

1 Supplementary

2 1 Supplementary Note

3 Supplementary Note 1

4 **Data Collection and Preprocessing** Metagenomic datasets were obtained from multiple repos-
5 itories. Freshwater samples (accession numbers: SRR26420192, ERR9631077, ERR4195020), human
6 fecal samples (SRR13060973, SRR13060977), and one plant sample (SRR10968246) were retrieved
7 from the Sequence Read Archive database. Marine samples (1021526, 1102218, 1125694) were sourced
8 from [1]. An additional human fecal sample (K0268, accession CNPO000334) was obtained from the
9 China National GeneBank (CNGB).

10 **Assembly and Gene Prediction** All metagenomic samples were assembled using metaSPAdes
11 with default parameters. Microbial diversity indices were calculated for each sample using Nonpareil
12 with the parameter settings ‘-T alignment -f fasta’. Complete protein-coding genes within assembled
13 contigs were identified using Prodigal under default configuration.

14 **Statistical Analysis** Total base pairs and gene counts were calculated separately for short and
15 long contigs. Four linear regression models were constructed using the ‘lm’ function in R to examine
16 relationships between microbial diversity and: (1) summed base pairs in short contigs, (2) summed
17 base pairs in long contigs, (3) number of complete genes in short contigs, and (4) number of complete
18 genes in long contigs.

19 **Comparative Analysis of Contig Length Distributions** To assess the impact of microbial com-
20 munity complexity on assembly outcomes, we performed a comparative analysis of contig length distri-
21 butions between metagenomic samples exhibiting distinct levels of microbial diversity. Sample K0268
22 (microbial diversity = 17.15) was selected as a representative of low-complexity microbial communi-
23 ties, whereas sample ERR9631077 (microbial diversity = 21.07) was designated as a high-complexity
24 representative. The distribution of contig lengths following assembly was visualized for both samples
25 (**Supplementary Figure 9**). To evaluate statistical differences between the contig length distri-
26 butions of low- and high-complexity samples, we employed the Wilcoxon rank-sum test (one-tailed;
27 $p < 2.2e^{-16}$). Our results demonstrated that high-complexity metagenomic samples yielded a greater
28 proportion of short contigs compared to low-complexity samples following assembly, indicating that
29 increased microbial diversity negatively impacts assembly contiguity.

Supplementary Note 2

Sample Selection and Assembly Four representative environmental samples were selected for coverage analysis: one freshwater sample (ERR4195020), one plant sample (SRR10968246), one marine sample (1021526), and one cold seep sediment sample (CS01, accession CNR1069767). All samples were assembled using metaSPAdes with default parameters.

Read Mapping and Coverage Calculation Reads were aligned to their respective assembled contigs using BWA-MEM. Coverage calculations were performed using PySAM to determine the sequencing depth for each contig. Average coverage values were computed separately for short and long contigs within each dataset.

Statistical Analysis of Coverage Distributions and Results Coverage distributions between short and long contigs were compared using the Wilcoxon rank-sum test (two-tailed, $\alpha = 0.05$). Statistical analyses were conducted to assess whether significant differences in coverage existed between short and long contigs across all environmental sample types. Analysis revealed that short contigs exhibited significantly lower average coverage compared to long contigs across all tested environmental samples, including plant, freshwater, marine, and cold seep sediment datasets (Wilcoxon rank-sum test, $p < 2.2e^{-16}$ for all comparisons; **Figure 1c**).

Supplementary Note 3

Simulation of Contigs from Reference Genomes Simulated contigs were generated from two reference genomes: *Halomonas elongata* DSM 2581 and *Halomonas* sp. R57-5, both derived from the CAMI II marine dataset. Contigs were extracted at eight predefined length thresholds: 10,000 bp, 7,500 bp, 5,000 bp, 2,500 bp, 1,000 bp, 800 bp, 500 bp, and 250 bp. For each length threshold, 1,000 independent simulations were performed to ensure statistical robustness. Contigs were randomly sampled from the reference genomes according to the specified length parameters.

Dimensionality Reduction and Clustering Analysis Tetranucleotide frequency (TNF) features were extracted from all simulated contigs. To visualize the distribution of TNF features across length thresholds, t-distributed stochastic neighbor embedding (t-SNE) [2] was applied to project the high-dimensional TNF vectors into two-dimensional space (**Figure 1d**). Silhouette coefficients were calculated for each length threshold to quantitatively assess the separability of contigs originating from different reference genomes. The relationship between silhouette coefficients and contig length was evaluated using linear regression analysis.

The t-SNE visualization demonstrated that TNF profiles from the two distinct species became progressively less distinguishable as contig length decreased (**Figure 1d**). Quantitative analysis revealed a significant positive correlation between silhouette coefficient and contig length (linear regression, $p = 3.9 \times 10^{-4}$; **Figure 1e**), indicating that TNF-based species discrimination is substantially compromised in shorter contigs.

Supplementary Note 4

Dataset Selection and Abundance Definition The CAMI II marine dataset (10 samples) was used to investigate the relationship between microbial abundance and assembly quality. For each sample, its reference genomes were ranked according to their abundance setting in the simulation procedure in descending order. Reference genomes in the upper 50th percentile were classified as high-abundance genomes, while those in the lower 50th percentile were designated as low-abundance genomes.

Assembly and Contig Mapping All samples were assembled using MegaHit with default parameters. Assembled contigs were aligned to their corresponding reference genomes using Minimap2 to determine their genomic origins. Contigs were subsequently categorized based on whether they originated from high-abundance or low-abundance genomes.

Statistical Analysis of Contig Length Distributions Contig length distributions were compared between high-abundance and low-abundance genome categories using the Wilcoxon rank-sum test (one-tailed, $\alpha = 0.05$). Contigs originating from high-abundance genomes were significantly longer than those from low-abundance genomes (Wilcoxon rank-sum test, $p < 2.2 \times 10^{-16}$; **Supplementary Figure 12**), confirming that low-abundance microorganisms exhibit more fragmented genome assemblies.

Supplementary Note 5

To quantify the specific contributions of CompleteBin to genome recovery and the enhancement of MAG quality, we systematically compared the performance of CompleteBin and COMEBin using contigs longer than 1,000 bp on the CAMI II marine dataset.

