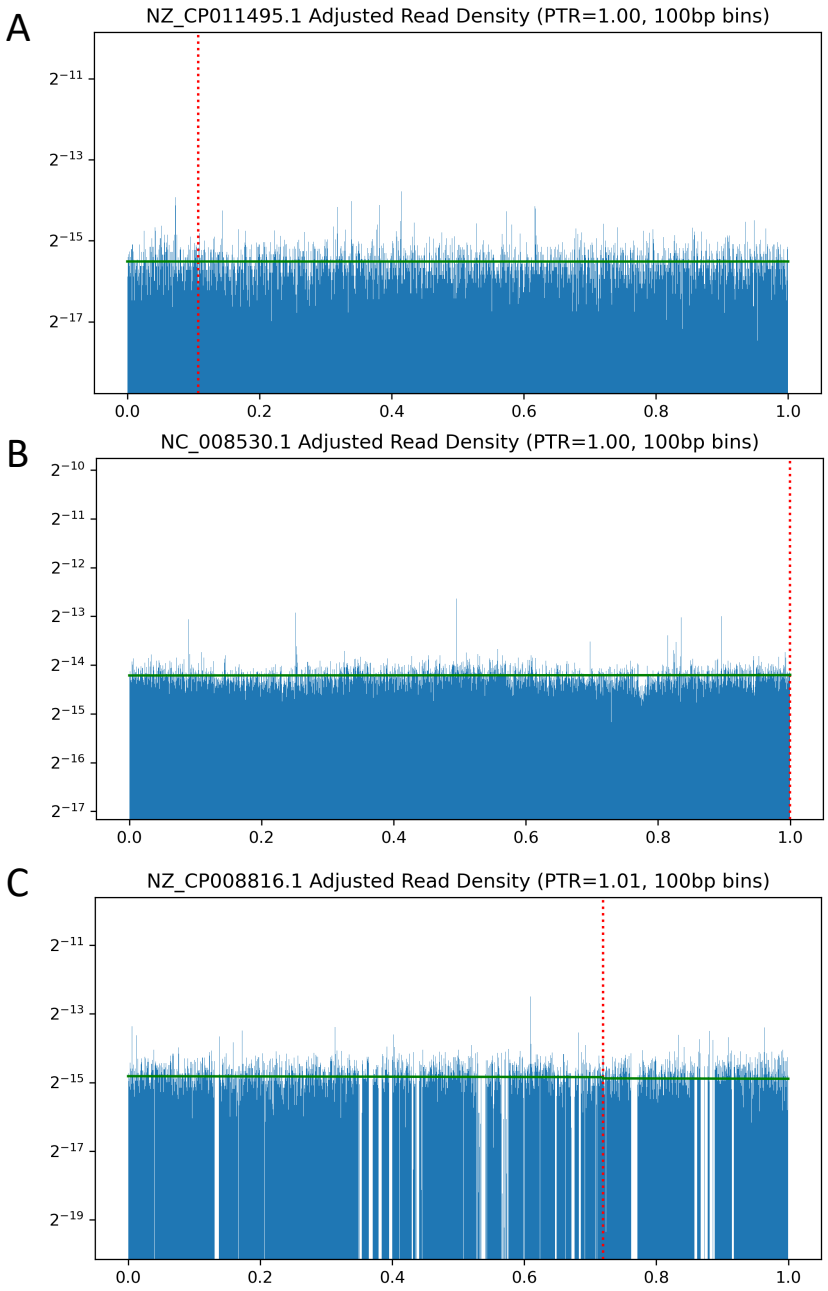
Supplemental Figures S1-S12: Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates

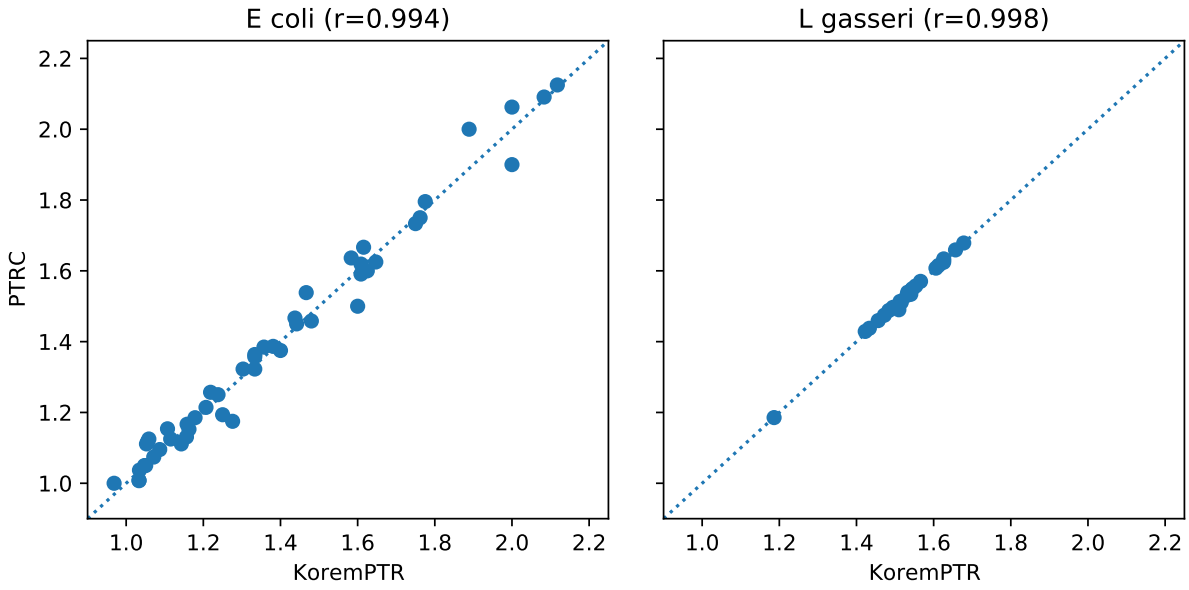
Tyler A. Joseph[1], Philippe Chlenski[1], Aviya Litman[2], Tal Korem[*2,3,4], and Itsik Pe'er [*1,2,5]

[1]*Department of Computer Science, Columbia University, New York, NY, USA*
[2]*Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA*
[3]*Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York, NY, USA*
[4]*CIFAR Azrieli Global Scholars Program, CIFAR, Toronto, Ontario, Canada*
[5]*Data Science Institute, Columbia University, New York, NY, USA*

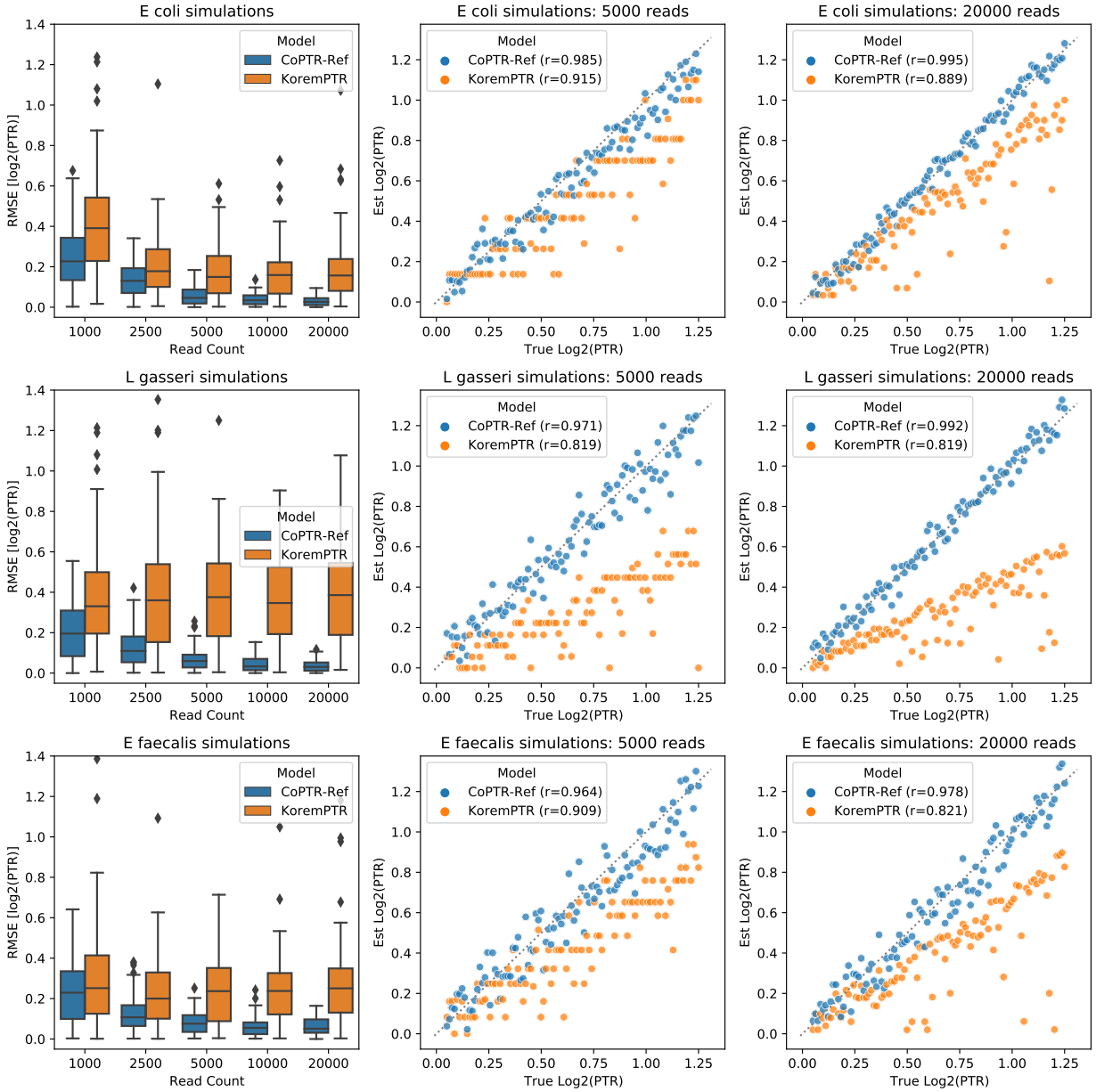
[*]These authors contributed equally to this work. Correspondence: `tal.korem@columbia.edu`, `itsik@cs.columbia.edu`.



