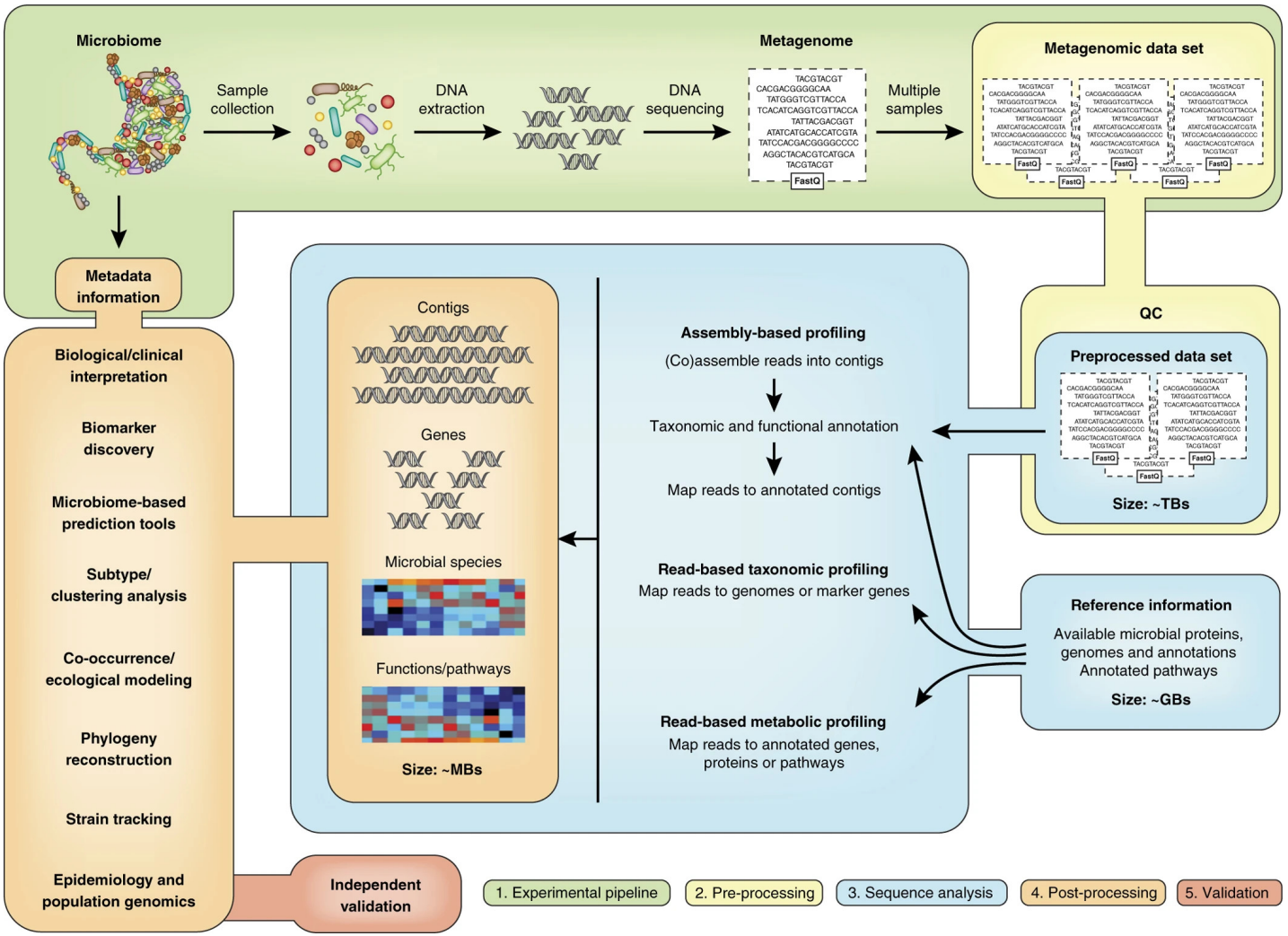
Shotgun metagenomics sequencing

Shotgun metagenomics sequencing

- Untargeted ('shotgun') sequencing of all ('meta-') microbial genomes 'genomics' present in a sample
- Could profile taxonomic composition and functional potential, and to recover whole genome sequences
- Other advanced use of metagenomics: Quantify bacterial growth rate.

