SUPPLEMENTARY INFORMATION

SUPPLEMENTARY FIGURES

Supplementary Figure S1. The pipeline of taxonomic/functional classification on short NGS reads by MetaCV.

Supplementary Figure S2. Comparisons of read number assigned under or upon genus level by using MetaCV and BlastX.

Supplementary Figure S3. Identity distribution of paired homologous genes between *Anabaena variabilis* and two reference organisms.

Supplementary Figure S4. Display of the relationship between average identities of homologous protein pairs and its classification accuracy MetaCV achieved.

Supplementary Figure S5. Display of the relationship between composition score of MetaCV and BlastX identities on 100 bp short reads classification.

Supplementary Figure S6. Distribution of composition scores of simulated short reads from human chromosome 1 against prokaryotic genes.

Supplementary Figure S7. Relative abundance of taxonomic groups on 109 cohorts.

Supplementary Figure S8. Relative abundance of 30 most enriched genera on 16 cohorts.

Supplementary Figure S9. Relative abundance of 30 most enriched functional categories from KEGG on 16 cohorts.

Supplementary Figure S10. Relative abundance of top 10 enriched genera and functional groups for 70 cohorts from healthy people.

Supplementary Figure S11. Correlation analysis between pairs of function groups and pairs of genera on 70 cohorts from guts of healthy human.

Supplementary Figure S12. Comparison of sensitivities and specificities of MetaCV, Phymm, BlastX and RAPSearch2 on a genus-mask testing.

Supplementary Figure S13. Time consumed by MetaCV on classification of short reads by using different oligopeptide lengths.

Supplementary Figure S14. Comparison of sensitivities and specificities of MetaCV by using different oligopeptide lengths.

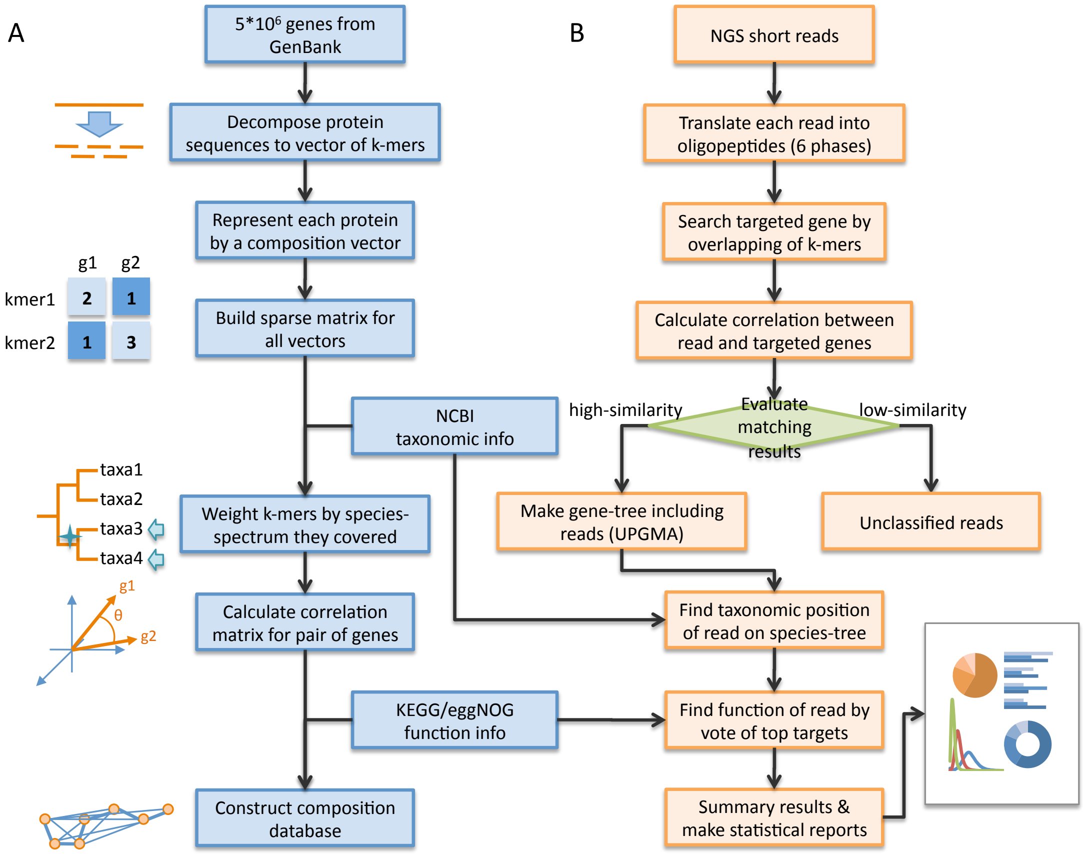
SUPPLEMENTARY TABLES

Supplementary Table S1. Performance comparison of MetaCV, Phymm, BlastX and RAPSearch2 on the classification of 100bp simulated reads from 154 genera.

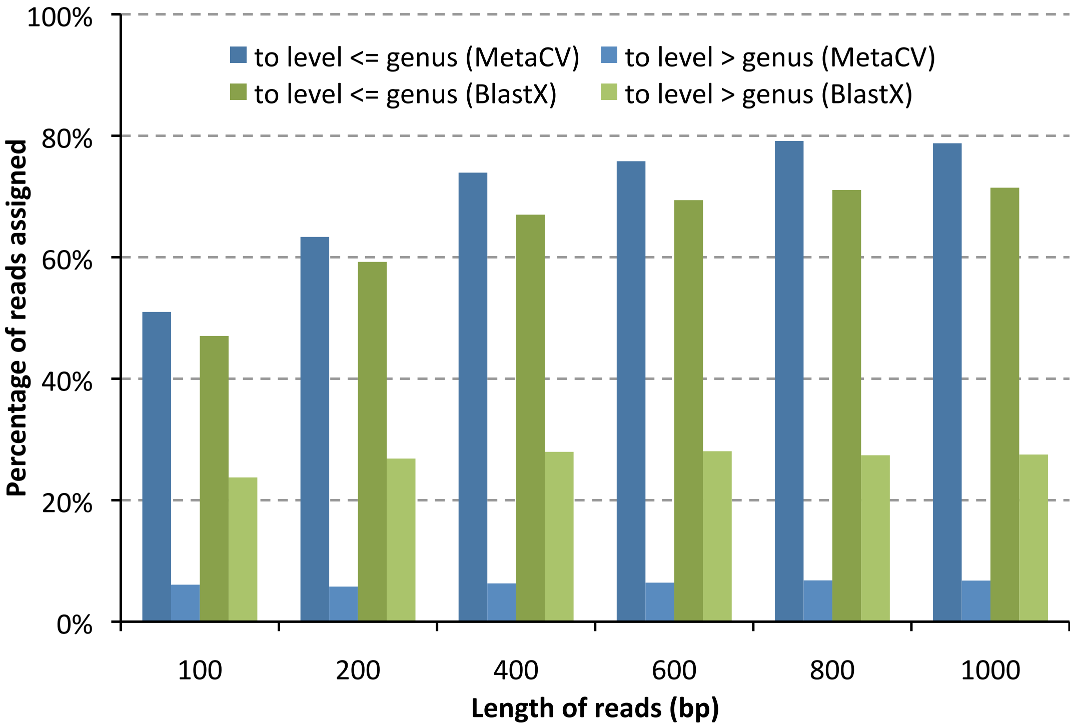
Supplementary Table S2. Species-mask comparison of MetaCV, Phymm, BlastX and RAPSearch2 on the classification of simulated reads of various lengths.

Supplementary Table S3. Performance comparison of MetaCV, Phymm, BlastX and RAPSearch2 on the classification of “randomized metagenomic” data sets.

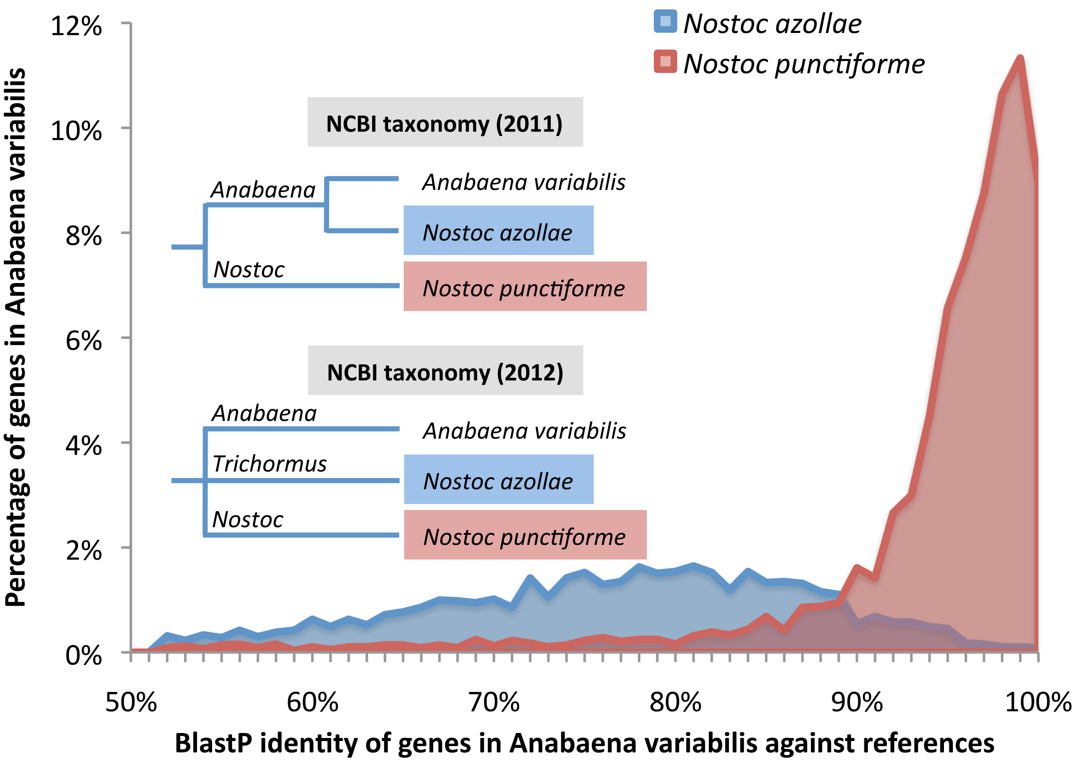
Supplementary Table S4. A list of 44 most abundant genera by MetaCV on 109 cohorts of human gut samples.



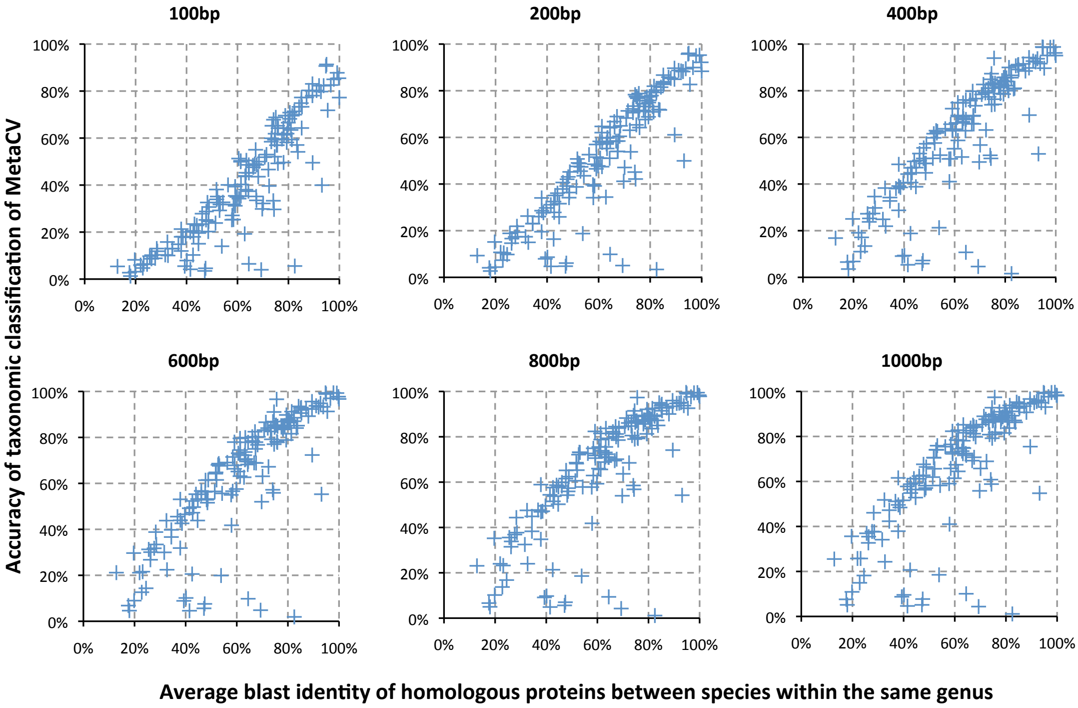
Supplementary Figure S1. The pipeline of taxonomic/functional classification on short NGS reads by MetaCV. (A) The workflow of reference database construction. (B) The comparison steps of NGS reads against reference genes.



