Improving bacterial metagenomic research through long read sequencing

Noah Greenman[1], Sayf Al-Deen Hassouneh[1], Catherine Johnston[1], Taj Azarian[1, *]

1 College of Medicine, University of Central Florida, Orlando, Florida, USA

* Correspondence: taj.azarian@ucf.edu

Abstract

Metagenomic sequencing analysis is central to investigating microbial communities in clinical and environmental studies. Short read sequencing remains the primary data type for metagenomic research, however, long read sequencing promises advantages of improved metagenomic assembly and resolved taxonomic identification. To assess the comparative performance of short and long read sequencing data for metagenomic analysis, we simulated short and long read datasets using increasingly complex metagenomes comprised of 10, 20, and 50 microbial taxa. In addition, an empirical dataset of paired short and long read data from mouse fecal pellets was generated to assess feasibility. We compared metagenomic assembly quality, taxonomic classification capabilities, and metagenome-assembled genome recovery rates for both simulated and real metagenomic sequence data. We show that long read sequencing data significantly improves taxonomic classification capabilities and assembly quality. For simulated long read datasets, metagenomic assemblies were completer and more contiguous with higher rates of metagenome-assembled genome recovery. This resulted in more precise taxonomic classifications. Analysis of empirical data demonstrated that sequencing technology directly affects compositional results. Overall, we highlight strengths of long read sequencing for metagenomic studies of microbial communities over traditional short read approaches.

Keywords:
Metagenomics, Shotgun Metagenomic Sequencing, Next-generation Sequencing, Third generation sequencing, Taxonomic classification, Metagenome-assembled genomes, Simulation

Introduction

Metagenomics has enabled the study of microbial communities for agricultural, ecological, and clinical applications without the need for isolation or lab cultivation of bacteria (Chen and Pachter 2005; Cheng et al. 2019; Song et al. 2022; Suttner et al. 2020). Advancement in the field has been driven by the continued improvement of sequencing technologies, which have enabled generation of high-quality and low-cost genomic data for studying microbial populations with greater resolution (Wallen et al. 2022). Currently, the most widely used next-generation sequencing (NGS) technology are short-read sequencers, such as those from Illumina™. Short-read sequencing platforms generate 75-300 bp single or paired-end reads with per-base accuracy estimated at 99.9% (Latorre-Pérez et al. 2020). Before the recent reduction in cost attributed to the widespread availability of NGS technology, so-called microbiome studies often employed metabarcoding, where small portions of the 16S rRNA gene is sequenced for taxonomic classification (Begmatov et al. 2022; Oyserman et al. 2022; Sun et al. 2022). Although effective for genus level classification, metabarcoding provides limited species level resolution (Gupta et al. 2019). Where metabarcoding uses small regions of a single gene, shotgun metagenomics involves the sequencing of whole genome fragments, which can be used to reconstruct partial or whole microbial genomes through reference-based or *de novo* assembly approaches. Offering greater taxonomic resolution, *de novo* assembly is particularly useful for the identification of species that are non-culturable or lack robust representation in genomic databases. Shotgun metagenomic sequencing can also characterize variation in gene content, single nucleotide polymorphisms, and mobile genetic elements (Berglund et al. 2019; Sanderson et al. 2020).

In recent years, 3rd generation "long-read" sequencing platforms such as those from Oxford Nanopore Technologies™ (ONT) and PacBio have increasingly been applied to clinical and biomedical research (Charalampous et al. 2019; Sanderson et al. 2020). These platforms use a single-molecule sequencing approach that generates significantly longer DNA sequences (Petersen et al. 2019). A major strength of 3rd generation sequencers is their ability to span regions of genomes that are difficult for short-read sequencers to resolve, including common repetitive elements and other low complexity regions (Koren and Phillippy 2015). Functional analyses are further improved through the use of long reads as they can capture whole genes and preserve genetic structures (Singleton et al. 2021). Although there are numerous benefits, some notable drawbacks of long-read sequencing include lower per base accuracy. However, recent versions promise improvement of accuracy to 99% (Amarasinghe et al. 2020; Sereika et al. 2022). In addition, available tools for long-read data analysis are relatively limited when compared to tools for analysis of short-read data. More recently, the availability of tools has improved, with the number long-read sequencing programs more than quadrupling since 2018 (Amarasinghe et al., 2021, 2023).

Due to its high accuracy, wide-spread availability, and well-established published pipelines, short-read sequencing is typically employed over long read sequencing for metagenomic studies (Govender et al. 2021). However, the advantages afforded by long-read

sequencing makes it an approach particularly well-suited for metagenomic research. In this study, we sought to investigate the relative performance of short- and long-read sequencing for metagenomic analysis. To this end, we simulated metagenomic data from synthetic microbial communities of increasing complexity and measured precision, recall, and F-score. We then applied paired short- and long-read sequencing in parallel to mouse fecal samples, and empirically compared the relative performance.

Materials and Methods

Simulated metagenomes of synthetic microbial communities

Nine data sets of simulated metagenomic sequencing reads were generated (Figure 1). Each set consisted of 10 unique, simulated metagenomic samples of varying complexity. Complexity was defined by the number of comprising taxa, the relative abundance of organisms in a metagenome, and the presence or absence of simulated sequencing errors. The size of a simulated metagenome consisted of genomes from 10, 20, or 50 taxa with abundances either evenly distributed or randomly generated and simulated sequencing errors present or absent. The three types of sequence data included "Perfect", "Uneven", and "True" datasets. Perfect datasets consisted of reads with no errors and even organism abundance distribution, uneven datasets without errors and random abundance values for each organism, and true datasets with simulated errors specific to each sequencing platform for a given sequence type and random abundance values for each organism. Simulated long-reads were generated to reflect sequence data from ONT's MinION™ and short-reads mirrored sequence data from Illumina's™ HiSeq™ sequencer.

Construction of synthetic metagenomes was performed using an assembly structure report downloaded from NCBI (Downloaded 04/18/2022; Figure S1). This report contained all bacteria present in RefSeq with the following criteria for inclusion: 1) Genomes must be part of the latest RefSeq database, 2) they must have a complete genome assembly, and 3) all anomalous assemblies were excluded.

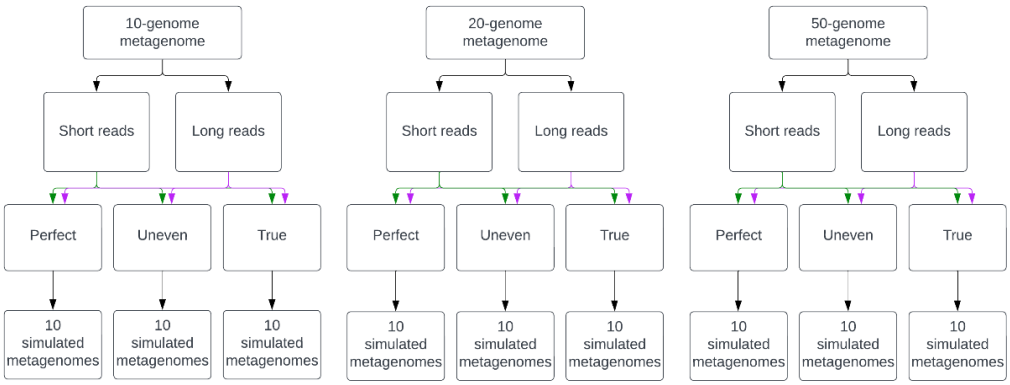


Figure 1. **Simulation of metagenomic datasets of increasing complexity. "**Perfect" metagenomic samples were simulated with no sequencing errors and an even abundance distribution of organisms. "Uneven" samples with no sequencing errors and randomly assigned abundance values for organisms in the metagenome, and "True" datasets with sequencing errors specific to each read type based on the platform they would be generated from and randomly assigned abundance values for members of each metagenome.

Simulation of ONT sequence data was performed using NanoSim version 3.0.0 (Yang et al. 2021). NanoSim input files (genome, abundance, and DNA lists) were created using a custom python script specifying parameters according to the level of metagenome complexity. The genome list (-gl) contained randomly selected taxa from the assembly structure report. The number of taxa selected was based on the desired size of the metagenome: 10, 20, or 50. Similar organisms could be present in a metagenome, such as different strains of the same species. The abundance list (-al) contained each organism and a value defining the abundance of that organism in the metagenome. This value was either equal across all organisms or randomly generated such that the total abundance of all organisms was 100%. The DNA list (-dl) contained all genetic molecules present in each organism's genome. This included an organism's chromosome and any plasmids, as well as whether the molecule was circular or linear. For simulating errors (-c), the pre-trained model 'metagenome_ERR3152364_Even' supplied by NanoSim was used. The basecaller (-b) specified was *guppy*. The '--perfect' option was specified for datasets without simulated errors, including "Perfect" and "Uneven" datasets. This option was not used for "True" datasets. All simulated reads were generated in FASTQ format (--fastq). The total number of reads generated per metagenome was 650,000, with reads having an N50 of approximately 5.2 Kbps. Total bases for each sample were approximately 2.67 Gbps. These values were selected to mimic realistic outputs from nanopore sequencing. For long reads, filtering was performed using Filtlong version 0.2.1 (R. Wick 2021). Long reads were filtered by minimum read length of 1000 bp (--min_length 1000).

Simulation of short read sequence data from an Illumina™ HighSeq™ was done using the randomreads.sh script from BBMap version 37.93 (Bushnell 2020). Ten million paired-end reads were generated for each simulated metagenome. When creating metagenomic short read data, the abundance of each genome was determined from the abundance list used for simulating long reads. Dividing 10 million by the abundance for each genome yielded the number of short reads to simulate. Each iteration of randomreads.sh took a single genome and the number of reads specific to that genome and output a FASTQ file containing interleaved paired-end short reads. All datasets consisted of paired-end reads 150 bp in length (paired=t, len=150), named according to their genomic origin and without coordinates (simplenames=t), read quality was set to 36 (q=36), and a single insert size distribution was specified (superflat=t). Error rates were controlled by specifying the maximum number of SNPs (maxsnps), insertions (maxinss), insertion rate (insrate), deletions (maxdels), deletion rate (delrate), substitutions (maxsubs), and substitution rate (subrate). For "Perfect" and "Uneven" datasets, all values for controlling errors were set to 0 and an additional argument controlling inclusion of errors was set to false (adderrors=f). "True" datasets used the following values for error-controlling parameters: maxsnps=3, maxinss=2, maxdels=2, maxsubs=2, insrate=0.00000315, delrate=0.000005, subrate=0.0033, adderrors=t. The chosen error rates are based off average error rates observed by Schrimer and colleagues in their paper on Illumina™ error profiles from metagenomic data (Schirmer et al. 2016). Simulated reads for each genome of a metagenome were concatenated. Then, to induce random ordering of reads, the tool shuffle.sh from BBMap was used with default parameters. Read pairs were then separated into two FASTQ files using BBMap's reformat.sh tool, specifying the single, shuffled, interleaved file be split into R1 and R2 FASTQ files. No filtering was performed on short reads as all reads possessed a q-score of 36.

Metagenome assembly and quality assessment

Long read assembly was performed using Flye version 2.9-b1768 (Kolmogorov et al. 2020). Default parameters for Flye were used with the addition of '--nano-raw' and '--meta' to denote assembly of uncorrected reads from a metagenomic sample. SPAdes version 3.15.5 was used for metagenomic assembly of simulated short read data (Nurk et al., 2017). Default options for SPAdes were used except for the following: '--meta' was specified to indicate the sequence data were from a metagenomic sample, kmer sizes (-k) consisted of a list of integers: 21, 33, 55, 77, 99, 127, '--phred-offset' was set to 33 due to simulated short reads not possessing defined signatures of the PHRED quality, and '--only-assembler' was specified to disable read error correction. Metagenomic assembly quality was evaluated using metaQUAST version 5.0.2 (Mikheenko, Saveliev, and Gurevich 2016). Default parameters were used except for supplying the directory containing the reference genomes for each metagenome to '-r'.

Metagenome-assembled genome recovery and taxonomic classification

Metagenome-assembled genomes (MAGs) were generated using the following steps. First, long reads were mapped back to the assemblies using minimap2 version 2.24-r1122. For long reads, the flag '-x' was given the 'map-ont' argument to optimize alignment parameters for nanopore sequence data (Li 2018). Short read headers were shortened using the tool seqtk-1.4 version r122's 'rename' function (Li, 2023). Short reads were then mapped using bwa-mem2 version 2.2.1 by first creating an index from a metagenome's assembly file using 'bwa-mem2 index', and then mapping the reads back using 'bwa-mem2 mem'. Each alignment process produced an unsorted SAM file which was then converted into sorted BAM files using Samtools version 1.16.1 with options 'samtools view -b' for creating the BAM file, and 'samtools sort' for sorting the BAM file (Danecek et al. 2021). A coverage file was generated for each read type using CoverM version 0.6.1 using the default '-m mean' for coverage estimation (Woodcroft 2021). For long reads, the mapper option for CoverM was set to '-p minimap2-ont', and for short reads it was '-p bwa-mem'. Each coverage file was used to bin each read type's respective assemblies using MetaBAT 2 version 2.15 with default parameters except for the addition of the '--cvExt' argument to denote coverage files originated from a third-party tool (Kang et al. 2019). Binned contigs were then assessed for completeness and contamination using CheckM2 version 1.0.1 with default parameters (Chklovski et al. 2022). Taxonomic classification was performed using Kraken2 version 2.1.2 for both short and long read metagenomic assemblies (Wood, Lu, and Langmead 2019). Default parameters were used with the '--use-names' flag included. The kraken2 report and output files were kept for further analysis.

Result evaluation

For visual assessment of metagenome assembly quality, assembly graphs were visualized using Bandage version 0.8.1 (R. R. Wick et al. 2015). Resulting values for genome fraction recovery, NGA50, and number of misassemblies from metaQUAST were analyzed using the *statannotations* python package (Charlier et al., 2022). Statistical significance was evaluated using a Mann-Whitney U test with Bonferroni correction. A p-value threshold of 0.05 was used to determine significance.

Performance evaluation of taxonomic classification from reads and assemblies was accomplished by calculating precision, recall, and F-score. Precision was defined as the rate of successful taxonomic assignment of a read/contig to the correct taxon within the metagenome over the total number of taxonomic assignments. Also referred to as the false positive rate, the equation used was $Precision = \frac{TP}{(TP+FP)}$ where TP is the true positive rate and FP is the false positive rate, or an incorrect classification of a read/contig at a given taxonomic rank. Recall, also defined as the rate of false negatives, was determined using the equation $Recall = \frac{TP}{total}$ where 'total' is the expected number of distinct organisms at a specified taxonomic rank. F-score, or the harmonic mean, generates an overall score factoring in both precision and recall. F-score was calculated using the equation $F1 = \frac{2*(precision*recall)}{(precision+recall)}$. A statistical comparison of precision, recall, and F-score was performed using a one-way ANOVA for assemblies and Mann-Whitney U test with Bonferroni correction for reads with the *statannotations* python package. A p-value threshold of 0.05 was used to determine significance.

Estimation of relative abundance of taxa at the genus and species level was carried out using Bracken version 2.8 (Lu et al., 2017). Short and long reads were processed through Kraken2 to generate report files that were then used as inputs for Bracken. Bracken then returns a report file containing estimates of relative abundance. For each genome in a metagenome, the predicted abundance was plotted against the actual abundance in a scatterplot along with a linear regression line using a linear regression function from SciPy version 1.10.1 (Virtanen et al., 2020). R-values, slope, and p-values for each read type's line were reported.

MAG recovery rates were determined by comparing the total number of MAGs recovered by the binning of contigs from short and long read metagenomic assemblies. A MAG's quality is determined by the completeness and contamination values of contigs assigned to a bin by MetaBAT2 and assessed by CheckM2. CheckM2 defines completeness as the percentage of universal, bacteria-specific gene markers present in a given bin. Contamination for CheckM2 is based on how many of these markers exist in multiple copies which indicates the presence of contigs that do not belong in a particular bin. For the minimum threshold of a bin to be considered a MAG, the cutoffs were defined as a completeness ≥50% and contamination ≤10% (A. Murat Eren 2016). For each of the 9 sets of metagenomes, the average number of MAGs recovered from short and long read assemblies was recorded. A Mann-Whitney U Test with Bonferroni correction was applied using the *statannotations* python package to compare differences in MAG recovery rates between short and long read assemblies. A p-value threshold of 0.05 was used to determine significance. Bins determined to be MAGs were also evaluated on quality which was defined by completeness only. High quality MAGs had a completeness ≥90%, medium quality MAGs had a completeness ≥70% and <90%, and low-quality MAGs had a completeness ≥50% and <70%, respectively.

Comparison of experimental metagenomic data

To examine whether read type affects assessment of microbial composition, we generated short- and long-read sequencing data from 19 mouse fecal pellets collected during ongoing experimental studies. Long-read sequencing data were generated using the ONT GridION. Prior

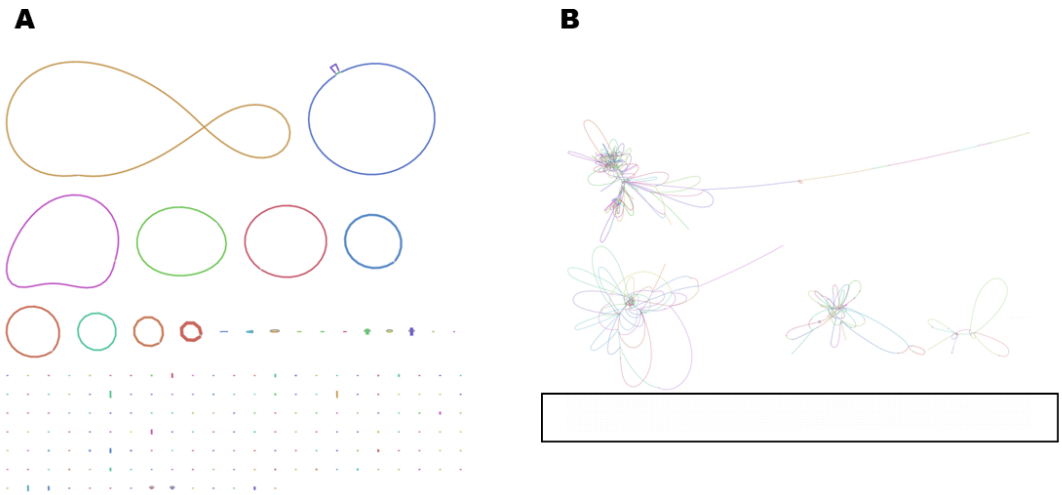
to DNA extraction, host DNA was depleted using an enzymatic approach as described previously with some modifications (Charalampous et al., 2019). Each sample extraction used 1-2 mouse fecal pellets. Fecal pellets were homogenized in 1 mL of InhibitEX buffer (Cat. No./ID: 19593, Qiagen, Hilden, Germany). After pelleting the homogenate and aspirating off the liquid, 250 µL of 1X PBS and 250 µL 4.4% saponin was added and mixed thoroughly before resting at room temperature for 10 minutes. Nuclease-free water (350 µL) along with 12 µL of 5M NaCl were added to induce osmotic lysis. After pelleting, the liquid was again aspirated off and the pellet was resuspended in 100 µL of 1X PBS. One-hundred microliters of HL-SAN enzyme buffer added (5.5 M NaCl and 100 mM MgCl2 in nuclease-free water), followed by 10 µL of HL-SAN endonuclease (Article No. 70910-202, ArcticZymes, Tromsø, Norway) before being incubated on a shaking incubator at 37°C for 30 minutes at 1000 rpm. Two final washes of the pellet were performed using 800 and 1000 µL of 1X PBS before proceeding to extraction. Genomic DNA (gDNA) was extracted and purified using the New England Biolabs Monarch® Genomic DNA Purification Kit (New England Biolabs, Ipswich, MA) following manufacturer instructions. Quality and concentration of gDNA was assessed using Agilent 4200 TapeStation System and Qubit 4 Fluorometer. Sequencing libraries for nanopore sequencing were prepared using the Rapid PCR Barcoding Kit (SQK-RPB004). Using an Oxford Nanopore GridION™, libraries were sequence on a R9.4.1 flow cell generating an average of 1.84 Gbps of sequence data with a mean read length of 3,491.62 bps and a mean read quality of 12.9. The longest read produced was 42,266 bps. Short read sequencing was performed by SeqCenter LLC (SeqCenter, Pittsburgh, PA) using an Illumina™ NovaSeq™ with Nextera Flex Library Preparation Kit® and V3 flow cell chemistry to produce an average of 2.53 Gbps of paired-end 2x150 bp reads across all samples. Illumina data were filtered using Trimmomatic v0.39 for paired-end reads with the following trimming settings: SLIDINGWINDOW:4:20 MINLEN:100.