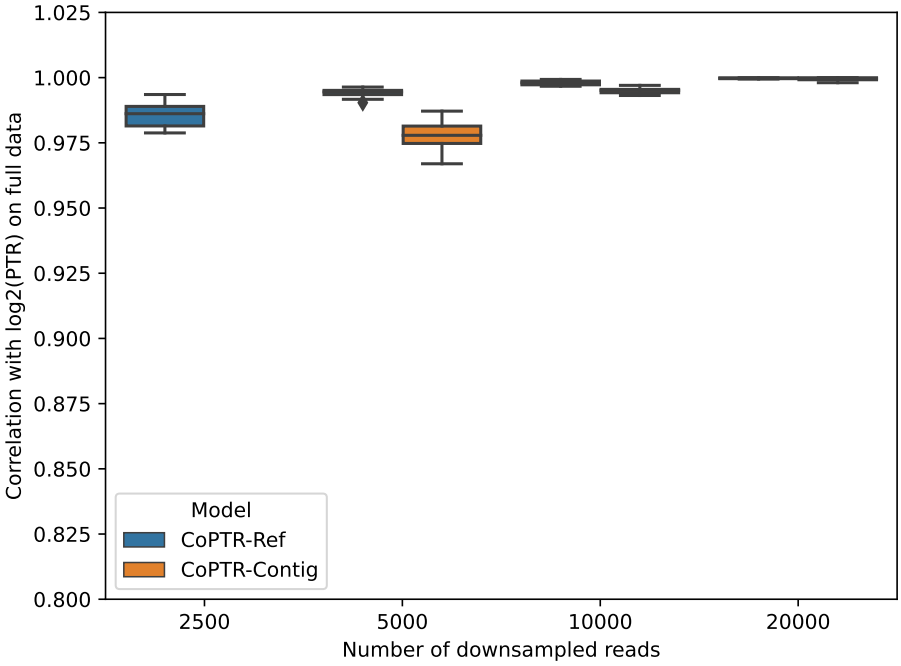
Supplemental Figure S1: Adjusted coverage maps used for simulated data. Coverage maps in 100bp windows for three species: **(A)** *E. coli*, **(B)** *L. gasseri*, and **(C)** *E. faecalis*. Reported PTRs are computed after adjustment for the estimated PTR on each dataset. Dashed red lines denote the replication origin from the DoriC database Luo and Gao (2019).



