Evidence tiers for strain-resolved long-read metagenomics

Yanhua Han[1#], Yunjuan Peng[2#], Minghui Li[1#], Xiaoyuan Wei[3], Jianmin Chai[1], Jiangchao Zhao[2], Ying Li[1*], Feilong Deng[1*]

1 School of Animal Science and Technology, Foshan University, Foshan, 528225, China
2 College of Animal Science, South China Agricultural University, Guangzhou, 510642, China
3 Morsani College of Medicine, University of South Florida, Tampa, 33612, USA

Yanhua Han, Yunjuan Peng, Minghui Li, Xiaoyuan Wei, Jianmin Chai, Jiangchao Zhao, Ying Li, et al. 2026. Evidence tiers for strain-resolved long-read metagenomics. ***iMetaOmics*** 3: e70130. https://doi.org/10.1002/imo2.70130

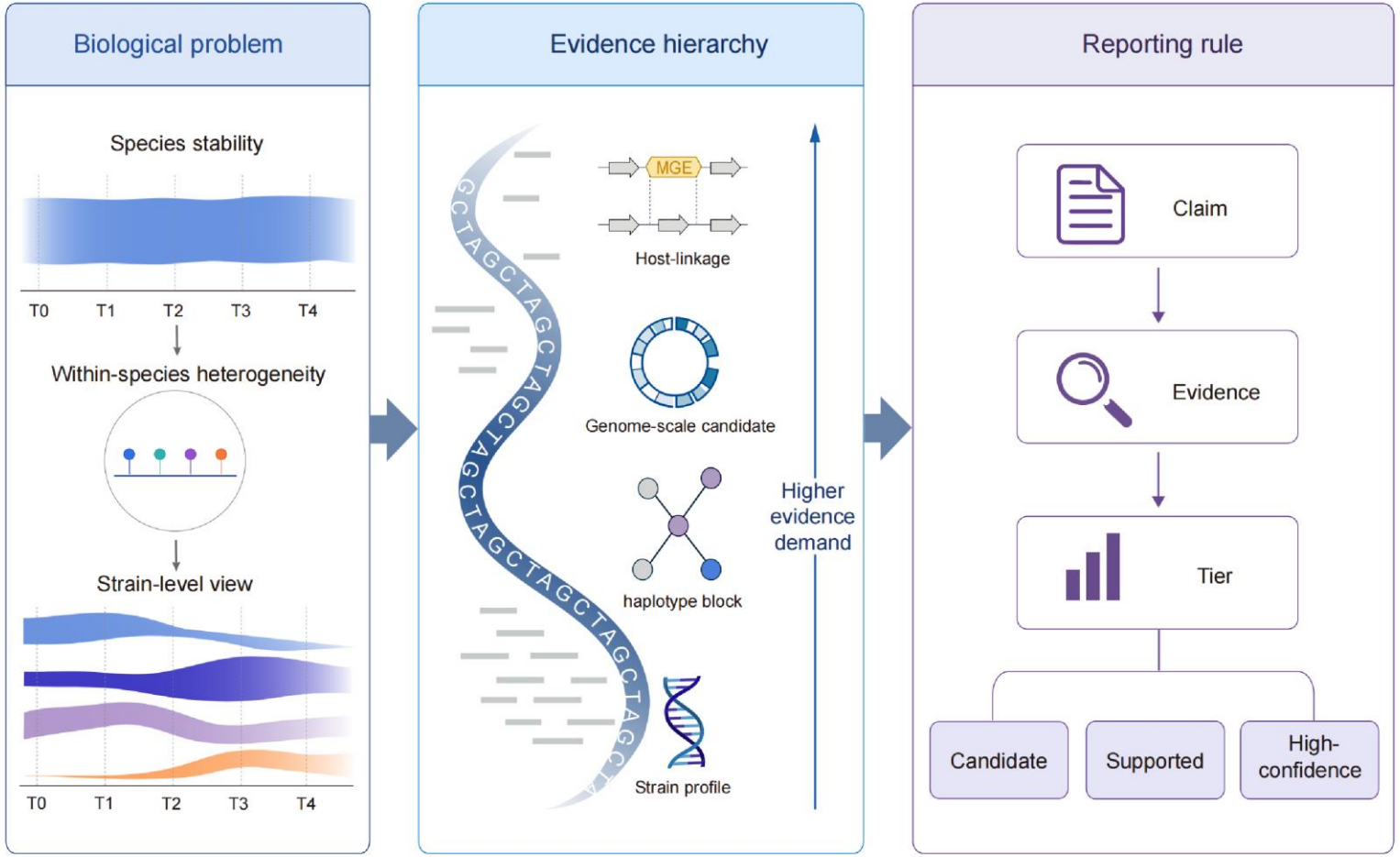
Introduction

01 Metagenomic analysis is moving from species-level abundance toward strain composition and dynamics.

02 A stable species profile may still hide strain replacement, persistence, or colonization.

03 Long reads can link variants, repeats, and mobile genetic elements, but local or incomplete linkage may still be overinterpreted.