Shotgun metagenomics sequencing



Quince et al., *Nature Biotechnology*, 2017

Shotgun metagenomics sequencing

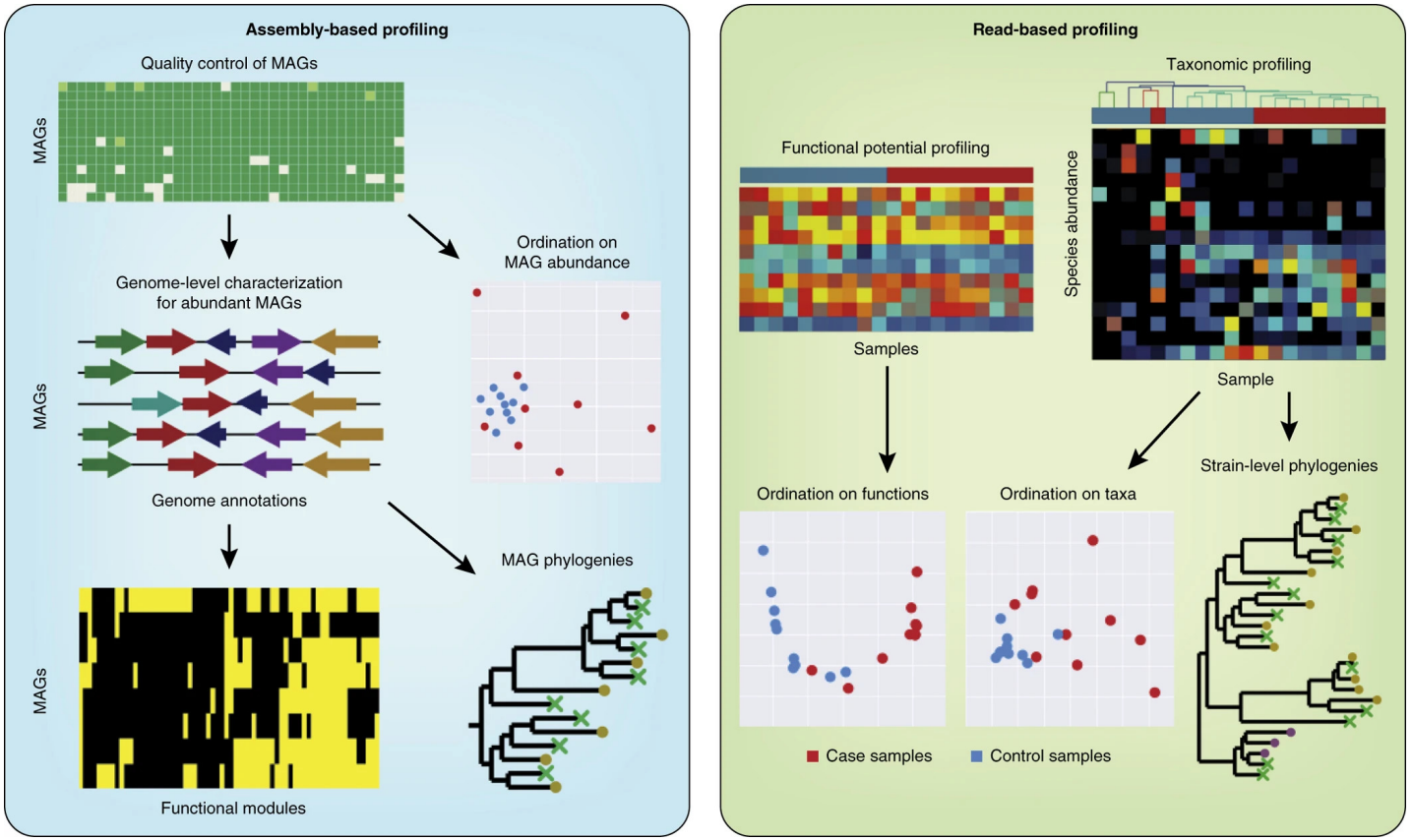


Figure: Integrative analysis leveraging information from multiple studies

Quince et al., *Nature Biotechnology*, 2017

Infer bacterial growth rate from shotgun metagenomics sequencing

MICROBIOME

Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples

Tal Korem,[1,2]* David Zeevi,[1,2]* Jotham Suez,[3]* Adina Weinberger,[1,2]* Tali Avnit-Sagi,[1,2] Maya Pompan-Lotan,[1,2] Elad Matot,[1,2] Ghil Jona,[4] Alon Harmelin,[5] Nadav Cohen,[1,2] Alexandra Sirota-Madi,[6] Christoph A. Thaiss,[3] Meirav Pevsner-Fischer,[3] Rotem Sorek,[7] Ramnik J. Xavier,[6] Eran Elinav,[3]† Eran Segal[1,2]†

PTRs extracted from these samples varied across individuals, in the range of 1 to 2.4, resembling the 1 to 2.6 range of ratios measured in vitro (Fig. 1B). Ratios higher than 2 are indicative of multifork replication, previously documented for *E. coli* (*18*, *22*).

To examine whether PTRs provide a quantitative measure of growth rate, we calculated the temporal growth rate of *E. coli* at different times during its growth experiment as the derivative of its abundance across time (*23*). PTRs were correlated with the measured growth rate, preceding it by 30 min ($R = 0.95$, $P < 10^{-4}$) (Fig. 1D), indicating that PTR predicts the change in abundance. To determine whether PTRs accurately

nature methods

BRIEF COMMUNICATION

https://doi.org/10.1038/s41592-018-0182-0

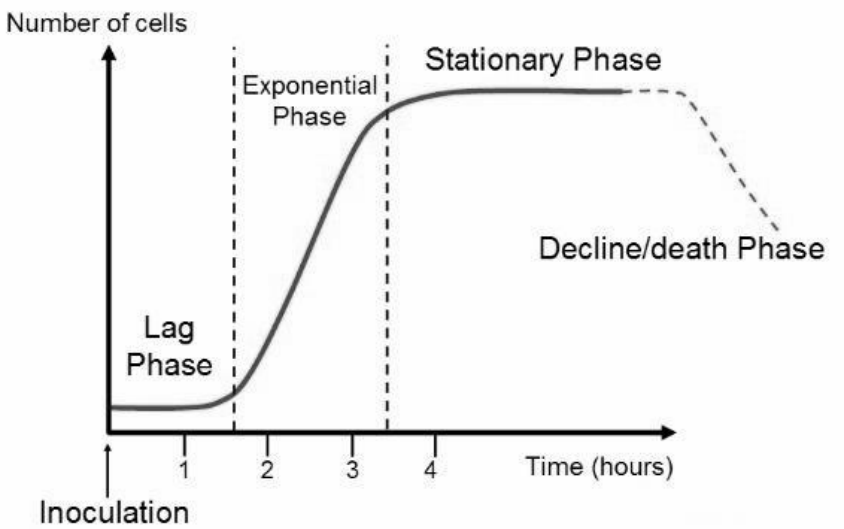
Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples

