



CGAS (Chloroplast Genome Analysis Suite): An Automated Python Pipeline for Comprehensive Comparative Chloroplast Genomics

Abdullah^{1,2*}, Rushan Yan^{1,2}, Xiaoxuan Tian^{1,2*}

¹State Key Laboratory of Chinese Medicine Modernization,
Tianjin University of Traditional Chinese Medicine

²Haihe Laboratory of Modern Chinese Medicine



Abdullah, Rushan Yan, Xiaoxuan Tian. 2026. CGAS (Chloroplast Genome Analysis Suite):
An Automated Python Pipeline for Comprehensive Comparative Chloroplast Genomics.
iMetaOmics 3: e70093. <https://doi.org/10.1002/imo2.70093>



The Challenge

Fragmented chloroplast genome analysis workflow



Multiple Tools

Fragmented workflow
requiring many
independent tools



Manual Formatting

Complex data
formatting and
conversion steps



Complex Pipelines

Significant
bioinformatics
expertise required

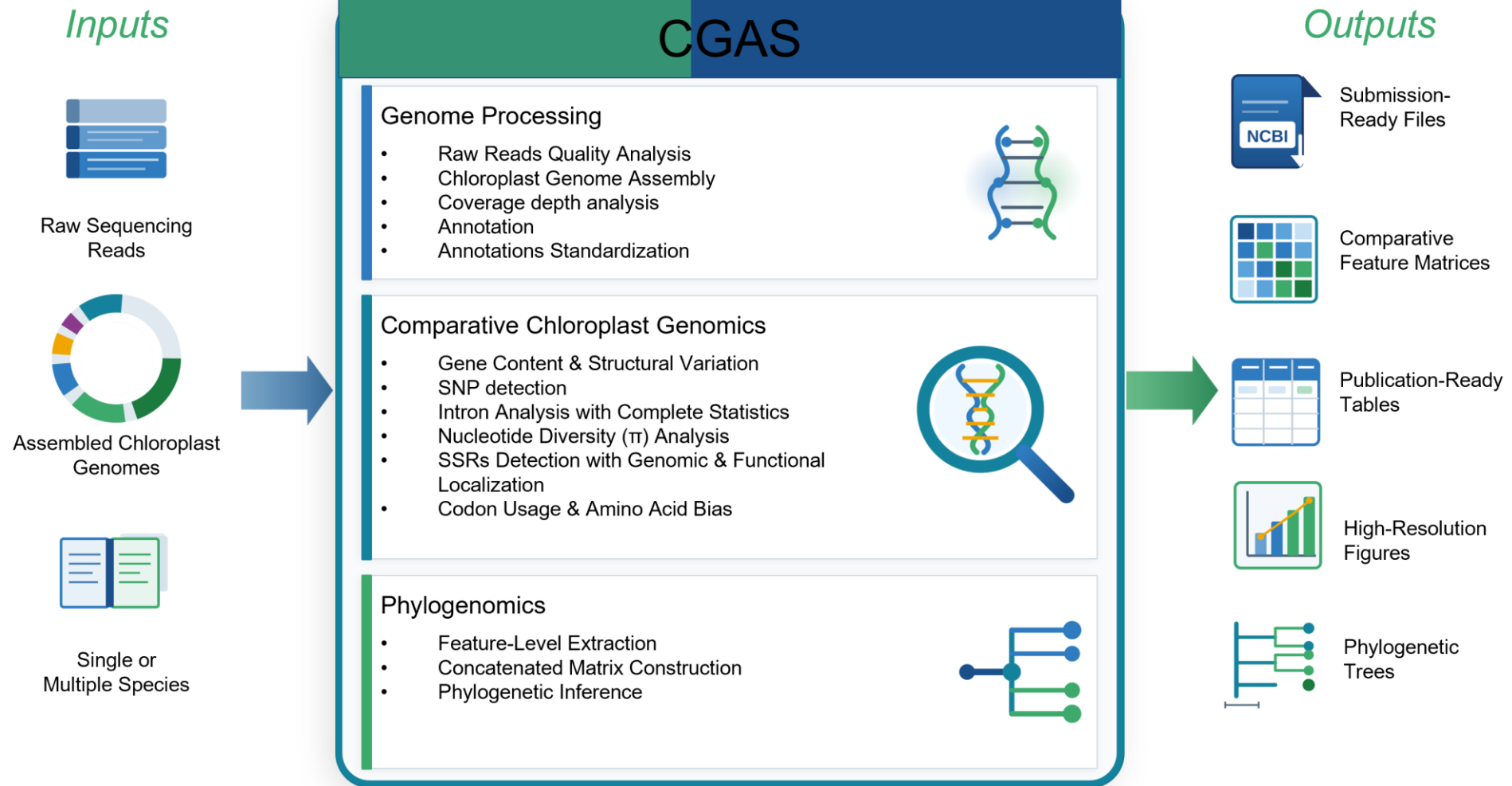
Time-consuming • Difficult to reproduce • High technical barrier