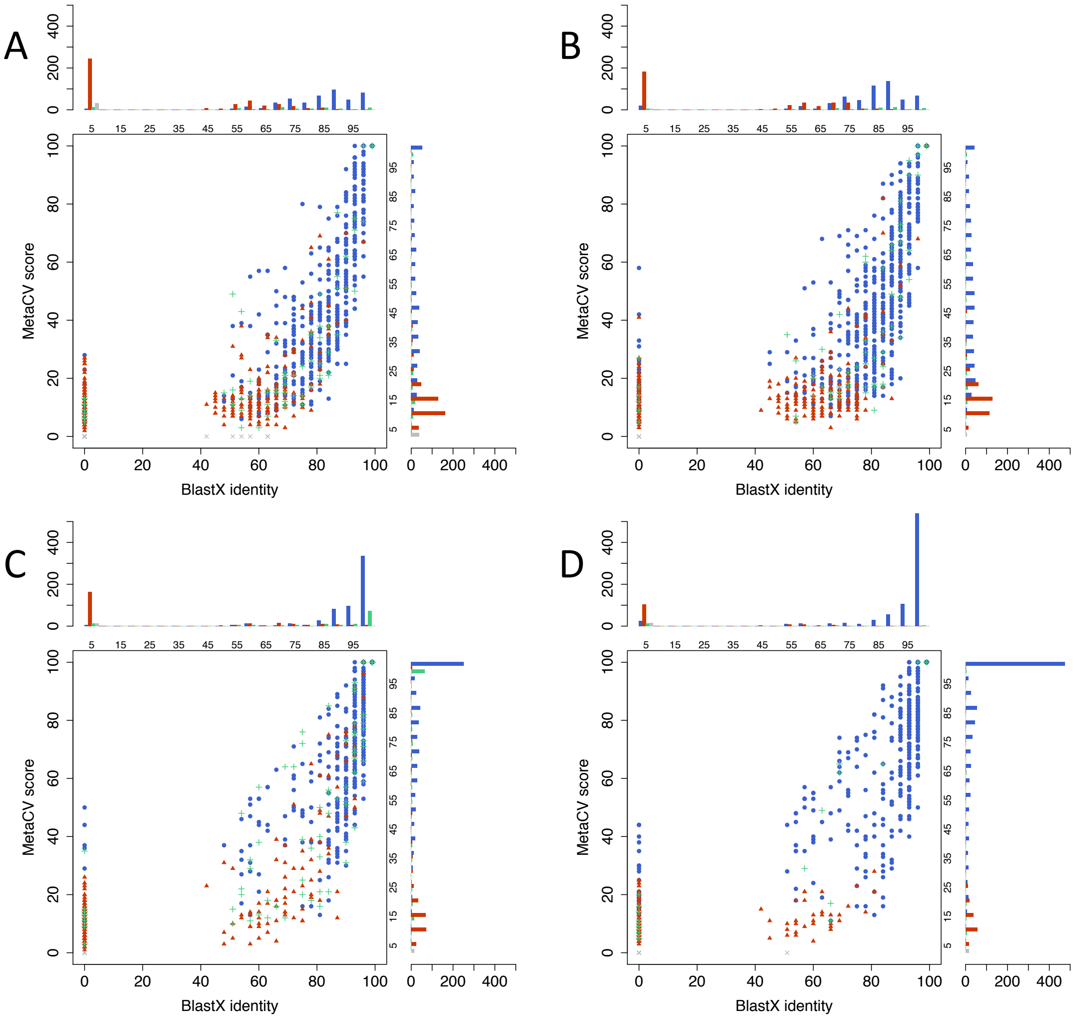
Supplementary Figure S2. Comparisons of read number assigned under or upon genus level by using MetaCV (blue) and BlastX (green) on the species-mask testing (the Dataset 1, see Materials and Methods).



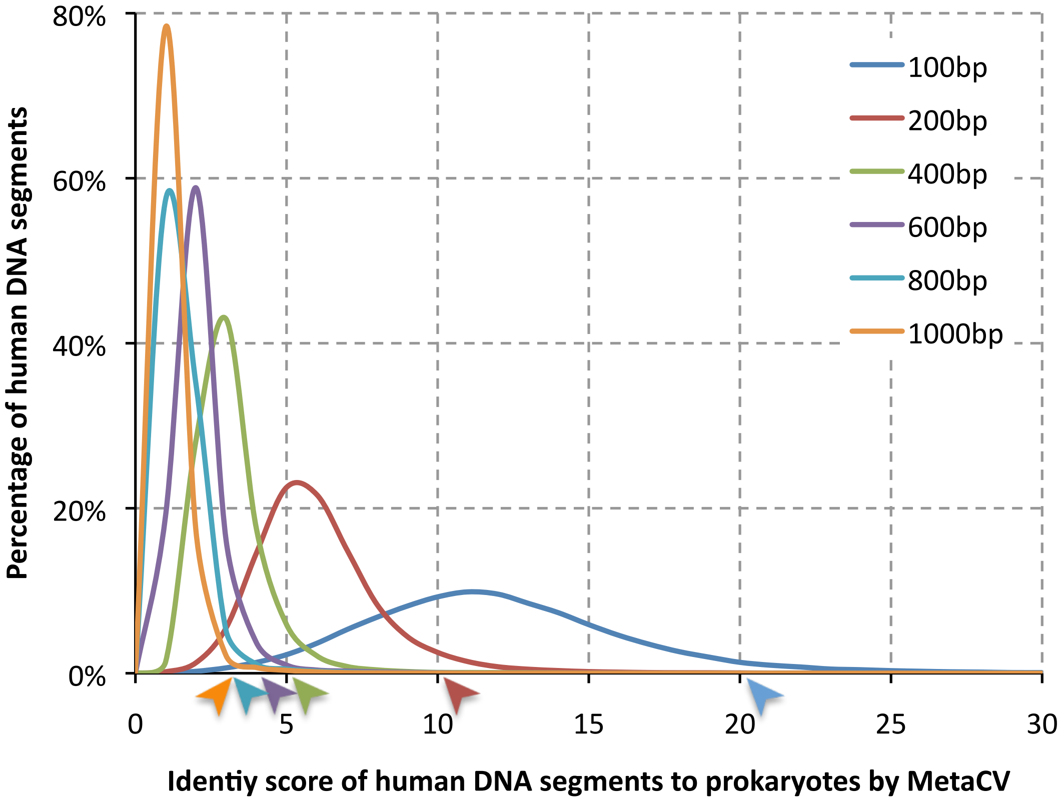
Supplementary Figure S3. Identity distribution of paired homologous genes between *Anabaena variabilis* ATCC 29413 and two reference organisms, *Nostoc azollae* 0708 (in blue) and *Nostoc punctiforme* PCC 73102 (in red). Taxonomic relationships of the three organisms from NCBI in both 2011 and 2012 are given above.



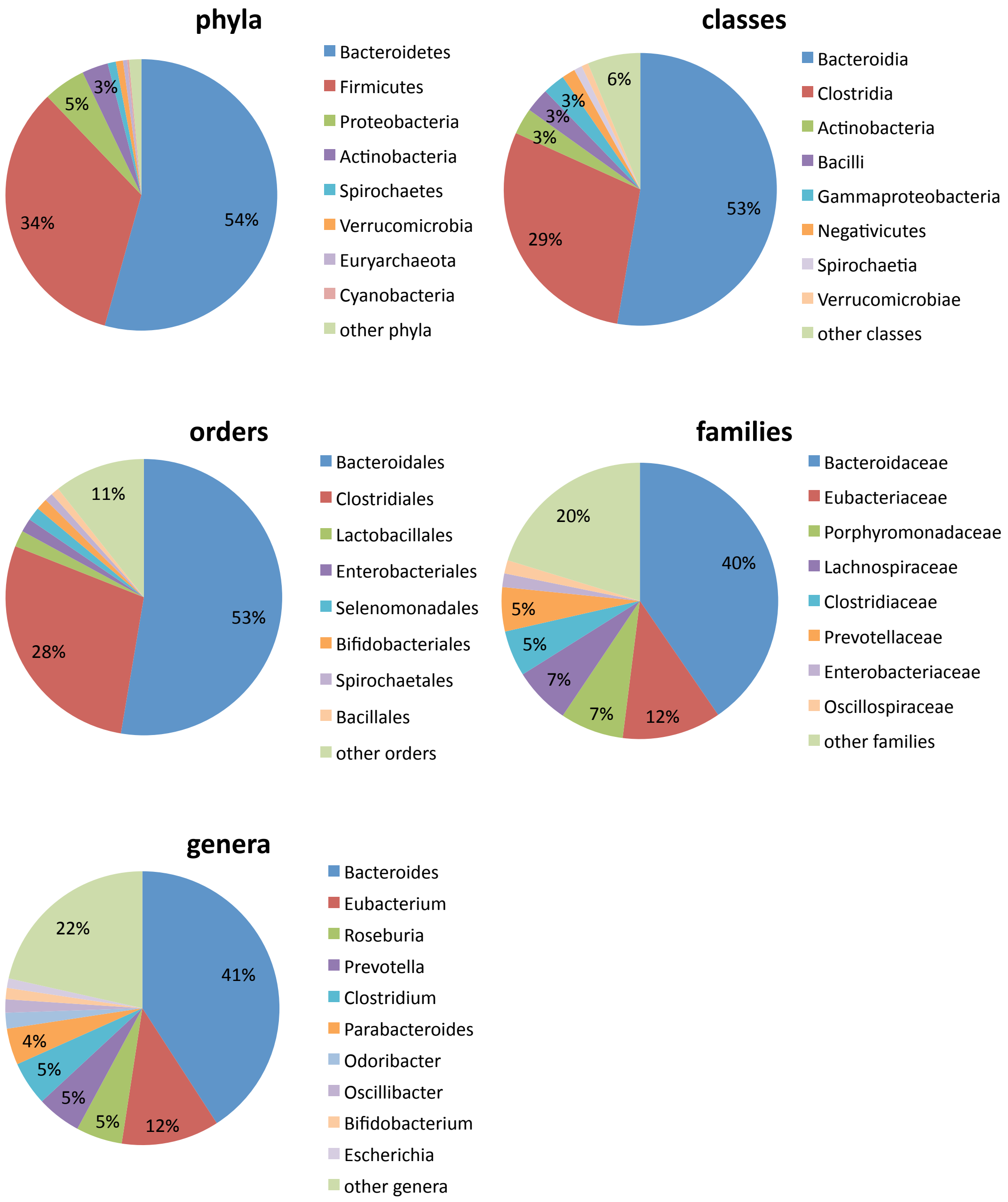
Supplementary Figure S4. Display of the relationship between average identities of homologous protein pairs (by BlastP) and its classification accuracy MetaCV achieved. Each dot denotes one species. The more similar of the sequences between the species within a genus, the more accurately MetaCV could assign the reads.