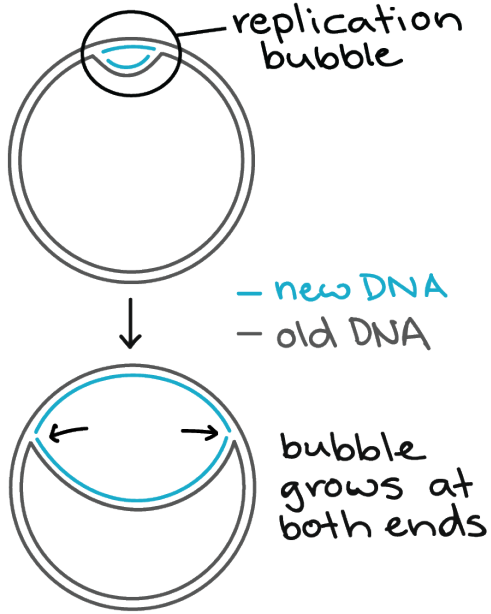
Yuan Gao and Hongzhe Li*

CSH PRESS GENOME RESEARCH

Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth

Bacteria growth



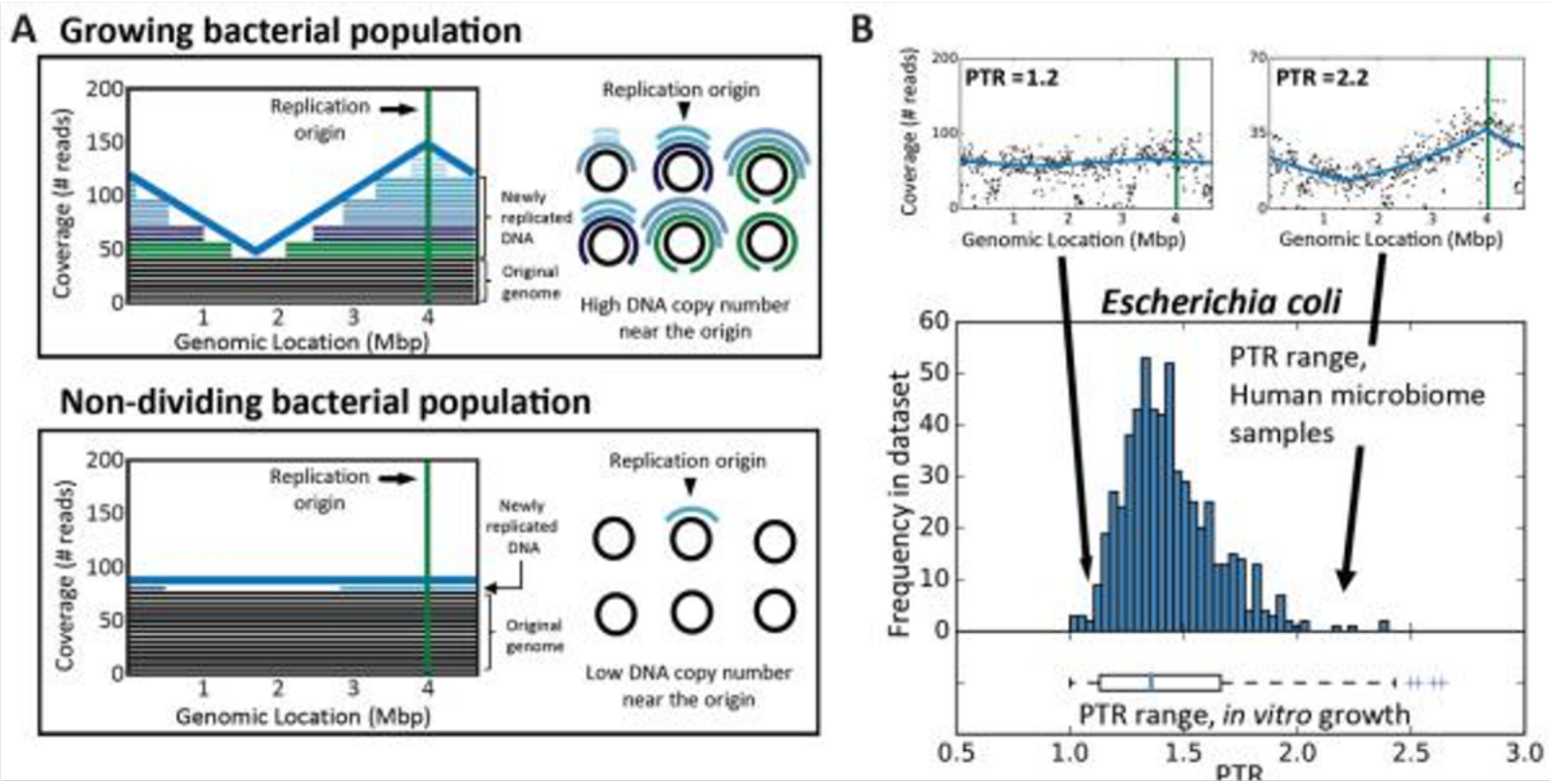


https://orbitbiotech.com/bacterial-growth-curve-generation-time-lag-phase-log-phase-exponential-phase-decline-phase/

FAVPNG

How do we use metagenomics sequencing to infer bacterial growth rate.

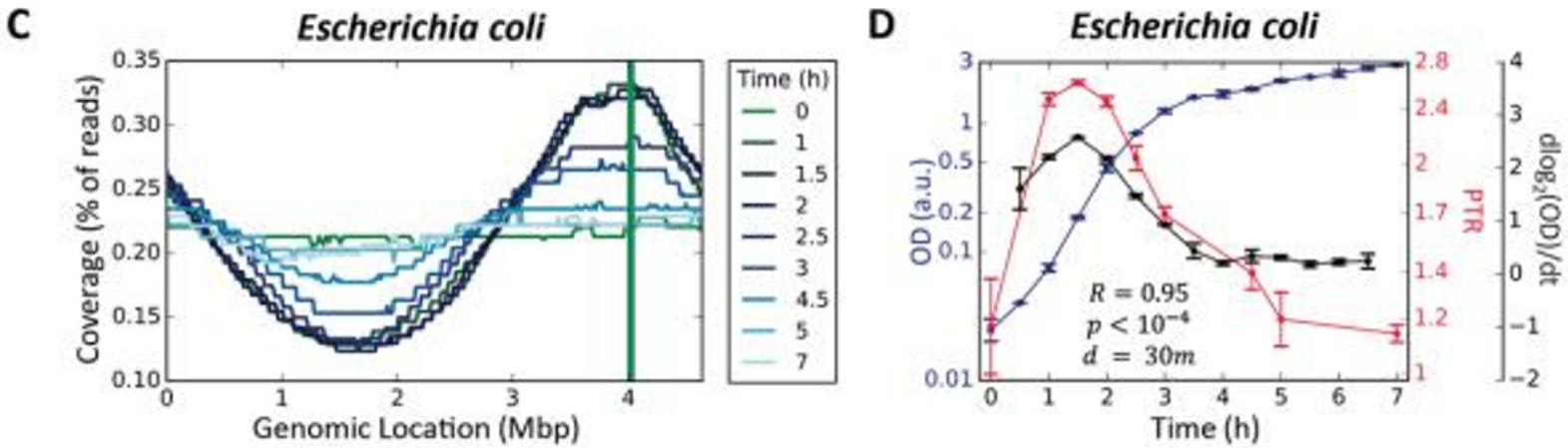
Peak-to-trough coverage ratio (PTR)