Highlights

Chloroplast Genome Analysis Suite (CGAS)

Integrated Platform for Comparative Chloroplast Genomics





CGAS Architecture

Three major analytical phases

1

Phase 1: Preparation

Genome assembly, Quality control,
Annotations and NCBI submission

2

Phase 2: Comparative Analysis

Comprehensive
chloroplast genomics

3

Phase 3: Phylogenetic Analysis

Evolutionary
relationships



Phase 1: Preparation Modules

Chloroplast Genome Analysis Suite (CGAS)

Unified Pipeline for Comparative Chloroplast Genomics

Input Data

Modules 1-4 (Preparation)

- FASTQ files (raw sequencing reads)
- Paired-end/Single-end

Modules 5-14 (Analysis & Phylogeny)

- Annotated GenBank files (.gb, .gbk)
- Multiple genomes (batch processing)

Phase 1: Preparation Modules (Run Individually)

1

Assembly & QC

- Quality Control (fastp)
- Adapter Trimming
- Genome Assembly (GetOrganelle)
- Coverage Analysis

Output: Assembled genomes

2

Annotation

- Plastome Gene Annotation
- PGA Pipeline
- Reference-guided Annotation

Output: GenBank files

3

Gene Comparison

- Normalize Gene Names
 - Standardize Annotations
- Cross-genome Consistency

Output: Normalized GenBank

4

Format Conversion

- Convert to NCBI Format
- Generate FASTA + TBL files
- Submission-ready

Output: NCBI submission files

Output: Assembled genomes → GenBank files → Normalized annotations → Annotation validation and NCBI submission files



Phase 2: Comparative Analysis Modules

Phase 2: Main Analysis Modules

(Run as Pipeline: `cgas 5,6,7,8,9,10,11,12,13`)

5 Gene Comparative Analysis

Compare gene content across species
Identify conserved and unique genes
Gene presence/absence matrix

Output: Comparative tables

6 Gene Content Tables

Publication-ready Word documents
Formatted gene tables
Journal-ready format

Output: .docx tables

7 Genome Structure

LSC/SSC/IR region analysis
Functional gene classification
GC content distribution

Output: Structure data

8 Codon Usage (RSCU)

Relative synonymous codon usage
Codon bias patterns
R-based visualization

Output: RSCU data + plots

9 Amino Acid Analysis

Amino acid composition
Patterns across genomes
R-based visualization

Output: AA data + plots

10 SNP Analysis

Single nucleotide polymorphisms
Substitution detection
Quality visualization

Output: SNP data + plots

11 Intron Analysis

Gene and tRNA intron detection
Intron-exon structure
Comprehensive comparison

Output: Intron data

12 SSR Analysis

Simple sequence repeat detection
Genomic/functional region mapping
High-quality visualization

Output: SSR data + plots

13 Nucleotide Diversity

Pi (π) diversity assessment
Region-wise analysis
High-quality visualization

Output: Diversity plots

Pipeline Command: `cgas --modules 5,6,7,8,9,10,11,12,13`
All 9 modules run together for comprehensive comparative analysis

Note: Modules 8, 9, 10, 12, 13 include R-based visualizations (optional, requires R installation)



Phase 3: Phylogenetic Analysis

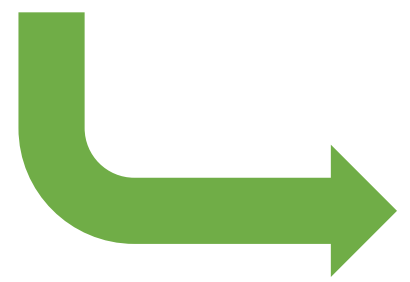
14

Module 14: Phylogenetic Analysis (Run Separately)

