


Xiao-Ning Zhang

Download this Curriculum Vitae:
 (Last updated June 6, 2022.)

Basic Info

✉ : tanklovermermaid@gmail.com

💎 : <https://tank.netlify.app/>

👤 : github.com/TankMermaid

🌱 : [tank_mermaid](#)

- Bioinformatics Data Scientist and Developer of Bioinformatic Utility Tools
- Developer of Bioinformatic Pipeline and Maintainer of Debian-based distribution of HPC @BaiLab
- Algorithms Optimizing for Genomics

Education and Institutions

2015–2022	Ph.D. in Bioinformatics <i>Institute of Genetics and Developmental Biology(IGDB), Chaoyang</i> Focusing on Machine Learning approach on Microbiome. Advisor: Yang Bai .
2011–2015	B.S. in Quantitative Biology <i>Liangxi Innovation Class, Beijing Forestry University(BJFU), Haidian</i> Major in Quantitative Biology. Minor in Math.

Dissertations

PHD DISS.	Title: Host Associated Quantitative Abundance Profiling and Establishment of Rice Root-associated Microbial Metagenomic Catalogue Advisor: Yang Bai <ul style="list-style-type: none">• We used the Spike-in strategy to establish a Host-Associated Quantitative Abundance Profiling workflow to accurately detect the total amount of microorganisms• We have applied a large-scale genome-resolved metagenomic approach to establish a gene and genomic catalogue of the rice root microbiome• We have developed some tool packages for bioinformatic analysis, such as Kraken2 customized database, NCBI genomes downloader and Rarefaction,etc.
-----------	--

Publications

2020	Xiaoxuan Guo#, Xiaoning Zhang #, Yuan Qin#, Yong-Xin Liu, Jingying Zhang, Na Zhang, Kun Wu, Baoyuan Qu, Zishan He, Xin Wang, Xinjian Zhang, Stephane Hacquard, Xiangdong Fu* & Yang Bai*. Host-Associated Quantitative Abundance Profiling Reveals the Microbial Load Variation of Root Microbiome. <i>Plant Communications</i> . 2020, doi:10.1016/j.xplc.2019.100003 HAQAP Method for Absolute Quantification
2019	Jingying Zhang#, Yong-Xin Liu#, Na Zhang#, Bin Hu#, Tao Jin#, Haoran Xu, Yuan Qin, Pengxu Yan, Xiaoning Zhang , Xiaoxuan Guo, Jing Hui, Shouyun Cao, Xin Wang, Chao Wang, Hui Wang, Baoyuan Qu, Guangyi Fan, Lixing Yuan, Ruben Garrido-Oter, Chengcai Chu* & Yang Bai*. NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. <i>Nature Biotechnology</i> . 2019, doi:10.1038/s41587-019-0104-4 RandomForest Classification in Crop Microbiome

2018	<p>Jingying Zhang#, Na Zhang#, Yong-Xin Liu#, Xiaoning Zhang, Bin Hu, Yuan Qin, Haoran Xu, Hui Wang, Xiaoxuan Guo, Jingmei Qian, Wei Wang, Pengfan Zhang, Tao Jin*, Chengcai Chu* & Yang Bai*. <i>Root microbiota shift in rice correlates with resident time in the field and developmental stage</i>. <i>Science China Life Sciences</i>. 2018, doi:10.1007/s11427-018-9284-4</p> <p>Time Course and Machine Learning in Microbiome</p>
------	--

Languages

HUMAN	Mandarin, English(CET6), some Russian and Japanese.
MACHINE	<ul style="list-style-type: none"> • Compiled: C++, Java and some knowledge of Golang; • Script: Python/Perl, R/Rstudio, zsh/shell and some JavaScript • Markup L.: \LaTeX/X₃TeX, R Markdown, HTML, CSS. • Automation: GNU Make and Snakemake • Container: Docker and Singularity • Database: MySQL, MongoDB, SQLite, Redis, etc. • Genomics: Deep into source code of GATK, Bowtie2, Kraken2, Minhash etc.