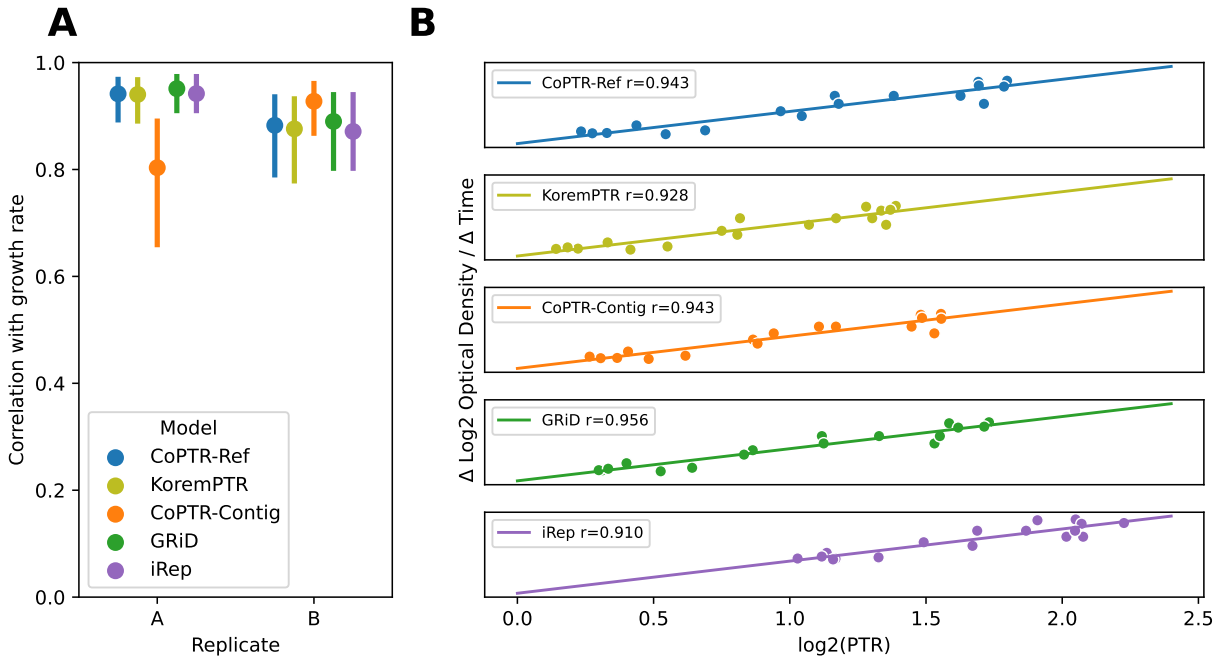
Supplemental Figure S2: Comparison between PTRC and KoremPTR on two datasets. Left: 55 genomic samples of *E. coli*. Right: 25 genomic samples of *L. gasseri*.



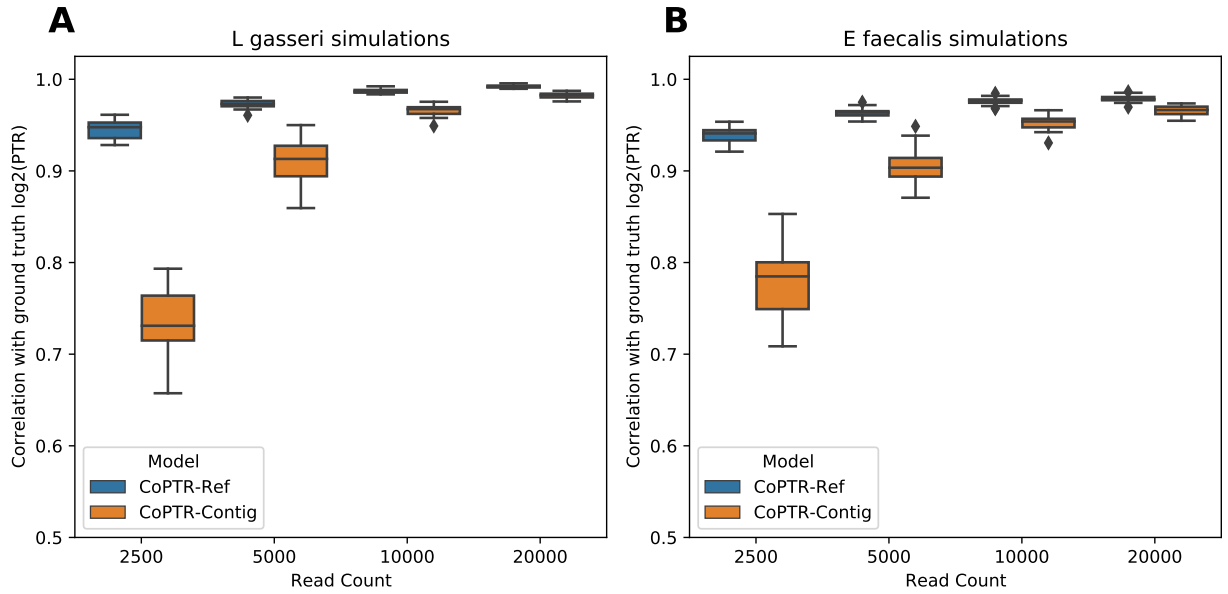
Supplemental Figure S3: Simulation comparison between CoPTR-Ref and KoremPTR. Top: Simulations based on the *E. coli* density map. The first and last columns recapitulate Figure 2. Middle: Simulations based on the *L. gasseri* density map. Bottom: Simulations based on the *E. faecalis* density map.