Phylogenetic matrix construction • MACSE/MAFFT alignment • IQ-TREE inference
Outgroup support • Complete genome or gene-specific phylogenies

Output: Phylogenetic trees (.treefile) + alignment matrices

Preparation Analysis Phylogeny



Produce publication-ready results



Structured Tables

Gene content
Comparative analysis
Statistical summaries



Visualizations

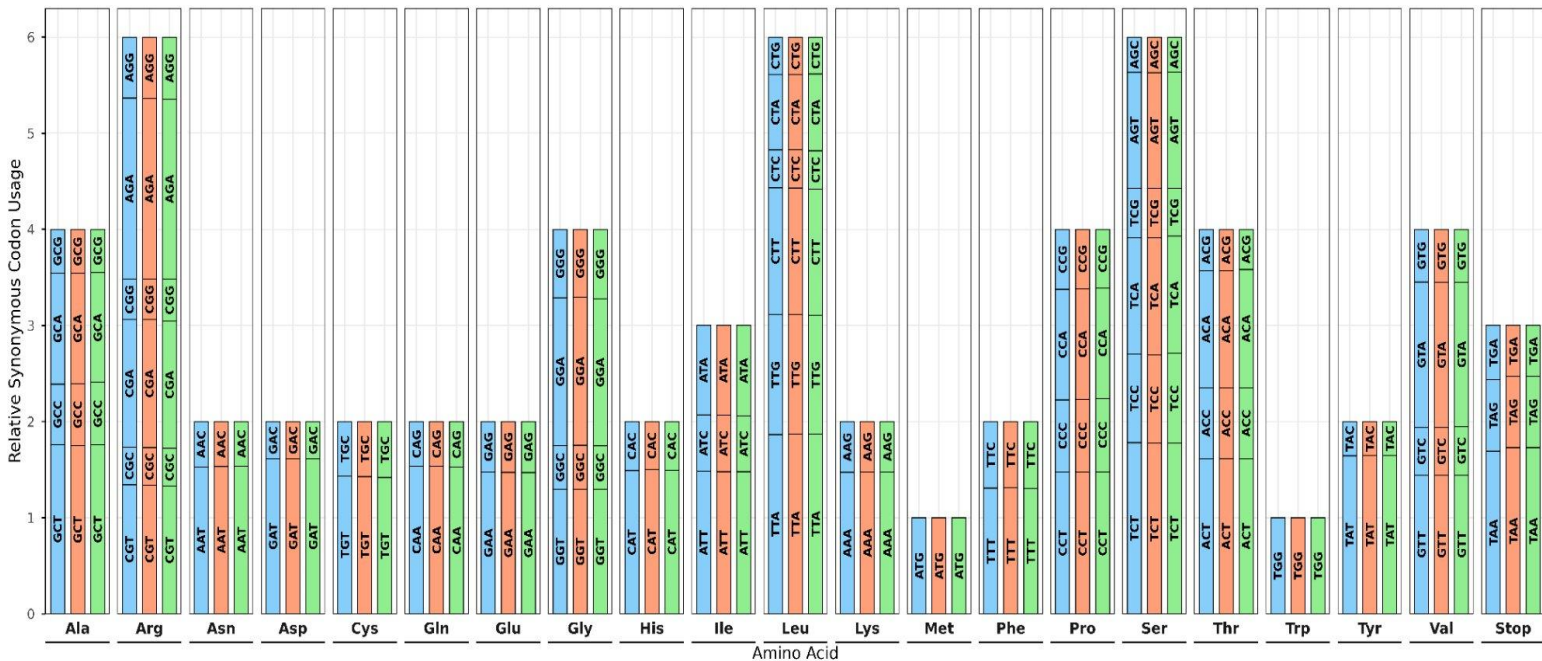
SNP plots
SSR distributions
Diversity analyses



