nature microbiology

SUPPLEMENTARY INFORMATION

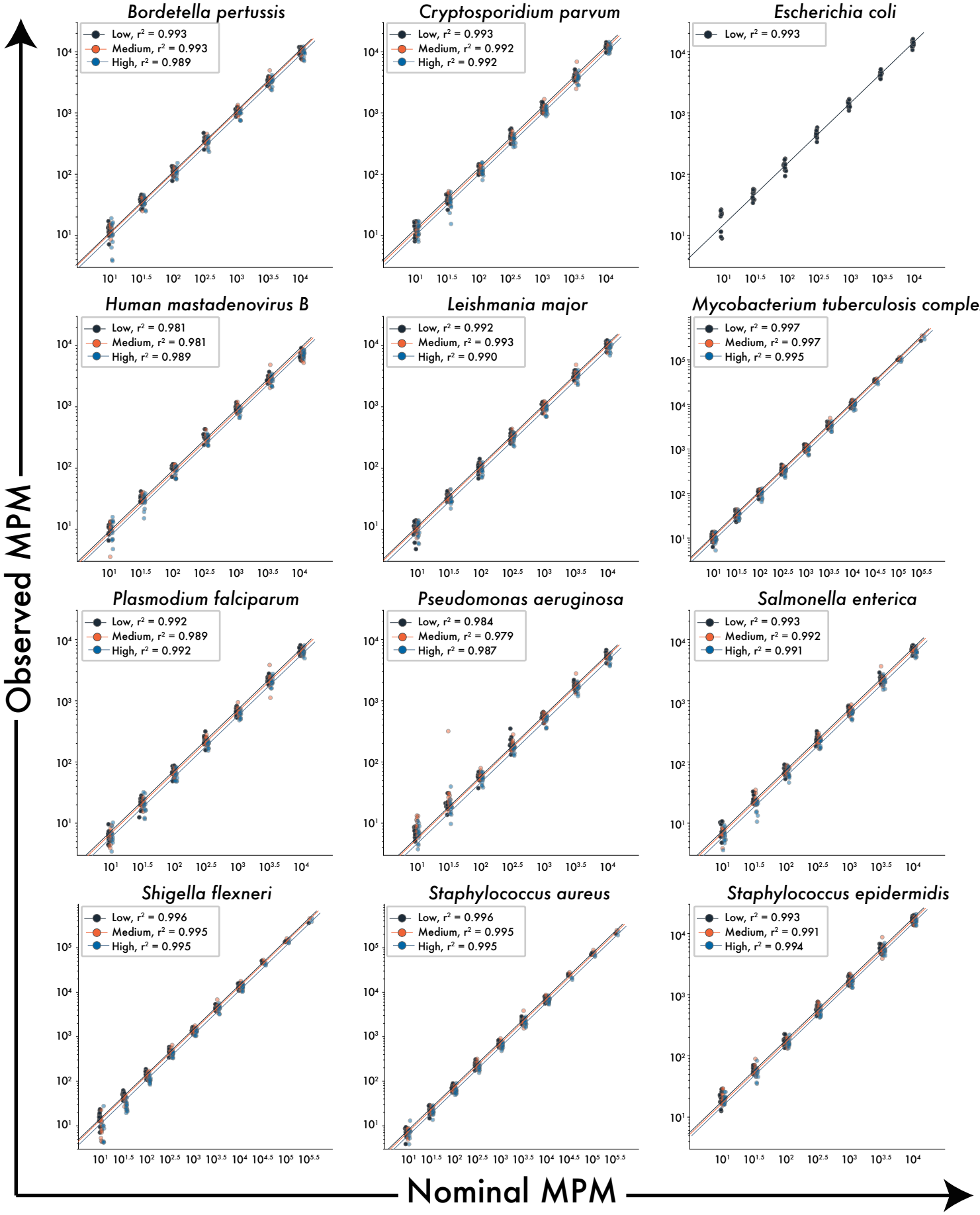
https://doi.org/10.1038/s41564-018-0349-6

In the format provided by the authors and unedited.