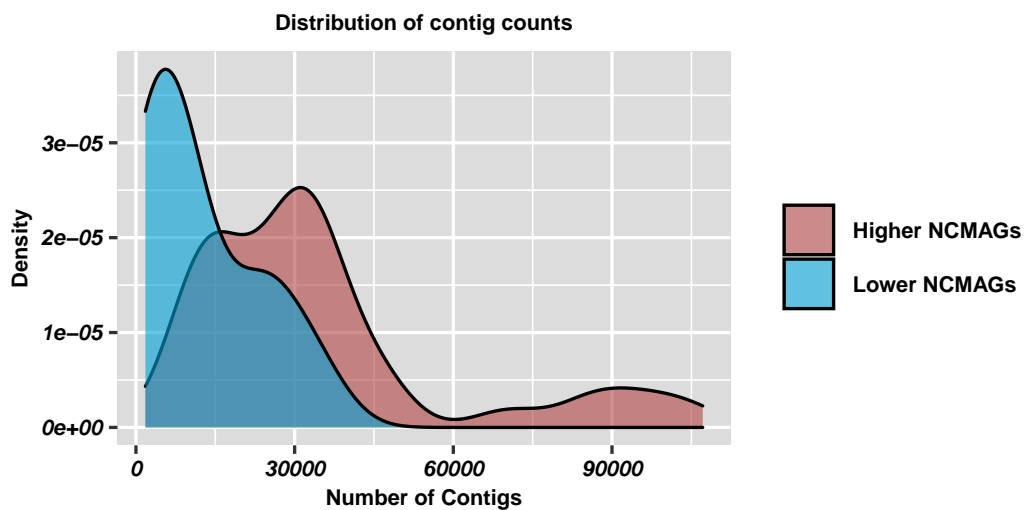
Of the 977 reference genomes in the CAMI II marine dataset, CompleteBin exclusively recovered 93 genomes, whereas COMEBin exclusively identified only 7 genomes. Genomes uniquely recovered by CompleteBin exhibited significantly lower abundance than genomes detected by both approaches (Wilcoxon rank-sum test, $p < 2.2 \times 10^{-16}$; **Supplementary Figure 2**). MAGs from these unique

low-abundance genomes recovered by CompleteBin demonstrated significantly reduced N50 values compared to MAGs from genomes identified by both approaches (Wilcoxon rank-sum test, $p < 2.2 \times 10^{-16}$; for both comparisons; **Supplementary Figure 13**). These findings demonstrate that CompleteBin recovers a greater number of low-abundance reference genomes by more effectively handling short contigs.

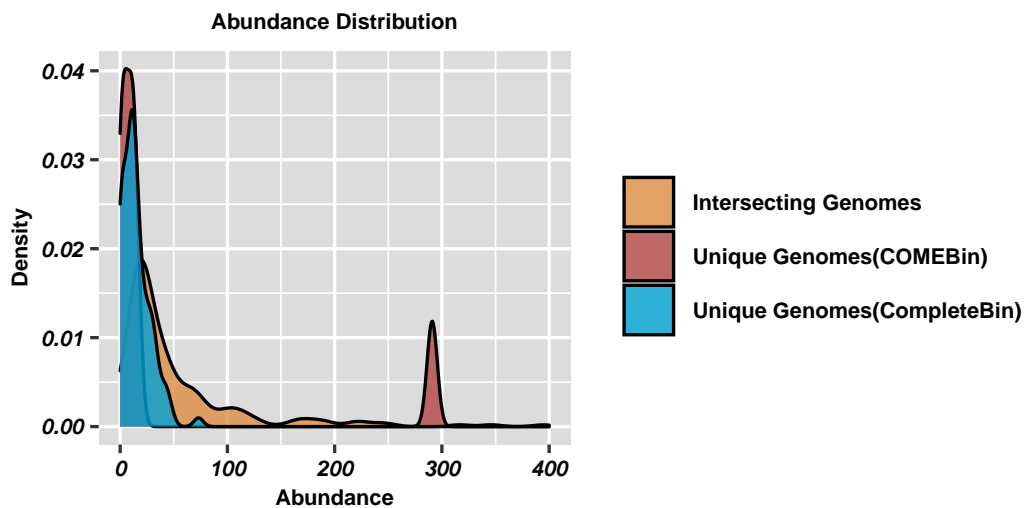
Beyond discovering novel genomes, CompleteBin substantially improved the quality of existing MAG reconstructions. Among shared genomes, CompleteBin enhanced the quality of 34 MAGs, upgrading LQMAGs to MQMAGs or MQMAGs to NCMAGs, whereas COMEBin upgraded only 12 MAGs between these quality categories.