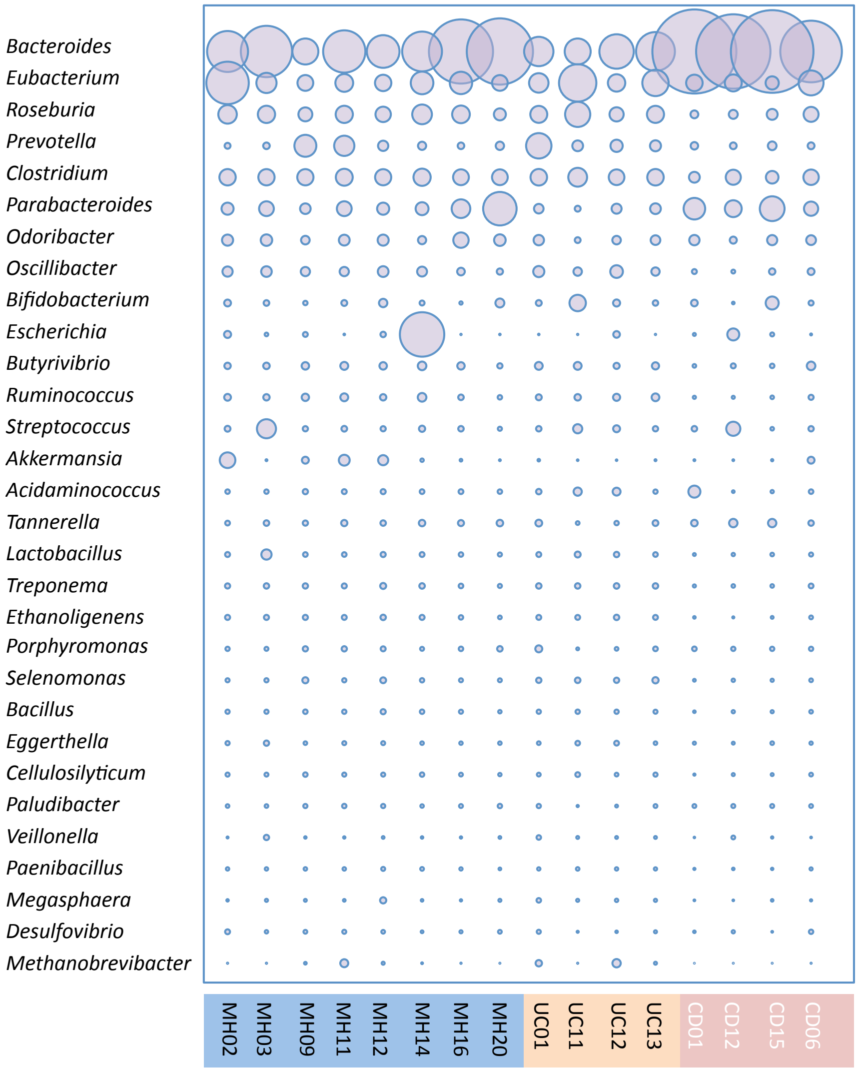
Supplementary Figure S5. Display of the relationship between composition scores of MetaCV and BlastX identities on 100 bp short reads classification (on the species-mask comparison) from (A) *Leuconostoc* which has a 60% protein similarity on average to its sister species, while (B) *Pyrobaculum*, (C) *Pectobacterium* and (D) *Caldicellulosiruptor* have 70%, 80% and 90% to their sister species respectively. Reads, correctly assigned by MetaCV, are represented by blue dots, while red and green dots denote reads mistakenly classified or assigned to taxa higher than genus level, respectively. Gray dots denote reads with no hits by MetaCV. In each sub-figure, distributions of reads on BlastX identities are displayed on the top and that on MetaCV scores are given in the right.



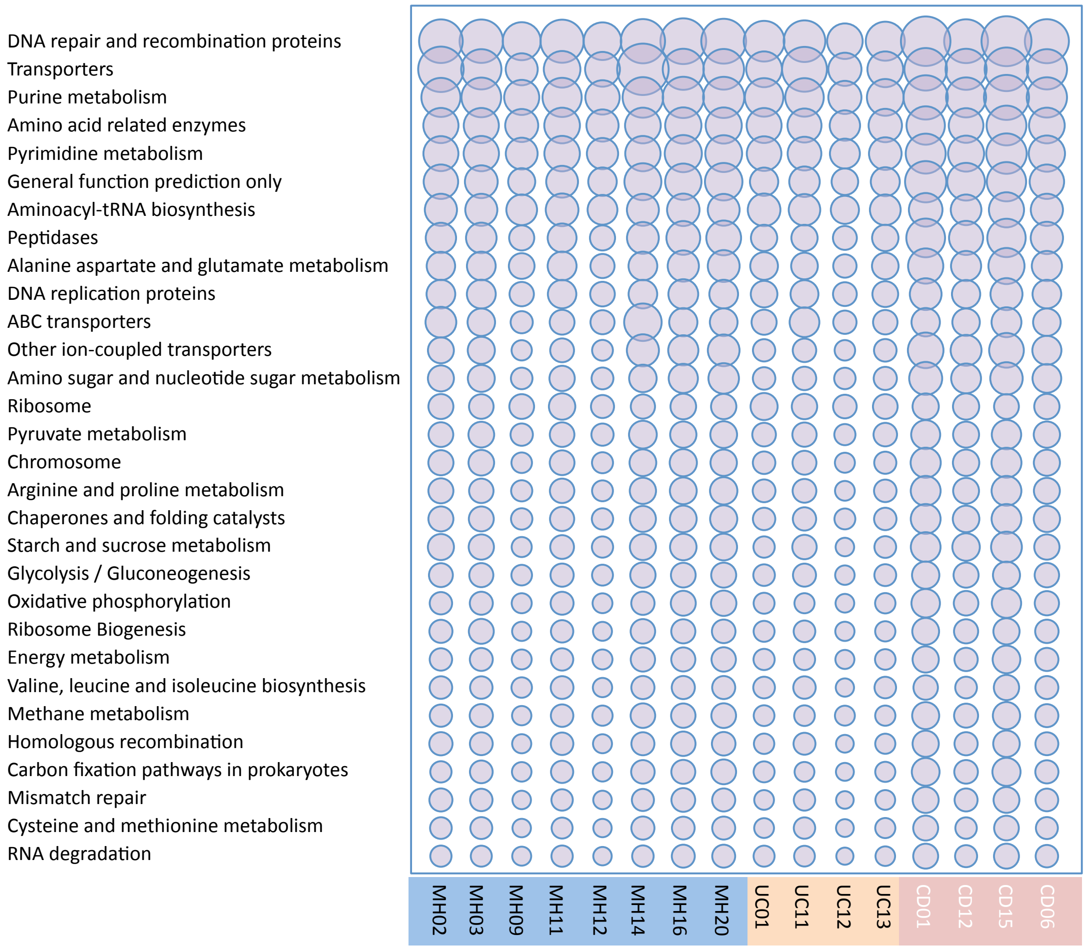
Supplementary Figure S6. Distribution of composition scores of simulated short reads from human chromosome 1 against $5*10^6$ genes from 1691 prokaryotic organisms assigned by MetaCV. Cutoff value of composition score could be set as 20 for 100 bp, 10 for 200 bp, 5 for 400 bp, 4 for 600 bp, 3 for both 800 and 1 kb read binning, as indicated by the arrows below, to control the false positive rate to lower than 5%.



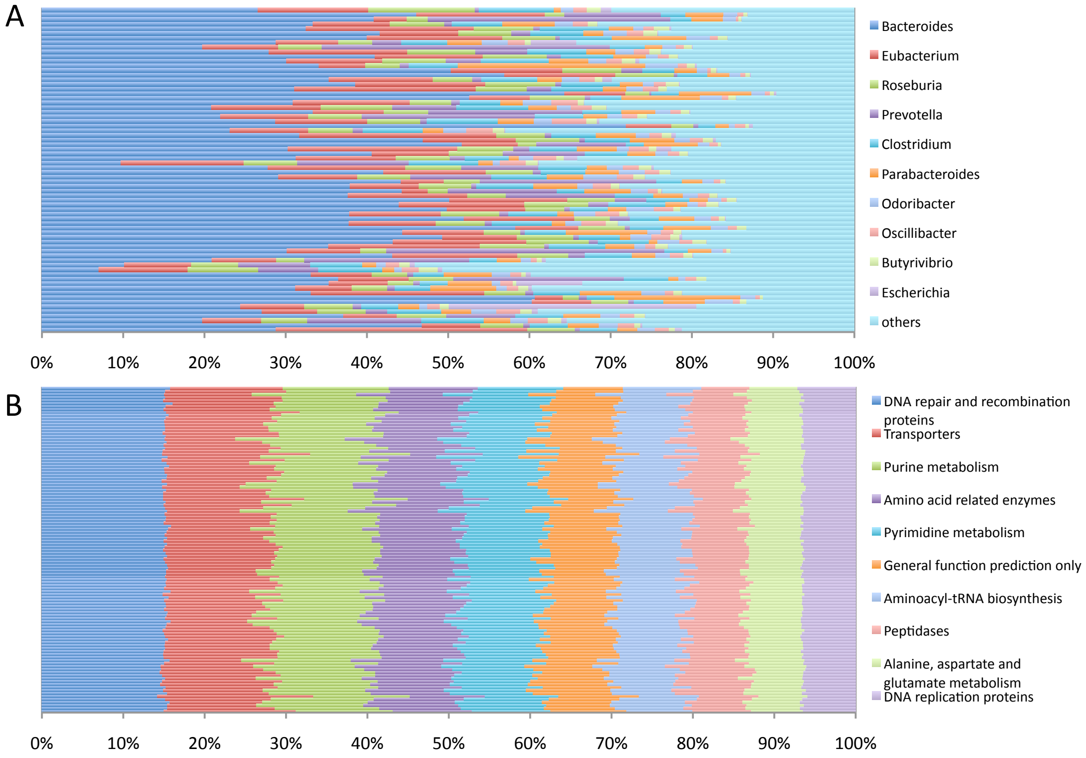
Supplementary Figure S7. Relative abundance of taxonomic groups from phylum to genus on 109 cohorts of human gut samples on average.



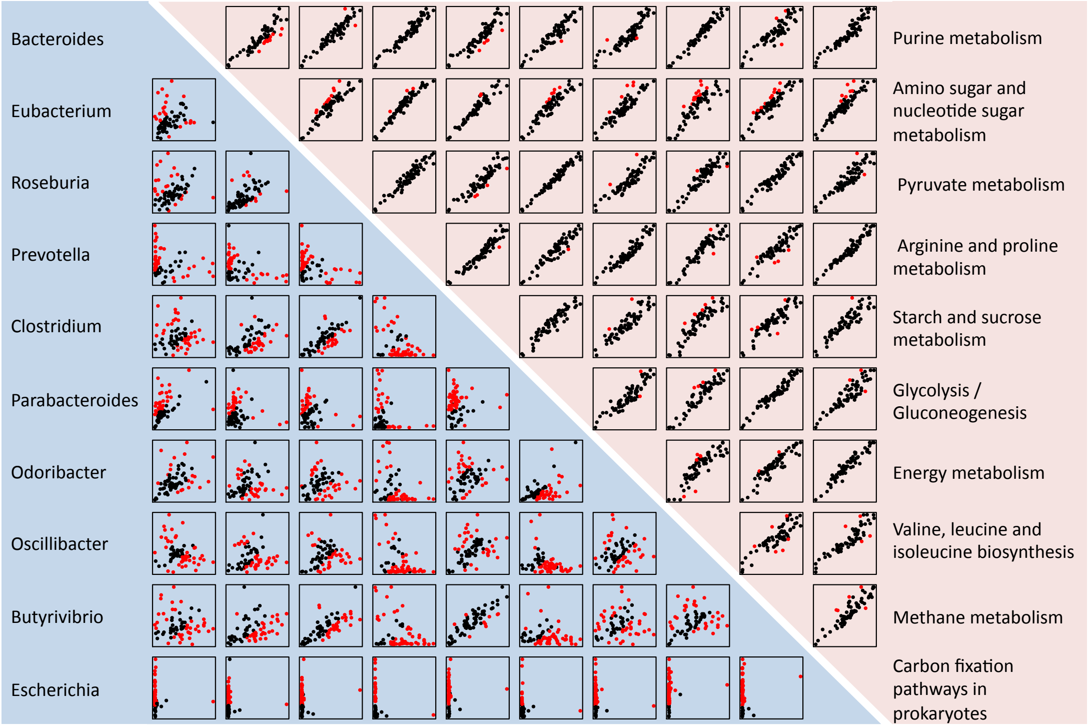
Supplementary Figure S8. Relative abundance of 30 most enriched genera on 16 cohorts, among which 8 samples are from healthy human (in blue), 4 samples are from ulcerative colitis patients (UC in orange) and 4 from Crohn's disease (CD in pink).