PTR accurately measures in vitro growth rates of *Escherichia coli*

Korem et al., *Science*, 2015

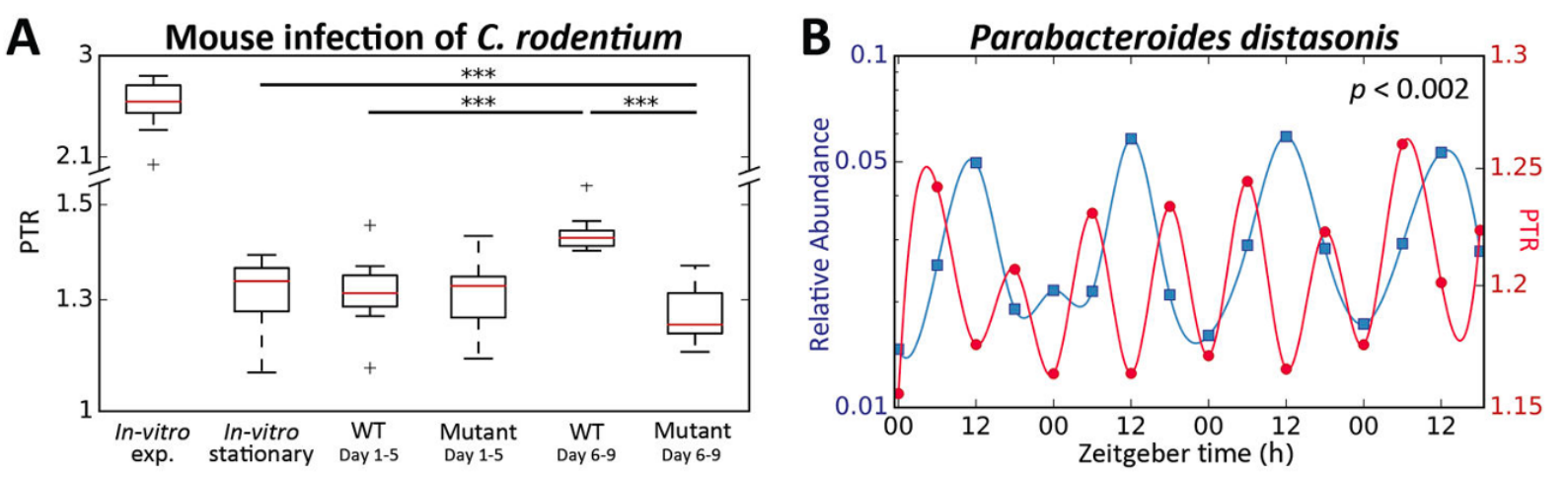
PTRs provide a quantitative measure of growth rate