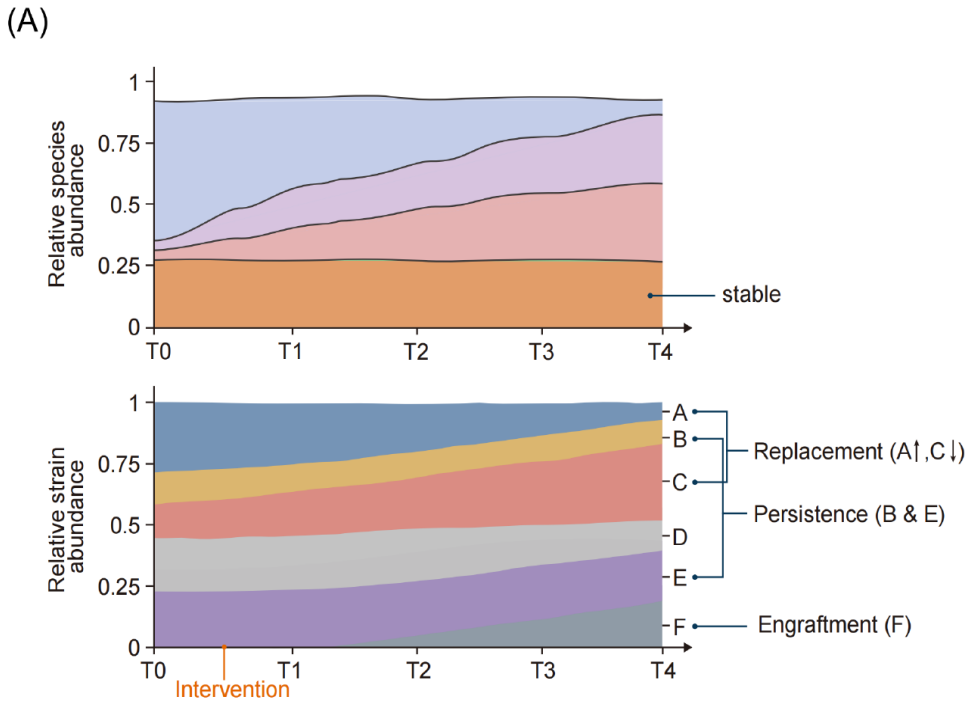
Introduction

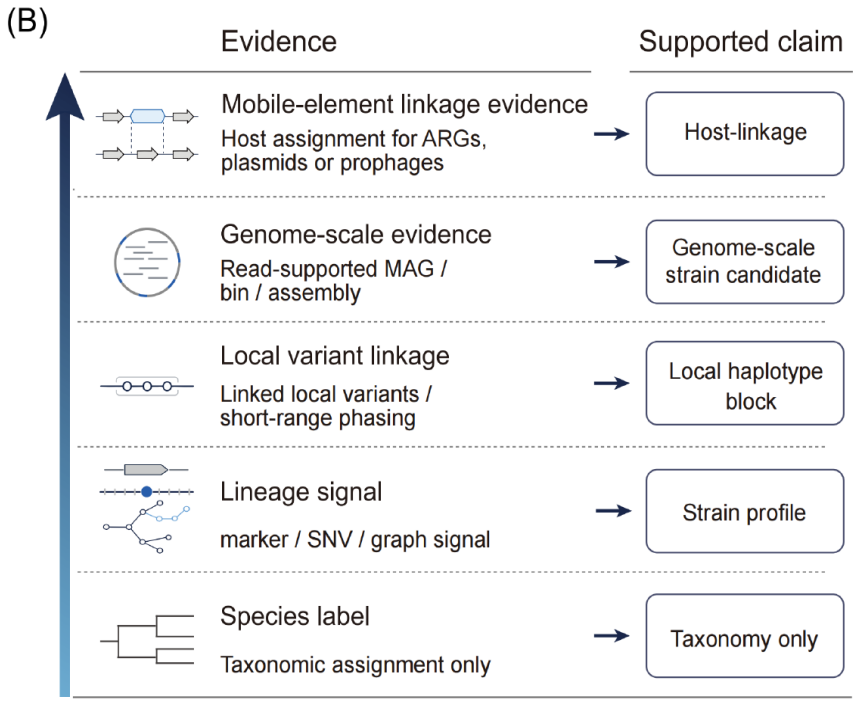


- Different strain-level outputs differ in genomic span and evidence requirements.
- Evidence type and confidence level should be reported separately, rather than treated as the same hierarchy.
- When ambiguous assignment, contamination, consensus collapse, coverage breaks, reference bias, or unresolved linkage is present, claim strength should be downgraded.
