

Supplemental Table S1. The read counts of the 27 stool samples subjected to metagenomics sequencing. MAL = Mali, MOZ = Mozambique, IND = India.

Sample	Sample ID	Site	Culture	qPCR	Total high quality reads
1	202260	MAL	+	+	255853444
2	203032	MAL	+	+	370748896
3	203745	MAL	+	+	336959130
4	300244	MOZ	+	+	267173944
5	302912	MOZ	+	+	318564818
6	303061	MOZ	+	+	275670712
7	500242	IND	+	+	251507850
8	504039	IND	+	+	322336854
9	503617	IND	+	+	259519220
10	202455	MAL	-	+	273141504
11	202742	MAL	-	+	304390100
12	203725	MAL	-	+	389080142
13	300055	MOZ	-	+	293291234
14	300716	MOZ	-	+	254650088
15	302016	MOZ	-	+	347812514
16	500620	IND	-	+	282549862
17	503417	IND	-	+	231606090
18	504637	IND	-	+	282331220
19	200714	MAL	-	-	373848930
20	201361	MAL	-	-	357411820
21	202587	MAL	-	-	271984392
22	300056	MOZ	-	-	387004856
23	300763	MOZ	-	-	244645304
24	302001	MOZ	-	-	409672600
25	503034	IND	-	-	327307332
26	503626	IND	-	-	258594624
27	504241	IND	-	-	335360728

Supplemental Table S2. The interrogated region of 6 *Shigella* virulence genes.

<b>Gene</b>	<b>Function</b>	<b>Accession number</b>	<b>Location</b>
<i>ipaH</i>	invasion plasmid antigen H	CP000039	186587 - 185588
<i>ipaH3</i>	invasion plasmid antigen H3	CP000038	1391081 - 1391611
<i>ial</i>	invasion associated protein	D50601	17450 - 18449
<i>ShET2</i>	enterotoxin Set2	Z54211	491 - 1490
<i>virA</i>	microtubule disruption	D26468	1101 - 2100
<i>virG</i>	encode icsA involved in actin based mobility and intercellular spread	M22802	2060 - 3059

Supplemental Table S3. Pathogen detection in 27 selected diarrheal samples, shown as quantification cycle (Cq). Red represents highly diarrhea-associated quantities, orange diarrhea-associated quantities, yellow represents detection with no association or for pathogens where no Cq cutoff exists (Liu, 2016), and grey denotes no detection.

Sample	Diarrhea associated pathogen with a Cq cutoff													Pathogen without a Cq cutoff or diarrhea association									
	Shigella/EIEC	adenovirus 40/41	C. jejuni/coli	H. pylori	Salmonella	V. cholerae	HEPEC	ST-EPEC	astrovirus	norovirus GII	rotavirus	sapovirus	Cryptosporidium	norovirus GI	Aeromonas	B. fragilis	C. difficile	EAEC	aEPEC	LT-EPEC	Ascaris	E. bieneusi	Giardia
Shigella culture + / qPCR +	19	-	23	-	-	-	21	-	28	-	-	-	-	-	-	-	-	18	-	-	-	30	20
	19	-	34	-	-	-	34	-	-	-	28	-	-	-	-	-	-	-	-	34	-	-	-
	20	-	-	-	-	-	-	25	-	29	-	-	29	-	-	-	-	21	-	-	-	-	31
	18	34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33	-	-	-	34
	17	15	-	35	-	-	27	24	-	-	-	33	32	-	-	27	-	29	-	-	-	23	18
	20	26	-	-	-	-	-	31	-	-	-	-	-	-	32	-	-	22	31	-	-	-	32
	19	26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32	-	-	-	-	30
	17	35	-	-	-	-	-	-	19	-	35	-	-	-	21	-	-	-	-	-	-	-	33
	19	28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	31
	Shigella culture - / qPCR +	18	-	-	-	-	-	18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19		-	-	-	-	-	13	35	-	31	-	-	-	-	-	33	-	28	-	-	-	-	35
20		-	-	-	-	-	29	-	-	-	-	-	-	-	-	-	-	23	-	-	-	-	31
19		-	-	-	-	-	34	-	-	-	-	-	-	-	-	-	-	16	-	34	-	-	32
17		32	-	-	-	35	34	-	-	-	-	-	-	-	-	-	-	29	-	-	-	-	15
20		-	-	-	-	-	34	-	-	-	-	-	-	-	-	-	-	18	-	-	-	-	-
18		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	31	-	-	21
19		-	-	34	-	-	-	-	-	-	-	-	-	-	-	21	-	18	-	-	-	-	-
19		-	25	-	-	-	-	-	-	-	-	-	32	-	-	27	-	-	26	27	29	-	19
Shigella culture - / qPCR -		-	-	-	-	-	-	-	-	-	-	-	29	30	-	-	34	-	-	-	-	-	-
	-	29	26	-	-	-	32	-	-	33	29	-	-	-	-	-	-	-	-	28	-	33	15
	-	-	-	-	34	-	-	-	-	-	24	-	17	35	-	-	-	35	-	-	-	-	20
	-	33	-	-	-	-	32	-	-	-	-	29	-	-	-	-	-	19	-	-	-	35	32
	-	-	-	-	-	-	-	-	-	-	24	31	-	-	-	-	-	34	30	-	-	35	24
	-	31	-	-	-	-	35	32	-	-	-	-	24	-	-	-	-	16	-	-	-	-	24
	-	-	-	-	-	-	-	-	21	-	-	-	-	-	-	-	-	22	23	-	-	-	34
	-	-	19	28	-	-	-	17	-	26	-	-	-	-	32	-	-	-	22	-	-	-	32
	-	17	28	-	-	-	19	-	-	-	-	31	-	-	-	-	-	19	-	-	-	-	25