$$PTR = 2^{C/G}; 1/\log 2(PTR) \propto G/C$$

- C: replication time
- G: generation time

PTR reflects the growth dynamics of *in vivo* microbial communities



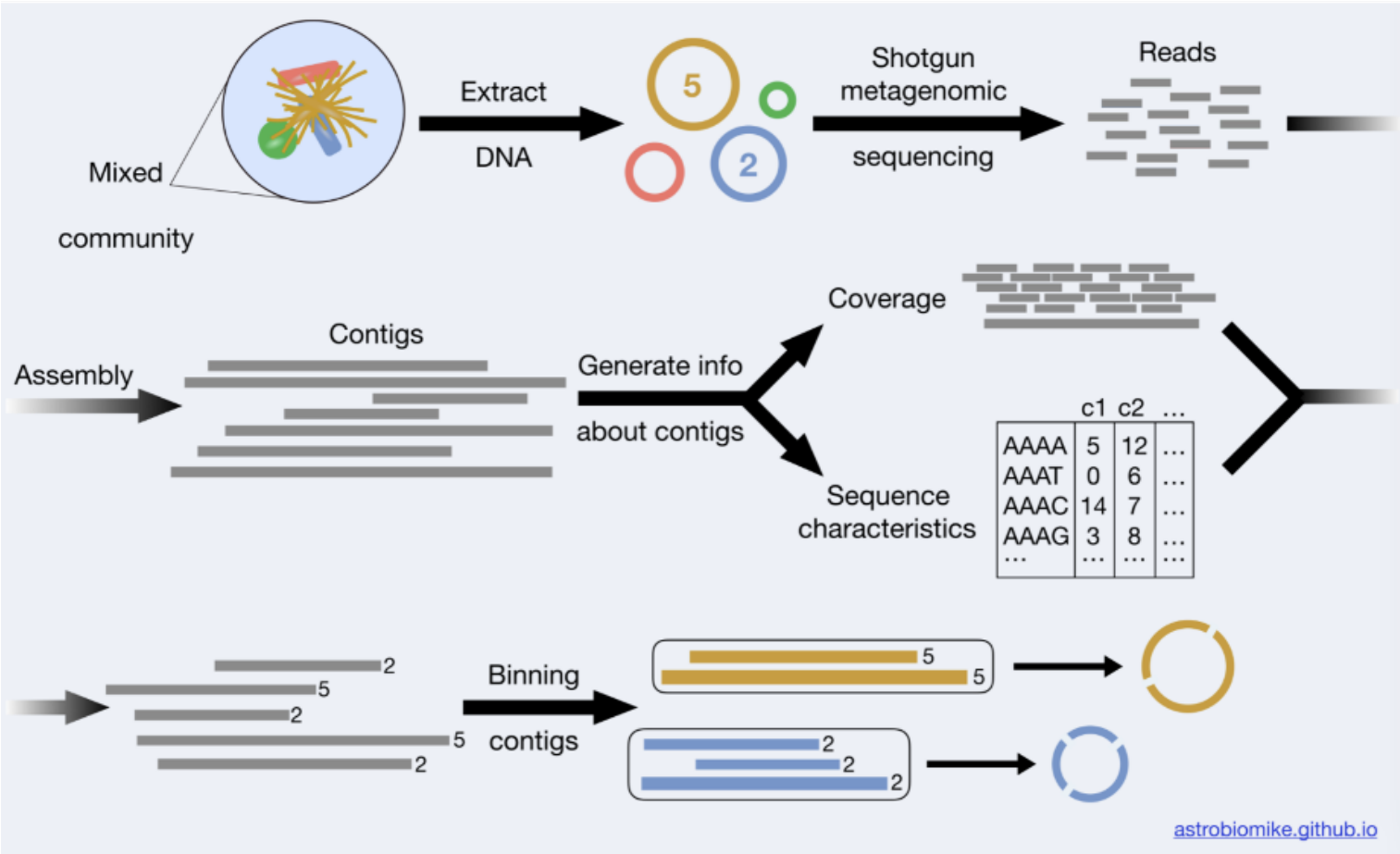
Korem et al., *Science*, 2015

Analysis pipeline

- Reads were mapped to full bacterial genomes. (requires complete reference genomes)
- Total sequencing reads were summed into non-overlapping 10Kbp bins.
- Predict replication origin: Fit $\log_2(C)$ (coverage) into a piecewise linear model. Within each sample, the highest and lowest are the origin and terminus of replication.
- $PTR = \frac{PredC_{peak}}{PredC_{trough}}$: $PredC$ predicted coverage.

When we do not have complete reference genome?

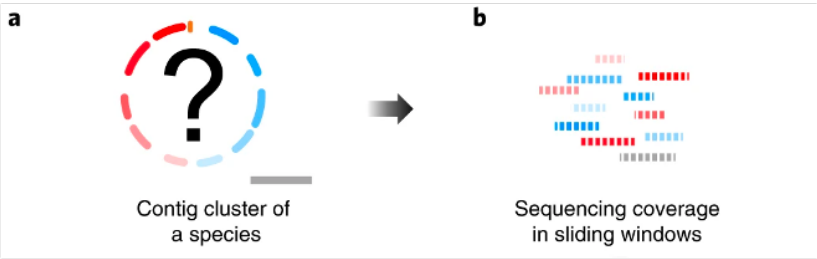
Assemble genomes from metagenomic samples! MAG (metagenomics assymbly genome)

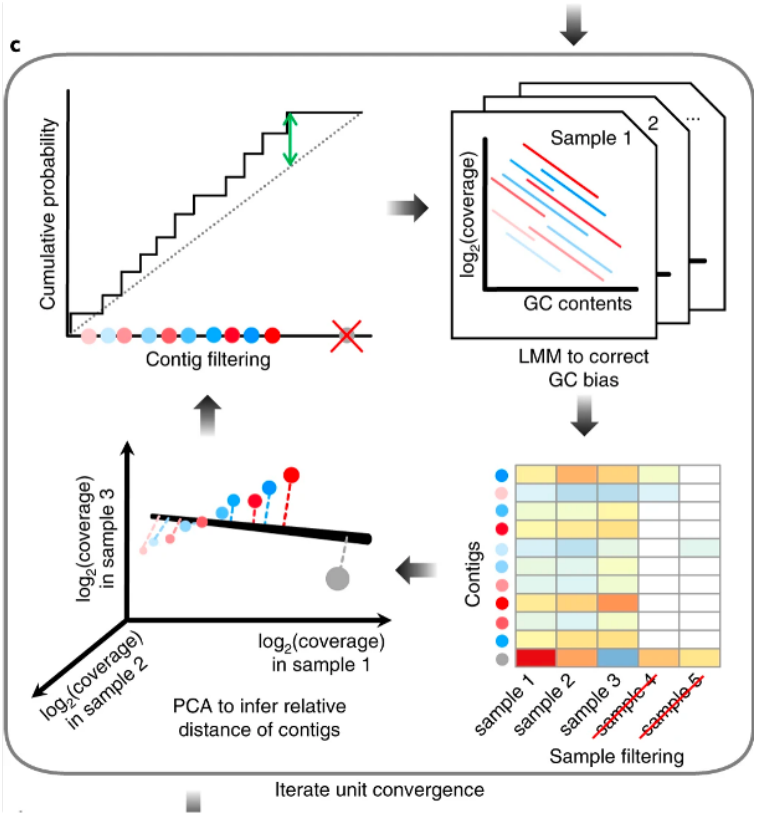


Challenges

- For species with only genome assemblies, the accurate locations of the assemblies on the original genome are unknown
- The contigs assembled from metagenomic data are highly fragmented
- Binning algorithms could fail to cluster all contigs from the same species into one group (incomplete contig clusters), or mistakenly include a fraction of contigs from other species (contig contamination).