Analysis followed the methodological approaches described previously for short- and long-reads, respectively. Reads were first classified using Kraken2 and relative abundances were estimated using Bracken. Before analysis, bracken reports were trimmed to remove all taxonomic identifications where the number of reads assigned to an organism fell below a given threshold of 1/20,000 (0.00005). To our knowledge, no studies using shotgun metagenomic sequencing have given the expected taxonomic diversity of the mouse gut microbiome at the genus or species level. The Murine Microbiome Database consists of 1732 genera and 4703 species, however these data come from 16S sequence analysis (Yang et al., 2019). An alternative database, the Comprehensive Mouse Microbiota Genome (CMMG), which derived 1,573 species from mouse gut metagenomes combined with known reference metagenome-assembled genomes at the time of its creation (Kieser et al., 2022). In this study, we empirically defined a cutoff where the count of unique taxa was closer to that of the CMMG, as 16S sequence often reports higher diversity. Initial plotting of the number unique taxa showed counts of over 1750 genera and over 6000 species for both short- and long-read data when no abundance cutoff was applied. Using a cutoff of 1/20,000 yielded unique taxa counts for species similar to that of the CMMG. Short- and long-read data for a group were then compared to each other using the beta_diversity.py script from the KrakenTools suite. Qualitative assessment was performed by generating a hierarchically clustered heatmap of the resulting Bray-Curtis Dissimilarity matrix. For quantitative analysis, a principal component analysis (PCA) was performed. First, data were

transformed using a center-log ratio transformation (Aitchison, 1986). Then, PCA was run using the PCA function from scikit-learn's decomposition library (Pedregosa et al., 2011). Briefly, sample IDs were separated as features and abundance values for each taxon at the genus and species level were retained. Abundance values for unique taxa were then given to the PCA object which created a fit and subsequently applied dimensional reduction to the abundance data. The resulting principal components were kept in a table from which the top two principal components were plotted.

Results

Comparison of metagenomic assembly completeness



Supplemental Figure 1. **Assembly graphs from a "True" 10-genome metagenome visualized by Bandage.** 2A. Bandage graph generated from metagenomic assembly of long read sequence data. 2B. Bandage graph generated from metagenomic assembly of short read sequence data. The black box highlights small, disconnected contigs.

Metagenomic assemblies for each synthetic dataset were qualitatively assessed through visualization (Figure S1). Long read assemblies produced graphs with greater contiguity and often resulted in circularization of unitigs, an indication they represented complete chromosome or, in the case of smaller, circularized contigs, plasmids.

Assemblies from long reads demonstrated significantly higher genome fraction coverage than their short read counterparts (Figure 2A, D, and G; Supplementary Table S1). This was observed across all metagenome sets, regardless of complexity, with the exception of the synthetic dataset composed of 50 organisms with perfect reads (Figure 2A). Similar results were observed when comparing assemblies by NGA50 values, with long read data showing

significantly higher scores (Figure 2B, E, and H; Supplementary Table S1). With "Perfect" data, short reads produced significantly fewer misassemblies, however this difference was not significant for larger "Uneven" and all "True" datasets (Figure 2C, F, and I; Supplementary Table S1).

Table S1. Results of Mann-Whitney U testing for genome fraction, NGA50, and number of misassemblies.

[1] "Perfect" indicates reads had no errors and abundances were evenly distributed. "Uneven" denotes reads had no errors, but variable abundances for each organism. "True" reads had simulated sequencing errors based on the read type and variable abundances for the organisms.

[2]P-value from Mann-Whitney U test with Bonferroni correction

simulated metagenome	complexity[1]	metaQUAST metric	statistic	p-value[2]
10-genome metagenome	perfect	genome fraction	7600.5	1.79E-10
10-genome metagenome	perfect	NGA50	8465	2.56E-17
10-genome metagenome	perfect	number of misassemblies	6473	5.46E-08
20-genome metagenome	perfect	genome fraction	27671.5	3.04E-11
20-genome metagenome	perfect	NGA50	30657.5	3.03E-20
20-genome metagenome	perfect	number of misassemblies	29583.5	1.24E-26
50-genome metagenome	perfect	genome fraction	134023.5	0.04813137
50-genome metagenome	perfect	NGA50	156277	7.40E-12
50-genome metagenome	perfect	number of misassemblies	194570	3.23E-67
10-genome metagenome	uneven	genome fraction	7433	2.59E-09
10-genome metagenome	uneven	NGA50	8152	1.35E-14
10-genome metagenome	uneven	number of misassemblies	6008	0.0044121
20-genome metagenome	uneven	genome fraction	28562.5	1.17E-13
20-genome metagenome	uneven	NGA50	32254	2.99E-26
20-genome metagenome	uneven	number of misassemblies	20066	0.94863897
50-genome metagenome	uneven	genome fraction	137225	0.00742218
50-genome metagenome	uneven	NGA50	169597	1.32E-22
50-genome metagenome	uneven	number of misassemblies	133230.5	0.0645971
10-genome metagenome	true	genome fraction	7812	5.66E-12
10-genome metagenome	true	NGA50	8341	3.26E-16
10-genome metagenome	true	number of misassemblies	5768.5	0.01977121
20-genome metagenome	true	genome fraction	26882.5	2.51E-09
20-genome metagenome	true	NGA50	29641.5	7.20E-17
20-genome metagenome	true	number of misassemblies	20560	0.583517
50-genome metagenome	true	genome fraction	138987.5	0.00218585
50-genome metagenome	true	NGA50	170176	3.94E-23
50-genome metagenome	true	number of misassemblies	122815	0.62025684

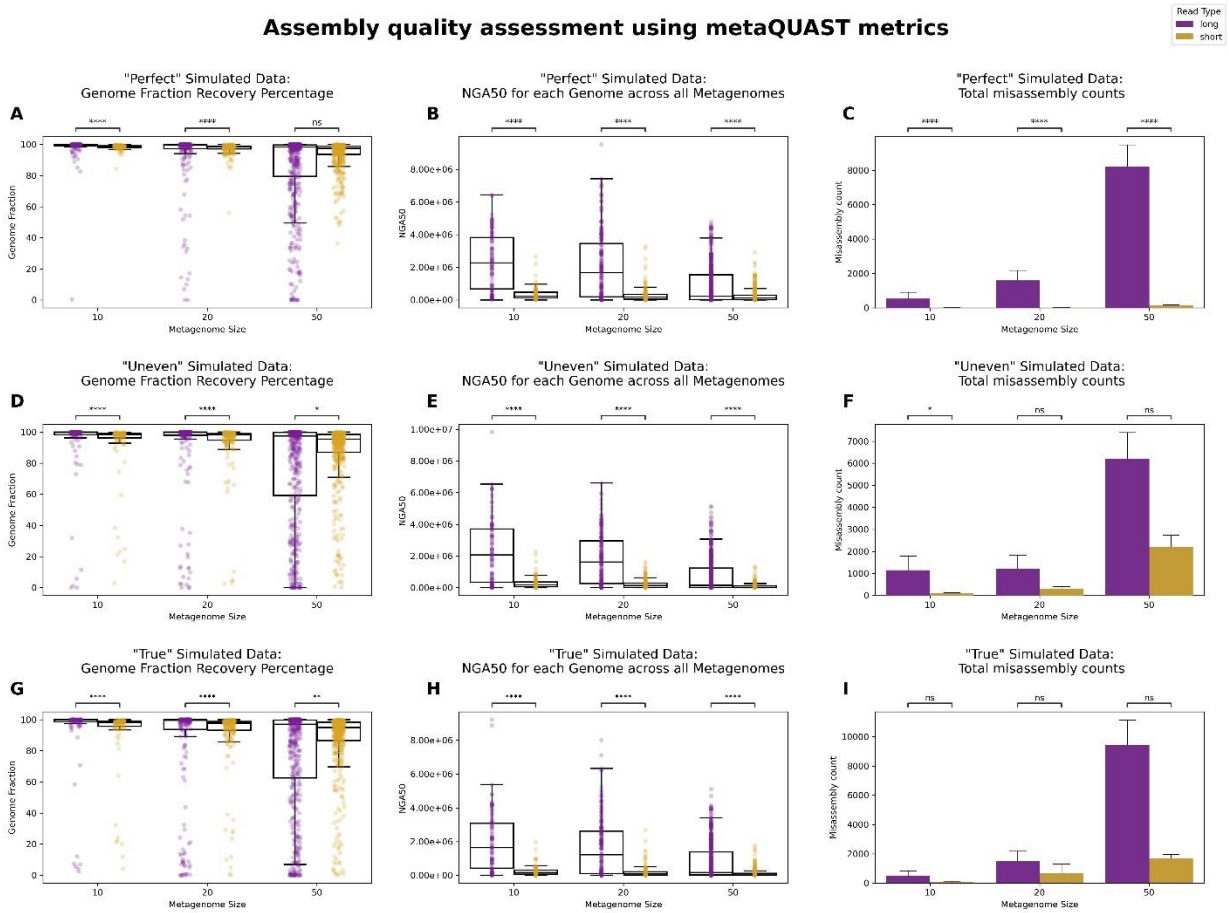


Figure 2. **Comparisons of assembly quality metrics.** Boxplots depicting differences in genome fraction values (A, D, G) and NGA50 (B, E, H) from all genomes in a set of metagenomes. Bar plots depicting total misassembly counts across all genomes in a set of metagenomes (C, F, I). "Perfect" denotes simulated data have no errors and abundances are evenly distributed for each organism in the metagenome. "Uneven" denotes simulated data have no errors, but variable abundances for each organism. "True" denotes simulated data have simulated sequencing errors based on the read type, as well as variable abundances for each organism. Significance was calculated using a Mann-Whitney U Test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

Evaluating taxonomic classification

Evaluation of taxonomic classifications using assemblies at the genus and species-level was assessed for precision, recall, and F-score. Neither read types significantly differed when measuring recall of taxa (Figure 3B, E, and H; Figure 4B, E, and H). However, long reads significantly outperformed short reads in precision across almost all levels of metagenome complexity at both the genus and species level (Figure 3A, D, and G; Figure 4A, D, and G). The F-score, or harmonic mean of the precision and recall, was calculated to measure the accuracy of classification results using either short or long reads. Long reads demonstrated significantly

higher F-scores than short reads across almost all levels of complexity, indicating that long reads provide greater accuracy for taxonomic classification.

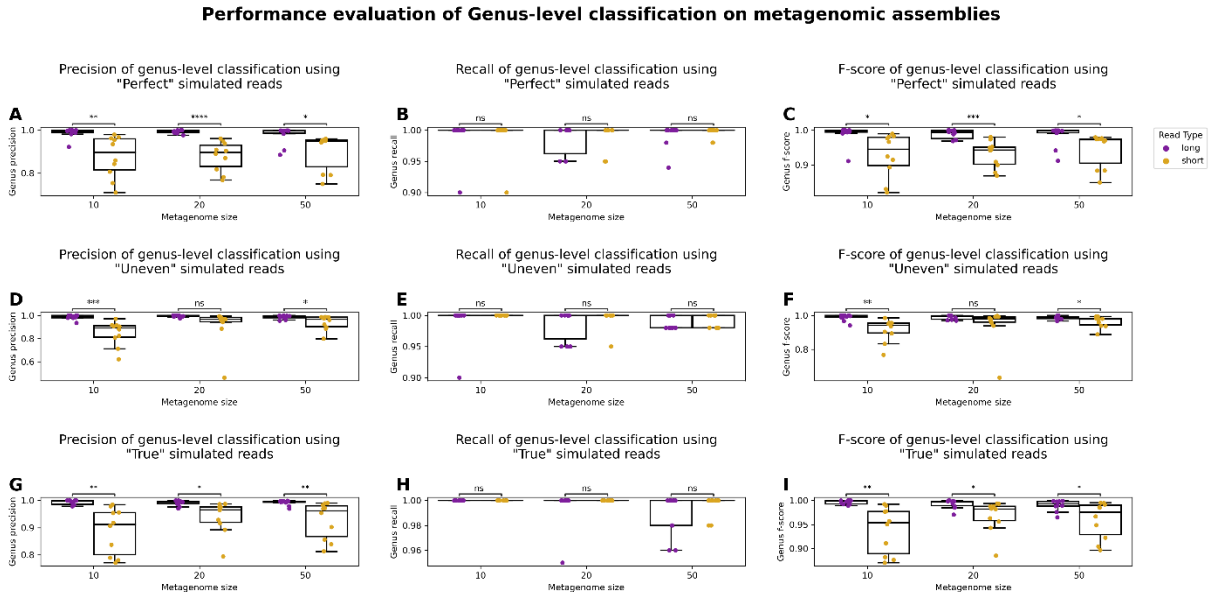


Figure 3. **Performance evaluation of genus-level classification on metagenomic assemblies using short or long reads.** Boxplots showing precision, recall, and F-score metrics for metagenomes of varying size and complexity. "Perfect" simulated reads have no errors and even abundances across organisms (A, B, C), "Uneven" have no errors and randomly varied abundances across organisms (D, E, F), and "True" have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a One-way ANOVA. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

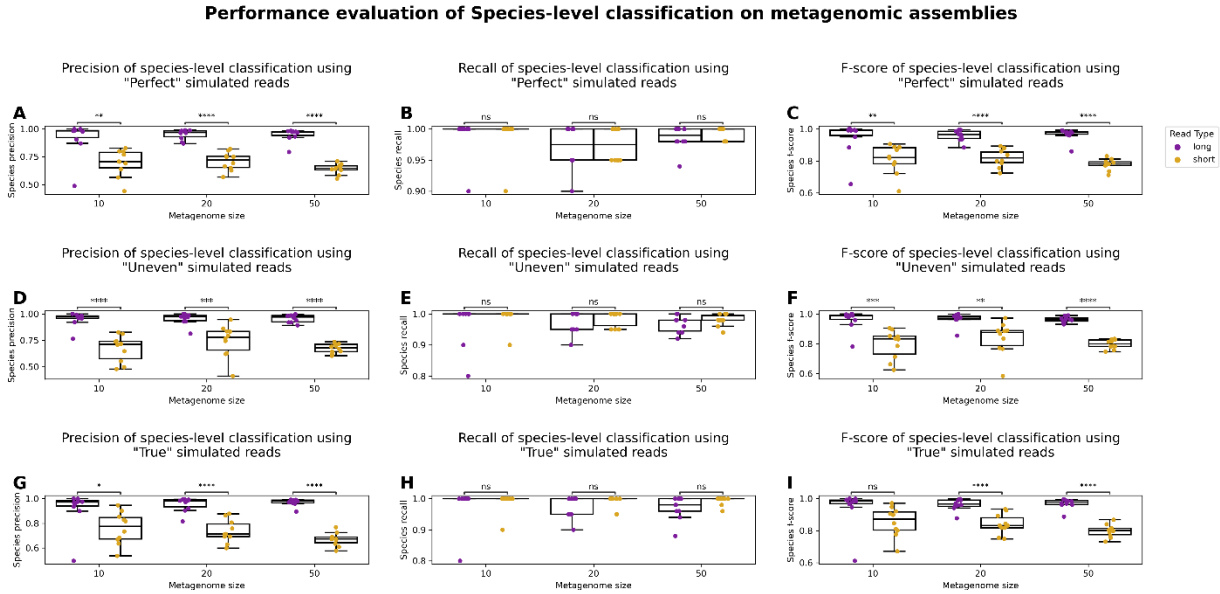
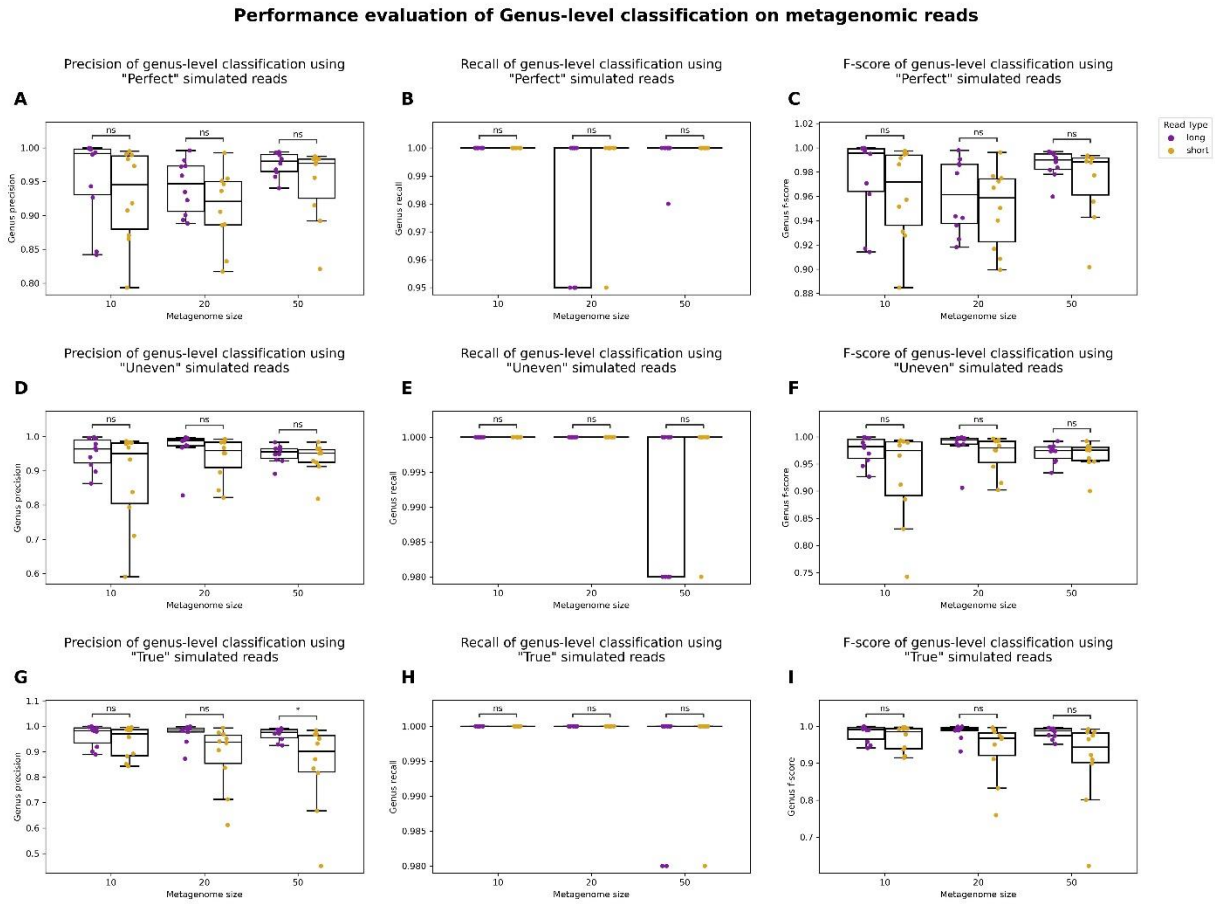