Supplementary Note 6

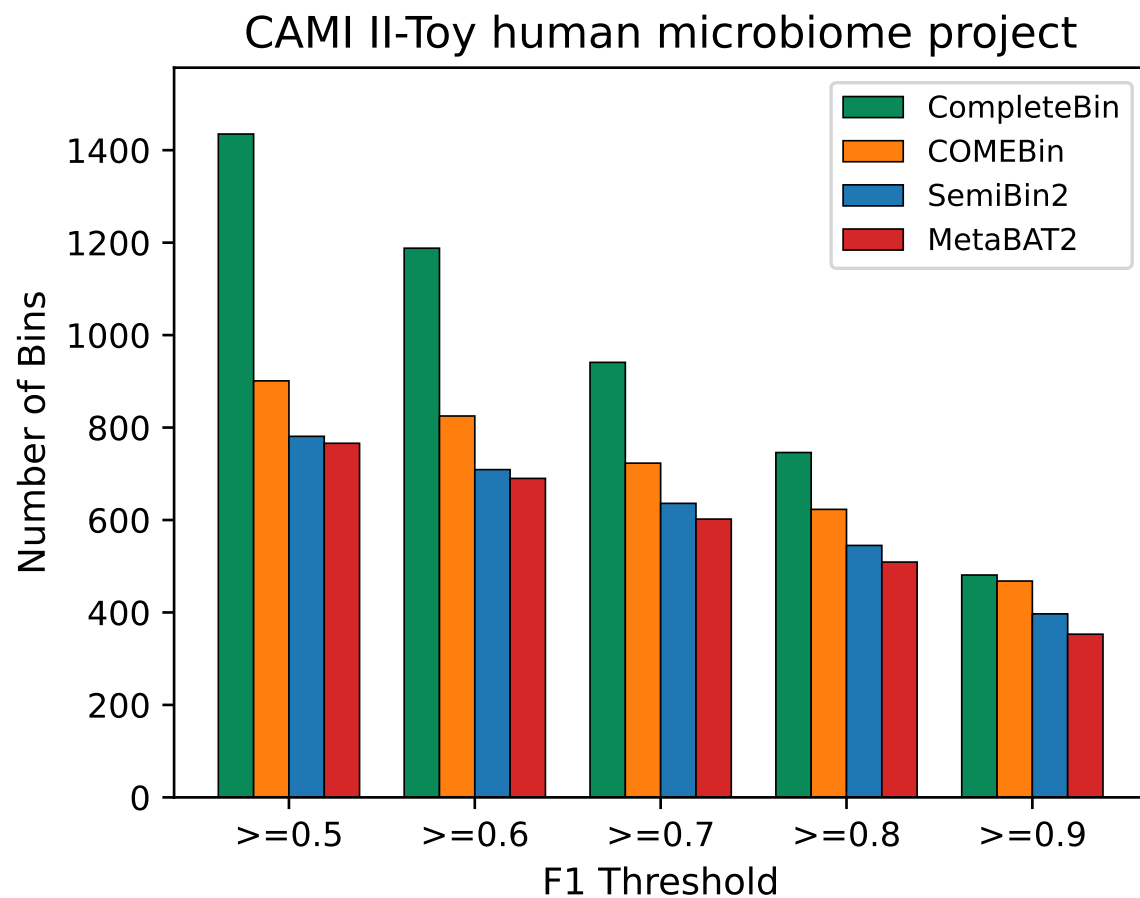
For a subset of metagenomic samples from the CAMI II human microbiome dataset, CompleteBin yielded fewer NCMAGs than COMEBin. To investigate this observation, we compared contig count distributions between samples where CompleteBin outperformed COMEBin and samples where it underperformed, based on the number of NCMAGs recovered. Samples in which CompleteBin produced fewer NCMAGs contained significantly fewer contigs than samples in which CompleteBin produced more NCMAGs (Wilcoxon rank-sum test, two-tailed, $p = 3.42 \times 10^{-5}$; **Supplementary Figure 1**). This finding suggests that CompleteBin performance may be limited in datasets with low contig counts, potentially due to insufficient training iterations when the number of contigs is small. Specifically, CompleteBin performance appeared constrained in datasets containing fewer than 8,162 contigs, the median contig count among samples where CompleteBin yielded fewer NCMAGs than COMEBin.



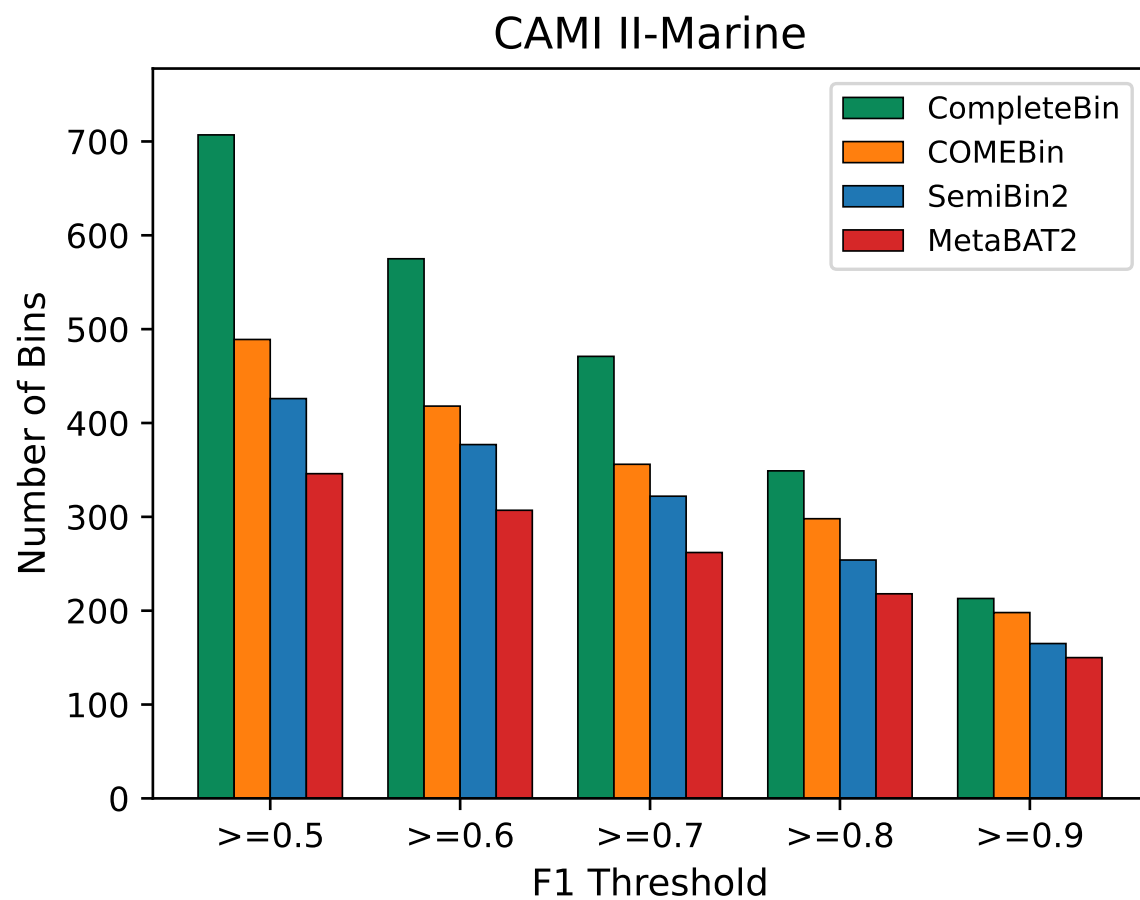
Supplementary Figure 1: Contig count distribution for metagenomic samples in the CAMI II toy human microbiome project after assembly. Red indicates samples where CompleteBin yielded more near-complete MAGs (NCMAGs) than COMEBin; blue indicates samples where CompleteBin yielded fewer NCMAGs than COMEBin.