- This makes statements about strain reconstruction, host assignment, persistence, replacement, and colonization more accurate, comparable, and reviewable.

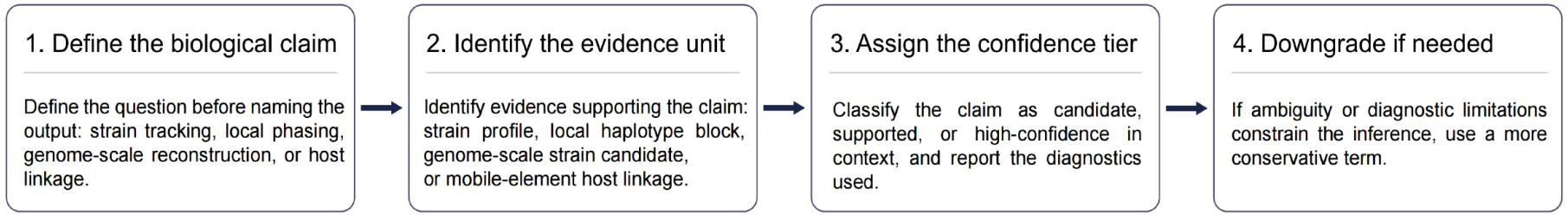
Main Results

◆ Evidence units and reporting workflow





(C)



(A) Species stability ≠ strain stability

- Species abundance may remain similar, while within-species lineages shift.
- Strain-level analysis can identify replacement, persistence, and colonization.