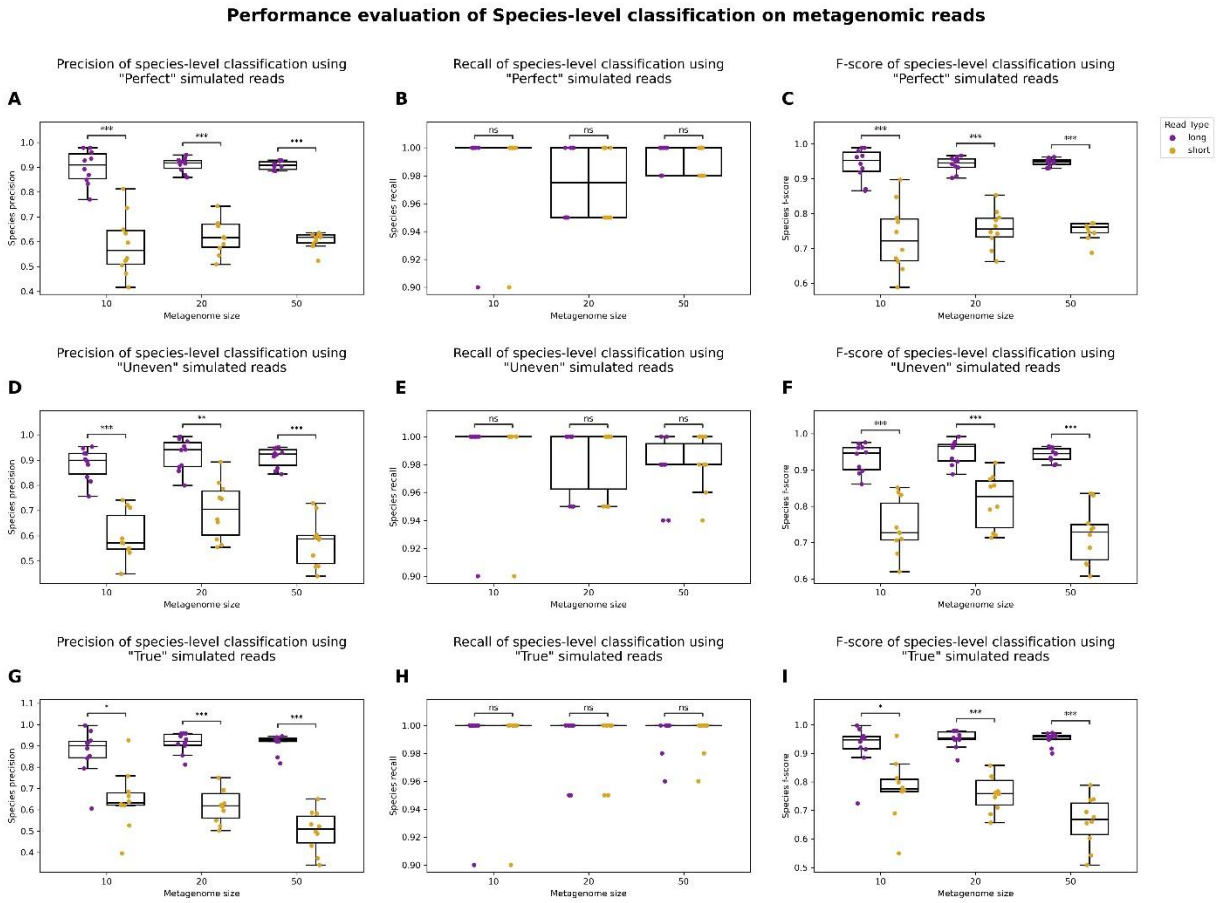
Figure 4. **Performance evaluation of species-level classification on metagenomic assemblies using short or long reads.** Boxplots showing precision, recall, and F-score metrics for metagenomes of varying size and complexity. "Perfect" simulated reads have no errors and even abundances across organisms (A, B, C), "Uneven" have no errors and randomly varied abundances across organisms (D, E, F), and "True" have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a One-way ANOVA. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

Estimating relative abundance using short and long read data

Before analyzing the performance of each read type when used for relative abundance estimation, the classification capabilities of the reads were first assessed. Precision, recall, and F-score of genus-level classification was not significantly affected by read type in most cases, which differed from classification using metagenomic assemblies (Supplementary Figure S2). However, the performance for each read type on species-level classification was significantly improved when using long reads (Supplementary Figure S3). Both precision and F-score were higher for classifications using long reads and recall was not significantly different between the two. Examining accuracy in estimation of relative abundance found long reads outperformed short reads in most cases at the genus and species level (Figure S4; Figure 5).



Supplemental Figure 2. **Performance evaluation of genus-level classification using short and long reads.** Boxplots showing precision, recall, and F-score metrics for metagenomes of varying complexity. "Perfect" simulated reads have no errors and even abundances across organisms (A, B, C), "Uneven" have no errors and randomly varied abundances across organisms (D, E, F), and "True" have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a Mann-Whitney U test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.



Supplemental Figure 3. **Performance evaluation of species-level classification using short and long reads.** Boxplots showing precision, recall, and F-score metrics for metagenomes of varying complexity. "Perfect" simulated reads have no errors and even abundances across organisms (A, B, C), "Uneven" have no errors and randomly varied abundances across organisms (D, E, F), and "True" have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a Mann-Whitney U test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

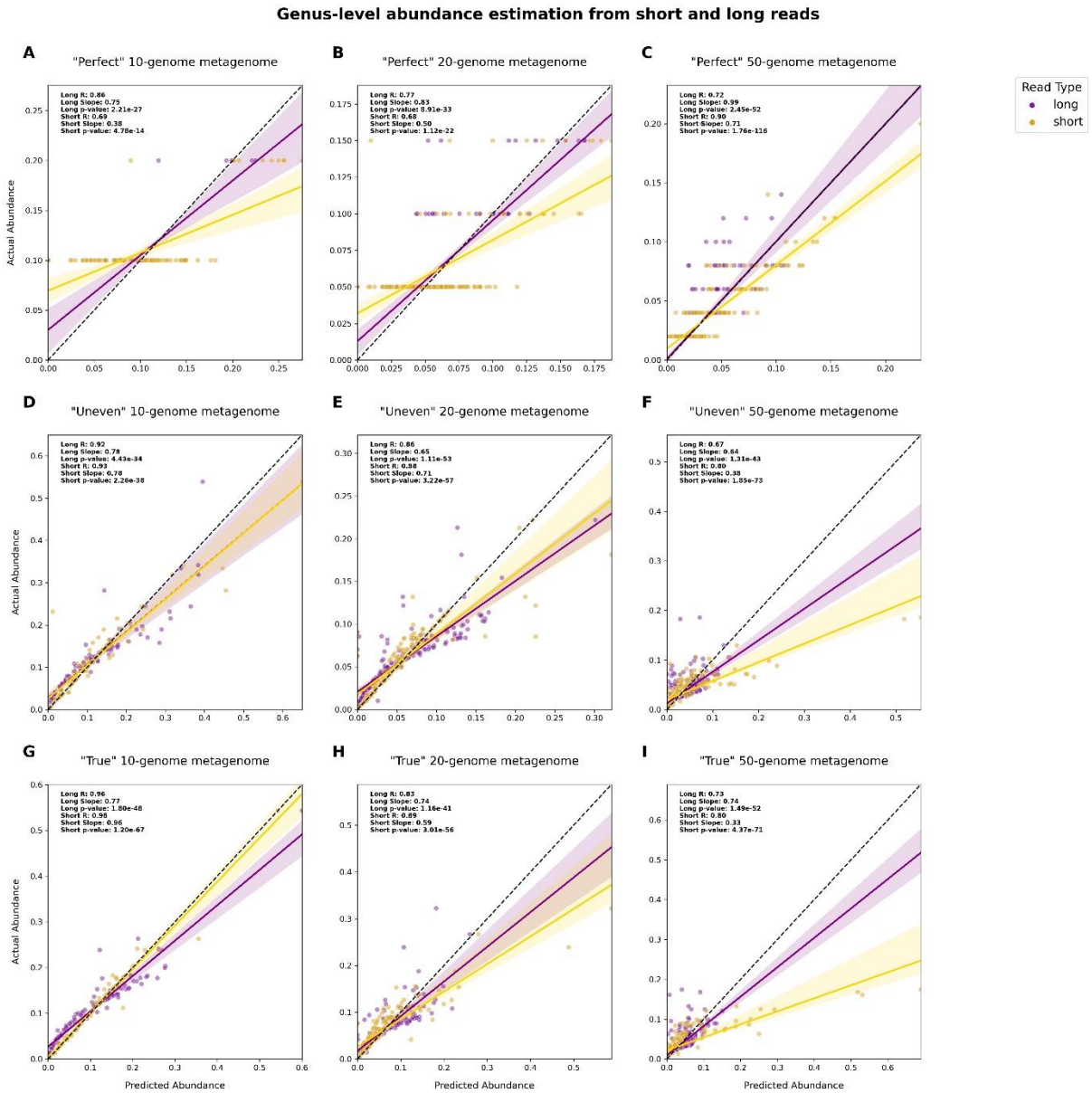


Figure S4. **Comparison of short and long read capacity for genus-level relative abundance estimation.** Scatterplots depicting predicted versus actual abundance values for genomes present in simulated metagenomes from short and long read data. Read types were compared across metagenomes of varying complexity at both the genus and species level. "Perfect" datasets consisted of reads without sequencing errors and organism abundance was evenly distributed (5A-C), "Uneven" consisted of reads without sequencing errors and randomly distributed abundances of organisms (5D-F), and "True" consisted of reads with simulated errors and randomly distributed abundances of organisms (5G-I). A linear regression line was plotted for each read type. Each line has its reported R-value, slope, and p-value. The dotted line represents the 1-to-1 line where values of predicted abundance match values of actual abundance.

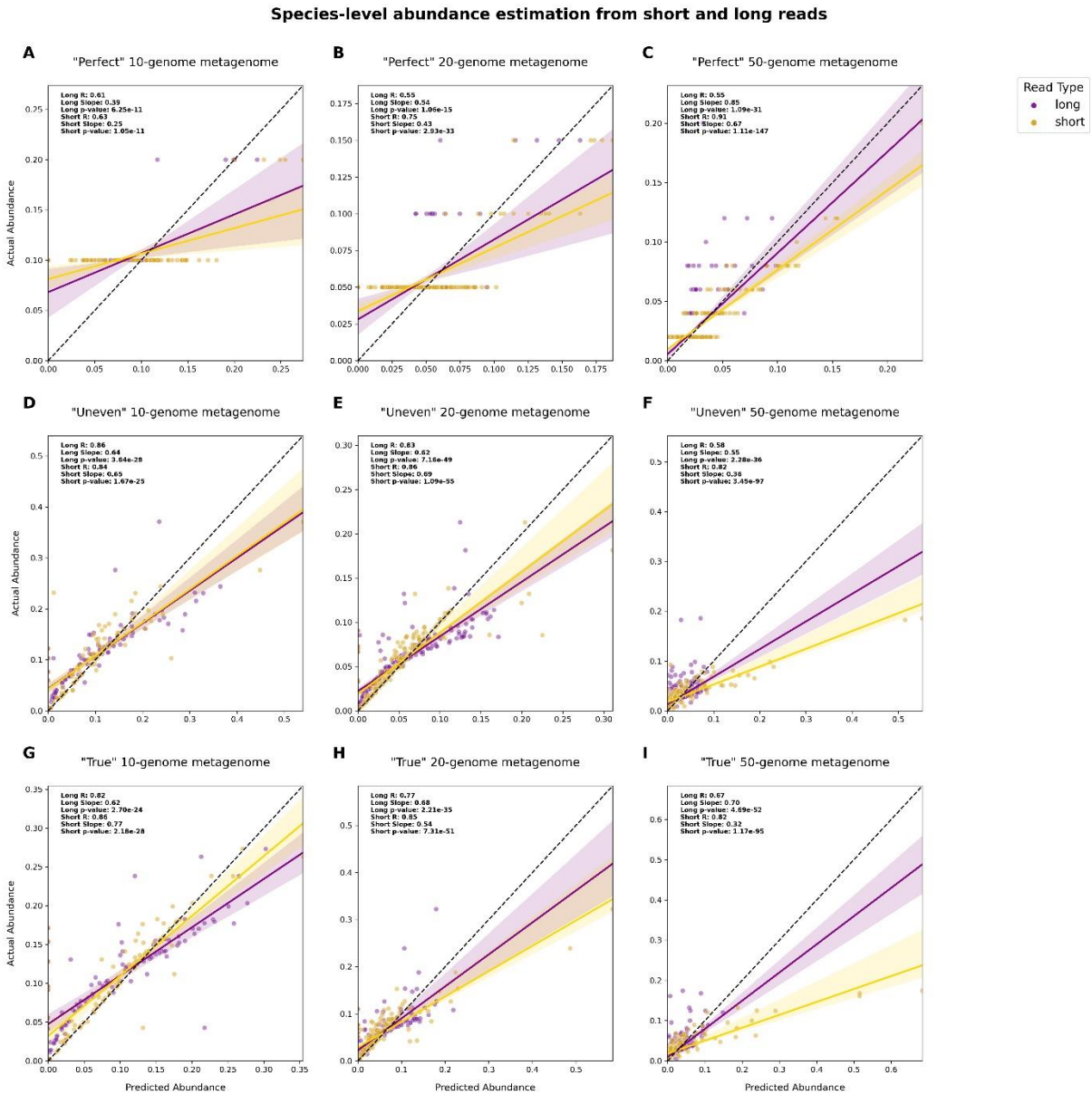


Figure 5. **Comparison of short and long read capacity for species-level relative abundance estimation.** Scatterplots depicting predicted versus actual abundance values for genomes present in simulated metagenomes from short and long read data. Read types were compared across metagenomes of varying complexity at both the genus and species level. "Perfect" datasets consisted of reads without sequencing errors and organism abundance was evenly distributed (5A-C), "Uneven" consisted of reads without sequencing errors and randomly distributed abundances of organisms (5D-F), and "True" consisted of reads with simulated errors and randomly distributed abundances of organisms (5G-I). A linear regression line was plotted for each read type. Each line has its reported R-value, slope, and p-value. The dotted line represents the 1-to-1 line where values of predicted abundance match values of actual abundance.

Metagenome-assembled genome (MAG) recovery from short- and long-read data

In most cases, neither read type resulted in more MAGs being recovered except for short reads producing more MAGs from "True" 50-genome metagenomes (Figure 6). It was found that

the binning of contigs from long read assemblies generally produced more total bins than assemblies produced from short reads (Supplementary Table S2). Following this observation, the quality of MAGs recovered were evaluated. Bins were assigned a quality of high, medium, or low, depending on their completeness. Long reads did not result in more high-quality MAGs at any level of complexity; however, they did show significantly more medium and low-quality MAGs in some cases (Supplementary Figure 4B, C, F). Complementing Figure 5, short reads recovered significantly high-quality MAGs from the most complex datasets (Figure 5I).

Table S2. **Bin and metagenome-assembled genome (MAG) counts from metagenomic samples.** Total bins recovered for each metagenomic sample and counts of MAGs identified from those bins. MAGs were further separated by quality as described in the methods section.

[1]Samples consist of the 10 metagenomes simulated for a given complexity.

[2]A metagenome set is the given complexity level for a collection of simulated metagenomes.

[3a]Sum of all bins predicted by MetaBAT2 from long read assemblies.

[3b]Sum of all bins predicted by MetaBAT2 from short-read assemblies.