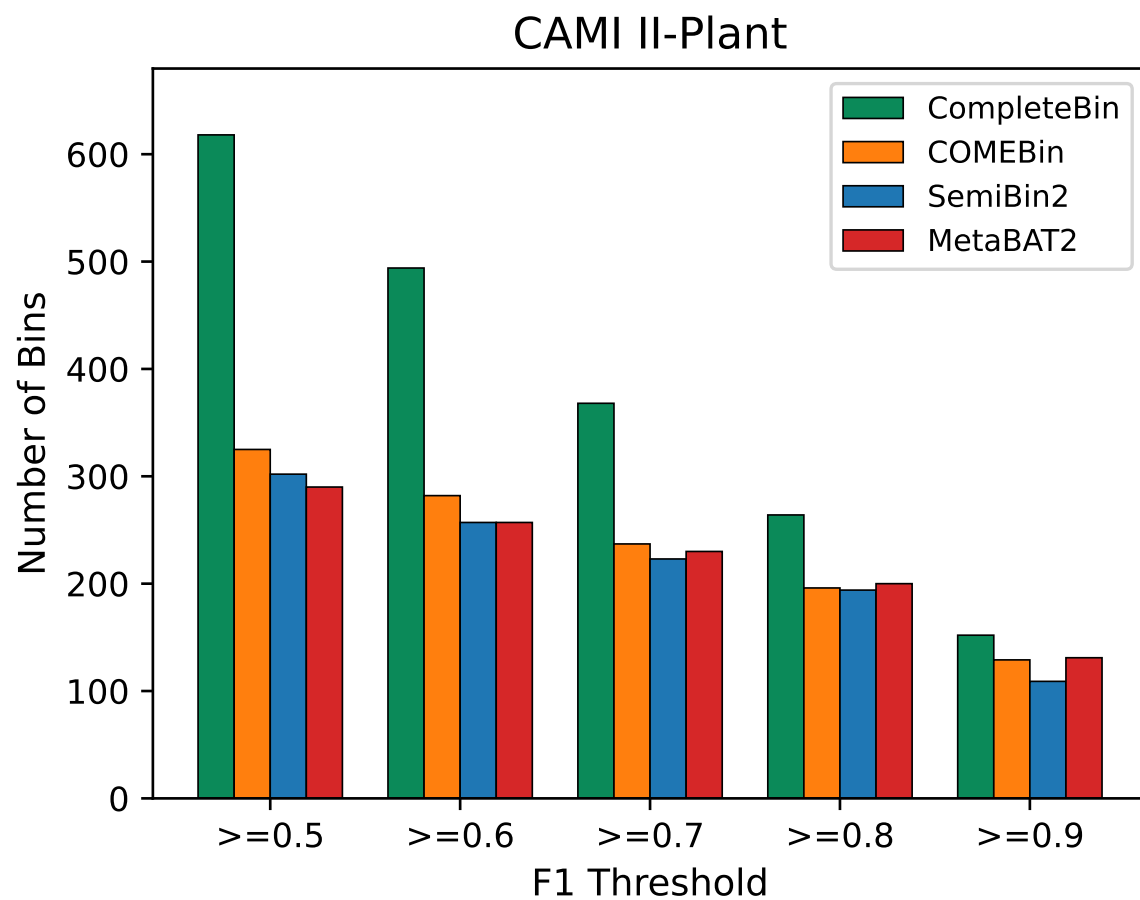
Supplementary Figure 2: Abundance distributions of recovered genomes across three categories: genomes identified by both binning approaches (intersecting genomes), genomes exclusively recovered by CompleteBin (unique genomes (CompleteBin)), and genomes exclusively recovered by COMEBin (unique genomes (COMEBin)).



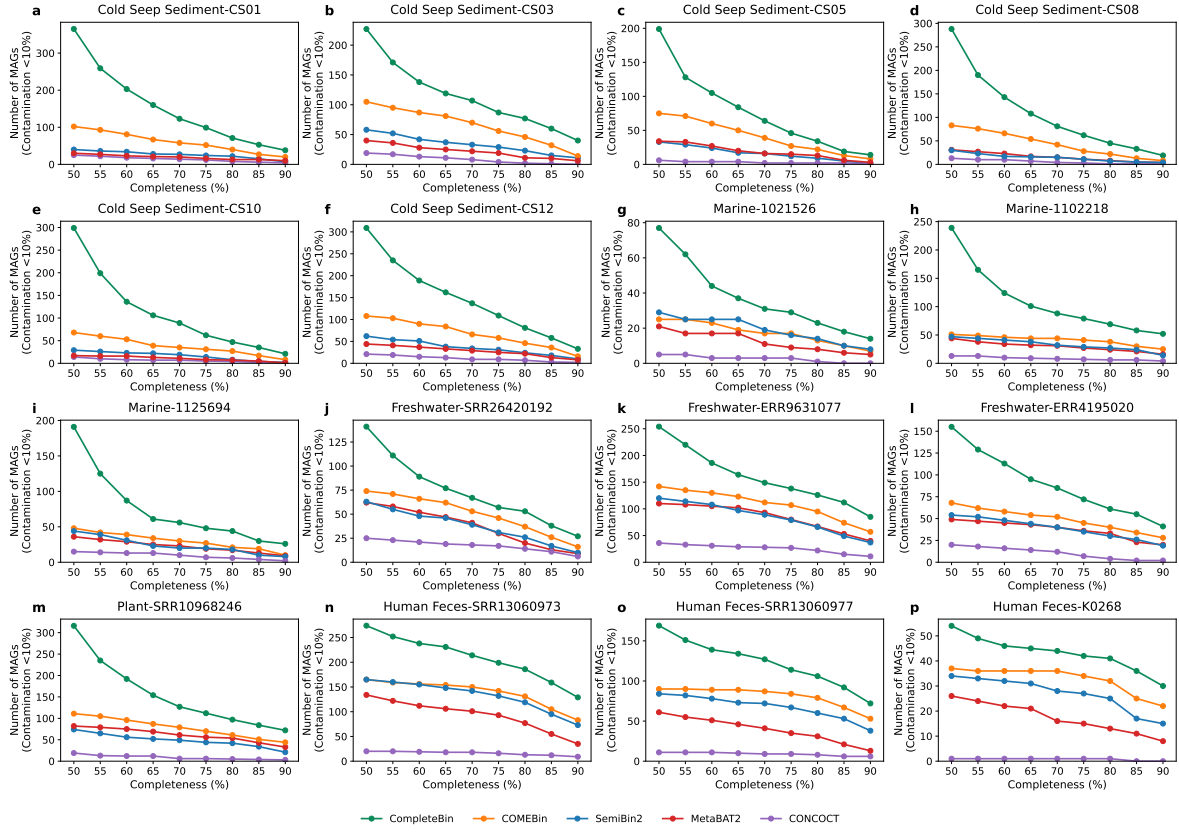
Supplementary Figure 3: Number of MAGs exceeding F1 score (bp) thresholds of 0.5, 0.6, 0.7, 0.8, and 0.9 for CompleteBin, COMEBin, SemiBin2, and MetaBAT2 on the CAMI II toy human microbiome project dataset.