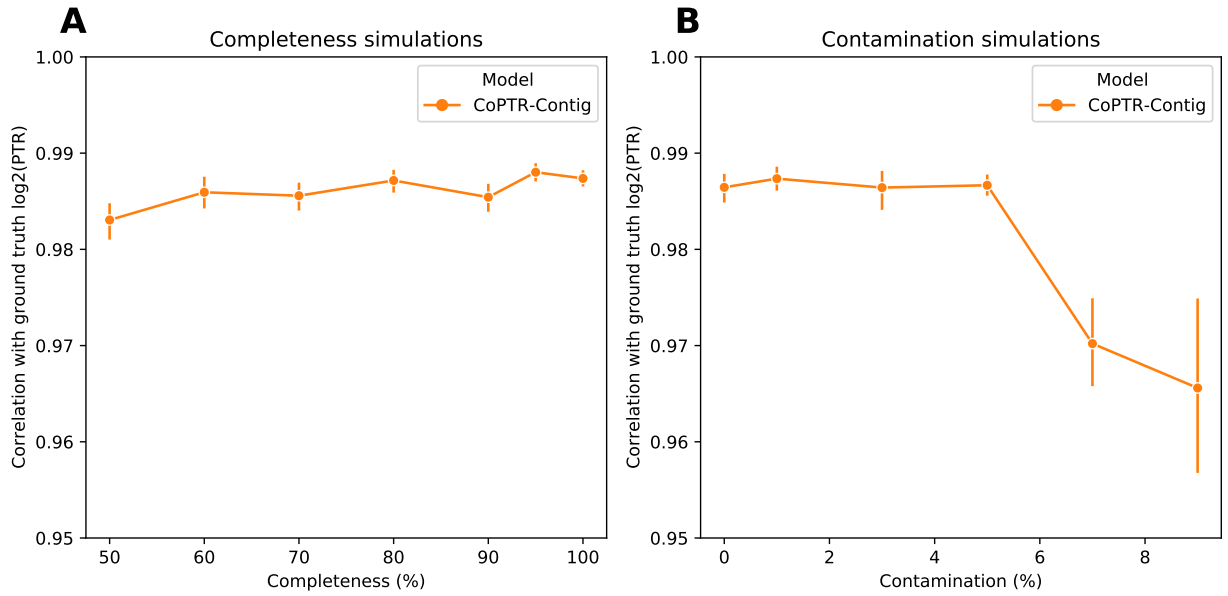
Supplemental Figure S4: Evaluation of required sequencing coverage for CoPTR-Ref and CoPTR-Contig on the *E. coli* genomic dataset. Coverage requirements were evaluated by downsampling the total number of reads to a specified amount (x-axis), and computing the correlation between the downsampled $\log_2$(PTR) and the estimate on the full data set.