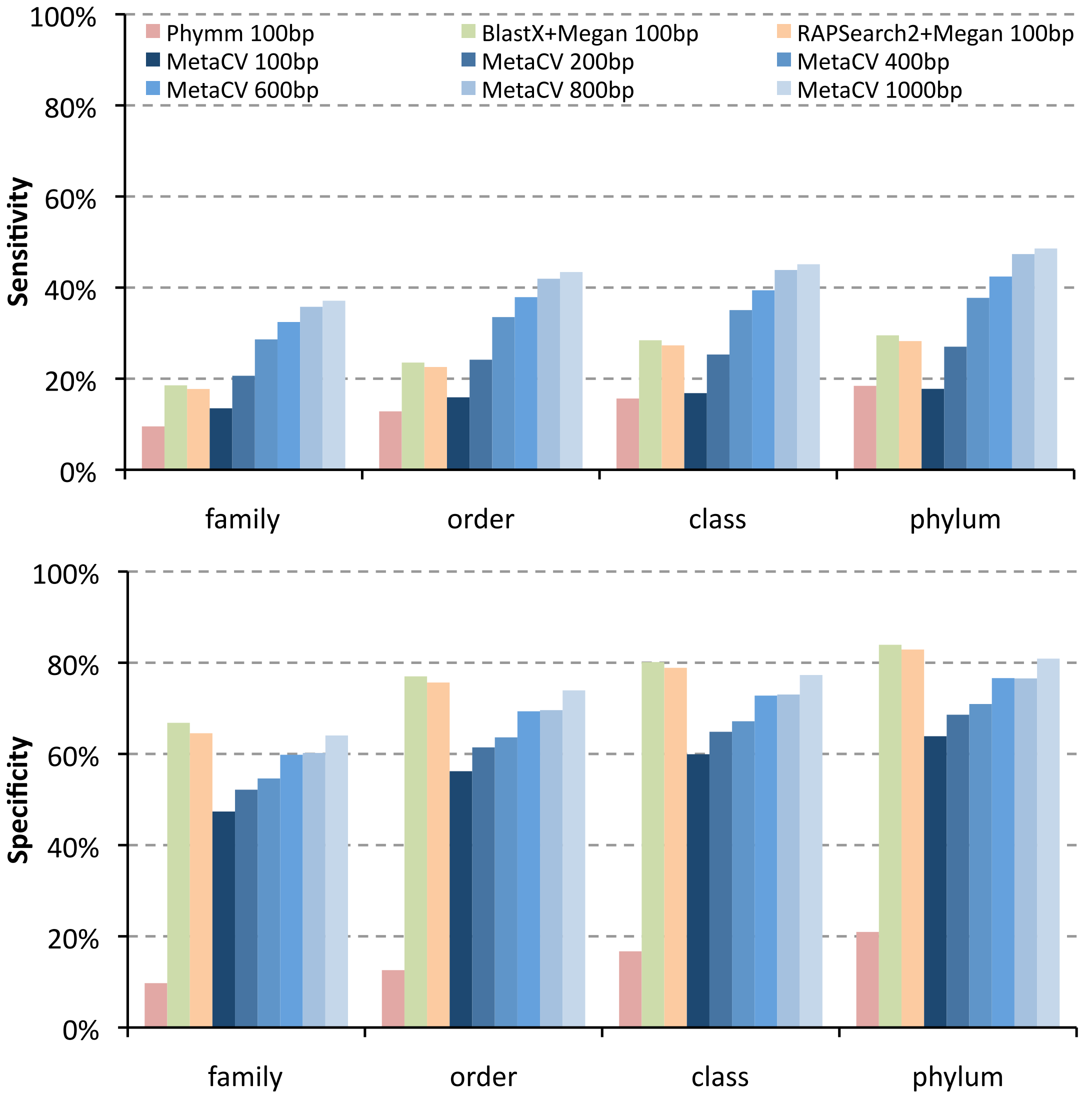
Supplementary Figure S9. Relative abundance of 30 most enriched functional categories from KEGG on 16 cohorts, samples are the same with those in Figure S8.



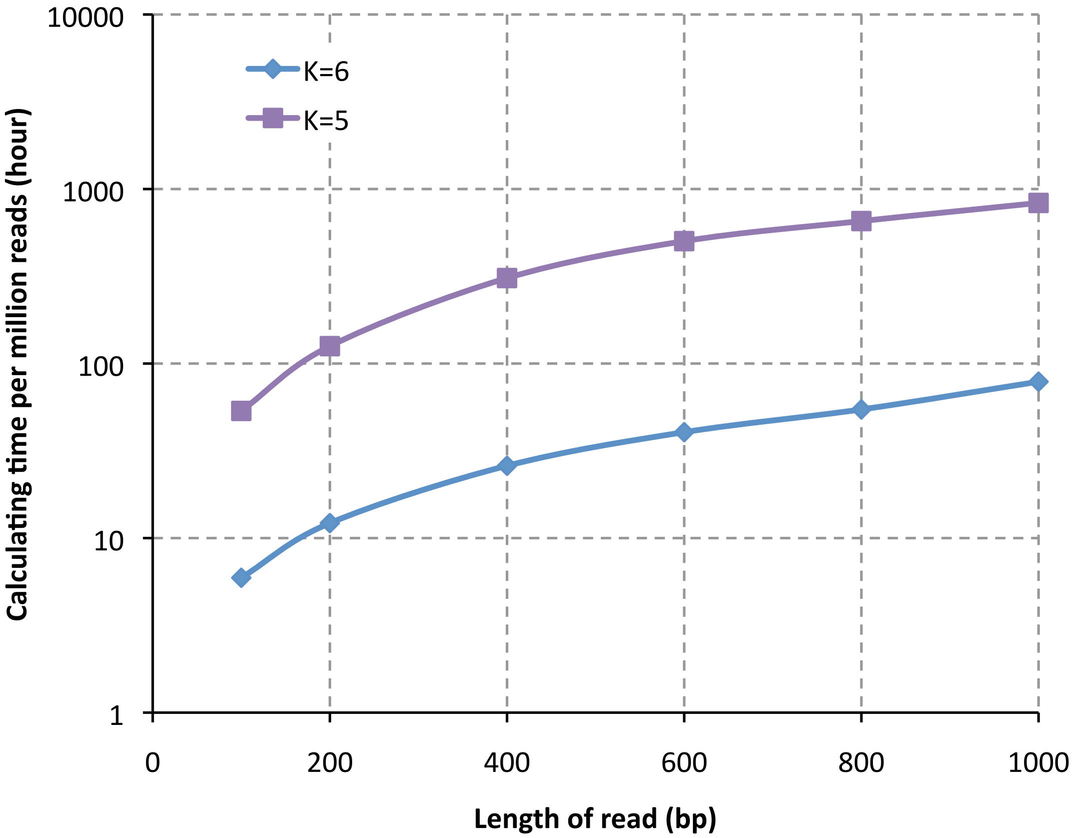
Supplementary Figure S10. Relative abundance of top 10 enriched genera (upper figure) and functional groups (lower figure) for 70 cohorts of healthy people. Each line stands for one sample.



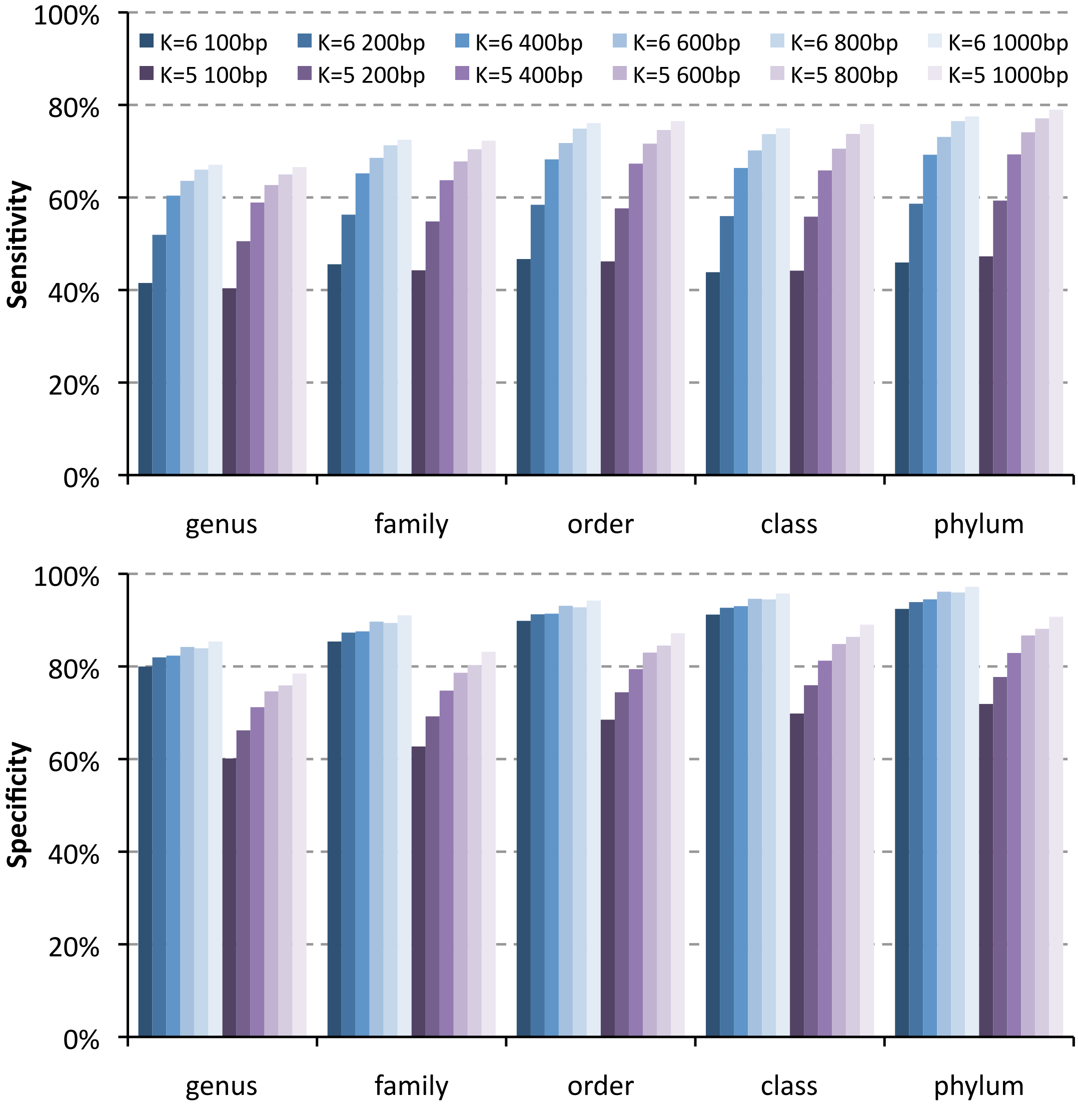
Supplementary Figure S11. Correlation analysis between pairs of function groups (upright panel) and pairs of genera (downleft panel) on 70 cohorts from gut of healthy human. Each dot denotes a sample, and black dots are samples with linear correlation between pairs of groups (with 5% divergence tolerance) and red dots are samples without proper linear correlation. Note that only top 10 most enriched genera/function groups are displayed.



Supplementary Figure S12. Sensitivity and specificity comparisons of MetaCV, Phymm, BlastX and RAPSearch2 on a genus-mask testing. Ratio values of MetaCV, Phymm, BlastX and RAPSearch2 are colored by blue, red, green and organge, respectively. Four groups from left to right are the results of these methods on different taxonomic level (family to phylum, respectively). Only MetaCV is applied to classify simulated reads longer than 100 bp.



Supplementary Figure S13. Time consumed by MetaCV on the classification of short reads by using different oligopeptide lengths, of which K=6 was colored by blue and K=5 by purple. K values longer than 6 are not considered in this analysis because the dimensions of compositions are too high to make databases with affordable size.



Supplementary Figure S14. Sensitivity and specificity comparisons of MetaCV by using different oligopeoptide lengths on the species-mask testing (the Dataset 1), where K=6 was colored by blue and K=5 by purple. Five groups from left to right are the results on different taxonomic levels (from genus to phylum, respectively).

Supplementary Table S1. Performace comparison of MetaCV, Phymm, BlastX and RAPSearch2 on the classification of 100bp simulated reads from 154 genera. For each genus, a genome from one species is used to simulate reads while the organisms from other species are for reference profiles.