sample[1]	metagenome set[2]	bin count (long)[3a]	bin count (short)[3b]	mag count (long)	mag count (short)	high (long)	high (short)	medium (long)	medium (short)	low (long)	low (short)
mock_metagenome_1	10_genomes_perfect	19	12	10	10	8	10	2	0	0	0
mock_metagenome_2	10_genomes_perfect	11	11	9	9	6	8	3	0	0	1
mock_metagenome_3	10_genomes_perfect	10	11	10	10	10	10	0	0	0	0
mock_metagenome_4	10_genomes_perfect	12	10	9	9	7	7	1	2	1	0
mock_metagenome_5	10_genomes_perfect	17	10	9	7	4	6	5	1	0	0
mock_metagenome_6	10_genomes_perfect	17	9	8	8	7	8	1	0	0	0
mock_metagenome_7	10_genomes_perfect	12	10	10	8	8	8	2	0	0	0
mock_metagenome_8	10_genomes_perfect	9	10	9	9	8	8	1	1	0	0
mock_metagenome_9	10_genomes_perfect	14	11	9	9	6	8	1	1	2	0
mock_metagenome_10	10_genomes_perfect	16	13	10	9	8	8	2	1	0	0
mock_metagenome_1	10_genomes_true	11	11	9	10	8	9	0	1	1	0
mock_metagenome_2	10_genomes_true	11	11	9	10	8	10	1	0	0	0
mock_metagenome_3	10_genomes_true	11	8	7	8	6	8	1	0	0	0
mock_metagenome_4	10_genomes_true	12	12	9	9	8	9	0	0	1	0
mock_metagenome_5	10_genomes_true	11	8	7	8	6	8	1	0	0	0
mock_metagenome_6	10_genomes_true	12	12	7	9	6	7	1	1	0	1
mock_metagenome_7	10_genomes_true	13	10	7	8	5	8	2	0	0	0
mock_metagenome_8	10_genomes_true	15	10	9	10	7	10	1	0	1	0
mock_metagenome_9	10_genomes_true	9	9	7	5	5	3	1	1	1	1
mock_metagenome_10	10_genomes_true	14	10	9	10	7	10	1	0	1	0
mock_metagenome_1	10_genomes_uneven	13	10	9	9	8	9	1	0	0	0
mock_metagenome_2	10_genomes_uneven	15	9	9	8	6	7	2	1	1	0
mock_metagenome_3	10_genomes_uneven	10	11	9	9	8	9	1	0	0	0
mock_metagenome_4	10_genomes_uneven	10	12	9	10	8	10	1	0	0	0
mock_metagenome_5	10_genomes_uneven	12	9	8	7	7	7	0	0	1	0
mock_metagenome_6	10_genomes_uneven	16	17	7	6	6	6	0	0	1	0
mock_metagenome_7	10_genomes_uneven	18	10	7	9	3	9	2	0	2	0
mock_metagenome_8	10_genomes_uneven	15	13	9	9	6	7	1	2	2	0
mock_metagenome_9	10_genomes_uneven	9	9	7	8	7	6	0	1	0	1
mock_metagenome_10	10_genomes_uneven	10	10	10	10	10	10	0	0	0	0
mock_metagenome_1	20_genomes_perfect	26	18	18	13	14	13	4	0	0	0
mock_metagenome_2	20_genomes_perfect	22	18	17	15	14	14	0	1	3	0
mock_metagenome_3	20_genomes_perfect	24	19	17	14	13	13	3	1	1	0
mock_metagenome_4	20_genomes_perfect	19	20	18	17	16	15	2	0	0	2
mock_metagenome_5	20_genomes_perfect	22	21	19	17	17	16	1	1	1	0
mock_metagenome_6	20_genomes_perfect	26	18	14	13	10	12	2	1	2	0
mock_metagenome_7	20_genomes_perfect	33	32	11	13	10	10	1	3	0	0
mock_metagenome_8	20_genomes_perfect	22	22	14	13	10	12	3	0	1	1
mock_metagenome_9	20_genomes_perfect	27	17	19	13	14	13	4	0	1	0
mock_metagenome_10	20_genomes_perfect	22	21	18	16	14	15	4	1	0	0
mock_metagenome_1	20_genomes_true	17	20	11	14	11	12	0	1	0	1
mock_metagenome_2	20_genomes_true	24	17	14	15	12	15	1	0	1	0
mock_metagenome_3	20_genomes_true	23	23	14	17	12	13	1	2	1	2
mock_metagenome_4	20_genomes_true	24	19	16	18	12	15	3	2	1	1
mock_metagenome_5	20_genomes_true	24	20	14	14	14	12	0	2	0	0
mock_metagenome_6	20_genomes_true	26	20	14	13	7	12	5	0	2	1
mock_metagenome_7	20_genomes_true	15	21	9	9	7	7	2	2	0	0
mock_metagenome_8	20_genomes_true	19	21	12	12	9	12	3	0	0	0
mock_metagenome_9	20_genomes_true	23	22	14	14	10	13	2	0	2	1
mock_metagenome_10	20_genomes_true	21	20	19	19	14	19	3	0	2	0
mock_metagenome_1	20_genomes_uneven	26	18	18	16	14	16	3	0	1	0
mock_metagenome_2	20_genomes_uneven	20	18	16	17	12	17	3	0	1	0
mock_metagenome_3	20_genomes_uneven	23	19	16	16	14	16	2	0	0	0
mock_metagenome_4	20_genomes_uneven	19	20	16	14	13	13	3	1	0	0
mock_metagenome_5	20_genomes_uneven	26	21	19	19	15	17	3	2	1	0
mock_metagenome_6	20_genomes_uneven	27	23	18	19	17	16	1	3	0	0
mock_metagenome_7	20_genomes_uneven	23	20	16	16	13	14	3	2	0	0
mock_metagenome_8	20_genomes_uneven	22	21	19	20	14	16	4	3	1	1
mock_metagenome_9	20_genomes_uneven	25	22	15	16	10	13	4	3	1	0
mock_metagenome_10	20_genomes_uneven	20	19	15	15	12	13	3	2	0	0
mock_metagenome_1	50_genomes_perfect	57	49	33	30	19	22	9	5	5	3
mock_metagenome_2	50_genomes_perfect	47	48	35	36	28	31	6	4	1	1
mock_metagenome_3	50_genomes_perfect	51	43	32	32	22	28	5	3	5	1
mock_metagenome_4	50_genomes_perfect	50	50	29	31	19	27	6	2	4	2
mock_metagenome_5	50_genomes_perfect	56	41	29	33	21	29	3	3	5	1
mock_metagenome_6	50_genomes_perfect	47	46	33	32	23	28	5	4	5	0
mock_metagenome_7	50_genomes_perfect	48	57	33	30	24	27	4	2	5	1
mock_metagenome_8	50_genomes_perfect	49	44	26	30	17	25	6	5	3	0
mock_metagenome_9	50_genomes_perfect	44	44	34	33	28	28	4	3	2	2
mock_metagenome_10	50_genomes_perfect	54	43	33	27	28	24	1	3	4	0
mock_metagenome_1	50_genomes_true	52	39	26	33	16	28	4	3	6	2
mock_metagenome_2	50_genomes_true	50	38	28	30	19	24	4	4	5	2
mock_metagenome_3	50_genomes_true	52	38	25	29	18	23	3	5	4	1
mock_metagenome_4	50_genomes_true	56	40	26	32	14	27	5	3	7	2
mock_metagenome_5	50_genomes_true	50	37	23	24	16	19	2	4	5	1
mock_metagenome_6	50_genomes_true	51	42	22	28	18	18	4	7	0	3
mock_metagenome_7	50_genomes_true	52	41	24	31	18	26	3	1	3	4
mock_metagenome_8	50_genomes_true	41	42	24	28	16	20	6	4	2	4
mock_metagenome_9	50_genomes_true	39	42	25	23	13	18	8	4	4	1
mock_metagenome_10	50_genomes_true	46	35	23	30	20	27	2	2	1	1
mock_metagenome_1	50_genomes_uneven	49	38	31	32	22	27	6	3	3	2
mock_metagenome_2	50_genomes_uneven	44	32	28	27	19	24	6	2	3	1
mock_metagenome_3	50_genomes_uneven	47	43	33	28	20	21	8	4	5	3
mock_metagenome_4	50_genomes_uneven	52	37	33	27	22	26	10	0	1	1
mock_metagenome_5	50_genomes_uneven	39	32	26	23	22	20	2	2	2	1
mock_metagenome_6	50_genomes_uneven	56	39	32	29	16	24	12	3	4	2
mock_metagenome_7	50_genomes_uneven	48	37	31	28	23	22	5	4	3	2
mock_metagenome_8	50_genomes_uneven	53	39	32	30	19	28	7	2	6	0
mock_metagenome_9	50_genomes_uneven	48	41	28	29	23	22	3	3	2	4
mock_metagenome_10	50_genomes_uneven	61	47	38	36	26	29	7	3	5	4

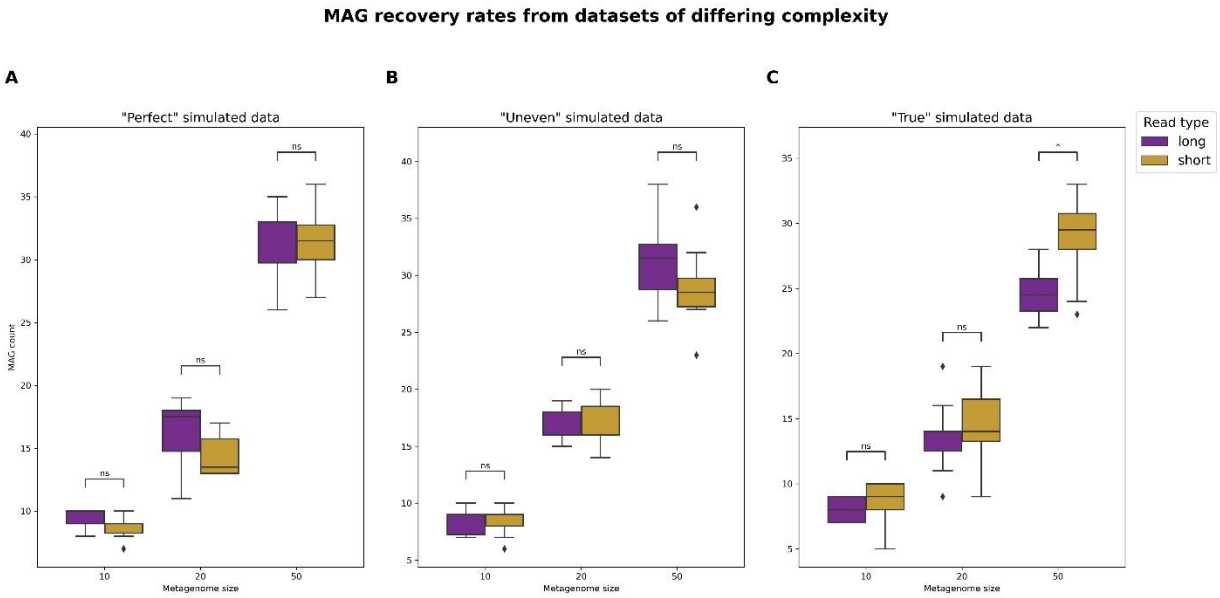
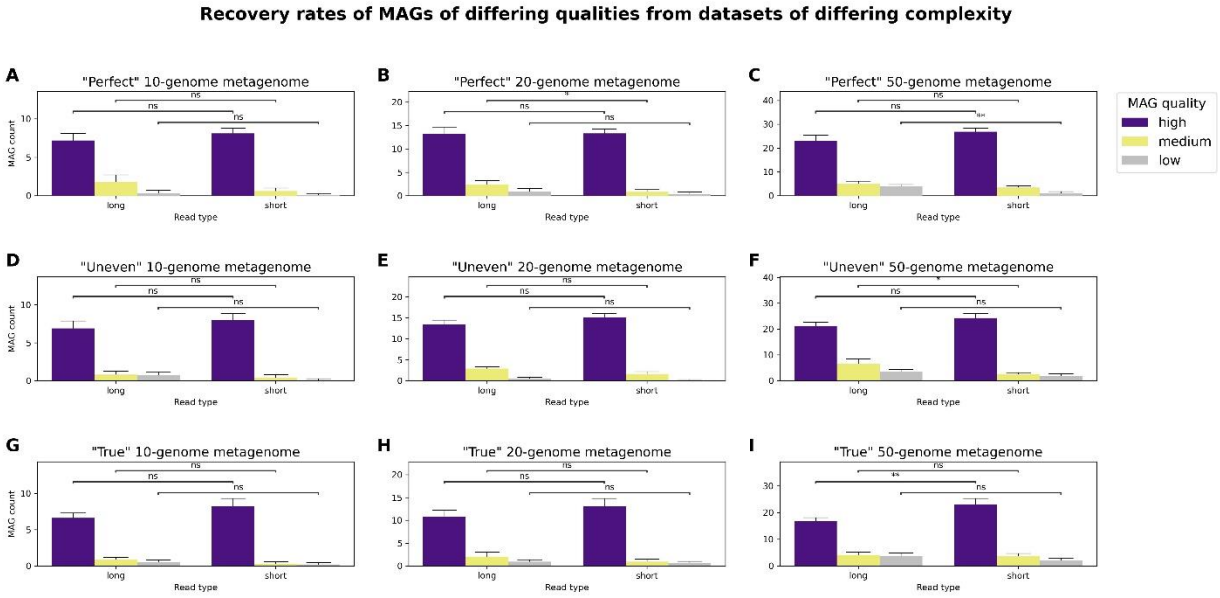


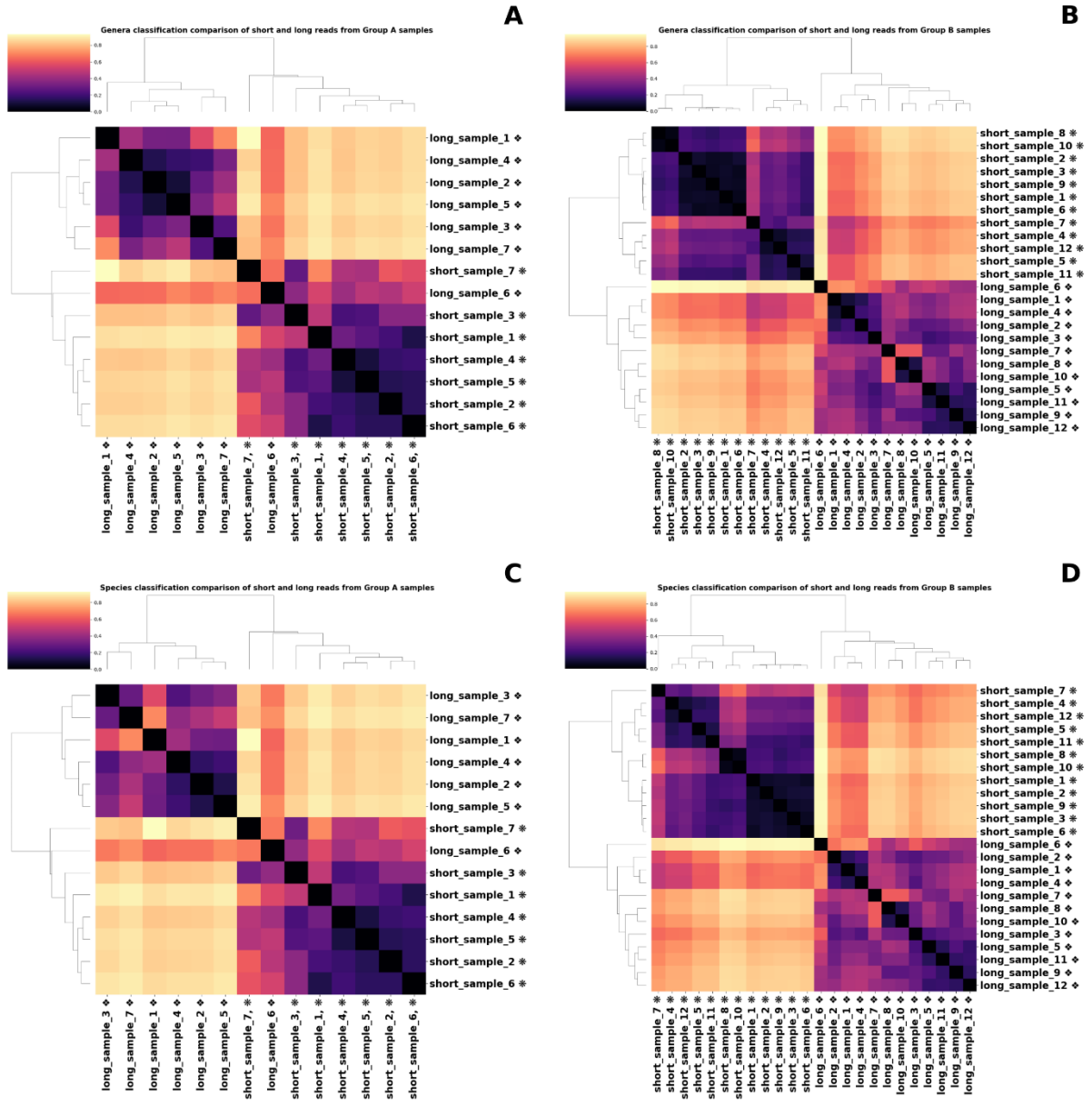
Figure 6. **Comparison of metagenome-assembled genome (MAG) recovery capabilities from assemblies produced using short or long reads.** Total number of MAGs recovered from assemblies of simulated short or long read sequence data. "Perfect" datasets consisted of reads without sequencing errors and organism abundance being evenly distributed (6A), "Uneven" consisted of reads without sequencing errors and organism abundance being randomly assigned (6B), and "True" consisted of reads with simulated sequencing errors and organism abundance being randomly assigned (6C). Significance was calculated using a Mann-Whitney U Test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.



Supplemental Figure 5. **Comparison of metagenome-assembled genome (MAG) counts recovered from short and long read assemblies by quality.** Barplots depict the total number of MAGs of a given quality recovered from both short and long read assemblies. "Perfect" denotes simulated data have no errors and abundances are evenly distributed for each organism in the metagenome (S4A-C), "Uneven" have no errors, but variable abundances for each organism (S4D-F), and "True" have simulated sequencing errors based on the read type, as well as variable abundances for each organism (S4G-I). Significance was calculated using a Mann-Whitney U Test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

Microbial compositional differences between short and long reads

Using experimental data obtained from mouse fecal pellets, we examined whether read type significantly altered inferences of microbial composition for a metagenomic sample. Qualitative assessment showed clustering of results around similar read types instead of around sample types (Figure S4). This trend was also observed when results were run through a PCA where samples clustered around similar read types instead of around similar sample types (Figure 8).



Supplemental Figure 6. **Hierarchically-clustered heatmap of Bray-Curtis Dissimilarity.** Heatmap visualizing the dissimilarity of samples based on the read type used in beta-diversity assessment. Brighter values indicate lower similarity (or higher Bray-Curtis dissimilarity), darker values indicate higher similarity (or lower Bray-Curtis dissimilarity). ❖: Samples consist of long read data. ❋: Samples consist of short read data

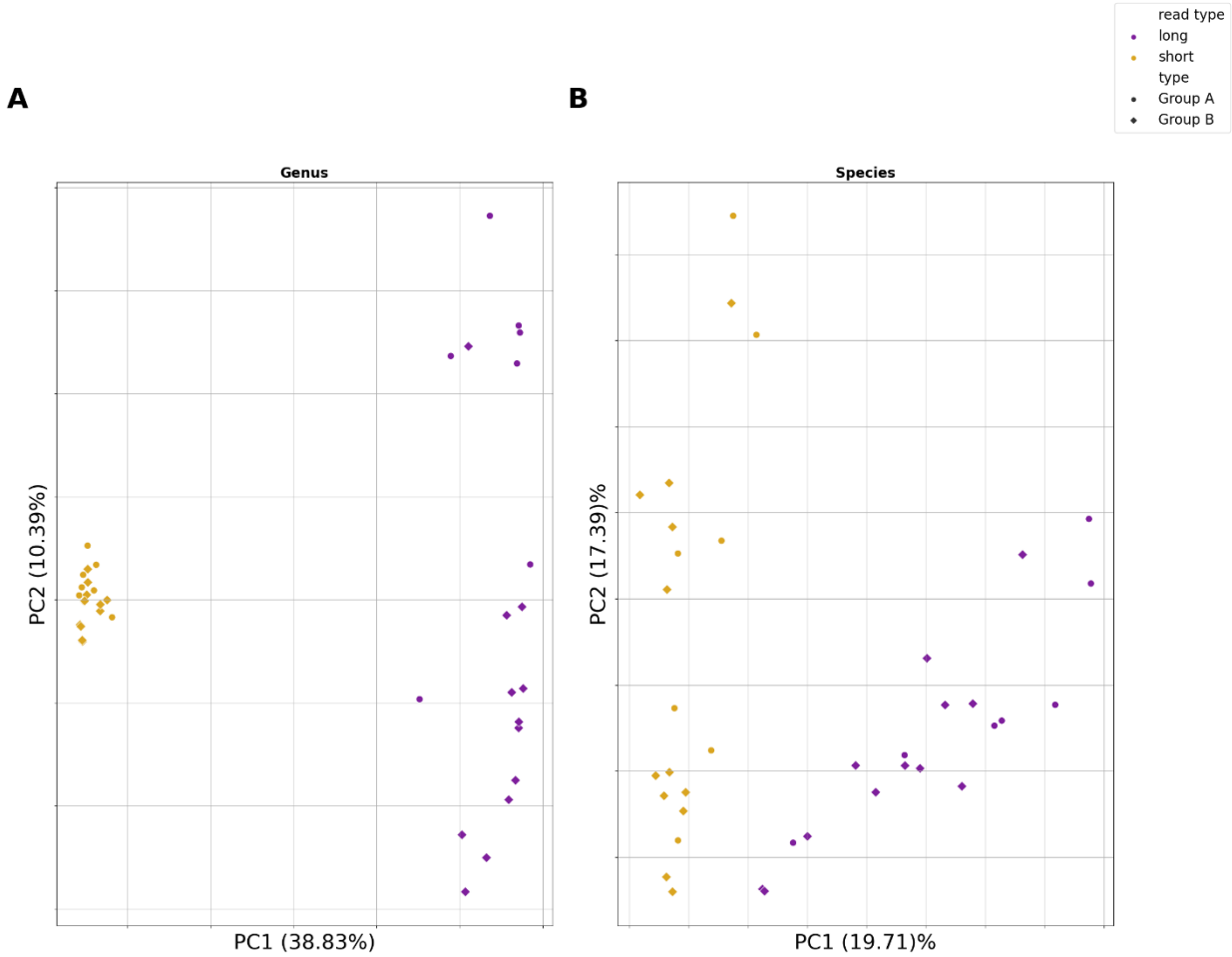


Figure 7. **Principal Component Analysis (PCA) results from short and long read sequence data analysis of empirical metagenomic samples from mouse fecal pellets.** PCA plot depicting the distribution of metagenomic samples according to their two main principal components. Coloration indicates the read type: purple for long read data and yellow for short read data. Shape indicates the sample type. Principal components 1 and 2 are on the x and y-axis respectively with their explained variance ratios written as a percentage.