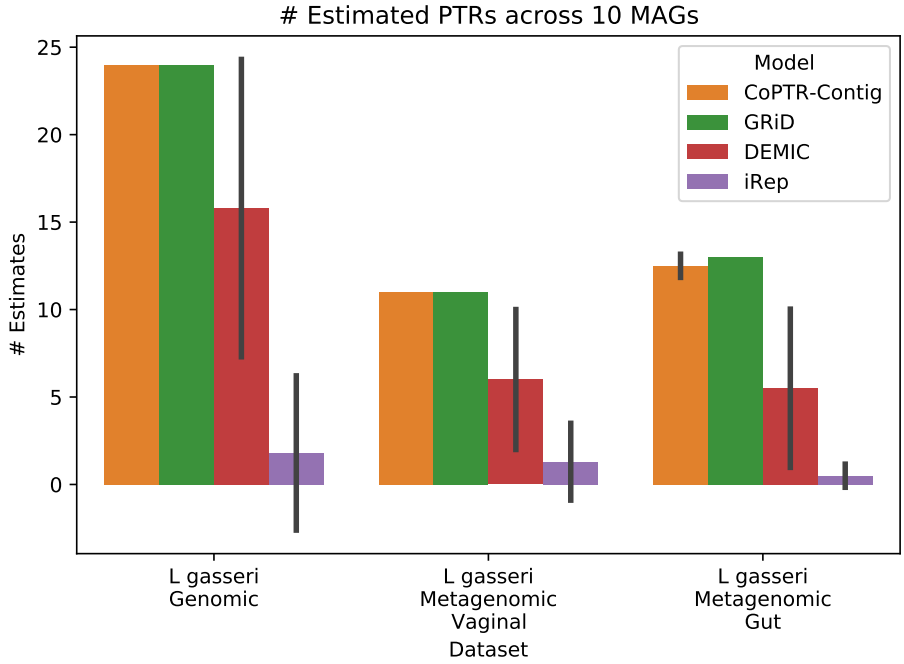
Supplemental Figure S5: Correlation of $\log_2$(PTR)s with (A) growth rates and (B) changes in abundance. (A) Correlation between $\log_2$(PTR)s and E. coli growth rates (defined to be $\frac{1}{\tau}$). Error bars represent 1 standard deviation, computed using Fisher's z-transformation. (B) Correlation between $\log_2$(PTR)s and changes in abundance of E. coli in an unrestricted growth experiment. The theory suggests that log2(PTR)s should be correlated with both quantities. We did not observe a significant difference between each model. DEMIC did not produce estimates on either dataset.



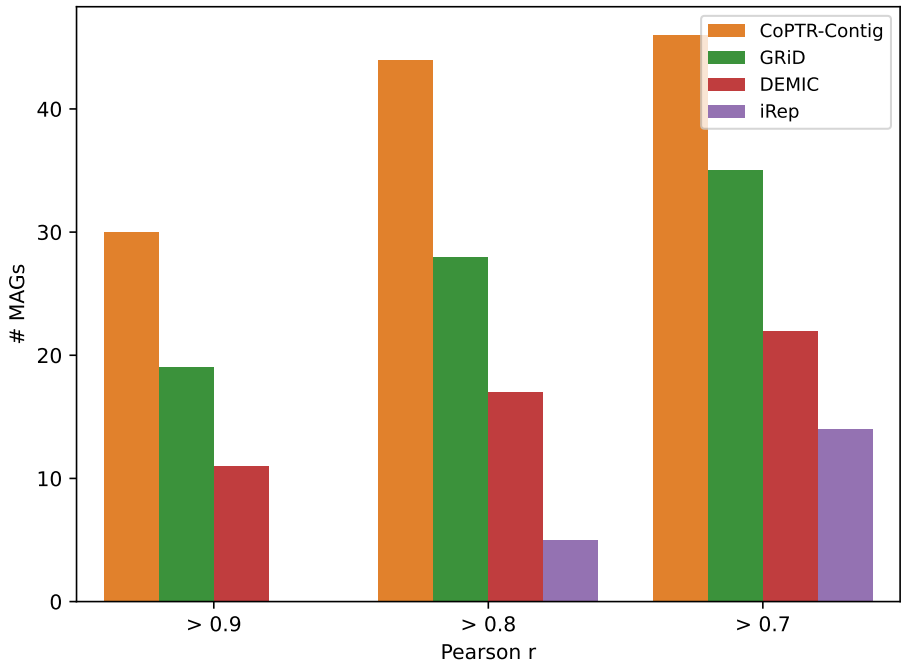
Supplemental Figure S6: Simulation comparison between CoPTR-Ref and CoPTR-Contig. **(A)** Comaparison on simulations from *L. gasseri* density maps. **(B)** Comparison on simulations from *E. faecalis* density maps.