(B) Different claims require different evidence

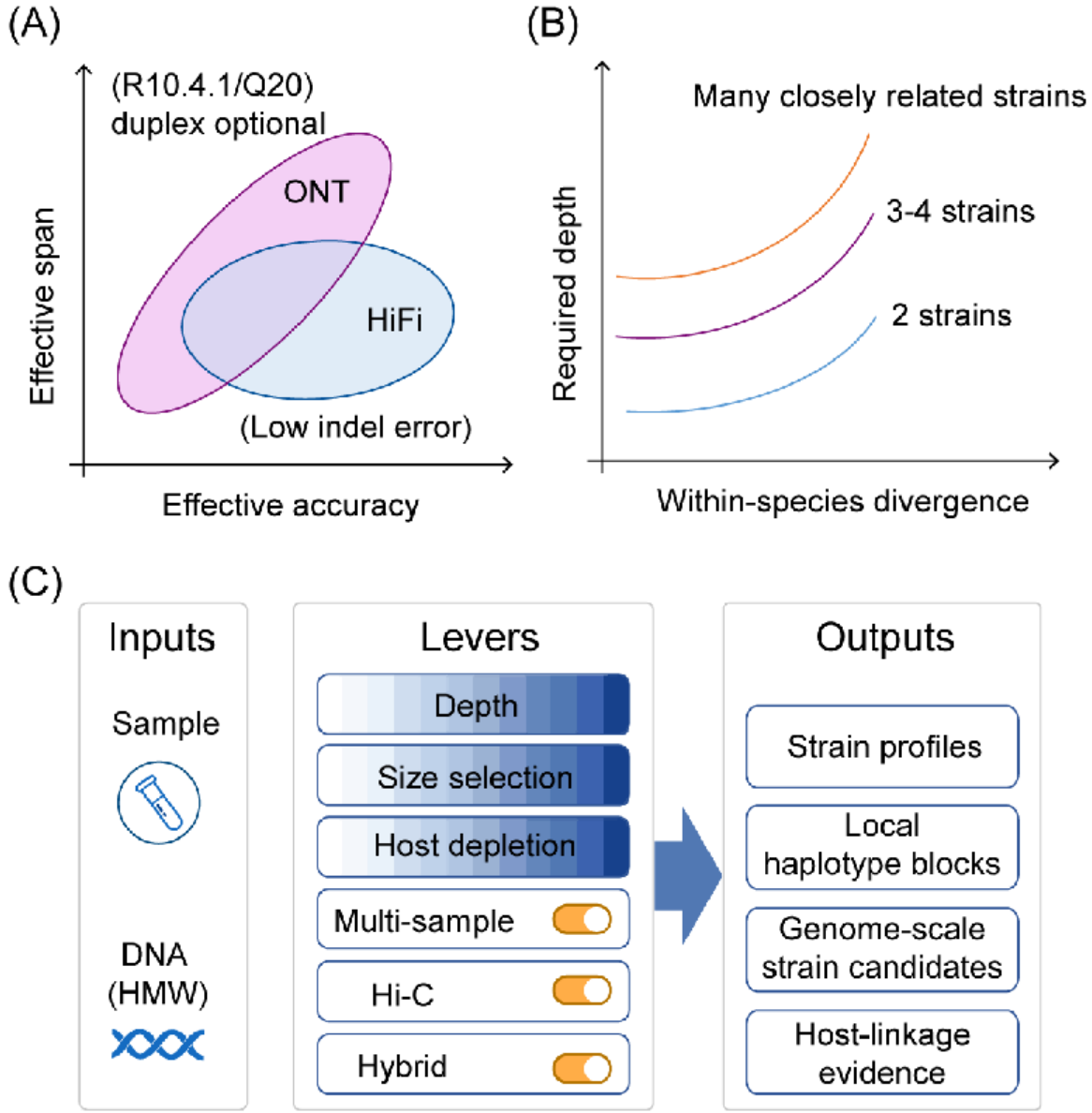
- Strain profiles, local haplotypes, genome-scale strain candidates, and host-linkage claims need different evidence.

(C) Evidence-guided reporting

- Define the question, identify the evidence unit, assign confidence, and downgrade claims when evidence is limited.

Main Results

Design space for strain-resolved long-read metagenomic claims



(A) Accuracy + span define information capacity

- HiFi supports allele-level inference and genome recovery; ONT supports long-range linkage. They are overlapping, not fixed-ranked, regimes.

