

# **MANUAL v1.0**

Web site: <https://gcmeta.wdcm.org/>

Microbial Resource and Big Data Center & State Key Laboratory of Microbial Resources,  
Institute of Microbiology, Chinese Academy of Sciences

# Initial Step: Login

## registration

*An account already, please click here!*

### New User

User name\*:

Password\*:

Repeat password\*:

E-Mail\*:

Given Name\*:

Family Name\*:

Title\*:

Organization\*:

Department\*:

Country\*:

Address\*:

Research Interests:

FAX:

**After verification, an confirm email will be received.**

## login



微生物组数据云平台

Microbiome Data Cloud Platform

Administrator  Common User

guest

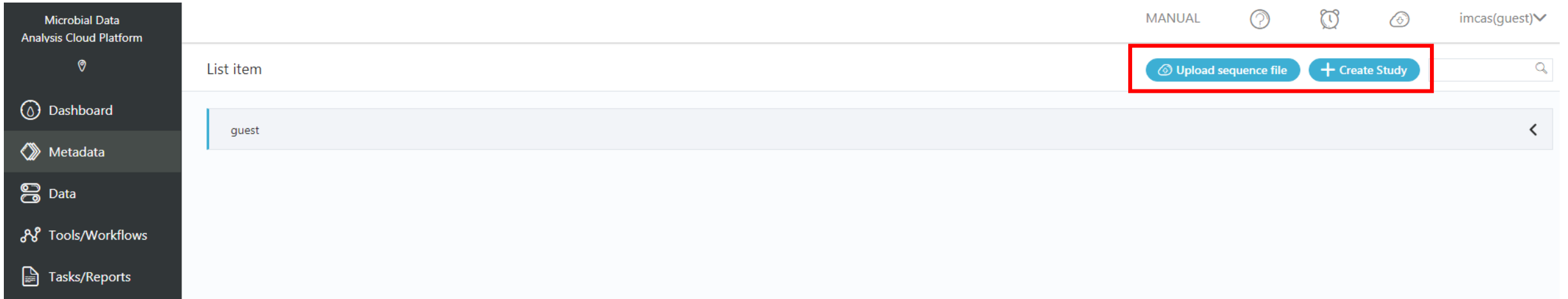
Password\*

.....

\*For data protection, login is required before data submission and exploring the full functions. We provide a temporary guest account ( Username : guest Password: guest ) effective for 24 hours along with any submissions, uploaded files and analysis results.

Login

# Metadata management



List item / Add study

\* StudyName:  
Text

\* Description:  
Text

ParentStudy:  
Option

Study Funding:  
Text

**Publicity**

private Private data can be viewed and accessed by you and your organization only. The quantity calculation ranks behind.

public Public data will be open to the public and is a part of the global establishment of global microbiological metadata. Data calculation will be prioritized.

Cancel Submit

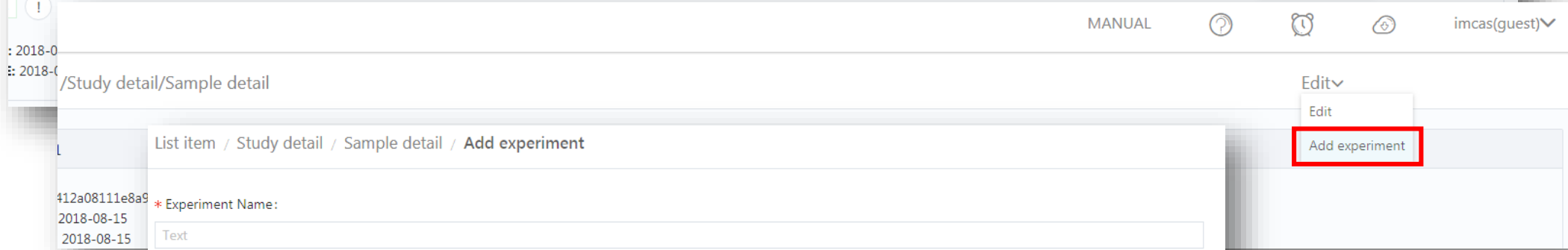
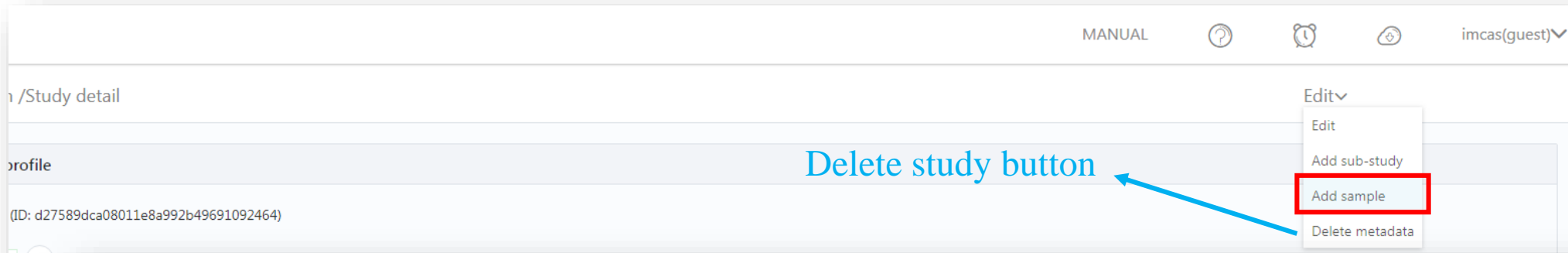
**Principal Investigator Information**