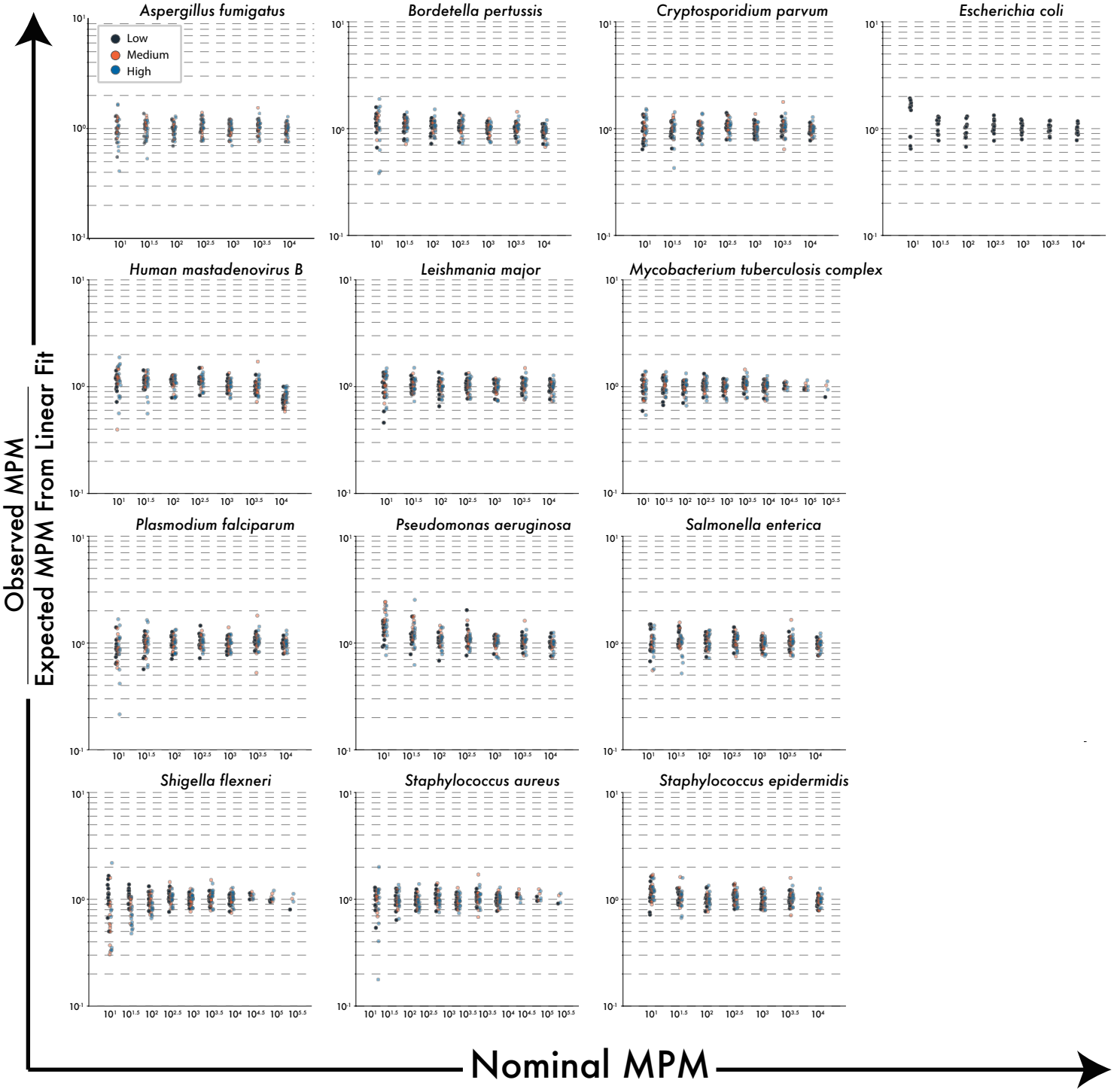
Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease

Timothy A. Blauwkamp[1,3]*, Simone Thair[2,3], Michael J. Rosen[1], Lily Blair[1], Martin S. Lindner[1], Igor D. Vilfan[1], Trupti Kawli[1], Fred C. Christians[1], Shivkumar Venkatasubrahmanyam[1], Gregory D. Wall[1], Anita Cheung[1], Zoë N. Rogers[1], Galit Meshulam-Simon[1], Liza Huijse[1], Sanjeev Balakrishnan[1], James V. Quinn[2], Desiree Hollemon[1], David K. Hong[1], Marla Lay Vaughn[1], Mickey Kertesz[1], Sivan Bercovici[1], Judith C. Wilber[1,3] and Samuel Yang[2,3]

[1]Karius, Inc., Redwood City, CA, USA. [2]Department of Emergency Medicine, Stanford University, Palo Alto, CA, USA. [3]These authors contributed equally: Timothy A. Blauwkamp, Simone Thair, Judith C. Wilber, Samuel Yang. *e-mail: tim.blauwkamp@kariusdx.com