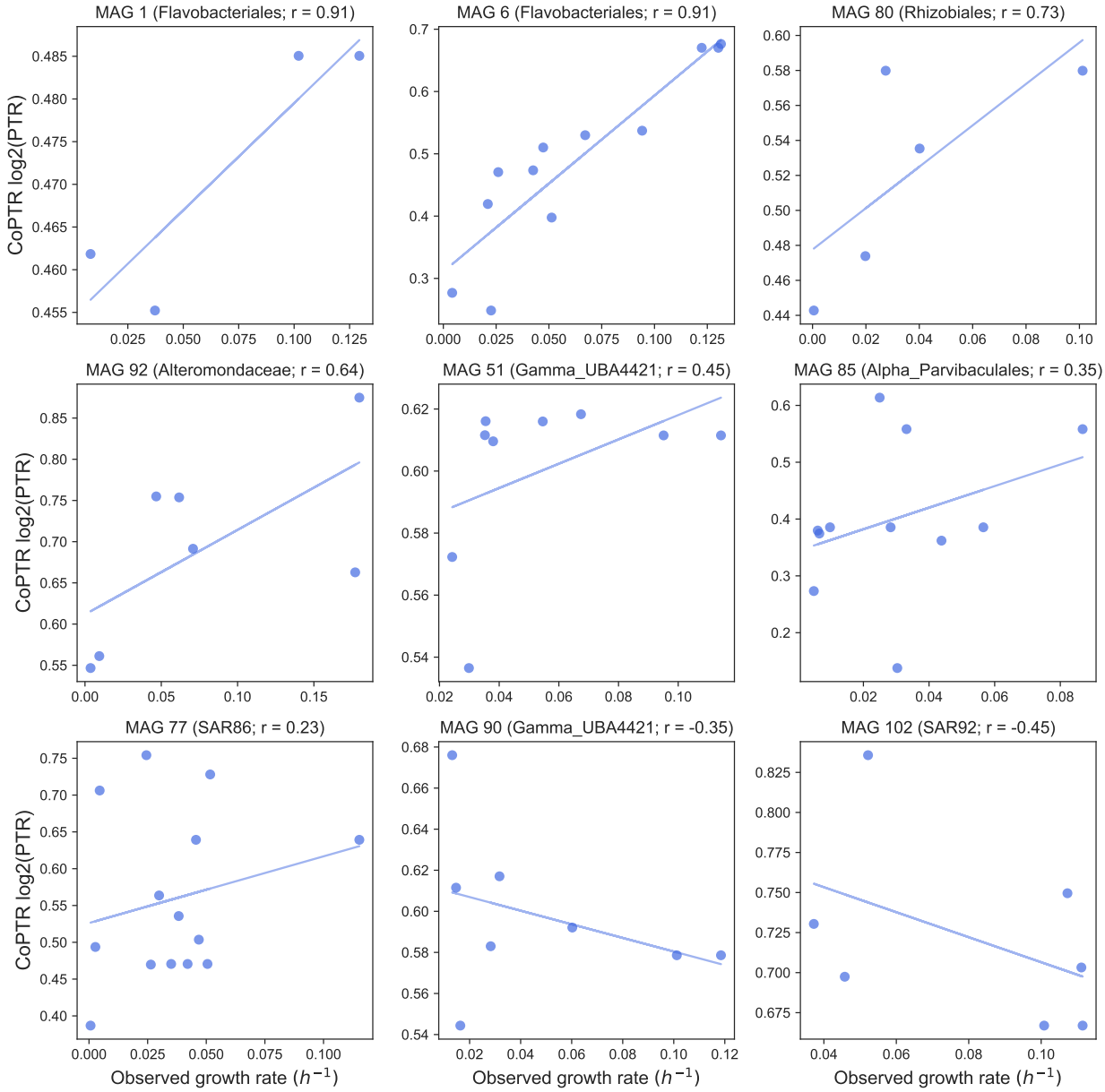
Supplemental Figure S7: CoPTR-Contig is robust to genome completeness and contamination. Evaluation of genome completeness **(A)** and contamination **(B)** on PTR estimation. Error bars depict one standard deviation across 20 simulation replicates, each replicate consisting of 100 simulated PTRs from a simulated assembly.