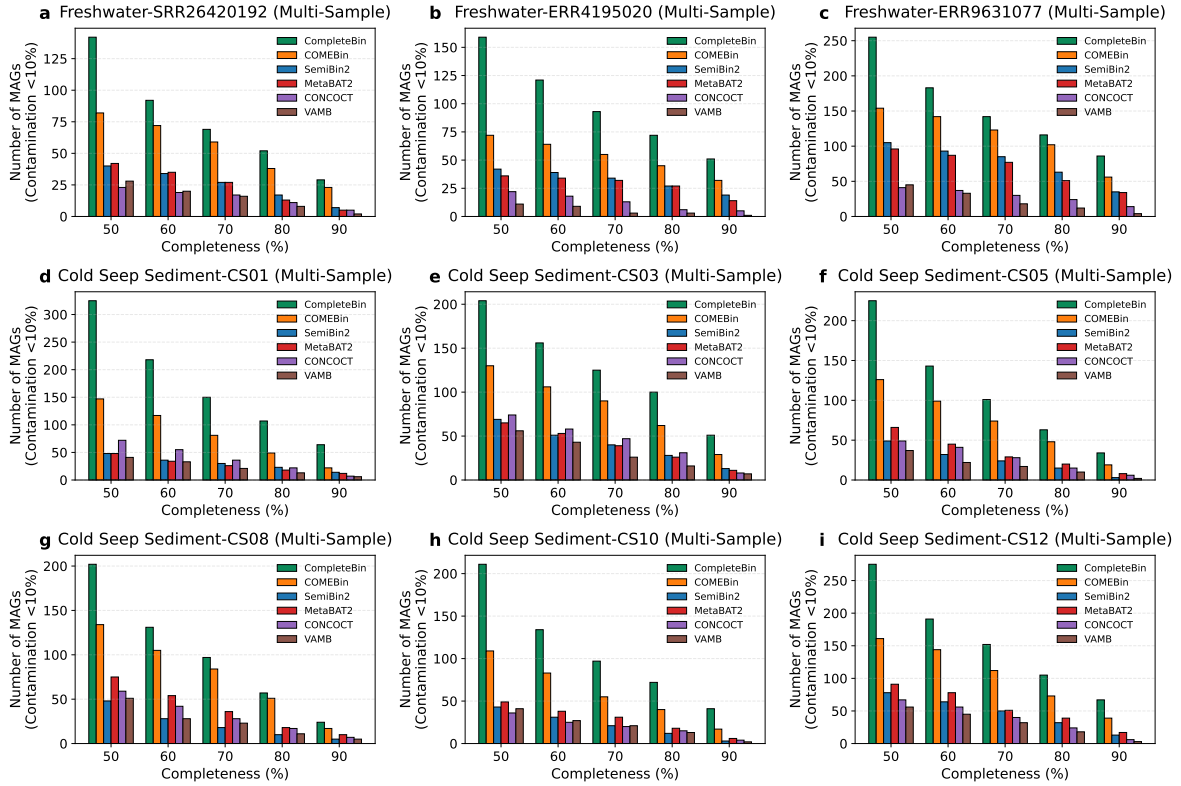
Supplementary Figure 4: Number of MAGs exceeding F1 score (bp) thresholds of 0.5, 0.6, 0.7, 0.8, and 0.9 for CompleteBin, COMEBin, SemiBin2, and MetaBAT2 on the CAMI II marine dataset.



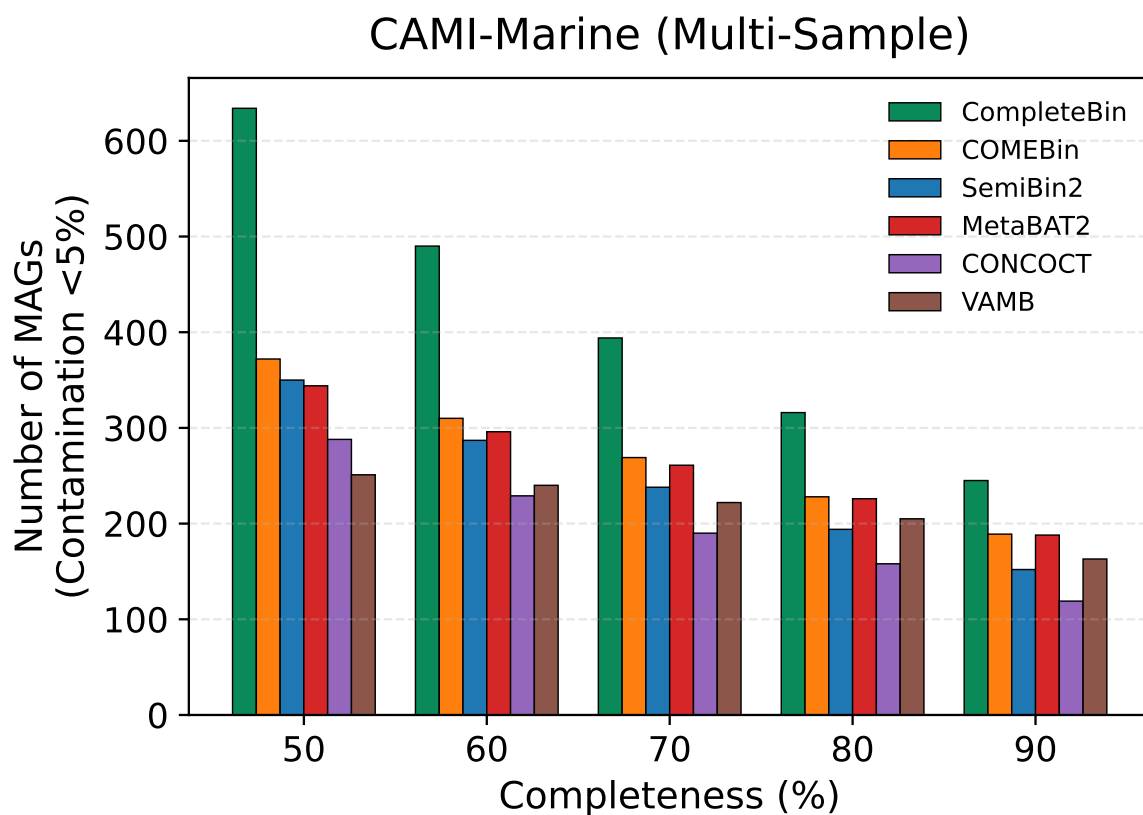
Supplementary Figure 5: Number of MAGs exceeding F1 score (bp) thresholds of 0.5, 0.6, 0.7, 0.8, and 0.9 for CompleteBin, COMEBin, SemiBin2, and MetaBAT2 on the CAMI II plant dataset.