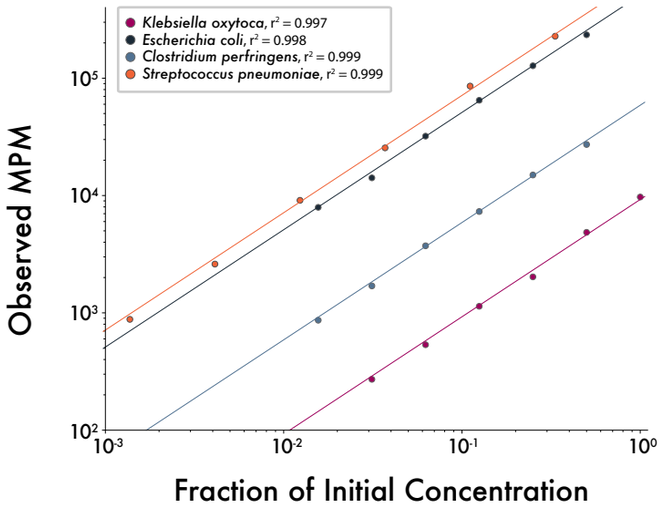


Supplementary Figure 1. Linearity. The correlation between nominal microbial cfDNA concentration and observed concentration is shown for each of the representative microbes not shown in Figure 3d. Each replicate (N=12) in each of the three backgrounds is represented by a dot (N=36 data points at each nominal concentration). Linear best fit regression of log-transformed MPM values showed all R^2 values were greater than 0.979, with an average R^2 of 0.991.

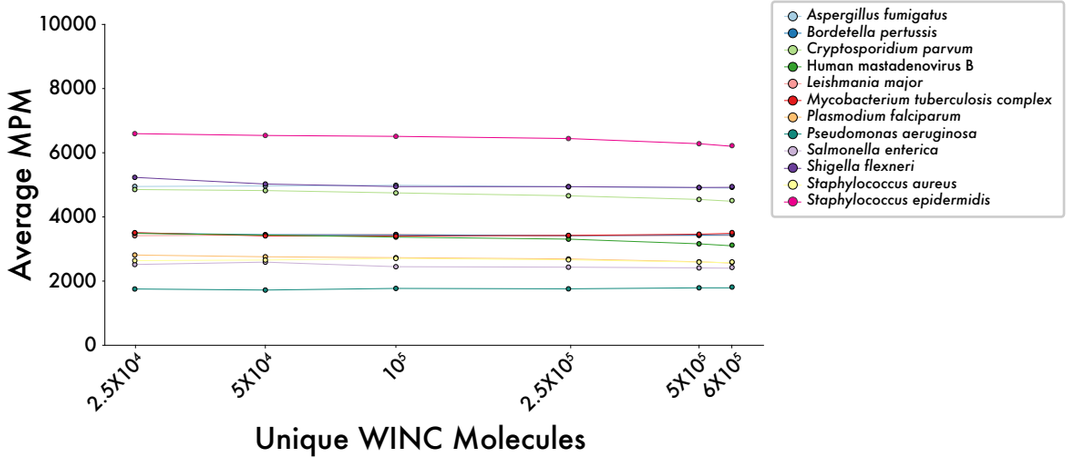


Supplementary Figure 2. Deviation from Linearity. The deviation from linearity as a function of nominal input concentration was assessed by plotting the ratio between observed MPM and that expected from the linear best fit regression calculated in Supplementary Figure 1 and Figure 3d. Data is shown for each of the representative microbes. Each replicate (N=12) in each of the three backgrounds is represented by a dot (N=36 data points at each nominal concentration in each panel). All microbes in all human matrices showed no systematic deviation from linearity at any concentration tested.

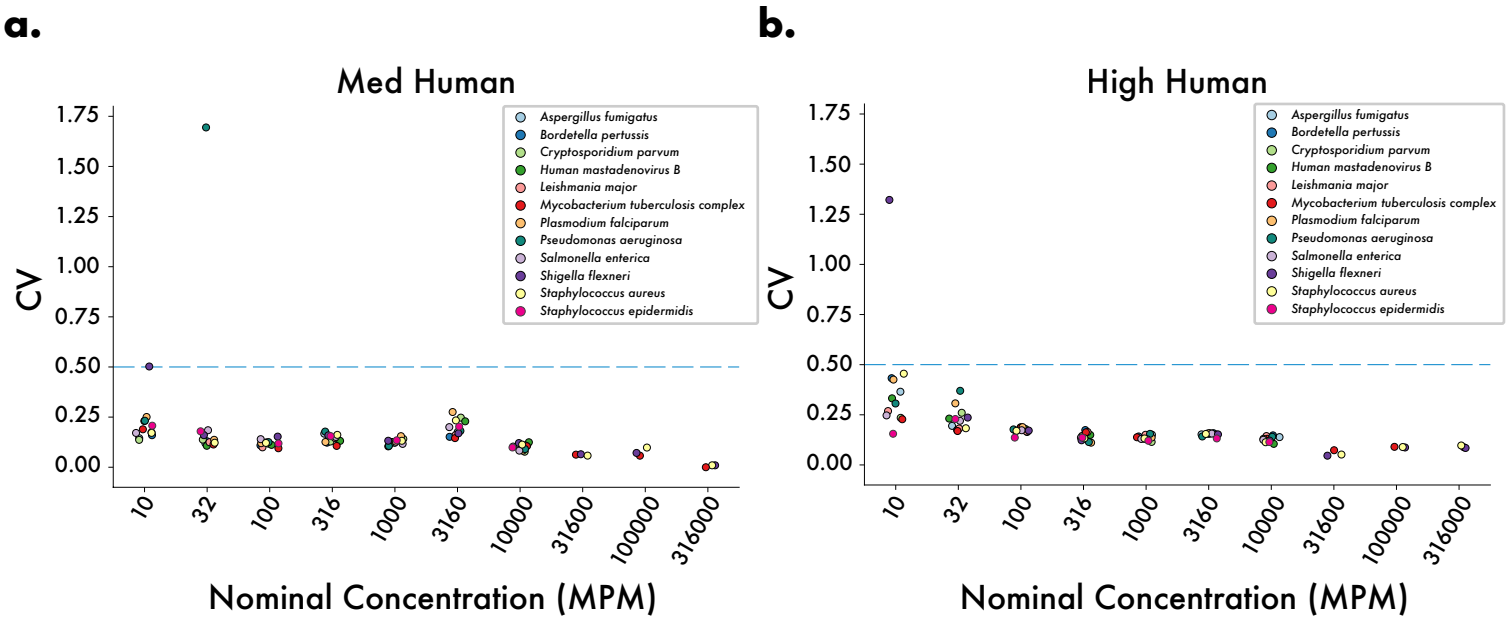
a.