(B) Depth should be lineage-level

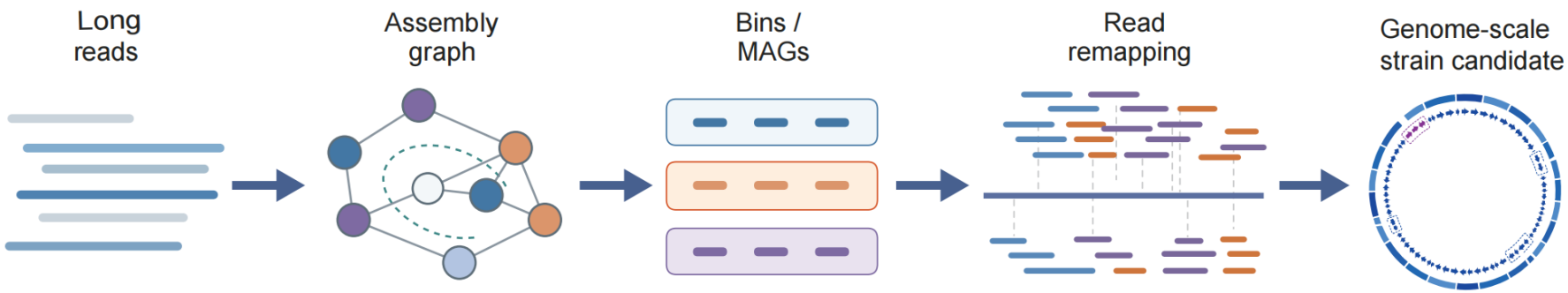
- More closely related coexisting strains require more usable per-lineage depth. High total depth may still be insufficient for minor lineages.

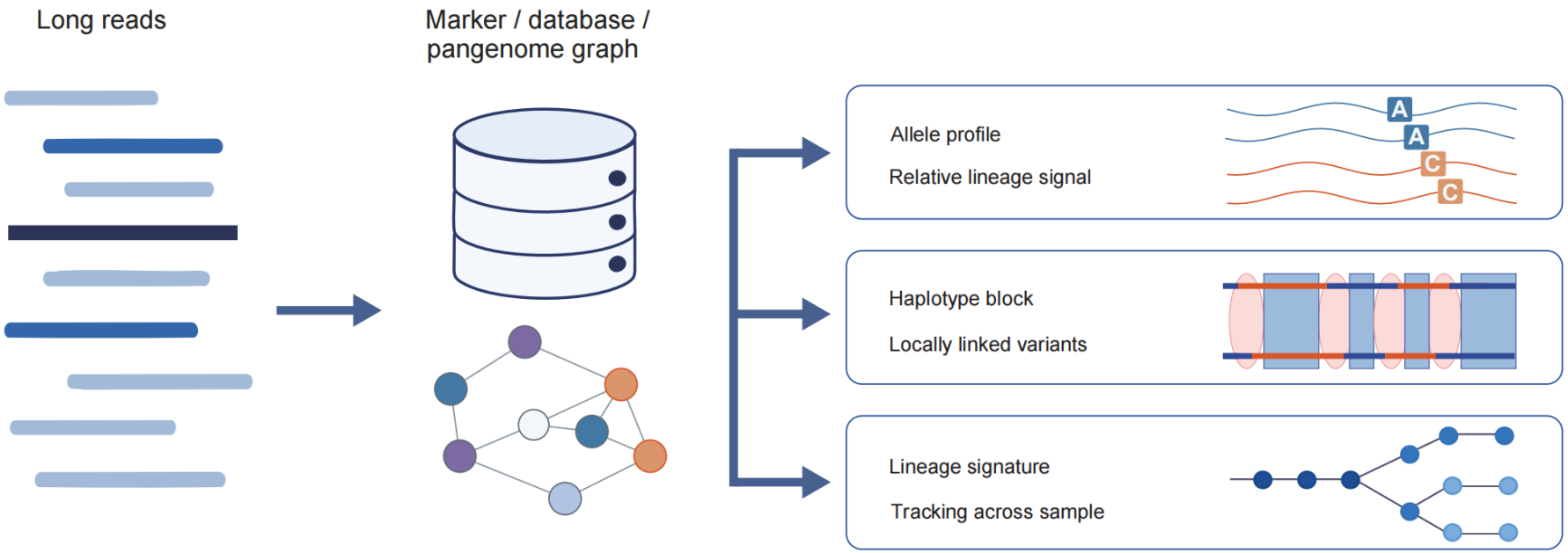
(C) Design determines reportable tiers

- HMW DNA, depth, size selection, host depletion, multi-sample design, Hi-C, and hybrid sequencing can strengthen usable evidence.