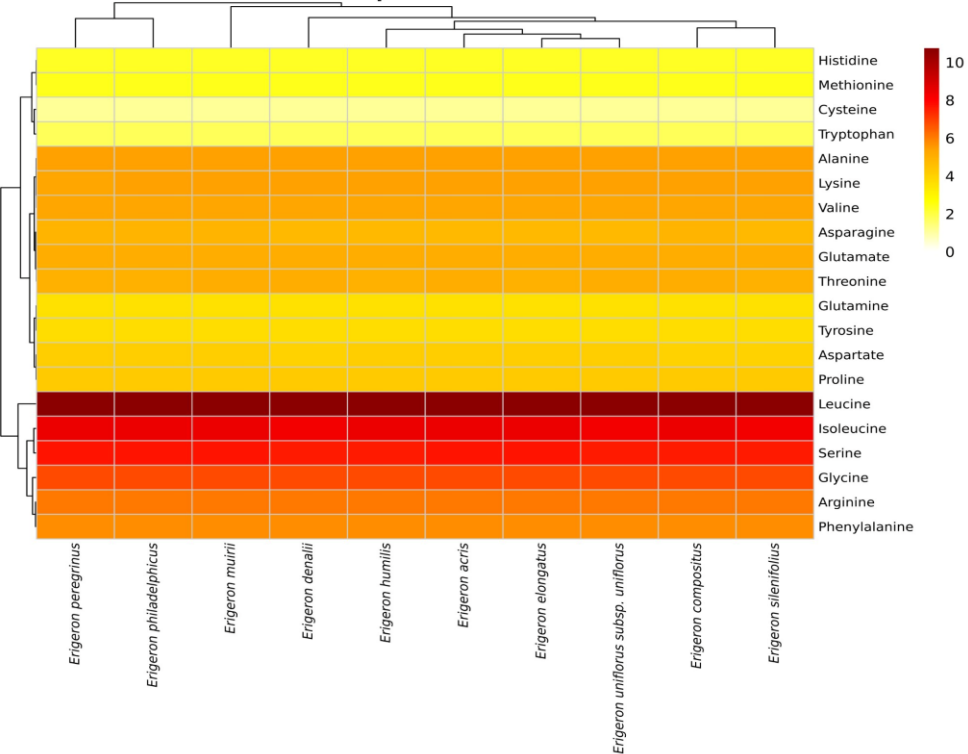
Partial results figures

Codon Usage Analysis

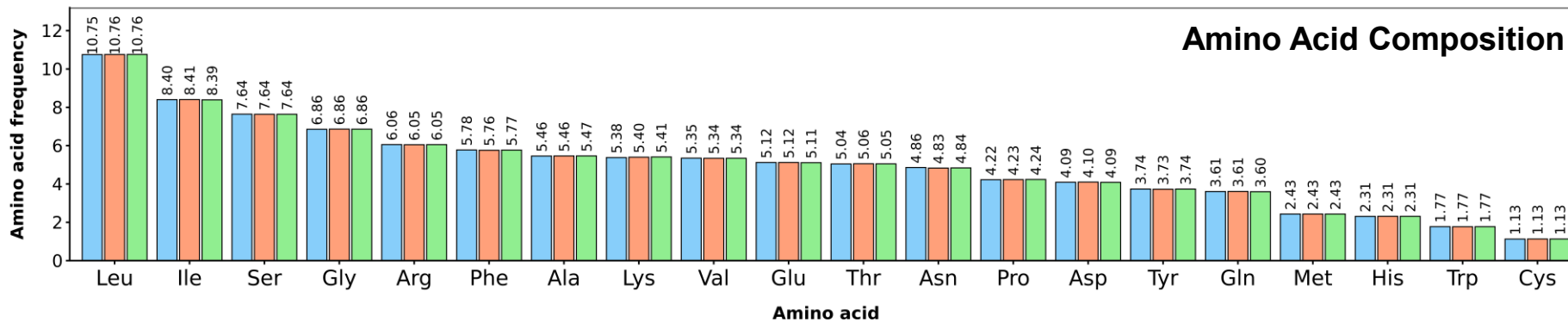
Erigeron compositus Erigeron silenifolius Erigeron peregrinus