b.



Supplementary Figure 3. Linearity Verification. (a) Linearity of the assay was confirmed on clinical samples using serial 2-fold or 3-fold dilutions of clinical samples with blood culture confirmed infections in healthy plasma (N=6 per sample). r^2 was calculated as described in figure 3d. (b) The microbial cfDNA concentration reported by the assay at different levels of sequencing coverage, generated by subsampling, is shown for each of 12 microbes spiked into medium human background plasma matrix. Sequencing depth does not significantly affect quantification accuracy.

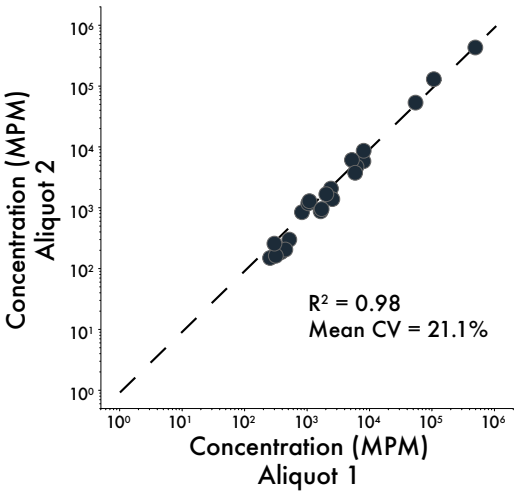


Supplementary Figure 4. Limit of quantitation precision. (a) Precision as a function of sequencing depth for each reference microbe is shown in medium (a) and high (b) human plasma matrices. Precision within the twelve replicate measurements is shown in a different color for each representative microbe at each nominal concentration (12 dots at each nominal concentration instead of 13 because *Escherichia coli* was excluded from these analyses due to significant *E. coli* DNA contamination of the purified mono nucleosomal DNA used to generate these human plasma matrices).

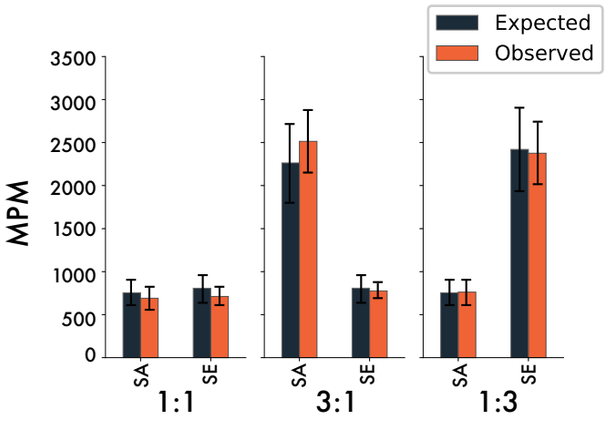
a.

	Qualitative Reproducibility	Quantitative Precison	
		Within Run (CV)	Within Lab (CV)
Plasmodium falciparum	100%	17.5%	18.9%
Cryptosporidium parvum	100%	16.7%	18.1%
Staphylococcus aureus	100%	17.4%	17.9%
Staphylococcus epidermidis	100%	18.5%	19.2%
Aspergillus fumigatus	100%	17.6%	20.3%
Escherichia coli	100%	18.7%	20.9%
Shigella flexneri	100%	18.7%	19.7%
Bordetella pertussis	100%	18.9%	22.2%
Salmonella enterica	100%	18.9%	18.9%
Leishmania major	100%	18.4%	21.3%
Human mastadenovirus B	100%	18.1%	18.9%
Mycobacterium tuberculosis	100%	18.3%	21.5%
Pseudomonas aeruginosa	100%	18.4%	21.2%
Mean	**100%**	**18.2%**	**19.9%**