\* PI Name:

The form with ‘\*’ marks the required fields.

All the fields must follow the data standards which have been stated in the gcMeta article.

# Metadata management



List item / Study detail / Sample detail / Add experiment

412a08111e8a9  
2018-08-15  
2018-08-15

\* Experiment Name:

Text

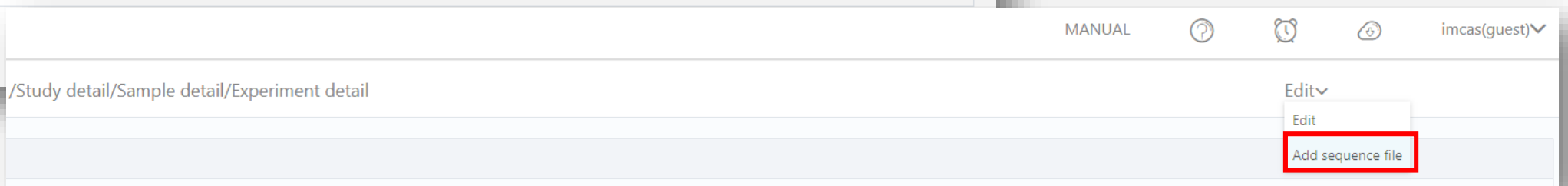
Description:

Text

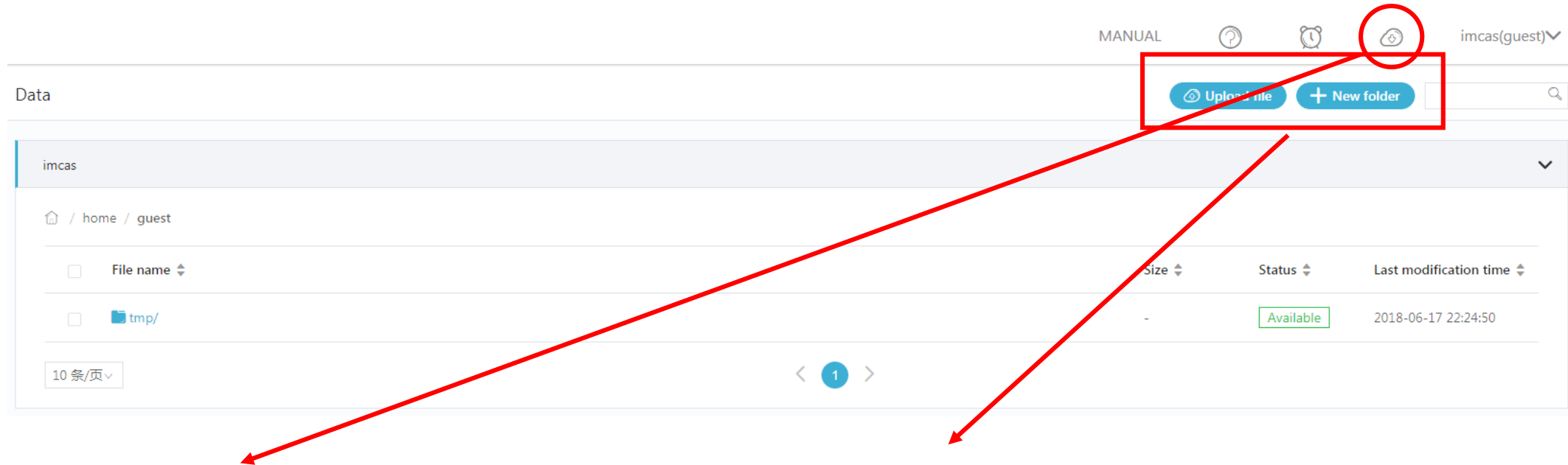
\* Investigation Type:

Option ▼

**After the study is created, sample, experiment, sequence and report can be created subsequently.**



# Data management



## Data client download:

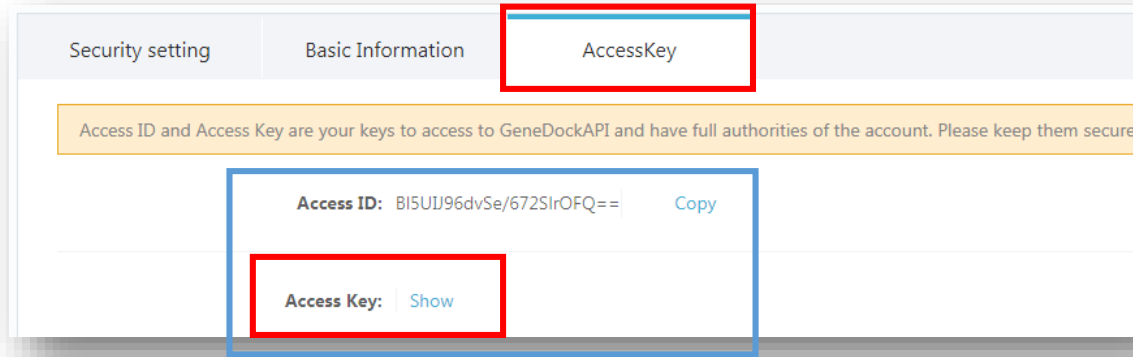
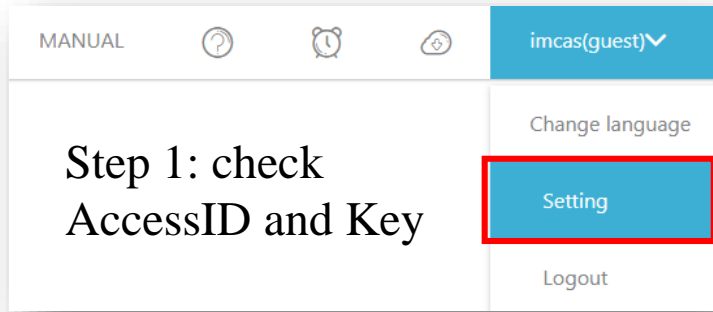
- Command-line client (Linux x64)
- Command-line client (MacOS x64)
- Command-line client (Windows x64)
- Interactive client (MacOS x64)
- Interactive client (Windows x64)
- Interactive client (Windows x86)

Web server upload directly  
(the file must be <100MB file)