Discussion

For metagenomic studies of microbial communities, long-read sequencing platforms potentially offer numerous advantages, yet short read sequencing remains the most commonly applied technology (Watson et al. 2021). We sought to quantitatively assess the use of long-read sequence data by using simulated metagenomes with increasing complexity. Evaluation of read type performance across several metrics showed that long and short reads possess respective strengths. We also used paired long and short read metagenomic data from murine fecal samples to assess the generalizability of our findings. Overall, we highlight that choice of sequence type has a direct impact on the ability to accurately assemble metagenomes, identify comprising

members of a diverse microbial community, and recover MAGs from metagenomic sequence data.

For optimal performance, sequencing data must be capable of producing high quality assemblies for taxonomic classification, relative abundance estimation, functional annotation, and other related analyses. We show that in simulated long read assemblies had significantly higher percent genome coverages than short read assemblies. Analyses that rely on examining contiguous sequences, such as gene annotation, would thus benefit from assemblies generated using long-read data. Short reads are known to struggle with repetitive regions that are common in many bacterial genomes (Hu et al., 2021), which likely explains the lower genome coverages and greater discontiguity observed from short read assemblies. Our results also suggest long reads would perform better at recovering novel genomes; however, that was not directly tested here. Higher NGA50 values indicate greater contiguity, further supporting the findings highlighted by higher genome fraction observed. Regarding misassembly rates, short reads did produce fewer misassemblies in all "Perfect" datasets (Figure 2C). However, in more complex datasets, the observed differences were found to not be significant in most cases (Figures 2F, I). One explanation is that metagenomic assemblers attempt to balance accuracy and contiguity, thus increasing the possibility of misassemblies (Kolmogorov et al., 2020). The lower per-base accuracy of long-read sequencers has long been a reason researchers tend to favor short reads. Our results highlight the shrinking gap in result quality of assemblies produced with either short or long reads.

Taxonomic classification is a central to metagenomic studies and several tools exist for use on filtered reads or assembled contigs. Previous studies have compared classification using partial and full 16S sequencing using short and long read platforms respectively (Hahn et al., 2016). Here, we demonstrate the capabilities of each read type when classifying whole metagenomic sequence data. For long-read assemblies, we found that classifications demonstrated significantly higher precision at both the genus and species level overall. The greater precision likely stems from long read assemblies being more contiguous, thus having fewer small fragments that could be improperly classified like in short read assemblies. When classifying reads at the genus level, both long and short reads performed similarly (Figure S2). This was partly explained by long reads having a lower precision compared to classifications with contigs. This is expected given there are more reads than contigs, thus more chances for reads to be misassigned to an improper taxon. When analyzing reads at a finer taxonomic resolution, long reads performed significantly better. For differentiation of species within a genus, larger reads provide more information for distinguishing the genome of origin. For assemblies and filtered reads, we observed no difference in sensitivity. These findings are relevant for the potential application of long-read metagenomic sequencing to infectious disease diagnostics (Bastida et al., 2021; Hoang et al., 2022; Huang et al., 2019). For example, the ability of ONT sequencing platforms to generate sequencing reads in real time, which can be used for taxonomic classification without sacrificing precision or sensitivity, is invaluable and avoids lengthy culturing, isolation, and molecular diagnostics.

In addition to taxonomic classification, assessment of differences in the relative abundance of comprising species is of great interest. This generally involves alignment of reads to a reference genome database and subsequent estimation of abundances from the coverage, which can be performed using both short and long reads (Verma & Sharma, 2020; Warwick-Dugdale et al., 2019). Our results showed that predicted abundances from long read data were more likely to be close to the actual abundance data for both genus and species-level classification (Figure 5; S4). This was likely due to the role of Kraken2's results in Bracken's estimation of organismal relative abundances. Bracken takes the initial classifications and then attempts to re-assign reads to proper taxa using a Bayesian-based approach. Having greater accuracy in these classifications enables Bracken to more likely assign reads to taxa accurately. Assessment of relative abundance thus benefitted from the higher precision of long reads. Research that derives conclusions based on significant shifts in abundances of all taxa or specific organisms can thus benefit from reliance on abundance estimation. Also, in clinical diagnostics, accurate estimation of abundance may enable clinicians to identify the etiologic microbe associated with the disease of interest (Wen et al., 2017).

A considerable challenge in microbial research is that most bacteria are non-culturable, which makes identification of organisms of potential importance difficult. Using shotgun sequencing and *de novo* metagenomic assembly, it is possible to recover both known and novel genomes and infer the identity and functional capacities of the organisms (Parks et al. 2017). Assemblies produced from either read type resulted in similar MAG recovery rates except short reads did identify more MAGs in "True" 50-genome metagenomes (Figure 6). A key tool in the binning process was MetaBAT2, which is highly dependent on assembly quality for performance. As a result, while long reads were more contiguous and often more complete, misassembly rates may have impacted read alignment during the MAG identification process. This is supported by our observation that long read assemblies contained more misassembles, even though misassembly rates were not significantly different between read types (Supplementary Table S2). This likely resulted in more bins with fewer contigs spread among them and therefore lower completeness due to the cutoffs used for MAG reporting. This also explains why long reads produced at times more medium or low-quality MAGs, since MAG quality was primarily determined by completeness, thus having contigs inadequately binned would result in lower completeness of MAGs.

Selection of sequencing platform has previously been shown to directly impact the taxonomic assignment of partial/total 16S rRNA sequences (Stevens et al., 2023). In the present study, bacterial metagenomic DNA extracted from murine fecal samples were sequenced using both short- and long-read sequencing platforms. Qualitative assessment measuring the beta diversity among samples of the same type, but different sequence data, revealed samples clustered by sequence type, not sample type (Supplementary Figure 4). Comparisons using PCA revealed the same trend of samples clustering by sequence type, rather than by sample type (Figure 8). This is likely due to an inflation of misclassified taxa in short read assemblies, causing clustering to be associated with those unique taxa. This is consistent with our findings that simulated short read data produced more false positives than long reads. This, in turn,

artificially increased the diversity of the sample and explains clustering due to read type instead of sample type.

Taken together, our results show that long read data offers certain advantages towards accurate representation of a metagenome. Specifically, long reads resulted in greater accuracy when taxonomically classifying contigs and reads, and their assemblies were more contiguous and returned greater genome fractions for genomes within a metagenome. While this study provides evidence supporting the use of long read sequencing for metagenomic research, it does have some limitations. The most complex metagenomes that were simulated consisted of 50 organisms with variable abundance and simulated sequencing errors specific to each read type, but empirical data may consist of hundreds or even thousands of unique organisms (Bastida et al. 2021). Future work could employ simulations that mimic high diversity as well as abundance distributions observed in real metagenomic samples; however, we feel that increasing taxa would only increase the disparity between sequence data types. Another consideration is that pipelines for processing and analyzing metagenomic data can vary considerably. Various tools exist for read correction, assembly, polishing, taxonomic classification, and binning, all of which can affect results. Steps such as read error correction and assembly polishing are often used in metagenomic studies; however, we did not include them here to reduce computational requirement and to limit variables in the comparison, especially if they are unique to one read type. Another limitation was that binning algorithms are usually designed with short reads in mind, leveraging the higher coverage of data that accompanies short read sequencing. In recent years, tools such as a LRBinner have emerged which offer better binning for error-prone long reads (Wickramarachchi and Lin 2022), but those were not applied here. Overall, we aimed to use the most commonly employed tools used for these studies at the time of writing. Last, we used a limited number of real-world samples, which could be expanded in future studies.

Short and long read sequencing have their respective places in metagenomic research. Short read sequencing does have its strengths, especially with pipelines for analyzing short read data being more established. Our results highlight that both read types perform comparably well across several metagenomic analyses, in particular assembly quality and MAG recovery. For studies concerned with misassemblies, such as those focused on specific species of interest, short reads produce significantly fewer and may be preferable. For studies concerned with microbial composition, long reads provided higher precision with similar sensitivity to short reads despite lower sequencing depths. They also provide better estimations of organismal abundance within metagenomic samples. Leveraging the strengths of both read types is also possible. Tools such as OPERA-MS combine long reads containing more genetic information per-read with high accuracy short reads (Bertrand et al., 2019). Other programs like PolyPolish enable the use of short reads for improving long read assembly quality (Wick & Holt, 2022). Further advancement may shift practice towards use of hybrid approaches as has occurred with genome sequencing. In conclusion, our results show long read sequence data outperforms short read sequence data when used for taxonomic classification and metagenomic assembly contiguity. This work provides evidence for the consideration of long read sequencing for research focused on accurate reconstruction of microbial populations and its strength in clinical diagnostics.

References

A. Murat Eren. 2016. “Assessing Completion and Contamination of Metagenome-Assembled Genomes.” Https://Merenlab.Org/2016/06/09/Assessing-Completion-and-Contamination-of-MAGs/. June 9, 2016.

Aitchison, J. (1986). The statistical analysis of Compositional Data. Chapman and Hall.

Amarasinghe, Shanika L., Shian Su, Xueyi Dong, Luke Zappia, Matthew E. Ritchie, and Quentin Gouil. 2020. “Opportunities and Challenges in Long-Read Sequencing Data Analysis.” *Genome Biology* 21 (1): 30. https://doi.org/10.1186/s13059-020-1935-5.

Amarasinghe, S. L., Ritchie, M. E., & Gouil, Q. (2021). Long-read-tools.org: An interactive catalogue of analysis methods for long-read sequencing data. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab003.

Amarasinghe, S. L., Ritchie, M. E., & Gouil, Q. (2023). *Long-read-tools*. Long Read Tools. https://long-read-tools.org/analysis.html

Bastida, Felipe, David J. Eldridge, Carlos García, G. Kenny Png, Richard D. Bardgett, and Manuel Delgado-Baquerizo. 2021. “Soil Microbial Diversity–Biomass Relationships Are Driven by Soil Carbon Content across Global Biomes.” *The ISME Journal* 15 (7): 2081–91. https://doi.org/10.1038/s41396-021-00906-0.

Begmatov, Shahjahon, Alexander G. Dorofeev, Vitaly V. Kadnikov, Alexey V. Beletsky, Nikolai V. Pimenov, Nikolai V. Ravin, and Andrey V. Mardanov. 2022. “The Structure of Microbial Communities of Activated Sludge of Large-Scale Wastewater Treatment Plants in the City of Moscow.” *Scientific Reports* 12 (1): 3458. https://doi.org/10.1038/s41598-022-07132-4.

Berglund, Fanny, Tobias Österlund, Fredrik Boulund, Nachiket P. Marathe, D. G. Joakim Larsson, and Erik Kristiansson. 2019. “Identification and Reconstruction of Novel Antibiotic Resistance Genes from Metagenomes.” *Microbiome* 7 (1): 52. https://doi.org/10.1186/s40168-019-0670-1.

Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A. H., Kumar, M. S., Li, C., Dvornicic, M., Soldo, J. P., Koh, J. Y., Tong, C., Ng, O. T., Barkham, T., Young, B., Marimuthu, K., Chng, K. R., Sikic, M., & Nagarajan, N. (2019). Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nature Biotechnology*, *37*(8), 937–944. https://doi.org/10.1038/s41587-019-0191-2.

Bushnell, Brian. 2020. “BBTools.” Sourceforge.Net/Projects/Bbmap/. Walnut Creek: sourceforge.net/projects/bbmap/. June 13, 2020. sourceforge.net/projects/bbmap/.

Charalampous, Themoula, Gemma L Kay, Hollian Richardson, Alp Aydin, Rossella Baldan, Christopher Jeanes, Duncan Rae, et al. 2019. “Nanopore Metagenomics Enables Rapid Clinical Diagnosis of Bacterial Lower Respiratory Infection.” *Nature Biotechnology* 37 (7): 783–92. https://doi.org/10.1038/s41587-019-0156-5.

Charlier, F., Weber, M., Izak, D., Harkin, E., Magnus, M., Lalli, J., Fresnais, L., Chan, M., Markov, N., Amsalem, O., Proost, S., Krasoulis, A., getzze, & Repplinger, S. (2022). *Statannotations* (0.5). Zenodo. https://doi.org/https://doi.org/10.5281/zenodo.7213391.

Chen, Kevin, and Lior Pachter. 2005. "Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities." *PLoS Computational Biology* 1 (2): e24. https://doi.org/10.1371/journal.pcbi.0010024.

Cheng, Jun, Huan Hu, Wei Fang, Duozhi Shi, Chen Liang, Yang Sun, GuoFeng Gao, et al. 2019. "Detection of Pathogens from Resected Heart Valves of Patients with Infective Endocarditis by Next-Generation Sequencing." *International Journal of Infectious Diseases* 83 (June): 148–53. https://doi.org/10.1016/j.ijid.2019.03.007.

Chklovski, Alex, Donovan H. Parks, Ben J. Woodcroft, and Gene W. Tyson. 2022. "CheckM2: A Rapid, Scalable and Accurate Tool for Assessing Microbial Genome Quality Using Machine Learning." *BioRxiv*, July. https://doi.org/https://doi.org/10.1101/2022.07.11.499243.

Danecek, Petr, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10 (2). https://doi.org/10.1093/gigascience/giab008.

Govender, Kumeren N., Teresa L. Street, Nicholas D. Sanderson, and David W. Eyre. 2021. "Metagenomic Sequencing as a Pathogen-Agnostic Clinical Diagnostic Tool for Infectious Diseases: A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies." *Journal of Clinical Microbiology* 59 (9). https://doi.org/10.1128/JCM.02916-20.

Gupta, Shashank, Martin S. Mortensen, Susanne Schjørring, Urvish Trivedi, Gisle Vestergaard, Jakob Stokholm, Hans Bisgaard, Karen A. Krogfelt, and Søren J. Sørensen. 2019. "Amplicon Sequencing Provides More Accurate Microbiome Information in Healthy Children Compared to Culturing." *Communications Biology* 2 (1): 291. https://doi.org/10.1038/s42003-019-0540-1.

Hahn, Andrea, Amit Sanyal, Geovanny F. Perez, Anamaris M. Colberg-Poley, Joseph Campos, Mary C. Rose, and Marcos Pérez-Losada. 2016. "Different next Generation Sequencing Platforms Produce Different Microbial Profiles and Diversity in Cystic Fibrosis Sputum." *Journal of Microbiological Methods* 130 (November): 95–99. https://doi.org/10.1016/J.MIMET.2016.09.002.

Hoang, M. T., Irinyi, L., Hu, Y., Schwessinger, B., & Meyer, W. (2022). Long-reads-based metagenomics in clinical diagnosis with a special focus on fungal infections. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.708550.

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. Human Immunology, 82(11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Huang, B., Jennison, A., Whiley, D., McMahon, J., Hewitson, G., Graham, R., De Jong, A., & Warrilow, D. (2019). Illumina sequencing of clinical samples for virus detection in a public health laboratory. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41830-w.

Kang, Dongwan D., Feng Li, Edward Kirton, Ashleigh Thomas, Rob Egan, Hong An, and Zhong Wang. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies." *PeerJ* 7 (July): e7359. https://doi.org/10.7717/peerj.7359.

Kieser, S., Zdobnov, E. M., & Trajkovski, M. (2022). Comprehensive mouse microbiota genome catalog reveals major difference to its human counterpart. PLOS Computational Biology, 18(3). https://doi.org/10.1371/journal.pcbi.1009947

Kolmogorov, Mikhail, Derek M. Bickhart, Bahar Behsaz, Alexey Gurevich, Mikhail Rayko, Sung Bong Shin, Kristen Kuhn, et al. 2020. "MetaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs." *Nature Methods* 17 (11): 1103–10. https://doi.org/10.1038/s41592-020-00971-x.

Koren, Sergey, and Adam M Phillippy. 2015. "One Chromosome, One Contig: Complete Microbial Genomes from Long-Read Sequencing and Assembly." *Current Opinion in Microbiology* 23 (February): 110–20. https://doi.org/10.1016/j.mib.2014.11.014.

Latorre-Pérez, Adriel, Pascual Villalba-Bermell, Javier Pascual, and Cristina Vilanova. 2020. "Assembly Methods for Nanopore-Based Metagenomic Sequencing: A Comparative Study." *Scientific Reports* 10 (1). https://doi.org/10.1038/s41598-020-70491-3.

Li, Heng. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H. (2023, May 19). Lh3/SEQTK: Toolkit for processing sequences in FASTA/Q Formats. GitHub. https://github.com/lh3/seqtk

Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*, *3*, e104. https://doi.org/10.7717/peerj-cs.104.

Lu, J., Rincon, N., Wood, D., Breitwieser, F., Pockrandt, C., Langmead, B., Salzberg, S., & Steineggar, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*, *17*, 2815–2839. https://doi.org/https://www.nature.com/articles/s41596-022-00738-y

Mikheenko, Alla, Vladislav Saveliev, and Alexey Gurevich. 2016. "MetaQUAST: Evaluation of Metagenome Assemblies." *Bioinformatics* 32 (7): 1088–90. https://doi.org/10.1093/bioinformatics/btv697.

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). Metaspades: A new versatile metagenomic assembler. Genome Research, 27(5), 824–834. https://doi.org/10.1101/gr.213959.116

Oyserman, Ben O., Stalin Sarango Flores, Thom Griffioen, Xinya Pan, Elmar van der Wijk, Lotte Pronk, Wouter Lokhorst, et al. 2022. "Disentangling the Genetic Basis of Rhizosphere

Microbiome Assembly in Tomato." *Nature Communications* 13 (1): 3228. https://doi.org/10.1038/s41467-022-30849-9.

Parks, Donovan H., Christian Rinke, Maria Chuvochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz, and Gene W. Tyson. 2017. "Recovery of Nearly 8,000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life." *Nature Microbiology* 2 (11): 1533–42. https://doi.org/10.1038/s41564-017-0012-7.

Petersen, Lauren M., Isabella W. Martin, Wayne E. Moschetti, Colleen M. Kershaw, and Gregory J. Tsongalis. 2019. "Third-Generation Sequencing in the Clinical Laboratory: Exploring the Advantages and Challenges of Nanopore Sequencing." *Journal of Clinical Microbiology* 58 (1). https://doi.org/10.1128/JCM.01315-19.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Sanderson, Nicholas D., Jeremy Swann, Leanne Barker, James Kavanagh, Sarah Hoosdally, Derrick Crook, Teresa L. Street, and David W. Eyre. 2020. "High Precision *Neisseria Gonorrhoeae* Variant and Antimicrobial Resistance Calling from Metagenomic Nanopore Sequencing." *Genome Research* 30 (9): 1354–63. https://doi.org/10.1101/gr.262865.120.

Schirmer, Melanie, Rosalinda D'Amore, Umer Z. Ijaz, Neil Hall, and Christopher Quince. 2016. "Illumina Error Profiles: Resolving Fine-Scale Variation in Metagenomic Sequencing Data." *BMC Bioinformatics* 17 (1): 125. https://doi.org/10.1186/s12859-016-0976-y.

Sereika, Mantas, Rasmus Hansen Kirkegaard, Søren Michael Karst, Thomas Yssing Michaelsen, Emil Aarre Sørensen, Rasmus Dam Wollenberg, and Mads Albertsen. 2022. "Oxford Nanopore R10.4 Long-Read Sequencing Enables the Generation of near-Finished Bacterial Genomes from Pure Cultures and Metagenomes without Short-Read or Reference Polishing." *Nature Methods* 19 (7): 823–26. https://doi.org/10.1038/s41592-022-01539-7.