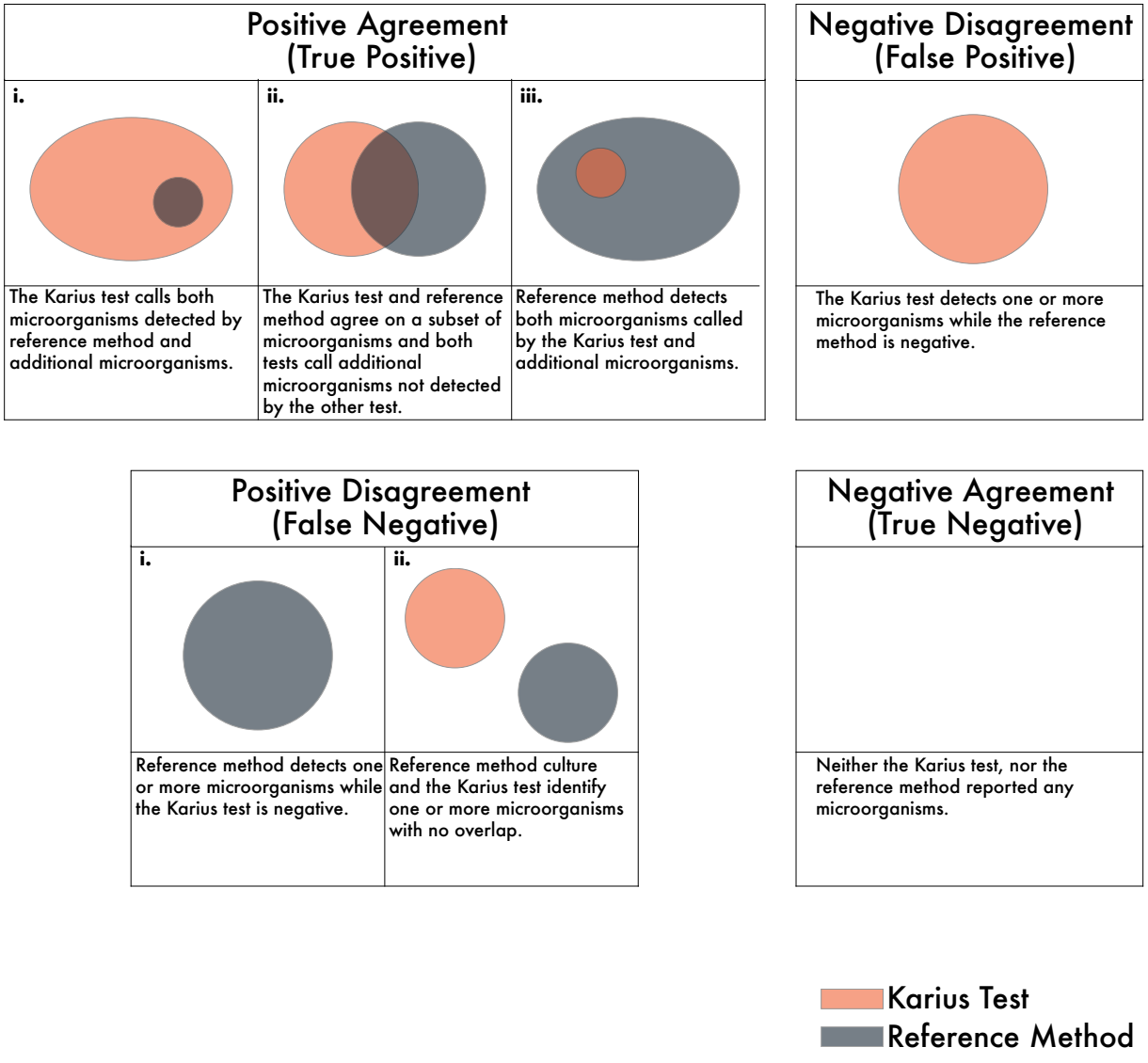
b.