**The sequence file can be upload to gcMeta for storage, management and analysis.**

# Data management

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

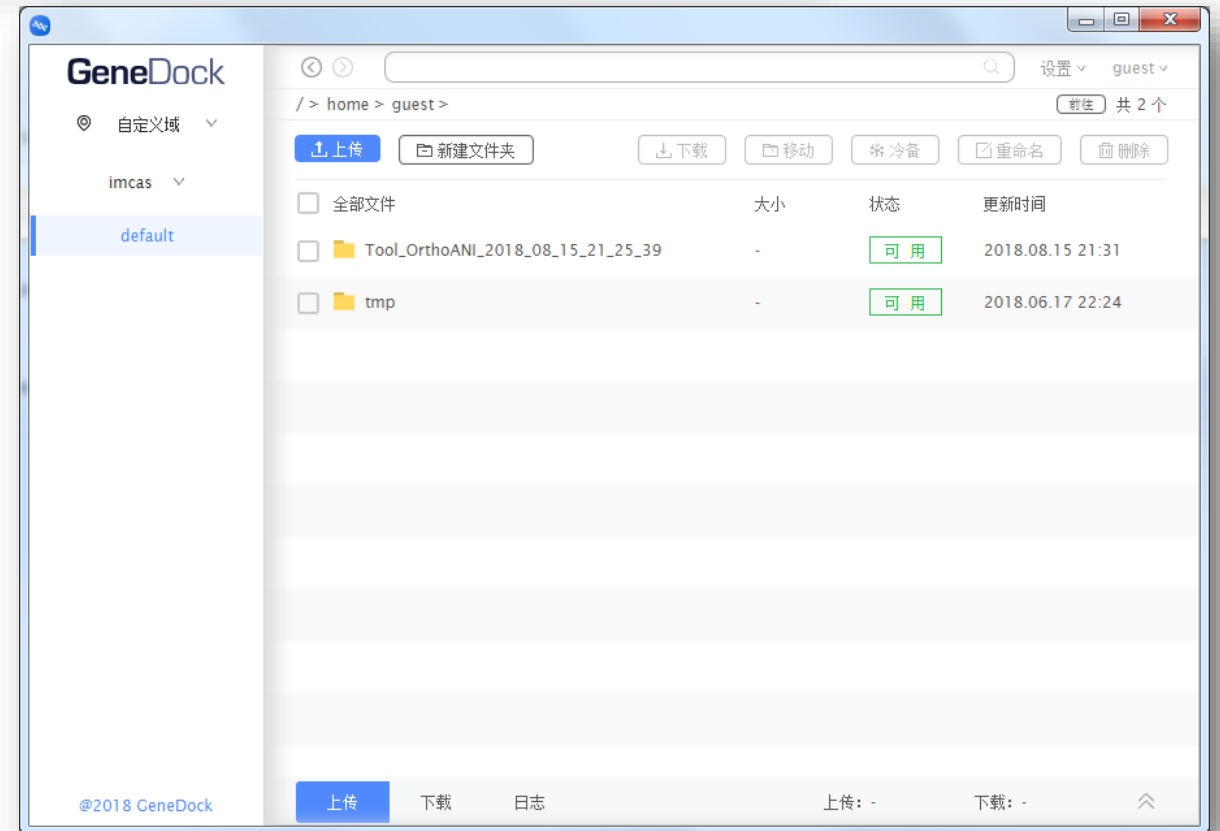


Step 2: modify the entry point

<https://gcmeta-api.wdcm.org>

Step 3: login with accessID and Key

copy & paste



# Data analysis: ENTRY①

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

The screenshot displays the gcMeta web interface. On the left is a dark sidebar with navigation options: Microbial Data Analysis Cloud Platform, Dashboard, Metadata, Data, Tools/Workflows (highlighted with a red box), and Tasks/Reports. The main content area shows a grid of tools and workflows, each with a title, a brief description, and 'Edit' and 'Execute' buttons. The tools listed include:

- Pipeline\_RNA\_seq\_analysis**: RNA-seq analysis pipeline (quality control--alignment--assembly--differential expression analysis)
- Pipeline\_RNAIdentification**: RNAIdentification\_pipeline is a workflow for 16S rRNA gene sequence analysis. It includes several steps: sequence quality control, preprocessing, low quality sequence deletion, rRNA sequence searching, and functional annotation. After all the analysis, a online report will be provided for
- Pipeline\_RNAIdentification\_pair...**: RNA identification workflow for paired fq files.
- Pipeline\_StrainEst**: StrainEst is a novel, reference-based method that uses the Single Nucleotide Variants (SNV) profiles of the available genomes of selected species to determine the number and identity of coexisting strains and their relative abundances in mixed metagenomic samples.
- Pipeline\_trim\_hisat\_samtools**
- Pipeline\_Trimming\_and\_Mapping**: Taking the raw reads contained in the fastq file, trimming them to remove adapter and low quality ends and map them against the reference genome
- Pipeline\_CONCOCT**: CONCOCT "bins" metagenomic contigs. Metagenomic binning is the process of clustering sequences into clusters corresponding to operational taxonomic units of some level.
- Pipeline\_QIIME2**: QIIME2 pipeline for 16s RNA sequences.
- Tool\_AbundanceBin**: AbundanceBin is an abundance-based tool for binning metagenomic sequences, such that the reads classified in a bin belong to species of identical or very similar abundances. AbundanceBin also gives estimations of species abundances and their genome sizes—two important characteristic
- Tool\_ART**: ART is a set of simulation tools to generate synthetic next-generation sequencing reads. ART simulates sequencing reads by mimicking real sequencing process with empirical error models or quality profiles summarized from large recalibrated sequencing data. ART can also simulate reads using
- Tool\_Ballgown**
- Tool\_BBMap**
- Tool\_BBMerge**
- Tool\_BCFTools\_view**
- Tool\_BLASR**