Singleton, Caitlin M., Francesca Petriglieri, Jannie M. Kristensen, Rasmus H. Kirkegaard, Thomas Y. Michaelsen, Martin H. Andersen, Zivile Kondrotaite, et al. 2021. "Connecting Structure to Function with the Recovery of over 1000 High-Quality Metagenome-Assembled Genomes from Activated Sludge Using Long-Read Sequencing." *Nature Communications* 12 (1): 2009. https://doi.org/10.1038/s41467-021-22203-2.

Song, Ruiping, Yuanze Sun, Xinfei Li, Changfeng Ding, Yi Huang, Xinyu Du, and Jie Wang. 2022. "Biodegradable Microplastics Induced the Dissemination of Antibiotic Resistance Genes and Virulence Factors in Soil: A Metagenomic Perspective." *Science of The Total Environment* 828 (July): 154596. https://doi.org/10.1016/j.scitotenv.2022.154596.

Stevens, B. M., Creed, T. B., Reardon, C. L., & Manter, D. K. (2023). Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and Agricultural Soil. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-36101-8.

Sun, Xinhe, Yaoyao Cai, Wenwen Dai, Weiwei Jiang, and Weibing Tang. 2022. “The Difference of Gut Microbiome in Different Biliary Diseases in Infant before Operation and the Changes after Operation.” *BMC Pediatrics* 22 (1): 502. https://doi.org/10.1186/s12887-022-03570-1.

Suttner, Brittany, Eric R. Johnston, Luis H. Orellana, Luis M. Rodriguez-R, Janet K. Hatt, Diana Carychao, Michelle Q. Carter, Michael B. Cooley, and Konstantinos T. Konstantinidis. 2020. “Metagenomics as a Public Health Risk Assessment Tool in a Study of Natural Creek Sediments Influenced by Agricultural and Livestock Runoff: Potential and Limitations.” *Applied and Environmental Microbiology* 86 (6). https://doi.org/10.1128/AEM.02525-19.

Verma, S. K., & Sharma, P. C. (2020). NGS-based characterization of Microbial Diversity and functional profiling of solid tannery waste metagenomes. *Genomics*, *112*(4), 2903–2913. https://doi.org/10.1016/j.ygeno.2020.04.002.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wallen, Zachary D., Ayse Demirkan, Guy Twa, Gwendolyn Cohen, Marissa N. Dean, David G. Standaert, Timothy R. Sampson, and Haydeh Payami. 2022. “Metagenomics of Parkinson’s Disease Implicates the Gut Microbiome in Multiple Disease Mechanisms.” *Nature Communications* 13 (1): 6958. https://doi.org/10.1038/s41467-022-34667-x.

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A. C., Allen, M. J., Sullivan, M. B., & Temperton, B. (2019). Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining Genomic Islands. *PeerJ*, *7*. https://doi.org/10.7717/peerj.6800.

Watson, Christopher M., Laura A. Crinnion, Helen Lindsay, Rowena Mitchell, Nick Camm, Rachel Robinson, Caroline Joyce, et al. 2021. “Assessing the Utility of Long-Read Nanopore Sequencing for Rapid and Efficient Characterization of Mobile Element Insertions.” *Laboratory Investigation* 101 (4): 442–49. https://doi.org/10.1038/s41374-020-00489-y.

Wen, C., Zheng, Z., Shao, T., Liu, L., Xie, Z., Le Chatelier, E., He, Z., Zhong, W., Fan, Y., Zhang, L., Li, H., Wu, C., Hu, C., Xu, Q., Zhou, J., Cai, S., Wang, D., Huang, Y., Breban, M., Qin, N., & Ehrlich, S. D. (2017). Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis. Genome Biology, 18(1). https://doi.org/10.1186/s13059-017-1271-6

Wick, Ryan. 2021. “Filtlong.” Https://Github.Com/Rrwick/Filtlong. https://github.com/rrwick/Filtlong. June 30, 2021. https://github.com/rrwick/Filtlong#acknowledgements.

Wick, R. R., & Holt, K. E. (2022). Polypolish: Short-read polishing of long-read bacterial genome assemblies. *PLOS Computational Biology*, *18*(1). https://doi.org/10.1371/journal.pcbi.1009802

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. "Bandage: Interactive Visualization of *de Novo* Genome Assemblies." *Bioinformatics* 31 (20): 3350–52. https://doi.org/10.1093/bioinformatics/btv383.

Wood, Derrick E, Jennifer Lu, and Ben Langmead. 2019. "Improved Metagenomic Analysis with Kraken 2." *Genome Biology* 20 (1): 257. https://doi.org/10.1186/s13059-019-1891-0.

Woodcroft, Ben. 2021. "CoverM." Https://Github.Com/Wwood/CoverM. February 24, 2021. https://github.com/wwood/CoverM.Yang, Chen, Theodora Lo, Ka Ming Nip, Saber Hafezqorani, Rene L Warren, and Inanc Birol. 2021. "Meta-NanoSim_2021.11.19.469328v1.Full." *BioRxiv*, November. https://doi.org/https://doi.org/10.1101/2021.11.19.469328.

Yang, J., Park, J., Park, S., Baek, I., & Chun, J. (2019). Introducing murine microbiome database (MMDB): A curated database with taxonomic profiling of the healthy mouse gastrointestinal microbiome. Microorganisms, 7(11), 480. https://doi.org/10.3390/microorganisms7110480.

Figure Legends

Figure 1. Flow chart detailing the set up for simulated metagenomic datasets. Perfect metagenomic samples had simulated reads with no sequencing errors and an even abundance distribution of organisms. Uneven samples had simulated reads with no sequencing errors and randomly assigned abundance values for organisms in the metagenome. True datasets had simulated reads with simulated sequencing errors specific to each read type based on the platform they would be generated from and randomly assigned abundance values for members of each metagenome.

Figure 2. Assembly graphs from a "True" 10-genome metagenome visualized by Bandage. 2A. Bandage graph generated from metagenomic assembly of long read sequence data. 2B. Bandage graph generated from metagenomic assembly of short read sequence data. The black box highlights small, disconnected contigs.

Figure 3. Comparisons of assembly quality from metaQUAST reports. Violin plots depict the differences in genome fraction values from all genomes in a set of metagenomes (A, D, G). Stripplots depict the differences in NGA50 across all genomes in a set of metagenomes (B, E, H). Barplots depict total misassembly counts across all genomes in a set of metagenomes (C, F, I). "Perfect" denotes simulated data have no errors and abundances are evenly distributed for each organism in the metagenome. "Uneven" denotes simulated data have no errors, but variable abundances for each organism. "True" denotes simulated data have simulated sequencing errors based on the read type, as well as variable abundances for each organism. Significance was calculated using a Mann-Whitney U Test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure 4. Performance evaluation of genus-level classification on metagenomic assemblies using short or long reads. Stripplots and boxplots were used to visualize the performance metrics (precision, recall, and F-score) on metagenomes of varying size and complexity.

“Perfect” simulated reads have no errors and even abundances across organisms (A, B, C). “Uneven” simulated reads have no errors and randomly varied abundances across organisms (D, E, F). “True” simulated reads have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a One-way ANOVA. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure 5. Performance evaluation of species-level classification on metagenomic assemblies using short or long reads. Stripplots and boxplots were used to visualize the performance metrics (precision, recall, and F-score) on metagenomes of varying size and complexity. “Perfect” simulated reads have no errors and even abundances across organisms (A, B, C). “Uneven” simulated reads have no errors and randomly varied abundances across organisms (D, E, F). “True” simulated reads have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a One-way ANOVA. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure 6. Comparison of short and long read capacity for relative abundance estimation. Overlapping stripplots and boxplots were used to portray the differences in relative abundance estimation of organisms present in metagenomic samples when using short or long reads. Read types were compared across metagenomes of varying complexity at both the genus and species level. “Perfect” datasets consisted of reads without sequencing errors and organism abundance was evenly distributed (6A, 6D). “Uneven” datasets consisted of reads without sequencing errors and randomly distributed abundances of organisms (6B, 6E). “True” datasets consisted of reads with simulated errors and randomly distributed abundances of organisms (6C, 6F). Significance was calculated using a Mann-Whitney U Test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure 7. Comparison of MAG recovery capabilities from assemblies produced using short or long reads. Data depicted show the total number of MAGs recovered from assemblies of simulated short or long read sequence data. “Perfect” datasets consisted of reads without sequencing errors and organism abundance being evenly distributed (7A). “Uneven” datasets consisted of reads without sequencing errors and organism abundance being randomly assigned (7B). “True” datasets consisted of reads with simulated sequencing errors and organism abundance being randomly assigned (7C). Significance was calculated using a Mann-Whitney U Test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure 8. Principal Component Analysis results from short and long read sequence data analysis of real-world metagenomic samples. PCA plot depicting the distribution of metagenomic samples according to their two main principal components. Coloration indicates the read type, purple for results from long read data and yellow for results from short read data. Shape indicates the sample type, circles for control data and diamonds for experimental data. Principal components 1 and 2 are on the x and y-axis respectively with their explained variance ratios written as a percentage

Figure S1. Performance evaluation of genus-level classification using short and long reads. Overlapping stripplots and boxplots were used to visualize the performance metrics (precision, recall, and F-score) on metagenomes of varying complexity. "Perfect" simulated reads have no errors and even abundances across organisms (A, B, C). "Uneven" simulated reads have no errors and randomly varied abundances across organisms (D, E, F). "True" simulated reads have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a Mann-Whitney U test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure S2. Performance evaluation of species-level classification using short and long reads. Overlapping stripplots and boxplots were used to visualize the performance metrics (precision, recall, and F-score) on metagenomes of varying complexity. "Perfect" simulated reads have no errors and even abundances across organisms (A, B, C). "Uneven" simulated reads have no errors and randomly varied abundances across organisms (D, E, F). "True" simulated reads have simulated errors specific to each read type and varied abundances across organisms (G, H, I). Significance was calculated using a One-way ANOVA. ns: non-significant p-value ($p > 0.05$; blank means $p = 1$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure S3. Comparison of MAG counts recovered from short and long read assemblies by quality. Barplots depict the total number of MAGs of a given quality recovered from both short and long read assemblies. "Perfect" denotes simulated data have no errors and abundances are evenly distributed for each organism in the metagenome. "Uneven" denotes simulated data have no errors, but variable abundances for each organism. "True" denotes simulated data have simulated sequencing errors based on the read type, as well as variable abundances for each organism. Significance was calculated using a Mann-Whitney U Test with Bonferroni correction. ns: non-significant p-value ($p > 0.05$), *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$

Figure S4. Hierarchically-clustered heatmap of Bray-Curtis Dissimilarity. Heatmap visualizing the dissimilarity of control and experimental samples based on the read type used in beta-diversity assessment. Brighter values indicate lower similarity, darker values indicate higher similarity.

Table S1. Results of Mann-Whitney U testing for genome fraction, NGA50, and number of misassembly values from metaQUAST.

[1]"Perfect" indicates reads had no errors and abundances were evenly distributed. "Uneven" denotes reads had no errors, but variable abundances for each organism. "True" reads had simulated sequencing errors based on the read type and variable abundances for the organisms.

[2]P-value from Mann-Whitney U test with Bonferroni correction applied

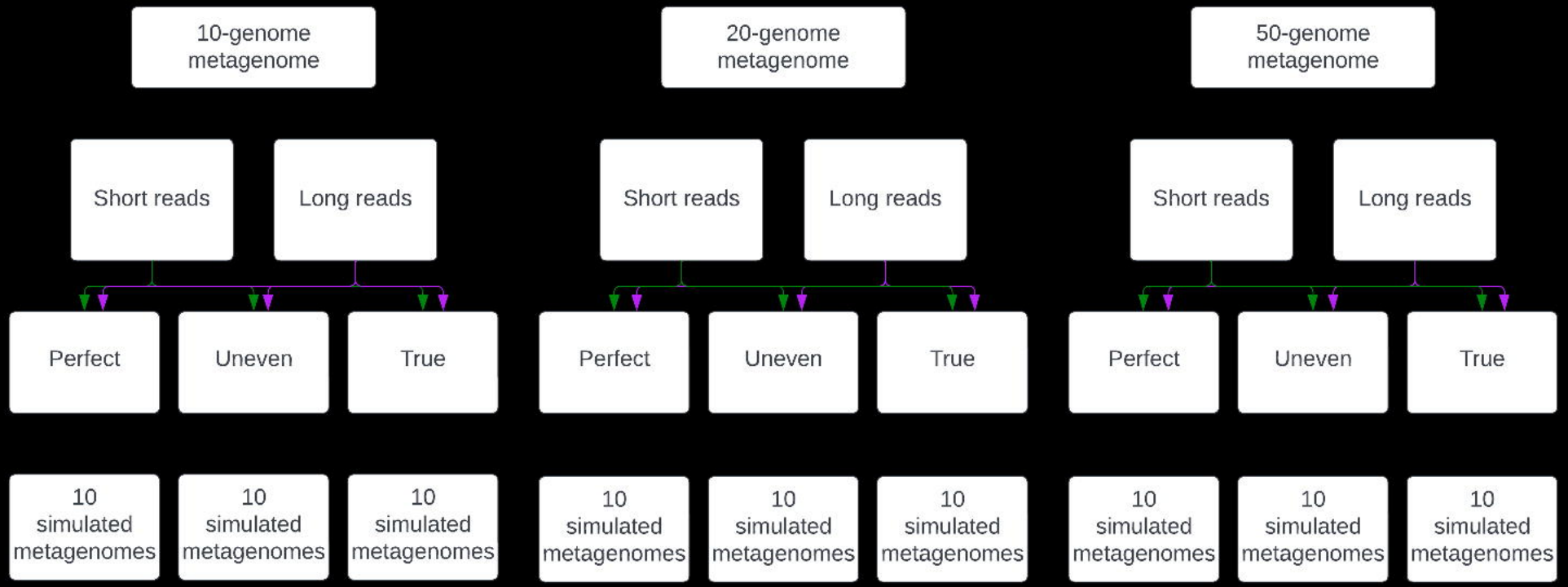
Table S2. Bin and MAG counts from metagenomic samples. Total bins recovered for each metagenomic sample and counts of MAGs identified from those bins. MAGs were further separated by quality as described in the methods section.

[1]Samples consist of the 10 metagenomes simulated for a given complexity

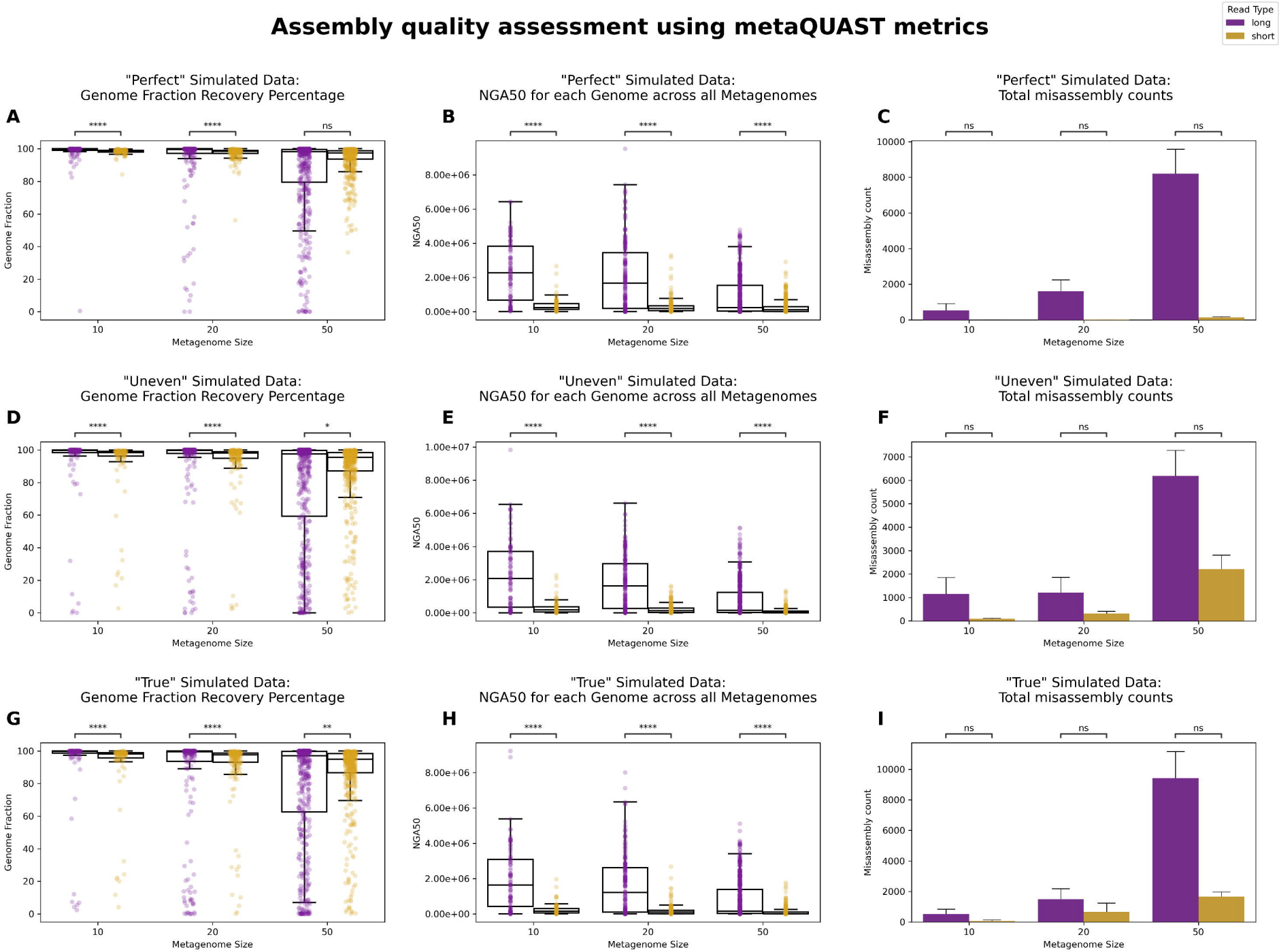
[2]A metagenome set is the given complexity level for a collection of simulated metagenomes