	MetaCV		Phymm		BlastX+Megan		RAPSearch2+Megan	
Genus	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Specificity	Specificity
Treponema	92.7 ± 0.5%	80.4 ± 0.7%	81.3 ± 1.4%	36.8 ± 0.6%	90.0 ± 1.1%	78.5 ± 1.0%	89.1 ± 0.9%	78.2 ± 1.1%
Rickettsia	89.5 ± 0.6%	98.2 ± 0.2%	87.3 ± 1.1%	27.1 ± 0.6%	88.2 ± 0.8%	99.1 ± 0.2%	88.9 ± 0.6%	99.1 ± 0.2%
Mycobacterium	91.7 ± 0.4%	71.5 ± 1.1%	71.7 ± 0.5%	18.0 ± 0.4%	84.7 ± 0.8%	85.6 ± 0.9%	79.0 ± 1.3%	85.6 ± 0.3%
Ureaplasma	70.7 ± 1.4%	99.6 ± 0.1%	36.0 ± 0.9%	71.5 ± 2.3%	83.7 ± 1.2%	100.0 ± 0.0%	80.7 ± 0.9%	99.9 ± 0.1%
Borrelia	79.7 ± 0.9%	98.8 ± 0.3%	54.0 ± 0.8%	30.0 ± 0.6%	82.4 ± 0.9%	99.9 ± 0.2%	80.3 ± 0.9%	99.9 ± 0.2%
Francisella	84.5 ± 0.2%	97.7 ± 0.7%	55.4 ± 0.3%	52.6 ± 1.5%	81.4 ± 1.0%	99.2 ± 0.3%	80.0 ± 1.1%	98.9 ± 0.5%
Bacillus	85.5 ± 0.5%	76.7 ± 1.0%	68.9 ± 1.9%	13.2 ± 0.4%	79.2 ± 1.0%	84.2 ± 1.5%	78.1 ± 0.9%	83.1 ± 1.2%
Caldicellulosiruptor	82.2 ± 1.1%	93.7 ± 0.7%	63.3 ± 0.4%	33.6 ± 0.6%	78.7 ± 1.2%	95.4 ± 0.3%	77.8 ± 1.0%	95.4 ± 0.7%
Neorickettsia	74.7 ± 1.2%	99.0 ± 0.4%	26.0 ± 2.1%	62.0 ± 2.3%	78.2 ± 1.6%	99.9 ± 0.1%	77.0 ± 1.2%	99.9 ± 0.1%
Brachyspira	76.5 ± 1.5%	95.8 ± 0.5%	52.7 ± 1.7%	48.4 ± 1.4%	77.5 ± 0.7%	98.6 ± 0.6%	75.7 ± 0.9%	98.2 ± 0.7%
Burkholderia	83.7 ± 0.6%	56.2 ± 0.7%	83.0 ± 1.2%	24.0 ± 0.4%	77.4 ± 0.6%	68.2 ± 0.8%	72.9 ± 1.1%	66.2 ± 0.8%
Acidiphilium	80.9 ± 1.2%	94.3 ± 0.7%	42.7 ± 1.8%	60.8 ± 0.8%	75.2 ± 1.4%	96.6 ± 0.5%	71.4 ± 1.3%	96.3 ± 0.5%
Thermoanaerobacter	87.9 ± 1.0%	91.6 ± 0.7%	77.5 ± 0.9%	44.5 ± 0.9%	75.1 ± 1.1%	90.7 ± 0.5%	73.9 ± 0.9%	90.4 ± 0.4%
Listeria	79.0 ± 0.5%	96.2 ± 0.4%	26.1 ± 1.5%	29.2 ± 0.8%	74.9 ± 0.9%	96.6 ± 0.6%	74.1 ± 1.2%	96.4 ± 0.7%
Brucella	89.5 ± 1.1%	96.4 ± 1.2%	91.2 ± 0.7%	44.6 ± 0.3%	74.6 ± 1.4%	97.7 ± 0.2%	73.2 ± 1.3%	97.6 ± 0.6%
Myxococcus	69.2 ± 2.1%	91.3 ± 1.1%	47.2 ± 0.2%	66.2 ± 1.5%	74.4 ± 2.1%	97.9 ± 0.6%	70.8 ± 2.1%	97.9 ± 0.4%
Bordetella	83.3 ± 0.9%	90.6 ± 0.9%	58.5 ± 1.4%	35.4 ± 0.9%	74.3 ± 1.1%	95.6 ± 1.0%	71.1 ± 0.6%	95.0 ± 1.1%
Streptococcus	75.2 ± 1.2%	79.5 ± 1.3%	55.2 ± 0.2%	9.7 ± 0.1%	74.2 ± 1.0%	84.4 ± 1.1%	74.0 ± 1.2%	83.6 ± 1.0%
Dictyoglomus	62.8 ± 1.4%	97.4 ± 0.5%	17.8 ± 1.5%	45.7 ± 2.7%	73.9 ± 0.4%	98.7 ± 0.4%	72.3 ± 0.7%	98.6 ± 0.5%
Haloarcula	70.3 ± 0.9%	98.2 ± 0.7%	42.3 ± 1.0%	65.8 ± 2.2%	73.9 ± 1.0%	99.6 ± 0.2%	71.2 ± 1.1%	99.5 ± 0.3%
Helicobacter	69.0 ± 1.3%	90.4 ± 1.4%	71.7 ± 1.7%	49.4 ± 1.1%	73.1 ± 0.6%	96.3 ± 0.5%	71.6 ± 0.8%	95.7 ± 0.6%
Methanothermobacter	60.6 ± 1.4%	95.8 ± 0.8%	53.0 ± 0.6%	67.6 ± 1.4%	72.7 ± 1.2%	97.0 ± 0.6%	70.6 ± 1.6%	96.5 ± 0.5%
Neisseria	78.8 ± 1.5%	96.2 ± 0.5%	76.2 ± 1.7%	65.4 ± 0.6%	72.5 ± 0.9%	96.5 ± 0.6%	72.2 ± 0.9%	96.7 ± 0.6%
Sulfolobus	68.6 ± 1.6%	89.6 ± 0.9%	29.2 ± 1.2%	31.5 ± 1.4%	72.0 ± 1.9%	92.4 ± 0.6%	70.7 ± 2.3%	91.8 ± 0.6%
Geobacillus	81.5 ± 0.5%	94.6 ± 1.0%	77.3 ± 1.5%	42.9 ± 0.4%	71.8 ± 1.1%	96.9 ± 0.9%	70.8 ± 1.4%	96.7 ± 0.9%
Geobacter	72.2 ± 1.3%	76.3 ± 1.2%	37.8 ± 1.3%	18.3 ± 0.8%	71.7 ± 1.3%	82.8 ± 1.4%	69.8 ± 1.3%	81.7 ± 1.1%
Methanocaldococcus	71.5 ± 1.2%	95.3 ± 1.1%	45.1 ± 0.9%	39.3 ± 0.8%	70.7 ± 1.3%	97.0 ± 0.2%	69.2 ± 1.2%	96.6 ± 0.2%
Staphylothermus	65.7 ± 1.8%	97.0 ± 0.1%	46.9 ± 1.5%	62.7 ± 1.9%	70.0 ± 2.2%	98.8 ± 0.4%	68.7 ± 1.8%	98.6 ± 0.4%

Roseobacter	71.5 ± 1.4%	90.5 ± 1.5%	22.7 ± 0.8%	46.8 ± 1.7%	69.3 ± 1.5%	93.6 ± 0.9%	68.0 ± 1.7%	93.4 ± 0.7%
Anaplasma	63.2 ± 1.3%	97.5 ± 0.6%	36.3 ± 1.5%	43.7 ± 1.5%	68.9 ± 1.1%	99.9 ± 0.1%	67.1 ± 1.4%	99.9 ± 0.1%
Chloroflexus	52.5 ± 1.5%	93.5 ± 0.7%	15.8 ± 1.1%	32.0 ± 1.2%	68.6 ± 2.0%	97.7 ± 0.7%	64.5 ± 1.4%	97.2 ± 0.8%
Candidatus	61.6 ± 1.2%	99.0 ± 0.5%	52.9 ± 0.3%	40.4 ± 0.7%	67.6 ± 1.8%	100.0 ± 0.0%	64.8 ± 2.0%	100.0 ± 0.1%
Bifidobacterium	60.4 ± 0.4%	93.2 ± 0.5%	26.6 ± 0.6%	14.9 ± 0.4%	66.6 ± 0.3%	97.7 ± 0.6%	64.4 ± 0.7%	97.3 ± 0.4%
Lactobacillus	57.7 ± 1.4%	63.0 ± 1.7%	27.5 ± 1.3%	6.7 ± 0.4%	64.3 ± 1.7%	72.6 ± 0.9%	62.4 ± 1.4%	71.0 ± 1.1%
Shewanella	65.4 ± 1.3%	69.6 ± 2.2%	43.1 ± 1.6%	13.3 ± 0.4%	63.8 ± 1.2%	83.2 ± 1.3%	62.9 ± 1.0%	81.5 ± 1.7%
Aeromonas	68.9 ± 0.6%	92.2 ± 1.0%	28.0 ± 0.7%	36.8 ± 0.7%	62.4 ± 0.8%	96.9 ± 0.8%	60.4 ± 1.1%	96.5 ± 0.3%
Wolbachia	51.0 ± 1.1%	97.7 ± 0.5%	12.2 ± 0.4%	21.8 ± 1.8%	62.3 ± 0.8%	99.6 ± 0.2%	60.6 ± 1.1%	99.6 ± 0.2%
Vulcanisaeta	49.2 ± 1.4%	97.5 ± 0.6%	53.9 ± 1.2%	71.2 ± 1.2%	62.3 ± 1.9%	98.2 ± 0.4%	60.4 ± 2.1%	98.1 ± 0.7%
Methylobacterium	65.6 ± 1.4%	75.8 ± 1.1%	41.8 ± 1.3%	26.0 ± 0.7%	62.0 ± 1.4%	86.9 ± 0.4%	59.0 ± 1.7%	86.2 ± 0.5%
Psychrobacter	64.6 ± 1.0%	95.4 ± 0.5%	29.7 ± 1.3%	45.5 ± 2.1%	61.7 ± 0.8%	98.0 ± 0.3%	60.6 ± 0.6%	97.8 ± 0.4%
Caulobacter	55.5 ± 0.4%	88.4 ± 1.8%	30.6 ± 0.6%	37.5 ± 1.1%	61.7 ± 1.2%	94.0 ± 0.9%	58.4 ± 1.4%	93.9 ± 0.8%
Acidovorax	65.6 ± 1.3%	90.0 ± 1.2%	29.1 ± 1.0%	28.7 ± 1.2%	61.0 ± 1.2%	95.2 ± 0.5%	58.0 ± 0.6%	95.6 ± 0.9%
Taylorella	53.8 ± 1.1%	98.5 ± 0.4%	8.9 ± 1.5%	32.7 ± 4.1%	59.2 ± 0.7%	99.7 ± 0.2%	58.0 ± 0.9%	99.6 ± 0.3%
Xanthomonas	61.9 ± 0.8%	71.2 ± 1.5%	33.7 ± 1.0%	24.8 ± 0.6%	58.7 ± 1.4%	76.2 ± 0.6%	55.5 ± 1.7%	75.8 ± 1.1%
Mesorhizobium	64.9 ± 1.0%	85.2 ± 0.9%	19.3 ± 1.9%	27.1 ± 2.4%	58.7 ± 0.4%	88.5 ± 1.4%	56.7 ± 0.6%	88.0 ± 0.9%
Pseudomonas	58.9 ± 1.4%	64.0 ± 1.7%	43.2 ± 2.4%	10.7 ± 0.5%	58.4 ± 1.8%	80.2 ± 1.2%	56.6 ± 1.8%	77.6 ± 0.6%
Streptomyces	50.6 ± 1.9%	53.6 ± 1.1%	45.0 ± 1.8%	22.1 ± 0.6%	57.1 ± 2.1%	77.2 ± 1.1%	54.5 ± 2.4%	77.5 ± 3.3%
Acidithiobacillus	50.6 ± 0.8%	91.7 ± 0.8%	8.3 ± 0.6%	16.5 ± 1.0%	57.1 ± 1.8%	96.6 ± 0.5%	55.1 ± 2.1%	96.2 ± 0.8%
Pyrobaculum	50.4 ± 1.5%	45.0 ± 1.1%	54.8 ± 0.8%	35.2 ± 0.6%	56.6 ± 0.6%	49.1 ± 0.7%	59.7 ± 1.7%	49.7 ± 1.3%
Staphylococcus	46.3 ± 1.3%	89.4 ± 0.9%	20.6 ± 1.9%	14.7 ± 1.1%	55.5 ± 2.0%	97.3 ± 0.7%	54.6 ± 2.1%	96.8 ± 0.7%
Thermus	48.7 ± 1.0%	91.2 ± 1.3%	73.5 ± 1.0%	82.2 ± 1.0%	55.2 ± 1.9%	93.5 ± 0.9%	50.2 ± 1.8%	92.7 ± 1.2%
Salinispora	50.8 ± 1.1%	92.7 ± 2.0%	14.6 ± 1.4%	46.6 ± 1.6%	54.8 ± 1.4%	98.0 ± 0.5%	50.0 ± 1.6%	97.6 ± 0.6%
Serratia	78.4 ± 2.1%	90.4 ± 1.3%	50.7 ± 1.7%	37.9 ± 0.8%	54.7 ± 1.5%	94.5 ± 1.0%	54.0 ± 1.9%	94.1 ± 0.5%
Thermoplasma	32.1 ± 1.9%	98.3 ± 0.2%	3.4 ± 0.8%	13.1 ± 2.7%	53.3 ± 1.0%	99.2 ± 0.4%	52.0 ± 1.1%	99.1 ± 0.4%
Sphingobium	51.7 ± 1.5%	89.7 ± 1.6%	33.5 ± 1.1%	36.2 ± 1.2%	52.9 ± 1.9%	94.7 ± 0.6%	51.2 ± 1.6%	94.6 ± 0.7%
Corynebacterium	39.7 ± 1.3%	72.1 ± 1.5%	28.7 ± 0.9%	10.4 ± 0.5%	52.8 ± 1.7%	88.3 ± 1.0%	50.7 ± 1.9%	87.1 ± 1.1%
Desulfurococcus	32.3 ± 1.4%	96.9 ± 1.0%	23.9 ± 1.5%	62.8 ± 2.2%	52.5 ± 1.7%	99.3 ± 0.5%	50.6 ± 1.0%	99.4 ± 0.5%
Methanosarcina	40.7 ± 0.6%	88.5 ± 1.4%	15.4 ± 0.5%	25.0 ± 1.0%	51.1 ± 1.2%	89.2 ± 1.4%	49.5 ± 1.2%	88.6 ± 0.7%
Prevotella	44.6 ± 2.0%	90.0 ± 1.7%	3.4 ± 0.5%	6.7 ± 0.9%	50.5 ± 2.8%	95.7 ± 1.2%	49.3 ± 3.1%	94.7 ± 1.0%
Yersinia	73.7 ± 2.0%	86.7 ± 0.8%	71.3 ± 1.7%	40.4 ± 1.0%	50.2 ± 1.3%	88.0 ± 1.8%	49.3 ± 0.9%	87.4 ± 1.0%
Metallosphaera	35.8 ± 1.4%	95.7 ± 1.0%	10.9 ± 0.6%	27.5 ± 0.6%	50.1 ± 1.1%	97.0 ± 1.0%	48.9 ± 1.3%	97.2 ± 1.3%
Chlamydophila	5.8 ± 0.5%	49.6 ± 4.9%	7.2 ± 0.7%	17.3 ± 1.1%	49.5 ± 1.5%	63.1 ± 2.1%	1.8 ± 0.4%	38.8 ± 7.8%
Nitrosomonas	44.2 ± 1.4%	87.0 ± 0.9%	7.1 ± 1.5%	11.0 ± 2.2%	49.5 ± 1.0%	92.7 ± 0.8%	48.2 ± 0.9%	92.2 ± 0.9%