**Various web-based tools and workflows can be chose to perform data analysis based on your demand and data type.**

# Data analysis: ENTRY②

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

**PAGE①**

Tools

click











- Preprocessing
- Assembly
- Structure
- Annotation
- Metagenome Analysis
- Comparative Analysis**
- Visualization

**PAGE②**

- Preprocessing
- Assembly
- Structure
- Annotation
- Metagenome Analysis
- Comparative Analysis**
- Visualization
- Demo

Comparative Analysis

Tools home page and reference

Tools	Key words	Links
<b>click</b> orthoANI	distance calculation	 
CD-hit	protein clustering	 
MUMmer	long sequence alignment	 
BWA ( ALN , MEM , SW )	reads mapping	 
Bowtie2	reads mapping	 

**PAGE③**

Workflow > Execute workflow

Execute: Tool\_OrthoANI\_2018\_08\_15\_20\_25\_09

\* Task name:  
Tool\_OrthoANI\_2018\_08\_15\_20\_25\_09

set input file, output file and parameter

1 Input file      2 Output file      3 Parameter



# Data analysis: ENTRY③

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

## PAGE②

Preprocessing

Assembly

Structure

Annotation











Metagenome Analysis

**Comparative Analysis**






Visualization

**Demo** **click**

### Comparative Analysis

Tools	Key words	Links
orthoANI	distance calculation	 
CD-hit	protein clustering	 
MUMmer	long sequence alignment	 
BWA ( ALN , MEM , SW )	reads mapping	 
Bowtie2	reads mapping	 

### Demo

Tools	Key words	Links
orthoANI	distance calculation	
RNAmmmer	RNA prediction	
R_plot_heatmap	ggplot	
R_plot_PCoA	ggplot	
R_plot_boxplot	ggplot	<b>click</b> 

## PAGE③

Preprocessing

Assembly

Structure

Annotation

Metagenome Analysis

Comparative Analysis

Visualization

**Demo**

### Tool\_boxplot

#### boxplot\_input

Select File

or else, upload your own data and set parameter to rub

#### box\_plot

markerx:

3

outlier:

1

markercol:

1

skippedcolumn:

3

markerfacet:

2

datafunc: **click to load the demo setting including input file and parameters**

none

datamulti:

1

Run

Default Parameters

# Data analysis: DEMO

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

## PAGE waiting

Task: Tool\_boxplot\_2018\_08\_15\_21\_15\_14 Status

App Name	Start Time	End Time	Status
R_plot_boxplot	-	-	waiting

## PAGE result

File Name	File Status	Action
output.pdf	available	download

click to view the result

click to download the file



## Input file

SampleID	Visit	Site	Mother	g_Lactobacillus
NCS.001	DELIVERY	STOOL	INFANT	0
NCS.001	DELIVERY	ORAL	INFANT	0
NCS.001	DELIVERY	INTROITUS	MOTHER	65.2
NCS.001	DELIVERY	ORAL	MOTHER	0
NCS.001	DELIVERY	STOOL	MOTHER	0
NCS.001	DELIVERY	POSTERIOR FORNIX	MOTHER	59.11
NCS.003	DELIVERY	NASAL	INFANT	88.4
NCS.003	DELIVERY	ORAL	INFANT	84.07
NCS.003	DELIVERY	SKIN	INFANT	94.06
NCS.003	DELIVERY	SKIN	INFANT	0.01
NCS.003	DELIVERY	INTROITUS	MOTHER	89.74
NCS.003	DELIVERY	NASAL	MOTHER	0.19
NCS.003	DELIVERY	ORAL	MOTHER	0
NCS.003	DELIVERY	SKIN	MOTHER	0.09
NCS.003	DELIVERY	STOOL	MOTHER	0.4
NCS.003	DELIVERY	STOOL	INFANT	0.02
NCS.003	DELIVERY	POSTERIOR FORNIX	MOTHER	98.81
NCS.004	6 WKS POSTPARTUM	NASAL	INFANT	0
NCS.004	DELIVERY	ORAL	INFANT	84.37
NCS.004	6 WKS POSTPARTUM	ORAL	INFANT	0
NCS.004	6 WKS POSTPARTUM	SKIN	INFANT	0
NCS.004	DELIVERY	INTROITUS	MOTHER	5.54
NCS.004	6 WKS POSTPARTUM	INTROITUS	MOTHER	0
NCS.004	DELIVERY	NASAL	MOTHER	4.23
NCS.004	6 WKS POSTPARTUM	NASAL	MOTHER	0
NCS.004	DELIVERY	ORAL	MOTHER	0.03
NCS.004	6 WKS POSTPARTUM	ORAL	MOTHER	0.01
NCS.004	DELIVERY	SKIN	MOTHER	8.94

skippedcolumn=4

markerx

markercol

markerfacet

## Arguments setting

The usage of the tools and workflows can be seen in manuals.

## Data analysis

### Input setting

The screenshot displays the 'Input setting' section of the gcMeta platform. It features a task execution window titled 'Execute: Tool\_OrthoANI\_2018\_08\_15\_21\_25\_39'. The 'Task name' field contains 'Tool\_OrthoANI\_2018\_08\_15\_21\_25\_39'. Under the 'Input file' section, there are two input fields: 'genome1' and 'genome2'. The 'genome1' field contains the path 'imcas:/test\_kira/OrthoANI/Staphylococcus\_aureus.fa', which is highlighted with a red box. A red text overlay reads: 'Input or paste from manual either input full path of the file or click to choose the file is OK~'. The 'genome2' field contains the path 'imcas:/test\_kira/OrthoANI/Bacillus\_anthraxis.fna'. A 'Next step' button is located at the bottom of the input setting area, also highlighted with a red box. A 'Choose file (fa,fasta,fna)' dialog box is open in the foreground, showing a file browser interface. The dialog has a search bar and a list of files. The current directory is 'imcas'. The list shows two folders: 'test\_kira/' and 'home/'. The 'test\_kira/' folder is selected, and a red arrow points to a folder icon in the background interface with the label 'click'. The dialog has 'Cancel' and 'Confirm' buttons at the bottom.

Execute: Tool\_OrthoANI\_2018\_08\_15\_21\_25\_39

\* Task name:  
Tool\_OrthoANI\_2018\_08\_15\_21\_25\_39

Input file

genome1 **Input or paste from manual**  
imcas:/test\_kira/OrthoANI/Staphylococcus\_aureus.fa

**either input full path of the file or click to choose the file is OK~**

genome2  
imcas:/test\_kira/OrthoANI/Bacillus\_anthraxis.fna

Next step

Choose file (fa,fasta,fna)

Data

imcas

File name	Size	Status	Last modification time
test_kira/	-	Available	2018-01-05 17:22:25
home/	-	Available	2017-05-09 10:09:32

10 条/页 < 1 >

click

## Data analysis

### Output setting

✓ Input file ————— ✓ Output file ————— ③ Parameter

OrthoANI\_output


### Parameter setting

✓ Input file ————— ✓ Output file ————— ③ Parameter

OrthoANI

args\_num\_threads ⓘ :

### Run!!

User	Task name	Creation time	End time	Status
○ guest	Tool_OrthoANI_2018_08_15_21_25_39	2018-08-15 21:33:22	--	▬ Running 

< ① >

# Data analysis

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

The screenshot shows the gcMeta dashboard interface. On the left, a dark sidebar contains navigation options: Microbial Data, Analysis Cloud Platform, Dashboard, Metadata, Data, Tools/Workflows, and **Tasks/Reports** (highlighted with a red box). The main area displays a 'Task/report' table with columns for User, Task name, Creation time, End time, and Status. A task named 'Tool\_OrthoANI\_2018\_08\_15\_21\_25\_39' is highlighted with a red box. The table also includes a filter for dates (2018-08-08 to 2018-08-15) and a search icon. The status of the task is indicated by a green checkmark.