[3a]Sum of all bins predicted by MetaBAT2 from long-read assemblies

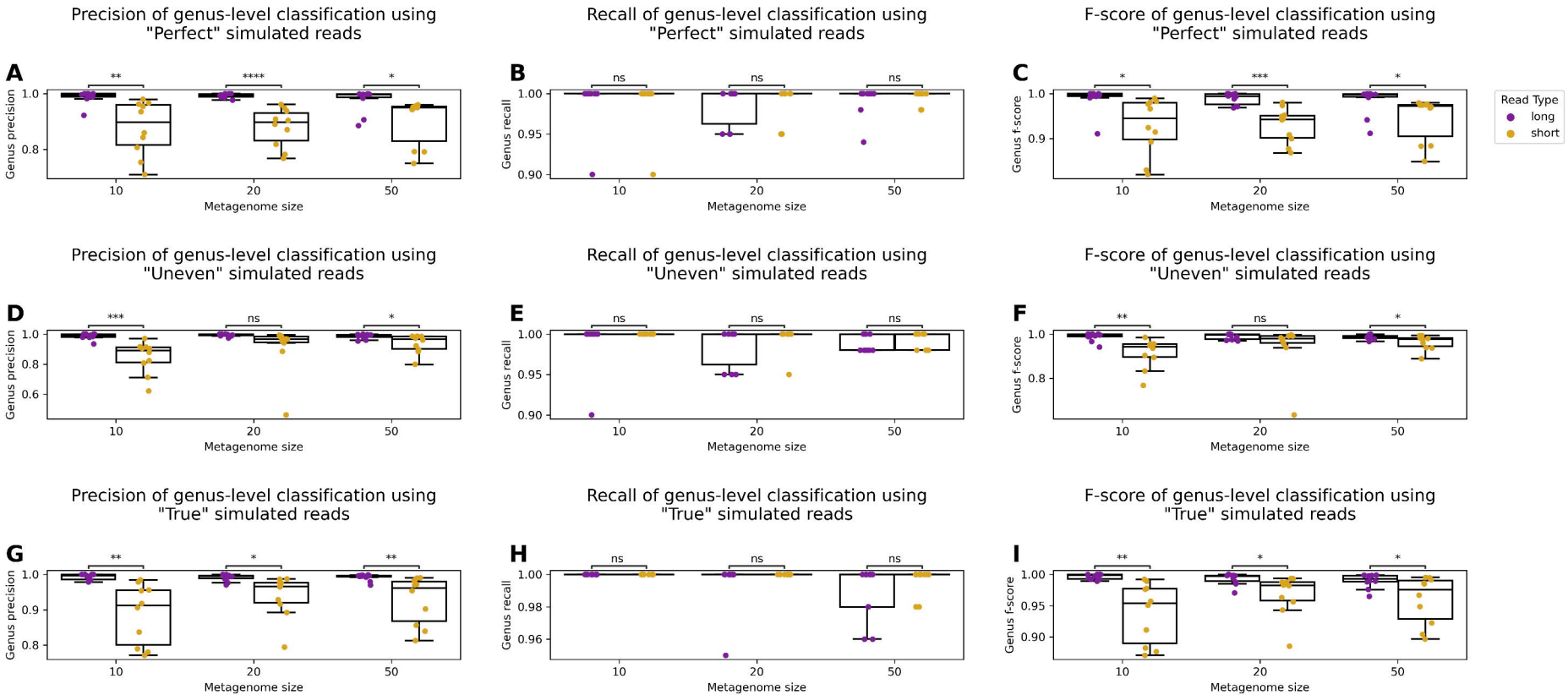
[3b]Sum of all bins predicted by MetaBAT2 from short-read assemblies


10-genome metagenome
20-genome metagenome
50-genome metagenome
Short reads
Long reads
Short reads
Long reads
Short reads
Long reads
Perfect
Uneven
True
Perfect
Uneven
True
Perfect
Uneven
True
10 simulated metagenomes
10 simulated metagenomes
10 simulated metagenomes
10 simulated metagenomes
10 simulated metagenomes
10 simulated metagenomes
10 simulated metagenomes
10 simulated metagenomes
10 simulated metagenomes

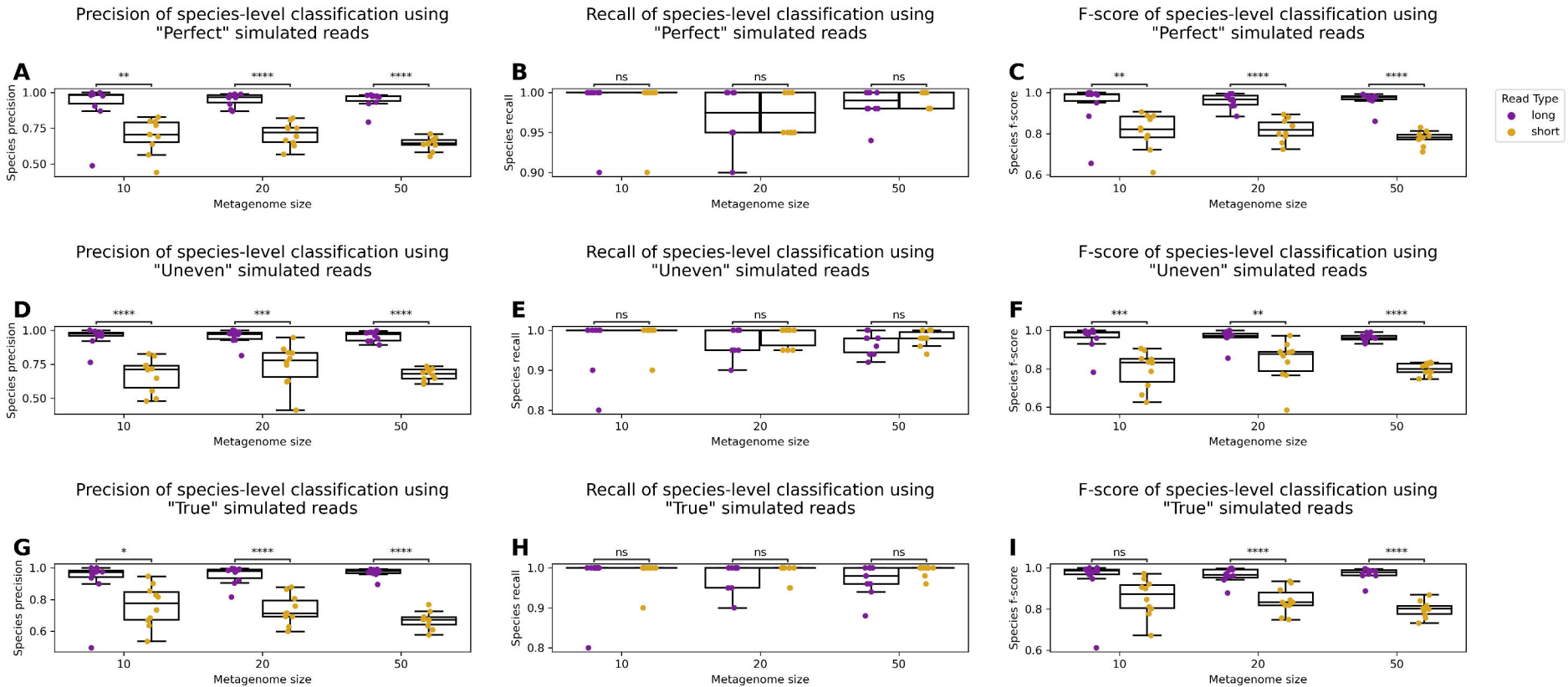
Assembly quality assessment using metaQUAST metrics



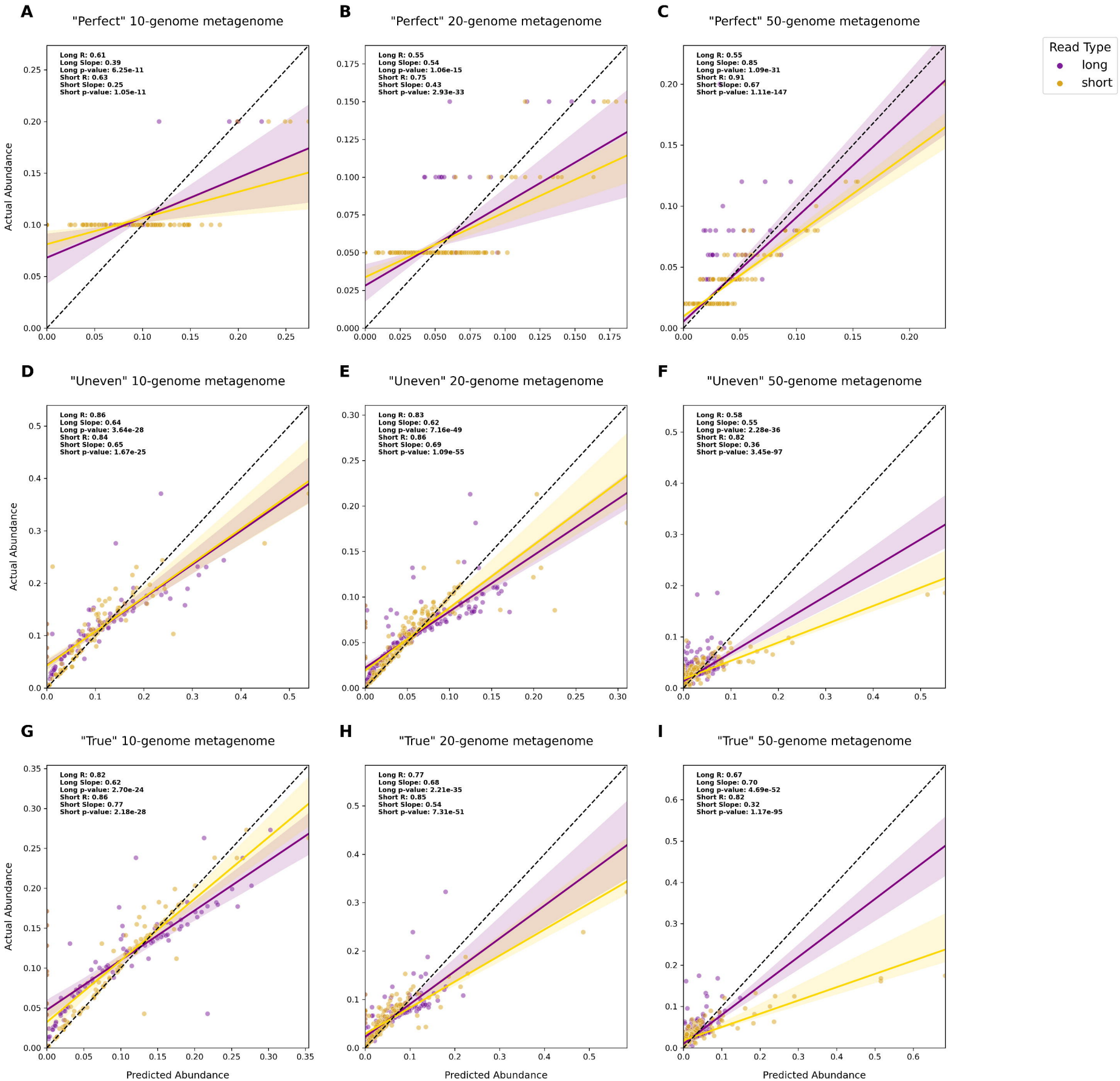
Performance evaluation of Genus-level classification on metagenomic assemblies



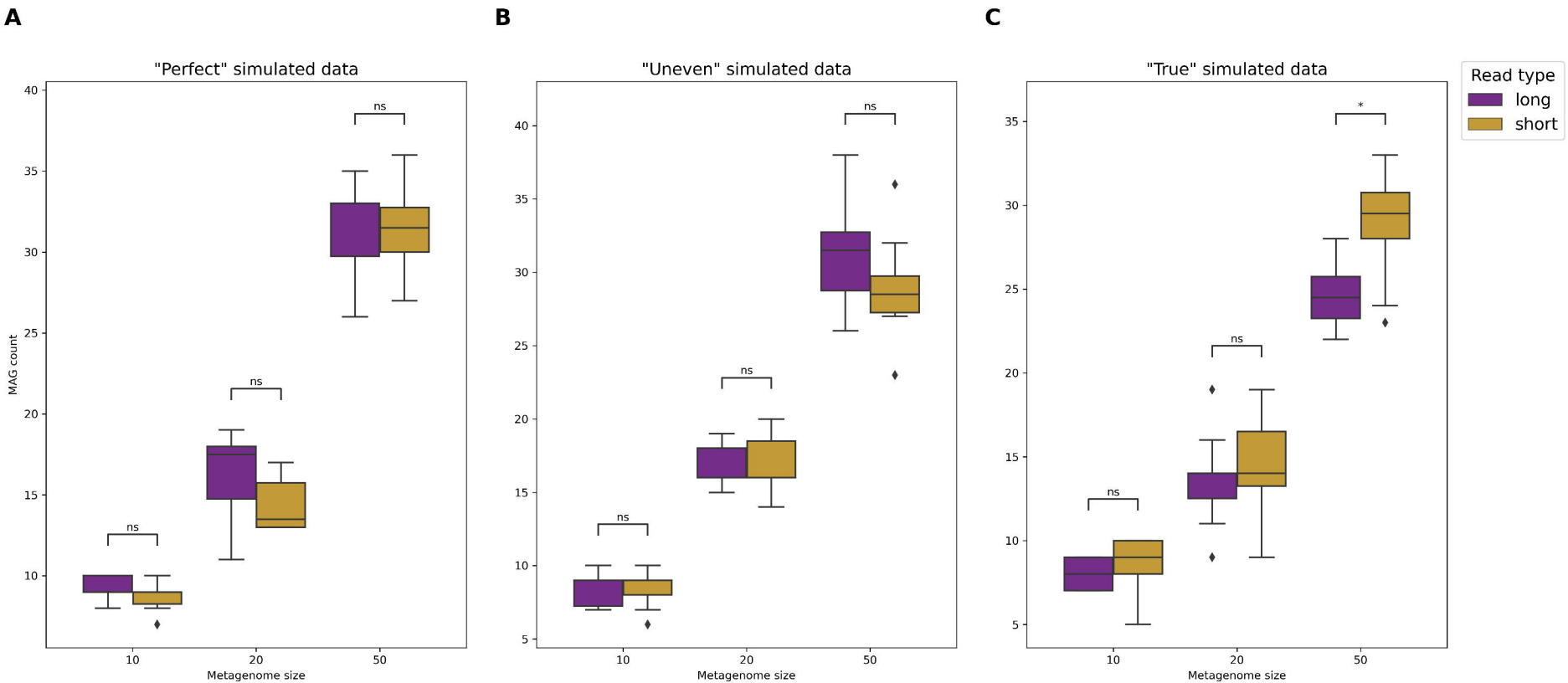
Performance evaluation of Species-level classification on metagenomic assemblies

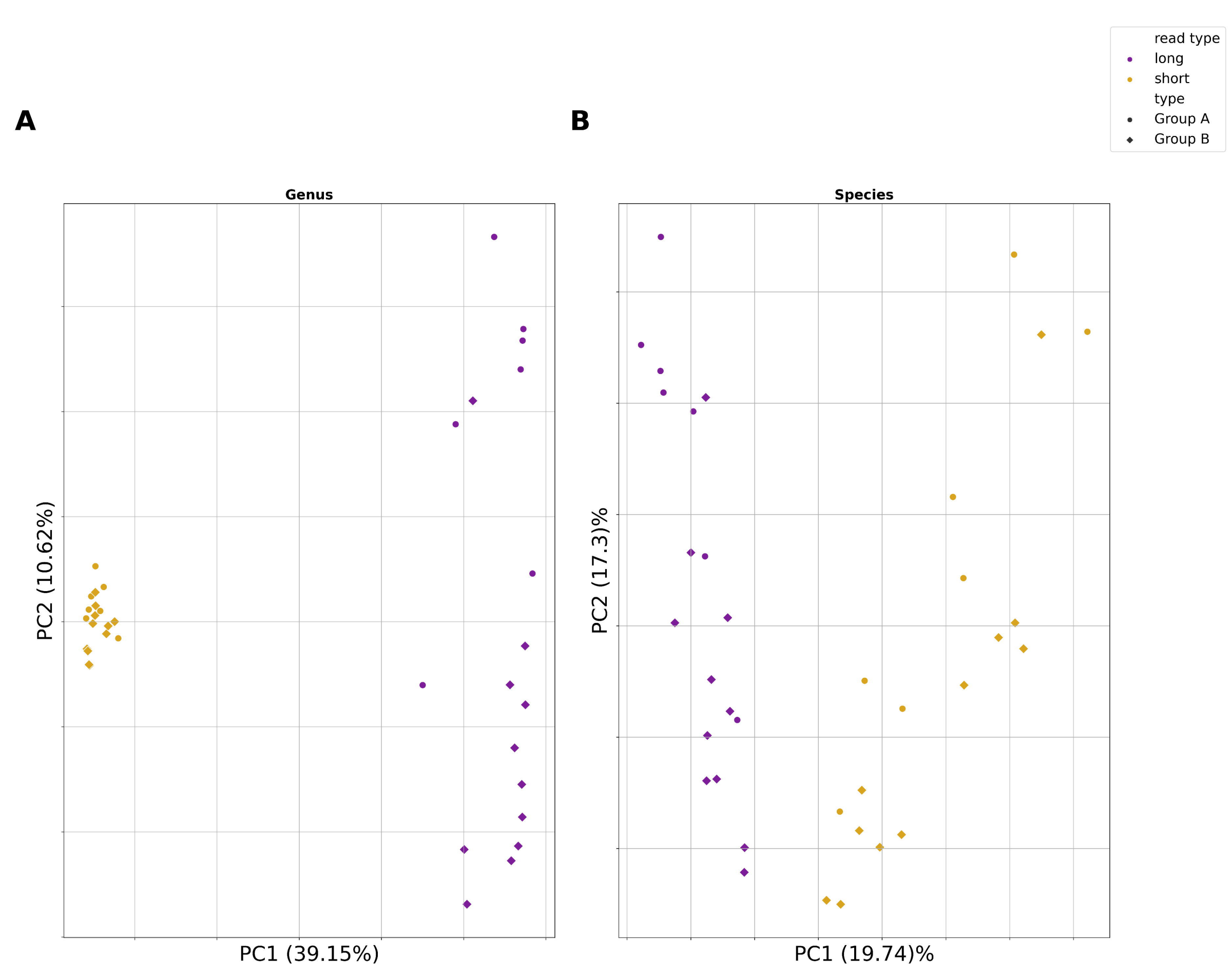


Species-level abundance estimation from short and long reads

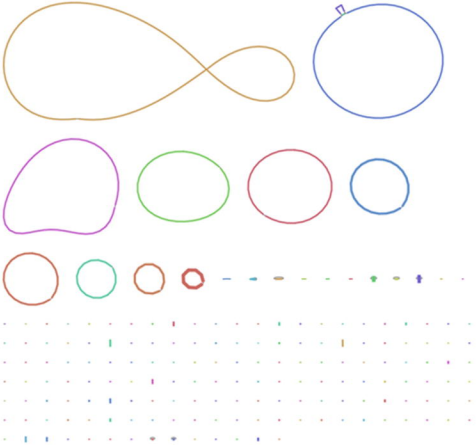


MAG recovery rates from datasets of differing complexity



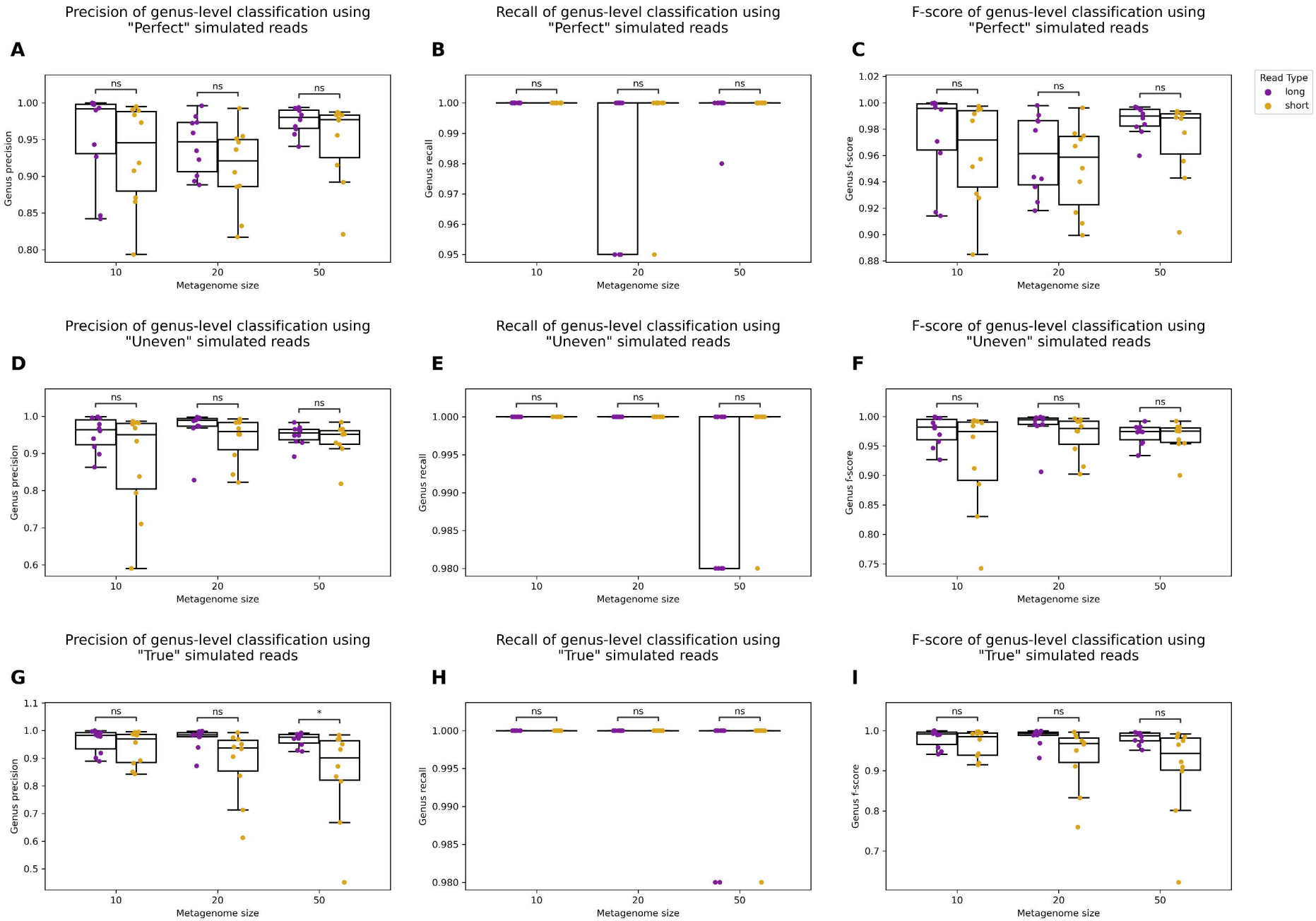

read type
long
short
type
Group A
Group B
A
B
Genus
Species
PC2 (10.62%)
PC1 (39.15%)
PC2 (17.3)%
PC1 (19.74)%