Bartonella	47.3 ± 0.9%	94.2 ± 0.8%	13.3 ± 1.0%	13.9 ± 0.8%	49.4 ± 1.9%	99.1 ± 0.4%	48.4 ± 1.4%	99.1 ± 0.3%
Candidatus	33.8 ± 1.0%	98.6 ± 0.7%	4.4 ± 0.8%	20.3 ± 3.1%	49.4 ± 2.6%	99.9 ± 0.2%	47.3 ± 2.7%	99.8 ± 0.3%
Deinococcus	34.0 ± 1.6%	81.8 ± 1.8%	15.1 ± 1.0%	15.1 ± 1.1%	48.5 ± 1.2%	93.8 ± 1.3%	46.0 ± 1.3%	92.6 ± 1.1%
Bacteroides	38.6 ± 1.3%	68.9 ± 1.5%	14.4 ± 0.8%	14.7 ± 0.8%	48.1 ± 1.3%	74.6 ± 2.0%	47.3 ± 1.3%	73.1 ± 2.0%
Methanococcus	38.8 ± 1.7%	93.1 ± 1.0%	18.8 ± 0.6%	21.3 ± 1.0%	47.5 ± 2.5%	97.2 ± 1.0%	45.7 ± 2.0%	97.6 ± 0.7%
Fervidobacterium	34.6 ± 1.3%	84.8 ± 1.4%	5.4 ± 0.4%	19.2 ± 1.9%	46.9 ± 1.3%	91.2 ± 0.6%	46.1 ± 1.3%	90.6 ± 1.1%
Arcobacter	48.4 ± 1.3%	89.7 ± 2.0%	34.1 ± 1.0%	44.2 ± 1.7%	46.2 ± 2.3%	94.6 ± 1.3%	60.8 ± 1.9%	94.1 ± 0.9%
Leuconostoc	40.7 ± 0.9%	94.5 ± 0.6%	14.0 ± 0.9%	15.7 ± 0.9%	45.6 ± 1.9%	97.6 ± 0.9%	48.1 ± 1.4%	97.1 ± 0.4%
Campylobacter	29.7 ± 2.0%	85.7 ± 2.1%	30.6 ± 0.6%	29.0 ± 0.7%	45.1 ± 1.0%	95.8 ± 1.2%	43.3 ± 0.8%	95.5 ± 0.8%
Thermosipho	31.5 ± 0.7%	89.1 ± 1.2%	6.6 ± 0.2%	23.6 ± 1.1%	43.7 ± 0.8%	95.7 ± 1.0%	42.0 ± 1.1%	95.5 ± 1.2%
Cupriavidus	49.3 ± 1.3%	78.6 ± 1.1%	16.3 ± 0.8%	21.0 ± 0.4%	42.8 ± 1.8%	89.6 ± 0.8%	40.1 ± 1.6%	89.0 ± 1.0%
Edwardsiella	59.1 ± 0.9%	93.7 ± 0.9%	24.0 ± 1.3%	49.3 ± 1.9%	42.3 ± 0.7%	97.7 ± 0.8%	41.4 ± 0.8%	97.1 ± 0.7%
Pyrococcus	52.9 ± 0.4%	78.4 ± 1.3%	37.5 ± 1.0%	26.7 ± 0.8%	42.1 ± 1.6%	87.2 ± 1.2%	41.9 ± 1.1%	85.2 ± 1.2%
Thermoanaerobacterium	46.0 ± 1.5%	97.0 ± 0.6%	7.0 ± 0.3%	27.4 ± 1.3%	41.5 ± 1.6%	98.5 ± 0.6%	40.9 ± 1.6%	98.1 ± 0.4%
Vibrio	44.3 ± 1.0%	66.5 ± 1.0%	21.7 ± 1.3%	11.4 ± 0.6%	41.0 ± 0.7%	78.2 ± 0.5%	41.0 ± 0.6%	76.4 ± 1.5%
Erwinia	57.8 ± 0.7%	82.7 ± 1.2%	20.4 ± 0.4%	23.3 ± 0.9%	40.1 ± 1.7%	90.4 ± 2.2%	38.5 ± 1.3%	89.9 ± 1.7%
Pectobacterium	60.1 ± 1.6%	87.5 ± 1.4%	18.6 ± 1.1%	29.7 ± 1.8%	39.7 ± 1.3%	92.2 ± 1.6%	39.1 ± 1.8%	91.3 ± 1.3%
Legionella	23.0 ± 1.2%	90.0 ± 0.8%	1.9 ± 0.3%	11.5 ± 2.0%	38.9 ± 1.9%	97.4 ± 1.0%	37.0 ± 2.3%	96.9 ± 1.1%
Rhizobium	49.0 ± 1.6%	66.7 ± 1.8%	17.1 ± 1.9%	23.6 ± 2.6%	38.5 ± 1.9%	70.4 ± 1.5%	37.5 ± 1.7%	70.4 ± 1.5%
Desulfovibrio	21.5 ± 0.4%	56.7 ± 1.4%	8.5 ± 0.5%	4.8 ± 0.4%	36.9 ± 1.8%	86.2 ± 0.9%	35.4 ± 2.0%	84.6 ± 1.5%
Rothia	29.0 ± 2.7%	91.9 ± 1.4%	7.6 ± 1.2%	24.4 ± 3.1%	36.6 ± 3.4%	98.0 ± 0.4%	35.1 ± 3.3%	97.7 ± 0.6%
Ehrlichia	25.7 ± 0.8%	96.5 ± 0.4%	17.2 ± 0.6%	32.8 ± 0.5%	36.0 ± 1.1%	99.7 ± 0.2%	34.6 ± 0.6%	99.7 ± 0.3%
Chlorobaculum	37.8 ± 0.7%	85.2 ± 2.4%	12.1 ± 1.0%	34.3 ± 2.9%	35.7 ± 0.8%	93.1 ± 0.8%	34.6 ± 1.0%	91.9 ± 1.6%
Acidaminococcus	23.6 ± 1.3%	91.0 ± 1.3%	2.7 ± 0.4%	10.2 ± 1.8%	35.3 ± 2.0%	97.0 ± 0.3%	33.9 ± 1.8%	96.8 ± 1.0%
Thermotoga	19.7 ± 2.1%	74.4 ± 2.2%	9.0 ± 0.7%	8.0 ± 0.7%	34.4 ± 1.8%	88.0 ± 1.0%	33.5 ± 1.4%	86.9 ± 1.9%
Nitrobacter	33.0 ± 1.8%	92.9 ± 2.1%	5.5 ± 0.9%	21.4 ± 3.0%	34.4 ± 1.6%	97.2 ± 0.9%	33.5 ± 1.7%	97.1 ± 0.9%
Methanocella	33.1 ± 1.4%	88.9 ± 1.4%	13.9 ± 0.9%	25.1 ± 1.1%	34.3 ± 0.6%	94.3 ± 1.2%	49.3 ± 1.9%	95.2 ± 1.2%
Mycoplasma	13.6 ± 0.6%	71.1 ± 1.8%	33.1 ± 1.6%	6.8 ± 0.4%	33.8 ± 1.7%	96.9 ± 1.0%	32.2 ± 1.8%	96.2 ± 0.7%
Sulfurimonas	20.9 ± 1.1%	91.2 ± 2.3%	4.3 ± 1.2%	18.9 ± 5.5%	32.9 ± 1.1%	97.5 ± 0.6%	31.8 ± 1.0%	97.5 ± 0.3%
Dickeya	49.9 ± 1.4%	92.8 ± 1.6%	10.1 ± 1.2%	28.4 ± 3.4%	32.2 ± 0.7%	95.4 ± 0.8%	31.0 ± 0.4%	95.0 ± 1.1%
Methylotenera	28.3 ± 0.9%	93.8 ± 0.7%	6.2 ± 0.9%	23.7 ± 3.1%	32.1 ± 0.8%	98.4 ± 0.5%	31.2 ± 1.3%	97.7 ± 0.6%
Cronobacter	54.8 ± 1.4%	89.6 ± 1.2%	30.2 ± 2.2%	47.6 ± 2.8%	32.0 ± 1.6%	91.4 ± 2.5%	31.4 ± 1.3%	90.7 ± 1.5%
Clostridium	20.4 ± 1.3%	30.2 ± 1.6%	34.6 ± 1.3%	9.2 ± 0.3%	31.3 ± 1.2%	49.6 ± 2.0%	31.0 ± 1.0%	46.6 ± 1.7%
Flavobacterium	23.8 ± 0.7%	58.0 ± 1.6%	11.0 ± 0.7%	10.9 ± 0.7%	31.1 ± 1.5%	72.3 ± 2.3%	30.3 ± 1.0%	70.2 ± 1.5%
Aliivibrio	36.1 ± 1.4%	93.2 ± 1.2%	4.3 ± 0.7%	18.6 ± 2.2%	29.9 ± 2.0%	98.5 ± 0.7%	29.2 ± 1.8%	97.9 ± 1.4%