User	Task name	Creation time	End time	Status
guest	Tool_OrthoANI_2018_08_15_21_25_39	2018-08-15 21:33:22	2018-08-15 21:34:13	✓

**check status of the job**

The screenshot shows the task detail page for 'Tool\_OrthoANI\_2018\_08\_15\_21\_25\_39'. A 'Log' button is highlighted with a white box. Below it, a 'Detail log' window is open, displaying the following command sequence:

```
1 + cd /var/data
2 + java -jar /home/bioapp/ANI/OAT_cmd.jar -blastplus_dir
  /home/bioapp/ncbi-blast-2.7.1+/bin -method ani -fastal
  /var/data/5b5ac29d1d41c8006b043a33.fa -fasta2
  /var/data/5b5ac2941d41c80025042be6.fna -num_threads 4
3 + java -jar /home/bioapp/ANI/OAT_cmd.jar -blastplus_dir
  /home/bioapp/ncbi-blast-2.7.1+/bin -method ggdc -fastal
  /var/data/5b5ac29d1d41c8006b043a33.fa -fasta2
  /var/data/5b5ac2941d41c80025042be6.fna -num_threads 4
4 + grep GGDC GGDC.txt
5 + awk '{printf $1$2"\t"%".4f\n", $3*100}'
```

The task name 'OrthoANI' is also highlighted with a white box. The status is 'success' and a 'Download log' button is visible.

**check the command log**

# Data analysis

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

The screenshot shows the gcMeta interface. On the left is a dark sidebar with navigation options: Microbial Data, Analysis Cloud Platform, Dashboard, Metadata, Data, Tools/Workflows, and **Tasks/Reports** (highlighted with a red box). The main area is titled 'Task/report' and includes a filter for dates (2018-08-08 to 2018-08-15) and a search bar. Below is a table with columns: User, Task name, Creation time, End time, and Status. A red box highlights the task name 'Tool\_OrthoANI\_2018\_08\_15\_21\_25\_39' in the first row.

User	Task name	Creation time	End time	Status
○ guest	<b>Tool_OrthoANI_2018_08_15_21_25_39</b>	2018-08-15 21:33:22	2018-08-15 21:34:13	✓

**check status of the job**

The screenshot shows a text file named 'res.txt' with the following content:

```
GGDC1: 98.7345
GGDC2: 15.0929
GGDC3: 98.5095
ANI: 66.5128
(Recommended Use GGDC2. The GGDC is a state-of-the-art in silico method
for genome-to-genome comparison, thus reliably mimicking conventional
DDH, except for its pitfalls.)
References:
1.Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M. Genome
sequence-based species delimitation with confidence intervals and
improved distance functions. BMC Bioinformatics 14:60, 2013.
2.Imchang Lee, Yeong Ouk Kim, Sang-Cheol Park, Jongsik Chun: OrthoANI:
An improved algorithm and software for calculating average nucleotide
identity. International Journal of Systematic and Evolutionary
Microbiology, 2016.
```

The screenshot shows a file management interface with a table. The table has columns: Log, Parameter, Report, **Relevant file download and preview** (highlighted with a red box), and an unlabeled column. The table contains one row with the file name 'res.txt', size '661.00B', and a 'Download' button (highlighted with a red box).

Log	Parameter	Report	<b>Relevant file download and preview</b>		
			File name	Size	操作
			<b>res.txt</b>	661.00B	<b>Download</b>

**click to view  
the result**

**click to download the file**

**click to visit the manual  
page to view the explanation  
of input, output, arguments  
and the example**

The screenshot shows the gcMeta interface with the 'MANUAL' button highlighted with a red box. Other elements include a question mark icon, an alarm clock icon, a refresh icon, and the user name 'imcas(guest)'.

# Manual

# gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

all the manuals are in pdf format including introduction, input&output, example

**PILER-CR (CRISPR element detection)**

Introduction: PILER-CR is public domain software for finding CRISPR repeats.

Input: DNA sequences (in fasta format).  
Output: Direct repeat sequence, CRISPR arrays structure (in text format).