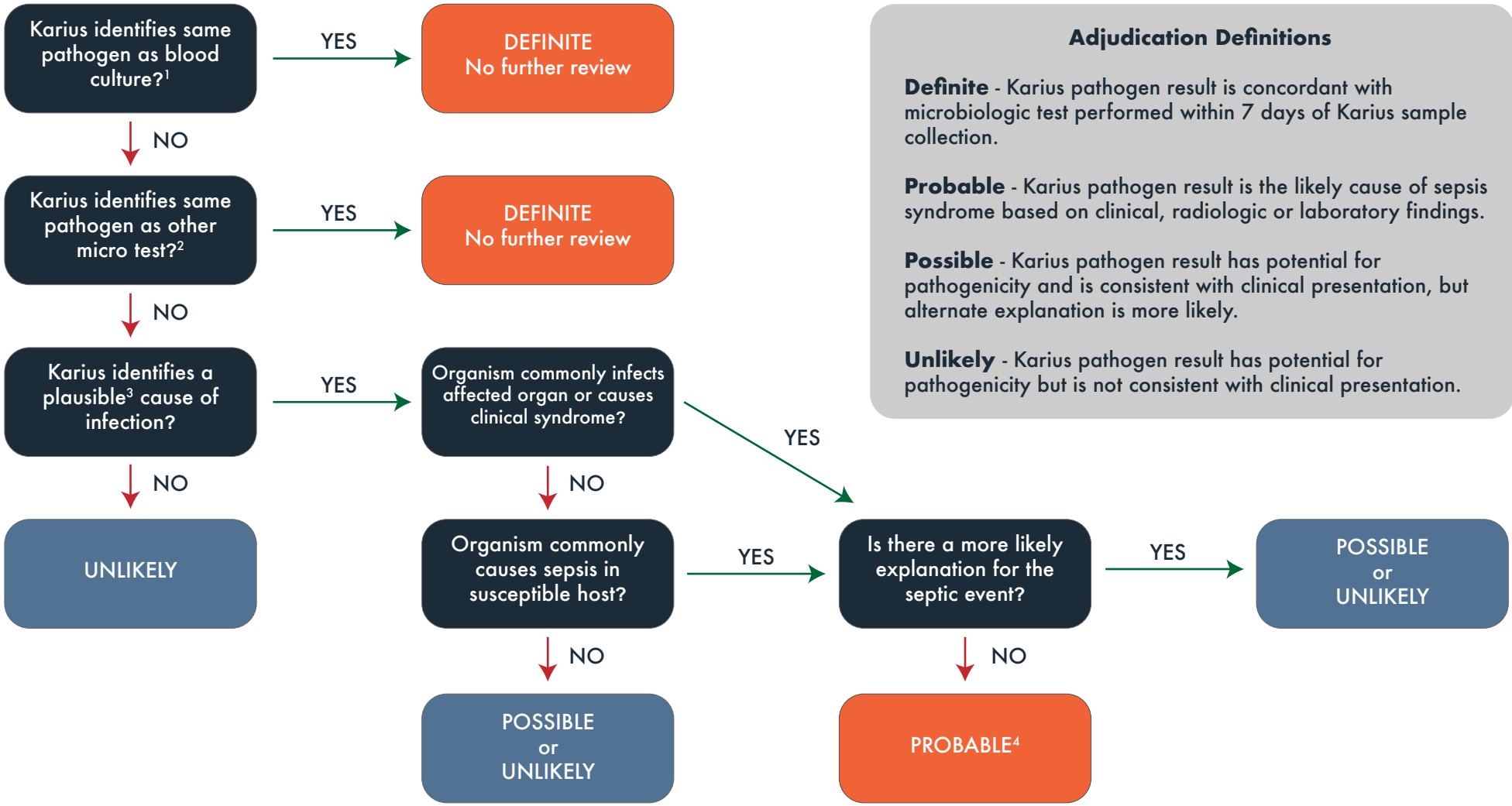
Supplementary Figure 5. Within run and within laboratory precision. (a) Within-run and within-laboratory precision were assessed by measuring microbial cfDNA abundance in a low human plasma matrix spiked with the representative microbial cfDNA reference materials at a nominal concentration of 1,000 MPM. Samples were processed in duplicate on each of 20 runs performed on 20 separate days by different operators using different instruments. Qualitative reproducibility indicates the fraction of tested samples with a positive result (N=40 for each microbe). Within-run repeatability and within-lab reproducibility indicate precision of the cfDNA concentration and were calculated according to CLSI EP5-A. (b) Reproducibility of the cfDNA concentration reported by the assay in clinical samples was assessed by testing 2 different aliquots of 20 different samples with known microbial infections on separate days (N=22). r^2 was calculated from linear regression.



Supplementary Figure 6. Interference. Potential interference during co-infection was assessed using healthy human plasma samples spiked with various ratios of genetically similar microbial cfDNA reference materials. Reported plasma cfDNA concentrations are shown for *Staphylococcus epidermidis* and *Staphylococcus aureus* when co-infection is contrived at ratios of 1:1, 3:1, and 1:3, respectively. Expected plasma cfDNA concentration is calculated from samples where each organism is spiked into a sample alone. Mean is shown and error bars show the sample standard deviation of the three replicates. All P-values >0.05 in two-tailed T-test. P-values for *S. epidermidis* were 0.28, 0.87, 0.67 respectively and *S. aureus* were 0.54, 0.42, 0.95 respectively.



Supplementary Figure 7. Scoring structure for microbial cfDNA sequencing comparisons to blood culture and all microbiological testing.



Supplementary Figure 8. SEP-SEQ Adjudication algorithm.

[1] Blood cultures with skin contaminant pathogens and no clinical evidence of infection will be considered negative.

[2] Microbiologic tests include blood culture, antibody, antigen, and nucleic acid-based testing, positive tests concordant with Karius will not be adjudicated.