Xenorhabdus	36.1 ± 1.2%	91.1 ± 1.2%	7.2 ± 1.0%	26.4 ± 4.1%	29.9 ± 1.1%	95.5 ± 1.1%	29.2 ± 1.2%	94.5 ± 0.6%
Sinorhizobium	35.2 ± 1.8%	74.5 ± 1.2%	11.8 ± 0.9%	13.6 ± 0.7%	29.9 ± 1.1%	81.4 ± 1.3%	28.2 ± 1.4%	80.1 ± 2.1%
Meiothermus	20.4 ± 2.3%	82.8 ± 2.9%	20.8 ± 1.0%	52.8 ± 1.0%	29.2 ± 0.6%	91.2 ± 1.7%	27.4 ± 1.1%	90.1 ± 1.3%
Archaeoglobus	15.9 ± 0.6%	86.8 ± 2.0%	17.1 ± 1.8%	23.8 ± 2.5%	29.2 ± 1.2%	93.9 ± 1.8%	27.6 ± 1.0%	93.8 ± 1.3%
Ralstonia	32.9 ± 0.7%	91.2 ± 1.4%	4.3 ± 1.0%	11.4 ± 2.2%	29.1 ± 1.1%	96.1 ± 1.6%	27.5 ± 0.7%	95.7 ± 1.0%
Acinetobacter	71.8 ± 0.4%	94.6 ± 1.0%	21.3 ± 1.2%	27.8 ± 0.9%	29.0 ± 1.1%	95.6 ± 1.9%	67.9 ± 0.9%	97.4 ± 0.8%
Aggregatibacter	40.4 ± 1.8%	81.8 ± 1.0%	14.1 ± 1.1%	36.8 ± 2.7%	28.6 ± 2.0%	89.1 ± 2.0%	27.8 ± 2.1%	89.6 ± 1.4%
Sphaerochaeta	17.9 ± 1.3%	87.6 ± 2.2%	7.5 ± 0.1%	21.4 ± 0.9%	27.9 ± 1.4%	96.1 ± 1.2%	26.9 ± 0.9%	95.8 ± 0.9%
Gordonia	24.0 ± 1.6%	85.3 ± 2.0%	7.0 ± 1.5%	29.0 ± 5.7%	27.4 ± 0.8%	97.0 ± 0.7%	25.5 ± 1.3%	96.2 ± 0.9%
Rhodococcus	25.4 ± 1.4%	56.8 ± 2.0%	13.2 ± 1.6%	18.7 ± 2.1%	25.9 ± 0.7%	77.0 ± 0.8%	23.8 ± 1.0%	74.6 ± 1.2%
Klebsiella	50.6 ± 1.3%	81.2 ± 1.4%	43.0 ± 1.9%	41.8 ± 1.2%	24.9 ± 1.3%	85.1 ± 2.2%	24.3 ± 1.3%	84.2 ± 1.3%
Nocardia	24.7 ± 1.8%	78.5 ± 3.4%	8.1 ± 1.1%	28.2 ± 2.8%	24.6 ± 2.2%	92.4 ± 1.6%	22.4 ± 1.4%	90.7 ± 1.3%
Cellulophaga	18.2 ± 2.0%	78.8 ± 3.8%	5.8 ± 0.8%	21.2 ± 1.4%	24.1 ± 1.8%	87.8 ± 1.9%	23.0 ± 1.8%	87.1 ± 1.2%
Gluconacetobacter	18.1 ± 1.3%	87.4 ± 2.8%	2.2 ± 0.3%	7.0 ± 0.9%	23.9 ± 1.5%	96.0 ± 0.6%	22.7 ± 1.5%	96.0 ± 0.8%
Rhodospirillum	14.8 ± 1.1%	68.7 ± 4.6%	7.8 ± 0.8%	10.6 ± 1.1%	22.1 ± 1.1%	90.9 ± 0.4%	20.2 ± 1.5%	88.6 ± 1.6%
Selenomonas	13.0 ± 1.4%	81.1 ± 3.1%	0.6 ± 0.3%	2.1 ± 1.0%	21.3 ± 1.8%	92.9 ± 1.1%	20.0 ± 1.8%	92.3 ± 1.5%
Lactococcus	18.1 ± 1.3%	93.4 ± 3.6%	1.8 ± 0.4%	11.2 ± 2.1%	21.3 ± 1.3%	96.8 ± 0.6%	20.0 ± 1.8%	96.7 ± 0.5%
Thermococcus	27.6 ± 1.0%	75.5 ± 2.5%	8.2 ± 0.6%	5.8 ± 0.3%	21.2 ± 1.1%	80.6 ± 2.1%	23.3 ± 1.1%	79.7 ± 1.4%
Shigella	41.0 ± 2.0%	52.2 ± 1.1%	54.4 ± 2.5%	26.1 ± 1.4%	21.1 ± 0.7%	55.5 ± 3.0%	21.6 ± 0.8%	55.9 ± 3.0%
Leptospira	8.2 ± 0.9%	81.0 ± 4.5%	11.5 ± 1.1%	16.6 ± 1.7%	20.4 ± 0.7%	93.5 ± 1.2%	19.7 ± 0.7%	93.6 ± 2.2%
Candidatus	19.4 ± 1.1%	96.3 ± 1.9%	13.4 ± 0.5%	22.1 ± 1.2%	19.6 ± 1.0%	100.0 ± 0.0%	19.4 ± 0.7%	99.9 ± 0.2%
Pedobacter	9.9 ± 0.5%	71.5 ± 1.7%	2.0 ± 0.3%	10.8 ± 1.1%	19.2 ± 1.1%	91.7 ± 1.4%	18.1 ± 1.0%	89.4 ± 0.9%
Haemophilus	30.3 ± 0.9%	40.9 ± 1.2%	21.4 ± 1.8%	21.6 ± 1.9%	19.2 ± 0.9%	35.8 ± 0.8%	19.0 ± 0.8%	36.0 ± 0.4%
Methanosaeta	8.9 ± 1.0%	86.6 ± 2.7%	5.3 ± 0.9%	16.8 ± 2.5%	18.4 ± 1.4%	97.1 ± 1.5%	17.2 ± 1.4%	96.4 ± 1.3%
Pantoea	32.5 ± 0.6%	81.1 ± 1.2%	8.2 ± 0.4%	12.4 ± 0.9%	18.1 ± 1.5%	85.9 ± 2.6%	17.8 ± 1.4%	84.4 ± 3.1%
Desulfotomaculum	10.2 ± 0.5%	54.5 ± 2.4%	7.9 ± 1.1%	8.4 ± 1.2%	16.5 ± 0.7%	81.1 ± 3.3%	15.7 ± 0.8%	79.6 ± 2.2%
Capnocytophaga	12.2 ± 0.9%	82.3 ± 3.2%	4.5 ± 0.8%	17.9 ± 2.7%	16.0 ± 1.0%	91.0 ± 1.8%	14.8 ± 1.0%	90.4 ± 2.3%
Pediococcus	15.3 ± 1.0%	93.1 ± 0.6%	1.4 ± 0.2%	10.2 ± 2.4%	15.4 ± 1.1%	97.7 ± 1.2%	15.3 ± 1.1%	97.4 ± 1.0%
Paenibacillus	12.0 ± 1.0%	54.6 ± 2.8%	5.7 ± 0.3%	6.1 ± 0.3%	15.1 ± 1.1%	80.2 ± 1.6%	16.6 ± 1.3%	77.5 ± 2.1%
Enterobacter	33.7 ± 1.0%	65.4 ± 1.4%	12.0 ± 0.9%	12.8 ± 1.0%	14.9 ± 1.1%	70.2 ± 1.5%	14.5 ± 1.0%	69.0 ± 1.7%
Methanobrevibacter	8.7 ± 1.3%	86.6 ± 2.5%	3.1 ± 0.5%	11.9 ± 1.3%	14.7 ± 0.3%	97.4 ± 0.6%	14.4 ± 0.4%	96.8 ± 1.9%
Enterococcus	13.1 ± 1.3%	86.3 ± 3.0%	2.0 ± 0.6%	9.6 ± 2.6%	14.5 ± 1.3%	95.3 ± 1.9%	13.9 ± 1.3%	94.3 ± 1.2%
Planctomyces	5.4 ± 0.7%	60.8 ± 5.5%	5.8 ± 0.6%	17.5 ± 1.6%	14.2 ± 0.3%	95.2 ± 1.1%	12.7 ± 0.6%	94.2 ± 1.2%
Salmonella	30.0 ± 1.3%	77.2 ± 1.4%	11.9 ± 0.6%	33.7 ± 3.0%	13.0 ± 1.1%	87.5 ± 3.4%	13.0 ± 1.0%	87.9 ± 3.1%
Arthrobacter	10.3 ± 0.6%	51.3 ± 3.3%	4.6 ± 0.6%	5.1 ± 0.5%	13.0 ± 0.4%	76.7 ± 2.4%	12.7 ± 0.5%	74.8 ± 3.0%

Marinomonas	25.5 ± 1.6%	82.3 ± 1.5%	4.2 ± 0.7%	11.8 ± 1.4%	11.4 ± 0.8%	88.4 ± 3.5%	32.1 ± 0.9%	93.4 ± 1.4%
Rhodobacter	10.1 ± 0.7%	78.6 ± 2.2%	7.7 ± 1.2%	20.1 ± 2.5%	10.2 ± 1.4%	90.4 ± 3.3%	9.6 ± 0.8%	88.7 ± 3.5%
Alkaliphilus	6.1 ± 0.7%	70.8 ± 2.7%	2.3 ± 0.4%	11.5 ± 1.9%	9.6 ± 0.4%	86.6 ± 3.2%	9.4 ± 0.7%	85.5 ± 2.5%
Propionibacterium	6.4 ± 0.9%	78.9 ± 5.6%	0.9 ± 0.3%	4.3 ± 1.4%	8.8 ± 1.2%	95.0 ± 1.6%	8.4 ± 1.3%	93.5 ± 2.7%
Cellulomonas	30.7 ± 1.1%	85.1 ± 1.7%	62.6 ± 1.3%	62.9 ± 2.0%	8.6 ± 0.8%	89.0 ± 2.9%	32.6 ± 1.3%	94.1 ± 1.4%
Blattabacterium	57.8 ± 1.2%	98.3 ± 0.5%	40.5 ± 1.2%	39.5 ± 0.4%	8.5 ± 0.7%	99.8 ± 0.5%	71.6 ± 1.5%	99.8 ± 0.1%
Nitrosococcus	35.5 ± 1.5%	90.5 ± 1.5%	8.7 ± 1.0%	21.2 ± 1.4%	7.7 ± 0.9%	86.8 ± 1.5%	41.5 ± 0.7%	95.2 ± 0.6%
Chlorobium	10.4 ± 1.3%	21.8 ± 2.5%	9.3 ± 0.8%	9.6 ± 0.8%	7.4 ± 0.9%	22.9 ± 1.9%	7.9 ± 1.4%	24.2 ± 3.2%
Eubacterium	6.4 ± 1.2%	57.3 ± 3.2%	6.9 ± 1.5%	13.2 ± 2.6%	6.5 ± 0.6%	74.5 ± 2.9%	6.4 ± 0.4%	71.9 ± 5.4%
Chlamydia	14.2 ± 1.3%	20.7 ± 1.3%	22.1 ± 3.3%	12.5 ± 1.5%	6.2 ± 0.9%	19.5 ± 2.4%	8.6 ± 0.7%	25.7 ± 1.2%
Porphyromonas	4.7 ± 0.9%	77.6 ± 6.8%	2.3 ± 0.5%	11.8 ± 2.4%	5.9 ± 1.1%	92.5 ± 3.8%	5.6 ± 1.1%	89.7 ± 3.5%
Agrobacterium	7.9 ± 0.5%	42.0 ± 1.9%	4.8 ± 0.5%	8.8 ± 0.8%	4.4 ± 0.6%	42.8 ± 1.3%	4.4 ± 0.6%	41.7 ± 0.7%
Thermoproteus	4.2 ± 0.2%	52.4 ± 4.1%	19.4 ± 0.6%	41.8 ± 1.5%	3.6 ± 0.4%	74.5 ± 6.5%	3.5 ± 0.4%	67.0 ± 9.0%
Actinobacillus	4.6 ± 0.7%	40.6 ± 4.3%	13.8 ± 1.3%	32.1 ± 2.0%	3.5 ± 0.8%	59.5 ± 7.7%	3.3 ± 0.7%	57.9 ± 6.9%
Spirochaeta	2.8 ± 0.7%	21.4 ± 6.2%	2.8 ± 0.3%	2.5 ± 0.3%	2.9 ± 0.6%	37.0 ± 4.2%	2.9 ± 0.6%	36.3 ± 4.7%
Pseudoxanthomonas	5.5 ± 0.7%	71.9 ± 4.0%	10.9 ± 0.9%	25.6 ± 1.3%	2.9 ± 0.7%	77.1 ± 3.8%	3.0 ± 0.5%	79.7 ± 5.2%
Pseudoalteromonas	3.2 ± 0.6%	49.7 ± 6.0%	3.9 ± 0.3%	8.7 ± 0.5%	2.8 ± 0.6%	67.0 ± 7.3%	2.8 ± 0.6%	65.9 ± 7.3%
Pelodictyon	3.5 ± 0.4%	36.4 ± 5.2%	1.2 ± 0.2%	4.2 ± 0.5%	2.3 ± 0.5%	33.2 ± 4.0%	2.5 ± 0.4%	33.8 ± 3.3%
Escherichia	4.1 ± 0.9%	37.8 ± 4.1%	8.2 ± 0.9%	22.0 ± 2.3%	2.0 ± 0.4%	39.7 ± 5.7%	2.0 ± 0.4%	38.4 ± 5.4%
Pelobacter	1.3 ± 0.3%	31.6 ± 4.5%	0.9 ± 0.2%	3.7 ± 0.9%	1.7 ± 0.2%	60.6 ± 4.5%	1.4 ± 0.4%	55.5 ± 5.3%
Citrobacter	6.5 ± 0.5%	45.3 ± 4.3%	5.2 ± 0.7%	14.6 ± 1.7%	1.3 ± 0.3%	24.9 ± 5.9%	1.3 ± 0.2%	24.0 ± 3.4%

Supplementary Table S2. Species-mask comparison of MetaCV, Phymm, BlastX and RAPSearch2 on the classification of simulated reads with length ranging from 100 bp to 1 kb from 154 genera. Evaluations are performed on levels of genus, family order class and phylum.