A

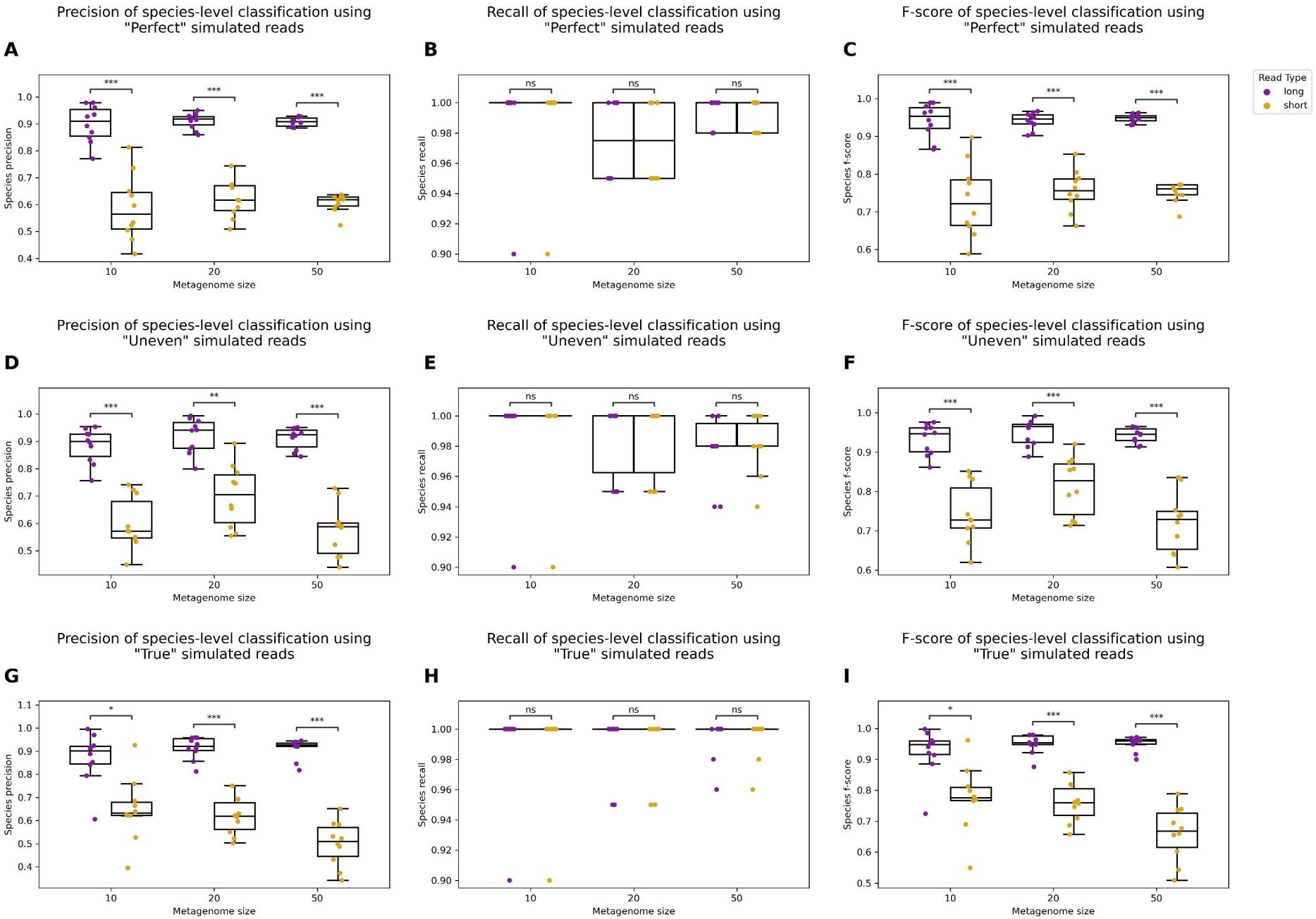


B

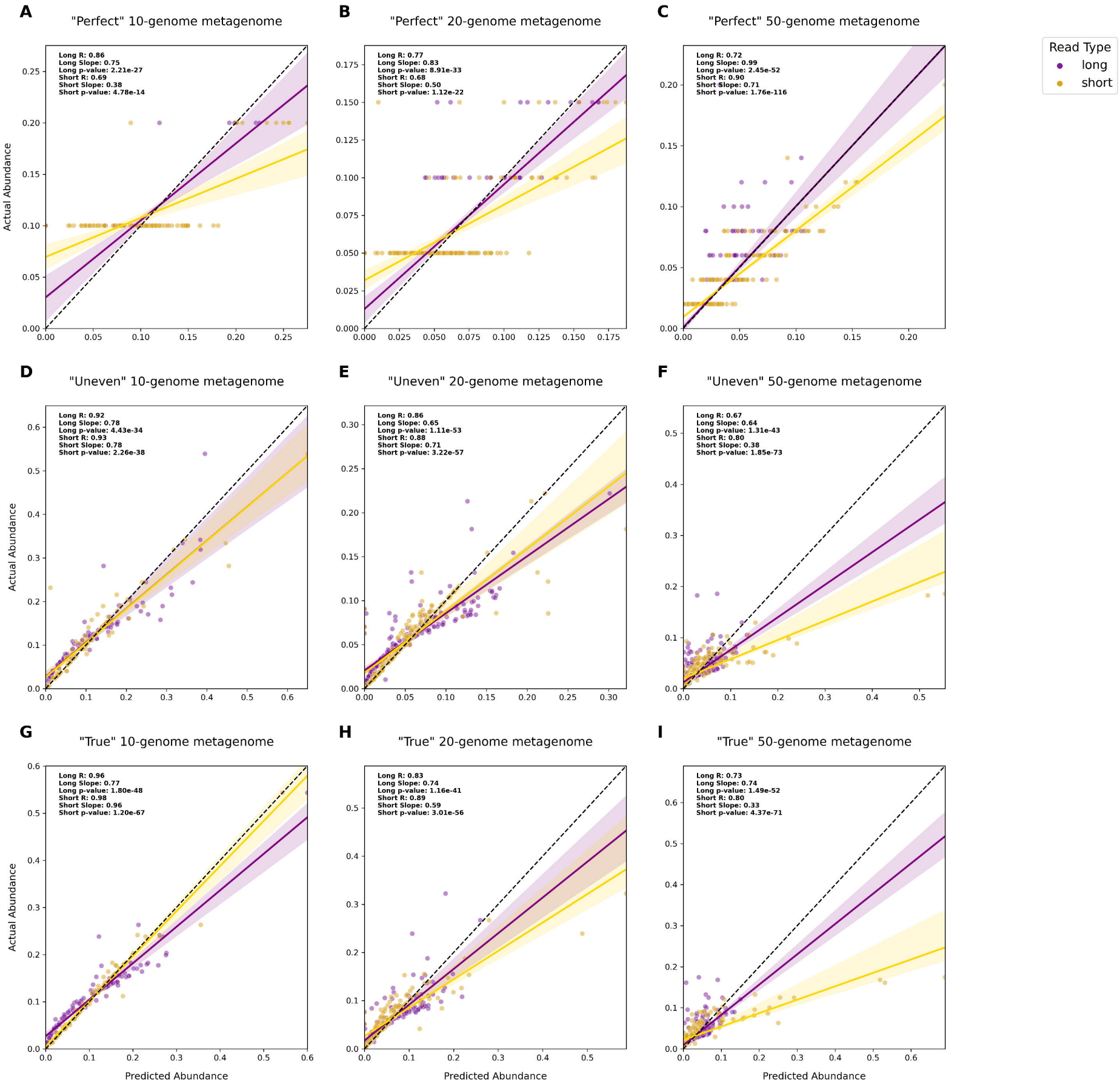
Performance evaluation of Genus-level classification on metagenomic reads



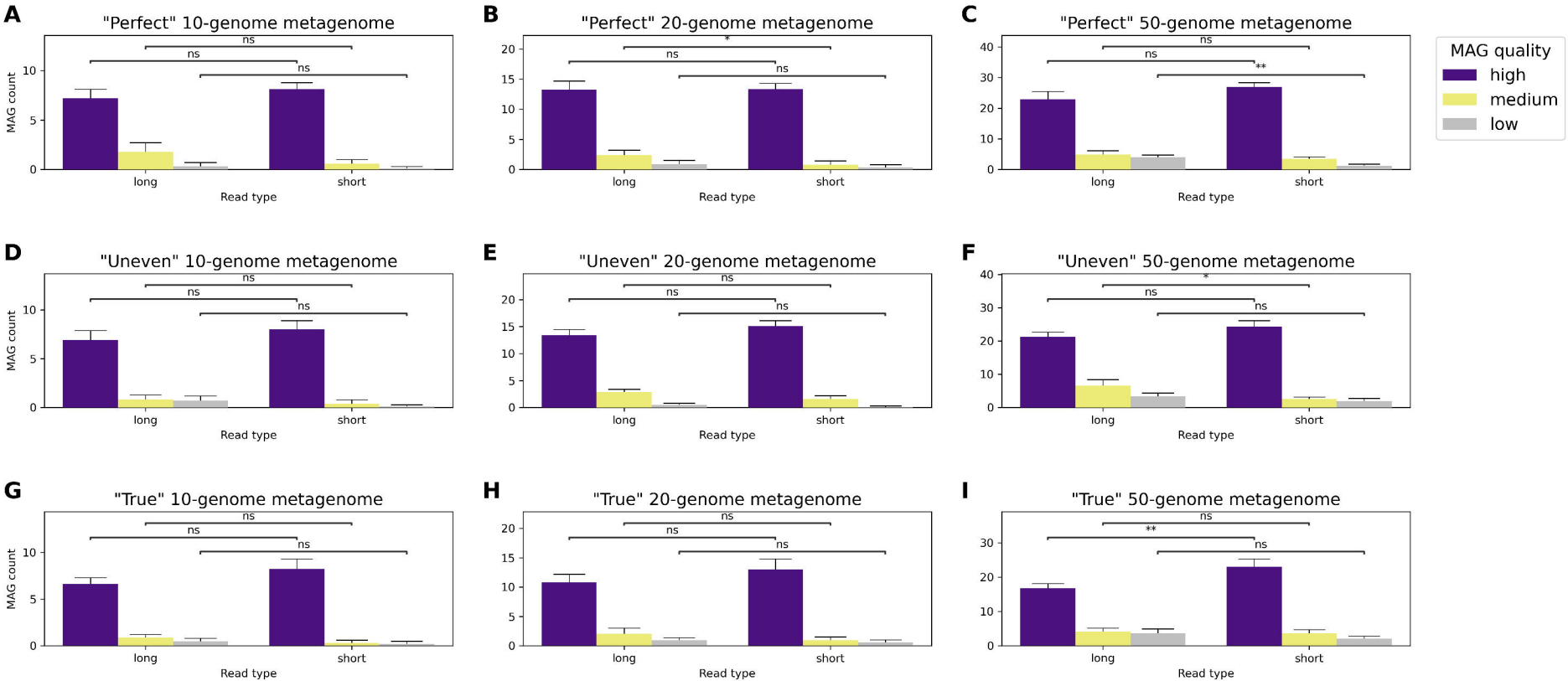
Performance evaluation of Species-level classification on metagenomic reads

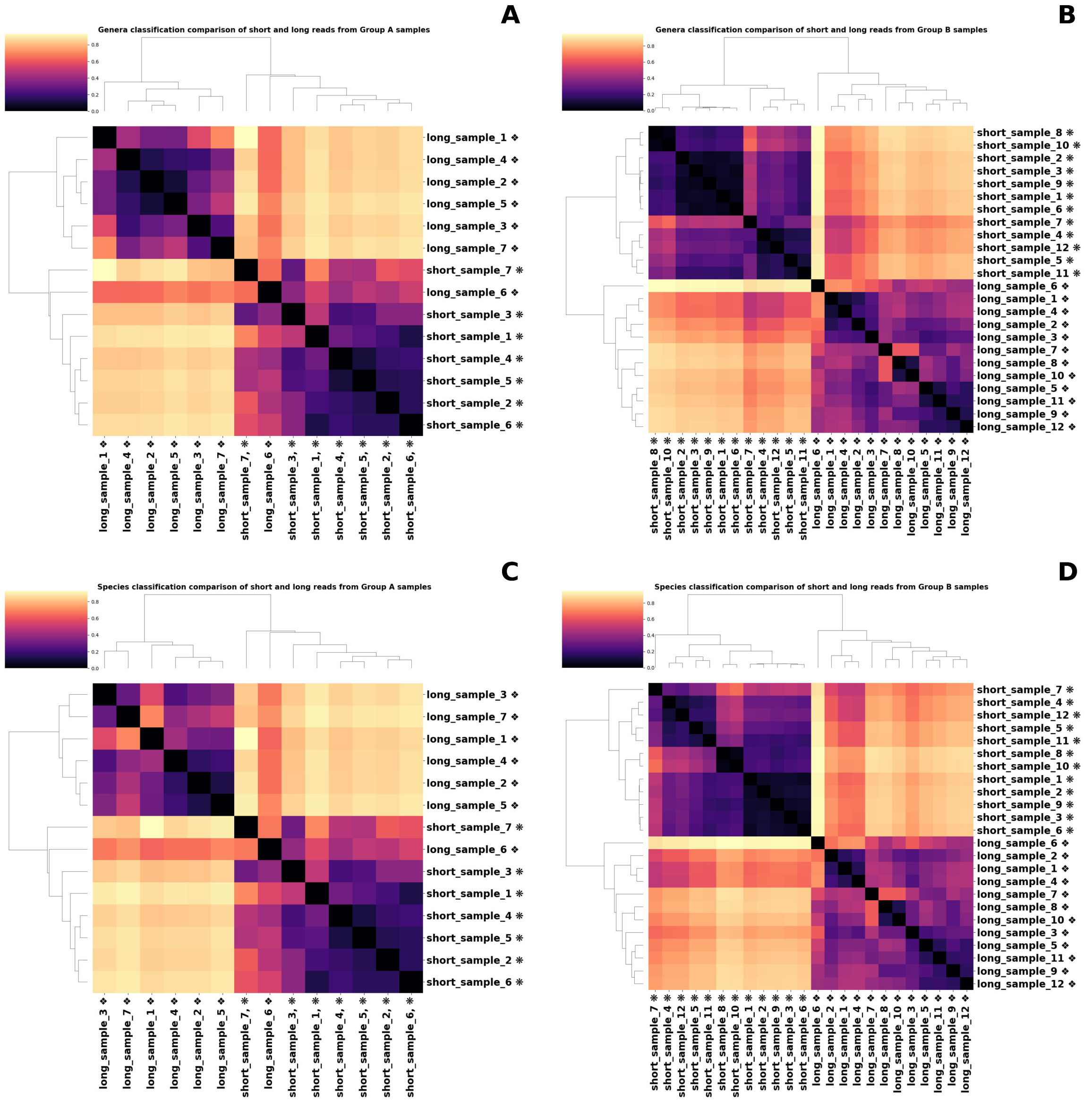


Genus-level abundance estimation from short and long reads



Recovery rates of MAGs of differing qualities from datasets of differing complexity




A
Genera classification comparison of short and long reads from Group A samples
0.8
0.6
0.4
0.2
0.0
long_sample_1 ❖
long_sample_4 ❖
long_sample_2 ❖
long_sample_5 ❖
long_sample_3 ❖
long_sample_7 ❖
short_sample_7 ✳
long_sample_6 ❖
short_sample_3 ✳
short_sample_1 ✳
short_sample_4 ✳
short_sample_5 ✳
short_sample_2 ✳
short_sample_6 ✳
B
Genera classification comparison of short and long reads from Group B samples
0.8
0.6
0.4
0.2
0.0
short_sample_8 ✳
short_sample_10 ✳
short_sample_2 ✳
short_sample_3 ✳
short_sample_9 ✳
short_sample_1 ✳
short_sample_6 ✳
short_sample_7 ✳
short_sample_4 ✳
short_sample_12 ✳
short_sample_5 ✳
short_sample_11 ✳
long_sample_6 ❖
long_sample_1 ❖
long_sample_4 ❖
long_sample_2 ❖
long_sample_3 ❖
long_sample_7 ❖
long_sample_8 ❖
long_sample_10 ❖
long_sample_5 ❖
long_sample_11 ❖
long_sample_9 ❖
long_sample_12 ❖
C
Species classification comparison of short and long reads from Group A samples
0.8
0.6
0.4
0.2
0.0
long_sample_3 ❖
long_sample_7 ❖
long_sample_1 ❖
long_sample_4 ❖
long_sample_2 ❖
long_sample_5 ❖
short_sample_7 ✳
long_sample_6 ❖
short_sample_3 ✳
short_sample_1 ✳
short_sample_4 ✳
short_sample_5 ✳
short_sample_2 ✳
short_sample_6 ✳
D
Species classification comparison of short and long reads from Group B samples
0.8
0.6
0.4
0.2
0.0
short_sample_7 ✳
short_sample_4 ✳
short_sample_12 ✳
short_sample_5 ✳
short_sample_11 ✳
short_sample_8 ✳
short_sample_10 ✳
short_sample_1 ✳
short_sample_2 ✳
short_sample_9 ✳
short_sample_3 ✳
short_sample_6 ✳
long_sample_6 ❖
long_sample_2 ❖
long_sample_1 ❖
long_sample_4 ❖
long_sample_7 ❖
long_sample_8 ❖
long_sample_10 ❖
long_sample_3 ❖
long_sample_5 ❖
long_sample_11 ❖
long_sample_9 ❖
long_sample_12 ❖