Research experience

- **2018-2022 PhD Graduate Student with Dr. Yang Bai**
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
 - **Project 1: Compositional Data Analysis(CoDA) and applications in root microbiota associations and pathogen diagnostics**
Plant-associated microbes are critical for plant growth and survival under natural environmental conditions. To date, most plant microbiome studies involving high-throughput amplicon sequencing have focused on the relative abundance of microbial taxa. However, this technique does not assess the total microbial load and the abundance of individual microbes relative to the amount of host plant tissues. Here, we report the development of a host-associated quantitative abundance profiling (HA-QAP) method that can accurately examine total microbial load and colonization of individual root microbiome members relative to host plants by the copy-number ratio of microbial marker gene to plant genome. We validate the HAQAP method using mock experiments, perturbation experiments, and metagenomic sequencing. The HA-QAP method eliminates the generation of spurious outputs in the classical method based on microbial relative abundance, and reveals the load of root microbiome to host plants. Using the HA-QAP method, we found that an increase in total microbial load represents a key feature of changes in root microbiome of rice plants exposed to drought stress and of wheat plants with root rot disease, which significantly influences patterns of differential taxa and species interaction networks. Given its accuracy and technical feasibility, HA-QAP would facilitate our understanding of genuine interactions between root microbiome and plants
Have done:
 - * Mathematical principle parsing for dealing with Compositional Data
 - * Developing the spike-in based reference workflow for curation of composition profiling(HAQAP)
 - * Co-occurrence Network Analysis in microbiome data
 - **Project 2: Establishment of Rice Root-associated Microbial Metagenomic Catalogue**
Rice root microbiota play an important role in agriculture and ecology, and participate in the nitrogen

and phosphorus recycle between soil and rice. However, in nature, although a large proportion of bacteria are cultivated in reproducible, high-abundance taxa based on 16s rDNA species classification, there is still a high proportion of bacteria in the same taxon based on 16s rDNA, of whose genetic diversity was not captured. The independent-culture genome-resolved metagenome method can reconstruct and deconvolute complex genomes from microbial communities through sequence assembly and binning, and obtain Metagenome Assembled Genomes (MAGs). In turn, microbiome species and functional are profiled with annotation. However, in the field of plant root microorganisms, there is no standard golden microbial database that can integrate both taxonomic and functional composition. Therefore, the establishment of microbial reference gene catalogue and genome catalogue is an important and urgent step to accurately characterize the taxonomy and function of root microbiome species in rice and even other crops (wheat, maize, etc.).

Have done:

- * Metagenome analysis pipeline construction based on snakemake
- * We applied a large-scale genome-resolved metagenomic approach to establish a gene catalogue of the rice root microbiome, and deconvolute 849 MAGs of medium and high quality.
- * The relevant public gold database is integrated into the non-redundant integrated microbial genome catalogue as NRIMG
- * Identified some gene/pathways/taxa ecogroup between root-associated and soil-associated with NRIMG

– **Project 3: Kraken2 customized database Construction and Microbiome-Oriented software packages development**

Since most MAGs lack species annotation information in NCBI RefSeq, an imbalance was introduced in the proportion of species annotated with Kraken2 software and functional annotation with KEGG. GTDB (Genome Taxonomy Database) which emerges as next-generation genome-based taxonomy classification database could provide the possible access to annotation for MAGs. Meanwhile, how to embed more and more MAGs into the Kraken2 database to improve the performance of the reference database, we have established a set of Snakemake-based workflow to build Kraken2 customized database, custom database significantly improves the rate of species classification. In addition, metagenomic data as a whole presents the characteristics of big data and high computing power consumption. Rarefaction is a commonly used standardization method for microbiome data analysis. Most software cannot handle large data sets friendly. We developed a Cython-based package rarefaction for generating rarefaction-accumulation curve for this purpose.

Have done:

- * Building Kraken2 custom database and Optimizing Kraken2 software written in C++.
- * Constructing NCBI-formatted taxonomy tree de novo with custom python package
- * Developing a Cython-based package rarefaction for generating rarefaction-accumulation curve
- * Developing utility tools to handle concurrent tasks such as ncbi-genome downloading with Golang.

• **2016-2018 Master Graduate Student with Dr. Yang Bai**

Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

– **Project 1: Time-course analysis and Machine Learning approach on root microbiota composition**

Land plants in natural soil form intimate relationships with the diverse root bacterial microbiota. A growing body of evidence shows that these microbes are important for plant growth and health. Root microbiota composition has been widely studied in several model plants and crops; however, little is known about how root microbiota vary throughout the plant's life cycle under field conditions. We performed longitudinal dense sampling in field trials to track the time-series shift of the root microbiota from two representative rice cultivars in two separate locations in China. We found that the rice root microbiota varied dramatically during the vegetative stages and stabilized from the beginning of the reproductive stage, after which the root microbiota underwent relatively minor changes until rice ripening. Notably, both rice genotype and geographical location influenced the patterns of root microbiota shift

that occurred during plant growth. The relative abundance of Deltaproteobacteria in roots significantly increased overtime throughout the entire life cycle of rice, while that of Betaproteobacteria, Firmicutes, and Gammaproteobacteria decreased. By a machine learning approach, we identified biomarker taxa and established a model to correlate root microbiota with rice resident time in the field.