	MetaCV		Phymm		BlastX+Megan		RAPSearch2+Megan	
Genus	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
100 bp	41.5 + 0.1%	80.0 + 0.2%	24.3 + 0.1%	26.4 + 0.1%	41.4 + 0.1%	87.1 ± 0.2%	41.7 ± 0.1%	86.5 ± 0.1%
200 bp	51.9 + 0.0%	82.0 + 0.1%	37.3 + 0.1%	40.5 + 0.1%	50.8 + 0.1%	85.4 ± 0.1%	51.5 ± 0.1%	85.4 ± 0.1%
400 bp	60.4 + 0.1%	82.3 + 0.1%	49.5 + 0.1%	54.2 + 0.1%	56.8 + 0.1%	84.6 ± 0.2%	57.9 ± 0.2%	84.4 ± 0.2%
600 bp	63.6 + 0.1%	84.2 + 0.1%	55.9 + 0.1%	61.0 + 0.2%	59.2 + 0.2%	84.8 ± 0.2%	60.5 ± 0.2%	84.2 ± 0.1%
800 bp	66.0 + 0.1%	83.9 + 0.2%	59.8 + 0.1%	64.9 + 0.1%	60.4 + 0.1%	84.2 ± 1.4%	61.8 ± 0.1%	84.2 ± 0.2%
1000 bp	67.1 + 0.1%	85.4 + 0.1%	62.4 + 0.1%	67.3 + 0.1%	61.1 + 0.1%	85.1 ± 0.1%	62.7 ± 0.1%	84.3 ± 0.2%
Family								
100 bp	45.6 + 0.2%	85.4 + 0.2%	27.0 + 0.1%	27.8 + 0.1%	50.3 ± 0.1%	93.0 ± 0.2%	49.8 ± 0.1%	92.5 ± 0.1%
200 bp	56.3 + 0.1%	87.3 + 0.1%	40.4 + 0.1%	42.6 + 0.1%	61.2 ± 0.1%	91.0 ± 0.1%	61.1 ± 0.1%	91.2 ± 0.1%
400 bp	65.2 + 0.1%	87.6 + 0.2%	53.2 + 0.1%	56.6 + 0.1%	68.2 ± 0.1%	90.3 ± 0.2%	68.4 ± 0.1%	90.3 ± 0.2%
600 bp	68.6 + 0.1%	89.7 + 0.1%	59.9 + 0.2%	63.5 + 0.3%	70.7 ± 0.1%	90.5 ± 0.1%	71.3 ± 0.1%	90.0 ± 0.1%
800 bp	71.3 + 0.1%	89.4 + 0.1%	64.0 + 0.1%	67.5 + 0.1%	71.3 ± 1.6%	90.0 ± 1.1%	72.7 ± 0.1%	90.0 ± 0.1%
1000 bp	72.5 + 0.1%	91.0 + 0.1%	66.6 + 0.2%	70.0 + 0.2%	73.1 ± 0.1%	90.8 ± 0.1%	73.9 ± 0.1%	90.1 ± 0.1%
Order								
100 bp	46.7 + 0.2%	89.8 + 0.1%	28.5 + 0.1%	32.8 + 0.1%	54.7 ± 0.1%	95.9 ± 0.1%	53.6 ± 0.2%	95.7 ± 0.1%
200 bp	58.4 + 0.1%	91.3 + 0.2%	42.7 + 0.1%	49.1 + 0.1%	66.6 ± 0.1%	94.3 ± 0.1%	65.6 ± 0.1%	94.4 ± 0.1%
400 bp	68.2 + 0.2%	91.4 + 0.1%	56.6 + 0.1%	64.0 + 0.2%	73.9 ± 0.1%	93.6 ± 0.1%	73.4 ± 0.1%	93.7 ± 0.0%
600 bp	71.8 + 0.1%	93.1 + 0.1%	63.8 + 0.1%	71.0 + 0.3%	76.5 ± 0.1%	93.6 ± 0.1%	76.3 ± 0.1%	93.5 ± 0.1%
800 bp	74.9 + 0.1%	92.8 + 0.1%	68.3 + 0.2%	75.0 + 0.1%	77.0 ± 2.1%	93.3 ± 0.6%	77.9 ± 0.0%	93.4 ± 0.1%
1000 bp	76.1 + 0.1%	94.2 + 0.1%	71.2 + 0.2%	77.5 + 0.2%	79.0 ± 0.1%	93.9 ± 0.0%	79.1 ± 0.2%	93.5 ± 0.1%
Class								
100 bp	43.8 + 0.1%	91.2 + 0.2%	30.2 + 0.1%	34.9 + 0.2%	55.0 ± 0.1%	96.9 ± 0.1%	54.3 ± 0.1%	96.7 ± 0.1%
200 bp	56.0 + 0.1%	92.7 + 0.2%	43.2 + 0.1%	51.9 + 0.1%	67.8 ± 0.1%	95.5 ± 0.1%	66.8 ± 0.1%	95.6 ± 0.0%
400 bp	66.4 + 0.2%	93.0 + 0.1%	56.0 + 0.1%	67.2 + 0.2%	75.1 ± 0.1%	94.8 ± 0.1%	74.6 ± 0.2%	95.1 ± 0.1%
600 bp	70.2 + 0.1%	94.6 + 0.2%	62.9 + 0.2%	74.3 + 0.5%	77.7 ± 0.2%	94.8 ± 0.1%	77.7 ± 0.1%	94.8 ± 0.1%
800 bp	73.7 + 0.2%	94.5 + 0.1%	67.3 + 0.2%	78.3 + 0.2%	78.3 ± 2.1%	94.9 ± 0.2%	79.4 ± 0.2%	94.8 ± 0.1%
1000 bp	75.0 + 0.1%	95.7 + 0.1%	70.0 + 0.2%	80.7 + 0.2%	80.3 ± 0.1%	95.1 ± 0.0%	80.7 ± 0.1%	94.9 ± 0.1%

Phylum								
100 bp	45.9 + 0.2%	92.4 + 0.3%	35.2 + 0.2%	39.5 + 0.2%	59.0 ± 0.1%	98.1 ± 0.1%	58.0 ± 0.1%	97.9 ± 0.1%
200 bp	58.7 + 0.1%	93.9 + 0.4%	48.4 + 0.1%	57.6 + 0.3%	72.2 ± 0.2%	97.0 ± 0.2%	70.9 ± 0.2%	97.2 ± 0.1%
400 bp	69.2 + 0.2%	94.5 + 0.1%	61.2 + 0.1%	73.5 + 0.2%	79.6 ± 0.1%	96.4 ± 0.1%	78.7 ± 0.2%	96.8 ± 0.1%
600 bp	73.1 + 0.1%	96.1 + 0.2%	68.5 + 0.4%	80.7 + 0.4%	82.4 ± 0.2%	96.5 ± 0.1%	82.0 ± 0.2%	96.6 ± 0.1%
800 bp	76.5 + 0.1%	96.0 + 0.1%	72.9 + 0.2%	84.5 + 0.1%	82.6 ± 2.6%	96.7 ± 0.1%	83.6 ± 0.2%	96.7 ± 0.0%
1000 bp	77.5 + 0.1%	97.2 + 0.2%	75.8 + 0.2%	87.0 + 0.2%	84.7 ± 0.2%	96.8 ± 0.1%	84.8 ± 0.2%	96.8 ± 0.1%

Supplementary Table S3. Performance comparison of MetaCV, Phymm, BlastX and RAPSearch2 on the classification of "randomized metagenomic" data sets. Phymm and BlastX were applied to 100 bp simulated reads while MetaCV was also applied to all reads of various lengths. Evaluations are performed on levels of genus, family, order, class and phylum.

	MetaCV		Phymm		BlastX+Megan		RAPSearch2+Megan	
Genus	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
100 bp	58.4 ± 0.0%	89.3 ± 0.1%	28.3 ± 0.1%	46.4 ± 0.1%	61.0 ± 0.4%	93.7 ± 0.1%	60.4 ± 0.0%	93.3 ± 0.1%
200 bp	61.5 ± 0.1%	88.3 ± 0.1%	_	_	_	_	_	_
400 bp	60.8 ± 0.1%	85.2 ± 0.0%	_	_	_	_	_	_
600 bp	60.3 ± 0.0%	83.9 ± 0.1%	_	_	_	_	_	_
800 bp	57.5 ± 0.1%	81.1 ± 0.1%	_	_	_	_	_	_
1000 bp	57.5 ± 0.0%	81.0 ± 0.1%	_	_	_	_	_	_
Family								
100 bp	52.4 ± 0.1%	88.6 ± 0.1%	32.6 ± 0.0%	40.5 ± 0.1%	55.0 ± 0.3%	94.7 ± 0.0%	54.4 ± 0.0%	94.2 ± 0.1%
200 bp	59.5 ± 0.1%	88.4 ± 0.0%	_	_	_	_	_	_
400 bp	64.5 ± 0.1%	85.1 ± 0.1%	_	_	_	_	_	_
600 bp	66.2 ± 0.0%	83.7 ± 0.0%	_	_	_	_	_	_
800 bp	66.5 ± 0.1%	80.3 ± 0.1%	_	_	_	_	_	_
1000 bp	67.1 ± 0.1%	80.2 ± 0.1%	_	_	_	_	_	_
Order								
100 bp	52.9 ± 0.1%	89.2 ± 0.1%	39.0 ± 0.1%	43.1 ± 0.0%	56.5 ± 0.4%	95.7 ± 0.0%	55.9 ± 0.1%	95.3 ± 0.0%
200 bp	61.1 ± 0.1%	89.4 ± 0.1%	_	_	_	_	_	_
400 bp	67.9 ± 0.1%	86.3 ± 0.1%	_	_	_	_	_	_
600 bp	70.3 ± 0.1%	85.0 ± 0.0%	_	_	_	_	_	_
800 bp	71.7 ± 0.1%	81.3 ± 0.1%	_	_	_	_	_	_
1000 bp	72.4 ± 0.0%	81.2 ± 0.1%	_	_	_	_	_	_
Class								
100 bp	53.1 ± 0.1%	90.4 ± 0.1%	44.6 ± 0.1%	48.1 ± 0.0%	59.0 ± 0.4%	96.5 ± 0.0%	58.5 ± 0.1%	96.1 ± 0.0%
200 bp	61.7 ± 0.1%	90.9 ± 0.0%	_	_	_	_	_	_
400 bp	69.3 ± 0.1%	88.2 ± 0.1%	_	_	_	_	_	_
600 bp	71.9 ± 0.1%	86.9 ± 0.0%	_	_	_	_	_	_
800 bp	73.8 ± 0.1%	83.4 ± 0.0%	_	_	_	_	_	_
1000 bp	74.5 ± 0.0%	83.3 ± 0.1%	_	_	_	_	_	_

Phylum								
100 bp	55.2 ± 0.1%	92.6 ± 0.0%	54.9 ± 0.0%	56.5 ± 0.0%	62.1 ± 0.4%	97.5 ± 0.0%	61.6 ± 0.1%	97.3 ± 0.0%
200 bp	64.4 ± 0.1%	93.2 ± 0.0%	_	_	_	_	_	_
400 bp	73.2 ± 0.1%	90.9 ± 0.1%	_	_	_	_	_	_
600 bp	76.4 ± 0.0%	89.7 ± 0.0%	_	_	_	_	_	_
800 bp	79.2 ± 0.1%	86.4 ± 0.0%	_	_	_	_	_	_
1000 bp	80.0 ± 0.0%	86.2 ± 0.0%	_	_	_	_	_	_

Supplementary Table S4. A list of 44 most abundant genera by MetaCV on 109 cohorts of human gut samples. Cohorts which have over 400 matched reads per million raw sequences are counted. "Meta-HIT" has 13 genera with at least one whole-genome-sequenced (WGS) species summarized by Qin et al., all of them are detected by MetaCV (marked with "Y"). In addition, another 14 genera were further proved as gut-associated bacteria by "HMP" (Human Microbiome Project of NIH), and were marked with "Y".

Genus	Ave. read num per million	Num of cohorts	Meta-HIT	HMP
Bacteroides	128,113	109	Y	Y
Eubacterium	36,196	109	Y	Y
Roseburia	17,091	109	Y	Y
Prevotella	16,382	109	Y	Y
Clostridium	16,315	109	Y	Y
Parabacteroides	13,490	109	Y	Y
Odoribacter	5,862	109	_	Y
Oscillibacter	4,922	109	_	_
Bifidobacterium	4,195	107	Y	Y
Escherichia	3,521	45	Y	Y
Butyrivibrio	3,073	109	Y	Y
Ruminococcus	2,868	109	Y	Y
Streptococcus	2,774	109	Y	Y
Akkermansia	2,550	49	Y	_
Acidaminococcus	2,379	108	_	Y
Tannerella	2,012	109	_	Y
Lactobacillus	2,011	108	_	Y
Treponema	1,672	109	_	_
Ethanoligenens	1,290	101	_	_
Porphyromonas	1,281	106	_	Y
Selenomonas	1,211	108	_	_
Bacillus	1,129	108	_	Y
Eggerthella	1,005	99	_	Y
Cellulosilyticum	951	99	_	_
Paludibacter	870	99	_	_
Veillonella	781	38	_	Y
Paenibacillus	695	95	_	Y
Megasphaera	665	51	_	_
Desulfovibrio	660	67	_	Y
Methanobrevibacter	627	22	_	Y
Enterococcus	623	76	Y	Y
Desulfotomaculum	613	82	_	_
Desulfitobacterium	564	69	_	Y
Pedobacter	556	66	_	_
Burkholderia	506	38	_	_
Alkaliphilus	493	58	_	_
Flavobacterium	475	39	_	_
Spirochaeta	473	48	_	_
Haemophilus	464	15	_	_
Slackia	452	38	_	Y
Acetobacterium	448	40	_	_
Fibrobacter	434	27	_	_
Sphaerochaeta	432	22	_	_
Mycoplasma	428	13	_	Y