Supplemental Table S5. The raw read counts of *Shigella* species in culture+/qPCR+ samples. Highest read counts are bolded if correctly classified the stool culture result and underlined if misclassified the stool culture result.

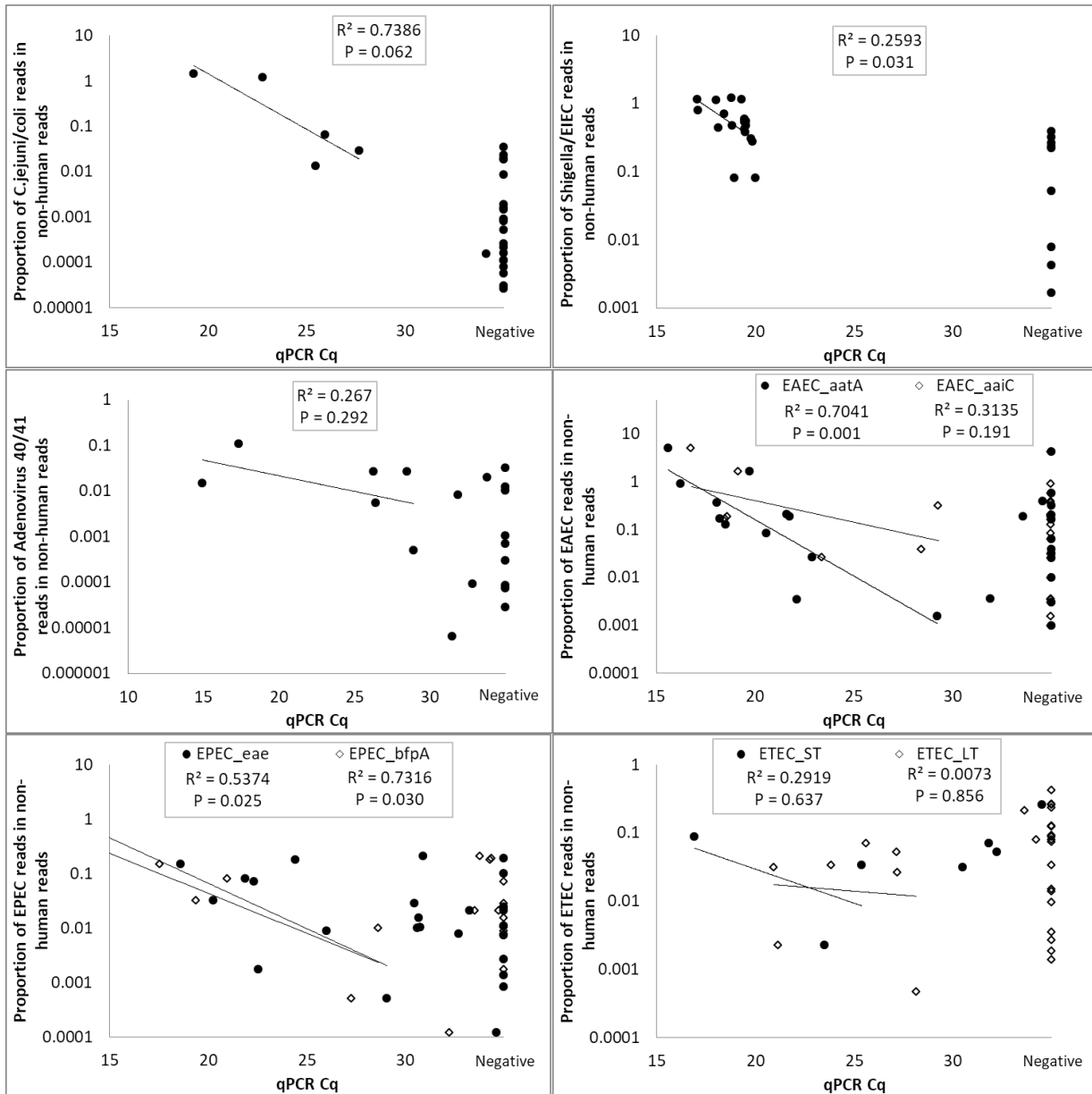
Culture identification	<i>S. flexneri</i>	<i>S. boydii</i>	<i>S. sonnei</i>	<i>S. dysenteriae</i>
<i>S. flexneri</i> serotype 6	8606	<u>41114</u>	10176	5659
<i>S. flexneri</i> serotype 3a	1083131	455181	739325	<u>1124807</u>
<i>S. sonnei</i>	32255	58086	<b>929797</b>	25220
<i>S. flexneri</i> serotype 2a	<b>32371</b>	668	2009	1528
<i>S. sonnei</i>	6125	6917	<b>164642</b>	2292
<i>S. flexneri</i> serotype 2a	<b>51050</b>	10431	14378	23872
<i>S. flexneri</i> serotype 3a	<b>21916</b>	661	2422	400
<i>S. sonnei</i>	8523	30985	<b>706606</b>	7810
<i>S. sonnei</i>	2738	3079	<b>62117</b>	591