Main Results

Two analysis workflows





(A) Assembly-first: genome-scale strain candidates

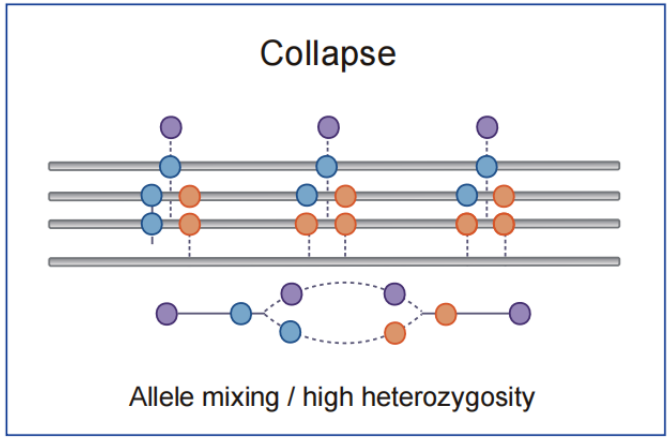
- Long reads are used for assembly graphs, binning, read mapping, and consistency checks.
- Reliability depends on graph ambiguity, bin contamination, allele consistency, coverage continuity, and boundary support.

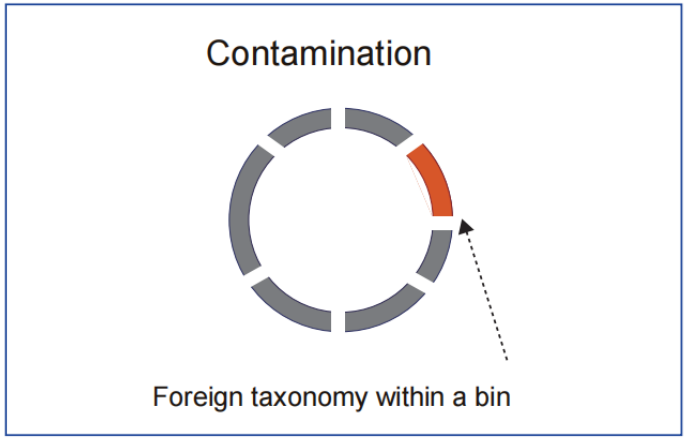
(B) Read-, graph-, and database-first: strain feature signals

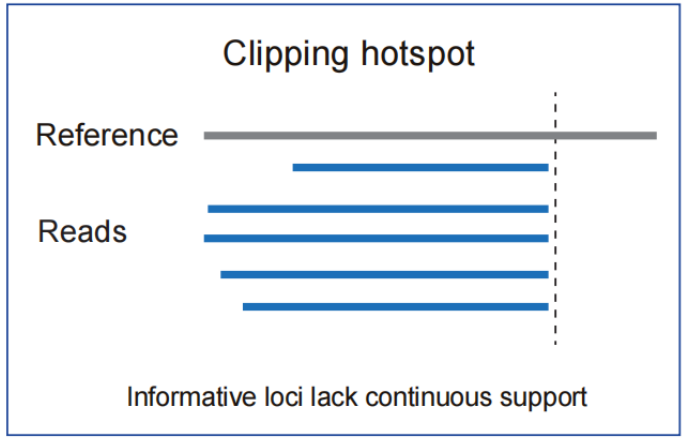
- Markers, databases, or pangenome graphs provide strain profiles, local haplotypes, and lineage signals.
- Without genome-scale linkage, these should not be called whole-genome strain reconstruction.