Have done:

- * 16S rRNA gene sequencing analysis pipeline construction
 - * Analyzed the rice root microbiota data with time-series analysis tools
 - * Fit the data with regression model
 - * Building machine learning approach, we identified biomarker taxa and established a model to correlate root microbiota with rice resident time
- **Project 2: Investigating host genetics shaping the microbiota with Genome-wide Analysis approach**

Nitrogen-use efficiency of indica varieties of rice is superior to that of japonica varieties. We apply 16S ribosomal RNA gene profiling to characterize root microbiota of 68 indica and 27 japonica varieties grown in the field. We find that indica and japonica recruit distinct root microbiota. Notably, indica-enriched bacterial taxa are more diverse, and contain more genera with nitrogen metabolism functions, than japonica-enriched taxa. Using genetic approaches, we provide evidence that NRT1.1B, a rice nitrate transporter and sensor, is associated with the recruitment of a large proportion of indica-enriched bacteria. Metagenomic sequencing reveals that the ammonification process is less abundant in the root microbiome of the nrt1.1b mutant. We isolated 1,079 pure bacterial isolates from indica and japonica roots and derived synthetic communities (SynComs). Inoculation of IR24, an indica variety, with an indica-enriched SynCom improved rice growth in organic nitrogen conditions compared with a japonica-enriched SynCom. The links between plant genotype and root microbiota membership established in this study will inform breeding strategies to improve nitrogen use in crops

Have done:

- * Population and quantitative genetics theory construction
 - * GWAS Pipeline Setup and Maturation
 - * Although We failed in discovery candidate genes shaping microbiota with GWAS approach, we identified the specific gene(NRT1.1B) could recruit specific bacteria
 - * Skilled at partial regression and LASSO/Ridge regression and Mixed Generalized Linear Model(GLM) in GWAS-based studies
 - * Skilled at GATK in SNP calling and association analysis with Plink or Tassel
- **2015-2016 Master Graduate Student**
Intensive Teaching in Huairou, Chinese Academy of Sciences, Beijing, China

Have done:

- P-10 Program for Advanced Computer Science

Online Classes Taught

2019 SPRING	Statistics – “Elements in Statistics” – Graduate Assistant <i>Institute of Genetics and Developmental Biology(IGDB)</i>
2019 FALL	Statistics – “Mathematical Statistics II” – Online Class <i>Institute of Genetics and Developmental Biology(IGDB)</i>
2018 FALL	Statistics – “Multivariate Statistical Analysis I” Online Class <i>Institute of Genetics and Developmental Biology(IGDB)</i>
2019 FALL	ML and DL – “Artificial Intelligence & Deep Learning” – Online Class <i>Institute of Genetics and Developmental Biology(IGDB)</i>

2016 SPRING	ML and DL – “Machine Learning on Big Data” Teaching Assistant <i>University of Chinese Academy of Sciences(UCAS) Huairou</i>
2019 SPRING	Data Analysis – “Time Series Analysis and Application” <i>Institute of Genetics and Developmental Biology(IGDB)</i>
2017 SPRING	Bioinformatics – “ Computational Genomics” – Teaching Assistant <i>University of Chinese Academy of Sciences(UCAS)</i>
2016 SPRING	Bioinformatics – “Algorithms in Bioinformatics” – Teaching Assistant <i>University of Chinese Academy of Sciences(UCAS)</i>
2017 SPRING	Network Science– “Intro to Multi-agent System” – Teaching Assistant <i>University of Chinese Academy of Sciences(UCAS)</i>

References

- Dr. Yang Bai PhD Advisor Email: ybai@genetics.ac.cn Phone: +86-10-64804201
Address: Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
- Dr. Fei Lu PhD Co-Advisor flu@genetics.ac.cn
Address: Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
- Dr. Jun Wang PhD Co-Advisor junwang@im.ac.cn
Address: Institute of Microbiology, Chinese Academy of Sciences, Beijing, China