Test inputs:  
input\_seq (fasta/fn/fno)  
imcas:/test\_kira/Lactobacillus.fasta

Test example:

*.txt	7.86KB
*.fasta	120.00B

copy and paste to test it

```
1542443 27 100.0 34 GGATCAATGA ..... CTTTAACTGACCAAAAGCGCTGTACAGCA
1542504 27 100.0 34 GGTACAGCA ..... GGAGCAAAAGGGCGAAAGCTAATTTATGAG
1542565 27 100.0 34 ATTATAGCG ..... CTTGCTAGTCCGCTGTGCTGATGCAATG
1542626 27 96.3 34 ATTGAGTCA ..... CCCAGCCACATATTTGTGCGAATAGCGTAA
1542687 27 100.0 34 ATAGCGTANG ..... CTTGCGCAAAATATGAGCGAACTTGAAATCA
1542748 27 100.0 34 TGGAAATCA ..... CAGAGCTGCAAAATAGCGCTTCTCTAATA
1542809 27 100.0 34 TCCATTAANA ..... CAGAGTTAGTCCGCAATGTTGTCTCAAGA
1542870 27 100.0 34 GTTCAATGA ..... CTAATAGTAGTCCGCAAGCTTGTCCGATTA
1542931 27 100.0 34 TCCGATTTA ..... CATAAATATAGAGGAAACCCCTTTCTCTA
1542992 27 100.0 ..... CACTAAAGGA
-----
31 27 34 GGATCACTCCACATACGTGGAGAAA
```

SUMMARY BY SIMILARITY

Array	Sequence	Position	Length	# Copies	Repeat	Spacer	Consensus
1	NC_066814.3	Lact	1541161	1858	31	27	34 + GGATCACTCCACATACGTGGAGAAA

SUMMARY BY POSITION

Array	Sequence	Position	Length	# Copies	Repeat	Spacer	Distance	Consensus
1	NC_066814.3	Lact	1541161	1858	31	27	34	GGATCACTCCACATACGTGGAGAAA

>NC\_066814.3 Lactobacillus acidophilus MCFM chromosome, complete genome

Array	Sequence	Position	Length	# Copies	Repeat	Spacer	Distance	Consensus
1	NC_066814.3	Lact	1541161	1858	31	27	34	GGATCACTCCACATACGTGGAGAAA

Figure: CRISPR arrays structure

input: data matrix (tab-separated value, in txt format)

Test inputs:  
input\_data (txt)  
imcas:/test\_kira/nm.4272-S2.txt

Test arguments:  
skippedline (first N rows will be skipped as marker, integer, [1] 4)  
topvector (first N rows sorted by average abundance will be calculated, integer, [80] 80)  
datamulti (data will be multiple by N, integer, [11] 1)  
markercol (color marker, see below, integer, [0] 3)  
markershape (shape marker, see below, integer, [0] 2)  
markeralpha (integer, [0] 0)  
markersize (integer, [0] 0)  
markerlabel (integer, [0] 0)  
vegmethod (string=mahalanobis, cao, chao, binomial, raup, mountford, horn, morisita, altGower, gower, jaccard, kulczynski, canberra, manhattan, euclidean, Bray, ... , [bray])

Test example:

Site	WSP_001	WSP_002	WSP_003	WSP_004	WSP_005	WSP_006	WSP_007	WSP_008	WSP_009	WSP_010	WSP_011	WSP_012	WSP_013	WSP_014	WSP_015	WSP_016	WSP_017	WSP_018	WSP_019	WSP_020
INTRACUTANEOUS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STOOL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
POSTNATALINTESTITIS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ORAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SKIN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STOOL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
POSTNATAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure: input data

sample meta data  
(argument skippedline = 4)

markershape (different shapes)  
markercol (different colors)

Site	WSP_001	WSP_002	WSP_003	WSP_004	WSP_005	WSP_006	WSP_007	WSP_008	WSP_009	WSP_010	WSP_011	WSP_012	WSP_013	WSP_014	WSP_015	WSP_016	WSP_017	WSP_018	WSP_019	WSP_020
INTRACUTANEOUS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STOOL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
POSTNATALINTESTITIS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ORAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SKIN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STOOL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
POSTNATAL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure: arguments explanation