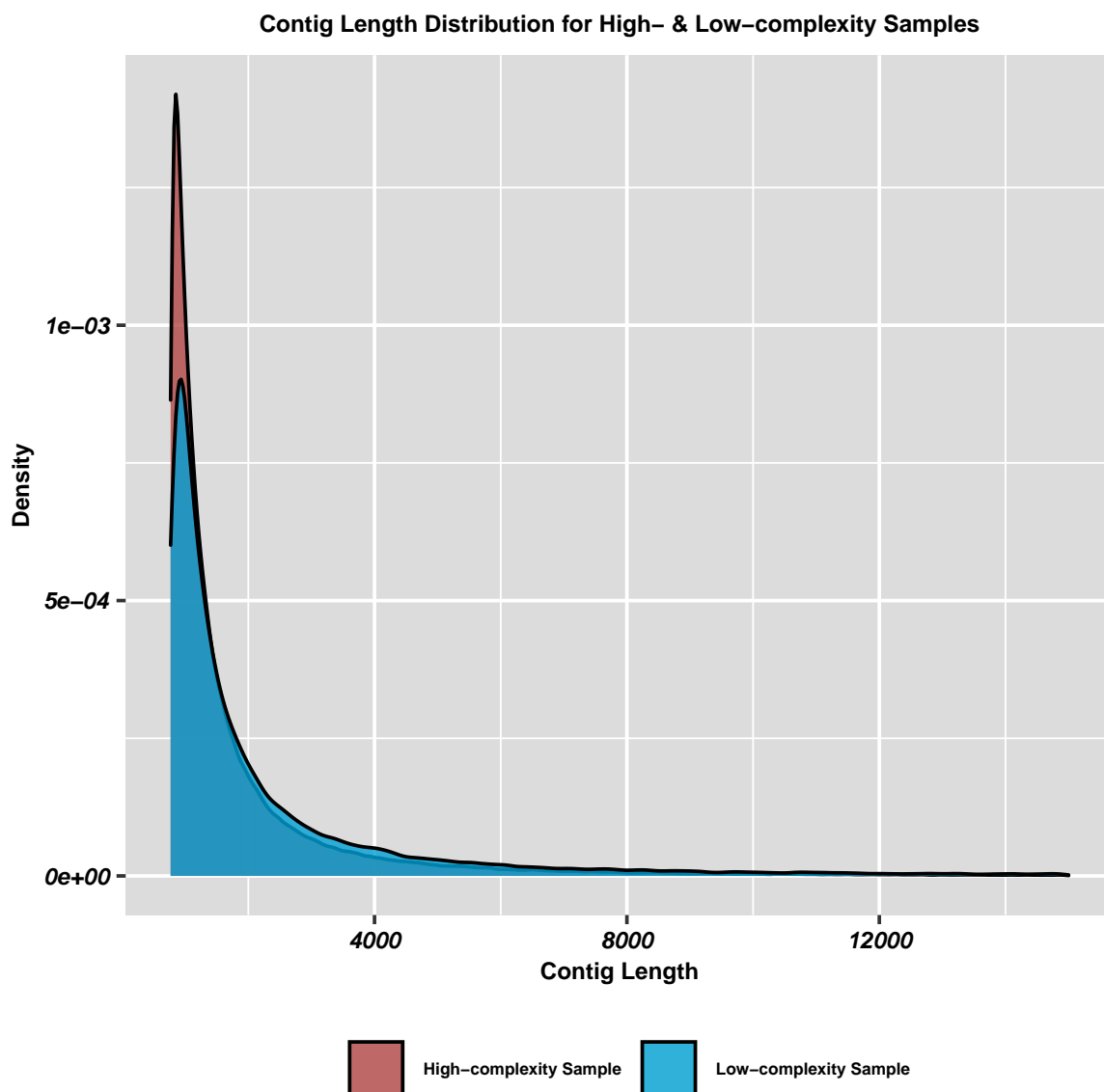
Supplementary Figure 6: Comparison of binning methods on the cold seep sediments (a-f), marine (g-i), freshwater (j-l), plant-associated (m), and human fecal (n-p) sequencing samples. The number of recovered MAGs with contamination thresholds of $\leq 10\%$ is shown for varying completeness thresholds.