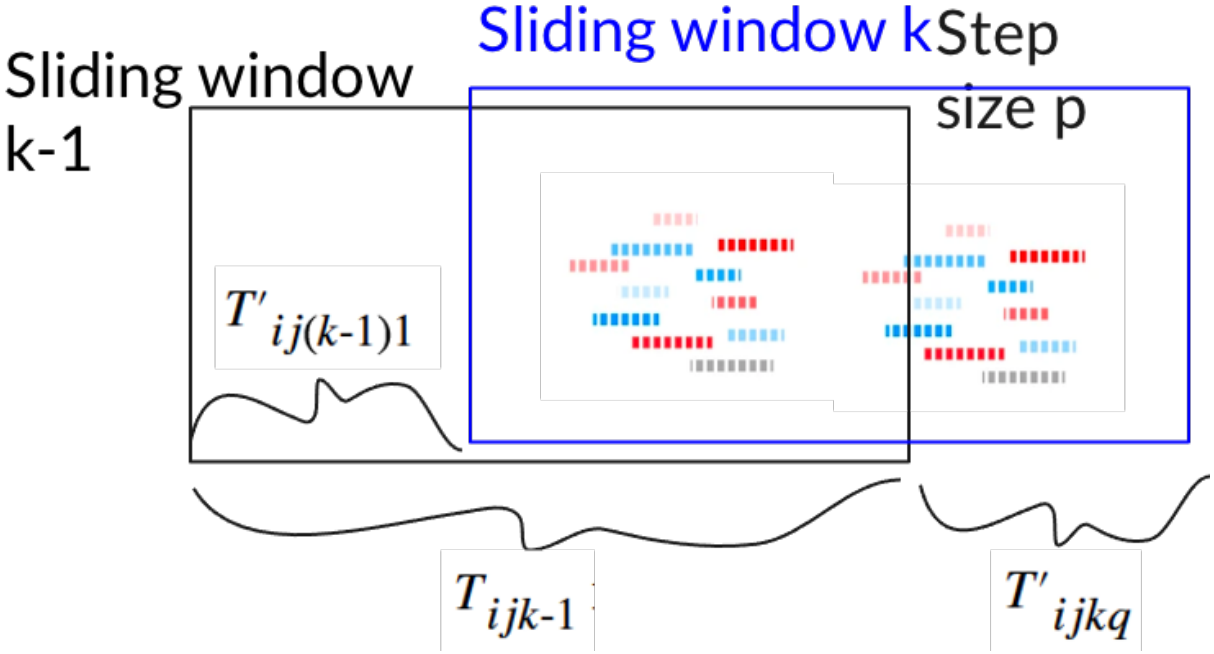
Overview of DEMIC





Gao & Li, *Nature Method*, 2018

Calculation of contig coverages for sliding windows



Average coverage of the k-th window for the j-th contig in the i-th sample:

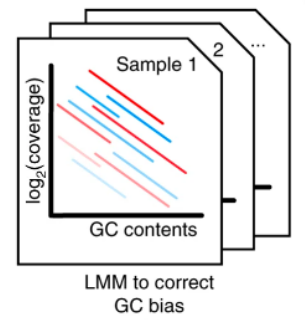
$$Y_{ijk} = \frac{1}{l'}(T_{ijk-1} + T'_{ijkq} - T'_{ij(k-1)1)})$$

Gao & Li, *Nature Method*, 2018

Correcting for sequencing bias using LMM

For the j-th contig of the i-th sample:

$$log_2 Y_{ijk} = a_0 + \left(X_{jk} - \bar{X}\right)a + Z_{ij} + e_{ijk}$$



Average sequencing coverage for the k-th window

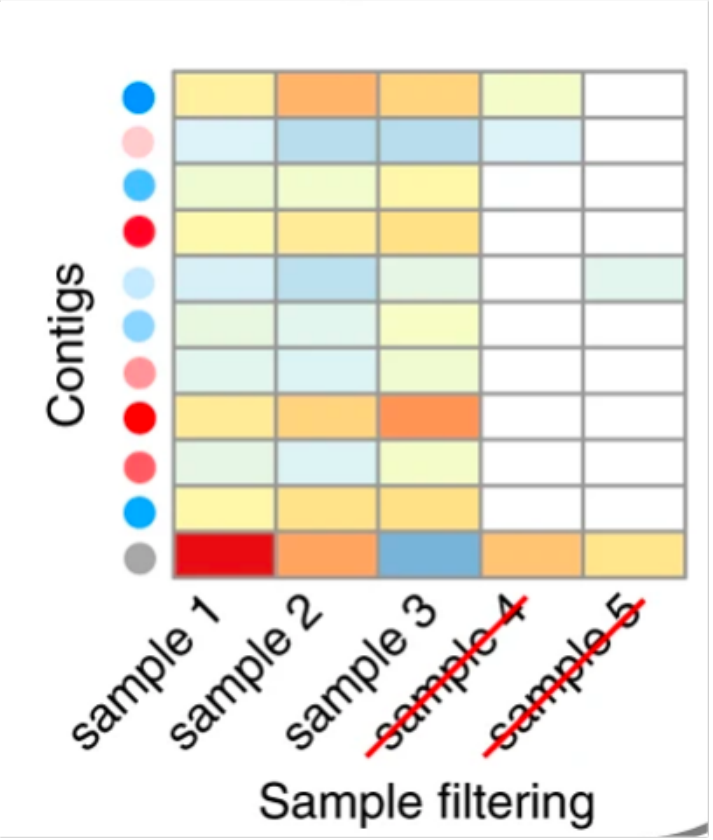
GC content for the k-th window

Sample- and contig-specific random intercept

GC-adjusted log-transformed coverage of sample i and contig j:

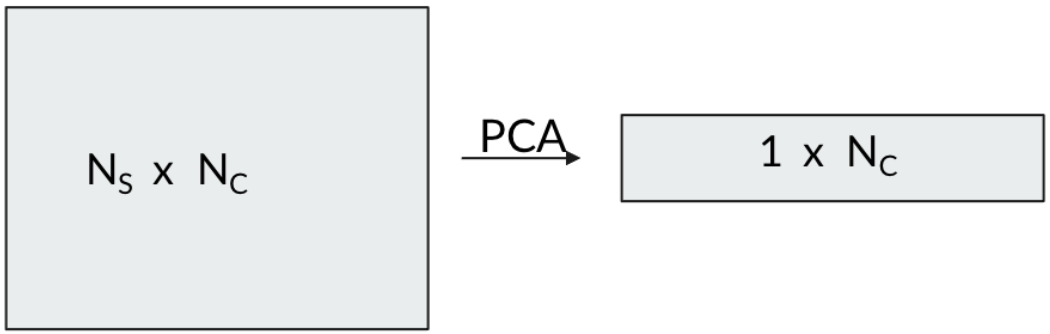
$Y'_{ij} = \hat{a}_0 + \hat{Z}_{ij}$

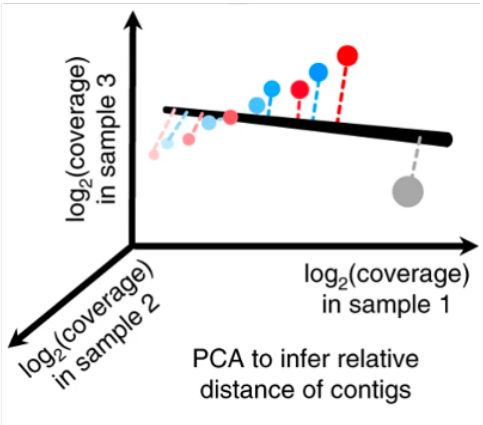
Sample filtering



- Exclude samples with low coverage of the given species (no count for more than half of the contigs within the cluster).

PCA to infer relative distance of contigs

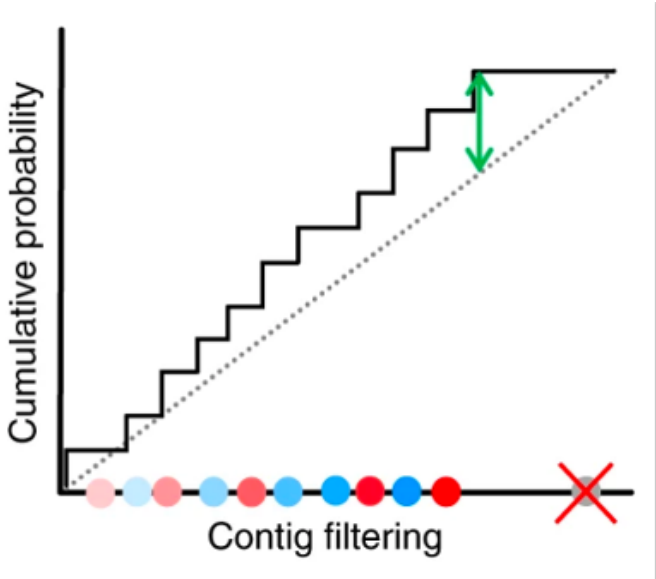




N_s: number of samples; N_c: number of contigs.

The (i, j)-th entry contains GC-adjusted log-transformed coverage of sample i and contig j.

Contig filtering



- Compare the distribution of PC1 values of all contigs against the uniform distribution, Unif(min(U), max(U)).
- If there is a significant difference in distribution, remove one of the two contigs with maximum and minimum PC1 values.
  - Compare the two contigs with respect to their distance from the adjacent contig.
  - Discard the one with the larger distance

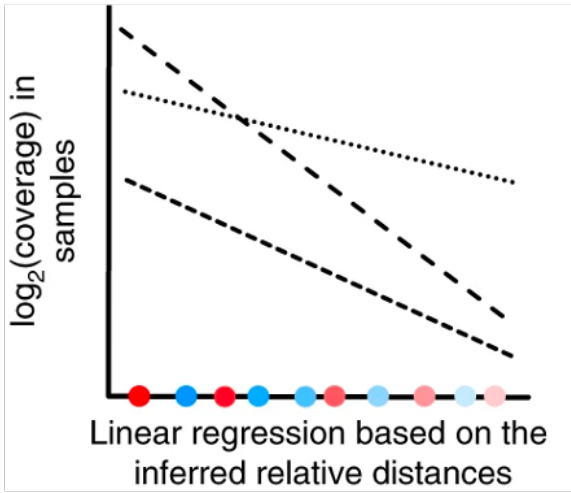
Linear regression based on inferred relative distances

For the i-th sample:

$$Y'_{ij} = b_{0i} + b_i U_j, \quad j = 1, 2, \ldots, N_c$$

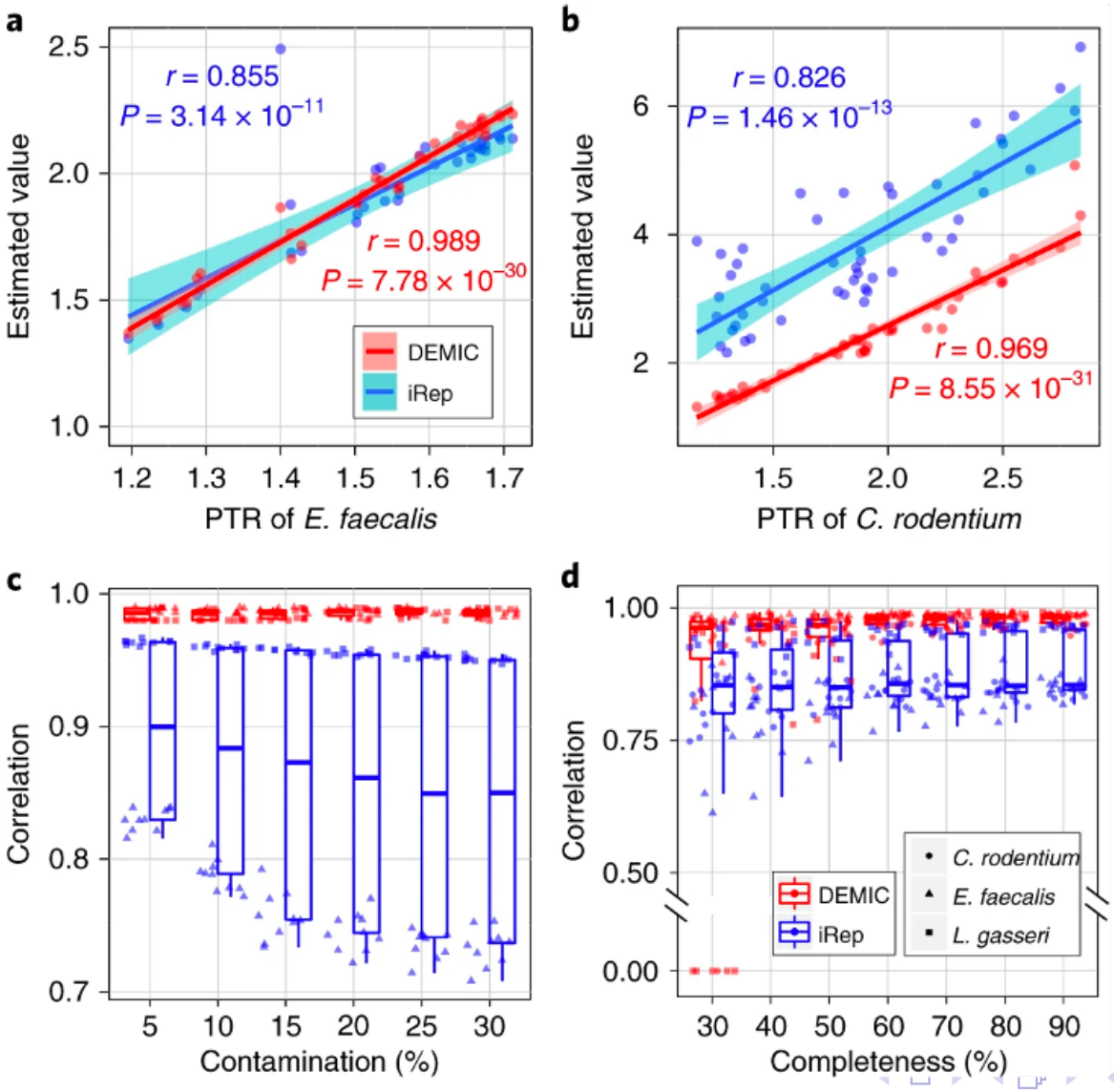
GC-adjusted log-transformed coverage of contig j

PC1 value of contig j

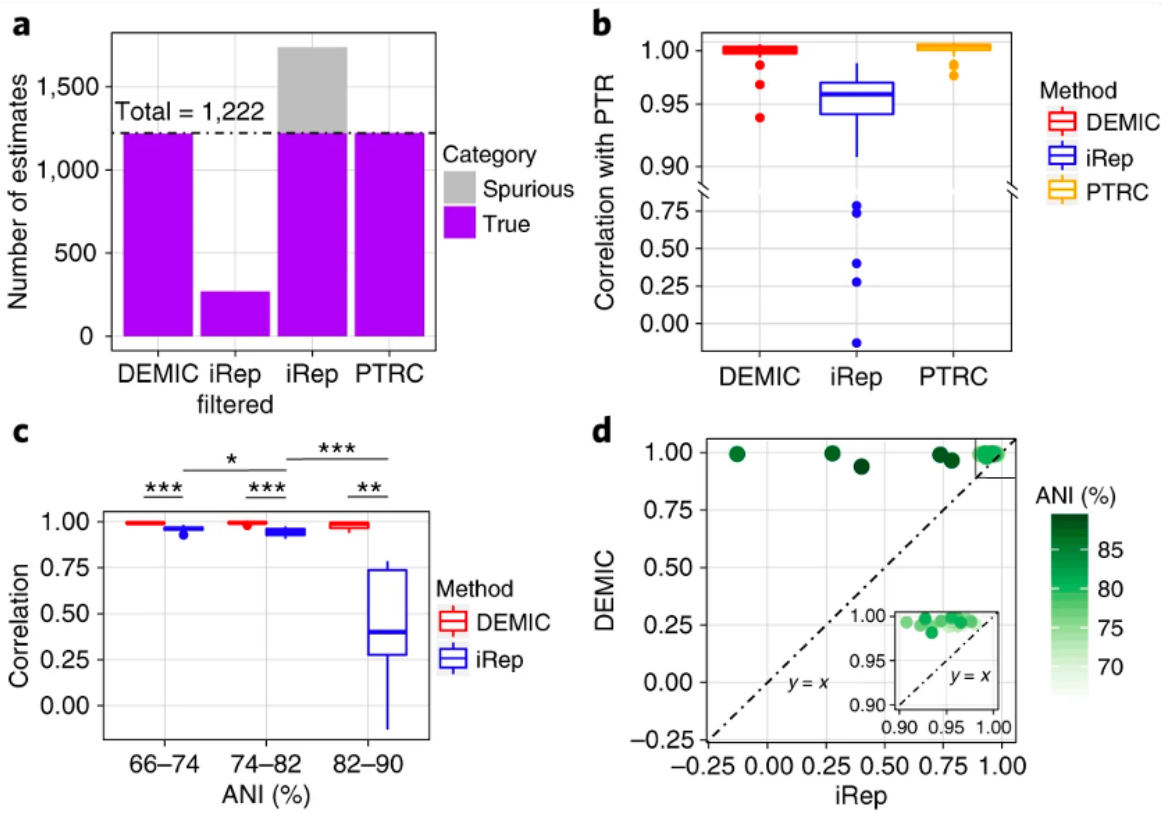


Estimated PTR $ePTR_i = \frac{\exp\left(\hat{b}_{0i} + \hat{b}_i U_{(N_c)}\right)}{\exp\left(\hat{b}_{0i} + \hat{b}_i U_{(N_1)}\right)}, \; i = 1, 2, \cdots, N_s$

Simulated results



Simulated results



Real data analysis

Metagenomic data sets of fecal samples from 26 healthy children and 86 children with Crohn's diseases.