[3] "Plausible" is defined as consistent with clinical presentation and with similar syndromes reported in the literature, all plausible cases will be adjudicated.

[4] Patients with neutropenic fever where Karius identified a plausible pathogen, and with no documented local source of infection, may be included in Probable category.

Category	Phyla	Families	Genera	CRR taxa
Bacteria	8	81	186	757
Fungi	4	67	119	332
Viruses	N/A	8	22	102
Nematodes	1	10	14	17
Platyhelminthes	1	6	8	11
Protozoans	1	10	14	30
Archaea	1	1	1	1
Total	16	183	364	1250

Supplementary Table 1. The clinically reportable range for this assay spans 1,250 taxa including culturable bacteria, additional fastidious and unculturable bacteria, fungi, DNA viruses and eukaryotic pathogens.

Representative Microbial DNA	Pathogen Class	% GC	Genome Size	Common Commensal	Common EC	Source
Aspergillus fumigatus	Eukaryotic Mold	50	29 Mb	No	No	ATCC
Bordetella pertussis	Gram Neg Bacteria	68	4.10 Mb	No	No	ATCC
Cryptosporidium parvum	Eukaryotic Parasite	30	9.10 Mb	No	No	ATCC
Escherichia coli	Gram Neg Bacteria	51	5.62 Mb	Yes, Gut	Yes	ATCC
Human mastadenovirus B	dsDNA Virus	51	35 kb	No	No	ATCC
Leishmania major	Eukaryotic Parasite	63	33 Mb	No	No	ATCC
Mycobacterium tuberculosis	Acid Fast Bacteria	66	4.41 Mb	No	No	ATCC
Plasmodium falciparum	Eukaryotic Parasite	19	22.90 Mb	No	No	ATCC
Pseudomonas aeruginosa	Gram Neg Bacteria	66	6.30 Mb	Yes, Skin	Yes	NIST
Salmonella enterica	Gram Neg Bacteria	52	4.77 Mb	No	No	NIST
Shigella flexneri	Gram Neg Bacteria	51	4.60 Mb	No	No	ATCC
Staphylococcus aureus	Gram Pos Bacteria	32	2.80 Mb	Yes, Nose	No	NIST
Staphylococcus epidermidis	Gram Pos Bacteria	32	2.56 Mb	Yes, Skin	Yes	ATCC

Supplementary Table 2. A panel of thirteen microbes was used to model the effects of several key determinants of analytical performance, including GC-content, genome size, super-kingdom, background signal from commensals, and background signal from environmental contamination. Together, this panel spans the breadth of key performance determinants found among the 1,250 microbes probed by the assay.

	LoD at 300,000 WINCs			LoD at 25,000 WINCs		
	Low Human	**Medium Human**	**High Human**	**Low Human**	**Medium Human**	**High Human**
Aspergillus fumigatus	32.99	32.99	38.62	325.6	342.7	357.2
Bordetella pertussis	37.39	36.18	38.62	357.2	368.7	414.3
Cryptosporidium parvum	34.87	32.99	40.02	325.6	368.7	368.6
Escherichia coli	73.54	N/A	N/A	595.7	N/A	N/A
Human mastadenovirus B	37.39	40.01	44.01	357.2	414.3	434.0
Leishmania major	37.40	40.88	41.29	382.8	394.0	401.7
Mycobacterium tuberculosis	38.62	37.39	40.02	357.2	394.0	414.3
Plasmodium falciparum	73.54	73.54	73.54	595.7	595.7	595.7
Pseudomonas aeruginosa	415.41	373.42	131.79	4159.1	5789.0	1340.9
Salmonella enterica	44.01	45.19	73.54	595.7	595.7	595.7
Shigella flexneri	34.87	37.39	73.54	342.7	357.2	368.6
Staphylococcus aureus	73.54	73.54	103.04	595.7	595.7	1031.0
Staphylococcus epidermidis	36.18	41.29	44.01	342.7	394.0	401.7
Median	**37.39**	**40.45**	**44.01**	**357.2**	**394.0**	**414.3**