Supplemental Table S6. Comparison of precision and recall on the genus level between Kraken and Clark MetaPhlan2, Kaiju on the *Shigella* isolates. The highest precision or sensitivity for each isolate was bolded.

Isolate ID	Monophyletic lineage	Genus precision				Genus sensitivity			
		Kraken	Clark	MetaPhlan2	Kaiju	Kraken	Clark	MetaPhlan2*	Kaiju
PK1010072_SF	S1	<b>86.8</b>	84.2	5.9	54.9	<b>38.4</b>	23.5	1.7	8.6
PK1010438_SF	S1	<b>86.7</b>	83.8	35.8	55.6	<b>37.4</b>	22.7	2.3	8.4
PK1010893_SF	S1	<b>87.3</b>	85.2	29.8	56.6	<b>39.3</b>	24.6	2.1	8.9
PK1011266_SB	S1	<b>88.7</b>	87.4	29.2	60.1	<b>40.0</b>	24.8	2.2	9.2
PK1010319_SS	S2	96.9	<b>97.1</b>	55.7	63.5	36.0	21.9	<b>47.6</b>	7.4
PK1010319_SF	S5	<b>96.2</b>	94.5	94.4	75.4	50.6	33.0	<b>90.6</b>	12.7
PK1010339_SF	S5	99.2	<b>99.4</b>	93.5	81.9	54.1	35.9	<b>81.8</b>	14.2
PK1010355_SF	S5	<b>94.9</b>	94.0	89.1	73.3	46.7	29.2	<b>68.6</b>	11.2
PK1010912_SF	S5	<b>95.1</b>	92.7	90.6	70.5	47.6	29.8	<b>81.6</b>	11.4
PK1010943_SF	S5	<b>91.5</b>	89.4	92.5	66.0	45.9	28.6	<b>65.0</b>	10.9
PK1011037_SF	S5	<b>95.8</b>	93.6	90.7	71.4	47.9	30.3	<b>80.6</b>	11.6
PK1011283_SF	S5	99.1	<b>99.4</b>	93.6	81.1	53.5	35.6	<b>77.6</b>	14.2

\*metaphlan2 gives an estimation of the number of reads that should match that taxa based on the marker gene detection

Supplemental Figure S1. Comparison of metagenomics sequencing reads versus qPCR Cq values. Only DNA pathogens with more than 5 positives were plotted, i.e. qPCR positive with Cq < 35. Regression was fit for Ct < 30.



## Reference

**Hazen TH, Leonard SR, Lampel KA, Lacher DW, Maurelli AT, Rasko DA.** 2016. Investigating the relatedness of Enteroinvasive Escherichia coli to other E. coli and Shigella Isolates by using comparative genomics. *Infect Immun* **84**:2362-2371.

**Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, Operario DJ, Uddin J, Ahmed S, Alonso PL, Antonio M, Becker SM, Blackwelder WC, Breiman RF, Faruque AS, Fields B, Gratz J, Haque R, Hossain A, Hossain MJ, Jarju S, Qamar F, Iqbal NT, Kwambana B, Mandomando I, McMurry TL, Ochieng C, Ochieng JB, Ochieng M, Onyango C, Panchalingam S, Kalam A, Aziz F, Qureshi S, Ramamurthy T, Roberts JH, Saha D, Sow SO, Stroup SE, Sur D, Tamboura B, Taniuchi M, Tennant SM, Toema D, Wu Y, Zaidi A, Nataro JP, Kotloff KL, Levine MM, Hupt ER.** 2016. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet* **388**:1291-1301.