Amino Acid Composition Across Genomes



Erigeron peregrinus Erigeron humilis Erigeron silenifolius



Amino Acid Composition



Summary

Simplifies Chloroplast Comparative Genomics

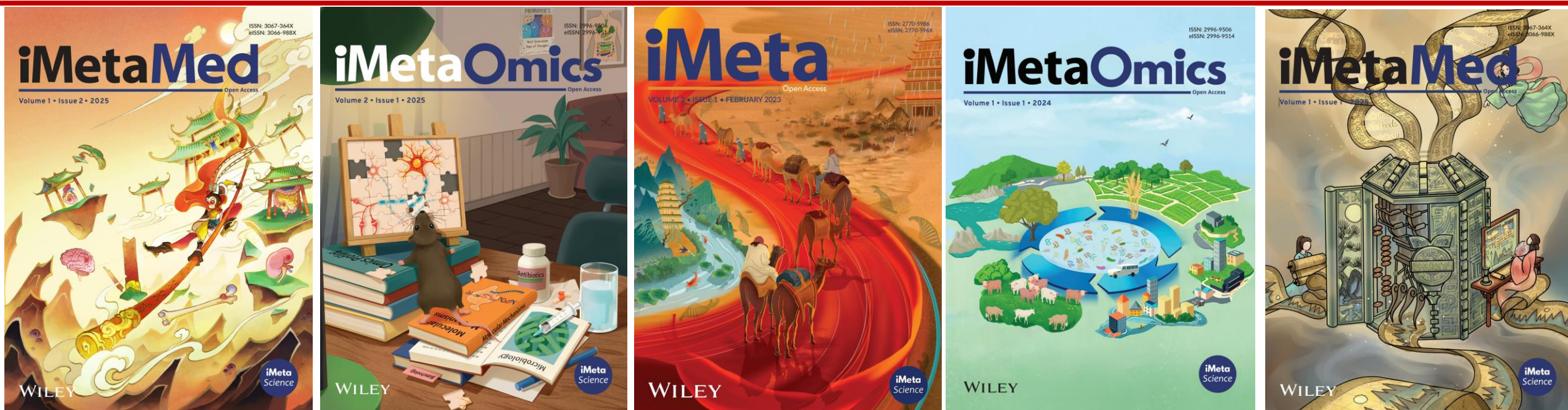
Accelerating Chloroplast Genomics Research

- ❑ Integrated 14-module pipeline
- ❑ Automated analysis workflow
- ❑ GitHub: github.com/abdullah30/Chloroplast-Genome-Analysis-Suite-CGAS
- ❑ Publication-ready outputs
- ❑ Reproducible research

Abdullah, Rushan Yan, Xiaoxuan Tian. 2026. CGAS (Chloroplast Genome Analysis Suite): An Automated Python Pipeline for Comprehensive Comparative Chloroplast Genomics. *iMetaOmics* 3: e70093. <https://doi.org/10.1002/imo2.70093>

iMeta: To be top journals in biology and medicine

WILEY



“**iMeta**” launched in 2022 by iMeta Science Society, **impact factor (IF) 33.2**, ranking **top 65/22249 in world and 2/161 in the microbiology**. It aims to publish innovative and high-quality papers with broad and diverse audiences. **Its scope is similar to Cell, Nature Biotechnology/Methods/Microbiology/Medicine/Food**. Its unique features include video abstract, bilingual publication, and social media with 600,000 followers. Indexed by **SCIE/ESI, PubMed, Google Scholar** etc.

“**iMetaOmics**” launched in 2024, with a **target IF>10**, and its scope is similar to **Nature Communications, Cell Reports, Microbiome, ISME J, Nucleic Acids Research, Briefings in Bioinformatics**, etc.

“**iMetaMed**” launched in 2025, with a **target IF>15**, similar to **Med, Cell Reports Medicine, eBioMedicine, eClinicalMedicine** etc.



Society: <http://www.imeta.science>

Publisher: <https://wileyonlinelibrary.com/journal/imeta>

iMeta: <https://wiley.atyponrex.com/journal/IMT2>

Submission: iMetaOmics: <https://wiley.atyponrex.com/journal/IMO2>

iMetaMed: <https://wiley.atyponrex.com/journal/IMM3>



[iMetaScience](#)



[iMetaScience](#)



office@imeta.science
imetaomics@imeta.science



[Promotion Video](#)

Update
2025/7/6