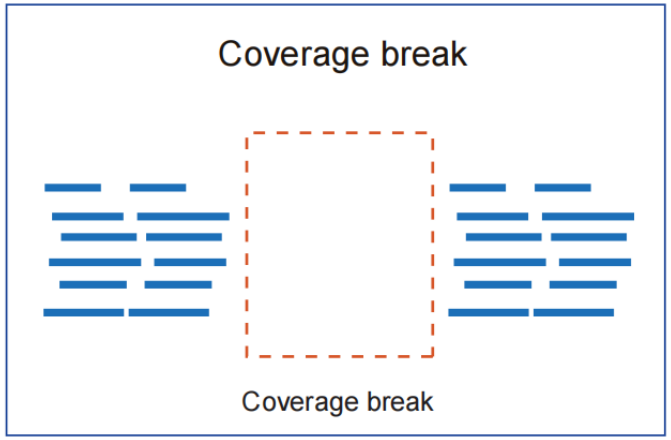
Main Results

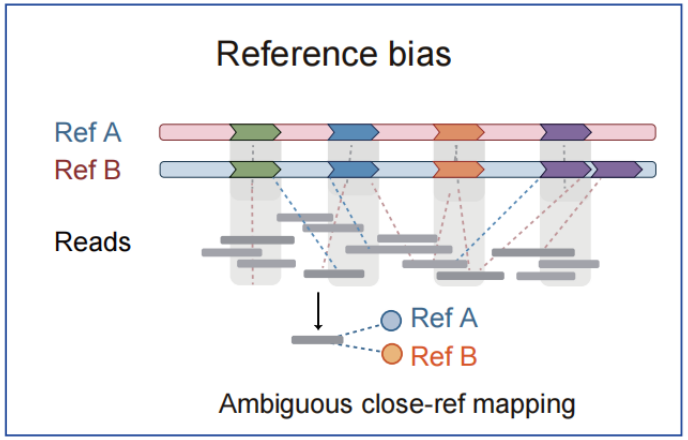
◆ Triggers for claim downgrading

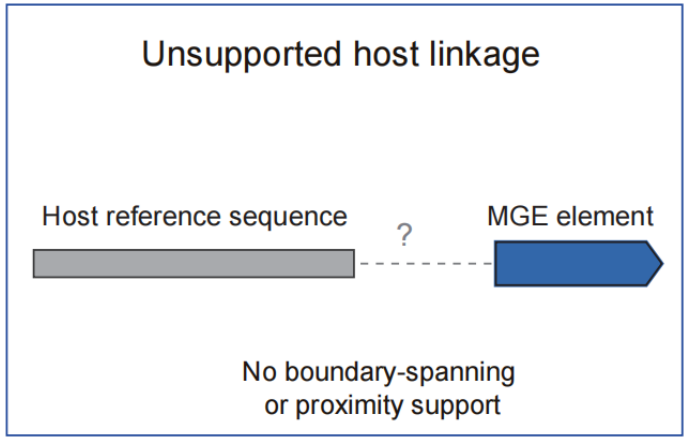












- Consensus collapse, contamination, clipping hotspots, coverage breaks, and reference bias can make strain reconstruction unreliable.
- Without boundary-spanning reads or independent linkage evidence, MGEs should not be assigned to a definite host.
- Use conservative terms such as strain candidate, local linkage, or putative host association.

Summary

Key Findings:

Long-read metagenomics expands the observable range of strain-level features, but each result should be named according to the scale actually supported by evidence, with confidence reported separately.

Implications:

This framework provides a reporting path for strain reconstruction, host assignment, persistence, replacement, and colonization, reducing overstatement and improving reliability, comparability, and reviewability.

Yanhua Han, Yunjuan Peng, Minghui Li, Xiaoyuan Wei, Jianmin Chai, Jiangchao Zhao, Ying Li, et al. 2026. Evidence tiers for strain-resolved long-read metagenomics. ***iMetaOmics*** 3: e70130. https://doi.org/10.1002/imo2.70130

iMeta: Biology and Medicine in Info and Meta Era

WILEY

"***iMeta***" launched in 2022 by iMeta Science Society, impact factor (IF) **44.4**, ranking top 47/22643 in the world. It aims to publish innovative and high-quality papers with broad and diverse audiences. Its scope is similar to *Cell*, *Nature* and *Science*. Its unique features include video abstract, bilingual publication, and social media with 800,000 followers. Indexed by **SCIE/ESI**, **PubMed**, **Google Scholar** etc.

"***iMetaOmics***" launched in 2024, indexed by **ESCI**, **PubMed**, **Google Scholar**, with a **target IF>15, and its scope is similar to** *Nature Communications*, *Science Advances*, *Advanced Science*, *Nucleic Acids Research*, 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
Submission: iMeta: https://wiley.atyponrex.com/journal/IMT2
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
2026/6/17