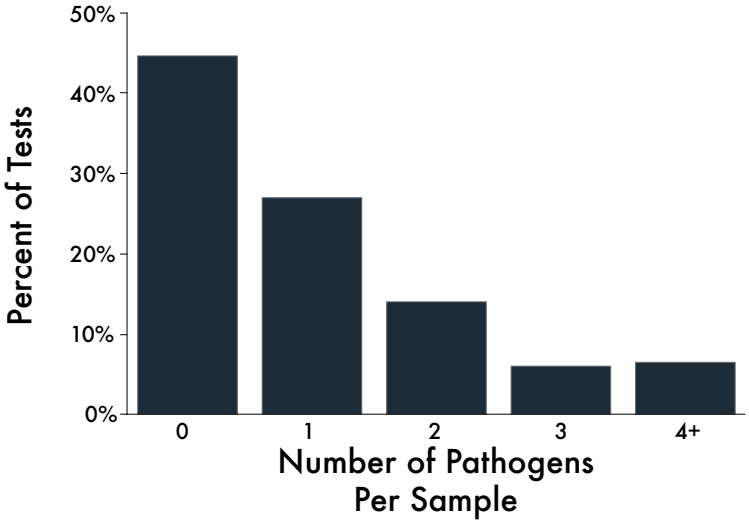
Supplementary Table 3. Limit of detection per microbe. The limit of detection (LoD) was determined using sheared microbial cfDNA reference materials for each of the 13 representative pathogens mixed together at nominally identical concentrations and spiked into low, medium, and high human matrices over seven 0.5-log serial dilutions ranging from 10,000 molecules per microliter plasma (MPM) to 10 MPM. Twelve replicates of each concentration in each healthy plasma matrix were measured over twelve different days, by different operators, using different instruments, to determine the fraction of expected microbes that were detected. The LoD determined for each representative microbe, in each plasma matrix, after down-sampling sequencing depth to levels typically found in clinical batches (300,000 unique WINC molecules) or minimum acceptable sequencing depth (25,000 unique WINC molecules) are listed.

Characteristic	Data (N=350)
Age, median (range), years	54 (18-97)
Sex, n (%)	
Male	179 (51.1)
Female	171 (48.9)
Race, n (%)	
White	197 (56.3)
Asian	74 (21.1)
Black or African American	15 (4.3)
Native Hawaiian or other Pacific Islander	7 (2)
American Indian or Alaskan Native	1 (0.3)
Not reported	55 (15.7)
Medical Comorbidities, n (%)	
≥ 1 concurrent chronic medical condition	227 (64.9)
Hypertension	97 (27.7)
Diabetes mellitus	61 (17.4)
Chronic heart disease	54 (15.4)
Hyperlipidemia	53 (15.1)
Lenght of Hospital Stay	
Mean length of stay in days, n (range)	4.7 (1-117)
Median length of stay in days, n (IQR)	3 (1-5)
Hospitalization Survival Status, n (%)	
Discharged	346 (98.9)
Died	4 (1.1)
Antimicrobial treatment[1] within 2 weeks of sepsis alert	97 (27.7)

Supplementary Table 4. Summary of patient demographics and clinical characteristics.
[1]Antimicrobial treatment count includes patients with antibiotic, antiviral, and antifungal therapy and one patient not included in the primary analysis.

CRR Taxon Name	Mean MPM	Median MPM	97.5th Percentile of MPM	Fraction of Samples With a Positive Call
Helicobacter pylori	3.95	0.00	47.43	0.07
Klebsiella pneumoniae	3.99	0.00	27.63	0.02
Haemophilus influenzae	2.64	0.00	23.53	0.02
Enterobacter cloacae complex	4.05	1.05	27.78	0.02
Aureobasidium pullulans	1.44	0.00	2.49	0.01
Bacteroides ovatus	1.09	0.00	2.30	0.01
Micrococcus lylae	0.90	0.00	1.34	0.01
Fusobacterium necrophorum	1.94	0.00	0.95	0.01
Agrobacterium tumefaciens	22.73	2.99	91.62	0.01
Pseudomonas putida	31.75	26.23	65.14	0.01
Escherichia coli	6.94	4.22	15.15	0.01
Streptococcus mitis	3.47	1.70	11.43	0.01
Rothia mucilaginosa	1.38	0.73	6.75	0.01
Pseudomonas fluorescens	0.99	0.00	4.25	0.01
Gemella haemolysans	0.70	0.00	3.97	0.01
Staphylococcus aureus	0.35	0.00	2.22	0.01
Corynebacterium kroppenstedtii	0.39	0.00	1.91	0.01
Aeromonas caviae	0.41	0.00	1.68	0.01
Pseudomonas oryzihabitans	0.27	0.00	1.27	0.01
Human herpesvirus 4	0.27	0.00	1.18	0.01
Acinetobacter radioresistens	0.23	0.00	1.13	0.01
Debaryomyces hansenii	0.29	0.00	0.99	0.01
Lactobacillus plantarum	0.27	0.00	0.90	0.01
Neisseria gonorrhoeae	0.52	0.00	0.43	0.01
[Clostridium] clostridioforme	0.25	0.00	0.00	0.01
Pichia kudriavzevii	0.08	0.00	0.00	0.01

Supplementary Table 8. Reference Interval Microbes Detection Frequency. Reference interval data from 167 asymptomatic individuals were characterized, including mean MPM, median MPM, 97.5th percentile MPM, and fraction of samples with a positive call for each analyte. All analytes that were called in any asymptomatic sample are included in this table.



Supplementary Figure 9. Analysis of the first 2,000 samples in clinical practice. Frequency of polymicrobial results in the first 2,000 samples.