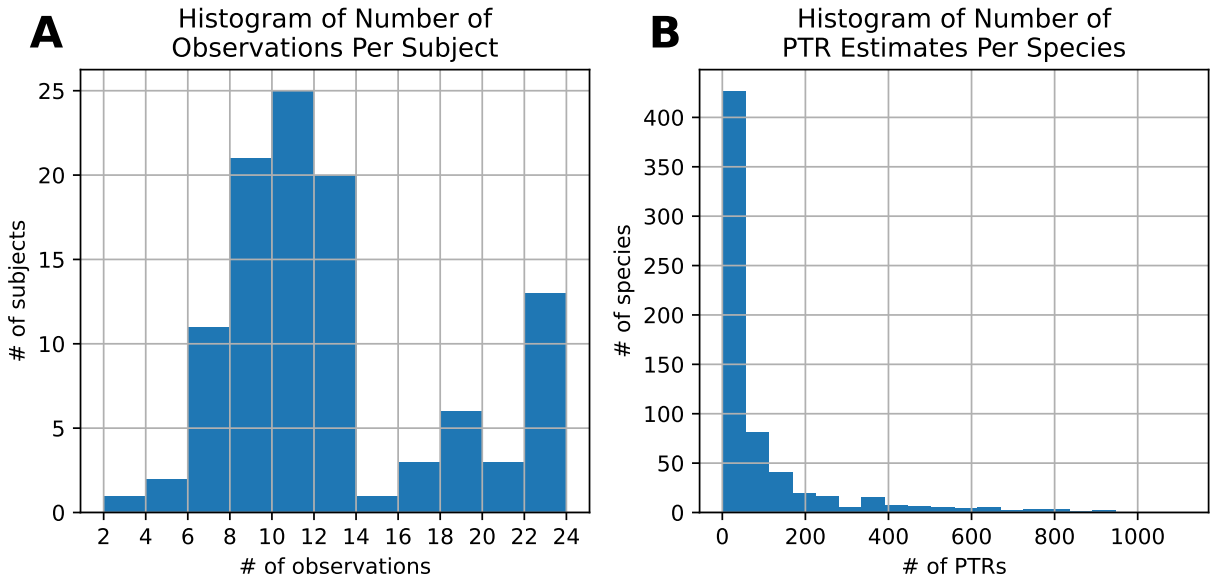
Supplemental Figure S8: Number of PTR estimates output by each model. Mean number of PTR estimates across 10 high quality MAGs. Error bars depict one standard deviation.



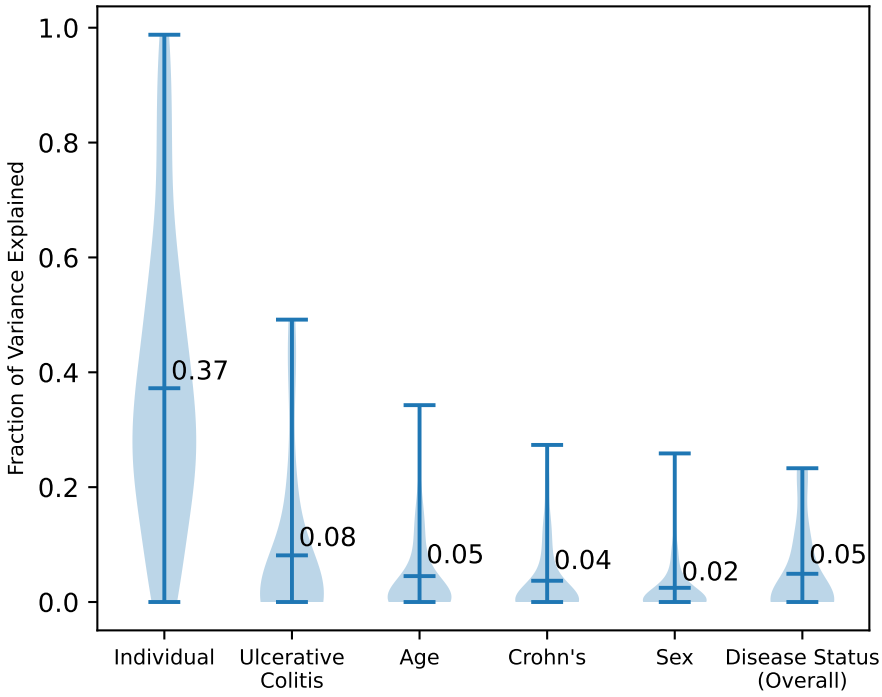
Supplemental Figure S9: Number of MAGs with high correlation to ground truth. Total number of MAGs (y-axis) where the correlation with the ground truth was greater than a thresholded amount (x-axis). Results correspond to the MAGs from Figure 3B.



Supplemental Figure S10: CoPTR values generally correlate with observed growth rates for MAGs with $>$ 90% completeness. Individual correlations for a MAGs in Figure 4A,B with $>$ 90% completeness.



Supplemental Figure S11: Distribution of number of observations per subject (left) and number of PTR estimates per species (right). PTR estimates are sparse across species.



Supplemental Figure S12: Fraction of variance explained when considering the top 50% of subjects with the most observations.

References

Luo H and Gao F. 2019. Doric 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic acids research* **47**: D74–D77.