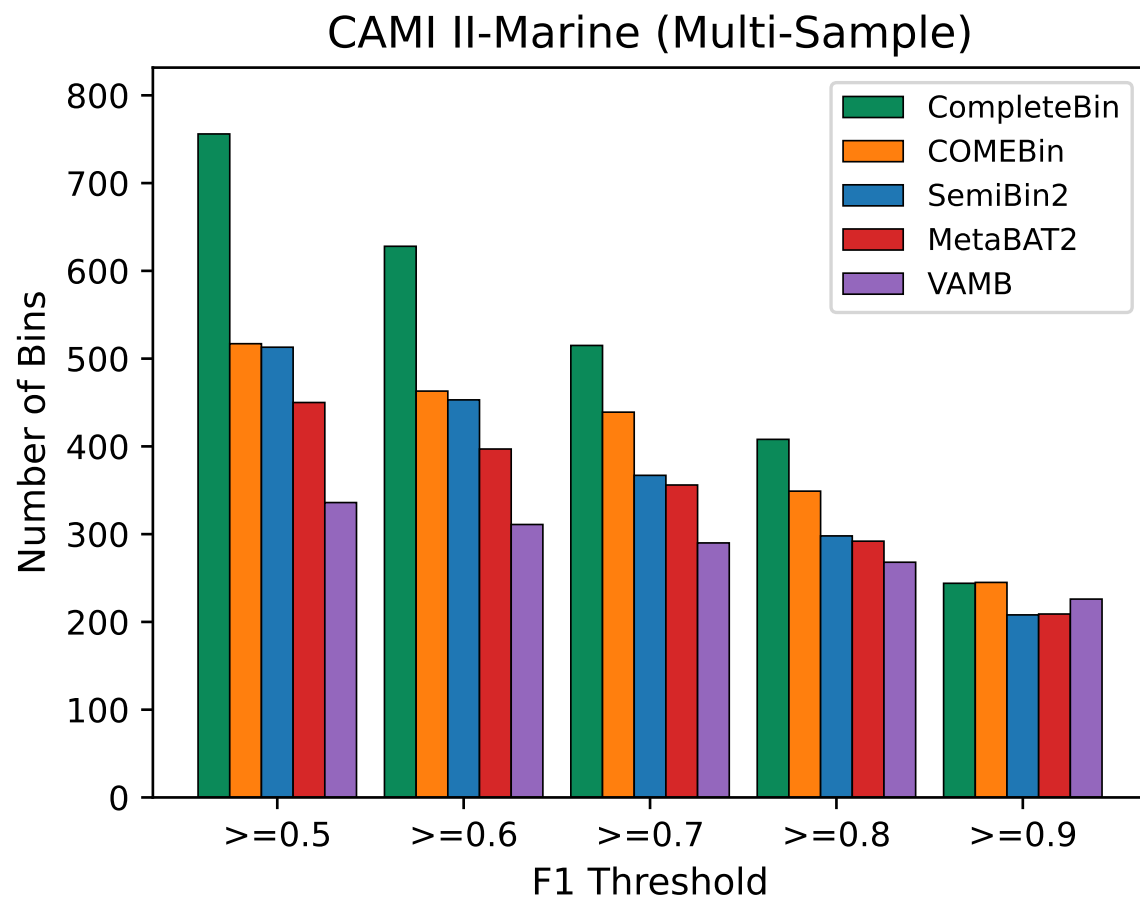
Supplementary Figure 7: Comparison of binning methods on the three freshwater (**a-c**), and six cold seep sediment (**d-i**) sequencing samples with multi-sample binning mode. The number of recovered MAGs with contamination thresholds of $\leq 10\%$ is shown for varying completeness thresholds.



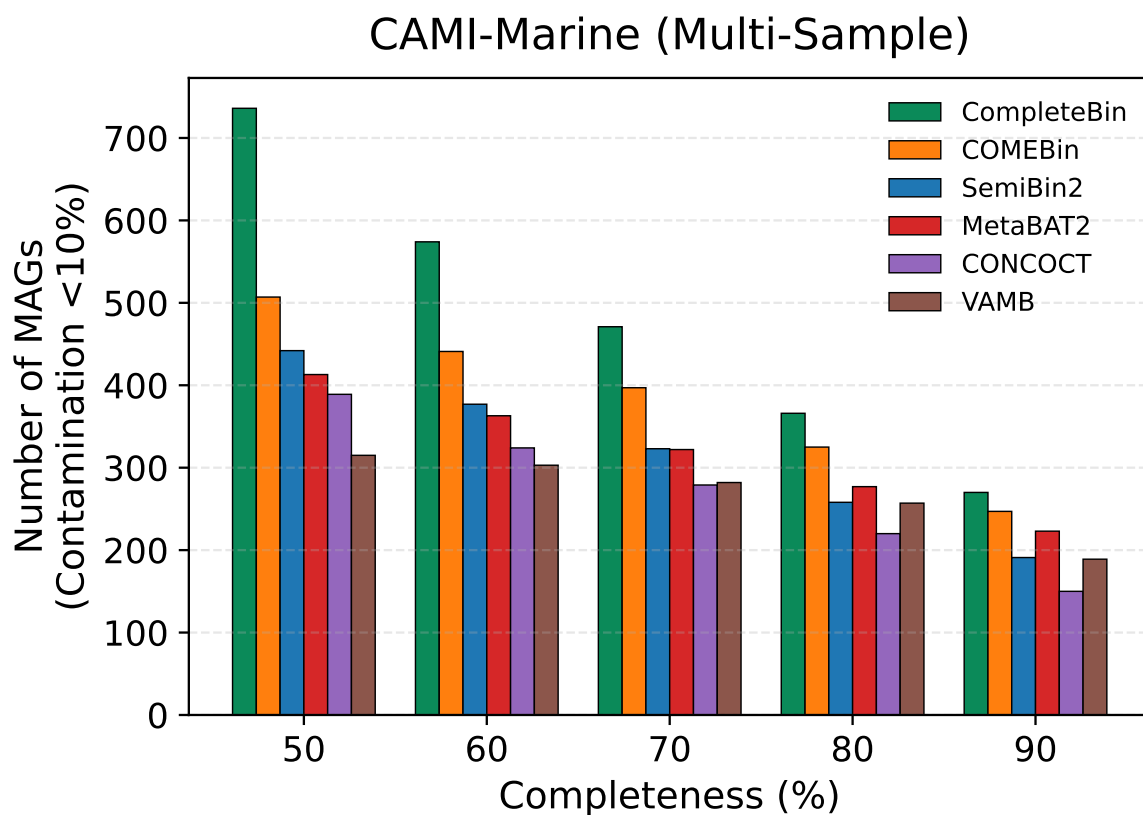
Supplementary Figure 8: Comparison of binning methods on the CAMI II marine dataset with multi-sample binning mode. The number of recovered MAGs with contamination thresholds of $\leq 5\%$ is shown for varying completeness thresholds.



