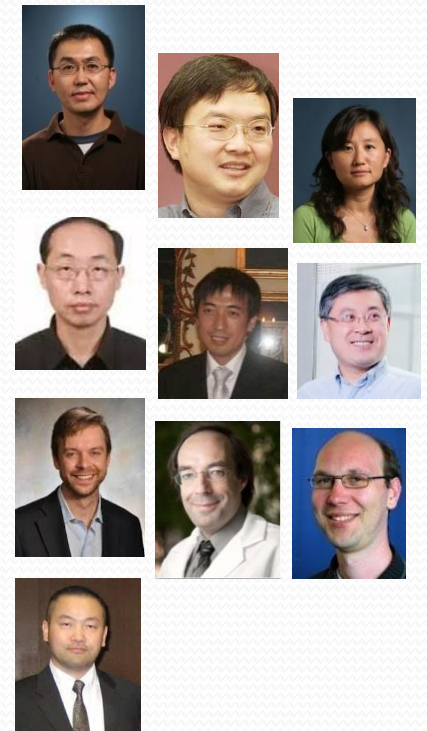
Members:

- **Staff:** H Bai, YM Zhao
- **Students:** MZ Han, Q Yao, CY Chen, CF Zhong, CY Tan, PS Yang, MY Cheng, X Gao, YG Zha, X Zhao, X Zhu



Collaborators:

- Jack Gilbert (U. Chicago) (on metagenomics)
- JIANG Tao (UC riverside, USA; ACM Fellow) (on metagenomics)
- WONG Limsoon (NUS, Singapore) (on omics)
- CUI Xingping (UC riverside, USA) (on biostatistics and metagenomics)
- BU DB(CAS-ICT) (on computational algorithms)
- ZHANG Yang (U. Michigan) (on protein structure)
- Alexey Nesvizhskii (U. Michigan, USA) (on proteomics)
- Ansgar Poetsch (RUB, Germany) (on proteomics)





Thank you!