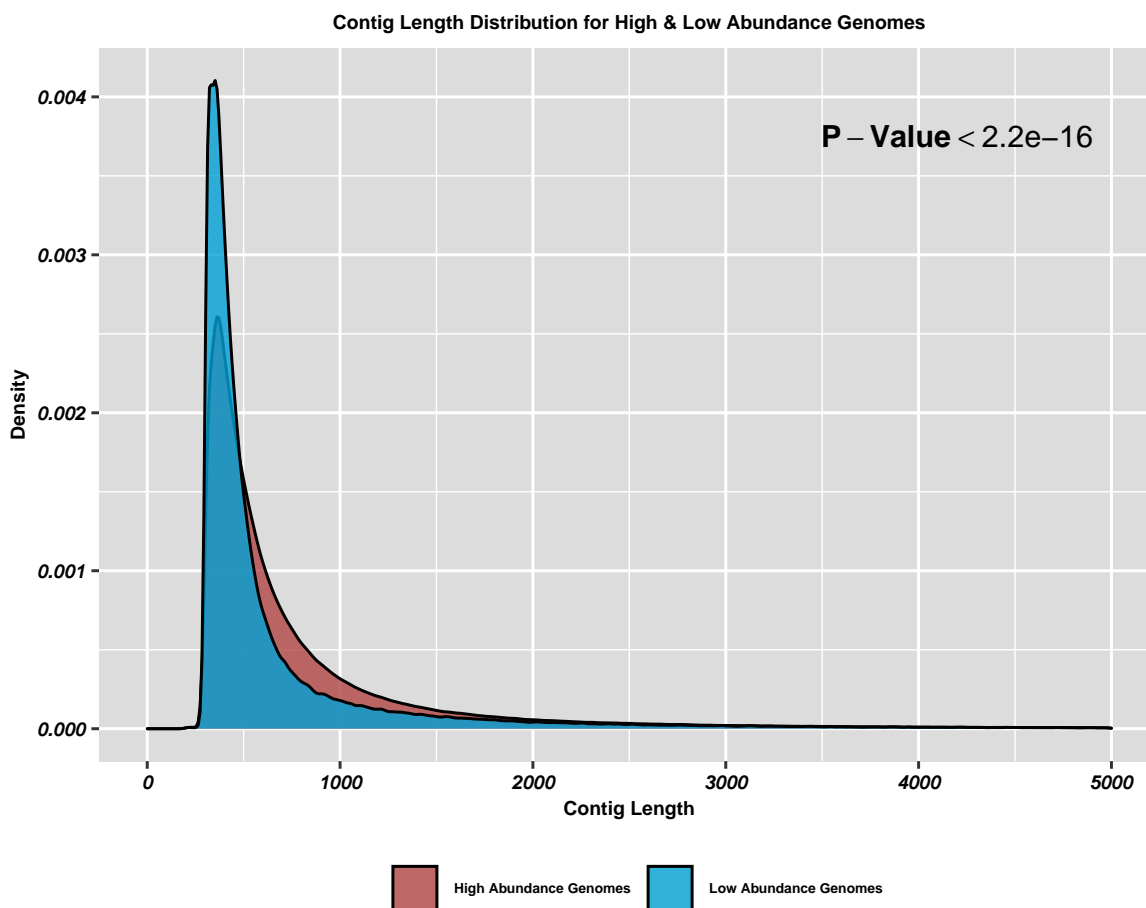
Supplementary Figure 9: The distribution of contig length for high- and low-complexity samples after assembly.



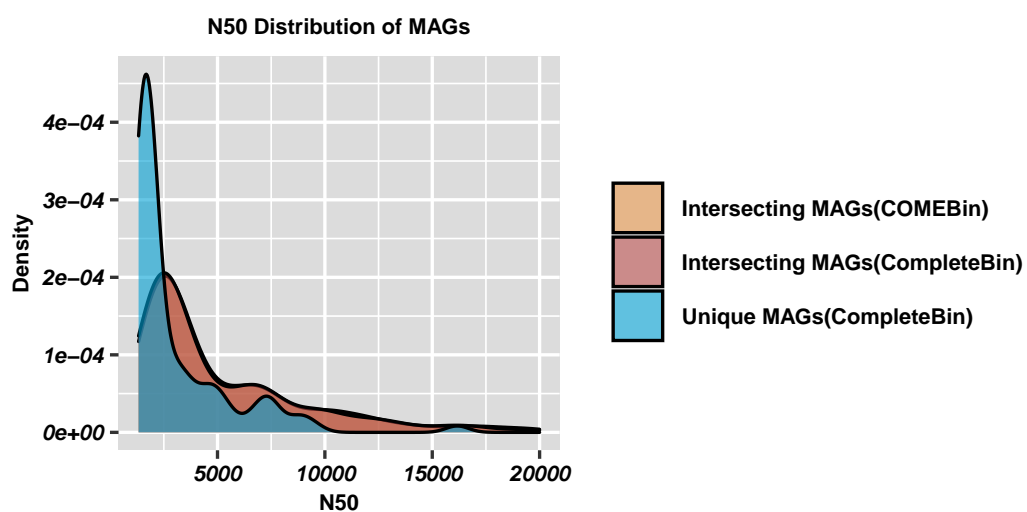
Supplementary Figure 10: Number of MAGs exceeding F1 score (bp) thresholds of 0.5, 0.6, 0.7, 0.8, and 0.9 for CompleteBin, COMEBin, SemiBin2, MetaBAT2, and VAMB on the CAMI II marine dataset with multi-sample binning mode.



Supplementary Figure 11: Comparison of binning methods on the CAMI II marine dataset with multi-sample binning mode. The number of recovered MAGs with contamination thresholds of $\leq 10\%$ is shown for varying completeness thresholds.



Supplementary Figure 12: The contig length distribution for the high- and low-abundance reference genomes after assembly.



Supplementary Figure 13: N50 distributions of MAGs across three categories: MAGs from genomes identified by both binning approaches (intersecting MAGs from COMEBin and CompleteBin), and MAGs from genomes exclusively recovered by CompleteBin (unique MAGs (CompleteBin)).

111 **References**

- 112 [1] Duncan, A. *et al.* Metagenome-assembled genomes of phytoplankton microbiomes from the arctic
113 and atlantic oceans. *Microbiome* **10**, 67 (2022).
- 114 [2] LJPvd, M. & Hinton, G. Visualizing high-dimensional data using t-sne. *J Mach Learn Res* **9**, 9
115 (2008).